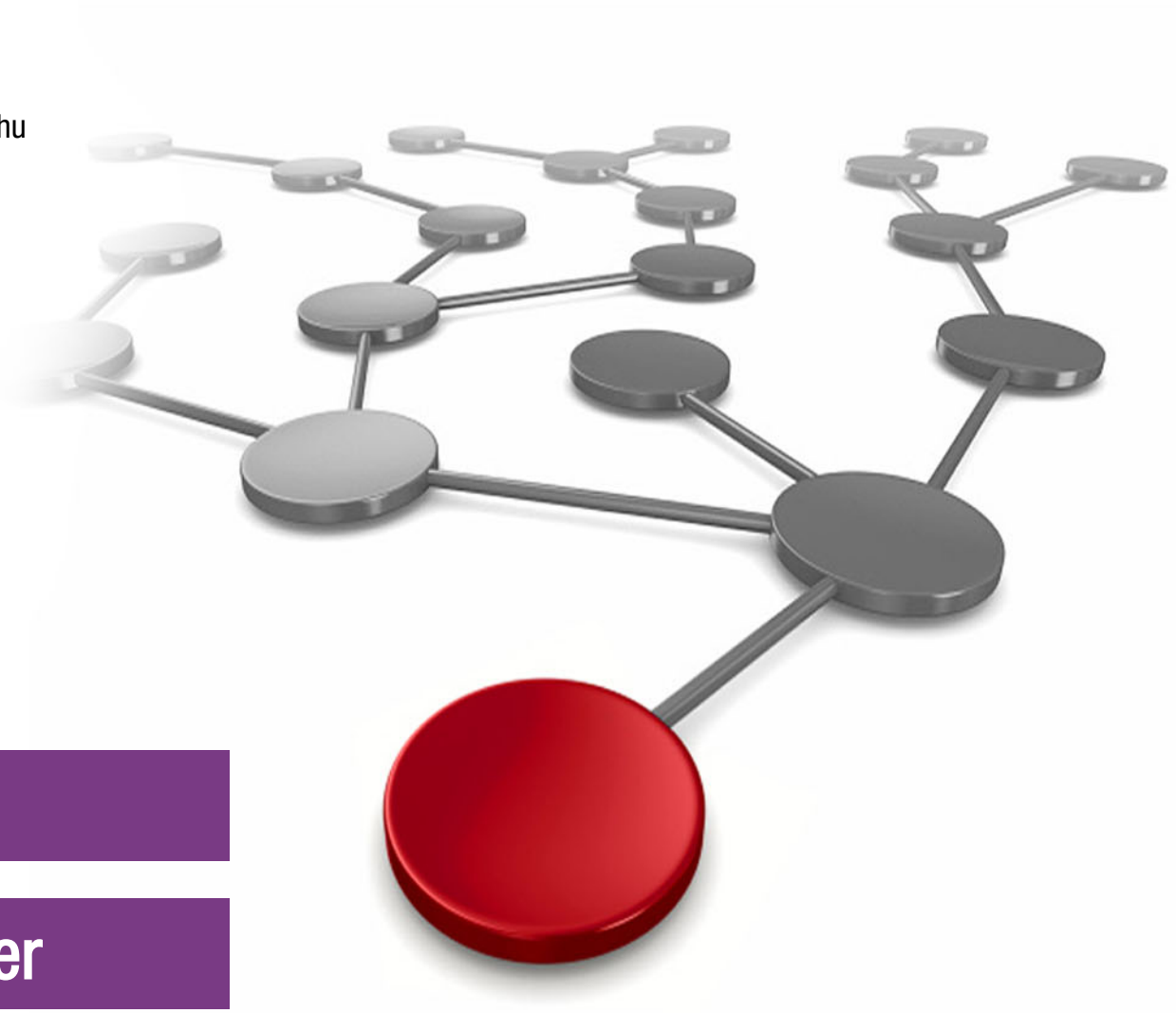
Redbooks
ibm.com/redbooks

Introduction to IBM Power Virtual Server Private Cloud

Tim Simon
Vijaybabu Ananimuthu
Samvedna Jha
Marius le Roux
Raydo Matthee
Arnold Ness
Adrian Orben
Borislav Stoymirski
Keith Uplinger
Henry Vo



Cloud

IBM Power

IBM

Redpaper

IBM Redbooks

Power Virtual Server Private Cloud

January 2025

REDP-5745-00

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

First Edition (January 2025)

This edition applies to IBM Power Virtual Server (December 2024).

This document was created or updated on January 29, 2025.

Contents

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®
Aspera®
CICS®
DB2®
Db2®
DS8000®
Guardium®
IBM®
IBM Cloud®
IBM Cloud Satellite®
IBM FlashSystem®
IBM Instana™
IBM Security®
IBM Spectrum®
IBM Z®
Instana®
POWER®
Power8®
Power9®
PowerHA®
PowerVM®
QRadar®
Rational®
Redbooks®
Redbooks (logo)  ®
Satellite™
System z®
SystemMirror®
XIV®

The following terms are trademarks of other companies:

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.

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.

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.

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

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

Preface

Enterprises adopting hybrid cloud and AI technologies require a flexible infrastructure for performance, security, and interoperability. This adaptability enables organizations to remain competitive, cut costs, and better respond to customer demands. To achieve this, enterprises need platforms that allow for quick adjustments in workload management and placement, focusing on business outcomes over infrastructure maintenance. IBM's Hybrid Cloud strategy prioritizes flexibility and choice.

IBM Power Virtual Server has provided infrastructure-as-a-service globally since its 2019 launch, combining compute, storage, and networking in IBM data centers. This reliable architecture supports mission-critical workloads and facilitates easy cloud migration without refactoring. IBM has introduced a new Private Cloud version of IBM Power Virtual Server, which installs infrastructure in a client data center, connecting a local zone to IBM Cloud® and offering a consistent as-a-service experience. This model enhances agility, reliability, and security, optimizing IT operations for diverse enterprise needs.

IBM Power Virtual Server Private Cloud helps customers leverage hybrid cloud architecture benefits, even when workloads are regulated, or data must stay in a specific country due to government regulations.

Authors

This paper was produced by a team of specialists from around the world working at IBM Redbooks, Remote Center.

Tim Simon is an IBM® Redbooks® Project Leader in Tulsa, Oklahoma, USA. He has over 40 years of experience with IBM, primarily in a technical sales role working with customers to help them create IBM solutions to solve their business problems. He holds a BS degree in Math from Towson University in Maryland. He has worked with many IBM products and has extensive experience creating customer solutions using IBM Power, IBM Storage, and IBM System z® throughout his career.

Vijaybabu Ananimuthu is a Technical Consultant at IBM Systems Experts Labs in India. He holds a bachelor's degree (BE) in Electrical and Electronics Engineering from Anna University, Chennai. He has over 16 years of experience working with customers designing and deploying solutions on IBM Power server and AIX®. He focuses on areas such as IT Infrastructure Enterprise Solutioning, technical enablement and implementations relative to IBM Power servers, Enterprise Pools, Performance, and automation. His areas of expertise include capacity planning, migration planning, system performance and automation.

Samvedna Jha is a Senior Technical Staff Member in the IBM Power Systems organization, Bengaluru, India. She holds a masters degree in Computer Application and has more than twenty years of work experience. In her current role as Security Architect, IBM Power, she has worldwide technical responsibility to handle security and compliance requirements of Power products. Samvedna is a recognized speaker in conferences, has authored blogs and published disclosures. She is also the security focal point for Power products secure release process.

Marius le Roux is an adept IBM i professional with a robust background in operations, development, and integration since 2006. His expertise extends beyond technical skills to a

profound understanding of overall enterprise concepts, enabling him to make impactful decisions that enhance system functionality and business alignment. Through his comprehensive experience with IBM i environments, Marius effectively bridges the gap between complex IT solutions and strategic enterprise objectives. This unique blend of technical acumen and business insight ensures that he not only executes projects but also innovates and optimizes them for long-term success.

Raydo Matthee is a South African technology leader and Solutions Architect at Skunkworks (Pty) Ltd, with over 15 years of experience in cloud computing, security, and enterprise integration. Specializing in IBM Power Virtual Server and hybrid cloud solutions, Raydo helps businesses integrate cloud technologies with a focus on security, compliance, and scalability. He has authored multiple courses on topics like OWASP, HashiCorp, and cloud security, reflecting his commitment to advancing IT education and consultancy.

Arnold Ness is a Senior Power Technology Sales Leader in Canada. He has 40 years of experience in IBM and Ciena working with customers designing and implementing business solutions leveraging technology. He holds an MBA in Information Technology Management from Royal Roads University, an Electrical Engineering degree from the University of Alberta and completed IBM's Client Executive program at Harvard. His areas of expertise include solution design and development across x86, IBM Power and IBM 390 platforms. He was awarded the Lou Gerstner Award for Client Excellence in 2022 and has implemented solutions with clients across North and South America. His interests lie in technology innovation, sustainability, Hybrid cloud, Artificial Intelligence, Cybersecurity, Internet of Things (IoT), and Quantum-Safe Computing.

Adrian Orben joined IBM in 2003. Since 2006, he has been a member of the Storage Support Team that is responsible for IBM high-end disk products. His focus is supporting IBM DS8000®, IBM XIV®, and IBM Flash Systems. During this time, Adrian achieved several storage-related certifications. Adrian holds a bachelor’s degree in Business Informatics from the University of Applied Science Mainz, Germany.

Borislav Stoymirski is an IBM Power Servers Hardware Product Engineer at IBM Bulgaria, specializing in solving complex hardware and software issues on IBM Power Servers, including IBM AIX, VIOS, HMC, IBM i, PowerVC, PowerVM® NovaLink, PowerVM, Linux on IBM Power Servers, and Red Hat OpenShift. Since joining IBM in 2015, he has provided reactive break-fix, proactive, preventative, and cognitive support to several clients worldwide. Borislav earned a masters degree in Computer and Software Engineering from the Technical University of Sofia, Bulgaria, a masters degree in Transport Machinery and Technologies from the Technical University of Sofia, Bulgaria, and another masters degree in Ecology and Environmental Protection from the University of Chemical Technology and Metallurgy, Bulgaria. He has authored several publications and has led technical training programs both inside and outside IBM. His interests lie in Green Blockchain, Artificial Intelligence, Deep Learning, Machine Learning, Cybersecurity, Internet of Things (IoT), Edge Computing, Cloud and Quantum Computing.

