Redbooks

ibm.com/redbooks

IBM Power E1180

Introduction and Overview

Tim Simon

Henry Vo

Jordan Antonov

Ivaylo Bozhinov

Rohit Chauhan

Nicolae Chirea

Turgut Genc

Bartlomiej Grabowski

Andrey Klyachkin

Tsvetomir Spasov

IBM Power

Hybrid Cloud

IBM

Redbooks

IBM Redbooks

IBM Power E1180: Introduction and Overview

September 2025

SG24-8587-00

Note: Before using this information and the product it supports, read the information in “Notices” on page vii.

First Edition (September 2025)

This edition applies to IBM Power E1180 (9080-HEU).

© Copyright International Business Machines Corporation 2025. All rights reserved.
Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

© Copyright IBM Corp. 2025.

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

© Copyright IBM Corp. 2025.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at https://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®
Concert®
DS8000®
FlashCopy®
HashiCorp®
IBM®
IBM API Connect®
IBM Cloud®
IBM Cloud Pak®
IBM Cloud Satellite®
IBM Concert®
IBM FlashSystem®
IBM Garage®
IBM Instana®
IBM Research®
IBM Security®
IBM Services®
IBM Spyre™
IBM Turbonomic®
IBM Watson®
IBM Z®
Instana®
Integrated Language Environment®
Language Environment®
Micro-Partitioning®
Passport Advantage®
Power Architecture®
Power8®
Power9®
PowerHA®
PowerVM®
Rational®
Redbooks®
Redbooks (logo) ®
Satellite®
Spyre™
SystemMirror®
Terraform®
Tivoli®
Turbonomic®
watsonx®
WebSphere®

The following terms are trademarks of other companies:

Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

LTO, Ultrium, the LTO Logo and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Red Hat, Ansible, Fedora, OpenShift, are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

VMware, VMware vSphere, and the VMware logo are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and/or other jurisdictions.

Other company, product, or service names may be trademarks or service marks of others.

Preface

This IBM Redbooks® publication provides a comprehensive overview of the IBM Power E1180, the flagship system in the IBM Power11 family. It describes the implementation of the Power11 processor, which is an evolution of the Power10 chip that is used in the IBM Power E1080, and examines the system architecture and performance capabilities in detail.

Engineered for the artificial intelligence (AI) era, the IBM Power11 platform is built on three foundational pillars: Trusted, Autonomous, and Modern.

- Trusted: Provides exceptional business continuity with 99.9999% availability, zero planned downtime (ZPD) for maintenance, quantum-safe cryptography, and rapid system recovery.
- Autonomous: Improves operational efficiency through intelligent automation, energy-optimized performance modes, and minimal manual intervention.
- Modern: Supports AI application development and deployment across hybrid cloud environments (on-premises and in the cloud) by using off-chip AI acceleration through the IBM Spyre Accelerator™ and a robust partner ecosystem.

The Power11 architecture introduces several key hardware innovations, including 7 nm processor technology, increased core density, and expanded memory bandwidth with DDR5. It also offers advanced reliability, availability, and serviceability (RAS), flexible consumption models, and broad compatibility with modern software platforms such as Red Hat OpenShift and IBM watsonx®.

Overall, the Power E1180 is a high-performance, enterprise-grade platform that can support AI workloads, mission-critical applications, and digital transformation at scale.

Authors

This book was produced by a team of specialists from around the world working remotely.

Tim Simon is an IBM Redbooks Project Leader who is based in Tulsa, Oklahoma, US. He has more than 40 years of experience at IBM®, primarily in technical sales, and has built a distinguished career helping clients design and implement IBM solutions that align with their business goals. He holds a Bachelor of Science degree in Mathematics from Towson University in Maryland. Tim's expertise spans a broad range of IBM technologies, with deep experience in developing customer solutions that use IBM Power, IBM Storage, and IBM Z® platforms.

Henry Vo is an IBM Redbooks Project Leader with 11 years of experience at IBM. He brings strong technical expertise in business problem-solving, risk and root-cause analysis, and the development of technical strategies that align with business objectives. Throughout his career at IBM, Henry has held various roles, including Project Manager, ST/FT/ETE Tester, Back-End Developer, and Department of Labor (DOL) Agent for New York. He is certified in both IBM Z and IBM Power technologies. Henry earned his Master's degree in Management Information Systems from the University of Texas at Dallas in 2012 and continues to apply his diverse skill set to drive innovation and deliver value across IBM projects.

© Copyright IBM Corp. 2025.

Jordan Antonov is based in Bulgaria and has been with IBM for more than 10 years, primarily contributing to the IBM Power support team. He is part of the Product Engineering team, where he applies his deep technical knowledge and problem-solving skills. He holds a Bachelor's degree in Machine Engineering from the Technical University of Sofia. Jordan's areas of expertise include IBM Power hardware, IBM PowerVM®, root-cause analysis, and hardware issue resolution. He has authored numerous technical documents and developed several service education offerings, demonstrating his commitment to technical excellence and knowledge sharing.

Ivaylo Bozhinov is an IBM Power subject matter expert (SME) who is based in IBM Bulgaria. His primary area of expertise is resolving complex hardware and software issues on IBM Power products, IBM AIX®, Virtual I/O Server (VIOS), Hardware Management Console (HMC), IBM i, IBM PowerVM, and Linux on IBM Power servers. He has been with IBM since 2015 and provides reactive break-fix, proactive, preventive, and cognitive support.

Rohit Chauhan is an IBM Champion and a Senior Technical Specialist with expertise in the IBM i platform and Power servers. He is also an IBM Redbooks author and works at Tietoevry Tech Services in Stavanger, Norway, an IBM Business Partner and one of the largest IT service providers in the Nordics. He has more than 12 years of experience working on the IBM Power platform, focusing on the design, planning, and implementation of IBM i infrastructure, including high availability and disaster recovery (HADR) solutions for many customers during this time. Before his current position, Rohit worked for clients in Singapore and the UAE in technical leadership and security roles for the IBM Power domain. He is a member of Common Europe Norway with a strong focus on the IBM i platform and security. As the leader of the IBM TechXchange IBM i Security and Innovation user group, he shares his expertise through technical articles on iBlog, featured by Common Norway. His areas of expertise include IBM i, IBM HMC, security enhancements, IBM PowerHA®, systems performance analysis and tuning, Backup, Recovery, and Media Services (BRMS), external storage, PowerVM, and IBM i platform solutions. Also, he is a co-author of several IBM Redbooks® publications.

Nicolae Chirea began his journey with AS/400 in 1999 as an RPG and CL programmer before moving to system administration. By 2007, he was managing more than 120 IBM iSeries systems and actively participating in infrastructure upgrades, logical partition (LPAR) implementations, XSM solutions, and independent auxiliary storage pool (IASP) migrations. Since 2007, he has worked at Seidor, an IBM Business Partner, where he is involved in the full lifecycle of Power servers, from initial setup to deployment. His expertise includes version upgrades, high availability (HA) solutions with IBM PowerHA, BRMS, Advanced Job Scheduler, and security and performance optimization for IBM i and Db2 for i. He is also a recognized expert in IBM storage technologies, including IBM FlashSystem®, IBM DS8000®, and storage area network (SAN) networking. In recent years, Nicolae expanded his focus to Linux on IBM Power, leading major SAP HANA implementations in Spain. He actively promotes Ansible within the IBM Power community and has deployed modern monitoring tools such as Prometheus and Grafana. As an IBM Champion since 2021, he frequently speaks at industry events, sharing his expertise. In September 2024, he joined the Common Europe Advisory Council (CEAC), further strengthening his role in the IBM ecosystem.

Turgut Genc is the IBM Power and Storage Solution Architect for IBM Technology Expert Labs in the UK. He holds a Master of Science degree in Computer Science Engineering from Yildiz Technical University in Istanbul. He has more than 20 years of experience with Power servers and has served as the EMEA team leader for IBM Power Virtual Server (PowerVS), Power Enterprise Pools (PEP), performance, migration workshops, and Technology Expert Labs tools.

Bartlomiej Grabowski is an IBM Champion and a Principal Systems Specialist at DHL IT Services, where he supports Power servers globally. With more than 20 years of experience in enterprise IT solutions, Bartlomiej brings deep technical expertise and a passion for innovation. He holds a Bachelor's degree in Computer Science from the Academy of Computer Science and Management in Bielsko-Biala, Poland. His areas of specialization include Power servers, IBM i, IBM PowerHA, IBM PowerVM, system performance optimization, and storage solutions. Bartlomiej is also a Platinum IBM Redbooks author.

Andrey Klyachkin is a Solution Architect at eNFence, an IBM Business Partner in Germany. He has more than 25 years of experience in UNIX systems, designing and supporting AIX and Linux systems for customers worldwide. He has co-authored many IBM AIX and IBM Power certifications. Andrey is an IBM Champion and IBM AIX Community Advocate. He is also a Red Hat Certified Engineer and Red Hat Certified Instructor. He regularly attends international and local events for IBM Power, such as IBM TechXchange, GSE, and the Common Europe Congress.

Tsvetomir Spasov is a IBM Power Servers Hardware Product Engineer who is based in Sofia, Bulgaria. He has 9 years of experience with IBM in RTS, SME, and PE roles. His main areas of expertise include HMC, Flexible Service Processor (FSP), eBMC, PowerLC, and GTMS. Over the years, he has authored numerous technical documents and developed several service education offerings, demonstrating his commitment to technical excellence and knowledge sharing. He holds a Master's degree in Electrical Engineering from the Technical University of Sofia.

Thanks to the following people for their contributions to this project:

Nicole Nett, IBM Chief Architect IBM Power Servers
IBM Austin

Lyudmila Simeonova, IBM AIX I/O Device Drivers and PowerVM VIOS
IBM Bulgaria

Milen Paiakov, IBM Power Courseware Developer
IBM Bulgaria

Ruby Zgabay, IBM Power Chief Engineer
IBM Austin

Kaveh Naderi, IBM Power Chief Architect
IBM Austin

Paul Bastide, IBM Senior Software Engineer
IBM Cambridge

Tanya Koleva, IBM Technology Lifecycle Services
IBM Bulgaria

Manjunath Kumatagi, IBM Senior Software Engineer, Red Hat OpenShift on IBM Power
IBM Bangalore

Manoj Kumar, IBM Chief Engineer, Red Hat OpenShift, Red Hat OpenShift AI, IBM Turbonomic®, IBM Instana® on IBM Power
IBM Austin

Pablo Pereira Ruiz, Senior UNIX IBM AIX Administrator/Support en Interamericana de Cómputos S.A.
Uruguay

George Gaylord, Program Director, Offering Management
IBM US

Linda Hirsch, IBM i Offering Program Manager, IBM i Licensing, and Terms and Conditions
IBM US

Dan Sundt, IBM i Power as-a-Service Product Manager
IBM Rochester

Yan Zhuo, Product Management Leader, IBM i
IBM Chicago

Pete Heyrman, IBM PowerVM Hypervisor (PHYP)
IBM Rochester

Dirk Michel, STSM, IBM Power Enterprise Performance Architect
IBM Austin

John Folkerts, IBM Power I/O Product Management
IBM Rochester

Bob Kovacs, Product Manager - IBM Power Virtualization
IBM Austin

Hariganesh Muralidharan, Systems Management Architect (HMC, IBM Cloud Management Console (IBM CMC), Novalink), IBM Power
IBM Bangalore

Jakub Koch
Technical Solution Architect - Entitled Systems Support (ESS), CIO BPT
IBM Copenhagen

Ronnie Moodley
Global Principal Sales Leader, IBM Power - AI and Modernization
IBM Austin

To the authors of *IBM Power11 E1150 Introduction and Technical Overview,* SG24-8589 and *IBM Power11 Scale-Out Servers: Introduction and Overview,* SG24-8590 for sharing content freely for this publication.

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author – all at the same time! Join an IBM Redbooks residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, IBM Redbooks
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on LinkedIn:

  https://www.linkedin.com/groups/2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/subscribe

- Stay current on recent Redbooks publications with RSS Feeds:

  https://www.redbooks.ibm.com/rss.html

1

Overview of IBM Power11

IBM Power11 represents an advancement in enterprise computing. Building on the strengths of IBM Power10, Power11 introduces more cores per chip, higher clock speeds, and improved energy efficiency. It continues IBM's focus on reliability, availability, and serviceability (RAS) while integrating quantum-safe security features to make critical workloads future-ready. The processor can support demanding enterprise applications, particularly ones involving artificial intelligence (AI), analytics, and hybrid cloud environments.

Power11 uses Integrated Stack Capacitor (ISC) technology and advanced 2.5D packaging to boost performance and efficiency. These innovations, when combined with enhanced thermal management, such as more efficient fans and heat sinks, improve system density, cooling effectiveness, and overall compute capability.

A key feature of Power11 is its continued support for AI workloads through the Matrix Math Accelerator (MMA) architecture and the integration of the IBM Spyre Accelerator, which is optimized for generative AI and complex model inference, and positions Power11 as a strong alternative to traditional GPU-heavy AI infrastructures. Also, Power11 strengthens its virtualization capabilities with deeper Kernel-based Virtual Machine (KVM) integration, enhancing compatibility with Linux-native tools and hybrid cloud platforms and making Power11 a versatile choice for enterprises seeking scalable, AI-ready infrastructure.

The following topics are covered in this chapter:

- IBM Power11: Advancing enterprise computing
- IBM Power roadmaps
- Power E1180: Resilient, efficient, and scalable
- Power11: Trusted, autonomous, and modern
- Power11 processor architecture
- Operating system support
- The IBM Power firmware and Hardware Management Console
- Power E1180 rack integration and handling guidelines

© Copyright IBM Corp. 2025.

1.1 IBM Power11: Advancing enterprise computing

For more than 35 years, IBM Power servers have been a cornerstone of enterprise computing, delivering unmatched performance, reliability, and availability for mission-critical workloads worldwide. From the original RS/6000 systems that were powered by the IBM POWER processor to the advanced Power10 platforms, IBM Power consistently evolves to meet the demands of modern IT infrastructure.

The IBM Power Architecture® follows a consistent innovation cadence, introducing a new generation of processors approximately every 3 years. Each iteration delivers enhancements, such as more cores and threads, improved energy efficiency, increased memory bandwidth, and expanded I/O capabilities, which helps ensure that the platform remains at the forefront of enterprise computing.

The Power11 platform is built on three foundational principles:

- Trusted

  Provides continuous business operations with up to 99.9999% availability[1] (on IBM Power E1180), zero planned downtime (ZPD) for maintenance[2], quantum-safe cryptography, and accelerated system recovery.

- Autonomous

  Improves operational efficiency through intelligent automation, dynamic performance tuning, and reduced manual intervention.

- Modern

  Enables AI-infused application deployment across hybrid environments (on-premises and in the cloud) by using off-chip AI acceleration and seamless integration with platforms such as Red Hat OpenShift and IBM watsonx.

The Power11 family is segmented into three primary tiers to address a wide range of workload requirements:

- Entry-level: IBM Power S1122 and IBM Power S1124, and the Linux-based IBM Power L1122 and IBM Power L1124, which are optimized for departmental and edge workloads.
- Midrange: IBM Power E1150, which is ideal for enterprise consolidation and scalable virtualization.
- High-end: IBM Power E1180, which is designed for large-scale, mission-critical, and AI-intensive workloads.

Power11 introduces several key advancements across the portfolio:

- Up to 6 nines (99.9999%) availability on high-end systems.
- Enhanced security with quantum-safe cryptography and Secure Boot improvements.
- Performance gains of up to 15% per core and up to 25% per thread through combined hardware and software optimizations.

---

[1] The IBM Power E1180 is the most resilient system in the history of the IBM Power platform. On average, Power E1180 systems are estimated to meet a 99.9999% stand-alone availability standard, as defined in “Enterprise Single Server Availability Standard” in the Power11 Processor-Based Systems RAS white paper.

[2] Based on IBM internal testing of system upgrade scenarios, many updates, such as Virtual I/O Server (VIOS), hot-plug adapters, I/O adapter firmware, and concurrent system firmware updates, can be performed in place. Some updates, such as non-concurrent system firmware and hardware maintenance, might require Live Partition Mobility (LPM) support.

- Significant throughput improvements in midrange and entry-level systems.
- Up to a 50% increase in memory bandwidth by using DDR5 and advanced memory controllers.

Also, Power11 introduces serial number–preserving upgrades from Power10 systems (available starting in Q4 2025), enabling seamless transitions without disrupting asset tracking or licensing.

This Redbooks publication focuses on the Power E1180, the enterprise-grade offering in the Power11 lineup. It delivers unprecedented performance and scalability for large-scale, mission-critical, and AI-intensive workloads.

1.2 IBM Power roadmaps

IBM continues to advance its IBM Power Architecture to address the evolving demands of AI, hybrid cloud, and mission-critical enterprise workloads. As a core component of IBM's hybrid cloud strategy, the IBM Power platform is optimized for Red Hat OpenShift, enabling containerized workloads and cloud-native application development.

IBM Power Virtual Server (PowerVS) extends these capabilities to the cloud. It enables you to run IBM Power workloads on-premises or in IBM Cloud®, while benefiting from flexible consumption models and meeting data residency and compliance requirements. This hybrid approach helps ensure seamless workload portability and supports modernization initiatives.

The IBM Power platform also maintains a robust ecosystem of operating systems (OSs), including IBM AIX (a UNIX based OS with a roadmap extending beyond 2035); IBM i (an integrated OS known for its security, stability, and ease of use); and Linux on IBM Power, which supports distributions such as Red Hat Enterprise Linux (RHEL) and SUSE Linux Enterprise Server. This flexibility makes it ideal for open-source and cloud-native environments.

Complementing its technology stack, IBM built a strategic network of alliances that expand the reach and capabilities of the IBM Power platform. PowerVS is a certified platform for RISE with SAP, enabling high-performance, reliable SAP workload deployment in hybrid cloud environments. Red Hat technologies such as Red Hat OpenShift and Ansible are integrated into IBM Power, supporting automation and DevOps practices. IBM also collaborates with major cloud providers, including Microsoft, Amazon Web Services (AWS), and Google Cloud to deliver hybrid and multicloud solutions, enabling IBM Power workloads to extend seamlessly into public cloud infrastructures.

This section describes IBM's roadmap for the evolution of IBM Power based infrastructure.

1.2.1 IBM POWER processor roadmap

The IBM Power roadmap continues to evolve with the introduction of the Power11 generation. Building on the innovations of Power10, Power11 introduces enhancements across processor architecture, packaging, and energy efficiency. A key advancement is the integration of an additional silicon layer for improved energy management. This feature enables higher performance with reduced power consumption. Power11 also increases core strength and count, enabling it to handle more demanding workloads. IBM is refining its Matrix Math Accelerator (MMA), first introduced in Power10 to further accelerate AI inferencing directly on the chip. This capability eliminates the need for external GPUs in many scenarios.

Looking beyond Power11, the IBM roadmap emphasizes full-stack innovation and adaptability to emerging technologies. The Power11 platform is memory-neutral, supporting both DDR4 and DDR5 through the Open Memory Interface (OMI). It also introduces improved thermal infrastructure through advanced packaging and cooling technologies. These developments enhance performance and reduce operational costs in data centers. Future iterations of the IBM Power Architecture are expected to continue this trend, focusing on hybrid cloud readiness, AI integration, and support for open-source ecosystems. IBM's long-term vision includes deeper collaboration with the OpenPOWER community and a renewed emphasis on openness, flexibility, and sustainability in enterprise computing.

Figure 1-1 shows the generations of the POWER processor family and the continued plans for improvements. With the release of Power11, the next generation of POWER processors is already in development, and more processor designs are in the pipeline.

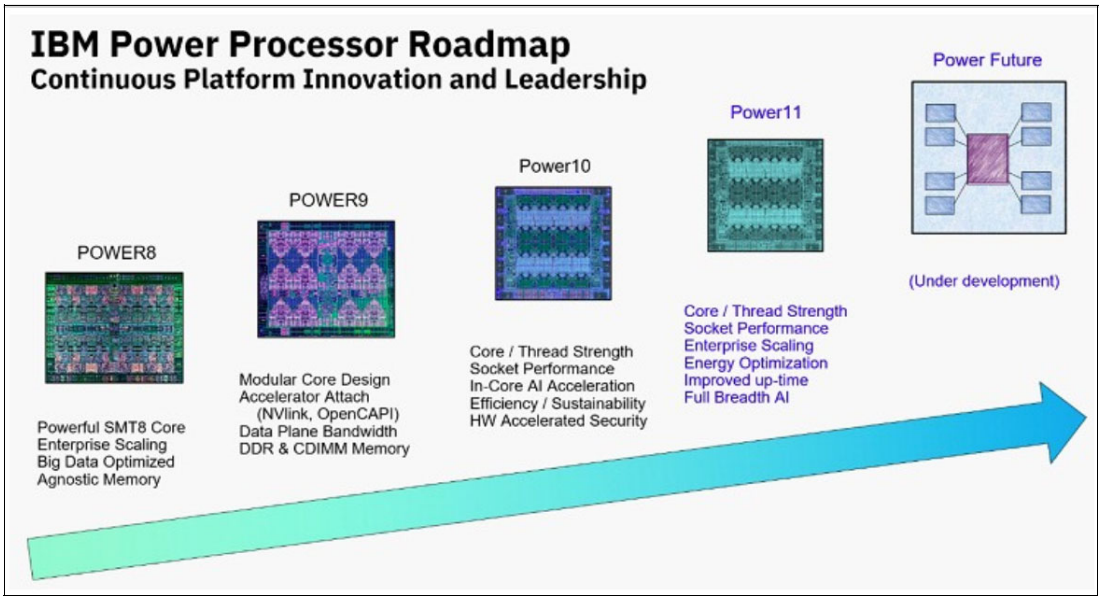

Figure 1-1 POWER processor roadmap

1.2.2 IBM AIX roadmap

IBM AIX is a proprietary UNIX based OS that is built on open standards and designed specifically for IBM Power servers.

For more than 38 years, AIX has powered mission-critical applications and databases in industries with high-performance computing (HPC) needs, such as finance, retail, healthcare, government, manufacturing, and insurance. Paired with IBM Power11 processor-based systems, AIX 7.3 (the latest version at the time of writing) expands its reach into new markets and workloads. It delivers unmatched performance, security, scalability, and reliability, while supporting digital transformation through flexible subscription models that are tailored to business needs. AIX also enables the adoption of technologies such as hybrid cloud, AI, and cloud-native applications.

AIX provides binary compatibility, enabling applications to run unchanged and without recompilation on the newest release. This capability safeguards IT investments in the platform.

IBM is committed to delivering an AIX release roadmap with continued innovation and purposeful support that extends beyond 2035. This commitment helps ensure a stable, future-ready platform for enterprise workloads.

Figure 1-2 shows the roadmap at the time of writing.

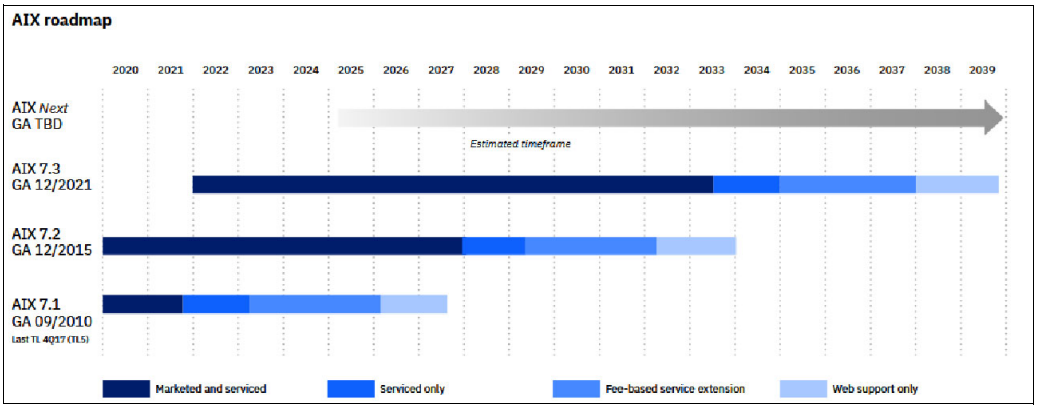

Figure 1-2 Support timeline for future generations of AIX

For more information about IBM's strategy and roadmap for IBM AIX, see Strategy and Roadmap for the IBM AIX Operating System. For more information about AIX, see 6.2, "AIX" on page 129.

1.2.3 IBM i roadmap

Since 2008, IBM has released new versions of IBM i approximately every 3 years, with IBM i 7.6 as the latest at the time of writing. Each release introduces new features, workload support, and language enhancements while maintaining compatibility with earlier versions. This compatibility enables you to preserve and extend existing investments.

IBM i is recognized as a secure, self-optimizing, and scalable OS. IBM continues to deliver updates and enhancements throughout each release lifecycle through Program Temporary Fixes (PTFs) and Technology Refreshes (TRs).

A key differentiator in the enterprise OS market is IBM's long-term commitment to IBM i. This commitment is demonstrated by its publicly available roadmap, which extends to 2035 and is updated annually. The roadmap underscores IBM's dedication to the platform's future.

Figure 1-3 on page 6 shows the roadmap at the time of writing.

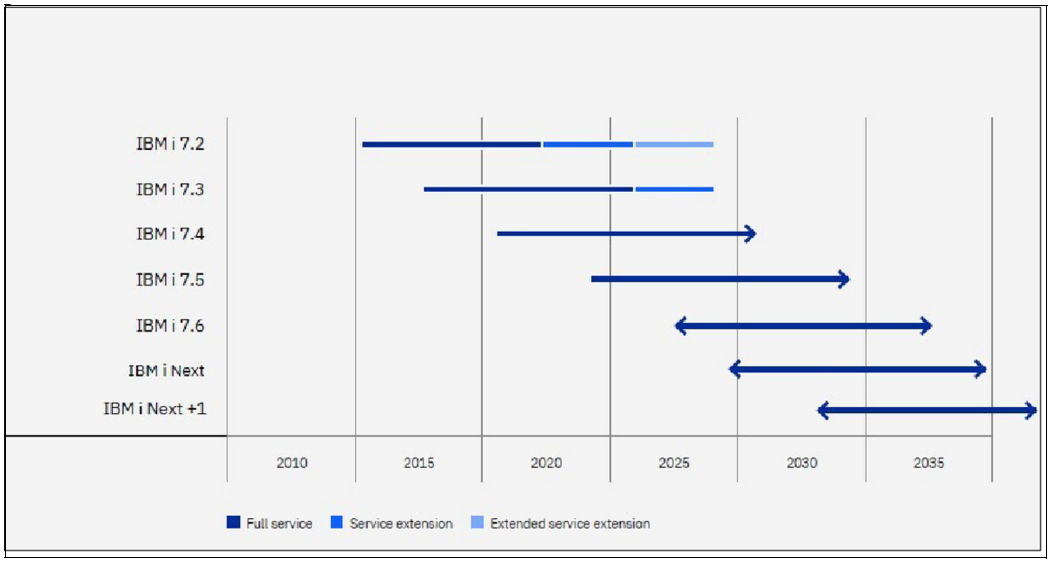

Figure 1-3 IBM i support matrix

1.2.4 PowerVM Virtual I/O Server roadmap

PowerVM VIOS is a critical component of the IBM Power virtualization stack. It enables shared access to physical I/O resources for client logical partitions (LPARs). Like all IBM software, VIOS follows a defined product lifecycle that includes general availability, support phases, and end-of-support milestones.

Table 1-1 shows the VIOS release schedule as of July 2025.

Table 1-1 VIOS release schedule

VIOS release	Release date	End of fix support[a]	Latest Fix Pack[b]	Next Fix Pack[c]
4.1.1	Dec 24	31 December 2027 (estimated)	VIOS_FP_4.1.1.0	23 July 2025
4.1.0	Nov 23	30 November 2026 (estimated)	VIOS_FP_4.1.0.30	23 July 2025
3.1.4	Dec 22	30 April 2026 (estimated)	VIOS_FP_3.1.4.50	23 July 2025
3.1.3	Sep 21	30-Sep-24	VIOS_FP_3.1.3.40	None
3.1.2	Nov 20	30-Nov-23	VIOS_FP_3.1.2.60	None

a. End of Fix Support (EoFS) marks the end of the maintenance period for a VIOS release. After EoFS, IBM does not provide Fix Packs or interim fixes for that release.
b. Latest Fix Pack is the most recent FP that is available for the VIOS release. It includes a link to that Fix Pack on IBM Fix Central.
c. Next Fix Pack indicates the target availability date for the upcoming Fix Pack for the specified VIOS release. After the final Fix Pack for a release becomes available, IBM provides further fixes as interim fixes.

Note: The end of service date for a VIOS release is not shown in this table. You can find end of service dates on the IBM Software Lifecycle page.

For more information about the PowerVM VIOS roadmap, see the PowerVM VIOS Lifecycle Information and System to PowerVM Virtual I/O Server maps.

1.2.5 Red Hat Enterprise Linux and IBM Power

Red Hat affirmed its strong commitment to the IBM Power platform, particularly as both companies align around hybrid cloud, AI, and open-source innovation. This partnership is rooted in a shared vision of delivering enterprise-grade solutions that are scalable, secure, and optimized for modern workloads.

At Red Hat Summit 2025, Red Hat emphasized its dedication to supporting IBM Power through continued enhancements to RHEL on IBM Power. These enhancements include performance tuning for Power10 and Power11 processors and deeper integration with Red Hat OpenShift for containerized and cloud-native applications. These efforts help ensure that you can use Red Hat's open hybrid cloud technologies across on-premises and cloud environments.

Red Hat also plays a pivotal role in enabling AI workloads on IBM Power. It provides support for AI inference and acceleration technologies that align with IBM hardware innovations such as the MMA and the upcoming IBM Spyre Accelerator. This synergy enables enterprises to deploy AI models efficiently on IBM Power infrastructure by using Red Hat's open-source toolchains and platforms.

In addition, Red Hat and IBM continue to collaborate on automation and DevOps through tools such as Ansible Automation Platform, which is optimized for managing IBM Power environments. This integration simplifies operations, enhances consistency, and accelerates application delivery across hybrid infrastructures.

For more information about RHEL support on IBM Power, see "Red Hat Enterprise Linux" on page 141.

1.2.6 SUSE Enterprise Linux and IBM Power

The partnership between SUSE Linux and IBM Power is a long-standing, strategic collaboration that is focused on delivering enterprise-grade Linux solutions for mission-critical workloads, hybrid cloud, and SAP environments.

SUSE and IBM have worked together for more than two decades to provide robust, scalable, and secure Linux solutions on IBM Power infrastructure. Their joint efforts focus on:

- SUSE Linux Enterprise Server for IBM Power
- SUSE Linux Enterprise Server for SAP Applications, which is optimized for SAP HANA on IBM Power
- Hybrid and multicloud enablement through container and virtualization technologies

SUSE's open-source model aligns with IBM's strategy to provide flexible, vendor-neutral solutions. This approach enables you to modernize IT environments while avoiding vendor lock-in.

For more information about SUSE Enterprise Linux, see "SUSE Linux Enterprise Server" on page 144.

1.3 Power E1180: Resilient, efficient, and scalable

The Power E1180 system introduces a comprehensive suite of advanced capabilities, all rooted in IBM's core strength: a full-stack, end-to-end design. By tightly integrating every layer of the infrastructure from the Power11 processor and system architecture to firmware, OSs, and cloud services, IBM Power delivers a synergistic platform that is built for autonomous IT.

This holistic approach drives measurable business outcomes across three foundational pillars:

- Unprecedented business continuity
- Enhanced productivity and efficiency
- Accelerated growth and scalability

Unprecedented business continuity

The Power E1180 system redefines operational resilience with features that are designed to eliminate downtime and enable continuous service delivery:

- Planned downtime elimination: By using technologies such as live updates, rolling upgrades, and autonomous patching, the Power E1180 enables maintenance without taking applications offline. These capabilities target zero hours of planned downtime for maintenance.
- Spare core technology: At the silicon level, the system includes spare processor cores that can be dynamically activated in response to hardware failures. This proactive fault-tolerance mechanism allows seamless failover without requiring system or application restarts, preserving compute capacity and system integrity.

  Note: The 16-core processor option (Feature Code EDQD) has no spare cores because all 16 cores are activated and available to you. A spare core is available on each chip for the 10-core and 12-core options (Feature Code EDQ0 and Feature Code EDQ9), providing four spares per node.

- IBM Power Cyber Vault: Advanced threat detection can identify ransomware attacks in under a minute. Combined with automated recovery mechanisms, this feature enhances cyber resiliency. It is offered in collaboration with IBM Storage and IBM Technology Expert Labs.
- Quantum-safe protection: The Power11 platform incorporates quantum-safe cryptography to secure system restarts and LPM, helping ensure future-ready data protection.
- Crypto compliance: Support for the IBM 4770 Crypto Card enables FIPS 140-3 Level 4 certification, enhancing compliance with stringent security standards.
- Automated diagnostics: Intelligent data collection accelerates error resolution, saving up to 8 hours per support ticket and reducing time to recovery.

Enhanced productivity and efficiency

The Power E1180 system is engineered to maximize IT efficiency and output:

- Autonomous IT operations (ITOps): Built-in automation and AI-driven management reduce manual intervention and operational impact.
- Integrated monitoring and management: Tools such as Hardware Management Console (HMC) and IBM Power Virtualization Center (PowerVC) provide centralized control, enabling streamlined operations across hybrid environments.

- Resource Groups: The innovative Resource Groups feature in Power11 enables more effective usage of Power11 resources across different environments or tenants.

Accelerated growth and scalability

The Power E1180 system supports dynamic scaling and hybrid cloud integration to meet evolving business demands:

- IBM PowerVS: Immediate availability of Power11 in IBM Cloud enables seamless extension of workloads to the cloud while maintaining consistent performance, security, and OS compatibility.
- Hybrid cloud agility: You can scale resources on demand, accelerate development and testing, and optimize costs through a consumption-based model.
- AI Acceleration with MMA: Power11 processor cores feature the MMA, which is a hardware acceleration unit that is optimized for AI and machine learning workloads. By running matrix operations directly on-chip, MMA reduces latency and boosts throughput, making the Power E1180 ideal for deep learning inference tasks.

Technical enhancements

Power11 based systems deliver enhanced performance through higher core frequencies, more cores per system, and improved memory bandwidth with reduced memory latency.

Table 1-2 shows some of the benefits that these improvements can bring to your enterprise.

Table 1-2 Power11 performance benefits

Improvement	Benefit
Up to 15 - 25% core performance increase with both hardware and software planned improvements	Exceed service-level agreements (SLAs) for response times and batch windows.
Performance increases due to larger core counts when using the new 16-Core Power11 Single-Chip Module (SCM)	Help meet increasing business workload demands by providing improved system performance.
Memory technology improvements with up to 50% higher data rates	Enhance performance and system scalability for memory-intensive workloads by optimizing both memory bandwidth and latency.

1.3.1 New Power11 system portfolio

The new Power11-based system portfolio includes the following offerings, giving you choices that are designed to meet specific business requirements:

- High-end: Power E1180
- Midrange: Power E1150
- Entry-level scale-out: Power S1124 and Power L1124, and Power S1122 and Power L1122

High-end: Power E1180

At the top of the portfolio is the Power E1180, a high-end enterprise system that is built for the most demanding workloads. It delivers exceptional performance, scalability, and resiliency, making it ideal for large-scale enterprise resource planning (ERP) systems, core banking, and high-throughput analytics. With support for massive memory footprints, advanced RAS features, and robust virtualization capabilities, the Power E1180 is designed for continuous availability and extreme workload consolidation.

Midrange: Power E1150

The Power E1150 serves as the midrange workhorse of the Power11 family. It brings many of the high-end capabilities of the Power E1180, such as spare core technology, pervasive memory encryption, and quantum-safe security into a more compact and cost-effective form factor.

The Power E1150 is ideal for organizations that need a balance between performance, scalability, and operational efficiency, especially in hybrid cloud environments.

Entry-level scale-out: Power S1124 and Power L1124, and Power S1122 and Power L1122

For smaller enterprises or distributed environments, IBM offers the Power S1124 and Power S1122 systems for general-purpose usage and the Power L1122 and Power L1124 systems, which are designed for the Linux marketplace. These entry-level, scale-out systems are optimized for cost-effective deployment of AIX, IBM i, and Linux workloads.

They are well suited for departmental applications, edge computing, and cloud-native workloads. Despite their compact size, these systems use the same Power11 processor technology, helping ensure consistent performance and security across the portfolio.

Figure 1-4 shows the new Power11 system portfolio.

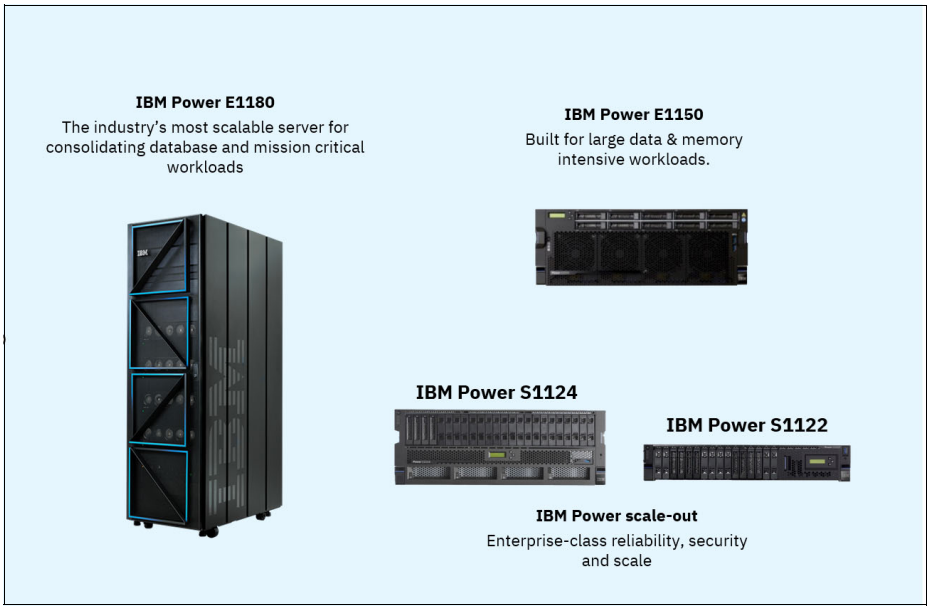

Figure 1-4 Power11 portfolio

Architectural technology Improvements

The Power11 portfolio introduces innovations across the product lineup and the underlying core technologies, all aimed at delivering enhanced performance and greater value. One of the most notable advancements is in the processor and memory architecture, where Power11 provides substantial improvements over the previous Power10 generation. These enhancements form the foundation for the next level of scalability, efficiency, and intelligent workload optimization.

Increased core counts and spare cores

Improved thermal dissipation solutions provide optimal cooling, supporting higher core densities and sustained performance.

Increased processor frequencies

ISC technology improves power delivery and stability at the silicon level, enabling higher performance under sustained workloads. Refined algorithms dynamically adjust processor frequency based on workload characteristics, maximizing performance while maintaining energy efficiency.

Increased memory frequencies

The IBM Power11 architecture introduces higher memory frequencies and increased memory bandwidth, resulting in substantial improvements in overall system performance. By supporting faster memory speeds and optimizing the memory subsystem, Power11 boosts memory throughput while reducing latency.

These enhancements provide quicker data access and processing, which is critical for memory-intensive workloads such as real-time analytics, large-scale databases, and AI applications. The combination of higher bandwidth and reduced latency enables smoother performance, better responsiveness, and more efficient use of compute resources across the entire system.

Table 1-3 provides a listing of the portfolio with a comparison to the Power10 offerings.

Table 1-3 Portfolio overview

Segment	Models	Footprint capacity	Power10 comparison
High-End (4S+)	Power E1180	► 4 nodes (central processing complex (CPC)) each 5U with 4 processor modules each ► Up to 64 TB of DDR5-based memory	► 15-core moving to 16-core ► Same memory capacity ► Introducing Resource Groups to maximize system usage ► In-place upgrades with serial number preservation
Mid-Range (4S4U)	Power E1150	► 2 - 4 processors in a 4U form factor ► Up to 30 cores per module ► Up to 16 TB of DDR5-based memory	► Up to 25% more core capacity ► Same memory capacity ► Introducing Resource Groups to maximize system usage ► In-place upgrades with serial number preservation

Segment	Models	Footprint capacity	Power10 comparison
Scale Out 2S4U	Power S1124 Power L1124	► Up to 2 processors in a 4U form factor ► Up to 30 cores per module ► Up to 8 TB of DDR5-based memory	► Up to 25% more core capacity and the same memory capacity ► Introducing Resource Groups to maximize system usage
Scale Out 2S2U	Power S1122 Power L1122	► Integrating the entry Single-Core Module (eSCM) modules into the Power S1122 family with 4 or 10 processors per eSCM ► Up to 30 cores per dual core module ► 2 processors in a 2U form factor ► Up to 4 TB of DDR5-based memory	► Up to 50% more core capacity ► Same memory ► Introducing Resource Groups to usage

Investment protection

IBM Power11 systems are designed with strong investment protection, enabling you to evolve your infrastructure without sacrificing prior investments. A key example is the ability to perform same-serial-number upgrades from the Power E1080 to the new Power E1180 and from the Power E1050 to the Power E1150. This upgrade path enables you to retain your existing system identity, simplifying software licensing, asset tracking, and operational continuity.

As part of this upgrade process, you can migrate existing memory modules and I/O adapters from your current systems, preserving valuable hardware investments. This compatibility reduces the cost and complexity of migrating to the latest technology while delivering the performance, scalability, and security benefits of the Power11 architecture.

Also, IBM supports IBM Power Enterprise Pools (PEP), including both Power E1080 and Power E1180 systems. This mixed-system support enables you to gradually migrate workloads from earlier to later systems within the same resource pool. It allows dynamic sharing of processor and memory activations across both generations, providing flexibility in capacity planning and workload placement. This approach protects existing investments and enables a smooth, nondisruptive path to modernization.

Portfolio simplification

One of the most notable changes in the Power11 generation is the streamlining of the scale-out system lineup. IBM reduced the number of scale-out models, simplifying your choices and reducing complexity in deployment and support. This focused approach makes it simpler to select the right system for your needs while still benefiting from the full-stack integration and innovation that IBM Power is known for.

Figure 1-5 on page 13 shows how you can migrate from existing IBM Power9® or Power10 systems to the Power11 portfolio.

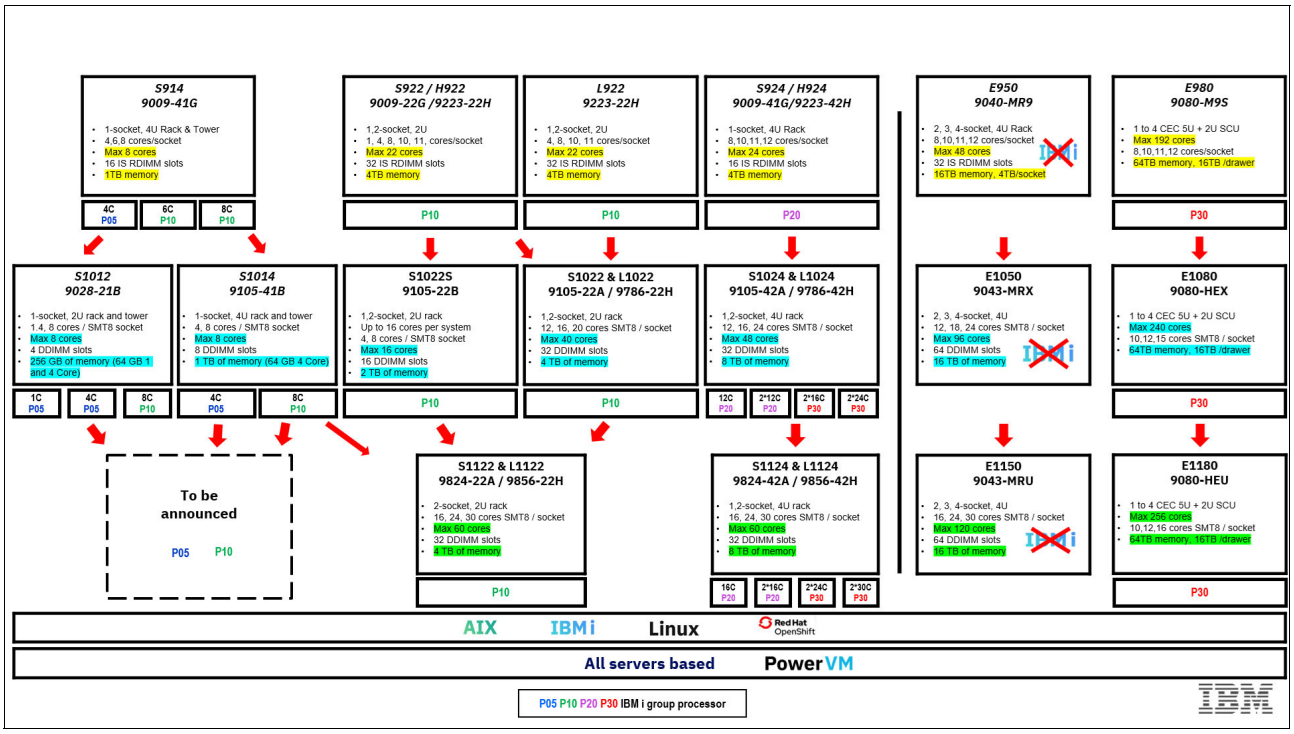

Figure 1-5 Migration path from Power9 to Power11

1.4 Power11: Trusted, autonomous, and modern

IBM's vision for IBM Power is autonomous IT, where Power11 processor-based systems can think, heal, and scale on their own. This vision supports business agility by enabling enterprises to respond quickly to market changes and customer demands, deploy new services rapidly, and reduce time to market for new products.

IBM is working toward this vision by building automation into the platform and combining that automation with AI-infused workflows.

This automation spans the day-0 to day-2 IT lifecycle for Power11 processor-based systems. It includes capabilities such as Infrastructure as Code (IaC) with IBM Terraform®, Application Configuration Management with Red Hat Ansible, and enhancements to existing platform tools such as the HMC and IBM Cloud Management Console. On the AI side, IBM is building AI-infused workflows with IBM watson.x and new tools such as IBM Concert®.

This section provides insights into the technologies and solutions that make Power11 a trusted, autonomous, and modern platform to support your business requirements.

1.4.1 Building a trusted infrastructure with Quantum-Safe Encryption

Infrastructure that is built on IBM Power processor-based systems benefits from robust security technologies that are integrated into both the hardware and software stacks. IBM Power processor-based systems provide advanced security features at every level of the system to help ensure comprehensive protection for sensitive data and applications. These features include advanced encryption technologies, Secure Boot capabilities, and integrated firmware updates. In addition, IBM Power processor-based systems use IBM expertise in securing mission-critical workloads, making them a preferred choice for organizations that require a secure environment for digital assets.

Workloads on Power11 processor-based systems gain significant advantages from improved cryptographic accelerator performance compared to previous generations. The Power11 chip supports accelerated cryptographic algorithms such as Advanced Encryption Standard (AES), SHA-2, and SHA-3, resulting in considerably higher per-core performance for these algorithms. This improvement enables features such as AIX Logical Volume Encryption to operate with minimal impact on system performance.

The processor-core technology of Power11 includes integrated security protections that are designed to deliver the following benefits:

- Improved cryptographic performance

  Integrated cryptographic support reduces the performance impact of encrypting and decrypting data, enabling you to make encryption pervasive to protect all critical data.
- Increased application security

  Hardened defenses help prevent return-oriented programming (ROP) attacks.
- Simplified hybrid cloud security

  Provides setup-free hybrid cloud security administration through a single interface.
- Enhanced virtual machine (VM) isolation

  Offers industry-leading VM isolation technology to defend against attacks that exploit operating system or application vulnerabilities in a VM to access other VMs or the host system.

Encryption technologies and their applications

Power11 emphasizes comprehensive security throughout its design and offers multiple encryption options. Key technologies include Transparent Memory Encryption (TME), Fully Homomorphic Encryption (FHE), and Quantum-Safe Encryption (QSE).

Transparent Memory Encryption

TME encrypts data in memory to protect it from unauthorized access and tampering during run time. Operating at the hardware level, TME uses the Power11 processor's cryptographic engines to perform encryption and decryption efficiently. TME provides pervasive protection of in-memory data with minimal impact on system performance because encryption and decryption occur at the chip level. Its integration into normal operations is seamless and automatic.

Fully Homomorphic Encryption

FHE enables computations to be performed directly on encrypted data without decrypting it first. This capability helps ensure that sensitive data remains confidential during processing. FHE operates at the software level and uses advanced mathematical algorithms to perform computations on ciphertext.

Implementing FHE requires specialized libraries and frameworks. However, FHE is computationally intensive and can introduce performance processing impact compared to conventional hardware-only encryption methods because of the complexity of the algorithms.

Quantum-Safe Encryption

QSE is designed to resist quantum attacks, protecting data against the computational capabilities of future quantum computers that might break current cryptographic algorithms. QSE uses cryptographic algorithms that are believed to be resistant to quantum attacks, such as lattice-based, hash-based, and multivariate-quadratic-equation-based cryptography.

Many quantum-safe algorithms are still undergoing testing and standardization to help ensure that they provide robust security against future quantum advancements. QSE is typically used to secure long-term data, sensitive communications, and critical infrastructure.

Table 1-4 shows the relevant features and differences in these technologies.

Table 1-4 Key differences

Feature	TME	FHE	QSE
Encryption scope	Secures data in memory.	Enables computations on encrypted data.	Prevents against future quantum computing threats.
Implementation level	Implemented in hardware.	Implemented by using a combination of hardware and software.	Implemented by using a combination of hardware and software.
Performance Impact	Hardware that is accelerated through the Power11 cryptographic engines and designed to have minimal performance impact.	Involves substantial computational impact.	The impact of QSE varies: some quantum-safe algorithms might introduce performance impact, which is a subject of ongoing research.
Use cases	Used to protect data in memory.	Used for performing secure computations on sensitive data without decrypting it.	Provide long-term data protection and secure communications in the future.

Quantum-safe compliance

QSE, also known as Post-Quantum Cryptography (PQC), refers to encryption methods that are secure against both classical and quantum computers. As quantum computing advances, it might threaten existing cryptographic systems and compromise their security. QSE is essential for protecting sensitive data, communication channels, and user identities in the era of quantum computing.

The urgency of adopting QSE stems from two primary concerns:

- Harvest now, decrypt later (HNDL) risk

  Advanced quantum computers might enable adversaries to intercept and store encrypted communications today and decrypt them later, even before reaching Q-Day. Q-Day refers to the anticipated point when quantum supremacy becomes widespread and many current encryption algorithms become ineffective.

- Lengthy migration timelines

  Migrating to QSE can take more than a decade because of the complexity of organizational structures and IT infrastructure.

Organizations should begin evaluating and implementing QSE solutions immediately to maintain protection and preserve stakeholder trust.

Delaying QSE adoption can have severe consequences. Cryptographic systems left unchanged might be compromised if there is a successful quantum attack, exposing sensitive data and jeopardizing confidential business transactions and individual privacy. Financial institutions, critical infrastructure providers, and government agencies might face challenges in maintaining operational integrity and confidentiality. Therefore, prioritizing QSE implementation is essential for long-term cybersecurity resilience.

Power11 is designed to support quantum-safe algorithms, helping ensure robust security as quantum computing advances. Power11 quantum-safe features provide the following protections:

- Data encryption breakage protection
  - Risk: Quantum computers might break widely used cryptographic algorithms such as RSA, Elliptic Curve Cryptography (ECC), and traditional Diffie-Hellman key exchange protocols. For example, Shor's algorithm can efficiently factor large integers and solve discrete logarithms, compromising these algorithms.
  - Protection: Power11 supports quantum-safe algorithms such as lattice-based, hash-based, code-based, and multivariate-quadratic-equation-based cryptography, which are believed to resist quantum attacks. The Power11 cryptographic engines enhance the performance of these algorithms, helping ensure secure encryption and key exchange with minimal performance impact.
- Secure communications
  - Risk: Quantum computers might intercept and decrypt secure communications, undermining protocols that rely on classical encryption methods.
  - Protection: Power11 secures communication channels with quantum-resistant protocols, maintaining end-to-end encryption throughout the data lifecycle, from storage to transmission by using algorithms that are resistant to quantum attacks.
- Data integrity and authenticity
  - Risk: Quantum computers might forge digital signatures or tamper with data.
  - Protection: Power11 supports quantum-safe digital signature algorithms such as eXtended Merkle Signature Scheme (XMSS) and Unbalanced Oil and Vinegar (UOV), providing strong protection against quantum attacks.

- Long-term data protection
  - Risk: Sensitive data stored today might be harvested and decrypted in the future as quantum computers become more powerful, threatening long-term confidentiality.
  - Protection: Implementing QSE methods protects data over time, even as quantum capabilities evolve. The Power11 architecture supports updates to cryptographic libraries and protocols, enabling adoption of new quantum-safe algorithms as they are developed and standardized.
- Physical and memory attack protection
  - Risk: Physical attacks on memory, such as cold start attacks, might expose sensitive data if the memory is not adequately protected.
  - Protection: TME encrypts data in memory, protecting it from physical attacks during run time.

Power11 supports the following quantum-safe algorithms:

- Lattice-based cryptography
  - Algorithms: ML-DSA (Dilithium), ML-KEM (Kyber), and NTRUEncrypt.
  - Characteristics: Secure against quantum attacks. Based on lattice problems such as Learning With Errors (LWE) and Ring-Learning With Errors (Ring-LWE). Relatively efficient for hardware and software implementations.
- Hash-based cryptography
  - Algorithms: Merkle Signature Scheme (MSS), XMSS, and SPHINCS+.
  - Characteristics: Security relies on hash functions. Typically produces larger signatures and keys.
- Code-based cryptography
  - Algorithms: McEliece Cryptosystem, Bit Flipping Key Encapsulation (BIKE), and Hamming Quasi-Cyclic (HQC).
  - Characteristics: Quantum-resistant based on decoding random linear codes. Public keys can be large.
- Multivariate quadratic equations
  - Algorithms: UOV, and Rainbow.
  - Characteristics: Secure against solving systems of multivariate quadratic equations. It is efficient in signature generation but might require larger key sizes.

Power11 implementation

Power11 processors support these quantum-safe algorithms through the following capabilities:

- Crypto engines

  Multiple engines per core enable efficient execution of cryptographic operations.
- Software updates

  The architecture supports updates to cryptographic libraries, helping ensure integration of new quantum-safe algorithms as they become standardized.

The Power11 design and capabilities deliver robust security against future quantum threats by combining hardware acceleration with flexible software updates, maintaining high security standards as the cryptographic landscape evolves.

Encryption enablement in hardware

There are two options for accelerating encryption in an IBM Power11 processor-based system:

- Built-in encryption acceleration

  The Power11 chip includes integrated encryption acceleration for efficient cryptographic operations.

- PCIe-based encryption accelerator

  IBM Power also supports a Peripheral Component Interconnect Express (PCIe)-based encryption accelerator for more performance.

On-chip encryption support in Power11

The Power11 processor is designed to support future encryption technologies, including F FHE and QSC to prepare for the quantum era. The Power11 Instruction Set Architecture (ISA) is optimized for these solutions' software libraries, which are available or will soon be available in open source communities at the time of writing.

Workloads on Power11 benefit from cryptographic algorithm acceleration, delivering higher per-core performance than Power10 processor-based systems for algorithms such as AES, SHA-2, and SHA-3. Features such as AIX Logical Volume Encryption can be enabled with minimal performance impact because of these enhancements.

With four times as many AES encryption engines, Power11 processor technology provides noticeably faster encryption performance. Power11 processor-based systems are more advanced than Power10 systems, meeting today's most stringent standards and supporting future cryptographic requirements, including post-quantum and FHE. Power11 also introduces improvements to container security. Through hardware-based features, TME simplifies encryption and supports end-to-end security without compromising performance.

1.4.2 Protecting, detecting, and recovering with IBM Power Cyber Vault

Regulatory frameworks such as the U.S. Securities and Exchange Commission (SEC) cybersecurity disclosure rules, and the Digital Operational Resilience Act (DORA) are no longer distant challenges, but are active requirements. More regulations are emerging, each demanding stronger risk management and governance, mechanisms to detect anomalous activities promptly, and resilience practices to minimize the risk of data corruption or loss. The bar is rising, and every organization's security posture is under scrutiny.

These frameworks go beyond checklists. They signal a fundamental shift: cybersecurity is no longer a compliance add-on but a core element of operational continuity, customer trust, and business strategy.

Traditional siloed security solutions can no longer keep pace

Most organizations still operate with fragmented security architectures, which are collections of tools, processes, and teams that function in silos. Although some integration has occurred over time, loosely connected security domains leave critical gaps in visibility, context, and response coordination.

A siloed approach results in the following challenges:

- Limited threat visibility across domains

  For example, linking identity misuse with network anomalies becomes difficult.
- Reactive security postures

  Success is measured by containment speed rather than business outcomes.
- Barriers to collaboration and innovation

  Teams struggle to securely share and access data.
- Missed opportunities to embed security and privacy

  Customer-facing products and services lack integrated protection.

Even more concerning, a lack of contextual awareness across tools and data streams can open the door to full-scale compromises, such as ransomware attacks that paralyze operations.

Security was never intended to be siloed. It evolved that way because of necessity and earlier architectures. That model is no longer sufficient. Forward-looking enterprises are rethinking cybersecurity as a defense mechanism and business enabler. They are integrating tools, processes, and insights across domains, from access and identity to threat detection, data governance, and compliance. They are moving from isolated defenses to connected intelligence.

Zero trust: A strategic shift

Many organizations are adopting zero trust frameworks as the foundation for next-generation security programs. Built on the principles of least-privilege access, continuous verification, and assumed breach, zero trust provides a compelling blueprint for modern cybersecurity.

Applying zero trust effectively requires the following items:

- A deep understanding of your organization’s risk surface
- Alignment with regulatory and operational requirements
- Integration with existing cloud and hybrid infrastructure
- Cultural and organizational readiness to adopt continuous governance and access control

Zero trust is not a product. It is an operating model for security that aligns with the way modern businesses function: dynamically, digitally, and in a distributed manner.

The reality of today's threats: Evolving, disruptive, and costly

Cyberattacks today are fast-moving, complex, and highly disruptive. A typical incident flow includes the following steps:

1. A service disruption is detected.
2. Immediate shutdowns are initiated to contain the spread.
3. A search for uncompromised backups begins.
4. Slow and costly recovery efforts start.

According to recent reports, 78% of organizations take more than 100 days to fully recover from a major cyber incident[3]. This delay represents more than downtime, which results in lost trust, revenue, and opportunity.

IBM provides a unified cyber resilient solution that includes IBM Power11 processor-based systems, IBM FlashSystem storage, advanced software, and services from IBM Technology Expert Labs. This solution, which is delivered through IBM Power Cyber Vault, helps organizations respond to evolving cyberthreats and meet regulatory standards.

IBM Power Cyber Vault

IBM Power Cyber Vault keeps your business running by combining the following features:

- IBM Power Cyber Vault protected data

  Automated immutable images that are stored in the IBM Power Cyber Vault protect client snapshots and backups.

- Continuous threat detection

  Near real-time file system monitoring, IBM FlashSystem inline detection, and recovery data scanning. Threat detection time can be as low as 60 seconds.

- Rapid clean room response

  Automated attack response creates IBM Power Cyber Vault clean rooms for testing and validation, accelerating recovery.

- Cyber Vault Assisted Recovery

  Automated multi-level testing and validation of IBM Power Cyber Vault images accelerate system recovery, enabling a return to full operations in hours instead of months.

IBM Power Cyber Vault solution components

IBM Power Cyber Vault is a complete cyber resilient solution that can be deployed in as little as 30 days. It includes the following components:

- IBM Power11 processor-based systems
  - Industry-leading threat resistance
  - PowerSC for security and compliance tracking
  - Built-in Secure Boot and memory encryption
  - Scalable, flexible architecture from entry-level to enterprise systems
- IBM FlashSystem Storage
  - Immutable copies for data protection
  - Ransomware detection for real-time threat identification
  - Flash-speed recovery for rapid restoration of snapshot data
  - Storage Insights Pro for comprehensive reporting
- IBM Technology Expert Labs – IBM Power Cyber Vault Deployment
  - A design workshop that is tailored to client-specific requirements
  - IBM Power Cyber Vault deployment to customize and enable the unified solution
- IBM Expertise Connect
  - Primary technical point of contact for IBM Power Cyber Vault
  - Dedicated to client success
  - Available postinstallation to guide client strategy and direction

[3] Source: https://www.cohesity.com/content/dam/cohesity/resource-assets/research-reports/cyber-resilience-global-survey-report-en.pdf

IBM Power Cyber Vault implementation

The implementation of the solution consists of several phases, each designed to support risk mitigation and enable the fastest possible recovery:

- Identify: Assess risks and design mitigation solutions
  - IBM Security® and Resilience Assessment.
    - Provides an overview of the organization, processes, and strategies for operational and cyber resilience.
    - Delivers a prioritized list of recommendations to improve overall resilience.
  - IBM Cyber Vault Solution and Design Workshop enables the creation of a detailed solution based on client cyber resiliency goals, infrastructure, and required services.
- Protect: Safeguarded Cyber Vault copies and backups

  Provides client-specific scheduling for creating and protecting immutable snapshots and backups to meet Recovery Point Objective (RPO) and Recovery Time Objective (RTO) requirements.
  - Snapshots: IBM FlashSystem and Storage Defender CSM save protected system snapshots into the Cyber Vault on a regular schedule.
  - Orchestration for crash consistency.
  - Application-specific customization is available.
- Detect: Real-time threat detection

  Monitors real-time threats and cyberattack vectors to enable rapid response by using PowerSC Real Time Compliance orchestration.
  - Zero Trust Execution detects, prevents, and reports unapproved executable files and applications on Power11 processor-based systems.
  - IBM FlashSystem integrated ransomware detection alerts on active attacks and reports to PowerSC through IBM Storage Insights Pro.
  - Security Information and Event Management (SIEM) integration accepts alerts and threat reports from SIEM systems.
- Respond and recover: real-time threat response

  When threats are detected, the response phase initiates auditing and recovery.
  - On receiving an integrity alert, IBM Power Cyber Vault automatically creates an immutable backup of the VM within seconds and spins up a copy in the clean room within minutes.
  - Runs multiple integrity checks in the clean room, including platform-specific tests.
  - Identifies corrupted copies and finds the newest clean copies.
  - After administrator approval, recovery to production meets RPO and RTO requirements.

1.4.3 Quantum-safe compliance with one-click inventory discovery

IBM PowerSC is a comprehensive security and compliance solution that is designed for virtualized environments running on Power servers with AIX, Linux, or IBM i OSs. It integrates deeply with the IBM Power platform to provide end-to-end protection through features such as security and compliance automation, trusted boot, trusted firewall, and trusted logging.

PowerSC helps organizations meet stringent regulatory standards such as PCI DSS, HIPAA, and SOX by automating the monitoring, auditing, and enforcement of security policies. It also includes advanced capabilities such as multi-factor authentication, intrusion detection, and fix management. With tools like PowerSC Trusted Surveyor, it helps ensure consistent network configuration and compliance across dynamic virtual environments, reducing administrative impact while maintaining strong security postures.

IBM is actively involved in developing and adopting quantum-safe cryptographic standards, including collaboration with NIST on its PQC Standardization project. PowerSC supports a seamless transition to NIST's forthcoming standards and offers capabilities to help verify and manage QSE.

The Quantum Safety Analysis feature evaluates the quantum-safe status of your VMs to identify areas of concern and assist with compliance with emerging quantum-safe cryptographic standards:

- Endpoint scanning

  PowerSC scans endpoints on your system to identify cryptographic artifacts and vulnerabilities that are related to quantum safety. Scans can also be scheduled.

- Encryption inventory

  Creates an inventory of where different encryption algorithms are implemented on your system.

- Cipher strength assessment

  Estimates the strength of ciphers, certificates, and keys on your AIX endpoints and categorizes them as weak, strong, quantum-safe, or unclassified. PowerSC is a security and compliance management tool, not a cryptographic module validation tool.

- Quantum Safety Analysis report

  Displays discovered cryptographic elements and their assessed strengths in the PowerSC GUI.

1.4.4 Zero planned downtime and end-to-end orchestration for maintenance events

Planned downtime remains one of the biggest challenges in ITOps, especially for a critical infrastructure such as Power11. IBM addresses this challenge by delivering *Autonomous IT* through extreme automation and AI-infused workflows that are embedded directly into the platform. This approach reduces operational effort, minimizes disruptions, and improves resilience.

Today's enterprise IT environments face increasing pressure to help ensure infrastructure reliability and security without disrupting operations. Key challenges include the following items:

- Operational complexity: Maintaining firmware, I/O adapters, and VIOSs is highly specialized and intricate.
- Service interruptions: Routine maintenance often requires downtime, which impacts application availability and business continuity.
- Security risks: Delayed updates expose systems to vulnerabilities, such as Common Vulnerabilities and Exposures (CVEs), which weaken compliance and increase risk.

- Resource drain: A significant amount of time is spent on maintenance tasks, such as fix evaluation, compatibility testing, and scheduling.
- High downtime costs: With downtime averaging $336,000 per hour, annual maintenance-related outages can exceed $2 million per system[4].

These issues force organizations to make a difficult tradeoff: stay current and risk downtime, or delay updates and increase exposure. This approach is unsustainable in today's always-on digital economy.

Defining infrastructure

In this context, *infrastructure* refers to the Power11 platform, which includes firmware, I/O components, VIOS, and the OS. To maintain performance and security, clients must complete the following tasks:

- Apply fixes and updates regularly.
- Remain on supported versions.
- Respond promptly to security advisories.
- Conduct preventive maintenance.

However, these actions typically require planned downtime. The core dilemma is whether businesses should accept regular service interruptions to stay current or defer updates and risk security and performance.

IBM Power11: A platform for autonomous IT

IBM Power11 is designed to resolve this dilemma by enabling continuous business operations while keeping infrastructure secure, up to date, and high performing. IBM Power11 introduces a transformative approach to infrastructure lifecycle management that is purpose-built for Autonomous IT. The platform operates with minimal human intervention by using embedded intelligence to monitor, diagnose, and remediate issues in real time. This capability enables an infrastructure that is inherently resilient, adaptive, and capable of maintaining continuous service availability.

At the core of Power11 is the *Automated Maintenance Framework*, which is orchestrated by IBM Concert, an AI-driven engine that automates the full maintenance lifecycle. This framework includes the following capabilities:

- Unified update orchestration for system firmware, VIOSs, and I/O adapters
- LPM to provide uninterrupted application availability during updates
- Prevalidation checks to confirm system readiness and compatibility
- Flexible update sourcing from IBM repositories, SFTP, NFS, USB, or HMC local file systems

Maintenance operations can run autonomously or manually through the HMC, providing both automation and administrative control. This architecture transforms maintenance from a disruptive, manual process into a seamless background operation.

