

IBM Power E1180: Introduction and Overview

Tim Simon

Jordan Antonov

Ivaylo Bozhinov

Rohit Chauhan

Nicolae Chirea

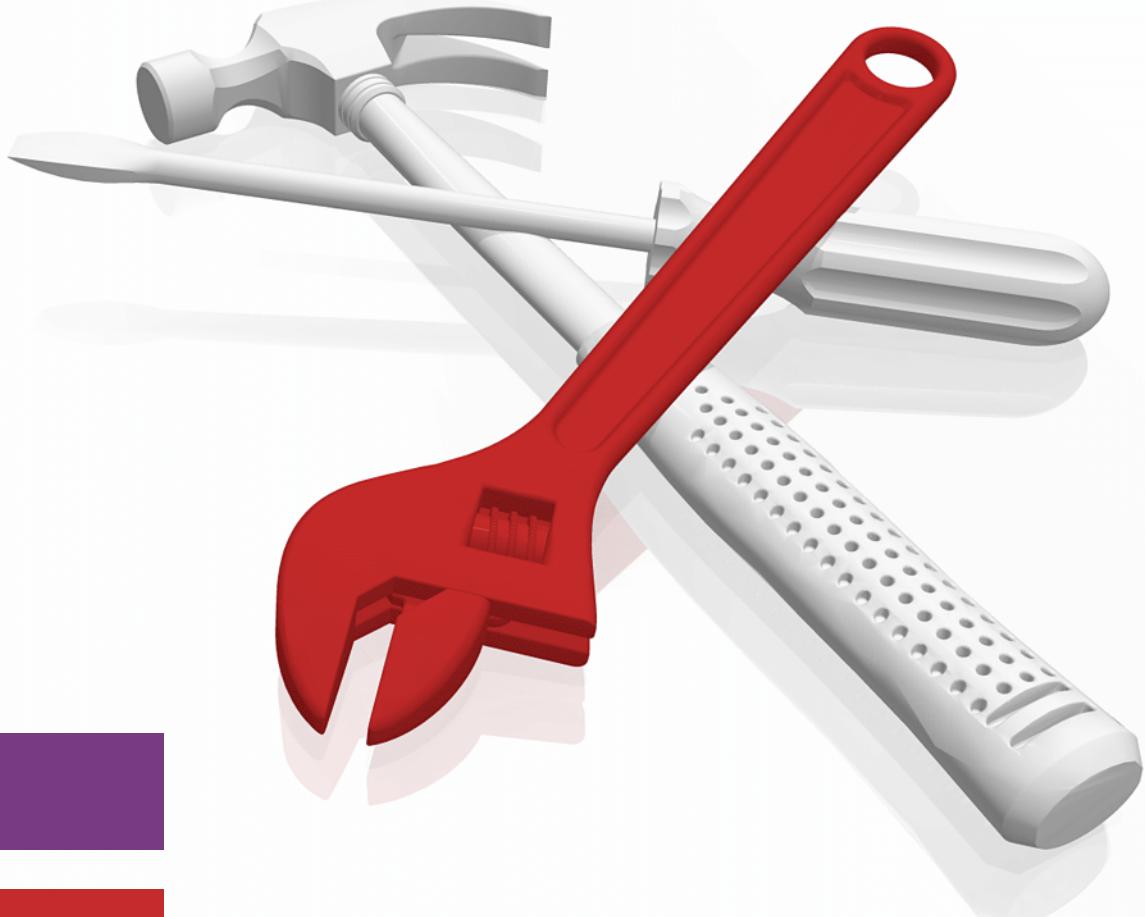
Turgut Genc

Bartłomiej Grabowski

Andrey Klyachkin

Tsvetomir Spasov

Henry Vo



IBM Power

 **Hybrid Cloud**

IBM
®

Redbooks



IBM Redbooks

IBM Power E1180

July 2025

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

First Edition (July 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

Notices	vii
Trademarks	viii
Preface	ix
Authors	ix
Now you can become a published author, too!	xii
Comments welcome	xii
Stay connected to IBM Redbooks	xiii
Chapter 1. Overview of Power11	1
1.1 IBM Power11: Advancing enterprise computing	2
1.2 IBM Power roadmaps	3
1.2.1 IBM Power Processor Roadmap	3
1.2.2 IBM AIX Roadmap	4
1.2.3 IBM i Roadmap	5
1.2.4 PowerVM VIOS roadmap	6
1.2.5 Red Hat Enterprise Linux and IBM Power	6
1.2.6 SUSE Enterprise Linux and IBM Power	7
1.3 IBM Power E1180: Resilient, efficient, and scalable	7
1.3.1 New Power11 server portfolio	9
1.4 Power11: Trusted, autonomous, and modern	12
1.4.1 Building a trusted infrastructure with Quantum Safe Encryption	12
1.4.2 Protect, detect, and recover with Power Cyber Vault	17
1.4.3 Quantum-safe compliance with one-click inventory discovery	20
1.4.4 Zero planned downtime and end-to-end orchestration for maintenance events	21
1.4.5 Automated data collection for faster error resolution	24
1.4.6 Increase efficiency and reduce energy consumption through smart scheduling options	25
1.4.7 Superior performance of up to 25% improvement for consolidation of Oracle 19c Workloads	29
1.4.8 Day 1 availability of Power11 in Power Virtual Server with hybrid billing	30
1.4.9 Off-chip AI acceleration with IBM Spyre	31
1.4.10 Red Hat OpenShift with support for expanded software ecosystem	32
1.5 Power11 processor architecture	33
1.6 Operating system support	36
1.6.1 AIX operating system	37
1.6.2 IBM i operating system	37
1.6.3 Linux operating system	37
1.6.4 VIOS	38
1.7 Firmware and Hardware Management Console	38
1.7.1 HMC requirements	39
1.8 IBM Power E1180 rack integration and handling guidelines	39
Chapter 2. IBM Power11 Enterprise Server – Power E1180	41
2.1 IBM Power E1180 overview	42
2.1.1 IBM Power11: Engineered for Unmatched Uptime	43
2.2 IBM Power E1180 components	45
2.2.1 System nodes	45
2.2.2 System control unit	46

2.3 Processors	47
2.4 Memory	48
2.4.1 Active Memory Mirroring	51
2.4.2 Pervasive memory encryption	52
2.4.3 Active Memory Expansion	52
2.5 PCIe Slots	53
2.6 Storage and USB features	54
2.7 I/O expansion drawers	54
2.8 Environmental requirements	55
2.8.1 Electrical characteristics	55
Chapter 3. I/O Subsystem.	61
3.1 Internal I/O	62
3.1.1 Internal PCIe subsystem	62
3.1.2 Internal NVMe storage subsystem	67
3.1.3 Internal USB subsystem	70
3.2 Enhancing I/O Scalability with Expansion Drawers	70
3.2.1 PCIe expansion card	71
3.2.2 Supported I/O Drawers	75
3.2.3 Non-supported drawers	75
3.2.4 PCIe Gen 4 I/O Expansion Drawer	76
3.2.5 NED24 NVMe Expansion Drawer	78
3.2.6 General rules for I/O drawers connection	82
3.3 Supported adapters	84
3.3.1 Supported Fibre Channel adapters	84
3.3.2 Network and RoCE adapters	85
3.3.3 Network and RoCE transceivers and cables	86
3.3.4 SAS and Tape adapters	87
3.3.5 NVMe	88
3.3.6 USB	90
3.3.7 Cryptographic adapters	90
3.4 Other device support	91
Chapter 4. Artificial Intelligence Support	93
4.1 On chip support	94
4.2 AI Acceleration with IBM Spyre adapter	94
4.2.1 Deploying AI in the Enterprise	95
4.2.2 Fit for Purpose AI Acceleration	95
4.2.3 IBM Spyre adapter	97
4.3 Solutions	98
4.3.1 Barriers to scaling AI and the IBM Power technology advantage	99
4.3.2 IBM Power AI Portfolio: from Data to Business Value	100
4.3.3 Strategic outcomes enabled by IBM Power	100
4.3.4 Uniqueness of the platform	100
Chapter 5. Automation and Management	103
5.1 Hardware Management Console overview	104
5.1.1 HMC Options	104
5.1.2 BMC network connectivity rules for 7063-CR2 HMC	106
5.1.3 High availability HMC configuration	107
5.1.4 HMC code level requirements	107
5.1.5 HMC currency	108
5.1.6 New features	108
5.1.7 Using the Automated Maintenance tool	108

5.2 Ansible	111
5.3 Terraform	114
5.4 PowerVC	117
Chapter 6. Operating Systems.....	119
6.1 Subscription licensing	121
6.1.1 AIX subscription licensing	121
6.1.2 IBM i subscription licensing	122
6.2 AIX	122
6.2.1 AIX 7.3 Key Features	122
6.2.2 Supported Levels	123
6.2.3 AIX Maintenance levels	123
6.2.4 Licensing.....	124
6.2.5 New AIX Editions	125
6.3 IBM i	126
6.3.1 Intro to IBM i	127
6.3.2 Supported levels.....	127
6.3.3 IBM i Software Tiers	127
6.3.4 New IBM i Features supported by Power11	127
6.3.5 IBM i Licensing	128
6.3.6 Subscription bundles.....	129
6.3.7 IBM i Software Maintenance (SWMA).....	133
6.3.8 CBU and DBM Licensing Models for IBM i	133
6.3.9 Upgrading an existing IBM i license to 7.6	133
6.3.10 IBM i Simplification	134
6.4 Linux on IBM Power	134
6.4.1 Supported distributions	135
6.4.2 Licensing.....	139
6.4.3 KVM support	140
6.5 PowerVM Virtual I/O Server	140
6.6 Setting your LPAR compatibility mode	141
Chapter 7. Enterprise Solutions	143
7.1 High Availability and Disaster Recovery Solutions	144
7.1.1 PowerHA SystemMirror for AIX.....	144
7.1.2 IBM PowerHA SystemMirror for i	144
7.1.3 VM Recovery Manager (VMMR).....	145
7.2 IBM Db2	146
7.3 Oracle	146
7.3.1 Running Oracle Standard Edition 2 on IBM Power.....	148
7.4 SAP HANA	149
7.5 Banking	150
7.6 Health Care	152
7.6.1 Epic	153
Chapter 8. Servicing Power11	155
8.1 IBM Maintenance	156
8.2 IBM Expert Care	156
8.3 IBM tools and interfaces	157
8.4 FSP	158
8.5 ASMi	159
8.6 Entitled System Support	159
8.7 Firmware Updates	160
8.7.1 Update Access Keys.....	167

Chapter 9. Virtualization	171
9.1 PowerVM	172
9.1.1 IBM PowerVM Hypervisor	172
9.1.2 Multiple shared processor pools	176
9.1.3 Virtual I/O Server	177
9.1.4 Live Partition Mobility	178
9.1.5 Active Memory Expansion	179
9.1.6 Remote Restart	179
9.1.7 Power processor modes	180
9.1.8 Single Root I/O Virtualization	181
9.1.9 More information about virtualization features	181
9.1.10 Resource groups	181
9.2 KVM support	187
Chapter 10. Hybrid Cloud Solutions	191
10.1 IBM Power Private Cloud with Shared Utility Capacity	192
10.1.1 IBM Cloud Management Console	193
10.2 Red Hat OpenShift	194
10.2.1 Red Hat OpenShift on Power	194
10.2.2 Red Hat OpenShift AI on IBM Power Servers	197
10.2.3 IBM Cloud Paks	198
10.2.4 Other cloud enablement solutions	200
10.3 Power Virtual Server	200
10.3.1 Power Virtual Server options	201
10.3.2 Power Virtual Server in the Cloud	203
10.3.3 Power Virtual Server Private Cloud	203
10.3.4 Introducing IBM Power11 in the Cloud	205
Related publications	207
IBM Redbooks	207
Online resources	207
Help from IBM	208

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.

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®	IBM Security®	PowerVM®
Concert®	IBM Spyre™	Rational®
Db2®	IBM Watson®	Redbooks®
DS8000®	IBM watsonx®	Redbooks (logo)  ®
FlashCopy®	IBM Z®	Satellite™
IBM®	Instana®	Spyre™
IBM API Connect®	Integrated Language Environment®	System z®
IBM Cloud®	Language Environment®	SystemMirror®
IBM Cloud Pak®	Micro-Partitioning®	Tivoli®
IBM Concert®	Passport Advantage®	Turbonomic®
IBM FlashSystem®	Power8®	watsonx®
IBM Garage™	Power9®	WebSphere®
IBM Research®	PowerHA®	

The following terms are trademarks of other companies:

Intel, Intel Xeon, 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.

Ansible, Fedora, OpenShift, Red Hat, 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.

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

Preface

This IBM Redbook provides a comprehensive overview of the IBM Power E1180 – the flagship server in the IBM Power11 family. It highlights the implementation of the Power11 processor, an evolution of the Power10 chip used in the IBM Power E1080, and explores the system's architecture and performance capabilities in depth.

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

- ▶ Trusted: Delivers exceptional business continuity with 99.9999% availability, zero planned downtime for maintenance, quantum-safe cryptography, and rapid system recovery.
- ▶ Autonomous: Enhances 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 – leveraging off-chip AI acceleration via IBM Spyre™ and a robust partner ecosystem.

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

Altogether, the IBM Power E1180 emerges as a high-performance, enterprise-grade platform designed to power 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 based in Tulsa, Oklahoma, USA. With over 40 years of experience at IBM, primarily in technical sales, Tim has built a distinguished career centered around 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 spectrum of IBM technologies, with deep experience in developing customer solutions using IBM Power, IBM Storage, and IBM System z® platforms.

Jordan Antonov is based in Bulgaria and has been with IBM for over 10 years, primarily contributing to the IBM Power Servers support team. He is currently part of the Product Engineering team, where he continues to apply 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. Over the years, 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 with IBM Bulgaria. His main area of expertise is solving complex hardware and software issues on IBM

Power products, IBM AIX, Virtual I/O Server (VIOS), HMC, IBM i, IBM PowerVM, and Linux on Power servers. He has been with IBM since 2015, and provides reactive break-patch, proactive, preventive, and cognitive support.

Rohit Chauhan is an IBM Champion and a Senior Technical Specialist with expertise in IBM i platform and IBM Power Servers. He is also an IBM Redbooks author and works at Tietoevry Tech Services, Stavanger, Norway, which is an IBM Business Partner and one of the biggest IT service providers in the Nordics. He has over 12 years of experience working on the IBM Power platform with design, planning, and implementation of IBM i infrastructure, which includes high availability and disaster recovery (HADR) solutions for many customers during this tenure. Prior to 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 strong focus on the IBM i platform and security. As the leader of IBM TechXchange's "IBM i Security and Innovation" user group, he also 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 solutions to customers for the IBM i platform. Additionally, 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 transitioning to system administration. By 2007, he was managing over 120 iSeries systems and actively participating in infrastructure upgrades, LPAR implementations, XSM solutions, and iASP migrations. Since 2007, he has been working at Seidor, an IBM Business Partner, where he is involved in the full lifecycle of IBM Power servers, from initial setup to deployment. His expertise includes version upgrades, high-availability 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 FlashSystems, IBM DS8000®, and SAN networking. In recent years, Nicolae has expanded his focus to Linux on Power, leading major SAP HANA implementations in Spain. Passionate about automation, he actively promotes Ansible within the IBM® Power community and has deployed modern monitoring tools like 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 Power and Storage Solution Architect of IBM Technology Expert Labs in the UK. He holds an MSc degree in Computer Science Engineering from Yildiz Technical University in Istanbul. He has 20+ years of experience in IBM Power Servers and he has been the EMEA team leader for IBM PowerVS, Enterprise Pools, Performance, Migration workshops, and Technology Expert Labs tools.

Bartłomiej Grabowski is an IBM Champion and a Principal Systems Specialist at DHL IT Services, where he supports IBM Power Servers globally. With over 20 years of experience in enterprise IT solutions, Bartłomiej 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 IBM Power Servers, IBM i, IBM PowerHA, IBM PowerVM, system performance optimization, and storage solutions. Bartłomiej is also a Platinum IBM Redbooks author and shares his insights and technical knowledge through his blog at www.theibmi.org contributing actively to the IBM i community.

Andrey Klyachkin is a solution architect at eNFence, an IBM Business Partner in Germany. He has more than 25 years experience in UNIX systems, designing and supporting AIX® and Linux systems for different customers worldwide. He has co-authored many IBM AIX and IBM Power certifications. He is an IBM Champion and IBM AIX Community Advocate. He is a Red

Hat Certified Engineer and Red Hat Certified Instructor. He attends international and local events for IBM Power, such as IBM TechXchange, GSE, and Common Europe Congress.

Tsvetomir Spasov is a Power Servers Hardware Product Engineer in Sofia, Bulgaria. He has 9 years of experience with IBM in an RTS, SME and PE role. His main areas of expertise is HMC, 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 masters degree in Electrical Engineering from Technical University of Sofia.

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 a variety of 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.

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

Nicole Nett
IBM Chief Architect Power Servers, Austin, TX

Lyudmila Simeonova
IBM AIX IO Device Drivers/PowerVM VIOS, Sofia, Bulgaria

Milen Paiakov
IBM Power Courseware Developer, Sofia, Bulgaria

Ruby Zgabay
IBM Power Chief Engineer, Austin, TX

Kaveh Naderi
IBM Power Chief Architect, Austin, TX

Paul Bastide
IBM Senior Software Engineer, Cambridge, MA

Tanya Koleva
IBM Technology Lifecycle Services, Sofia, Bulgaria

Manjunath Kumtagi
IBM Senior Software Engineer, OpenShift on Power, Bangalore, KA, India

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

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

George Gaylord
Program Director, Offering Management, USA

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

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

Yan Zhuo
Product Management Leader, IBM i, Chicago, IL

Pete Heyrman
PowerVM Hypervisor, Rochester, MN

Dirk Michel
STSM, Power Systems Enterprise Performance Architect, Austin, TX

John Folkerts
Power Systems IO Product Management, Rochester, MN

Bob Kovacs
Product Manager - Power Virtualization, Austin, TX

Hariganesh Muralidharan
Systems Management Architect (HMC/CMC/Novalink), Power Systems, Bangalore, IN

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

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

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>



Overview of Power11

IBM Power11 represents a significant leap in enterprise computing. Building on the strengths of Power10, Power11 introduces up to 25% more cores per chip, higher clock speeds, and enhanced energy efficiency. It continues IBM's focus on reliability, availability, and serviceability (RAS) while also integrating quantum-safe security features to future-proof critical workloads. The processor is designed to support demanding enterprise applications, particularly those involving AI, analytics, and hybrid cloud environments.

Power11 leverages Integrated Stacked Capacitor (ISC) technology and advanced 2.5D packaging to boost performance and efficiency. These innovations, combined with enhanced thermal management – such as more efficient fans and heatsinks – significantly improve system density, cooling effectiveness, and overall compute capability.

A standout feature of Power11 is its continued support for AI workloads through the Matrix-Math Assist (MMA) architecture and the integration of the IBM Spyre Accelerator, which is optimized for generative AI and complex model inference. This positions Power11 as a strong alternative to traditional GPU-heavy AI infrastructures. Additionally, Power11 strengthens its virtualization capabilities with deeper KVM integration, enhancing compatibility with Linux-native tools and hybrid cloud platforms. This makes it 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
- ▶ IBM Power E1180: Resilient, efficient, and scalable
- ▶ Power11: Trusted, autonomous, and modern
- ▶ Power11 processor architecture
- ▶ Operating system support
- ▶ Firmware and Hardware Management Console
- ▶ IBM Power E1180 rack integration and handling guidelines

1.1 IBM Power11: Advancing enterprise computing

For over 35 years, IBM Power servers have been a cornerstone of enterprise computing, delivering unmatched performance, reliability, and availability for mission-critical workloads across the globe. From the original RS/6000 systems powered by the IBM Power processor to the advanced Power10-based platforms, IBM Power has consistently evolved to meet the demands of modern IT infrastructure.

The Power architecture follows a consistent innovation cadence, introducing a new generation of processors approximately every three years. Each iteration brings significant enhancements – more cores and threads, improved energy efficiency, increased memory bandwidth, and expanded I/O capabilities – ensuring the platform remains at the forefront of enterprise computing.

The Power11 platform is built on three foundational principles:

- ▶ Trusted
 - Delivers continuous business operations with up to 99.9999% availability (on Power E1180), zero planned downtime for maintenance, quantum-safe cryptography, and accelerated system recovery.
- ▶ Autonomous
 - Boosts 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 – leveraging off-chip AI acceleration and seamless integration with platforms like 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: Power S1122 and Power S1124 (as well as the Linux varieties Power L1122 and Power L1124) optimized for departmental and edge workloads
- ▶ Midrange: Power E1150 – ideal for enterprise consolidation and scalable virtualization
- ▶ High-end: Power E1180 – 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 enhancements
- ▶ Performance gains of up to 15% per core and up to 25% per thread with combined hardware and software optimizations
- ▶ Significant throughput improvements in midrange and entry-level systems
- ▶ Up to 50% increase in memory bandwidth, leveraging DDR5 and advanced memory controllers

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

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

1.2 IBM Power roadmaps

IBM continues to advance its Power architecture to address the evolving demands of artificial intelligence, hybrid cloud, and mission-critical enterprise workloads. Central to IBM's hybrid cloud strategy, the Power platform is optimized for Red Hat OpenShift, enabling containerized workloads and cloud-native application development.

IBM Power Virtual Server extends these capabilities into the cloud, allowing clients to run Power workloads either on-premises or in IBM Cloud®, while benefiting from flexible consumption models and meeting data residency and compliance requirements. This hybrid approach ensures seamless workload portability and supports modernization initiatives.

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

Complementing its technology stack, IBM has built a strategic network of alliances that significantly expand the reach and capabilities of the Power platform. Notably, IBM Power Virtual Server is a certified platform for RISE with SAP, enabling high-performance, reliable SAP workload deployment in hybrid cloud environments. Red Hat technologies like OpenShift and Ansible are deeply integrated into IBM Power, supporting automation and DevOps practices. Additionally, IBM collaborates with major cloud providers – including Microsoft, AWS, and Google Cloud – to deliver hybrid and multicloud solutions, allowing Power workloads to extend seamlessly into public cloud infrastructures.

This section discusses IBM's roadmaps for evolution for the IBM Power-based infrastructure.

1.2.1 IBM Power Processor Roadmap

IBM's Power roadmap continues to evolve with the addition of the Power11 generation. Building on the innovations of Power10, Power11 introduces enhancements across the processor architecture, packaging, and energy efficiency. A key advancement is the integration of an additional silicon layer for improved energy management, allowing for higher performance with lower power consumption. This generation also boosts core strength and count, enabling it to handle even more demanding workloads. IBM is also refining its matrix math accelerators (MMAs), which were first introduced in Power10, to further accelerate AI inferencing directly on the chip – eliminating the need for external GPUs in many scenarios.

Looking beyond Power11, IBM's roadmap emphasizes full-stack innovation and adaptability to emerging technologies. The Power11 platform is designed to be memory-agnostic, supporting both DDR4 and DDR5 via the Open Memory Interface (OMI), and it introduces improved thermal infrastructure through advanced packaging and cooling technologies. These developments not only enhance performance but also reduce operational costs in data centers. Future iterations of the Power architecture are expected to continue this trend, with a focus 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 different generations of the Power processor family and shows the continued plans for improvements. With the release of Power11, the next generation of Power processors is already in development and additional processor designs are in the pipeline after that generation.

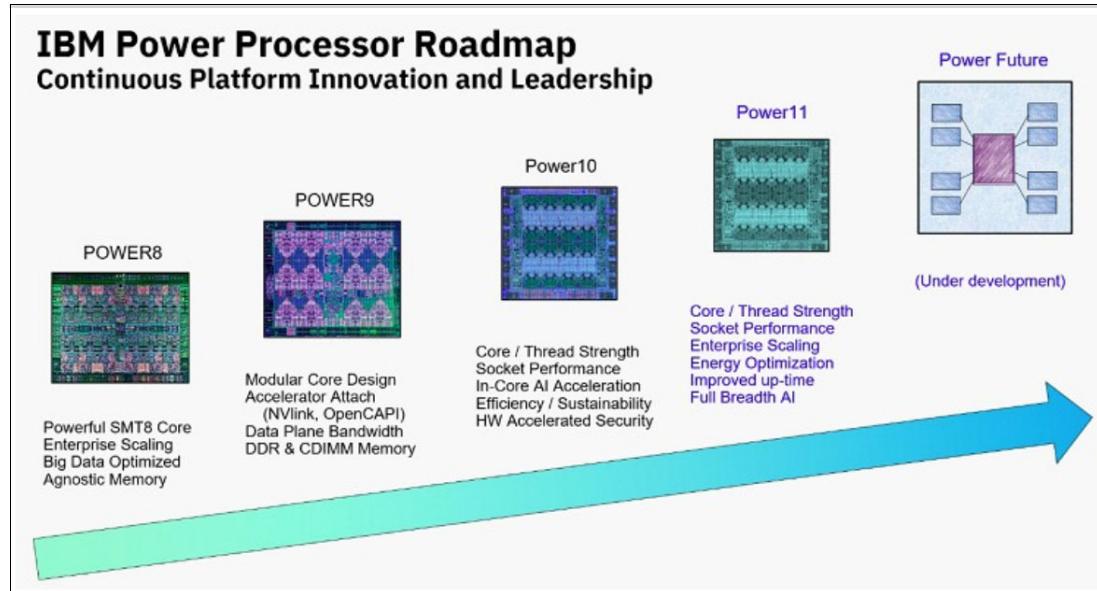


Figure 1-1 Power processor roadmap

1.2.2 IBM AIX Roadmap

IBM AIX (Advanced Interactive eXecutive) operating system is a proprietary UNIX-based operating system built on open standards and specifically designed for IBM Power servers.

For over 38 years, AIX has powered mission-critical applications and databases in industries with high-performance computing needs like finance, retail, healthcare, government, manufacturing, and insurance. Paired with IBM Power11 processor-based systems, the latest AIX 7.3 expands its reach into new markets and workloads to deliver unmatched performance, security, scalability, and reliability, all while supporting digital transformation through flexible subscription models tailored to business needs, and enabling adoption of technologies like hybrid cloud, AI, and cloud-native applications.

AIX's proven binary compatibility allows applications to run unchanged and without recompilation on the newest release, safeguarding IT investments in the platform.

IBM is strongly committed to delivering an AIX release roadmap of further innovations and purposeful support that extends beyond 2035 ensuring a stable, future-proof platform for enterprise workloads.

The current roadmap is shown in Figure 1-2.

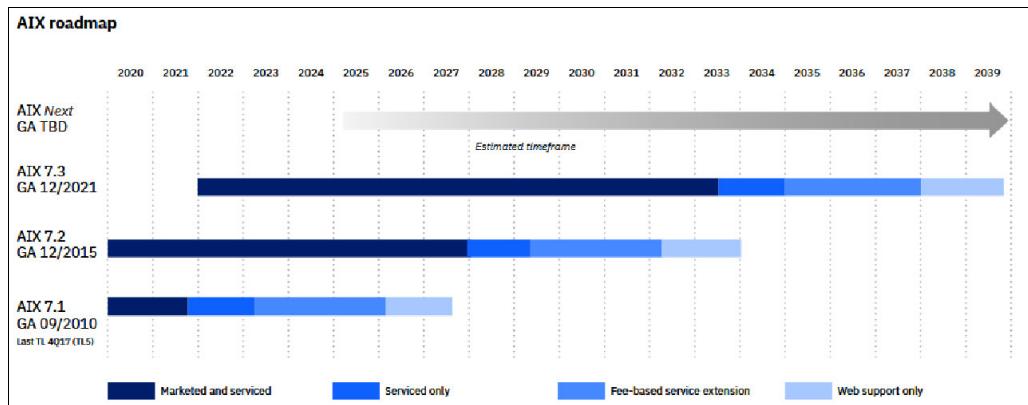


Figure 1-2 Support time line for future generations of AIX

For more information on IBM's strategy and roadmap for IBM AIX see [Strategy and Roadmap for the IBM AIX Operating System](#).

Additional information about AIX can be found in section 6.2, "AIX" on page 122.

1.2.3 IBM i Roadmap

Since 2008, IBM has consistently released new versions of IBM i approximately every three years, with IBM i 7.6 being the latest. Each release introduces new features, workload support, and language enhancements while maintaining backward compatibility – allowing customers to preserve and extend their existing investments. IBM i is recognized as one of the most secure, self-optimizing, and scalable operating systems in the world. IBM continues to deliver updates and enhancements throughout each release lifecycle via 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, demonstrated by its publicly available roadmap, which currently extends to 2035 and is updated annually – underscoring IBM's dedication to the platform's future. The roadmap is shown in Figure 1-3.

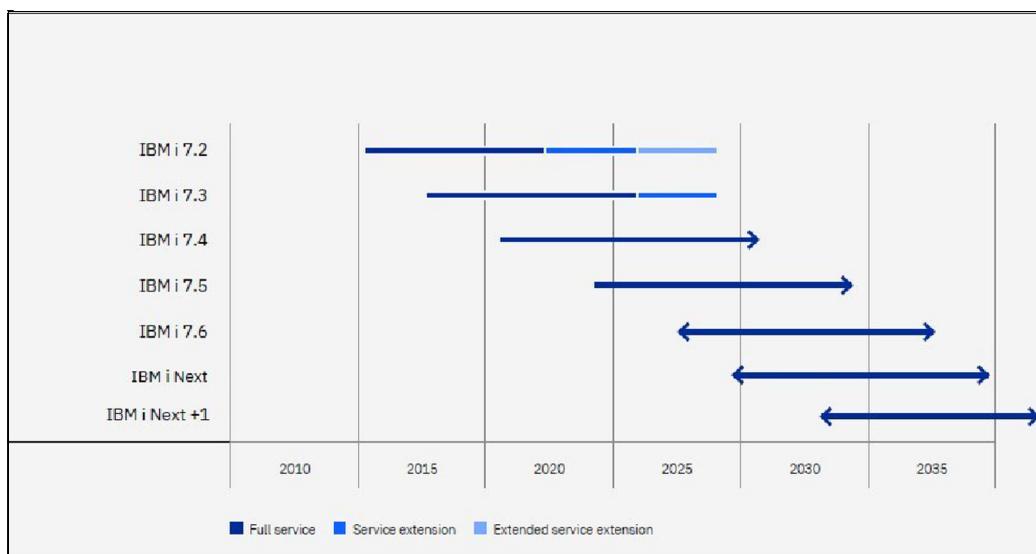


Figure 1-3 IBM i support matrix

1.2.4 PowerVM VIOS roadmap

PowerVM Virtual I/O Server (VIOS) is a critical component of the IBM Power virtualization stack, enabling 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 (EOS) 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 FP ^b	Next FP ^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) is the end of the maintenance period for a VIOS Release. Fix Packs and interim fixes will not be created for a VIOS Release after EoFS.
- b. Latest FP is the most recent FP available for the VIOS Release and links to that FP on Fix Central.
- c. Next FP is the target availability date for the next Fix Pack for the given VIOS Release. After the final Fix Pack for a VIOS Release is available, further fixes for the release are provided as interim fixes.

Note: The End of Service (EOS) date for a VIOS release is not indicated in this table. EOS dates can be found on the [IBM Software Lifecycle](#) page.

For additional details regarding the PowerVM Virtual I/O Server (VIOS) roadmap, please refer to 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 has 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 deeply rooted in their shared vision of delivering enterprise-grade solutions that are scalable, secure, and optimized for modern workloads.

At the Red Hat Summit 2025, Red Hat emphasized its dedication to supporting IBM Power through continued enhancements to Red Hat Enterprise Linux (RHEL) on IBM Power. This includes performance tuning for Power10 and Power11 processors, as well as deeper integration with Red Hat OpenShift for containerized and cloud-native applications. These efforts ensure that Power users can fully leverage Red Hat's open hybrid cloud technologies across both on-premises and cloud environments.

Red Hat also plays a pivotal role in enabling AI workloads on Power, with support for AI inference and acceleration technologies that align with IBM's hardware innovations like Matrix Math Assist (MMA) and the upcoming Spyre AI Accelerator. This synergy allows enterprises

to deploy AI models efficiently on Power infrastructure using Red Hat's open-source toolchains and platforms.

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

For more information on Red Hat Enterprise Linux support on IBM Power see “Red Hat Enterprise Linux” on page 135.

1.2.6 SUSE Enterprise Linux and IBM Power

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

SUSE and IBM have worked together for over two decades to provide robust, scalable, and secure Linux solutions on IBM Power infrastructure. Their joint efforts are centered around:

- ▶ SUSE Linux Enterprise Server (SLES) for Power
- ▶ SLES for SAP Applications, optimized for SAP HANA on 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 enables customers to modernize their IT environments while avoiding vendor lock-in.

For more information on SUSE Enterprise Linux see “SUSE Linux Enterprise Server” on page 137

1.3 IBM Power E1180: Resilient, efficient, and scalable

The IBM 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, operating systems, and cloud services – IBM Power delivers a synergistic platform built for autonomous IT. This holistic approach consistently drives tangible business outcomes across three foundational pillars: unprecedented business continuity, enhanced productivity and efficiency, and accelerated growth and scalability.

Unprecedented business continuity

The Power E1180 redefines operational resilience with features designed to eliminate downtime and help continuous service delivery:

- ▶ Planned Downtime Elimination: Leveraging technologies such as live updates, rolling upgrades, and autonomous patching, the Power E1180 enables maintenance without taking applications offline – targeting 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.

- ▶ Power Cyber Vault: Advanced threat detection can identify ransomware attacks in under a minute. Combined with automated recovery mechanisms, this feature significantly 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 reboots and live partition mobility, ensuring 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 is engineered to maximize IT efficiency and output:

- ▶ Autonomous IT Operations: With built-in automation and AI-driven management, the system reduces manual intervention and operational overhead.
- ▶ Integrated Monitoring and Management: Tools like HMC and PowerVC provide centralized control, enabling streamlined operations across hybrid environments.
- ▶ Resource Groups: Using innovative Resource Groups feature in Power11 enables even more effective usage of Power11 resources between different environments or tenants.

Accelerated growth and scalability

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

- ▶ IBM Power Virtual Servers (PowerVS): Immediate availability of Power11 in IBM Cloud enables seamless extension of workloads to the cloud, maintaining consistent performance, security, and OS compatibility.
- ▶ Hybrid Cloud Agility: Enterprises 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 Matrix Multiply Assist (MMA) – a hardware acceleration unit optimized for AI and machine learning workloads. By executing matrix operations directly on-chip, MMA reduces latency and boosts throughput, making the Power E1180 ideal for deep learning inference tasks.

Technical enhancements

Along with the benefits listed above, the Power11 based systems provide enhanced performance benefits from higher core frequencies, additional cores per server 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 HW and SW planned improvements	Exceed SLAs for response times and batch windows.
30-45% system performance increase with the new larger 30-Core Power11 DCMs	Assist in meeting 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 server portfolio

The new Power11 server portfolio consists of the following offerings, providing clients the choices designed to meet their specific business requirements.

High-End: IBM Power E1180

At the top of the portfolio is the IBM Power E1180, a high-end enterprise server built for the most demanding workloads. It delivers exceptional performance, scalability, and resiliency, making it ideal for large-scale 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 engineered for continuous availability and extreme workload consolidation.

Midrange: IBM 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 seeking a balance between performance, scalability, and operational efficiency, especially in hybrid cloud environments.

Entry-Level scale-out: IBM Power S1124/L1124 and Power S1122/L1122

For smaller enterprises or distributed environments, IBM offers the Power S1124 and Power S1122 servers for general usage and the Power L1122 and Power L1124 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 benefit from the same Power11 processor technology, ensuring consistent performance and security across the portfolio.

Figure 1-4 shows the new Power11 server portfolio.

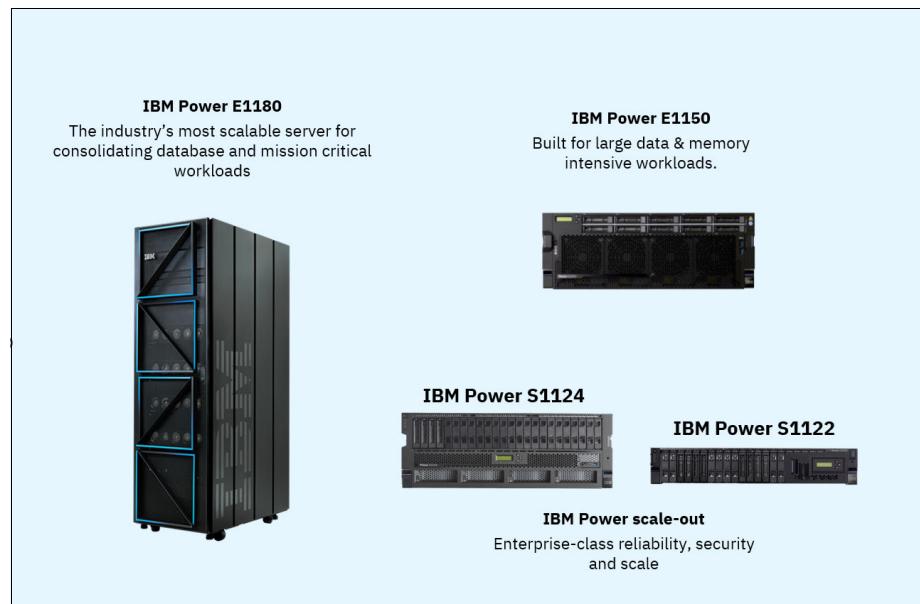


Figure 1-4 Power11 portfolio

Architectural technology Improvements

The Power11 portfolio introduces significant innovations across both the product lineup and the underlying core technologies, all aimed at delivering enhanced performance and greater value. One of the most notable advancements lies in the processor and memory architecture, where Power11 brings 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 assist in providing optimal cooling, supporting higher core densities and sustained performance.

Increased processor frequencies

Integrated Stacked Capacitor 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 significantly boosts memory throughput while simultaneously reducing latency. These enhancements provide quicker data access and processing, which is especially critical for memory-intensive workloads such as real-time analytics, large-scale databases, and AI applications. The combination of higher bandwidth and lower 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 Compare
High-End (4S+):	IBM Power E1180	<ul style="list-style-type: none">▶ 4 nodes (CEC) each 5U with 4 processor modules each▶ Up to 64 TB – DDR5 based memory	<ul style="list-style-type: none">▶ 15-core moving to 16-core▶ Same memory capacity▶ Introducing Resource Groups to maximize system utilization▶ In-place upgrades with serial number preservation
Mid-Range (4S4U)	IBM Power E1150	<ul style="list-style-type: none">▶ 2-4 processors in a 4U form factor▶ Up to 30 cores per module▶ Up to 16 TB – DDR5 based memory	<ul style="list-style-type: none">▶ Up to 25% more core capacity▶ Same memory capacity▶ Introducing Resource Groups to maximize system utilization▶ In-place upgrades with serial number preservation

Segment	Models	Footprint Capacity	Power10 Compare
Scale Out 2S4U	IBM Power S1124 IBM Power L1124	<ul style="list-style-type: none"> ▶ Up to 2 processors in a 4U form factor ▶ Up to 30 cores per module ▶ Up to 8 TB – DDR5 based memory 	<ul style="list-style-type: none"> ▶ Up to 25% more core capacity and same memory capacity ▶ Introducing Resource Groups to maximize system utilization
Scale Out 2S2U	IBM Power S1122 IBM Power L1122	<ul style="list-style-type: none"> ▶ Integrating the eSCM modules into the S1122 family - four or 10 processors per eSCM ▶ Up to 30 cores per dual core module ▶ 2 processors in a 2U form factor ▶ Up to 4 TB – DDR5 based memory 	<ul style="list-style-type: none"> ▶ Up to 50% more core capacity ▶ Same memory ▶ Introducing Resource Groups to maximize system utilization

Investment protection

IBM Power11 systems are designed with strong investment protection in mind, ensuring that customers can evolve their infrastructure without sacrificing prior investments. A key example of this 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 allows organizations to retain their existing system identity, simplifying software licensing, asset tracking, and operational continuity.

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

Additionally, IBM supports Power Enterprise Pools (PEP) that include both Power E1080 and Power E1180 systems. This mixed-system support enables organizations to gradually migrate workloads from older to newer systems within the same resource pool. It allows for dynamic sharing of processor and memory activations across both generations, providing flexibility in capacity planning and workload placement. This approach not only protects existing investments but also enables a smooth, non-disruptive path to modernization.

Portfolio simplification

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

Figure 1-5 shows how customers can migrate from their existing IBM Power9 or Power10 servers to the new Power11 portfolio.

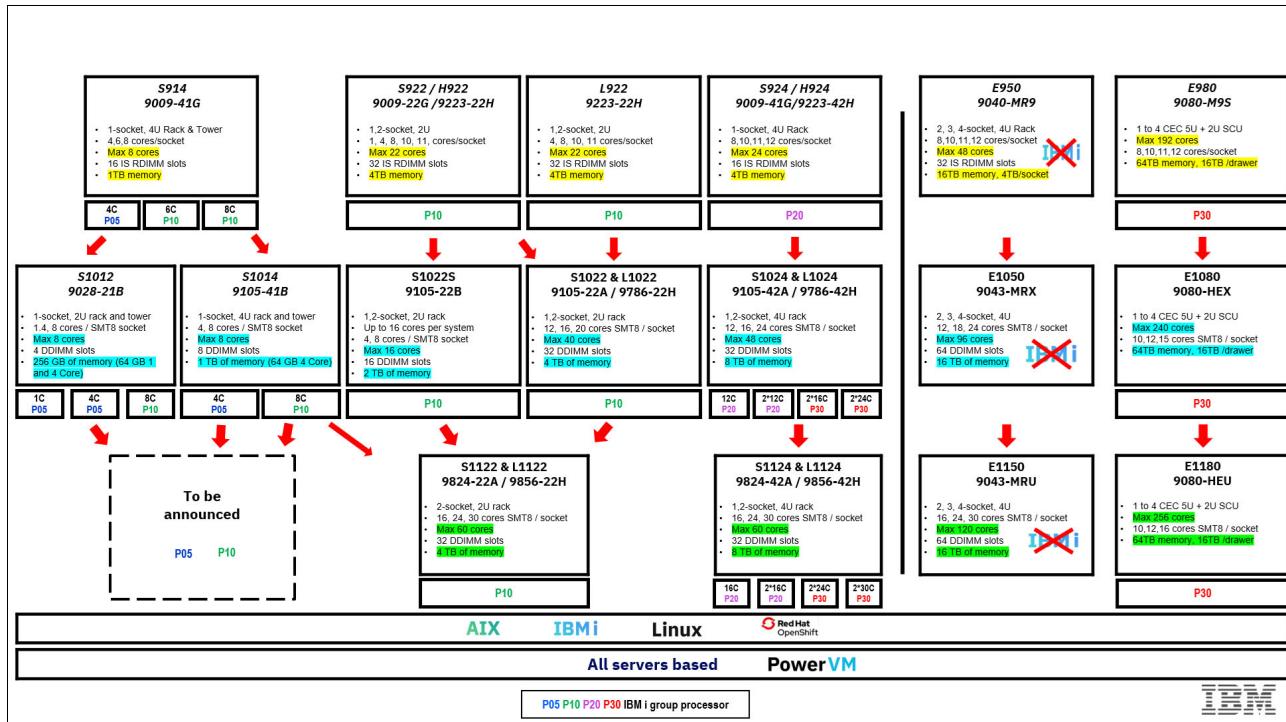


Figure 1-5 Migration path Power9 to Power11

1.4 Power11: Trusted, autonomous, and modern

Our vision for Power is autonomous IT, where Power11 processor-based servers can think, heal and scale on their own. This in turn supports business agility so that enterprises can respond more quickly to market changes and customer demands, rapidly deploy new services, and reduce the time to market for new products.

We are working towards 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 servers and includes capabilities such as Infrastructure as Code with Terraform, Application Configuration Management with Red Hat Ansible and enhancements to existing platform tools such as the Hardware Management Console and Cloud Management Console. On the AI side we are building AI infused workflows with IBM Watson.x and new tools like IBM Concert®.

This section provides insights into some of 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 built using IBM Power processor-based servers benefits from the robust security technologies integrated into both the hardware and software stacks. Power processor-based servers offer advanced security features at every level of the system, ensuring comprehensive

protection for sensitive data and applications. These features include advanced encryption technologies, secure boot capabilities, and integrated firmware updates. Additionally, Power processor-based servers leverage IBM's extensive expertise in securing mission-critical workloads, making them a popular choice for organizations seeking a secure environment for their digital assets.

Workloads on Power11 processor-based servers see significant benefits from improved cryptographic accelerator performance compared to previous generations. Specifically, the Power11 chip supports accelerated cryptographic algorithms such as AES, SHA2, and SHA3, resulting in considerably higher per-core performance for these algorithms. This enhancement allows features like AIX Logical Volume Encryption to operate with minimal impact on system performance.

The processor-core technology of Power11 incorporates integrated security protections aimed at:

- ▶ Improved Cryptographic Performance

Integrated cryptographic support reduces the performance impact of encrypting and decrypting your data, allowing you to make encryption pervasive to protect all your critical data.

- ▶ Increased Application Security:

Hardened defenses against return-oriented programming (ROP) attacks.

- ▶ Simplified Hybrid Cloud Security

Easy-to-use, setup-free hybrid cloud security administration with a single interface.

- ▶ Enhanced Virtual Machine Isolation

Providing the industry's most secure virtual machine isolation technology, this technology defends against attacks that exploit operating system or application vulnerabilities in the virtual machine to access other virtual machines or the host system.

Encryption technologies and their applications

Power11 emphasizes comprehensive security throughout its design, offering multiple encryption options. Key among these is Transparent Memory Encryption (TME), Fully Homomorphic Encryption (FHE), and Quantum-Safe Encryption (QSE).

Transparent Memory Encryption

Transparent Memory Encryption (TME) encrypts data in memory to protect it from unauthorized access and tampering during runtime. Operating at the hardware level, TME utilizes the Power11 processor's cryptographic engines to perform encryption and decryption tasks efficiently. TME provides pervasive protection of data in memory with minimal impact on system performance, as encryption and decryption are managed at the chip level. Its integration into normal operations is seamless and automatic.

Fully Homomorphic Encryption

Fully Homomorphic Encryption (FHE) allows computations to be performed directly on encrypted data without decrypting it first. This helps that sensitive data remain confidential even during processing. FHE operates at the software level and involves sophisticated mathematical algorithms to enable computations on ciphertexts.

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

Quantum-Safe Encryption

Quantum-Safe Encryption (QSE) is designed to be resistant to quantum attacks, securing data against the computational capabilities of future quantum computers that could potentially break current cryptographic algorithms. QSE employs cryptographic algorithms 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 show they provide robust security in the face of future quantum advancements. QSE is typically used for securing long-term data, sensitive communications, and critical infrastructure.

The relevant features and differences in these technologies is shown in Table 1-4

Table 1-4 Key Differences

Feature	TME	FHE	QSE
Encryption Scope	Secures data in memory	Allows computations on encrypted data	Prevents against future quantum computing threats
Implementation Level	Implemented in hardware	Implemented using a combination of hardware and software	Implemented using a combination of hardware and software
Performance Impact	Hardware accelerated through the Power11 cryptographic engines and designed to have minimal performance impact	Involves substantial computational overhead	The impact of QSE varies; some quantum-safe algorithms may introduce performance overhead; this 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

Quantum-Safe Encryption, also known as Post-Quantum Cryptography (PQC), refers to encryption methods that are secure against both classical and quantum computers. As quantum computers advance, they may pose a threat to existing cryptographic systems, potentially compromising their security. Quantum-Safe Encryption (QSE) is essential for protecting sensitive data, communication channels, and user identities in the age of quantum computing.

The urgency of adopting QSE stems from two primary concerns.

- Advanced quantum computers could allow adversaries to intercept and decrypt protected digital communications through Harvest Now, Decrypt Later (HNDL) strategies, even before reaching Q-Day. Q-Day is the anticipated point in time when quantum supremacy becomes widespread and many of the current encryption algorithms are no longer effective.
- Transitioning to QSE might require over a decade due to the complexities of organizational structures and IT infrastructure.

Consequently, organizations should start evaluating and implementing QSE solutions immediately to provide continued protection and maintain trust among their stakeholders.

Delaying QSE adoption could have severe consequences. Legacy cryptographic systems left unaltered could be compromised in the event of a successful quantum attack, exposing sensitive data and risking confidential business transactions and individual privacy. Financial institutions, critical infrastructure providers, and government agencies would face significant challenges in maintaining operational integrity and confidentiality. Therefore, prioritizing QSE implementation is crucial for long-term cybersecurity resilience.

Power11 is designed to support these quantum-safe algorithms, ensuring robust security even as quantum computing advances. Power11's support of Quantum-Safe features provide the following.

- ▶ Data Encryption Breakage Protection
 - **Risk:** Quantum computers could break widely used cryptographic algorithms such as RSA, ECC (Elliptic Curve Cryptography), and traditional Diffie-Hellman key exchange protocols. Shor's algorithm, for instance, could efficiently factor large integers and solve discrete logarithms, compromising the security of these algorithms.
 - **Protection:** Power11 is capable of supporting quantum-safe algorithms like lattice-based, hash-based, code-based, and multivariate quadratic equations-based cryptography, which are believed to be resistant to quantum attacks. The crypto engines in Power11 enhance the performance of these algorithms, ensuring secure encryption and key exchange processes with minimal performance degradation.
- ▶ Secure Communications
 - **Risk:** Quantum computers could intercept and decrypt secure communications, undermining protocols that currently rely on classical encryption methods.
 - **Protection:** Power11 secures communication channels with quantum-resistant protocols, ensuring data confidentiality even in the presence of quantum adversaries. End-to-end encryption is maintained throughout the data lifecycle, from storage to transmission, using algorithms resistant to quantum attacks.
- ▶ Data Integrity and Authenticity
 - **Risk:** Quantum computers could forge digital signatures or tamper with data.
 - **Protection:** Power11 is capable of using quantum-safe digital signature algorithms such as XMSS (eXtended Merkle Signature Scheme) and UOV (Unbalanced Oil and Vinegar), providing strong security against quantum attacks.
- ▶ Long-Term Data Protection
 - **Risk:** Sensitive data stored today could be harvested and decrypted in the future as quantum computers become more powerful, threatening long-term confidentiality.
 - **Protection:** Implementing quantum-safe encryption methods protects data, keeping it secure over time, even as quantum computing capabilities evolve. Power11's architecture supports updates to cryptographic libraries and protocols, enabling the 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 boot attacks, could expose sensitive data if not adequately protected.
 - **Protection:** Power11's Transparent Memory Encryption (TME) ensures that data in memory is encrypted, protecting it from physical attacks during runtime.

Quantum-Safe algorithms supported by Power11:

The following quantum-safe algorithms are supported by Power11:

- ▶ Lattice-Based Cryptography
 - **Algorithms:** ML-DSA (Dilithium), ML-KEM (Kyber) and NTRUEncrypt.
 - **Characteristics:** Secure against quantum attacks. Based on lattice problems – 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 (eXtended Merkle Signature Scheme), and SPHINCS+.
 - **Characteristics:** Secure based on hash functions, though generally produces larger signatures and keys.
- ▶ Code-Based Cryptography
 - **Algorithms:** McEliece Cryptosystem, BIKE (Bit Flipping Key Encapsulation), and HQC (Hamming Quasi-Cyclic).
 - **Characteristics:** Quantum-resistant based on decoding random linear codes, though public keys can be large.
- ▶ Multivariate Quadratic Equations
 - **Algorithms:** Unbalanced Oil and Vinegar (UOV), Rainbow.
 - **Characteristics:** Secure against solving systems of multivariate quadratic equations, generally efficient in signature generation but may involve larger key sizes.

Power11 implementation

Power11 processors support these quantum-safe algorithms through:

- ▶ Crypto Engines: Multiple engines per core enable efficient execution of cryptographic operations.
- ▶ Software Updates: The architecture allows updates to cryptographic libraries, ensuring the integration of new quantum-safe algorithms as they become standardized.

Power11's design and capabilities provide robust security against future quantum threats by leveraging hardware acceleration and 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 server. The first option is to use the built in encryption acceleration built into the Power11 chip. In addition, IBM Power supports a PCIe based encryption accelerator.

On-chip encryption support in Power11

The Power11 processor is designed to effectively support future encryption – including Fully Homomorphic Encryption (FHE) and Quantum-Safe Cryptography – in order to be ready for the Quantum age. The Power11 processor-chip instruction set architecture (ISA) is tailored for these solutions' software libraries, which are currently available or will soon be made available in the corresponding open source communities.

Workloads on Power11 benefit from cryptographic algorithm acceleration, enabling much higher per-core performance than Power10 processor-based servers for algorithms like Advanced Encryption Standard (AES), SHA2, and SHA3. Features like AIX Logical Volume Encryption can be activated with minimal performance overhead thanks to this performance enhancement.

With four times as many AES encryption engines, Power11 processor technology is designed to offer noticeably faster encryption performance. Power11 processor-based servers are more advanced than Power10 processor-based servers, with updates for the most stringent standards of today as well as future cryptographic standards including post-quantum and fully homomorphic encryption. It also introduces additional improvements to container security. Through the use of hardware features for a seamless user experience, transparent memory encryption aims to simplify encryption and support end-to-end security without compromising performance.

1.4.2 Protect, detect, and recover with 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 – they are active requirements. More regulations are coming, each demanding higher levels of risk management and governance, mechanisms to promptly detect anomalous activities, and resilience practices to minimize the risk of data corruption or loss. The bar is rising, and the spotlight is now on every organization's security posture.

These frameworks go beyond checklists. They are signaling a shift: cybersecurity is no longer a compliance add-on, but a foundational 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 – collections of tools, processes, and teams that operate in silos. While some integration has occurred over time, loosely connecting security domains still leaves critical gaps in visibility, context, and response coordination.

A siloed approach results in:

- Limited understanding of threats across domains (e.g., linking identity misuse with network anomalies).
- Reactive security postures measured by containment speed – not business outcomes.
- Barriers to collaboration and innovation, as teams struggle to securely share and access data.
- Missed opportunities to embed privacy and security into customer-facing products and services.

Even more concerning, lack of contextual awareness across tools and data streams can open the door to full-blown compromises – like ransomware attacks that paralyze operations.

Security was never meant to be siloed. It became that way out of necessity and legacy architectures. But that model is no longer sufficient. Forward-looking enterprises are rethinking their cybersecurity strategy not just as a defense mechanism, but as a business enabler. They're integrating tools, processes, and insights across domains-from access and identity to threat detection, data governance, and compliance. They're moving from isolated defenses to connected intelligence.