Keith Uplinger is an IT Architect for Power Virtual Server Private Cloud in Dallas, Texas, USA. He has over 20 years of experience in developing full stack software solutions as well as designing custom hardware solutions. He holds a degree in Computer Science from Texas Tech University. He was the lead on IBM World Community Grid for many years and worked towards optimizing research applications to further research in fields of Cancer, Weather, Water Filtration and Covid. Keith was crucial to the development of Asynchronous Replication Exchange for molecule folding application used for drug discovery. Most recently Keith has been the primary developer on Power Virtual Server Private Cloud integrating many different software requirements in automated fashion to accelerate the delivery time for a Pod ordered by a customer.

Henry Vo is an IBM Redbooks Project Leader with more than 10 years of experience in IBM. He has technical expertise in business problem solving, risk/root-cause analyze, and writing technical plans for business. He has had multiple roles in IBM such as Project management, ST/FT/ETE Test, Back End Developer, DOL agent for NY and is certified in IBM zOS Mainframe Practices, IBM Z® System programming, Agile, and Telecommunication Development Jumpstart. Henry holds a Master of MIS (Management Information System) from the University of Texas at Dallas since 2012.

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

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

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 papers to be as helpful as possible. Send us your comments about this paper or other IBM Redbooks publications in one of the following ways:

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

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

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

Stay connected to IBM Redbooks

- Find us on LinkedIn:

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

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

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

- Stay current on recent Redbooks publications with RSS Feeds:

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

1

Hybrid by Design

IBM continues to invest in concepts that bring value to enterprises seeking to increase their return on investment for infrastructure and cloud implementations. Hybrid by Design, a proven architectural framework, leverages hybrid cloud technology to maximize business value. It provides the agility, speed, and integration necessary for achieving future business outcomes.

Hybrid by Design, originally rooted in cloud architecture, outlines how organizations can shape their hybrid cloud environments based on business priorities. By combining public and private clouds with on-premises data centers, enterprises enhance their agility, speed, and ability to scale initiatives. With the rise of generative AI, Hybrid by Design extends beyond cloud computing to encompass the entire enterprise technology landscape—including platforms, security, AI, cloud, and data. This approach transforms diverse technologies into a cohesive system, amplifying business outcomes through thoughtful design and integration.

The IBM Power Virtual Server Private Cloud offering, based on Hybrid by Design, allows enterprises to install infrastructure within their own data centers, providing added security and control of their data. This solution allows the use of cloud resources with no upfront costs and pay-as-you-go pricing.

This chapter introduces the Power Virtual Server Private Cloud and discusses the benefits it brings to enterprise IT environments.

The following topics are included in this chapter:

1.1 The IBM Hybrid by Design approach

As enterprises embark on their journeys to adopt innovative hybrid cloud and AI technologies, its increasingly vital for these organizations to have a flexible infrastructure that delivers performance, security, and interoperability to support evolving IT needs. To remain competitive, reduce technology costs, and improve the ability to respond to changing customer needs, enterprises will need platforms that allow workload consumption to ramp up or down quickly, place workloads on cloud or on-site, with a focus on business outcomes and less on infrastructure management. Generative AI also has the potential to be a key factor in business transformation, and enterprises that take a more intentional approach with hybrid cloud could be better positioned to accelerate their impact.[1]

The IBM Hybrid by Design[2] approach is a comprehensive framework for the strategic implementation of a hybrid cloud architecture with 5 Key Priorities:

1. Drive product-centric mindset to enable business priorities.
2. Build intentional architecture to accelerate & scale business competencies.
3. Create a consistent development and operational experience across platforms.
4. Empower product teams to exploit hybrid capabilities.
5. Harness the power of all data and scale Gen AI deployment.

Figure 1-1 presents the results of a current IBM Institute for Business Values study, which reveals that enterprises are working toward operating at enterprise scale. The study shows that enterprises expect to significantly increase their platform operational capabilities, and hybrid by design platforms are tailored to meet this growth requirement.

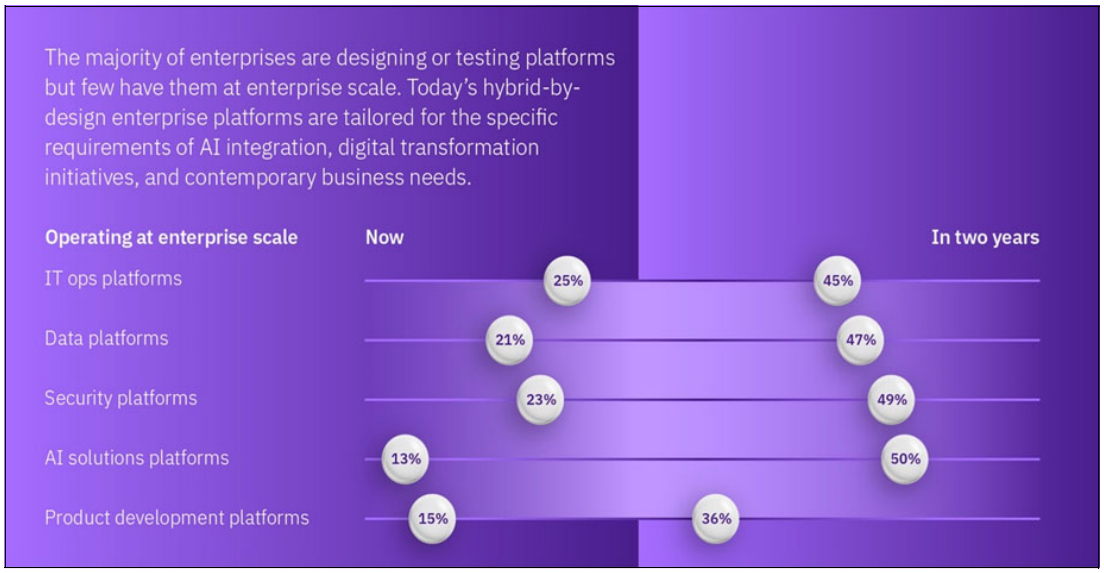


Figure 1-1 Results of IBM Institute for Business Value study

[1] https://newsroom.ibm.com/blog-infuse-ai-into-ibm-power-mission-critical-workloads-on-the-cloud-at-our-site-or-yours

[2] https://www.forbes.com/sites/stevemcdowell/2024/06/18/hybrid-by-design-a-new-paradigm-for-enterprise-cloud-success/

The Hybrid by Design versus a Hybrid by Default approach involves:

- Business-Centric Product Mindset[3] - The framework emphasizes a business-centric product mindset, prioritizing business objectives and enhancing productivity. This approach ensures that technology investments yield tangible business benefits and fosters a consistent development and operational environment across different platforms using common technology standards.
- Intentional Design - The core principle of IBM's Hybrid by Design framework is intentionality. Each technical component supports specific business objectives versus adopting cloud solutions haphazardly to avoid silos and incompatible islands of solutions.
- Scalability and Flexibility - Focus on creating systems that can scale with evolving business needs. The architecture is designed to adapt to changes in demand, technological advancements, and market conditions, enabling companies to remain agile and responsive.
- Value Creation[4] - Leverage a common automation strategy for installation, operations and development across all platforms. An enterprise API catalog supercharges innovation and productivity that can utilize Gen AI over time.
- Security and Compliance - Security is built into the product from the beginning. AI is used to consistently apply the highest standards for security across all environments and provide visibility across platforms with embedded automation.
- Operational Efficiency - Automate and streamline processes using standardized tools and processes to increase development productivity and reduce time to market.
- AI Integration[5] - IBM Hybrid by Design emphasizes the integration of AI within the hybrid cloud framework to drive transformative outcomes, optimize operations, and maintain a competitive edge in an increasingly digital landscape.
- Open and Integrated Foundation - Building an open, integrated foundation across all digital environments that leverages expertise with hybrid multi-clouds to maximize ROI.

IBM’s Hybrid Cloud strategy is focused on harnessing the power of choice and flexibility for clients. That’s why we recently extended the capabilities of Power Virtual Server to deploy in a client data center. This delivers flexibility around consumption and management capabilities of the cloud while keeping the data on-premises to help address regional compliance and governance requirements of the business.

1.2 Introduction to IBM Power Virtual Server Private Cloud

Since 2019, IBM has offered IBM Power Virtual Server to provide IBM Power resources in the cloud. The offering has grown to 21 data centers worldwide (with additional locations expected) and combines an infrastructure as a service (IAAS) model which includes IBM Power compute nodes with SAN attached IBM Storage and associated network connections within IBM Cloud locations. And now, as part of IBM’s distributed hybrid infrastructure strategy, 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.

[3] https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/hybrid-by-design/hybrid-by-design-architecting-for-agility
[4] https://www.ibm.com/consulting/hybrid-by-design
[5] https://www.ibm.com/hybrid-cloud

Figure 1-2 illustrates the comprehensive Power Virtual Server offering, spanning both IBM data centers and client locations. Power Virtual Server provides a platform for hosting IBM Power in a fully hybrid cloud enabled solution. This allows you to make the choice whether to place your workload in the IBM data centers or in your own data center (with an infrastructure owned and managed by IBM).