[4] Source: https://itic-corp.com/itic-2024-hourly-cost-of-downtime-part-2/#:~:text=ITIC's%2011th%20annual%20Hourly%20Cost,costs%20their%20company%20over%20%24100%2C000

IBM Concert: AI-powered application management for the modern enterprise

IBM Concert addresses the persistent challenge of siloed application data by using AI to deliver intelligent, prioritized recommendations. By streamlining issue identification and resolution, Concert reduces the mean time to resolution (MTTR) for critical risk factors such as the following ones:

- CVEs
- Certificate management issues
- Application compliance challenges

What sets IBM Concert apart

Built on IBM watsonx, IBM Concert uses advanced generative AI to analyze complex environments and deliver actionable insights.

- Data-neutral integration

  IBM Concert supports a wide range of data sources across networks, infrastructure, and application architectures.

- Hybrid cloud by design

  IBM Concert seamlessly supports hybrid cloud environments to meet the requirements of modern enterprise IT.

- AI-infused orchestration for proactive maintenance

  IBM Concert orchestrates maintenance operations through intelligent workflows and end-to-end automation, including the following capabilities:

  - Inventory and risk discovery: Automatically identifies assets and uncovers hidden risks across the environment.
  - Risk analysis and remediation planning: Delivers prioritized, actionable plans to address vulnerabilities and compliance gaps.
  - Automated risk resolution: Enables automated actions to remediate issues, reducing manual effort and downtime.

Automation can run directly from the IBM Concert solution or from the HMC, providing flexibility and administrative control.

Key benefits of IBM Power11 autonomous maintenance

The IBM Power11 integrated solution for ZPD delivers measurable benefits across operational, security, and financial dimensions:

- Business continuity: Updates apply without interrupting workloads, preserving uptime and service-level commitments.
- Operational efficiency: Automation streamlines the maintenance lifecycle, reducing manual effort and administrative impact.
- Risk reduction: AI-driven orchestration provides timely updates, minimizing exposure to vulnerabilities and compliance risks.
- Productivity gains: IT staff benefit from a simplified, intelligent interface that reduces time on task from weeks to minutes.
- Cost optimization: Eliminating downtime and manual labor reduces the total cost of infrastructure operations.

Together, these capabilities position IBM Power11 as a foundational platform for secure, scalable, and self-managing enterprise IT infrastructure that aligns with the demands of modern digital transformation.

The new maintenance model

From the IT administrator's perspective, maintenance is now part of a unified, intelligent flow that includes the following steps:

- Inventory discovery and risk detection
- Smart planning
- Automated download and application of updates
- Validation
- VM (LPAR) evacuation and reintegration

This approach minimizes disruptions, helps ensure service-level agreement (SLA) compliance, and enables faster, more timely maintenance. Systems remain secure, stable, and aligned with evolving requirements without downtime or performance impact.

The Power11 built-in Automated Maintenance Tool enables platform maintenance without affecting applications. This automation can run through IBM Concert orchestration or be initiated manually from the HMC, independent of IBM Concert.

IBM Concert includes the following capabilities:

- Ability to update system firmware, VIOSs, and I/O adapter microcode from a single update flow, supporting both concurrent and disruptive updates
- Validation for LPM and VIOS redundancy (VIOS maintenance readiness check)
- Ability to automatically migrate partitions and return them as part of the update process:
  - Option to return to the source system or remain on the target system
  - Option to evacuate all LPARs or select a subset of LPARs and define the order
  - Option to choose the order of updates
  - Option to download only or download and update
- Support for multiple update sources:
  - IBM website (preferred and recommended for end-to-end automation)
  - SFTP server
  - HMC file system
  - NFS server
  - USB

The new tool appears as a Licensed Capability of the system, as shown in Figure 1-6.

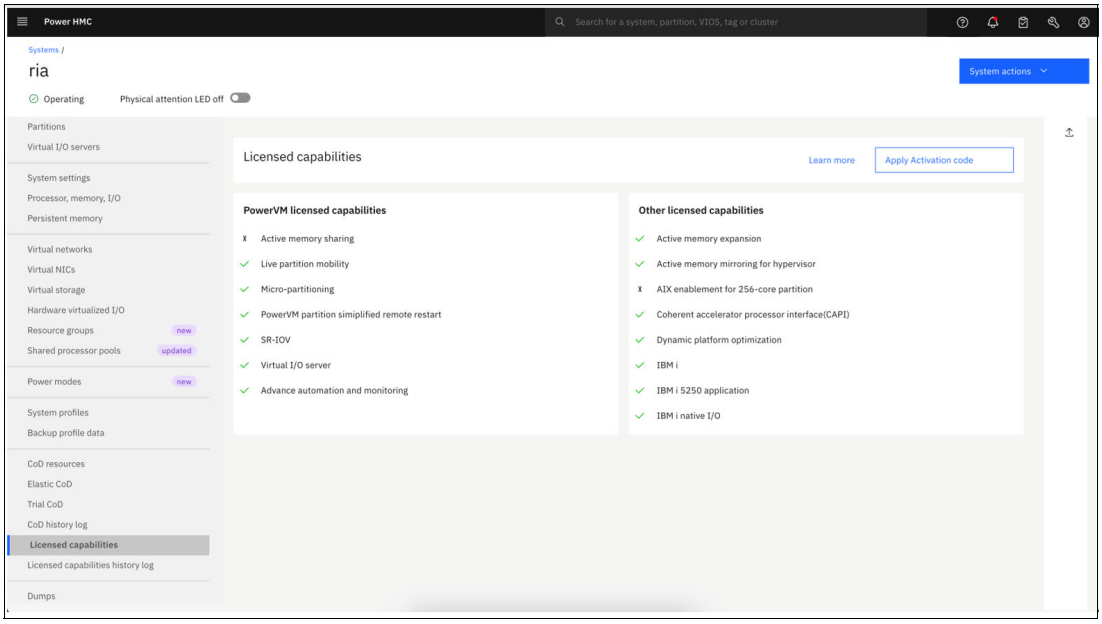

Figure 1-6 IBM Power licensed capabilities

For more information about using the Automated Maintenance Tool, see 5.1.7, “Using the Automated Maintenance Framework tool” on page 114.

1.4.5 Automated data collection for faster error resolution

Enterprises are increasingly adopting hybrid cloud environments, with 82% reporting usage as of 2024 (IBM Cloud Survey). This shift introduces operational complexity, particularly in managing infrastructure and resolving issues. Manual error resolution continues to overwhelm IT teams, with 60% of organizations citing insufficient staffing for incident management. Moreover, the financial impact of downtime is escalating with over 90% of midsize and large enterprises reporting that a single hour of downtime now exceeds $300,000 in cost[5].

To address these challenges, IBM introduced enhanced support capabilities in the Power11 HMC. These capabilities include a streamlined case creation process and automated log collection through Call Home. You can initiate support cases directly from the HMC, and the system automatically gathers and transmits relevant First Failure Data Capture (FFDC) logs to IBM Support.

Supported log types include:

- System firmware (dumps)
- HMC (pedbg)
- VIOS (snap)
- Cloud connector logs
- Partition migration logs

Call Home configuration is required to create a case in CSP and upload logs. If Call Home is not configured, the HMC creates a problem record, but it does not create a case. Logs remain stored in the HMC and are not uploaded automatically.

[5] Source: https://uptimeinstitute.com/2019-data-center-industry-survey-results

Note: As a best practice, enable Call Home because it is also required for Autonomous Error Resolution (AER) data collection.

If the HMC is a Virtual HMC and the problem is an HMC issue, the system collects logs when the issue occurs. However, it does not initiate a Call Home, and the logs are not sent automatically.

Use case: automated support workflow

Consider Louis, a system administrator at a large insurance company that runs critical back-end applications on IBM Power infrastructure. During a routine maintenance task involving LPM, Louis encounters a migration validation error. Previously, resolving such issues required manually opening a support case, coordinating with IBM Support, and collecting FFDC data. This process might take 5 - 8 hours before investigation begins.

With Power11, you can now open a support case directly from the HMC. The system automatically collects and submits the necessary diagnostic data, eliminating manual steps and reducing turnaround time. IBM Support uses AI-enhanced diagnostic tools that are trained on IBM Power processor-based systems to accelerate root cause analysis and resolution.

Benefits of the enhanced support model

Implementing this new automated data collection capability with applied AI workflows in IBM Support provides the following benefits:

- Reduced time to resolution: Automated log collection and AI-assisted diagnostics accelerate issue resolution.
- Improved productivity: System administrators save hours by avoiding manual data gathering and coordination.
- Reduced operational risk: Faster resolution minimizes downtime and its associated business impact.
- Enhanced support experience: A simplified, integrated workflow improves responsiveness and reduces stress for IT teams.

1.4.6 Increasing efficiency and reducing energy consumption through smart scheduling options

In recent years, sustainability has become a central strategic priority in the IT industry due to regulatory pressure and business value. Over the past 5 years, climate commitments among companies increased by 800%, reflecting broader awareness and the recognition that sustainable practices can enhance profitability. At the time of writing, 83% of organizations are investing in research and development for low-carbon products and services, and products with sustainability attributes are achieving revenue growth rates more than 25% higher than traditional products[6].

Although 95% of companies established operational sustainability goals, only 41% achieved measurable progress[7]. Many organizations face challenges in starting the implementation process. This situation highlights the need for explicit guidance, stronger governance frameworks, and better integration of sustainability metrics into core business operations.

[6] Source: https://www.pwc.com/us/en/services/esg/library/assets/pwc-sustainability-decarbonization-2025.pdf

[7] Ibiid

IBM Power processor-based systems have consistently improved energy efficiency per watt with each generation, and Power11 continues this trend. Power11 enhances efficiency, reducing energy use, carbon emissions, and the data center energy footprint.

Clients upgrading from Power9 to Power11 can reduce energy consumption by up to 60% for the same performance. Power11 outperforms x86 systems, delivering twice the performance per watt. Compared to Power10, Power11 offers up to 33% better performance per watt, with the Power S1122 model showing the largest gain[8].

Figure 1-7 illustrates natural IT efficiency improvements from Power9 to Power11, offering up to a 3:1 consolidation ratio, reducing energy use and carbon emissions by approximately 60%, and cutting the data center footprint by two-thirds.

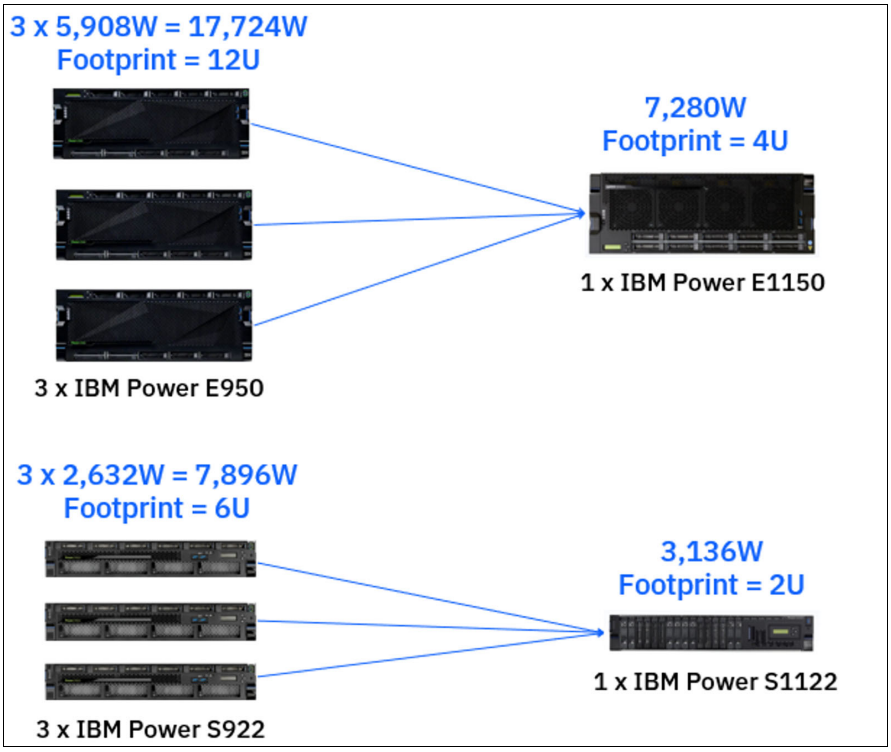

Figure 1-7 Power9 to Power11 energy consumption reduction

In addition to the base energy savings that are delivered by Power11, new features were introduced to address client feedback and further enhance operational efficiency.

Partition level monitoring

At the time of writing, The HMC supports monitoring and reporting capabilities to track energy usage, carbon emissions, and other environmental factors for existing Power servers. Power11 extends this function by delivering insights at the partition level.

At the time of writing, the HMC provides monitoring and reporting capabilities to track energy usage, carbon emissions, and other environmental factors for existing Power servers. Power11 extends these functions by delivering insights at the partition level, as shown in Figure 1-8 on page 29.

[8] Source: https://techchannel.com/power11/server-models/

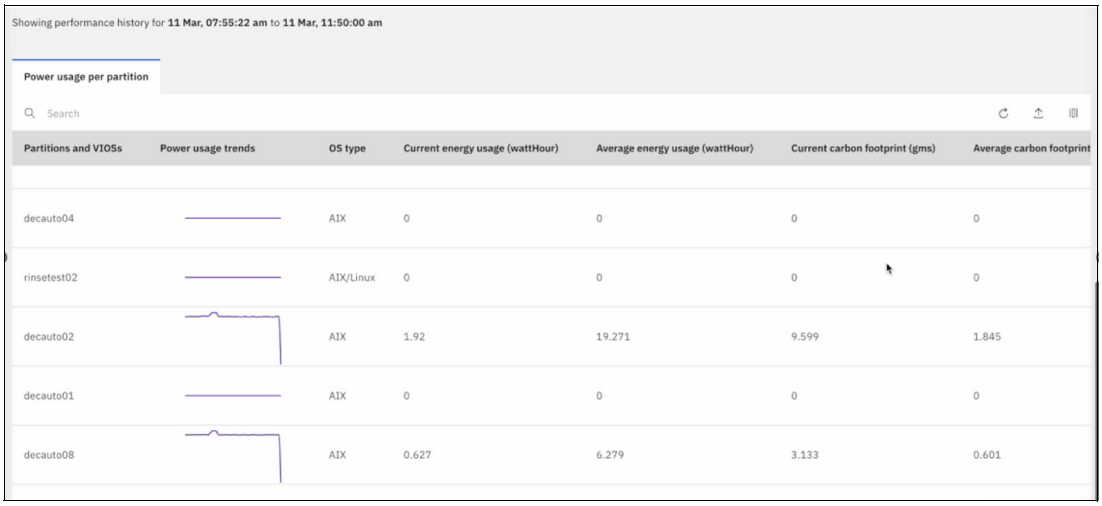

Figure 1-8 Partition level monitoring

Energy efficient mode

Power11 introduces a new energy efficient mode that, when enabled, can reduce energy consumption by up to 30%, with only an approximate 10% reduction in performance[9]. The impact varies depending on system configuration. You can dynamically configure the power mode through the HMC GUI, as illustrated in Figure 1-9.

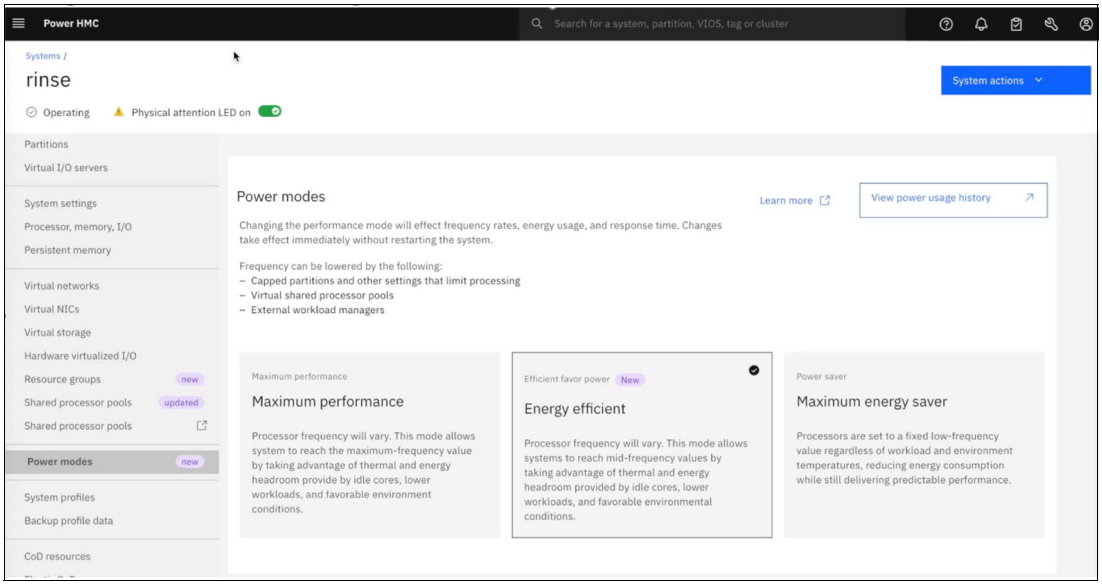

Figure 1-9 IBM Power modes

The environmental dashboard shows that in maximum performance mode, energy consumption remains almost constant even when core usage decreases.

[9] Based on IBM measurements on servers comparing Maximum Performance Mode to Energy Efficient mode while running compute-based, disk-based, and memory-based workloads.

Figure 1-10 illustrates energy usage when the processor operates in maximum performance mode.

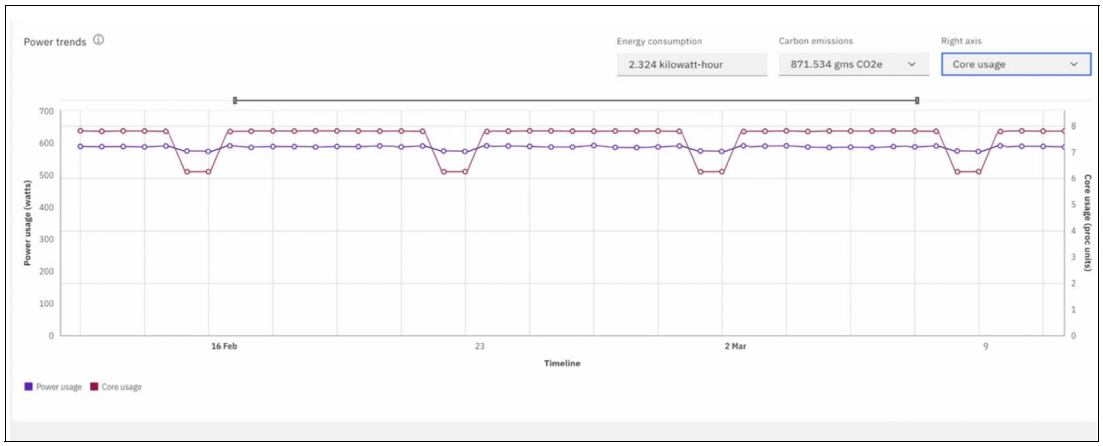

Figure 1-10 Energy consumption in max performance mode

Switching to Energy efficient mode, which is applied instantly, results in only a 10% reduction in performance while decreasing energy consumption by up to 30%, as shown in Figure 1-11.

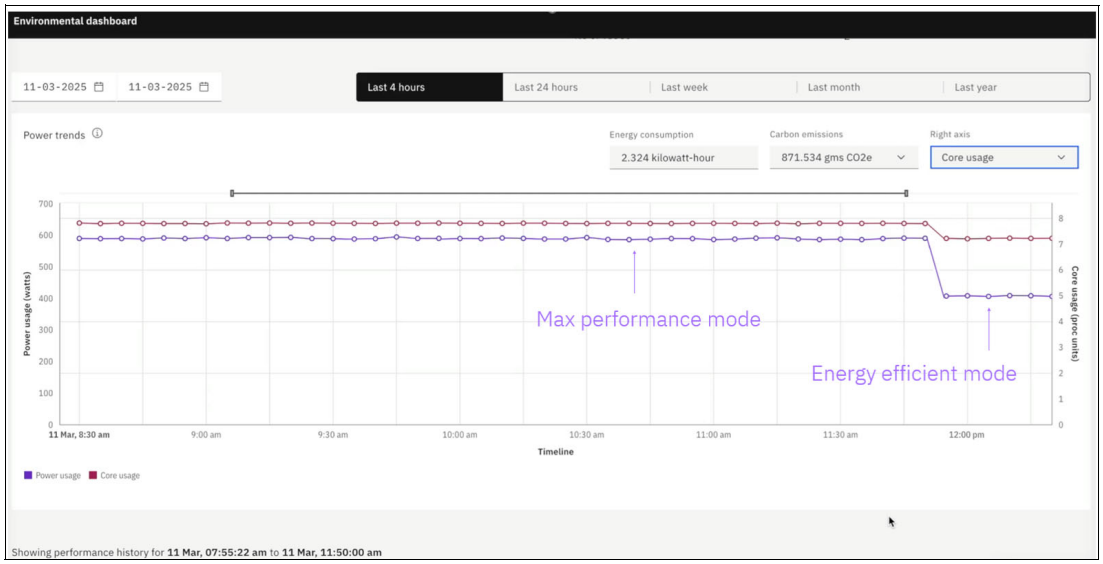

Figure 1-11 Energy efficient mode

IBM Power mode scheduling

If system usage follows a predictable pattern, you can use the new scheduling power mode feature to optimize energy efficiency. This feature enables the system to activate Energy efficient mode during designated time periods.

As illustrated in Figure 1-12 on page 31, Energy efficient mode is engaged over the weekend, starting at 8:00 PM on Friday and ending at 5:00 AM on Monday.

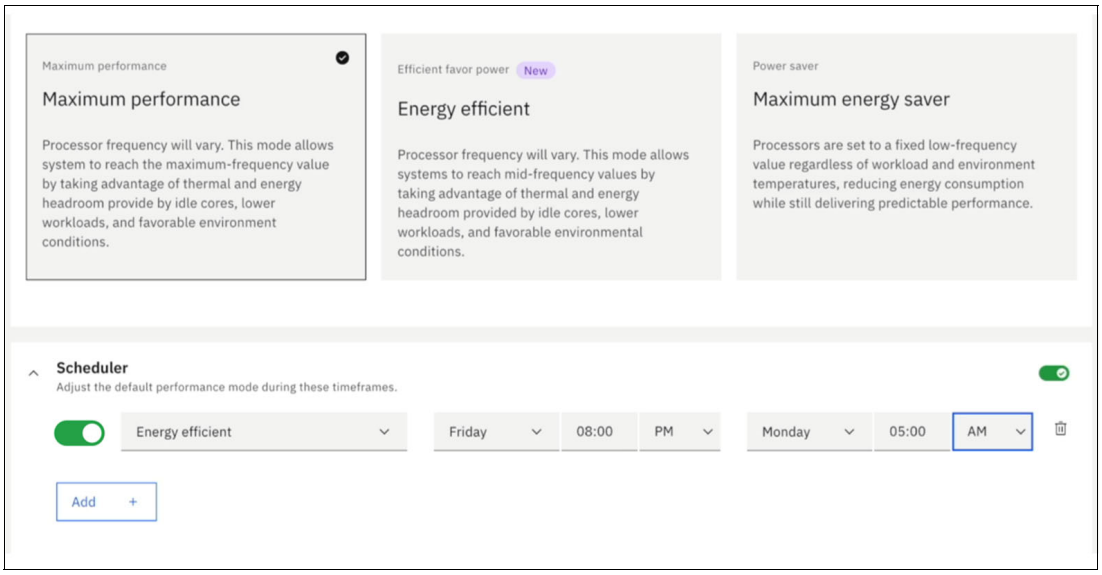

Figure 1-12 IBM Power mode scheduling

The new environmental dashboard enables energy savings whenever opportunities arise. Automated scheduling capabilities allow the system to migrate between power modes efficiently.

1.4.7 Consolidation of Oracle 19c workloads

At the time of writing, Oracle Database 19c is the latest long-term release, and Version 19.12 is the most recent Release Update for the AIX operating system. Oracle Database 19c introduces numerous enhancements over previous versions. Oracle Database versions for IBM AIX are compatible with later AIX Technology Levels (TLs) within the same major version. For certification details, see the “Certification” section at the My Oracle Support portal.

Starting with Oracle Database 11g Release 2 (11.2.0.4), Oracle Database Instant Client is supported on Linux on Power (32-bit and 64-bit). This support continues with Oracle Database 19c. Beginning with Oracle Database 12c Release 1, Little-Endian support was introduced for the Instant Client on Linux on Power. As of Oracle Database 12c Release 2, only Little-Endian is supported.

Oracle Database Instant Client is a lightweight version of the full Oracle Database Client binary files. It is available on Linux on Power, enabling applications that are deployed on Linux on Power to connect to Oracle Database instances regardless of the platform on which the database is deployed (AIX, IBM Power, or other platforms).

Although systems based on IBM Power8® processor technology reached end of support, many clients have not completed a TR. IBM demonstrated that migrating Oracle Database workloads from Power8 to Power11 can reduce processor core requirements by up to 50%[10]. This reduction directly lowers the number of required Oracle licenses, decreasing total cost of ownership (TCO).

[10] Source: Based on performance results published https://www.ibm.com/downloads/documents/us-en/10c31775c5d40fed

The IBM Power11 processor delivers substantial performance improvements, enabling up to a 25% increase in Oracle Database 19c workload consolidation compared to Power10[11]. These gains enable organizations to optimize infrastructure efficiency, reduce software licensing costs through improved core usage, and enhance overall system throughput while maintaining enterprise-grade RAS features.

Oracle is committed to release their latest products on IBM Power. For more information, see the Oracle statement about support for 23ai on IBM Power.

1.4.8 Day 1 availability of Power11 in IBM Power Virtual Server with hybrid billing

As hybrid cloud adoption reaches 82% among enterprises and the number of applications is projected to exceed 1 billion by 2028[12], organizations face increasing pressure to manage complexity, address skill shortages, and control infrastructure costs. IBM Power11, combined with PowerVS, offers a secure, scalable, and high-performance hybrid cloud platform for mission-critical workloads.

A key differentiator is the day-one availability of Power11 in both public and private PowerVS environments. This immediate access enables you to deploy virtual servers for development, testing, production, and disaster recovery (DR) without delay. With provisioning times under 10 minutes, you can accelerate time-to-value and reduce migration risk through a consistent architecture and user experience across on-premises and cloud environments.

Using Power11 in PowerVS also unlocks advanced enterprise-grade security features, which include quantum-safe firmware signing, encrypted LPM, and a reduced number of hypervisor vulnerabilities (CVEs) compared to x86 alternatives. These capabilities are essential for organizations operating in regulated industries or managing sensitive data.

In addition, PowerVS integrates natively with IBM watsonx services through IBM Satellite® Connector, enabling you to seamlessly infuse AI into your workloads and accelerate digital transformation initiatives.

PowerVS now offers three generations of IBM Power servers, providing flexibility for clients with diverse requirements:

- Power9
  - Soft landing for clients moving to cloud, with substantial support life remaining (through October 2028).
  - Lowest cost option with special pricing on existing capacity.
  - Clients can continue to provision and use Power9.
- Power10
  - Ideal for business-critical applications such as SAP RISE.
  - Broad capacity availability across data centers.
  - Excellent price and performance for AIX and IBM i workloads.
  - Clients can continue to provision and use Power10.

[11] The actual performance improvement depends on the specific configuration of the Power11 system. Not all configurations achieve a 25% improvement. See https://www.ibm.com/downloads/documents/us-en/10c31775c5d40fed.

[12] Source: https://gitnux.org/hybrid-cloud-statistics/

- Power11
  - Immediate access to Power11 without waiting for on-premises delivery.
  - Provision in less than 10 minutes.
  - Suitable for development, testing, and application qualification.
  - The first cloud environment to provide Power11 in both IBM Cloud and on-premises (Private).

From a business value perspective, PowerVS delivers up to 25% faster SAP workload migration compared to x86 platforms[13], which is driven by architectural consistency and superior compute performance. Flexible consumption models, shared processor pools (SPPs), and tiered storage options help reduce TCO. Automated DR, secure backup, and nondisruptive maintenance further enhance operational efficiency. The platform also supports regulatory compliance with standards such as SOC 2, HIPAA, and PCI-DSS.

Common use cases include:

- Application modernization through access to more than 250 IBM Cloud services
- Business resiliency with warm DR environments
- Data center optimization by consolidating or exiting physical infrastructure
- SAP transformation initiatives with unmatched availability and scalability for large-scale HANA deployments

PowerVS is available in 22 IBM Cloud data centers worldwide for public cloud use and can also be deployed in dedicated private cloud environments at client sites to meet stringent regulatory or operational requirements. Power11 will initially be available in four global data centers and will be fully enabled for automated DR operations. This broad availability enables organizations to adopt Power11 in alignment with their business goals and compliance mandates.

For more information about PowerVS, see 10.3, “IBM Power Virtual Server” on page 206.

1.4.9 Off-chip AI acceleration with the IBM Spyre Accelerator

The IBM Spyre Accelerator is a low-power, high-efficiency hardware module that is designed to complement Power11 processor-based systems for AI inference, analytics, and memory-intensive workloads. This feature is planned for support on the Power E1180 in the fourth quarter of 2025.

The card uses a system-on-chip architecture that is optimized for enterprise AI workloads. It includes 32 low-power AI cores and supports multi-precision operations for inference and training, including FP16/8 and INT8/4. The card is enabled for foundation models and is planned to support the Red Hat software stack (Open Compute Project Open Systems for AI). It will also support popular AI frameworks and libraries such as PyTorch and vLLM.

Using 5 nm technology, the IBM Spyre Accelerator delivers high performance with low power consumption, requiring only 75 W. These characteristics make it ideal for organizations seeking to accelerate AI workloads while maintaining energy efficiency.

[13] Source: https://www.ibm.com/new/product-blog/rise-with-sap-migration-ibm-power-virtual-server

Figure 1-13 shows the IBM Spyre Accelerator.

Figure 1-13 IBM Spyre Accelerator

Each IBM Spyre Accelerator operates at only 75 W, reducing the power footprint compared to traditional GPU-based accelerators, which often exceed 300 W per device. The IBM Spyre Accelerator is planned for support only inside the accelerator expansion drawer. Each expansion drawer supports up to eight IBM Spyre Accelerator cards.

Expansion drawer support on the Power E1180:

- At initial availability, a maximum of one accelerator expansion drawer is supported on a Power11 processor-based system.
- Support for more expansion drawers is planned for 2026.

Note: The IBM Spyre Accelerator requires an updated fan-out card in the expansion drawer.

Unlike GPUs, which typically rely on high-bandwidth memory that is integrated on the card, the IBM Spyre Accelerator is designed to access and use the system's main DDR5 memory modules. This design enables the use of terabytes of high-performance memory that is available within the Power11 processor-based system. It provides two primary advantages:

- Reduced latency data access through a shared memory model
- Elimination of memory duplication, avoiding the need to copy data between host and device memory, as is common with discrete GPUs

Using system RAM simplifies software development and memory management while contributing to a more energy-efficient and thermally balanced system architecture, which is an increasingly important factor in modern data centers and edge environments. This design makes the IBM Spyre Accelerator suitable for workloads that require frequent access to large memory spaces with minimal energy cost, such as real-time analytics, AI inferencing, and high-throughput data preprocessing.

Here are the minimum supported software stack levels for general availability (4Q 2025):

- FW1110.10
- HMC1111
- RHEL 9.6
- Spyre software stack container
- Red Hat OpenShift AI (technology preview in 4Q 2025, GA in 1Q 2026)

For more information about use cases and solutions with IBM Spyre Accelerator, see Chapter 4, “Artificial intelligence support” on page 99.

1.4.10 Red Hat OpenShift with support for an expanded software ecosystem

Red Hat OpenShift on IBM Power provides a robust, enterprise-grade container platform that enables you to modernize applications, deploy AI workloads, and integrate with a broad software ecosystem. Running Red Hat OpenShift on Power11 systems enables organizations to use the platform’s performance, scalability, and security while maintaining architectural consistency across hybrid cloud environments.

A key advantage of Red Hat OpenShift on Power is its support for an expanded software ecosystem, which includes IBM Cloud Pak® offerings, such as IBM Cloud Pak for Data and IBM Cloud Pak for Integration; Red Hat Runtimes like Quarkus and Spring Boot; and a wide range of open-source AI and ML frameworks, including PyTorch, TensorFlow, and ONNX. Many independent software vendor (ISV) applications are now certified for the Power Architecture, enabling financial institutions and other enterprises to run containerized workloads alongside traditional AIX or IBM i systems. This colocation reduces latency, improves data gravity, and simplifies integration.

Red Hat OpenShift on IBM Power also supports seamless hybrid cloud integration, particularly when deployed on IBM PowerVS. This integration provides access to IBM Cloud services such as IBM Cloud Object Storage, IBM Key Protect, IBM Event Streams (Kafka), and IBM API Connect®. These capabilities are essential for hybrid architectures where AI models may be trained on-premises and deployed in the cloud or vice versa, depending on data locality and compliance requirements.

For developers and IT architects, Red Hat OpenShift on Power offers a rich set of tools and resources. The Red Hat OpenShift Container Catalog provides pre-built images and Helm charts that are optimized for Power, while IaC templates that use Terraform simplify cluster provisioning on PowerVS. IBM also supports the ecosystem with community channels, technical workshops, and IBM Garage® engagements to accelerate adoption.

By combining the strengths of IBM Power hardware with the flexibility of Red Hat OpenShift and the breadth of the container ecosystem, organizations can build and scale AI-driven applications with confidence, agility, and enterprise-grade resilience.

1.5 Power11 processor architecture

The Power11 processor is designed to deliver higher clock speeds and up to 25% more cores per processor chip compared to IBM Power10 systems. Power11 builds on the key capabilities that were introduced with Power10, including stronger RAS features, improved energy efficiency, enhanced energy management, and advanced quantum-safe security.

Beyond the processor itself, here are more improvements to packaging, the memory architecture, and AI acceleration:

- Packaging innovation

  The Power11 processor uses new ISC technology and advanced 2.5D packaging, which are combined with cooling innovations such as improved heat sinks and more efficient fans. These enhancements optimize energy delivery, strengthen thread and core performance, and increase overall system capacity.

- Enhanced system architecture

  Power11 systems feature a robust memory architecture that is based on DDR5 DDIMMs and enhanced OMI technology. These improvements deliver higher memory reliability, capacity, and bandwidth. Because OMI is technology-neutral, Power11 systems also support OMI DDR4 memory that is migrated from Power10 high-end systems, protecting client investments in memory technology.

- AI acceleration

  Power11 continues to support a wide range of enterprise AI use cases through the MMA architecture. Improvements in processor core strength and system capacity enhance MMA performance for inference workloads. In addition, IBM plans to incorporate the IBM Spyre Accelerator into Power11 offerings to provide more AI inferencing capabilities. Together, IBM POWER processors and the IBM Spyre Accelerator enable next-generation infrastructure to scale demanding AI workloads for business.

For more information about AI and IBM Spyre Accelerator, see Chapter 4, “Artificial intelligence support” on page 99.

1.5.1 New Power11 processor details

Here are some details on the new Power11 processor design:

- Technology and packaging
  - 654 mm² 7 nm refined Samsung process (30 billion devices).
  - Advanced 2.5D packaging technology optimizes space and improves signal integrity.
  - ISC technology enhances power delivery and performance by integrating capacitors directly into the chip package.
  - Supports single-chip or dual-chip sockets.
- Computational capabilities
  - Up to 16 SMT8 cores (2 MB L2 cache per core).
  - Up to 128 MB L3 cache with low-latency non-uniform cache architecture management.
  - Enterprise performance focus
    - 3× core performance compared to Power9.
    - 2× thread strength compared to Power9.
    - 4× L2 cache and 4× MMU per core compared to Power9.
    - 4× cryptographic engine per core compared to Power9.
  - AI computational focus
    - 2× general SIMD per core compared to Power9.
    - 4× matrix SIMD per core compared to Power9.
    - New AI instructions and data types.

- Robust data plane
  - 2 TBps raw bandwidth (32 GTps) with PowerAXON and OMI signaling.
  - SMP interconnect supports up to 16 sockets.
  - 2× OMI memory bandwidth compared to Power10.
  - 64 TB OMI DDR large system memory capacity.

Figure 1-14 shows the Power11 processor.

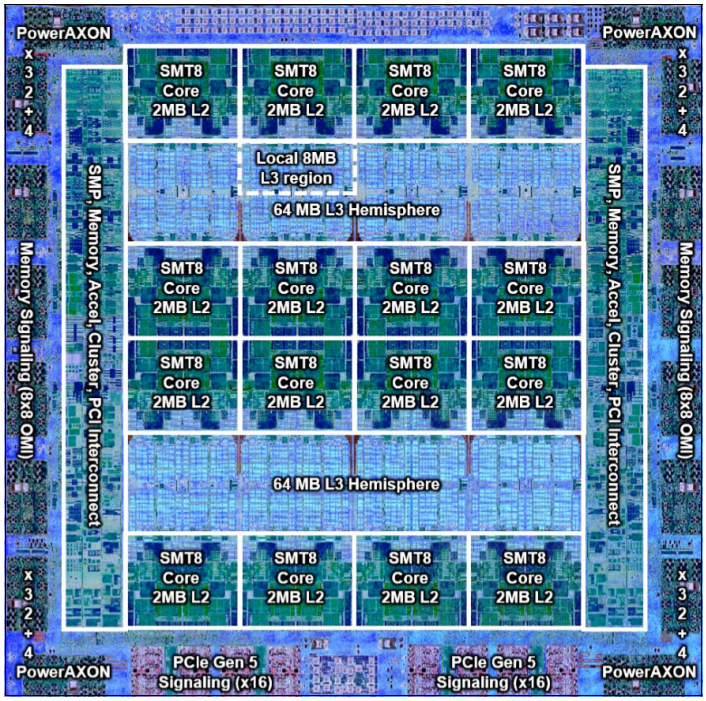

Figure 1-14 Power11 processor diagram

Each of the cores has a 64 KB data cache + 96 KB instruction cache for the L1 cache.

Comparing Power11, Power10, and Power9 processors

Table 1-5 shows a comparison of Power11 to the previous generations.

Table 1-5 Comparing the Power11 processor technology to previous processor generations

Characteristics	Power11	Power10	Power9
Technology	7 nm	7 nm	14 nm
Die size	654 mm^2	602 mm^2	693 mm^2
Processor module size	68.5 mm x 77.5 mm	68.5 mm x 77.5 mm	68.5 mm x 68.5 mm
Number of transistors	30 billion[a]	18 billion	8 billion
Maximum cores	16	15	12
Maximum hardware threads per core	8	8	8

Characteristics	Power11	Power10	Power9
Maximum static frequency and high-performance frequency range	3.8 - 4.4 GHz	3.75 - 4.15 GHZ	3.9 - 4.0 GHz
L2 cache per core	2048 KB	2048 KB	512 KB
L3 cache	8 MB of L3 cache per core with each core having access to the full 128 MB of L3 cache, on-chip high-efficiency SRAMa	8 MB of L3 cache per core with each core having access to the full 120 MB of L3 cache, on-chip high-efficiency SRAM	10 MB of L3 cache per core with each core having access to the full 120 MB of L3 cache, on-chip eDRAM
Memory technology	DDR5 and DDR4[b]	DDR4 and DDR5[c]	DDR4 and DDR3
I/O bus	PCIe Gen 5	PCIe Gen 5	PCIe Gen 4

a. IBM Power recently changed its methodology for counting transistors to be more in line with the industry.
b. DDR4 only as a migration from Power10 during the same serial number upgrade.
c. DDR5 support was added after GA.

Chip packaging options

To provide flexibility and scalability, the Power11 chip is packaged in several ways. For the high-end Power E1180, IBM uses an SCM, which provides one chip per socket and supports up to 16 sockets per system. For the midrange Power E1150, the Power11 chip is packaged with two chips per socket in a Dual-Chip Module (DCM) and supports up to four sockets per system. The DCM version is also used in the scale-out system line, including the Power S1122 and Power L1122, and Power S1124 and Power L1124, with up to two sockets per system. A third packaging option, the eSCM, is used in the entry-level processor features of the Power S1122 (the 4-core and 10-core options). The eSCM is a modified DCM in which the second chip in the module does not provide processor capability and only adds I/O functions.

Single-Chip Module

The Power E1180 system uses the Power11 SCM to deliver high performance, scalability, and energy efficiency for enterprise workloads. The SCM design integrates up to 16 cores per chip, each supporting simultaneous multithreading (SMT), which enables up to 128 threads per socket. This compact architecture enables the Power E1180 to scale up to 256 cores in a full system configuration, making it suitable for data-intensive applications such as AI, analytics, and large-scale transactional processing. The SCM design emphasizes high throughput and low latency, helping ensure that the system handles demanding workloads with consistent performance.

Figure 1-15 on page 39 shows the layout of the SCM that is used in the Power E1180.

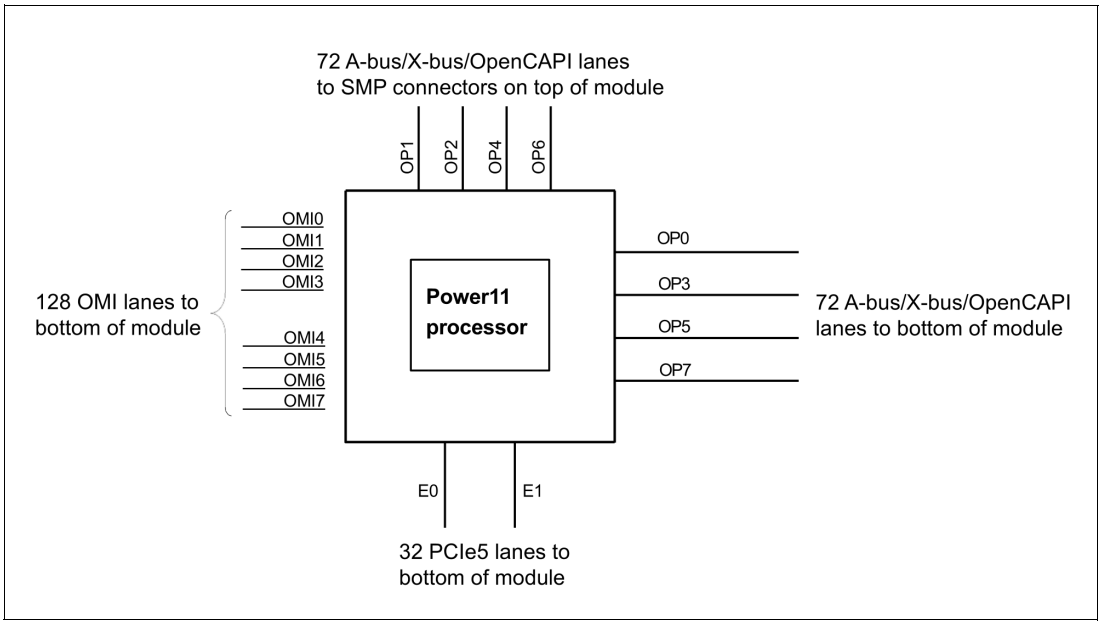

Figure 1-15 Single-Chip Module diagram

1.5.2 Spare core support

In previous generations of IBM POWER processors, IBM provided a proactive fault-tolerance mechanism to handle certain predictive core errors. When a suspect core was identified, PowerVM could automatically unconfigure it and, if spare capacity was available (either unused or unlicensed), it substituted these resources to maintain system processing capacity. This seamless failover occurred without requiring a system or application restart, helping ensure continued operation. This feature remains available in the Power11 processor.

Because the substituted resources were not guaranteed to be on the processor chip with the failing core, the spare might not share the same affinity as the original core. In addition, because the spare core was not a dedicated resource, there was a possibility that the core might eventually be needed or licensed. Therefore, the processor containing the suspect core was flagged for repair. As a result, a planned system outage was typically scheduled to replace the affected processor module and restore full system integrity.

Spare cores can also be used for processor core failures other than predictive errors. In these cases, the processor module is called out for repair. However, if the spare is successfully deployed, system owners might decide not to take the system down to repair the processor module because full processor capacity and performance are maintained for these events.

With Power11, spare cores (one spare in each SCM, or four spares per E1180 node, and two spares in each DCM) are allocated on each processor module. The exceptions are the 16-core SCM, which has all cores that are allocated for client usage, and the eSCM modules, where no spares are allocated. These cores are used only to substitute for a failing core on the same processor module, which resolves affinity issues and helps ensure that the spare core is not otherwise used. As a result, predictive processor core failures in the Power11 do not require system downtime to repair the processor module.

1.6 Operating system support

The Power E1180 system supports the following OSs:

- AIX
- IBM i
- Linux
- Red Hat OpenShift

In addition, a VIOS can be installed in special partitions that provide virtualization of I/O capabilities, such as network and storage connectivity. Multiple VIOS partitions can be installed to provide support and services to other partitions running AIX, IBM i, or Linux, such as virtualized devices and LPM capabilities.

For more information about the operating systems and other software that are available on IBM Power, see IBM Power. The minimum support levels of IBM AIX, IBM i, and Linux are described in the following sections.

IBM announces support for newer versions of the operating systems over time. Keep your operating systems up to date to enable new security capabilities and support new features.

1.6.1 AIX operating system

The Power E1180 systems support the following minimum levels of the AIX OS:

- When using virtual I/O

  The Power E1180 systems support the following minimum levels of AIX when installed with virtual I/O:

  - AIX 7.3 with Technology Level (TL) 7300-02 and Service Pack (SP) 7300-02-02-2420
  - AIX 7.3 with TL 7300-03 and SP 7300-03-00-2446
  - AIX 7.2 with TL 7200-05 and SP 7200-05-08-2420

- When using direct I/O connectivity

  The Power E1180 systems support the following minimum levels of AIX when installed by using direct I/O connectivity:

  - AIX 7.3 with TL 7300-02 and SP 7300-02-04-2520
  - AIX 7.3 with TL 7300-03 and SP 7300-03-01-2520
  - AIX 7.2 with TL 7200-05 and SP 7200-05-10-2520

Important:

- AIX 7.2 instances can use physical and virtual I/O adapters but must run in an LPAR in IBM Power9 compatibility mode.
- Starting with AIX 7.3 TL 3 SP 1, AIX 7.3 instances can use physical and virtual I/O adapters and can run in an LPAR in native Power11 mode.

For more information about AIX support, see 6.2, “AIX” on page 129.

1.6.2 IBM i operating system

The Power E1180 supports the following minimum levels of the IBM i OS both native and as a client of VIOS:

- IBM i 7.4 TR12
- IBM i 7.5 TR6
- IBM i 7.6 with the Power11 GA PTF Group

For more information about IBM i support, see 6.3, “IBM i” on page 134.

1.6.3 Linux operating system

The Linux distributions that are listed here are fully supported on the Power E1180. Other distributions, including open source releases, can run on these systems, but do not include any formal Enterprise Grade support.

Red Hat Enterprise Linux

The latest version of the RHEL distribution from Red Hat is supported in native Power11 mode, which enables it to access all features of the Power11 processor and platform.

At announcement, the Power E1180 systems support the following minimum levels of the RHEL OS:

- RHEL 8.10 for IBM Power LE or later (Power10 compatibility)
- RHEL 9.4 for IBM Power LE or later (Power10 compatibility)
- RHEL 9.6 for IBM Power LE or later (Power11 native)
- RHEL 10 for IBM Power LE 10.0, or later
- RHEL for SAP with RHEL 9 for IBM Power LE 9.6 or later

Red Hat OpenShift Container Platform

The Red Hat OpenShift Container Platform is supported on IBM Power servers, including the Power E1180. To run Red Hat OpenShift Container Platform in native mode on Power11, you must use Red Hat OpenShift Container Platform 4.19, which is based on RHEL 9.6. Previous versions run in Power10 compatibility mode.

The Power S1122 and Power S1124 systems support the following minimum levels of the OSs that are supported for Red Hat OpenShift Container Platform:

- Red Hat Enterprise Linux CoreOS (RHCOS) 4.18 (Power10 compatibility)
- RHCOS 4.19 (Power11 native)

SUSE Linux Enterprise Server

The latest version of the SUSE Linux Enterprise Server distribution is supported in native Power11 mode, which enables it to access all features of the Power11 processor and platform.

At announcement, the Power E1180 systems support the following minimum levels of the SUSE Linux Enterprise Server OS:

- SUSE Linux Enterprise Server for SAP with SUSE Linux Enterprise Server 15 SP 6 or later
- SUSE Linux Enterprise Server 15 SP 6 or later

For more information about Linux support, see 6.4, “Linux on IBM Power” on page 141.

1.6.4 VIOS

The minimum required level of VIOS for the Power E1180 is VIOS 4.1.1.10 when running in Power11 mode. The following levels are supported:

- VIOS 4.1.1.10 (SP)
- VIOS 4.1.0.40 (SP) (P10 compatibility mode)
- VIOS 3.1.4.60 (SP) (P9 compatibility mode)

1.7 The IBM Power firmware and Hardware Management Console

The firmware on an IBM Power server is the foundational software layer that initializes hardware components, manages system resources, and provides the interface between the hardware and the OS.

The IBM Power firmware includes components such as the Flexible Service Processor (FSP), PowerVM Hypervisor (PHYP), and system boot code. This firmware enables advanced features such as logical partitioning, dynamic resource allocation, and Secure Boot processes. It also supports system diagnostics, error reporting, and remote management capabilities, especially when integrated with tools such as the HMC. Regular firmware updates from IBM provide performance improvements, security fixes, and compatibility with new hardware or software features.

An HMC is a dedicated appliance that configures and manages system resources on IBM Power servers. It can be delivered as a physical system or as a virtual appliance running in either a PowerVM environment or an x86 VM. A GUI, a CLI, or Representational State Transfer (REST) application programming interfaces (APIs) are available.

1.7.1 HMC requirements

Supported HMCs are the 7063-CR2 and the Virtual HMC appliances.

The minimum required HMC version for the Power E1180 is V11R1 M1110. V11R1 M1110 is supported on the 7063-CR2, and the Virtual HMC appliances only. HMCs running Version 11 can manage Power11, Power10, and Power9 processor-based systems.

Restriction:

- The 7063-CR1 and 7042 HMCs are not supported as an HMC for Power11 systems.
- The HMC code at V11R1M1110 cannot manage Power8 or earlier processor-based systems.

1.8 Power E1180 rack integration and handling guidelines

The Power E1180 system is designed to fit seamlessly into a standard 19-inch rack and is fully tested and certified in the IBM 42U Enterprise Rack (model 7965-S42). Although clients can install the system in third-party racks, those racks must meet IBM requirements for strength, rigidity, depth, and hole pattern compatibility. As a best practice, consult IBM Services® to validate the suitability of any third-party rack.

For optimal installation, performance, and serviceability, order the Power E1180 with the IBM 42U Enterprise Rack. Initial system configurations are included in the 7965-S42 rack to streamline deployment, help ensure high-quality assembly and testing in IBM Manufacturing, and provide a complete, secure shipping solution. Clients who prefer to receive the system without a rack can select the factory deracking feature (Feature Code ER21). In this case, the system is tested in the rack and then shipped separately.

The 7965-S42 rack supports several front door options to meet different operational and aesthetic needs:

- Feature Code ECRA: Acoustic front door for noise reduction
- Feature Code ECRT: High-end appearance front door
- Feature Code ECRM: Cost-effective plain front door

Because each system node drawer weighs up to 86.2 kg (190 lb), multiple service personnel are required for manual installation or removal. To simplify this process, IBM offers a lightweight, cost-effective lift tool (Feature Code EB3Z) with a hand crank capable of lifting systems up to 400 lb, and an optional angled shelf kit (Feature Code EB4Z) for precise positioning during installation.

Important: The customer is responsible for ensuring that the installation of the drawer in the preferred rack or cabinet results in a configuration that is stable, serviceable, safe, and compatible with the drawer requirements for power, cooling, cable management, weight, and rail security.

Consider the following points when ordering racks:

- The IBM Enterprise 42U Slim Rack (7965-S42) offers 42 Electronic Industries Alliance (EIA) units (U) of space in a slim footprint.
- The 7014-T42 rack is no longer available for purchase with a Power E1180 system. Installing a Power E1180 system in this rack is still supported.

IBM Power distribution units

The Power E1180 system, when integrated into an IBM rack, uses horizontal power distribution units (PDUs) in the EIA drawer space of the rack instead of the typical vertical PDUs found in the side pockets of a rack. This design aids cable routing. Each horizontal PDU occupies 1U. Vertically mounting the PDUs to save rack space can create cable routing challenges and interfere with optimal service access.

When mounting horizontal PDUs, place them near the top or bottom of the rack, leaving 2U or more of space at the top or bottom for cable management. Mounting a horizontal PDU in the middle of the rack is not optimal for cable management.

Two PDU ratings are supported: 60 A (orderable in most countries) and 30 A.

- The 60 A PDU supports four system node power supplies and one I/O expansion drawer or eight I/O expansion drawers.
- The 30 A PDU supports two system node power supplies and one I/O expansion drawer or four I/O expansion drawers.

PCIe I/O expansion drawer and racks

IBM Manufacturing can factory-integrate the PCIe I/O expansion drawer (Feature Code ENZ0) with new system orders. Because expansion drawers can obstruct access to vertical PDUs if they are at the same height, install PDUs horizontally on racks that include one or more PCIe I/O expansion drawers. IBM Manufacturing assembles integrated rack configurations with horizontally mounted PDUs unless the CRSP feature (Feature Code 8453) is included in the order.

When specifying CRSP, you must provide the locations where the PCIe I/O expansion drawers should be placed and avoid positioning them next to vertical PDU locations EIA 6 - 16 and 21 - 31.

2

IBM Power E1180

This chapter focuses on the IBM Power E1180. The Power E1180, a follow-on to the IBM Power E980 and IBM Power E1080, is a scalable system for consolidating database and mission-critical workloads. The Power E1180 provides up to 16 sockets and up to 256 SMT8 cores with up to 64 TB of DDR5 memory. These capabilities enable the system to handle massive data volumes and complex applications efficiently. One of its standout features is Transparent Memory Encryption (TME), which secures data in memory without compromising performance. This feature is ideal for hybrid cloud environments. Also, the built-in artificial intelligence (AI) inference engine enables organizations to process and analyze data directly at the source, reducing latency and infrastructure complexity.

Beyond raw performance, the Power E1180 is engineered for resilience and efficiency. It offers twice the memory reliability of standard DIMMs and includes advanced diagnostic and recovery capabilities to minimize downtime. The system's energy-efficient 7 nm architecture reduces power consumption while delivering higher throughput per core. These features make it suitable for mission-critical workloads such as SAP HANA, Oracle databases, and healthcare applications that demand maximum uptime and security. With dynamic scaling and reduced software licensing costs through workload consolidation, the Power E1180 helps businesses optimize IT investments while maintaining agility.

The following topics are covered in this chapter:

- Power E1180 overview
- Power E1180 components
- Processors
- Memory
- PCIe slots
- Storage and USB features
- I/O expansion drawers
- Environmental requirements

© Copyright IBM Corp. 2025.

2.1 Power E1180 overview

The Power E1180 sets a new benchmark in enterprise computing through a full-stack innovation approach. Designed to deliver unmatched performance, security, and availability, it enables organizations to run mission-critical workloads with zero planned downtime (ZPD) for maintenance and industry-leading resilience.

Key capabilities include:

- Autonomous system maintenance with ZPD
- Unmatched resilience with 99.9999% uptime
- Spare core architecture: Up to four spare cores per drawer for enhanced availability
- IBM Power Cyber Vault: Sub-minute ransomware detection and automated recovery with IBM Storage
- Quantum-safe encryption (QSE): Secure Boot and Live Partition Mobility (LPM)
- Automated cryptographic inventory through IBM PowerSC
- Support for IBM 4770 Crypto Card, FIPS 140-3 Level 4 certified
- Main memory encryption for enhanced data protection
- Automated data collection for accelerated error resolution
- Energy-efficiency mode with advanced reporting and scheduling
- Flexible consumption models: Capacity on Demand and IBM Power Enterprise Pools 2.0 (PEP 2.0)
- Modular, scalable design: Up to four 5U CPC drawers plus one 2U System Control Unit
- High core density: Up to 256 Power11 SMT8 cores (10-, 12-, or 16-core Single-Chip Module (SCM) options)
- Massive memory capacity: Up to 64 TB DDR5 (16 TB per drawer)
- High-speed I/O: Eight Peripheral Component Interconnect Express (PCIe) Gen5 blind-swap slots per drawer
- Concurrent repair: Non-active SMP cables with improved fault isolation
- Integrated Non-Volatile Memory Express (NVMe) storage: Four 7 mm or two 15 mm U.2 bays per drawer
- Scalable expansion: Up to 16 I/O and NVMe drawers (four per CPC drawer)
- Seamless upgrades: Serial number-preserving upgrades from Power10 (DDR4 and DDR5 supported)

The Power E1180 is more than a system: It is a strategic platform for digital transformation. Whether deployed on-premises or in the cloud, it provides a secure, scalable, and AI-ready foundation for modern workloads. With its unmatched reliability and forward-looking architecture, it enables enterprises to reduce risk, accelerate innovation, and support the next generation of applications.

Figure 2-1 on page 47 shows the Power E1180 system unit from the front.

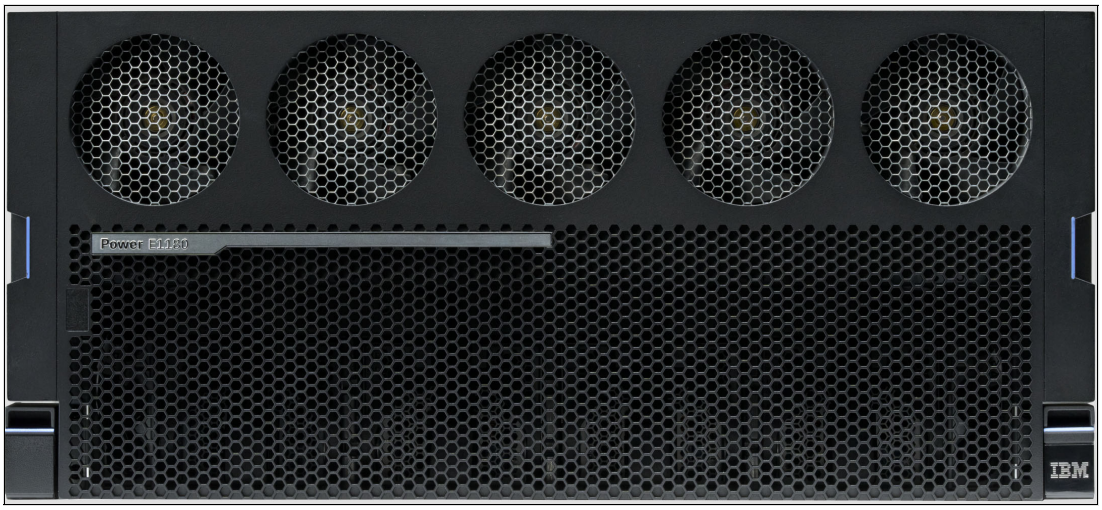

Figure 2-1 Power E1180 system unit

Table 2-1 lists the machine type and model for the Power11 E1180 system.

Table 2-1 Machine type and model of Power E1180 system

System name	Machine type and model
Power E1180	9080- HEU

2.1.1 IBM Power11: Engineered for unmatched uptime

The Power E1180 is built to support maximum reliability, availability, and serviceability (RAS). Reliability is achieved by minimizing potential errors and outages through error detection and self-healing. Availability is achieved by reducing unplanned downtime through self-diagnosing capabilities and by predicting and fixing errors to avoid outages. Serviceability provides accurate diagnostic capability by using the Call Home function and concurrent repair features that are built into the system.

Reliability

The Power11 is built to minimize potential errors and maximize system integrity through advanced design and manufacturing hardening.

- Error detection and self-healing: Proactively identifies and corrects faults before they impact operations.
- Spare core technology: Each system drawer includes up to four spare cores, enhancing fault tolerance and maintaining performance.
- Industry-leading uptime: Designed to be the most reliable system in IBM Power history, delivering 99.9999% availability.

Availability

The Power11 provides continuous operations even if there are hardware or software issues:

- Predictive Failure Analysis (PFA): Detects early signs of failure and takes corrective action to avoid outages.
- Self-diagnosing capabilities: Enables the system to identify and isolate issues without human intervention.
- Keep-running architecture: Maintains workload continuity despite underlying problems.

Serviceability

The IBM Power11 simplifies maintenance and reduces downtime through intelligent service features:

- Concurrent and guided repair: Enables components to be serviced or replaced without shutting down the system.
- Accurate problem diagnostics: Speeds issue resolution with precise fault identification.
- End-to-end service automation: From problem detection to Call Home and service closure, the system streamlines the entire support process.

The Power11 is designed to be the most reliable system in the history of the IBM Power platform, delivering 99.9999% uptime. The Power E1180 includes dual power cord redundancy, n+1 power supply redundancy, and n+1 fan rotor redundancy. Power supplies and fans are concurrently maintainable. The system service processor is redundant and supports dynamic failover. Each system node includes dual processor clock logic with dynamic failover capability. The system also supports concurrent maintenance of the real-time clock battery and the operator panel.

Table 2-2 provides an overview of the Power E1180 RAS characteristics.

Table 2-2 RAS features on IBM Power E1180

Feature name	Power E1180 support
Redundant / Hot Swap Power Supplies	Yes
Pluggable VRMs with redundant/spare voltage phases	N+2
OCC error handling with power safe mode	Yes
EPOW reporting	Yes
TOD Battery Concurrent Maintenance	Yes
Redundant Fans/Blowers	Yes
Hot Swap Fans/Blowers	Yes
NVMe drives or Media Concurrent Maintenance	Yes
PCI Adapter Concurrent Maintenance	Yes
Enhanced Error Handling (EEH) support	Yes
Concurrent Op Panel Repair	Yes
Redundant Service Processors	Yes
Dynamic Service Processor Failover	Yes
Auto Service Proc. Reset/Reload	Yes

Feature name	Power E1180 support
Redundant System Clocks	Yes
Dynamic System Clock Failover	Yes
Redundant VPD	Yes
Concurrent Firmware Update	Yes
Memory Preserving Initial Program Load (IPL)	Yes
Hardware and OS Dump	Yes
Processor Instruction Retry	Yes
Spare core	2
Predictive De-allocation (such as core)	Yes
Core checkstop	Yes
Dynamic Core Init. Changes	Yes
L2/L3 Line Delete	Yes
Fabric Bus CRC/Retry	Yes
SMP Cable Fault Isolation	Yes
Ext SMP cable concurrent maintenance	Yes
Memory ECC/Chip Kill Handling	Yes
Spare DRAM	2 spare DRAMs per rank
Integrated N+1 Power regulator	Yes
Memory Scrubbing	Yes
Processor to Memory Buffer Recovery	Yes
Memory (DRAM) Row Repair	Dynamic (detect and deploy at run time)
Predictive Dynamic Memory De-allocation	Yes
Active Memory Mirroring (AMM) for Hypervisor	Yes

2.2 Power E1180 components

The Power E1180 system (9080-HEU) is the most powerful and scalable system in the IBM Power portfolio. It is built from modular components, primarily central processing complex (CPC) enclosures known as *system nodes*, along with more units and expansion drawers.

2.2.1 System nodes

A system node houses the connections and electronics that are required to link the processors with memory, internal storage, adapters, and system interconnects. The Power E1180 supports configurations of 1 - 4 system nodes per system.

Each 5 Electronic Industries Alliance (EIA) unit, or 5U system node, contains four air-cooled SCMs that are optimized for performance and scalability. The Power E1180 SCMs can include 10-, 12-, or 16-core Power11 processors running at up to 4.40 GHz, with simultaneous multithreading (SMT) that runs up to eight threads (SMT8) per core. Each SCM has dual memory controllers that deliver up to 512 GBps of peak memory bandwidth per socket, or 2,048 GBps per node. Using PCIe Gen5 I/O controllers, which are also integrated onto each SCM to reduce latency, up to 576 GBps of peak I/O bandwidth is available per node. This design provides maximum processor performance, enabling applications to run faster and respond more efficiently.

Each system node contains:

- Four sockets for Power11 SCMs: Each socket supports 10-, 12-, or 16-core Power11 processors and includes integrated logic for power and data connectivity.
- 64 DDIMM slots for DDR5 memory: Supports DDR4 DIMMs when upgrading from an existing Power10 system.
- Up to 16 TB of system memory per drawer: Using the largest available DDIMMs, a fully configured four-node system can scale up to 64 TB of RAM.
- Four NVMe drive bays for internal boot storage: Extra capacity is available through expansion drawers.
- Eight PCIe Gen5 slots per node: Provides a maximum of 32 PCIe slots across a four-node configuration.

The processor configuration is determined by the selected processor feature, with each feature defining a set of four Power11 SCMs with the same core count (10, 12, or 16). All system nodes in a Power E1180 system must use the same processor feature to help ensure uniformity.

A single SCU is required per Power E1180 system. The SCU provides centralized system management through redundant Flexible Service Processors (FSPs). System management and configuration, including service actions, are performed through the Hardware Management Console (HMC).

2.2.2 System Control Unit

The SCU is implemented in a 2U-high chassis and provides system hardware, firmware, and virtualization control functions through a pair of redundant FSP devices. It also contains the operator panel and the electronics module that stores the system Vital Product Data (VPD). The SCU also supports USB connectivity for use by the Power E1180 system.

Figure 2-2 on page 51 shows the front and rear view of a SCU with the locations of the external ports and features.