Zero Trust: A strategic shift, not just a technology play

Many organizations are now adopting Zero Trust frameworks as the foundation for their next-generation security programs. Built on the principles of least privilege access, continuous verification, and assumed breach, Zero Trust offers a compelling blueprint.

But the blueprint alone is not enough. Applying Zero Trust effectively requires:

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

In essence, Zero Trust is not a product. It's an operating model for security that aligns with the way modern businesses function-dynamically, digitally, and distributed.

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

Cyber attacks today are fast-moving, complex, and deeply disruptive. Here's the typical incident flow:

1. Service disruption is noticed.
2. Immediate shutdowns are initiated to contain spread.
3. Search for uncompromised backups begins.
4. Slow and costly recovery efforts start.

According to recent reports, 78% of organizations take over 100 days to fully recover from a major cyber incident. That is not just downtime – it is lost trust, revenue, and opportunity.

IBM provides a unified cyber resilient solution which includes IBM Power11 processor-based servers, IBM FlashSystems storage, advanced software, and services from IBM Technology Expert Labs services to respond to evolving cyber threats and regulatory standards with IBM Power Cyber Vault.

IBM Power Cyber Vault

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

- ▶ Power Cyber Vault Protected Data
 - Automated immutable images stored in the Power Cyber Vault to protect clients' snapshots and backups
- ▶ Continuous Threat Detection
 - Near real-time filesystem monitoring, FlashSystem inline detection, and recovery data scanning. Threat Detection Time as low as 60 Seconds.
- ▶ Rapid Clean Room Response
 - Automated response to attacks, creating Power Cyber Vault clean rooms for testing and validation. Accelerated Recovery Response.
- ▶ Cyber Vault Assisted Recovery
 - Automated multi-level testing and validation of Power Cyber Vault images to accelerate system recovery. Return to full operations in hours versus months.

Power Cyber Vault solution components

IBM Power Cyber Vault is a complete Cyber Resilient® Solution deployed in as little as 30 days. It is composed of various elements:

- ▶ IBM Power11 processor based servers

IBM Power11 processor based servers provide:

- Industry leading threat resistance
- PowerSC for security and compliance tracking

- Secure boot and memory encryption built in
- Scalable and flexible architecture from entry to enterprise servers

► IBM FlashSystems Storage

IBM FlashSystems provide:

- Immutable Copies for Data Protection
- Ransomware Detection for real-time threat detection
- Flash-speed recovery for rapid restore of snapshot data
- Storage Insights Pro for comprehensive reporting

► IBM Technology Expert Labs Power Cyber Vault Deployment

IBM Technology Expert Labs provides:

- Design Workshop for client specific requirements
- Power Cyber Vault Deployment to customize and enable the unified Power Cyber Vault to meet client needs

► IBM Expertise Connect

IBM Expertise Connect is:

- Primary Technical Point of Contact for Power Cyber Vault
- Dedicated to client and client success
- Available after install to help guide client's directions

Power Cyber Vault implementation

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

1. Identify - Assess Risks and Design Mitigation Solutions

a. IBM Security® and Resilience Assessment

This assessment provides:

- A view of the organization, processes, and strategies regarding operational and cyber resilience
- A list of prioritized recommendations to improve overall resilience

b. IBM Cyber Vault Solution and Design Workshop

This workshop provides:

- Create detailed solution based on customer cyber resiliency goals, infrastructure, and required services

2. Protect - Safeguarded Cyber Vault Copies and Backups

Client specific scheduled creation and protection of immutable snapshots and backups to meet RPO and RTO requirements:

- Snapshots: FlashSystems and Storage Defender CSM save protected system snapshot state into the Cyber Vault on a regularly scheduled basis.
- Orchestration for crash consistency
- Application specific customization available

3. Detect - Real-Time Threat Detection

Monitoring real-time threats and cyber attack vectors to respond quickly as threats emerge via Power SC Real Time Compliance orchestration:

- Zero Trust Execution to detect, prevent, and report unapproved executables and applications from running on Power11 based-processor servers
- FlashSystem's integrated Ransomware Threat Detection - alerts on active attacks and reports to PowerSC via Storage Insights Pro

- SEIM Integration to accept alerts and threat reports from SEIM systems
4. Respond and Recover - Real Time Threat Response
- Routinely, or when threats are detected, a response phase is triggered to audit and recover the system:
- Upon receiving an integrity alert, IBM Power Cyber Vault automatically initiates the validation process by creating an immutable backup of the VM within seconds and spinning up a copy of that VM in the clean room within minutes.
 - Various types of Integrity Checks are run in the clean room (including platform specific tests)
 - Power Cyber Vault reports corrupted copies, and the newest clean copies are found
 - When approved by administrator, recovery to production will commence to recover and help meet RPO and RTO requirements

1.4.3 Quantum-safe compliance with one-click inventory discovery

IBM PowerSC is a comprehensive security and compliance solution designed specifically for virtualized environments running on IBM Power Servers with AIX, Linux, and IBM i. It integrates deeply with the 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 like PCI DSS, HIPAA, and SOX by automating the monitoring, auditing, and enforcement of security policies. It also includes advanced capabilities like multi-factor authentication, intrusion detection, and patch management. With tools like PowerSC Trusted Surveyor, it provides consistent network configuration and compliance across dynamic virtual environments, helping reduce administrative overhead while maintaining strong security postures

IBM is actively involved in the development and adoption of quantum-safe cryptographic standards, including working with NIST on their Post-Quantum Cryptography Standardization project. PowerSC supports a seamless transition to NIST's forthcoming standards.

IBM PowerSC offers capabilities to help verify and manage quantum-safe encryption.

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

- ▶ PowerSC scans endpoints on your server to identify cryptographic artifacts and vulnerabilities related to quantum safety. These can also be scheduled.
- ▶ It creates an inventory of where different encryption algorithms are implemented on your system.
- ▶ The analysis estimates the strength of ciphers, certificates, and keys on your AIX endpoints and categorizes them as weak, strong, quantum safe, or unclassified. Note that PowerSC is a security and compliance management tool, not a cryptographic module validation tool.
- ▶ An Quantum Safety Analysis report in the PowerSC GUI lists the discovered cryptographic elements and their assessed strengths.

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

Planned downtime remains one of the biggest challenges in IT operations – especially for critical infrastructure such as Power11. IBM's approach is to deliver *Autonomous IT* by embedding extreme automation and AI-infused workflows directly into the platform. The result: a significant reduction in operational effort, minimized disruptions, and improved resilience.

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

- ▶ Operational Complexity: Maintaining firmware, I/O adapters, and Virtual I/O Servers (VIOS) is highly specialized and intricate.
- ▶ Service Interruptions: Routine maintenance often requires downtime, impacting application availability and business continuity.
- ▶ Security Risks: Delayed updates expose systems to vulnerabilities (CVEs), weakening compliance and increasing risk.
- ▶ Resource Drain: IT teams spend 20–40% of their time on maintenance tasks like patch 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.

These issues force organizations into a difficult trade-off: stay current and risk downtime, or delay updates and increase exposure – an unsustainable model in today's always-on digital economy.

What we mean by “infrastructure”

In this context, infrastructure refers to the Power11 platform, encompassing firmware, I/O components, VIOS, and the operating system. To maintain performance and security, clients must:

- ▶ Apply patches and updates regularly
- ▶ Remain on supported versions
- ▶ Respond swiftly to security advisories
- ▶ Conduct preventive maintenance

However, these actions typically require planned downtime. The core dilemma: should businesses 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 – enabling continuous business operations while keeping infrastructure secure, up-to-date, and high-performing. IBM Power11 introduces a transformative approach to infrastructure lifecycle management, purpose-built for Autonomous IT. The platform is designed to operate with minimal human intervention, leveraging embedded intelligence to monitor, diagnose, and remediate issues in real time. This enables infrastructure that is inherently resilient, adaptive, and capable of maintaining continuous service availability.

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

- ▶ Unified update orchestration for system firmware, VIOS, and I/O adapters.
- ▶ Live Partition Mobility (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 filesystems.

Maintenance operations can be executed autonomously or manually via the Hardware Management Console (HMC), offering both automation and administrative control. This architecture transforms maintenance from a disruptive, manual process into a seamless, background operation.

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

IBM Concert addresses the persistent challenge of siloed application data by harnessing the power of AI to deliver intelligent, prioritized recommendations. By streamlining issue identification and resolution, Concert significantly reduces Mean Time to Resolution (MTTR) for critical risk factors such as:

- ▶ Critical Vulnerability Exploits (CVEs)
- ▶ Certificate management issues
- ▶ Application compliance challenges

What Sets IBM Concert apart

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

- ▶ Data-Agnostic Integration

Designed to meet you where you are, Concert supports a wide range of data sources across networks, infrastructure, and application architectures.

- ▶ Hybrid Cloud by Design

Tailored for the realities of modern enterprise IT, Concert seamlessly supports hybrid cloud environments.

- ▶ AI-Infused Orchestration for Proactive Maintenance

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

- ▶ 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 triggering of automated actions to remediate issues, reducing manual effort and downtime.

Automation can be triggered directly from IBM Concert solution or from the HMC, ensuring flexibility and ease of use.

Key benefits of IBM power11's autonomous maintenance

IBM Power11's integrated solution for Zero Planned Downtime (ZPD) delivers measurable benefits across operational, security, and financial dimensions:

- ▶ Business Continuity: Updates are applied without interrupting workloads, preserving uptime and service-level commitments.
- ▶ Operational Efficiency: Automation streamlines the maintenance lifecycle, reducing manual effort and administrative overhead.
- ▶ 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 significantly lowers 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 – aligned 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:

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

This approach minimizes disruptions, assuring SLA compliance, and allows faster and more timely maintenance. Systems stay secure, stable, and aligned with evolving requirements – without downtime or performance impact.

Power11's built-in Automated Maintenance Tool enables platform maintenance without application impact. This automation can be triggered via IBM Concert's orchestration, or it can be human-triggered from the Hardware Management Console (HMC), completely independent of IBM Concert.

IBM Concert includes:

- ▶ Ability to update System Firmware, VIOS and IO adapter microcode from a single update flow. This is supported for both concurrent and disruptive updates
- ▶ Validation for LPM and VIOS redundancy (VIOS maintenance readiness check)
- ▶ Ability to automatically migrate partitions and return as part of the update process
 - Option to choose to return to the source or leave on the target system
 - Option to evacuate all LPARs or choose a subset of LPARs and order of LPAR
 - Option to choose order of updates
- ▶ Option to either download only or download and update

Different sources of updates

- IBM Website (preferred and recommended for end-to-end automation)
- SFTP Server
- HMC filesystem

- NFS Server
- USB

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

Figure 1-6 Power server licensed capabilities

For more information on using the HMC automated maintenance tool see section 5.1.7, “Using the Automated Maintenance tool” on page 108’

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 significant 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—over 90% of mid-size and large enterprises report that a single hour of downtime now exceeds \$300,000 in cost.

To address these challenges, IBM has introduced enhanced support capabilities in the Power11 Hardware Management Console (HMC). This includes a streamlined case creation process and automated log collection via Call Home. Administrators can now initiate support cases directly from the HMC, with relevant FFDC (First Failure Data Capture) logs automatically gathered and transmitted to IBM Support.

Supported log types are:

- System FW (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, a problem will be created in the HMC, but a case will not be created. Logs will be stored in the HMC, but not uploaded automatically.

Note: IBM recommends enabling call home which is also needed for autonomous error resolution data collection.

If the HMC is a virtual HMC, and the problem is an HMC problem, logs will be collected when an issue occurs. However, a call home will not occur and they will not be automatically sent.

Use Case: automated support workflow

Consider Louis, a system administrator at a large insurance company running critical backend applications on IBM Power infrastructure. During a routine maintenance task involving live partition mobility, Louis encounters a migration validation error. Previously, resolving such issues required manually opening a support case, coordinating with IBM Support, and collecting FFDC data—a process that could take 5 to 8 hours before investigation could begin.

With Power11, Louis can now easily and quickly 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 leverages AI-enhanced diagnostic tools trained on Power processor-based servers to accelerate root cause analysis and resolution.

Benefits of the enhanced support model

Implementing this new automated data collection capability in conjunction with applied AI workflows in IBM Support results in:

- ▶ 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.
- ▶ Lower 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 Increase efficiency and reduce 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 five years, there has been an 800% increase in climate commitments among companies, indicating broader awareness and the realization that sustainable practices can enhance profitability. Currently, 83% of organizations are investing in R&D for low-carbon products and services, and products with sustainability attributes are achieving revenue growth rates over 25% higher than traditional ones.

Although 95% of companies have established operational sustainability goals, only 41% have achieved measurable progress. Many companies are facing challenges in starting the implementation process. This indicates a need for more explicit guidance, enhanced governance frameworks, and improved integration of sustainability metrics into core business operations.

Power processor-based servers have consistently improved energy efficiency per watt with each generation, and Power11 continues this trend. Power11 enhances efficiency, cutting energy use, carbon emissions, and the datacenter's energy footprint.

Power9® clients upgrading to Power11 can reduce energy consumption by up to 60% for the same performance. Power11 outperforms x86 systems, providing twice the performance per watt. Compared to Power10, Power11 offers up to 33% better performance per watt, with the S1122 showing the largest gain.

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 datacenter footprint by two-thirds.

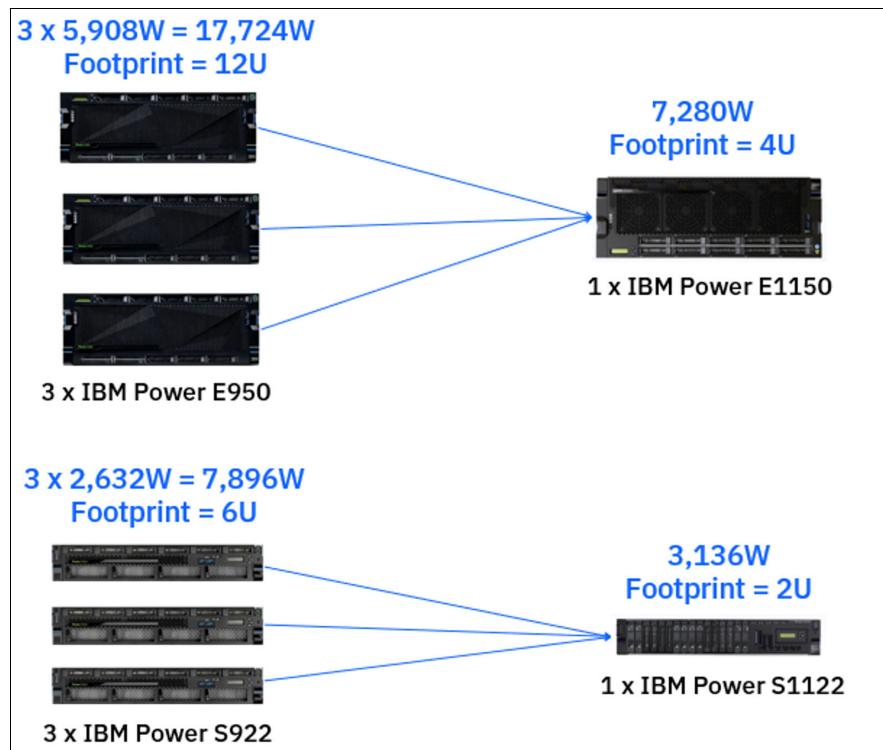


Figure 1-7 Power9 to Power11 energy consumption reduction

Along with the base energy savings delivered with Power11, there are some exciting new features that are introduced to address some of the other feedback received from clients.

Partition level monitoring

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

This is demonstrated in Figure 1-8.

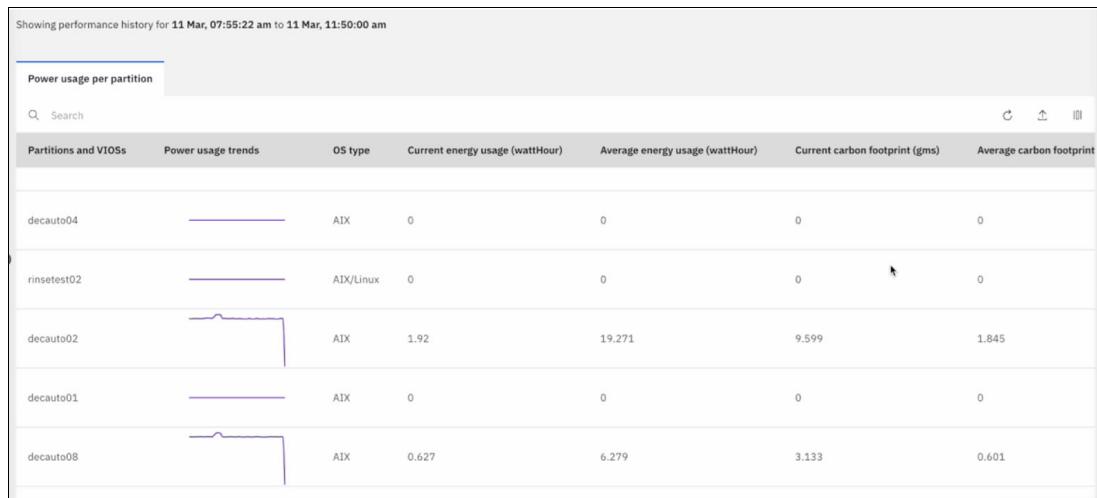


Figure 1-8 Partition level monitoring

Energy Efficient mode

Power11 has introduced a new Energy Efficient mode that, when enabled, can reduce energy consumption by up to 30%, with only an approximate 10% reduction in performance. This impact will vary depending on the system and configuration. The power mode can be dynamically established through the HMC GUI, as illustrated in Figure 1-9.

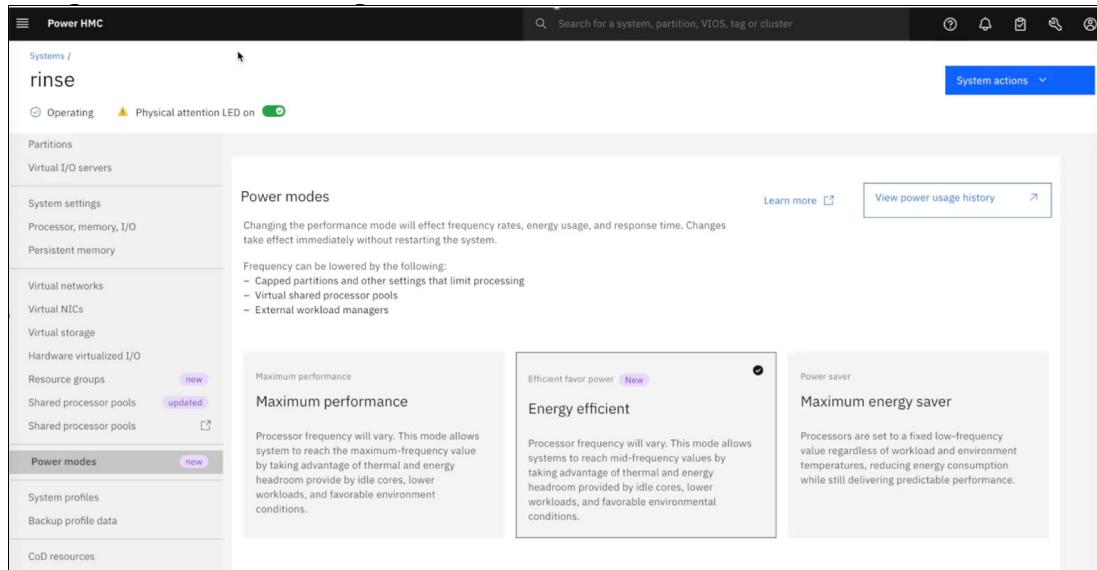


Figure 1-9 Power modes

In the environmental dashboard, it is observed that in maximum performance mode, energy consumption remains almost constant even when core usage decreases.

Figure 1-10 shows the energy usage when the processor is set to maximum performance.

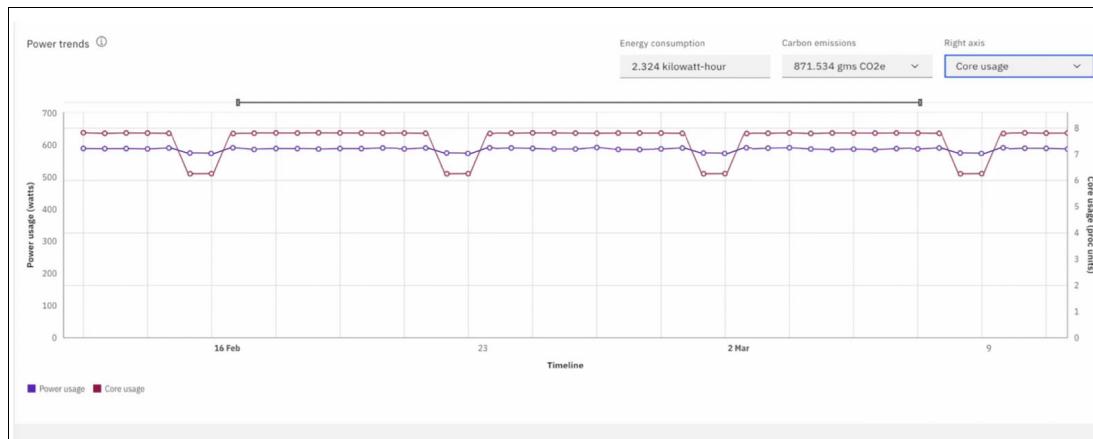


Figure 1-10 Energy consumption in max performance mode

Switching to energy efficiency mode, which is implemented instantaneously, results in only a 10% reduction in performance, while significantly decreasing energy consumption by up to 30%, as illustrated in Figure 1-11.

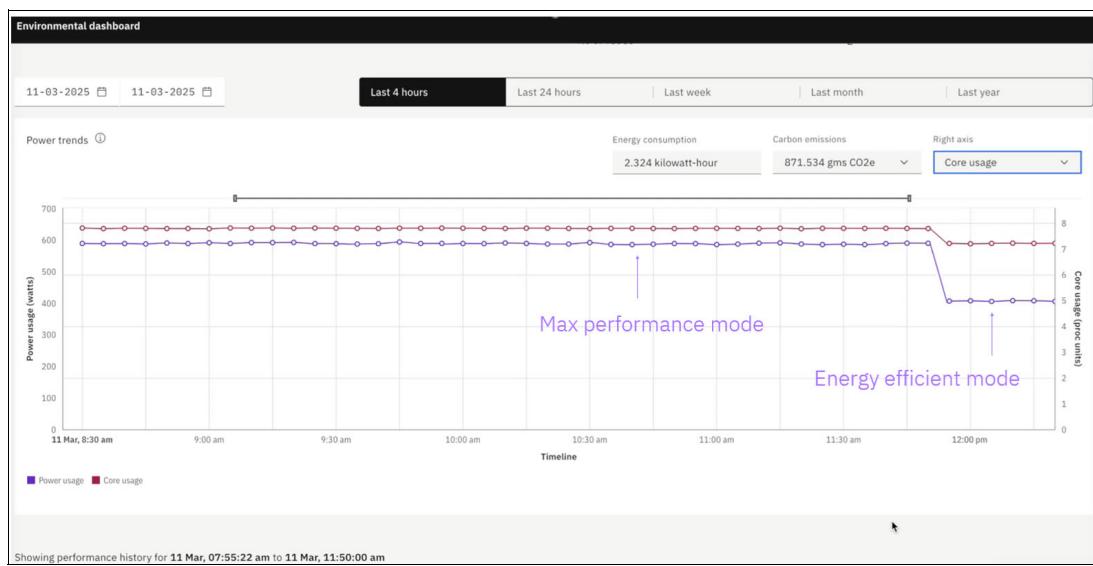


Figure 1-11 Energy efficient mode

Power mode scheduling

If there is a pattern in system use, the new scheduling power mode feature can be utilized. This enables the system to activate Energy Efficient mode during designated times. As illustrated in Figure 1-12 on page 29, the Energy Efficient mode will be engaged over the weekend, starting from 8:00 PM on Friday until 5:00 AM on Monday.

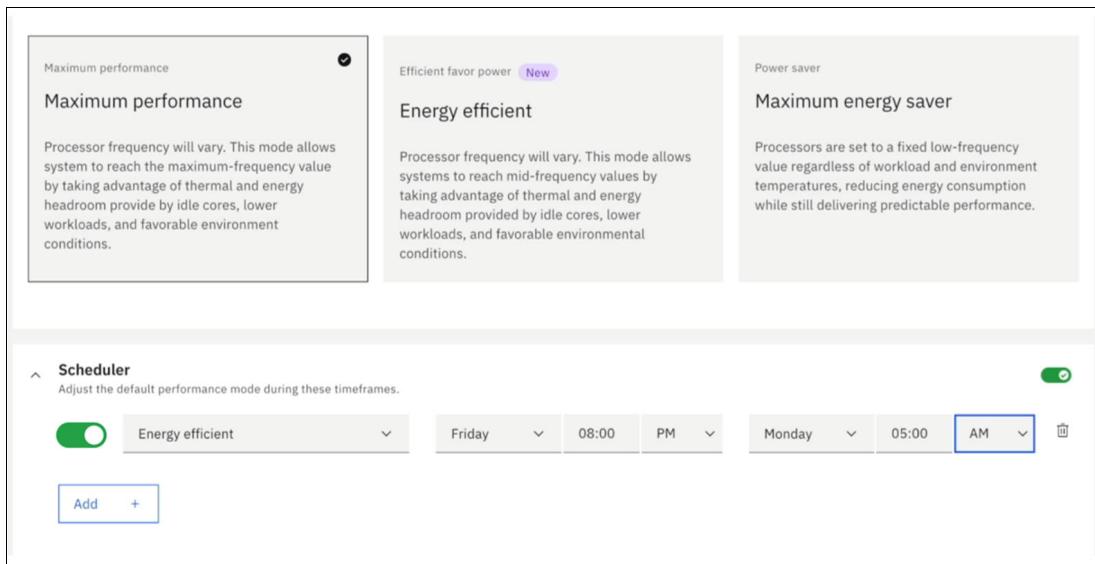


Figure 1-12 Power mode scheduling

The new environmental dashboard enables energy savings whenever potential opportunities arise. The automated scheduling capabilities allow the system to transition into and out of various modes efficiently and appropriately.

1.4.7 Superior performance of up to 25% improvement for 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 latest Release Update for the AIX operating system. Oracle Database 19c includes many features over its previous database versions. Oracle Database versions for IBM AIX are upward compatible with the higher versions of AIX Technology Levels (TL) within the same major version. For certification information of Oracle Database with IBM AIX, refer to [My Oracle Support portal](#) in the “Certification” section.

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). The Instant Client version of the Oracle Database is also supported on the 19c version of the database. With Oracle Database 12c Release 1, support for little-endian was introduced for the Instant Client running on Linux on Power and with 12c Release 2 support is for little-endian only.

Oracle Database Instant Client is a light version of Oracle Database Client Binaries. Oracle Database Instant Client is available on Linux on IBM Power to allow an application deployed on Linux on Power to connect to the Oracle Database no matter what platform the database is deployed on (AIX/Power or any other platform).

Although servers based on IBM POWER8 processor technology have reached end of support, many customers have yet to benefit from a technology refresh. IBM has already demonstrated and published that migrating Oracle Database workloads from a server with IBM POWER8 processors to one with IBM Power11 processors can halve the processor core requirements. The reduction in cores reduces the number of Oracle licenses by 50% which then reduces the Total Cost of Ownership (TCO). The IBM Power11 processor introduces significant performance advancements, enabling up to 25% improvement in the consolidation of Oracle Database 19c workloads compared to the previous IBM Power10 processor generation. This performance gain allows organizations to further optimize infrastructure

efficiency, reduce software licensing costs through better core utilization, and improve overall system throughput, all while having the best level of enterprise-grade reliability, availability, and serviceability (RAS) features.

Oracle is committed to release their latest products on IBM Power, to see the Oracle statement about support for 23ai on IBM Power see
<https://community.ibm.com/community/user/blogs/doug-bloom/2024/05/29/whats-the-latest-with-oracle-db-23ai-and-poweraix>.

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

As hybrid cloud adoption reaches 82% among enterprises and the number of applications is projected to surpass 1 billion by 2028, organizations are under pressure to manage growing complexities, skill shortages, and escalating infrastructure costs. IBM Power11, in combination with IBM Power Virtual Server (PowerVS), offers a compelling solution to these challenges by delivering a secure, scalable, and high-performance hybrid cloud platform purpose-built for mission-critical workloads.

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

Utilizing Power11 in PowerVS also unlocks advanced enterprise-grade security features. These include quantum-safe firmware signing, encrypted Live Partition Mobility (LPM), and a significantly lower number of hypervisor vulnerabilities (CVEs) compared to x86 alternatives. These capabilities are essential for organizations operating in regulated industries or managing sensitive data. Furthermore, PowerVS integrates natively with IBM WatsonX services via Satellite™ Connector, enabling enterprises to seamlessly infuse AI into their workloads and accelerate digital transformation initiatives.

PowerVS now has three generations of Power available:

- ▶ Power9
 - Soft landing for Power9 clients coming to cloud
 - Substantial support life remaining (through Oct. 2028)
 - Lowest cost
 - Special pricing on existing capacity
 - Clients can continue to provision and use Power9
- ▶ Power10
 - Work horse for business critical applications such as SAP RISE
 - Broad capacity availability across data centers
 - Excellent price/performance for AIX and IBM i workloads
 - Clients can continue to provision and use Power10
- ▶ Power11
 - No need to wait to receive Power11 on premises. Gain immediate access to Power11 via PowerVS
 - Provision in <10 minutes
 - Leverage for dev/test, application qualification
 - First cloud environment to provide Power11 both in IBM Cloud and on-premises (Private)

From a business value perspective, PowerVS delivers up to 25% faster SAP workload migration compared to x86 platforms, thanks to architectural consistency and superior compute performance. Flexible consumption models, shared processor pools, and tiered storage options help reduce total cost of ownership (TCO), while automated disaster recovery, secure backup, and non-disruptive maintenance enhance operational efficiency. The platform also supports regulatory compliance with standards such as SOC2, HIPAA, and PCI-DSS.

Common use cases include application modernization through access to over 250 IBM Cloud services, enhancing business resiliency with warm disaster recovery environments, optimizing data center operations by consolidating or exiting physical infrastructure, and supporting SAP transformation initiatives with unmatched availability and scalability for large-scale HANA deployments.

PowerVS is currently available in 22 IBM Cloud data centers globally 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 disaster recovery operations. This broad availability helps organizations adopt Power11 in a way that aligns with their specific business goals and compliance mandates. For more information on Power Virtual Server see section 10.3, “Power Virtual Server” on page 200.

1.4.9 Off-chip AI acceleration with IBM Spyre

IBM's Spyre accelerator card is a low-power, high-efficiency hardware module designed to complement Power11 processor-based servers for AI inference, analytics, and memory-intensive workloads. This feature is planned to be supported on the Power E1180 in fourth quarter of 2025.

The card is a system on a chip architecture optimized for enterprise AI workloads, with 32 low-power AI cores, and supports multi-precision for inference and training: FP16/8, INT8/4.

The card is enabled for foundation models and it is planned to support the Red Hat software stack (OCP AI). It will also support popular AI Frameworks and libraries (PyTorch, vLLM). Implemented in 5nm technology, the IBM Spyre card offers high performance and low-power characteristics with only 75W of power consumed.



Figure 1-13 Spyre card physical view

Since each IBM Spyre card operates at only 75 watts, it significantly reduces the power footprint compared to traditional GPU-based acceleration, which often exceeds 300 watts per device. IBM Spyre card is planned to be supported only inside the accelerator expansion drawer. Each accelerator expansion drawer supports a maximum of eight Spyre cards. Accelerator expansion drawer support on E1180 is as follows:

- ▶ At initial availability, a maximum of one accelerator expansion drawer will be supported on a Power11 processor-based server;
- ▶ Support for additional expansion drawers is planned in 2026.

Note: The Spyre card will require an updated fan-out card in the expansion drawer.

Unlike GPUs, which typically rely on high-bandwidth memory integrated on the card itself, the Spyre card is designed to access and leverage the system's main DDR5 memory modules. This enables the use of terabytes of high-performance memory available within the Power11 processor-based server. This architectural choice provides two main advantages:

- ▶ lower latency data access, thanks to the shared memory model;
- ▶ elimination of memory duplication, avoiding the need to copy data between host and device memory, as is common with discrete GPUs.

The use of system RAM not only simplifies software development and memory management but also contributes to a more energy-efficient and thermally balanced system architecture, an increasingly important factor in modern data centers and edge environments.

This design makes the Spyre card particularly well-suited 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 pre-processing.

Spyre card minimum supported levels of software stack for general availability and support on Power11 processor-based server in fourth quarter of 2025 are as follows:

- ▶ FW1110.10;
- ▶ HMC1111;
- ▶ RHEL 9.6;
- ▶ Spyre software stack container;
- ▶ OpenShift AI (Tech Preview 4Q 2025, GA 1Q 2026).

See Chapter 4, “Artificial Intelligence Support” on page 93 for more detailed information about use cases and solutions with the IBM Spyre card.

1.4.10 Red Hat OpenShift with support for expanded software ecosystem

Red Hat OpenShift on Power provides a robust, enterprise-grade container platform that empowers enterprises to modernize applications, deploy AI workloads, and integrate with a broad ecosystem of software. By running OpenShift on Power11 systems, organizations can take advantage of the platform’s performance, scalability, and security while maintaining architectural consistency across hybrid cloud environments.

One of the key advantages of OpenShift on Power is its support for a significantly expanded software ecosystem. This includes IBM Cloud Paks such as Cloud Pak for Data and Cloud Pak for Integration, Red Hat Runtimes like Quarkus and Spring Boot, and a wide range of open-source AI/ML frameworks including PyTorch, TensorFlow, and ONNX. Additionally, many independent software vendor (ISV) applications are now certified for the Power architecture, allowing banks to run containerized workloads alongside traditional AIX or IBM i systems. This co-location reduces latency, improves data gravity, and simplifies integration.

OpenShift on IBM Power also supports seamless hybrid cloud integration, particularly when deployed on IBM Power Virtual Server (PowerVS). This enables 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, OpenShift on Power offers a rich set of tools and resources. The OpenShift Container Catalog provides pre-built images and Helm charts optimized for Power, while infrastructure-as-code templates using 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 OpenShift and the breadth of the container ecosystem, financial institutions 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 adds up to 25% more cores per processor chip than comparable IBM Power10 systems. The Power11 processor builds upon the key capabilities we delivered with Power10 including stronger reliability, availability, and serviceability (RAS) characteristics, better energy efficiency, improved energy management, and improved quantum-safe security.

Beyond the processor itself, we are introducing the following improvements to packaging, memory architecture and AI acceleration:

- ▶ **Packaging Innovation**

The Power11 processor leverages new integrated stack capacitor (ISC) technology and advanced 2.5D packaging along with innovations in cooling such as improved heatsinks and more efficient fans to optimize energy delivery, improve thread and core strength and increase system capacity.

- ▶ **Enhanced System Architecture:**

The robust memory architecture for Power11 systems is based on the recently released DDR5 DDIMMs and enhanced OMI interfaces which enable improved memory reliability, capacity, and bandwidth. Given that OMI is technology agnostic, the Power11 portfolio built on the Power11 processor will also support OMI DDR4 memory migrated from Power10 high-end systems, enabling clients to protect their investments in memory technology.

- ▶ **AI Acceleration**

We continue to support a range of emerging enterprise AI use cases with the MMA architecture. Improvements to Power11 processor core strength and system capacity will improve the performance of the MMA for inferencing workloads. Furthermore, IBM intends to incorporate the IBM Spyre accelerator into Power11 offerings to provide additional AI inferencing capabilities. Working together, IBM Power processors and the IBM Spyre accelerator will enable the next generation infrastructure to scale demanding AI workloads for businesses. For more information on AI and the Spyre adapter see Chapter 4, “Artificial Intelligence Support” on page 93.

New Power11 processor details

Here are some details on the new Power11 processor design:

Technology and Packaging:

- 654 mm² 7nm refined Samsung (30 Billion Devices)
- Advanced 2.5D packaging technology optimizes space and improves signal integrity.
- Integrated Stack Capacitor technology provides enhanced power delivery and performance by integrating capacitors directly into the chip package
- Single-chip or Dual-chip sockets

Computational Capabilities:

- Up to 16 SMT8 Cores (2 MB L2 Cache / core)
- Up to 128 MB L3 cache (low latency non-uniform cache architecture management)
- Enterprise performance focus:
 - 3x core performance relative to Power9
 - 2x thread strength relative to Power9
 - 4x L2 cache, 4xMMU/ core relative to Power9
 - 4x crypto engine / core relative to Power9
- AI computational Focus:
 - 2x general SIMD / core relative to Power9
 - 4x matrix SIMD / core relative to Power9
 - New AI Instructions and data types

Robust Data Plane:

- 2 TB/s raw (32GT/s) PowerAXON + OMI signaling
- SMP interconnect for up to 16 sockets
- 2x OMI memory bandwidth relative 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 64KB data cache + 96KB instruction cache for L1 cache.

Comparing Power11, Power10, and Power9 processors

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

Table 1-5 Comparison of the Power11 processor technology to prior 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	18 billion	8 billion
Maximum cores	16	15	12
Maximum hardware threads per core	8	8	8
Maximum static frequency / 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 ^a	DDR4 and DDR5 ^b	DDR4 and DDR3
I/O bus	PCIe Gen 5	PCIe Gen 5	PCIe Gen 4

a. DDR4 only as migration from Power10 during same serial number upgrade

b. DDR5 support 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 utilizes a single chip module (SCM) which provides a single chip per socket and up to sixteen sockets per system. For the midrange Power E1150, the Power11 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 server line in the Power S1122/L1122 and Power S1124/L1124, with up to two sockets per system. There is a third packaging option, the entry Single Core Module (eSCM), which is used in the entry level processor features of the Power S1122 (the 4 core and the 10 core options). The eSCM is a modified DCM where the second core in the module does not have any processor capability and only adds I/O functionality.

Single Chip Module

The IBM Power E1180 server utilizes the Power11 Single-Chip Module (SCM) to deliver exceptional performance, scalability, and energy efficiency for enterprise workloads. The SCM design integrates up to 16 cores per chip, each capable of supporting simultaneous multithreading (SMT8), allowing up to 128 threads per socket. This compact yet powerful architecture enables the Power E1180 to scale up to 256 cores in a full system configuration, making it ideal for data-intensive applications such as AI, analytics, and large-scale transactional processing. The SCM's design emphasizes high throughput and low latency, ensuring that the server can handle demanding workloads with consistent performance.

Figure 1-15 shows the layout of the SCM used in the Power E1180.

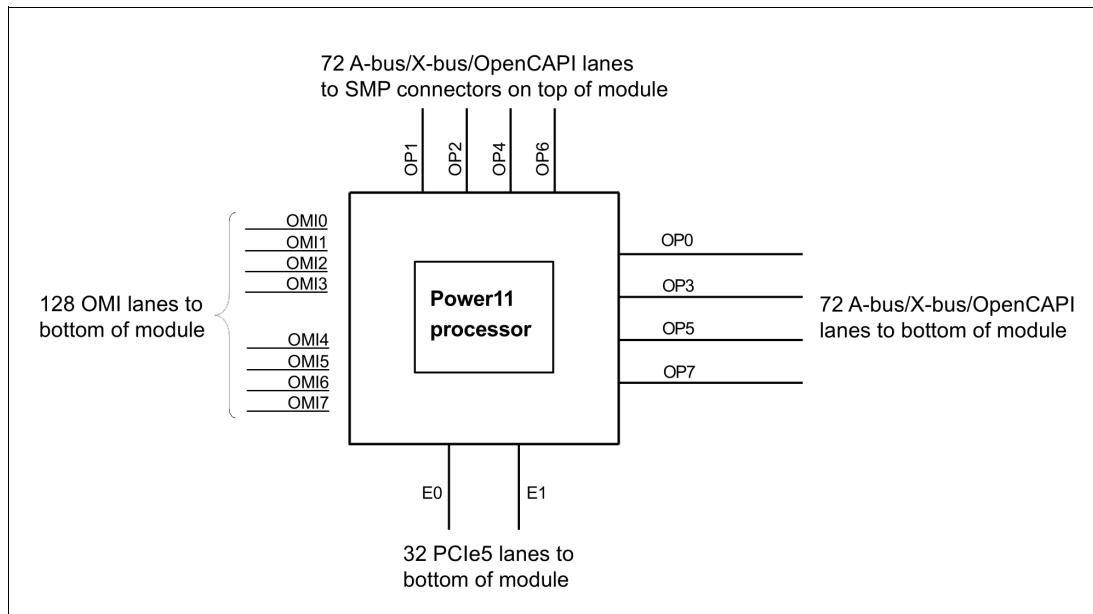


Figure 1-15 Single core module diagram

1.6 Operating system support

The IBM Power E1180 server supports the following families of operating systems:

- ▶ AIX
- ▶ IBM i
- ▶ Linux
- ▶ Red Hat OpenShift

In addition, the Virtual I/O Server (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 Live Partition Mobility capabilities.

For more information about the Operating System and other software that is available on Power, see [IBM Power](#). The minimum supported levels of IBM AIX, IBM i, and Linux are described in the following sections.

IBM makes announcements about support by newer versions of the operating systems over time. It is highly recommended that you keep your operating systems up to date to enable new security capabilities and for support of new features.

1.6.1 AIX operating system

The Power E1180 servers support the following minimum levels of the AIX operating system:

- ▶ When using virtual I/O

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

- AIX Version 7.3 with Technology Level 7300-02 and Service Pack 7300-02-02-2420
- AIX Version 7.3 with Technology Level 7300-03 and Service Pack 7300-03-00-2446
- AIX Version 7.2 with Technology Level 7200-05 and Service Pack 7200-05-08-2420

- ▶ When using direct I/O connectivity

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

- AIX Version 7.3 with Technology Level 7300-02 and Service Pack 7300-02-04-2520
- AIX Version 7.3 with Technology Level 7300-03 and Service Pack 7300-03-01-2520
- AIX Version 7.2 with Technology Level 7200-05 and Service Pack 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 TL3 SP1, 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 on AIX support, see section 6.2, “AIX” on page 122.

1.6.2 IBM i operating system

The Power E1180 supports the following minimum levels of the IBM i operating system 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 on IBM i support see section 6.3, “IBM i” on page 126.

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 servers, but do not include any formal Enterprise Grade support.

Red Hat Enterprise Linux

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

At announcement, the Power E1180 servers support the following minimum levels of the Red Hat Enterprise Linux operating system:

- ▶ Red Hat Enterprise Linux 8.10 for Power LE, or later (Power10 compatibility)
- ▶ Red Hat Enterprise Linux 9.4 for Power LE, or later (Power10 compatibility)
- ▶ Red Hat Enterprise Linux 9.6 for Power LE, or later (Power11 native)

- ▶ Red Hat Enterprise Linux 10 for Power LE, version 10.0, or later
- ▶ Red Hat Enterprise Linux for SAP with Red Hat Enterprise Linux 9 for Power LE version 9.6, or later

Red Hat OpenShift Container Platform

The Red Hat OpenShift Container Platform is supported to run on IBM Power servers, including the Power E1180. To run OCP in native mode on Power11, you need to run OCP 4.19 which is based off of Red Hat Enterprise Linux 9.6. Previous versions will run in Power10 compatibility mode.

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

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

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 allows it to access all of the features of the Power11 processor and platform.

At announcement, the Power E1180 servers support the following minimum levels of the SUSE Linux Enterprise Server operating system:

- ▶ SUSE Linux Enterprise Server for SAP with SUSE Linux Enterprise Server 15, Service Pack 6, or later
- ▶ SUSE Linux Enterprise Server 15, Service Pack 6, or later

For more information on Linux support see section 6.4, “Linux on IBM Power” on page 134.

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 (Service Pack)
- ▶ VIOS 4.1.0.40 (Service Pack) (P10 compatibility mode)
- ▶ VIOS 3.1.4.60 (Service Pack) (P9 compatibility mode)

1.7 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 operating system.

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

An HMC is a dedicated appliance to configure and manage system resources on IBM Power. It can be delivered as a physical server or as a virtual appliance running in either Power VM or an x86 VM. A graphical user interface (GUI), a Command-line interface (CLI), or REST API interfaces 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 V11 can manage Power11, Power10 and Power 9 based servers.

Restriction:

- ▶ The 7063-CR1 and 7042 HMCs are not supported as an HMC for Power11 servers.
- ▶ The HMC code at V11R1M1110 cannot manage POWER8 or earlier processor-based systems

1.8 IBM Power E1180 rack integration and handling guidelines

The IBM Power E1180 server is engineered to fit seamlessly into a standard 19-inch rack and has been fully tested and certified in the IBM 42U Enterprise Rack (model 7965-S42). While clients may choose to install the server in third-party racks, it is essential that those racks meet IBM's requirements for strength, rigidity, depth, and hole pattern compatibility. IBM strongly recommends consulting with IBM Service to validate the suitability of any non-IBM rack.

For optimal installation, performance, and serviceability, IBM highly recommends ordering the Power E1180 with the IBM 42U Enterprise Rack. Initial system configurations are shipped in the 7965-S42 rack to streamline client deployment, assure high-quality assembly and testing in IBM Manufacturing, and provide a complete, secure shipping solution. Clients who prefer to receive the server without a rack can opt for the factory deracking feature (ER21), in which case, the system is tested in the rack and then shipped separately.

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

- ▶ #ECRA – Acoustic front door for noise reduction
- ▶ #ECRT – High-end appearance front door
- ▶ #ECRM – Cost-effective plain front door

Due to the size and weight of the system node drawers – each weighing 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 (FC EB3Z) with a hand crank capable of lifting servers up to 400 lb, and an optional angled shelf kit (FC EB4Z) for precise positioning during installation.

Important: The customer is responsible for helping ensure 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 racks are ordered:

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

Power distribution units (PDUs)

The Power E1180 server factory that is integrated into an IBM rack uses horizontal PDUs located in the EIA drawer space of the rack instead of the typical vertical PDUs found in the side pockets of a rack. This is done to aid cable routing. Each horizontal PDU occupies 1U. Vertically mounting the PDUs to save rack space can cause cable routing challenges and interfere with optimal service access.

When mounting the horizontal PDUs, it is a good practice to place them almost at the top or almost at the bottom of the rack, leaving 2U or more of space at the very top or very bottom open for cable management. Mounting a horizontal PDU in the middle of the rack is generally not optimal for cable management.

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

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

PCIe I/O expansion drawer and racks

IBM Manufacturing can factory-integrate the PCIe I/O expansion drawer (#ENZ0) with new server orders. Because expansion drawers complicate the access to vertical PDUs if located at the same height, IBM recommends accommodating PDUs horizontally on racks that have one or more PCIe I/O expansion drawers. Following this recommendation, IBM Manufacturing will always assemble the integrated rack configuration with horizontally mounted PDUs unless CRSP (#8453) is on the order. When specifying CSRP, you must provide the locations where the PCIe I/O expansion drawers should be placed and avoid locating them adjacent to vertical PDU locations EIA 6 through 16 and 21 through 31.



IBM Power11 Enterprise Server – Power E1180

This chapter focuses on the Power11 Enterprise Server E1180. The IBM Power E1180, a follow on to the Power E980 and the Power E1080, is the industry's most scalable server for consolidating database and mission critical workloads. The Power E1180 provides up to sixteen sockets and up to 256 SMT8 cores with up to 64 TB of DDR5 memory, enabling it to handle massive data volumes and complex applications with ease. One of its standout features is transparent memory encryption, which secures data in memory without compromising performance – ideal for hybrid cloud environments. Additionally, its built-in AI inference engine allows 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 2x better memory reliability than standard DIMMs and includes advanced diagnostic and recovery capabilities to minimize downtime. The server's energy-efficient enhanced 7nm architecture reduces power consumption while delivering higher throughput per core. These features make it particularly well-suited for mission-critical workloads like SAP HANA, Oracle databases, and healthcare applications that demand maximum uptime and security. With dynamic scaling and lower 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:

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

2.1 IBM Power E1180 overview

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

Key capabilities include:

- ▶ Autonomous System Maintenance with zero planned downtime
- ▶ Unmatched Resilience: 99.9999% uptime – IBM Power's most reliable server ever
- ▶ Spare Core Architecture: Up to 4 spare cores per drawer for enhanced availability
- ▶ IBM Power Cyber Vault: Sub-minute ransomware detection and automated recovery (with IBM Storage)
- ▶ Quantum-Safe Encryption: Secure boot and Logical Partition Migration (LPM)
- ▶ Automated Cryptographic Inventory via 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 Enterprise Pools 2.0
- ▶ Modular, Scalable Design: Up to four 5U CEC drawers + 2U System Control Unit
- ▶ High-Core Density: Up to 256 Power11 SMT8 cores (10, 12, 16-core SCM options)
- ▶ Massive Memory Capacity: Up to 64TB DDR5 (16TB per drawer)
- ▶ High-Speed I/O: 8 PCIe Gen5 blind-swap slots per drawer
- ▶ Concurrent Repair: Non-active SMP cables with improved fault isolation
- ▶ Integrated NVMe Storage: 4x 7mm or 2x 15mm U.2 bays per drawer
- ▶ Scalable Expansion: Up to 16 I/O and NVMe drawers (4 per CEC drawer)
- ▶ Seamless Upgrades: Serial number-preserving upgrades from Power10 (DDR4 and DDR5 supported)

The IBM Power E1180 is more than a server – it's 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 with confidence. Figure 2-1 shows the Power E1180 system unit from the front



Figure 2-1 Power E1180 system unit

The machine types and model for the Power11 E1180 systems are listed in Table 2-1.

Table 2-1 Machine types and model of Power E1180 server

Server name	Machine type and model
Power E1180	9080- HEU

2.1.1 IBM Power11: Engineered for Unmatched Uptime

IBM Power E1180 is built to support maximum Reliability, Availability, and Serviceability (RAS). Reliability is provided by minimizing potential errors and outages through error detection and self-healing. Availability is provided by minimizing unplanned downtime through self-diagnosing capabilities as well as predicting and fixing errors to avoid downtime. Serviceability allows accurate diagnostic capability utilizing call home functionality and concurrent repair capabilities built into the system.

Reliability

IBM Power11 is built to minimize potential errors and maximize system integrity through advanced design and manufacturing hardening. Key features include:

- ▶ Error Detection and Self-Healing: Proactively identifies and corrects faults before they impact operations.
- ▶ Spare Core Technology: Each system drawer includes up to 4 spare cores, enhancing fault tolerance and maintaining performance.
- ▶ Industry-Leading Uptime: Designed to be the most reliable server in IBM Power history, delivering 99.9999% availability.

Availability

Power11 provides continuous operations even in the face of hardware or software issues:

- ▶ Predictive Failure Analysis: 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

Power11 simplifies maintenance and reduces downtime through intelligent service features:

- ▶ Concurrent and Guided Repair: Allows components to be serviced or replaced without shutting down the system.
- ▶ Accurate Problem Diagnostics: Speeds up 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.

Power11 is designed to be the most reliable server in the history of the IBM Power platform with 99.9999% of uptime.

The IBM Power E1180 comes with dual line cord redundancy along with n+1 power supply redundancy and n+1 fan rotor redundancy. Power supplies and fans are concurrently maintainable. The system service processor is redundant with a dynamic failover capability. Each system node has dual processor clock logic with a dynamic failover capability.

Concurrent maintenance of the real-time clock battery and the operator panel is also provided.

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

Table 2-2 RAS features on Power E1180

Feature name	IBM Power E1180 support
Redundant / Hot Swap Power Supplies	Yes
Pluggable VRMs with redundant/spare voltage phases	N+2
OCC error handling w/ power safe mode	Yes
EPOW reporting	Yes
TOD Battery Concurrent Maintenance	Yes
Redundant Fans/Blowers	Yes
Hot Swap Fans / Blowers	Yes
NVMe drives and/or Media Concurrent Maintenance	Yes
PCI Adapter Concurrent Maintenance	Yes
EEH (Enhanced Error Handling) Support	Yes
Concurrent Op Panel Repair	Yes
Redundant Service Processors	Yes
Dynamic Service Processor Failover	Yes
Auto Service Proc. Reset/Reload	Yes
Redundant System Clocks	Yes
Dynamic System Clock Failover	Yes
Redundant VPD	Yes
Concurrent FW Update	Yes
Memory Preserving IPL	Yes
HW and OS Dump	Yes
Processor Instruction Retry	Yes
Spare core	2
Predictive De-allocation (such as, core)	Yes
Core check stop	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

Feature name	IBM Power E1180 support
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 runtime)
Predictive Dynamic Memory De-allocation	Yes
Active Memory Mirroring for Hypervisor	Yes

2.2 IBM Power E1180 components

The IBM Power E1180 server, designated as machine type-model 9080-HEU, is the most powerful and scalable system in the IBM Power portfolio. It is built from modular components, primarily central electronic complex (CEC) enclosures known as system nodes, along with additional units and expansion drawers.

2.2.1 System nodes

A system node houses the connections and electronics needed to link the processors with memory, internal storage, adapters, and system interconnects. The Power E1180 supports configurations of one to four system nodes per server.

Each 5 EIA or 5U system node has four air-cooled SCMs optimized for performance and scalability. The Power E1180 SCMs can have ten, twelve, or sixteen Power11 cores running at up to 4.40 GHz and simultaneous multithreading that executes up to eight threads per core. Each SCM has dual memory controllers to deliver up to 512 GBps of peak memory bandwidth per socket or 2048 GBps per node. Using PCIe Gen5 I/O controllers, which are also integrated onto each SCM to further reduce latency, up to 576 GBps peak I/O bandwidth is available per node. Thus, a Power E1180 system bandwidth provides maximum processor performance, enabling applications to run faster and be more responsive.

Each system node contains:

- ▶ Four sockets for Power10 single-chip modules (SCMs), each of which supports 10, 12, or 16 Power11 processor cores and includes integrated logic for power and data connectivity.
- ▶ 64 DDIMM slots for DDR5 memory, with support for DDR4 DIMMs when upgrading from an existing Power10 system.
- ▶ Up to 16 TB of system memory per drawer when using the largest available DDIMMs, allowing a fully configured 4-node system to scale up to 64 TB of RAM.
- ▶ Four NVMe drive bays for internal boot storage, with additional capacity available via expansion drawers.
- ▶ Eight PCIe Gen5 slots per node, for a maximum of 32 PCIe slots across a 4-node configuration.

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 server must use the same processor feature to ensure uniformity.

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

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 VPD. The SCU is also prepared to facilitate USB connectivity that can be used by the Power E1180 server.