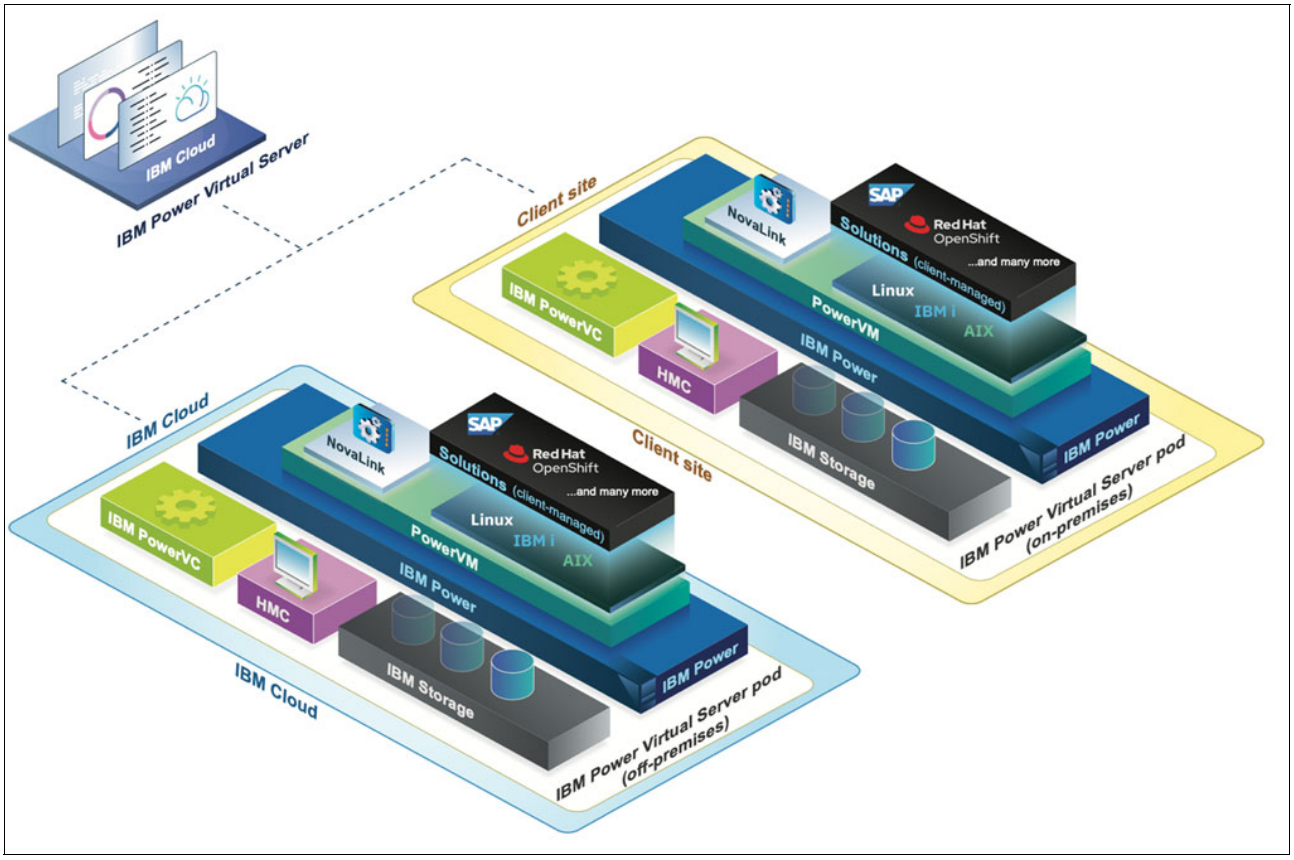


Figure 1-2 Power Virtual Server – a true hybrid cloud solution for your IBM Power workloads

Clients can achieve true hybrid cloud capabilities with their critical Power workloads either in IBM Cloud or on-premises for secure, sovereign, sensitive and regulated workloads. IBM Power Virtual Server Private Cloud provides the following benefits:

- Private and secure

  You will enjoy the flexibility of Infrastructure-as-a-Service and cloud management, while keeping your data on-premises to ensure security and data sovereignty, meeting regional compliance and governance requirements.

- Flexible and intentional placement of workloads

  Choose the best location to run each workload, off-premises in Power Virtual Server or on-premises in your data center based on your business requirements for security and data governance.

- Hybrid cloud enabled

  You will experience the same user interface available in Power Virtual Server and IBM Cloud, making hybrid cloud operations more efficient to manage your on-premises infrastructure.

- Metered usage-based pricing with flexible consumption and acquisition

  You pay only for what you use each month – compute, memory, storage, and operating system licenses are fully metered with no upfront costs[6].

- Streamlined IT operations and accelerated client Time to Value

  IBM provides the expertise to maintain and operate the infrastructure platform up to the Virtual Machine allowing clients to focus on their business outcomes.

- Constant improvements and enhancements

  IBM continues to improve Power Virtual Server with new capabilities, such as:

  - Identical management interfaces for workloads whether placed on Power Virtual Server in the cloud or on-premises.
  - Same architecture and support for a broad range of IBM Power servers, including scale up and scale out servers.
  - A single consolidated bill for IT services whether on-premises or off-premises.
  - Completely flexible consumption with the ability to grow to meet evolving business requirements.

Figure 1-3 shows how a unified platform with both a public and private cloud offering can allow customers to intentionally place their workloads in the best environment.

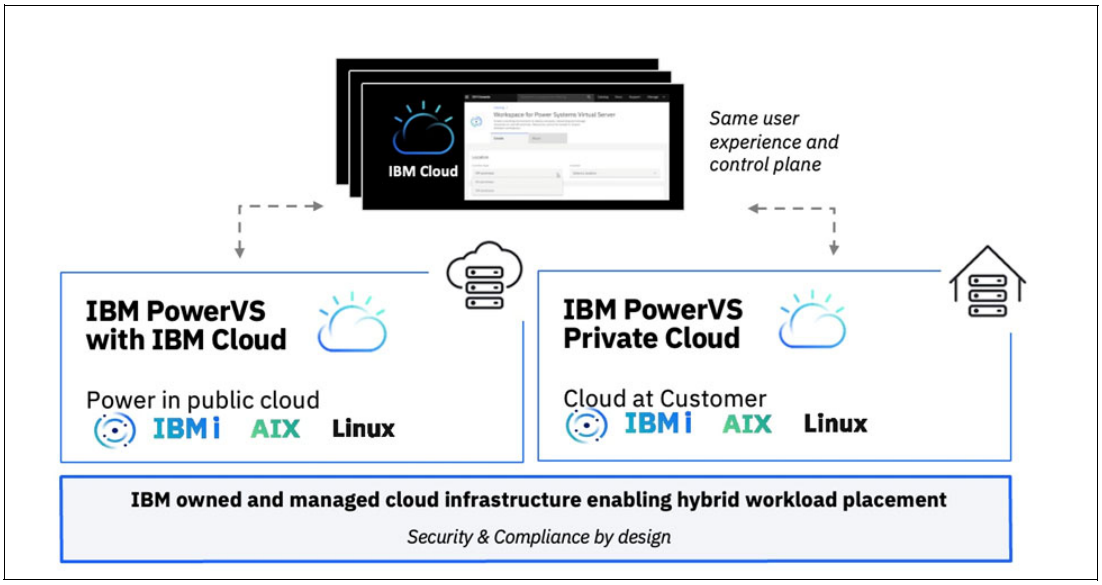


Figure 1-3 Hybrid cloud view of the Power Virtual Server offering

The IBM Power Virtual Server Private Cloud platform supports:

- Existing Power workloads with co-existence with new container based micro-services and modernization of legacy applications on a shared infrastructure.
- Support for products from independent software vendors (ISVs) running on IBM Power.
- Virtualized platform to replace VMware based Linux workloads to lower the total cost of delivery and increase the virtualization rate for applications.
- A cloud solution that provides the capability to meet specific regulatory requirements for data protection and data sovereignty.

[6] Clients are required to commit to a minimum monthly usage for the term (3 or 5 years) of the agreement and pay the greater of usage or the minimum commitment fees.

- Red Hat OpenShift 4.15 with Installer Provisioned Infrastructure (IPI) simplifies and automates deployment of OpenShift on IBM Power Virtual Server (~2 hours to setup OpenShift cluster vs. days/weeks)
- Infusing business applications on IBM Power with AI.

  Integrate AI onto IBM Power to run your business operations and AI integration on the same platform. This seamless integration enables you to incorporate AI into your existing workflows, where transactions are run, and data is located. Figure 1-4 illustrates this, showing how IBM Power Virtual Server applications integrate with **watsonx** to provide an enhanced AI experience.

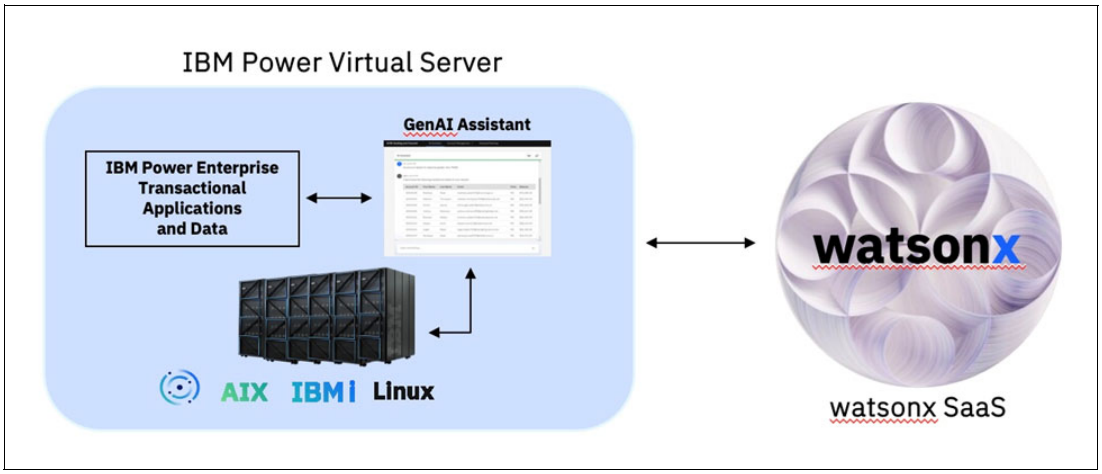


Figure 1-4 AI integration with Power Virtual Server and watsonx

With Power Virtual Server Private Cloud, customers can:

- Maintain customer data and workloads on their 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 require some data and workloads to stay in that 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 a predictable charging model with committed monthly spend combined with flexible consumption with metered usage-based pricing.

  Both IBM site and at client site offerings include compute, memory, storage, and operating system licenses that are 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 your 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.

1.3 Benefits of using IBM Power as your cloud platform

The same benefits that led to investment in IBM Power infrastructure in private data centers are provided by utilizing IBM Power Virtual Server infrastructure, either in the cloud or now a datacenter of choice. "Build on Power" for security, performance, reliability, availability, scalability, sustainability and a lower Total Cost of Ownership (TCO).