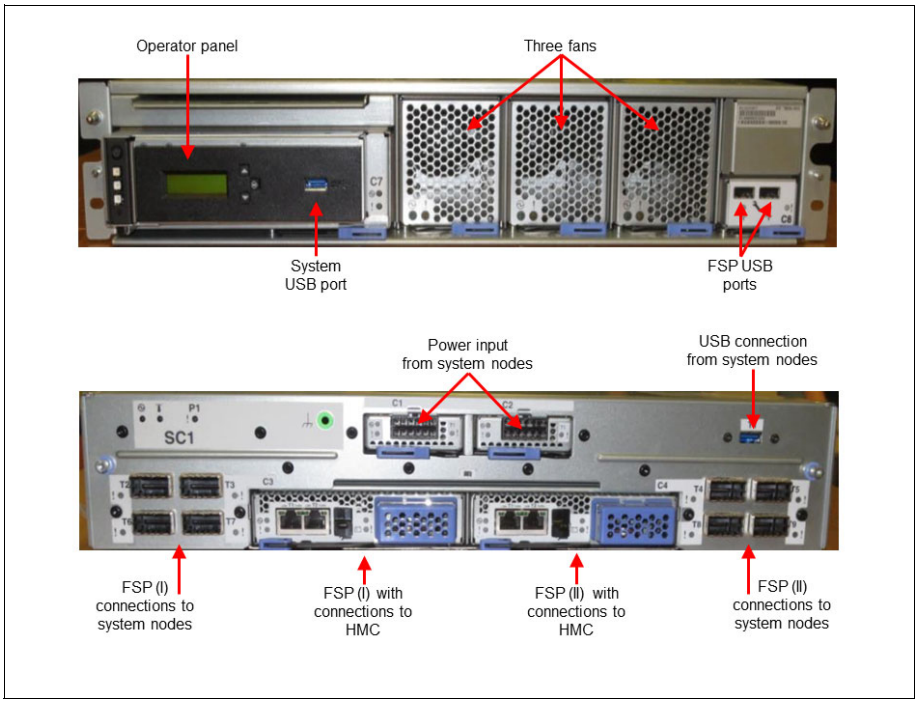

Figure 2-2 Front and rear view of the System Control Unit

One SCU is required for each Power E1180 system, regardless of the number of system nodes. The SCU is powered according to the following rules:

- Two universal power interconnect (UPIC) cables are used to provide redundant power to the SCU.
- Single-node configuration: Both UPIC cables are connected from the single system node to the SCU.
- Two-, three-, or four-node configuration: One UPIC cable is connected from the first system node, and the second UPIC cable is connected from the second system node to the SCU.

The set of two cables provides a 1+1 redundant power supply. If one cable fails, the remaining UPIC cable is sufficient to supply power to the SCU.

Two FSP cards are installed in the SCU and are ordered by using two mandatory Feature Code EDFB features. Each processor card provides two 1 Gb Ethernet ports for connection to the HMC. One port is used as the primary connection, and the second port can be used for redundancy. To enhance resiliency, it is a best practice to implement a dual-HMC configuration by attaching separate HMCs to each card in the SCU.

Each FSP provides four ports for redundant connections from the SCU to each system node. System nodes connect to the SCU by using cable features Feature Code EFCH, Feature Code EFCE, Feature Code EFCF, and Feature Code EFCG. Feature Code EFCH connects the first system node to the SCU and is included by default in every system node configuration. It provides FSP, UPIC, and USB cables, but no SMP cables, which are used to interconnect the processor cards in each node. All other cable features are added based on the number of extra system nodes and include FSP and SMP cables.

The SCU implementation also includes the following features:

- Elimination of clock cabling: Introduced with Power9 processor-based systems.
- Front-accessible USB port.
- Optimized UPIC power cabling.
- Optional external DVD drive.
- Concurrently maintainable time-of-day clock battery.

2.3 Processors

Each system node in an IBM Power E1180 system provides four sockets to accommodate Power11 SCMs. Processor features are available in 10-, 12-, and 16-core configurations.

Table 2-3 lists the available processor Feature Codes for a Power E1180 system.

Table 2-3 Processor features

Feature Code	Processor cores	Max system config	Frequency range (GHz)	Recommended usage
EDQD	16 cores	256 cores	3.80 - 4.3	Max throughput and density
EDQ9	12 cores	192 cores	3.90 - 4.4	Max core/thread performance
EDQ0	10 cores	160 cores	3.90 - 4.2	Entry price/performance

Each Feature Code provides four processor SCMs. The Feature Code identifies the number of cores on each SCM, and a system configuration requires one, two, three, or four of the same processor features. The number of Feature Codes corresponds to the number of system nodes.

All system nodes must use the same processor feature. Mixing processor features in a single Power E1180 system is not allowed.

Table 2-4 shows the Power11 processor frequencies for the Power E1180 offerings.

Table 2-4 Power11 processor frequencies

Fused core configuration	Power Save	Fixed frequency (nominal)	WOF base	Ultra Turbo	Power
16-core	3.6	3.8	3.95	4.3	440 W
12-core	3.6	3.9	4.05	4.4	400 W
10-core	3.6	3.9	4.05	4.2	355 W

Processor activations

Each Power E1180 system requires a minimum of 16 permanent processor core activations, which can be either static activations or Linux on IBM Power activations. This minimum applies to the entire system, not to individual nodes. The remaining cores can be activated permanently, activated temporarily, or left inactive until needed. Activations are not tied to specific hardware cores, SCMs, or nodes; instead, they are tracked by the system as a total number of activations, of various types, that the IBM Power Hypervisor (PHYP) allocates as needed. The exception is when using PEP 2.0, where a minimum of one activation is required.

Different activation options are available to accommodate various usage scenarios and pricing structures:

- Static activations: Permanent and support any type of application environment on the system.
- Mobile activations: Assigned to a specific system but can be transferred to any system within the PEP, supporting all types of applications.

2.4 Memory

The Power11 processor continues to use the native Open Memory Interface (OMI) introduced with Power10. This interface provides low-latency, high-bandwidth, and technology-neutral memory access, enabling support for both established and emerging memory technologies.

The Memory Controller Unit (MCU) in Power11 acts as the bridge between the on-chip SMP interconnect fabric and the OMI links. The processor includes eight on-chip MCUs, distributed across two physical blocks at opposite edges of the Power11 die. Each block supports 64 OMI lanes, which are organized into four ports. Each port contains two links, with eight lanes per link, operating at 32 Gbps. This configuration delivers high bandwidth and scalability with latency-optimized performance.

OMI and DDIMM configuration

The 16 OMI subchannels from the Power11 SCM connect directly to 16 memory-buffered DDIMM slots, with one DDIMM per subchannel. Each Power11 processor supports up to 16 DDIMMs, allowing a maximum of 64 DDIMMs per Power E1180 drawer.

OMI lane summary (per SCM):

- 128 OMI lanes total
- 32 Gbps signaling rate
- Eight lanes per OMI link
- Two OMI links per port (2 × 8 lanes)
- Eight OMI ports per SCM (16 × 8 lanes)

Figure 2-3 shows the memory subsystem architecture.

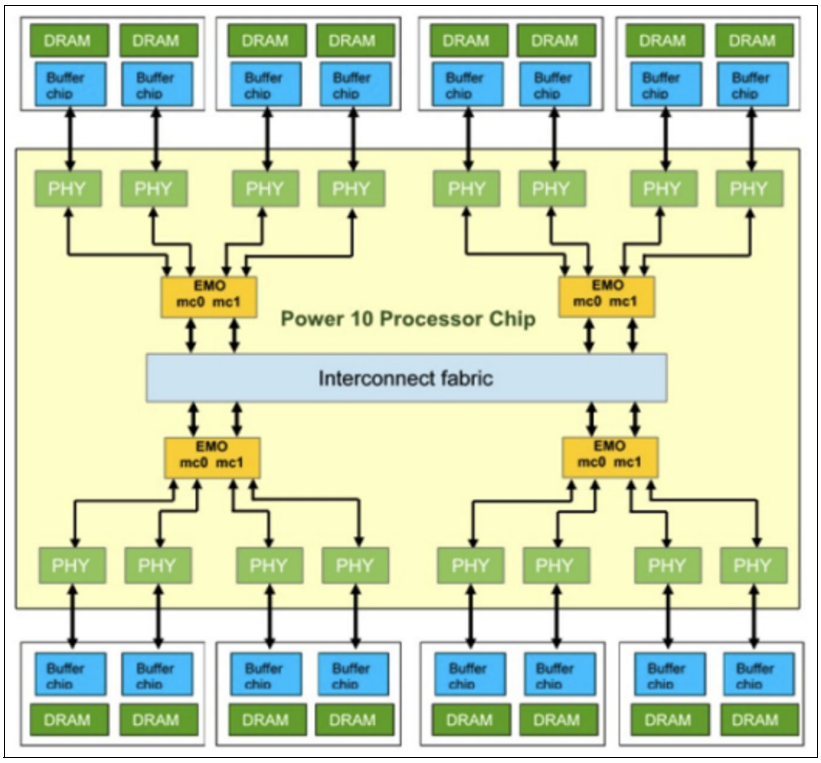

Figure 2-3 Power E1180 memory subsystem

DDR5 DDIMM support

The Power E1180 includes only 4U enterprise-level DDIMMs that use DDR5 DRAM. The supported DDIMM capacities are 32 GB, 64 GB, 128 GB, and 256 GB. Each DDIMM includes its own voltage regulators and is n+1 redundant. Two Power Management Integrated Circuits (PMICs) supply all the voltage levels that are required by the DDIMM.

Each 4U DDIMM contains four PMICs, organized into two redundant pairs. If one PMIC in each redundant pair remains functional, the DDIMM does not require replacement. Each DDR5 DDIMM operates at a data rate of 4000 MHz and includes a second DRAM port from the OMI buffer to the DRAM.

A DDR5 DDIMM:

- Increases sustained bandwidth by up to 50% under mixed read/write workloads compared to Power10 with DDR4 memory.
- Reduces memory latency by up to 15% under load compared to Power10 with DDR4 memory.

The maximum theoretical full-duplex bandwidth of the OMI interface reaches 1 TBps per Power11 SCM (2 × 512 GBps) when using DDR5 DDIMMs.

Note: While only DDR5 DDIMMs are orderable with new Power E1180 systems, customers upgrading from a Power E1080 with the same serial number can reuse their existing DDR4-based DDIMMs.

Memory capacity and activation

This configuration yields a total memory capacity of 16 DDIMMs per SCM and 64 DDIMMs per Power E1180 system node. The memory bandwidth depends on the DDIMM capacity that is configured for a specific Power E1180 system.

For each processor socket, a minimum of eight of the 16 DDIMMs must be populated.

- In a node, a minimum of 32 of the 64 DDIMMs must be populated.
- In a four-node system, a minimum of 128 of the 256 DDIMMs must be populated.

When filling the memory slots in an SCM, DDIMMs must be installed in groups of four. Therefore, the DDIMM slots of the SCMs are populated from 50% - 100%. The system node (four SCMs) can have 32, 36, 40, 44, 48, 52, 56, 60, or 64 DDIMMs physically installed.

To assist with the quad-plugging rules, four DDIMMs are ordered by using one memory Feature Code. The available options are shown in Table 2-5.

Table 2-5 Memory features

Memory capacity	Feature Code	Contents
128 GB	EMFM	(4 x 32 GB DDR5)
256 GB	EMFM	(4 x 64 GB DDR5)
512 GB	EMFP	(4 x 128 GB DDR5)
1024 GB	EMFQ	(4 x 256 GB DDR5)

All DDIMMs on the same SCM must be identical. For example, if you are using eight DDIMMs, both memory features on an SCM must be the same. A different SCM in the same system node can use a different memory feature. For example, one system node might use 128 GB, 256 GB, 512 GB, and 1024 GB memory features.

To provide more flexible pricing, memory activations are ordered separately from the physical memory and can be permanent or temporary. Activation features can be applied to DDR5 memory features of any size. Activations are not tied to a specific DDIMM; instead, they are tracked as a total quantity by the system. The PHYP determines which physical memory to use.

A minimum of 50% of the total physical memory capacity of a system must have permanent memory activations. For example, a system with 8 TB of physical memory must have at least 4 TB of permanent memory activations. These activations can be static, mobile, or Linux on IBM Power. At least 25% of the total physical memory must be static activations or Linux on IBM Power activations. For example, a system with 8 TB of physical memory must have at least 2 TB of static or Linux on IBM Power activations. The 50% minimum cannot be fulfilled by using mobile activations that are ordered on a different system. The exception is when using PEP 2.0, where the minimum memory activation is 256 GB.

Memory performance considerations

Memory performance depends on the number of DDIMMs populated and their size. Plugging DDIMMs of the same size provides the highest performance.

- System performance improves as more quads match each other.
- System performance improves as more processor DDIMMs match each other.
- System performance improves in a multidrawer system if memory capacity is balanced across drawers.
- System performance improves as memory is distributed across more DDIMM slots.

For example, in one drawer, if 1 TB of memory is required, it is better to use 64 × 32 GB DDIMMs than 32 × 64 GB DDIMMs. Although maximum memory bandwidth is achieved by filling all memory slots, plans for future memory expansion should be considered when selecting the memory feature size during the initial system order.

Note: Active Memory Expansion (AME) is an available option to increase memory capacity. AME is an AIX based function that uses data compression to expand the usable capacity of memory that is allocated to an AIX partition. For more information about AME, see 2.4.3, “Active Memory Expansion” on page 58.

DDIMM plug rules

The Power E1180 system drawers are offered with a base memory configuration of 32 DDIMMs each. In this base configuration, each processor module has eight of the 16 DDIMMs that are populated. Feature upgrades are available in increments of four DDIMMs, all of which must have the same capacity. These four DDIMMs must be installed in four of the eight remaining slots that are connected to a single processor module.

You may not install DDIMMs in a way that results in an odd number of populated slots behind a single processor module. You may not install DDIMMs in a configuration that results in 10 or 14 populated slots behind a single processor module. The only valid numbers of DDIMMs that are installed in slots that are connected to a processor module are 8, 12, or 16.

Here are the general DDIMM plug rules for the Power E1180 system:

- For the minimum memory that is required, see E1080 memory setup and optimization.
- All DDIMMs under each processor must have the same capacity.
- DDIMMs must be installed in quads (groups of four).
- DDR4 and DDR5 DDIMMs cannot be mixed in a single-node system.
- Each DDIMM quad must be the same size and type (identical to each other) and must match other DDIMM quads under the same socket, but can differ from DDIMM quads under other sockets.

  For example, in a one-node drawer:

- The first processor can have 8 × 32 GB DDIMMs.
- The second processor can have 8 × 64 GB DDIMMs.
- The third processor can have 8 × 128 GB DDIMMs.
- The fourth processor can have 8 × 256 GB DDIMMs.
- System performance improves when DDIMMs match each other.

The user has flexibility in assigning more memory beyond the minimum. In multidrawer Power E1180 systems, the user can either fully populate one drawer and partially populate the others or populate all drawers symmetrically.

For example, in a two-drawer system with 12 extra quads of memory (beyond the minimum), the user can install eight quads in one drawer and four quads in the other, or six quads in each drawer.

2.4.1 Active Memory Mirroring

AMM is a reliability feature in Power servers that enhances system availability by duplicating the memory that is used by the system hypervisor. It maintains two synchronized copies of this critical memory, enabling the system to switch instantly to the mirrored copy if a memory failure occurs. This seamless failover mechanism helps prevent system crashes and helps ensure continuous operation, making AMM especially valuable for enterprise environments that run mission-critical workloads.

AMM is available in all configurations of the Power E1180 system.

2.4.2 Pervasive memory encryption

The Power11 MCU provides the system memory interface between the on-chip SMP interconnect fabric and the OMI links. This design makes the MCU an ideal functional unit to implement memory encryption logic. The Power11 on-chip MCU encrypts and decrypts all traffic to and from system memory by using Advanced Encryption Standard (AES) technology.

The Power11 processor supports the following modes of operation:

- AES XTS mode

  XTS stands for *xor–encrypt–xor based tweaked-codebook mode* with ciphertext stealing. AES XTS provides a block cipher with strong encryption, which is useful for encrypting persistent memory.

  Persistent DIMM technology retains the data that is stored inside the memory DIMMs even if the power is turned off. A malicious attacker who gains physical access to the DIMMs can steal memory cards. If the data that is stored in the DIMMs is not encrypted, it can leave the data center in clear text.

  Memory cards that leave the data center for repair or replacement can also create a potential security breach because the attacker might have arbitrary access to the persistent DIMM data. Therefore, the stronger encryption of the AES XTS mode is required for persistent memory. The AES XTS mode of the Power10 processor is supported for future use if persistent memory solutions become available for Power servers.

- AES CTR mode

  CTR stands for *counter mode of operation* and designates a low-latency AES block cipher. Although the level of encryption is not as strong as with the XTS mode, the low-latency characteristics make it the preferred mode for encrypting volatile memory. AES CTR makes it more difficult to gain physical access to data through the memory card interfaces. The goal is to protect against physical attacks, which becomes increasingly important in the context of cloud deployments.

The Power E1180 systems support the AES CTR mode for pervasive memory encryption. Each Power11 processor holds a 128-bit encryption key that is used by the processor's MCU to encrypt the data of the DDIMMs that are attached to the OMI links.

The MCU crypto engine is transparently integrated into the data path, which means that the data fetch and store bandwidth is not compromised by the AES CTR encryption mode. Because the encryption has no noticeable performance effect and because of the obvious security benefit, pervasive memory encryption is enabled by default and cannot be turned off through any administrative interface.

2.4.3 Active Memory Expansion

AME is an innovative technology that supports the AIX OS. It enables the effective maximum memory capacity to exceed the physical memory limit. By compressing and decompressing memory content, AME can expand memory by 100% or more. This capability enables a partition to handle more work or support more users with the same physical memory. Similarly, it enables a system to run more partitions and perform more work without more physical memory.

AME uses CPU resources to compress and decompress memory contents. The tradeoff between memory capacity and processor cycles can be advantageous, but the degree of expansion depends on how compressible the memory content is. It also requires adequate spare CPU capacity for compression and decompression.

The Power E1180 includes a hardware accelerator that improves AME efficiency and reduces IBM Power core resource usage. You have extensive control over AME usage. Each AIX partition can enable or disable AME. Control parameters define the wanted expansion level for each partition to manage CPU usage by the AME function. An IPL is required for the partition that enables memory expansion. After activation, you can monitor AME through standard AIX performance tools, such as **`lparstat`**, **`vmstat`**, **`topas`**, and **`svmon`**.

A planning tool that is included with AIX enables you to sample actual workloads and estimate both memory expandability and CPU requirements. Any IBM Power model can run the planning tool. In addition, a one-time, 60-day AME trial is available to provide accurate memory expansion and CPU usage measurements. You can request the trial at the Capacity on Demand web page.

AME is enabled by the chargeable hardware feature Feature Code EM89, which you can order with the system or as a Miscellaneous Equipment Specification (MES) order. When you order the enablement feature, you receive a software key that applies to the system node. An IPL is not required to enable the system node. The key is specific to an individual system and is permanent. It cannot be transferred to another system.

The additional CPU resources that are used for memory expansion are part of the CPU resources that are assigned to the AIX partition running AME. Normal licensing requirements apply.

2.5 PCIe slots

Each Power E1180 system node is equipped with eight PCIe Gen 5 hot-plug enabled slots. The total number of available PCIe slots scales with the number of system nodes in the system, as shown in Table 2-6.

Table 2-6 PCIe slots available internally

Number of nodes	Number of PCIe slots
1	8
2	16
3	24
4	32

This modular design enables scalable expansion of PCIe resources based on system requirements. These buses provide high-bandwidth connectivity for performance-critical components and peripheral devices.

Each PCIe adapter installs in a dedicated slot within a PCIe extender cassette. These cassettes connect directly to the system backplane, helping ensure efficient signal routing and simplified maintenance. This design supports hot-plug capabilities and streamlines the process of upgrading or replacing PCIe cards without system downtime. The PCIe cassettes support half-height, half-length (HHHL) adapters, and all slots are concurrently maintainable. All slots support EEH.

For customers who require more PCIe adapter slots, the Power E1180 supports the PCIe Gen 4 I/O expansion drawer (Feature Code ENZ0). Each system node supports up to four PCIe Gen 4 I/O expansion drawers, and each drawer provides up to 12 Gen 4 PCIe adapter slots.

For more information about PCIe slots and I/O Expansion drawers, see Chapter 3, “I/O subsystem” on page 65.

2.6 Storage and USB features

The Power E1180 provides expanded PCIe connectivity, enabling the attachment of internal NVMe devices within each system node. Unlike some systems, it does not include an integrated USB port, reflecting many clients’ preference to minimize potential security risks that are associated with USB access. If you require USB functions, you can install a PCIe USB adapter in one of the available I/O slots, either within a system node or in an I/O expansion drawer.

NVMe devices

The Power E1180 provides four PCIe x4 Gen 4 buses that connect to internal NVMe storage devices. Each system node includes a carrier backplane that supports up to four 7 mm NVMe U.2 drives or up to two 15 mm NVMe U.2 drives. Feature Code EJBC provides the four-drive carrier, and Feature Code EJBD provides the two-drive carrier. Each system node requires one backplane, even if you do not select any NVMe U.2 drives.

If you require extra data storage capacity, the Power E1180 supports the NED24 I/O expansion drawer. Each NED24 drawer provides 24 NVMe slots that support small form factor (SFF) NVMe U.2 drives, delivering high-performance, scalable storage expansion.

For more information about NVMe support, see 3.1.2, “Internal NVMe storage subsystem” on page 71.

USB

The Power E1180 supports a single stand-alone external USB DVD drive (Feature Code EUA5), which includes a cable for connecting the drive to the front-accessible USB port on the system control unit (SCU).

The USB 3.0 adapter (Feature Code EC6J) can be assigned to any logical partition (LPAR) and migrated between operating LPARs as needed. This flexibility allows you to assign the DVD drive feature to any LPAR by using dynamic logical partitioning (DLPAR).

For more information about the USB features of Power E1180, see 3.1.3, “Internal USB subsystem” on page 75.

2.7 I/O expansion drawers

Each system node provides up to eight PCIe adapter slots and up to four internal NVMe solid-state drives (SSDs) installed in an internal NVMe enclosure. In a fully configured four-node system, the Power E1180 supports up to 32 PCIe Gen 5 adapters and 16 NVMe drives.

If you require extra PCIe slots or internal storage, the Power E1180 supports I/O expansion drawers:

- To add more PCIe slots beyond the system node slots, you can install a PCIe Gen 4 I/O expansion drawer (Feature Code ENZ0).
- To increase internal disk capacity, you can install a 24-bay NVMe enclosure, the NED24 NVMe expansion drawer (Feature Code ESR0).

I/O drawer connectivity

To enable connectivity to PCIe Gen 4 I/O expansion drawer or NED24 PCIe Gen 4 expansion drawers, install the PCIe Converter Adapter (Feature Code EJ24) in one of the PCIe extender slots at the rear of the system node. This adapter works with a new generation of high-speed active optical cables (AOCs) and copper cables that support the PCIe Gen 4 communication protocol, helping ensure high bandwidth and signal integrity over extended distances.

Table 2-7 shows the maximum number of PCIe slots that are available for installing the CXP converter adapter, the maximum number of PCIe Gen 4 I/O drawers, and the maximum number of NED24 drawers based on the system configuration.

Table 2-7 Drawer availability for different system node configurations

Number of system nodes	CXP converter adapter available slots per node	Maximum PCI Gen4 I/O drawers only	Maximum NED24 drawers only
1	8	4	2
2	16	8	4
3	24	12	6
4	32	16	8

2.8 Environmental requirements

The Power E1180 system specifications are essential for planning your system. For an initial assessment, this section provides an overview of the following topics:

- Electrical characteristics
- Environmental requirements and noise emissions
- Physical dimensions

For comprehensive Model 9080-HEU system specifications and product documentation, see IBM Documentation.

2.8.1 Electrical characteristics

Each Power E1180 system node includes four 1950 W bulk power supplies. The hardware design provides N+2 redundancy for the system power supply, enabling any node to operate at full function in nominal mode with any two power supplies active.

Depending on the specific Power E1180 configuration, the SCU receives power through two UPIC cables that are connected to one or two system nodes.

Table 2-8 lists the electrical characteristics for each Power E1180 system node. For planning purposes, use the maximum values that are provided. Actual power draw and heat load depend on the processor, memory, adapter, and expansion drawer configuration, and workload characteristics.

Table 2-8 Electrical characteristics per Power E1180 system node

Electrical characteristics	Properties
Operating voltage	200 - 208 / 220 - 240 V AC
Operating frequency	50 Hz or 60 Hz +/- 3 Hz AC
Maximum power consumption	4850 W
Maximum power source loading	5 kVA
Maximum thermal output	16548 BTU/h
Phase	Single

Note: The Power E1180 must be installed in a rack with a rear door and side panels to comply with electromagnetic compatibility requirements.

Environmental requirements and noise emission

The environmental requirements for Power E1180 systems are classified for both operating and non-operating environments. The operating environments are further segmented into recommended and allowable conditions.

Note: The IBM Systems Energy Estimator provides more accurate information about power consumption and thermal output based on a specific configuration.

The recommended operating environment designates the long-term conditions that deliver the greatest reliability and energy efficiency. The allowable operating environment represents the conditions under which the equipment is tested to verify functions. Because operating in the allowable range can stress the equipment, use these conditions only for short-term operation, not continuous operation.

The non-operating environment applies when the equipment is removed from its original shipping container and installed but powered down. The allowable non-operating environment defines the range that an unpowered system can experience short term without damage.

Table 2-9 lists the environmental requirements for the Power E1180 system, including temperature, humidity, dew point, and altitude. It also lists the maximum noise emission level for a fully configured Power E1180 system.

Table 2-9 Power E1180 environment requirements and noise emission

Property	Environment		
	Operating		Non-operating
	Recommended	Allowable	
Temperature	18.0°C - 27.0°C (64.4°F - 80.6°F)	5.0°C - 40.0°C[a] (41.0°F - 104.0°F)	5°C - 45°C (41°F - 113°F)
Low-end moisture	-9.0°C (15.8°F) dew point	-12.0°C (10.4°F) dew point and 8% relative humidity	N/A
High-end moisture	60% relative humidity and 15°C (59°F) dew point	85% relative humidity and 24.0°C (75.2°F) dew point	N/A
Relative humidity	N/A	N/A	8% to 85%
Maximum dew point	N/A	N/A	27.0°C (80.6°F)
Maximum altitude	N/A	3,050 m (10,000 ft)	N/A

a. As a best practice, do not operate above 27°C, but you can expect full performance up to 35°C. Above 35°C, the system can operate, but performance might be reduced to preserve component integrity. At temperatures above 40°C, component reliability can be affected.

A comprehensive list of noise emission values for various Power E1180 system configurations is available in the Power E1180 product documentation. For more information about noise emissions, search for “Model 9080-HEU server specifications” at IBM Documentation.

Note: Government regulations, such as those issued by the Occupational Safety and Health Administration (OSHA) or European Community Directives, can govern workplace noise exposure and might apply to your system installation. The Power E1180 system offers an optional acoustical door feature that helps reduce noise emissions.

Sound pressure levels in your installation depend on several factors, including the number of racks, the size, materials, and configuration of the room, noise from other equipment, ambient temperature, and employee location relative to the equipment.

Compliance with government regulations also depends on factors such as the duration of employee exposure and whether employees use hearing protection. As a best practice, consult qualified experts to help ensure compliance with applicable regulations.

Physical dimensions

The Power E1180 is a modular system that is built on a single SCU and one, two, three, or four system nodes.

Each component must be mounted in a 19-inch industry-standard rack. The SCU requires 2U of rack space, and each system node requires 5U.

Table 2-10 on page 63 shows the total space that is required based on the number of nodes.

Table 2-10 PCIe slots available internally

Number of nodes	Space required
1	7U
2	12U
3	17U
4	22U

Allow extra rack space for components such as PCIe I/O expansion drawers, an HMC, a flat-panel console kit, network switches, power distribution units (PDUs), and cable egress.

Table 2-11 lists the physical dimensions of the Power E1180 System Control Unit (SCU) and a Power E1180 system node. The component height is also provided in EIA rack units. (One EIA unit corresponds to one rack unit (U) and is defined as 1.75 inches or 44.45 mm.)

Table 2-11 Physical dimensions of the Power E1180 system components

Dimensions	Power E1180 System Control Unit	Power E1180 system node
Width	445.6 mm (17.54 in)	445 mm (17.51 in)
Depth	779.7 mm (30.7 in)	866.95 mm (34.13 in)
Height	86 mm (3.39 in) / 2 EIA units	217.25 mm (8.55 in) / 5 EIA units
Weight	22.7 kg (50 lb)	81.6 kg (180 lb)

Lift tools

As a best practice, keep a lift tool available at each site where one or more Power E1180 systems are installed to avoid delays during servicing. An optional lift tool (Feature Code EB2Z) is available for order with a Power E1180 system. One EB2Z lift tool can be shared among multiple systems and I/O drawers. The EB2Z lift tool includes a hand crank to lift and position up to 159 kg (350 lb) and measures 1.12 m × 0.62 m (44 in × 24.5 in).

Attention: A single system node can weigh up to 86.2 kg (190 lb). A lighter, lower-cost lift tool (Feature Code EB3Z) and a wedge shelf toolkit for EB3Z (Feature Code EB4Z) are also available.

Environmental operating envelope

Table 2-12 shows the environmental operating envelope for the Power E1180 system.

Table 2-12 ASHRAE Class A3 requirements

Status	Temperature (dry bulb)	Altitude	Relative humidity
Power Off	5°C - 45°C	N/A	8% - 80%
Shipping	-40°C - 60°C	N/A	5% - 100%

Status	Temperature (dry bulb)	Altitude	Relative humidity
Operating	5°C - 40°C	Up to 3050 meter @ 5°C - 28°C	10% - 80%
		Up to 2875 meter @ 29°C	
		Up to 2700 meter @ 30°C	
		Up to 2525 meter @ 31°C	
		Up to 2350 meter @ 32°C	
		Up to 2175 meter @ 33°C	
		Up to 2000 meter @ 34°C	
		Up to 1825 meter @ 35°C	
		Up to 1650 meter @ 36°C	
		Up to 1475 meter @ 37°C	
		Up to 1300 meter @ 38°C	
		Up to 1125 meter @ 39°C	
		Up to 950 meter @ 40°C	

3

I/O subsystem

The internal I/O system of IBM Power servers is designed for high-speed, low-latency data movement to help ensure optimal performance for enterprise workloads. At the core of this architecture is the Power11 processor, which integrates advanced I/O subsystems that support high throughput and efficient communication between processors, memory, and peripheral devices. IBM Power uses technologies such as Peripheral Component Interconnect Express (PCIe) Gen5 and Non-Volatile Memory Express (NVMe) to deliver fast data access and reduce bottlenecks.

IBM Power servers also use a modular and scalable I/O architecture that allows flexible configuration and expansion. The system supports Single Root I/O Virtualization (SR-IOV) for efficient sharing of I/O resources across virtual machines (VMs). PowerVM virtualization technology enables dynamic allocation of I/O bandwidth to match workload demands.

This internal I/O design improves performance and contributes to system reliability and availability by incorporating features such as redundant paths, hot-swappable components, and advanced error detection and correction. These capabilities make the IBM Power internal I/O system a key enabler of high-throughput, resilient, and scalable enterprise computing.

The following topics are covered in this chapter:

- Internal I/O
- Enhancing I/O scalability with expansion drawers
- Supported adapters
- Other device support

© Copyright IBM Corp. 2025.

3.1 Internal I/O

The IBM Power E1180, built on the Power11 architecture, includes an advanced internal I/O subsystem that supports high-throughput, low-latency data movement within the server. At the core of this subsystem is the Power11 processor's Open Memory Interface (OMI), which decouples memory from the processor and enables flexible, high-bandwidth memory configurations. This design supports DDR4 and DDR5 memory through memory buffer chips, providing faster data access and improved scalability. The Power E1180 also integrates PCIe Gen5 technology, which doubles the bandwidth of its predecessor and enables faster communication between internal components such as storage controllers and network adapters.

The internal I/O design of the Power E1180 emphasizes reliability and serviceability. It includes redundant paths and hot-swappable components to minimize downtime and maintain continuous operation in enterprise environments. The system supports NVMe drives natively, improving internal storage performance, and uses a modular design that enables expansion through I/O drawers. These drawers connect through high-speed links and provide extra PCIe slots, enabling the system to scale with growing workload demands. Overall, the internal I/O architecture of the Power E1180 is designed to deliver high performance, flexibility, and resilience for mission-critical applications.

3.1.1 Internal PCIe subsystem

Each Power E1180 system node includes eight PCIe Gen5 hot-plug enabled slots. The total number of available PCIe slots scales with the number of system nodes in the server:

- A single-node system provides eight PCIe slots.
- A two-node system provides 16 PCIe slots.
- A three-node system provides 24 PCIe slots.
- A four-node system provides 32 PCIe slots.

This modular design enables scalable expansion of PCIe resources based on system requirements. For customers who require extra PCIe adapter slots, the Power E1180 supports the PCIe Gen4 I/O expansion drawer (Feature Code ENZ0). Each system node supports up to four PCIe Gen4 I/O expansion drawers, and each drawer provides up to 12 Gen4 PCIe adapter slots.

PCIe bus layout in the Power E1180

The Power E1180 system supports a combination of PCIe buses to accommodate a wide range of PCIe adapters. Specifically, the system includes the following buses:

- Six PCIe x16 Gen4 or x8 Gen5 buses
- Two PCIe x8 Gen5 buses

Each system node also provides four PCIe x4 Gen4 buses that connect to the internal No NVMe storage bays. NVMe support is described in 3.1.2, "Internal NVMe storage subsystem" on page 71.

The internal I/O subsystem of the Power E1180 server connects to the PCIe Gen5 controllers on a Power11 chip in the system. Each Power11 processor module includes two PCIe Host Bridges (PHBs). The PHBs (PHB0 and PHB1) each provides 16 PCIe lanes.

- Each Power E1180 system node includes eight PCIe Gen5 hot-plug enabled slots. The PHBs provide different types of end-point connections.
- Both PHBs from Power11 CPUs with physical location codes C13 and C16 connect directly to a total of four PCIe Gen4 x16 or PCIe Gen5 x8 slots (two slots per CPU).
- On Power11 CPUs with physical location codes C14 and C15, one PHB per CPU connects directly to a PCIe Gen4 x16 or PCIe Gen5 x8 slot.
- The second PHB connects to a PCIe Gen5 x8 slot and two internal NVMe solid-state drive (SSD) PCIe Gen4 x4 slots.

Figure 3-1 shows the PCIe bus layout in the Power E1180.

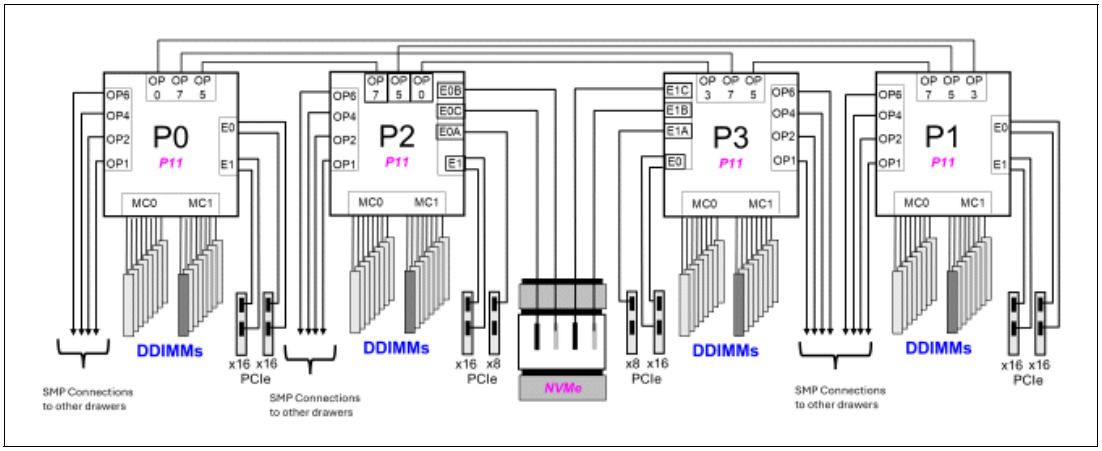

Figure 3-1 PCIe connectivity

Figure 3-2 shows how the PCIe connectivity relates to the Power11 chip connections.

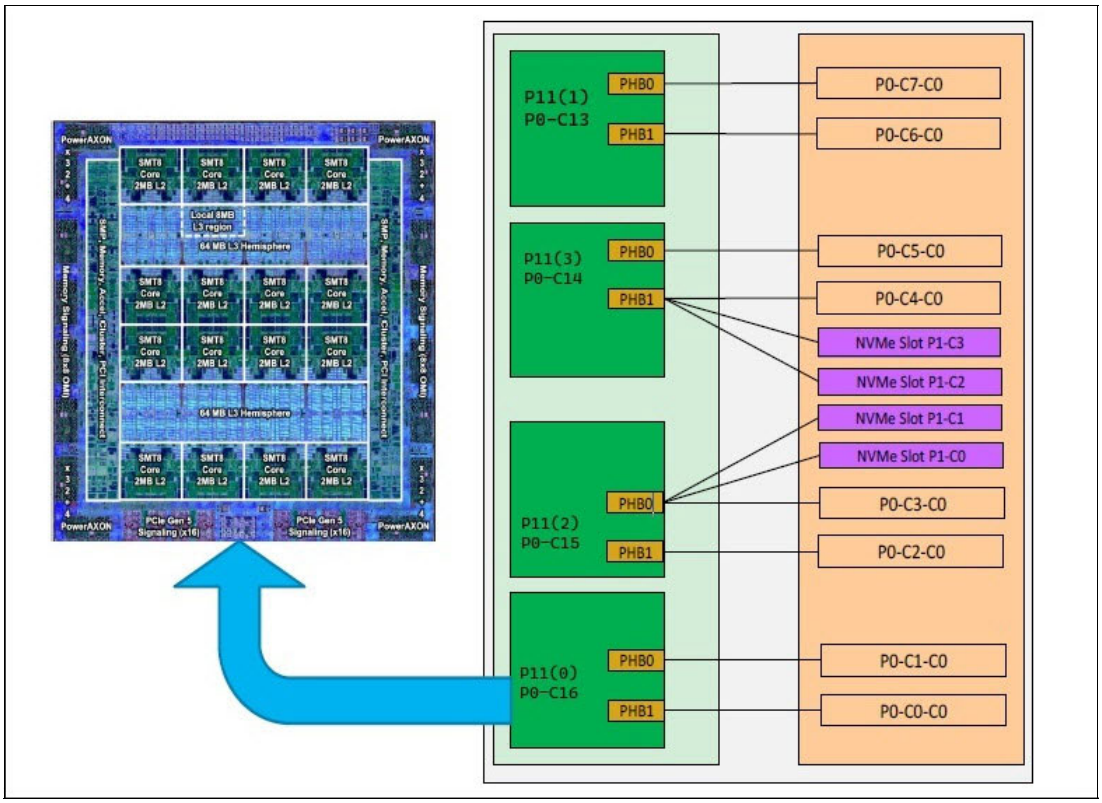

Figure 3-2 PCIe connectivity to the processor chip

The overall design provides each Power E1180 system node with the following components:

- Six PCIe Gen4 x16 or PCIe Gen5 x8 slots. These slots operate with eight lanes at Gen5 or 16 lanes at Gen4 and also support 16-lane Gen1 through Gen3 cards.
- Two PCIe Gen5 x8 slots. These slots operate with eight lanes at Gen5 and also support eight-lane Gen1 through Gen4 cards. A physical x16 connector is used, but only eight lanes are wired.
- Four internal NVMe SSDs, each using a PCIe Gen4 x4 connection.

I/O bandwidth calculations

Table 3-1 shows the maximum I/O bandwidth for each PCIe adapter slot type in the system node. These numbers are used inTable 3-2 on page 69 to calculate the maximum bandwidth of a single system node.

Table 3-1 Internal I/O component bandwidth

Connection	Type	Speed (duplex)
PCIe adapter slot	PCIe Gen4 x16	64
PCIe adapter slot	PCIe Gen5 x8	64
NVMe slot	PCIe Gen4 x4	16

Table 3-2 shows the theoretical maximum I/O bandwidth for a single system node.

Table 3-2 Maximum I/O bandwidth of a single system node

Number	Type	Subtotal (GBps)	Total (GBps)
Six	PCIe Gen 4 x16 / PCIe Gen 5 x8 slots at 64 GBps	384	384
Two	PCIe Gen 5 x8 Slots at 64 GBps	128	512
Four	NVMe slots at 16 GBps	64	576

Table 3-3 shows the maximum I/O throughput for a Power E1180 based on the number of system nodes that are configured.

Table 3-3 Maximum I/O bandwidth based on number of nodes

Number of nodes	Maximum bandwidth (GBps)
One node	576
Two nodes	1152
Three nodes	1728
Four nodes	2304

PCIe cassette

Each PCIe adapter is installed in a dedicated slot within a PCIe extender cassette. These cassettes connect directly to the system backplane, helping ensure efficient signal routing and simplified maintenance. This design supports hot-plug capabilities and enables upgrades or replacements of PCIe cards without system downtime.

Figure 3-3 shows the PCIe extender cassette with the PCIe adapter inside it.

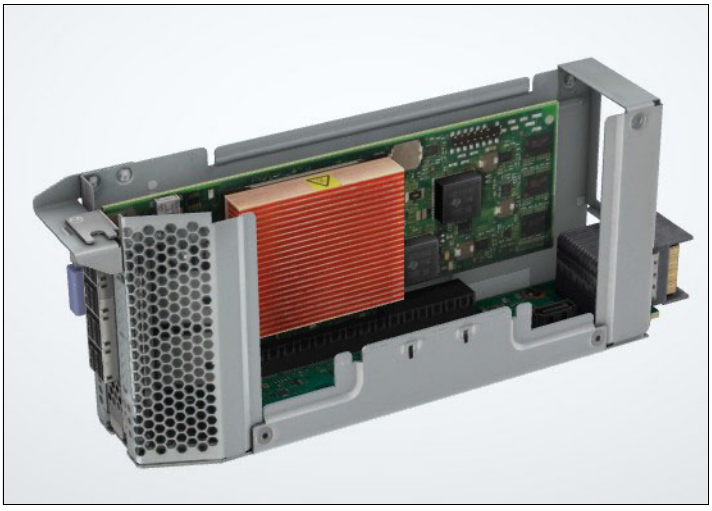

Figure 3-3 PCIe extender cassette with PCIe adapter

- The cassette locations are P0-C0 through P0-C7. The adapter locations are P0-C0-C0 through P0-C7-C0. All slots are low-profile (LP) type, which means half-height, half-length (HHHL).
- The slots support Enhanced Error Handling (EEH) and can be serviced while the system power is on. They also support Gen1 through Gen3 adapters. All PCIe slots support SR-IOV adapters.
- At the time of writing, no adapter can dissipate more than 25 watts. Special considerations are required to exceed this limit. The Power E1180 server supports concurrent maintenance (hot-plugging) of PCIe adapters.

Table 3-4 shows the slot location and capabilities for each slot.

Table 3-4 Internal PCIe slots

Location code	Description	Socket and PHB	Slot number	SRIOV
P0-C0	► PCIe Gen4 x16 ► PCIe Gen5 x8	P0/1	1	Yes
P0-C1	► PCIe Gen4 x16 ► PCIe Gen5 x8	P0/0	2	Yes
P0-C2	► PCIe Gen4 x16 ► PCIe Gen5 x8	P2/1	3	Yes
P0-C3	► PCIe Gen5 x8	P2/0	4	
P1-C0	► PCIe Gen 4 x4	P2/0	NVMe Slot 1	N/A
P1-C1	► PCIe Gen 4 x4	P2/0	NVMe Slot 2	N/A
P1-C2	► PCIe Gen 4 x4	P3/1	NVMe Slot 3	N/A
P1-C3	► PCIe Gen 4 x4	P3/1	NVMe Slot 4	N/A
P0-C4	► PCIe Gen 5 x8	P3/1	5	Yes
P0-C5	► PCIe Gen4 x16 ► PCIe Gen5 x8	P3/0	6	Yes
P0-C6	► PCIe Gen4 x16 ► PCIe Gen5 x8	P1/1	7	Yes
P0-C7	► PCIe Gen4 x16 ► PCIe Gen5 x8	P1/0	8	Yes

Figure 3-4 on page 71 shows a rear view of a Power E1180 system node that illustrates the locations that are shown in the table. The adapter cassette includes three LEDs to indicate status:

- A power and activity LED (green)
- An identify LED for the adapter (amber, labeled C0)
- A fault LED for the cassette (amber)

To find the server, use the blue identify LED on the enclosure.

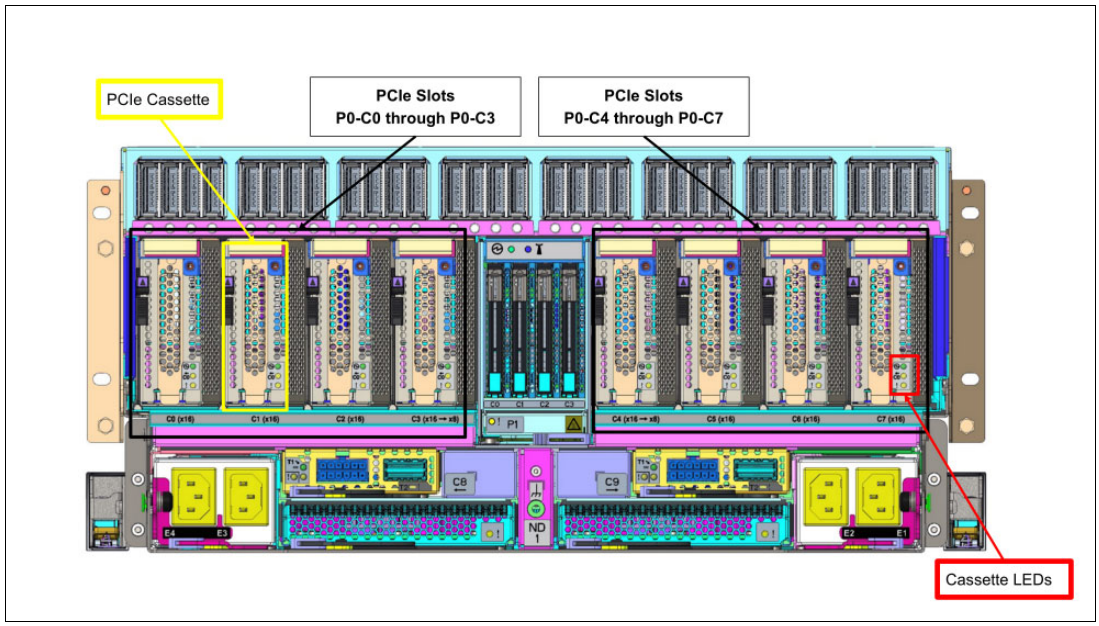

Figure 3-4 Rear view of a Power E1180 system node highlighting PCIe slots

Considerations for adapter placement

When choosing an adapter and selecting a slot, consider the following guidelines:

- PCIe x16 cards are supported only in PCIe x16 slots. The slot priority is 1, 3, 7, 2, 6, and 8.
- PCIe adapters with x8 or fewer lanes are supported in all PCIe slots. The slot priority is 1, 7, 3, 6, 2, 8, 4, and 5.
- Install x8 and x16 adapters in slots of the same size before mixing connector sizes with slot sizes.
- Adapters with smaller connectors can be installed in larger PCIe slots, but larger connectors are not compatible with smaller PCIe slots.

  For a list of adapters and their supported slots, see 3.3, “Supported adapters” on page 90.
- Place adapters that do not require high bandwidth in an external expansion drawer. Using a low-latency, high-bandwidth slot for a low-bandwidth adapter is not an efficient use of system resources.
- Use the high-profile version of all adapters when the system includes a PCIe expansion drawer. This configuration allows most adapters to be installed in the expansion drawer.
- Use node slots for cable adapters (Feature Code EJ24) or high-bandwidth PCIe Gen4 or PCIe Gen5 adapters.

3.1.2 Internal NVMe storage subsystem

The Power E1180 supports NVMe storage technology. The NVMe subsystem is available in each Power E1180 system node drawer. Each system node includes four internal NVMe U.2 drive slots in the center of the rear of the system.

These slots can be configured as either four 7 mm slots by using the Feature Code EBJC backplane or two 15 mm slots by using the Feature Code EJBD or EJBE backplane. The slots can be assigned independently to any partition, as shown in Figure 3-5.

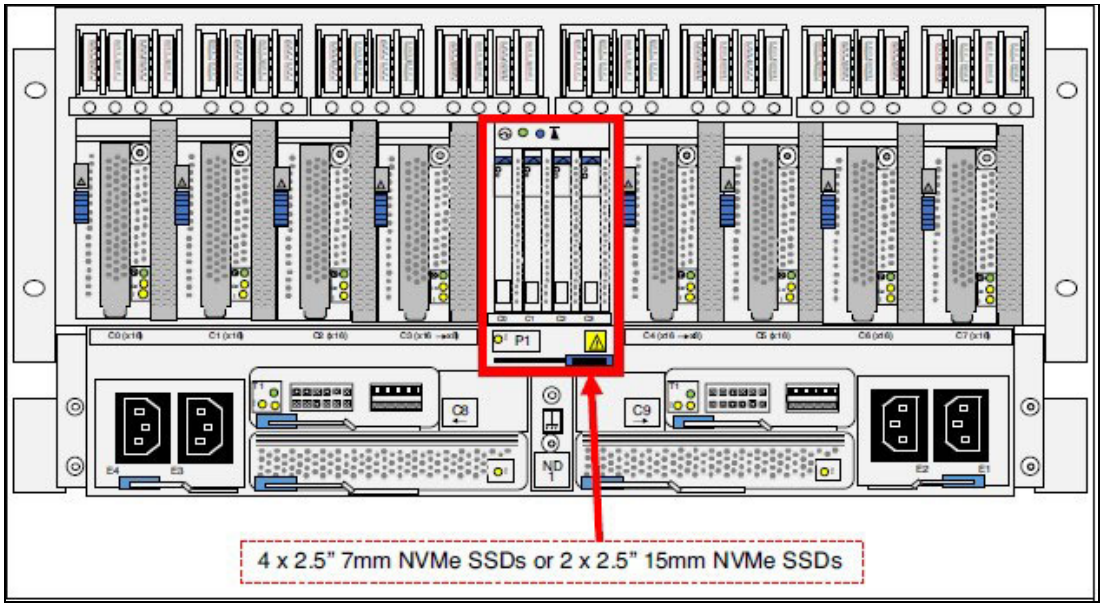

Figure 3-5 Rear view of a Power E1180 system node highlighting NVMe slots

The internal 7 mm NVMe devices support only one namespace. Their primary use is to host the Virtual I/O Server (VIOS). These devices do not support IBM i. IBM i support requires the Feature Code EJBD or Feature Code EJBE backplane and 15 mm NVMe devices.

The NVMe backplane has a location code of P1. From left to right, the location codes of each NVMe slot are P1-C0 through P1-C3. Slot locations P1-C2 and P1-C3 remain empty if the Feature Code EJBD backplane is selected to allow spacing for the 15 mm option.

Each slot uses an x4 PCIe connection. Two SSDs connect to the PHB on the P2 CPU socket, and two SSDs connect to the PHB on the P3 CPU socket, as shown in Figure 3-6 on page 73. The buses are routed through the system board from the CPU to the NVMe slot interface card, then through the SSD riser card and into the SSD.

Figure 3-6 shows the NVMe connectivity to the CPUs through the backplane.

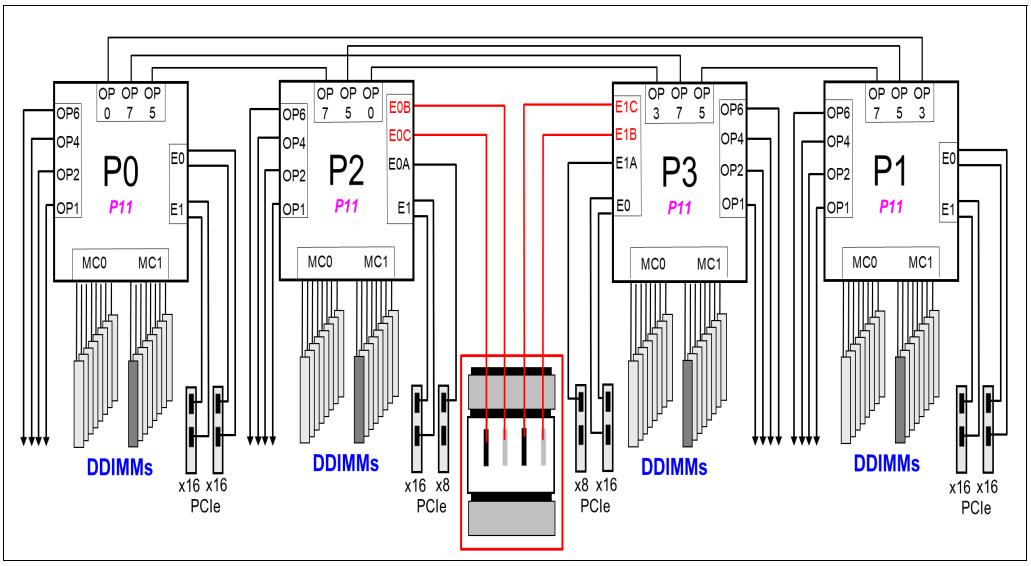

Figure 3-6 NVMe connectivity

The NVMe slots support U.2 NVMe flash SSDs. Each system node can use either four 2.5-inch, 7 mm form-factor SSDs or two 2.5-inch, 15 mm NVMe SSDs. The PCIe buses for the NVMe U.2 drives support up to PCIe Gen4.

Important: Choose the type of NVMe SSD carefully based on your planned usage. Match drive characteristics, such as Drive Writes Per Day (DWPD) and the number of namespaces, to your intended use case.

Internal SSD plug order

For redundancy, it is a best practice to distribute the NVMe drives across the system nodes as follows:

- Populate slot C0 in each system node, starting with node 1, and then populate other C0 slots if extra nodes are present.
- Populate slot C2 in each system node, starting with node 1, and then populate other C2 slots if extra nodes are present.
- Populate slot C1 in each system node, starting with node 1, and then populate other C1 slots if extra nodes are present.
- Populate slot C3 in each system node, starting with node 1, and then populate other C3 slots if extra nodes are present.

Each NVMe U.2 drive has two LEDs at the top of the drive that indicate the following status:

- A power and activity LED (green)
- An error and identify LED (amber)

Figure 3-7 shows the location of the NVMe U.2 drive LEDs in the system.

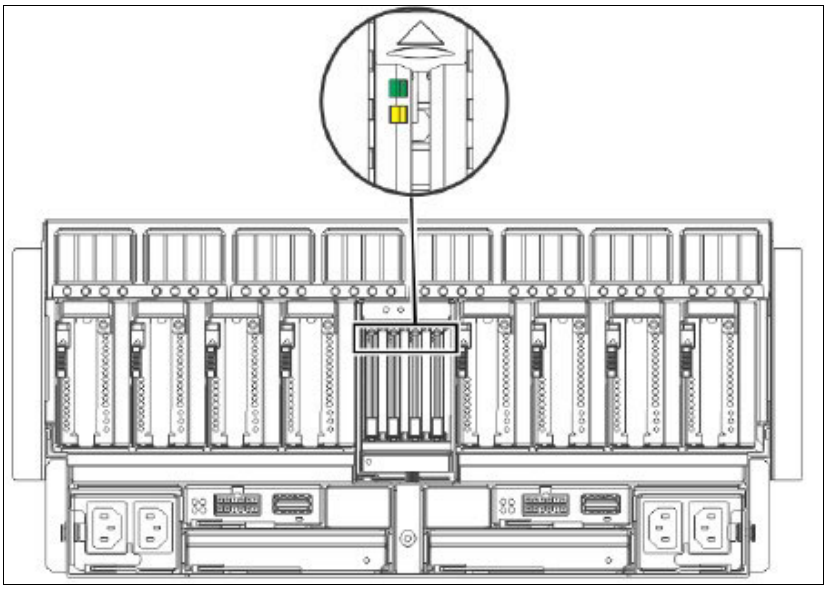

Figure 3-7 Rear view of a Power E1180 system node with the NVMe slot LED location

Maintaining NVMe drives

The lifecycle of an NVMe device depends on the number of write and erase cycles that its flash memory cells can endure. Each write operation slightly wears the memory cells, and this wear accumulates over time. Modern NVMe drives use wear-leveling algorithms and error correction to extend their usable life. However, they still have a finite endurance rating, typically measured in terabytes written (TBW). Check the specifications of any NVMe device that you use to determine its write endurance rating.

Concurrent maintenance of NVMe drives is supported. You can also check the remaining life of an NVMe device from the logical partition (LPAR) that owns the device. To determine the remaining life of an NVMe device, complete the following steps:

- For an IBM AIX OS:
  a. From the AIX CLI, enter **`diag`** and press Enter.
  b. From the Function Selection menu, select **Task Selection** → **NVMe general health information**.
  c. Select the NVMe device that you want to check and press Enter.
  d. View the “Percentage of NVMe subsystem life used” field.
- For a Linux OS:
  a. From the Linux CLI, enter the following command and press Enter:

     ```
     nvme smart-log /dev/nvmeX -H
     ```

     `nvmeX` is the resource name of the NVMe device.
  b. View the Percentage used field.

Replace an NVMe device that is approaching the end of its lifespan. As it nears the limit of supported write cycles, write performance degrades. Eventually, the device moves to a read-only state. When the OS attempts to write to a read-only device, those operations fail, and the system treats the device as failed. To maintain normal write functions, replace the NVMe device before it reaches this state.

3.1.3 Internal USB subsystem

The Power E1180 system does not include integrated USB controllers or ports in the system nodes. To enable USB access, you must use a USB PCIe adapter. The PCIe2 2-PORT USB 3.0 LP adapter (Feature Code EC6J) is supported in the system node. This adapter is typically installed in PCIe slot 4 or 5 of system node 1 or node 2.

A USB cable (Feature Code EC6N) connects the EC6J adapter port to the rear USB port in the System Control Unit (SCU). This connection is internally routed to a front-accessible USB port on the SCU, as shown in Figure 3-8.

Figure 3-8 Rear view of the SCU showing the USB port for connection

Figure 3-9 shows the front view of the SCU with the system USB port and the Flexible Service Processor (FSP) USB ports highlighted.

Figure 3-9 Front view of SCU showing the System USB port and FSP USB ports

Feature Code EUA5: Stand-alone USB DVD drive with cable is supported and available for the Power E1180 system.

Important: RDX USB external docking stations are not supported on IBM Power E1180 systems.

3.2 Enhancing I/O scalability with expansion drawers

Adding I/O drawers to a Power E1180 server enhances the system's scalability, flexibility, and performance in enterprise environments. The Power E1180 supports PCIe Gen4 I/O expansion drawers, which provide increased throughput and bandwidth for connected devices. These drawers enable the system to handle demanding workloads and a broader range of peripheral devices.

I/O drawers expand the available PCIe slots beyond what is natively supported on the system board. This expansion enables you to integrate more network adapters, storage controllers, and accelerators without compromising existing configurations. This capability is especially beneficial for data-intensive applications, such as artificial intelligence (AI) inferencing, high-speed networking, and large-scale database operations. The modular nature of I/O expansion supports a pay-as-you-grow model, aligning with dynamic business needs and reducing upfront capital costs.

By increasing the number and type of adapters that you can install, these drawers enable more flexible system configurations. This flexibility is useful in virtualized environments where multiple VMs or containers require dedicated I/O resources. The additional capacity supports higher VM density and more granular resource allocation, improving overall system usage. IBM expansion drawers include enterprise-grade reliability features, such as hot-plug capabilities, redundant paths, and integration with PowerVM and the Hardware Management Console (HMC). These features enable you to add or service I/O resources without system downtime, supporting continuous operations and minimizing the risk of service disruptions.

The I/O expansion capability aligns with the Power E1180 design goals of resilience, security, and performance. The system architecture, including its I/O subsystem, supports mission-critical workloads with features such as redundant paths, error recovery, and Secure Boot. The drawers integrate seamlessly with IBM HMC and PowerVM virtualization, providing centralized control and simplified management. This integration makes the Power E1180 with I/O drawers an ideal platform for enterprises that want to modernize their infrastructure while maintaining high availability (HA) and robust performance.

3.2.1 PCIe expansion card

The PCIe expansion function is enabled by using the PCIe Gen4 cable adapter (Feature Code EJ24). Each adapter occupies one PCIe Gen5 slot in the system node. You can install up to eight adapters in a system node and up to 32 adapters in a system with four system nodes. Each card connects to either one Feature Code ENZ0 fanout module or one NED24 Enclosure Services Manager (ESM). The cable adapter can be installed in any slot in the system node. For recommended adapter locations based on the expansion drawer being attached, see Table 3-6 on page 81.

Figure 3-10 shows the PCIe Gen4 cable adapter (Feature Code EJ24).

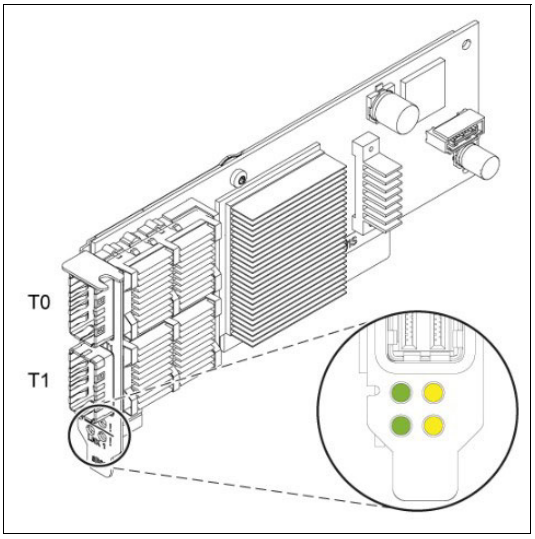

Figure 3-10 PCIe Gen4 cable adapter with Feature Code EJ24

The EJ24 card (Feature Code EJ24, CCIN 6B99) is a PCIe Gen4 x16 cable adapter that provides high-speed connectivity between the system and supported I/O expansion drawers. It is a single-wide, low-profile PCIe Gen4 cable adapter with two ports for connecting expansion drawer cables. The adapter can be installed in either PCIe Gen4 x16 or x8 slots. Installing the adapter in x8 slots allows for greater flexibility in connecting more drawers but delivers reduced performance compared to installation in full x16 slots.

This adapter supports the attachment of the following components:

- One ESM in an NED24 NVMe expansion drawer
- One PCIe Gen4 six-slot fanout module in an ENZ0 PCIe Gen4 expansion drawer

The EJ24 adapter extends the PCIe bus from the Power E1180 system to external I/O resources, enabling scalable configurations for high-performance workloads. It includes two status LEDs: a green LED that indicates link status and an amber LED that identifies the adapter during maintenance activities.

Each EJ24 adapter supports two CXP interface ports, which can use either optical or copper cables to connect to PCIe Gen4 expansion drawers. A single adapter connects to one fanout module within an expansion drawer. To use both fanout modules in a drawer, you need two EJ24 adapters. Connecting an NED24 expansion drawer also requires two adapters for full connectivity.

The EJ24 includes a built-in PCIe switch that extends the Power E1180 internal PCIe Gen4 bus to external I/O drawers while maintaining bandwidth and performance integrity.

Figure 3-11 illustrates the adapter's physical layout and connectivity.

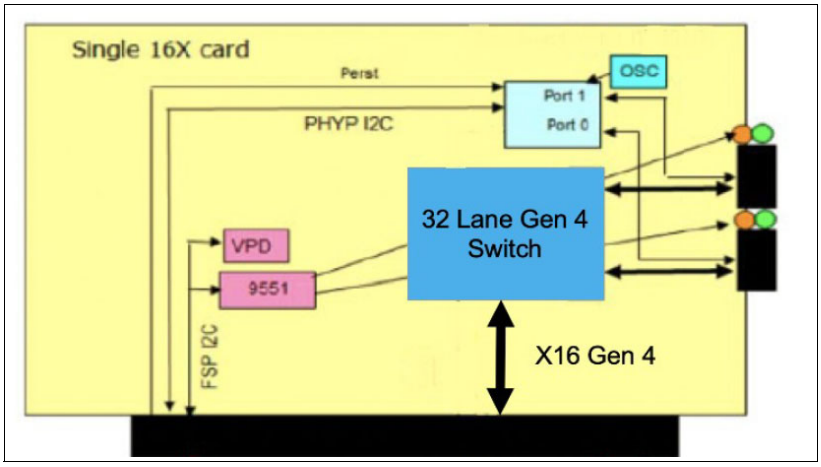

Figure 3-11 PCIe Expansion card

Table 3-5 lists the number of supported cable card adapters for the Power E1180 system.

Table 3-5 Supported cable card adapters for Power E1180

System name	Adapter Feature Code and CCIN
Power E1180	PCIe4 cable adapter (Feature Code EJ24; CCIN 6B99)

Cables: copper or active optical cable optical

The choice of copper or optical cables depends on your specific configuration and the required cable length. Use active optical cables (AOCs) for longer and thinner connections, such as Feature Code ECLX (3 m optical) and Feature Code ECLY (10 m optical). Each cable Feature Code includes two cables.

Active optical cables

The AOC consists of a fiber cable and two active electrical-to-optical converter modules that are combined into one assembly, also known as CXP converters. The AOC has a minimum bend radius of 1 in. (25 mm).

Figure 3-12 shows the AOC connector.

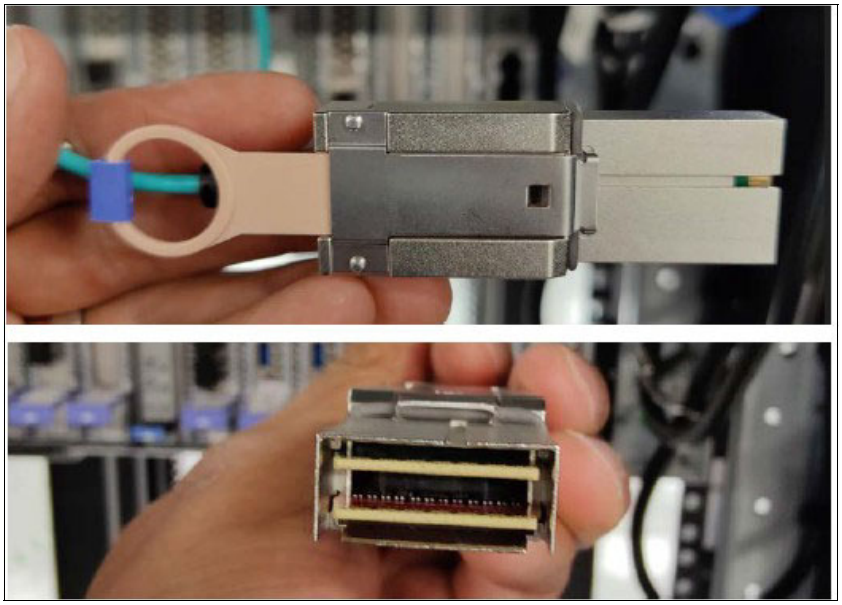

Figure 3-12 AOC connector

The following cables are available:

- Feature Code ECLS: 3.0 m CXP x16 copper cable pair for the PCIe Gen4 expansion drawer

  This 3.0 m cable pair connects a PCIe Gen4 fanout module in the PCIe Gen4 I/O expansion drawer to a PCIe Gen4 optical converter adapter in the system unit. The pair includes two identical copper cables, each with two CXP connectors. One cable connects the top CXP ports of the fanout module and the optical converter adapter. The other cable connects the bottom CXP ports.