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

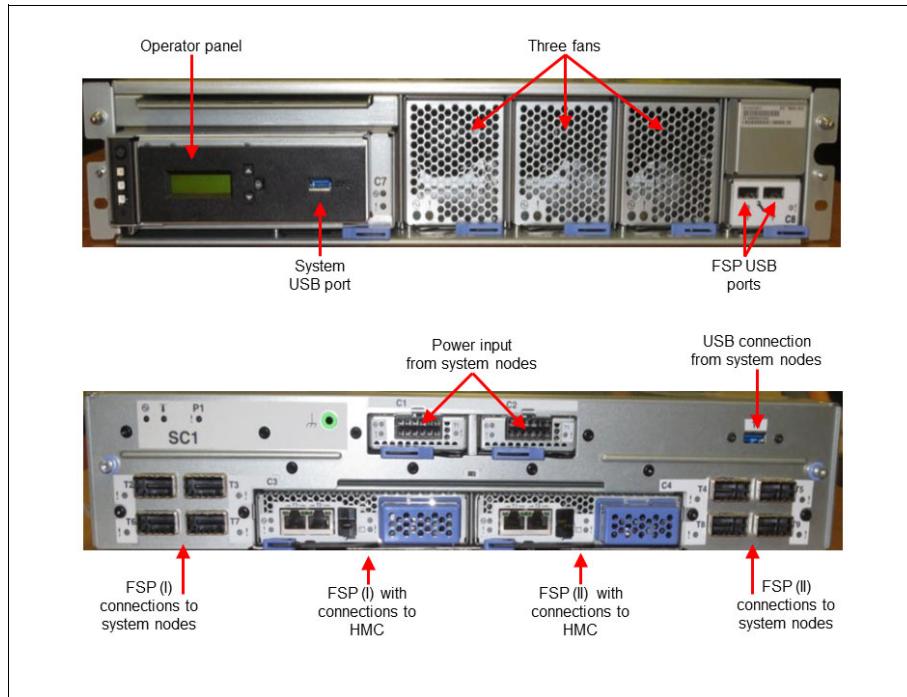


Figure 2-2 Front and rear view of the system control unit

One SCU is required for each Power E1180 server (any number of system nodes), and depending on 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.
- ▶ In a Power E1180 single system node configuration, both UPIC cables are provided from the single system node to be connected to the SCU.
- ▶ For a two, three, or four system nodes configuration, one UPIC cable is provided from the first system node and the second UPIC cable is provided from second system node to be connected to the SCU.

The set of two cables facilitate a 1+1 redundant electric power supply. If there is a failure of one cable, the remaining UPIC cable is sufficient to feed the needed power to the SCU.

Two service processor cards (FSPs) are installed in the SCU, ordered using two mandatory #EDFB features. Each processor card provides two 1 Gb Ethernet ports to connect to the Hardware Management Console (HMC) system management connection. One port is used as 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 of the cards in the SCU.

Four ports per FSP provide redundant connections from the SCU to each system node. System nodes connect to the SCU by using the cable features #EFCH, #EFCE, #EFCF, and #EFCG. Feature #EFCH connects the first system node to the SCU and it 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 of the nodes. All the other cable features are added depending on the number of extra system nodes that are configured and includes FSP and SMP cables.

The SCU implementation also includes the following highlights:

- ▶ Elimination of clock cabling since the introduction of Power9 processor-based servers
- ▶ Front-accessible USB port
- ▶ Optimized UPIC power cabling
- ▶ Optional external DVD
- ▶ Concurrently maintainable time of day clock battery

2.3 Processors

Each system node in a Power E1180 server provides four sockets to accommodate Power11 SCMs. Processor features are offered in 10-core, 12-core, and 16-core density. Table 2-3 lists the available processor Feature Codes for a Power E1180 server.

Table 2-3 Processor features

Feature Code	Processor Cores	Max System Config	Frequency Range (GHz)	Recommended Usage
EDQD	16 cores	256 cores	3.80 to 4.3	Max throughput and density
EDQ9	12 cores + 1 spare	192 cores	3.90 to 4.4	Max core/thread performance
EDQ0	10 cores + 1 spare	160 cores	3.90 to 4.2	Entry price/performance

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

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

The Power11 processor frequencies for the Power E1180 offerings are shown in Table 2-4

Table 2-4 Power11 Power & Frequencies.

Fused Core Configuration	Power Save	Fixed Frequency (Nominal)	WOF Base	Ultra Turbo	Power
16 Core	3.6	3.8	3.95	4.3	440W
12 Core	3.6	3.9	4.05	4.4	400W
10 Core	3.6	3.9	4.05	4.2	355W

Processor activations

Each Power E1180 server requires a minimum of 16 permanent processor core activations, which can be either static activations or Linux on Power activations. This minimum applies to the entire system, not to individual nodes. The remaining cores can be activated either permanently or temporarily, or left inactive until needed. The 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 Power hypervisor allocates as needed. The exception is when using Power Enterprise Pools 2.0 where a minimum of one activation is required.

There are different activation options available to accommodate various usage scenarios and pricing structures. Static-activations are permanent and can support any type of application environment on the server. Mobile activations are assigned to a specific server but can be transferred to any server within the Power Enterprise Pool, supporting all types of applications.

2.4 Memory

The Power11 processor continues to leverage the native Open Memory Interface (OMI) introduced with Power10. This interface provides low-latency, high-bandwidth, and technology-agnostic 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 8 on-chip MCUs, distributed across two physical blocks located on opposite edges of the Power11 die. Each block supports 64 OMI lanes, organized into four ports. Each port contains two links, with 8 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 sub-channels from the Power11 single-chip module (SCM) connect directly to 16 memory-buffered DDIMM slots, with one DDIMM per sub-channel. Each Power11 processor supports up to 16 DDIMMs, allowing for a maximum of 64 DDIMMs per Power E1180 drawer.

OMI Lane Summary (per SCM)

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

Figure 2-3 shows the memory subsystem architecture.

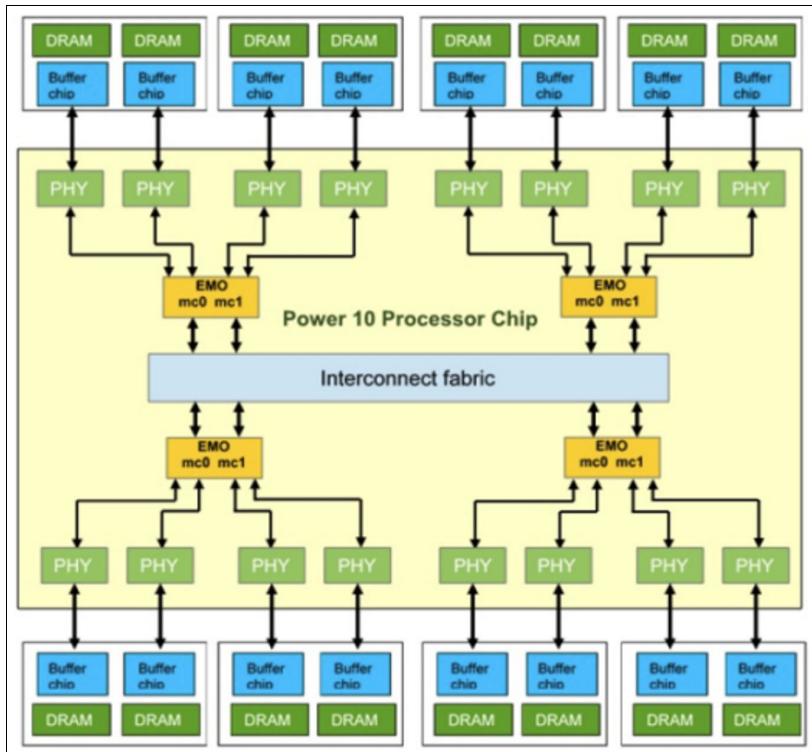


Figure 2-3 E1180 Memory Subsystem

DDR5 DDIMM Support

E1180 will only ship with the 4U Enterprise level DDIMMs utilizing DDR5 DRAMs. The DDIMM densities supported in E1180 are 32GB, 64GB, 128GB and 256GB. The DDIMMs have their own voltage regulators and are n+1 redundant. It takes two PMICs (Power Management IC) to supply all the voltage levels required by the DDIMM.

On the 4U DDIMM there are 4 PMICs total, made up of two redundant pairs. As long as one PMIC in each of the redundant pairs is still functional, the DDIMM is not called for 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. This design:

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

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 density that is configured for a specific Power E1180 server.

For each processor socket a minimum of 8 of the 16 DDIMMs are populated

- In a node, a minimum of 32 of 64 for the DDIMMs are populated
- In a 4 node system, a minimum of 128 of the 256 DDIMMs are populated

When filling the memory slots in the SCM, a quantity of four DDIMMs must be used. Thus, the DDIMM slots of the SCMs are from 50% to 100% filled. The system node (four SCMs) DDIMM slots can have 32, 36, 40, 44, 48, 52, 56, 60, and 64 DDIMMs physically installed.

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

Table 2-5 Memory features

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

All DDIMMs must be identical on the same SCM, so if you are using eight DDIMMs, both memory features on an SCM must be identical. A different SCM in the same system node can use a different memory feature. For example, one system node could technically 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 used on DDR5 memory features and used on any size memory feature. Activations are not specific to a DDIMM, but they are known as a total quantity to the server. The Power hypervisor determines what physical memory to use.

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

Memory Performance Considerations

Memory performance is dependent on the number of DDIMMs populated and the size of the DDIMMs.

Plugging DDIMMs that are all the same size will provide 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 on a multi drawer system if memory capacity between drawers are balanced.
- ▶ System performance improves as the amount of memory is spread across more DDIMM slots.

For example in one drawer if 1TB is needed, it is better to have 64 x 32GB DDIMMs than to have 32 x 64GB DDIMMs. Though maximum memory bandwidth is achieved by filling all the memory slots, plans for future memory additions should be considered when deciding which memory feature size to use at the time of 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 allocated to an AIX partition. For more information on AME see 2.4.3, “Active Memory Expansion” on page 52.

DDIMM Plug Rules

E1180 system drawers will be offered with a base memory configuration of 32 DDIMM's each. In this base configuration each processor module will have 8 of 16 DDIMM's populated. Feature upgrades will be offered in increments of 4 DDIMM's, all of which have the same capacity. These 4 DDIMMs must all be plugged into 4 of the 8 remaining sites connected to single processor module.

It is not allowed to plug DDIMM's in a way that ends up with an odd number of DDIMM sites populated behind a single processor module. Nor is it allowed to plug DDIMM's in a configuration that ends up with 10 or 14 DDIMM sites populated behind a single processor module. The only valid number of DDIMM's plugged into sites connected to a given processor module is 8 or 12 or 16.

The general DDIMM plug rules for the E1180 system are as follows:

- ▶ Minimum memory allowed (please see 3.5 Minimum Memory in E1180)
- ▶ All DDIMMs under each processor has to be same capacity
- ▶ DDIMMs must be plugged in Quads (group of 4)
- ▶ DDR4 and DDR5 DDIMMs can NOT be mixed in a single node system.
- ▶ Each DDIMM quad must all be the same size and type (identical to each other) and has to be same with other DDIMM quads under each socket, but can be different from other DDIMM quads from other sockets.

For example, in a one node drawer, first processor can have 8x32GB DDIMMS and second processor can have 8x64GB DDIMMS and third processor can have 8x128GB DDIMMS and forth processor can have 8x256GB DDIMMS.

- ▶ System performance improves when DDIMMS match each other.

Note that the end user is given flexibility on which processor is assigned the additional memory beyond the minimum memory. In multi drawer E1180 systems, the user can either fully populate one drawer and have the other drawers partially populated or have all the drawers symmetrically populated.

For example, in a two drawer system the end user has 12 additional quads of memory (beyond the minimum already installed). They can install eight quads in one drawer and four quads in the other drawer. Or they can install six quads in both drawers.

2.4.1 Active Memory Mirroring

Active Memory Mirroring (AMM) is a reliability feature in IBM Power Servers that enhances system availability by duplicating the memory used by the system hypervisor. It maintains two synchronized copies of this critical memory, allowing the system to instantly switch to the mirrored copy in the event of a memory failure. This seamless failover mechanism helps

prevent system crashes and provides continuous operation, making AMM especially valuable for enterprise environments running mission-critical workloads.

Active Memory Mirroring is available in all configurations of Power E1180.

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 qualifies the MCU as an ideal functional unit to implement memory encryption logic. The Power11 on-chip MCU encrypts and decrypts all traffic to and from system memory that is based on the AES technology.

The Power11 processor supports the following modes of operation:

- ▶ AES XTS mode

XTS is an abbreviation for the xor–encrypt–xor based tweaked-codebook mode with ciphertext stealing. AES XTS provides a block cipher with strong encryption, which is useful to encrypt 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. The data that is stored in the DIMMs can leave the data center in the clear if not encrypted.

Also, memory cards that leave the data center for repair or replacement can be a potential security breach. Because the attacker might have arbitrary access to the persistent DIMM data, 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 IBM Power servers.

- ▶ AES CTR mode

CTR stands for *Counter* mode of operation and designates a low-latency AES block cipher. Although the level of encrypting is not as strong as with the XTS mode, the low-latency characteristics make it the preferred mode for memory encryption of volatile memory. AES CTR makes it more difficult to physically gain 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 servers 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 are not compromised by the AES CTR encryption mode. Because the encryption has no noticeable performance effect and because of the obvious security benefit, the 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 supporting the AIX operating system that helps enable the effective maximum memory capacity to be larger than the true physical memory maximum. Compression and decompression of memory content can enable memory expansion up to 100% or more. This can enable a partition to do significantly more work or support more users with the same physical amount of memory. Similarly, it can enable a server to run more partitions and do more work for the same physical amount of memory.

AME uses CPU resource to compress and decompress the memory contents. The trade-off of memory capacity for processor cycles can be an excellent choice, but the degree of expansion varies on how compressible the memory content is. It also depends on having adequate spare CPU capacity available for this compression and decompression.

The Power E1180 includes a hardware accelerator designed to boost AME efficiency and uses less Power core resource. You have a great deal of control over AME usage. Each individual AIX partition can turn on or turn off AME. Control parameters set the amount of expansion desired in each partition to help control the amount of CPU used by the AME function. An IPL is required for the specific partition that is turning memory expansion. When turned on, monitoring capabilities are available in standard AIX performance tools, such as lparstat, vmstat, topas, and svmon.

A planning tool is included with AIX, enabling you to sample actual workloads and estimate both how expandable the partition's memory is and how much CPU resource is needed. Any Power model can run the planning tool. In addition, a one-time, 60-day trial of AME is available to enable more exact memory expansion and CPU measurements. You can request the trial using the [Capacity on Demand](#) web page.

AME is enabled by chargeable hardware feature EM89, which can be ordered with the initial order of the system or as an MES order. A software key is provided when the enablement feature is ordered, which is 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.

The additional CPU resource used to expand memory is part of the CPU resource 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 server 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 allows for scalable expansion of PCIe resources based on system requirements. These buses provide high-bandwidth connectivity for performance-critical components and peripherals.

Each PCIe adapter is installed in a dedicated slot located within a PCIe extender cassette. These cassettes are directly connected to the system backplane, ensuring efficient signal routing and ease of maintenance. This design supports hot-plug capabilities and simplifies the process of upgrading or replacing PCIe cards without system downtime. The PCIe cassettes support Half-Height, Half-Length adapters and all slots are concurrently maintainable. All slots support Enhanced Error Handling (EEH).

For customers requiring additional PCIe adapters slots, the Power E1180 supports the PCIe Gen 4 I/O Expansion Drawer (#ENZ0). Up to four PCIe Gen 4 I/O Expansion Drawers are supported on each system node and each drawer supports up to 12 Gen4 PCIe adapter slots. For more information on PCIe slots and I/O Expansion drawers see Chapter 3, “I/O Subsystem” on page 61.

2.6 Storage and USB features

The Power E1180 offers 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 the preference of many clients to minimize potential security vulnerabilities associated with USB access. For those who require USB functionality, a PCIe USB adapter can be installed in one of the available I/O slots, either within a system node or in an I/O expansion drawer.

NVMe Devices

Power E1180 provides 4 PCIe x4 Gen4 buses that connect to internal NVMe storage devices. The Power E1180 provides a carrier backplane in each system node 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 4-drive carrier, or EJBD provides the 2-drive carrier. Each system node requires one backplane, even if no NVMe U.2 drives are selected.

For customers requiring additional data storage capacity, the Power E1180 supports the attachment of the NED24 I/O expansion drawer. Each NED24 drawer provides 24 NVMe slots which support small form factor (SFF) NVMe U.2 drives, providing high-performance, scalable storage expansion.

More detail on the NVMe support is provided in section 3.1.2, “Internal NVMe storage subsystem” on page 67

USB

The IBM Power E1180 supports a single stand-alone external USB DVD (#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 LPAR and can be migrated from one operating LPAR to another as needed to assign the DVD drive feature to any LPAR according to the need using DLPAR functionality.

Section 3.1.3, “Internal USB subsystem” on page 70 provides more information related to the USB features of the system.

2.7 I/O expansion drawers

As previously discussed each system node provides up to eight PCIe adapter slots and up to four Internal NVMe SSDs installed in an internal NVMe enclosure. In a fully configured four-node system, the Power E1180 supports up to 32 PCIe Gen5 adapters and 16 NVMe drives.

If additional PCIe slots or additional internal storage is required the Power E1180 supports I/O expansion drawers:

- ▶ If more PCIe slots beyond the system node slots are required, the Power E1180 server supports the addition of a PCIe Gen 4 I/O Expansion Drawer (#ENZ0)
- ▶ To increase the amount of internal disk in the system, the Power E1180 supports a 24-bay NVMe enclosure - NED24 NVMe expansion drawer (#ESR0).

I/O Drawer connectivity

To enable connectivity to either the ENZ0 or NED24 PCIe Gen4 expansion drawers, the PCIe Converter Adapter (Feature Code EJ24) is available for installation in one of the PCIe extender slots located at the rear of the system node. In conjunction with the adapter, a new generation of high-speed Active Optical Cables (AOC) and copper cables is introduced to support the PCIe Gen4 communication protocol, ensuring 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 or PCI Gen4 I/O drawers only and the maximum number of NED24 drawers based on the system configuration.

Table 2-7 Drawer availability for different system nodes configuration

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 server specifications are essential to planning for your server. For a first assessment in the context of your planning effort, this section provides you with an overview that is related to the following topics:

- ▶ Electrical characteristics
- ▶ Environment requirements and noise emission
- ▶ Physical dimensions

For more information about the comprehensive Model 9080-HEU server specifications and product documentation, see [IBM Documentation](#).

2.8.1 Electrical characteristics

Each Power E1180 system node has four 1950 W bulk power supplies. The hardware design provides N+2 redundancy for the system power supply, and any node can continue to operate at full function in nominal mode with any two of the power supplies functioning.

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

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

Environmental requirements and noise emission

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

Note: The [IBM Systems Energy Estimator](#) tool can provide more accurate information about the power consumption and thermal output of systems based on a specific configuration.

The recommended operating environment designates the long-term operating environment that can result in the greatest reliability, energy efficiency, and reliability. The allowable operating environment represents where the equipment is tested to verify functions. Because of the stresses that operating in the allowable envelope can place on the equipment, these envelopes must be used for short-term operation, not continuous operation.

The condition of a non-operating environment pertains to the situation when equipment is removed from the original shipping container and is installed, but is powered down. The allowable non-operating environment is provided to define the environmental range that an unpowered system can experience short term without being damaged.

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

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)

Property	Environment		
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. IBM does not recommend operation above 27° C, but you can expect full performance up to 35° C for these systems. Above 35° C, the system can operate, but possible reductions in performance might occur to preserve the integrity of the system components. Above 40° C, there might be reliability concerns for components within the system.

A comprehensive list of noise emission values for various different Power E1180 server configurations is provided by 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 regulations that are issued by the Occupational Safety and Health Administration (OSHA) or Europe-an Community Directives, can govern noise level exposure in the work-place and might apply to you and your server installation. The Power E1180 server is available with an optional acoustical door feature that can help reduce the noise that is emitted from this system.

The sound pressure levels in your installation depend on various factors, including the number of racks in the installation; the size, materials, and configuration of the room where you designate the racks to be installed; the noise levels from other equipment; the room ambient temperature, and employees' location in relation to the equipment.

Also, compliance with such government regulations depends on various other factors, including the duration of employees' exposure and whether employees wear hearing protection. As a best practice, consult with qualified experts in this field to determine whether you are in compliance with the 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 system component must be mounted in a 19-inch industry standard rack. The SCU requires 2U and each system node requires 5U rack units. The space required based on the number of nodes is shown in Table 2-10.

Table 2-10 PCIe slots available internally

Number of Nodes	Space required
1	7U

Number of Nodes	Space required
2	12U
3	17U
4	22U

More rack space must be allotted; for example, to PCIe I/O expansion drawers, an HMC, flat panel console kit, network switches, power distribution units (PDUs), and cable egress space.

Table 2-11 lists the physical dimensions of the Power E1180 server control unit and a Power E1180 system node. The component height is also given in Electronic Industries Alliance (EIA) rack units. (One EIA unit corresponds to one rack unit (U) and is defined as 1 3/4 inch or 44.45 mm respectively).

Table 2-11 Physical dimensions of the Power E1180 server 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

It is a best practice to have a lift tool available at each site where one or more Power E1180 servers are located to avoid any delays when servicing systems. An optional lift tool #EB2Z is available for order with a Power E1180 server. One #EB2Z lift tool can be shared among many servers and I/O drawers. The #EB2Z lift tool provides a hand crank to lift and position up to 159 kg (350 lb). The #EB2Z lift tool is 1.12 meters x 0.62 meters (44 in x 24.5 in).

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

Environmental Operating Envelope

Table 2-12 shows the environmental envelope for the Power E1180/

Table 2-12 ASHRAE Class A3 Requirements

Status	Temperature (Dry Bulb)	Altitude	Relative Humidity
Power Off	5C to 45C	N/A	8% to 80%
Shipping	-40C to 60C	N/A	5% to 100%

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



I/O Subsystem

The internal I/O system of IBM Power servers is engineered for high-speed, low-latency data movement, ensuring optimal performance for enterprise workloads. At the heart of this architecture is the Power11 processor, which integrates advanced I/O subsystems that support massive throughput and efficient communication between processors, memory, and peripheral devices. IBM Power leverages technologies like PCIe Gen5, and NVMe to deliver ultra-fast data access and minimize bottlenecks.

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

This internal I/O design not only enhances performance but also contributes to system reliability and availability by incorporating features like redundant paths, hot-swappable components, and advanced error detection and correction. Together, these capabilities make the IBM Power internal I/O system a key enabler of high-throughput, resilient, and scalable enterprise computing.

This chapter contains the following topics:

- ▶ Internal I/O
- ▶ Enhancing I/O Scalability with Expansion Drawers
- ▶ Supported adapters
- ▶ Other device support

3.1 Internal I/O

The IBM Power E1180, built on the Power11 architecture, features a highly advanced internal I/O subsystem designed to support high-throughput, low-latency data movement within the server. At the heart of its internal I/O is the Power11 processor's Open Memory Interface (OMI), which decouples memory from the processor and allows for flexible, high-bandwidth memory configurations. This architecture supports DDR4 and DDR5 memory through memory buffer chips, enabling faster data access and improved scalability. Additionally, the Power E1180 integrates PCIe Gen5 technology, which doubles the bandwidth of its predecessor, allowing for faster communication between internal components such as storage controllers, and network adapters.

The internal I/O design of the Power E1180 also emphasizes reliability and serviceability. It includes redundant paths and hot-swappable components to minimize downtime and support continuous operation in enterprise environments. The system supports NVMe drives natively, enhancing internal storage performance, and features a modular design that allows for easy expansion through I/O drawers. These drawers connect via high-speed links and can house additional PCIe slots, enabling the system to scale with growing workload demands. Overall, the internal I/O architecture of the IBM Power E1180 is engineered to deliver exceptional performance, flexibility, and resilience for mission-critical applications.

3.1.1 Internal PCIe subsystem

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 server:

- ▶ A single-node system provides 8 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 allows for scalable expansion of PCIe resources based on system requirements. For customers requiring additional PCIe adapters slots, the Power E1180 supports the PCIe Gen 4 I/O Expansion Drawer (#ENZ0). Up to four PCIe Gen 4 I/O Expansion Drawers are supported on each system node and each drawer supports up to 12 Gen4 PCIe adapter slots.

PCIe Bus Layout in the Power E1180

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

- ▶ Six PCIe x16 Gen 4 / x8 Gen 5 buses
- ▶ Two PCIe x8 Gen 5 buses

In addition, each system node provides four PCIe x4 Gen4 buses that connect to the PCIe internal NVMe storage bays. The NVMe support is discussed in 3.1.2, “Internal NVMe storage subsystem” on page 67.

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

- ▶ Each Power E1180 system node is equipped with eight PCIe Gen 5 hot-plug enabled slots.

- ▶ The PHBs provide different type of end-point connections. Both PHBs from Power11 CPUs with physical location codes C13 and C16 are directly connected 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 is directly connected to a PCIe Gen4 x16 or PCIe Gen 5 x8 slot.
- ▶ The second PHB is split between a PCIe Gen5 x8 slot and two Internal Non-Volatile Memory Express (NVMe) solid-state drive (SSD) PCIe Gen4 x4 slots.

This is shown in Figure 3-1.

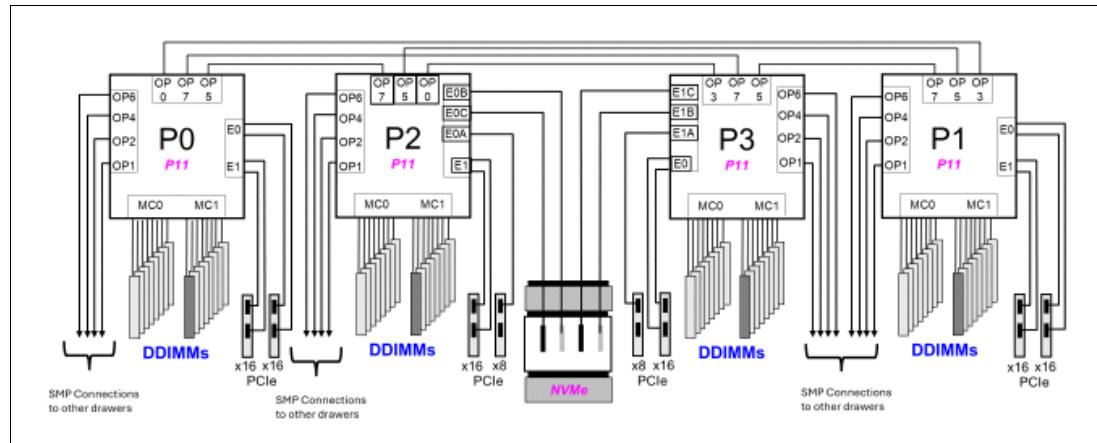


Figure 3-1 PCIe Connectivity

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

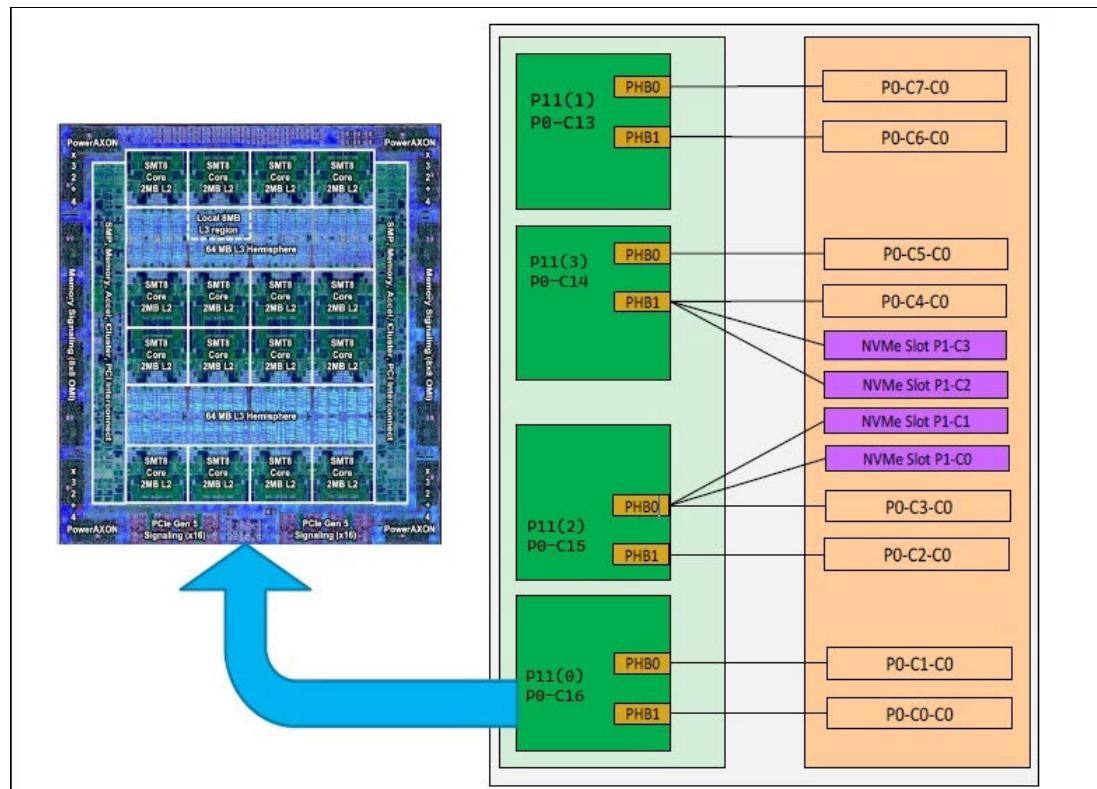


Figure 3-2 PCIe connectivity to the processor chip

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

- ▶ Six PCIe Gen 4 x16 or PCIe Gen 5 x8 - these slots are 8 lanes at Gen5 or 16 lanes at Gen4 and also support 16 lane Gen1-Gen 3 cards
- ▶ Two PCIe Gen 5 x8 - these slots are 8 lanes at Gen 5 and also support 8 lane Gen1-Gen4 cards. A physical x16 connector will be used here but only 8 lanes will be wired to it.
- ▶ Four Internal NVMe SSDs, each using a PCIe Gen 4 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 in Table 3-2 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	Sub Total (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 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 located within a PCIe extender cassette. These cassettes are directly connected to the system backplane, ensuring efficient signal routing and ease of maintenance. This design supports hot-plug capabilities and simplifies the process of upgrading or replacing PCIe cards without system downtime.

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

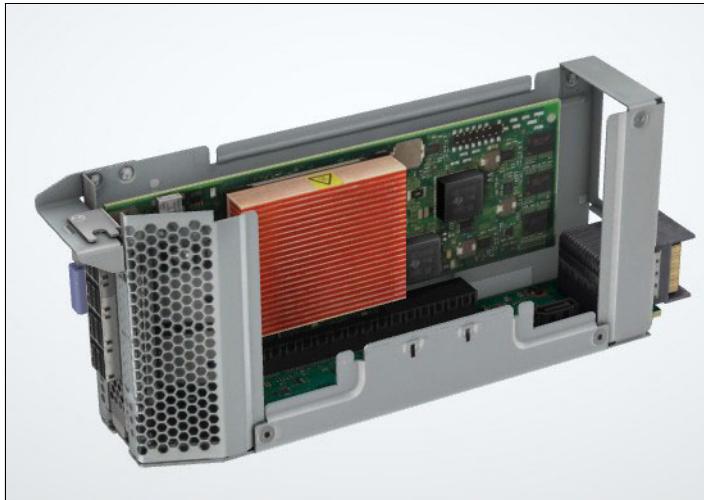


Figure 3-3 PCIe extender cassette with PCIe adapter

- ▶ The cassette locations are P0-C0 - P0-C7. The adapter locations are P0-C0-C0 - P0-C7-C0.
- ▶ All slots are of the low-profile (LP) type; that is, half-height, half-length (HHHL).
- ▶ They support Enhanced Error Handling (EEH) and can be serviced with the system power turned on. The slots also support Gen 1 to Gen 3 adapters.
- ▶ All PCIe slots support Single Root I/O Virtualization (SR-IOV) adapters
- ▶ Currently, no adapter can dissipate more than 25 watts. Special considerations are needed to exceed this limit.
- ▶ The Power E1180 server supports concurrent maintenance (hot plugging) of PCIe adapters.

Slot location and capabilities for each slot are shown in Table 3-4.

Table 3-4 Internal PCIe Slot

Location code	Description	Socket/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

Location code	Description	Socket/PHB	Slot number	SRIOV
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 shows a rear view of a Power E1180 system node illustrating the locations shown in the table. As seen in the figure, the adapter cassette is provided with three LEDs that to indicate status. The power and activity LED (green), the identify function LED for the adapter (amber and labeled C0), and fault function LED (amber) for the cassette. The server can be located by using the blue identify LED on the enclosure.

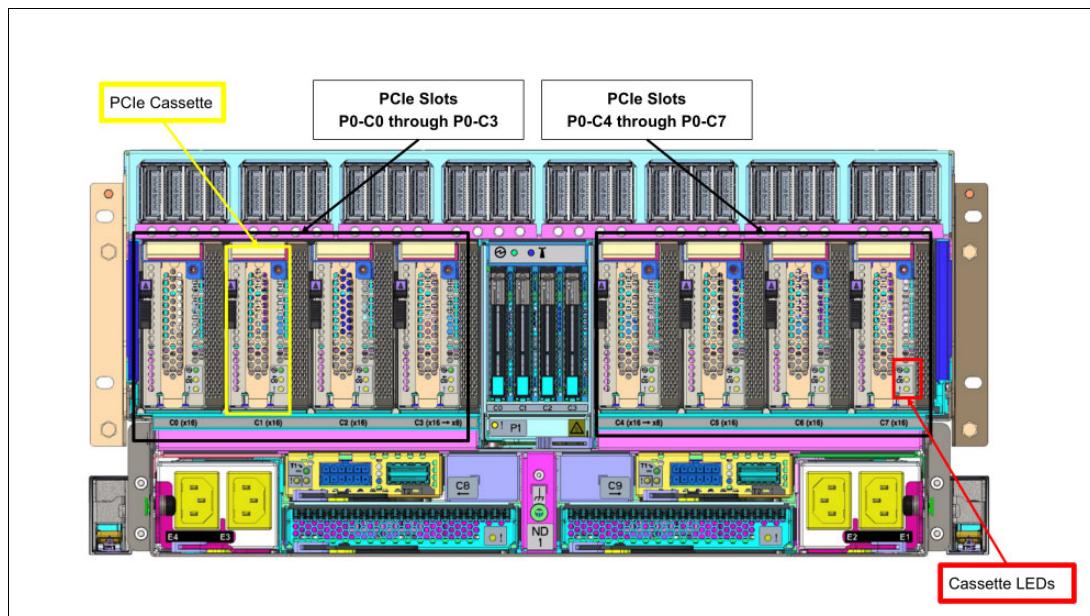


Figure 3-4 Rear view of a Power E1180 system node highlighting PCIe slots

Considerations for adapter placement

When choosing an adapter and choosing a slot to place it in consider the following:

- ▶ PCIe x16 cards are supported only in PCIe x16 slots and the slot priority is (1, 3, 7, 2, 6, 8). PCIe adapters with x8 and lower lanes are supported in all the PCIe slots and the slot priority is (1, 7, 3, 6, 2, 8, 4, 5).
- ▶ Place 8x and 16x adapters in same size slots first before mixing connector size with slot size.
- ▶ Adapters with smaller connectors are allowed in larger PCIe slot sizes, but larger connectors are not compatible in smaller PCIe slots sizes.
- ▶ For a list of adapters and their supported slots, see section 3.3, “Supported adapters” on page 84.
- ▶ As a best practice, the adapters that do not require high bandwidth should be placed in an external expansion drawer. Populating a low-latency, high-bandwidth slot with a low-bandwidth adapter is not the best use of system resources.

- As a best practice, use the high-profile version of all adapters whenever the system includes a PCIe expansion drawer. This configuration allows the user to place most of the adapters within the PCIe expansion drawer. Typically, it is advised to use node slots for the cable adapters (#EJ24) or high bandwidth PCIe Gen 4 or PCIe Gen 5 adapters

3.1.2 Internal NVMe storage subsystem

The Power E1180 supports NVMe storage technology. The NVMe subsystem is available in each of the Power E1180 system node drawers. Four internal NVMe U.2 drive slots are present in each of the system nodes in the center of the rear of the system. These slots can be configured as either 4 x 7 mm slots by using the #EBJC backplane or 2 x 15 mm slots by using the #EJBD or EJBE backplane. The slots can be independently assigned to any of the partitions. This is shown in Figure 3-5.

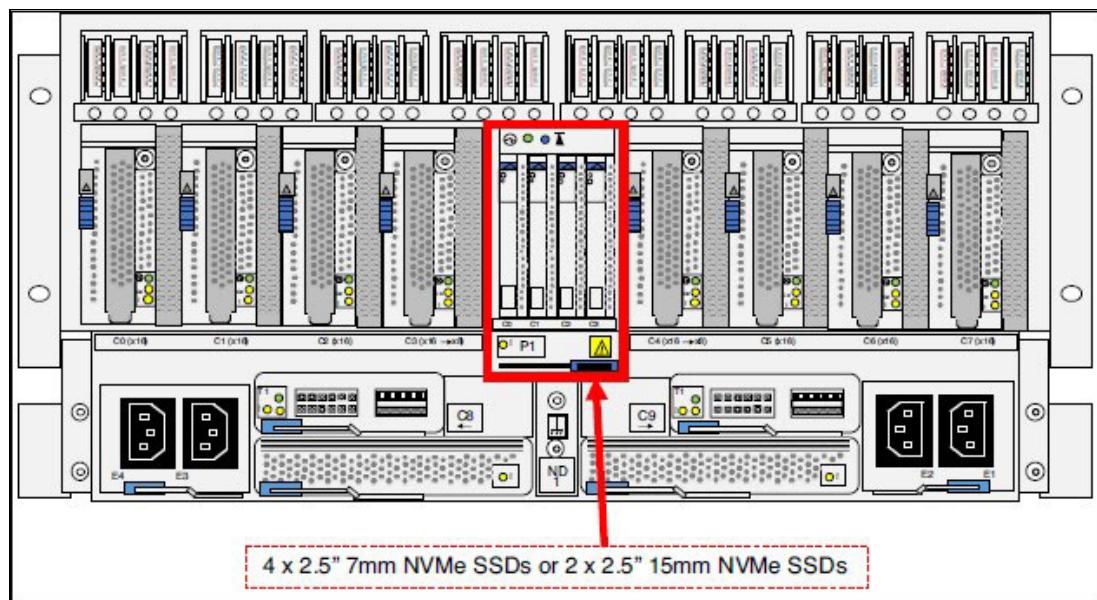


Figure 3-5 Rear view of a Power E1180 system node highlighting NVMe slots.

The internal 7 mm NVMe devices support only one namespace. The primary use of the devices is to host VIOS. None of the 7 mm devices support IBM i, IBM i support requires the EJBD or EJBE backplane and 15 mm NVMe devices.

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

Each slot is driven from an x4 PCIe connection, with two SSDs that are connected to the PHB on the P2 CPU socket and two SSDs that are connected to the PHB of the P3 CPU socket, as shown in Figure 3-6 on page 68. The buses are routed through the system board from the CPU to the NVMe slot interface card, and then up through the SSD riser card and into the SSD.

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

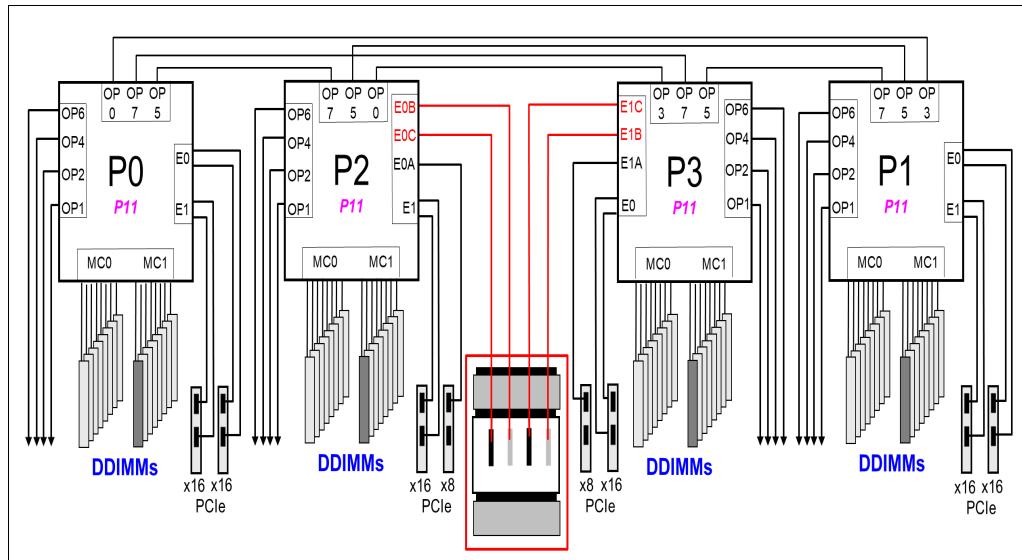


Figure 3-6 NVMe connectivity

The NVMe slots support U.2 NVMe flash SSD drives. Four 2.5-inch 7 mm form-factor SSDs or two 2.5-inch 15 mm NVMe SSDs can be used. The PCIe busses for the NVMe U.2 drives are only rated to run up to PCIe Gen4.

Important: Choose the type of NVMe SSD drives carefully based on your planned usage. Match the drive characteristics such as DWPD and number of namespaces to your intended use case.

Internal SSD plug order

For redundancy purposes, it is a best practice to distribute the NVMe drives across the system nodes if they are present as follows:

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

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

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

The location of the NVMe U.2 drive LEDs in the system is shown in Figure 3-7

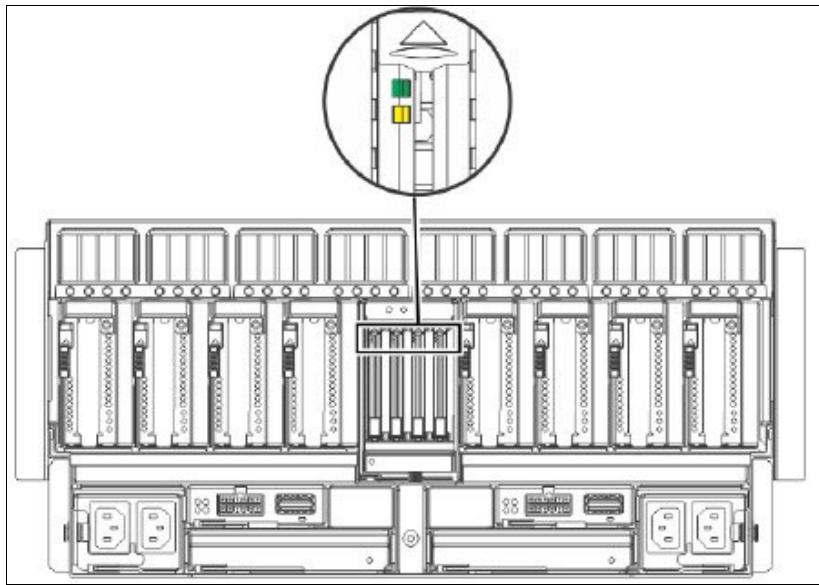


Figure 3-7 Rear view of a Power E1180 system node with the NVMe slot LED location

Maintaining NVMe drives

The lifecycle of an NVMe device is primarily determined by the number of write and erase cycles its flash memory cells can endure. Each time data is written, the memory cells wear down slightly, and over time, this wear accumulates. Modern NVMe drives use wear-leveling algorithms and error correction to extend their usable life, but they still have a finite endurance rating, typically measured in terabytes written (TBW). Check the specifications on any NVMe devices you are using to determine the drive write rating.

Concurrent maintenance of the NVMe drives is supported. In addition you can find the remaining life on a NVMe device from the LPAR that owns the device. To determine the remaining life of an NVMe device, complete the following steps:

- ▶ For an IBM AIX operating system:
 - a. From the AIX command line, 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 the remaining life for and press **Enter**.
 - d. View the Percentage of NVMe subsystem life used field.
- ▶ For a Linux operating system:
 - a. From the Linux command line, enter the following command and press **Enter**:

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

Where **nvmeX** is the resource name of the NVMe device.
 - b. View the Percentage used field.

An NVMe device approaching the end of its lifespan should be replaced. As it nears the limit of supported write cycles, write performance degrades. Eventually, the device transitions to a read-only state. When the operating system attempts to write to a read-only device, those operations fail, and the system treats the device as if it has failed. To maintain normal write functionality, the NVMe device must be replaced before it reaches this state.

3.1.3 Internal USB subsystem

The Power E1180 has no integrated USB controllers or ports in the system nodes. The USB PCIe adapter must be used for enabling USB access. The PCIe2 2-PORT USB 3.0 LP adapter (#EC6J) is supported in the system node. This adapter is typically placed in the PCIe Slot 4/5 of system node1 or node2.

A USB Cable (#EC6N) is used for connecting the #EC6J adapter port to the rear USB port in the system control unit (SCU), which is internally routed to a front-accessible USB port on the System Control Unit (SCU) see Figure 3-8.



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

The front view of the SCU in which the System USB port and Flexible Service Processor (FSP) USB ports are highlighted is shown in Figure 3-9.

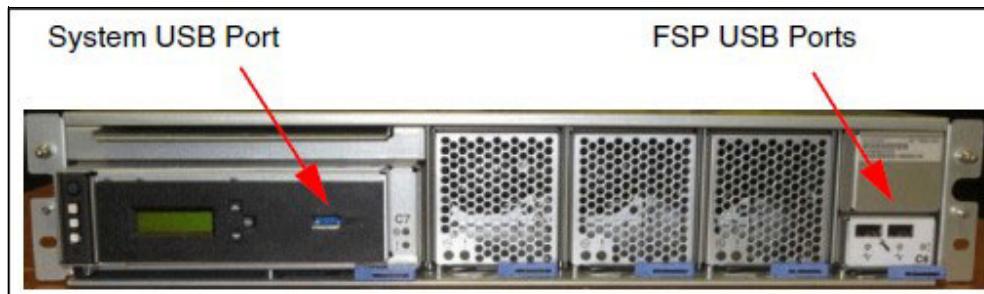


Figure 3-9 Front view of SCU showing the System USB port and FSP USB ports

The following USB devices are supported and available for the Power E1180

- ▶ #EUA5: Stand-alone USB DVD Drive with cable

Important: RDX USB External Docking Stations will not be supported with Power11 E1180 servers.

3.2 Enhancing I/O Scalability with Expansion Drawers

Adding I/O drawers to an IBM Power E1180 server significantly enhances the system's scalability, flexibility, and overall performance in enterprise environments. The Power E1180 supports PCIe Gen4 I/O Expansion Drawers, which provide increased throughput and bandwidth for connected devices, enabling the system to handle more demanding workloads and a broader range of peripherals.

These drawers expand the available PCIe slots beyond what is natively supported on the system board, allowing for additional network adapters, storage controllers, and accelerators to be integrated without compromising existing configurations. This is particularly beneficial for data-intensive applications, such as AI inferencing, high-speed networking, and large-scale database operations. Furthermore, the modular nature of I/O expansion supports a pay-as-you-grow model, aligning with dynamic business needs and reducing upfront capital expenditure.

By expanding the number and type of adapters that can be installed, these drawers enable more flexible system configurations. This is especially beneficial in virtualized environments where multiple virtual machines or containers require dedicated I/O resources. The additional capacity supports higher VM density and more granular resource allocation, improving overall system utilization. IBM's expansion drawers are designed with enterprise-grade reliability features, including hot-plug capabilities, redundant paths, and integration with PowerVM and the Hardware Management Console (HMC). This means that I/O resources can be added or serviced without system downtime, supporting continuous operations and minimizing the risk of service disruptions.

Finally, the I/O expansion capability aligns with the Power E1180's broader design goals of resilience, security, and performance. The system's architecture, including its I/O subsystem, is built to support mission-critical workloads with features like redundant paths, error recovery, and secure boot. The drawers integrate seamlessly with IBM's Hardware Management Console (HMC) and PowerVM virtualization, providing centralized control and simplified management. This makes the Power E1180 with I/O drawers an ideal platform for enterprises seeking to modernize their infrastructure while maintaining high availability and robust performance.

3.2.1 PCIe expansion card

The PCIe expansion function is enabled by using the PCIe Gen4 cable adapter (#EJ24). Each adapter occupies one of the PCIe Gen5 slots in the system node. A maximum of 8 adapters can be installed inside a system node and a maximum of 32 in a system with 4 system nodes. Each adapter card connects to either one ENZ0 fanout module or one NED24 Expansion Service Manager (ESM)). The cable adapter can be placed in any slot in the system node. See Table 3-6 on page 75 for the recommended adapter locations depending on the expansion drawer being attached. Figure 3-10 shows the PCIe Gen4 cable adapter (#EJ24).

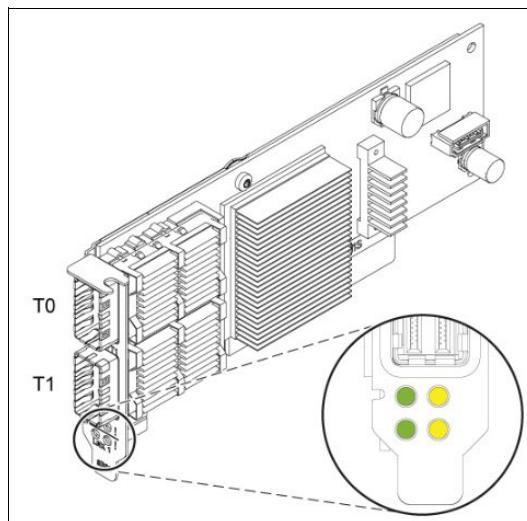


Figure 3-10 PCIe Gen4 cable adapter with FC EJ24

The EJ2A adapter card (Feature Code EJ2A, CCIN 6B99) is a PCIe Gen4 x16 cable adapter designed to provide high-speed connectivity between the system and supported I/O expansion drawers. It is a full-height, half-length card featuring two ports for connecting expansion drawer cables. The adapter can be installed in either PCIe Gen4 x16 or x8 slots; however, while using x8 slots allows for greater flexibility in connecting more drawers, it delivers lower performance compared to installation in full x16 slots.

This adapter supports the attachment of:

- One Enclosure Services Manager (ESM) in an NED24 NVMe expansion drawer
- One PCIe4 6-slot fanout module in an ENZ0 PCIe4 expansion drawer

The EJ2A adapter is a key component for extending the PCIe bus from the IBM Power E1180 system to external I/O resources, enabling scalable configurations for high-performance workloads. It is equipped with two status LEDs: a green LED that indicates link status and an amber LED used for identifying the adapter during maintenance activities.

Each EJ2A adapter supports two CXP interface ports, which can be used with either optical or copper cables to connect to PCIe Gen4 expansion drawers. A single adapter connects to one fan-out module within an expansion drawer; therefore, to utilize both fan-out modules in a drawer, two EJ2A adapters are required. In the case of connecting a NED24 expansion drawer, two adapters are also necessary for full connectivity.

Internally, the EJ2A includes a built-in PCIe switch, which enables it to extend the Power E1180's internal PCIe Gen4 bus to external I/O drawers while maintaining bandwidth and performance integrity. The adapter's physical layout and connectivity are illustrated in Figure 3-11.

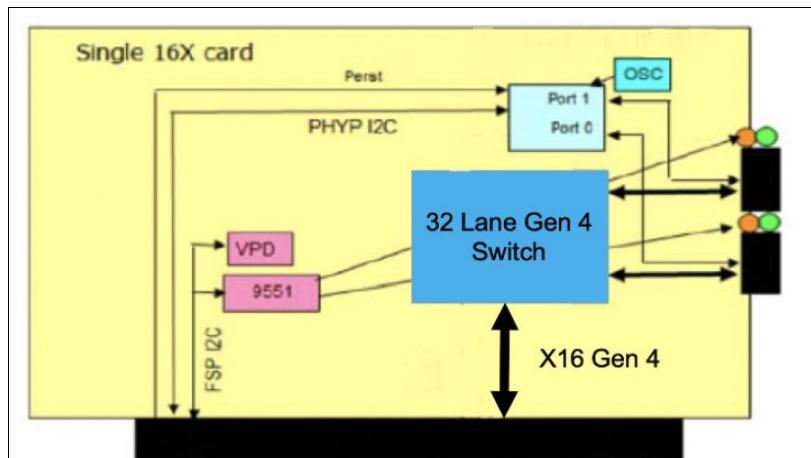


Figure 3-11 PCIe Expansion card

The number of supported cable card adapters for Power E1180 systems are listed in Table 3-5.

Table 3-5 Supported cable card adapters for Power E1180

System Name	Adapter feature code and CCIN
IBM Power E1180	PCIe4 cable adapter (FC EJ2A; CCIN 6B99)

Cables: copper or AOC optical

The choice of copper or optical cables is dependent on your specific configuration and length of cables required. Use optical AOC cables features for cables which are much thinner and

can be longer such as the features ECLX (3M optical), ECLY (10M optical). Each cable feature code contains two cables.

Active Optical Cables (AOC)

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

Figure 3-12 shows the AOC connector.

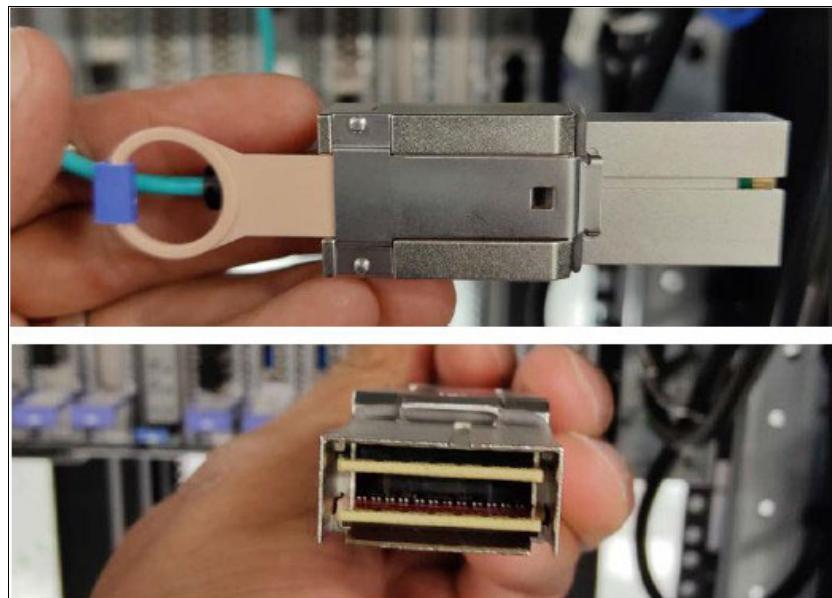


Figure 3-12 AOC connector

The following cables are available:

- ▶ (#ECLS) - 3.0M CXP x16 Copper Cable Pair for PCIe4 Expansion Drawer

This 3.0 meter cable pair connects a PCIe4 fan-out module in the PCIe Gen4 I/O Expansion Drawer to a PCIe4 Optical Converter Adapter in the system unit. There are two identical copper cables in the cable pair, each with two CXP connectors. One of the cables attaches to the top CXP port of the PCIe4 fan-out module and to the top CXP port of the PCIe4 Optical Converter Adapter. The other cable attaches to the bottom CXP ports.