Fewer cores = Fewer licenses = Lower TCO.

Beyond the TCO benefits of using IBM Power based cloud infrastructure, there are migration efficiencies, management advantages, reliability and availability advantages, and security benefits. This section discusses the benefits IBM Power can bring to your cloud platform.

1.3.1 TCO Benefits

Hardware is typically less than 15% of the cost of an IT solution – with software and services comprising 85%. IBM Power is designed to support more workload per core when compared to x86 architecture alternatives. The power of the IBM Power processor core often results in only needing between 25-50% of the number of licenses compared to an x86 solutions. This is shown in Figure 1-5.

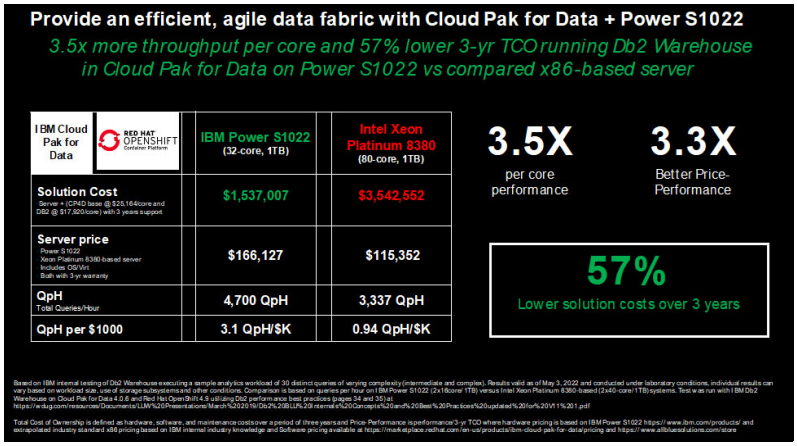


Figure 1-5 TCO comparison of Power and x86

Fewer cores equals fewer software licenses. The ongoing yearly support and service fees (which can be 20-25% of the initial purchase), and the additional cores required to support workloads on x86 can double your cost every 4-5 years. Choosing to implement a solution on

x86 could result in doubling the overall cost of the solution. Design for affordability by running on IBM Power servers to reduce TCO!

IBM Power servers are an integrated solution designed to provide:

- Reliability - Ranked #1 for the 15th year in a row by ITIC's global reliability report[7]
- Pervasive security - Most secure open system in the market:
  - Provides Transparent Memory Encryption,
  - 2.5X faster AES crypto performance per core with 4 crypto engines
  - Future ready for Quantum Safe Cryptography & Fully Homomorphic Encryption
- Performance:
  - Runs Oracle and other database workloads better than other platforms
  - 4.1 x more containerized throughput per core versus x86 running Red Hat OpenShift
  - Supports 8 threads/core and significantly higher utilizations
  - Significantly greater memory bandwidth versus x86
  - Collocate cloud applications and legacy applications on the same platform
  - AIX, IBM i, Linux (RHEL, SUSE), KVM, OpenShift can all be supported in separate partitions, running at the same time
- Sustainability - Reduce carbon footprint – 2x more capacity with the same energy usage
- AI-ready - 4 matrix math accelerators per core for faster inferencing

1.3.2 No refactoring

Prior to the introduction of IBM Power Virtual Server running in the cloud, migrating workloads from an existing IBM Power infrastructure to the cloud involved moving the applications and data to an x86 platform. This was complicated for several reasons:

- Compatibility Issues: Applications and workloads designed for IBM Power may not run natively on x86, requiring refactoring or recompiling.
- Performance Variations: Some workloads, particularly those that are compute intensive or require high throughput, may perform differently on x86 hardware.
- Data Migration: Moving data from one architecture to another can be complex, especially if it involves different database systems or file formats.

With IBM Power Virtual Server (either the public cloud or the new Private Cloud), the cloud target now uses the same technology. Migrating workloads are now much easier and there are a variety of tools that can be used to assist in the migration. While migration still takes careful planning, there is no need for refactoring or recompiling the applications.

The flexibility and ease of migration of applications and data makes IBM Power Virtual Server in the IBM Cloud an ideal solution for backup, disaster recovery, and development. Meanwhile, IBM Power Virtual Server Private Cloud offers secure, dedicated data centers tailored to each client’s needs.

1.3.3 Security and availability

Security and avail9ability are two important components of any enterprise application environment. Having a critical application down due to hardware issues can cost thousands of dollars an hour in some industries. In addition to the lost business caused by an unreliable infrastructure, your company’s reputation can be harmed. The IBM Power Virtual Server Private Cloud leverages IBM Power10 systems that provide industry leading availability as

[7] ITIC 2023 Global Server Hardware, Server OS Reliability Report

measured by the amount of unplanned downtime customers experienced. Security is also critical in an environment. Having a security breach can cost millions of dollars in lost business, loss of reputation and regulatory fines. IBM Power10 systems are designed with security in mind, with built-in encryption and available security tools to help secure and manage your environment.

Availability

The reliability of IBM Power10 is ranked #1 by ITIC with 99.999% availability as shown in Figure 1-6.

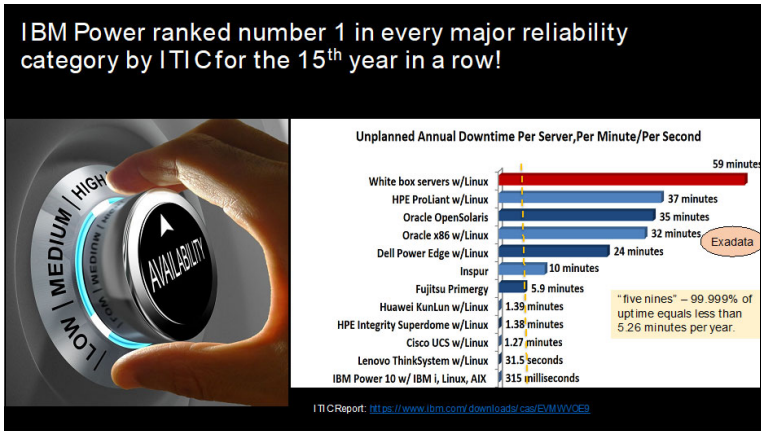


Figure 1-6 ITIC Reliability Report

In addition to the built-in reliability of the Power10 servers, IBM Power Virtual Server (both the public version and Private Cloud) are designed with reliability and availability built in. Redundant resources and the ability to restart your workloads automatically using those redundant components are designed into the solution. Also provided are storage solutions to keep your data accessible with backup and data replication services available to be configured to meet your availability requirements.

Another quote from Pfizer in reference to the requirement for availability and reliability in an ERP instance:

> *"To operate a single global ERP instance, you need a highly robust platform able to support very complex applications. Reliability, availability, serviceability, flexibility and similar strengths were all traits we needed to be ingrained into the platform—and IBM Power provided all of these."*[8]

Security

In today's digital landscape, IBM Infrastructure serves as a formidable shield against increasingly sophisticated cyber threats through its robust and integrated security solutions. IBM weaves security into the fabric of its systems and platforms, allowing businesses to operate confidently amid evolving risks.

At the heart of IBM's approach is the integration of security throughout its systems, building trust and resilience from the ground up. This includes safeguarding firmware integrity with secure boot processes and bolstering data protection through hardware-based encryption acceleration.

IBM goes beyond basic protection with a proactive cybersecurity strategy. IBM offers secure storage solutions and advanced threat prevention and detection mechanisms. In the event of

[8] Pfizer creates a healthier tomorrow with IBM Power

an incident, IBM provides rapid response and recovery options to minimize downtime and effectively manage operational risks.

Privacy and confidentiality are paramount, supported by IBM advanced encryption technologies. These include pervasive encryption throughout the data lifecycle and quantum-safe cryptography, designed to guard against emerging threats such as quantum computing. Figure 1-7 shows the IBM Power security architecture and how security is integrated across all of the layers.

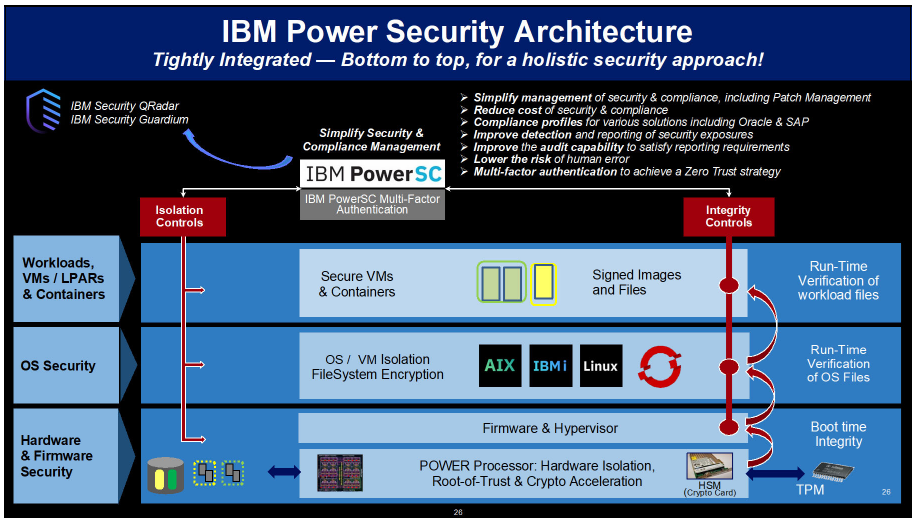


Figure 1-7 IBM Power Security Architecture

IBM simplifies regulatory compliance with continuous compliance and audit capabilities. Automated monitoring and enforcement tools ensure adherence to industry standards, while unified security management tools facilitate consistent governance across diverse IT environments.

Collaborating closely with ecosystem partners, IBM integrates security across hybrid cloud environments, networks, software systems, architectures, and chip designs. This comprehensive approach ensures holistic protection and resilience across all facets of IT infrastructure.