- Feature Code ECLX: 3.0 m active optical cable (AOC) x16 pair for PCIe Gen4 expansion drawer

  This 3.0 m AOC pair connects a PCIe Gen4 module in the PCIe Gen4 I/O expansion drawer to a PCIe Gen4 optical converter adapter in the system unit. The pair includes two identical cables, each with two CXP connectors. One cable connects the top CXP ports, and the other connects the bottom CXP ports.

- Feature Code ECLY: 10 m active optical cable (AOC) x16 pair for PCIe Gen4 expansion drawer

  This 10 m AOC pair connects a PCIe Gen4 module in the PCIe Gen4 I/O expansion drawer to a PCIe Gen4 optical converter adapter in the system unit. The pair includes two identical cables, each with two CXP connectors. One cable connects the top CXP ports, and the other connects the bottom CXP ports.
- Feature Code ECLZ: 20 m active optical cable (AOC) x16 pair for PCIe Gen4 expansion drawer

  This 20 m AOC pair connects a PCIe Gen4 module in the PCIe Gen4 I/O expansion drawer to a PCIe Gen4 optical converter adapter in the system unit. The pair includes two identical cables, each with two CXP connectors. One cable connects the top CXP ports, and the other connects the bottom CXP ports.

Note: Consider the following points:

- Use 3-m cables for intra-rack installations.
- Use 10-m cables for inter-rack installations.
- Do not mix copper and optical cables on the same PCIe Gen4 I/O drawer. Both fanout modules must use either copper cables or optical cables.

A PCIe Gen4 I/O expansion drawer with two I/O fanout modules connects to one host system node by using two PCIe Gen4 cable adapters and four expansion drawer cables (two cable pairs). Each cable pair connects to one of the PCIe Gen4 six-slot fanout modules.

Connectivity to PCIe expansion drawer

Figure 3-13 illustrates the connection of two expansion drawer cable pairs for two PCIe4 6-slot fanout modules.

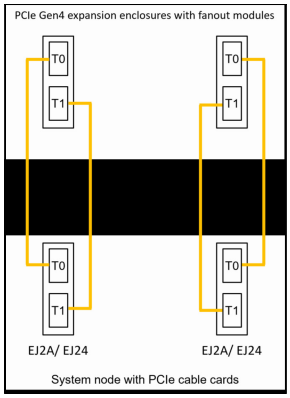

Figure 3-13 Cabling setup for PCIe Gen4 expansion drawer

Connectivity to NED24 expansion drawer

Figure 3-14 shows the connectivity to the NED24 expansion drawer. Both connections to the drawer must be populated and from the same system.

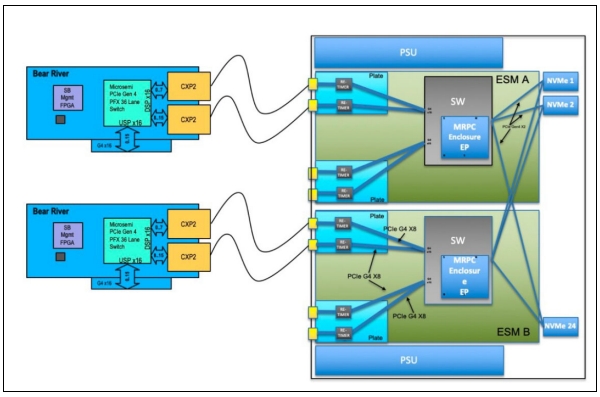

Figure 3-14 Connectivity to NED24 drawers

3.2.2 Supported I/O drawers

The Power E1180 supports the following I/O expansion drawers:

- PCIe Gen4 I/O expansion drawer (Feature Code ENZ0)

  The PCIe Gen4 I/O expansion drawer is a 4U-high, 19-inch-wide, rack-mountable I/O drawer based on PCIe Gen4 technology. It is available as a feature of Power10 servers and replaces the PCIe Gen3 I/O expansion drawer (Feature Code EMX0). There is no upgrade path from the PCIe Gen3 I/O expansion drawer to the PCIe Gen4 I/O expansion drawer.

- NED24 NVMe expansion drawer (Feature Code ESR0)

  The NED24 NVMe expansion drawer is a storage expansion enclosure with 24 U.2 NVMe bays. It supports up to 24 U.2 NVMe devices in 15 mm Gen3 carriers. The 15 mm carriers can accommodate either 7 mm or 15 mm NVMe devices.

Within the Power E1180 server, PCIe slots are used to install the PCIe x16 to CXP converter card (Feature Code EJ24) to attach expansion drawers. Two PCIe x16 to CXP converter cards are required for each expansion drawer. The maximum number of expansion drawers depends on the number of system nodes and the number of interconnect cards installed.

Table 3-6 shows the maximum supported configurations.

Table 3-6 Maximum I/O drawer configuration

Configuration	Max I/O drawers	Max NED24	Max ENZ0	Comments
Power E1180 1-system node	4	2[a]	4[b]	Slots C3 and C4 are Gen 5 x8[c]
Power E1180 2-system node	8	4[a]	8[b]	Slots C3 and C4 are Gen 5 x8[c]
Power E1180 3-system node	12	6[a]	12[b]	Slots C3 and C4 are Gen 5 x8[c]
Power E1180 4-system node	16	8[a]	16[b]	Slots C3 and C4 are Gen 5 x8[c]

a. Power E1180 slots for NED24: C2/C5, C0/C6, or C1/C7.
b. Power E1180 slots for ENZ0: C1/C7, C0/C6, C2/C5, C3/C4.
c. Using the x8 slots yields performance that is similar to the performance of the EMX0 PCIe3 expansion drawer. Using x16 slots yields improved performance.

3.2.3 Non-supported drawers

With the Power E1180 system, the following drawers are no longer supported:

- The PCIe Gen3 I/O expansion drawer (Feature Code EMX0) is not supported on Power11 systems. Fanout modules (Feature Code EMXF, Feature Code EMXG, and Feature Code EMXH) are also not supported in the PCIe Gen3 I/O expansion drawer (Feature Code ENZ0).
- The EXP24SX SAS storage enclosure (Feature Code ESLS and Feature Code ESLL) is not supported on Power11 systems. For extra internal storage, use the NED24 NVMe expansion drawer (Feature Code ESR0), which supports up to 24 NVMe drives.

3.2.4 PCIe Gen 4 I/O expansion drawer

The PCIe Gen4 I/O expansion drawer (Feature Code ENZ0) is a 4U-high, 19-inch-wide, rack-mountable I/O drawer that is available as a feature of Power11 servers. When configured with two PCIe fanout modules (Feature Code ENZF), the drawer provides 12 full-length, full-height PCIe Gen4 slots. Each fanout module provides six PCIe slots labeled C0 - C5. Slots C0 - C3 are x16, and slots C4 and C5 are x8. PCIe Gen1, Gen2, and Gen3 full-height adapters are also supported.

Important: The PCIe Gen3 I/O expansion drawer (Feature Code EMX0) is not supported on Power11 systems. Previous-generation fanout modules (Feature Code EMXF, Feature Code EMXG, and Feature Code EMXH) are also not supported in the PCIe Gen4 I/O expansion drawer (Feature Code ENZ0).

A blind swap cassette (BSC) houses the full-height adapters that are installed in these slots. The drawer has a full set of BSCs. A PCIe CXP converter adapter (Feature Code EJ24) occupies one PCIe Gen5 slot in the system node, and a pair of active optical or copper cables connects the system node to the fanout module. Both fanout modules are independent PCIe domains and can be serviced separately. A minimum of one fanout module is required in the Feature Code ENZ0 drawer, which is installed in location P0 on the left side of the drawer when viewed from the rear. Each PCIe Gen4 I/O expansion drawer includes two power supplies.

You can add drawers to a server dynamically. Concurrent repair and the addition or removal of expansion drawers and PCIe adapters are supported through HMC-guided menus or operating system utilities. Careful I/O balancing, assigning adapters across redundant Feature Code ENZ0 expansion drawers, and connecting to different system nodes provide HA for I/O resources that are assigned to LPARs.

Figure 3-15 shows the front view of a PCIe Gen 4 I/O expansion drawer.

Figure 3-15 PCIe Gen 4 I/O expansion drawer front view

Figure 3-16 shows the rear view of a PCIe Gen 4 I/O expansion drawer.

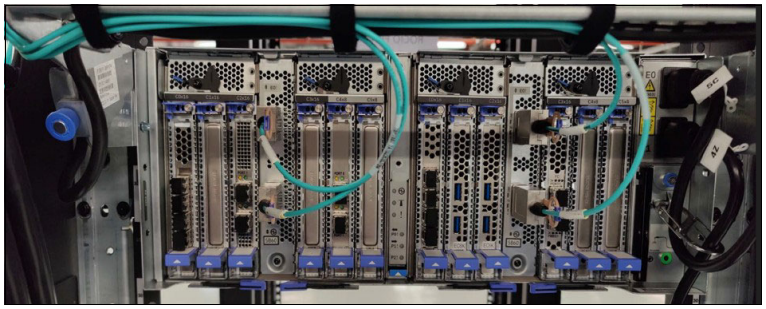

Figure 3-16 PCIe Gen 4 I/O expansion drawer rear view

Table 3-7 shows the PCIe4 expansion drawer dimensions and weight specifications.

Table 3-7 PCIe4 expansion drawer dimensions and weight

Description	Value
Width	Drawer: 44.45 cm (17.5 in) Rack and bezel: 48.2 cm (19 in)
Height	4U (17.3 cm or 6.8 in)
Depth	73.66 cm (29 in)
Weight (maximum configuration)	55 kg (121.2 lb)

Supported fanout modules

Each PCIe Gen4 I/O expansion drawer supports one or two fanout modules. These modules must connect to a PCIe expansion adapter in the system to extend the PCIe bus to the drawer. The number of supported fanout modules varies by system type and configuration, which directly affects how many PCIe Gen4 drawers you can connect.

For a detailed breakdown of the maximum number of supported fanout modules and total cable requirements for the Power E1180 server, see Table 3-8.

Table 3-8 Number of supported fanout modules per system

System name	Maximum fanout modules per node	Maximum number of I/O expansion drawers
Power E1180	8	4

PCI slots that are available in the PCIe Gen4 expansion drawer

Table 3-9 lists the PCI slots in the PCIe Gen4 I/O expansion drawer that is equipped with two PCIe4 6-slot fan-out modules.

Table 3-9 PCIe slot configuration in the PCIe Gen4 I/O expansion drawer

Slot	Location code	Description
Left I/O module Slot 0	P0-C0	PCIe x16 adapter
Left I/O module Slot 1	P0-C1	PCIe x16 adapter
Left I/O module Slot 2	P0-C2	PCIe x16 adapter
Left I/O module Slot 3	P0-C3	PCIe x16 adapter
Left I/O module Slot 4	P0-C4	PCIe x8 adapter
Left I/O module Slot 5	P0-C5	PCIe x8 adapter
Right I/O module Slot 0	P1-C0	PCIe x16 adapter
Right I/O module Slot 1	P1-C1	PCIe x16 adapter
Right I/O module Slot 2	P1-C2	PCIe x16 adapter
Right I/O module Slot 3	P1-C3	PCIe x16 adapter
Right I/O module Slot 4	P1-C4	PCIe x8 adapter
Right I/O module Slot 5	P1-C5	PCIe x8 adapter

Notes:

- All slots are PCIe Gen4 slots.
- All slots support full-length, full-height adapters or short form-factor adapters with a full-height tailstock in single-wide, generation 3 BSCs.
- Slots C0 - C3 in each PCIe Gen4 six-slot fanout module are PCIe Gen4 x16 buses, and slots C4 and C5 are PCIe Gen4 x8 buses.
- All slots support EEH.
- All PCIe slots can be serviced with the power on.
- All six slots in a PCIe Gen4 six-slot fanout module support SR-IOV shared mode.
- Only four adapters with Feature Code EC2S, Feature Code EC2U, or Feature Code EC72 adapters can operate in SR-IOV mode simultaneously per six-slot fanout module.

Table 3-10 lists the maximum number of Feature Code ENZ0 drawers, fanout modules, and PCIe slots that are supported based on the number of nodes.

Table 3-10 Maximum number of drawers, fanout modules, and PCIe slots, and their type

System node	ENZ0 drawers	Feature Code ENZF0 fanout module	Total Number of PCIe slots	x16 Gen4 slots	x8 Gen4 slots
1	4	8	48	32	16
2	8	16	96	64	32
3	12	24	144	96	48
4	16	32	192	128	64

3.2.5 NED24 NVMe expansion drawer

IBM continues to provide industry-leading I/O capabilities with a PCIe direct-attached expansion drawer that supports NVMe drive attachment. The NED24 NVMe expansion drawer (Feature Code ESR0) is a storage expansion enclosure with 24 U.2 NVMe bays.

Important: The EXP24SX SAS Storage Enclosure is no longer supported by Power11.

Figure 3-17 shows the front view of a NED24 drawer with 4 NVMe Drives at slots C8 - C11.

Figure 3-17 NED24 NVMe drawer front view

Figure 3-18 shows the rear view of the NED24 drawer with two power supplies and two ESMs.

Figure 3-18 NED24 NVMe drawer rear view

Each of the 24 NVMe bays in the NED24 drawer is individually addressable and can be assigned to a specific LPAR or VIOS to provide native boot support for up to 24 partitions. At the time of writing, each drawer supports up to 367 TB when populated with 15.3 TB flash drives.

The devices that are listed in Table 3-11 are supported in the NED24 drawer at the time of writing.

Note: Both ESMs must connect to the same server. Single connections and multiple server connections are not supported.

Table 3-11 NVMe devices that are supported in the NED24

Feature Code	Description
ES3A	Enterprise 800 GB SSD PCIe4 NVMe U.2 module for IBM i
ES3H	Enterprise 800 GB SSD PCIe4 NVMe U.2 module for IBM i
ES3B	Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module for AIX and Linux
ES3C	Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module for IBM i
ES3D	Enterprise 3.2 TB SSD PCIe4 NVMe U.2 module for AIX and Linux
ES3E	Enterprise 3.2 TB SSD PCIe4 NVMe U.2 module for IBM i
ES3F	Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module for AIX and Linux
ES3G	Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module for IBM i
ECT9	15.3 TB Mainstream NVMe U.2 SSD 4k for AIX and Linux

Each U.2 NVMe device is installed in the NED24 drawer by using 15 mm carriers. The 15 mm carriers can accommodate either 7 mm or 15 mm NVMe devices.

The NED24 drawer is supported on the Power E1180 system by using the same interconnect card as the PCIe Gen4 I/O expansion drawer. You can install a maximum of two NED24 NVMe expansion drawers per system node and up to eight drawers in a single system. When mixing different expansion drawers, the maximum number of supported drawers depends on the number of EJ24 fanout cards that are installed.

Table 3-12 shows the number of available expansion drawers for the different node configurations.

Table 3-12 Maximum expansion drawers

Configuration	Max I/O drawers	Max NED24 drawers	Max PCIe Gen 4 expansion drawers
Single node	4	2	4
Two nodes	8	4	8
Three nodes	12	6	12
Four nodes	16	8	16

Note:

- Slots for NED24 drawer attachment: C2/C5, C0/C6, or C1/C7

  Each NED24 drawer requires two slots. Single-attached drawers or drawers that are shared between systems are not supported.

- Slots for PCIe Gen4 expansion: C1/C7, C0/C6, C2/C5, C3/C4

  Slots C3 and C4 are PCIe Gen5 x8. Using these slots provides performance similar to PCIe Gen3 connectivity. Installing adapters in x16 slots improves performance.

Multipath support

The NED24 drawer offers two connectivity options for NVMe drives. In Mode 1 (single connect), the NVMe drives are configured as single-path devices, with only one ESM controlling each device (see Figure 3-19). The switch in each ESM is configured to logically manage 12 of the 24 NVMe drives. In this mode, no device failover capability is available.

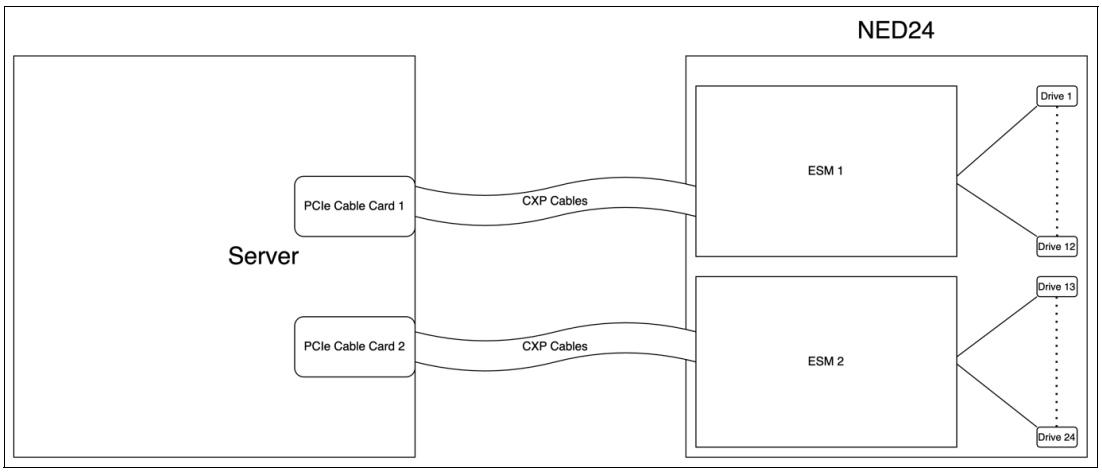

Figure 3-19 Single path mode

When running in Mode 1, OS-level mirroring is a best practice to avoid a single point of failure in the connection to the drives in the NED24 enclosures. For more information about drive locations within the drawer to improve availability and reliability, see “Drive installation order” on page 87. Both ESMs must connect to the same server. Single connections and multiple server connections are not supported.

The Power E1180 also supports the NED24 drawer in Mode 2 (multipath mode). In this mode, each drive has two connections because each port on the multiport drives connects through both ESMs (see Figure 3-20). This configuration provides more reliability, availability, and serviceability (RAS) and improves performance.

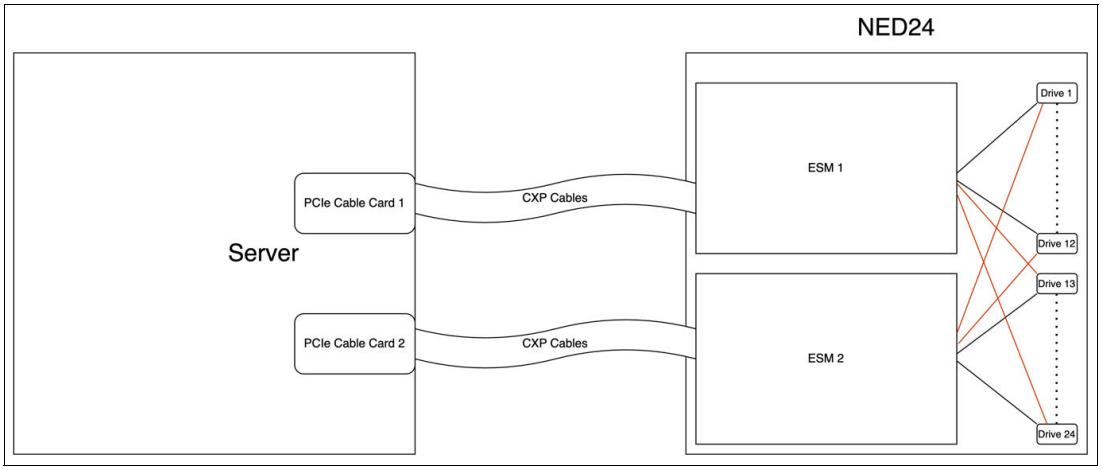

Figure 3-20 Multipath support

Multipath is automatically enabled when the appropriate OS level is installed. Multipath support is provided in the OSs that are shown in Table 3-13.

Important: Both ESMs must be connected to the same server. Single connections and multiple server connections are not supported.

Table 3-13 Multipath support by operating system

OS	Supported releases
AIX	► AIX 7.3 with the 7300-02 Technology Level (TL) and Service Pack (SP) 7300-02-02-2420 or later ► AIX 7.2 with the 7200-05 TL and SP 7200-05-08-2420 or later ► AIX 7.3 with the 7300-01 TL and SP 7300-01-04-2420 or later
IBM i	► IBM i 7.6, IBM i 7.5 Technology Refresh (TR) 4, and IBM i 7.4 TR 10
Linux	► SUSE Linux Enterprise Server 15.5 or later ► Red Hat Enterprise Linux (RHEL) 9.2 or later

Each NVMe U.2 drive in the NED24 NVMe expansion drawer includes a “-R1” or “-R2” suffix at the end of its physical location code. Both drive paths must be assigned to a single LPAR by using the HMC LPAR Physical I/O Adapters page or through the partition profile. The Partner Location Code column identifies the other path to the NVMe U.2 drive in a multipath configuration.

Drive installation order

There is no performance difference for drives in any of the NED24 slots. However, a recommended installation order exists for drives within the enclosure. This order helps maintain separation for the suggested mirroring between drives and helps ensure optimal cooling and airflow.

Figure 3-21 shows the recommended placement for the first four drives. Table 3-14 shows the recommended placement for all drives.

Figure 3-21 Drive installation order

Table 3-14 Recommended drive installation order

Drive pair	First drive slot	Second drive slot
1	1	13
2	7	19
3	2	14
4	8	20
5	3	15
6	9	21
7	4	16
8	10	22
9	5	17
10	11	23
11	6	18
12	12	24

3.2.6 General rules for I/O drawers connection

The PCIe Gen4 cable adapter (Feature Code EJ24) can be installed in any PCIe adapter slot in a Power E1180 system node. However, it is a best practice to populate the PCIe adapter slots in the order that is described in 3.2.2, “Supported I/O drawers” on page 80. If an I/O expansion drawer is present, give the PCIe Gen4 cable adapter (Feature Code EJ24) the highest priority.

Each processor module supports two PCIe Gen5 slots, and all slots have the same bandwidth characteristics. By following the slot priorities, you help ensure that one PCIe Gen5 slot per processor module is populated before you use the second PCIe Gen5 slot of each module.

Use 2-meter or 3-meter cables for intra-rack installations and 10-meter or 20-meter cables for inter-rack installations.

The correct cabling is as follows:

- Connect the top port (T0) of the EJ24 adapter to the top port (T0) of the fanout module of the PCIe Gen4 expansion drawer or the top port of an ESM.
- Connect the bottom port (T1) of the EJ24 adapter to the bottom port (T1) of the fanout module of the PCIe Gen4 expansion drawer or the bottom port of an ESM.

Important: The two cables that connect a host system node's PCIe Gen4 cable adapter to the PCIe Gen4 I/O expansion drawer's I/O module or NED24 ESM must be the same type and length.

Figure 3-22 shows the connection between a fully populated ENZ0 drawer and two EJ24 cards.

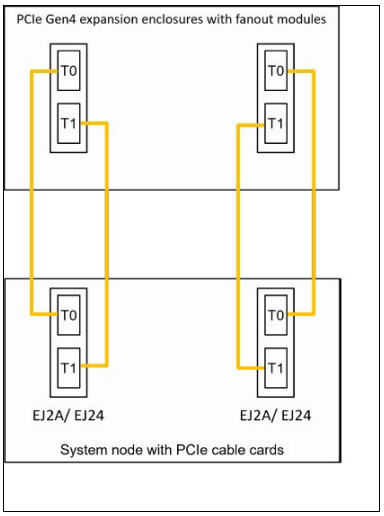

Figure 3-22 Connecting the expansion drawer cables from two cards to two fanout modules

Figure 3-23 shows the connectivity between two EJ24 cards in the system node and the two ESMs in a NED24 drawer. Both ESMs must be connected to the same system node.

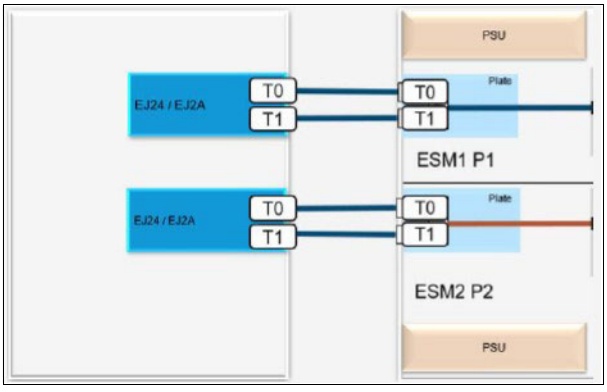

Figure 3-23 NED24 connectivity to the system node

3.3 Supported adapters

This section provides more information about supported PCIe I/O adapters on the Power E1180 server at general availability. This list is updated frequently as more PCIe I/O adapters are tested and certified.

For more information about supported storage combinations, see the IBM System Storage Interoperation Center (SSIC).

In addition, the Power E1180 sales manual provides a complete list of all supported devices and their Feature Codes.

3.3.1 Supported Fibre Channel adapters

Table 3-15 lists the supported Fibre Channel (FC) adapters.

Table 3-15 Supported Fibre Channel adapters on IBM Power E1180

Feature Code	CCIN	Description	Operating system (OS) support	Supported in ENZ0
EN1A	578F	PCIe3 x8 2-port Fibre Channel (32 Gbps) full-height	IBM i, AIX, VIOS, and Linux	Yes
EN1B	578F	PCIe3 x8 2-port Fibre Channel (32 Gbps) low-profile	IBM i, AIX, VIOS, and Linux	No
EN1J	579C	PCIe4 x8 2-port Fibre Channel (32 Gbps) full-height	IBM i, AIX, VIOS, and Linux	Yes
EN1K	579C	PCIe4 x8 2-port Fibre Channel (32 Gbps) low-profile	IBM i, AIX, VIOS, and Linux	No
EN1L	2CFC	PCIe4 x8 4-port Fibre Channel (32 Gbps) full-height	IBM i, AIX, VIOS, and Linux	Yes
EN1M	2CFC	PCIe4 x8 4-port Fibre Channel (32 Gbps) low-profile	IBM i, AIX, VIOS, and Linux	No
EN1N	2CFD	PCIe4 x8 2-port Fibre Channel (64 Gbps) full-height	IBM i, AIX, VIOS, and Linux	Yes
EN1P	2CFD	PCIe4 x8 2-port Fibre Channel (64 Gbps) low-profile	IBM i, AIX, VIOS, and Linux	No
EN2L	2F06	PCIe4 x16 4-port Fibre Channel (32 Gbps) full-height	IBM i, AIX, VIOS, and Linux	Yes
EN2M	2F06	PCIe4 x16 4-port Fibre Channel (32 Gbps) low-profile	IBM i, AIX, VIOS, and Linux	No
EN2N	2F05	PCIe4 x8 2-port Fibre Channel (64 Gbps) full-height	IBM i, AIX, VIOS, and Linux	Yes
EN2P	2F05	PCIe4 x8 2-port Fibre Channel (64 Gbps) low-profile	IBM i, AIX, VIOS, and Linux	No

In addition to Table 3-15 on page 90, the following resources supply information about which adapters support replaceable SFPs and which adapters do not:

- Transceiver component (SFP) *not* replaceable for IBM Power FC and Ethernet adapter features
- Transceiver component (SFP) is replaceable for following IBM Power FC and Ethernet adapter features

3.3.2 Network and RoCE adapters

Table 3-16 list the supported network adapters.

Table 3-16 Supported Network adapters on IBM Power E1180

Feature Code	CCIN	Description	OS support	Supported in ENZ0
EC2T	58FB	PCIe3 2-port 25/10 Gb NIC & RoCE SFP28 adapter low-profile	► NIC: IBM i, AIX, VIOS, and Linux. ► RoCe: AIX and Linux.	No
EC2U	58FB	PCIe3 2-port 25/10 Gb NIC & RoCE SFP28 adapter full-height	► NIC: IBM i, AIX, VIOS, and Linux. ► RoCe: AIX and Linux.	Yes
EC71	2CF9	PCIe4 2-port 25/10/1 Gb RoCE SFP28 adapter low-profile	IBM i, AIX, VIOS, and Linux.	No
EC72	2CF9	PCIe4 2-port 25/10/1 Gb RoCE SFP28 adapter full-height	IBM i, AIX, VIOS, and Linux.	Yes
EC73	2CF8	PCIe4 2-port 25/10/1 Gb RoCE SFP28 with Crypto adapter low-profile	At the time of writing, this information is N/A.	No
EC74	2CF8	PCIe4 2-port 25/10/1 Gb RoCE SFP28 with Crypto adapter full-height	At the time of writing, this information is N/A.	Yes
EC75	2CFB	PCIe4 2-port 100 GbE RoCE x16 adapter low-profile	IBM i, AIX, VIOS, and Linux.	No
EC76	2CFB	PCIe4 2-port 100 GbE RoCE x16 adapter full-height	IBM i, AIX, VIOS, and Linux.	Yes
EC85	EC2C	PCIe5 x16 2-port 200 GbE RoCE adapter	At the time of writing, this information is N/A.	No
EN24	EC2A	PCIe4 x16 4-port 25/10/1 GbE RoCE SFP28 adapter low-profile	► IBM i, AIX, VIOS, and Linux. ► RoCE function available only on IBM i.	No

Feature Code	CCIN	Description	OS support	Supported in ENZ0
EN26	EC2A	PCIe4 x16 4-port 25/10/1 GbE RoCE SFP28 adapter full-height	► IBM i, AIX, VIOS, and Linux. ► RoCe function available only on IBM i.	Yes
EN2W	2F04	PCIe3 4-port 10 GbE adapter full-height	IBM i (through VIOS), AIX, VIOS, and Linux.	Yes
EN2X	2F04	PCIe3 4-port 10 GbE adapter low-profile	IBM i (through VIOS), AIX, VIOS, and Linux.	No

3.3.3 Network and RoCE transceivers and cables

Table 3-17 provides information about all SFPs and cables that are used in NIC and RoCE adapters for Power11 systems.

Table 3-17 Supported transceivers and cables

Feature Code	Name	Connection	Speed in Gbps	Adapter
EB48	1 Gb RJ45 Transceiver	SFP and RJ45	1	► EC2T/U ► EC71/2/3/4 ► EN24/6
EB46	10 GbE Optical Transceiver SR	SFP+ and LC MMF	10	► EC2T/U ► EC71/2/3/4 ► EN24/6
EB47	25 GbE Optical Transceiver	SFP28 and LC MMF	25	► EC2T/U ► EC71/2/3/4 ► EN24/6
EB57	40 GbE Base-SR4 Transceiver	QSFP+ and MPO MMF	40	► EC75/6 ► EC85/6
EB59	100 GbE Base-SR4 Transceiver	QSFP28 and MPO MMF	100	► EC75/6 ► EC85/6
EB49	QSFP28 to SFP28 Connector	QSFP28	25	► EC75/6 ► EC85/6
EB4J	0.5M 25 GbE Passive CU Cable SFP+	SFP+	25	► EC2T/U ► EC71/2/3/4 ► EN24/6
EB4K	1M 25 GbE Passive CU Cable SFP+	SFP+	25	► EC2T/U ► EC71/2/3/4 ► EN24/6
EB4M	2M 25 GbE Passive CU Cable SFP+	SFP+	25	► EC2T/U ► EC71/2/3/4 ► EN24/6

Feature Code	Name	Connection	Speed in Gbps	Adapter
EB4P	QSFP28 Passive Copper 100 GbE to SFP28 4x25 GbE Split Cable	QSFP29 to SFP28	25	► EC2T/U ► EC71/2/3/4 ► EN24/6
EN01	1M 10 GbE CU Active Twinax Cable SFP+	SFP+	10	► EC2T/U ► EC71/2/3/4 ► EN24/6
EN02	3M 10 GbE CU Active Twinax Cable SFP+	SFP+	10	► EC2T/U ► EC71/2/3/4 ► EN24/6
EN03	5M 10 GbE CU Active Twinax Cable SFP+	SFP+	10	► EC2T/U ► EC71/2/3/4 ► EN24/6
EB5K	1M 100 GbE Passive CU Cable QSFP28	QSFP28	100	► EC75/6 ► EC85/6
EB5L	1.5M 100 GbE Passive CU Cable QSFP28	QSFP28	100	► EC75/6 ► EC85/6
EB5M	2M 100 GbE Passive CU Cable QSFP28	QSFP28	100	► EC75/6 ► EC85/6
EB5R	3M 100 GbE Active Optic Cable QSFP28	QSFP28	100	► EC75/6 ► EC85/6
EB5S	5M 100 GbE Active Optic Cable QSFP28	QSFP28	100	► EC75/6 ► EC85/6
EB5T	10M 100 GbE Active Optic Cable QSFP28	QSFP28	100	► EC75/6 ► EC85/6
EB5U	15M 100 GbE Active Optic Cable QSFP28	QSFP28	100	► EC75/6 ► EC85/6
EB5V	20M 100 GbE Active Optic Cable QSFP28	QSFP28	100	► EC75/6 ► EC85/6
EB5W	30M 100 GbE Active Optic Cable QSFP28	QSFP28	100	► EC75/6 ► EC85/6
EB5X	50M 100 GbE Active Optic Cable QSFP28	QSFP28	100	► EC75/6 ► EC85/6

3.3.4 SAS and tape adapters

Table 3-18 lists the supported SAS adapters.

Table 3-18 Supported SAS adapters on IBM Power E1180

Feature Code	CCIN	Description	OS support	Supported in ENZ0
EJ2B	57F2	PCIe3 x8 SAS quad-port 12 Gb tape full-height adapter	IBM i, AIX, and VIOS	Yes
EJ2C	57F2	PCIe3 x8 SAS quad-port 12 Gb tape low-profile adapter	IBM i, AIX, and VIOS	No

3.3.5 NVMe

Table 3-19 lists the supported NVMe drives.

Table 3-19 Supported NVMe adapters on IBM Power E1180

Feature Code	CCIN	Description	OS support	Placement
EC5J	59B4	Mainstream 800 GB SSD NVMe U.2 module	AIX and Linux	System Node
EC7Q	5948	800 GB Mainstream NVMe U.2 SSD 4k module	AIX and Linux	System Node
ECT9	5941	15.3 TB Mainstream NVMe U.2 SSD 4k module	AIX and Linux	System Node and NED24
ECTB	5941	15.3 TB Mainstream NVMe U.2 SSD 4k module	IBM i	System Node
EKF2	5B53	Enterprise 800 GB SSD PCIe4 NVMe U.2 module	IBM i	NED24
EKF3	5B52	Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
EKF4	5B52	Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	IBM i	NED24
EKF5	5B51	Enterprise 3.2 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
EKF6	5B51	Enterprise 3.2 TB SSD PCIe4 NVMe U.2 module	IBM i	NED24
EKF7	5B50	Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
EKF8	5B50	Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module	IBM i	NED24
EKF9		Enterprise 800 GB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
ES3A	5B53	Enterprise 800 GB SSD PCIe4 NVMe U.2 module	IBM i	NED24
ES3B	5B52	Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
ES3C	5B52	Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	IBM i	NED24
ES3D	5B51	Enterprise 3.2 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
ES3E	5B51	Enterprise 3.2 TB SSD PCIe4 NVMe U.2 module	IBM i	NED24
ES3F	5B50	Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
ES3G	5B50	Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module for IBM i	IBM i	NED24

Feature Code	CCIN	Description	OS support	Placement
ES3H		Enterprise 800 GB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
ES4B		Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	System Node
ES4C		Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	IBM i	System Node
ES4D		Enterprise 3.2 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	System Node
ES4E		Enterprise 3.2 TB SSD PCIe4 NVMe U.2 module	IBM i	System Node
ES4F		Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	System Node
ES4G		Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module	IBM i	System Node
ES5A	5B53	Enterprise 800 GB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
ES5B	5B53	Enterprise 800 GB SSD PCIe4 NVMe U.2 module	IBM i	NED24
ES5C	5B52	Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
ES5D	5B52	Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	IBM i	NED24
ES5E	5B51	Enterprise 3.2 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
ES5F	5B51	Enterprise 3.2 TB SSD PCIe4 NVMe U.2 module	IBM i	NED24
ES5G	5B50	Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module	AIX and Linux	NED24
ES5H	5B50	Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module	IBM i	NED24

Differences between mainstream and enterprise SSD

This section explains the differences between enterprise SSDs and mainstream SSDs (formerly known as read-intensive SSDs).

- Endurance and use case

  Enterprise SSDs are designed for high-endurance workloads, supporting up to 10 DWPD and typically using durable MLC NAND (formerly called eMLC). Mainstream SSDs usually support about 1 DWPD, making them suitable for read-heavy or less write-intensive applications.

- Over-provisioning and performance

  Enterprise SSDs include more over-provisioned NAND capacity, which helps manage garbage collection, reduces write amplification, and improves random write performance and endurance. Mainstream SSDs have less over-provisioning, which reduces cost but affects performance and lifespan during write-heavy operations.

- Cost

  Mainstream SSDs are more cost-effective because of reduced over-provisioning and endurance specifications. They offer a reduced cost per GB but reduced performance in write-intensive tasks.

- Mixing in RAID arrays

  IBM systems do not allow mixing mainstream and enterprise SSDs in RAID arrays. The data striping that is performed by the PCIe SAS adapter assumes uniform endurance across all drives.

- Monitoring and lifecycle management

  Mainstream SSDs must be monitored for end-of-life (EOL) indicators. When nearing EOL, a Predictive Failure Analysis (PFA) alert is triggered. Replace the drive promptly to avoid degraded performance or write failures.

- Warranty considerations

  Mainstream SSDs are not covered under warranty if they exceed their rated write lifecycle. Usage beyond rated endurance can result in degraded or failed write operations.

For more information, see the following resources:

- Mainstream solid-state drives
- IBM Power11 documentation

3.3.6 USB

Table 3-20 lists the supported USB adapters.

Table 3-20 Supported USB adapters

Feature Code	CCIN	Description	OS support	Supported in ENZ0
EC6J	590F	PCIe2 2-Port USB 3.0 Adapter low-profile	IBM i, AIX, VIOS, and Linux	No
EC6K	590F	PCIe2 2-Port USB 3.0 Adapter full-height	IBM i, AIX, VIOS, and Linux	Yes

3.3.7 Cryptographic adapters

Table 3-21 lists the supported cryptographic adapters.

Table 3-21 Supported cryptographic adapters

Feature Code	CCIN	Description	OS support	Supported in ENZ0
EJ37	4769	4769 Cryptographic Coprocessor full-height	IBM i, AIX, and Linux	Yes
EPG4	C138	4770 Cryptographic Coprocessor low-profile	IBM i, AIX, and Linux	No
EPG6	C138	4770 Cryptographic Coprocessor full-height	IBM i, AIX, and Linux	Yes

3.4 Other device support

This section provides an overview of the available media features on the Power E1180 server.

IBM System Storage 7226 Model 1U3 Multi-Media Enclosure

The IBM System Storage 7226 Model 1U3 Multi-Media Enclosure can accommodate up to two tape drives or up to four DVD-RAM drives.

The IBM System Storage 7226 Multi-Media Enclosure offers a customer-replaceable unit (CRU) maintenance service to install or replace drives efficiently. Other 7226 components are also designed for CRU maintenance.

The IBM System Storage 7226 Multi-Media Enclosure is compatible with most IBM Power8, Power9, Power10, and Power11 processor-based systems that run the current levels of AIX, IBM i, and Linux operating systems.

Restriction: IBM i does not support 7226 USB devices.

Figure 3-24 shows the 7226 configured with two DVDs and one tape drive.

Figure 3-24 7226-1U3 media drawer

Here are the options that are available with the enclosure:

- Half-High LTO Ultrium 5 (3.0 TB) SAS and Fibre Channel Tape Drive
- Half-High LTO Ultrium 6 (6.25 TB) SAS and Fibre Channel Tape Drive
- Half-High LTO Ultrium 7 (15.0 TB) SAS and Fibre Channel Tape Drive
- Half-High LTO Ultrium 8 (30.0 TB) SAS and Fibre Channel Tape Drive
- DVD-RAM: 9.4 GB Slim Optical Drive with SAS and USB interface options

For more information and technical descriptions, see these sales manual pages.

Stand-alone USB DVD drive (Feature Code EUA5)

The stand-alone USB DVD drive (Feature Code EUA5) is an optional external USB DVD device. It requires high current at 5 V and must use the front USB 3.0 port on the system or a USB port on a USB 3.0 adapter.

The USB DVD drive is supported by all operating systems. To use the drive, the USB adapter must be assigned to the corresponding logical partition (LPAR).

Note: If you have a 9080-HEU system, you must use the USB port on a USB 3.0 adapter to connect the stand-alone USB DVD drive.

Note: A USB extension cable (P/N 32N1311) is included. Use this cable when there are no safe, flat spots available in the rack. The cable allows the drive to reach the floor. Alternative or extra extension cables are not supported because the total USB cable length cannot exceed 3 meters.

RDX Removable Hard Disk Cartridge

The RDX Removable Hard Disk Cartridge is not supported by Power E1180. It cannot be purchased to use with this server.

There are alternative solutions to an RDX-based backup strategy:

- Cloud-based backup services
- On-premises entry tape drives and libraries

4

Artificial intelligence support

Artificial intelligence (AI) is a cornerstone of digital transformation across industries. It enables organizations to automate processes, gain deeper insights, and deliver personalized experiences. The demand for AI-driven solutions is growing rapidly, from predictive analytics and natural language processing to computer vision and generative AI. To support these workloads effectively, enterprises need infrastructure that can handle the computational intensity and data throughput that AI requires. IBM Power11 processor-based servers play a critical role in meeting these needs.

IBM Power is purpose-built for AI workloads with its high-performance architecture, including the latest Power11 processors that feature integrated AI acceleration. These processors handle the low-precision arithmetic operations that are commonly used in AI models, boosting inferencing speed without compromising accuracy. IBM Power also supports advanced technologies such as the IBM Spyre Accelerator, a Peripheral Component Interconnect Express (PCIe)-attached AI card that is optimized for enterprise AI workloads, and Power11 on-chip AI inferencing, which enables real-time decision-making at scale. These capabilities make IBM Power processor-based servers ideal for deploying AI models in production environments where performance, reliability, and scalability are essential.

Beyond hardware, IBM Power integrates with Red Hat OpenShift Container Platform AI, enabling organizations to build, train, and deploy AI models in a containerized, hybrid cloud environment. This combination supports consistent DevOps practices, efficient resource usage, and simplified management of AI workflows. With support for popular AI frameworks such as PyTorch, TensorFlow, and vLLM, IBM Power provides a complete, enterprise-ready platform for using AI. Whether running AI at the edge, in the data center, or across hybrid cloud environments, IBM Power delivers the performance and flexibility that are needed to turn AI potential into real-world impact.

The following topics are covered in this chapter:

- On-chip support
- AI Acceleration with the IBM Spyre Accelerator
- Solutions

© Copyright IBM Corp. 2025.

4.1 On-chip support

With the upcoming availability of the IBM Spyre Accelerator on IBM Power processor-based servers, starting with systems that are powered by the new IBM Power11 processor, organizations can use a new class of hardware acceleration to support a wide spectrum of enterprise workloads. The IBM Spyre Accelerator delivers performance benefits, particularly for compute-intensive and AI-driven use cases, while aligning with the architectural strengths of the IBM Power platform. The IBM Spyre Accelerator is not a GPU. It is a dedicated hardware device that is developed by IBM Research® to provide superior acceleration.

A common misconception among IT decision-makers is that one or more GPUs are always required for any workload involving AI. This perception often leads to infrastructure decisions that prioritize GPU integration by default, regardless of the nature or scope of the AI tasks. However, GPUs are not a prerequisite for all AI use cases. Including GPUs in every server configuration can result in disproportionate acquisition costs, elevated power consumption, and increased complexity in thermal and workload management, without benefits for the workload.

Beginning with IBM Power10 and improved with Power11, each core integrates four Matrix Math Accelerators (MMAs) that support a wide range of AI inference workloads directly on the CPU. This innovation enables you to run AI models natively on IBM Power cores without requiring a discrete GPU. The architecture is suitable for traditional AI use cases such as fraud detection, text extraction, document analysis, domain adaptation through retrieval-augmented generation (RAG), pattern recognition, forecasting, and image, video, or audio processing.

The IBM Power11 processor further enhances these capabilities. Although it does not always match the raw throughput of high-end GPUs in some generative AI scenarios, IBM Power11 technology provides excellent performance for tasks such as entity extraction, translation, summarization, and classification. It also offers reduced energy consumption and improved data protection. By enabling AI workloads to run closer to where the data is, IBM Power11 processor-based systems support secure, efficient, and scalable deployment of AI without unnecessary data movement.

When the IBM Spyre Accelerator becomes available on IBM Power11 processor-based systems, it will introduce extra acceleration capabilities that complement the on-chip features. IBM Power11 processor-based systems, with their integrated AI accelerators and future support for the IBM Spyre Accelerator, present a highly optimized, energy-aware, and secure platform for deploying AI across a broad range of use cases without a default dependency on discrete GPUs.

4.2 AI Acceleration with the IBM Spyre Accelerator

Improvements to the IBM Power11 processor core strength and system capacity boost the performance of the MMA for inferencing workloads. Adding IBM Spyre Accelerator adapters to Power11 systems provides more AI inferencing capabilities. Working together, IBM POWER processors and the IBM Spyre Accelerator enable next-generation infrastructure to scale demanding AI workloads for businesses.

The IBM Spyre Accelerator extends the capabilities of Power11 processor-based systems by offering a low-power, high-efficiency acceleration solution for workloads that require frequent memory access and streamlined data movement.

Because of the advanced virtualization and workload consolidation capabilities of IBM Power processor-based servers, Spyre based applications can be colocated with other mission-critical services within the same physical server. This approach allows AI inference engines, real-time analytics, or data preprocessing workloads that use the IBM Spyre Accelerator to run near databases or transactional systems that are hosted in logical partitions (LPARs) or containers on the same server. This architectural proximity reduces latency, improves throughput, and eliminates the extra impact that is typically associated with cross-node or cross-platform communication.

The IBM Spyre Accelerator is positioned as one of the only AI solutions that combines the following functions:

- Data privacy: Data and AI sovereignty on a reliable, trusted on-premises infrastructure.
- Skills: Ready-to-use enterprise AI services.
- Complexity: Accelerated plug-and-play AI for business workflows.

Important: Support for the IBM Spyre Accelerator on IBM Power is expected to be announced and available in late 2025. Statements regarding IBM's future direction and intent are subject to change or withdrawal without notice and represent goals only.

4.2.1 Deploying AI in the enterprise

AI is transforming enterprise operations across industries by enabling organizations to optimize decision-making, streamline processes, and deliver personalized experiences at scale. However, the path to successful AI integration in enterprise environments is complex. It requires the correct algorithms and models and a robust, secure, and scalable infrastructure that is tailored to diverse AI workloads. An AI-optimized IBM Power platform provides a unique approach to meet this challenge by offering both on-chip and off-chip acceleration capabilities for different levels of AI maturity.

Most enterprise AI adoption begins with experimentation and initial use cases such as proof-of-concept models for customer segmentation, log analysis, or anomaly detection. These workloads typically require limited computational resources and can be efficiently handled by the on-chip AI acceleration that is embedded in IBM Power11 processors, which includes Matrix Math Accelerator (MMA) engines and SIMD vector instructions that are combined with high memory bandwidth, which is a critical component for feeding data into AI models rapidly. On-chip acceleration enables real-time inference close to the data source, minimizing latency and reducing the need for extra hardware. It is particularly effective for traditional machine learning models, data warehouse analytics, and vector database operations such as RAG queries.

As enterprise AI use cases evolve to include more complex deep learning models such as image classification, video processing, and time-series forecasting, the demands on compute performance and memory bandwidth increase. At this stage, enterprises need a more flexible and powerful acceleration option, particularly as workloads move from experimentation to production.

4.2.2 Fit-for-purpose AI infrastructure

To address the need for a more powerful acceleration option, IBM introduced IBM Spyre Accelerator, a dedicated off-chip accelerator designed specifically for AI-intensive workloads. Compared to the on-chip accelerators in IBM Power11 processors, IBM Spyre Accelerator offers higher throughput, parallelism, and model capacity. This advantage is achieved through its dedicated memory architecture, optimized data paths, and ability to offload processing from the CPU. IBM Spyre Accelerator is ideal for large-scale transformer models that are used in generative AI applications such as translation, summarization, and sentiment analysis. This architecture supports massive parallel compute operations with higher efficiency than general-purpose CPUs.

A key advantage of the IBM Power platform is its ability to combine both acceleration strategies (on-chip and off-chip) within a unified, enterprise-grade ecosystem. Initial AI workloads can begin on existing IBM Power processor-based servers with no added hardware, by using the built-in accelerators for cost efficiency. As demands grow, the IBM Spyre Accelerator can be seamlessly integrated to boost performance without requiring a platform change or software rewrite. This flexibility protects existing investments while enabling future scaling, as shown in Figure 4-1.

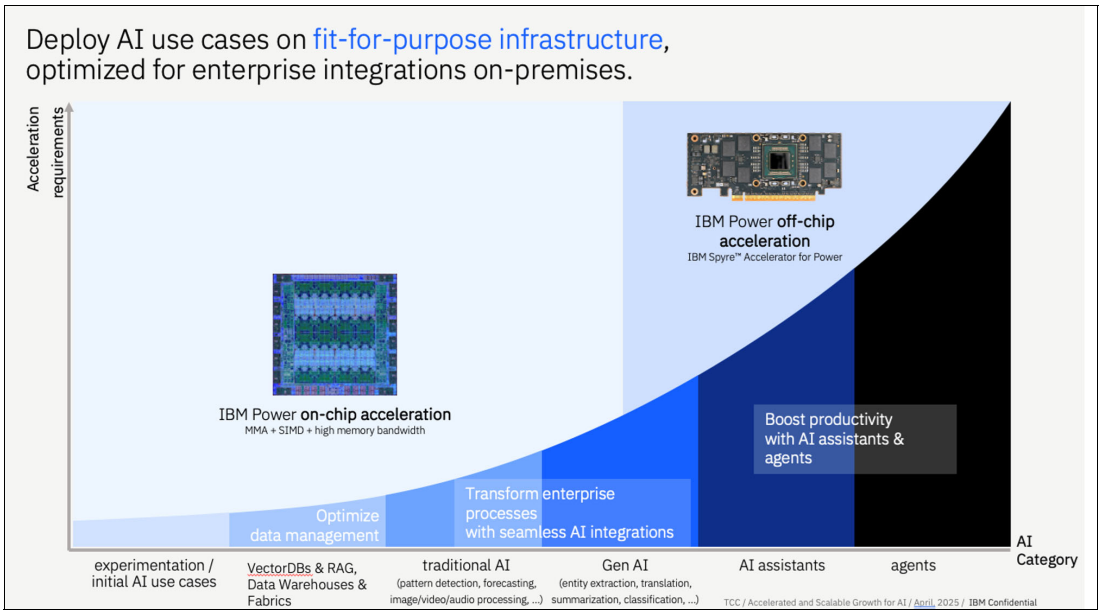

Figure 4-1 IBM's fit-for-purpose AI infrastructure for IBM Power

In high-value use cases such as AI assistants and autonomous agents, where real-time interaction, low-latency inference, and contextual awareness are essential, off-chip acceleration becomes a requirement. These applications often rely on large language models (LLMs) and must manage vast knowledge graphs or context windows that exceed the practical limits of CPU-based inference alone. Spyre is built for these scenarios. When paired with IBM Power11, it provides a hybrid architecture that is secure, high-performing, and manageable within enterprise IT constraints.

This fit-for-purpose infrastructure aligns well with existing on-premises environments, especially ones where data sovereignty, latency, and compliance are non-negotiable. Unlike cloud-only solutions, the Power11 and Spyre combination gives enterprises full control over their data and compute stack while supporting AI workloads that rival hyperscale offerings in performance.

This approach to enterprise AI infrastructure stands out for its modularity and adaptability. Whether you start with lightweight models or scale to multimodal AI agents, IBM Power11 processor-based servers enable you to deploy AI on your terms. With on-chip acceleration for efficient starting points and Spyre off-chip accelerators for high-performance growth, businesses can move from AI exploration to enterprise-wide transformation, all on a platform that is built for the future of AI.

4.2.3 IBM Spyre Accelerator

The IBM Spyre Accelerator is a purpose-built, enterprise-grade accelerator that offers scalable capabilities for complex AI models and generative AI use cases. The new accelerator features 32 individual accelerator cores and is mounted on a PCIe card. Jointly designed by IBM Research and IBM Infrastructure, the Spyre architecture is optimized for efficient AI computation. The chip sends data directly from one compute engine to the next, enabling efficient energy use. This family of processors also supports a range of lower-precision numeric formats such as INT4 and INT8 to make running AI models more energy efficient and less memory intensive.

Spyre uses a system-on-a-chip architecture that is optimized for enterprise AI workloads. It is implemented in 5 nm technology with a high-performance, low-power design that consumes 75 W and provides the following capabilities:

- Thirty-two low-power AI cores
- Multi-precision support for inference and training: FP16/8 and INT8/4
- Support for foundation models that are enabled in the Red Hat software stack (Open Compute Project Open Systems for AI)
- Support for popular AI frameworks and libraries such as PyTorch and vLLM

Figure 4-2 shows the IBM Spyre Accelerator.

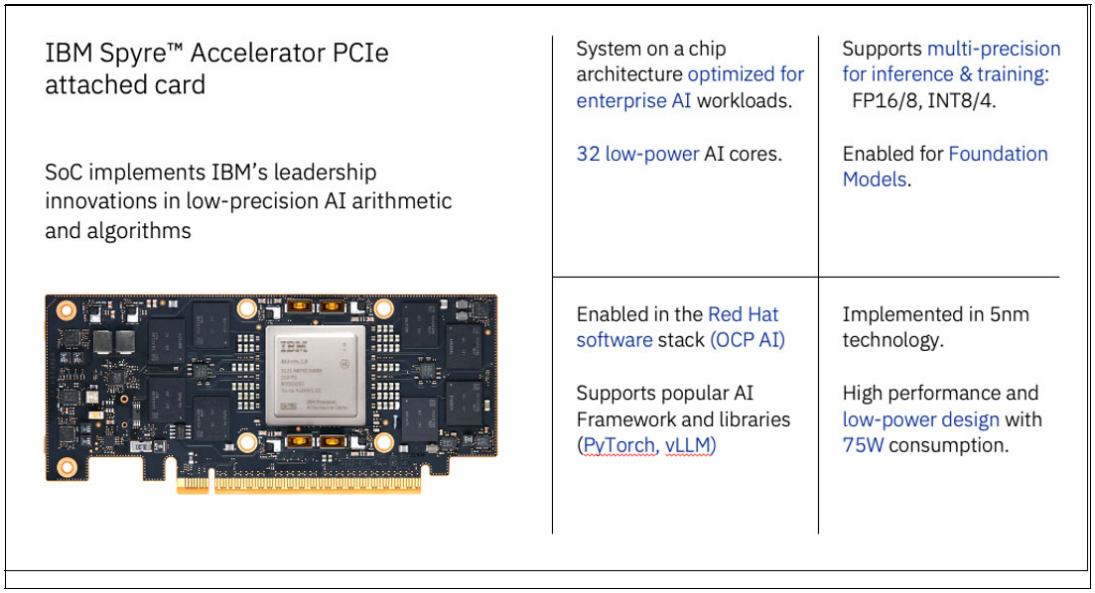

Figure 4-2 The IBM Spyre Accelerator

This dedicated, enterprise-grade AI acceleration chip sits on a 75 W PCIe adapter and is paired with 128 GB of LPDDR5 memory to support several LLMs for the heterogeneous workloads that are typically seen on IBM Power servers. A single IBM Spyre Accelerator does not provide enough compute capacity for most use cases. Therefore, the IBM solution uses current I/O expansion technology to attach a cluster of eight IBM Spyre Accelerator adapters in a single I/O expansion drawer, creating a logical cluster. The firmware on these eight cards coordinates compute distribution and data transfer among the cards. This architecture presents the cluster to the software as one high-performance compute engine with 1 TB of memory and 1.6 TBps of memory bandwidth.

Figure 4-3 shows this architecture.

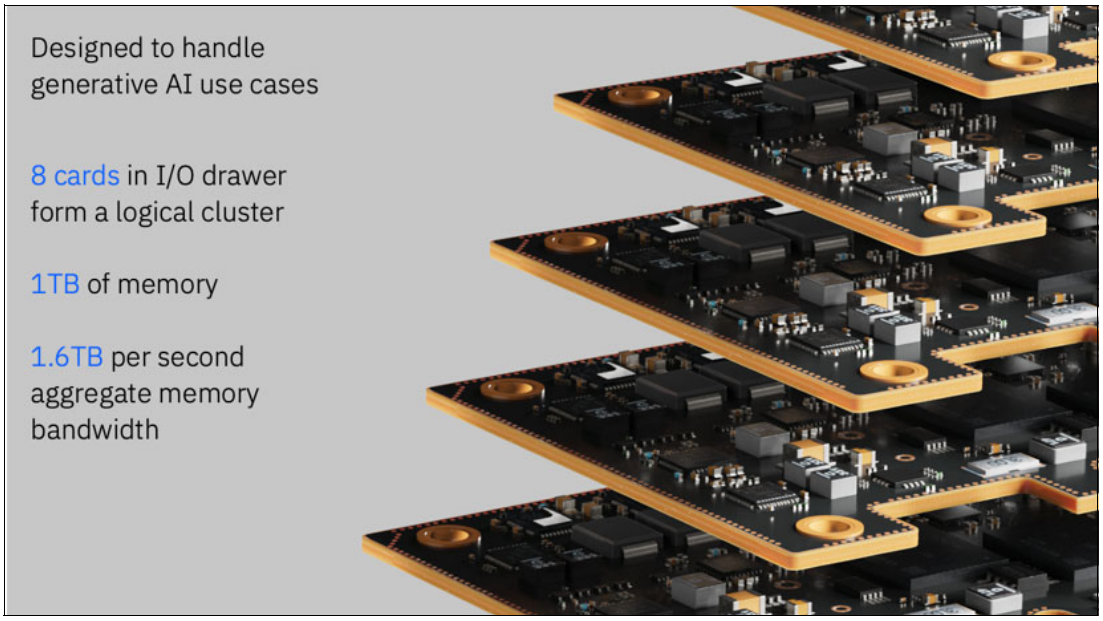

Figure 4-3 IBM Spyre Accelerator cluster in an expansion drawer

The Spyre expansion drawer can be attached to any announced IBM Power11 server. The number of supported drawers depends on the Power11 server model. At announcement, Spyre is planned to be available in a fixed configuration that includes eight IBM Spyre Accelerator adapters that are installed in a PCIe Gen4 I/O expansion drawer. The Spyre expansion drawer can be attached to any Power11 server: IBM Power S1122, IBM Power S1124, IBM Power E1150, or IBM Power E1180. One expansion drawer is supported on all models except the Power E1180, which supports up to two. New I/O drawer components, such as power supplies and fanout modules, are used to support the additional power requirements of the IBM Spyre Accelerator adapter configuration.

The initial supported software and hardware include:

- Firmware 1110.10
- Hardware Management Console (HMC) 1111
- Red Hat Enterprise Linux (RHEL) 9.6
- Spyre software stack container
- Red Hat OpenShift Container Platform AI (technology preview in 4Q 2025, GA in 1Q 2026)

4.3 Solutions

AI is no longer an emerging trend; it is a foundational pillar of competitive advantage. According to the IBM Institute for Business Value, about 72% of top-performing CEOs identify advanced generative AI as essential to their future success[1]. However, only a fraction of enterprises has successfully moved beyond pilot projects and proofs of concept (PoCs) into full production deployments.

The IBM Power11 processor-based server provides the AI-optimized infrastructure and integrated software stack that are required to scale AI workloads, secure enterprise data, and modernize mission-critical processes.

Organizations across industries are realizing tangible benefits from deploying AI on IBM Power:

- Logistics service providers can reduce order processing time by embedding generative AI into their enterprise resource planning (ERP) systems.
- Financial institutions can accelerate anomaly detection while reducing their total cost of ownership (TCO).
- Utility providers can reduce equipment downtime and energy consumption by using AI-powered visual inspection.

These outcomes highlight the value of deploying generative AI in real business workflows by using the trusted IBM Power platform.

IBM Power is optimized for a wide range of AI workloads across key sectors:

- Finance and ERP: Fraud detection, anti-money laundering (AML), order processing, and invoice compliance
- Healthcare: Medical transcription, claims-to-EHR matching, diagnostic imaging, and assistant bots
- IT and development: Predictive IT operations (ITOps), code assistants (RPG and Ansible), transcription, and documentation
- Cross-industry: Document digitization, summarization, know-your-customer (KYC), visual quality inspection, customer churn prediction, and business intelligence

Each use case uses the inherent strengths of IBM Power, such as resilient compute, secure data access, and scalable architecture.

4.3.1 Barriers to scaling AI and the IBM Power technology advantage

Although the promise of AI is significant, common enterprise challenges persist, as shown in Table 4-1.

Table 4-1 Challenges to scale AI

Barrier	Description
Data complexity	Fragmented formats, distributed sources, and lack of standardized connectors
AI integration difficulty	Earlier workflows, lack of AI skills, and software maturity

[1] Source: https://www.ibm.com/thought-leadership/institute-business-value/en-us/think-2025

Barrier	Description
Security and privacy	Concerns around data leakage, adversarial threats, and model protection
Infrastructure and cost concerns	High cost of AI development and fears of redesigning core systems
Talent and skill gaps	Shortage of data scientists and AI engineers

IBM Power11 addresses these challenges by offering a complete AI platform that is integrated into existing systems, with hardware acceleration, data fabric, and AI assistants.

4.3.2 IBM PowerAI portfolio: from data to business value

To scale AI in production, IBM Power provides an AI-ready portfolio that includes the following items:

- AI-ready infrastructure: Built-in accelerators and support for hybrid and cloud-native deployment (for example, IBM Power Virtual Server (PowerVS))
- Optimized software stack: Red Hat, IBM, open source, and certified independent software vendor (ISV) AI solutions
- Data fabric: Secure, unified access to enterprise data across silos
- AI assistants and agents: Tools to enhance developer productivity, automate IT operations (ITOps), and streamline business workflows

This stack enables organizations to infuse AI into core business processes while maintaining governance, performance, and security.

4.3.3 Strategic outcomes that are enabled by IBM Power

Table 4-2 shows example outcomes that are enabled by implementing AI solutions on IBM Power processor-based servers.

Table 4-2 Example outcomes that are enabled by IBM Power

Objective	Outcome example
Transform processes	Accelerate ERP or financial workflows that are integrated with AI services.
Boost productivity	Roll out new application features faster with generative AI code assistants.
Optimize data use	Reduce TCO and improve resource efficiency by using unified data governance.

IBM Power enables clients to put AI to work today: securely, efficiently, and at enterprise scale.

4.3.4 Uniqueness of the platform

With its AI-optimized infrastructure and enterprise architecture, IBM Power is uniquely positioned to help organizations move from experimentation to real AI-driven transformation. Its scalable ecosystem, which spans hardware, software, data, and automation, makes it a reliable platform for deploying production-grade AI at scale.

A key differentiator of the IBM Power11 processor-based platform is its ability to deliver AI capabilities without requiring external GPUs. Unlike conventional assumptions that all AI workloads demand dedicated GPU acceleration, IBM Power11 uses advanced, on-core accelerators that are integrated directly within each processor core, which are combined with a system-wide infrastructure design that is optimized for AI and data-intensive tasks.

This architectural approach enables the deployment of a wide range of AI use cases, such as inferencing, predictive analytics, and cognitive automation, while maintaining enterprise-grade attributes that include resilience, reliability, performance, security, and sustainability. By eliminating or reducing dependency on external GPUs, organizations benefit from simplified infrastructure, reduced energy consumption, and reduce TCO, all without compromising AI readiness. The IBM Power11 platform is purpose-built to support scalable and secure AI solutions across diverse industry environments.

5

Automation and management

Managing IBM Power processor-based servers effectively requires a combination of tools that provide automation, orchestration, and visibility across the infrastructure. At the core of this management stack is the Hardware Management Console (HMC), which serves as the central point for configuring and monitoring Power servers. The HMC enables you to manage logical partitions (LPARs), perform firmware updates, and monitor system health. It provides a CLI, a GUI, and a Representational State Transfer (REST) application programming interface (API), enabling integration with automation tools for scalable operations.

Ansible and Terraform bring Infrastructure as Code (IaC) capabilities to IBM Power environments. Ansible is widely used for configuration management and automation of tasks such as patching, user management, and software deployment. It supports IBM Power through modules and collections that are tailored for AIX, IBM i, and Linux on IBM Power. Terraform complements Ansible by enabling declarative provisioning of infrastructure components, including IBM Power Virtual Server (PowerVS) and IBM Power Virtualization Center (PowerVC) managed resources. With these tools, you can automate the full lifecycle of infrastructure from provisioning to configuration, which helps ensure consistency and reduces manual effort.

PowerVC plays a crucial role in managing virtualized environments on IBM Power. Built on OpenStack, PowerVC provides advanced virtualization management capabilities such as image management, dynamic resource allocation, and integration with cloud platforms. It supports automation through REST APIs and integrates with both Ansible and Terraform, enabling seamless orchestration of virtual machines (VMs) and workloads. By combining HMC, Ansible, Terraform, and PowerVC, you can build a robust, automated, and scalable management framework for IBM Power infrastructure.

The following topics are covered in this chapter:

- Hardware Management Console overview
- Ansible
- Terraform
- PowerVC

© Copyright IBM Corp. 2025.

5.1 Hardware Management Console overview

The HMC is a hardware or virtual appliance that you use to configure and manage systems. The HMC connects to one or more managed systems and provides capabilities for the following primary functions:

- Provide systems management functions, including these examples:
  - Power off
  - Power on
  - System settings
  - Capacity on Demand
  - Enterprise Pools
  - Shared processor pools (SPPs)
  - Performance and capacity monitoring
- Start the Advanced System Management Interface (ASMI) for managed systems.
- Deliver virtualization management by creating, managing, and deleting LPARs; performing Live Partition Mobility (LPM) and Remote Restart; configuring SR-IOV; managing Virtual I/O Servers (VIOSs); allocating resources dynamically; and providing operating system (OS) terminals.
- Act as the service focal point for systems and support service functions, including Call Home, dump management, guided repair and verification, concurrent firmware updates for managed systems, and around-the-clock error reporting with IBM Electronic Service Agent (ESA) for faster support.
- Provide appliance management capabilities for configuring the network, managing users on the HMC, and updating or upgrading the HMC.