- ▶ (#ECLX) - 3.0M Active Optical Cable x16 Pair for PCIe4 Expansion Drawer

The 3.0 meter active optical cable (AOC) x16 pair connects a PCIe4 module in the PCIe Gen4 I/O Expansion Drawer to a PCIe4 Optical Converter Adapter in the system unit. There are two identical cables in the cable pair, each with two CXP connectors. One of the cables attaches to the top CXP port of the PCIe4 module and to the top CXP port of the PCIe4 Optical Converter Adapter. The other cable attaches to the bottom CXP ports.

- ▶ (#ECLY) - 10M Active Optical Cable x16 Pair for PCIe4 Expansion Drawer

The 10 meter active optical cable (AOC) x16 pair connects a PCIe4 module in the PCIe Gen4 I/O Expansion Drawer to a PCIe4 Optical Converter Adapter in the system unit. There are two identical cables in the cable pair, each with two CXP connectors. One of the cables attaches to the top CXP port of the PCIe4 module and to the top CXP port of the PCIe4 Optical Converter Adapter. The other cable attaches to the bottom CXP ports.

- ▶ (#ECLZ) - 20M Active Optical Cable x16 Pair for PCIe4 Expansion Drawer

The 20 meter active optical cable (AOC) x16 pair connects a PCIe4 module in the PCIe Gen4 I/O Expansion Drawer to a PCIe4 Optical Converter Adapter in the system unit.

There are two identical cables in the cable pair, each with two CXP connectors. One of the cables attaches to the top CXP port of the PCIe4 module and to the top CXP port of the PCIe4 Optical Converter Adapter. The other cable attaches to the bottom CXP ports.

Note: Consider the following points:

- ▶ Use the 3 m cables for intra-rack installations. Use the 10 m cables for inter-rack installations.
- ▶ You cannot mix copper and optical cables on the same PCIe Gen4 I/O drawer. Both fan-out modules use copper cables or both use optical cables.

A PCIe Gen4 I/O expansion drawer with two I/O fanout modules is connected to one host system node via two PCIe4 cable adapters with four expansion drawer cables (two expansion drawer cable pairs). One pair is used for each of the PCIe4 6-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.

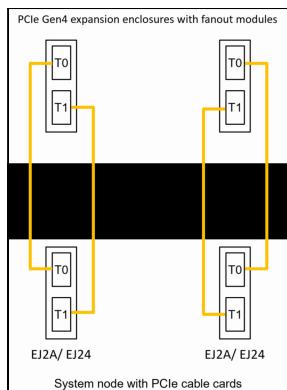


Figure 3-13 Cabling setup for PCIe Gen4 expansion drawer

Connectivity to NED24 expansion drawer

Figure 3-14 shows the connectivity to the NED24 expansion drawer. Note that both connections to the drawer need to be populated and be from the same system.

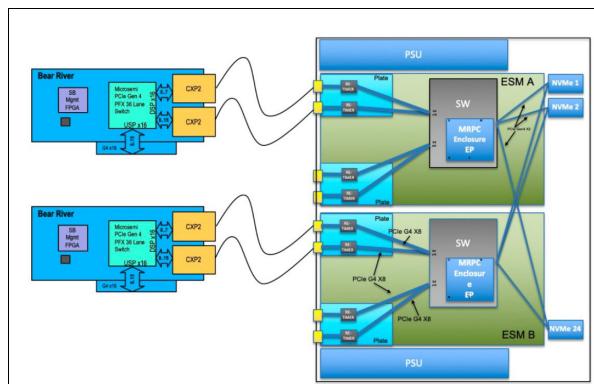


Figure 3-14 Connectivity to NED24 drawers

3.2.2 Supported I/O Drawers

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

- ▶ PCIe Gen4 I/O Expansion Drawer (#ENZ0)

The PCIe Gen4 I/O Expansion Drawer (#ENZ0) is a 4U high, 19-inch wide, PCIe Gen4 based rack mountable I/O drawer that is available as a feature of Power10 Servers. The PCIe Gen4 I/O Expansion Drawer (#ENZ0) replaces the PCIe Gen3 I/O Expansion Drawer (#EMX0). There is no upgrade path from PCIe Gen3 I/O Expansion Drawer (#EMX0) to PCIe Gen4 I/O Expansion Drawer (#ENZ0).

- ▶ NED24 NVMe Expansion Drawer (#ESR0)

The NED24 NVMe Expansion Drawer (#ESR0) 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 (FC EJ2A) in order to attach expansion drawers. Two PCIe x16 to CXP Converter Cards are used to attach each expansion drawer. The maximum number of expansion drawers attached is dependent on the number of system nodes that are configured and the number of interconnect cards installed. The maximums are shown in Table 3-6.

Table 3-6 Maximum IO Drawer configuration

Configuration	Max IO 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 IBM Power E1180, the following drawers are no longer supported:

- ▶ EMX0:

PCI Gen 3 I/O Expansion Drawer (#EMX0) is no longer supported with Power11. None of the Fanout Modules EMXF/G nor EMXH are supported within the new PCI Gen 3 I/O Expansion Drawer (#ENZ0).

- ▶ IBM EXP24SX:

The EXP24SX SAS storage enclosure (#ESLS/ ESLL) is no longer supported with Power11. For additional internal storage a new NED24 NVMe Expansion Drawer (#ESR0) is available, populated with up to 24 NVMe drives.

3.2.4 PCIe Gen 4 I/O Expansion Drawer

The PCIe Gen4 I/O Expansion Drawer (#ENZ0) is a 4U high, 19-inch wide, PCIe Gen4 based rack mountable I/O drawer that is available as a feature of Power11 Servers. The PCIe Gen 4 I/O Expansion Drawer (#ENZ0), when configured with two PCIe Fanout Modules (#ENZF), provides 12 PCIe I/O full-length, full-height Gen4 PCI slots. Each Fanout Module provides 6 PCIe slots that are labeled C0 - C5. C0 through C3 are x16 slots, and C4, C5 are x8 slots. PCIe Gen1, Gen2, and Gen3 full-high adapters are also supported.

Important: PCI Gen 3 I/O Expansion Drawer (#EMX0) is no longer supported with Power11. None of the previous generation Fanout Modules (EMXF/G and EMXH) are supported within the new PCI Gen 4 I/O Expansion Drawer (#ENZ0).

A blind swap cassette (BSC) houses the full-high adapters that go into these slots. The drawer is shipped with a full set of BSCs.

A PCIe CXP converter adapter (#EJ24) that occupies one of the PCIe Gen5 slots in the system node and a pair of Active Optical Cables (AOCs) or Copper Cable are used for system node to Fanout Module connection. Both Fanout Modules are completely independent PCIe domains and can be serviced independently. A minimum of one Fanout Module is required in ENZ0 drawer in location P0 being placed at the left side of the drawer when viewed from behind. Each PCIe Gen 4 I/O Expansion Drawer has two power supplies.

Drawers can be added to a server dynamically. Concurrent repair and adding or removing expansion drawers and PCIe adapters is done through HMC-guided menus or by operating system support utilities.

Careful balancing of I/O, assigning adapters through redundant ENZ0 expansion drawers, and connectivity to different system nodes provides high availability for I/O resources that are assigned to LPARs. Figure 3-15 shows a PCIe Gen 4 I/O Expansion Drawer front view.

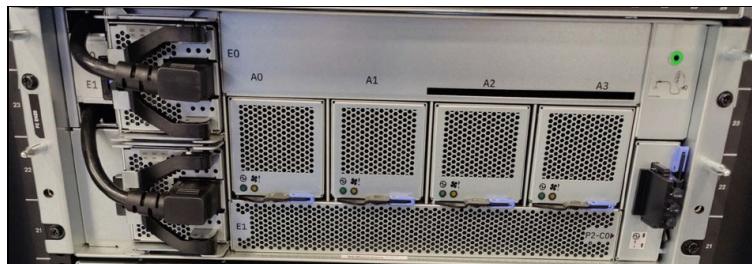


Figure 3-15 PCIe Gen 4I /O Expansion Drawer front view

Figure 3-16 shows the rear view of the PCIe Gen 4 I/O Expansion Drawer.



Figure 3-16 Rear view of a PCIe Gen 4 I/O Expansion Drawer

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 / 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)

Fanout modules supported by system

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

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

Table 3-8 Number of supported fanout modules per system

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

PCI Slots available in the PCIe Gen4 expansion drawer

Table 3.1 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

Note:

- All slots are PCIe4 slots.
- All slots support full-length, full-height adapters or short form-factor with a full-height tailstock in single-wide, generation 3, blind-swap cassettes.
- Slots C0 through C3 in each PCIe4 6-slot fanout module are PCIe4 x16 buses and slots C4 and C5 are PCIe4 x8 buses.
- All slots support enhanced error handling (EEH).
- All PCIe slots can be serviced with the power on.
- All six slots in a PCIe4 6-slot fanout module support Single Root I/O Virtualization (SR-IOV) shared mode.
- Only four FC EC2S, FC EC2U, or FC EC72 adapters can be in SR-IOV mode simultaneously per 6-slot fanout module.

Table 3-10 lists the maximum number of 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, PCIe slots and their type.

System Node	ENZ0 drawers	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 (#ESR0) is a storage expansion enclosure with 24 U.2 NVMe bays.

Important: The EXP24SX SAS Storage Enclosure is no longer supported with 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 Expansion Service Manager (ESM)



Figure 3-18 NED24 NVMe drawer rear view

Each of the 24 NVMe bays in the NED24 drawer is separately addressable and can be assigned to a specific LPAR or Virtual I/O Server (VIOS) to provide native boot support for up to 24 partitions. At the time of writing, each drawer can support up to 367 TB when utilizing the 15.3 TB flash drive. The devices that are shown in Table 3-9 are supported in the NED24 drawer at the time of writing.

Note: Both ESMs must be connected to the same server, single connections and multiple server connections are not supported.

Table 3-11 NVMe devices supported in the NED24

Feature Code	Description
ES3A	Enterprise 800GB 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/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/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/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/Linux

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

The NED24 drawer is supported in the Power E1180 using the same interconnect card that is used for the PCIe Gen 4 I/O drawer. A maximum of two NED24 NVMe expansion drawers is supported per system node in the Power E1180 with a maximum of eight in a single system. When mixing the different expansion drawers, the maximum number of drawers that are supported is based on the number of EJ24 fanout cards that are supported.

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

Table 3-12 Maximum expansion drawers

Configuration	Max IO Drawers	Max NED24 Drawers	Max PCIe Gen 4 Expansion Drawers
Single node	4	2	4
Two node	8	4	8
Three node	12	6	12
Four node	16	8	16

Note:

- ▶ Slots used for NED24 drawer attachment: C2/C5, C0/C6 or C1/C7
 - Each NED24 requires two slots. Single attached drawers or drawers shared between systems is not supported.
- ▶ Slots used for PCIe Gen 4 expansion: C1/C7, C0/C6, C2/C5, C3/C4
 - Slots C3/C4 are Gen 5 x8 - Using these slots yields performance similar to PCIe Gen 3 connectivity. Using the x16 slots improves performance.

Multipath support

The NED24 has two options for connectivity to the NVMe drives. In Mode 1 (single connect). In Mode 1, the NVMe drives are configured as single-path devices with only 1 ESM controlling each device. The switch in each of the ESMs is configured to logically drive only 12 of the 24 NVMe drives. No device failover capability is available.

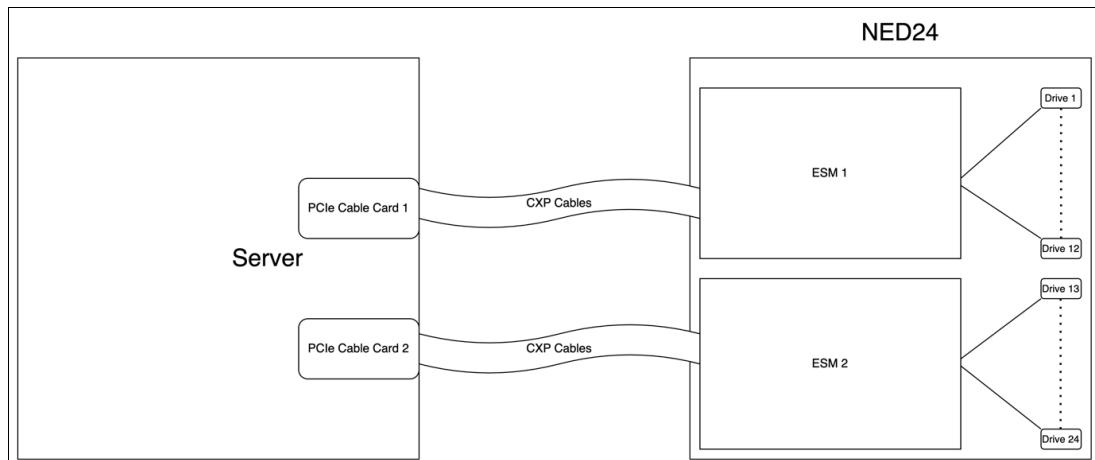


Figure 3-19 Single path mode

When running in Mode 1, OS level mirroring is recommended to avoid a single point of failure in the connection to the drives in the NED24 enclosures. See “Drive installation order” on page 82 for recommended drive locations within the drawer for availability and reliability. Both ESMs must be connected to the same server, single connections and multiple server connections are not supported.

The Power E1180 supports the NED24 drawer in Mode 2 multipath mode. The multipath functionality supports two connections for each drive as each of the ports on the multi-port drives is connected through both ESMs. This provides additional RAS and better performance.

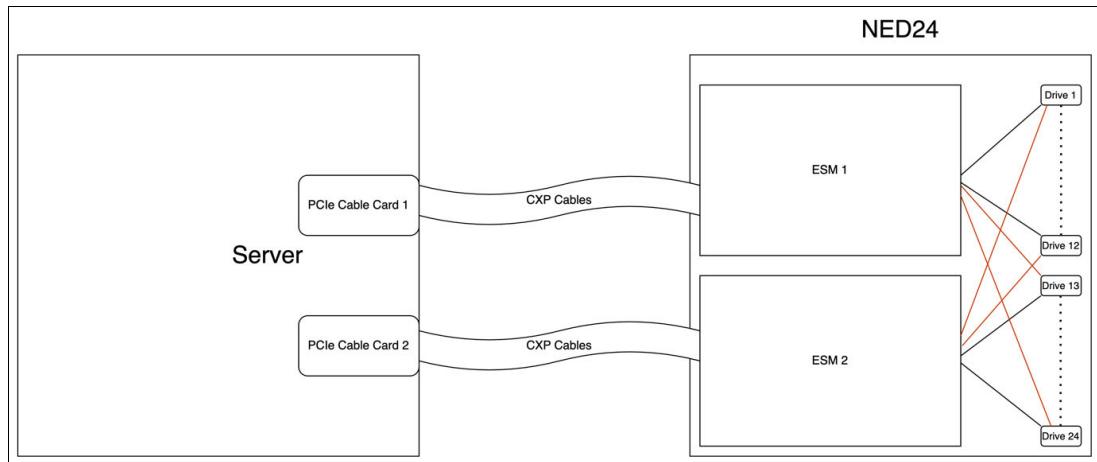


Figure 3-20 Multipath support

Multipath is automatically enabled and is enabled when the appropriate OS level is installed. Multipath support is provided in the operating systems 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

Operating System	Supported releases
AIX	<ul style="list-style-type: none"> ▶ AIX Version 7.3 with the 7300-02 Technology Level and Service Pack 7300-02-02-2420, or later ▶ AIX Version 7.2 with the 7200-05 Technology Level and Service Pack 7200-05-08-2420, or later ▶ AIX Version 7.3 with the 7300-01 Technology Level and Service Pack 7300-01-04-2420, or later
IBM i	<ul style="list-style-type: none"> ▶ IBM i 7.6, IBM i 7.5 TR 4, and IBM i 7.4 TR 10
Linux	<ul style="list-style-type: none"> ▶ SUSE Linux Enterprise Server 15.5 or later ▶ Red Hat Enterprise Linux 9.2 or later

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

Drive installation order

While there is no performance difference for drives in any of the NED24 slots, there is a recommended order for installation of drives within the enclosure. This is to provide good separation for the suggested mirroring between drives and to also provide optimal cooling and airflow within the enclosure.

Figure 3-21 shows the suggested placement for the first four drives and Table 3-14 shows the suggested placement of all drives.

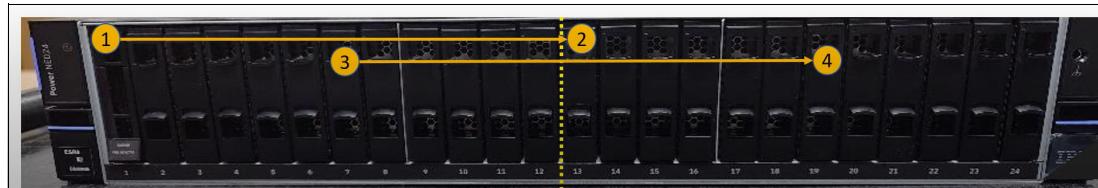


Figure 3-21 Drive installation order

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 Ge 4 Cable Adapter (#EJ24) can be in any of the PCIe adapter slots in a Power E1180 system node. However, it is a best practice to first populate the PCIe adapter slots as discussed in section 3.2.2. If an I/O expansion drawer is present, the PCIe Gen4 cable adapter (#EJ24) must be given the highest priority.

Each processor module drives two PCIe Gen5 slots, and all slots are equal regarding their bandwidth characteristics. If you first use the slots 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 the processor modules.

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

The correct cabling is as follows:

- ▶ Top port T0 of the EJ24 adapter goes to the top port T0 of the Fanout Module of the PCIe Gen4 Expansion drawer or top port of an Expansion Service Manager (ESM)
- ▶ Bottom port T1 of the EJ24 adapter goes to the bottom port T1 of the Fanout Module of the PCIe Gen4 Expansion drawer or bottom port of an Expansion Service Manager (ESM)

Important: The two cables used to connect a host system node's PCIe4 cable card to the PCIe Gen4 I/O expansion drawer's I/O module or NED24 ESM must be a cable pair that are of the same type and length.

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

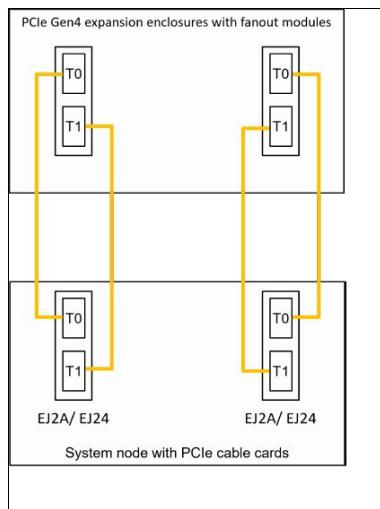


Figure 3-22 connecting the expansion drawer cables from two adapter cards to two fanout modules

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

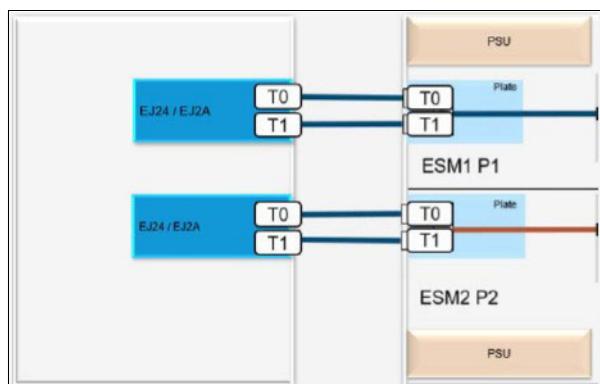


Figure 3-23 NED24 connectivity to the system node

3.3 Supported adapters

In this section, we provide more information about supported PCIe I/O adapters on Power E1180 server as on system general availability. This list tends to update frequently as more PCIe I/O adapters are tested and certified.

For more information regarding different supported storage combinations please visit the [IBM System Storage Interoperation Center \(SSIC\)](#):

In addition the Power E1180 sales manual provides complete list of all supported devices and their feature codes.

3.3.1 Supported Fibre Channel adapters

The supported Fibre Channel adapters are listed in Table 3-15

Table 3-15 supported Fibre Channel adapters on Power E1180

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

In addition to above table the following two links supply information on which adapters support replaceable SFPs and which adapter does not.

- ▶ Transceiver component (SFP) **not** replaceable for Power Fibre Channel and Ethernet adapter features - <https://www.ibm.com/support/pages/node/6593855>
- ▶ Transceiver component (SFP) is replaceable for following Power Fibre Channel and Ethernet adapter features - <https://www.ibm.com/support/pages/node/7148984>

3.3.2 Network and RoCE adapters

The supported Network adapters are listed in Table 3-16

Table 3-16 Supported Network adapters on 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 / Linux RoCe - AIX / Linux	No
EC2U	58FB	PCIe3 2-port 25/10 Gb NIC & RoCE SFP28 adapter full-height	NIC - IBM i / AIX / VIOS / Linux RoCe - AIX / Linux	Yes
EC71	2CF9	PCIe4 2-port 25/10/1 Gb RoCE SFP28 adapter low-profile	IBM i / AIX / VIOS / Linux	No
EC72	2CF9	PCIe4 2-port 25/10/1 Gb RoCE SFP28 adapter full-height	IBM i / AIX / VIOS / Linux	Yes
EC73	2CF8	PCIe4 2-port 25/10/1 Gb RoCE SFP28 with Crypto adapter low-profile	At the time of writing this Redbook the 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 Redbook the information is N/A	Yes
EC75	2CFB	PCIe4 2-port 100 GbE RoCE x16 adapter low-profile	IBM i / AIX / VIOS / Linux	No
EC76	2CFB	PCIe4 2-port 100 GbE RoCE x16 adapter full-height	IBM i / AIX / VIOS / Linux	Yes
EC85	EC2C	PCIe5 x16 2-port 200 GbE RoCE adapter	At the time of writing this Redbook the information is N/A	No
EN24	EC2A	PCIe4 x16 4-port 25/10/1 GbE RoCE SFP28 adapter low-profile	IBM i / AIX / VIOS / Linux RoCE function available only on IBM i	No
EN26	EC2A	PCIe4 x16 4-port 25/10/1 GbE RoCE SFP28 adapter full-height	IBM i / AIX / VIOS / Linux RoCe function available only on IBM i	Yes

Feature Code	CCIN	Description	OS support	Supported in ENZ0
EN2W	2F04	PCIe3 4-port 10 GbE adapter full-height	IBM i (through VIOS) / AIX / VIOS / Linux	Yes
EN2X	2F04	PCIe3 4-port 10 GbE adapter low-profile	IBM i (through VIOS) / AIX / VIOS / Linux	No

3.3.3 Network and RoCE transceivers and cables

Table 3-17 provides information related to all SFPs and Cables used in NIC and RoCE adapters supplied for Power11 systems

Table 3-17 Supported transceivers and cables

Feature Code	Name	Connection	Speed in Gbs	Adapter
EB48	1Gb RJ45 Transceiver	SFP/RJ45	1	EC2T/U EC71/2/3/4 EN24/6
EB46	10GbE Optical Transceiver SR	SFP+/LC MMF	10	EC2T/U EC71/2/3/4 EN24/6
EB47	25GbE Optical Transceiver	SFP28/LC MMF	25	EC2T/U EC71/2/3/4 EN24/6
EB57	40GbE Base-SR4 Transceiver	QSFP+/ MPO MMF	40	EC75/6 EC85/6
EB59	100GbE Base-SR4 Transceiver	QSFP28/MPO MMF	100	EC75/6 EC85/6
EB49	QSFP28 to SFP28 Connector	QSFP28	25	EC75/6 EC85/6
EB4J	0.5M 25GbE Passive CU Cable SFP+	SFP+	25	EC2T/U EC71/2/3/4 EN24/6
EB4K	1M 25GbE Passive CU Cable SFP+	SFP+	25	EC2T/U EC71/2/3/4 EN24/6
EB4M	2M 25GbE Passive CU Cable SFP+	SFP+	25	EC2T/U EC71/2/3/4 EN24/6
EB4P	QSFP28 Passive Copper 100GbE to SFP28 4x25Gbe Split Cable	QSFP29->SF P28	25	EC2T/U EC71/2/3/4 EN24/6
EN01	1M 10GbE CU Active Twinax Cable SFP+	SFP+	10	EC2T/U EC71/2/3/4 EN24/6

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

3.3.4 SAS and Tape adapters

Supported SAS adapters are listed in Table 3-14

Table 3-18 Supported SAS adapters on 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 / VIOS	Yes
EJ2C	57F2	PCIe3 x8 SAS quad-port 12 Gb tape low-profile adapter	IBM i / AIX / VIOS	No

3.3.5 NVMe

The supported NVMe drives are listed in Table 3-19

Table 3-19 Supported NVMe adapters on Power E1180

Feature Code	CCIN	Description	OS support	Placement
EC5J	59B4	Mainstream 800 GB SSD NVMe U.2 module	AIX/Linux	System Node
EC7Q	5948	800GB Mainstream NVMe U.2 SSD 4k module	AIX/Linux	System Node
ECT9	5941	15.3 TB Mainstream NVMe U.2 SSD 4k module	AIX/Linux	System Node and NED24
ECTB	5941	15.3 TB Mainstream NVMe U.2 SSD 4k module	IBM i	System Node
EKF2	5B53	Enterprise 800GB SSD PCIe4 NVMe U.2 module	IBM i	NED24
EKF3	5B52	Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	AIX/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/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/Linux	NED24
EKF8	5B50	Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module	IBM i	NED24
EKF9		Enterprise 800GB SSD PCIe4 NVMe U.2 module	AIX/Linux	NED24
ES3A	5B53	Enterprise 800GB SSD PCIe4 NVMe U.2 module	IBM i	NED24
ES3B	5B52	Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	AIX/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/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/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 800GB SSD PCIe4 NVMe U.2 module	AIX/Linux	NED24
ES4B		Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	AIX/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/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/Linux	System Node
ES4G		Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module	IBM i	System Node
ES5A	5B53	Enterprise 800GB SSD PCIe4 NVMe U.2 module	AIX/Linux	NED24
ES5B	5B53	Enterprise 800GB SSD PCIe4 NVMe U.2 module	IBM i	NED24
ES5C	5B52	Enterprise 1.6 TB SSD PCIe4 NVMe U.2 module	AIX/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/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/Linux	NED24
ES5H	5B50	Enterprise 6.4 TB SSD PCIe4 NVMe U.2 module	IBM i	NED24

Differences between mainstream and enterprise SSD

In this section, you will learn about 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 Drive Writes Per Day (DWPD), typically using durable MLC NAND (formerly called eMLC). Mainstream SSDs usually support ~1 DWPD, making them suitable for read-heavy or less write-intensive applications.

- ▶ **Over-Provisioning and Performance:**

Enterprise SSDs have 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 impacts performance and lifespan during write-heavy operations.

► **Cost:**

Mainstream SSDs are more cost-effective due to lower over-provisioning and endurance specs, offering lower cost per GB but reduced performance in write-intensive tasks.

► **Mixing in RAID Arrays:**

IBM systems do not allow mixing of mainstream and enterprise SSDs in RAID arrays. The data striping 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. The drive should be replaced 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 may result in degraded or failed write operations.

For more information see:

<https://www.ibm.com/docs/en/power10?topic=drives-mainstream-solid-state>
<https://www.ibm.com/docs/en/power11?topic=drives-mainstream-solid-state>

3.3.6 USB

The supported USB adapter is listed in Table 3-20

Table 3-20

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

3.3.7 Cryptographic adapters

The supported crypto adapters are listed in Table 3-21

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 / Linux	Yes
EPG4	C138	4770 Cryptographic Coprocessor low-profile	IBM i / AIX / Linux	No
EPG6	C138	4770 Cryptographic Coprocessor full-height	IBM i / AIX / Linux	Yes

3.4 Other device support

This section provides an overview of the available media features with Power E1180

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 help make the installation or replacement of new drives efficient. 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 offer current level 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

The following are the options available with the enclosure:

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

For more information and technical description reference these [sales manual pages](#).

Stand-alone USB DVD drive (FC EUA5)

The stand-alone USB DVD drive (FC EUA5) is an optional, stand-alone 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 with all operating systems. For an OS to use the drive, the USB adapter must be assigned to the corresponding 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 is included (P/N 32N1311). The USB extension cable is to be used when there are no safe, flat spots available in the rack. This cable allows the drive to reach the floor. Alternative or additional extension cables are not supported as the total USB cable length can be no longer than 3 meters.

RDX Removable Hard Disk Cartridge

The RDX removable Hard Disk Cartridge is not supported with Power E1180. It can not be purchased to be used with this server.

There are alternative solutions to an RDX based Backup strategy that are being offered. Clients should consider:

- ▶ Cloud based Backup Services or
- ▶ On-Premise Entry Tape Drives and Libraries



Artificial Intelligence Support

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

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

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

This chapter contains the following topics:

- ▶ On chip support
- ▶ AI Acceleration with IBM Spyre adapter
- ▶ Solutions

4.1 On chip support

With the upcoming availability of the Spyre card on IBM Power-based servers, starting with systems powered by the new IBM Power11 processor, organizations will be able to take advantage of a new class of hardware acceleration designed to support a wide spectrum of enterprise workloads. The Spyre card will deliver significant performance benefits, particularly for compute-intensive and AI-driven use cases, while aligning with the architectural strengths of the IBM Power platform. The Spyre card is not a GPU, but a dedicated hardware developed by IBM Research® Lab to provide superior acceleration.

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

Beginning with IBM Power10 chip and improved in IBM Power11, each core integrates four Matrix Math Accelerators (MMAs) capable of supporting a wide range of AI inference workloads directly on the CPU. This innovation enables customers to run AI models natively on IBM Power cores without requiring a discrete GPU. The architecture is particularly well-suited to “traditional AI” use cases, such as fraud detection, text extraction, document analysis, domain adaptation via Retrieval-Augmented Generation (RAG), pattern recognition, forecasting, and image/video/audio processing.

The IBM Power11 processor further enhances these capabilities. While not always matching the raw throughput of high-end GPUs in certain generative AI (GenAI) scenarios, IBM Power11 technology provides excellent performance for tasks such as entity extraction, translation, summarization, and classification, all while offering lower energy consumption and improved data protection. By enabling AI workloads to run closer to where the data resides, IBM Power11 processor-based systems support secure, efficient, and scalable deployment of AI without unnecessary data movement.

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

4.2 AI Acceleration with IBM Spyre adapter

Improvements to Power11 processor core strength and system capacity boost the performance of the MMA (Matrix-Math Accelerator) for inferencing workloads. Furthermore, with IBM Spyre accelerator into Power11 provide additional AI inferencing capabilities. Working together, IBM Power processors and the IBM Spyre accelerator will enable the next generation infrastructure to scale demanding AI workloads for businesses.

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

Thanks to the advanced virtualization and workload consolidation capabilities of IBM Power processor-based servers, Spyre-based applications can be co-located with other mission-critical services within the same physical server. This allows AI inference engines, real-time analytics, or data preprocessing workloads that leverage IBM Spyre card to run in close proximity to databases or transactional systems hosted in LPARs or containers on the same server. This architectural proximity reduces latency, improves throughput, and eliminates the overhead typically associated with cross-node or cross-platform communication.

The IBM Spyre card will be uniquely positioned as the one of the only AI solutions that combines:

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

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

4.2.1 Deploying AI in the Enterprise

Artificial intelligence (AI) is transforming enterprise operations across industries, 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 far from trivial. It demands not only the right algorithms and models but also a robust, secure, and scalable infrastructure tailored to diverse AI workloads. IBM's AI-optimized Power platform provides a unique approach to meet this challenge, 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 embedded in IBM Power11 processors. This includes Matrix Math Assist (MMA) engines and SIMD vector instructions, combined with high memory bandwidth-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 additional hardware. It is particularly effective for traditional machine learning models, data warehouse analytics, and vector database operations (like RAG-style 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 significantly. 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 Acceleration

To address this, IBM introduced IBM Spyre, a dedicated off-chip accelerator designed specifically for AI-intensive workloads. Compared to the on-chip accelerators in Power11, Spyre offers significantly higher throughput, parallelism, and model capacity, thanks to its dedicated memory architecture, optimized data paths, and ability to offload processing from the CPU. Spyre is ideal for large-scale transformer models, such as those used in generative AI applications-translation, summarization, sentiment analysis, and more. 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. Early AI workloads can begin on existing Power processor-based servers with no added hardware, leveraging the built-in accelerators for cost-efficiency. As demands grow, the Spyre accelerators can be seamlessly integrated to boost performance without requiring a platform change or software rewrite. This flexibility protects existing investments while enabling future scaling this is shown in Figure 4-1.

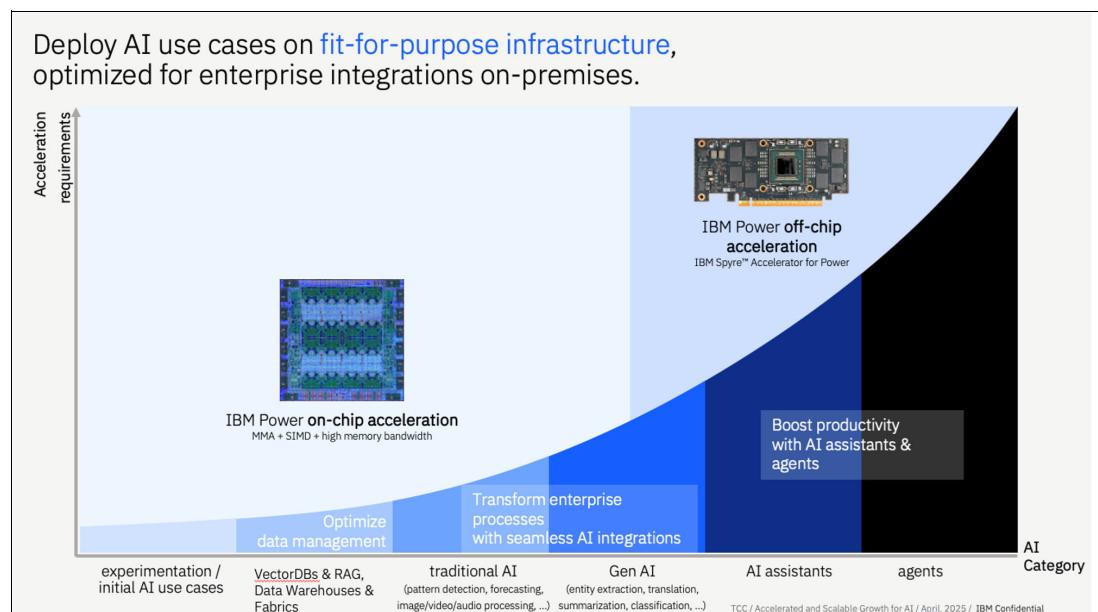


Figure 4-1 IBM's fit for purpose AI architecture for Power

In high-value use cases like 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 need to manage vast knowledge graphs or context windows, which exceed the practical limits of CPU-based inference alone. IBM Spyre is built precisely for such scenarios, and when paired with Power11 it provides a hybrid architecture that is secure, performant, and manageable within enterprise IT constraints.

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

In conclusion, the approach to enterprise AI infrastructure stands out for its modularity and adaptability. Whether you're starting with lightweight models or scaling to multi-modal AI agents, IBM Power processor-based servers allow 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 gain the ability to move from AI exploration to enterprise-wide transformation-all on a platform built for the future of AI.

4.2.3 IBM Spyre adapter

The IBM Spyre Accelerator is a purpose-built enterprise-grade accelerator offering scalable capabilities for complex AI models and generative AI use cases. The new accelerator features 32 individual accelerator cores onboard, and each Spyre is mounted on a PCIe card. Jointly designed by IBM Research and IBM Infrastructure, Spyre's architecture is designed for more efficient AI computation. Notably, the chip will send data directly from one compute engine to the next, leading to an efficient use of energy. This family of processors also uses a range of lower precision numeric formats (such as int4 and int8), to make running an AI model more energy efficient and far less memory intensive.

IBM Spyre Is designed with a system-on-a-chip architecture optimized for enterprise AI workloads. Spyre is implemented in 5nm technology with a high performance and low-power design with 75W consumption and provides the following capabilities:

- 32 low-power AI cores.
- Supports multi-precision for inference and training: FP16/8, INT8/4.
- Enabled for Foundation Models and enabled in the Red Hat software stack (OCP AI)
- Supports popular AI Framework and libraries (PyTorch, vLLM)

Figure 4-2 shows the IBM Spyre adapter.

IBM Spyre™ Accelerator PCIe attached card SoC implements IBM's leadership innovations in low-precision AI arithmetic and algorithms	System on a chip architecture optimized for enterprise AI workloads. 32 low-power AI cores.	Supports multi-precision for inference & training: FP16/8, INT8/4 . Enabled for Foundation Models .
	Enabled in the Red Hat software stack (OCP AI) Supports popular AI Framework and libraries (PyTorch , vLLM)	Implemented in 5nm technology. High performance and low-power design with 75W consumption .

Figure 4-2 The IBM Spyre adapter

This dedicated, enterprise-grade AI acceleration chip sits on a 75W PCIe adapter, surrounded by 128 GB of LPDDR5 memory - which is large memory to hold a wide variety of LLMs in support of the heterogeneous workloads that are typically seen on IBM Power. As a single Spyre adapter will not provide enough compute capacity for most use cases, the IBM solution will utilize current I/O expansion technology to attach a cluster of eight Spyre adapters in a single I/O expansion drawer, creating a logical cluster. The firmware on those eight cards will coordinate the distribution of compute, and transfer of data amongst the cards. Making the clusters appear to the software as one high-performance compute engine, with 1 TB of memory and 1.6TB/s of memory bandwidth. this is shown in Figure 4-3 on page 98

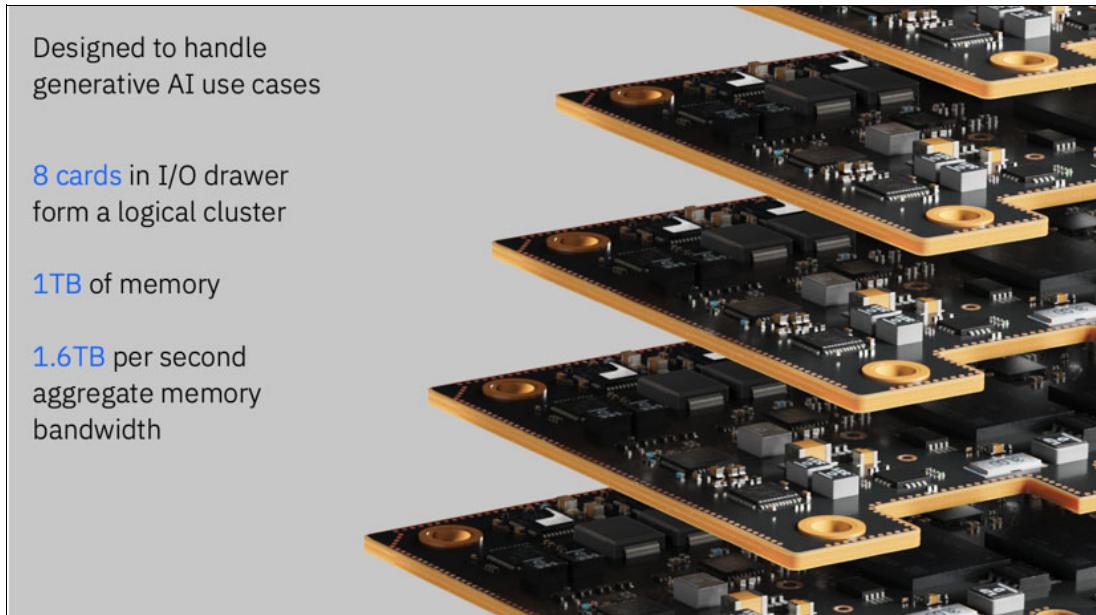


Figure 4-3 *Spyre cluster in an expansion drawer*

This expansion drawer with Spyre adapters can be attached to any of the announced IBM Power11 servers, the number of supported drawers is dependent on the Power11 server that it is attached to.

At announcement, Spyre is planned to be available in a fixed configuration which includes eight Spyre adapters installed in the PCIe Gen4 I/O expansion drawer. The Spyre expansion drawer can be attached to any Power11 server: S1122, S1124, E1150, E1180. One expansion drawer will be supported on all models except the E1180 which will support up to two. New I/O drawer components such as power supplies and fan-out modules will be utilized to support the additional power requirements of the Spyre adapter configuration.

The initial supported software and hardware will be:

- FW1110.10
- HMC1111
- RHEL 9.6
- Spyre software stack container
- OpenShift AI (Tech Preview 4Q 2025, GA 1Q 2026)

4.3 Solutions

Artificial Intelligence is no longer an emerging trend, but it is a foundational pillar of competitive advantage. According to IBM Institute for Business Value, at the time of writing, about 72% of top-performing CEOs identify advanced generative AI as essential to their future success. However, only a fraction of enterprises have successfully moved beyond pilots and proofs of concept (PoCs) into full production deployments.

IBM Power11 processor-based server provides the AI-optimized infrastructure and integrated software stack 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 GenAI into their ERP system;
- ▶ Financial institutions can accelerate anomaly detection by four times, while reducing TCO;
- ▶ Utility providers can reduce equipment downtime, and lower energy consumption using AI-powered visual inspection.

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

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

- ▶ Finance and ERP:
Fraud detection, AML, order processing, invoice compliance;
- ▶ Healthcare:
Medical transcription, claims-EHR matching, diagnostic imaging, assistant bots;
- ▶ IT and Development:
Predictive ITOps, code assistants (RPG, Ansible), transcription, documentation;
- ▶ Cross-Industry:
Document digitization, summarization, KYC, visual quality inspection, customer churn prediction, business intelligence;

Each use case leverages 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

While the promise of AI is great, common enterprise challenges persist, as described in the following table:

Table 4-1 challenges to scale AI

Barrier	Description
Data complexity	Fragmented formats, distributed sources, lack of standardized connectors
AI integration difficulty	Legacy workflows, lack of AI skills, software maturity
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 Power addresses these challenges by offering a complete AI platform, integrated into existing systems, with hardware acceleration, data fabric, and AI assistants.

4.3.2 IBM Power AI Portfolio: from Data to Business Value

To scale AI in production, IBM Power provides an AI-ready portfolio including:

- ▶ AI-Ready Infrastructure:
Built-in accelerators, support for hybrid and cloud-native deployment (e.g., PowerVS);
- ▶ Optimized software stack:
Red Hat, IBM, open source, and certified ISV AI solutions;
- ▶ data fabric:
Secure, unified access to enterprise data across silos;
- ▶ AI assistants and agents:
Tools to enhance developer productivity, automate ITOps, and streamline business workflows.

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

4.3.3 Strategic outcomes enabled by IBM Power

The following table describes outcome examples enabled by the implementation of AI-solutions on IBM Power processor-based server:

Table 4-2 outcomes examples enabled by IBM Power

Objective	Outcome example
Transform processes	Accelerate ERP or financial workflows integrated with AI services
Boost productivity	Roll out new application features faster with GenAI code assistants
Optimize data use	Reduce TCO and improve resource efficiency via 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-proven architecture, IBM Power is uniquely positioned to help organizations move from experimentation to real AI-driven transformation. Its scalable ecosystem (spanning 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 lies in its ability to deliver AI capabilities without requiring external GPUs. Unlike conventional assumptions that all AI workloads demand dedicated GPU acceleration, IBM Power11 leverages advanced on-core accelerators integrated directly within each processor core, combined with a system-wide infrastructure design 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 including resilience, reliability, performance, security, and sustainability. By eliminating or reducing dependency on external GPUs, organizations

benefit from simplified infrastructure, reduced energy consumption, and lower total cost of ownership (TCO), all without compromising AI readiness. The IBM Power11 platform is purpose-built to support scalable and secure AI solutions across diverse industry environments.



Automation and Management

Managing IBM Power 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. HMC enables administrators to manage logical partitions (LPARs), perform firmware updates, and monitor system health. It provides both a graphical interface and a REST API, allowing integration with automation tools for more 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 tailored for AIX, IBM i, and Linux on Power. Terraform complements Ansible by enabling declarative provisioning of infrastructure components, including Power Virtual Servers and PowerVC-managed resources. Together, they allow teams to automate the full lifecycle of infrastructure – from provisioning to configuration – ensuring consistency and reducing manual effort.

PowerVC (Power Virtualization Center) 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 and workloads. By combining HMC, Ansible, Terraform, and PowerVC, organizations can build a robust, automated, and scalable management framework for their IBM Power infrastructure.

This chapter contains the following topics:

- ▶ Hardware Management Console overview
- ▶ Ansible
- ▶ Terraform
- ▶ PowerVC

5.1 Hardware Management Console overview

The hardware management console (HMC) is a hardware or virtual appliance that is used to configure and manage your systems. The HMC connects to one or more managed systems and provides capabilities for the following primary functions:

- ▶ Provide systems management functions, including the following examples:
 - Power off
 - Power on
 - System settings
 - Capacity
 - on Demand
 - Enterprise
 - Pools
 - Shared Processor Pools
 - Performance and Capacity Monitoring
- ▶ Starting Advanced System Management Interface (ASMI) for managed systems
- ▶ Deliver virtualization management through support for creating, managing, and deleting Logical Partitions, Live Partition Mobility, Remote Restart, configuring SRIOV, managing Virtual IO Servers, dynamic resource allocation, and operating system terminals.
- ▶ Acts as the service focal point for systems and supports service functions, including call home, dump management, guided repair and verify, concurrent firmware updates for managed systems, and around-the-clock error reporting with Electronic Service Agent for faster support.
- ▶ Provides appliance management capabilities for configuring network, users on the HMC, and updating and upgrading the HMC.

5.1.1 HMC Options

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

Restriction: The 7063-CR1 is not supported by the Power11 servers.

HMC 7063-CR2

The 7063-CR2 IBM Power HMC (see Figure 2-1) is a second-generation Power processor-based HMC. It includes the following features:

- 6-core IBM Power9 130W processor chip
- 64 GB (4x16 GB) or 128 GB (4x32 GB) of memory
- 1.8 TB of internal disk capacity with RAID1 protection
- 4-ports 1 Gbps Ethernet (RJ-45), 2-ports 10 Gbps Ethernet (RJ-45), two USB 3.0 ports (front side) and two USB 3.0 ports (rear side), and 1 Gbps IPMI Ethernet (RJ-45)
- Two 900W power supply units
- Remote Management Service: IPMI port (OpenBMC) and Redfish application programming interface (API)

The base Warranty is 1-year 9x5 with available optional upgrades. A USB Smart Drive is not included.

Figure 5-1 is a picture of 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 are withdrawn from marketing

Virtual HMC

Initially, the HMC was sold only as a hardware appliance, including the HMC firmware installed. However, IBM extended this offering to allow the purchase of the hardware appliance or a virtual appliance that can be deployed on ppc64le architectures or x86 platforms.

Any customer with a valid contract can download the HMC from the IBM Enterprise Storage Server®, or it can be included within an initial Power E1180 order.

The virtual HMC supports the following hypervisors:

- ▶ On x86 processor-based servers
 - KVM
 - Xen
 - VMware
- ▶ On Power processor-based servers
 - IBM PowerVM

The following minimum requirements must be met to install the virtual HMC:

- 16 GB of Memory
- 4 virtual processors
- 2 network interfaces (maximum 4 allowed)
- 1 disk drive (500 GB available disk drive)

For an initial Power E1180 order with the IBM configurator (e-config), the HMC virtual appliance can be found by selecting **Add software → Other System Offerings** (as product selections) and then choosing:

- 5765-VHP for IBM HMC Virtual Appliance for Power V10 or
- 5765-VHX for IBM HMC Virtual Appliance x86 V10

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 requirements, see [Installing the HMC virtual appliance](#).

5.1.2 BMC network connectivity rules for 7063-CR2 HMC

The 7063-CR2 HMC features a baseboard management controller (BMC), which is a specialized service processor that monitors the physical state of the system by using sensors. OpenBMC that is used on 7063-CR2 provides a graphical user interface (GUI) that can be accessed from a workstation that includes network connectivity to the BMC. This connection requires an Ethernet port to be configured for use by the BMC.

The 7063-CR2 provides two network interfaces (eth0 and eth1) for configuring network connectivity for BMC on the appliance.

Each interface maps to a different physical port on the system. Different management tools name the interfaces differently. The HMC task **Console Management → Console Settings → Change BMC/IPMI Network Settings** modifies only the Dedicated interface.

The BMC ports are listed in Table 5-1.

Table 5-1 BMC ports

Management tool	Logical port	Shared/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 BMC/IPMI Network settings)	lan2	Dedicated	Management port only

Figure 1-15 shows the BMC interfaces of the HMC.

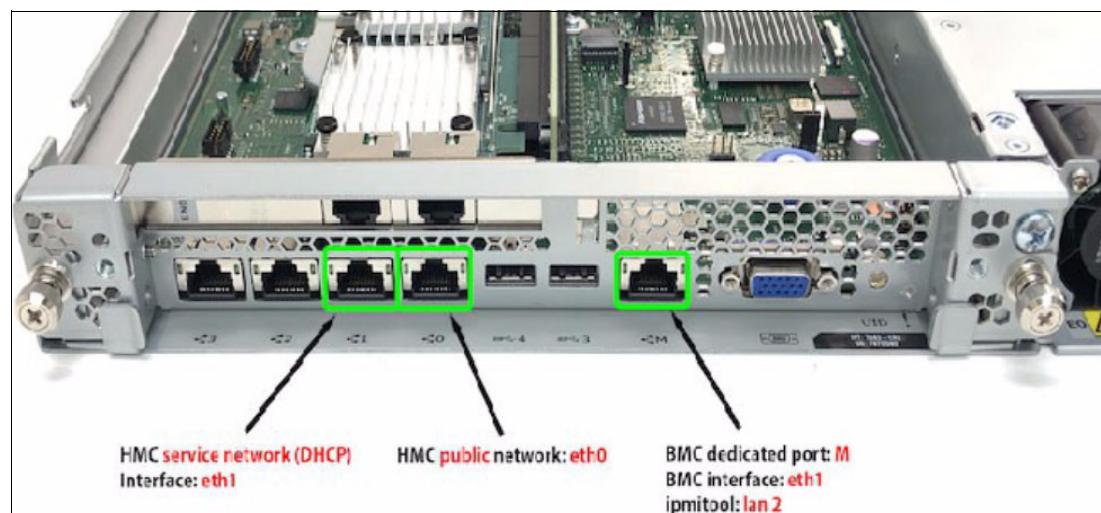


Figure 5-2 BMC interfaces

The main difference is that the shared and dedicated interface to the BMC can coexist. Each has its own LAN number and physical port. Ideally, the customer configures one port, but both can be configured. The rules for connecting Power servers to the HMC remain the same as for previous versions.

5.1.3 High availability HMC configuration

For the best manageability and redundancy, a dual HMC configuration is suggested. This configuration can be two hardware appliances, or one hardware appliance and one virtual appliance or two virtual appliances.

The following requirements must be met:

- Two HMCs are at the same version.
- The HMCs use different subnets to connect to the FSPs.
- The HMCs can communicate with the servers' partitions over a public network to allow for full synchronization and function.

5.1.4 HMC code level requirements

The minimum required HMC version for the Power E1180 is V11R1 M1110. V11R1 M1110 is supported on 7063-CR2, and Virtual HMC appliances only. It is not supported on the 7063-CR1 or the 7042 machine types.

Note: HMC with V11R1M1110 cannot manage POWER8 or earlier processor-based systems.

An HMC that is 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
 - Increase in platform keystore size
 - Remove support for vTPM 1.2
 - Quantum safe LPM
 - Minimum Affinity Score and actions
- ▶ Console Management and User Experience Improvements
 - User experience improvements spanning
 - Network Topology
 - Trusted Keystore
 - Import Certificate
 - Multi-factor Authentication Allow list
 - Ability to advertise device information via LLDP
- ▶ Power Infrastructure Maintenance and Automated Power Platform Updates
 - Execute 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
 - Ability to collect platform logs (FW, hypervisor, HMC, VIOS) from a single interface
 - Ability to create a case and upload logs to the case from the HMC
- ▶ Sustainability
 - New energy efficiency mode – higher performance/ watt
 - Partition level energy monitoring
 - Real time monitoring and reporting of energy and carbon emissions at the VM/partition level
 - Scheduling of energy modes

5.1.5 HMC currency

In recent years, cybersecurity emerged as a national security issue and an increasingly critical concern for CIOs and enterprise IT managers.

The IBM Power processor-based architecture has always 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 Power environment is ensuring that the HMC (or virtual HMC) is current and fully supported (including hardware, software, and Power firmware updates).

Outdated or unsupported HMCs represent a technology risk that can quickly and easily be mitigated by upgrading to a current release.

5.1.6 New features

The Hardware Management Console appliance, and the Virtual HMC have the following enhancements:

- Support for managing the new generation of Power11 processor-based systems
- Support for new IO adapters
- Resource Groups
- IBM i Secure Boot
- Quantum safe LPM
- Minimum Affinity Score and Actions
- Ability to advertise device information via LLDP
- Power Infrastructure Maintenance and Automated Power Platform Update Tool
- Automated FFDC Log collection
- Sustainability
- Energy Efficiency mode
- Scheduling of energy modes
- Partition Level Energy Monitoring

5.1.7 Using the Automated Maintenance tool

This section shows using the new automated maintenance tool provided in the IBM Power11 HMC.

Launch Point

Prior to launching the new automated maintenance, the system should have the latest 11.10 firmware installed.

- To start the process, we choose the option “**Update system, VIOS, adapter levels**”. which is a new action menu introduced with Firmware 11.10. This is shown in Figure 5-3.