The IBM Power security architecture provides safe, simple security with isolation and integrity controls between each layer of components. This architecture supports the industry evolution to Software Bill of Materials (SBOMs) to allow open-source industry collaboration while ensuring we can validate the source and integrity of each component. IBM works closely with National Institute of Standards and Technology (NIST) to track and fix vulnerabilities and is working to support the evolution to Post Quantum Cryptography in the future. IBM is also working on “Fully Homomorphic Encryption” (FHE) which enables applications to work with encrypted data to further enhance security. [9]

By consolidating security insights across various domains, IBM enables informed decision-making and proactive threat management. This integrated approach dissolves traditional security silos, turning security into a catalyst for innovation and business growth.

In summary, IBM Infrastructure sets a high standard for security excellence by embedding advanced features into its solutions and equipping businesses to address both current and future cybersecurity challenges with confidence. Through collaborative efforts with ecosystem partners and a focus on regulatory compliance, IBM delivers secure, resilient, and compliant infrastructure solutions, empowering businesses to thrive in the digital age amidst evolving cyber

[9] https://research.ibm.com/topics/fully-homomorphic-encryption

threats. The IBM PowerSC architecture provides safe, simple security with isolation and integrity controls between layers and components.

1.3.4 IBM Power Virtual Server Benefits

IBM Power Virtual Server has offered a cloud-based, off-premises solution since 2019, allowing customers to dynamically provision IBM Power servers and storage with a pay-as-you-go charging model. This enables clients to quickly migrate existing IBM Power-based workloads to the cloud.

IBM Power Virtual Server Private Cloud extends this offering, providing enterprises with secure, integrated data center services at their chosen location, also on a pay-as-you-go basis. Leveraging the data center expertise gained through IBM Power Virtual Server, the IBM Power Virtual Server Private Cloud solution delivers performance, scalability, flexibility, security, and industry-leading reliability. It integrates servers, storage, network, security, and solution patterns to enable self-service capabilities in the client's data center.

Replicating these cloud capabilities and maintaining this integrated solution would require an enterprise to invest millions of dollars and months or years of development time.

Fully metered consumption

IBM Power Virtual Server Private Cloud provides fully metered consumption, allowing you to pay for the infrastructure as you use it and eliminating capital expense for the infrastructure. The terms of the offering are:

- 3-year or 5-year term with 1-year renewal
- Pay-as-you-use monthly billing with Committed Monthly Spend (CMS)
- IBM owned and managed

Lowest TCO, fastest Time to Value

Moving IBM Power workloads to the cloud can be challenging and expensive. Prior to the availability of Power Virtual Server, that migration usually meant an expensive and risky refactoring of the applications to be able to run them on x86 base hardware.

IBM Power Virtual Server Private Cloud offers the best TCO value for customers wishing to move to on-premises cloud solutions compared to our competitors as shown in Table 1-1.

Table 1-1 TCO comparison

IBM Power Virtual Server at client ► Highest security servers ► Highest reliability servers ► High performance processors ► Metered usage billing **Lowest TCO**	**AWS Outposts** ► Higher energy costs ► Higher software costs ► Lower performance processors ► Delivered capacity billing **TCO 1.5x higher**
AWS Outposts ► Higher energy costs ► Higher software costs ► Lower performance processors ► Delivered capacity billing **TCO 1.2x higher**	**AWS Outposts** ► Higher energy costs ► Higher software costs ► Lower performance processors ► Metered usage billing **TCO 1.2x higher**

IBM offers Power Virtual Server Private Cloud with no upfront capital on a pay-as-you-go consumption model with a monthly minimum fee for a 3-year or 5-year term. The configuration can be designed as a small or medium Pod depending on the client workloads.

If you are choosing whether to use Power Virtual Server versus choosing to build your own infrastructure consider the points illustrated in Figure 1-8. Beyond just the up-front capital expense for the equipment, you need to consider the costs of creating the support structure for all of the other components involved in the solution, for example storage, networking, security and monitoring.

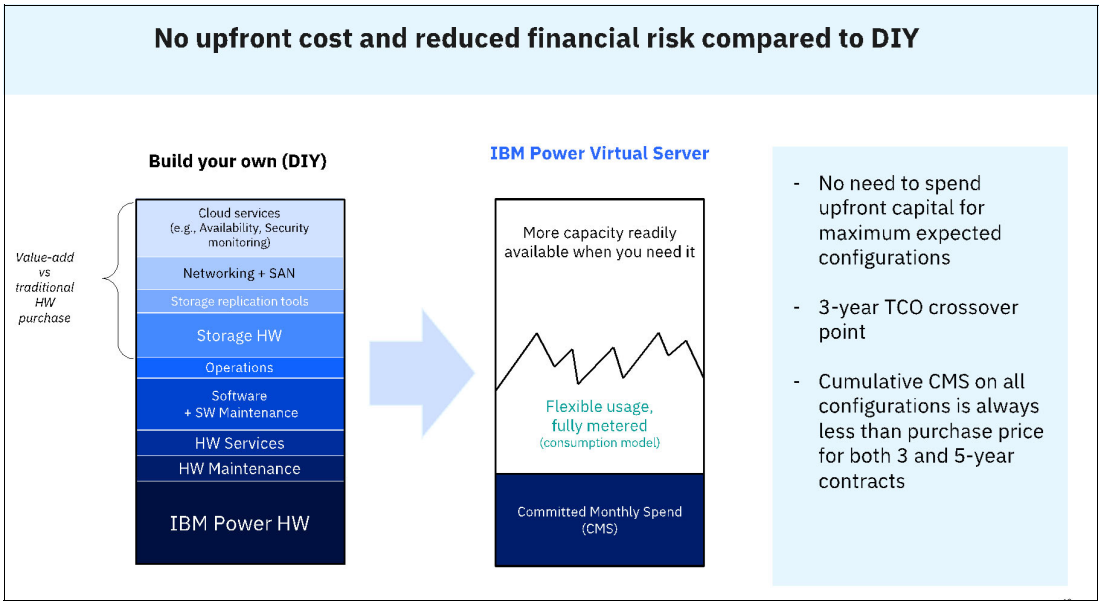


Figure 1-8 DIY comparison

Choosing IBM Power Virtual Server – either in IBM data centers or client locations – provides an excellent option to implement a Power based infrastructure with no up-front investment and with usage-based costs.

1.4 Cloud offerings for IBM Power systems

IBM has a range of cloud solutions designed to provide flexible acquisition and consumption options for IBM Power servers.

- IBM Power Virtual Server in IBM data centers

  IBM Power Virtual Server is a cloud-based offering running in IBM data centers with IBM Power servers and storage can be dynamically defined. There is no customer up-front investment, and you pay only for resources consumed.

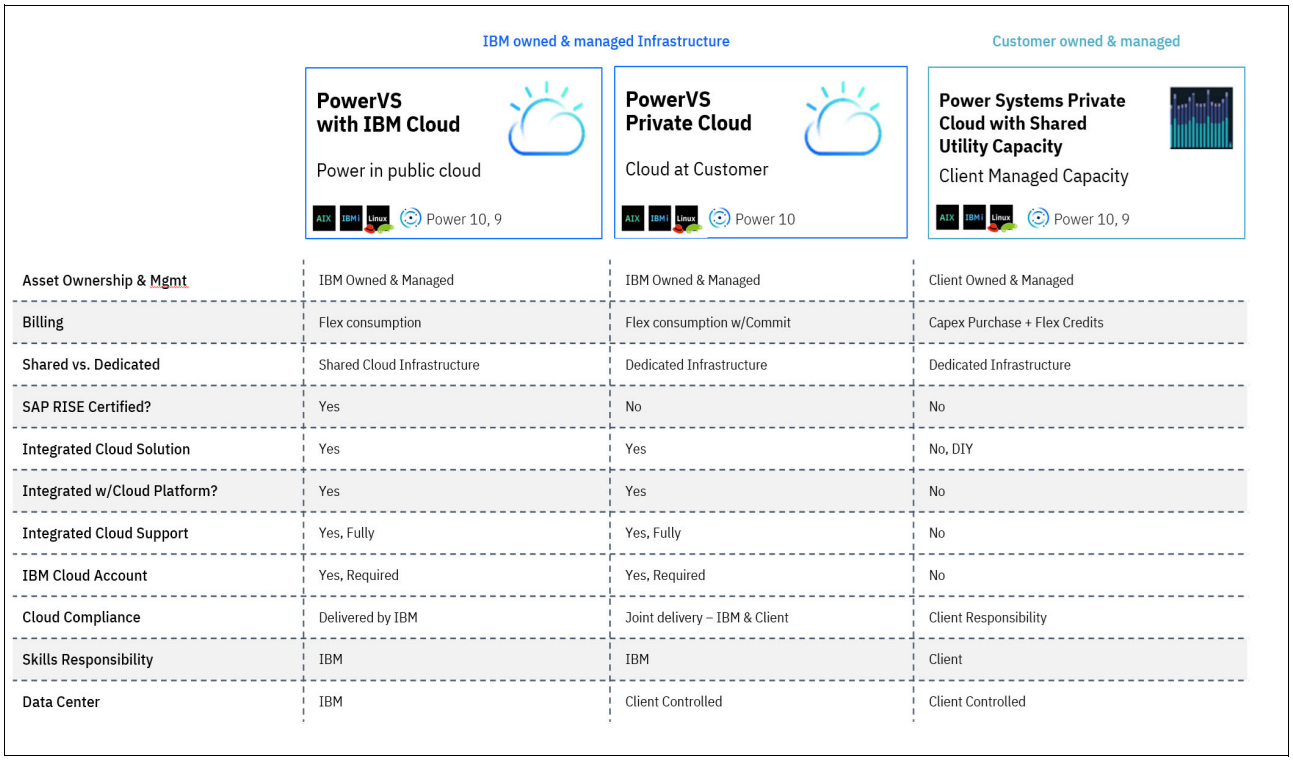
- IBM Power Virtual Server at client locations (i.e., IBM Power Virtual Server Private Cloud)

  IBM Power Virtual Server Private Cloud is an extension of IBM Power Virtual Server. In this case, the customer is responsible for providing data center space, and IBM is responsible for installing and managing the servers, storage, and networking for the customer's usage. There is no customer up-front investment, and you pay for resources consumed with a minimum monthly commitment.