5.1.1 HMC options

Power11 servers can connect to either the 7063-CR2 HMC or a Virtual HMC.

Restriction: Power11 servers do not support the 7063-CR1 HMC.

7063-CR2 HMC

The 7063-CR2 HMC is a second-generation IBM Power processor-based HMC. It includes the following features:

- A six-core IBM Power9 130 W processor chip
- 64 GB (4 × 16 GB) or 128 GB (4 × 32 GB) of memory
- 1.8 TB of internal disk capacity with RAID1 protection
- Four 1 Gbps Ethernet ports (RJ-45), two 10 Gbps Ethernet ports (RJ-45), two USB 3.0 ports (front), two USB 3.0 ports (rear), and one 1 Gbps IPMI Ethernet port (RJ-45)
- Two 900 W power supply units
- Remote Management Service: IPMI port (OpenBMC) and Redfish API

The base warranty is 1 year, 9 × 5, with optional upgrades available. A USB Smart Drive is not included.

Figure 5-1 shows a 7063-CR2 HMC.

Figure 5-1 HMC 7063-CR2

The 7063-CR2 is compatible with flat panel console kits 7316-TF3, TF4, and TF5.

Note: The 7316-TF3 and TF4 were withdrawn from marketing.

Virtual HMC

Initially, IBM sold the HMC only as a hardware appliance with the HMC firmware installed. IBM later extended this offering to include a hardware appliance or a virtual appliance that you can deploy on ppc64le architectures or x86 platforms.

If you have a valid contract, you can download the HMC from the IBM Enterprise Storage Server, or it can be included with an initial Power E1180 order.

The virtual HMC supports the following hypervisors:

- On x86 processor-based servers:
  - Kernel-based Virtual Machine (KVM)
  - Xen
  - VMware
- On IBM Power processor-based servers: IBM PowerVM

The following minimum requirements apply to install the virtual HMC:

- 16 GB of memory
- Four virtual processors
- Two network interfaces (maximum of four allowed)
- One disk drive with 500 GB of available space

For an initial Power E1180 order with the IBM configurator (e-config), you can find the HMC virtual appliance by selecting **Add software** → **Other System Offerings** and then selecting one of the following options:

- 5765-HMD: IBM HMC Virtual Appliance for Power V11
- 5765-HMU: IBM HMC Virtual Appliance for x86 V11

For more information and an overview of the Virtual HMC, see Virtual HMC appliance (vHMC) overview.

For more information about how to install the virtual HMC appliance and all the requirements, see Installing the HMC virtual appliance.

5.1.2 BMC network connectivity rules for 7063-CR2 HMC

The 7063-CR2 HMC includes a baseboard management controller (BMC), which is a specialized service processor that monitors the physical state of the system by using sensors. The OpenBMC implementation on the 7063-CR2 provides a GUI that you can access from a workstation with network connectivity to the BMC. This connection requires an Ethernet port that is configured for BMC use.

The 7063-CR2 HMC provides two network interfaces (eth0 and eth1) for configuring BMC network connectivity on the appliance. Each interface maps to a different physical port on the system. Different management tools might use different names for these interfaces. Selecting **Console Management** → **Console Settings** → **Change BMC/IPMI Network Settings** modifies only the Dedicated interface.

Table 5-1 lists the BMC ports.

Table 5-1 BMC ports

Management tool	Logical port	Shared or dedicated	CR2 physical port
OpenBMC UI	eth0	Shared	eth0
OpenBMC UI	eth1	Dedicated	Management port only
`ipmitool`	lan1	Shared	eth0
`ipmitool`	lan2	Dedicated	Management port only
HMC task (change the BMC/IPMI Network settings	lan2	Dedicated	Management port only

Figure 5-2 shows the BMC interfaces of the HMC.

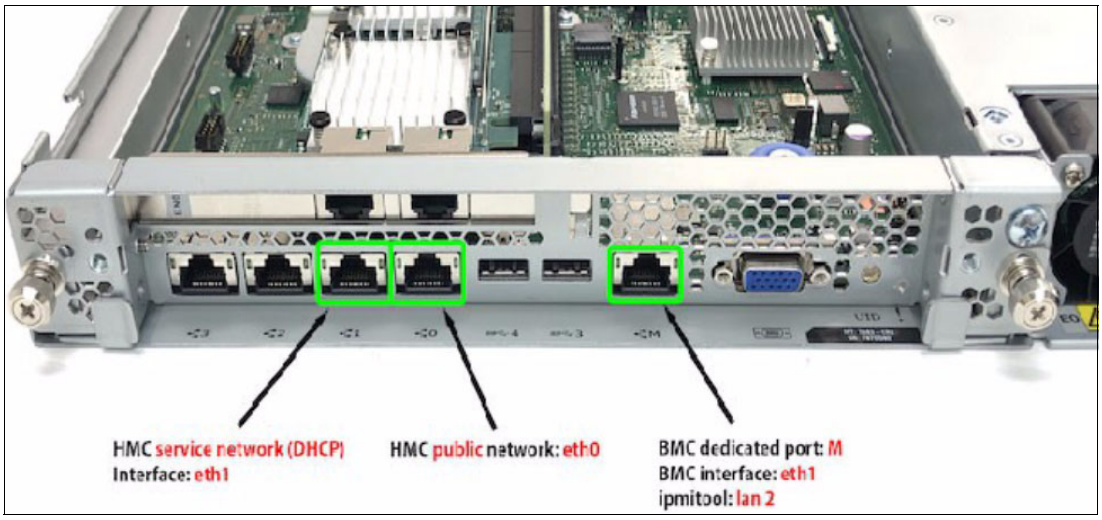

Figure 5-2 BMC interfaces

The main difference is that the shared and dedicated interfaces to the BMC may coexist. Each interface has its own LAN number and physical port. Ideally, you configure one port, but

both can be configured. The rules for connecting IBM Power servers to the HMC remain the same as in previous versions.

5.1.3 High availability HMC configuration

For the best manageability and redundancy, use a dual HMC configuration. This configuration can include two hardware appliances, one hardware appliance and one virtual appliance, or two virtual appliances.

The following requirements apply:

- Both HMCs must run the same version.
- The HMCs must use different subnets to connect to the Flexible Service Processors (FSPs).
- The HMCs must communicate with the servers' partitions over a public network to enable full synchronization and function.

5.1.4 HMC code level requirements

The minimum required HMC version for the Power E1180 is V11R1M1110. This version is supported only on the 7063-CR2 HMC and Virtual HMC appliances. It is not supported on the 7063-CR1 or 7042 HMCs.

Note: An HMC running V11R1M1110 cannot manage Power8 or earlier processor-based systems.

An HMC running V11R1M1110 includes the following features:

- Support for managing Power11 systems
- Support for new I/O adapters
- VIOS management enhancements
  - Resource groups
  - IBM i Secure Boot
  - Increased platform keystore size
  - Removal of support for vTPM 1.2
  - Quantum-safe LPM
  - Minimum affinity score and actions
- Console management and user experience improvements
  - Network topology
  - Trusted keystore
  - Import certificate
  - Multi-factor authentication allowlist
  - Ability to advertise device information by using Link Layer Discovery Protocol (LLDP)

- IBM Power infrastructure maintenance and automated IBM Power platform updates
  - Run IBM Power platform updates with minimal touch points to enhance simplicity and reduce the risk of human error
  - Experience seamless platform updates with one-touch solutions
  - Automatic operational recovery and resiliency
- Autonomous Error Resolution (AER)
  - Ability to collect platform logs (firmware, hypervisor, HMC, or VIOS) from a single interface
  - Ability to create a case and upload logs to the case from the HMC
- Sustainability
  - New Energy efficiency mode for higher performance per watt
  - Partition-level energy monitoring with real-time reporting of energy and carbon emissions at the VM or partition level
  - Scheduling of energy modes

5.1.5 HMC currency

In recent years, cybersecurity has emerged as a national security issue and an increasingly critical concern for CIOs and enterprise IT managers. The IBM Power Architecture has consistently ranked highly in terms of end-to-end security, which is why it remains a platform of choice for mission-critical enterprise workloads.

A key aspect of maintaining a secure IBM Power environment is helping ensure that the HMC or Virtual HMC is current and fully supported, including hardware, software, and IBM Power firmware updates. Outdated or unsupported HMCs represent a technology risk that you can quickly mitigate by upgrading to a current release.

5.1.6 New features

The HMC and the Virtual HMC have the following enhancements:

- Support for managing the new generation of Power11 systems
- Support for new I/O adapters
- Resource Groups
- IBM i Secure Boot
- Quantum-safe LPM
- Minimum Affinity Score and Actions
- Ability to advertise device information through LLDP
- IBM Power Infrastructure Maintenance and Automated Power Platform Update Tool
- Automated First Failure Data Capture (FFDC) Log collection
- Sustainability
- Energy efficiency mode
- Scheduling of energy modes
- Partition Level Energy Monitoring

5.1.7 Using the Automated Maintenance Framework tool

This section describes how to use the new Automated Maintenance Framework tool that is provided in the IBM Power11 HMC.

Launchpoint

Before you start the tool, install the latest 11.10 firmware by completing the following steps:

1. Select **Update system** → **VIOS** → **Adapter levels**, as shown in Figure 5-3.

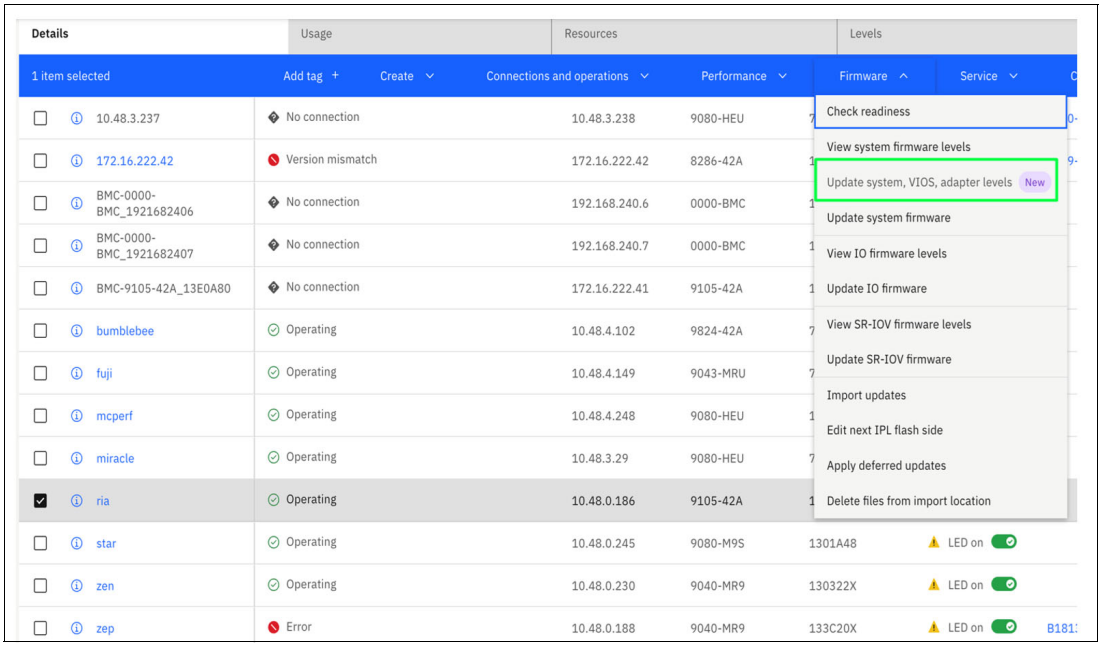

Figure 5-3 Starting the update

2. Complete the following steps:
   a. For the Import update files option, select **Import files to HMC filesystem and perform update** or **Import files only**.
   b. Check the system readiness. The system must be in the “Ready” state to proceed with update process.
   c. Select the source file location.

3. Select the **Update/Upgrade** type, as shown in Figure 5-4.

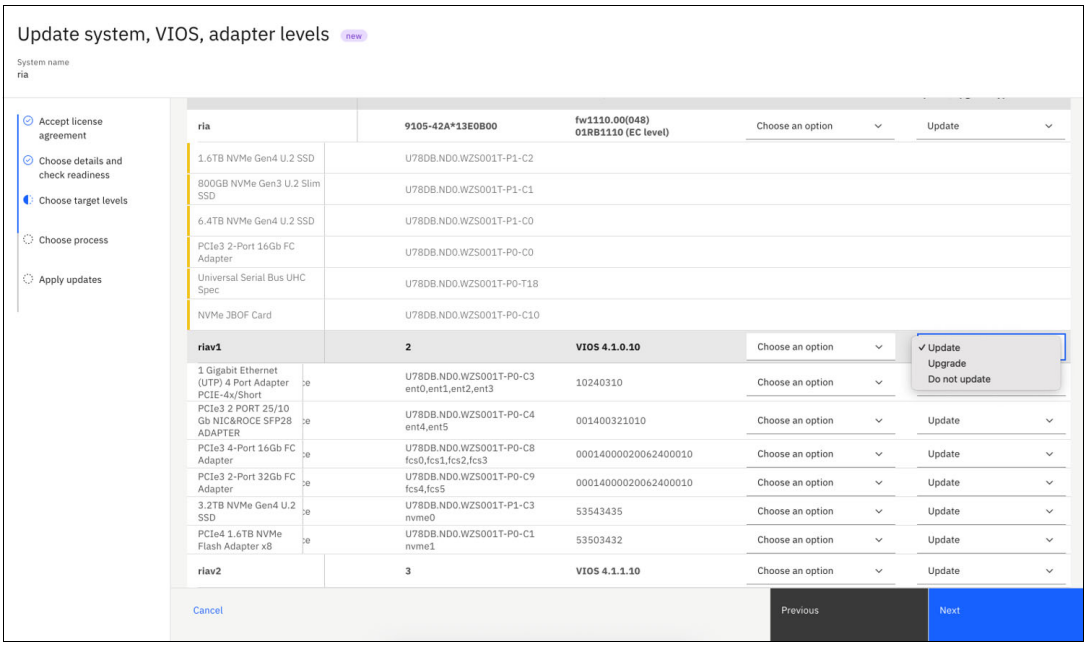

Figure 5-4 Upgrade type

4. The partitions migration window opens, as shown in Figure 5-5. Use the options in this window to preserve system availability during disruptive upgrades or updates.

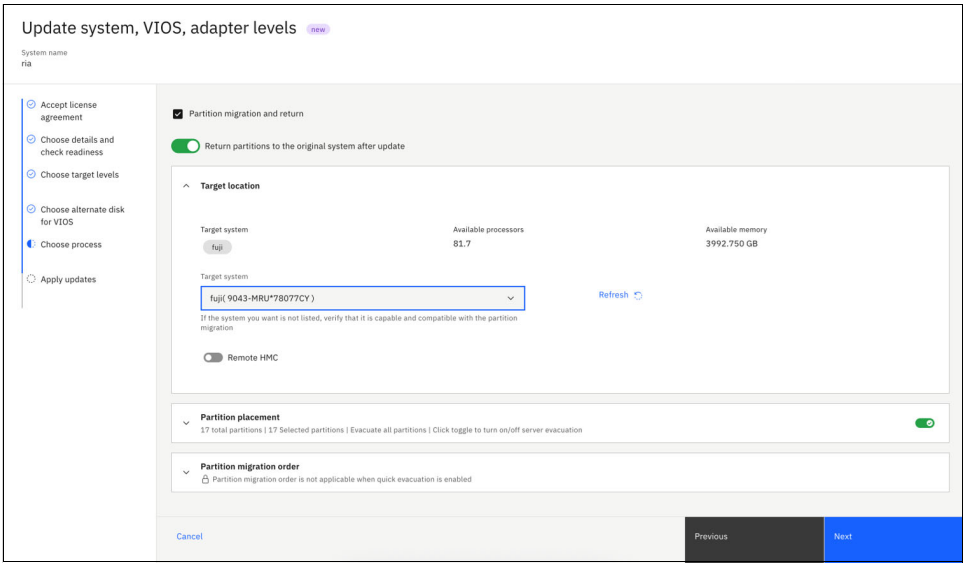

Figure 5-5 Choose Process

5. Select one of the following options:
   - **Remote HMC switch**: Enables you to migrate partitions to a system that is managed by a remote HMC. When this option is selected, partitions are not automatically migrated back to the original system after the update.
   - **Server evacuation switch**: By default, all partitions are selected for migration. You can select individual partitions for migration by toggling this switch.

If you select specific partitions for migration, you can specify the migration sequence for a set of partitions that must be migrated first.

The update process completes automatically, and you can view the results as shown in Figure 5-6.

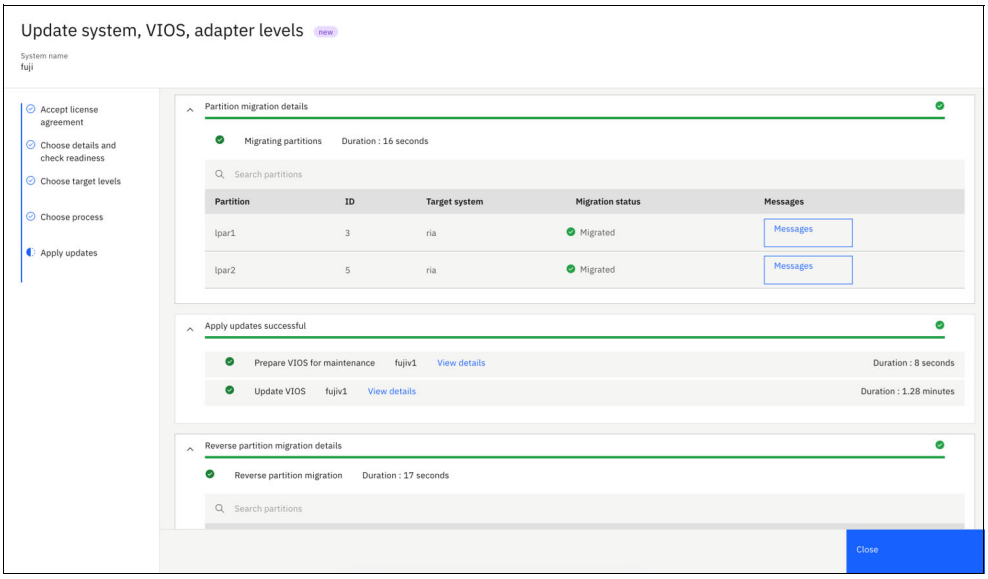

Figure 5-6 Update status

In conclusion, this solution provides unmatched business resiliency with zero planned downtime (ZPD) for maintenance, which is delivered through platform automation capabilities that are integrated with IBM Concert artificial intelligence (AI)-infused workloads. It enables faster and more frequent maintenance updates, keeping systems secure, stable, and compatible with evolving software and hardware requirements, while reducing the risk of performance degradation, security breaches, or unplanned downtime.

5.2 Ansible

Ansible is an open-source, cross-platform tool for resource provisioning and automation that DevOps professionals use for continuous deployment (CD) of software by using an IaC approach. The Ansible automation platform evolved to deliver sophisticated automation solutions for operators, administrators, and IT decision-makers across various technical disciplines. It is an enterprise automation solution with a strong open-source foundation. Ansible operates on several UNIX like platforms and can manage systems such as UNIX and Microsoft architectures. It uses a descriptive language for defining system settings.

Because of its broad adoption, open-source design, and extensive support for many devices and platforms, Ansible is becoming a dominant tool in the market. However, it is also common to use other automation tools with Ansible to achieve more complex automation. For example, many organizations use Ansible with Terraform to automate infrastructure provisioning.

Ansible architecture

As shown in Figure 5-7, the Ansible architecture consists of an Ansible controller and one or more Ansible client hosts. The controller runs automation tasks and stores Ansible collections, which contain modules, plug-ins, and roles that define the actions Ansible can perform on client nodes.

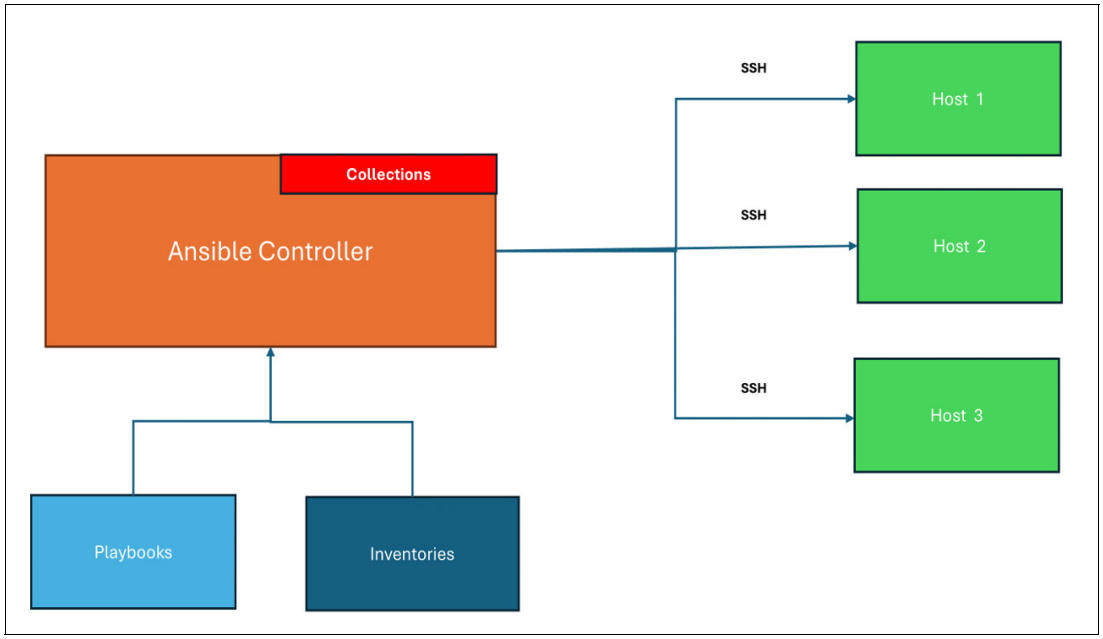

Figure 5-7 Simplified Ansible architecture

Playbooks

The core of Ansible automation is the Ansible playbook. Playbooks are YAML files that define sequences of tasks to run on remote hosts. These tasks can include installing packages, configuring services, or copying files. Playbooks enable IT teams to automate infrastructure provisioning, configuration management, application deployment, and more.

Why choose Ansible

Ansible offers numerous benefits for IT professionals who want to improve efficiency, scalability, and consistency in their infrastructure. Here are some key advantages:

- Versatility: Ansible supports a wide range of devices and can scale to accommodate growing environments and automation needs.
- Agentless architecture: Ansible manages devices by using Secure Shell (SSH), which eliminates the need for agents on target systems.
- Flexibility: Ansible can handle simple CLI tasks and complex workflows that are defined in playbooks.
- Extensive module library: Ansible provides a rich collection of modules for managing various systems, cloud infrastructures, and OpenStack.
- Declarative approach: With Ansible's declarative syntax, you define the wanted state of a system, and Ansible takes the necessary steps to achieve it.
- Ease of learning: The YAML-based syntax and minimal learning curve make Ansible accessible to IT professionals at all levels.

Ansible is a powerful automation tool that helps organizations improve efficiency, scalability, and reliability in their IT infrastructure. By using Ansible playbooks, IT teams can streamline routine tasks, automate complex workflows, and help ensure consistent configurations across environments.

Options for implementing Ansible

As you plan to implement Ansible for IT management, it is essential to select the correct product and support level to meet your organization's needs. This section describes some of the available options.

Ansible Community

Here are the primary community versions of Ansible:

- Ansible Core

  Ansible Core is the foundation of Ansible. It provides the core automation engine and includes the basic functions for configuration management, application deployment, and task automation. Ansible Core includes modules, plug-ins, and the CLI that are needed to run playbooks and manage configurations.

- AWX

  AWX is the upstream, open-source project that serves as the community version of Red Hat Ansible Tower. AWX provides a web-based user interface (UI), REST API, and task engine for managing Ansible automation at scale. It offers role-based access control (RBAC), job scheduling, graphical inventory management, and more to help users manage and scale automation efforts.

- Ansible Collections

  Ansible Collections are prepackaged modules, roles, and plug-ins that are created and shared by the community. Collections extend Ansible functions with extra content that is often maintained by the community or specific organizations. You can download collections from Ansible Galaxy, a community hub for sharing and discovering Ansible content.

- Ansible Galaxy

  Ansible Galaxy is a repository for sharing and discovering Ansible roles and collections. It is a community-driven platform where users can find reusable Ansible content to simplify automation tasks. It provides a searchable repository of roles and collections that are created by the Ansible community and can be integrated into your automation workflows.

These community versions are suitable for individual users, small teams, and development environments but lack the formal support and advanced features that are provided by the Red Hat Ansible Automation Platform.

Ansible Automation Platform

Ansible Automation Platform is a subscription-based enterprise solution that combines more than 20 community projects into a fully supported automation platform. It provides curated, certified, and validated Ansible Collections and roles from partners such as IBM, Juniper, Cisco, and public cloud providers.

Here are the key considerations for choosing Ansible Automation Platform:

- Support level: Ansible Automation Platform offers enterprise-grade support, including service-level agreements (SLAs) for security, compatibility, and upgrades. Community options might have limited support.
- Features: The platform includes features beyond Ansible Core, such as a web interface and integration with other tools.

- Cost: Ansible Automation Platform is subscription-based, but community options are available at no charge.
- Scale and complexity: For large organizations with complex automation needs, Ansible Automation Platform might be the better choice because of its enterprise-grade features and support.

By evaluating these factors, you can select the Ansible offering that best aligns with your organization's goals, budget, and support requirements.

Using Ansible to automate your IBM Power infrastructure

Ansible is a powerful automation tool that delivers value for managing IBM Power environments. Here are some key use cases where Ansible enhances efficiency, consistency, and scalability:

- System configuration and provisioning

  Ansible automates the setup and configuration of AIX, IBM i, and Linux on IBM Power. This process includes tasks such as user and group management, network configuration, software installation, and system tuning. By using Ansible playbooks, you can help ensure consistent configurations across multiple systems, reducing manual errors and accelerating provisioning.

- Fix management and compliance

  Keeping systems up to date is critical for security and stability. Ansible automates the patching process for AIX and IBM i, including downloading fixes, applying them, and verifying system health after updates. It also supports compliance checks by validating system configurations against predefined baselines, helping organizations meet regulatory and security standards.

- Integration with HMC and PowerVC

  Ansible collections for IBM Power include modules that interact with the HMC and PowerVC, which enable automation of tasks such as LPAR creation, VIOS configuration, and VM lifecycle management. Combined with dynamic inventory capabilities, Ansible can discover and manage IBM Power infrastructure components in real time.

For more information about implementing automation with Ansible in an IBM Power environment, see *Using Ansible for Automation in IBM Power Environments*, SG24-8551.

5.3 Terraform

Terraform is an open-source tool that was originally developed by IBM HashiCorp® and is now owned by IBM. It is written in the Go programming language and compiles into an executable files that is named Terraform. Terraform is an IaC tool that enables you to build, change, and create versions of cloud and on-premises resources safely and efficiently. It provides a mechanism to access any API for any cloud provider to manage infrastructure as a service (IaaS).

Figure 5-8 on page 121 shows the process of calling the API. The APIs to call are defined in configuration files, which serve as the code that is referenced in the IaC approach.

Figure 5-8 Terraform functions

With Terraform, you can define and provision infrastructure by using a high-level configuration language that is called HashiCorp Configuration Language (HCL). With Terraform, you can describe infrastructure components such as servers, databases, networking, and storage in code, enabling version control, collaboration, and repeatability. This approach eliminates the need for manual setup and reduces the risk of configuration drift across environments.

Terraform operates through a workflow that includes writing configuration files, initializing the working directory, planning changes, and applying them:

- The **`terraform init`** command sets up the environment by downloading the necessary provider plug-ins.
- The **`terraform plan`** command creates an execution plan that shows the actions Terraform takes to reach the wanted state.
- The **`terraform apply`** command runs the plan, making the changes to the infrastructure.

Terraform maintains a state file that tracks the state of the infrastructure, which is essential for determining what changes must be made during future runs.

One of Terraform's key strengths is its provider ecosystem, which you use to manage resources across a wide range of platforms, including PowerVS, Amazon Web Services (AWS), Microsoft Azure, Google Cloud, Kubernetes, and many others. This ecosystem makes Terraform a powerful tool for managing hybrid and multi-cloud environments. Also, Terraform supports modules, which are reusable configurations that promote consistency and reduce duplication. By codifying infrastructure, Terraform enables DevOps practices such as continuous integration and delivery (CI/CD), infrastructure testing, and automated deployments.

Figure 5-9 shows how Terraform works by defining your intended end state through configuration files, which are then processed into a plan and applied through infrastructure providers.

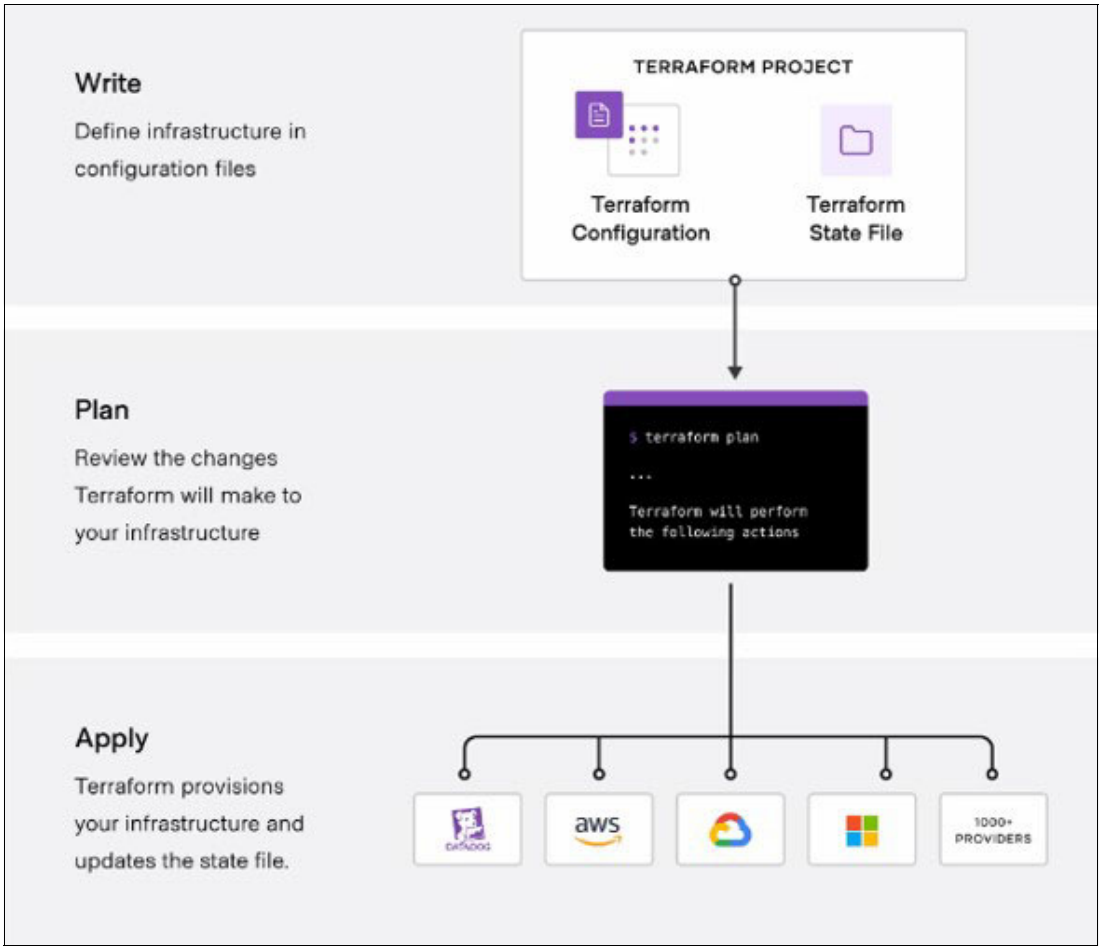

Figure 5-9 Terraform process

- Tracking your infrastructure

  Terraform generates a plan and prompts you for approval before modifying your infrastructure. The state of your infrastructure is stored in a file that is named `terraform.tfstate`, which can be maintained in Git, GitLab, or HCP Terraform to version, encrypt, and securely share it with your team. This file acts as the single source of truth for your environment.

- Automating changes

  Terraform configuration files are declarative and describe the wanted end state. You can automate them with tools such as Ansible.

- Standardizing configurations

  Terraform provides standardization through modules. A module consists of a collection of `.tf` and `.tf.json` files that are stored together in a directory. Modules are the primary method for packaging and reusing resource configurations with Terraform.

- Collaborating

  Because Terraform configurations can be distributed as files and version-controlled in systems such as Git, GitHub, and HCP Terraform, these platforms are ideal for sharing and collaboration.

- Combining Ansible and Terraform

  Terraform connects to any provider, including Ansible, to manage your infrastructure. Browse the Terraform Registry for available providers. The Terraform Provider for Ansible offers a more straightforward and robust method of running Ansible automation from Terraform rather than using `local-exec`.[1]

Figure 5-10 shows the Ansible provider entry in the registry.

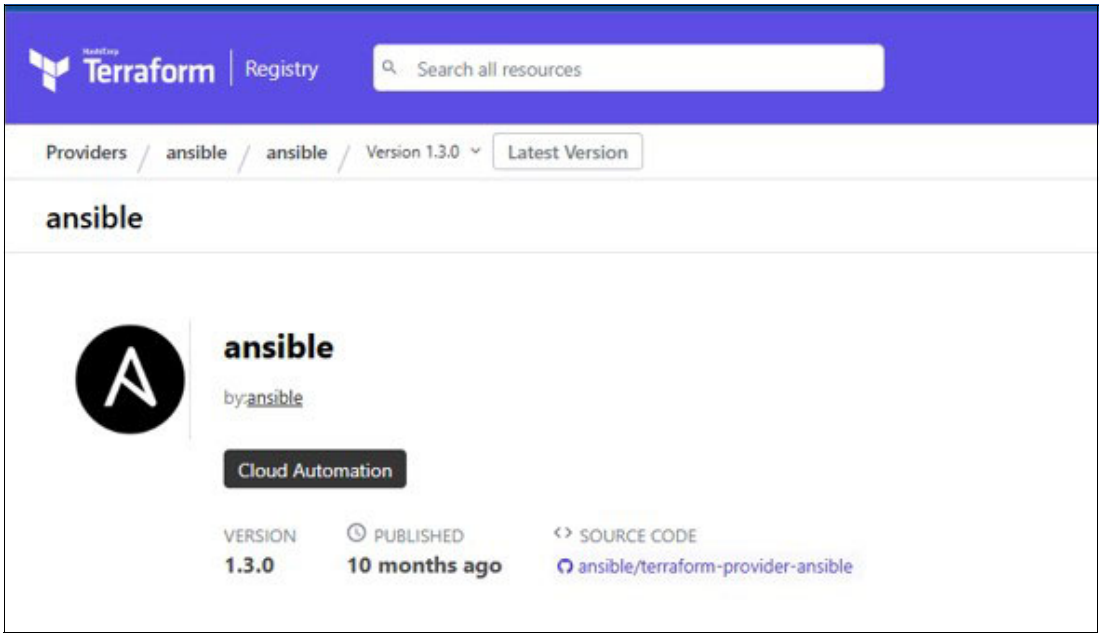

Figure 5-10 Terraform Ansible provider

The prerequisites for using the Ansible provider are as follows;

1. Install Go.

   For installation instructions, see the official installation guide.

2. Install Terraform.

   Install the ppc64le version for operation on IBM Power. For installation instructions, see this GitHub registry.

3. Install Ansible.

   To install Ansible, see the Ansible official installation guide.

For more information about Terraform in the IBM Power environment, see *Modernization Techniques for IBM Power*, SG24-8582.

[1] Source: https://developer.hashicorp.com/terraform/language/resources/provisioners/local-exec

5.4 PowerVC

IBM PowerVC is an advanced virtualization and cloud management platform that is built on OpenStack. It is designed for managing virtualized workloads on IBM Power servers. PowerVC supports AIX, IBM i, and Linux OSs and provides a unified interface for deploying, managing, and automating VMs across IBM Power environments. It simplifies private cloud creation and integrates with higher-level cloud orchestrators such as Ansible, Terraform, and Red Hat OpenShift, enabling hybrid cloud strategies.

PowerVC offers two editions: Standard Edition and Private Cloud Edition.

- Standard Edition includes core virtualization management features such as the following ones:
  - VM image capture, import and export, and deployment
  - Policy-based VM placement for optimized resource usage
  - Snapshots and cloning for backup and testing
  - Live VM mobility and remote restart for high availability (HA)
  - RBAC and automated I/O configuration
- Private Cloud Edition adds the features:
  - A self-service portal for user VM provisioning
  - Approval workflows for provisioning requests
  - Pre-built deployment templates
  - Cloud management policies and metering for chargeback

PowerVC supports Dynamic Resource Optimization (DRO), which automatically balances workloads based on CPU and memory usage, and Simplified Remote Restart, which enables VMs to restart on alternative hosts if there is a failure. These features reduce administrative impact and improve system resilience.

NovaLink is a lightweight, Linux-based virtualization management interface that acts as a bridge between PowerVC and PowerVM Hypervisor (PHYP), which is the hypervisor on Power servers. It enables direct, scalable, and efficient control of VMs and system resources without relying solely on the HMC.

Installed on a Power server, NovaLink runs as a service on a dedicated LPAR on the same Power server that it manages. This LPAR communicates directly with the PHYP through hypervisor APIs. NovaLink provides faster and more granular control of virtualization tasks such as VM creation, deletion, migration, and resource allocation. This approach improves PowerVC responsiveness and scalability, especially in environments with many VMs or frequent provisioning changes.

Table 5-2 shows the comparison of PowerVC with HMC management.

Table 5-2 Comparing PowerVC with HMC management:

Feature	PowerVC	HMC
Policy-based placement	Automatically places VMs based on resource availability and policies.	Manual placement.
Integration with Cloud Tools	Integrates with OpenStack, Red Hat OpenShift, and hybrid cloud platforms.	No native integration.
Live VM Migration	Fully supported and automated.	Supported but requires manual steps.
DRO	Automatically balances workloads across hosts.	Not available.

Feature	PowerVC	HMC
Simplified networking and storage setup	Automated configuration of virtual I/O, storage, and networking.	Manual configuration required.
NovaLink support	Works directly with NovaLink for faster, scalable VM operations.	HMC is external and less scalable.
Scalability	Designed for large-scale environments with many VMs.	Best suited for smaller or static environments.
User interface	Modern, web-based UI with dashboards and templates.	Traditional interface, more administrative.

6

Operating systems

IBM Power servers support three robust operating systems (OSs): IBM i, AIX, and Linux. Each OS is designed to meet specific enterprise computing requirements.

IBM i is an integrated operating environment that is known for reliability, security, and simplicity of management. It includes a built-in database (IBM Db2 for i), middleware, and development tools, making it suitable for running mission-critical business applications with minimal administrative effort. Its tight integration and compatibility with earlier version make it popular in industries such as finance, manufacturing, and retail.

AIX, IBM's UNIX like operating system, is designed for high-performance, scalable enterprise workloads. Built on a UNIX System V base and enhanced with IBM innovations, AIX offers advanced features such as dynamic system tuning, workload partitioning, and strong security. It is widely used in environments that require stability and performance, including enterprise resource planning (ERP) systems, databases, and analytics platforms. AIX is optimized for IBM Power hardware, helping ensure seamless integration and efficient resource usage.

IBM Power also supports enterprise-grade Linux distributions, including Red Hat Enterprise Linux (RHEL) and SUSE Linux Enterprise Server. These Linux environments provide flexibility and open-source innovation, making them ideal for cloud-native applications, containerized workloads, and artificial intelligence (AI) or machine learning (ML) development. Running Linux on IBM Power enables organizations to use the performance advantages of the IBM Power Architecture while maintaining compatibility with the broader Linux ecosystem. Together, these operating systems make IBM Power a versatile platform that supports a wide range of modern and earlier workloads.

Complementing these operating systems is Red Hat OpenShift, which is a Kubernetes based container platform that runs natively on IBM Power. Red Hat OpenShift enables organizations to build, deploy, and manage containerized applications across hybrid cloud environments with consistency and scalability. It provides a unified DevOps experience and supports modern application architectures, including microservices and AI workloads. This integration helps enterprises modernize their IT infrastructure while using the performance, security, and reliability of IBM Power.

© Copyright IBM Corp. 2025.

The following topics are covered in this chapter:

- Subscription licensing
- AIX
- IBM i
- Linux on IBM Power
- PowerVM Virtual I/O Server
- Setting your LPAR compatibility mode

6.1 Subscription licensing

IBM offers subscription-based licensing models for AIX and IBM i. These models provide flexibility, predictability, and cost efficiency. Organizations can license the OS and associated software on a monthly or annual basis instead of using traditional perpetual licenses. This approach aligns with modern IT consumption trends. It enables businesses to scale usage up or down based on workload demands and simplifies budgeting and procurement processes. Subscription licensing also includes access to updates, fixes, and IBM Support to help ensure that systems remain secure and current.

For IBM i, subscription licensing includes the core operating system along with integrated middleware such as Db2, security features, and application development tools. AIX subscription licensing covers the base OS and can be bundled with other IBM software components. These subscriptions are available for deployment on IBM Power11 and IBM Power10 systems. They are suitable for hybrid cloud environments where agility and operational efficiency are essential. By adopting subscription licensing, organizations can modernize their licensing strategy while continuing to use the reliability and performance of AIX and IBM i.

6.1.1 AIX subscription licensing

AIX Standard Edition, AIX Enterprise Edition, IBM Private Cloud Edition, and IBM Private Cloud Edition with AIX are available under a subscription licensing model. This model provides access to an IBM program and IBM software maintenance for a specified subscription term of one or three years. The subscription term begins on the start date and ends on the expiration date, which is shown on the IBM Entitled Systems Support (ESS) website.

Customers are licensed to run the product through the expiration date of the subscription term and can renew at the end of the term to continue using the product. This model offers flexible and predictable pricing over a defined period with a reduced up-front cost.

Another benefit of this model is that the licenses are customer-number entitled, which means they are not tied to a specific hardware serial number as with perpetual licenses. Therefore, the licenses can move between on-premises and cloud environments if needed, which is important for hybrid workloads.

Table 6-1 shows the product IDs for the subscription licenses.

Table 6-1 Subscription license product IDs (one or three year terms)

Product ID	Description
5765-2B1	IBM AIX 7 Standard Edition Subscription 7.3.0
5765-2E1	IBM AIX Enterprise Edition Subscription 1.10.0
5765-2C1	IBM Private Cloud Edition with AIX 7 Subscription 1.10.0
5765-6C1	IBM Private Cloud Edition Subscription 1.10.0 (without AIX)

The subscription licenses are orderable through IBM configuration tools. The AIX perpetual and monthly term licenses for Standard Edition are still available.

6.1.2 IBM i subscription licensing

IBM offers several subscription licensing options for IBM i to provide flexibility and align with modern IT consumption models. These subscriptions are available on Power servers, including Power11, and are ideal for clients who want to simplify licensing, reduce up-front costs, and scale easily.

- The core offering is the IBM i Operating System Subscription, which includes the IBM i OS along with integrated middleware such as Db2 for i, security features, and development tools. This subscription is licensed on a per-core basis and is available in monthly or annual terms. It helps organizations align software costs with usage. Support and updates are included to help ensure that systems remain secure and current without separate maintenance agreements.
- In addition to the base OS subscription, IBM provides IBM i Value Packages. These packages bundle the OS with other IBM software such as IBM Rational® Development Studio for i, IBM Db2 Mirror for i, and IBM i Access Client Solutions. They are tailored to support specific workloads or modernization efforts.
- For cloud-based deployments, the IBM i Cloud Subscription offers a flexible, usage-based model that includes the OS, support, and infrastructure when running IBM i on platforms such as IBM Power Virtual Server (PowerVS). This option is beneficial for organizations adopting hybrid or cloud-first strategies because it enables scalable, enterprise-grade IBM i environments without the complexity of traditional licensing.

6.2 AIX

The AIX OS is a secure, scalable, and robust UNIX like system that is based on open standards. For more than 30 years, AIX has been the cornerstone of mission-critical computing for enterprise organizations in complex industries. It continues to evolve by introducing new hybrid cloud and open-source capabilities.

AIX 7.3 is the latest release that is available. It builds on a solid foundation by offering new functions and capabilities that enhance performance, scalability, availability, and security. It also preserves application-binary compatibility to protect IT investments in AIX.

When combined with IBM Power11 processor-based systems, AIX 7.3 provides an optimized and resilient computing platform that adapts to changing business demands. It supports new cloud use cases and delivers improved economics.

6.2.1 AIX 7.3 key features

The following list summarizes key features in the latest AIX release:

- Workload scalability and automation: AIX 7.3 offers enhanced workload scalability, improved cloud automation through Ansible, and more than 300 open-source packages in the AIX Toolbox for Open Source Software. These capabilities enable modern application development.
- Live Kernel Update: Introduced in AIX 7.2 and enhanced in AIX 7.3, this feature enables interim fixes, Service Packs (SPs), and Technology Level (TL) updates without requiring system restarts. It supports IBM Power Virtualization Center (PowerVC) managed environments and IBM Power Enterprise Pool (PEP) systems for resource optimization.
- High availability and disaster recovery (HADR): IBM PowerHA and VM Recovery Manager (VMRM) provide automated recovery and multi-site replication. These features minimize downtime and help ensure business continuity in hybrid or public cloud environments.
- AI integration: AIX workloads are a natural source for AI because they host large volumes of high-quality data on customer behavior and transactional information. This data can be used for AI development, enabling machine learning (ML) and deep learning for actionable insights on a unified platform.
- Cloud flexibility: AIX supports private, on-premises cloud transformation with PowerVC, offering hybrid cloud functions for seamless AIX virtual machine (VM) import and export. It also provides software-defined infrastructure for storage area network (SAN)-less DevOps environments. Available through PowerVS, AIX supports mission-critical databases with enhanced scalability, cloud automation, robust security, and flexible licensing. Workloads can run in hybrid or public clouds without refactoring, helping ensure reliability and efficiency.

6.2.2 Supported levels

At the time of writing, IBM Power11 processor-based servers support the following minimum AIX OS levels when installed by using direct I/O connectivity:

- AIX 7.3 with TL 7300-03 and SP 1 or later
- AIX 7.3 with TL 7300-02 and SP 4 or later
- AIX 7.2 with TL 7200-05 and SP 10 or later

For installations that use virtual I/O, the following minimum AIX levels are supported:

- AIX 7.3 with TL 7300-03 and SP 0 or later
- AIX 7.3 with TL 7300-02 and SP 2 or later
- AIX 7.2 with TL 7200-05 and SP 8 or later

Important:

- Starting with AIX 7.3 TL 3 SP 1, AIX logical partitions (LPARs) operate in native Power11 processor compatibility mode, fully using Power11 capabilities.
- LPARs running versions earlier than AIX 7.3 TL 3 SP 1 operate in Power9 or Power10 processor compatibility mode on Power11 systems.
- LPARs running AIX 7.2 can operate in Power9 or Power10 processor compatibility mode.

6.2.3 AIX maintenance levels

IBM periodically releases maintenance packages, such as SPs and TLs, for the AIX operating system. For more information about these packages, including how to download and install them, see IBM Fix Central. For more information about hardware feature compatibility and the corresponding AIX TLs, see IBM Support.

The Service Update Management Assistant (SUMA), which is included in the base OS, can help automate the process of checking for and downloading updates. For more information about the **suma** command, see Service Update Management Assistant (SUMA).

The Fix Level Recommendation Tool (FLRT) provides cross-product compatibility information and fix recommendations for IBM products. Use FLRT to plan component upgrades or verify the current health of a system.

The IBM AIX Operating System Service Strategy and Best Practices resource is available at no cost to AIX clients. It provides insight into the AIX service strategy and offers lifecycle information to help maintain your AIX version. For more information about AIX, see AIX on IBM Power.

For more information, see the following resources:

- AIX support lifecycle information
- System Software Maps
- System to AIX maps

6.2.4 Licensing

The AIX OS is available in the following editions:

- AIX Standard Edition
- AIX Enterprise Edition
- AIX Cloud Edition

The Enterprise and Cloud Editions include IBM Power related software that is typically required to manage larger IBM Power environments, including hybrid clouds.

AIX offers two licensing models:

- CPU-based licensing
- Subscription licensing

With CPU-based licensing, AIX licenses can be ordered with the server or purchased later as a Miscellaneous Equipment Specification (MES) upgrade. This license grants the right to use AIX on a specific server. To receive AIX support, you must have a valid Software Maintenance Agreement (SWMA).

Subscription licensing is a newer model that offers greater flexibility. The subscription license includes access to IBM software maintenance for a specified term of one or three years. After the term ends, you can renew the subscription to continue using AIX. This model provides predictable pricing over a defined period with reduced up-front costs.

Another benefit of subscription licensing is that the licenses are customer-number-entitled, which means you can use the licenses on any IBM Power server in your environment and move them between on-premises, private cloud, and public cloud environments as needed.

6.2.5 New AIX editions

The IBM AIX OS is an open standards-based UNIX like system that has been the foundation of mission-critical workloads and databases for more than 35 years. AIX provides an enterprise-class IT infrastructure that delivers the reliability, availability, performance, and security that are required for organizations to succeed in a global economy.

IBM now offers the following updates and enhancements to the AIX OS and products that include AIX for IBM Power8, Power9, Power10, and Power11 processor-based servers:

- IBM AIX 7 Enterprise Edition 1.13
- IBM Private Cloud Edition 1.13
- IBM Private Cloud Edition with AIX 1.13
- VM Recovery Manager (VMRM) 1.9

IBM AIX 7 Enterprise Edition 1.13

IBM has updated AIX 7 Enterprise Edition and its corresponding subscription offering to Version 1.13. Here are the bundled software components that are included with AIX 7 Enterprise Edition 1.13 (5765-CD3 and 5765-2E1):

- IBM AIX 7.3 TL 3 or IBM AIX 7.2 TL 5
- IBM PowerSC 2.3
- IBM PowerVC for Private Cloud 2.3.1
- IBM VM Recovery Manager (VMRM) HA 1.9
- IBM Tivoli® Monitoring 6.3

The bundle components were updated as follows:

- AIX 7.3 TL 3 was updated with SP 1.
- AIX 7.2 TL 5 was updated with SP 10.
- IBM PowerSC was updated from Version 2.2 to Version 2.3.
- IBM PowerVC for Private Cloud was updated from Version 2.3.0 to Version 2.3.1.
- IBM VMRM HA was updated from Version 1.8 to Version 1.9.

Additional information

- Clients with an active SWMA or subscriptions for earlier versions of AIX Standard Edition or AIX Enterprise Edition are entitled to upgrade at no additional cost. To update, download, or install, see the IBM ESS website.
- Clients can choose either AIX 7.3 TL 3 or AIX 7.2 TL 5. Clients who select AIX 7.2 TL 5 can upgrade to AIX 7.3 TL 3 at any time if the SWMA or subscription remains current.
- Clients with AIX Enterprise Edition can trade up to IBM Private Cloud Edition with AIX.

IBM Private Cloud Edition 1.13

IBM has updated the Private Cloud Edition and its corresponding subscription offering to Version 1.13. Here are the bundled software components that are included with Private Cloud Edition 1.13 (5765-ECB and 5765-6C1):

- IBM PowerSC 2.3
- IBM PowerVC for Private Cloud 2.3.1
- IBM VM Recovery Manager (VMRM) DR 1.9
- IBM Tivoli Monitoring 6.3

The following components were updated in this release:

- IBM PowerSC was updated from Version 2.2 to Version 2.3.
- IBM PowerVC for Private Cloud was updated from Version 2.3.0 to Version 2.3.1.
- IBM VMRM DR was updated from Version 1.8 to Version 1.9.
- IBM Cloud Management Console (IBM CMC) was updated from Version 1.22 to Version 1.23.

IBM Private Cloud Edition with AIX 1.13

IBM has updated Private Cloud Edition with AIX and its corresponding subscription offering to Version 1.13. Here are the bundled software components that are included with Private Cloud Edition with AIX 1.13 (5765-CBA and 5765-2C1):

- IBM AIX 7.3 TL3 or IBM AIX 7.2 TL5
- IBM PowerSC 2.3
- IBM PowerVC for Private Cloud 2.3.1
- IBM VM Recovery Manager (VMRM) DR 1.9
- IBM Tivoli Monitoring 6.3

The following components were updated in this release:

- AIX 7.3 TL3 was updated with Service Pack (SP) 1.
- AIX 7.2 TL5 was updated with SP 10.
- IBM PowerSC was updated from Version 2.2 to Version 2.3.
- IBM PowerVC for Private Cloud was updated from Version 2.3.0 to Version 2.3.1.
- IBM VMRM DR was updated from Version 1.8 to Version 1.9.
- IBM CMC was updated from Version 1.22 to Version 1.23.

Additional information for IBM Private Cloud Edition

- Private Cloud Edition 1.13 and Private Cloud Edition 1.13 with AIX 7 include an entitlement for an IBM CMC subscription (5765-CMT) for the same term as the SWMA.
- Clients with an active SWMA or subscription for earlier versions of Private Cloud Edition or Private Cloud Edition with AIX are entitled to upgrade at no additional cost. To update, download, or install, see the IBM ESS website.
- Clients can choose either AIX 7.3 TL 3 or AIX 7.2 TL 5. Clients who select AIX 7.2 TL 5 can upgrade to AIX 7.3 TL 3 at any time if the SWMA or subscription is current.
- Clients with AIX Enterprise Edition can trade up to Private Cloud Edition with AIX.

VM Recovery Manager 1.9

VMRM provides automated VM management in the data center and disaster recovery (DR) management between sites. In PowerVS public cloud environments, the product is called DR Automation. DR Automation provides automated DR management between geographically dispersed PowerVS data centers.

Here are some extra capabilities:

- Workgroup support for HADR
- Failover rehearsal for PowerVS

6.3 IBM i

IBM i is a powerful, integrated OS that is designed to run on IBM Power hardware. It is known for robustness, security, and simplicity of management. Over the years, IBM i has evolved through several naming transitions, reflecting both technological advancements and IBM's strategic branding.

IBM i remains a premier enterprise computing platform, and is recognized for strong security, scalability, and seamless integration. Its ability to support modern workloads while maintaining compatibility with earlier applications makes it essential in industries such as finance, retail, and manufacturing, where reliability and performance are critical. IBM continues to innovate, helping ensure that IBM i, a key component of IBM Power, enables enterprises to manage and deploy their most critical applications efficiently while building on its stability and business continuity.

6.3.1 Introducing IBM i

IBM i is a highly integrated and dependable platform that streamlines IT infrastructure by combining the OS, Db2 for i database, security, and middleware into a single environment. This integration reduces the need for complex configurations and third-party components, making it simpler to deploy, manage, and scale applications. As a result, organizations benefit from a reduced total cost of ownership (TCO) and reduced administrative impact, which is especially valuable in environments with limited IT resources.

One IBM i key strength is its reliability, availability, and serviceability (RAS). Designed for mission-critical workloads, it offers features such as automatic error detection, self-healing capabilities, and robust HADR options. These capabilities make IBM i a trusted platform in industries such as finance, healthcare, and manufacturing, where system uptime and data integrity are essential. In addition, IBM i supports both traditional and modern application development, enabling businesses to run traditional RPG and COBOL applications alongside modern languages such as Java, Python, and Node.js, while integrating with open-source tools and cloud-native services.

For more information about IBM i and its latest functions, see *IBM i 7.6 features and function*, SG24-8588.

6.3.2 Supported levels

The Power E1180 supports the following minimum levels of the IBM i OS both native and as a client of Virtual I/O Server (VIOS):

- IBM i 7.4 Technology Refresh (TR)12
- IBM i 7.5 TR6
- IBM i 7.6

6.3.3 IBM i software tiers

For the Power E1180, all configurations are at the P30 IBM i Software Tier.

6.3.4 New IBM i features that are supported by Power11

The IBM i 7.6 Power11 GA Program Temporary Fix (PTF) Group, IBM i 7.5 Hardware TR 6, and IBM i 7.4 Hardware TR 12 all support the new IBM Power servers with Power11 technology and their new I/O offerings.

Several new features are available only with Power11 processor-based servers and newer IBM i code levels:

- NVMe Hot Spare (IBM i 7.6): Adds the ability to immediately replace a mirror-protected disk unit under a failing Non-Volatile Memory Express (NVMe) device. For more information, see NVMe Hot Spare.
- Secure Boot (IBM i 7.6): Verifies the authenticity of the boot process. Security administrators can extend existing Secure Boot technology, which validates PowerVM firmware (including hostboot, PHYP, and partition firmware [PFW]) to also include IBM i Licensed Internal Code (LIC) through digital signatures. The IBM i OS verifies the integrity of the boot chain up to the LIC before running by using a digital signature and a hardware-protected asymmetric key.
- Common Cryptographic Architecture Cryptographic Service Provider (CCA CSP): Delivered as IBM i Option 35 with IBM i 7.6, 7.5, and 7.4. Supports the IBM 4770 Cryptographic Coprocessor with release 8.4.x for CCA.
- NX-accelerated compression (IBM i 7.5 and 7.6): Enables NX-accelerated compression for PASE programs on Power11 servers. Using the zlibNX library, IBM i accelerates DEFLATE compression in PASE, with the pigz program benefiting automatically. This feature can reduce CPU usage by up to 95% and speed up compression tasks.
- Tape Library Virtualization (IBM i 7.6): Extends support to IBM 3584 tape libraries when running on Power10 and Power11 servers. This enhancement builds on existing virtualization capabilities for libraries such as the IBM 3572, IBM 3573, and IBM 3555. Tape Library Virtualization enables multiple IBM i partitions to share a single physical tape library, reducing infrastructure costs by eliminating the need for a dedicated tape drive per LPAR. Support also extends to Virtual Tape Libraries (VTLs) that emulate the IBM 3584 model. For more information, see IBM i Removable Media: Tape Library Virtualization.

6.3.5 IBM i licensing

On IBM Power11 systems, IBM i licenses are available only through subscription, except for customers performing MES or serial number upgrades from Power E1080 to Power E1180, where licensing remains unchanged.

For new IBM Power10 purchases, IBM announced the withdrawal of perpetual IBM i licenses for the P20 and P30 software tiers in July 2025, effective January 1, 2026. If you purchase a new Power10 system before January 2026, you have the following options:

- You may purchase new subscription licenses.
- You may purchase new perpetual licenses.
- You may transfer existing perpetual licenses.

IBM previously withdrew new perpetual licenses for P05 and P10 tiers.

When an IBM i license is acquired, license keys are generated that are unique to the server on which the license is installed. These licenses fall into two categories:

- Non-expiring (perpetual): License keys have no expiration date.
- Subscription: License keys have an expiration date that matches the subscription term.

IBM i licenses are also categorized into Software Tiers, also known as Processor Groups, which correspond to the size of the hardware platform. The tiers are P05, P10, P20, and P30. The IBM i licensing metric varies by tier; for example, the P30 tier is based on the number of cores.

For many customers, IBM i 7.6 is their first experience with subscription-based licensing. IBM is moving toward a subscription-based license structure across the entire IBM i environment.

Note: IBM's general direction is that all software tiers will move to this subscription model in the future.

For more information about IBM i subscription licensing, including a comprehensive FAQ section, see the *IBM i Subscription Transformation FAQ*.

Perpetual licenses withdrawal: IBM i perpetual licenses were withdrawn for new orders. However, existing customers can migrate perpetual IBM i licenses from Power E1080 to Power E1180 models.

When purchasing new subscription licenses, P30 customers may purchase stand-alone subscriptions for each component or they can purchase new subscriptions bundles.

6.3.6 Subscription bundles

IBM introduces the IBM i P30 Enterprise Edition, a strategic subscription bundle that delivers comprehensive IBM i enterprise capabilities:

- IBM i: An integrated OS that includes database, middleware, security administration, and more.
- Advanced HADR: Maximizes uptime for mission-critical workloads.
- Automation-driven operational management: Increases IT productivity and minimizes disruptions.
- Application development and modernization tools: Accelerates innovation.
- Expert guidance: Helps organizations adopt and realize value quickly.

Table 6-2 shows the products and services that are included with the IBM i P30 Enterprise Edition offering.

Table 6-2 Enterprise Edition subscription bundle

Product name	Description
IBM i	IBM i is a fully integrated OS, meaning that the database, middleware, security, run time, and hypervisor are all included in a single stack and licensed together. This integration helps clients reduce TCO, simplify systems management, and achieve more with fewer resources.
IBM Backup, Recovery, and Media Services for i (BRMS)	BRMS helps implement a disciplined approach to managing backups and provides an orderly method for retrieving lost or damaged data. BRMS is IBM's strategic solution for planning and managing save and restore operations on IBM i.

Product name	Description
IBM PowerHA SystemMirror® for i	IBM PowerHA technology enables deployment of a high availability (HA) solution that addresses storage and availability requirements in a single integrated configuration with a simplified user interface. IBM Power continues to invest in solutions that are designed to keep IT environments resilient.
IBM Db2 Mirror for i	IBM Db2 Mirror for i enables continuous availability for mission-critical applications and provides a Recovery Time Objective (RTO) of zero. It synchronously mirrors database updates between two separate nodes by using Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE) networks. Applications can run in an active-active configuration or in an active-passive configuration with read access on the secondary node.
IBM PowerSC	IBM PowerSC provides endpoint protection for IBM i by enabling real-time security monitoring for continuous visibility and compliance automation. It integrates IBM i security with enterprise Security Information and Event Management (SIEM) systems to support a broader enterprise security strategy. PowerSC operates on top of the IBM Power server stack, integrating security features across multiple layers.
PowerHA Tools for IBM i	PowerHA Tools for IBM i provide customizable templates for HA solutions that are designed to help ensure business continuity by minimizing downtime and simplifying HADR processes: ► IASP Manager ► Full System FlashCopy® (FSFC) Manager ► Full System Replication (FSR) Manager ► Smart Assist for PowerHA on IBM i ► Safeguarded Copy Toolkit PowerHA Tools for IBM i are part of IBM Technology Expert Labs and can be implemented with other chargeable services that are available from IBM Technology Expert Labs.
IBM i Migrate While Active	IBM i Migrate While Active orchestrates the migration of an IBM i LPAR to a different location. The target location can be within the same server, in the same data center, or at a remote site such as PowerVS in IBM Cloud. The process begins by creating a copy of the source LPAR running IBM i 7.4 or later and migrating data from the source to the copy. While the source node remains active, changed data is continuously migrated to the copy node. When the two nodes are synchronized, a cut-over to the new node completes the migration. Migrate While Active is optimized for scenarios where a physical tape cannot be used for a traditional D-mode initial program load (IPL) and restore. It manages the save to optical media so that a network-based installation can be performed.
iDoctor toolkit	iDoctor is a suite of dynamic tools that quickly identify performance issues on IBM i systems. You can monitor overall system health at a high level and use advanced drill-down capabilities to analyze specific issues. iDoctor is an IBM tool that you can combine with other chargeable performance services from IBM Technology Expert Labs.

Product name	Description
IBM Rational Development Studio for i	IBM Rational Development Studio for i is an application development package that includes all server tools and compilers that are required to develop and maintain applications for IBM i. Rational Development Studio for i provides these three features: ► IBM Integrated Language Environment® (ILE) compilers ► Heritage compilers ► Application Development ToolSet (ADTS)
IBM i Extensions for VS Code Enterprise Support	This enterprise support offering provides a curated set of VS Code extensions that enable you to develop IBM i applications by using RPG, COBOL, CL, and SQL in a modern development environment.
IBM i Advisory Connect	IBM i Advisory Connect is designed for large enterprises to provide rapid access to IBM Technology Expert Labs experts for advice and insights about databases, performance, resiliency, security, and automation. In addition to ongoing advice as needed, you can use one remote annual technology assessment for database, performance, resiliency, security, or automation capabilities and practices.
Red Hat Ansible Automation	Ansible Automation Platform is a foundation for building and operating automation across an organization. The platform includes all tools that are required to implement enterprise-wide automation. Ansible managed endpoints can include LPARs on IBM Power servers running IBM AIX, IBM i, or Linux. Ansible Automation Platform provides enterprise subscription and support for Ansible deployments, including certified endpoint module collections for AIX, IBM i, Linux, VIOS, and IBM Power Hardware Management Consoles (HMCs). If you purchase 10 processor cores or more of IBM i P30 Enterprise Edition in a single transaction, you receive a 1-year subscription to Red Hat Ansible Automation Platform for up to 100 managed nodes. After the first year, you can renew the subscription directly from IBM Power or from Red Hat.

IBM i P30 Enterprise Edition with the subscription bundle can be ordered with commitment for 1, 2, 3, 4, or 5 years.

Subscription bundles and IBM Power Enterprise Pools 2.0

IBM Power Enterprise Pools 2.0 (PEP 2.0) enables IBM i OS and PowerHA SystemMirror sharing across systems. For more information about IBM Power Enterprise Pools, see 10.1, “IBM Power Private Cloud with Shared Utility Capacity” on page 198.

In a pool, you can mix systems with the following licenses or subscriptions:

- Subscription licenses
- Subscription bundles: IBM i P30 Enterprise Edition
- Stand-alone subscriptions
- Perpetual licenses

Each system must use a single license type, either subscription or perpetual. Licenses are shared as base capacity or available as metered capacity by the minute in a pool.

Other IBM i licensed program products (LPPs) must be licensed per system based on peak usage. They are not shared as base capacity or available as metered capacity by the minute in a pool.

Note: For a specific serial number, a system can have either an OS subscription or an OS perpetual license, but not both.

IBM Rational Development Studio for i named user license

IBM Rational Development Studio for i named user licenses that are acquired through the IBM i P30 Enterprise Edition can be shared across all systems with Enterprise Edition licenses.

For example, if you have three systems with 10 cores each (30 total cores), you are entitled to 150 named users (five users per core). These 150 named users can be assigned to any or all three systems.

IBM i license terms

There is no change in IBM i license terms for subscriptions, subscription bundles, or perpetual licenses.

Determine the number of processors that is used by the program in a machine by using the smaller value from one of the following methods:

- The total number of activated processors in the machine.
- The sum of a and b is as follows (round any remaining fraction of a processor up to a full processor in the final aggregation):
  a. When IBM i runs in partitions with dedicated processors, use the sum of the processors that are assigned to those partitions.
  b. When IBM i runs in partitions that are members of a shared processing pool (SPP), for each SPP, use the smaller of i or ii:
     i. The maximum capacity of the SPP.
     ii. The sum of the virtual processors of each uncapped partition plus the processing units that are assigned to each capped partition running the program.