Details		Usage	Resources		Levels	
1 item selected		Add tag +	Create ▾	Connections and operations ▾	Performance ▾	Firmware ▾
<input type="checkbox"/>	10.48.3.237	No connection	10.48.3.238	9080-HEU	7	Check readiness
<input type="checkbox"/>	172.16.222.42	Version mismatch	172.16.222.42	8286-42A	1	View system firmware levels
<input type="checkbox"/>	BMC-0000-BMC_1921682406	No connection	192.168.240.6	0000-BMC	1	Update system, VIOS, adapter levels New
<input type="checkbox"/>	BMC-0000-BMC_1921682407	No connection	192.168.240.7	0000-BMC	1	Update system firmware
<input type="checkbox"/>	BMC-9105-42A_13E0A80	No connection	172.16.222.41	9105-42A	1	View IO firmware levels
<input type="checkbox"/>	bumblebee	Operating	10.48.4.102	9824-42A	7	Update IO firmware
<input type="checkbox"/>	fuji	Operating	10.48.4.149	9043-MRU	7	View SR-IOV firmware levels
<input type="checkbox"/>	mperf	Operating	10.48.4.248	9080-HEU	1	Update SR-IOV firmware
<input type="checkbox"/>	miracle	Operating	10.48.3.29	9080-HEU	7	Import updates
<input checked="" type="checkbox"/>	ria	Operating	10.48.0.186	9105-42A	1	Edit next IPL flash side
<input type="checkbox"/>	star	Operating	10.48.0.245	9080-M9S	1301A48	LED on <input checked="" type="checkbox"/> B181
<input type="checkbox"/>	zen	Operating	10.48.0.230	9040-MR9	130322X	LED on <input checked="" type="checkbox"/> B181
<input type="checkbox"/>	zep	Error	10.48.0.188	9040-MR9	133C20X	LED on <input checked="" type="checkbox"/> B181

Figure 5-3 Start update

This will start a wizard that will guide us through all the process:

- Import update files option: “Import files to HMC filesystem and perform update” or “Import files only”
- Check System readiness - System needs to be in “Ready” state in order to proceed with update process
- Select the source file location

- Select “**Update/Upgrade**” type as shown in Figure 5-4.

Update system, VIOS, adapter levels new

System name: ria

Accept license agreement
 Choose details and check readiness
 Choose target levels
 Choose process
 Apply updates

System	Processor	fw1110.00(048) 01RB110 (EC level)	Action
ria	9105-42A*13E0B00	Choose an option ▾	Update ▾
	1.6TB NVMe Gen4 U.2 SSD	U78DB.NDO.WZS001T-P1-C2	
	800GB NVMe Gen3 U.2 Slim SSD	U78DB.NDO.WZS001T-P1-C1	
	6.4TB NVMe Gen4 U.2 SSD	U78DB.NDO.WZS001T-P0-C0	
	PCIe3 2-Port 16Gb FC Adapter	U78DB.NDO.WZS001T-P0-C0	
	Universal Serial Bus UHC Spec	U78DB.NDO.WZS001T-P0-T1B	
	NVMe JBOD Card	U78DB.NDO.WZS001T-P0-C10	
rlav1	2	VIOS 4.1.0.10	Choose an option ▾
	1 Gigabit Ethernet (UTP) 4 Port Adapter PCIe-4.0 x8	U78DB.NDO.WZS001T-P0-C3 ent0,ent1,ent2,ent3	Choose an option ▾
	PCIe 2 Port 25/10 Gb NIC&ROCE SFP28 ADAPTER	U78DB.NDO.WZS001T-P0-C4 ent4,ent5	Choose an option ▾
	PCIe3 4-Port 16Gb FC Adapter	U78DB.NDO.WZS001T-P0-C8 fcs0,fcs1,fcs2,fcs3	Choose an option ▾
	PCIe3 2-Port 32Gb FC Adapter	U78DB.NDO.WZS001T-P0-C9 fcs4,fcs5	Choose an option ▾
	3.2TB NVMe Gen4 U.2 SSD	U78DB.NDO.WZS001T-P1-C3 nvme0	Choose an option ▾
	PCIe4 1.6TB NVMe Flash Adapter x8	U78DB.NDO.WZS001T-P0-C1 nvme1	Choose an option ▾
rlav2	3	VIOS 4.1.1.10	Choose an option ▾

Update
 Upgrade
 Do not update

Cancel Previous Next

Figure 5-4 Upgrade type

3. In order to preserve the system availability, for disruptive upgrades/updates, the partitions migration panel will be shown as in Figure 5-5.

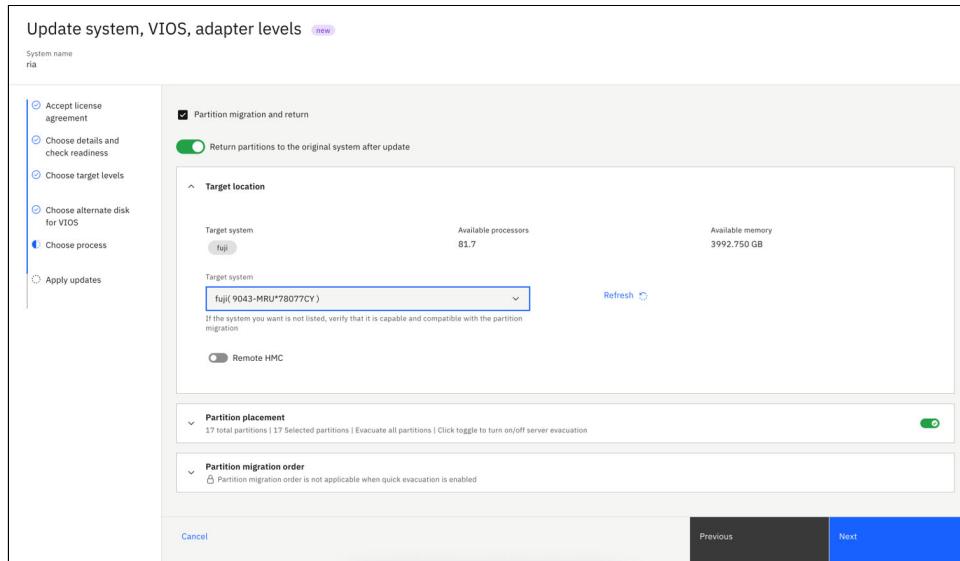


Figure 5-5 Choose Process

4. These are the options that can be selected:

- Remote HMC switch enables to migrate the partitions to a system managed by remote HMC. When selected partitions are not automatically migrated back to the system after the update.
- By default, all the partitions are selected for migration. Partitions can be individually selected to migrate by toggling Server evacuation switch.
- In case of selected partitions for migration, partition migration sequence can be specified for a set of partitions that needs to be migrated initially.

Finally, all the update process is automatically done, and we can see the results as shown in Figure 5-6.

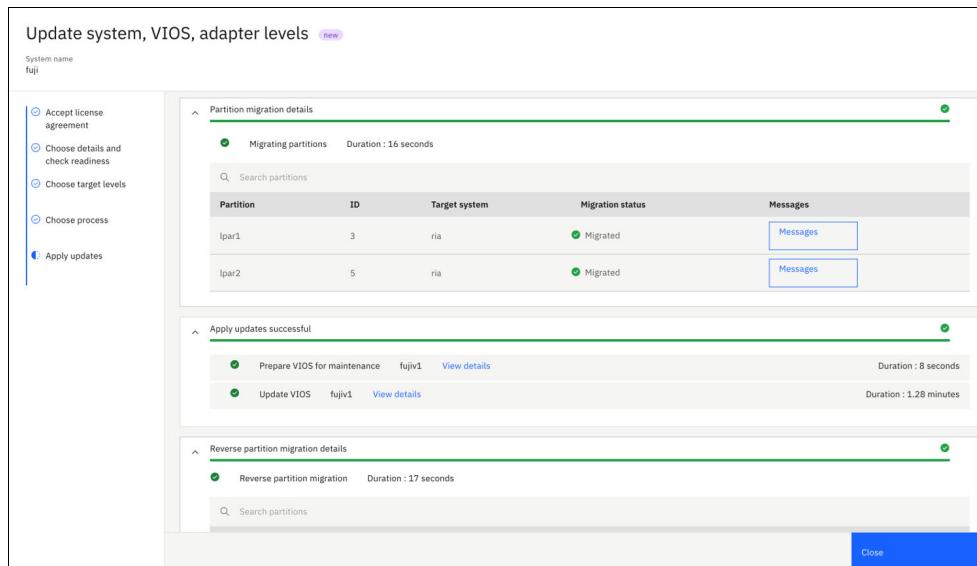


Figure 5-6 Figure 5 – Update status

In conclusion this solution provides an unmatched Business Resiliency with Zero Planned Downtime for maintenance, delivered through platform automation capabilities integrated with IBM Concert's AI-infused workloads. It enables faster and more frequent maintenance updates, keeping systems secure, stable, and compatible with evolving software and hardware requirements, lowering risks of performance degradation, security breaches or unplanned downtimes.

5.2 Ansible

Ansible is an open-source, cross-platform tool for resource provisioning automation that DevOps professionals use for continuous deployment (CD) of software code by leveraging an IaC approach. The Ansible automation platform has evolved to deliver sophisticated automation solutions for operators, administrators, and IT decision-makers across various technical disciplines. It is an enterprise automation solution with flourishing open-source software. It operates on several UNIX like platforms, and can manage systems like UNIX and Microsoft architectures. It comes with descriptive language for describing system settings.

Because of the broad acceptance of the Ansible platform, its open-source design, and its wide support for many devices and platforms, it is becoming a dominant tool in the market. However, it is also common to use other automation tools with Ansible to do more complex automation. For example, many companies use Ansible with Terraform to provide automatic provisioning of their infrastructure.

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 houses Ansible collections, which contain modules, plug-ins, and roles defining the actions Ansible can perform on client nodes.

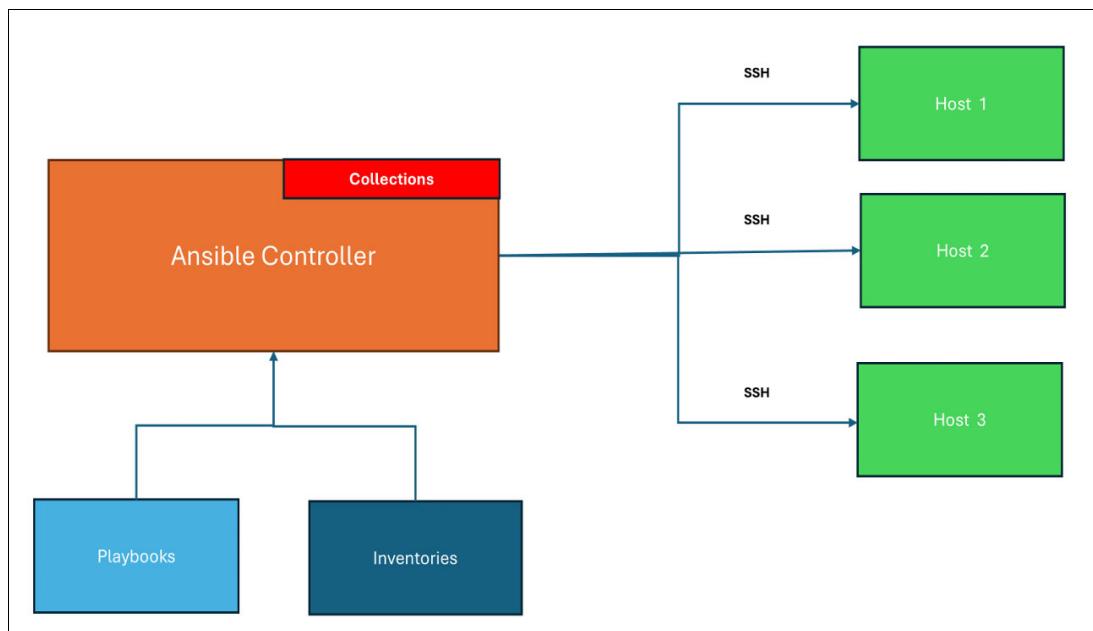


Figure 5-7 Simplified Ansible architecture

Playbooks

The heart of Ansible Automation Ansible playbooks are YAML files that define sequences of tasks to run on remote hosts. These tasks can range from installing packages to 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 seeking 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 be used for 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 the Ansible declarative syntax, you can define the state of a system, and Ansible takes the necessary steps to achieve it.
- ▶ **Ease of learning:** The Ansible YAML syntax and minimal learning curve make it accessible to IT professionals at all levels.

Ansible is a powerful automation tool that can help organizations improve efficiency, scalability, and reliability in their IT infrastructure. By leveraging Ansible playbooks, IT teams can streamline routine tasks, automate complex workflows, and help provide consistent configurations across their environments.

Options for implementing Ansible

As you decide 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 options that are available to you.

Ansible Community

The community versions of Ansible primarily include the following ones:

- ▶ **Ansible Core**

Ansible Core is a fundamental part of Ansible. It provides the core automation engine. It is an open-source tool that includes the basic functions for configuration management, application deployment, and task automation. Ansible Core includes modules, plug-ins, and the CLI that is 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 UI, Representational State Transfer (REST) API, and task engine for managing Ansible automation at scale. AWX offers role-based access control (RBAC), job scheduling, graphical inventory management, and more. It helps users manage and scale automation efforts.

- ▶ Ansible Collections

Ansible Collections are pre-packaged modules, roles, and plug-ins that are created and shared by the community. With Collections, users can extend Ansible functions with more content that is often maintained by the community or specific organizations. Collections can be downloaded 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, which 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 Red Hat Ansible Automation Platform.

Ansible Automation Platform

Ansible Automation Platform is a subscription-based enterprise solution that combines over 20 community projects into a fully supported automation platform. Ansible Automation Platform provides curated, certified, and validated Ansible Collections and roles from partners like 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, which includes SLAs for security, compatibility, and upgrades. Community options might have limited support.
- ▶ Features: Ansible Automation Platform includes features beyond Ansible Core, such as a web interface and integration with other tools.
- ▶ Cost: Ansible Automation Platform is a subscription-based product, 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 due to its enterprise-grade features and support.

By carefully 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 brings significant value to managing IBM Power. Here are several key use cases where Ansible enhances efficiency, consistency, and scalability in Power environments:

1. System Configuration and Provisioning

Ansible automates the setup and configuration of AIX, IBM i, and Linux on IBM Power. This includes tasks like user and group management, network configuration, software installation, and system tuning. By using Ansible playbooks, administrators can provide consistent configurations across multiple systems, reducing manual errors and speeding up provisioning.

2. Patch Management and Compliance

Keeping systems up to date is critical for security and stability. Ansible can automate the patching process for AIX and IBM i, including downloading patches, applying them, and verifying system health post-update. It also supports compliance checks by validating system configurations against predefined baselines, helping organizations meet regulatory and security standards.

3. Integration with HMC and PowerVC

Ansible collections for IBM Power include modules that interact with the Hardware Management Console (HMC) and PowerVC. This enables automation of tasks such as LPAR creation, VIOS configuration, and virtual machine lifecycle management. Combined with dynamic inventory capabilities, Ansible can discover and manage Power infrastructure components in real time.

For more information on implementing automation with Ansible in an IBM Power environment refer to this IBM Redbook: *Using Ansible for Automation in IBM Power Environments*, SG24-8551

5.3 Terraform

Terraform is an open source tool originally developed by HashiCorp and now owned by IBM. It is written in the Go programming language and compiles down into an executable named Terraform. Terraform is an infrastructure as code tool that lets you build, change, and version cloud and on-premises resources safely and efficiently. Terraform provides a mechanism to access any API for any cloud provider to manage infrastructure as a service (IaaS).

Figure 5-8 shows the process involved in calling the API. The definition of which APIs to call is defined in configuration files. These configuration files are the code in that is referenced in Infrastructure as code.



Figure 5-8 Terraform functionality

Terraform allows users to define and provision infrastructure using a high-level configuration language called HashiCorp Configuration Language (HCL). With Terraform, infrastructure components such as servers, databases, networking, and storage can be described 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 necessary provider plugins. The `terraform plan` command then creates an execution plan, showing what actions Terraform will take to reach the desired state. Finally, `terraform apply` executes the plan, making the actual changes to the infrastructure. Terraform maintains a state file that tracks the current state of the

infrastructure, which is essential for determining what changes need to be made during future runs.

One of Terraform's key strengths is its provider ecosystem, which allows it to manage resources across a wide range of platforms, including AWS, Azure, Google Cloud, Kubernetes, and many others. This makes it a powerful tool for managing hybrid and multi-cloud environments. Additionally, 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 through defining your intended end point through configuration files which are then placed into a plan and final applied through infrastructure providers.

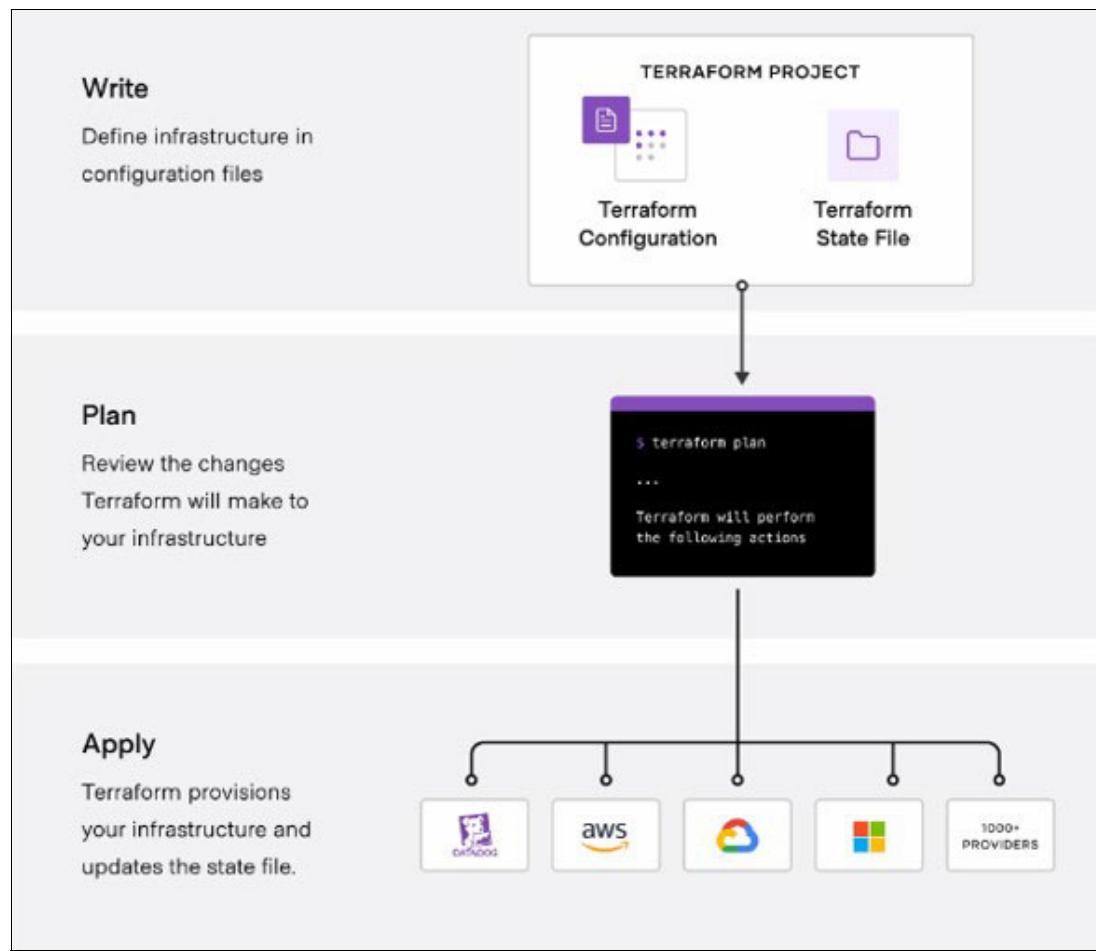


Figure 5-9 Terraform process

Track your infrastructure

Terraform generates a plan and prompts you for your approval before modifying your infrastructure. The state of your infrastructure is kept in a file named "terraform.tfstate", which can be held in Git, Gitlab or HCP Terraform to version, encrypt, and securely share it with your team. This acts as a single source of truth for your environment.

Automate Changes

Terraform configuration files are declarative, describing the end state. So are easy to automate with tools like Ansible.

Standardize configurations

Terraform provides standardization in modules. A module consists of a collection of `.tf` and `.tf.json` files kept together in a directory. Modules are the main way to package and reuse resource configurations with Terraform.

Collaborate

Since Terraform can be distributed as configuration files and version controlled in applications like git, GitHub and HCP Terraform, these are ideal locations to share and collaborate.

Combining Ansible and Terraform

Terraform connects to any provider, like Ansible, to manage your infrastructure. Browse the [Terraform registry](#) for providers. The Terraform Provider for Ansible provides a more straightforward and robust means of executing Ansible automation from Terraform rather than local-exec¹. Figure 5-10 shows the Ansible provider entry in the registry.

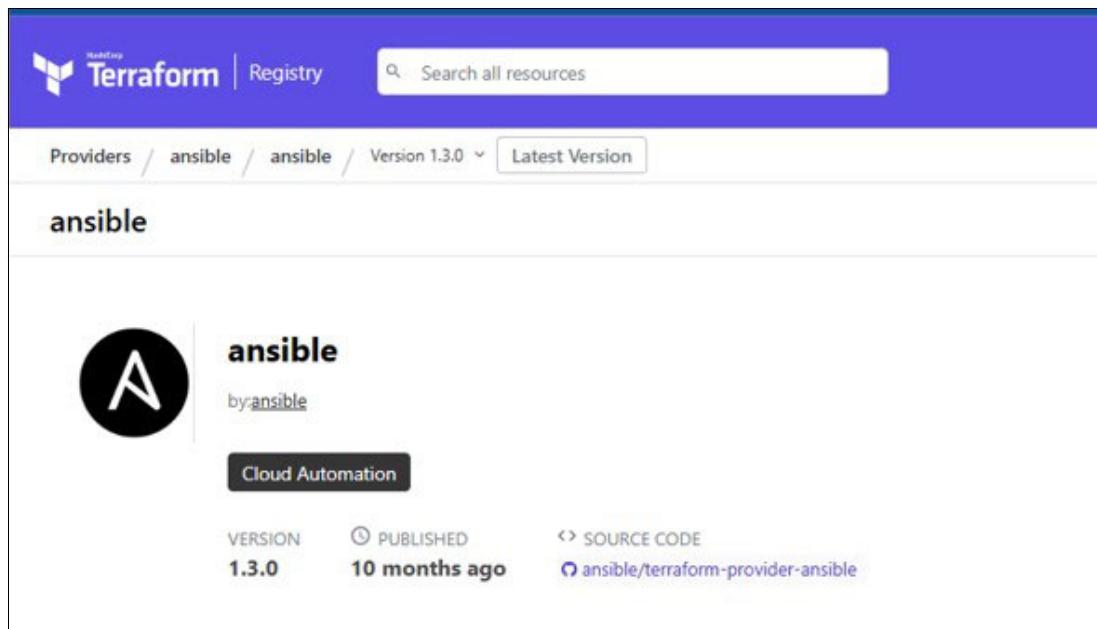


Figure 5-10 Terraform Ansible provider

The prerequisites for using the Ansible provider are as follows;

1. Install Go

For installation instructions, refer to the [official installation guide](#).

2. Install Terraform:

Install the ppc64le version for operation on IBM Power. Installation instructions are found on the [GitHub registry](#).

3. Install Ansible

To install Ansible refer to the [Ansible official installation guide](#)

¹ <https://developer.hashicorp.com/terraform/language/resources/provisioners/local-exec>

For more information on Terraform in the IBM Power environment, refer to the IBM Redbook publication *Modernization Techniques for IBM Power*, SG24-8582

5.4 PowerVC

IBM PowerVC is an advanced virtualization and cloud management platform built on OpenStack, specifically designed for managing virtualized workloads on IBM Power. It supports AIX, IBM i, and Linux operating systems and provides a unified interface for deploying, managing, and automating virtual machines (VMs) across Power environments. PowerVC simplifies the creation of private clouds and integrates seamlessly 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. The Standard Edition includes core virtualization management features such as:

- ▶ VM image capture, import/export, and deployment
- ▶ Policy-based VM placement for optimized resource utilization
- ▶ Snapshots and cloning for backup and testing
- ▶ Live VM mobility and remote restart for high availability
- ▶ Role-based access control and automated I/O configuration

The Private Cloud Edition builds on this by adding:

- ▶ A self-service portal for end users to provision VMs
- ▶ Approval workflows for provisioning requests
- ▶ Pre-built deployment templates
- ▶ Cloud management policies and metering for chargeback

PowerVC also supports Dynamic Resource Optimization (DRO), which automatically balances workloads based on CPU and memory usage, and Simplified Remote Restart, which means VMs can be restarted on alternate hosts in case of failure. These features reduce administrative overhead and improve system resilience.

NovaLink is a lightweight, Linux-based virtualization management interface that acts as a bridge between PowerVC (IBM's cloud and virtualization manager) and PowerVM (the hypervisor on IBM Power Servers). It enables direct, scalable, and efficient control of virtual machines (VMs) and system resources without relying solely on the traditional Hardware Management Console (HMC).

Installed on the Power Server: NovaLink runs as a service on a dedicated Linux partition (LPAR) on the same Power server it manages. This LPAR communicates directly with the PowerVM hypervisor the HMC, which manages systems externally. NovaLink has direct access to the hypervisor APIs. This allows for faster and more granular control of virtualization tasks like VM creation, deletion, migration, and resource allocation. This makes PowerVC more responsive and scalable, especially in environments with many VMs or frequent provisioning changes.

Table 5-2 Comparisons of PowerVC with HMC management:

Feature/ Capability	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
Dynamic Resource Optimization (DRO)	Automatically balances workloads across hosts	Not available
Simplified Networking & 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



Operating Systems

IBM Power Servers support a trio of robust operating systems – IBM i, AIX, and Linux – each tailored to meet distinct enterprise computing needs. IBM i is an integrated operating environment known for its exceptional reliability, security, and ease of management. It includes a built-in database (Db2 for i), middleware, and development tools, making it ideal for running mission-critical business applications with minimal administrative overhead. Its tight integration and backward compatibility make it a favorite for industries like finance, manufacturing, and retail.

AIX, IBM's UNIX 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 like dynamic system tuning, workload partitioning, and robust security. It is widely used in environments that demand stability and performance, such as ERP systems, databases, and analytics platforms. AIX is optimized for IBM Power hardware, ensuring seamless integration and efficient resource utilization.

IBM Power also supports enterprise-grade Linux distributions, including Red Hat Enterprise Linux (RHEL) and SUSE Linux Enterprise Server (SLES). These Linux environments provide flexibility and open-source innovation, making them ideal for cloud-native applications, containerized workloads, and AI/ML development. Running Linux on Power enables organizations to leverage the performance advantages of the Power architecture while maintaining compatibility with the broader Linux ecosystem. Together, these operating systems make IBM Power a versatile platform capable of supporting a wide range of modern and legacy workloads.

Complementing these operating systems is Red Hat OpenShift, a Kubernetes-based container platform that runs natively on IBM Power. OpenShift enables organizations to build, deploy, and manage containerized applications across hybrid cloud environments with consistency and scalability. OpenShift brings a unified DevOps experience and supports modern application architectures, including microservices and AI workloads. This integration empowers enterprises to modernize their IT infrastructure while leveraging the performance, security, and reliability of IBM Power.

This chapter provides the following topics:

- ▶ 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 both AIX and IBM i, providing customers with greater flexibility, predictability, and cost efficiency. These models allow organizations to license the operating system and associated software on a monthly or annual basis, rather than through traditional perpetual licenses. This approach aligns with modern IT consumption trends, enabling businesses to scale usage up or down based on workload demands while simplifying budgeting and procurement processes. Subscription licensing also includes access to updates, patches, and IBM support, ensuring systems remain secure and up to date.

For IBM i, subscription licensing includes the core operating system along with integrated middleware such as Db2, security features, and application development tools. Similarly, AIX subscription licensing covers the base OS and can be bundled with additional IBM software components. These subscriptions are available for deployment on IBM Power11 and Power10, and are particularly well-suited for hybrid cloud environments where agility and operational efficiency are key. By adopting subscription licensing, organizations can modernize their licensing strategy while continuing to leverage 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 that provides access to an IBM program and IBM software maintenance for a specified subscription term (one or three years). The subscription term begins on the start date and ends on the expiration date, which is reflected at the IBM ESS website.

Customers are licensed to run the product through the expiration date of the one- or three-year subscription term and then can renew at the end of the subscription to continue using the product. This model provides flexible and predictable pricing over a specific term, with lower 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 be moved between on-premises and cloud if needed, something that is becoming more of a requirement with hybrid workloads.

The product IDs for the subscription licenses are listed in Table 6-1.

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, designed to provide flexibility and align with modern IT consumption models. These subscriptions are typically available on IBM Power Servers, including Power11, and are ideal for clients looking to simplify licensing, reduce upfront costs, and scale more 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, making it easier for organizations to align software costs with usage. Support and updates are included, ensuring systems remain secure and current without the need for separate maintenance agreements.
- ▶ In addition to the base OS subscription, IBM provides IBM i Value Packages, which bundle the operating system with additional IBM software like Rational® Development Studio for i, IBM Db2 Mirror for i, and IBM i Access Client Solutions. These packages 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 like IBM Power Virtual Server. This is particularly beneficial for organizations adopting hybrid or cloud-first strategies, as it enables scalable, enterprise-grade IBM i environments without the complexity of traditional licensing.

6.2 AIX

The AIX operating system is a secure, scalable, and robust open standards-based UNIX system. For over thirty years, AIX has been the cornerstone of mission-critical computing for enterprise organizations in highly complex industries, evolving to introduce a wealth of new hybrid cloud and open-source capabilities.

AIX 7.3 is the latest AIX release available in the market. It builds on a solid foundation by offering new functions and capabilities that further enhance performance, scalability, availability, and security, all while preserving application-binary compatibility to safeguard IT investments in AIX.

Coupled with the IBM Power11 processor-based systems, AIX 7.3 provides an optimized and more resilient computing platform that adapts to changing business demands, including new cloud use cases and improved economics.

6.2.1 AIX 7.3 Key Features

Here is a list of the key features in the latest AIX release.

- ▶ Workload Scalability & Automation: AIX 7.3 offers enhanced workload scalability, improved cloud automation via Ansible, and over 300 open-source packages through the AIX Toolbox for Open Source Software, enabling modern application development.
- ▶ Live Kernel Update: Introduced in AIX 7.2 and enhanced in 7.3, this feature allows interim fixes, service packs, and technology level updates without reboots, supporting PowerVC-managed environments and Power Enterprise Pool systems for resource optimization.

- ▶ High Availability & Disaster Recovery: IBM PowerHA and VM Recovery Manager provide automated recovery and multi-site replication, minimizing downtime and ensuring business continuity in hybrid or public cloud environments.
- ▶ AI Integration: AIX workloads are a natural source for AI. These systems host a tremendous amount of high-quality data on customer behavior and transactional information that can be further leveraged for AI development, enabling machine and deep learning for actionable insights on a unified platform.
- ▶ Cloud Flexibility: AIX enables private, on-premises cloud transformation with PowerVC, offering hybrid cloud functionality for seamless AIX VM import/export and software-defined infrastructure for SAN-less DevOps environments. Also available via IBM Power Virtual Server, 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, ensuring reliability and efficiency.

6.2.2 Supported Levels

Currently, the IBM Power11 processor-based server supports the following minimum levels of the AIX operating system when installed by using direct I/O connectivity:

- AIX 7.3 with the 7300-03 Technology Level and Service Pack 1 or later
- AIX 7.3 with the 7300-02 Technology Level and Service Pack 4 or later
- AIX 7.2 with the 7200-05 Technology Level and Service Pack 10 or later

Currently, the IBM Power11 processor-based server supports the following minimum levels of AIX operating system when installed using virtual I/O:

- AIX 7.3 with the Technology Level 7300-03 and Service Pack 0 or later
- AIX 7.3 with the Technology Level 7300-02 and Service Pack 2 or later
- AIX 7.2 with the Technology Level 7200-05 and Service Pack 8 or later

Important:

- Starting with AIX Release 7.3 Technology Level 3 Service Pack 1, AIX logical partitions operate in native Power11 processor compatibility mode, fully leveraging Power11's advanced capabilities.
- Logical partitions running versions older than AIX Release 7.3 Technology Level 3 Service Pack 1 operate in Power9 or Power10 processor compatibility mode on Power11 systems.
- Logical partitions running AIX Release 7.2 can operate in Power9 or Power10 processor compatibility mode.

6.2.3 AIX Maintenance levels

IBM periodically releases maintenance packages (service packs (SPs) or technology levels (TLs) for the AIX operating system. For more information about these packages, downloading, and obtaining the installation packages, see [IBM Fix Central](#). For more information about hardware features compatibility and the corresponding AIX Technology Levels, see [IBM Support](#).

The Service Update Management Assistant (SUMA), which can help you automate the task of checking and downloading operating system downloads, is part of the base operating system. 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 upgrades of key components or to verify the current health of a system.

The [IBM AIX Operating System Service Strategy and Best Practices](#) is a free resource available to AIX clients and gives insight into the AIX service strategy, while also providing helpful lifecycle information to best maintain your version of AIX. For more information on AIX, visit [AIX on IBM Power](#).

For additional information:

- ▶ AIX support lifecycle information
<https://www.ibm.com/support/pages/aix-support-lifecycle-information>
- ▶ System Software Maps
<https://www.ibm.com/support/pages/system-software-maps>
- ▶ System to AIX maps
<https://www.ibm.com/support/pages/system-aix-maps>

6.2.4 Licensing

The AIX operating system is available in the following editions:

- ▶ AIX Standard Edition
- ▶ AIX Enterprise Edition
- ▶ AIX Cloud Edition

The Enterprise and Cloud Editions include Power-related software, typically required to manage larger IBM Power environments, including those in hybrid clouds.

There are two licensing models for AIX:

- ▶ CPU-based licensing
- ▶ Subscription licensing

In the first case, AIX licenses can be ordered together with the server or purchased later as a MES upgrade for the server. The license grants you the right to use AIX on a specific server. If you require support for AIX, you must have a valid Software Maintenance Agreement (SWMA) for AIX.

Subscription licensing is a new model that offers greater flexibility in AIX acquisition. The subscription license includes access to IBM software maintenance for a specified subscription term (1 or 3 years). After this term, you can renew the subscription if you continue using AIX. This model provides flexible and predictable pricing over a specific term with lower upfront acquisition costs.

Another benefit is that the licenses are customer-number-entitled, meaning you can use the licenses on any IBM Power server you have in your environment. You can even move them between on-premises, your private cloud, and public cloud if needed.

6.2.5 New AIX Editions

The IBM AIX operating system is an open standards-based UNIX operating system that has been the foundation of mission-critical workloads and databases for tens of thousands of customers for over 35 years. AIX provides an enterprise-class IT infrastructure that delivers the reliability, availability, performance, and security that is required for organizations to be successful in a global economy.

Now, IBM offers the following updates and enhancements to the AIX operating system and products that include AIX for IBM Power8, Power9, Power10 and Power11 technology 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 1.9

IBM AIX 7 Enterprise Edition 1.13

IBM has updated the AIX 7 Enterprise Edition and its corresponding subscription offering to version 1.13. The bundled software components offered with AIX 7 Enterprise Edition 1.13 (5765-CD3 and 5765-2E1) now include:

- ▶ 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 HA 1.9
- ▶ IBM Tivoli® Monitoring 6.3

The bundle components are updated as follows:

- ▶ AIX 7.3 TL3 has been updated with service pack 1
- ▶ AIX 7.2 TL5 has been updated with service pack 10
- ▶ IBM PowerSC 2.3 has been updated from 2.2 to 2.3
- ▶ IBM PowerVC for Private Cloud has been updated from 2.3.0 to 2.3.1
- ▶ IBM VM Recovery Manager HA has been updated from 1.8 to 1.9

Additional information

- ▶ Clients with active Software Maintenance (SWMA) or subscriptions for earlier versions of AIX Standard Edition or AIX Enterprise Edition are entitled to upgrade at no charge. To update, download, or install, see the IBM Entitled Systems Support website.
- ▶ Clients can choose either AIX 7.3 TL3 or AIX 7.2 TL5.
- ▶ Clients selecting AIX 7.2 TL5 can later upgrade to AIX 7.3 TL3 at any time, provided SWMA or subscription is 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. The bundled software components offered with Private Cloud Edition 1.13 (5765-ECB and 5765-6C1) now include:

- ▶ IBM PowerSC 2.3
- ▶ IBM PowerVC for Private Cloud 2.3.1
- ▶ IBM VM Recovery Manager DR 1.9
- ▶ IBM Tivoli Monitoring 6.3

The bundle components are updated as follows:

- ▶ IBM PowerSC 2.3 has been updated from 2.2 to 2.3
- ▶ IBM PowerVC for Private Cloud has been updated from 2.3.0 to 2.3.1
- ▶ IBM VM Recovery Manager DR has been updated from 1.8 to 1.9
- ▶ Cloud Management Console (CMC) has been updated from 1.22 to 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. The bundled software components offered with Private Cloud Edition with AIX 1.13 (5765-CBA and 5765- 2C1) now include:

- ▶ 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 DR 1.9
- ▶ IBM Tivoli Monitoring 6.3

The bundle components are updated as follows:

- ▶ AIX 7.3 TL3 has been updated with service pack 1
- ▶ AIX 7.2 TL5 has been updated with service pack 10
- ▶ IBM PowerSC 2.3 has been updated from 2.2 to 2.3
- ▶ IBM PowerVC for Private Cloud has been updated from 2.3.0 to 2.3.1
- ▶ IBM VM Recovery Manager DR has been updated from 1.8 to 1.9
- ▶ Cloud Management Console (CMC) has been updated from 1.22 to 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 subscription to IBM Cloud Management Console (5765-CMT) for the same term as their SWMA.
- ▶ Clients with active SWMA or subscriptions for earlier versions of Private Cloud Edition or Private Cloud with AIX are entitled to upgrade at no charge. To update, download, or install, see the [IBM Entitled Systems Support](#) website.
- ▶ Clients can choose either AIX 7.3 TL3 or AIX 7.2 TL5.
- ▶ Clients selecting AIX 7.2 TL5 can later upgrade to AIX 7.3 TL3 at any time, provided SWMA or subscription is current.
- ▶ Clients with AIX Enterprise Edition can trade up to Private Cloud Edition with AIX.

VM Recovery Manager 1.9

VM Recovery Manager provides automated VM management in the data center and disaster recovery management between sites. In IBM Power Virtual Server public the product is called DR Automation. DR Automation provides automated disaster recover management between geographically dispersed Power Virtual Server data centers.

- ▶ Work group support for HA-DR-HA
- ▶ Failover rehearsal for IBM Power Virtual Server

6.3 IBM i

IBM i is a powerful, integrated operating system designed to run on IBM Power hardware. It is known for its robustness, security, and ease of management. Over the years, it has evolved through several naming transitions, reflecting both technological advancements and IBM's strategic branding changes.

IBM i remains a premier enterprise computing platform, renowned for its robust security, scalability, and seamless integration. Its ability to support modern workloads while preserving legacy application compatibility makes it a cornerstone in industries like finance, retail, and manufacturing, where unwavering reliability and performance are paramount. IBM's ongoing innovation ensures IBM i, a key component of IBM Power, continues to empower enterprises to efficiently manage and deploy their most critical applications, building on its legacy of stability and business continuity.

6.3.1 Intro to IBM i

IBM i is a highly integrated and dependable platform that streamlines IT infrastructure by combining the operating system, Db2 for i database, security, and middleware into a single, cohesive environment. This deep integration minimizes the need for complex configurations and third-party components, making it easier to deploy, manage, and scale applications. As a result, organizations benefit from a lower total cost of ownership (TCO) and reduced administrative overhead – especially valuable in environments with limited IT resources.

One of IBM i's standout strengths is its exceptional reliability, availability, and serviceability (RAS). Designed to support mission-critical workloads, it offers features such as automatic error detection, self-healing capabilities, and robust high availability and disaster recovery options. These capabilities make IBM i a trusted platform across industries like finance, healthcare, and manufacturing, where system uptime and data integrity are paramount. Additionally, IBM i supports both legacy and modern application development, allowing businesses to run traditional RPG and COBOL applications alongside modern languages like Java, Python, and Node.js, while integrating with open-source tools and cloud-native services.

For more information on IBM i and its latest functions refer to the IBM Redbook *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 operating system both native and as a client of VIOS:

- ▶ IBM i - 7.4 TR12
- ▶ IBM i - 7.5 TR6
- ▶ IBM i - 7.6

6.3.3 IBM i Software Tiers

For the Power E1180, all configurations are P30 IBM i Software Tier

6.3.4 New IBM i Features supported by Power11

IBM i 7.6 Power11 GA PTF Group, IBM i 7.5 Hardware Technology Refresh 6, and IBM i 7.4 Hardware Technology Refresh 12 all support the new IBM Power servers with Power11 technology and their new I/O offerings.

There are several new features that are available only with the new servers with Power11 technology and newer code levels for IBM i:

- ▶ IBM i 7.6 adds an NVMe Hot Spare feature to allow immediate replacement of a mirror-protected disk unit under a failing NVMe device. For more information on this feature see [NVMe Hot Spare](#).
- ▶ IBM i 7.6 Secure boot feature is used to verify the authenticity of the boot process. Security administrators can extend the existing secure boot technology used to verify the integrity of PowerVM firmware, including hostboot, Power hypervisor (PHYP), and partition firmware (PFW) to also include IBM i Licensed Internal Code (LIC) through digital signatures. The IBM i operating system can verify the integrity of the boot chain up through the IBM i LIC prior to its execution by using a digital signature and a hardware-protected asymmetric key.
- ▶ IBM i 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.
- ▶ IBM i 7.5 and 7.6 enable NX-accelerated compression for PASE programs on Power11 servers. Using the zlibNX library, as supported in AIX, IBM i now accelerates DEFLATE compression in PASE, with the pigz program automatically benefiting. This can reduce CPU usage by up to 95% and significantly speed up compression tasks.
- ▶ With IBM i 7.6, support for Tape Library Virtualization has been extended to include IBM 3584 tape libraries when running on Power10 and Power11 servers. This enhancement builds on existing virtualization capabilities previously available for libraries such as the 3572, 3573, and 3555. Tape Library Virtualization allows multiple IBM i partitions to share a single physical tape library, significantly reducing infrastructure costs by eliminating the need for a dedicated tape drive per LPAR. Additionally, this support extends to Virtual Tape Libraries (VTLs) that emulate the 3584 model. This enhancement is described in the [IBM i Removable Media: Tape Library Virtualization](#) landing page.

6.3.5 IBM i Licensing

On Power11, IBM i licenses will only be available via subscription, with the exception of customers doing MES or serial number upgrades from E1080 to E1180 where there will be no change in licensing.

For Power10 new purchases, IBM announced the withdrawal of perpetual IBM i licenses for P20 and P30 software tiers in July 2025 with an effective date of January 1st 2026. New Power10 purchases prior to January 2026 can:

- Purchase new subscription licenses
- Purchase new perpetual licenses
- Transfer existing perpetual licenses

IBM had previously withdrawn new perpetual licenses for P05 and P10 tiers.

When an IBM i license is acquired, license keys are generated which are unique to the server on which the license is acquired. These licenses can be grouped into two categories with respect to the license keys that are fulfilled:

- ▶ Non-expiring (also called “perpetual”) - meaning that the license keys have no expiration date, and
- ▶ Subscription - meaning that the license keys have an expiration date for the duration of the subscription term.

IBM i licenses are also categorized into Software Tiers, also known as Processor Groups. These tiers relate to the size of the hardware platform on which the operating system runs. The Software Tiers are P05, P10, P20, and P30. The IBM i licensing metric is different between the various tiers, for the P30 tier it is based on number of cores.

For many customers, IBM i 7.6 will be their first experience of 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 on IBM i subscription licensing, including a comprehensive Frequently Asked Questions (FAQ) section, refer to:

<https://public.dhe.ibm.com/systems/support/planning/transfer/IBM.i.Transformation.FAQ.pdf>

Perpetual licenses withdrawal: IBM i perpetual licenses have been withdrawn for new orders. However, it is possible to migrate perpetual IBM i licenses for existing customers from E1080 to E1180 models.

When purchasing new subscription licenses, P30 customers can purchase standalone subscriptions for each component or they can choose to take advantage of new subscriptions bundles.

6.3.6 Subscription bundles

IBM introduces IBM i P30 Enterprise Edition, a strategic subscription bundle that offers comprehensive IBM i enterprise capabilities:

- ▶ IBM i, the integrated operating system that includes database, middleware, security administration, and more.
- ▶ Advanced HA/DR capabilities to maximize uptime for mission-critical workloads.
- ▶ Automation-driven operational management to increase IT productivity and minimize disruptions.
- ▶ Application development and modernization tools to accelerate innovation.
- ▶ Expert guidance to accelerate value adoption

IBM i P30 Enterprise Edition offering includes the following products and services:

Table 6-1 Enterprise Edition subscription bundle

Product name	Description
IBM i	Fully integrated operating system, meaning the database, middleware, security, runtime, and hypervisor are all integrated into the stack and licensed as one. This integration helps clients lower TCO, simplify systems management, and do more with fewer resources.

Product name	Description
IBM Backup, Recovery and Media Services for i (BRMS)	IBM Backup, Recovery, and Media Services (BRMS) helps you implement a disciplined approach to managing your backups and provides you with an orderly way to retrieve lost or damaged data. BRMS is the IBM strategic solution for planning and managing save and restore operations on IBM i.
IBM PowerHA SystemMirror® for i	IBM PowerHA technology positions you to deploy an HA solution that addresses storage and high availability requirements with one integrated configuration, with a simplified user interface. IBM Power is committed to investing in and bringing to market solutions designed to keep your IT environments resilient.
IBM Db2® Mirror for i	IBM Db2 Mirror for i enables continuous availability for your mission-critical applications. It provides a Recovery Time Objective (RTO) of zero. Db2 Mirror for i synchronously mirrors database updates between two separate nodes via remote direct memory access (RDMA) over Converged Ethernet (RoCE) network. Applications can be deployed in an active-active or active-passive (with read access on the secondary) node.
IBM PowerSC	PowerSC provides end-point protection for IBM i, enabling real-time security monitoring for continuous visibility and compliance automation. It integrates IBM i security with enterprise SIEMs for your broader enterprise security strategy. It sits on top of the IBM Power server stack, integrating security features built at different layers.
PowerHA Tools for IBM i	<p>PowerHA Tools for IBM i provide customizable templates for high availability solutions designed to provide business continuity by minimizing downtime and simplifying HA/DR processes. The tools include:</p> <ul style="list-style-type: none"> ▶ IASP Manager ▶ Full System FlashCopy (FSFC) Manager ▶ Full System Replication (FSR) Manager ▶ Smart Assist for PowerHA on IBM i ▶ Safeguarded Copy Toolkit <p>PowerHA Tools for IBM i are IBM Technology Expert Labs tools and can be implemented with additional chargeable services available from IBM Technology Expert Labs.</p>
IBM i Migrate While Active	IBM i Migrate While Active orchestrates the migration of an IBM i logical partition (LPAR) to a different location. That location can be within the same server, same data center, or to a remote location such as IBM Power Virtual Server (IBM Power VS) in IBM Cloud. The process starts with taking a source LPAR at IBM i 7.4, or later, making a copy of that LPAR, and migrating data from the source to the copy. While the source node is still active, changed data will be continuously migrated to the copy node. When the two nodes are “in sync”, a cutover to the new node can be performed, which completes the migration. Migrate While Active is optimized for migration where a physical tape cannot be used for a traditional D Mode IPL and restore. It handles the save to optical media so a network install may be performed.

Product name	Description
iDoctor toolkit	iDoctor is a suite of dynamic tools that identify performance issues quickly on IBM i systems. Monitor your overall system health at a high level and take advantage of advanced drill-down capabilities for specific issues. iDoctor is an IBM tool that can be leveraged with additional chargeable performance services available from IBM Technology Expert Labs.
IBM Rational Development Studio for i	<p>IBM Rational Development Studio for i is an application development package that includes all of the server tools and compilers needed for developing and maintaining applications for IBM i.</p> <p>Rational Development Studio for i provides all three features:</p> <ul style="list-style-type: none"> ▶ Integrated Language Environment (ILE) compilers ▶ Heritage compilers ▶ Application Development ToolSet (ADTS)
IBM i Extensions for VS Code Enterprise Support	An enterprise support offering for a curated set of VS Code extensions that enable developers to develop IBM i applications using RPG, COBOL, CL and SQL within a modern development environment
IBM i Advisory Connect	IBM i Advisory Connect service is designed for large enterprises to provide rapid access to IBM Technology Expert Labs experts for advice and insights on database, performance, resiliency, security, and automation. In addition to accessing ongoing advice as needed, clients can utilize one remote annual technology assessment for database, performance, resiliency, security, or automation capabilities and practices.
Red Hat Ansible Automation	<p>The Ansible Automation Platform is a foundation for building and operating automation across an organization. The platform includes all the tools required to implement enterprise-wide automation. Ansible-managed endpoints can include logical partitions (LPARs) on Power servers running IBM AIX, IBM i, or Linux. The 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). Clients purchasing 10 processor cores or more of IBM i P30 Enterprise Edition are granted to get a one-year subscription of Red Hat Ansible Automation Platform for up to 100 managed nodes. After the first year, renewals are available directly from IBM Power or from Red Hat.</p> <p>Clients purchasing 10 processor cores or more of IBM i P30 Enterprise Edition are granted to get a one-year subscription of Red Hat Ansible Automation Platform for up to 100 managed nodes. After the first year, renewals are available directly from IBM Power or from Red Hat.</p> <p>NOTE: The 10 processor cores purchase must be in a single transaction to be eligible for this benefit.</p>

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 (PEP2.0) enables IBM i OS and PowerHA SystemMirror sharing across systems. For information about IBM Power Enterprise Pools see section 10.1, “IBM Power Private Cloud with Shared Utility Capacity” on page 192.

- ▶ In a Pool, clients can mix systems with
 - Subscription licenses
 - Subscription Bundles: IBM i P30 Enterprise Edition
 - Standalone Subscriptions
 - Perpetual licenses
- Each system must be of a single license type, Subscription or Perpetual. They will be shared as Base Capacity or available as Metered Capacity by the minute in a Pool.
- ▶ Other IBM i Licensed Program Products (LPP) must be licensed per system based on their peak usage
 - They are not shared as Base Capacity or available as Metered Capability by the minute in a Pool

Note: On a given serial number, a system can have either OS subscription or OS perpetual license. but not both.

Rational Development Studio for i (RDS) named user license

RDS named user licenses acquired through IBM i P30 Enterprise Edition may be shared across each system with Enterprise Edition licenses.

For example:

- ▶ A client has 3 systems with 10 cores each: 30 total cores.
 - They are entitled to 150 total named users (5 users per core).
 - These 150 named users can be on any or all 3 systems.

IBM i License Terms

There is no change in IBM i License Terms based on subscriptions, subscription bundles, and perpetual licenses.

Determine the number of processors used by the Program in a machine using the smaller value of either of the following methods:

1. The total number of activated processors in the machine.
2. The sum of **a** and **b** as follows (any remaining fraction of a processor must be rounded up to a full processor in the final aggregation):
 - a. When IBM i is run in partitions with dedicated processors, the sum of the processors assigned to those partitions.
 - b. When IBM i is run in partitions that are members of a shared processing pool, for each shared processor pool, the sum of the smaller of **i** or **ii**:
 - i. The maximum capacity of the shared processor pool.
 - ii. The sum of the virtual processors of each uncapped partition plus the processing units assigned in each capped partition running the Program

6.3.7 IBM i Software Maintenance (SWMA)

IBM i Software Maintenance (SWMA) is included in the Subscription licenses. On the other hand, for non-expiring licenses, one year of SWMA is included with the purchase of the license, and then additional years of SWMA are charged separately.

IBM i SWMA is important for your business. Not only does it allow you to engage with IBM's support centers around the world but it also entitles you to software updates (which on IBM i are called Program Temporary Fixes, or PTFs) and new releases of the operating system. These PTFs provide a variety of new functions and enhancements across the IBM i portfolio, e.g., security, database, and high availability enhancements, performance improvements, and more.

6.3.8 CBU and DBM Licensing Models for IBM i

In addition to the IBM i licensing model described above. IBM has additional licensing models that IBM i customers.

- ▶ Capacity Back Up (CBU)

The CBU offering provides the ability for a customer, when ordering a new IBM i server with the CBU feature, that will facilitate the temporary transfer of IBM i license entitlements between an IBM i production server and the CBU. The production and CBU servers must both be owned by the same company. CBU provides a cost-effective backup server For information on the CBU offering, refer to [Capacity Backup](#).

- ▶ Dedicated Backup Machine (DBM)

This is licensing scenarios that facilitates backup environments. One key difference is that a DBM licensed machine does not have to belong to the same company as the production server for which it is backing up. Another difference between CBU and DBM machines is that for DBM server, IBM i license entitlement is acquired for all cores which use IBM i on the server.

For information regarding temporary keys for IBM i Licensed Program Products (LPPs) for CBU and DBM, refer to Request for designation of a primary and backup pairing at <https://www.ibm.com/support/pages/ibm-power-systems-capacity-backup>

6.3.9 Upgrading an existing IBM i license to 7.6

If you have active IBM i SWMA, you can upgrade your IBM i license to 7.6 at no extra charge.

If you have IBM i Subscription licenses at 7.5 or earlier, for upgrade information, refer to the [IBM i Subscription Transformation FAQ](#), and search for “upgrade”.

If you have IBM i non-expiring licenses at 7.5 or earlier, upgrade IBM i via the [IBM Entitled Systems Support \(ESS\) Website](#), using **My Entitled Software → Software Updates** function. Here you will select your IBM System via IBM customer number server model and serial number, then take the option to upgrade it from the current release to 7.6. Once this process completes, you will be directed to the Software Download section of this site and you can download your installation media, ready to install or upgrade your system.

Once you have upgraded your IBM i non-expiring or subscription licenses to 7.6, use the ESS website to download copies of your IBM i license keys for 7.6. Until license keys are installed, a 70-day evaluation period is available. Failure to install the license keys can result in the IBM i and/or LPPs being no longer usable until the appropriate license keys are installed.

Your licensing model will not change by the upgrade process. Regardless of whether you have a non-expiring or subscription license before the upgrade, you will have the same license type afterwards.

When performing the upgrade to IBM i 7.6 on your server, upgrades to IBM i 7.6 are supported from IBM i 7.4 and 7.5.

6.3.10 IBM i Simplification

IBM i and many of the IBM i portfolio LPPs require license keys to assure that the software is properly licensed. With recent simplification, the number of products requiring keys has been reduced. For more information on the IBM i Simplification, visit [IBM i Portfolio Simplification](#)

6.4 Linux on IBM Power

Linux is a powerful, open-source, cross-platform operating system that runs on a broad spectrum of hardware – from embedded devices to mainframes – delivering a consistent, UNIX-like environment across diverse architectures. Its compatibility with IBM Power offers a flexible and cost-effective alternative for running a wide range of applications, while taking full advantage of the platform's renowned performance, availability, and reliability.