- IBM Power Systems Private Cloud with Shared Utility Capacity

The Power System Private Cloud with Shared Utility Capacity (also known as Power Enterprise Pools 2 or PEP2) is a client managed offering. The Power servers are installed and managed by the client in their data center. The Shared Utility Capacity allows flexibility of paying some portion of the cost of the Power processors and memory that are installed in the servers initially and then paying for additional capacity as it is consumed. Capacity utilization is pooled across all the servers in the PEP2 pool, providing the ability to move workloads to any server based on business demand.

The three options are shown in Figure 1-9



	IBM owned & managed Infrastructure		Customer owned & managed
	PowerVS with IBM Cloud Power in public cloud AIX IBMi Linux Power 10, 9	**PowerVS Private Cloud** Cloud at Customer AIX IBMi Linux Power 10	**Power Systems Private Cloud with Shared Utility Capacity** Client Managed Capacity AIX IBMi Linux Power 10, 9
Asset Ownership & Mgmt	IBM Owned & Managed	IBM Owned & Managed	Client Owned & Managed
Billing	Flex consumption	Flex consumption w/Commit	Capex Purchase + Flex Credits
Shared vs. Dedicated	Shared Cloud Infrastructure	Dedicated Infrastructure	Dedicated Infrastructure
SAP RISE Certified?	Yes	No	No
Integrated Cloud Solution	Yes	Yes	No, DIY
Integrated w/Cloud Platform?	Yes	Yes	No
Integrated Cloud Support	Yes, Fully	Yes, Fully	No
IBM Cloud Account	Yes, Required	Yes, Required	No
Cloud Compliance	Delivered by IBM	Joint delivery – IBM & Client	Client Responsibility
Skills Responsibility	IBM	IBM	Client
Data Center	IBM	Client Controlled	Client Controlled

Figure 1-9 Comparison of IBM Power based cloud options

With these options, IBM provides a full range of cloud solutions, giving you the ability to choose where you want to deploy the infrastructure for your IBM Power based workloads:

- Off-premises, managed by IBM
- On-premises, managed by IBM
- On-premises, fully managed by you

The cost of running an IT infrastructure consists of more than just the acquisition costs and initial capital investment required and also includes the costs of operating and maintaining the infrastructure. For each option you are considering, you need to understand those additional management costs and who is responsible for the operations of each component. Figure 1-10 on page 14 illustrates who provides support for the many layers and components of your application infrastructure from the server hardware, to the data, to the applications, and including the networking and storage infrastructure.

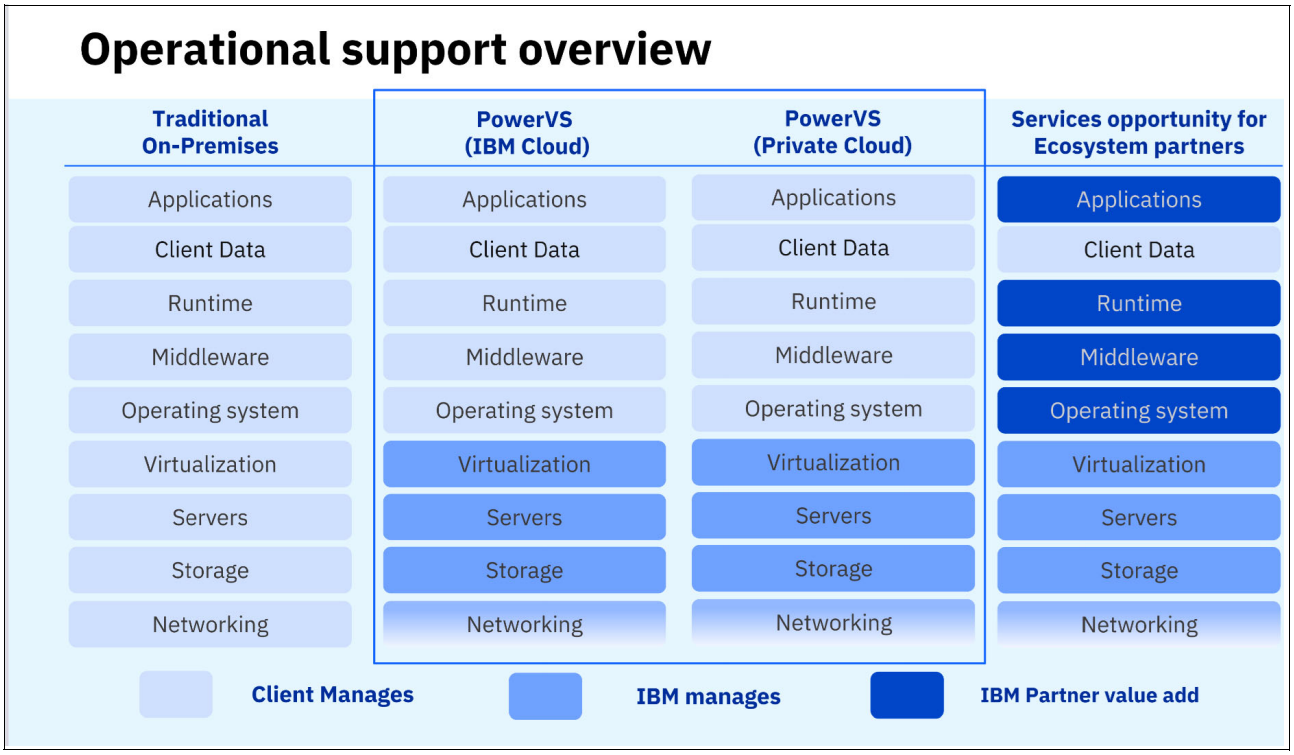


Figure 1-10 Operational support assignments across cloud implementation models

For the PEP2 option, the responsibilities are shown in the first column. The Power Virtual Server options are shown in the next two columns. The ability to have IBM manage the server layer, the virtualization layer, and the storage layer in the Power VS offerings, allows you to focus on running your business instead of managing the infrastructure.

The next sections discuss each of the options in more detail.

1.4.1 IBM Power Systems Private Cloud with Shared Utility Capacity

IBM Power Systems Private Cloud with Shared Utility Capacity – previously named Power Enterprise Pools 2.0 – provides the ability to pool resources across multiple IBM Power systems within an enterprise. The value proposition of Private Cloud with Shared Utility Capacity is the ability to move to a pay-as-you-go model, reducing the initial investment in hardware and replacing it with a metered consumption offering. This solution is significantly more flexible in comparison to other Capacity on Demand (CoD) offerings that are available on IBM Power servers.

The Shared Utility Capacity solution is managed through the use of the IBM Cloud Management Console (CMC), which is used to meter the usage across the Shared Utility Capacity pool.

As more enterprises look at ways to reduce costs while maintaining the scalability that is required to stay competitive in today's fast paced business environment, pay-as-you-go or consumption-based infrastructure models have grown increasingly popular. These purchase models enable clients to quickly scale up or down their IT infrastructure to optimize costs while quickly adapting to dynamic business environments. To maintain a competitive advantage, use an IT model that helps you adapt rapidly to the changing tech landscape.

A consumption-based IT model thrives in this environment because it has the following characteristics:

- Balances the flexibility of a cloud infrastructure with the control, security, and reliability that you expect from on-premises data centers.
- Pays for IT resources when they are consumed, which reduces the up-front capital expense and other costs that are associated with traditional procurement processes.
- Enables rapid infrastructure expansion to quickly accommodate the needs of new projects and workloads.

All these benefits contribute to a greater alignment between business and IT leaders, and when both sides of an organization are aligned, you are better prepared to deliver innovative products and services to your clients. Regardless of industry or company size, the shift to consumption-based IT is clear.

Flexible consumption offerings provide many of the attributes that clients like about public cloud in an on-premises, private cloud with better control and security.

IBM Power Systems Private Cloud with Shared Capacity provides greater flexibility and scalability within your data centers. This new end-to-end solution enables you to take advantage of cloud agility and economics while getting the same business continuity and security that you are used to from IBM Power. There are three core components that make up the IBM Power Systems Private Cloud Solution:

- Dynamic cost optimization with pay-as-you-go pricing.
- Automated, consistent enterprise IT management.
- Increased flexibility for hybrid cloud.

Pool Options

Power Enterprise Pools 2.0 provides enhanced multisystem resource sharing and by-the-minute consumption of on-premises compute resources to clients who deploy and manage a private cloud infrastructure.

There are multiple pools supported in PEP2 where servers of similar capabilities can be mixed within a pool. For example, Power E1080 and Power E980 can be in a pool, but Power E1080 and Power E1050 must be in separate pools. Currently the following types of pools are available in Power Enterprise Pools 2.0:

- Power E1080 and Power E980
- Power E1050 and Power E950
- Power S1022, Power S1024, Power S922, and Power S924

Activations and usage charges

Base Processor Activation features and corresponding software license entitlements are purchased for each server in a Power Enterprise Pool 2.0. Base Memory Activation features are purchased for each enterprise server in the pool (E1080, E1050, E980 and E950). Base Memory Activation features are not supported for the scale out servers (S1022, S1024, S922, and S924) as all installed memory is included with the purchase of these systems. However, memory usage can still be tracked.

When acquiring systems to include in the Shared Utility Capacity offering, you can purchase as little as one base processor activation and 256 GB of base memory activation (for the enterprise servers, for the scale out servers all installed memory is purchased up-front). This allows acquisition of server capacity with minimal capital expense replaced by a

pay-as-you-go model. The offering allows you to balance the initial capital investment with a metered approach to meet your specific business requirements.

All installed processors and memory on servers in a Power Enterprise Pool 2.0 are activated and made available for immediate use when a pool is started. Processor and memory usage on each server are tracked by the minute and aggregated across the pool. Any usage under the base capacity across the pool is not charged. This allows the resources to be shared across the pool without having to move them from server to server as workloads expand or decrease.

Any unpurchased capacity in the pool can be used on a pay-as-you-go basis. Resource usage that exceeds the pool's aggregated base resources is charged as metered capacity by the minute, on a real-time basis. Metered capacity can either be debited against prepurchased capacity credits or it can be billed monthly.

Processor usage is tracked based on actual consumption by partitions. Memory usage is tracked based on the assignment of memory to active partitions and is not based on Operating System (OS) usage of the memory. Both processor and memory usage are tracked and charged by the minute. Both processor and memory usage are based on the average usage for one minute and not the peak usage during the minute.