6.3.7 IBM i Software Maintenance Agreement

IBM i SWMA is included with subscription licenses. For non-expiring licenses, 1 year of SWMA is included with the license purchase, and extra years are charged separately.

SWMA is important for your business. It enables you to engage with IBM Support centers worldwide and entitles you to software updates (called PTFs on IBM i) and new OS releases. These PTFs provide new functions and enhancements across the IBM i portfolio, including security, database, HA, and performance improvements.

6.3.8 Capacity BackUp and Dedicated Backup Machine Licensing Models for IBM i

In addition to the IBM i licensing model, IBM offers extra licensing models for IBM i customers:

- Capacity BackUp (CBU)

  With the CBU offering, you can temporarily transfer IBM i license entitlements between an IBM i production server and a CBU server when you order a new IBM i server with the CBU feature. The production and CBU servers must both be owned by the same company. CBU provides a cost-effective backup server. For more information, see Capacity Backup.

- Dedicated Backup Machine (DBM)

  This licensing model supports backup environments. One key difference from CBU is that a DBM-licensed machine does not have to belong to the same company as the production server it backs up. Another difference is that for a DBM server, you must acquire IBM i license entitlements for all cores that use IBM i on the server.

For more information about temporary keys for IBM i LPPs for CBU and DBM, see IBM Power Systems: Capacity BackUp.

6.3.9 Upgrading an existing IBM i license to 7.6

If you have an active IBM i SWMA, you can upgrade your IBM i license to Version 7.6 at no additional charge.

If you have IBM i subscription licenses at Version 7.5 or earlier, see the IBM i Subscription Transformation FAQ, and search for "upgrade".

If you have IBM i non-expiring licenses at Version 7.5 or earlier, upgrade IBM i through the IBM ESS website by selecting **My Entitled Software** → **Software Updates**. Select your IBM system by IBM customer number, server model, and serial number. Then, choose the option to upgrade from the current release to Version 7.6. After the process completes, you are directed to the Software Download section, where you can download the installation media to install or upgrade your system.

After you upgrade your IBM i non-expiring or subscription licenses to Version 7.6, use the ESS website to download your IBM i license keys for Version 7.6. Until you install the license keys, a 70-day evaluation period is available. Failure to install the license keys can result in IBM i and LPPs becoming unusable until the appropriate license keys are installed.

Your licensing model does not change during the upgrade process. If you had a non-expiring or subscription license before the upgrade, you have the same license type afterward.

When you upgrade to IBM i 7.6 on your server, upgrades are supported from IBM i 7.4 and 7.5.

6.3.10 IBM i simplification

IBM i and many IBM i portfolio LPPs require license keys to help ensure that the software is properly licensed. Recent simplification reduced the number of products that require keys.

For more information about IBM i portfolio simplification, see IBM i Portfolio Simplification.

6.4 Linux on IBM Power

Linux is a powerful, open-source, cross-platform operating system that runs on a wide range of hardware, from embedded devices to mainframes. It delivers a consistent, UNIX like environment across diverse architectures. Its compatibility with IBM Power provides a flexible and cost-effective option for running many applications while taking full advantage of the platform's performance, availability, and reliability.

As a no-charge and open-source solution, Linux reduces TCO by eliminating licensing fees and enabling deep customization. It supports modern technologies such as containers, Kubernetes, and cloud-native applications, making it ideal for building scalable, agile IT environments. Its portability enables it to run on architectures such as x86, ARM, and IBM Power. A large global community supports Linux by providing documentation, tools, and peer assistance.

IBM has advocated for Linux for more than two decades, integrating it into its enterprise ecosystem, including IBM Power, IBM Z mainframes, and IBM Cloud. IBM collaborates with major Linux distributions such as Red Hat, SUSE, and Canonical to deliver optimized performance and seamless integration. Through initiatives such as LinuxONE and contributions to the Linux kernel and open-source projects, IBM continues to position Linux as a foundation for secure, modern infrastructure.

Security is a key strength of Linux. Its user privilege model, which combined with community-driven updates and tools such as SELinux and AppArmor, helps protect systems from vulnerabilities and malware. Linux is efficient and lightweight, capable of running on devices from low-powered IoT systems to high-performance enterprise servers. Its modular design enables users to tailor the system to specific needs, from minimal installations to full-featured desktop environments.

Linux also provides access to a broad ecosystem of tools and software, some of which are exclusive or better supported on Linux than on other OSs. Its open-source nature offers flexibility and adaptability, making it simple to align with business or technical requirements. Linux serves as a strong foundation for hybrid cloud infrastructure, supporting application modernization and deployment at scale. Its scalability, reliability, and cost-effectiveness make it suitable for large enterprise environments and mission-critical workloads.

6.4.1 Supported distributions

The Linux distributions that are described in this section are supported on the Power E1180. Other distributions, including open-source releases, can run on these servers but do not include formal enterprise-grade support.

Red Hat Enterprise Linux

RHEL on IBM Power servers delivers scalability and performance for high-demand workloads such as big data, cloud, and SAP HANA. The efficiency of Power servers enables organizations to optimize resource usage and reduce costs by using fewer processor cores.

RHEL can be deployed in on-premises and cloud environments, giving organizations control over their data and infrastructure with the flexibility to choose their preferred platform.

The latest version of RHEL is supported in native Power11 mode, which enables access to all features of the Power11 processor and platform.

At launch, IBM Power11 servers support the following RHEL versions:

- RHEL 10 with native Power11 support, fully using the architecture's advanced features
- RHEL 9.6 for Power LE or later (Power11 native support)
- RHEL 9.4 for Power LE or later (Power10 compatibility)
- RHEL 8.10 for Power LE or later (Power10 compatibility)

Red Hat Enterprise Linux roadmap

IBM Power servers support Linux workloads while using the performance, reliability, and availability of IBM POWER hardware. This section outlines supported enterprise Linux distributions. Other distributions might run on IBM Power, but only select enterprise versions include formal support.

RHEL is a leading open-source platform for Linux, hybrid cloud, containers, and Kubernetes. It provides secure, transparent, and proactive lifecycle management, supporting autonomous operations. Lifecycle planning is essential for customers, partners, independent software vendors (ISVs), and the broader ecosystem. Starting with RHEL 8, the lifecycle includes three phases: Full Support, Maintenance Support, and Extended Life. Red Hat also shares projected release timelines and extended support details for minor versions. RHEL 8, 9, and 10 each offer a 10-year lifecycle across the first two phases, followed by Extended Life, as shown in Figure 6-1.

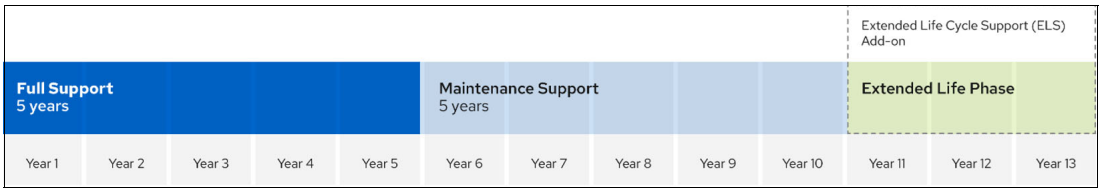

Figure 6-1 RHEL support lifecycle

Note: The RHEL lifecycle phases are designed to minimize changes within each major release and help ensure predictable availability and content.

- Full Support Phase

  During this phase, Red Hat provides updates for the following items:

  - Security issues (Common Vulnerabilities and Exposures (CVEs) with a CVSS score of 7 or higher) through Red Hat Security Advisories
  - Urgent and select high-priority bugs through Red Hat Bug Advisories
  - More errata as needed

  New or improved hardware support and select software enhancements might be included, typically in minor releases. These minor releases are cumulative and focus on resolving medium- or higher-priority issues. Updated installation images are also provided.

- Maintenance Support Phase

  For RHEL 8, 9, and 10, Red Hat continues to release the following updates:

  - Security updates (CVEs with a CVSS score of 7 or higher)
  - Urgent and select high-priority bug fixes

  New features and hardware support are not included during this phase.

- Extended Life Phase

  Customers retain access to existing content through the Red Hat Customer Portal, including documentation and migration guidance. Only limited technical support is available. No new fixes, hardware support, or root-cause analysis are provided. Support applies to existing installations only, and Red Hat may end support at its discretion.

Release cadence

To provide predictability, minor RHEL releases are scheduled for every 6 months during the Full Support Phase. Details about Extended Update Support (EUS) and SAP-specific services are included with each release.

Figure 6-2 shows the planned release schedule for RHEL 8.

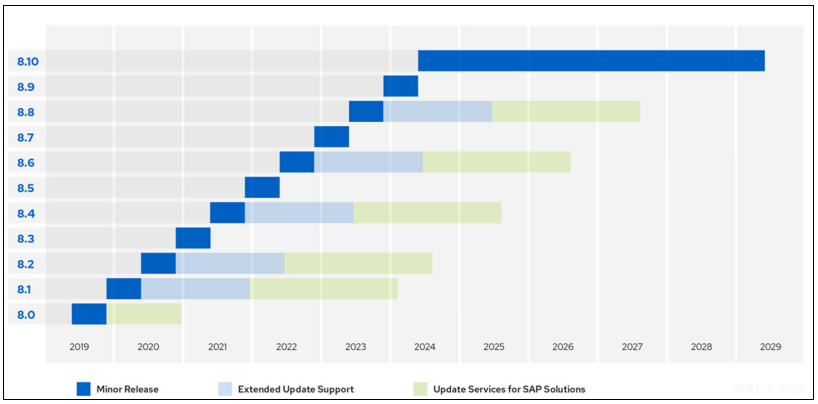

Figure 6-2 Release schedule for RHEL 8

Figure 6-3 shows the planned release schedule for RHEL 9.

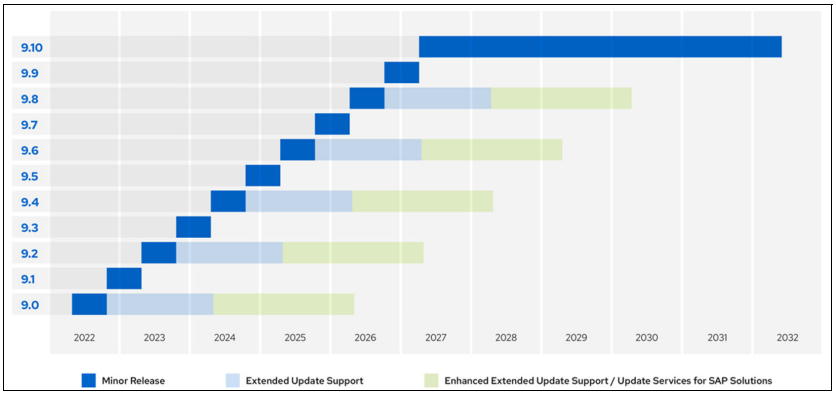

Figure 6-3 Release schedule for RHEL 9

Figure 6-4 shows the release schedule for RHEL 10.

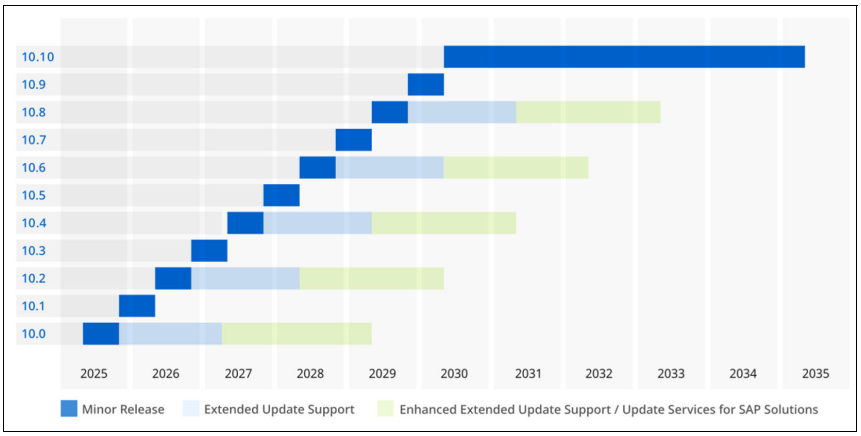

Figure 6-4 Release schedule for RHEL 10

For more information about the RHEL roadmap, see the Red Hat Enterprise Linux Life Cycle. To find the latest Red Hat certified IBM Power servers, see the Red Hat Certified Hardware catalog and System to Red Hat Enterprise Linux maps.

SUSE Linux Enterprise Server

The latest version of the SUSE Linux Enterprise Server distribution of Linux from SUSE is supported in native Power11 mode, which enables it to access all features of the Power11 processor and platform.

At announcement, the Power E1180 servers support SUSE Linux Enterprise Server 15 SP 6 or later (Power11 native)

SUSE roadmap

SUSE Linux Enterprise Server is designed for long-term enterprise stability and support. Its product lifecycle provides predictable and reliable maintenance phases, enabling organizations to plan deployments, updates, and migrations effectively.

- Lifecycle duration

  SUSE Linux Enterprise Server offers a total lifecycle of 13 years per major release, which is divided into two main phases:

  - General Support (10 years): This phase includes full maintenance, security patches, bug fixes, hardware enablement, and new certified third-party software. It is ideal for production environments that require continuous updates and active support.
  - Extended Support (3 years): After General Support ends, Extended Support provides critical security updates and selected bug fixes, enabling customers more time to migrate to newer versions while maintaining operational security.

- Release cadence

  - Major releases: Every 4 years, introducing new features, platform support, and architectural improvements.
  - Service Packs (SPs): Released every 12 - 14 months, SPs include incremental updates, feature enhancements, and hardware enablement.

  Each SP is supported for 6 months after the release of the subsequent SP, providing customers time to validate and upgrade within a consistent and predictable window.

Figure 6-5 shows the SP schedule.

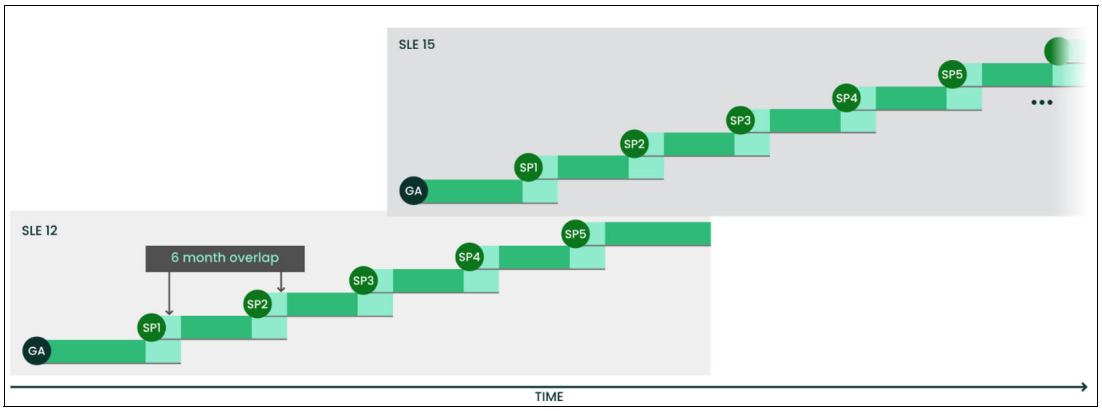

Figure 6-5 SUSE Linux Enterprise Server major releases and Service Packs

Figure 6-6 shows the release lifecycle, including long term SP support.

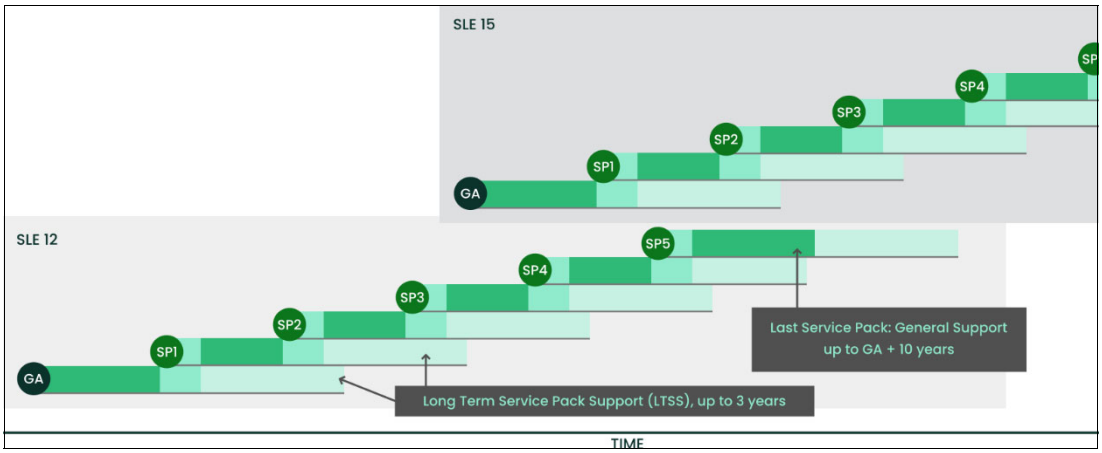

Figure 6-6 SUSE Linux Enterprise Server long-term Service Pack support.

For more information about the SUSE roadmap, see the SUSE Linux Enterprise Server Documentation and Product Support Lifecycle. To find the latest SUSE certified IBM Power servers, see the System to SUSE Linux Enterprise Server maps and SUSE YES Certified Hardware - Bulletin Search.

Linux and Power11 technology

The Power11-specific toolchain is available in IBM Advance Toolchain for Linux 15.0. This toolchain enables customers and developers to use all new Power11 processor-based technology instructions when programming. Cross-module function call processing impact is reduced because of a new PC-relative addressing mode.

One key benefit of Power11 technology is a 10x to 20x performance advantage over Power9 processor-based technology for AI inferencing workloads. This improvement is due to increased memory bandwidth and new instructions. For example, the new special-purpose Matrix Math Accelerator (MMA) is designed for machine learning and deep learning inference and supports many AI data types.

6.4.2 Licensing

Linux licensing on IBM Power follows the same foundational principles as on other platforms, with considerations that are tailored to enterprise environments. The Linux kernel is licensed under the GNU General Public License Version 2 (GPLv2), which helps ensure that the software remains no-charge and open source, enabling users to run, study, modify, and distribute it. On IBM Power, several major Linux distributions are supported, including RHEL, SUSE Linux Enterprise Server, and Ubuntu. Each distribution adheres to open-source licensing models but might differ in how it packages and delivers software.

For example, RHEL and SUSE Linux Enterprise Server provide access to source code under open-source licenses but require a paid subscription for access to precompiled binary files, updates, and enterprise support. This model enables organizations to benefit from open-source flexibility while receiving the stability and support required for mission-critical workloads.

On IBM Power, these distributions are optimized to use the architecture's performance, scalability, and reliability features. Licensing terms typically include support for virtualization technologies such as PowerVM.

Red Hat licensing

RHEL is sold on a subscription basis, with initial subscriptions and support that are available for 1, 3, or 5 years. Support is available directly from Red Hat or through IBM Technical Support Services.

RHEL 8 for Power LE subscriptions cover up to four cores and up to four LPARs and can be stacked to cover more cores or LPARs.

When you order RHEL from IBM, a subscription activation code is automatically published in IBM ESS. After retrieving this code from IBM ESS, use it to establish proof of entitlement and download the software from Red Hat.

For more information, see the Red Hat Enterprise Linux subscription guide.

SUSE Linux Enterprise Server licensing

SUSE Linux Enterprise Server is sold on a subscription basis, with initial subscriptions and support that are available for 1, 3, or 5 years. Support is available directly from SUSE or through IBM Technical Support Services.

SUSE Linux Enterprise Server 15 subscriptions cover one socket or one LPAR and can be stacked to cover more sockets or LPARs.

When you order SUSE Linux Enterprise Server from IBM, a subscription activation code is automatically published in IBM ESS. After retrieving this code from IBM ESS, use it to establish proof of entitlement and download the software from SUSE.

Clients must register the purchased Linux offering on the distributor's website by using the activation code. After registration, clients can download the software packages electronically and obtain the latest available upgrades for the purchased product.

For more information about SUSE Linux Enterprise Server licensing information, see SUSE Licensing.

6.4.3 Kernel-based Virtual Machine support

Kernel-based Virtual Machine (KVM) is a widely used virtualization technology on x86_64 systems that can also be used on IBM Power servers. IBM remains committed to PowerVM as the premier enterprise virtualization software in the industry.

With KVM on IBM Power, IBM targets x86 customers on entry servers but offers both KVM and PowerVM to meet the varying virtualization needs of PowerLinux customers. KVM provides an opportunity to simplify virtualization infrastructure by using a single hypervisor and management software across multiple platforms.

KVM guests can run within a PowerVM LPAR. These guests are VMs that run on top of the LPAR and use its existing resources.

One benefit of this approach is that it combines the advantages of PowerVM virtualization capabilities with the flexibility of KVM. KVM guests running in a PowerVM LPAR have a unique runtime architecture that differs from other virtualization mechanisms on IBM Power.

For more information about KVM running on IBM Power11, see 9.2, “KVM support” on page 194.

6.5 PowerVM Virtual I/O Server

IBM PowerVM software is a virtualization environment that can run AIX, IBM i, and Linux VMs on IBM Power servers. Businesses use server virtualization to consolidate multiple workloads onto fewer systems, increase server usage, and reduce costs. PowerVM provides a secure and scalable virtualization environment for applications that is built on the advanced RAS features and performance of the IBM Power platform.

PowerVM is designed to protect and isolate critical workloads through a highly secure, enterprise-grade hypervisor. It enforces strong workload isolation and I/O integrity, helping ensure the reliability of mission-critical applications. With robust automation capabilities, PowerVM accelerates service delivery by streamlining the provisioning and management of VMs and storage resources, making it suitable for cloud-based infrastructures. It also enhances operational efficiency and maximizes return on investment through features such as Live Partition Mobility (LPM), which enables zero-downtime workload migration, and resource optimization tools that improve the usage of compute and storage resources.

The VIOS is part of the PowerVM Editions hardware feature. VIOS is software that runs in an LPAR and facilitates the sharing of physical I/O resources between client LPARs within the server.

At the time of writing, IBM Power11 processor-based servers support the following minimum levels of the PowerVM VIOS OS:

- VIOS 4.1.1.10
- VIOS 4.1.0.40
- VIOS 3.1.4.60

6.6 Setting your LPAR compatibility mode

IBM Power Processor Compatibility Mode is a feature that enables newer IBM Power servers to run applications and OSs that were originally compiled for earlier generations of POWER processors. This capability is especially important when migrating to the latest hardware, such as Power11, where some OSs and applications might not yet support native run times. In such cases, a compatibility mode enables these workloads to run in a Power10 environment, helping ensure continuity and minimizing disruption during upgrades.

Compatibility mode is defined at the LPAR level during LPAR creation. Administrators can specify which processor generation that the partition should emulate, enabling multiple workloads to operate in different compatibility modes on the same physical server. This flexibility is useful in mixed environments where not all software is updated to support the latest architecture.

Although running in a previous-generation compatibility mode helps maintain application portability and simplifies migration, it might restrict access to newer processor features and performance enhancements. For this reason, use compatibility mode as a temporary solution. The long-term goal should be to recompile applications and upgrade OSs to use the advanced capabilities and performance improvements of the latest Power Architecture.

Figure 6-7 shows how to set compatibility mode in your LPAR definition.

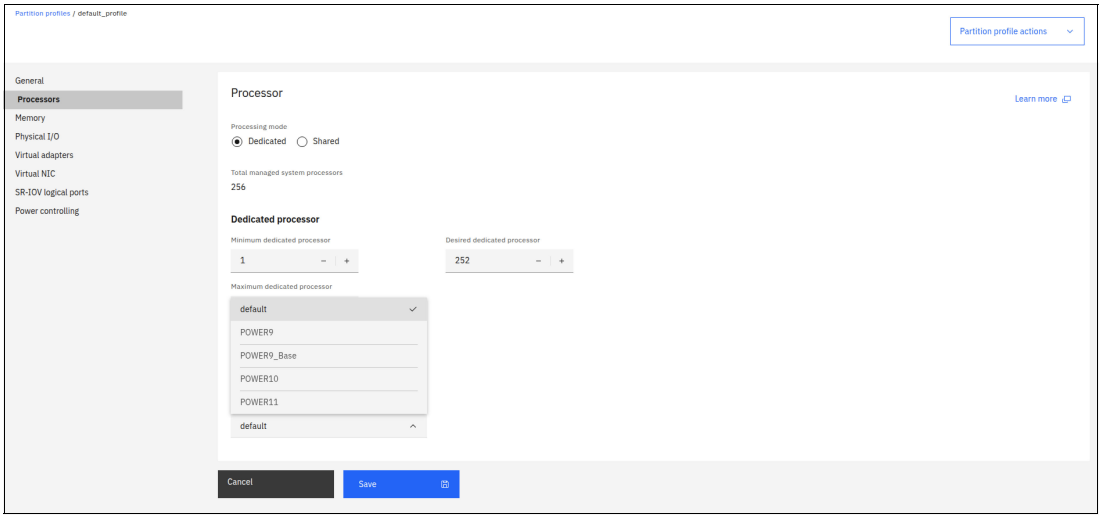

Figure 6-7 Setting the LPAR processor compatibility mode

7

Enterprise solutions

IBM Power servers support a wide range of enterprise workloads that require high availability (HA), performance, and scalability. These systems are ideal for mission-critical applications such as SAP HANA, Oracle, and IBM Db2, where consistent uptime and fast transaction processing are essential. With advanced reliability, availability, and serviceability (RAS) features and capabilities such as Live Partition Mobility (LPM), Power servers minimize downtime and enable seamless workload management. They are also suitable for enterprise resource planning (ERP) and customer relationship management (CRM) systems, providing the processing power and memory capacity that are needed to handle large-scale, real-time data operations.

In addition to traditional workloads, IBM Power servers meet modern enterprise requirements, including artificial intelligence (AI), machine learning, and hybrid cloud deployments. Built-in accelerators and support for containerized environments such as Red Hat OpenShift enable organizations to run AI models and cloud-native applications efficiently. These servers also perform well in high-performance computing (HPC) scenarios, such as financial modeling, scientific research, and healthcare analytics, because of their high core counts and memory bandwidth. With robust security features such as Transparent Memory Encryption (TME) and Secure Boot, IBM Power provides a trusted platform for industries with strict compliance requirements. Overall, Power servers deliver a flexible, secure, and cost-effective foundation for evolving enterprise IT landscapes.

This chapter contains the following topics:

- High availability and disaster recovery solutions
- IBM Db2
- Oracle
- SAP HANA
- Banking
- Healthcare

© Copyright IBM Corp. 2025.

7.1 High availability and disaster recovery solutions

IBM offers a suite of high availability and disaster recovery (HADR) solutions for the IBM Power platform, each designed to meet different operational needs and workloads. PowerHA SystemMirror for AIX is a clustering solution that improves application uptime through an active-standby model, enabling fast failover and minimal disruption during outages or maintenance. For IBM i environments, PowerHA SystemMirror for i provides integrated HADR capabilities by using features such as independent auxiliary storage pools (IASPs) and real-time data replication to protect mission-critical workloads. VM Recovery Manager (VMRM) complements these solutions as a flexible option that supports AIX, IBM i, and Linux workloads by managing full logical partition (LPAR) recovery across systems. VMRM is suitable for environments where full system recovery is acceptable, offering automated failover and disaster recovery (DR) without requiring application-level configuration. Together, these tools deliver a comprehensive HADR strategy across the IBM Power ecosystem.

7.1.1 PowerHA SystemMirror for AIX

PowerHA SystemMirror for AIX, commonly known as PowerHA, is a HA clustering solution that improves application uptime and provides fast recovery. It uses an active-standby or active-passive model of application availability. This model is also called an application restart model, where an application and its resources (such as a file system with data, an IP address, application scripts, and a volume group) are logically combined and operate as a single entity on a virtual machine (VM) or LPAR, known as a node, in a cluster. The resource group moves as a unit to a standby VM or LPAR, and the application restarts.

During a planned outage, PowerHA can gracefully quiesce the application on the primary node. The outage lasts until the node acquires the resources and restarts the application, which is typically measured in minutes. This approach is considered a warm standby method of HA because the target node already runs a version of AIX.

PowerHA for AIX can also operate across data centers, where data is replicated through the storage subsystem or, by using the AIX Geographic Logical Volume Manager (GLVM) subsystem, over IP to the target node. In both single-site and multisite configurations, the failover process can be automated to reduce application outages. Applications can move to another running node to allow maintenance, enabling near-zero downtime, although the PowerHA application itself can be maintained with a live kernel update.

PowerHA for AIX integrates tightly with the AIX operating system's inherent reliability and IBM Power hardware features such as virtual I/O, concurrent hardware maintenance, and dynamic resizing of LPARs. PowerHA supports critical enterprise applications such as Oracle, SAP, Db2, and Epic systems. Licensing follows an *n*+1 model per cluster, where *n* is the number of cores in the production online copy of the application, with one additional core for the standby. For example, on a 24-core Power server, if two clusters are created with one supporting an LPAR with four cores and another with five, a total of 11 cores must be licensed (4 + 1 = 5 and 5 + 1 = 6).

7.1.2 IBM PowerHA SystemMirror for i

PowerHA for IBM i is IBM’s HADR solution for the IBM i operating system. It provides automated failover and data replication to maintain business continuity during planned or unplanned outages. PowerHA integrates with IBM i native features, such as IASPs to enable fast and efficient switching between systems with minimal downtime.

At the core of PowerHA is its ability to replicate data in real time between primary and backup systems by using technologies such as Geographic Mirroring or IBM storage-based replication. This capability helps ensure that critical applications and data remain available even when the primary system becomes unavailable. PowerHA supports both local HA within a single data center and remote DR across geographically dispersed locations, offering flexibility for various business continuity strategies.

PowerHA also includes tools for monitoring, automation, and management, which simplify configuration and maintenance of HADR environments. Features such as role-based access, cluster resource groups, and automated failover policies help reduce the complexity of managing HA systems. PowerHA is a key component for organizations that rely on IBM i to run mission-critical workloads and require continuous access to data and applications.

7.1.3 VM Recovery Manager

VMRM operates on an active-inactive model, where the entire definition of a VM or LPAR is re-created on a target system and then restarted. It supports any workload running on IBM Power, including AIX, IBM i, and Linux. Recovery takes longer than an application restart model because the entire LPAR must be restarted after being redefined, rather than restarting only the application.

In HA mode, VMRM uses LPM to dynamically move an LPAR to a system that shares a Hardware Management Console (HMC). In DR mode, it manages the data replication process from the storage provider to a set of disks at the target system. The LPAR exists in only one location at a time, except in limited circumstances where a DR exercise can be tested on the target system.

Like PowerHA, VMRM automates DR by detecting when a VM or an entire system is down and initiating a policy-based failover to a single target or a group of targets. It can coexist with PowerHA, allowing a cluster of LPARs to fail over as a group to systems at an alternative site. VMRM is often used to provide simple DR support for Linux VMs because it requires no software installation or OS configuration: The Linux VM simply restarts.

The ideal use cases for VMRM include development, test, and sandbox environments that do not require the same recovery time objective (RTO) as production systems, and Linux on IBM Power. It integrates with solutions such as SAP HANA and Oracle. No configuration work is required on the individual OSs that are managed, although optional agents can be installed to check for application failures and trigger an LPAR failover.

Similar to PowerHA, VMRM licensing is based on the number of production cores in use, not the entire system. However, unlike PowerHA, no standby licenses are required. For example, on a 24-core system, if the VMs requiring VMRM support use nine cores, only nine cores must be licensed.

7.2 IBM Db2

In today’s data-driven world, organizations across industries, including manufacturing, healthcare, and the public sector, must extract maximum value from rapidly growing volumes of information. Whether the goal is to uncover actionable insights from customer data or accelerate the processing of high-volume online transactions, businesses need solutions that are optimized for performance, scalability, and cost efficiency. These solutions must support mission-critical workloads with HA while keeping operational costs under control.

By combining IBM Db2 with IBM Power servers that are built on the advanced Power11 architecture, organizations can meet these demands with confidence. This integrated, workload-optimized stack delivers exceptional performance for both transactional and analytical workloads. Db2 uses advanced database innovations to maximize efficiency and throughput, earning top rankings in industry benchmarks such as Transaction Processing Performance Council Benchmark C (TPC-C), Transaction Processing Performance Council Benchmark H (TPC-H), and SAP SD 3-Tier. IBM Power provides the robust hardware foundation, offering the ability to handle more than one million transactions per minute at a cost of less than $1 per transaction[1]. Db2 automatically uses Power11 parallelism and large page sizes, simplifying deployment and enabling cost-effective scaling for web applications, messaging backbones, and workload consolidation.

Beyond performance, the combination of Db2 and IBM Power delivers HA and operational efficiency. Power servers include built-in redundancy, error handling, and reliability features, while IBM PowerHA SystemMirror adds automated monitoring and recovery to minimize downtime. Cost control is further enhanced through Db2 Deep Compression, which reduces storage requirements, and IBM PowerVM virtualization, which enables efficient resource usage and workload consolidation. These capabilities reduce hardware, energy, and management costs, creating a more agile and cost-effective IT infrastructure. Together, Db2 and IBM Power provide a powerful, scalable, and resilient platform for modern data-driven enterprises.

For more information about Db2 HA, see *IBM PowerHA SystemMirror for AIX Cookbook*, SG24-7739.

7.3 Oracle

For more than 35 years, clients have relied on IBM Power to deploy Oracle database and application workloads. Organizations of all sizes can use IBM Power reliability and security, and its advanced recovery, self-healing, and diagnostic capabilities to reduce application downtime.

IBM Power11 servers, fully certified by Oracle software, enable organizations to consolidate multiple workloads on fewer servers, increasing overall system usage and reducing costs. This higher usage enables clients to maximize Oracle license usage while reducing IT costs by managing fewer physical machines.

The cost of security breaches continues to grow, averaging USD 4 million per breach[2]. IBM Power technology helps protect businesses against cyberthreats with end-to-end security protocols, including Transparent Memory Encryption (TME), which does not impact performance. Power servers are considered more secure than other servers based on the number of security incidents that are reported. Combined with IBM AIX Trusted Execution (TE), IBM Power servers enable you to verify system integrity and implement advanced security policies to enhance overall system trust.

As data volumes continue to expand, organizations often struggle to modernize their IT infrastructure. IBM Power servers and the IBM AIX OS provide a solid foundation for modernizing traditional Oracle database workloads, developing new applications, and consolidating workloads. With the Red Hat Ansible Automation Platform, clients can manage Oracle workloads on IBM Power servers as part of a broader enterprise automation strategy.

[1] Source: https://www.tpc.org/tpcc/results/tpcc_results5.asp?print=false&orderby=priceperf&sortby=asc
[2] Source: https://www.ibm.com/think/insights/cost-of-a-data-breach-2024-financial-industry

Mission-critical workloads require a server and operating system with reliability, HA, and the ability to scale without impacting performance. IBM Power servers deliver up to 2.5 times better per-core performance compared to x86 servers[3], enabling Oracle workloads to scale linearly without bottlenecks.

Oracle certifies its products on IBM Power, delivering benefits such as comprehensive end-to-end support, portability, and efficiency. IBM Power servers provide 99.999% reliability to maintain maximum availability. The combined design of the AIX operating system on IBM Power servers with IBM PowerHA technology delivers exceptional uptime and the ability to manage and monitor availability to prevent both planned and unplanned outages.

Benefits of running Oracle on Power11

Here are some benefits of running your Oracle database workloads on IBM Power11:

- Industry-leading security

  IBM Power and AIX keep critical Oracle workloads protected and available while reducing costs.

- Simplified management

  Automatically deploy a cloud-ready operating system that meets private cloud requirements with IBM Power Virtualization Center (PowerVC).

- Unmatched uptime

  IBM Power supports the most demanding workloads and provides 99.999% reliability to maintain maximum availability.

- Improved workload cost-effectiveness

  By using Power LPARs and dynamic logical partition (DLPAR) eligibility for Oracle hard partitioning, clients can license only the Power cores that are available to Oracle software.

- Standards-based automation

  Improve manageability and scalability while helping ensure consistent, repeatable outcomes by using IBM's enhanced automation portfolio for Oracle workloads, built on Ansible.

- Data protection

  Benefit from end-to-end security in Power with trusted boot, main memory encryption, runtime verification of OS files, and role-based access and execution control with PowerSC.

For more information about running Oracle on IBM Power, see *Oracle on IBM Power Systems*, SG24-8485.

7.3.1 Running Oracle Standard Edition 2 on IBM Power

Oracle Standard Edition 2 (SE2) is a database option that is suitable for many business needs and can run on IBM Power servers. SE2 offers a full-featured database with simplicity of use, performance, and flexibility, including support for relational, JSON, and XML data handling, and Oracle Real Application Clusters for clustering services. SE2 is licensed on servers with up to two sockets, with licensing costs remaining the same regardless of the number of cores within each socket.

[3] Source: https://newsroom.ibm.com/2021-09-08-IBM-unveils-new-generation-of-IBM-Power-servers-for-frictionless,-scalable-hybrid-cloud

Key benefits of using Oracle Standard Edition 2 on IBM Power servers

Here are some key benefits of running Oracle SE2 on IBM Power:

- Licensing

  SE2 can be licensed with either the processor metric or the Named User Plus metric. Licensing rules include a maximum of two occupied sockets and a maximum of 16 CPU threads per socket, effectively limiting the number of cores.

- IBM Power compatibility

  Oracle Database SE2 is compatible with IBM Power. IBM has designed specific Power10 servers, such as IBM Power S1012 and IBM Power S1014, to compete with x86 systems in the two-socket space, addressing Oracle's requirement for SE2 to remain cost-competitive with SQL Server. IBM is expected to offer a Power11 server in the future that will also qualify for Oracle Standard Edition 2 licensing.

Benefits of IBM Power for Oracle Workloads

IBM Power offers several advantages for Oracle databases, including main memory encryption, runtime verification of operating system files, role-based access control (RBAC), and advanced recovery, self-healing, and diagnostic capabilities.

- HA

  Starting with Oracle Database 19c Release Update 19.7, Standard Edition High Availability is supported on IBM AIX. This feature provides cluster-based database failover for Oracle Database 19c Standard Edition.

- Other features

  Oracle SE2 includes a range of capabilities for building business applications, such as support for relational, JSON, XML, spatial, graph, and unstructured data, and Oracle Multitenant Architecture.

Licensing and implementation

Consider the following items when planning your Oracle SE2 implementation:

- Although the number of cores in each socket does not affect the license price, as a best practice, adhere to the maximum number of CPUs that are allowed for the cluster, which applies to the entire cluster, not per node.
- Standard Edition High Availability provides cluster-based failover for single-instance Oracle Database Standard Edition environments by using Oracle Clusterware.
  - Oracle Standard Edition High Availability benefits from the cluster capabilities and storage solutions that are part of Oracle Grid Infrastructure, such as Oracle Clusterware, Oracle Automatic Storage Management (Oracle ASM), and Oracle ASM Cluster File System (Oracle ACFS).
  - Using integrated, shared, and concurrently mounted storage, such as Oracle ASM and Oracle ACFS for database files and unstructured data, enables Oracle Grid Infrastructure to restart an Oracle Database on a failover node faster than cluster solutions that rely on failing over and remounting volumes and file systems.
  - Starting with Oracle Database 19c Release Update (19.13), Standard Edition High Availability is supported on IBM AIX (64-bit).

7.4 SAP HANA

IBM Power offers several options to meet clients wherever they are on their journey to SAP S/4HANA. Whether you are considering SAP HANA or the next-generation S/4HANA, IBM Power provides on-premises, off-premises, or fully managed (RISE with SAP) solutions. These solutions are built to run mission-critical applications such as SAP, accelerate ERP and application deployments, and maximize the impact on data management, data integration, automation, and business processes.

IBM and SAP have been partners for more than 50 years and produced the following results:

- Over 4,800 clients run SAP HANA on IBM Power servers.
- More than 120 external client references for SAP HANA on IBM Power.
- 39 SAP Pinnacle Awards won by IBM.
- 30,000 organizations run essential workloads such as SAP on IBM Power.

The key to IT efficiency and business continuity is a platform that integrates with your infrastructure while supporting digital transformation. IBM Power servers are designed for data-intensive applications such as SAP HANA and S/4HANA that require large amounts of in-memory computing while maintaining the HA and flexibility that are required for hybrid cloud environments.

Benefits of running SAP HANA on IBM Power

With global data volumes projected to exceed 180 zettabytes in 2025, organizations across every sector face tremendous pressure to manage, process, store, and extract valuable insights from critical data. By running SAP HANA on IBM Power servers, businesses can achieve the following benefits:

- Provision faster.

  Simplify system management and boost business agility. Create environments flexibly by allocating resources incrementally, starting with a minimum of 0.01 cores and 1 GB of memory.

- Maximize uptime.

  Minimize disruption to business operations with best-in-class reliability for 15 consecutive years[4].

- Cut energy usage.

  Reduce data center costs and enhance environmental sustainability. IBM Power delivers better performance while using less energy compared to x86-based solutions[5].

- Scale affordably.

  Reduce the risk of overprovisioning with flexible hybrid cloud solutions, instant scaling, and pay-per-use consumption options. Scale up to 40 TB, which is the largest capacity that is supported for SAP S/4HANA and SAP BW.

[4] Source: https://www.itjungle.com/2024/10/21/ibm-nears-the-end-of-the-road-for-server-reliability-improvements/

[5] Source: https://www.techspot.com/news/108599-ibm-introduces-power11-servers-boosts-uptime-security-energy.html

- Strengthen security.

  Protect critical data and applications from cyberthreats with end-to-end security, including TME with no performance impact. Power servers are considered more secure than other servers based on the number of security incidents that are reported.

- Gain faster insights.

  Make rapid decisions to maximize business efficiency with up to 2.5 times better per-core performance compared to x86 servers[6].

7.5 Banking

As financial institutions accelerate digital transformation, IT architects must design platforms that are secure, scalable, and AI-ready. IBM Power servers, especially models that are based on Power10 and Power11, provide a robust foundation for deploying AI workloads in banking environments. With built-in AI acceleration, enterprise-grade RAS features, and seamless integration with hybrid cloud and container platforms, IBM Power enables architects to modernize their infrastructure while embedding intelligence into core banking workflows.

Hybrid cloud deployment models for AI in banking

Hybrid cloud is the preferred architecture for modern banking platforms because it enables agility, regulatory compliance, and cost optimization. IBM Power supports hybrid cloud deployments by using the following products:

- IBM Power Virtual Server (PowerVS) for public cloud scalability
- Red Hat OpenShift for container orchestration across environments
- IBM Cloud Pak for Data and watsonx for AI lifecycle management

Banking use cases on IBM Power that use AI

Here is a list of some of the use cases for AI for banks and other financial institutions:

- On-premises AI training with cloud-based inference

  Banks often must train AI models on-premises to comply with strict data residency and regulatory requirements. Using IBM Power10 or Power11 systems, financial institutions can securely train models on sensitive customer or transactional data. Once trained, these models can be deployed to PowerVS for inference, enabling scalable, real-time decision-making in the cloud. This hybrid approach is ideal for applications such as real-time credit scoring or fraud detection, where data privacy is critical but rapid response times are also essential.

- End-to-end AI on PowerVS with secure on-premises data access

  In this model, banks run the full AI lifecycle (training, tuning, and inference) on PowerVS while maintaining secure access to on-premises data sources. This setup enables institutions to use the scalability and flexibility of the cloud without moving sensitive data offsite. For example, a bank might deploy a generative AI model on PowerVS to automate the summarization of complex financial reports, pulling data securely from internal systems. This approach helps ensure compliance with data governance policies while accelerating insights and reporting.

[6] Source: https://newsroom.ibm.com/2021-09-08-IBM-unveils-new-generation-of-IBM-Power-servers-for-frictionless,-scalable-hybrid-cloud

- Distributed AI microservices across hybrid cloud

  Some banks adopt a microservices architecture, distributing AI components across both on-premises IBM Power and PowerVS in the cloud. This model supports modular, scalable AI applications that operate across environments. For example, asset valuation engines or personalized client engagement tools can run inference in the cloud while accessing real-time data from on-premises systems. This approach enables agility and performance, especially for institutions managing diverse workloads worldwide or across business units.

- AI for risk and compliance on IBM Power

  Global Tier-1 banks use AI to automate risk management and compliance processes. By deploying generative AI and machine learning models on IBM Power, these institutions can continuously monitor transactions, flag anomalies, and generate regulatory reports with minimal human intervention. This model supports high-throughput, low-latency processing, making it ideal for meeting stringent regulatory requirements such as anti-money laundering (AML), know-your-customer (KYC), and Basel III compliance.

- AI-enhanced core banking applications

  IBM Power also supports the integration of AI directly into core banking platforms. This approach enables real-time intelligence within mission-critical applications such as loan underwriting, treasury operations, and fraud analytics. For example, a bank might use AI models running on Power11 to detect unusual transaction patterns or predict liquidity needs, enhancing operational efficiency and decision-making. This model provides AI capabilities that are embedded where they deliver the most immediate business value.

Integrating IBM watsonx with IBM Power

IBM watsonx is a modular AI and data platform that is designed to accelerate the deployment of enterprise-grade AI. When integrated with IBM Power10 and Power11, watsonx enables financial institutions to build, deploy, and govern AI models across hybrid cloud environments, including PowerVS.

Here are the components of watsonx:

- IBM watsonx.ai: Model training and inference
- IBM watsonx.data: Data lakehouse integration
- IBM watsonx.governance: Model risk and compliance
- IBM watsonx.orchestrate: AI agents and workflow automation

IBM Power servers are high-performance systems that are optimized for data-intensive workloads. The integration between watsonx and IBM Power provides the following benefits:

- Accelerated AI workloads: Power10 processors are optimized for AI inference and training, making them ideal for running watsonx models efficiently.
- Hybrid cloud flexibility: watsonx can run on IBM Power in hybrid cloud environments, enabling businesses to keep sensitive data on-premises while using cloud-native AI capabilities.
- Enterprise-grade security and reliability: IBM Power delivers robust security and uptime, aligning with the watsonx focus on trustworthy AI.
- Tight integration with Red Hat OpenShift: Both watsonx and IBM Power support Red Hat OpenShift, enabling containerized AI workloads to run seamlessly across environments.

7.6 Healthcare

IBM Power provides a secure, high-performance, and resilient infrastructure that enables healthcare organizations to manage and protect digital health data, drive data-informed decisions, and meet regulatory requirements with confidence. Designed for mission-critical workloads, IBM Power helps healthcare providers improve patient outcomes, enhance operational efficiency, and reduce risk.

IBM Power enables the following benefits for healthcare:

- Security and compliance

  Healthcare organizations handle sensitive patient data that must be protected under regulations such as HIPAA. IBM Power delivers advanced security features, including Secure Boot, RBAC, and encryption at rest and in transit. With quantum-safe firmware signing and encrypted LPM, IBM Power helps ensure data integrity and privacy. According to Information Technology Intelligence Consulting (ITIC), organizations that use IBM Power experience an average of just 3.3 minutes of unplanned downtime per year due to security issues, demonstrating its industry-leading resilience[7].

- Unmatched reliability for clinical systems

  IBM Power is engineered for continuous availability of mission-critical applications such as electronic health records (EHRs), PACS, and healthcare information systems (HISs). Features such as Predictive Failure Analysis (PFA), dynamic resource allocation, and system redundancy contribute to its 99.9999% availability rating, as reported by 1,900 C-level executives in ITIC's global reliability survey[8]. IBM Power has been the most reliable non-mainframe server platform for more than 15 years.

- High-performance computing for healthcare analytics

  Healthcare workloads, from medical imaging to genomics and real-time analytics, demand exceptional compute power. Power11 processor-based servers deliver up to 2.5 times more performance per core than previous generations[9], enabling faster insights for research, drug discovery, and personalized medicine.

- Scalable and flexible infrastructure

  Healthcare environments often face unpredictable demand. IBM Power supports dynamic scaling to accommodate workload fluctuations, whether driven by seasonal surges or new clinical initiatives. Its support for multiple operating systems and virtualization technologies enable healthcare IT teams to run diverse workloads efficiently on a single platform.

- AI-ready for clinical innovation

  IBM Power is optimized for AI and machine learning workloads, with on-chip Matrix Math Accelerators (MMAs) that enable real-time inferencing at the point of care. Power S1022 servers can process up to 42% more batch queries per second than comparable x86 systems under peak load, with subsecond inferencing latency[10], which is ideal for applications such as diagnostic imaging, predictive analytics, and treatment optimization.

[7] Source: https://itic-corp.com/itic-2022-global-server-reliability-results/f
[8] Source: https://www.itjungle.com/2024/10/21/ibm-nears-the-end-of-the-road-for-server-reliability-improvements/
[9] Source: https://newsroom.ibm.com/2025-07-08-ibm-power11-raises-the-bar-for-enterprise-it
[10] Source: https://www.ibm.com/new/announcements/combining-ai-with-a-trusted-data-approach-on-ibm-power-to-fuel-business-outcomes

- Seamless integration with healthcare ecosystems

  IBM Power integrates with a wide range of healthcare applications, including electronic medical records (EMRs), imaging platforms, analytics tools, and telemedicine systems. It supports interoperability and data exchange across systems and works closely with leading independent software vendors (ISVs) such as Epic to certify hardware for industry-standard healthcare software.

7.6.1 Epic

Epic is a HIS provider that develops and delivers a comprehensive EMR platform covering all aspects of the healthcare profession. The Epic solution includes several applications for areas such as medical billing, emergency care, radiology, outpatient services, inpatient care, and ambulatory care.

IBM and Epic Systems maintain a strategic partnership that enables healthcare organizations to run Epic's EHR platform on IBM infrastructure, particularly IBM Power and IBM Storage solutions. This collaboration is designed to support the performance, security, and compliance requirements of healthcare providers managing mission-critical clinical workloads.

IBM Power servers are certified to run Epic Operational Database (ODB) workloads, with best practices that are published for tuning AIX environments to meet Epic's stringent performance requirements. These configurations are optimized for high throughput, low latency, and HA, which are critical for real-time clinical operations. At the time of writing, IBM Power10 servers are certified by Epic Systems, supporting key infrastructure components on both AIX and Linux for large healthcare deployments, primarily in the database and back-end services tier. These platforms host InterSystems Caché or IRIS for Health, which serve as the core databases behind Epic's EHR system.

IBM and Epic are expected to certify IBM Power11 configurations after the appropriate testing is completed.

Supported products

Here is a list of supported Epic products at the time of writing:

- InterSystems Caché and IRIS for Health
  - Officially supported by Epic on both AIX and Linux (Red Hat Enterprise Linux (RHEL)) on Power10.
  - Deployed in production for the Epic ODB (Chronicles).
- Epic Production and Reporting Database Tiers

  Most often run on AIX or RHEL on IBM Power10 for performance and reliability.
- Other back-end services, such as database copies that are used for Clarity extracts, may also run on AIX or Linux, depending on the customer's architecture.

Note: The Epic application layer (Hyperspace, Interconnect, and others) typically runs on Windows or Linux on x86. IBM Power processor-based AIX or Linux deployments are focused on the infrastructure and database tiers.

Here are the required OS and firmware levels for Power10:

- AIX
  - AIX 7.2 TL5 SP4 or later
  - AIX 7.3 TL1 or later

  Certified by both Epic and InterSystems for Power10 compatibility.
- Linux

  RHEL 8.6 or later on Power10 (ppc64le)
  - Certified by InterSystems for IRIS for Health on IBM Power.
  - Required for newer installations that use containerized services or performance-optimized Linux environments.
- Virtualization
  - Both AIX and Linux deployments use IBM PowerVM for LPAR management.
  - Linux LPARs can also be deployed by using Red Hat OpenShift on IBM Power for modern container-based workloads.
- Storage

  High-performance storage area network (SAN) such as IBM FlashSystem is commonly used, with tuning for Caché and IRIS.
- HADR
  - PowerHA for AIX.
  - Linux HA tools (including IBM VMRM).
  - InterSystems mirroring for resilience.

8

Servicing IBM Power11

The goal of serviceability is to enable efficient system repair while minimizing or avoiding disruption to operations. Serviceability includes system installation, upgrades or reversion (Miscellaneous Equipment Specification (MES)), and ongoing maintenance or repair activities. Depending on the system configuration and warranty agreement, the client, an IBM technician, or an authorized service provider performs service tasks.

IBM Power servers include advanced serviceability features that support an efficient maintenance environment. These features streamline service operations and reduce downtime by incorporating the following key attributes:

- Simplified installation and upgrade processes
- Support for concurrent maintenance and guided repair
- Automated diagnostics and error reporting
- End-to-end service workflows, from issue detection to resolution

The following topics are covered in this chapter:

- IBM maintenance
- IBM Power Expert Care
- IBM tools and interfaces
- Flexible Service Processor
- ASMI
- IBM Entitled System Support
- Firmware updates

© Copyright IBM Corp. 2025.

8.1 IBM maintenance

IBM offers a comprehensive global maintenance and support framework for IBM Power servers. This framework is designed to provide high availability (HA), rapid issue resolution, and minimal disruption to business operations. Maintenance services are available under various service-level agreements (SLAs) that are tailored to meet the needs of different environments, from standard business hours to mission-critical 24x7 operations. the IBM Support infrastructure includes remote diagnostics, onsite service, and proactive monitoring, all backed by a global network of skilled service professionals and parts depots.

You can choose between 8x5 and 24x7 service coverage. The 8x5 option provides support during standard business hours (typically 8 AM to 5 PM, Monday through Friday, excluding holidays). This option is ideal for non-critical systems or environments with internal IT support. For systems that require continuous uptime, the 24x7 service option provides around-the-clock support, including weekends and holidays. This level of coverage is essential for industries such as finance, healthcare, and manufacturing, where downtime can cause operational or financial consequences.

IBM also defines response time targets based on the selected service level. For example, under a 24x7 agreement, IBM might commit to a 2-hour or 4-hour onsite response for critical hardware issues, depending on the location and contract terms. These response times are supported by IBM's global logistics and service infrastructure, which includes strategically located parts centers and field engineers in more than 170 countries.

This worldwide coverage helps ensure that IBM Power customers receive consistent, high-quality support regardless of their location. IBM's maintenance services are enhanced by features such as Call Home diagnostics, automated error reporting, and remote problem determination. These features accelerate issue resolution and reduce the need for manual intervention. Combined, these capabilities provide a robust and reliable maintenance ecosystem that supports both traditional on-premises deployments and modern hybrid cloud environments.

8.2 IBM Power Expert Care

Managing your IBM Power environment should be seamless and efficient. IBM Power Expert Care provides immediate access to a curated bundle of high-value services that are trusted by the IBM Power community. This approach eliminates the delays and complexity of traditional procurement processes. With this preselected service package, you benefit from streamlined support that is enhanced by artificial intelligence (AI)-driven tools that accelerate response times, improve case resolution, and increase overall satisfaction.

IBM Power Expert Care Premium focuses on mission-critical environments where downtime is not an option. Backed by the same experts who assisted with the design and build of the Power11 platform, this service tier delivers infrastructure support with unmatched technical depth and operational excellence.

Here are the key features of IBM Power Expert Care Premium:

- 30-minute response time for both hardware and software support cases
- 4-hour onsite response target for urgent issues
- System and microcode compatibility guidance to help ensure optimal performance
- Dedicated Technical Account Manager (TAM) and mission-critical support resources

- Automated case creation and log analysis for faster issue resolution
- Health checks for hardware, operating system (OS), and applications (available as add-ons)
- Zero planned downtime (ZPD) for maintenance support with proactive planning and TAM assistance

For non-mission-critical systems, IBM offers IBM Power Expert Care Advanced, which provides 24x7 standard response times for repair and maintenance cases.

Technical Account Manager

The TAM plays a vital role in delivering proactive, product-based support for IBM Power. Acting as the primary point of contact, the TAM provides strategic guidance and direct engagement for both hardware and software within the scope of your support agreement. TAM services are delivered in English during your business hours. Support in other languages is available on request and mutual agreement, subject to availability.

TAMs help streamline operations and reduce downtime through the following responsibilities:

- Enabling Call Home for proactive error reporting, Autonomous Error Resolution (AER), and ZPD (available on Power11)
- Activating Support Insights for predictive analytics and delivering monthly reports
- Providing firmware and microcode compatibility analysis (exclusive to IBM Power Expert Care Premium clients)
- Sharing software lifecycle and roadmap updates
- Delivering HIPER alerts to help avoid high-impact issues
- Offering best practice documentation and priority handling of Severity 1 and 2 cases
- Leading complex case resolution and managing case progression
- Coordinating welcome calls, support planning, monthly reporting, and quarterly reviews
- Engaging IBM resources for Remote Code Load on Power10 and Power11 systems (onsite code load is available separately for Power11)
- Supporting change management by communicating planned events to relevant teams

For clients running SAP HANA on Power10 or Power11, the TAM also provides the following benefits:

- Advises on SAP HANA best practices.
- Coordinates troubleshooting across the full software and hardware stack.
- Provides technical recommendations for error identification, environment optimization, and known defect mitigation.

8.3 IBM tools and interfaces

Servicing IBM Power servers relies on a robust combination of proactive maintenance, intelligent diagnostics, and modern remote management tools, all designed to provide HA and system reliability. At the center of this ecosystem is the Hardware Management Console (HMC), which serves as the central hub for monitoring system health, managing logical partitions (LPARs), and coordinating service activities. Complementing the HMC is the service processor, a dedicated hardware component that operates independently of the main system. It continuously monitors hardware status, logs errors, and enables early detection of potential issues. Features such as Call Home support and automated error reporting help streamline the support process by alerting IBM Support teams to critical events, often before they affect operations.

You can access and manage the service processor through the HMC or directly through the Advanced System Management Interface (ASMI), which provides low-level control and extra troubleshooting capabilities. For modern, automated environments, the service processor also supports DMTF Redfish application programming interfaces (APIs), enabling secure, standardized remote management and integration with broader data center orchestration tools. Together, these technologies form a resilient and intelligent service infrastructure that supports both on-premises and hybrid cloud deployments, helping
IBM Power servers remain secure, efficient, and continuously operational.

8.4 Flexible Service Processor

The Flexible Service Processor (FSP) is a dedicated, always-on management controller that is embedded in the IBM Power E1180 system. It operates independently of the main system processors and OS, providing out-of-band management capabilities. With this design, you can monitor, configure, and service the system even when the OS is not running or the system is in a fault state.

The Power E1180 includes redundant system master FSPs that are in the System Control Unit (SCU). These FSPs operate in an active-standby configuration to help ensure continuous system management:

- Automatic failover: If the active FSP fails, the standby FSP takes over seamlessly, maintaining uninterrupted control and monitoring.
- Resilience: Both FSPs can manage the system independently, eliminating single points of failure in the management plane.
- Enhanced uptime: Redundancy supports 24x7 availability by enabling fault isolation and recovery without requiring a full system shutdown.

The FSP performs a wide range of critical system management tasks:

- Hardware monitoring and control
  - Control power and monitors consumption.
  - Monitors the environment (temperature, fan speeds, and voltage).
  - Handles hardware error logging and alerting.
  - Manages Vital Product Data (VPD).
- System configuration and initialization
  - Hosts the ASMI for hardware-level configuration.
  - Manages the boot process and initial program load (IPL).

- Diagnostics and serviceability
  - Captures First Failure Data Capture (FFDC) for rapid fault analysis.
  - Supports guided service procedures for hardware maintenance.
- Security and connectivity
  - Operates a secure Linux based environment with support for encrypted protocols.
  - Manages digital certificates and secure access policies.
  - Interfaces with the HMC and supports Common Information Model (CIM), Intelligent Platform Management Interface (IPMI), and Redfish APIs for integration with modern data center tools.

8.5 ASMI

The ASMI is a GUI that is part of the service processor firmware. The ASMI manages and communicates with the service processor. Use the ASMI to set up the service processor and perform service tasks, such as reading service processor error logs, reading VPD, and controlling system power.

You can access the ASMI by using the HMC GUI or CLI, or by directly attaching a laptop to the FSP port with a crossed Ethernet cable or through a switch (the same network subnet is required). Some ASMI functions are restricted to IBM service representatives and require celogin or dev user credentials during repair activities.

You can use the control panel functions to interface with the server. These functions range from simple status displays (such as IPL speed) to advanced service functions that only IBM service representatives can access.

For a complete list of ASMI functions, see Managing the Advanced System Management Interface.

The HMC GUI and CLI remain the primary management interfaces for clients.

8.6 IBM Entitled System Support

You can use IBM Entitled System Support to view and manage IBM Power and IBM Storage software and hardware. Most products that IBM Systems offers and that you purchase through IBM Digital Sales representatives or IBM Business Partners are available on this site when you use the IBM Configurator.

The site includes the following three main sections:

- My entitled software

  This section lists activities that are related to IBM Power and IBM Storage software. You can download licensed, no-charge, and trial software media, place software update orders, and manage software keys.
- My entitled hardware

  This section lists activities that are related to IBM Power and IBM Storage hardware. You can renew Update Access Keys (UAKs), buy and use Elastic Capacity on Demand, assign or buy credits for new and existing pools in a Power Private Cloud environment (Power Enterprise Pools 2.0 (PEP 2.0)), download Storage Capacity on Demand codes, and manage Hybrid Capacity credits.

- My inventory

  This section lists activities that are related to IBM Power and IBM Storage inventory. You can browse software licenses, software maintenance, and hardware inventory, manage inventory retrievals by using Base Composer, and generate several types of reports.

8.7 Firmware updates

Firmware management on IBM Power servers is a critical component of system reliability, security, and performance. It includes deploying, updating, and monitoring system firmware, such as the Power Hypervisor (PHYP), service processor firmware, and device microcode. You typically manage firmware through the HMC or the API, with support for both concurrent (live) and disruptive updates, depending on the component. IBM provides tools such as the Fix Level Recommendation Tool (FLRT) and IBM Electronic Service Agent (ESA) to help you identify and apply the latest firmware levels. These tools help ensure that systems remain secure and optimized for enterprise workloads.

Updating system firmware from the HMC

To update or upgrade the system firmware by using the HMC, complete the following steps:

1. In the System View, select the system that you want to update.
2. Click the **Firmware** menu.
3. Select **Update System Firmware**.

The Update System Firmware wizard guides you through the necessary steps to complete the firmware update process, as shown in Figure 8-1.

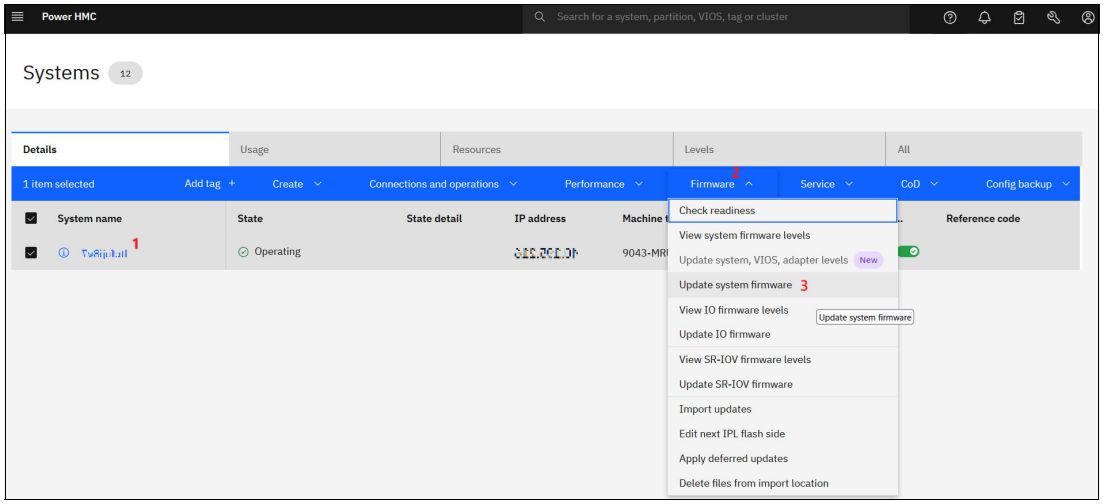

Figure 8-1 Starting the Update system firmware wizard

The Firmware update wizard guides you through updating your system. The process begins with accepting the license agreement and performing a system health check by using the Readiness Check. Next, you provide the required firmware files and select the firmware level to apply. The final step is to monitor the update progress and confirm its successful completion. Figure 8-2 on page 167 through Figure 8-5 on page 168 illustrate these key steps in the update process.

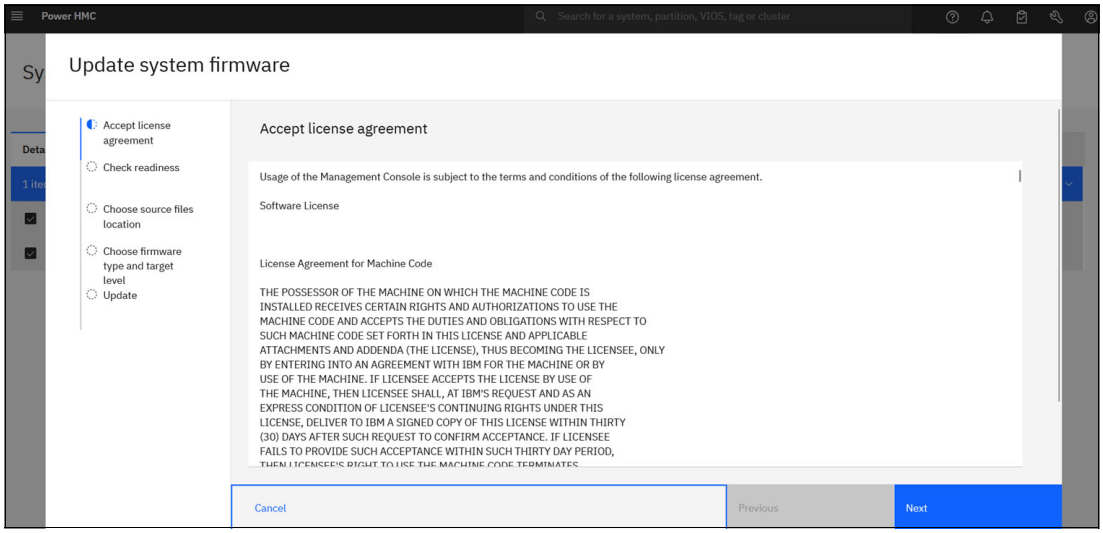

Figure 8-2 License agreement

Figure 8-3 shows that the System Readiness Check completed successfully for the IBM Power11 system firmware update process.