As a free and open-source solution, Linux significantly reduces total cost of ownership 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 high portability allows it to run seamlessly on architectures like x86, ARM, and IBM Power, and it is supported by a vast, active global community that provides extensive documentation, tools, and peer support.

IBM has been a leading advocate for Linux for over two decades, integrating it deeply into its enterprise ecosystem – including IBM Power, IBM Z® mainframes, and IBM Cloud. The company collaborates closely with major Linux distributions such as Red Hat, SUSE, and Canonical to provide optimized performance and seamless integration. Through initiatives like LinuxONE and ongoing contributions to the Linux kernel and open-source projects, IBM continues to position Linux as a cornerstone of secure, modern infrastructure.

Security is another key strength of Linux. Its robust user privilege model, combined with community-driven updates and tools like SELinux and AppArmor, helps protect systems from vulnerabilities and malware. Linux is also highly efficient and lightweight, capable of running on everything from low-powered IoT devices to high-performance enterprise servers. Its modularity allows users to tailor the system to their specific needs – whether that means a minimal installation or a fully featured desktop environment.

Additionally, Linux provides access to a vast ecosystem of tools and software, some of which are exclusive or better supported on Linux than on other operating systems. Its open-source nature allows for greater flexibility and adaptability, making it easier to align with specific business or technical requirements. Linux also serves as a solid foundation for hybrid cloud infrastructure, supporting application modernization and deployment at scale. Its scalability, reliability, and cost-effectiveness make it especially well-suited for large enterprise environments and mission-critical workloads.

6.4.1 Supported distributions

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

Red Hat Enterprise Linux

RHEL on IBM Power servers delivers excellent scalability and performance for high-demand workloads such as big data, cloud, and SAP HANA. Power servers efficiency allows organizations to optimize resource usage and reduce costs by using fewer processor cores.

Red Hat can be deployed in both on-premises and cloud environments, giving organizations full control over their data and infrastructure with the flexibility to choose their preferred platform.

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

At launch, IBM Power11 servers support:

- ▶ RHEL 10 with native Power11 support, fully leveraging 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 leveraging POWER® hardware's performance, reliability, and availability. This section outlines supported enterprise Linux distributions. While other distributions may run on Power, only select enterprise versions offer formal support.

Red Hat Enterprise Linux (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, 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 time lines 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.



Figure 6-1 Red Hat Enterprise Linux support lifecycle

Note: The Red Hat Enterprise Linux life cycle phases are designed to minimize changes within each major release over time and ensure predictable availability and content.

- ▶ Full Support Phase

During this phase, Red Hat provides updates for:

- Security issues (CVEs with CVSS greater or equal to 7) via RHSAs
- Urgent and select high-priority bugs via RHBAs
- Additional errata as needed

New or improved hardware support and select software enhancements may 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 releasing:

- Security updates (CVEs with CVSS ??)
- Urgent and selected high-priority bug fixes

New features and hardware support are not included during this phase.

► Extended Life Phase

Customers retain access to existing content via 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 is 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 every six months during the Full Support Phase. Details on Extended Update Support (EUS) and SAP-specific services are included per release. Figure 6-2 shows the planned release schedule for RHEL8.

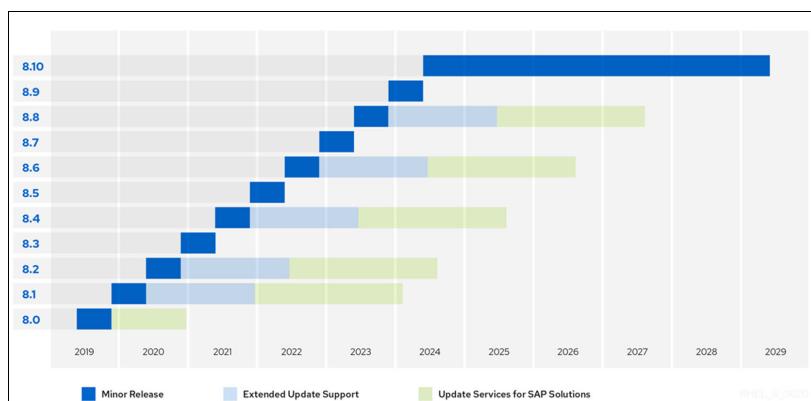


Figure 6-2 Release schedule for RHEL8

Figure 6-3 shows the planned release schedule for RHEL9.

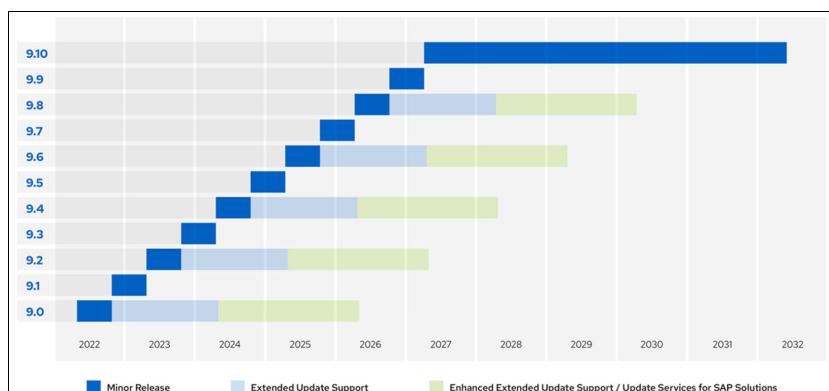


Figure 6-3 Release schedule for RHEL9

Figure 6-4 shows the release schedule for RHEL10.

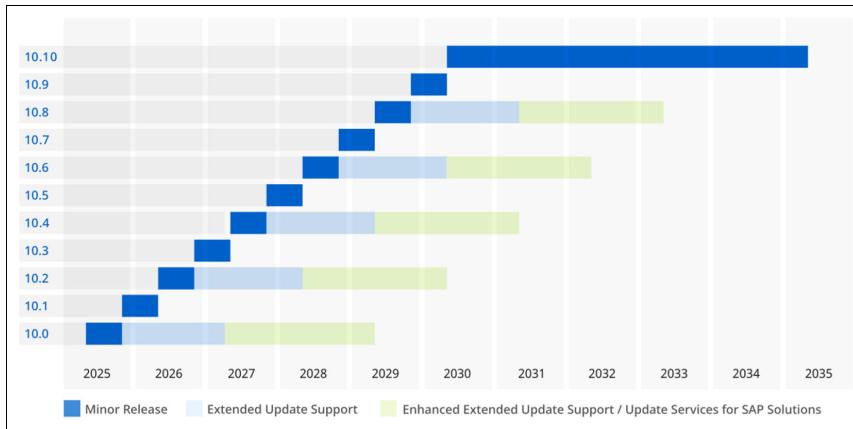


Figure 6-4 Release schedule for RHEL10

For additional details regarding the RHEL roadmap, please refer to the [Red Hat Enterprise Linux Life Cycle](#). To find the latest Red Hat certified IBM Power servers, please refer to 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 allows it to access all of the features of the Power11 processor and platform.

At announcement, the Power E1180 servers support the following minimum levels of the SUSE Linux Enterprise Server operating system:

- ▶ SUSE Linux Enterprise Server 15 Service Pack 6, or later (Power11 native)

SUSE roadmap

SUSE Linux Enterprise Server (SLES) is designed with long-term enterprise stability and support in mind. Its product lifecycle provides predictable and reliable maintenance phases, enabling organizations to plan deployments, updates, and migrations effectively.

- ▶ Lifecycle Duration

SLES offers a total lifecycle of 13 years per major release, 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 requiring continuous updates and active support.
- Extended Support (3 years): After General Support ends, Extended Support offers critical security updates and selected bug fixes, allowing customers more time to transition 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 to 14 months, these include incremental updates, feature enhancements, and hardware enablement.

Each Service Pack is supported for six 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 service pack schedule.

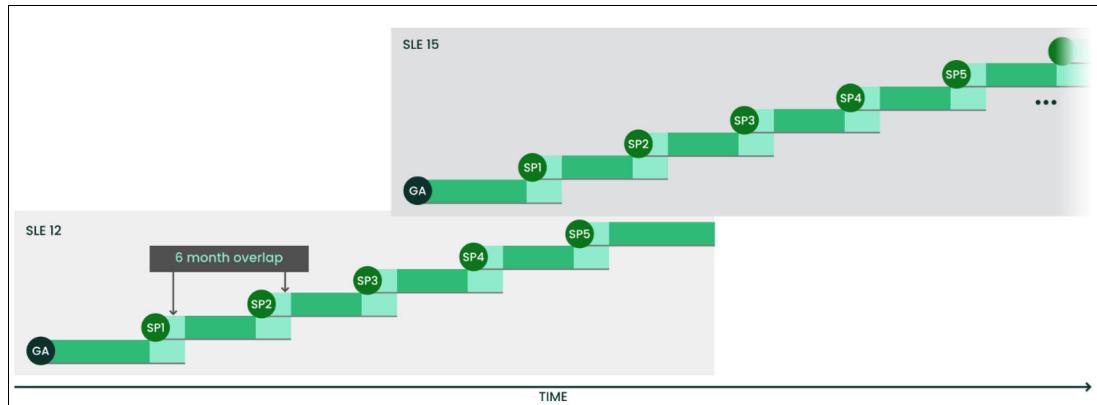


Figure 6-5 SLES major releases and service packs

Figure 6-6 shows the release lifecycle, including long term service pack support.

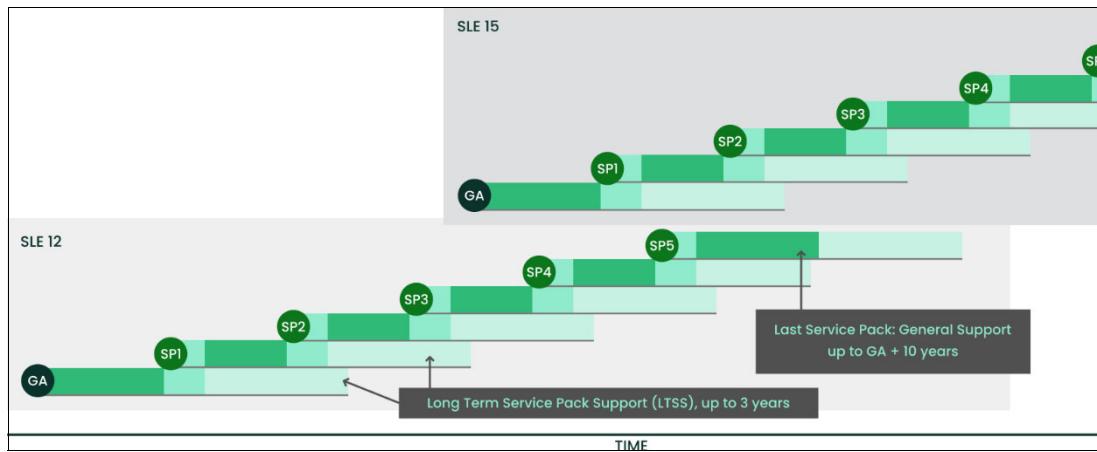


Figure 6-6 SLES long-term service pack support.

For additional details regarding the SUSE roadmap, please refer to the [SUSE Linux Enterprise Server Documentation](#) and [Product Support Lifecycle](#). To find the latest SUSE certified IBM Power servers, please refer to 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 the IBM Advance Toolchain for Linux version 15.0, which allows customers and developers to use all new Power11 processor-based technology instructions when programming. Cross-module function call overhead was reduced because of a new PC-relative addressing mode.

One specific benefit of Power11 technology is a 10x to 20x advantage over Power9 processor-based technology for AI inferencing workloads because of increased memory bandwidth and new instructions. One example is the new special purpose-built matrix math accelerator (MMA) that was tailored for the demands of machine learning and deep learning inference. It also supports many AI data types.

6.4.2 Licensing

Linux licensing on IBM Power follows the same foundational principles as on other platforms, but with considerations tailored to enterprise environments. The Linux kernel itself is licensed under the GNU General Public License version 2 (GPLv2), which helps the software remain free and open-source, allowing users to run, study, modify, and distribute it. On IBM Power, several major Linux distributions are supported, including Red Hat Enterprise Linux (RHEL), SUSE Linux Enterprise Server (SLES), and Ubuntu. Each of these distributions adheres to open-source licensing models but may differ in how they package and deliver their software.

For example, RHEL and SLES provide access to source code under open-source licenses but require a paid subscription for access to precompiled binaries, updates, and enterprise support. This model allows organizations to benefit from open-source flexibility while receiving the stability and support needed for mission-critical workloads.

On IBM Power, these distributions are optimized to take advantage of the architecture's performance, scalability, and reliability features, and licensing terms typically include support for virtualization technologies like PowerVM.

Red Hat Licensing

Red Hat Enterprise Linux is sold on a subscription basis, with initial subscriptions and support available for one, three, or five years. Support is available directly from Red Hat or IBM Technical Support Services.

Red Hat Enterprise Linux 8 for Power LE subscriptions covers up to four cores and up to four LPARs, and can be stacked to cover a larger number of cores or LPARs.

When you order RHEL from IBM, a subscription activation code is automatically published in Enterprise Storage Server. After retrieving this code from Entitled Systems Support (ESS), you use it to establish proof of entitlement and download the software from Red Hat.

Red Hat Licensing information can be found [here](#).

SUSE Linux Enterprise Server Licensing

SUSE Linux Enterprise Server is sold on a subscription basis, with initial subscriptions and support available for one, three, or five years. Support is available directly from SUSE or from IBM Technical Support Services.

SUSE Linux Enterprise Server 15 subscriptions cover up to one socket or one LPAR, and can be stacked to cover a larger number of sockets or LPARs.

When you order SLES from IBM, a subscription activation code is automatically published in Entitled Systems Support (ESS), you use it to establish proof of entitlement and download the software from SUSE.

Clients are required to register the Linux offering purchased at the Distributor's website with the activation code. After registered, clients are able to download the software packages electronically and obtain the latest upgrades available for the product purchased.

SLES Licensing Information can be found [here](#).

6.4.3 KVM support

KVM stands for Kernel Virtual Machine, and it's a widely spread virtualization technology in X86_64, and also can be used in Power Servers. Even when IBM remains committed to the PowerVM being the premier enterprise virtualization software in the industry.

With KVM on Power, IBM will be targeting x86 customers on entry servers but will offer both KVM and PowerVM to meet the varying virtualization needs Power Linux customers. However, KVM virtualization technology represents an opportunity to simplify customer's virtualization infrastructure with a single hypervisor and management software across multiple platform.

The KVM Guests may run within a PowerVM LPAR, KVM can be used to run KVM guests. These guests are essentially VMs that run on top of the LPAR, using the existing resources of the LPAR.

One of the Benefits is that this approach combines the advantages of PowerVM's virtualization capabilities with the power and flexibility of KVM.

KVM guests running in a PowerVM LPAR have a unique runtime architecture, different from other virtualization mechanisms on IBM Power. For more details on KVM running on IBM Power11 see 9.2, "KVM support" on page 187.

6.5 PowerVM Virtual I/O Server

IBM PowerVM software is a virtualization environment that can run AIX, IBM i and Linux virtual machines on IBM Power servers. Businesses are turning to server virtualization to consolidate multiple workloads onto fewer systems, increase server utilization and reduce costs. PowerVM provides a secure and scalable server virtualization environment for your applications, built upon the advanced RAS features and leading 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, ensuring the reliability of mission-critical applications. With robust automation capabilities, PowerVM accelerates service delivery by streamlining the provisioning and management of virtual machines and storage resources, making it well-suited for cloud-based infrastructures. It also enhances operational efficiency and maximizes return on investment through features like live partition mobility, which enables zero-downtime workload migration, and resource optimization tools that improve the utilization of compute and storage infrastructure.

The Virtual I/O Server (VIOS) is part of the PowerVM Editions hardware feature. The VIOS is a software that is located in a logical partition and facilitates the sharing of physical I/O resources between client logical partitions within the server.

Currently, the IBM Power11 processor-based server supports the following minimum levels of the PowerVM VIOS operating system:

- ▶ 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 operating systems originally compiled for earlier generations of Power processors. This is especially important when transitioning to the latest hardware, such as Power11, where some operating systems and applications may not yet support native execution. In such cases, compatibility mode allows these workloads to run in a Power10 environment, ensuring continuity and minimizing disruption during upgrades.

The compatibility mode is defined at the logical partition (LPAR) level during the creation of the LPAR. Administrators can specify which processor generation the partition should emulate, allowing multiple workloads to operate in different compatibility modes on the same physical server. This flexibility is particularly useful in mixed environments where not all software has been updated to support the latest architecture.

While running in a previous-generation compatibility mode helps maintain application portability and eases the migration process, it may restrict access to newer processor features and performance enhancements. For this reason, IBM recommends using compatibility mode as a transitional solution. The long-term goal should be to recompile applications and upgrade operating systems to fully leverage the advanced capabilities and performance improvements of the latest Power processor architecture.

Figure 6-7 shows how to set compatibility mode in your LPAR definition.

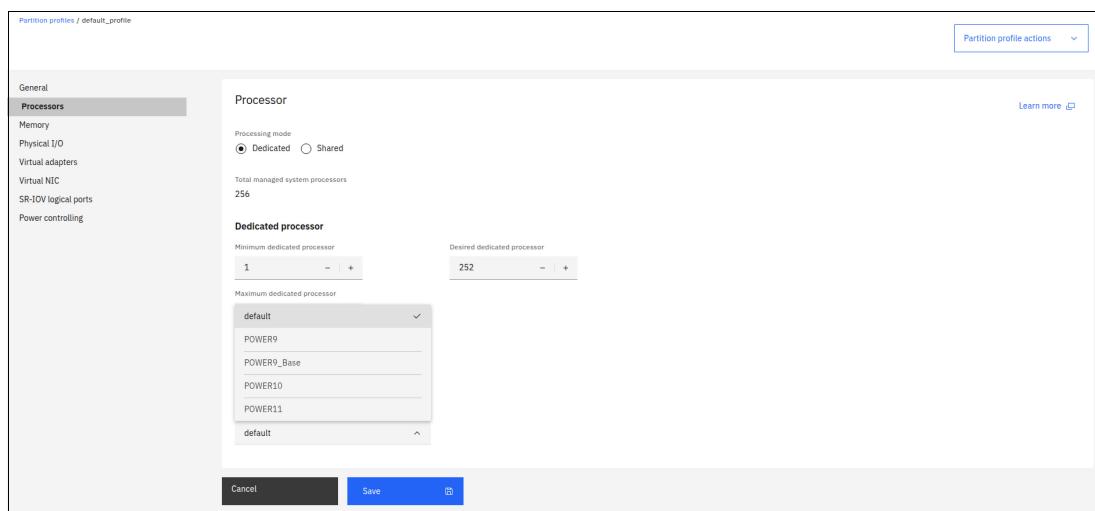


Figure 6-7 Set Lpar processor compatibility mode.



Enterprise Solutions

IBM Power servers are engineered to support a wide range of enterprise workloads that demand high availability, 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 RAS (Reliability, Availability, and Serviceability) features and capabilities like live partition mobility, Power servers provide minimal downtime and seamless workload management. They are also well-suited for enterprise resource planning (ERP) and customer relationship management (CRM) systems, offering the processing power and memory capacity needed to handle large-scale, real-time data operations.

In addition to traditional workloads, IBM Power servers are optimized for modern enterprise needs, including artificial intelligence (AI), machine learning, and hybrid cloud deployments. Built-in accelerators and support for containerized environments like Red Hat OpenShift allow organizations to run AI models and cloud-native applications efficiently. These servers also excel in high-performance computing (HPC) scenarios, such as financial modeling, scientific research, and healthcare analytics, thanks to their high core counts and memory bandwidth. With robust security features like transparent memory encryption and secure boot, IBM Power is a trusted platform for industries with strict compliance requirements. Overall, Power servers provide 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
- ▶ Health Care

7.1 High Availability and Disaster Recovery Solutions

IBM offers a suite of high availability and disaster recovery (HA/DR) solutions tailored to its IBM Power platform, each designed to meet different operational needs and workloads. PowerHA SystemMirror for AIX is a clustering solution that enhances 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 tightly integrated HA/DR capabilities using features like Independent Auxiliary Storage Pools (I ASPs) and real-time data replication to protect mission-critical workloads. Complementing these is VM Recovery Manager (VMRM), a flexible solution that supports AIX, IBM i, and Linux workloads by managing full LPAR recovery across systems. VMRM is ideal for environments where full system recovery is acceptable, offering automated failover and disaster recovery without requiring application-level configuration. Together, these tools provide a comprehensive HA/DR strategy across the IBM Power ecosystem.

7.1.1 PowerHA SystemMirror for AIX

PowerHA SystemMirror for AIX, generally known as PowerHA, is a high availability clustering solution to enhance application uptime and provide fast recovery. It is considered an active-standby or active-passive system of application availability. It is also known as an application restart model, where an application and its resources (a file system with data, an IP address, application scripts, a volume group, etc.) are logically combined and operate as a single entity on a VM or LPAR (known as a “node”) in a cluster. That group of resources are moved as a unit to a standby VM or LPAR, and the application is restarted. In the event of a planned outage - PowerHA can gracefully quiesce the application on the primary node. In any case, the outage lasts just as long as it takes to acquire the resources and restart the application - this might be measured in minutes. It is also considered to be a warm standby method of high availability, in that the target node already has a running version of AIX.

PowerHA for AIX can also be used across data centers, where data is replicated through the storage subsystem or, using the AIX Geographic Logical Volume Manager (GLVM) subsystem, over IP to the target node.

In any case - within or across data centers - the failover process can be automated to reduce application outages. Applications can be moved to another running node to allow for maintenance, allowing for near-zero downtime – although the PowerHA application can be maintained with a live kernel update.

PowerHA for AIX is tightly integrated with the AIX operating system's inherent reliability, and IBM Power's hardware features such as virtual I/O, concurrent hardware maintenance, dynamic resizing of LPARs, etc. PowerHA supports critical enterprise apps, like Oracle, SAP, Db2, and Epic systems - all common workloads. Licensing is on a n+1 model per cluster, where n is the number of cores in the production online copy of the application, with 1 additional core added for the standby. For example, on a 24-core Power server, if two clusters are created, one supporting an LPAR with 4 cores and another with 5, a total of 11 cores will need to be licensed – 4+1=5 and 5+1=6.

7.1.2 IBM PowerHA SystemMirror for i

PowerHA for IBM i is IBM's high-availability and disaster recovery (HA/DR) solution designed specifically for the IBM i operating system. It provides automated failover capabilities and data replication to provide business continuity in the event of planned or unplanned outages. PowerHA integrates tightly with IBM i's native features, such as the Independent Auxiliary

Storage Pool (IASP), 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 using technologies like Geographic Mirroring or IBM's Storage-based replication. This means that critical applications and data remain available even if the primary system becomes unavailable. PowerHA supports both local high availability within a single data center and remote disaster recovery across geographically dispersed locations, offering flexibility for various business continuity strategies.

PowerHA also includes tools for monitoring, automation, and management, making it easier for IT administrators to configure and maintain HA/DR environments. With features like role-based access, cluster resource groups, and automated failover policies, PowerHA helps reduce the complexity of managing high-availability systems. It is a key component for organizations that rely on IBM i to run mission-critical workloads and require continuous access to their data and applications.

7.1.3 VM Recovery Manager (VMRM)

VM Recovery Manager (aka VMR or VMRM) operates on an active-inactive model, where the entire definition of an VM or LPAR is recreated onto a target system and then restarted. It supports any workload running on IBM Power - AIX, IBM i, and Linux. Recovery will take longer than an application restart model because the entire LPAR must be restarted after being redefined, instead of just restarting the application.

In HA mode, VMRM uses the Live Partition Mobility to dynamically move an LPAR to a system that shared an HMC. In DR mode, it will manage the data replication process from the storage provider to a set of disks at the target system. The LPAR only exists in one place at any given time, except in some limited circumstances where a disaster exercise can be tested on the target system.

As with PowerHA, VMRM helps automate disaster recovery - it can detect when a VM or entire system is down and initiate policy-based failover to a single target or to a group of targets. It can coexist with PowerHA - where a cluster of LPARs can be failed as a group over to systems at an alternate site. It is often used to provide simple DR support for Linux VMs - no software installation or OS configuration required; the Linux VM will be restarted.

The ideal use cases for VM Recovery Manager are development, test, and sandbox environments that may not require the same RTO requirements as production systems, and Linux on Power. It integrates well with solutions like SAP HANA and Oracle. No configuration work is required on the individual operating systems being managed, although there are some agents that can be installed to check for application failures and trigger a failover of the LPAR.

Similarly to PowerHA; it is licensed based on the number of production cores in use, not on an entire system. However, unlike PowerHA, no standby licenses are required. On a 24-core system, if the VMs needing VMRM support have 9 cores, then only 9 cores need to be procured.

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 high availability while keeping operational costs under control.

By combining IBM Db2 with IBM Power Servers 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 leverages cutting-edge database innovations to maximize efficiency and throughput, earning top rankings in industry benchmarks such as TPC-C, TPC-H, and SAP SD 3-Tier. IBM Power provides the robust hardware foundation, offering superior price/performance and the ability to handle over a million transactions per minute at a cost of less than \$1 per transaction. Db2 automatically exploits Power11's 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 provides high availability and operational efficiency. Power servers are engineered with 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 needs, and IBM PowerVM virtualization, which allows for efficient resource utilization and workload consolidation. These capabilities reduce hardware, energy, and management costs, enabling a more agile and cost-effective IT infrastructure. Together, Db2 and IBM Power offer a powerful, scalable, and resilient platform for modern data-driven enterprises.

For more information on Db2 high availability see the IBM Redbook *IBM PowerHA SystemMirror for AIX Cookbook*, SG24-7739

7.3 Oracle

For over 35 years, clients have relied on IBM Power to deploy their Oracle database and application workloads. Organizations, both big and small, can take advantage of Power's class leading reliability and security as well as its advanced recovery, self-healing and diagnostic capabilities designed to reduce application downtime.

IBM Power11 servers, fully certified by Oracle software, can enable organizations the ability to consolidate multiple workloads on fewer servers – increasing overall system utilization and lowering overall costs. This efficiency can also lead to less Oracle licenses required. For example, IBM Power S1124 servers running Oracle Database SE2 can help reduce application cost per database instance by up to 33% compared to fourth- generation Intel Xeon scalable processors, and reduce the overall number of servers needed to improve energy costs.

The cost of security breaches continues to grow, averaging at USD 4 million per breach.² IBM Power technology is built to protect businesses against cyberthreats with end-to-end security protocols, including new transparent memory encryption that doesn't impact performance. Power servers may be considered 60 times more secure than unbranded white box servers.² IBM Power servers, combined with IBM AIX Trusted Execution (TE), you can verify the

integrity of your system and implement advance security policies to enhance the trust level of the complete system.

As data continues to rapidly expand, organizations may struggle when it comes to modernizing their IT infrastructure. IBM Power servers and the IBM AIX operating system (OS) create a solid foundation for the modernization of traditional Oracle database workloads, new application developments and workload consolidation. With the Red Hat Ansible Automation Platform, clients can manage Oracle workloads on IBM Power servers as part of their wider enterprise automation strategy.

Mission-critical workloads need a server and OS with reliability, high availability and the ability to scale without impacting mission-critical performance. IBM Power servers can offer 2.5 times better core than compared to x86 servers, enabling Oracle workloads to grow linearly without hitting bottlenecks.

Oracle certifies its products on IBM Power, delivering a host of benefits – including comprehensive end-to-end support, portability and efficiency. IBM Power servers provide 99.999% of reliability to maintain maximum availability. The combined design of the AIX OS on Power servers with IBM PowerHA technology can bring clients stunning 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 your critical Oracle workloads protected and available while reducing costs.²
- ▶ Simplified management
 - Automatically deploy a “cloud ready” OS capable of meeting any organization’s private cloud requirements with IBM PowerVC.
- ▶ Unmatched uptime
 - IBM Power supports the most demanding workloads and provides 99.999% of reliability to maintain maximum availability.¹
- ▶ Improved workload cost-effectiveness
 - Leverage Power LPAR and DLPAR eligibility for Oracle hard partitioning, clients can license only the Power cores available to Oracle software.
- ▶ Standards-based automation
 - Improve manageability and scalability ensuring consistent and repeatable outcomes by leveraging our enhanced and expanded automation portfolio for Oracle workloads, built on Ansible.
- ▶ Data Protection
 - Benefit from Top-to-Bottom security in Power with trusted boot, main memory encryption, run-time verification of OS files, and role-based access and execution control with PowerSC

For additional information on running Oracle on Power reference the IBM Redbook publication *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 suitable for many business needs and can be used on IBM Power Servers. SE2 offers a full-featured database with ease of use, power, and performance, including features like relational, JSON, and XML data handling, and 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.

Key aspects of using Oracle Standard Edition 2 on IBM Power servers

Here are some key benefits of running Oracle Standard Edition 2 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 Standard Edition 2 is compatible with IBM Power, and IBM has even designed specific Power10 servers (Power S1012 and Power S1014) to compete with x86 engines in the two-socket space, catering to Oracle's need for SE2 to be cost-competitive with SQL Server. It is expected that IBM will offer a Power11 server in the future that will also qualify for Oracle Standard Edition2 licensing.

Benefits of IBM Power for Oracle Workloads:

IBM Power offers several advantages for Oracle databases, including main memory encryption, runtime verification of OS files, role-based access control, and advanced recovery, self-healing, and diagnostic capabilities.

- ▶ High Availability:

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 Standard Edition Oracle Databases 19c.

- ▶ Other Features:

SE2 offers a range of features for building business applications, including support for relational, JSON, XML, spatial, graph, and unstructured data, as well as Oracle Multi-tenant Architecture.

Licensing and implementation

Consider the following as you plan for your Oracle Standard Edition 2 implementation:

- ▶ While the number of cores in each socket does not affect the license price, it is important to adhere to the maximum number of CPUs allowed for the cluster, which is not per node.
- ▶ Standard Edition High Availability provides cluster-based failover for single-instance Standard Edition Oracle Databases using Oracle Clusterware.
 - Oracle Standard Edition High Availability benefits from the cluster capabilities and storage solutions that are already 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 as well as for unstructured data, enables Oracle Grid

- Infrastructure to restart an Oracle Database on a failover node much faster than any cluster solution that relies 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 has a variety of options to meet clients where they are on their journey to SAP S/4HANA. Whether you are considering SAP HANA or the next-generation S/4HANA, IBM Power offers on-premises, off-premises or fully-managed (RISE with SAP) solutions, built to run mission-critical applications like SAP, help accelerate your ERP and application deployments, and help maximize the impact they can have on your data management, data integration, automation and business processes.

IBM and SAP have partnered for over 50 years:

- Over 4,800 clients are running SAP HANA on IBM Power servers
- Over 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 current infrastructure while simultaneously supporting digital transformation. IBM Power servers are purpose built for data-intensive applications such as SAP HANA and S/4HANA that require large amounts of in-memory computing but still let you maintain the high availability and flexibility required for your hybrid cloud.

Benefits of running SAP HANA on IBM Power

With global data volumes set to grow to more than 180 zettabytes in 2025, organizations across every sector are facing tremendous pressure to manage, process, store and extract valuable insights from their critical data. By running SAP HANA on IBM Power servers, businesses can:

- Provision Faster
 - Simplify system management and boost business agility. 0.01 cores, 1GB memory Create new environments flexibly by allocating incrementally starting as low as 0.01 cores and 1GB memory.
- Maximize uptime
 - Minimize disruption to business-as-usual activities. #1 Best-in-class reliability for 15 years. 99.999% Five-nines server reliability score achieved during independent testing. 2x Two-times better memory RAS than Industry-Standard DIMMs
- Cut Energy Usage
 - Reduce datacenter costs and enhance environmental sustainability. IBM Power provides better performance while using less energy when compared to existing x86 based solutions.
- Scale affordably
 - Reduce the risk of over-provisioning with flexible hybrid cloud solutions, instant scaling, and pay-per-use consumption options. 40TB Scale up capacity – the largest supported for SAP S/4HANA and SAP BW.

- ▶ Strengthen security

Protect critical data and applications from cyberthreats with end-to-end security, including new transparent memory encryption with no performance impact. 60x More secure than unbranded commodity servers.

- ▶ Gain Faster Insights

Make rapid decisions to maximize business efficiency. 2.5x Better per core performance than compared x86 servers.

7.5 Banking

As financial institutions accelerate digital transformation, IT architects are tasked with designing platforms that are secure, scalable, and AI-ready. IBM Power Servers – particularly Power10 and Power11 based models – offer 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 legacy 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, enabling agility, regulatory compliance, and cost optimization. IBM Power supports hybrid cloud deployments through:

- ▶ 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 utilizing AI on IBM Power

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 need to 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 are deployed to IBM Power Virtual Server (PowerVS) for inference, enabling scalable, real-time decision-making in the cloud. This hybrid approach is ideal for applications like real-time credit scoring or fraud detection, where data privacy is paramount but rapid response times are also critical.

- ▶ End-to-End AI on Power Virtual Server with Secure On-Premises Data Access

In this model, banks run the full AI lifecycle – training, tuning, and inference – on IBM PowerVS, while maintaining secure access to on-premises data sources. This setup allows institutions to leverage 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 can provide compliance with data governance policies while accelerating insights and reporting.

- ▶ 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 can operate across environments. For instance, 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 across geographies or business units.

- ▶ **AI for Risk and Compliance on IBM Power**

Global Tier-1 banks are increasingly turning to 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 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 can deliver the most immediate business value.

Integrating IBM watsonx with IBM Power

IBM watsonx is a modular AI and data platform 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 IBM PowerVS.

Here are the components of watsonx:

- ▶ `watsonx.ai`: Model training and inference
- ▶ `watsonx.data`: Data lakehouse integration
- ▶ `watsonx.governance`: Model risk and compliance
- ▶ `watsonx.orchestrate`: AI agents and workflow automation

IBM Power Servers are high-performance servers optimized for data-intensive workloads. The integration between watsonx and IBM Power enables:

- ▶ **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, allowing businesses to keep sensitive data on-premises while leveraging cloud-native AI capabilities.
- ▶ **Enterprise-Grade Security and Reliability:** IBM Power provides robust security and uptime, aligning with watsonx's focus on trustworthy AI.
- ▶ **Tight Integration with Red Hat OpenShift:** Both watsonx and IBM Power support OpenShift, enabling containerized AI workloads to run seamlessly across environments.

7.6 Health Care

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 empowers healthcare providers to improve patient outcomes, enhance operational efficiency, and reduce risk.

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, role-based access control, and encryption at rest and in transit. With quantum-safe firmware signing and encrypted Live Partition Mobility (LPM), IBM Power provides data integrity and privacy. According to ITIC, organizations using IBM Power experience an average of just 3.3 minutes of unplanned downtime per year due to security issues – demonstrating its industry-leading resilience.

Unmatched Reliability for Clinical Systems

IBM Power is engineered for continuous availability of mission-critical applications such as EHRs, PACS, and healthcare information systems. Features like predictive failure analysis, 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. IBM Power has held the title of the most reliable non-mainframe server platform for over 15 years.

High-Performance Computing for Healthcare Analytics

Healthcare workloads – from medical imaging to genomics and real-time analytics – demand exceptional compute power. Power11-based servers deliver up to 2.5x more performance per core than previous generations, 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 allows 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 sub-second inferencing latency – ideal for applications like diagnostic imaging, predictive analytics, and treatment optimization.

Seamless Integration with Healthcare Ecosystems

IBM Power integrates with a wide range of healthcare applications, including EMRs, imaging platforms, analytics tools, and telemedicine systems. It supports interoperability and data exchange across systems, and works closely with leading ISVs such as Epic to certify hardware for industry-standard healthcare software.

7.6.1 Epic

Epic is a Healthcare Information System (HIS) provider that develops and delivers a comprehensive electronic medical record (EMR) system covering all aspects of the medical healthcare profession. The Epic solution includes a variety of applications that cover areas, such as medical billing, emergency room, radiology, outpatient, inpatient, and ambulatory care.

IBM and Epic Systems maintain a strategic partnership that enables healthcare organizations to run Epic's electronic health record (EHR) platform on IBM infrastructure, particularly IBM Power solutions and IBM Storage solutions. This collaboration is designed to support the performance, security, and compliance needs of healthcare providers managing mission-critical clinical workloads.

IBM Power Servers are certified to run Epic Operational Database (ODB) workloads, with best practices published for tuning AIX environments to meet Epic's stringent performance requirements. These configurations are optimized for high throughput, low latency, and high availability – critical for real-time clinical operations. Currently, 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 backend services tier. These platforms are used to host InterSystems Caché or IRIS for Health, which serve as the core databases behind Epic's EHR system.

It is expected that IBM and Epic will partner to certify IBM Power11 configurations after the appropriate testing is completed.

Supported Products

Here is a list of currently supported Epic products:

- ▶ InterSystems Caché / IRIS for Health
 - Officially supported by Epic on both AIX and Linux (RHEL) on Power10.
 - Deployed in production for the Epic operational database (Chronicles).
- ▶ Epic Production and Reporting Database Tiers
 - Most often run on AIX or RHEL on IBM Power10 for performance and reliability.
- ▶ Other backend services, such as database copies used for Clarity extracts, may also run on AIX or Linux, depending on the customer's architecture.

Note: The Epic application layer (Hyperspace, Interconnect, etc.) typically runs on Windows or Linux on x86. IBM Power-based AIX or Linux deployments are focused on the infrastructure and database tiers.

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:
 - Red Hat Enterprise Linux (RHEL) 8.6 or later on Power10 (ppc64le)
 - Certified by InterSystems for IRIS for Health on Power.
 - Required for newer installations using 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 using OpenShift on Power for modern container-based workloads.
- ▶ Storage:
 - High-performance SAN such as IBM FlashSystem is commonly used, with tuning for Caché/IRIS.
- ▶ HA/DR:
 - PowerHA for AIX,
 - Linux HA tools (including IBM VM Recovery Manager),
 - InterSystems mirroring for resilience.



Servicing Power11

The goal of serviceability is to enable efficient system repair while minimizing or avoiding any disruption to operations. It encompasses system installation, upgrades or downgrades (Miscellaneous Equipment Specification or MES), and ongoing maintenance or repair activities. Depending on the system configuration and warranty agreement, service tasks may be performed by the client, an IBM technician, or an authorized service provider.

IBM Power Servers are designed with advanced serviceability features to support a highly efficient maintenance environment. These features help 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 Expert Care
- ▶ IBM tools and interfaces
- ▶ FSP
- ▶ ASMi
- ▶ Entitled System Support
- ▶ Firmware Updates

8.1 IBM Maintenance

IBM offers a comprehensive global maintenance and support framework for its IBM Power Servers, designed to provide high availability, rapid issue resolution, and minimal disruption to business operations. Maintenance services are available under various service level agreements (SLAs), tailored to meet the needs of different environments – from standard business hours to mission-critical 24/7 operations. IBM's support infrastructure includes remote diagnostics, on-site service, and proactive monitoring, all backed by a global network of skilled service professionals and parts depots.

Customers can choose between 8x5 and 24x7 service coverage. The 8x5 option provides support during standard business hours (typically 8 a.m. to 5 p.m., Monday through Friday, excluding holidays), which 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 have significant operational or financial consequences.

IBM also defines response time targets based on the selected service level. For example, under a 24x7 agreement, IBM may commit to a 2-hour or 4-hour on-site response for critical hardware issues, depending on the location and contract terms. These response times are supported by IBM's extensive global logistics and service infrastructure, which includes strategically located parts centers and field engineers in over 170 countries.

This worldwide coverage means that IBM Power customers receive consistent, high-quality support regardless of their geographic location. IBM's maintenance services are further enhanced by features such as call-home diagnostics, automated error reporting, and remote problem determination, which help 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 Expert Care

Managing your IBM Power environment should be seamless and efficient. Power Expert Care delivers immediate access to a curated bundle of high-value services trusted by the Power community – eliminating the delays and complexity of traditional procurement processes. With this preselected service package, clients benefit from streamlined support, enhanced by AI-driven tools that accelerate response times, improve case resolution, and elevate overall satisfaction.

What sets Power Expert Care Premium apart is its focus 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 provides infrastructure support with unmatched technical depth and operational excellence.

Key Features of the Premium Bundle:

- ▶ 30-minute response time for both hardware and software support cases
- ▶ 4-hour on-site response target for urgent issues
- ▶ System and microcode compatibility guidance to provide optimal performance
- ▶ Dedicated Technical Account Manager and mission-critical support resources

- ▶ Automated case creation and log analysis for faster issue resolution
- ▶ Health checks for hardware, OS, and applications (available as add-ons)
- ▶ Zero planned downtime for maintenance support with proactive planning and TAM assistance

For non-mission-critical systems IBM offers the Advanced Expert Care bundle providing 24x7 standard response times on repair and maintenance cases.

Technical Account Manager

The Technical Account Manager (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 the client's business hours, with support in other languages available upon request and mutual agreement, subject to availability.

TAMs help streamline operations and reduce downtime through a wide range of responsibilities, including:

- ▶ Enabling Call Home for proactive error reporting, Autonomous Error Resolution (AER), and Zero Planned Downtime (ZPD) – available on Power11
- ▶ Activating Support Insights for predictive analytics and delivering monthly reports
- ▶ Providing firmware and microcode compatibility analysis (exclusive to Power11 Premium clients)
- ▶ Sharing software lifecycle and roadmap updates
- ▶ Delivering HIPER alerts to help avoid high-impact issues
- ▶ Offering best practices 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 (on-site code load 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:

- ▶ Advises on SAP HANA best practices
- ▶ Coordinates troubleshooting across the full software/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 high availability and system reliability. At the heart 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 like call-home support and automated error reporting further streamline the support process by alerting IBM support teams to critical events – often before they impact operations.

The service processor can be accessed and managed through the HMC, but also directly via the Advanced System Management Interface (ASMI), which provides low-level control and additional troubleshooting capabilities. For modern, automated environments, the service processor also supports DMTF Redfish 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 FSP

The Flexible Service Processor (FSP) is a dedicated, always-on management controller embedded in the IBM Power E1180 system. It operates independently of the main system processors and operating system, providing out-of-band management capabilities. This allows administrators to monitor, configure, and service the system even when the operating system is not running or the system is in a fault state.

The Power E1180 is equipped with redundant system master FSPs, housed within the System Control Unit (SCU). These FSPs operate in an active-standby configuration, ensuring continuous system management:

- ▶ Automatic Failover: In the event of a failure in the active FSP, the standby FSP takes over seamlessly, maintaining uninterrupted control and monitoring.
- ▶ Resilience: Both FSPs are capable of managing the system independently, eliminating single points of failure in the management plane.
- ▶ Enhanced Uptime: Redundancy supports 24/7 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:
 - Power control and consumption monitoring
 - Environmental monitoring (temperature, fan speeds, voltage)
 - Hardware error logging and alerting
 - Management of Vital Product Data (VPD)
- ▶ System Configuration and Initialization:
 - Hosts the Advanced System Management Interface (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 Hardware Management Console (HMC) and supports CIM, IPMI, and Redfish APIs for integration with modern data center tools.

8.5 ASMi

ASMi is a graphical interface that is part of the service processor firmware. The ASMi manages and communicates with the service processor. The ASMi is required to set up the service processor and to perform service tasks, such as reading service processor error logs, reading VPD, and controlling the system power.

ASMi can be accessed by using the HMC GUI or CLI or directly attaching a laptop to the FSP port with an Ethernet cable that is crossed or through a switch (the same network subnet is required). Some of the functions in the ASMi are limited only to IBM service representatives, which require celogin or dev user credentials during repair activities.

The control panel functions allow you to interface with the server. Control panel functions range in complexity from functions that display a status (such as IPL speed) to service functions that only service representatives can access.

For more information about a full list of all functions, see [this web page](#).

The HMC GUI/CLI is the main manageable interface for customers.

8.6 Entitled System Support

IBM Enterprise Storage Server is available to view and manage Power and Storage software and hardware. In general, most products that are offered by IBM Systems that are purchased through our IBM Digital Sales representatives or Business Partners are accessed on this site when the IBM Configurator is used.

The site features the following three main sections:

- ▶ My entitled software

Activities are listed that are related to Power and Storage software, including the ability to download licensed, free, and trial software media, place software update orders, and manage software keys.
- ▶ My entitled hardware

Activities are listed related to Power and Storage hardware including the ability to renew Update Access Keys, buy and use Elastic Capacity on Demand, assign or buy credits for new and existing pools in a Power Private Cloud environment (Enterprise Pools 2.0), download Storage Capacity on-Demand codes, and manage Hybrid Capacity credits.
- ▶ My inventory

Activities related to Power and Storage inventory including the ability to browse software license, software maintenance, and hardware inventory, manage inventory retrievals by way of Base Composer or 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 involves the deployment, updating, and monitoring of system firmware, including the Power hypervisor, service processor firmware, and device microcode. Administrators typically manage firmware through the Hardware Management Console (HMC) or the Advanced System Management Interface (ASMI), with support for both concurrent (live) and disruptive updates depending on the component. IBM also provides tools like Fix Level Recommendation Tool (FLRT) and IBM Electronic Service Agent to help identify and apply the latest recommended firmware levels, ensuring systems remain secure and optimized for enterprise workloads.

Updating System Firmware from the HMC

To update or upgrade the system firmware using the Hardware Management Console (HMC), follow the steps below:

1. In the System View, select the system you want to update.
2. Click on the Firmware menu.
3. Choose Update System Firmware.

The Update System Firmware Wizard will guide you through the necessary steps to complete the firmware update process, as shown in Figure 8-1.

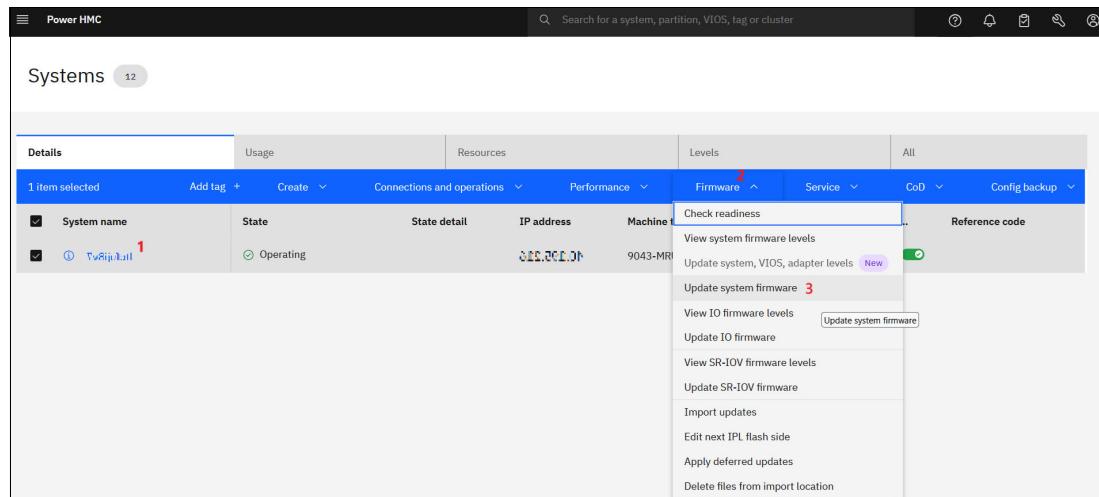


Figure 8-1 Launch the Update System Firmware wizard.

The Firmware Update Wizard guides you through the process of updating your system. It begins with accepting the license agreement, followed by performing a system health check using the Readiness Check. Next, you will provide the required firmware files and select the firmware level to be applied. The final step involves monitoring the update progress and confirming its successful completion. Figure 8-2 on page 161 through 3-6 illustrate these key steps in the update process.

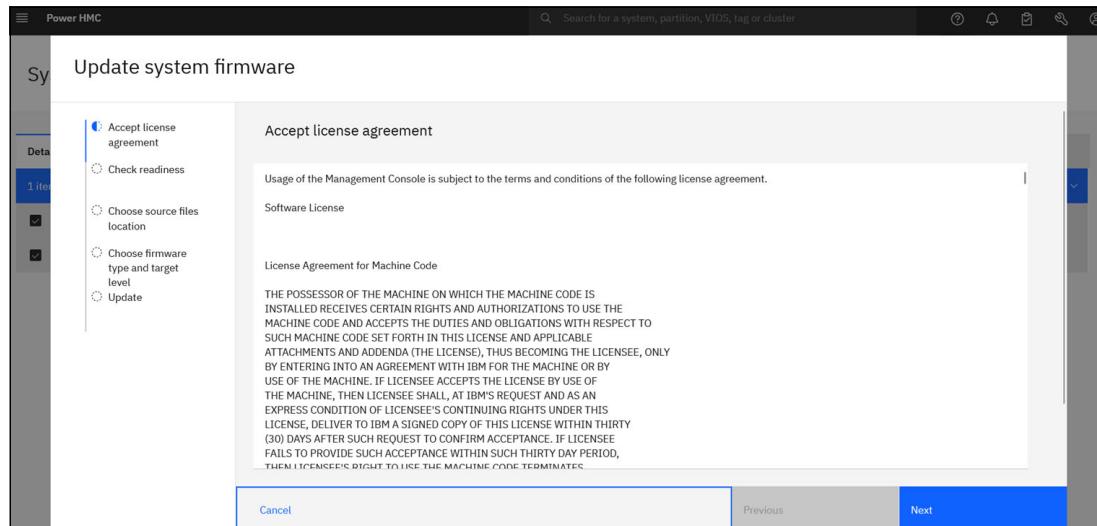


Figure 8-2 The license agreement is displayed for the user to review and accept the terms.

Figure 8-3 shows that the System Readiness Check has completed successfully for the IBM Power11 system firmware update process.

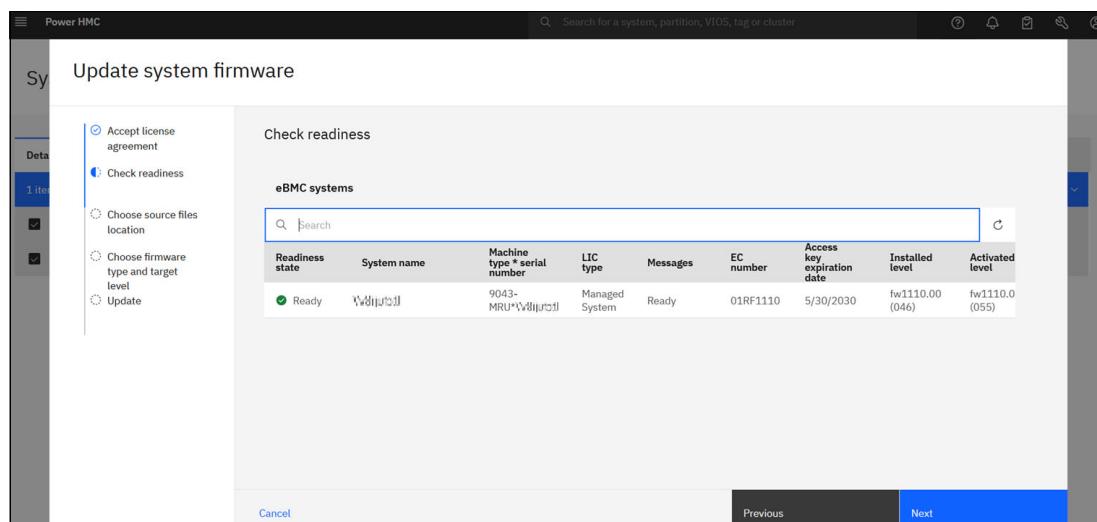


Figure 8-3 shows that the System Readiness Check completed successfully.

Figure 8-4 on page 162 illustrates the “Choose source files location” drop-down menu, which is used to update the system firmware. Several options are available for loading the required update files, including:

- ▶ IBM Website
- ▶ FTP Server
- ▶ SFTP Server
- ▶ Mount Point on the HMC
- ▶ CD/DVD
- ▶ USB Drive

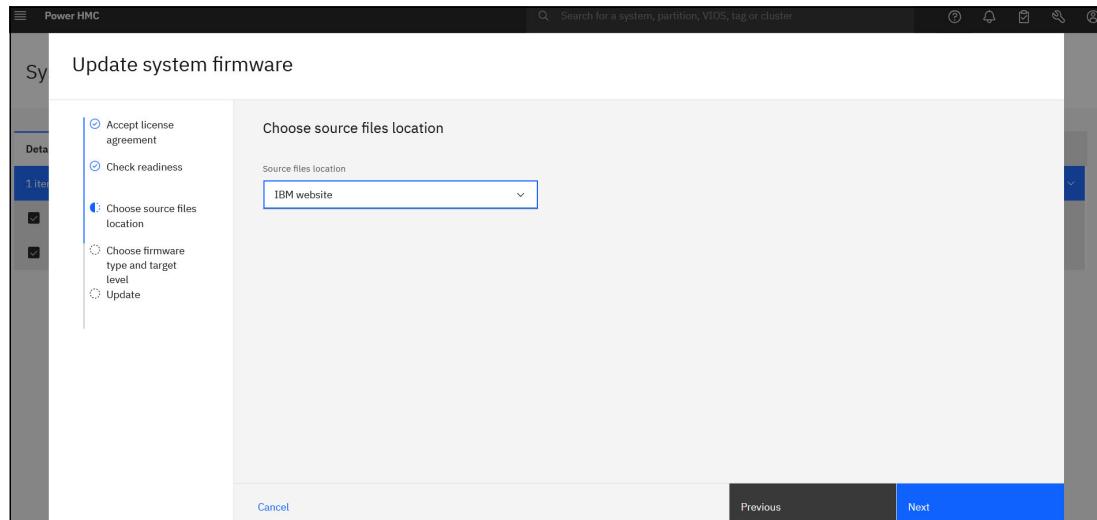


Figure 8-4 the “Choose Source File Location” drop-down menu used for updating the system firmware.

As shown in Figure 8-5, click the “Search Available Levels” button to locate the new system firmware on the installation media. On the following screen, select the desired system firmware level.