The cloud-like agility and resilience that is offered with the IBM Power Systems Private Cloud Solution on-premises can ensure the business continuity and security that you look for and always count on from IBM Power servers with new economics around the ability to pay per use, by the minute, for compute resources to optimize your costs. IBM CMC provides robust monitoring and drill-down of resource consumption in real time and historically for systems within a pool.

Summary

Shared Utility Capacity delivers enhanced multi-system resource sharing and by-the-minute consumption of on-premises compute resources for clients deploying and managing a private cloud infrastructure. This flexibility combined with IBM Power security, reliability, and scalability provides clients with the consumption-based IT solution that is required in today's dynamic, hyper-competitive market. Shared Utility Capacity simplifies system management, so clients can focus on optimizing their business results instead of moving resources and applications around within their data center. Resources are easily tracked and monitored by IBM CMC, which automatically tracks usage by the minute and debits against Capacity Credits, which are based on actual usage. With Shared Utility Capacity, you no longer need to worry about over provisioning capacity to support growth because all resources are activated on all systems in a pool. Purchased Base Activations can be seamlessly shared between systems in a pool, and all unpurchased capacity can be used on a pay-per-use basis.

1.4.2 IBM Power Virtual Server

IBM Power Virtual Server is a cloud-based infrastructure service that allows businesses to run workloads on IBM Power Systems hardware. It combines the performance and reliability of IBM's Power architecture with the flexibility and scalability of a virtualized environment. Key features include:

- Flexibility: Users can scale resources up or down based on demand, making it suitable for various workloads, including enterprise applications and databases.
- Performance: Leveraging IBM Power processors, the service is designed to be for high performance computing tasks.
- Hybrid Cloud Integration: It supports seamless integration with on-premises systems and other cloud services, facilitating a hybrid cloud approach.

- Security: Built-in security features help protect sensitive data and comply with industry standards.
- Managed Services: IBM provides management options, allowing users to focus on their applications rather than infrastructure.

IBM Power Virtual Server is targeted at organizations looking for robust, enterprise grade cloud solutions that can support demanding applications. IBM Power Virtual Server resources reside in IBM data centers with dedicated networking and storage area network attached Fibre Channel storage. You can choose one of the regions that is listed in the specifications that are nearest to your data center. There are currently 21 data centers located around the world hosting Power Virtual Server. IBM Power clients who rely on private cloud infrastructure can now quickly and economically extend their Power IT resources on the cloud.

In the data centers, the Power Virtual Servers are separated from the rest of the IBM Cloud servers with separate networks and direct attached storage. The internal networks are fenced but offer connectivity options to IBM Cloud infrastructure or private cloud environments. This infrastructure design enables Power Virtual Server to maintain key enterprise software certification and support as the Power Virtual Server architecture is identical to certified private cloud infrastructure.

Power Virtual Server is an infrastructure as a service offering (IaaS) where there are no up-front costs for deploying resources and resources are paid for based on usage. Power Virtual Server uses a monthly billing rate that includes the licenses for the AIX, IBM i, or Linux operating systems. The monthly billing rate is prorated by the hour based on the resources that are deployed to the Power Virtual Server instance for the month. When you create the Power Virtual Server instance, you can see the total cost for your configuration based on the options that you specify. You can quickly identify what configuration options provide you with the best value for your business needs. There is also an option for a “bring your own Linux image”. In this case the customer is responsible for acquiring the subscription and paying the Linux distributor for licenses used.

You can configure and customize the following options when you create a Power Virtual Server:

- Number of virtual server instances
- Number of cores
- Amount of memory
- Data volume size and type
- Network interfaces

Configuration, billing and management are all handled by IBM Cloud through a cloud based graphical user interface and through defined APIs.

1.4.3 IBM Power Virtual Server Private Cloud

IBM Power Virtual Server private cloud is an extension of IBM Power Virtual Server where the resources utilized in your cloud environment are located on your data center floor instead of using resources in the IBM Power Virtual Server data centers.

This offering is designed to provide a private cloud option for clients that need to meet specific regulations about data privacy and data sovereignty. The equipment that is installed in the client data center is the same set of Power servers, SAN attached storage devices, and networking equipment that is installed in the Power Virtual Server public cloud and is managed using the same Power Virtual Server management tools.

This is an IaaS solution with no up-front capital expense and the on-premises equipment is owned and managed by IBM. Pricing for the offering is usage based like the public cloud version of Power Virtual Server, with a minimum monthly spending commitment.

The client works closely with the IBM team in selecting the appropriate configuration from a list of supported options and then that configuration is assembled and delivered to the client data center. The client is responsible for providing the necessary floor space, power, heat and air movement, and networking connections from the client data center to one of the IBM data centers that support connectivity to a Power Virtual Server Private Cloud location.

Once the Power Virtual Server Point of Delivery (Pod) is delivered to the client data center, IBM service representatives install the Pod and connect it to the IBM cloud. At that point the customer manages the resources using the same management interface used for any Power Virtual Server implementation.

1.4.4 Choosing your cloud option

When businesses decide to move workloads to the cloud, selecting the right cloud environment is a critical decision. It involves balancing factors such as cost, security, performance, compliance, and scalability. IBM Power cloud solutions offer several options, including **IBM Power Systems Private Cloud with Shared Utility Capacity**, **Power Virtual Server**, and **Power Virtual Server Private Cloud**. Each of these options provides unique benefits suited to different business needs, ranging from hybrid cloud flexibility to on-premises control. Table 1-2 provides a comparison of the cloud options by feature. This is a starting point for making your decision.

Table 1-2 Feature comparison of the cloud options

Feature	IBM Power Systems Private Cloud with Shared Utility Capacity	Power Virtual Server	Power Virtual Server Private Cloud
Cost Model	Pay-as-you-go, no upfront cost	Pay-as-you-go, metered hourly	Pay-as-you-go, metered hourly
Security	Highest level of control and security	Enterprise-grade security, encryption	Full control with IBM management
Compliance	Full compliance for sensitive data	Complies with major standards	Fully compliant for in-country hosting
Performance	High-performance, low latency	Elastic, global cloud infrastructure	Low latency, high performance
Hybrid Cloud Integration	Seamless with IBM Cloud and others	Public cloud hybrid capabilities	Hybrid cloud with on-premises control
Use Cases	Mission-critical, on-premises apps	Scalable, flexible cloud workloads	Hybrid cloud with strict data control

As an example, a financial services company with highly sensitive customer data and compliance needs may choose Power Virtual Server Private Cloud for their core banking systems. This option allows them to host critical applications on private infrastructure within their own data center, ensuring compliance with data sovereignty regulations, while using Power Virtual Server for less sensitive workloads such as development and testing.

Factors to consider when choosing your cloud option

There are many factors involved in choosing which cloud option fits your requirements. Each option has benefits across each of the factors and depending on which factors are most important for the specific workloads you are considering, these factors will have different weights. Consider the following factors:

- Cost Efficiency
  - IBM Power Systems Private Cloud with Shared Utility Capacity

    For businesses that want the flexibility of a private cloud without large upfront costs, IBM Power Systems Private Cloud with Shared Utility Capacity offers a pay-as-you-go model. This makes it ideal for businesses looking to reduce capital expenditure (CapEx) while still enjoying cloud flexibility. The OPEX model also allows businesses to align costs with usage, making it easier to scale resources as needed.
  - Power Virtual Server

    For businesses that need scalability and flexibility, Power Virtual Server offers an elastic compute infrastructure, where businesses can scale up or down based on demand. It is suited for workloads that experience fluctuations, such as seasonal applications or testing environments.
  - Power Virtual Server Private Cloud

    This option provides the same pay-as-you-go flexibility as Power Virtual Server but within a fully private, on-premises environment. This is ideal for businesses that need to maintain physical control over their infrastructure but want the benefits of cloud billing models.
- Security and Compliance
  - IBM Power Systems Private Cloud with Shared Utility Capacity

    This solution is designed for businesses that need complete control over their environment but still want to benefit from cloud technologies. Since the infrastructure is hosted within the business's own data center or collocation facility, it provides the highest level of security and control, ensuring compliance with strict regulatory standards.
  - Power Virtual Server Private Cloud

    For industries that handle sensitive data (e.g., finance, healthcare), Power Virtual Server Private Cloud offers cloud services that meet stringent security and compliance requirements. The private infrastructure is fully managed by IBM but hosted on the client's premises, giving businesses full control over data locality and privacy.
  - Power Virtual Server

    While offering the flexibility of a public cloud, Power Virtual Server includes enterprise-grade security measures such as encryption, identity management, and compliance certifications, making it a good option for businesses needing to comply with global regulatory standards.
- Performance and Latency:
  - IBM Power Systems Private Cloud with Shared Utility Capacity

    For businesses that require low latency and high-performance workloads, especially those running mission-critical applications or AI workloads, IBM Power Systems Private Cloud with Shared Utility Capacity ensures that applications run on dedicated infrastructure with consistent performance.

- Power Virtual Server

  Power Virtual Server offers elastic computing resources spread across multiple global regions. This can be beneficial for workloads where performance is less critical but global reach and availability are important. However, businesses with strict latency requirements should evaluate the proximity of their users or systems to IBM Cloud regions to ensure optimal performance.

- Power Virtual Server Private Cloud

  By offering cloud flexibility in an on-premises setup, Power Virtual Server Private Cloud minimizes latency for locally critical workloads while providing the scalability and flexibility of a cloud solution. This is particularly important for businesses that cannot tolerate delays due to remote connections but still want cloud-like capabilities.

- Hybrid Cloud Capabilities
  - IBM Power Systems Private Cloud with Shared Utility Capacity

    Designed for businesses looking for hybrid cloud solutions, IBM Power Systems Private Cloud with Shared Utility Capacity allows seamless integration between on-premises and public cloud environments. Businesses can extend workloads to IBM Cloud or other public clouds when needed, while maintaining core applications and data on private infrastructure.

  - Power Virtual Server

    Power Virtual Server is a fully public cloud solution that can integrate with existing on-premises environments to create a hybrid cloud. It is well-suited for businesses that need a flexible approach to scaling out workloads to the cloud while maintaining some control over core infrastructure.