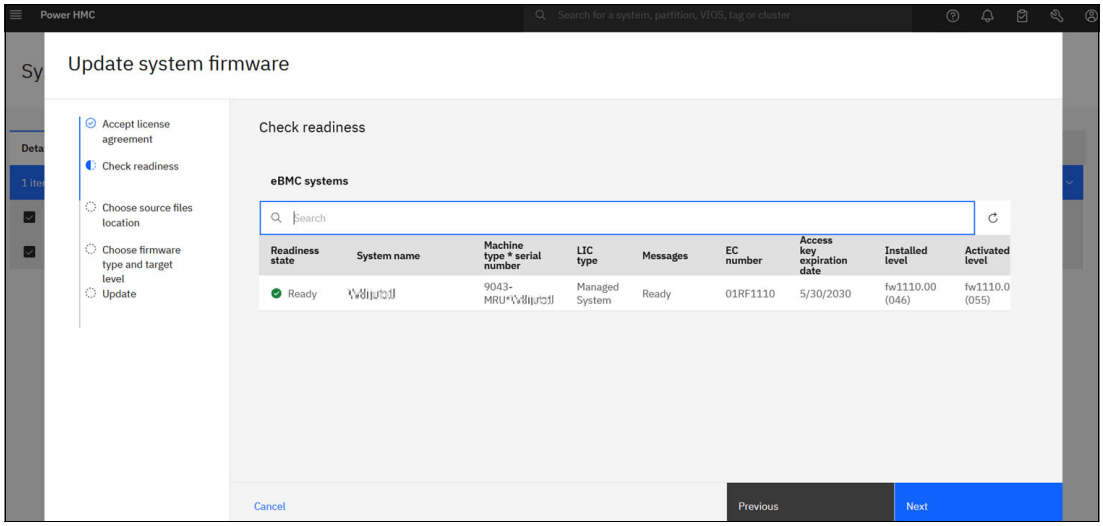

Figure 8-3 The system readiness check completed successfully

Figure 8-4 illustrates the “Choose source files location” menu, which is used to update the system firmware. Several options are available for loading the required update files:

- IBM Website
- FTP Server
- SFTP Server
- Mount Point on the HMC
- CD-ROM/DVD
- USB Drive

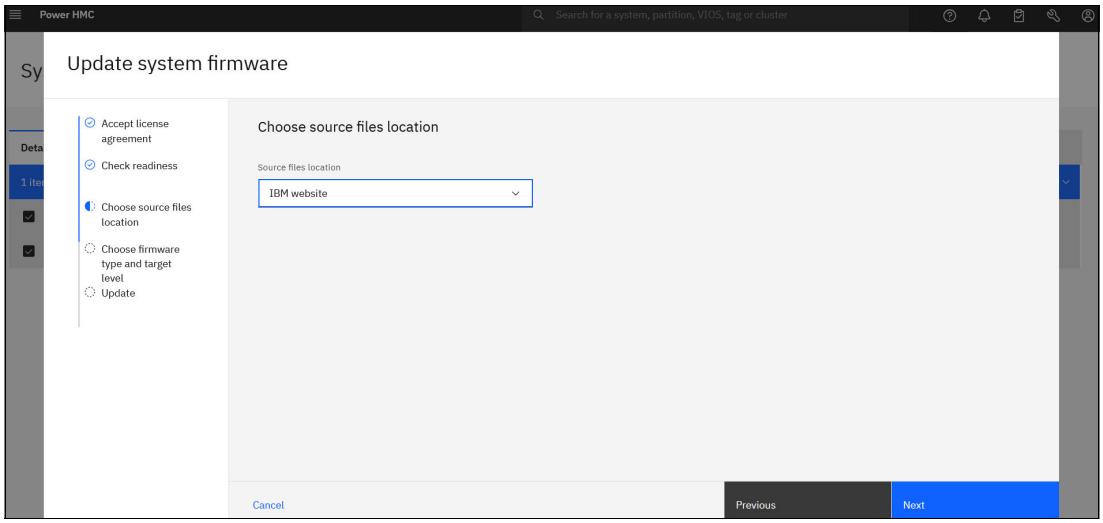

Figure 8-4 Choose source file location window

Click Search Available Levels to find the new system firmware on the installation media, As shown in Figure 8-5. The, select the system firmware level.

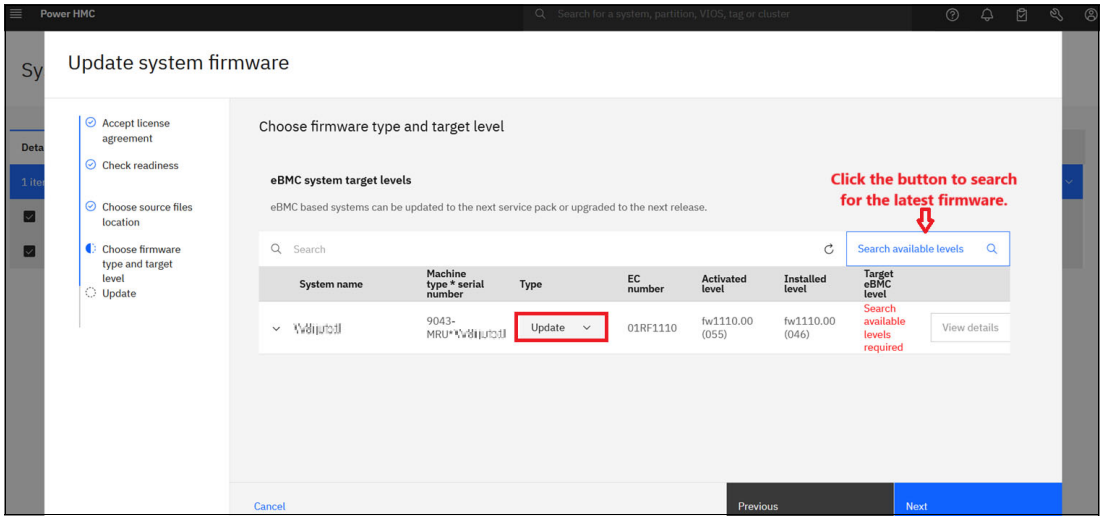

Figure 8-5 System firmware target level

Importing update files from HMC

The Import update files wizard in the HMC GUI, as shown in Figure 8-6 on page 169, guides you through the necessary steps to import firmware update or upgrade files. These files can then be reused to update the firmware on other IBM Power11 systems.

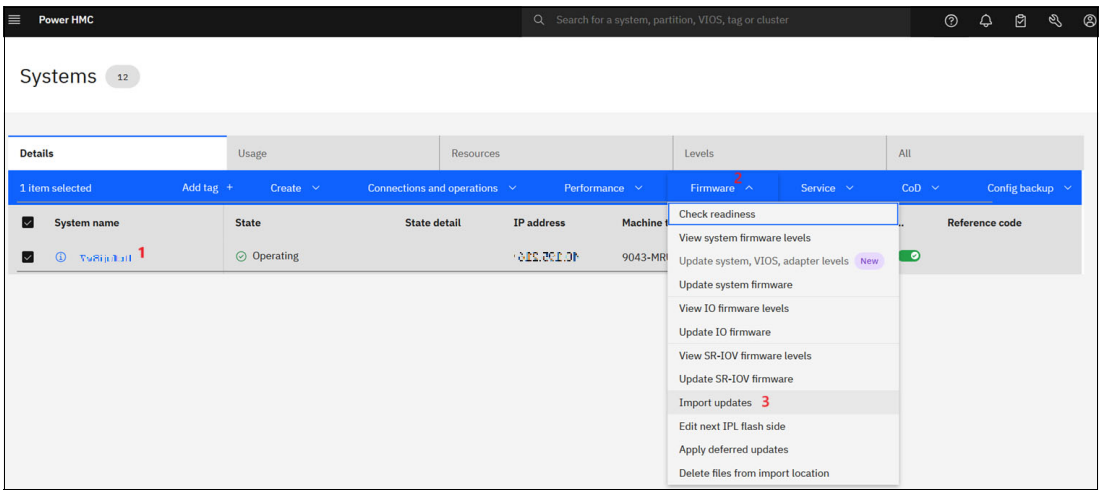

Figure 8-6 The Import update files option in the HMC GUI

Delete files from import location

The Delete files from import location wizard in the HMC GUI, as shown in Figure 8-7, guides you through the steps to delete imported firmware files so that you can manage and clean up firmware files that are stored on the HMC.

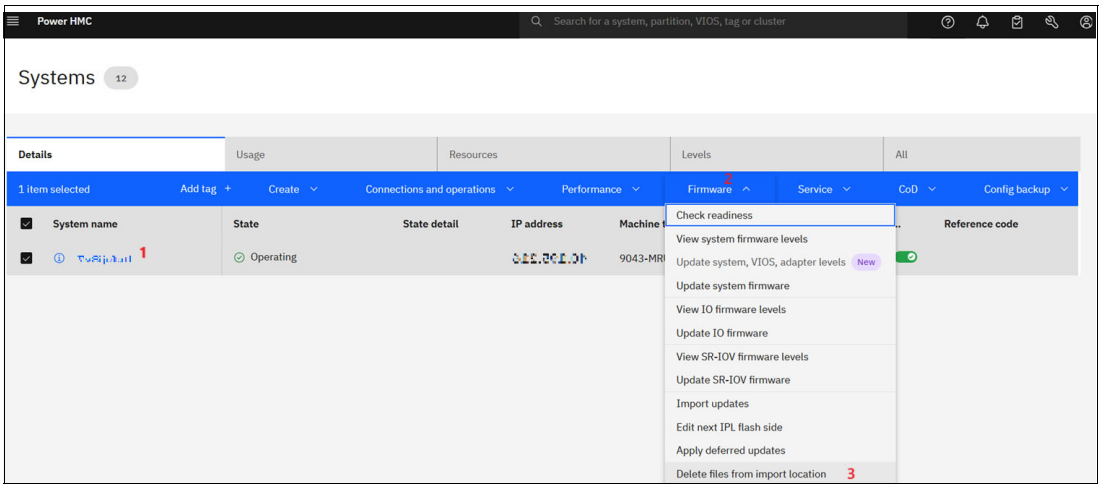

Figure 8-7 The “Delete files from import location” option in the HMC GUI

SR-IOV Shared mode adapter firmware management

Single Root I/O Virtualization (SR-IOV)–capable adapters that run in Shared mode use a different firmware mechanism to create versions than SR-IOV–capable adapters that run in Dedicated mode.

SR-IOV configuration requirements

- Ensure that the managed system includes Peripheral Component Interconnect Express (PCIe) adapters that support SR-IOV functions.
- Each SR-IOV–capable adapter must be installed in a PCIe slot that supports SR-IOV. For more information, see the adapter placement guidelines in IBM Documentation.

- Verify that the system is running a supported level of the following components:
  - System firmware
  - HMC or PowerVM NovaLink
  - OS (with SR-IOV driver support)
- Configure the adapters in SR-IOV Shared mode by using either the HMC or NovaLink interface.

In SR-IOV (Shared mode)

- The adapter firmware and driver are integrated with the system firmware.
- Firmware updates and the accompanying readme documentation are available at IBM Fix Central.
- When an adapter is configured in SR-IOV Shared mode, its firmware and driver are automatically updated to the latest versions that are included with the system firmware. This update occurs during standard maintenance operations, such as a system IPL, adapter replacement, or when you switch the adapter mode between Shared and Dedicated.
- If system firmware updates are installed concurrently, you can perform selective manual updates of SR-IOV adapter firmware by using the HMC. For more information, see Update SR-IOV Firmware.

Note: You can update either the adapter firmware alone or both the adapter driver and firmware.

During the firmware update process, network traffic on the configured logical ports of the SR-IOV adapter might be temporarily disrupted.

Updating each SR-IOV adapter typically takes 2 - 5 minutes. The update is performed sequentially across all SR-IOV adapters in the system.

Figure 8-8 shows the “View SR-IOV firmware levels” and “Update SR-IOV firmware” options that are available in the HMC GUI.

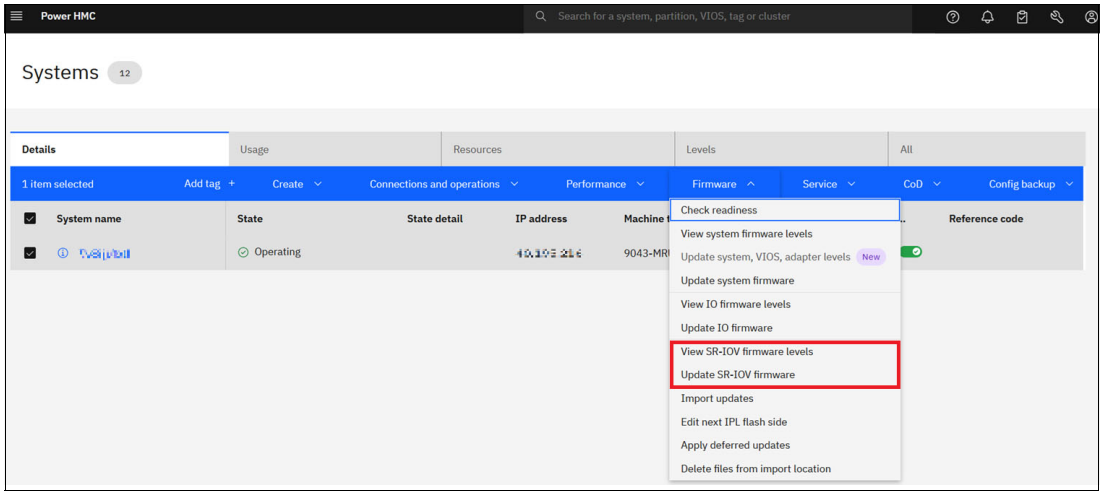

Figure 8-8 “View SR-IOV firmware levels” and “Update SR-IOV firmware” options in the HMC GUI

I/O microcode update strategy

Microcode updates are the responsibility of the customer unless the services are provided under a Microcode Support Services contract.

You can update device microcode by using the following methods:

- HMC

  Use the Update I/O Firmware feature in the HMC GUI to apply microcode updates to supported I/O devices.

- Diagnostic menus (AIX and Virtual I/O Server (VIOS))

  You can perform microcode updates by using either the system diagnostics or stand-alone diagnostics utilities from a USB key or a network location. This method applies to AIX and VIOS environments.

- Linux

  See the microcode readme files on IBM Fix Central for detailed instructions about updating microcode on Linux systems. These instructions might include the usage of other vendor-provided tools, depending on the I/O devices.

  IBM i (Program Temporary Fixes (PTFs))

  This method applies only to IBM i systems. Microcode updates are delivered through PTFs, which you can download from IBM Fix Central.

Viewing or updating I/O firmware from the HMC

You can use the HMC to view the current I/O firmware levels for a system and to update those levels by using a firmware repository.

The HMC does not have direct access to log in to the partitions to perform I/O firmware updates. Instead, it relies on the Resource Monitoring and Control (RMC) infrastructure, which is also used for features such as dynamic logical partitioning (DLPAR) and Service Focal Point to facilitate communication between the HMC and the partitions.

On each partition, the **`invscout`** command handles the query and update process. The **`invscout`** command exchanges inventory and update-related files with the HMC over the RMC interface.

Figure 8-9 shows the “View I/O firmware levels” and “Update I/O firmware” options that are available in the HMC GUI.

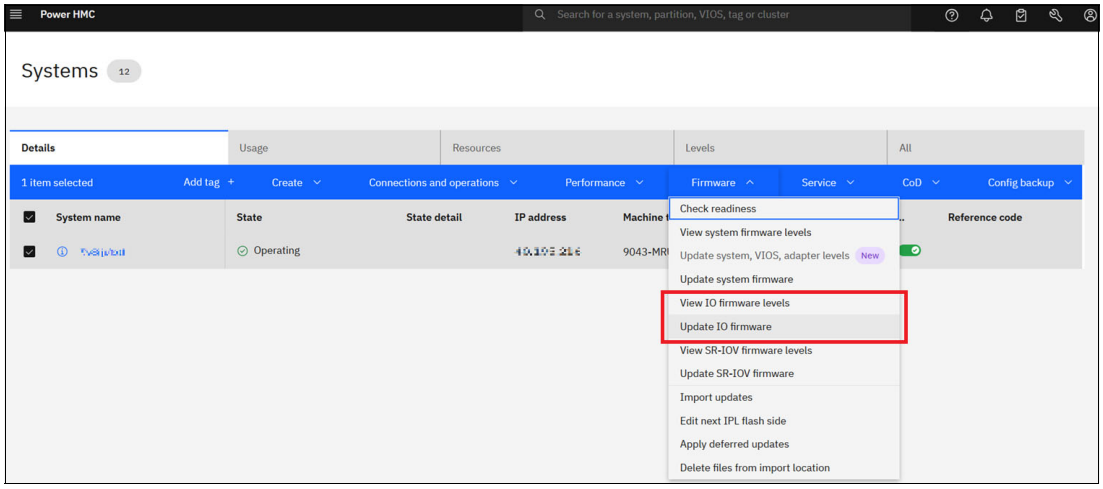

Figure 8-9 “View I/O firmware levels” and “Update I/O firmware” options in the HMC GUI

Figure 8-10 shows the “View I/O firmware levels” window, which shows I/O firmware information for both FSP and eBMC-based systems, as available in the HMC GUI.

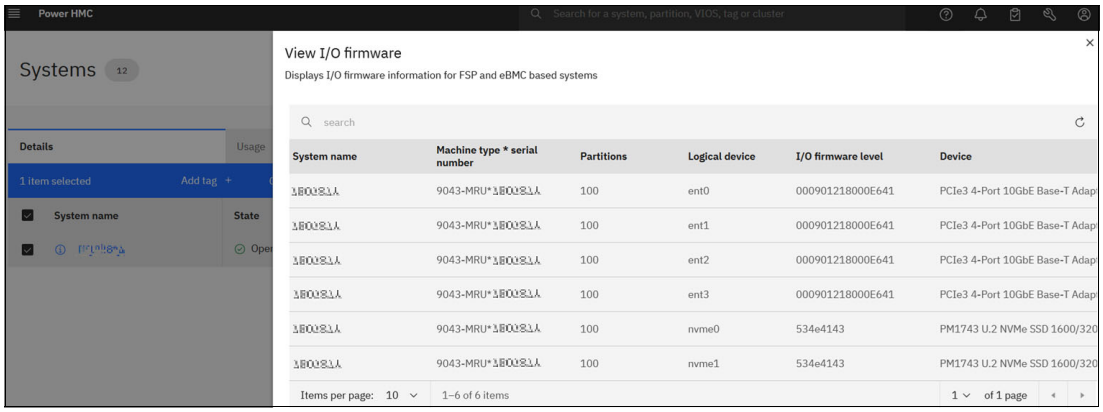

Figure 8-10 “View I/O firmware levels” window in the HMC GUI

Figure 8-11 shows the “Choose source files location” menu, which is used to specify the source location for updating the microcode on I/O devices. Several options are available for loading the required update files:

- IBM Website
- FTP Server
- SFTP Server
- Mount Point on the HMC
- CD-ROM/DVD
- USB Drive

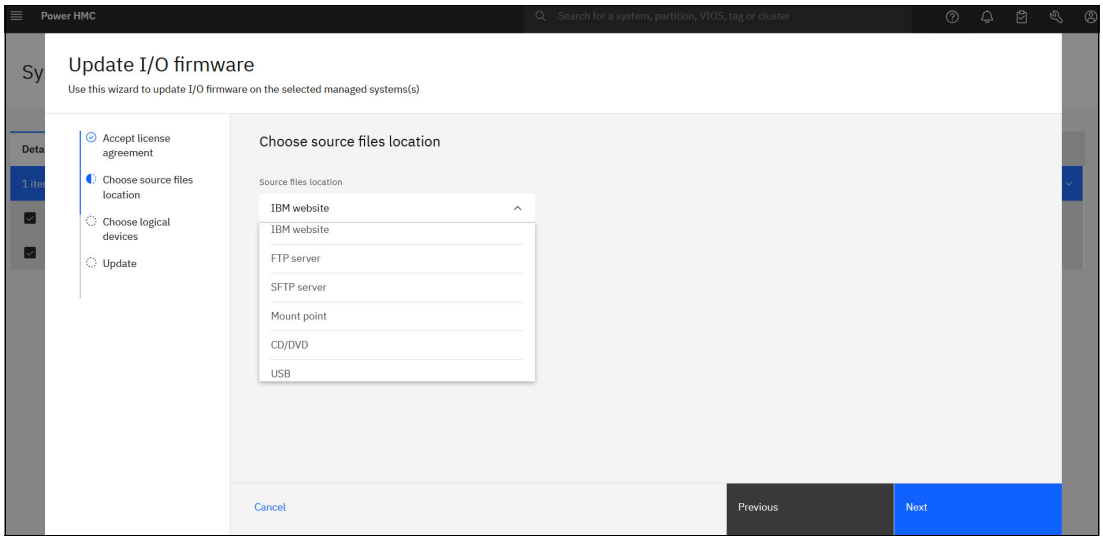

Figure 8-11 Update I/O firmware window

For more information about how to view or update the I/O firmware from the HMC, see View or Update the I/O Firmware from the HMC.

8.7.1 Update Access Keys

Since the introduction of IBM Power8 processor-based servers, IBM has used the concept of a UAK for each server.

When you apply system firmware updates, the UAK and its expiration date are checked. System firmware updates include a release date. If the release date of the firmware updates is later than the UAK expiration date, the updates are not applied.

Managing Update Access Keys

The system uses a UAK to control the application of system firmware updates. Each UAK includes an expiration date, and each firmware update includes a release date. When a firmware update is applied, the system checks whether the update's release date is later than the UAK's expiration date. If it is, the firmware update is not processed.

When a UAK expires, you must replace it by using the HMC. The expiration of a UAK does not affect I/O microcode updates; you can still apply these updates even if the UAK has expired. You can obtain a new UAK by opening a case with IBM Support and requesting a renewal key or by downloading it directly from the IBM Entitled Systems Support (ESS) website.

By default, newly delivered systems include a UAK that expires after 3 years. Thereafter, you can extend the UAK every 6 months, but only if a current hardware maintenance contract exists for that server. You can verify the contract on the Enterprise Storage Server web page.

You can check the validity and expiration date of the current UAK by using either the HMC GUI or the CLI. You can also retrieve the expiration date from the OS.

Verifying the expiration date of UAKs by using the HMC

You can see the current UAK expiration date on the "View current system firmware levels" window of the HMC, as shown in Figure 8-12.

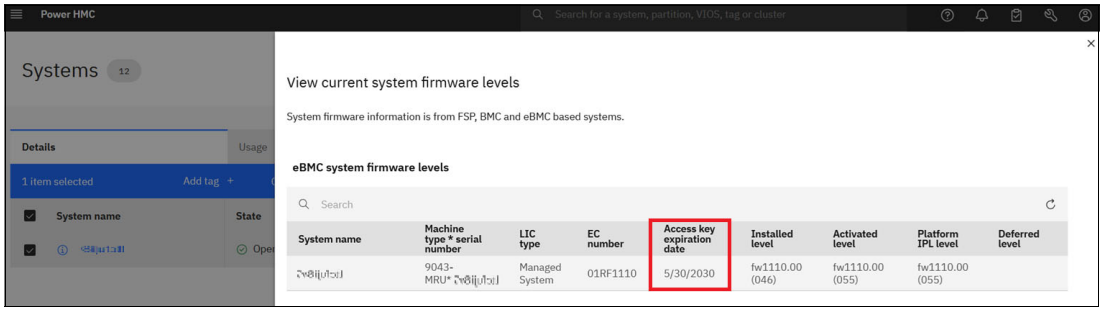

Figure 8-12 UAK expiration date on the "View current system firmware levels" window

You can use the HMC or Virtual HMC to configure the Automatic Firmware UAK. For more information, see How to Automatically Update the Access Key.

Verifying the expiration date of UAKs by using AIX

There are multiple methods of checking the UAK expiration date within AIX:

- Run the following command:

  ```
  lscfg -vpl sysplanar0 |grep -p "System Firmware"
  ```

  The output is shown in Example 8-1.

Example 8-1 Output of the lscfg command to check the UAK expiration date

```
lscfg -vpl sysplanar0 |grep -p "System Firmware"
      System Firmware:
        Code Level, LID Keyword.....Phyp_1 21372025061280A00701
        Code Level, LID Keyword.....PFW 21212025060681CF0681
        Code Level, LID Keyword.....FSP_Fil 16112025061681E00109
        Code Level, LID Keyword.....FipS_BU 16112025061681E00208
        Microcode Image.............RK110_058 RK110_057 RK110_058
        Microcode Level.............FW110.00 FW110.00 FW110.00
        Microcode Build Date........20250725 20250725 20250725
        Update Access Key Exp Date..20301014
        Hardware Location Code......U9080.HEU.DEN0013-Y1
      Physical Location: U9080.HEU.DEN0043-Y1
```

- Run the **grep** command on the UAK, as shown in Example 8-2.

Example 8-2 Output of the lscfg command to check the UAK expiration date

```
# lscfg -vpl sysplanar0 | grep "Access Key"
        Update Access Key Exp Date..20301014
```

- In AIX 7.3, run the **lparstat** command, as shown in Example 8-3.

Example 8-3 Output of the lparstat command to check the UAK expiration date

```
# lparstat -u
FW Update Access Key Expiration  (YYYYMMDD):  20301014
AIX Update Access Key Expiration (YYYYMMDD):  20301014
AIX Image Date                   (YYYYMMDD):  20250725
```

Verifying the expiration date of UAKs by using Linux

There is no single Linux command to view a UAK directly. Instead, UAKs are managed through IBM tools and interfaces, such as ESA, ASMI, and HMC.

Managing the UAK with IBM i

ESA on IBM i introduced a new feature to help you manage the refresh of UAKs for IBM i stand-alone partitions. You can use this UAK management feature in two ways:

- Manual

  A system administrator can manually start the process of checking the UAK expiration date, downloading the UAK, and applying the UAK. To do this task, run the **WRKSRVAGT TYPE(*UAK)** command.

  You can also run the **GO SERVICE** command to access the ESA on IBM i main menu and then select Option 20: Check and refresh Update Access Key, as shown in Figure 8-13 on page 175.

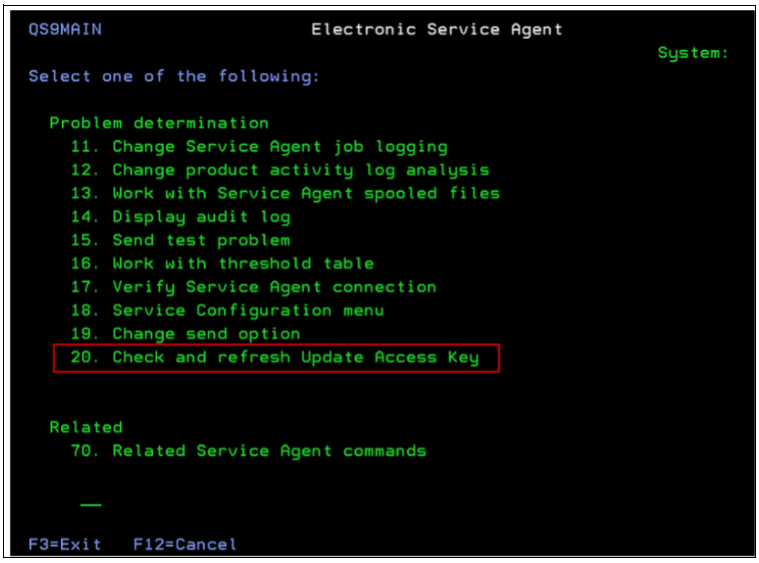

Figure 8-13 Electronic Service Agent on IBM i main menu option 20

The help instructions for Option 20 provide more information about this feature, as shown in Figure 8-14.

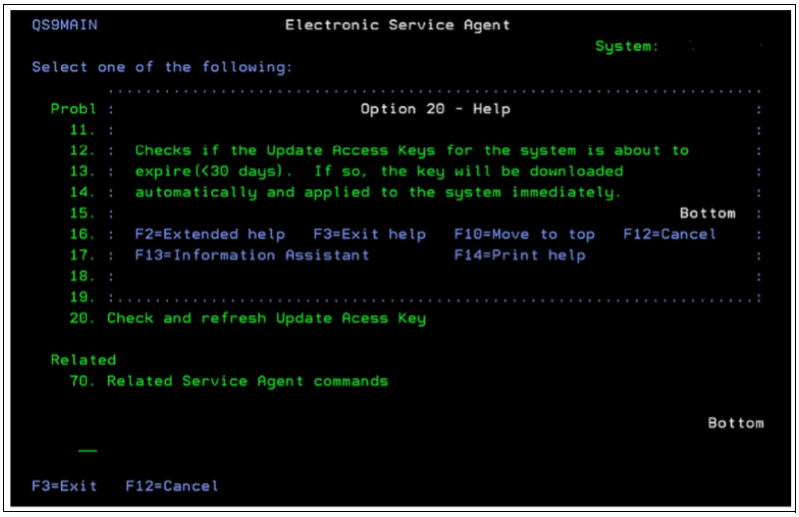

Figure 8-14 Help information on Electronic Service Agent on IBM i main menu option 20

- Automatic

  You can use the Change Service Agent Attribute (**CHGSRVAGTA**) command to enable automatic checking. The **REFRESHUAK** parameter of the **CHGSRVAGTA** command establishes automatic UAK management. By default, a job that is named QS9UAK runs every Sunday, but you can customize the schedule.

To complete this task, complete the following steps:

a. Enter **CHGSRVAGTA** and press F4.

b. Enter Yes for Enable.

c. Press Page Down.

The Change Service Agent Attribute panel opens, as shown in Figure 8-15. Specify the weekday and time for UAK refresh processing.

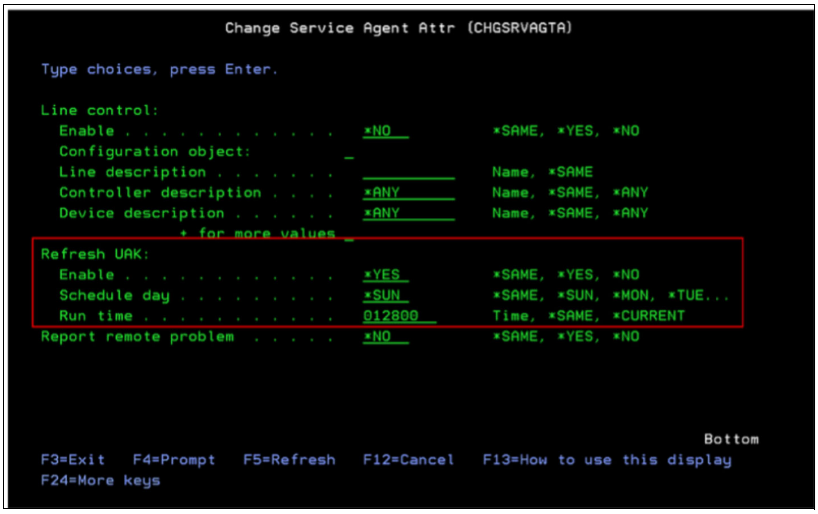

Figure 8-15 Change Service Agent Attribute: Refresh UAK configuration panel

9

Virtualization

Virtualization on IBM Power is a cornerstone of enterprise computing. It provides robust, scalable, and secure environments for running multiple workloads on a single physical server. At the core of this capability is IBM PowerVM, IBM enterprise-grade virtualization technology. With PowerVM, you can create logical partitions (LPARs) so that multiple operating systems (OSs), such as AIX, IBM i, and Linux, may run concurrently on the same hardware. It supports advanced features such as Live Partition Mobility (LPM), dynamic resource allocation, and IBM Micro-Partitioning®. These features help maximize hardware usage and reduce operational costs.

In addition to PowerVM, IBM Power servers support Kernel-based Virtual Machine (KVM), an open-source virtualization option. KVM on IBM Power provides a flexible and cost-effective alternative for Linux-based workloads, particularly in cloud-native and containerized environments. It integrates with modern orchestration tools and supports features such as nested virtualization and Single Root I/O Virtualization (SR-IOV) for high-performance networking. KVM is ideal for organizations that want to use open-source technologies while benefiting from the performance and reliability of IBM Power hardware.

A key component of virtualization efficiency on IBM Power is the use of shared processor pools (SPPs). With SPPs, multiple LPARs may share a pool of physical processors, enabling dynamic allocation of CPU resources based on workload demand. With IBM Power11, PowerVM adds support for Resource Groups to enhance processor sharing capabilities and improve efficiency. Processor pools extend the capabilities of SPPs to improve processor usage and enhance isolation and control by capping the maximum CPU resources that are available to each pool. When combined with PowerVM or KVM, SPPs help optimize performance, reduce licensing costs, and provide consistent service levels across virtual environments.

The following topics are covered in this chapter:

- PowerVM
- KVM support

© Copyright IBM Corp. 2025.

9.1 PowerVM

The PowerVM platform is a family of technologies, capabilities, and offerings that deliver industry-leading virtualization for enterprises. It is the umbrella term for Power processor-based server virtualization, including IBM PowerVM Hypervisor (PHYP), logical partitioning, IBM Micro-Partitioning, Virtual I/O Server (VIOS), LPM, and more. PowerVM combines hardware enablement and software.

Note: A PowerVM Enterprise Edition license entitlement is included with each IBM Power E1180 server. PowerVM Enterprise Edition is available as a hardware feature (Feature Code EPVT). It supports up to 20 partitions per core, VIOS, and Resource Groups, which can be combined with multiple shared processor pools (MSPPs). It also offers LPM.

9.1.1 IBM PowerVM Hypervisor

Power processor-based servers combined with PowerVM technology offer the following key capabilities to help consolidate and simplify IT environments:

- Improve server usage and share I/O resources to reduce the total cost of ownership (TCO) and make better usage of IT assets.
- Improve business responsiveness and operational speed by dynamically reallocating resources to applications as needed. This approach helps match changing business needs or handle unexpected changes in demand.
- Simplify IT infrastructure management by making workloads independent of hardware resources. This independence enables the usage of business-driven policies to deliver resources based on time, cost, and service-level requirements.

Combined with features in the Power E1180, the PHYP delivers functions that enable other system technologies, including LPARs, virtualized processors, IEEE virtual local area network (VLAN)-compatible virtual switches, virtual SCSI adapters, virtual Fibre Channel (FC) adapters, and virtual consoles.

The PHYP is a core component of the system firmware and offers the following functions:

- Provide an abstraction between the physical hardware resources and the LPARs that use them.
- Enforce partition integrity by providing a security layer between LPARs.
- Control the dispatch of virtual processors to physical processors.
- Save and restore all processor state information during a logical processor context switch.
- Control hardware I/O interrupt management facilities for LPARs.
- Provide VLAN channels between LPARs to reduce the need for physical Ethernet adapters for inter-partition communication.
- Monitor the Flexible Service Processor (FSP) and perform a reset or reload if it detects the loss of an FSP, notifying the OS if the problem is not corrected.

The PHYP is always active, regardless of the system configuration or whether it is connected to the managed console. It requires memory to support the resource assignment of the LPARs on the server. The amount of memory that the PHYP firmware requires varies according to several factors:

- Memory usage for hardware page tables (HPTs)
- Memory usage to support I/O devices
- Memory usage for virtualization

Memory usage for hardware page tables

Each partition on the system includes its own HPT, which contributes to hypervisor memory usage. The HPT is used by the OS to translate effective addresses (EAs) to physical real addresses in the hardware. This translation allows multiple OSs to run simultaneously in their own logical address space. Whenever a virtual processor for a partition is dispatched on a physical processor, the hypervisor indicates to the hardware the location of the partition HPT for address translation.

The amount of memory for the HPT is based on the maximum memory size of the partition and the HPT ratio. The default HPT ratio is 1:128 (for AIX, VIOS, and logical partitions (LPARs)) of the maximum memory size of the partition. AIX, VIOS, and Linux use larger page sizes (16 KB and 64 KB) instead of 4 KB pages. Using larger page sizes reduces the overall number of pages that must be tracked; therefore, the overall size of the HPT can be reduced. For example, the HPT is 2 GB for an AIX partition with a maximum memory size of 256 GB.

When defining a partition, the maximum memory size that you specify is based on the amount of memory that can be dynamically added to the dynamic logical partition (DLPAR) without changing the configuration and restarting the partition.

In addition to setting the maximum memory size, you can configure the HPT ratio. Use the **`hpt_ratio`** parameter with the **`chsyscfg`** Hardware Management Console (HMC) command to define the HPT ratio for a partition profile. The valid values are `1:32`, `1:64`, `1:128`, `1:256`, or `1:512`.

Specifying a smaller absolute ratio (1:512 is the smallest value) decreases the overall memory that is assigned to the HPT. Testing is required when changing the HPT ratio because a smaller HPT might increase CPU consumption. The OS might need to reload the entries in the HPT more frequently. Most customers choose to use the IBM provided default values for the HPT ratios.

Memory usage for I/O devices

To support I/O operations, the hypervisor maintains structures that are called *translation control entries* (TCEs), which provide an information path between I/O devices and partitions. The TCEs include the address of the I/O buffer, indications of read/write requests, and other I/O-related attributes. Many TCEs are used per I/O device, so multiple requests can be active simultaneously on the same physical device. To improve performance and provide better affinity, the TCEs are distributed across multiple processor chips or drawers.

For physical I/O devices, the base amount of space for the TCEs is defined by the hypervisor based on the number of supported I/O devices. A system that supports high-speed adapters can also be configured to allocate more memory to improve I/O performance. Linux is the only OS that uses these extra TCEs. If the system uses only AIX, the memory can be freed for use by partitions.

Memory usage for virtualization features

Virtualization requires additional memory to be allocated by the PHYP for hardware statesave areas and various virtualization technologies. For example, on Power11 processor-based systems, each processor core supports up to eight simultaneous multithreading (SMT8) threads of execution, and each thread contains more than 80 different registers.

The PHYP must reserve save areas for the register contents for the maximum number of virtual processors that are configured. The greater the number of physical hardware devices, virtual devices, and virtualization features, the more hypervisor memory is required. For efficient memory consumption, the wanted and maximum values for various attributes (processors, memory, and virtual adapters) must be based on business needs and not set higher than actual requirements.

Predicting memory that is used by the PowerVM Hypervisor

The IBM System Planning Tool (SPT) is a resource that you can use to estimate the amount of hypervisor memory that is required for a specific server configuration. After you download and install the SPT executable file, you can define a configuration by selecting the correct hardware platform, installed processors, and memory, and by defining partitions and partition attributes. SPT can estimate the amount of memory that is assigned to the hypervisor, which helps when you change a configuration or deploy new servers.

The PHYP provides the following types of virtual I/O adapters:

- Virtual SCSI

  The PHYP provides a virtual SCSI mechanism for the virtualization of storage devices. Storage virtualization uses two paired adapters: a virtual SCSI server adapter and a virtual SCSI client adapter.

- Virtual Ethernet

  The PHYP provides a virtual Ethernet switch function that partitions use to communicate quickly and securely on the same server without requiring a physical interconnection or external connectivity. If a Layer 2 bridge to a physical Ethernet adapter is set in one VIOS partition, this configuration is known as a Shared Ethernet Adapter (SEA).

- Virtual FC

  A virtual FC adapter provides client LPARs with an FC connection to a storage area network (SAN) through the VIOS partition. The VIOS partition connects the virtual FC adapters on the VIOS partition to the physical FC adapters on the managed system.

- Virtual (tty) Console

  Each partition must have access to a system console. Tasks such as OS installation, network setup, and problem analysis require a dedicated system console. The PHYP provides the virtual console by using a virtual tty or serial adapter and a set of hypervisor calls. Virtual tty does not require the purchase of any additional features or software, such as PowerVM Edition features.

Logical partitions

LPARs and virtualization increase the usage of system resources and add configuration flexibility.

Logical partitioning enables a server to operate as two or more independent servers. When you use logical partitioning on a server, you divide its resources into subsets that are called LPARs. You can install software on an LPAR, and the LPAR runs as an independent logical server with the resources that you allocate to it.

Some documentation refers to an LPAR as a *virtual machine* (VM), which makes it appear similar to what other hypervisors offer. However, LPARs provide a higher level of security, isolation, and other features that are described in this chapter.

You can assign processors, memory, and I/O devices to LPARs. AIX, IBM i, Linux, and VIOS can run on LPARs. VIOS provides virtual I/O resources to other LPARs that run general-purpose OSs.

LPARs share a few system attributes, such as the system serial number, system model, and processor feature codes. All other system attributes can vary from one LPAR to another.

Micro-Partitioning

When you use Micro-Partitioning technology, you can allocate fractions of processors to an LPAR. An LPAR that uses fractions of processors is also known as a shared processor partition or micropartition. Micropartitions run on a set of processors that are called an SPP. Virtual processors enable the OS to manage the fractions of processing power that are assigned to the LPAR.

From an OS perspective, a virtual processor cannot be distinguished from a physical processor unless the OS is enhanced to determine the difference. Physical processors are abstracted into virtual processors that are available to partitions.

On a Power11 processor-based server, you can define a partition with a processor capacity as small as 0.05 processing units. This value represents 0.05 of a physical core. Each physical core can be shared by up to 20 shared processor partitions, and the partition entitlement can be incremented fractionally by as little as 0.05 of a processor. The shared processor partitions are dispatched and time-sliced on the physical processors under the control of the PHYP. The HMC creates and manages the shared processor partitions.

The Power E1180 supports up to 256 cores in a single system and 1,000 micropartitions, which is the maximum that PowerVM supports.

Note: Although the Power E1180 supports up to 1,000 micropartitions, the actual limit depends on the application workload demands on the server.

Processing mode

When you create an LPAR, you can assign entire processors for dedicated use or allocate partial processing units from an SPP. This setting defines the processing mode of the LPAR.

Dedicated mode

In dedicated mode, physical processors are assigned as a whole to partitions. The simultaneous multithreading (SMT) feature in the Power11 processor core enables the core to run instructions from two, four, or eight independent software threads simultaneously.

Shared dedicated mode

On Power11 processor-based servers, you can configure dedicated partitions to donate idle processors that they own to an SPP. This feature enables the donation of spare CPU cycles from dedicated processor partitions to an SPP. The dedicated partition maintains absolute priority for its dedicated CPU cycles. Enabling this feature can help increase system usage without compromising the computing power for critical workloads in a dedicated processor mode LPAR.

Shared mode

In shared mode, LPARs use virtual processors to access fractions of physical processors. Shared partitions can define any number of virtual processors, up to a maximum of 20 times the number of processing units that are assigned to the partition. The PHYP dispatches virtual processors to physical processors according to the partition's processing unit entitlement. One processing unit represents the processing capacity of one physical processor. All partitions receive a total CPU time equal to their processing unit entitlement.

Logical processors are defined on top of virtual processors. Therefore, even with a virtual processor, the concept of a logical processor exists. The number of logical processors depends on whether SMT is enabled.

9.1.2 Multiple shared processor pools

MSPPs are supported on Power11 processor-based servers. With this capability, a system administrator can create a set of micropartitions to control the processor capacity that can be used from the physical SPP.

Micropartitions are created and then assigned to either the default processor pool or a user-defined SPP. The PHYP monitors the virtual processors within the set of micropartitions. The processor capacity is managed according to user-defined attributes.

If the Power server is under heavy load, each micropartition within an SPP is assured of its processor entitlement, plus any capacity that might be allocated from the reserved pool capacity if the micropartition is uncapped.

If specific micropartitions in an SPP do not use their processing capacity entitlement, the unused capacity is ceded. Other uncapped micropartitions within the same SPP can use the extra capacity according to their uncapped weighting. In this way, the entitled pool capacity of an SPP is distributed among the micropartitions within that SPP.

All Power servers that support the MSPP capability have a minimum of one SPP (the default) and up to a maximum of 64 SPPs.

This capability helps reduce the TCO when the cost of software or database licenses depends on the number of assigned processor cores.

Note: Shared processor pools can be used with Resource Groups. This configuration can improve performance and isolation.

9.1.3 Virtual I/O Server

VIOS is part of PowerVM. It is the appliance that enables the sharing of physical resources among LPARs to improve efficiency, such as through system consolidation. The VIOS owns the physical I/O resources (SCSI and FC) network adapters, and optical devices) and enables client partitions to share access to them. This approach minimizes and optimizes the number of physical adapters in the system.

VIOS eliminates the requirement for every partition to own a dedicated network adapter, disk adapter, and disk drive. It supports OpenSSH for secure remote logins and provides a firewall for limiting access by ports, network services, and IP addresses.

Figure 9-1 shows an overview of a VIOS configuration.

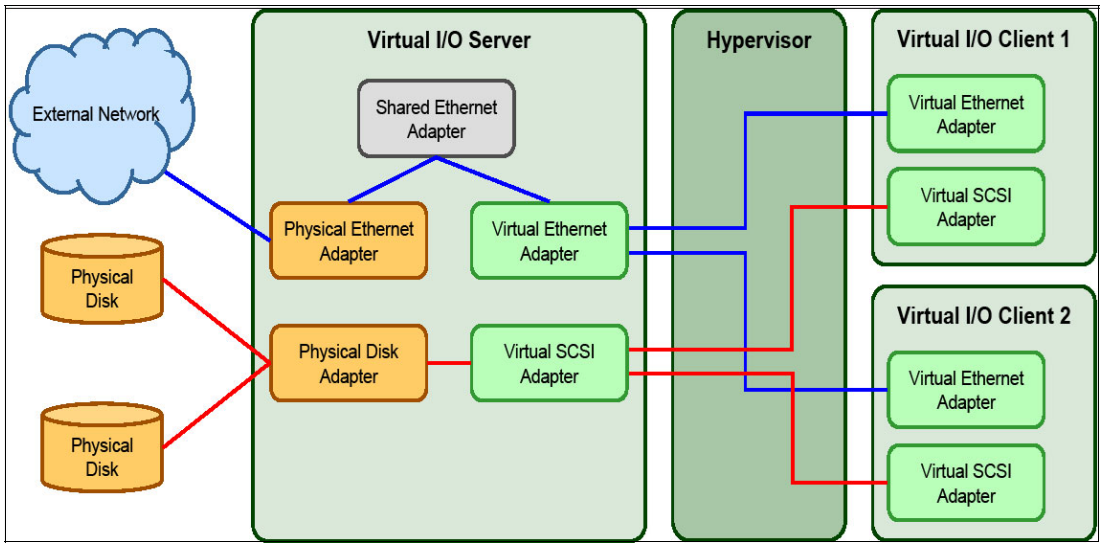

Figure 9-1 Architectural view of VIOS

It is a best practice to run dual VIOSs per physical server.

Shared Ethernet Adapter

A SEA can connect a physical Ethernet network to a virtual Ethernet network. The SEA provides this access by connecting the PHYP VLANs to the VLANs on external switches. Because the SEA processes packets at Layer 2, the original MAC address and VLAN tags of the packet are visible to other systems on the physical network. IEEE 802.1 VLAN tagging is supported.

By using the SEA, several client partitions can share one physical adapter. You can also connect internal and external VLANs by using a physical adapter. The SEA service can be hosted only in VIOS, not in a general-purpose AIX partition or LPAR. The SEA acts as a Layer 2 network bridge to securely transport network traffic between virtual Ethernet networks (internal) and one or more physical network adapters (external) that are configured as an Etherchannel. These virtual Ethernet network adapters are defined by the PHYP on VIOS.

Virtual SCSI

Virtual SCSI provides a virtualized implementation of the SCSI protocol and is based on a client/server model. The VIOS LPAR owns the physical I/O resources and acts as a server, or in SCSI terms, a target device. Client LPARs access the virtual SCSI-backing storage devices that VIOS provides.

The virtual I/O adapters (a virtual SCSI server adapter and a virtual SCSI client adapter) are configured by using the HMC. The virtual SCSI server (target) adapter runs any SCSI commands that it receives and is owned by the VIOS partition. The virtual SCSI client adapter enables a client partition to access physical SCSI and SAN-attached devices and LUNs that are mapped for client use. VIOS provides the provisioning of virtual disk resources.

Internet Small Computer Systems Interface

The internet Small Computer Systems Interface (iSCSI) disk is supported in VIOS 3.1.0 and later. The iSCSI disk provides block-level access to storage devices by carrying SCSI commands over an IP network. It facilitates data transfers over the internet by using TCP, a reliable transport mechanism that supports both IPv6 and IPv4 protocols. The iSCSI disk enables storage management over long distances.

The iSCSI support in VIOS allows iSCSI disks to be exported to client LPARs as virtual SCSI (vSCSI) disks. This support is available in VIOS 3.1 and later.

VIOS 3.1 enables Multipath I/O (MPIO) support for the iSCSI initiator. With MPIO support, you can configure and create multiple paths to an iSCSI disk, similar to other protocols. The client LPAR can run either an AIX or Linux operating system.

VIOS 3.1.1 adds support for multiple iSCSI initiators on the VIOS. This enhancement also includes performance improvements for the iSCSI driver. With multiple iSCSI initiator support, you can create multiple iSCSI software initiator devices on a single AIX OS instance.

N_Port ID Virtualization

N_Port ID Virtualization (NPIV) is a technology that enables multiple LPARs to access one or more external physical storage devices through the same physical FC adapter. This adapter is attached to a VIOS partition that acts only as a pass-through to manage the data transfer through the PHYP.

Each partition includes one or more virtual FC adapters, each with its own pair of unique worldwide port names. This configuration enables you to connect each partition to independent physical storage on a SAN. Unlike virtual SCSI, only the client partitions see the disk.

For more information and requirements for NPIV, see *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590.

9.1.4 Live Partition Mobility

LPM enables you to move a running LPAR from one system to another without disruption. With inactive partition mobility, you can move a powered-off LPAR between systems.

LPM provides systems management flexibility and improves system availability by avoiding the following situations:

- Planned outages for hardware upgrades or firmware maintenance.
- Unplanned downtime. With preventive failure management, if a server indicates a potential failure, you can move its LPARs to another server before the failure occurs.

HMC 10 release 1 (HMC V10R1) and VIOS 3.1.3 or later provide the following enhancements to the LPM feature:

- Automatically chooses the fastest network for LPM memory transfer.
- Enables LPM when a virtual optical device is assigned to a partition.

9.1.5 Active Memory Expansion

Active Memory Expansion (AME) is an optional feature (Feature Code EM89) under the PowerVM umbrella that enables memory expansion on the system. AME is an innovative technology that supports the AIX OS. It enables the effective maximum memory capacity to be larger than the physical memory maximum. Compression and decompression of memory content can enable memory expansion of up to 100% or more. This capability enables a partition to complete more work or support more users with the same physical amount of memory. Similarly, it enables a server to run more partitions and perform more work without adding physical memory.

AME uses CPU resources to compress and decompress memory contents. The tradeoff of memory capacity for processor cycles can be beneficial, but the degree of expansion depends on how compressible the memory content is and whether adequate spare CPU capacity is available for compression and decompression.

The Power E1180 includes a hardware accelerator that boosts AME efficiency and uses fewer processor core resources. Each AIX partition can enable or disable AME. Control parameters define the wanted amount of expansion for each partition to help manage CPU usage by the AME function.

An initial program load (IPL) is required for the specific partition that enables memory expansion. When AME is enabled, monitoring capabilities are available in standard AIX performance tools, such as **`lparstat`**, **`vmstat`**, **`topas`**, and **`svmon`**.

A planning tool is included with AIX to sample workloads and estimate how expandable the partition memory is and how many CPU resources are needed. The feature can be ordered with the initial Power E1180 order or as a Miscellaneous Equipment Specification (MES) order. A software key is provided when the enablement feature is ordered and applied to the system node. An IPL is not required to enable the system node. The key is specific to an individual system and is permanent. It cannot be moved to a different server.

IBM i does not support AME.

9.1.6 Remote Restart

Remote Restart is a high availability (HA) option for partitions. If an error occurs that causes a server outage, a partition that is configured for Remote Restart can be restarted on a different physical server. Sometimes, restarting the server might take longer, so the Remote Restart function can be used for faster reprovisioning of the partition. Typically, Remote Restart is faster than restarting the failed server and then restarting its partitions.

The Remote Restart function relies on technology that is similar to LPM. A partition must be configured with storage on a SAN that is shared and accessible by the server that hosts the partition.

HMC V10R1 provides an enhancement to the Remote Restart feature that enables remote restart when a virtual optical device is assigned to a partition.

9.1.7 POWER processor modes

Although they are not virtualization features, POWER processor modes are described here because they affect various virtualization capabilities.

On Power servers, partitions can be configured to run in several modes:

- Power9

  This native mode for Power9 processors implements Version 3.0 of the IBM Power Instruction Set Architecture (ISA).

- Power10

  This native mode for Power10 processors implements Version 3.1 of the IBM Power ISA.

- Power11

  This native mode for Power11 processors implements Version 3.1 of the IBM Power ISA.

For more information, see the Power11 documentation.

Figure 9-2 shows the available processor modes on a Power E1180.

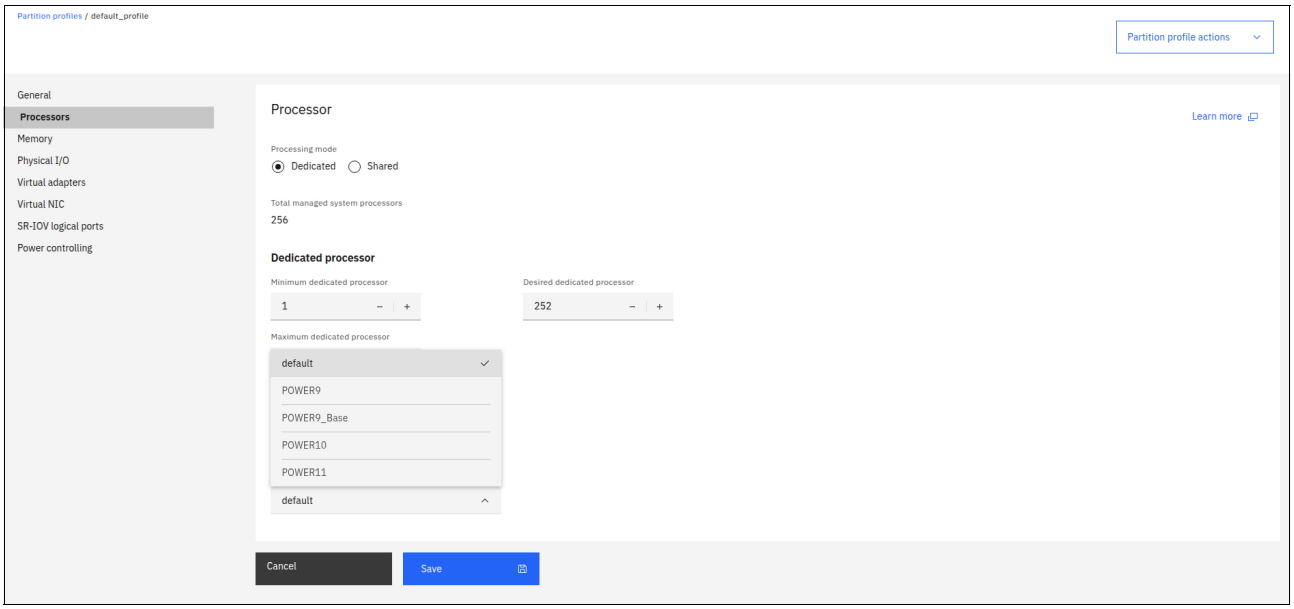

Figure 9-2 Processor modes

Processor compatibility mode is important when planning LPM migrations between different generations of servers. An LPAR that might be migrated to a system with a processor from another generation must be activated in a specific compatibility mode.

9.1.8 Single Root I/O Virtualization

SR-IOV is an extension to the Peripheral Component Interconnect Express (PCIe) specification that enables multiple OSs to share a PCIe adapter simultaneously with little or no runtime involvement from a hypervisor or other virtualization intermediary.

SR-IOV is a PCI-standard architecture that enables PCIe adapters to become self-virtualizing. It supports adapter consolidation through sharing, similar to how logical partitioning enables server consolidation. With an SR-IOV-capable adapter, you can assign virtual slices of a single physical adapter to multiple partitions through logical ports without using a VIOS.

9.1.9 More information about virtualization features

The following IBM Redbooks publications provide more information about virtualization features:

- *IBM PowerVM Best Practices*, SG24-8062
- *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- *IBM Power Systems SR-IOV: Technical Overview and Introduction*, REDP-5065

9.1.10 Resource Groups

Resource Groups is a new feature that is introduced with IBM Power11 to enhance system performance by up to 25% through improved workload optimization and resource affinity. Although SPPs already provide compute capacity isolation by capping the maximum resources that are available to each pool, Resource Groups extend this capability by introducing advanced affinity-based optimizations for more efficient workload dispatching.

Early performance evaluations indicate that when configured effectively, Resource Groups can deliver shared processor performance that closely matches the performance of dedicated processors, particularly in large, partitioned environments.

Resource Groups properties

With the introduction of Resource Groups in IBM Power11, system administrators gain a powerful tool for organizing and optimizing compute resources. Designed for flexibility and performance, Resource Groups enable the grouping of both dedicated and shared processor partitions, enabling more granular control over resource allocation and workload management. This feature enhances system efficiency, supports dynamic reconfiguration, and integrates seamlessly with existing technologies such as SPPs, LPM, and PEP 2.0.

The core features of Resource Groups include the following items:

- Performance optimization

  Resource Groups improve workload affinity, enabling shared processor performance that is nearly equivalent to dedicated processors.

- Flexible configuration
  - Can include both dedicated and shared processor partitions.
  - The configuration specifies the number of general-purpose cores (AIX, IBM i, Linux, and VIOS) and Integrated Facility for Linux (IFL) cores (Linux and VIOS).
  - Powered-off partitions can be reassigned between groups.

- Resources that are not assigned to a user-defined group are placed in the default Resource Group.
- Cores can be dynamically reallocated among groups.

- Affinity and isolation
  - Each Resource Group has its own set of SPPs.
  - Enhances compute capacity isolation and affinity-based dispatching.
- Monitoring and management
  - Usage metrics are available for Resource Groups, SPPs, and the overall system.
  - Dynamic Platform Optimizer (DPO) can run at the system or group level.
- Mobility and compatibility
  - When using LPM, the Resource Group on the target system can be selected.
  - Fully compatible with PEP 2.0.

Resource Groups use cases

There are multiple scenarios in which Resource Groups can deliver significant benefits. This section provides several example use cases. Beyond these examples, use cases can be combined to maximize functions and tailor performance to specific business needs.

Consolidating across multiple lines of business

Figure 9-3 shows how resources are allocated and shared across multiple lines of business on Power10 systems. With the introduction of Resource Groups in Power11, resource allocation is more efficient. This feature enables better workload isolation and improves overall system performance.

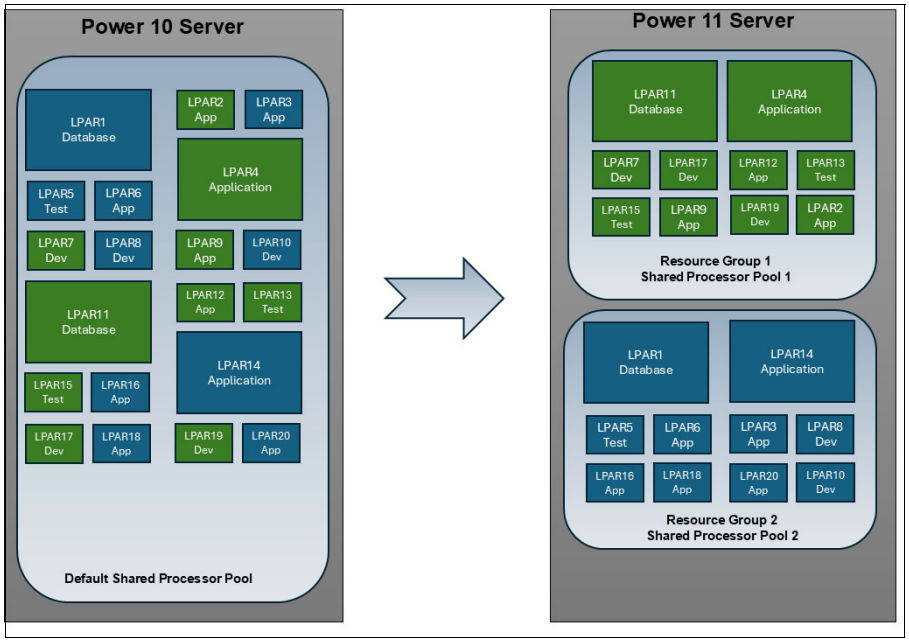

Figure 9-3 Example where Resource Groups consolidate multiple lines of business

Isolating production workloads from test or dev workloads

In this example, the development and test LPARs use resources only from the default Resource Group. This restriction limits their access to shared system resources and isolates them from production workloads.

Figure 9-4 shows how Resource Groups are used to provide this isolation of development and test workloads.

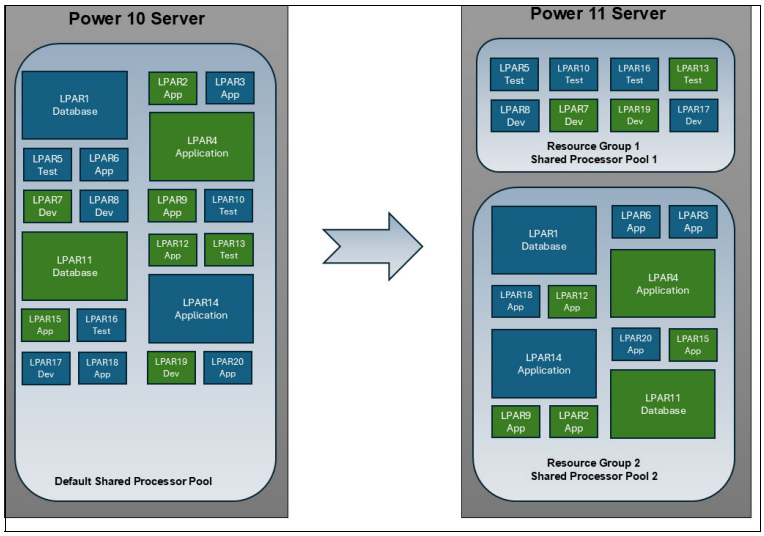

Figure 9-4 Example where Resource Groups isolate workloads

Improving application performance by grouping workload tiers into Resource Groups

In this example, performance improves through better affinity between the database and the application server. This configuration enables more efficient resource usage and faster communication within the same Resource Group.

Figure 9-5 shows how grouping workloads into different Resource Groups can improve application affinity and enhance performance.

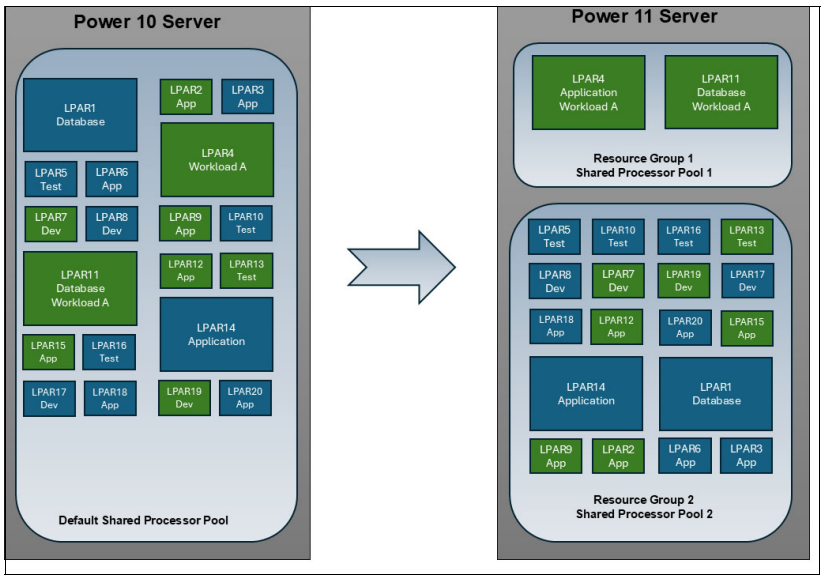

Figure 9-5 Example where Resource Groups improve performance by grouping workload tiers

System-level isolation in multi-server consolidation scenarios

Resource Groups enable effective workload isolation on IBM Power11 systems. They enable organizations to consolidate multiple smaller servers into a single, more powerful Power11 system while logically separating workloads into distinct groups. Each group is managed independently to help ensure performance, security, and resource control across different business functions or environments.

Figure 9-6 shows how using Resource Groups helps maintain system-level isolation during server consolidation.

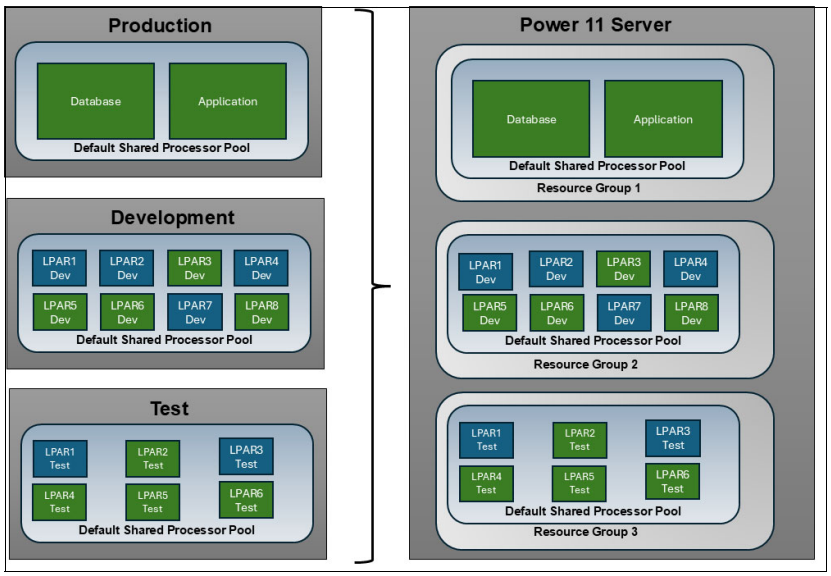

Figure 9-6 Example where Resource Groups provide an extra layer of workload isolation

Improving performance by mapping shared processor pools into Resource Groups

SPPs and Resource Groups work together on IBM Power11 systems to optimize application performance and reduce TCO. This combination provides precise control over resource allocation and workload isolation, helping ensure efficient usage of compute capacity while maintaining performance consistency across diverse workloads.

Figure 9-7 on page 191 shows how SPPs and Resource Groups enhance performance and reduce TCO.