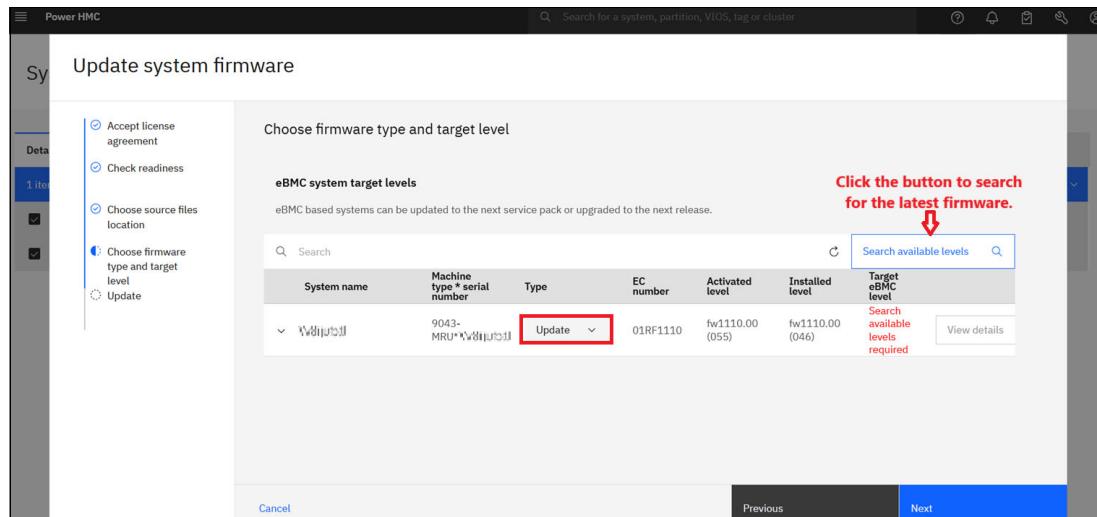


Figure 8-5 System Firmware Target Level Selection Screen. This screen allows you to choose between a firmware update or upgrade.

Import Update Files from HMC

The Import Update Files wizard in the HMC GUI, as shown in Figure 8-6 on page 163, 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.

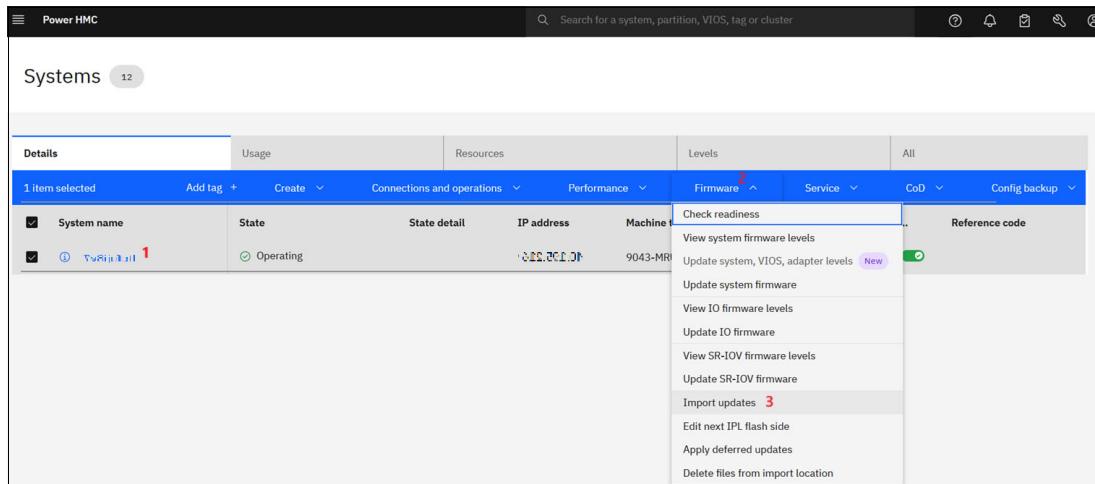


Figure 8-6 The “Import Update Files” option in the HMC GUI.

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. This allows you to manage and clean up firmware files stored on the HMC.

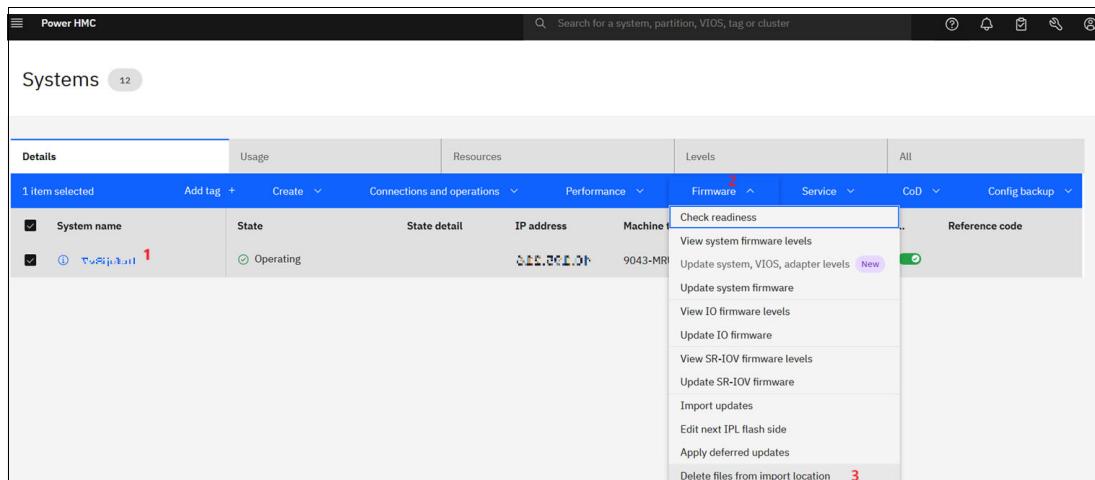


Figure 8-7 The “Delete Files from Import Location” option in the HMC GUI.

SR-IOV Shared Mode Adapter Firmware Management

SR-IOV capable adapters running in Shared mode use a different firmware versioning mechanism compared to those operating in Dedicated mode.

SR-IOV Configuration Requirements

- ▶ Ensure that the managed system includes PCIe adapter(s) that support SR-IOV functionality.
- ▶ Each SR-IOV-capable adapter must be installed in a PCIe slot that supports SR-IOV. Refer to the adapter placement guidelines in the IBM Documentation.
- ▶ Verify that the system is running a supported level of:
 - System firmware
 - HMC or PowerVM NovaLink
 - Operating system (with SR-IOV driver support)

- ▶ Configure the adapter(s) in SR-IOV Shared Mode 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 accompanying ReadMe documentation are available on IBM Fix Central.
- ▶ When an adapter is configured in SR-IOV Shared mode, its firmware and driver are automatically updated to the latest versions included with the system firmware. This update occurs during standard maintenance operations such as a system IPL, adapter replacement, or when switching the adapter mode between Shared and Dedicated.
- ▶ Additionally, if system firmware updates are installed concurrently, selective manual updates of SR-IOV adapter firmware can be performed using the Hardware Management Console (HMC).

<https://www.ibm.com/docs/en/power10/7063-CR1?topic=firmware-update-sr-iov>

Note: You may 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 may be temporarily disrupted. Updating each SR-IOV adapter typically takes between 2 to 5 minutes. The update is performed sequentially across all SR-IOV adapters in the system.

Figure 8-8 displays the “View SR-IOV Firmware Levels” and “Update SR-IOV Firmware” options available in the HMC GUI.

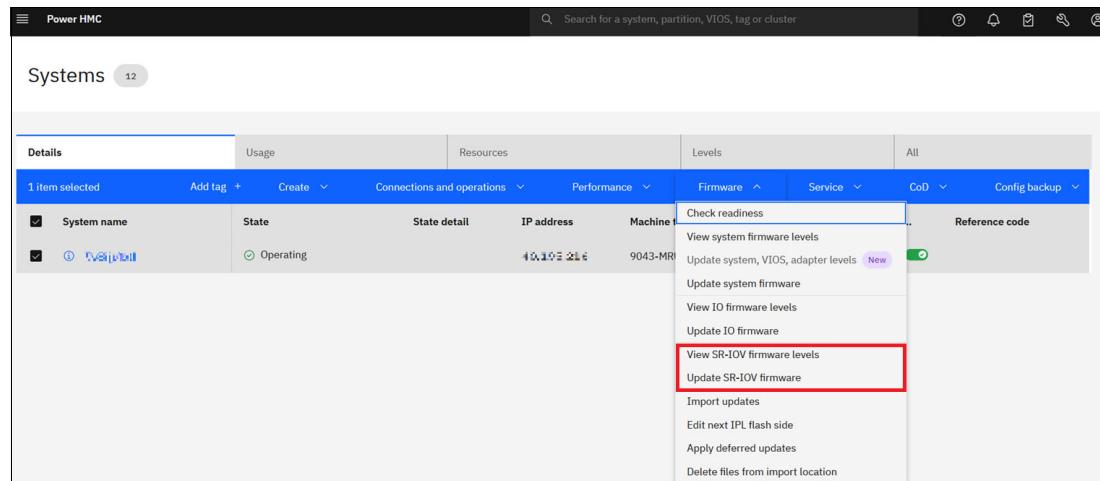


Figure 8-8 “View SR-IOV Firmware Levels” and “Update SR-IOV Firmware” options available in the HMC GUI.

I/O Microcode Update Strategy

Microcode updates are the responsibility of the customer, unless the services are provided under a Microcode Support Services contract.

Device microcode can be updated using the following methods:

▶ **HMC (Hardware Management Console):**

The “Update I/O Firmware” feature available in the HMC GUI allows administrators to apply microcode updates to supported I/O devices.

► **Diagnostic Menus (AIX and VIOS):**

Microcode updates can be performed using either the system diagnostics or stand-alone diagnostics utilities, from a USB key or a network location. This method is applicable to AIX and VIOS environments.

► **Linux:**

Refer to the microcode README files available on IBM FixCentral for detailed instructions on updating microcode on Linux systems. These instructions may include the use of additional vendor-provided tools, depending on the I/O devices.

► **IBM i (PTFs):**

This method is applicable only to IBM i systems. Microcode updates are delivered through Program Temporary Fixes (PTFs), which can be downloaded from IBM FixCentral.

Viewing or Updating I/O Firmware from the HMC

The Hardware Management Console (HMC) can be used to view the current I/O firmware levels for a system and to update those levels 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 also used for features like 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. Invscout exchanges inventory and update-related files with the HMC over the RMC interface.

Figure 8-9 displays the “View I/O Firmware Levels” and “Update I/O Firmware” options available in the HMC GUI.

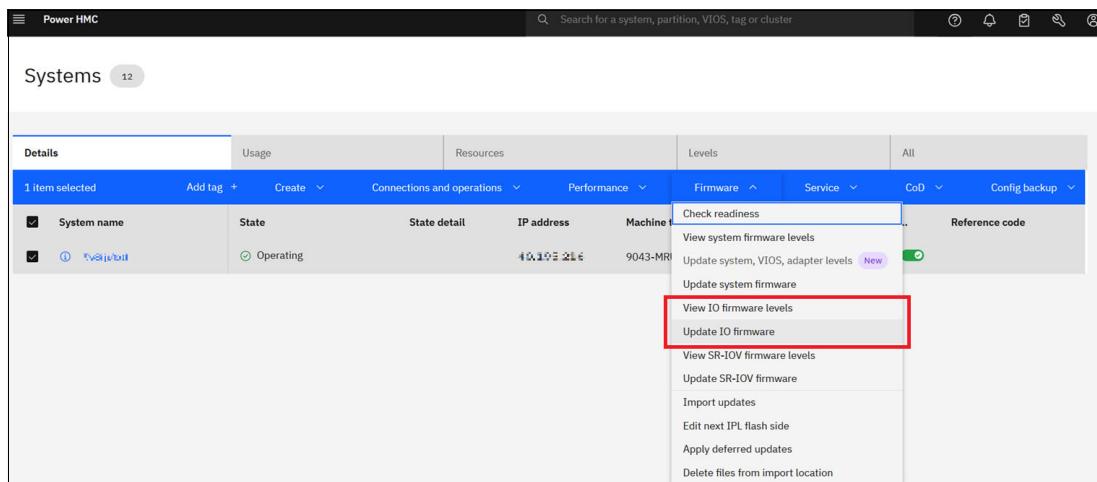


Figure 8-9 “View I/O Firmware Levels” and “Update I/O Firmware” options available in the HMC GUI.

Figure 8-10 displays the “View I/O Firmware Levels” panel, which shows I/O firmware information for both FSP and eBMC-based systems, as available in the HMC GUI.

System name	Machine type * serial number	Partitions	Logical device	I/O firmware level	Device
1B00831A	9043-MRU*1B00831A	100	ent0	000901218000E641	PCIe3 4-Port 10GbE Base-T Adapter
1B00831A	9043-MRU*1B00831A	100	ent1	000901218000E641	PCIe3 4-Port 10GbE Base-T Adapter
1B00831A	9043-MRU*1B00831A	100	ent2	000901218000E641	PCIe3 4-Port 10GbE Base-T Adapter
1B00831A	9043-MRU*1B00831A	100	ent3	000901218000E641	PCIe3 4-Port 10GbE Base-T Adapter
1B00831A	9043-MRU*1B00831A	100	nvme0	534e4143	PM1743 U.2 NVMe SSD 1600/320
1B00831A	9043-MRU*1B00831A	100	nvme1	534e4143	PM1743 U.2 NVMe SSD 1600/320

Figure 8-10 “View I/O Firmware Levels” panel available in the HMC GUI.

Figure 8-11 illustrates the “Choose Source Files Location” drop-down 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, including:

- ▶ IBM Website
- ▶ FTP Server
- ▶ SFTP Server
- ▶ Mount Point on the HMC
- ▶ CD/DVD
- ▶ USB Drive

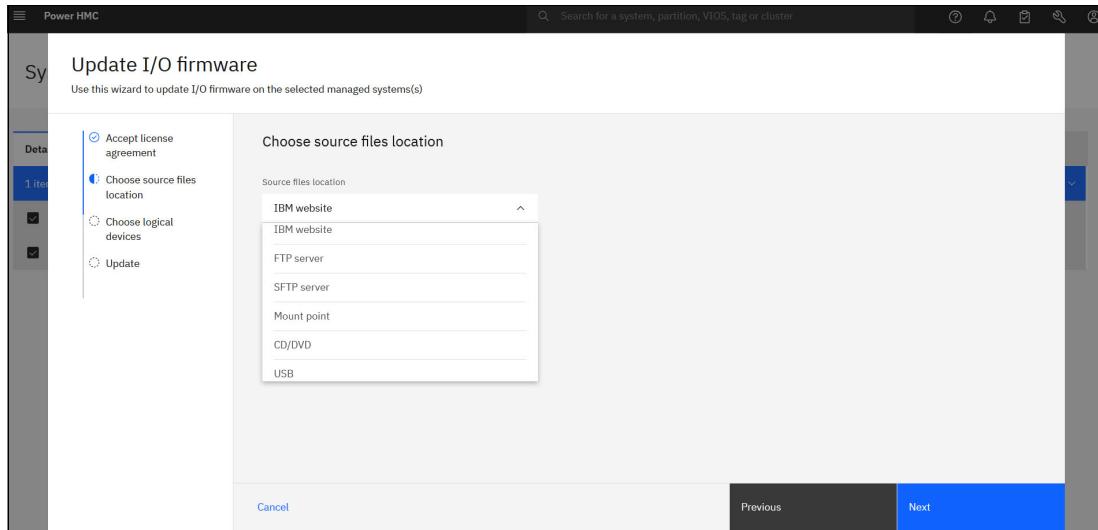


Figure 8-11 Update I/O firmware screen

For additional information on how to view or update the I/O firmware from the HMC, please refer to the following technical document: [“View or Update the I/O Firmware from the HMC”](#).

8.7.1 Update Access Keys

Since the introduction of the IBM Power8 processor-based servers, IBM uses the concept of an Update Access Key (UAK) for each server.

When system firmware updates are applied to the system, the UAK and its expiration date are checked. System firmware updates include a release date. If the release date for the firmware updates is past the expiration date for the update access key when attempting to apply system firmware updates, the updates are not processed.

Managing Update Access Keys

The system uses an Update Access Key (UAK) to control the application of system firmware updates. Each UAK includes an expiration date, while firmware updates include 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 will not be processed. When a UAK expires, it must be replaced using either the Hardware Management Console (HMC). Importantly, the expiration of a UAK does not affect I/O microcode updates; these can still be applied even if the UAK has expired. Customers 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 an UAK that expires after three years. Thereafter, the UAK can be extended every six months, but only if a current hardware maintenance contract exists for that server. The contract can be verified on the Enterprise Storage Server web page.

The validity and expiration date of the current UAK can be checked using either the HMC graphical interface, or through the command-line interface. Additionally, the expiration date can also be retrieved from the operating system level.

Verifying the Expiration Date of UAKs Using HMC

The current Update Access Key (UAK) expiration date is visible on the View current system firmware levels of the HMC, as shown in Figure 8-12.

System name	Machine type * serial number	LIC type	EC number	Access key expiration date	Installed level	Activated level	Platform IPL level	Deferred level
9043-MRU* 1234567890	9043-MRU* 1234567890	Managed System	01RF1110	5/30/2030	fw1110.00 (046)	fw1110.00 (055)	fw1110.00 (055)	

Figure 8-12 UAK expiration date displayed on the HMC View current system firmware levels page.

The Hardware Management Console (HMC) or vHMC (Virtual HMC Software Appliance) can be used to configure the Automatic Firmware Update Access Key. For more information, refer to the documentation titled [“How to Automatically Update the Access Key”](#).

Verifying the Expiration Date of UAKs Using AIX

There are multiple methods of checking the UAK expiration date within AIX:

- ▶ The first option utilizes the lscfg command. Use the following command:

```
lscfg -vpl sysplanar0 |grep -p "System Firmware"
```

The output is similar to the output that is shown in Example 8-1.

Example 8-1 Output of the lscfg command to check 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.....RK1110_058 RK1110_057 RK1110_058
    Microcode Level.....FW1110.00 FW1110.00 FW1110.00
    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
```

- ▶ Alternatively, you can grep on Access Key as shown in Example 8-2.

Example 8-2 Output of the lscfg command to check UAK expiration date

```
# lscfg -vpl sysplanar0 | grep "Access Key"
  Update Access Key Exp Date..20301014
```

- ▶ A third option on AIX 7.3 is the lparstat command which is shown in Example 8-3.

Example 8-3 Output of the lparstat command to check 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 Using Linux

There isn't a single Linux command to view a UAK directly. Instead, UAKs are managed through IBM's tools and interfaces, like ESA, ASMI and HMC.

Managing the UAK with IBM i

IBM Electronic Service Agent (ESA) on IBM i has introduced a new feature to help clients easily manage the refreshment of their update access keys (UAK) for IBM i stand-alone partitions. There are two ways of using this UAK management feature:

Manual: A system administrator can manually launch the process of checking UAK's expiration date, downloading UAK, and applying UAK. This is accomplished by running the WRKSRVAGT TYPE(*UAK) command. This can also be accomplished by running the GO SERVICE command to access the Electronic Service Agent on IBM i main menu, then selecting Option 20.

Check and refresh Update Access Key, as shown in Figure 8-13 and Figure 8-14.

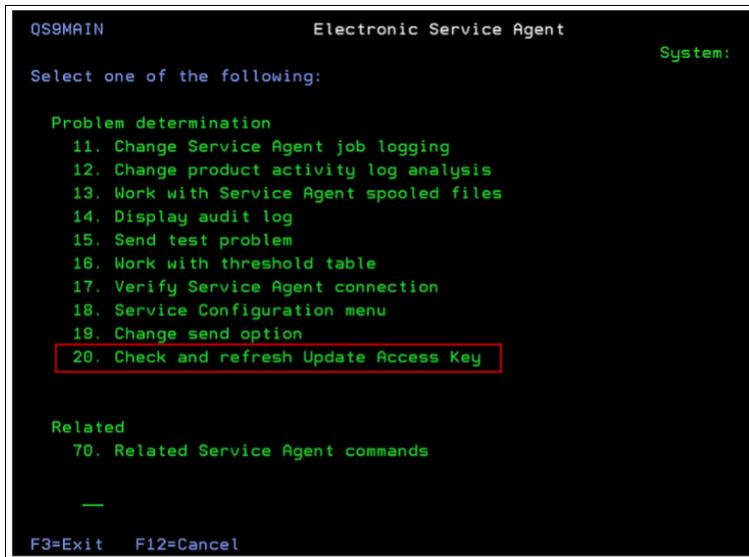


Figure 8-13 Electronic Service Agent on IBM i main menu option 20

Help instructions on option 20 can provide clients additional information about this feature as shown in Figure 8-14.

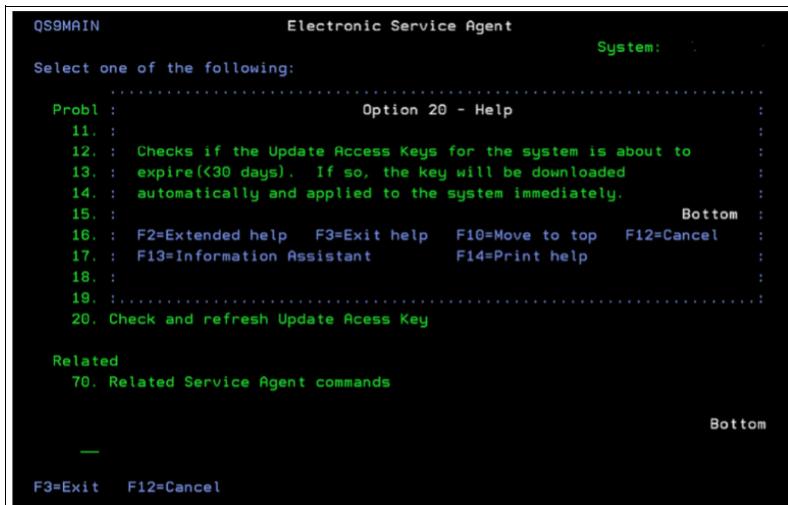


Figure 8-14 Help information on Electronic Service Agent on IBM i main menu option 20

Automatic: The Change Service Agent Attribute (CHGSRVAGTA) command can be used to enable automatic checking. The CHGSRVAGTA command parameter (REFRESHUAK) can be used to establish automatic UAK management. By default, a job named QS9UAK, will be run every Sunday, but can be easily customized.

Enablement steps:

- ▶ Enter CHGSRVAGTA and press F4.
- ▶ Enter Yes for Enable.
- ▶ Press Page Down.

At this point, the Change Service Agent Attribute panel appears as shown in Figure 8-15 on page 170. Specify the weekday and time to use for UAK refresh processing.

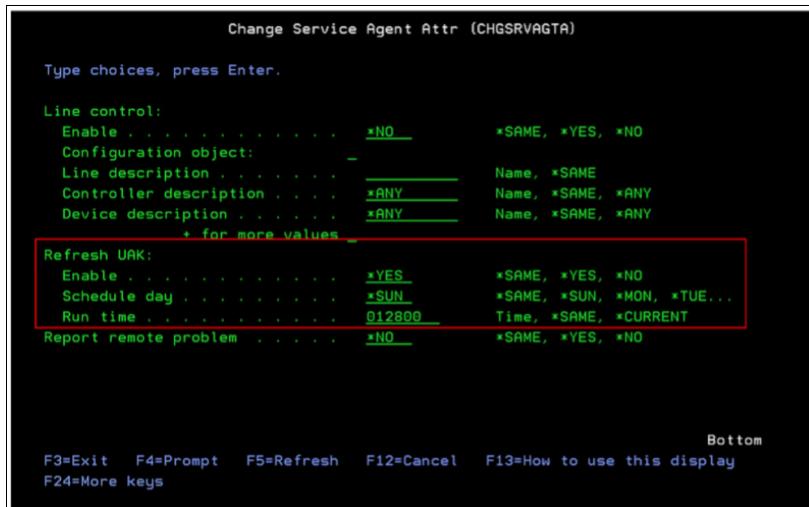


Figure 8-15 Change Service Agent Attribute – refresh UAK configuration panel



Virtualization

Virtualization on IBM Power is a cornerstone of its enterprise computing strategy, offering robust, scalable, and secure environments for running multiple workloads on a single physical server. At the heart of this capability is PowerVM, IBM's enterprise-grade virtualization technology. PowerVM enables the creation of logical partitions (LPARs), allowing multiple operating systems such as AIX, IBM i, and Linux to run concurrently on the same hardware. It supports advanced features like live partition mobility, dynamic resource allocation, and micro-partitioning, which help maximize hardware utilization and reduce operational costs.

In addition to PowerVM, IBM Power Servers also support KVM (Kernel-based Virtual Machine), 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 well with modern orchestration tools and supports features like nested virtualization and SR-IOV for high-performance networking. KVM is ideal for organizations looking to leverage open-source technologies while still 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). SPPs allow multiple LPARs to share a pool of physical processors, enabling dynamic allocation of CPU resources based on workload demand. With Power11 PowerVM add support for Resource Groups to further enhance the processor sharing capabilities and add additional efficiencies. Processor Pools add to the capabilities of Shared Processor Pools to improve processor utilization and enhance the isolation and control by capping the maximum CPU resources 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.

This chapter contains the following topics:

- ▶ PowerVM
- ▶ KVM support

9.1 PowerVM

The PowerVM platform is the family of technologies, capabilities, and offerings that delivers industry-leading virtualization for enterprises. It is the umbrella branding term for Power processor-based server virtualization, that is, IBM PowerVM Hypervisor, logical partitioning, IBM Micro-Partitioning®, Virtual I/O Server (VIOS), Live Partition Mobility (LPM), and more. PowerVM is a combination of hardware enablement and software.

Note: PowerVM Enterprise Edition License Entitlement is included with each Power E1180 server. PowerVM Enterprise Edition is available as a hardware feature (#EPVT), supports up to 20 partitions per core, VIOS, Resource Groups which can be combined with shared processor pools (MSPPs) and also offers LPM.

9.1.1 IBM PowerVM Hypervisor

Power processor-based servers are combined with PowerVM technology and offer the following key capabilities that can help to consolidate and simplify IT environments:

- ▶ Improve server usage and share I/O resources to reduce the total cost of ownership (TCO) and better use IT assets.
- ▶ Improve business responsiveness and operational speed by dynamically reallocating resources to applications as needed to better match changing business needs or handle unexpected changes in demand.
- ▶ Simplify IT infrastructure management by making workloads independent of hardware resources so that business-driven policies can be used to deliver resources that are based on time, cost, and service-level requirements.

Combined with features in the Power E1180, the IBM PowerVM Hypervisor delivers functions that enable other system technologies, including logical partition (LPAR) technology, virtualized processors, IEEE virtual local area network (VLAN)-compatible virtual switch, virtual SCSI adapters, virtual Fibre Channel adapters, and virtual consoles.

The PowerVM Hypervisor is a basic component of the system's firmware and offers the following functions:

- ▶ Provides an abstraction between the physical hardware resources and the LPARs that use them.
- ▶ Enforces partition integrity by providing a security layer between LPARs.
- ▶ Controls the dispatch of virtual processors to physical processors.
- ▶ Saves and restores all processor state information during a logical processor context switch.
- ▶ Controls hardware I/O interrupt management facilities for LPARs.
- ▶ Provides VLAN channels between LPARs that help reduce the need for physical Ethernet adapters for inter-partition communication.
- ▶ Monitors the Flexible Service Processor (FSP) and performs a reset or reload if it detects the loss of one of the FSP, notifying the operating system if the problem is not corrected.

The PowerVM Hypervisor 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 is required by the PowerVM Hypervisor firmware 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 that contributes to hypervisor memory usage. The HPT is used by the operating system to translate from effective addresses (EAs) to physical real addresses in the hardware. This translation from effective to real addresses allows multiple operating systems 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 that can be used when translating addresses.

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 1/128th (for AIX, VIOS, and Linux partitions) of the maximum memory size of the partition. AIX, VIOS, and Linux use larger page sizes (16 KB and 64 KB) instead of using 4 KB pages. The use of 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 is specified 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, the HPT ratio can be configured. The `hpt_ratio` parameter for the `chsyscfg` Hardware Management Console (HMC) command can be issued to define the HPT ratio that is used 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 incur more CPU consumption because the operating system 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

In support of I/O operations, the hypervisor maintains structures that are called the *translation control entities* (TCEs), which provide an information path between I/O devices and partitions. The TCEs provide the address of the I/O buffer, indications of read versus write requests, and other I/O-related attributes. Many TCEs are used per I/O device, so multiple requests can be active simultaneously to the same physical device. To provide better affinity, the TCEs are spread across multiple processor chips or drawers to improve performance while accessing the TCEs.

For physical I/O devices, the base amount of space for the TCEs is defined by the hypervisor that is based on the number of I/O devices that are supported. A system that supports high-speed adapters can also be configured to allocate more memory to improve I/O performance. Linux is the only operating system that uses these extra TCEs so that the memory can be freed for use by partitions if the system uses only AIX.

Memory usage for virtualization features

Virtualization requires more memory to be allocated by the PowerVM Hypervisor for hardware statesave areas and various virtualization technologies. For example, on Power11 processor-based systems, each processor core supports up to eight simultaneous multithreading (SMT) threads of execution, and each thread contains over 80 different registers.

The PowerVM Hypervisor must set aside save areas for the register contents for the maximum number of virtual processors that are configured. The greater the number of physical hardware devices, the greater the number of virtual devices, the greater the amount of virtualization, and the more hypervisor memory is required. For efficient memory consumption, wanted and maximum values for various attributes (processors, memory, and virtual adapters) must be based on business needs, and not set to values that are higher than actual requirements.

Predicting memory that is used by the PowerVM Hypervisor

The IBM System Planning Tool (SPT) is a resource that can be used to estimate the amount of hypervisor memory that is required for a specific server configuration. After the SPT executable file is downloaded and installed, you can define a configuration by selecting the correct hardware platform and the installed processors and memory, and defining partitions and partition attributes. SPT can estimate the amount of memory that is assigned to the hypervisor, which assists you when you change a configuration or deploy new servers.

The PowerVM Hypervisor provides the following types of virtual I/O adapters:

- ▶ Virtual SCSI

The PowerVM Hypervisor provides a virtual SCSI mechanism for the virtualization of storage devices. The storage virtualization is accomplished by using two paired adapters: a virtual SCSI server adapter and a virtual SCSI customer adapter.

- ▶ Virtual Ethernet

The PowerVM Hypervisor provides a virtual Ethernet switch function that allows partitions fast and secure communication on the same server without any need for physical interconnection or connectivity outside of the server if a Layer 2 bridge to a physical Ethernet adapter is set in one VIOS partition, also known as *Shared Ethernet Adapter* (SEA).

- ▶ Virtual Fibre Channel

A virtual Fibre Channel adapter is a virtual adapter that provides customer LPARs with a Fibre Channel connection to a storage area network through the VIOS partition. The VIOS partition provides the connection between the virtual Fibre Channel adapters on the VIOS partition and the physical Fibre Channel adapters on the managed system.

- ▶ Virtual (TTY) console

Each partition must have access to a system console. Tasks, such as operating system installation, network setup, and various problem analysis activities, require a dedicated system console. The PowerVM Hypervisor provides the virtual console by using a virtual TTY or serial adapter and a set of hypervisor calls to operate on them. Virtual TTY does not require the purchase of any other features or software, such as the PowerVM Edition features.

Logical partitions

LPARs and virtualization increase the usage of system resources and add a level of configuration possibilities.

Logical partitioning is the ability to make a server run as though it were two or more independent servers. When you logically partition a server, you divide the resources on the server into subsets, called *LPARs*. You can install software on an LPAR, and the LPAR runs as an independent logical server with the resources that you allocated to the LPAR.

LPAR is also referred to in some documentation as a *virtual machine* (VM), which makes it look similar to what other hypervisors offer. However, LPARs provide a higher level of security and isolation and other features that are described in this chapter.

Processors, memory, and I/O devices can be assigned to LPARs. AIX, IBM i, Linux, and VIOS can run on LPARs. VIOS provides virtual I/O resources to other LPARs with general-purpose operating systems.

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 the 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 over a set of processors that is called a *shared processor pool* (SPP), and virtual processors are used to enable the operating system manage the fractions of processing power that are assigned to the LPAR.

From an operating system perspective, a virtual processor cannot be distinguished from a physical processor, unless the operating system is enhanced to determine the difference. Physical processors are abstracted into virtual processors that are available to partitions.

On the Power11 processor-based server, a partition can be defined with a processor capacity as small as 0.05 processing units. This number represents 0.05 of a physical core. Each physical core can be shared by up to 20 shared processor partitions, and the partition's entitlement can be incremented fractionally by as little as 0.05 of the processor. The shared processor partitions are dispatched and time-sliced on the physical processors under the control of the PowerVM Hypervisor. The shared processor partitions are created and managed by the HMC.

The Power E1180 supports up to 256 cores in a single system and 1000 micropartitions (1000 is the maximum that PowerVM supports).

Note: Although the Power E1180 supports up to 1000 micropartitions, the real limit depends on application workload demands in use on the server.

Processing mode

When you create an LPAR, you can assign entire processors for dedicated use, or you can assign 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 SMT feature in the Power11 processor core allows 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 become processor donors for idle processors that they own, which allows for the donation of spare CPU cycles from dedicated processor partitions to an SPP. The dedicated partition maintains absolute priority for 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 (the maximum number is 20 times the number of processing units that are assigned to the partition). The PowerVM Hypervisor dispatches virtual processors to physical processors according to the partition's processing units entitlement. One processing unit represents one physical processor's processing capacity. All partitions receive a total CPU time equal to their processing unit's entitlement. The logical processors are defined on top of virtual processors. Therefore, even with a virtual processor, the concept of a logical processor exists, and the number of logical processors depends on whether SMT is turned on or off.

9.1.2 Multiple shared processor pools

MSPPs are supported on Power11 processor-based servers. This capability allows a system administrator to create a set of micropartitions with the purpose of controlling the processor capacity that can be used from the physical SPP.

Micropartitions are created and then identified as members of the default processor pool or a user-defined SPP. The virtual processors that exist within the set of micropartitions are monitored by the PowerVM Hypervisor. 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 and 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 to the set of micropartitions within that SPP.

All Power servers that support the MSPP capability have a minimum of one (the default) SPP and up to a maximum of 64 SPPs.

This capability helps customers reduce TCO significantly when the cost of software or database licenses depends on the number of assigned processor-cores.

Shared Processor Pools: Can be used in conjunction with Resource Groups. Such configuration can improve performance and isolation significantly.

9.1.3 Virtual I/O Server

The VIOS is part of PowerVM. It is the specific appliance that allows the sharing of physical resources among LPARs to allow more efficient usage (for example, consolidation). In this case, the VIOS owns the physical I/O resources (SCSI, Fibre Channel, network adapters, or optical devices) and allows customer partitions to share access to them, which minimizes and optimizes the number of physical adapters in the system.

The VIOS eliminates the requirement that every partition owns a dedicated network adapter, disk adapter, and disk drive. The VIOS supports OpenSSH for secure remote logins. It also provides a firewall for limiting access by ports, network services, and IP addresses.

Figure 9-1 shows an overview of a VIOS configuration.

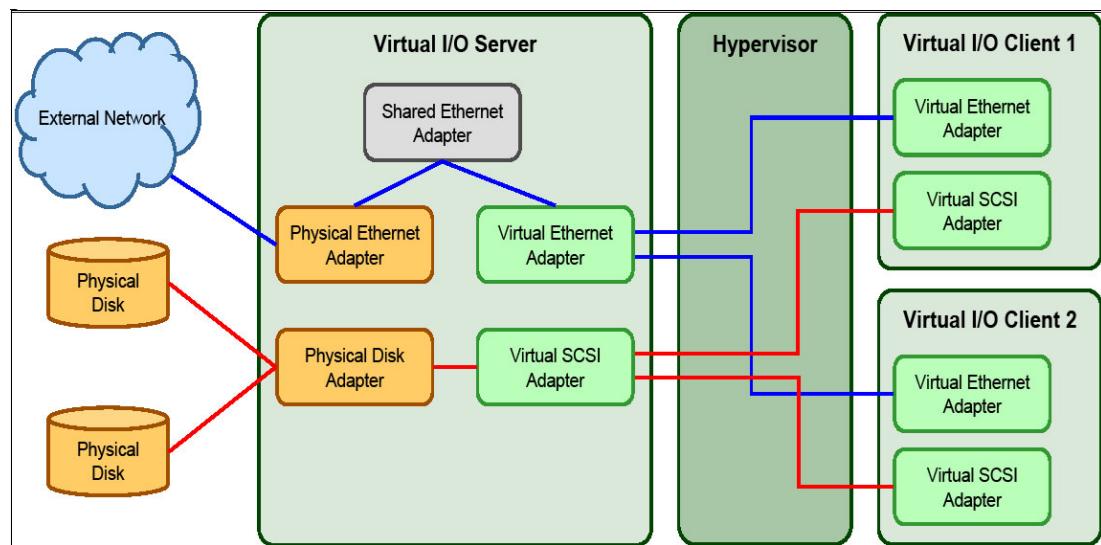


Figure 9-1 Architectural view of the VIOS

It is a best practice to run dual VIO servers per physical server.

Shared Ethernet Adapter

A SEA can be used to connect a physical Ethernet network to a virtual Ethernet network. The SEA provides this access by connecting the PowerVM Hypervisor VLANs to the VLANs on the 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 customer 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 the VIOS (not in a general-purpose AIX or Linux partition) and acts as a Layer 2 network bridge to securely transport network traffic between virtual Ethernet networks (internal) and one or more (Etherchannel) physical network adapters (external). These virtual Ethernet network adapters are defined by the PowerVM Hypervisor on the VIOS.

Virtual SCSI

Virtual SCSI is used to view a virtualized implementation of the SCSI protocol. Virtual SCSI is based on a customer/server relationship. The VIOS LPAR owns the physical I/O resources and acts as a server or, in SCSI terms, a target device. The client LPARs access the virtual SCSI backing storage devices that are provided by the VIOS as clients.

The virtual I/O adapters (a virtual SCSI server adapter and a virtual SCSI client adapter) are configured by using an HMC. The virtual SCSI server (target) adapter is responsible for running any SCSI commands that it receives, and is owned by the VIOS partition. The virtual SCSI client adapter allows a client partition to access physical SCSI and SAN-attached devices and LUNs that are mapped to be used by the client partitions. The provisioning of virtual disk resources is provided by the VIOS.

iSCSI

The Internet Small Computer Systems Interface (iSCSI) disk is supported in the Virtual I/O Server (VIOS) 3.1.0, or later

The Internet Small Computer Systems Interface (iSCSI) disk provides block-level access to storage devices by carrying SCSI commands over an Internet Protocol network. The iSCSI disk is used to facilitate data transfers over the internet by using TCP, a reliable transport mechanism that uses either IPV6 or IPV4 protocols. The iSCSI disk is used to manage storage over long distances.

The iSCSI support in VIOS allows iSCSI disks to be exported to client logical partitions as virtual disks (vSCSI disks). This support is available in VIOS version 3.1, and later,

VIOS version 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 logical partition can run either an AIX or Linux operating system.

VIOS version 3.1.1 enables support for multiple iSCSI initiators on the VIOS. This support also includes performance enhancements for the iSCSI driver. With multiple iSCSI initiator support, you can create multiple iSCSI software initiator devices on a single AIX operating system instance.

N_Port ID Virtualization

N_Port ID Virtualization (NPIV) is a technology that allows multiple LPARs to access one or more external physical storage devices through the same physical Fibre Channel adapter. This adapter is attached to a VIOS partition that acts only as a pass-through that manages the data transfer through the PowerVM Hypervisor.

Each partition features one or more virtual Fibre Channel adapters, each with their 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*<Default -1 Font>, SG24-7590.

9.1.4 Live Partition Mobility

LPM enables you to move a running LPAR from one system to another without disruption. Inactive partition mobility allows you to move a powered-off LPAR from one system to another one.

LPM provides systems management flexibility and improves system availability by avoiding the following situations:

- ▶ Planned outages for hardware upgrade 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.

For more information and requirements for LPM, see *IBM PowerVM Live Partition Mobility*, SG24-7460.

HMC V10R1 and VIOS 3.1.3 or later provide the following enhancements to the LPM Feature:

- ▶ Automatically choose the fastest network for LPM memory transfer.
- ▶ Allow 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 Code (#EM89) that belongs to the technologies under the PowerVM umbrella and enables memory expansion on the system.

AME is an innovative technology that supports the AIX operating system. It helps enable the effective maximum memory capacity to be larger than the true physical memory maximum. Compression and decompression of memory content can enable memory expansion up to 100% or more. This expansion can enable a partition to complete more work or support more users with the same physical amount of memory. Similarly, it can enable a server to run more partitions and do more work for the same physical amount of memory.

AME uses CPU resources to compress and decompress the memory contents. The trade-off of memory capacity for processor cycles can be an excellent choice, but the degree of expansion varies about how compressible the memory content is. It also depends on having adequate spare CPU capacity available for this compression and decompression.

The Power E1180 includes a hardware accelerator that is designed to boost AME efficiency and uses less processor core resources. Each AIX partition can turn on or turn off AME. Control parameters set the amount of expansion that is wanted in each partition to help control the amount of CPU used by the AME function.

An IPL is required for the specific partition that is turning memory expansion. When enabled, monitoring capabilities are available in standard AIX performance tools, such as lparstat, vmstat, topas, and svmon.

A planning tool is included with AIX, which enables you to sample workloads and estimate how expandable the partition's memory is and much CPU resource is needed. The feature can be ordered with the initial order of the Power E1180 or as a Miscellaneous Equipment Specification (MES) order. A software key is provided when the enablement feature is ordered, which is 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 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. At times, it might take longer to start the server, in which case the Remote Restart function can be used for faster reprovisioning of the partition. Typically, Remote Restart can be done faster than restarting the server that stopped and then restarting the partitions. The Remote Restart function relies on technology that is similar to LPM where a partition is configured with storage on a SAN that is shared (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, the Power processor modes are described here because they affect various virtualization features.

On Power servers, partitions can be configured to run in several modes, including the following modes:

- ▶ Power9

This native mode for Power9 processors implements Version 3.0 of the IBM Power ISA.
For more information, see [IBM Documentation](#).

- ▶ Power10

This native mode for Power10 processors implements Version 3.1 of the IBM Power ISA.
For more information, see [IBM Documentation](#).

- ▶ Power11

This native mode for Power11 processors implements Version 3.X of the IBM Power ISA.
For more information, see [IBM Documentation](#).

Figure 9-2 shows the available processor modes on a Power E1180.

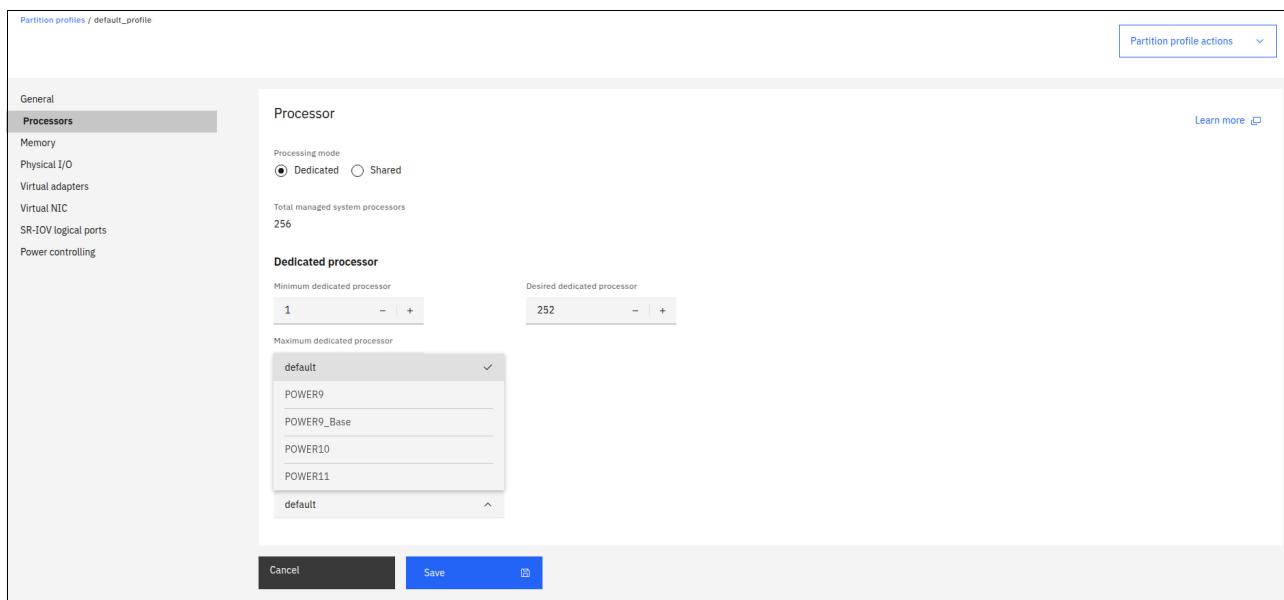


Figure 9-2 Processor modes

Processor compatibility mode is important when LPM migration is planned between different generations of servers. An LPAR that might be migrated to a machine that is managed by a processor from another generation must be activated in a specific compatibility mode.

Note: Migrating an LPAR from a Power9 processor-based server to Power E1180 by using LPM is not supported; however, the following steps can be completed to accomplish this task:

1. Migrate LPAR from Power9 processor-based server to Power10 processor-based server by using LPM.
2. Migrate then the LPAR from Power10 processor-based server to Power11.

9.1.8 Single Root I/O Virtualization

Single Root I/O Virtualization (SR-IOV) is an extension to the Peripheral Component Interconnect Express (PCIe) specification that allows multiple operating systems to simultaneously share a PCIe adapter 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 enables adapter consolidation through sharing much like logical partitioning enables server consolidation. With an adapter capable of SR-IOV, 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 the 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 are a new feature introduced with IBM Power11, designed to enhance system performance by up to 25% through improved workload optimization and resource affinity. While Shared Processor Pools (SPPs) already provide compute capacity isolation by capping the maximum resources available to each pool, Resource Groups take this further 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 that 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 new tool for organizing and optimizing compute resources. Designed for flexibility and performance, Resource Groups allow for 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 like Shared Processor Pools, Live Partition Mobility, and PEP 2.0.

The core features of resource groups are:

- ▶ Performance Optimization:
 - Resource Groups improve workload affinity, enabling shared processor performance nearly equivalent to dedicated processors.
- ▶ Flexible Configuration:
 - Can include both dedicated processor partitions and shared processor partitions.
 - Resource group configuration will specify the number of general purpose cores (AIX/IBM i/Linux/VIOS) and IFL cores (Linux/VIOS).
 - Powered-off partitions can be reassigned between groups.

- Resources 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 shared processor pools.
 - Enhances compute capacity isolation and affinity-based dispatching.
- ▶ Monitoring and Management:
 - Utilization metrics available for Resource Groups, SPPs, and the overall system.
 - Dynamic Platform Optimizer (DPO) can run at system or group level.
- ▶ Mobility and Compatibility:
 - When using Live Partition Mobility 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 to customers. In this section, we provide several example use cases, but beyond the examples provided, use cases can also be combined to maximize functionality and tailor performance to specific business needs.

Consolidation across multiple lines of business

Figure 9-3 illustrates how resources are allocated and shared across multiple lines of business on Power10 systems. With the introduction of Resource Groups in Power11, resource allocation becomes significantly more efficient, enabling better workload isolation and improved performance across the system.

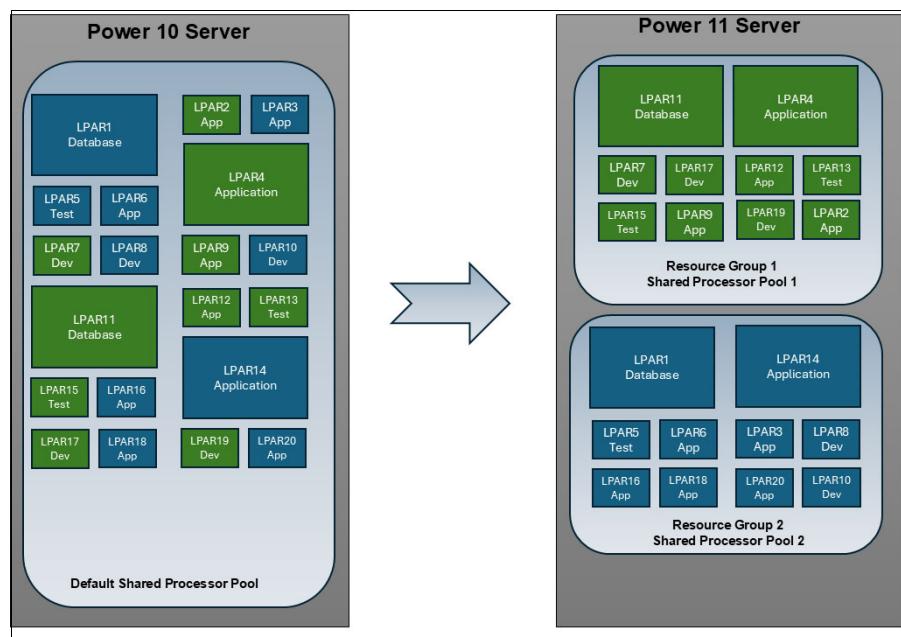


Figure 9-3 Example where Resource Groups consolidate multiple lines of business

Isolation of production workloads from test/dev workloads

In this example, the development and test LPARs are restricted to using resources only from the default Resource Group, limiting their access to shared system resources and isolating them from production workloads.

Figure 9-4 shows how resource groups are used to provide this isolation of development and test workloads.

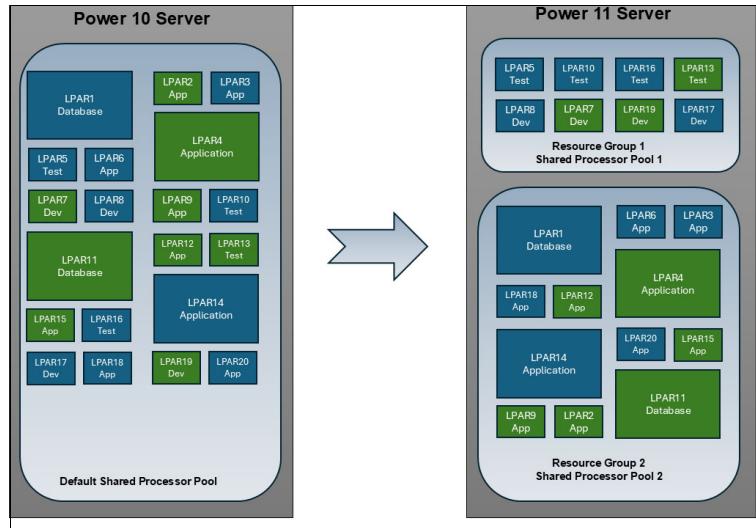


Figure 9-4 Example where Resource Groups isolate workloads

Improve application performance by grouping workload tiers into resource groups

In this example, performance is enhanced through improved affinity between the database and application server, allowing for more efficient resource utilization and faster communication within the same Resource Group. Figure 9-5 shows how grouping workloads into different resource groups can improve application affinity and improve performance.

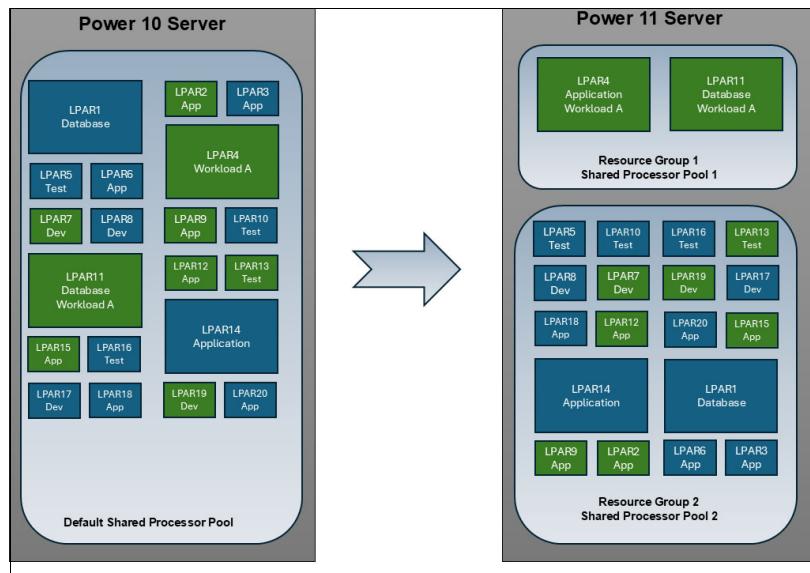


Figure 9-5 Example where Resource Groups improve performance by grouping workload tiers

System-level isolation in multi-server consolidation scenarios

Resource Groups enable effective workload isolation on IBM Power11 systems. This allows organizations to consolidate multiple smaller servers into a single, more powerful Power11 system while logically separating workloads into distinct groups. Each group can be managed independently, ensuring performance, security, and resource control across different business

functions or environments. Figure 9-6 shows how using resource groups can help maintain system level isolation while doing server consolidation.

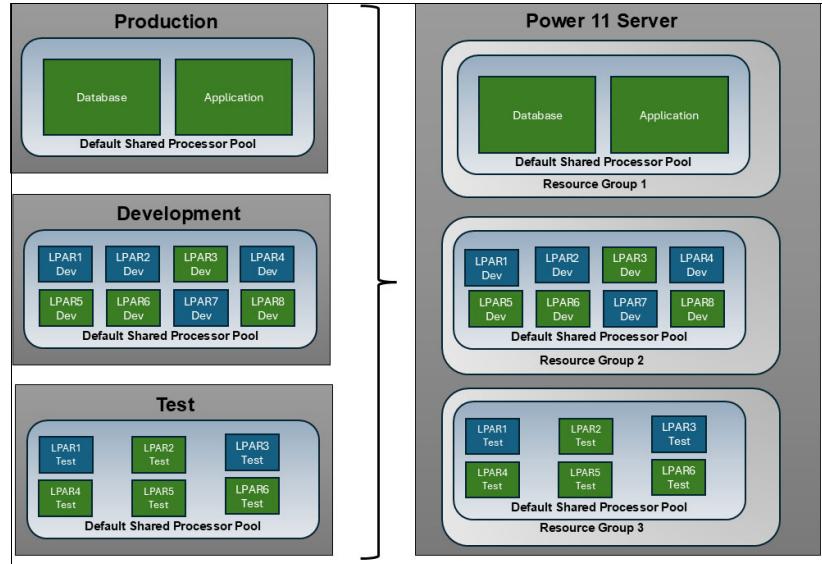


Figure 9-6 Example where Resource Groups provides additional layer of workload isolation

Improved performance by mapping Shared Processor Pools into Resource Groups

Shared Processor Pools (SPPs) and Resource Groups can be used together on IBM Power11 systems to simultaneously optimize application performance and reduce total cost of ownership (TCO). This combination allows for precise control over resource allocation and workload isolation, ensuring efficient utilization of compute capacity while maintaining performance consistency across diverse workloads. Figure 9-7 shows how SPPs and Resource Groups enhance your performance and reduces your TCO.

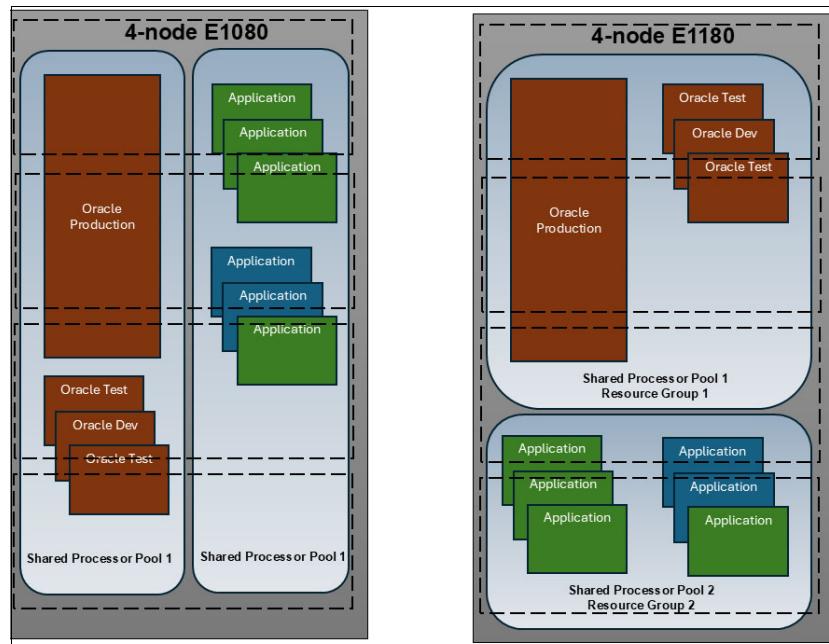


Figure 9-7 Example where Resource Groups improve performance by mapping Shared Processor Pools into Resource Groups.

Resource Groups Advisor

Resource Groups Advisor (RGA) is a web-based tool designed to assist with the configuration of resource groups on IBM Power. Offered as a free service, RGA analyzes a customer's server setup and provides tailored recommendations for optimal resource group configurations. It helps model and validate configurations to ensure efficient resource allocation and system performance.

RGA is available at:

<https://www.ibm.com/it-infrastructure/resources/resource-groups-advisor>,

The welcome screen for RGA is shown in Figure 9-8.

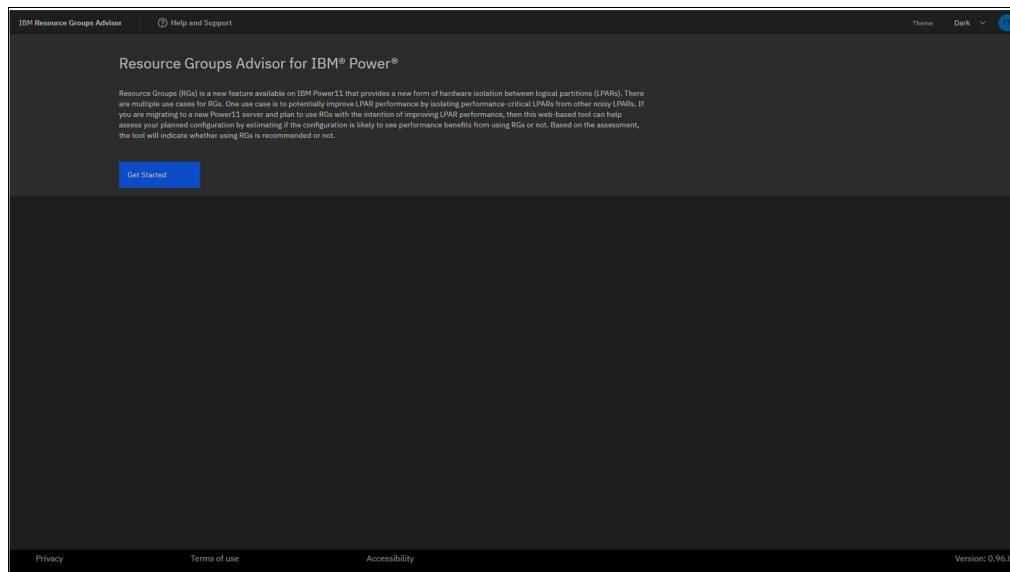


Figure 9-8 Welcome screen

After selecting Getting Started, you see a screen where you can choose to create a new configuration or upload a previously saved version. Selecting **New Configuration** brings up the screen shown in Figure 9-9.

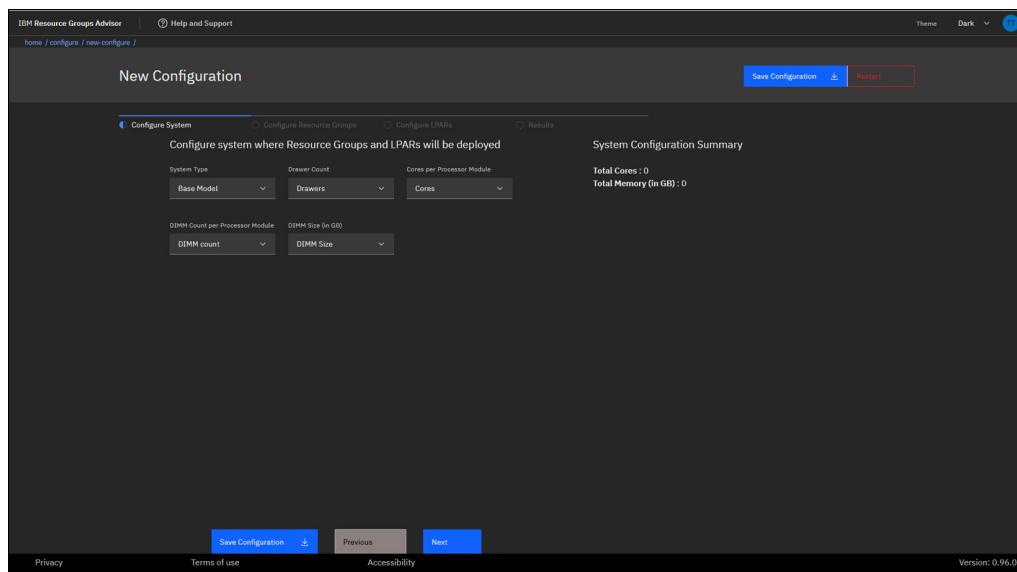


Figure 9-9 New configuration screen

After providing the system configuration parameters, press **Next** to be presented with the resource group configuration planned for your system. This is shown in Figure 9-10. The default resource group is always configured, but you can choose to add additional resource groups to your configuration.

Resource Group Name	Processor cores	Affinity Priority
Default Resource Group	192	128

Figure 9-10 Resource group definitions

After defining your resource groups, you will be given the opportunity to define the LPAR details. Selecting **Add LPAR** allows you to enter the details about each LPAR such as whether it is running in Shared or Dedicated mode, the processor and memory allocations, and which resource group the LPAR is assigned to. LPARs can also be added from the [IBM System Planning Tool](#).

When all of your LPARs are defined, you will see an LPAR summary screen similar to Figure 9-11.

LPAR Name	OS	Processor Mode	SPUs	Desired VPs	Desired Processors	Desired Memory(GB)	Max Memory(GB)	Virtual Persistent Memory(GB)	Huge Pages	Resource Group	Critical
lp1	AIX	shared	4	8	NA	100	100	0	0	Default Resource Group	yes
lp2	AIX	shared	8	8	NA	1000	1000	0	0	rgt	yes
lp3	AIX	shared	8	16	NA	1000	1000	0	0	rgt	yes
lp4	AIX	shared	16	32	NA	1000	1000	0	0	rgt	yes

Figure 9-11 LPAR configuration

After entering all of the LPAR information, press **Submit** to get the model's output similar to what is shown in Figure 9-12.

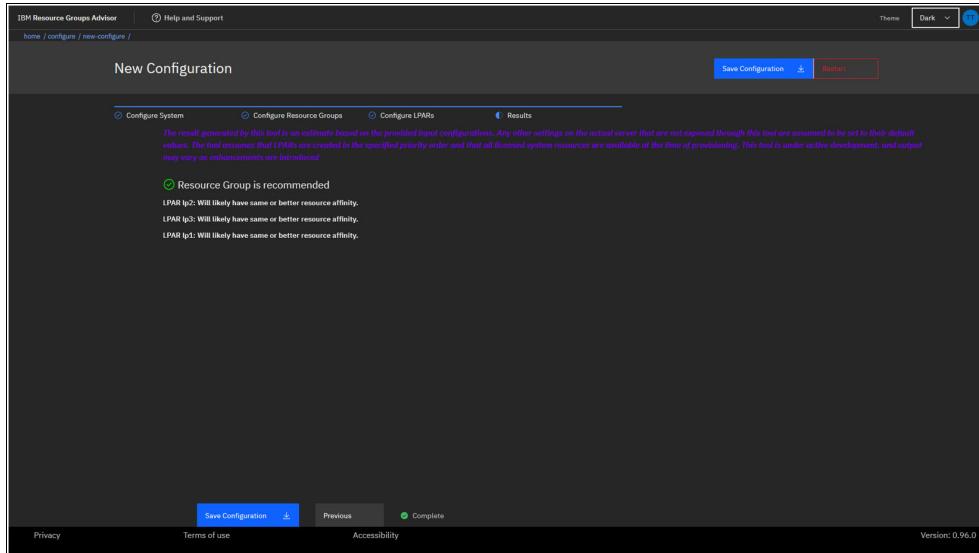


Figure 9-12 Output from the tool

You can save the configuration at this point, or start a new model.

Note: RGA simulates how PowerVM would create resource groups and allocate resources to LPARs on a system, all based on the provided input configurations. No “customer data” is being pulled off actual servers and assessed.

9.2 KVM support

IBM Power11 processor based servers can utilize Kernel-based Virtual Machine (KVM) within a PowerVM logical partition (LPAR). This allows for the creation and management of lightweight Linux virtual machines (VMs) using standard KVM tools, while still leveraging the existing resources of the PowerVM LPAR. Essentially, KVM becomes an additional virtualization option alongside PowerVM on IBM Power.

KVM on IBM Power11 processor based servers is not a replacement for PowerVM, but rather an additional capability that brings the power, speed, and flexibility of the KVM virtualization technology to a PowerVM logical partition (LPAR).

KVM-enabled LPARs can host PPC64-LE KVM guests, which are essentially Linux VMs. The KVM guests within the LPAR utilize resources (CPU, memory, I/O) that have been allocated to the LPAR by the PowerVM hypervisor. This approach offers flexibility in deploying Linux workloads and can be more cost-effective than other virtualization solutions, especially for organizations already invested in the Linux ecosystem. KVM's integration with the Linux kernel can lead to high performance, especially when running Linux-based workloads. This setup enables use cases such as running standard Linux distributions, containers, and other workloads that benefit from the KVM virtualization stack and the benefit to consolidate in one or more IBM Power11 processor based servers different kind of workloads (see Figure 9-13 on page 188 for an example). In essence, KVM on Power10 provides a way to leverage the strengths of both PowerVM and KVM, offering a powerful and flexible virtualization environment for Linux workloads.

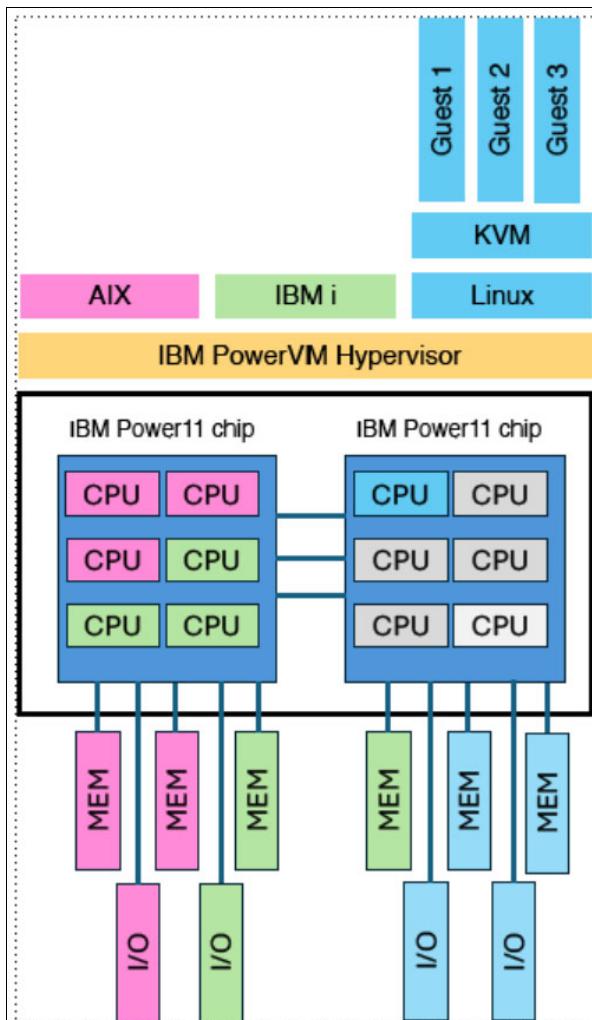


Figure 9-13 IBM Power stack for KVM

About capabilities, KVM in a PowerVM LPAR utilizes the industry standard Linux KVM virtualization stack and can easily integrate within an existing Linux virtualization ecosystem. KVM in an LPAR is enabled by:

- ▶ IBM Power architecture and Power11 implementation has advanced virtualization capabilities to run multiple operating system (OS) instances that share the same hardware resources while providing isolation. The Radix MMU architecture provides the capability to independently manage page tables for the LPAR and the KVM guest instances on the LPAR.
- ▶ PowerVM industry-leading virtualization stack provides new functions to create and manage KVM guests. These changes extend the Power platform architecture to include new hypervisor interfaces.
- ▶ Linux kernel that includes the KVM kernel module (KVM) provides core virtualization infrastructure to run multiple virtual machines in a Linux host LPAR. Upstream kernels and enabled downstream distributions such as Fedora and Ubuntu use the newly introduced Power architecture extensions to create and manage KVM guests in the Power Linux LPAR.

- ▶ QEMU: user space component that implements virtual machines on the host that use KVM functions.
- ▶ LibVirt provides a toolkit for virtual machine management.

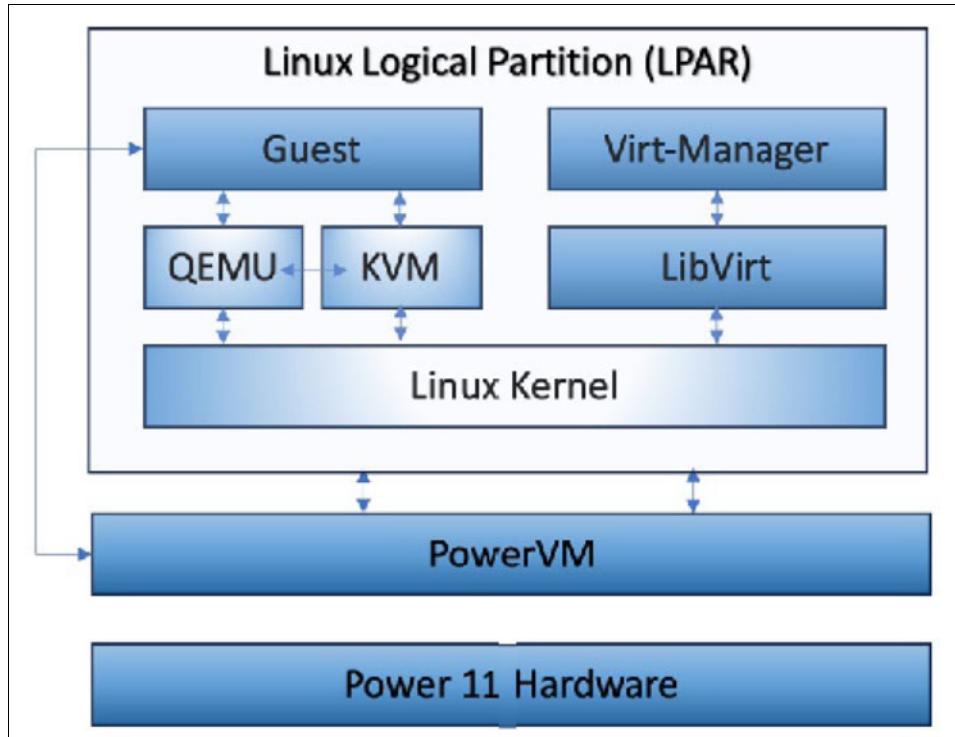


Figure 9-14 Industry Standard Linux Virtualization Stack

KVM support on IBM Power11 processor based server requires:

- ▶ Partition must be a Linux partition that runs in Power10 processor compatibility mode;
- ▶ Partition must be enabled for KVM:
 - For HMC-managed systems, you must set the partition to KVM Capable;
 - For unmanaged systems, you must set the default partition environment to Linux KVM on the BMC;
- ▶ Partition must be running in Radix mode (default MMU mode for Linux LPARs);
- ▶ partition must be assigned with dedicated CPUs with processor sharing set to Never Allow;

The following features are not supported on KVM logical partitions:

- ▶ Shared processors
- ▶ vPMEM LUNs
- ▶ Platform keystore
- ▶ Live partition migration (LPM)
- ▶ Dynamic platform optimization
- ▶ Add or Remove memory, processor, and I/O DLPAR

The following KVM features for guests are not supported:

- ▶ PCI pass-through of LPAR-attached PCI devices to KVM guests. This feature will be supported in future releases.

See <https://www.ibm.com/docs/en/linux-on-systems?topic=servers-kvm-in-powervm-lpar> for additional references.



Hybrid Cloud Solutions

This chapter explores IBM Power servers as a foundational platform for hybrid cloud environments, emphasizing their ability to support mission-critical workloads with exceptional performance, security, and flexibility. Built on the advanced Power11 architecture, IBM Power Servers deliver significant improvements in compute efficiency and operational cost savings. These systems are engineered to integrate seamlessly across on-premises, private, and public cloud infrastructures, supporting a wide range of enterprise operating systems including AIX, IBM i, and Linux.

The chapter also highlights IBM's strategic approach to hybrid cloud, showcasing how Power servers enable scalable, cloud-native application development using Kubernetes and Red Hat OpenShift. Key management tools such as IBM PowerVC, IBM Cloud Management Console, and Power Enterprise Pools 2.0 are discussed for their roles in simplifying virtualization, optimizing resource allocation, and enhancing operational agility. Additionally, the chapter introduces IBM Power Virtual Server (PowerVS), a cloud-based extension of IBM Power that allows organizations to modernize at their own pace. PowerVS offers flexible, pay-as-you-go access to virtualized IBM Power infrastructure, enabling seamless workload migration, rapid provisioning, and robust business continuity capabilities.

This chapter contains the following topics:

- ▶ IBM Power Private Cloud with Shared Utility Capacity
- ▶ Red Hat OpenShift
- ▶ Power Virtual Server

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 innovation allows 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 enhances operational agility by supporting dynamic, minute-by-minute consumption of compute resources.

In this model, each system within the pool is provisioned with a set number of base processor and memory activations. When configured as a Power Enterprise Pool, these base activations and their associated operating system entitlements are aggregated to define the pool's baseline capacity. Simultaneously, all hardware resources across the pool are fully activated and made available for workload deployment – allowing usage to exceed the base capacity when needed. IBM Cloud Management Console 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, offering granular resource sharing, improved cost efficiency, and simplified management for enterprise IT environments.

Power Enterprise Pools with Shared Utility Capacity on 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 E1080 systems within a single Power Enterprise Pools (2.0). Shared Utility Capacity provides a complete range of 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 Power Enterprise Pools 2.0 of Power E1080 systems can simply add one or more Power E1180 systems into their pool and migrate to them at the rate and pace of their choosing, as any Power E1180 and Power E1080 systems may seamlessly inter-operate and share compute resources within the same pool. Clients with Mobile Capacity on a Power E1180 may easily upgrade their system to support Power Enterprise Pools 2.0 to leverage Shared Utility Capacity resources.

A Power infrastructure consolidated onto Power E1180 systems has the potential to greatly simplify system management so IT teams can focus on optimizing their business results instead of moving resources around within their data center. Shared Utility Capacity resources are easily tracked by virtual machine (VM) and monitored by a CMC, which integrates with local HMCs to manage the pool and track resource use by system and VM, by the minute, across a pool.

Clients need not over-provision capacity on each individual system to support growth, as 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 now offered through the purchase of a Power Enterprise Pools Subscription (5765-P2E) for each Power E1180 system being configured for use in a Power Enterprise Pool.

When a Power Enterprise Pool 2.0 pool is started, each eligible Power E1180 system's purchased processor activations, memory activations, and supported operating system entitlement resources become Base Capacity resources as part of the Power Enterprise Pool and are aggregated across a defined pool of systems for consumption monitoring. Metered Capacity is the additional installed processor and memory resource above each system's Base Capacity. It is activated and made available for immediate use when a pool is started, then monitored by the minute by a CMC.

Metered resource usage is charged only for minutes exceeding the pool's aggregate Base resources, and usage charges are debited in real-time against a client's Capacity Credits (5819-CRD) on account or may be billed monthly, in arrears, where available.

Important: Note: Only two consecutive generations of Power servers are supported in the same Power Enterprise Pool. Customers who want to add Power11 to their existing pool with Power9 and Power10 processor-based systems, are recommended to reach out to their IBM representatives or IBM Technology Expert Labs to discuss transition planning and best practices.

10.1.1 IBM Cloud Management Console

As private and hybrid cloud deployments continue to expand, enterprises require deeper management insights into these increasingly complex environments. Tools that deliver unified analytics and consolidated information are essential for ensuring smooth and efficient infrastructure operations.

The IBM Cloud Management Console for Power offers a comprehensive view of your systems cloud environment – regardless of the number of systems or data centers involved. It provides:

- ▶ Centralized inventory management of systems and virtual components
- ▶ Consolidated performance metrics to help optimize resource utilization and system performance across all data centers
- ▶ Aggregated logging and analytics for enhanced operational insights

This unified approach empowers IT teams to manage their infrastructure more effectively and make informed decisions with confidence. The IBM Cloud Management Console (CMC) now offers full support for IBM Power11 systems, ensuring compatibility with the latest advancements in Power architecture.

The IBM Cloud Management Console (CMC) is also used to monitor and manage Power Enterprise Pool 2.0 pools in your enterprise. The CMC Enterprise Pools 2.0 application helps you to monitor base and metered capacity across a Power Enterprise Pool 2.0, with summary and sophisticated drill-down views of real-time and historical resource consumption by the logical partition.

CMC provides a cloud-based interface to monitor, manage, and optimize IBM Power infrastructure. It supports AIX, IBM i, and Linux VM workloads, and is accessible through a secure, web-based dashboard. The IBM Cloud Management Console is a cloud-based portal that allows clients to:

- ▶ View 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 those looking to streamline infrastructure oversight, CMC offers a way to perform essential monitoring and management tasks without requiring complex, locally hosted solutions.

CMC integrates with the Hardware Management Console (HMC) and provides advanced insights into performance, capacity, firmware compliance, patch planning, and system health. With Power11's scalability and built-in virtualization, CMC helps administrators maintain control over increasingly complex hybrid IT environments, whether on-premises or connected to IBM Cloud or other supported clouds.

CMC is offered through a software-as-a-service (SaaS) model and is licensed on a per-server subscription basis. Power11 systems are registered to IBM CMC through their HMCs, which securely transmit telemetry and configuration data to the cloud portal. IBM provides both a no-cost base offering and optional chargeable add-ons that include more advanced features, such as predictive analytics, patch automation, and longer data retention for historical performance views. Customers can choose monthly or annual billing, and entitlement can be managed through IBM Passport Advantage® or directly through the IBM Cloud portal.

10.2 Red Hat OpenShift

With the growing demand for digital services and the continuous release of new offerings, organizations face the challenge of delivering exceptional customer experiences while maintaining resilient and highly available services. To meet these demands, they must innovate – by developing new applications, modernizing existing ones, and ensuring seamless scalability.

So how can organizations achieve this? By embracing cloud technologies – whether through full cloud migration, developing cloud-native applications, or adopting a hybrid cloud approach.

Red Hat OpenShift on IBM Power empowers both developers and IT operations teams with the flexibility and speed needed to build, deploy, and manage applications across diverse environments – on-premises, in the cloud, or across multiple infrastructures. This platform accelerates digital transformation by simplifying scalability and enhancing security.

10.2.1 Red Hat OpenShift on Power

Red Hat OpenShift Container Platform is fully supported on IBM Power Servers, offering a robust and scalable environment for running cloud-native applications and modernizing traditional workloads on the IBM 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 OpenShift.

OpenShift on IBM Power provides a flexible and efficient platform for hybrid IT environments, enabling the seamless coexistence of cloud-native applications with traditional VM-based workloads. By leveraging the same infrastructure, it allows collocation of containerized workloads alongside existing applications running on AIX, IBM i, or Linux, ensuring low-latency communication between applications and data. IBM Power servers provide an extremely resilient and secure platform and how they can be an excellent platform for your cloud implementation using OpenShift. We then describe the benefits of using a multi-architecture cluster and provide implementation guidelines and advice to assist the reader in implementing a multi-architecture cluster using IBM Power control plane nodes and x86 or AMD-based worker nodes.

Red Hat OpenShift is a leading enterprise Kubernetes platform that provides a robust foundation for developing, deploying, and scaling cloud-native applications. It extends Kubernetes with additional features and tools to enhance productivity and security, making it an ideal choice for businesses looking to leverage container technology at scale.

Red Hat OpenShift is a unified platform to build, modernize, and deploy applications at scale. Work smarter and faster with a complete set of services for bringing apps to market on your choice of infrastructure. Red Hat OpenShift delivers a consistent experience across public cloud, on-premise, hybrid cloud, or edge architecture.

Red Hat OpenShift offers you a unified, flexible platform to address a variety of business needs spanning from an enterprise-ready Kubernetes orchestrator to a comprehensive cloud-native application development platform that can be self-managed or used as a fully managed cloud service

Red Hat OpenShift clusters

A Red Hat OpenShift Container Platform cluster consists of multiple nodes, which can run on either physical or virtual machines. 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 used during installation to host configuration and installation files

In 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 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 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 utilizing different types of hardware. A prime example includes data centers using x86_64 servers and Power servers, while edge locations might feature ARM-based devices due to their power efficiency. We discussed why IBM Power is an ideal choice for running Red Hat OpenShift clusters. However, it is possible that some applications within the environment aren't compatible with IBM Power architecture nodes. You can retain the benefits of IBM Power while retaining simpler cluster management and efficient resource utilization by incorporating both IBM Power and x86_64 architecture nodes in the same cluster.

The release of Red Hat OpenShift 4.14 brought the OpenShift Container Platform Multiple-Architecture Compute feature to IBM Power. Multi-Arch Compute provides a single heterogeneous cluster, enabling fit-for-purpose computing so clients can align tasks and applications to CPU strengths and software availability rather than one architecture. This support was expanded in Red Hat OpenShift 4.15 which enabled a Red Hat OpenShift cluster to support an IBM Power control plane and add x86 architecture worker nodes.

Multi-Arch Compute for OpenShift Container Platform lets you use a pair of compute architectures, such as ppc64le and amd64, within a single cluster. This exciting feature opens new possibilities for versatility and optimization for composite solutions that span multiple architectures.

Multi-architecture capability offers several benefits:

- ▶ Platform independence

Multi-Arch Compute enables applications to function flawlessly across various hardware platforms, including Intel servers in data centers, ARM-based Raspberry Pis in remote locations, and IBM Power Servers in corporate settings.

- ▶ Reduced complexity through standardizing application deployment across architectures

Streamlines operations and eliminates the need to maintain separate stacks for different hardware. Cost efficiency: Differing hardware architectures provide varying cost-to-performance ratios, which can be exploited to minimize overall infrastructure expenses.
- ▶ Multi-architecture support facilitates optimal resource usage

Organizations can select the most cost-effective architecture for each specific workload. For instance, ARM servers could be more affordable for lightweight services, whereas x86_64 or IBM Power servers might excel at handling heavy computational tasks.
- ▶ Energy efficiency

Specific architectures, like ARM, are renowned for their low power consumption, which can substantially decrease energy costs, notably in scale-out scenarios like IoT and mobile services.
- ▶ Scalability and flexibility

Implementing applications on multiple architectures enhances scalability and operational agility, which is vital for businesses experiencing fluctuating loads.

Multi-architecture support allows companies to:

- ▶ Scale elastically across platforms: Multi-Arch Compute enables businesses to dynamically allocate resources among different types of hardware to handle surges in demand without being restricted to a single architecture.
- ▶ Escape vendor lock-in: With Multi-Arch Compute, corporations are unshackled from a single supplier or type of hardware, thus avoiding vendor lock-in and empowering more bargaining power during procurement decisions.
- ▶ Optimize performance: Each architecture boasts unique strengths and weaknesses contingent upon the application or workload. Multi-Arch Compute maximizes performance by aligning application requirements with the architectural advantages.
- ▶ Create tailored solutions: Select applications might gain from the high I/O throughput of IBM Power Servers, while others could perform optimally on the high-throughput, multi-core processors of x86_64 architectures.

Meet specialized computing needs: Certain tasks may necessitate specialized hardware, such as GPUs for machine learning work flows, which might be more accessible or better supported on particular architecture.

Single Node clusters

For a simpler and more accessible deployment experience, OpenShift now supports a Single Node cluster where all of the management and worker functions are installed on a single node. This approach is especially suitable for development, testing, or edge computing environments.

However, it's important to recognize the inherent limitations of a single-node configuration – most notably, the absence of built-in high availability. If the node fails, the entire cluster becomes unavailable, potentially resulting in downtime and data loss. That said, the robust high availability features of IBM Power infrastructure, when properly architected with redundant power, networking, and storage, can significantly mitigate these risks.

10.2.2 Red Hat OpenShift AI on IBM Power Servers

OpenShift AI brings enterprise-grade capabilities for artificial intelligence and machine learning directly into the Red Hat OpenShift platform, enabling organizations to build, train, deploy, and monitor AI models within a unified, containerized environment. Red Hat OpenShift AI is built on Red Hat OpenShift, an enterprise-grade Kubernetes platform designed to support AI/ML workloads across hybrid and multi-cloud infrastructures. By integrating AI workflows into the same infrastructure used for traditional applications, OpenShift AI simplifies operations, accelerates development cycles, and provides consistent governance and security across the entire AI lifecycle. Red Hat OpenShift AI (RHOAI) enables organizations to build, deploy, and manage AI-enabled applications at scale across hybrid cloud environments.

OpenShift AI (RHOAI) Self-Managed is available by installing the Red Hat OpenShift AI Operator and configuring it to manage the standalone components of the platform. RHOAI Self-Managed is supported on Red Hat OpenShift Container Platform across multiple architectures, including ppc64le (IBM Power), s390x (IBM Z) and x86_64. This includes support for the following 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
- ▶ Google Cloud Platform
- ▶ Microsoft Azure
- ▶ VMware vSphere
- ▶ Oracle Cloud

IBM Power Servers, renowned for their high performance, scalability, and reliability, are particularly well-suited for compute-intensive AI workloads. The integration of RHOAI with IBM Power Servers enables enterprises to take advantage of both platforms' capabilities for robust, production-grade AI deployments.

As of RHOAI version 2.20 (Self-Managed), the IBM Power architecture (ppc64le) is available as a Technology Preview. Currently, model deployments on IBM Power are supported only in standard mode. RHOAI provides an integrated platform for developing, training, serving, and monitoring AI/ML models. Support for vLLM is also available on the IBM Power architecture as a Technology Preview, with vLLM runtime templates accessible for experimentation and development.

When deployed on IBM Power, 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 will further be enhanced with the planned availability and support of the IBM Spyre Accelerator in IBM Power11 servers. The Spyre adapter is a PCIe-attached AI card engineered for low-precision AI arithmetic and enterprise-grade AI algorithms, making it ideal for deploying large-scale models and generative AI frameworks.

Together, 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 operationalize AI with confidence – ensuring performance, efficiency, and flexibility across diverse workloads and environments.

10.2.3 IBM Cloud Paks

IBM Cloud Paks are a suite of modular, containerized software solutions designed to accelerate digital transformation. Built on Red Hat OpenShift, they enable organizations to modernize applications, automate business operations, manage data, and integrate AI across hybrid cloud environments with consistency and scalability.

Each Cloud Pak includes a combination of IBM middleware, open-source technologies, Kubernetes operators, and enterprise-grade security, providing a robust foundation for innovation and agility in the cloud. IBM Cloud Paks take a bundled approach that allows you to accelerate your modernization journey by packaging everything you need to get started.

There are three main benefits of IBM Cloud Paks:

- ▶ They are comprehensive and easy to use.
- ▶ They are supported by IBM.
- ▶ They run anywhere Red Hat OpenShift runs.

IBM Cloud Paks on IBM Power

Optimized for deployment on IBM Power, IBM Cloud Paks deliver high performance and efficiency for containerized workloads. They support a wide range of use cases – from data and automation to AI and security – while ensuring seamless integration across on-premises and cloud environments.

IBM Power enhances the value of Cloud Paks by offering superior performance, scalability, and cost-efficiency for modern workloads. Combined with Red Hat's open-source ecosystem, it empowers businesses to both modernize legacy applications and build new cloud-native solutions on a unified, enterprise-ready platform.

There are several IBM Cloud Paks that are currently supported on IBM Power, each focused on a specific domain:

- ▶ Cloud Pak for Applications
- ▶ Cloud Pak for Data
- ▶ Cloud Pak for Integration
- ▶ Cloud Pak for AIOps

IBM Cloud Pak for Applications

IBM Cloud Pak for Applications is an enterprise-ready, containerized software solution designed to modernize existing applications and develop new cloud-native apps. Built on IBM WebSphere® offerings and Red Hat OpenShift Container Platform, it provides a comprehensive set of tools to help organizations transition between public, private, and hybrid clouds.

IBM Cloud Pak for Applications includes IBM Cloud Transformation Advisor, an AI-powered tool which assists in refactoring and rearchitecting legacy applications. The solution includes automated vulnerability assessment and identification, ensuring 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 runtimes and DevOps workflows. This integration streamlines the development process and enhances productivity.

IBM Cloud Pak for Data

IBM Cloud Pak for Data allows you to unify and simplify 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 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, Amazon Web Services (AWS), and Microsoft Azure.

The platform allows secure access to data at its source, eliminating the need for data migration and reducing data silos, ensuring seamless data integration. It creates a trusted, business-ready analytics foundation, simplifying data preparation, policy enforcement, security, and compliance, while automating data governance and the AI lifecycle. IBM Cloud Pak for Data provides 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 operationalizing AI throughout the business with trust and transparency, the platform supports the end-to-end AI workflow, ensuring 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, while its scalable Kubernetes environment allows organizations to grow their data and AI capabilities as needed. Supporting multi-cloud 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 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, ensuring scalability and security. The platform includes components such as IBM API Connect for managing APIs, IBM App Connect for no-code integration, and IBM Event Streams for real-time data processing. By leveraging AI and automation, IBM Cloud Pak for Integration accelerates the integration process, reduces manual workflows, and enhances responsiveness to real-time events. This makes it an ideal solution for organizations looking 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 designed to automate work and accelerate business growth. Built for any hybrid cloud, it simplifies complex workflows, facilitates records management, and enhances overall productivity. The platform uses AI to identify gaps and build low-code and no-code automations, making it easier 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 designed to enhance IT operations (ITOps) by leveraging artificial intelligence and machine learning. It integrates seamlessly with existing ITOps toolchains to provide comprehensive visibility, proactive incident management, and automated remediation. By analyzing data from various 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 their 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 overall operational efficiency. With its collaborative tools and real-time insights, IBM Cloud Pak for AIOps empowers IT teams to innovate faster, reduce operational costs, and enhance the reliability of mission-critical workloads.

For more information about IBM Cloud Paks see [this IBM documentation](#).

10.2.4 Other cloud enablement solutions

From a Red Hat software perspective, there is also a comprehensive set of software solutions to accelerate your modernization efforts, including Red Hat Runtimes, Red Hat 3scale API Management, Red Hat Fuse and Red Hat AMQ.

10.3 Power Virtual Server

IBM Power Virtual Server is a cloud-based infrastructure-as-a-service (IaaS) offering that allows businesses to run IBM Power workloads in a flexible, scalable, and secure virtual environment. Built on the same architecture as IBM Power Servers used on-premises, it enables seamless hybrid cloud integration, making it ideal for enterprises looking to modernize their IT infrastructure without completely abandoning their existing investments. With support for AIX, IBM i, and Linux operating systems, Power Virtual Server provides a versatile platform for mission-critical applications. It provides an excellent platform for moving AIX, IBM i, and Linux on Power workloads to the cloud without a time consuming and risky migration of those workloads to a different platform.

One of the key advantages of IBM Power Virtual Server is its ability to deliver high performance and reliability for enterprise workloads. It leverages IBM's 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 options. These capabilities help organizations maintain business continuity and meet demanding service-level agreements (SLAs).

Additionally, IBM Power Virtual Server integrates with a wide range of IBM Cloud services and third-party tools, enabling users to build and manage hybrid cloud environments with ease. It supports DevOps practices, containerization with OpenShift, and AI/ML 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, IBM Power Virtual Server offers a cost-effective and agile solution for businesses aiming to innovate and grow in a digital-first world.

Power Virtual Server allows you to run IBM i, AIX, and Linux workloads in a cloud experience giving you fast, self-service provisioning, flexible compute, memory, and storage resources. PowerVS is designed to help organizations modernize their infrastructure, migrate to hybrid cloud models, and optimize their data center resources.

Key features and benefits of PowerVS include:

- ▶ **Flexible and Scalable:**
Users can easily adjust compute, memory, and storage resources on demand, scaling up or down as needed.
- ▶ **Pay-as-you-use:**
Billing is based on consumption, allowing organizations to manage cost effectively.
- ▶ **Hybrid Cloud Integration:**
PowerVS facilitates seamless integration of AIX and IBM i workloads with cloud-native applications and services.
- ▶ **Performance and Security:**
PowerVS leverages the performance and security features of IBM Power while offering the agility of cloud computing.
- ▶ **Managed Infrastructure:**
PowerVS offers managed infrastructure, where IBM handles the hardware maintenance, allowing organizations to focus on their workloads from a software and application standpoint.
- ▶ **Disaster Recovery:**
PowerVS can be used to implement disaster recovery 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 Power Virtual Server options

Power Virtual Server is a single offering that can be delivered in two variations; off-premises, where the infrastructure components are located in IBM data centers, and on-premises, where the infrastructure components are located 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, where you pay for resources as they are consumed, and are managed with the same management interfaces.

Figure 10-1 shows the two different implementations of IBM Power Virtual Server.

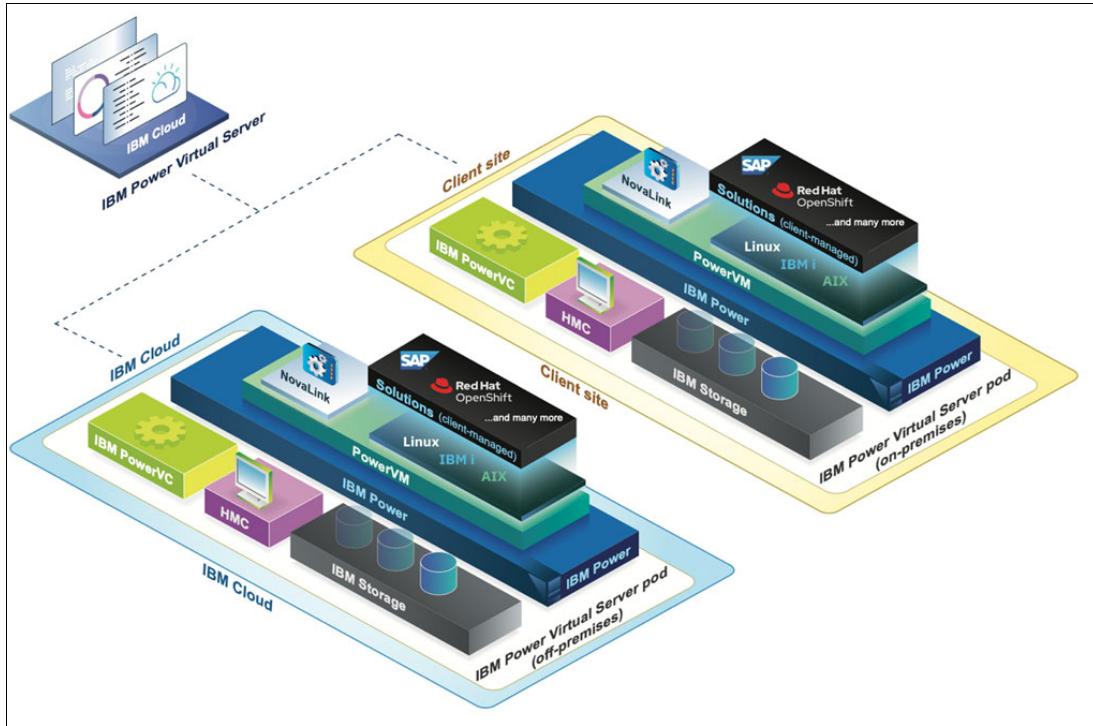


Figure 10-1 IBM Power Virtual Server options

Power Virtual Server Private Cloud is discussed in section 10.3.3, “Power Virtual Server Private Cloud” on page 203.

Use cases for Power Virtual Server

IBM Power Virtual Server supports a wide range of use cases, particularly for enterprises that rely on IBM Power Servers for mission-critical workloads. Here are some of the most common and impactful scenarios:

1. Hybrid Cloud Modernization

Organizations with on-premises IBM Power can extend their infrastructure to the cloud using Power Virtual Server. This enables a hybrid cloud model where workloads can be moved or replicated to the cloud for scalability, disaster recovery, or testing – without needing to replatform or rewrite applications. It's especially useful for businesses running AIX or IBM i environments that want to modernize gradually.

2. Disaster Recovery and High Availability

Power Virtual Server provides a reliable platform for disaster recovery (DR) by allowing businesses to replicate their on-prem workloads to the cloud. In the event of a failure, workloads can be quickly restored in the cloud, minimizing downtime. It also supports high availability configurations, ensuring business continuity for critical applications.

3. Development and Testing Environments

Developers can use Power Virtual Server to quickly spin up isolated environments for application development, testing, and quality assurance. This is particularly valuable for teams working on AIX, IBM i, or Linux-based applications, as it eliminates the need for dedicated on-prem hardware and allows for faster iteration and deployment cycles.

4. SAP HANA and ERP Workloads

IBM Power Virtual Server is certified to run SAP HANA and other SAP ERP applications. Enterprises can migrate or extend their SAP environments to the cloud for better scalability, performance, and cost-efficiency – especially when paired with IBM's high-performance Power10 processors.

5. AI and Machine Learning

Power Virtual Server supports AI/ML workloads, particularly those requiring high compute power and memory bandwidth. It integrates with IBM Watson® and Red Hat OpenShift, allowing data scientists to build, train, and deploy models using familiar tools while leveraging the performance of IBM Power.

6. Database Hosting

Many organizations use Power Virtual Server to host enterprise-grade databases like IBM Db2, Oracle, or PostgreSQL. The platform offers high I/O throughput and reliability, making it ideal for transaction-heavy applications such as banking, retail, and logistics.

7. Legacy Application Hosting

For businesses with legacy applications running on AIX or IBM i, Power Virtual Server provides a cloud-based alternative to aging on-prem hardware. This allows companies to extend the life of critical applications while gradually modernizing their infrastructure.

8. Compliance and Data Sovereignty

Organizations in regulated industries (e.g., healthcare, finance, government) use Power Virtual Server to meet strict compliance and data residency requirements. IBM offers regional data centers and compliance certifications to help meet these needs.

10.3.2 Power Virtual Server in the Cloud

In this variation, IBM Power Virtual Server resources reside in IBM data centers with dedicated networking and storage area network (SAN) attached storage. Customers can choose from one of the currently 21 different data centers, choosing the data center that is closest to their users. IBM Power clients who rely on private cloud infrastructure can now quickly and economically extend their Power IT resources on the cloud.

In these IBM data centers, Power Virtual Server resources are separated from the rest of IBM cloud servers with separate networks and direct attached storage. In this offering, customers are given flexibility to choose from different IBM Power server models, an IBM supplied OS image, and different tiers for storage based on your specific workload requirements. Customers are also given to choose their own image as Bring Your Own License (BYOL) so that they can bring already configured images from their enterprise infrastructure onto Power Virtual Server. This offering also supports distinctive features such as shared processor pools, public connectivity of VMs, placement groups, and global replication service (GRS) which can be consumed based on customer's requirements.

For more information on IBM PowerVS: Getting Started with IBM Power Virtual Server

10.3.3 Power Virtual Server Private Cloud

For those clients that want a cloud experience running on IBM Power servers, but have requirements to keep control of their data and applications (whether those requirements are regulatory, security based, or perhaps performance based). For those clients, IBM designed an offering where we bring the Power Virtual Server offering to a data center of your choice.

Power Virtual Server Private Cloud extends all the benefits of IBM Power Virtual Server into your (or a partner's) data center. The enhanced capabilities of IBM Power Virtual Server Private Cloud provide managed infrastructure as a service at client locations, with metered consumption and no upfront costs to support Hybrid by Design delivery of services.

The IBM Power Virtual Server Private Cloud offering is engineered to:

- ▶ Maintain customer data and workloads on your own site.

Enterprises may have workloads or data that is regulated and cannot be hosted off-premises. In some cases, enterprises can have workloads that are sensitive or with ultra-short latency requirements that are better served on site and in very close proximity with other on-site workloads.

- ▶ Maintain customer data in region and specific geographies in the location of their choice.

Country sovereignty regulations are requiring some data and workloads to stay in country. According to a recent IBM Institute of 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.

- ▶ Provide a seamless hybrid cloud experience.

Enterprises can foster a unified hybrid cloud landscape by seamlessly integrating Power Virtual Server running both at an IBM site and at a client site location with the ability to manage all the virtual machines (VMs) and infrastructure effortlessly through a unified user interface. Clients can receive the flexibility utilizing as-a-service with intentional workload placement on and off premises.

- ▶ Deliver predictable charging model with committed monthly spend combined with flexible consumption with metered usage-based pricing.

Both Power Virtual Server offerings, off-premises running at an IBM location and on-premises running at the client site, include compute, memory, storage, and operating system licenses that are fully metered by the hour allowing clients to pay for how much they use each month with no upfront payment.

- ▶ Streamline IT operations.

Whether in the cloud or at an enterprise's site, IBM manages the infrastructure, freeing enterprises to focus on business outcomes and less on managing infrastructure. IBM will own, deliver, and set up the Power Virtual Server in datacenter of choice, and provide a fully managed solution, including monitoring, security, firmware updates, and infrastructure management.

- ▶ Provide enhanced security and control of data.

IBM Power Virtual Server is designed to provide comprehensive security for IBM Power infrastructure by integrating with IBM Cloud tooling to manage security. This alleviates the need to manage Power infrastructure security with the added benefit of maintaining sensitive data and workload on-premises.

The physical infrastructure is delivered as a point of delivery (pod) which will be deployed in customer's data center. A pod is the physical component which resides within the client datacenter and contains the compute, storage and network components. A pod contains one or more racks where each of the components are installed. The racks are interconnected to provide a completely self-contained infrastructure, including both customer usable components, spare components, and management components.

This pod will be maintained by IBM site reliability engineering (SREs) and managed through IBM Cloud platform. Each pod is associated with an IBM Cloud satellite location that is owned by customer's IBM cloud account. This architecture provides the ability to scale your private cloud infrastructure horizontally by adding more pods to meet your workloads requirements.

For more information about Power Virtual Server Private Cloud refer to the IBM Redpaper *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 IBM Power Virtual Server and IBM Power Virtual Server Private Cloud. With the launch of IBM Power11, clients and ISVs gain immediate access to the latest Power hardware in the cloud – enabling faster innovation and greater flexibility. Whether you're modernizing applications or enhancing business continuity, Power11 in PowerVS delivers a seamless hybrid cloud experience.

Key benefits include:

- ▶ Rapid Deployment – Launch 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 shows the initial implementation of Power11 processors in Power Virtual Server on announcement day.

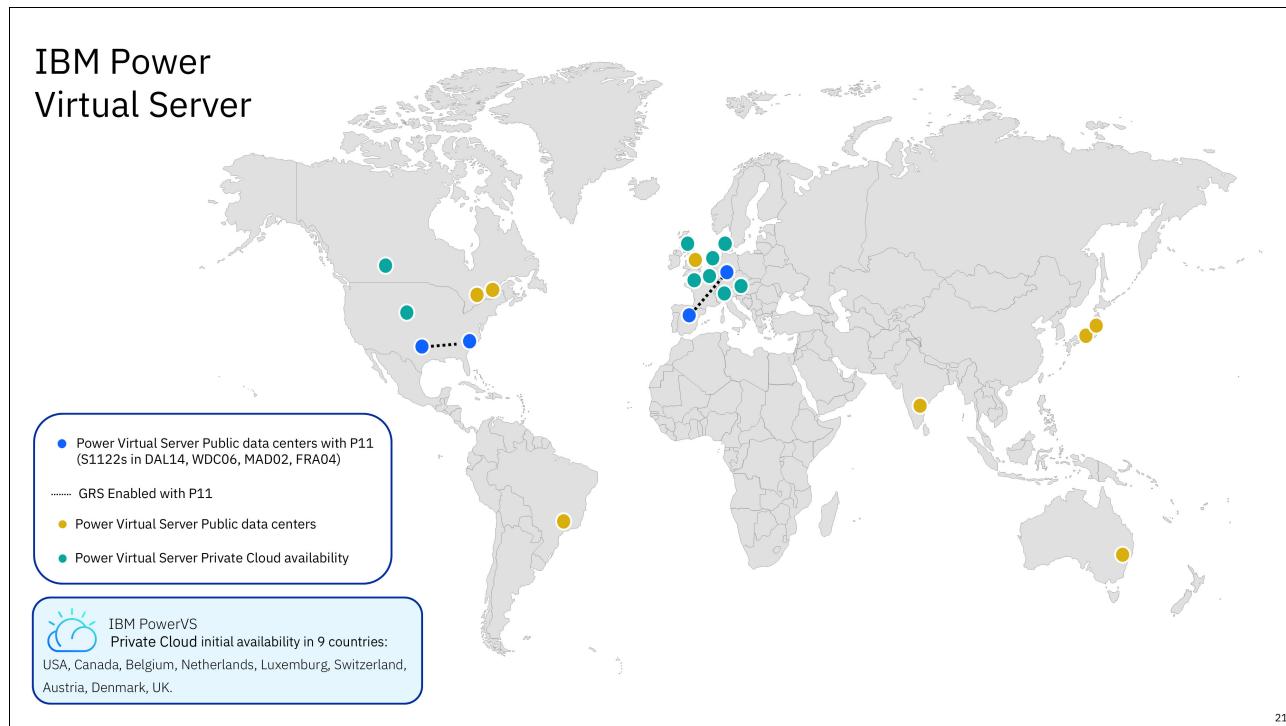


Figure 10-2 Availability of Power11 in Power Virtual Server

21

Power11 technology in Power Virtual Server allows you to immediately take advantage of the new functionality without a requirement for initial capital investment. You can safely implement workloads taking advantage of the Trusted security capabilities of Power11 like:

- ▶ Quantum-safe infrastructure compliance with optional workload cryptographic inventory discovery
- ▶ A unified cyber resiliency solution responding to evolving cyber threats and regulatory standards with optional 3rd party services

In addition, you can immediately take advantage of the improved performance of Power11 as shown in Figure 10-3.

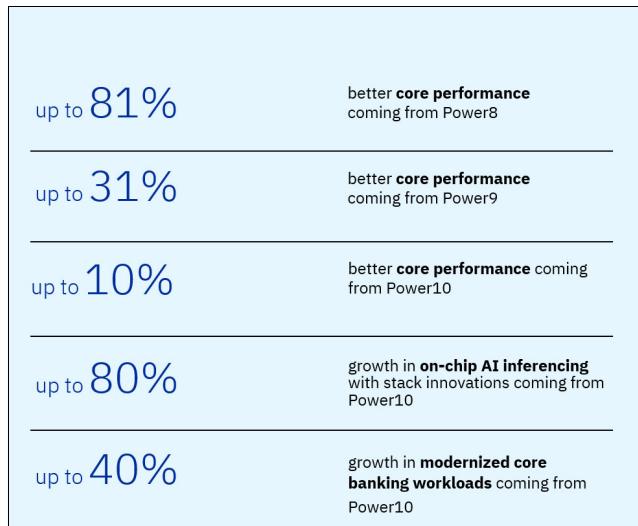


Figure 10-3 Potential performance benefits from Power in PowerVS

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- ▶ *Modernization Techniques for IBM Power*, SG24-8582
- ▶ *IBM i 7.6 features and function*, SG24-8588
- ▶ *IBM Power Security Catalog*, SG24-8568
- ▶ *Creating OpenShift Multiple Architecture Clusters with IBM Power*, SG24-8565
- ▶ *Introduction to IBM Power Virtual Server Private Cloud*, REDP-5745
- ▶ *Using Ansible for Automation in IBM Power Environments*, SG24-8551
- ▶ *IBM PowerHA SystemMirror and IBM VM Recovery Manager Solutions Updates*, REDP-5694

You can search for, view, download or order these documents and other Redbooks, Redpapers, Web Docs, draft 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: Supported Configurations
<https://access.redhat.com/articles/rhoai-supported-configs>
- ▶ Red Hat OpenShift AI Self-Managed 2.20 - Documentation
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 Self-Managed 2.20
https://docs.redhat.com/en/documentation/red_hat_openshift_ai_self-managed/2.20
- ▶ Red Hat OpenShift AI Service Definition
<https://access.redhat.com/articles/7102731>
- ▶ OpenShift Operator Life Cycles
https://access.redhat.com/support/policy/updates/openshift_operators

- ▶ RHEL Versions Utilized by RHEL CoreOS and OCP
<https://access.redhat.com/articles/6907891>
- ▶ 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>

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services



IBM Power E1180: Introduction and Overview

SG24-8587-00

ISBN



(1.5" spine)
1.5" <-> 1.998"
789 <-> 1051 pages



IBM Power E1180: Introduction and Overview

SG24-8587-00

ISBN



(1.0" spine)
0.875" <-> 1.498"
460 <-> 788 pages



IBM Power E1180

SG24-8587-00

ISBN



(0.5" spine)
0.475" <-> 0.873"
250 <-> 459 pages



IBM Power E1180

(0.2"spine)
0.17" <-> 0.473"
90 <-> 249 pages

(0.1"spine)
0.1" <-> 0.169"
53 <-> 89 pages



Redbooks

IBM Power E1180: Introduction and Overview

SG24-8587-00

ISBN



(2.0" spine)
2.0" <-> 2.498"
1052 <-> 1314 pages



Redbooks

SG24-8587-00

ISBN



(2.5" spine)
2.5" <-> nnn.n"
1315 <-> nnnn pages



SG24-8587-00

ISBN

Printed in U.S.A.

Get connected