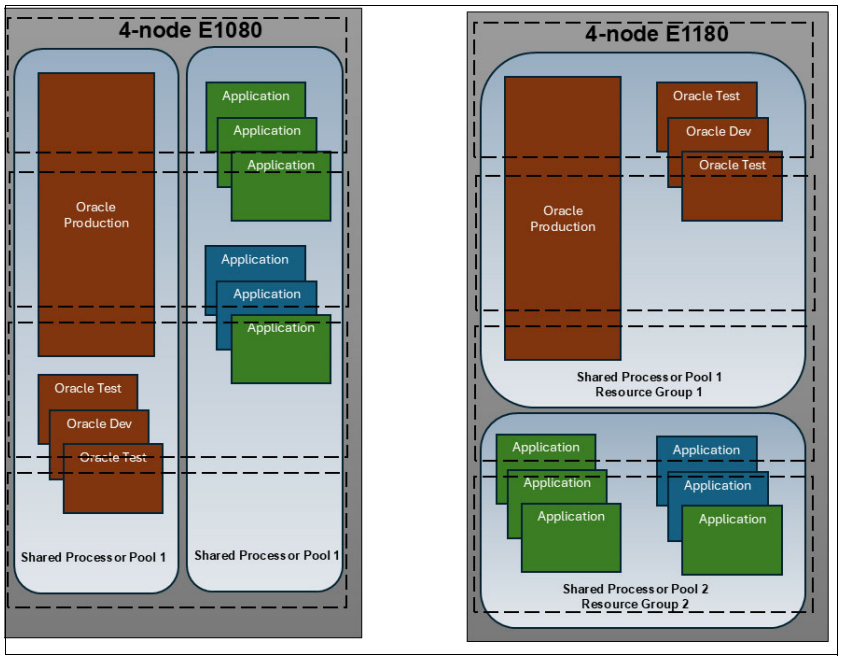

Figure 9-7 Improving performance by mapping SPPs into Resource Groups

Resource Groups Advisor

Resource Groups Advisor (RGA) is a web-based tool that helps with configuring Resource Groups on IBM Power servers. Offered as a no-charge service, RGA analyzes a customer's server setup and provides tailored recommendations for optimal Resource Group configurations. It helps model and validate configurations to help ensure efficient resource allocation and system performance.

To use RGA, see Resource Groups Advisor for IBM Power (IBM login required).

Figure 9-8 shows the welcome window for RGA.

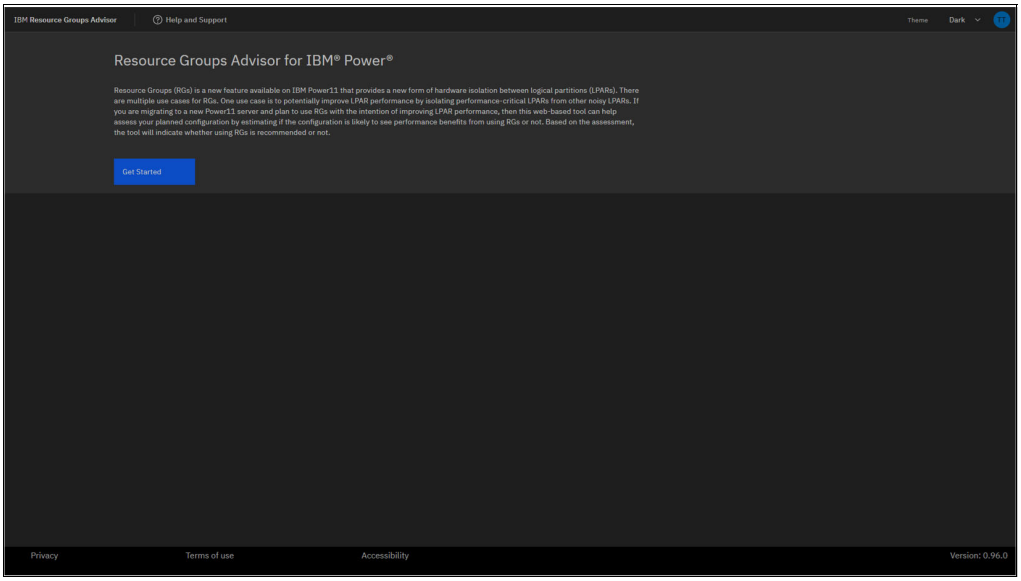

Figure 9-8 Welcome window

To use RGA, complete the following steps:

1. After you select **Getting Started**, a window opens where you can create a configuration or upload a previously saved configuration. Selecting **New Configuration** opens the window that is shown in Figure 9-9.

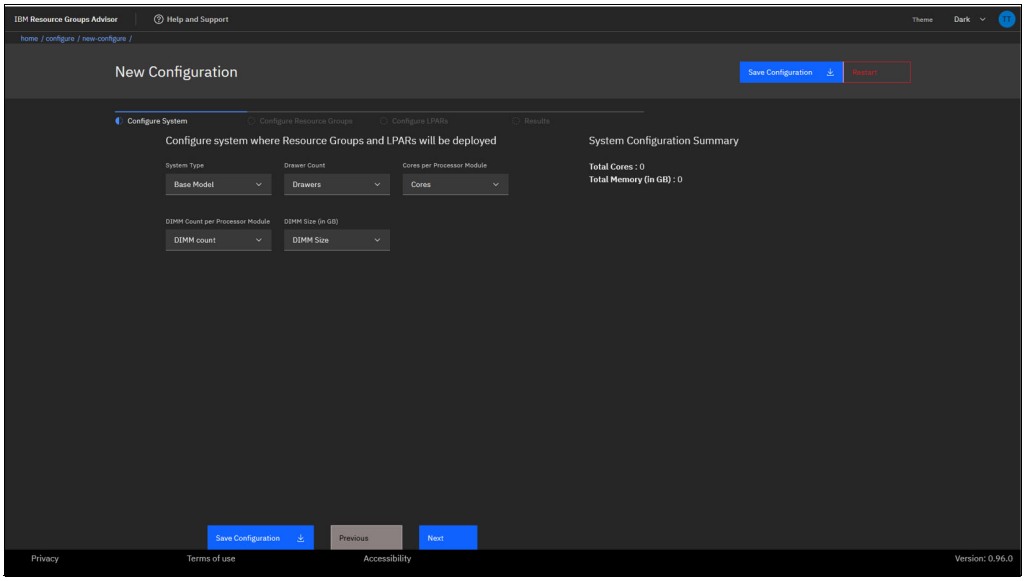

Figure 9-9 New configuration window

2. After you provide the system configuration parameters, click **Next** to view the planned Resource Group configuration for your system, as shown in Figure 9-10. The default Resource Group is always configured, but you can add more Resource Groups to your configuration.

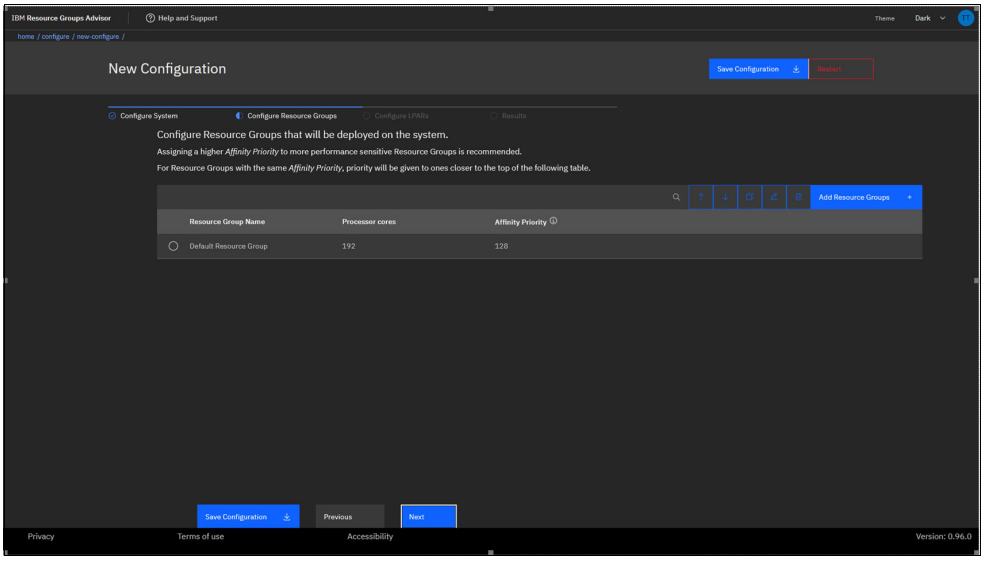

Figure 9-10 Resource Group definitions

3. After you define your Resource Groups, you can specify the LPAR details. Select **Add LPAR** to enter details for each LPAR, such as whether it runs in Shared or Dedicated mode, the processor and memory allocations, and the Resource Group to which the LPAR is assigned. You can also add LPARs from the IBM System Planning Tool.

When all of your LPARs are defined, you see an LPAR summary window similar to Figure 9-11.

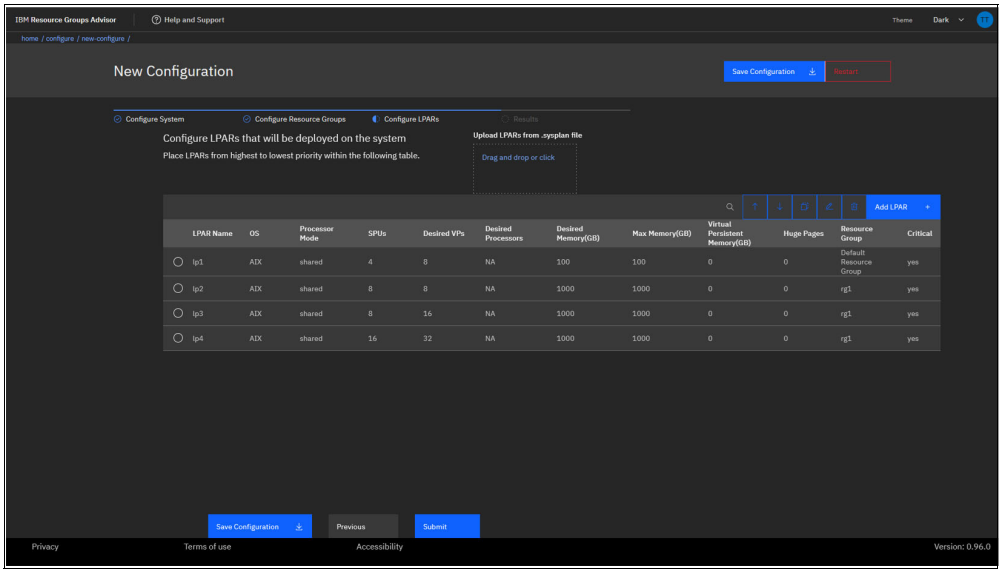

Figure 9-11 LPAR configuration

4. After entering the LPAR information, click **Submit** to get the model's output, similar to what is shown in Figure 9-12.

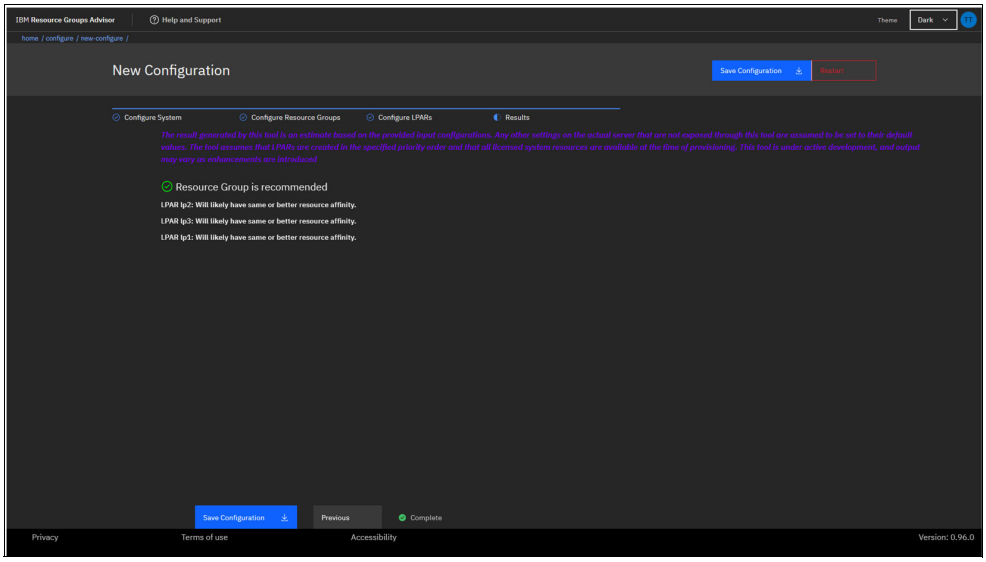

Figure 9-12 Output from the tool

You can save the configuration, or start a new model.

Note: RGA simulates how PowerVM creates Resource Groups and allocates resources to LPARs on a system based on the provided input configurations. No customer data is pulled from actual servers or assessed.

9.2 KVM support

IBM Power11 processor-based servers can run KVM within a PowerVM LPAR. This capability enables the creation and management of lightweight Linux VMs by using standard KVM tools while using the existing resources of the PowerVM LPAR. KVM becomes an extra virtualization option alongside PowerVM on IBM Power servers.

KVM on IBM Power11 processor-based servers is not a replacement for PowerVM. It is an extra capability that brings the power, speed, and flexibility of KVM virtualization technology to a PowerVM LPAR.

KVM-enabled LPARs can host PPC64-LE KVM guests, which are Linux VMs. These guests use the resources (CPU, memory, and I/O) that are allocated to the LPAR by the PHYP. This approach offers flexibility in deploying Linux workloads and can be more cost-effective than other virtualization solutions, especially for organizations that are already invested in the Linux ecosystem. KVM's integration with the Linux kernel can deliver high performance, particularly for Linux-based workloads. This setup supports use cases such as running standard Linux distributions, containers, and other workloads that benefit from the KVM virtualization stack. It also enables consolidation of different workloads on one or more IBM Power11 processor-based servers (see Figure 9-13 for an example).

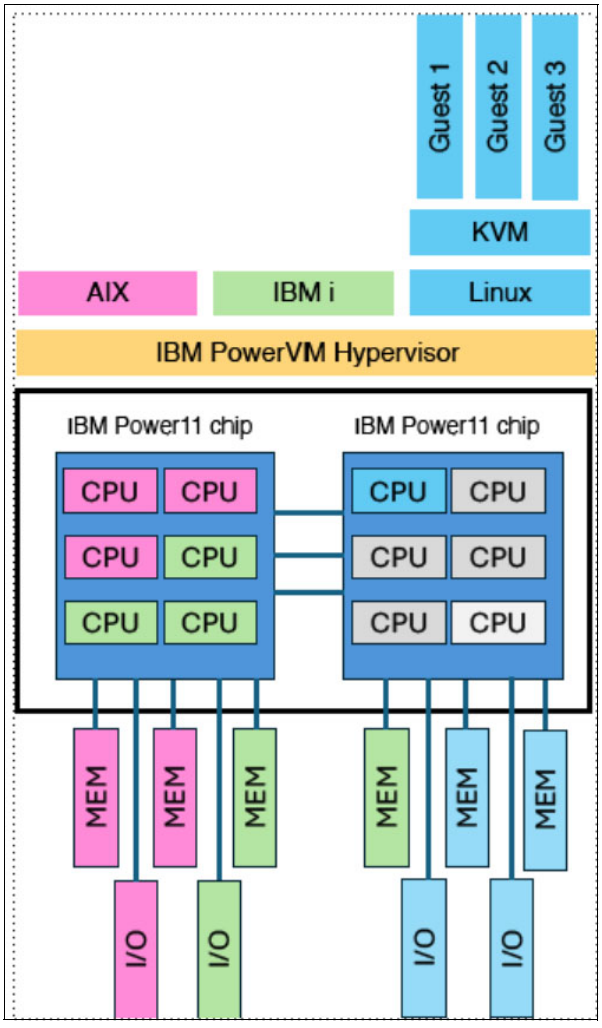

Figure 9-13 IBM Power stack for KVM

In essence, KVM on Power11 provides a way to use the strengths of both PowerVM and KVM, offering a powerful and flexible virtualization environment for Linux workloads.

KVM in a PowerVM LPAR uses the industry-standard Linux KVM virtualization stack (see Figure 9-14) and integrates easily with an existing Linux virtualization environment. KVM in an LPAR is enabled by the following components:

- IBM Power Architecture and Power11 implementation: Provides advanced virtualization capabilities to run multiple OS instances that share hardware resources while maintaining isolation. The Radix MMU architecture enables independent management of page tables for the LPAR and the KVM guest instances on the LPAR.
- PowerVM virtualization stack: Delivers new functions to create and manage KVM guests. These enhancements extend the Power Architecture with new hypervisor interfaces.
- Linux kernel with the KVM module: Provides core virtualization infrastructure to run multiple VMs in a Linux host LPAR. Upstream kernels and supported downstream distributions, such as Fedora and Ubuntu, use the new Power Architecture extensions to create and manage KVM guests in the PowerLinux LPAR.
- QEMU: A user-space component that implements VMs on the host by using KVM functions.
- Libvirt: Provides a toolkit for VM management.

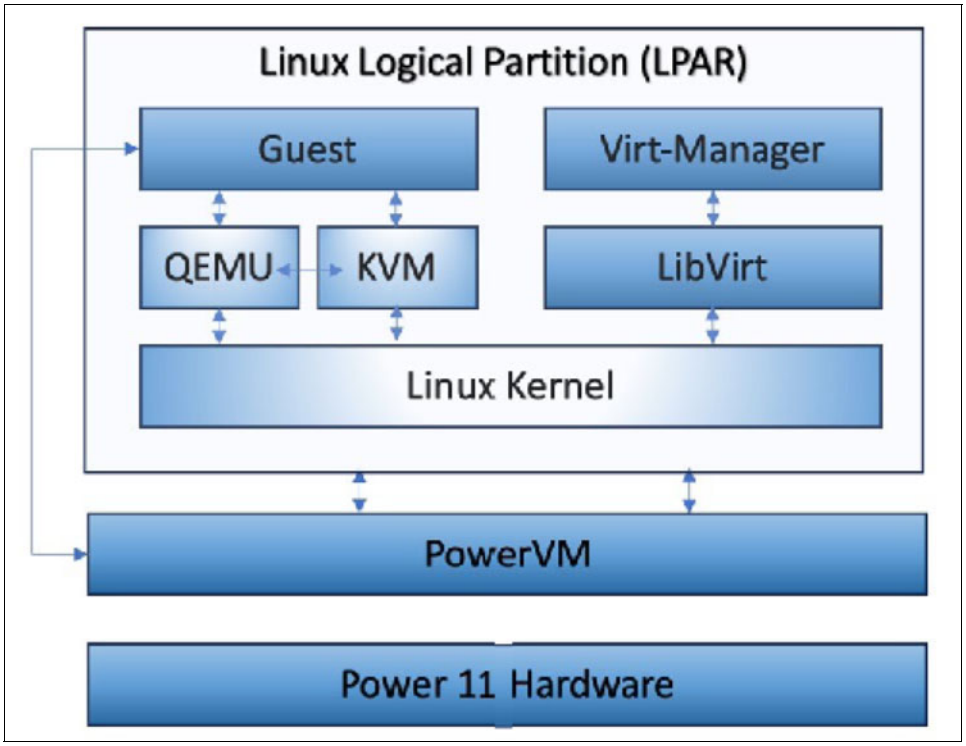

Figure 9-14 Industry-standard Linux virtualization stack

KVM support on IBM Power11 processor-based servers requires the following items:

- The partition must be an LPAR that runs in Power10 processor compatibility mode.
- The partition must be enabled for KVM:
  - For HMC-managed systems, set the partition to KVM Capable.
  - For unmanaged systems, set the default partition environment to Linux KVM on the baseboard management controller (BMC).

- The partition must run in Radix mode (the default MMU mode for Linux LPARs).
- The partition must be assigned to dedicated CPUs with processor sharing set to Never Allow.

The following features are not supported on KVM LPARs:

- Shared processors.
- vPMEM LUNs.
- Platform keystore.
- LPM.
- Dynamic platform optimization.
- Add or remove memory, processor, and I/O DLPAR.

PCI pass-through of LPAR-attached PCI devices to KVM guests for guests are not supported (this feature will be supported in future releases).

For more information, see KVM in a PowerVM LPAR.

10

Hybrid cloud solutions

This chapter describes IBM Power servers as a foundational platform for hybrid cloud environments. These servers support mission-critical workloads with high performance, security, and flexibility. Built on the advanced Power11 architecture, IBM Power servers improve compute efficiency and reduce operational costs. They integrate seamlessly across on-premises, private, and public cloud infrastructures and support a wide range of enterprise operating systems (OSs), including AIX, IBM i, and Linux.

The chapter also explains IBM's strategic approach to hybrid cloud. It shows how Power servers enable scalable, cloud-native application development by using Kubernetes and Red Hat OpenShift. Key management tools, such as IBM Power Virtualization Center (PowerVC), IBM Cloud Management Console (IBM CMC), and Power Enterprise Pools (PEP) 2.0, simplify virtualization, optimize resource allocation, and improve operational agility. The chapter also introduces IBM Power Virtual Server (PowerVS), a cloud-based extension of IBM Power that allows organizations to modernize at their own pace. PowerVS provides flexible, pay-as-you-go access to a virtualized IBM Power infrastructure, enabling seamless workload migration, rapid provisioning, and strong business continuity capabilities.

The following topics are covered in this chapter:

- IBM Power Private Cloud with Shared Utility Capacity
- Red Hat OpenShift
- IBM Power Virtual Server

© Copyright IBM Corp. 2025.

10.1 IBM Power Private Cloud with Shared Utility Capacity

IBM Power Private Cloud with Shared Utility Capacity introduces a flexible and efficient approach to resource management through Power Enterprise Pools (PEP) 2.0. This capability enables multiple IBM Power servers to operate as a unified resource pool, enabling shared compute capacity across systems. Designed for clients deploying private cloud infrastructure, PEP 2.0 improves operational agility by supporting dynamic, minute-by-minute consumption of compute resources.

In this model, each system in the pool is provisioned with a set number of base processor and memory activations. When configured as a PEP, these base activations and their associated OS entitlements are aggregated to define the pool's baseline capacity. At the same time, all hardware resources across the pool are fully activated and available for workload deployment, enabling usage to exceed the base capacity when needed. IBM CMC continuously monitors average resource usage across the pool and compares it to the base entitlement. Any usage beyond the base is billed as metered capacity, either through Prepaid Capacity Credits or monthly billing. PEP 2.0 eliminates the need for manual reallocation of activations between systems, providing granular resource sharing, improved cost efficiency, and simplified management for enterprise IT environments.

PEP with Shared Utility Capacity on IBM Power E1180 systems provides enhanced multi-system resource sharing and by-the-minute tracking and consumption of compute resources across a collection of Power E1180 and IBM Power E1080 systems within a single PEP 2.0. Shared Utility Capacity offers flexibility to tailor initial system configurations with the right mix of purchased and pay-for-use consumption of processor, memory, and software. Clients with existing PEP 2.0 configurations of Power E1080 systems can add one or more Power E1180 systems to their pool and migrate workloads at their own pace because Power E1180 and Power E1080 systems can interoperate and share compute resources within the same pool. Clients with Mobile Capacity on a Power E1180 can upgrade their system to support PEP 2.0 and use Shared Utility Capacity resources.

A Power infrastructure that is consolidated onto Power E1180 systems can simplify system management so IT teams can focus on optimizing business results instead of moving resources within the data center. Shared Utility Capacity resources are tracked by virtual machine (VM) and monitored by IBM CMC, which integrates with local Hardware Management Consoles (HMCs) to manage the pool and track resource usage by system and VM, by the minute, across a pool.

Clients do not need to overprovision capacity on each system to support growth because all available processor and memory on all systems in a pool are activated and available for use. On Power E1180 systems, Mobile and Shared Utility Capacity capabilities are available through the purchase of a Power Enterprise Pools Subscription (5765-P2E) for each Power E1180 system that is configured for use in a PEP.

When a PEP 2.0 pool starts, each eligible Power E1180 system's purchased processor activations, memory activations, and supported OS entitlement resources become Base Capacity resources as part of the pool and are aggregated across the defined pool for consumption monitoring. Metered Capacity is the additional installed processor and memory resource beyond each system's Base Capacity. It is activated and available for immediate use when a pool starts, then monitored by the minute by IBM CMC.

Metered resource usage is charged only for minutes exceeding the pool's aggregate Base resources. Usage charges are debited in real time against a client's Capacity Credits (5819-CRD) on account or billed monthly, where available.

Important: Only two consecutive generations of Power servers are supported in the same PEP. Clients who want to add Power11 to an existing pool with Power9 and Power10 processor-based systems should contact their IBM representative or IBM Technology Expert Labs to discuss migration planning and best practices.

10.1.1 IBM Cloud Management Console

As private and hybrid cloud deployments expand, enterprises need deeper management insights into these increasingly complex environments. Tools that deliver unified analytics and consolidated information are essential for efficient infrastructure operations.

IBM CMC for Power provides a comprehensive view of the cloud environment, regardless of the number of systems or data centers. It offers the following tools:

- Centralized inventory management of systems and virtual components
- Consolidated performance metrics to help optimize resource usage and system performance across all data centers
- Aggregated logging and analytics for enhanced operational insights

This unified approach helps IT teams manage infrastructure more effectively and make informed decisions with confidence. IBM CMC now supports IBM Power11 systems, helping ensure compatibility with the latest Power Architecture.

IBM CMC also monitors and manages PEP 2.0 pools in the enterprise. The IBM CMC Enterprise Pools 2.0 application monitors base and metered capacity across a PEP 2.0, with summary and detailed views of real-time and historical resource consumption by logical partition (LPAR).

IBM CMC provides a cloud-based interface to monitor, manage, and optimize an IBM Power infrastructure. It supports AIX, IBM i, and Linux VM workloads and is accessible through a secure, web-based dashboard. IBM CMC allows clients to accomplish the following tasks:

- View the health and performance of IBM Power servers.
- Monitor capacity and performance metrics.
- Get insights into firmware levels and configurations.
- Receive proactive support and alerts.

For organizations with limited IT staff or seeking to streamline infrastructure oversight, IBM CMC enables essential monitoring and management without requiring complex, locally hosted solutions.

IBM CMC integrates with the HMC and provides advanced insights into performance, capacity, firmware compliance, fix planning, and system health. With Power11 scalability and built-in virtualization, IBM CMC helps administrators maintain control over complex hybrid IT environments, whether on-premises or connected to IBM Cloud or other supported clouds.

IBM CMC is delivered as a software-as-a-service (SaaS) offering and licensed on a per-server subscription basis. Power11 systems register to IBM CMC through their HMCs, which securely transmit telemetry and configuration data to the cloud portal. IBM provides a no-cost base offering and optional chargeable add-ons, including predictive analytics, fix automation, and extended historical performance data retention. Clients can choose monthly or annual billing, and entitlement can be managed through IBM Passport Advantage® or the IBM Cloud portal.

10.2 Red Hat OpenShift

As the demand for digital services grows and new offerings are released, organizations face the challenge of delivering exceptional customer experiences while maintaining resilient and highly available (HA) services. To meet these demands, they must innovate by developing new applications, modernizing existing ones, and helping ensure seamless scalability.

Organizations can achieve these goals by adopting cloud technologies through full cloud migration, developing cloud-native applications, or implementing a hybrid cloud approach.

Red Hat OpenShift on IBM Power provides developers and IT operations (ITOps) teams with the flexibility and speed to build, deploy, and manage applications across diverse environments, whether on-premises, in the cloud, or across multiple infrastructures. This platform accelerates digital transformation by simplifying scalability and improving security.

10.2.1 Red Hat OpenShift on IBM Power

Red Hat OpenShift Container Platform is fully supported on IBM Power servers, providing a robust and scalable environment for running cloud-native applications and modernizing traditional workloads on the Power Architecture (ppc64le). This integration enhances the IBM Power software ecosystem by enabling enterprise-grade container orchestration for building, deploying, and managing containerized applications with Red Hat OpenShift.

Red Hat OpenShift on IBM Power offers a flexible and efficient platform for hybrid IT environments, enabling cloud-native applications to coexist with traditional VM-based workloads. By using the same infrastructure, it supports the colocation of containerized workloads alongside existing applications running on AIX, IBM i, or Linux, helping ensure low-latency communication between applications and data. IBM Power servers provide a resilient and secure platform for cloud implementations by using Red Hat OpenShift. This section also describes the benefits of using a multi-architecture cluster and provides implementation guidelines to help you deploy a multi-architecture cluster with IBM Power control plane nodes and x86- or AMD-based worker nodes.

Red Hat OpenShift is an enterprise Kubernetes platform that provides a strong foundation for developing, deploying, and scaling cloud-native applications. It extends Kubernetes with more features and tools to improve productivity and security, making it an ideal choice for organizations adopting container technology at scale.

Red Hat OpenShift is a unified platform for building, modernizing, and deploying applications at scale. It includes a complete set of services for bringing applications to market on your choice of infrastructure. Red Hat OpenShift delivers a consistent experience across public cloud, on-premises, hybrid cloud, and edge environments.

Red Hat OpenShift offers a unified, flexible platform to address many business needs, from an enterprise-ready Kubernetes orchestrator to a comprehensive cloud-native application development platform that can be self-managed or delivered as a fully managed cloud service.

Red Hat OpenShift clusters

A Red Hat OpenShift Container Platform cluster consists of multiple nodes that can run on physical servers or VMs. At a minimum, the following node requirements apply:

- Three control plane nodes to manage and maintain the cluster
- Two worker nodes to run containerized workloads
- A temporary bootstrap node that is used during installation to host configuration and installation files

In Red Hat OpenShift Container Platform 4.18, the bootstrap and control plane nodes must run Red Hat Enterprise Linux CoreOS (RHCOS), a minimal, immutable container host OS that is derived from Red Hat Enterprise Linux (RHEL). Compute nodes can run either RHCOS or standard RHEL, depending on the deployment architecture. RHEL machines are deprecated in Red Hat OpenShift Container Platform 4.16 and will be removed in a future release.

Multiple-Architecture Compute clusters

Enterprises expanding their operations often deploy applications in heterogeneous environments that use different types of hardware. For example, data centers might use x86_64 servers and IBM Power servers, but edge locations might include ARM-based devices for power efficiency. IBM Power is an ideal choice for running Red Hat OpenShift clusters. However, some applications in the environment might not be compatible with Power Architecture nodes. You can retain the benefits of IBM Power while simplifying cluster management and improving resource usage by incorporating both IBM Power and x86_64 architecture nodes in the same cluster.

The release of Red Hat OpenShift 4.14 introduced the Multiple-Architecture Compute feature for IBM Power. Multiple-Architecture Compute enables a single heterogeneous cluster, enabling fit-for-purpose computing so clients can align tasks and applications to CPU strengths and software availability rather than a single architecture. Red Hat OpenShift 4.15 expanded this support to enable a cluster to use an IBM Power control plane and x86 architecture worker nodes.

With Multiple-Architecture Compute for Red Hat OpenShift Container Platform, you can use two compute architectures, such as ppc64le and amd64, within a single cluster. This feature provides versatility and optimization for solutions that span multiple architectures.

Multi-Architecture Compute offers several benefits:

- Platform independence

  Applications can run across various hardware platforms, including Intel servers in data centers, ARM-based devices in remote locations, and IBM Power servers in corporate environments.

- Reduced complexity

  Standardizing application deployment across architectures streamlines operations and eliminates the need to maintain separate stacks for different hardware.

- Cost efficiency

  Different hardware architectures offer varying cost-to-performance ratios, which can reduce overall infrastructure expenses.

- Optimal resource usage

  Organizations can select the most cost-effective architecture for each workload. For example, ARM servers might be suitable for lightweight services, and x86_64 or IBM Power servers can handle compute-intensive tasks.

- Energy efficiency

  Architectures such as ARM are known for low power consumption, which can reduce energy costs in scale-out scenarios like IoT and mobile services.

- Scalability and flexibility

  Multi-architecture deployments enhance scalability and operational agility, which is critical for businesses with fluctuating workloads.

Multi-architecture support enable companies to accomplish the following goals:

- Elastic scaling across platforms

  Businesses can dynamically allocate resources among different hardware types to handle demand surges without being restricted to a single architecture.

- Avoiding vendor lock-in

  Multiple-Architecture Compute reduces dependency on a single hardware supplier, providing greater flexibility and negotiating power.

- Performance optimization

  Each architecture has unique strengths. Multiple-Architecture Compute aligns application requirements with the most suitable architecture.

- Tailored solutions

  Some applications benefit from the high I/O throughput of IBM Power servers, and others perform best on the multi-core processors of x86_64 architectures.

- Specialized computing support

  Certain workloads, such as machine learning, might require GPUs or specialized hardware that are better supported on specific architectures.

Single-node clusters

For a simpler and more accessible deployment experience, Red Hat OpenShift supports a single-node cluster in which the management and worker functions run on a single node. This approach is suitable for development, testing, or edge computing environments.

However, a single-node configuration has inherent limitations, most notably the absence of built-in high availability (HA). If the node fails, the entire cluster becomes unavailable, which can result in downtime and data loss. The HA features of IBM Power infrastructure, when designed with redundant power, networking, and storage, can reduce these risks.

10.2.2 Red Hat OpenShift AI on IBM Power Servers

Red Hat OpenShift AI delivers enterprise-grade capabilities for AI and machine learning within the Red Hat OpenShift platform. It enables organizations to build, train, deploy, and monitor AI models in a unified, containerized environment. Red Hat OpenShift AI is based on Red Hat OpenShift, an enterprise-grade Kubernetes platform that is designed to support AI and ML workloads across hybrid and multicloud infrastructures. By integrating AI workflows into the same infrastructure that is used for traditional applications, Red Hat OpenShift AI simplifies operations, accelerates development cycles, and provides consistent governance and security across the AI lifecycle. It enables organizations to build, deploy, and manage AI-enabled applications at scale across hybrid cloud environments.

Red Hat OpenShift AI Self-Managed is available by installing the Red Hat OpenShift AI Operator and configuring it to manage the stand-alone components of the platform. It is supported on Red Hat OpenShift Container Platform across multiple architectures, including ppc64le (IBM Power), s390x (IBM Z), and x86_64. Here are the supported infrastructure providers:

- IBM Power (technology preview)
- IBM Z (technology preview)
- IBM Cloud
- Red Hat OpenStack
- Bare Metal
- Hosted control planes on Bare Metal
- Amazon Web Services (AWS)
- Google Cloud Platform
- Microsoft Azure
- VMware vSphere
- Oracle Cloud

IBM Power servers, which are known for high performance, scalability, and reliability, are suitable for compute-intensive AI workloads. Integrating Red Hat OpenShift AI with IBM Power servers enables enterprises to use both platforms for robust, production-grade AI deployments.

As of Red Hat OpenShift AI Self-Managed 2.20, the IBM Power Architecture (ppc64le) is available as a technology preview. At the time of writing, model deployments on IBM Power are supported only in standard mode. Red Hat OpenShift AI provides an integrated platform for developing, training, serving, and monitoring AI and ML models. Support for vLLM is also available on the IBM Power architecture as a technology preview, with accessible vLLM runtime templates for experimentation and development.

When deployed on IBM Power, Red Hat OpenShift AI benefits from the platform's high-performance architecture, which is optimized for data-intensive and compute-heavy workloads. The latest IBM POWER processors feature built-in AI inferencing capabilities, delivering faster insights and reduced latency for real-time applications. This capability will be enhanced with the planned availability of the IBM Spyre Accelerator in IBM Power11 servers. The IBM Spyre Accelerator is a Peripheral Component Interconnect Express (PCIe)-attached AI card that is designed for low-precision AI arithmetic and enterprise-grade AI algorithms, making it ideal for deploying large-scale models and generative AI frameworks.

Together, Red Hat OpenShift AI and IBM Power provide a scalable, production-ready foundation for modern AI initiatives. Whether running models built with PyTorch, TensorFlow, or vLLM, this integrated stack supports hybrid cloud deployments and enables enterprises to use AI with confidence, helping ensure performance, efficiency, and flexibility across diverse workloads and environments.

10.2.3 IBM Cloud Pak

IBM Cloud Pak is a suite of modular, containerized software solutions that are designed to accelerate digital transformation. Built on Red Hat OpenShift, these solutions enable organizations to modernize applications, automate business operations, manage data, and integrate AI across hybrid cloud environments with consistency and scalability.

Each IBM Cloud Pak includes IBM middleware, open-source technologies, Kubernetes operators, and enterprise-grade security. This combination provides a robust foundation for innovation and agility in the cloud. IBM Cloud Pak solutions use a bundled approach that you can use to accelerate your modernization journey by packaging everything that you need to get started.

IBM Cloud Pak solutions offer three main benefits:

- They are comprehensive and simple to use.
- They are supported by IBM.
- They run anywhere Red Hat OpenShift runs.

IBM Cloud Pak solutions on IBM Power

Optimized for deployment on IBM Power, an IBM Cloud Pak solution delivers high performance and efficiency for containerized workloads. These solutions support a wide range of use cases, including data, automation, AI, and security, while helping ensure seamless integration across on-premises and cloud environments.

IBM Power enhances the value of an IBM Cloud Pak solution by providing superior performance, scalability, and cost efficiency for modern workloads. Combined with Red Hat's open-source ecosystem, it enables businesses to modernize applications and build new cloud-native solutions on a unified, enterprise-ready platform.

Several IBM Cloud Pak solutions are supported on IBM Power at the time of writing, each focused on a specific domain:

- IBM Cloud Pak for Applications
- IBM Cloud Pak for Data
- IBM Cloud Pak for Integration
- IBM Cloud Pak for AIOps

IBM Cloud Pak for Applications

IBM Cloud Pak for Applications is an enterprise-ready, containerized software solution that is designed to modernize existing applications and develop new cloud-native applications. Built on IBM WebSphere® offerings and Red Hat OpenShift Container Platform, it provides a comprehensive set of tools to help organizations transition across public, private, and hybrid clouds.

IBM Cloud Pak for Applications includes IBM Cloud Transformation Advisor, an AI-powered tool that helps refactor applications. The solution provides automated vulnerability assessment and identification to help ensure continuous security compliance across all deployment environments. It also automates audit reporting, simplifying compliance management. Developers can use their preferred IDEs to build and deploy applications, with support for modern run times and DevOps workflows. This integration streamlines the development process and improves productivity.

IBM Cloud Pak for Data

IBM Cloud Pak for Data unifies and simplifies data collection, organization, and analysis. It is ideal for AI and analytics workloads. IBM Cloud Pak for Data is a unified, preintegrated data and AI platform that is designed to help organizations collect, organize, analyze, and infuse AI into their data. Running natively on the Red Hat OpenShift Container Platform, it supports deployment across various cloud environments, including IBM Cloud, AWS, and Microsoft Azure.

The platform provides secure access to data at its source, eliminating the need for data migration and reducing data silos to help ensure seamless data integration. It creates a trusted, business-ready analytics foundation by simplifying data preparation, policy enforcement, security, and compliance, while automating data governance and the AI lifecycle. IBM Cloud Pak for Data includes tools for building, deploying, and managing AI and machine learning models, scaling these capabilities consistently across the organization to enable comprehensive data analysis and insights.

By using AI throughout the business with trust and transparency, the platform supports the end-to-end AI workflow and helps ensure effective integration of AI into business processes. Offering a single interface for end-to-end analytics with built-in governance, IBM Cloud Pak for Data simplifies the management of data and AI capabilities. Its scalable Kubernetes environment enables organizations to expand their data and AI capabilities as needed. Supporting multicloud deployments, it provides agility and avoids vendor lock-in, making it a powerful tool for accelerating the journey to AI and unlocking the value of data for AI-driven digital transformation.

IBM Cloud Pak for Integration

IBM Cloud Pak for Integration is a comprehensive, AI-powered hybrid integration platform that is designed to connect applications, data, systems, and services across any environment. It provides a unified experience with a suite of integration tools that streamline the creation, management, and deployment of integrations. Running on Red Hat OpenShift, IBM Cloud Pak for Integration supports both cloud and on-premises deployments, helping ensure scalability and security.

The platform includes components such as IBM API Connect for managing application programming interfaces (APIs), IBM App Connect for no-code integration, and IBM Event Streams for real-time data processing. By using AI and automation, IBM Cloud Pak for Integration accelerates the integration process, reduces manual workflows, and improves responsiveness to real-time events. It is an ideal solution for organizations that want to modernize their integration capabilities and drive digital transformation.

IBM Cloud Pak for Business Automation

IBM Cloud Pak for Business Automation is a modular set of integrated software components that is designed to automate work and accelerate business growth. Built for any hybrid cloud, it simplifies complex workflows, facilitates records management, and improves overall productivity. The platform uses AI to identify gaps and build low-code and no-code automations, making it simpler to streamline operations. Running on Red Hat OpenShift, IBM Cloud Pak for Business Automation supports containerized deployments across various cloud environments, providing flexibility and scalability.

Key features include automating case and process workflows, converting unstructured content into valuable data, and using software robots to complete tasks based on AI insights. This comprehensive automation solution helps organizations improve efficiency, reduce operational costs, and drive continuous process improvements.

IBM Cloud Pak for AIOps

IBM Cloud Pak for AIOps is an advanced, AI-powered platform that is designed to enhance ITOps by using AI and machine learning. It integrates seamlessly with existing ITOps toolchains to provide comprehensive visibility, proactive incident management, and automated remediation. By analyzing data from sources such as logs, metrics, and events, IBM Cloud Pak for AIOps helps IT teams predict and resolve issues before they impact business operations.

The platform supports hybrid cloud environments, enabling organizations to manage IT infrastructure across on-premises, cloud, and containerized environments. Key features include event correlation and compression, anomaly detection, root cause analysis, and automated runbooks, all aimed at reducing mean time to resolution (MTTR) and improving operational efficiency. With collaborative tools and real-time insights, IBM Cloud Pak for AIOps enables IT teams to innovate faster, reduce operational costs, and improve the reliability of mission-critical workloads.

For more information about IBM Cloud Pak, see IBM Documentation.

10.2.4 Other cloud enablement solutions

From a Red Hat software perspective, a comprehensive set of solutions is available to accelerate modernization efforts, including Red Hat Runtimes, Red Hat 3scale API Management, Red Hat Fuse, and Red Hat AMQ.

10.3 IBM Power Virtual Server

PowerVS is a cloud-based infrastructure as a service (IaaS) offering that enables businesses to run IBM Power workloads in a flexible, scalable, and secure virtual environment. Built on the same architecture as IBM Power servers use on-premises, it supports seamless hybrid cloud integration, making it ideal for enterprises that want to modernize their IT infrastructure without abandoning existing investments. With support for AIX, IBM i, and Linux OSs, PowerVS provides a versatile platform for mission-critical applications. It offers an efficient way to move AIX, IBM i, and Linux on IBM Power workloads to the cloud without a time-consuming and risky migration to a different platform.

One of the key advantages of PowerVS is its ability to deliver high performance and reliability for enterprise workloads. It uses IBM POWER processors, which are optimized for data-intensive tasks and high-throughput computing. The platform also offers features such as dynamic resource allocation, automated scaling, and integrated backup and disaster recovery (DR) options. These capabilities help organizations maintain business continuity and meet demanding service-level agreements (SLAs).

PowerVS integrates with a wide range of IBM Cloud services and third-party tools, enabling users to build and manage hybrid cloud environments. It supports DevOps practices, containerization with Red Hat OpenShift, and AI and machine learning workloads, making it suitable for both traditional enterprise applications and modern cloud-native development. With its pay-as-you-go pricing model and global availability, PowerVS offers a cost-effective and agile solution for businesses aiming to innovate and grow in a digital-first world.

PowerVS provides a cloud experience for running IBM i, AIX, and Linux workloads, offering fast, self-service provisioning and flexible compute, memory, and storage resources. It is designed to help organizations modernize their infrastructure, migrate to hybrid cloud models, and optimize data center resources.

Here are the key features and benefits of PowerVS:

- Flexible and scalable: Users can adjust compute, memory, and storage resources on demand, scaling up or down as needed.
- Pay-as-you-use: Billing is based on consumption, enabling organizations to manage costs effectively.

- Hybrid cloud integration: PowerVS facilitates seamless integration of AIX and IBM i workloads with cloud-native applications and services.
- Performance and security: PowerVS uses the performance and security features of IBM Power while offering the agility of cloud computing.
- Managed infrastructure: IBM handles hardware maintenance, enabling organizations to focus on workloads from a software and application perspective.
- DR: PowerVS can be used to implement DR solutions for critical applications.
- Simplified cloud adoption: PowerVS streamlines the cloud adoption process for IBM i, AIX, and other IBM Power workloads.

10.3.1 IBM Power Virtual Server options

PowerVS is a single offering that is available in two variations: off-premises, where the infrastructure components are in IBM data centers, and on-premises, where the infrastructure components are in the client's data center. The on-premises variation is known as IBM Power Virtual Server Private Cloud. Both variations provide a cloud-based consumption model in which you pay for resources as they are used and are managed through the same interfaces.

Figure 10-1 shows the two different implementations of PowerVS.

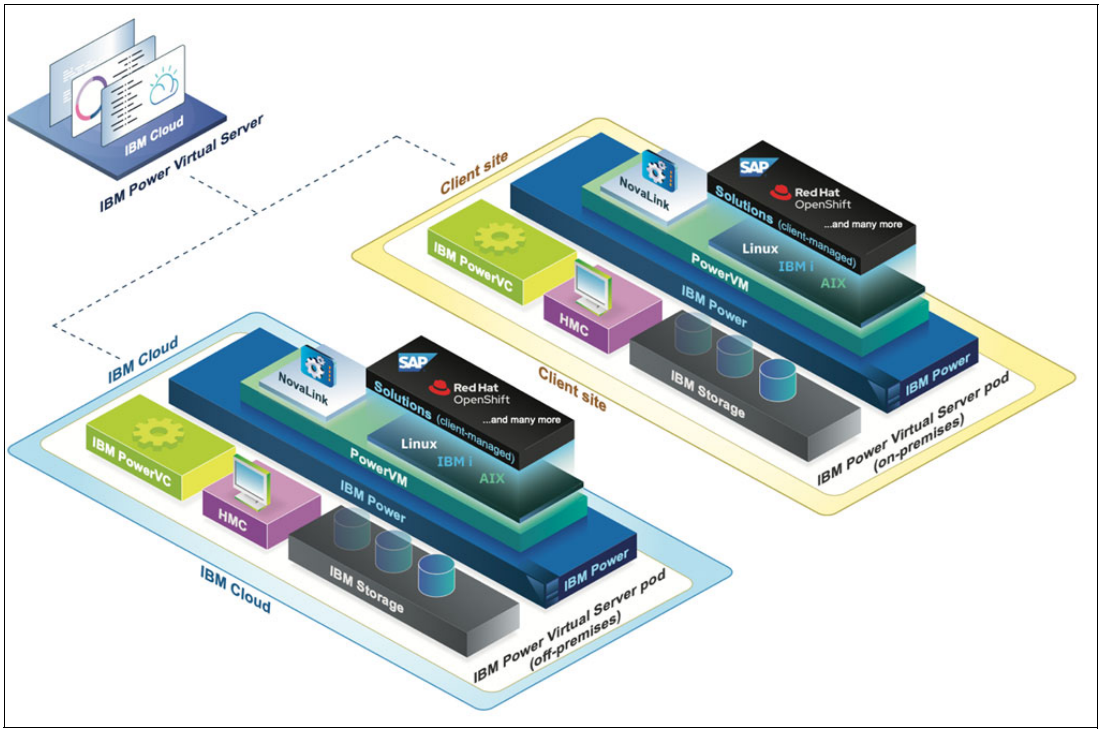

Figure 10-1 IBM Power Virtual Server options

PowerVS Private Cloud is described in 10.3.3, "IBM Power Virtual Server Private Cloud" on page 209.

Use cases for IBM Power Virtual Server

PowerVS supports a wide range of use cases, particularly for enterprises that rely on IBM Power servers for mission-critical workloads. Here are some common and impactful scenarios:

- Hybrid cloud modernization

  Organizations with on-premises IBM Power can extend their infrastructure to the cloud by using PowerVS. This approach enables a hybrid cloud model in which workloads can be moved or replicated to the cloud for scalability, DR, or testing without re-platforming or rewriting applications. It is especially useful for businesses running AIX or IBM i environments that want to modernize gradually.

- High availability and disaster recovery (HADR)

  PowerVS provides a reliable platform for DR by allowing businesses to replicate on-premises workloads to the cloud. If there is a failure, workloads can be quickly restored in the cloud, minimizing downtime. It also supports HA configurations to help ensure business continuity for critical applications.

- Development and testing environments

  Developers can use PowerVS to quickly create isolated environments for application development, testing, and quality assurance. This capability is valuable for teams working on AIX, IBM i, or Linux based applications because it eliminates the need for dedicated on-premises hardware and enables faster iteration and deployment cycles.

- SAP HANA and enterprise resource planning (ERP) workloads

  PowerVS is certified to run SAP HANA and other SAP ERP applications. Enterprises can migrate or extend their SAP environments to the cloud for improved scalability, performance, and cost efficiency, especially when paired with IBM Power10 processors.

- AI and machine learning

  PowerVS supports AI and machine learning workloads that require high compute power and memory bandwidth. It integrates with IBM Watson® and Red Hat OpenShift, enabling data scientists to build, train, and deploy models by using familiar tools while using the performance of IBM Power.

- Database hosting

  Many organizations use PowerVS to host enterprise-grade databases such as IBM Db2, Oracle, or PostgreSQL. The platform offers high I/O throughput and reliability, making it ideal for transaction-heavy applications in industries such as banking, retail, and logistics.

- Application hosting

  For businesses with applications running on AIX or IBM i, PowerVS provides a cloud-based alternative to on-premises hardware. This approach enables companies to extend the life of critical applications while gradually modernizing their infrastructure.

- Compliance and data sovereignty

  Organizations in regulated industries such as healthcare, finance, and government use PowerVS to meet strict compliance and data residency requirements. IBM offers regional data centers and compliance certifications to support these needs.

10.3.2 IBM Power Virtual Server in the cloud

In this variation, PowerVS resources are in IBM data centers with dedicated networking and storage area network (SAN)-attached storage. Customers can choose from the 21 data centers that are available at the time of writing by selecting the location that is closest to their users. IBM Power clients that rely on a private cloud infrastructure can quickly and economically extend their Power IT resources to the cloud.

In these IBM data centers, PowerVS resources are isolated from other IBM Cloud servers through separate networks and direct-attached storage. This offering provides flexibility to choose from different IBM Power server models, IBM-supplied OS images, and storage tiers that are based on specific workload requirements. Customers can also bring their own image under a Bring Your Own License (BYOL) model, enabling them to migrate preconfigured images from their enterprise infrastructure to PowerVS.

This offering supports features such as shared processor pools (SPPs), public connectivity for VMs, placement groups, and global replication service (GRS), which are used based on customer requirements.

For more information about IBM PowerVS, see Getting Started with IBM Power Virtual Server.

10.3.3 IBM Power Virtual Server Private Cloud

For clients that want a cloud experience running on IBM Power servers but must maintain control of their data and applications, whether for regulatory, security, or performance reasons, IBM offers an option to bring the PowerVS offering to a data center of the client's choice.

IBM Power Virtual Server Private Cloud extends all the benefits of PowerVS into a client or other data center. This offering provides managed IaaS at client locations with metered consumption and no upfront costs, supporting Hybrid by Design delivery of services.

The IBM Power Virtual Server Private Cloud offering has the following features:

- Maintain customer data and workloads on site.

  Some enterprises have workloads or data that are regulated and cannot be hosted off-premises. Others have sensitive workloads or ultra-low latency requirements that are best served on site and near other workloads.

- Maintain customer data in region and specific locations.

  Country sovereignty regulations require some data and workloads to remain in the country. According to a recent IBM Institute for Business Value study, 61% of cloud leaders cite security or compliance as reasons for moving certain workloads from public clouds to private clouds or on-premises data centers[1].

- Provide a seamless hybrid cloud experience.

  Enterprises can create a unified hybrid cloud landscape by integrating PowerVS running at both IBM and client sites. A single interface enables management of all VMs and infrastructure, with the flexibility to use as-a-service models and intentional workload placement on-premises or off-premises.

[1] Source: https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/data-story-hybrid-cloud-ai

- Deliver a predictable charging model.

  Both PowerVS offerings (off-premises at an IBM location and on-premises at a client site) include compute, memory, storage, and OS licenses that are fully metered by the hour. Clients pay only for what they use each month, with no upfront payment.

- Streamline ITOps.

  IBM manages the infrastructure, freeing enterprises to focus on business outcomes. IBM owns, delivers, and sets up PowerVS in the client's data center and provides a fully managed solution, including monitoring, security, firmware updates, and infrastructure management.

- Provide enhanced security and control of data.

  PowerVS integrates with IBM Cloud tools to manage security, reducing the need for clients to manage Power infrastructure security while maintaining sensitive data and workloads on premises.

The physical infrastructure is delivered as a point of delivery (pod) that is deployed in the client's data center. A *pod* contains compute, storage, and network components that are installed in one or more racks. These racks are interconnected to provide a self-contained infrastructure, including customer-usable components, spare components, and management components.

IBM site reliability engineers (SREs) maintain the pod, which is managed through the IBM Cloud platform. Each pod is associated with an IBM Cloud Satellite® location that is owned by the client's IBM Cloud account. This architecture enables horizontal scaling of private cloud infrastructure by adding more pods to meet workload requirements.

For more information about IBM Power Virtual Server Private Cloud, see *Introduction to IBM Power Virtual Server Private Cloud*, REDP-5745.

10.3.4 Introducing IBM Power11 in the cloud

Power11 will be available on the day of announcement in PowerVS and IBM Power Virtual Server Private Cloud. With the launch of IBM Power11, clients and independent software vendors (ISVs) gain immediate access to the latest Power hardware in the cloud, enabling faster innovation and greater flexibility. Whether modernizing applications or enhancing business continuity, Power11 in PowerVS delivers a seamless hybrid cloud experience.

Here are some key benefits:

- Rapid deployment: Start Power11 virtual servers in the cloud in under 10 minutes.
- Accelerated modernization: Enable agile development and testing on the latest Power infrastructure.
- Enhanced business resiliency: Strengthen continuity strategies with cloud-based Power11 resources.
- Data sovereignty and compliance: Meet regulatory and sensitive data requirements with on-premises options.
- Consistent hybrid cloud experience: Enjoy unified operations across on-premises and cloud environments.

Figure 10-2 on page 211 shows the initial implementation of Power11 processors in PowerVS on announcement day.

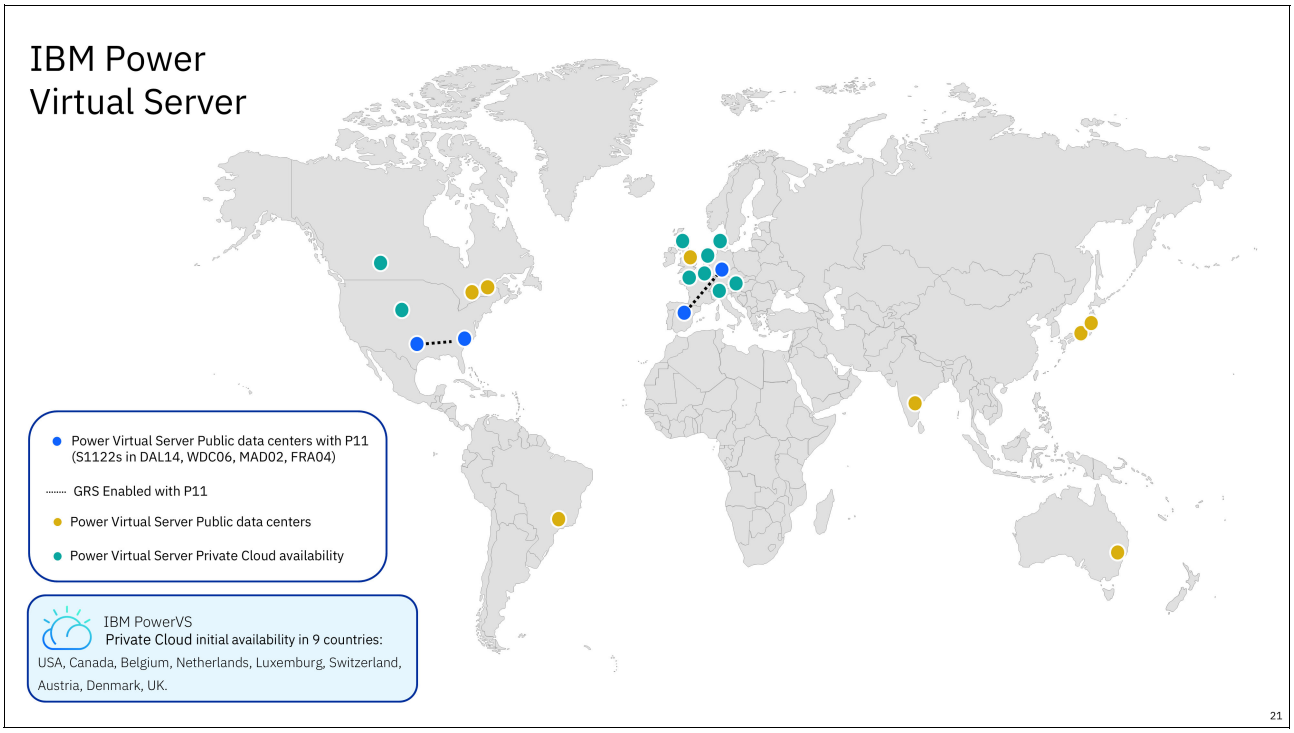

Figure 10-2 Availability of Power11 in IBM Power Virtual Server

Power11 technology in PowerVS enables clients to use new functions immediately without an initial capital investment. Workloads can be implemented securely by using the trusted security capabilities of Power11:

- Quantum-safe infrastructure compliance with optional workload cryptographic inventory discovery
- A unified cyber resiliency solution that responds to evolving cyberthreats and regulatory standards, with optional third-party services

In addition, clients can benefit from the improved performance of Power11, as shown in Figure 10-3.

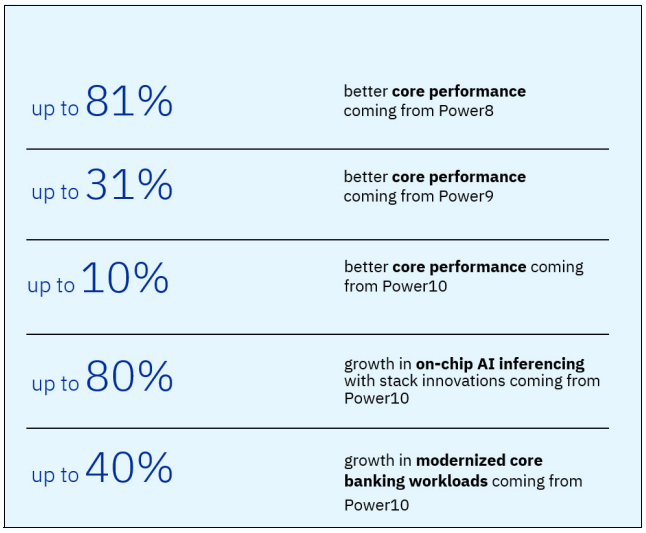

Figure 10-3 Potential performance benefits from Power 11 in PowerVS

Abbreviations and acronyms

AER	Autonomous Error Resolution
AES	Advanced Encryption Standard
AI	artificial intelligence
AME	Active Memory Expansion
AML	anti-money laundering
AMM	Active Memory Mirroring
AOC	active optical cable
API	application programming interface
ASMI	Advanced System Management Interface
AWS	Amazon Web Services
BMC	baseboard management controller
BRMS	Backup, Recovery, and Media Services
BSC	blind swap cassette
BYOL	Bring Your Own License
CBU	Capacity BackUp
CD	continuous deployment
CEAC	Common Europe Advisory Council
CPC	central processing complex
CRM	customer relationship management
CRU	customer-replaceable unit
CVEs	Common Vulnerabilities and Exposures
DBM	Dedicated Backup Machine
DCM	Dual-Chip Module
DLPAR	dynamic logical partitioning or dynamic logical partition
DOL	Department of Labor
DORA	Digital Operational Resilience Act
DPO	Dynamic Platform Optimizer
DR	disaster recovery
DRO	Dynamic Resource Optimization
DWPD	Drive Writes Per Day
EA	effective address
EEH	Enhanced Error Handling
EHR	electronic health record
EIA	Electronic Industries Alliance
EMR	electronic medical record
EoFS	End of Fix Support
EOL	end-of-life
ERP	enterprise resource planning
ESA	Electronic Service Agent
eSCM	entry Single Core Module
ESM	Enclosure Services Manager
ESS	Entitled Systems Support
EUS	Extended Update Support
FC	Fibre Channel
FFDC	First Failure Data Capture
FHE	Fully Homomorphic Encryption
FLRT	Fix Level Recommendation Tool
FSFC	Full System FlashCopy
FSP	Flexible Service Processor
FSR	Full System Replication
GLVM	Geographic Logical Volume Manager
GPLv2	GNU General Public License version 2
GRS	global replication service
HA	high availability
HADR	high availability and disaster recovery
HCL	HashiCorp Configuration Language
HHHL	half-height, half-length
HIS	Healthcare Information System
HMC	Hardware Management Console
HNDL	harvest now, decrypt later
HPC	high-performance computing
HPT	hardware page table
IaaS	infrastructure as a service
IaC	Infrastructure as Code
IASP	independent auxiliary storage pool
IBM	International Business Machines Corporation
IBM CMC	IBM Cloud Management Console
ILE	Integrated Language Environment
IPL	initial program load
ISA	Instruction Set Architecture
ISC	Integrated Stack Capacitor
iSCSI	internet Small Computer Systems Interface
ISV	independent software vendor

© Copyright IBM Corp. 2025.

ITIC	Information Technology Intelligence Consulting (ITIC)
ITOps	IT operations
KVM	Kernel-based Virtual Machine
KYC	know-your-customer
LIC	Licensed Internal Code
LLDP	Link Layer Discovery Protocol
LLM	large language model
LP	low-profile
LPAR	logical partition
LPM	Live Partition Mobility
LPP	licensed program product
LWE	Learning With Errors
MCU	Memory Controller Unit
MES	Miscellaneous Equipment Specification
MMA	Matrix Math Accelerator
MPIO	Multipath I/O
MSS	Merkle Signature Scheme
MTTR	mean time to resolution
NPIV	N_Port ID Virtualization
NVMe	Non-Volatile Memory Express
ODB	Operational Database
OMI	Open Memory Interface
OS	operating system
OSHA	Occupational Safety and Health Administration
PCIe	Peripheral Component Interconnect Express
PDU	power distribution unit
PEP	Power Enterprise Pools
PFA	Predictive Failure Analysis
PFW	partition firmware
PHB	PCIe Host Bridge
PHYP	Power Hypervisor or PowerVM Hypervisor
PoCs	proofs of concept
PowerVC	IBM Power Virtualization Center
PowerVS	IBM Power Virtual Server
PQC	Post-Quantum Cryptography
PTF	Program Temporary Fix
QSE	Quantum-Safe Encryption
RAG	retrieval augmented generation
RAS	reliability, availability, and serviceability
RBAC	role-based access control
RDMA	remote direct memory access
REST	Representational State Transfer
RGA	Resource Groups Advisor
RHCOS	Red Hat Enterprise Linux CoreOS
RHEL	Red Hat Enterprise Linux
Ring-LWE	Ring-Learning With Errors
RMC	Resource Monitoring and Control
ROP	return-oriented programming
RPO	Recovery Point Objective
RTO	Recovery Time Objective
SaaS	software as a service
SAN	storage area network
SCM	Single-Chip Module
SCU	System Control Unit
SE2	Standard Edition 2
SEA	Shared Ethernet Adapter
SEC	Securities and Exchange Commission
SFF	small form factor
SIEM	Security Information and Event Management
SLA	service-level agreement
SME	subject matter expert
SMT	simultaneous multithreading
SP	Service Pack
SPP	shared processor pool
SPT	System Planning Tool
SR-IOV	Single Root I/O Virtualization
SRE	site reliability engineer
SSD	solid-state drive
SSH	Secure Shell
SSIC	System Storage Interoperation Center
SUMA	Service Update Management Assistant
SWMA	Software Maintenance Agreement
TAM	Technical Account Manager
TBW	terabytes written
TCE	translation control entity
TCO	total cost of ownership
TE	Trusted Execution
TL	Technology Level
TME	Transparent Memory Encryption
TR	Technology Refresh
UAK	Update Access Key

UOV	Unbalanced Oil and Vinegar
UPIC	universal power interconnect
VIOS	Virtual I/O Server
VLAN	virtual local area network
VM	virtual machine
VMRM	VM Recovery Manager
VPD	Vital Product Data
VTL	Virtual Tape Library
XMSS	eXtended Merkle Signature Scheme
ZPD	zero planned downtime

Related publications

The publications that are listed in this section are considered suitable for a more detailed description of the topics that are covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topics in this document. Some publications that are referenced in this list might be available in softcopy only.

- *Creating Red Hat OpenShift Multiple Architecture Clusters with IBM Power*, SG24-8565
- *IBM i 7.6 features and function*, SG24-8588
- *IBM PowerHA SystemMirror and IBM VM Recovery Manager Solutions Updates*, REDP-5694
- *IBM Power Security Catalog*, SG24-8568
- *Introduction to IBM Power Virtual Server Private Cloud*, REDP-5745
- *Modernization Techniques for IBM Power*, SG24-8582
- *Using Ansible for Automation in IBM Power Environments*, SG24-8551

You can search for, view, download, or order these documents and other Redbooks, Redpapers, Web Docs, drafts, and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- Red Hat OpenShift AI Self-Managed 2.20

  https://docs.redhat.com/en/documentation/red_hat_openshift_ai_self-managed/2.20
- Red Hat OpenShift AI Self-Managed 2.20 Release Notes

  https://docs.redhat.com/en/documentation/red_hat_openshift_ai_self-managed/2.20/pdf/release_notes/Red_Hat_OpenShift_AI_Self-Managed-2.20-Release_notes-en-US.pdf
- Red Hat OpenShift AI Self-Managed Life Cycle

  https://access.redhat.com/support/policy/updates/rhoai-sm/lifecycle
- Red Hat OpenShift AI Service Definition

  https://access.redhat.com/articles/7102731
- Red Hat OpenShift AI: Supported Configurations

  https://access.redhat.com/articles/rhoai-supported-configs

© Copyright IBM Corp. 2025.

- Red Hat OpenShift for IBM Power

  https://www.redhat.com/en/resources/openshift-ibm-power-systems-datasheet#:~:text=Red%20Hat%20OpenShift%20and%20IBM,on%20IBM%20Power%20are%20supported
- Red Hat OpenShift Operator Life Cycles

  https://access.redhat.com/support/policy/updates/openshift_operators
- RHEL Versions Utilized by RHEL CoreOS and Red Hat OpenShift Container Platform

  https://access.redhat.com/articles/6907891

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Redbooks

IBM Power E1180: Introduction and Overview

(0.2”spine)
0.17”<->0.473”
90<->249 pages

SG24-8587-00

ISBN 0738462284

Printed in U.S.A.

Get connected