  - Power Virtual Server Private Cloud

    This option provides the best of both worlds: cloud features within a private environment. It is ideal for businesses with hybrid cloud strategies that want to keep sensitive workloads on-premises while extending less sensitive workloads to public cloud environments.

- Workload-Specific Considerations
  - Development and Testing

    If the cloud is primarily required for development and testing, then Power Virtual Server offers the fastest provisioning, flexible scaling, and cost-efficiency. Development teams can spin up and tear down virtual servers as needed, making this option ideal for agile testing cycles.

  - Mission-Critical Workloads

    For production and mission-critical workloads that require high performance, low latency, and disaster recovery features, IBM Power Systems Private Cloud with Shared Utility Capacity or Power Virtual Server Private Cloud provide the security, performance, and resilience necessary to maintain continuous operation.

Table 1-3 provides a summary of the factors listed above and the strengths of each of the options.

Table 1-3 Summary of factors to consider in choosing your cloud option

	IBM Power Systems Private Cloud with Shared Utility Capacity	Power Virtual Server	Power Virtual Server Private Cloud
Cost efficiency	► Pay-as-you-go model ► Reduced Capex ► OPEX model	► Pay-as-you-go model ► Elastic compute ► Scale up/down ► Flexible resources	► On-premises flexibility ► Pay-as-you-go model ► Private infrastructure
Security and compliance	► Complete control ► On-premises or collocation ► Highest security	► Enterprise grade security ► Encryption ► Compliance certification	► Sensitive data handling ► On-premises control ► Full data locality
Performance	► Low latency ► High performance ► Dedicated infrastructure	► Global reach ► Multiple regions ► Elastic resources	► On-premises performance ► Low latency ► Cloud-like scalability
Hybrid cloud	► Seamless integration ► Extend to public cloud ► Core apps on-premises	► Public cloud solution ► Integrate with on-premises ► Flexible scaling	► Best of both worlds ► Sensitive workloads on-premises ► Extend to public cloud
Workload specific considerations: ► Development and testing consider Power Virtual Server for fast provisioning and flexible scaling. ► Mission critical workloads consider Power Private Cloud or Power Virtual Server Private Cloud for high performance, low latency and security.			

Choosing the right cloud option for IBM Power workloads depends on the specific requirements of the business. Factors such as security, compliance, performance, and cost must all be carefully considered. IBM offers a range of cloud solutions that cater to different needs, from fully private clouds to highly flexible public cloud options. Businesses should evaluate their workloads and operational requirements to make an informed decision about which cloud model best supports their strategic goals.

1.5 Government regulations and compliance

Government regulations and data protection laws play a pivotal role in how organizations approach cloud adoption and management. These rules, particularly in industries like healthcare, finance, and government, determine where and how sensitive data must be stored, processed, and transmitted. Understanding and adhering to these regulations is crucial when selecting a cloud provider.

IBM Power Virtual Server is designed to provide you with a solution that enables you to meet regulatory requirements for data processing and has already earned several certifications for

differing regulations. such as General Data Protection Regulation (GDPR), System and Organization Controls (SOC), and PCI-DSS. IBM continues to conduct regular compliance audits to ensure continuous compliance with legal and regulatory standards. By taking advantage of the on-premises Power Virtual Server Private Cloud, you can meet additional regulatory standards – specifically in data control and data sovereignty. Figure 1-11 shows how Power Virtual Server can help with your compliance requirements.

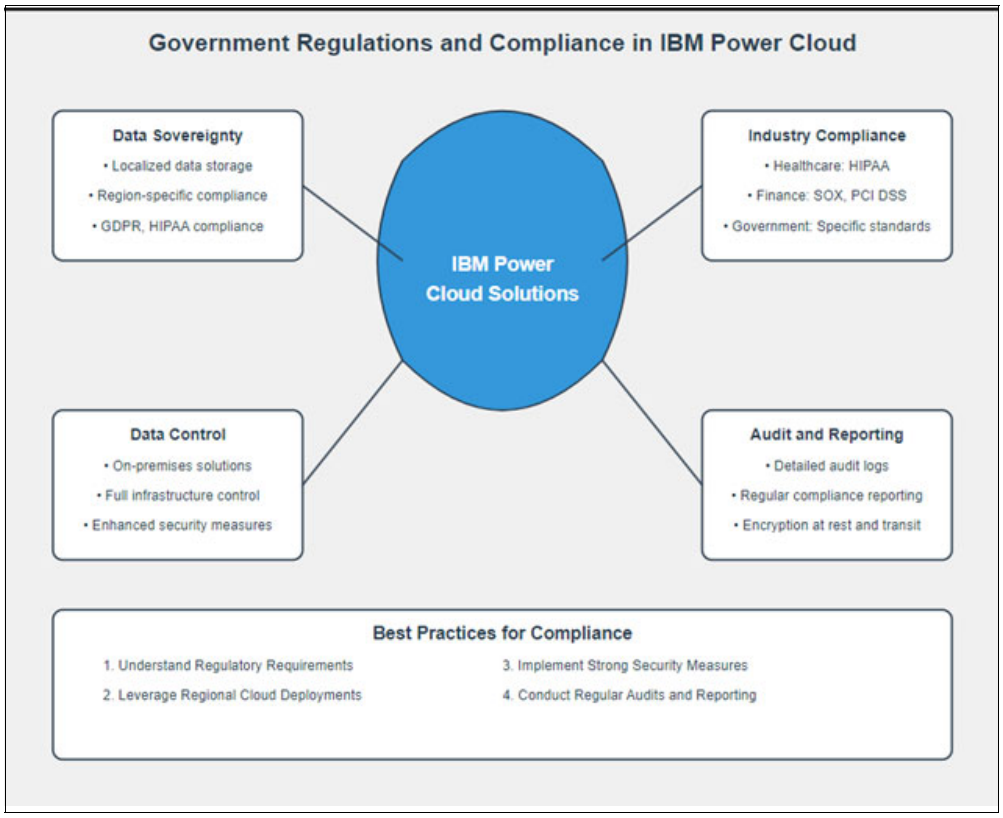


Figure 1-11 Government regulations and compliance

Monitoring and compliance tools

Power Virtual Server provides several key monitoring, security, and auditing tools. These tools allow organizations to track system performance, security, and data access, ensuring that any compliance gaps or risks are identified and addressed promptly. Figure 1-12 shows the IBM compliance and audit tools found in the IBM Cloud.

Figure 1-12 IBM Cloud compliance tools for Power Virtual Server

Below are the essential tools available within IBM Power Virtual Server Private Cloud to assist in compliance and monitoring:

- IBM Cloud Monitoring

  Provides real-time insights into the health, performance, and availability of cloud resources.

  How it Helps:

  - Tracks system performance, resource utilization, and network metrics.
  - Monitors the status of virtual machines, applications, and infrastructure components.
  - Allows businesses to set up alerts based on custom thresholds, helping identify compliance violations or performance degradation.

  Ensures continuous operational compliance by monitoring the system and identifying issues before they impact service availability.

- IBM Cloud Security and Compliance Center

  A centralized platform that monitors and manages compliance with industry standards and regulations.

  How it Helps:

  - Automates compliance checks by comparing infrastructure settings and configurations against regulatory standards like HIPAA, PCI DSS, and GDPR.
  - Provides compliance dashboards and generates audit reports to ensure that security configurations meet legal requirements.
  - Helps detect misconfiguration, vulnerabilities, and potential risks that could lead to non-compliance.

  Regularly audits the system's compliance status and provides actionable insights to maintain regulatory adherence.

- IBM QRadar® (SIEM - Security Information and Event Management)

  A comprehensive security monitoring tool that detects and responds to security incidents across the entire environment.

  How it Helps:

  - Collects and analyzes log data from various sources, including network traffic, user activity, and system events.
  - Identifies potential security breaches, unauthorized access attempts, and other suspicious activities in real-time.
  - Provides audit trails and incident reports to support security audits and investigations.

  Helps organizations meet compliance standards by detecting and reporting security violations and incidents in real-time, ensuring prompt remediation.

- IBM Spectrum Protect

  A data protection and backup tool that ensures business continuity and compliance with data retention policies.

  How it Helps:

  - Automates backup processes and ensures that data is securely stored in line with regulatory requirements.
  - Provides detailed reports on data backups, recovery points, and retention periods to support audit processes.

- Offers data encryption both at rest and during transmission, ensuring compliance with standards that mandate data protection.

Supports compliance with data protection laws by ensuring that backup and recovery policies align with retention and encryption standards (e.g., GDPR, HIPAA).

- IBM Guardium®

  A security solution focused on data protection and database activity monitoring.

  How it Helps:

  - Monitors access to sensitive data and logs all database activity, including user access, changes, and queries.
  - Provides detailed audit trails of database transactions, ensuring that sensitive data is accessed and modified in accordance with security policies.
  - Enforces data access controls and generates reports for compliance audits.

  Compliance Role: Ensures compliance with data protection regulations like PCI DSS and SOX, which mandate strict controls over access to sensitive data.

- IBM Log Analysis

  Provides centralized logging for tracking and analyzing system and application logs across Power Virtual Server environments.

  How it Helps:

  - Collects logs from various cloud components and aggregates them for detailed analysis.
  - Identifies anomalies, errors, and security events based on logs, providing a clear audit trail.
  - Generates compliance reports and ensures that logs are stored securely for later retrieval.

  Supports compliance efforts by maintaining detailed log records for auditing purposes, ensuring traceability and accountability for system activities.

- IBM PowerVC (Power Virtualization Center)

  A cloud management tool for managing and monitoring virtualized Power systems.

  How it Helps:

  - Manages resource allocation, provisioning, and monitoring of virtual machines.
  - Provides real-time monitoring of cloud performance and infrastructure status.
  - Supports compliance by offering visibility into infrastructure configurations, ensuring alignment with regulatory requirements.

  Helps monitor compliance by tracking system configurations, resource allocations, and changes within the virtualized environment.

Table 1-4 provides a summary of the tools previously discussed.

Table 1-4 Compliance and monitoring tools summary

Tool	Primary Use	Compliance Role
IBM Cloud Monitoring	Monitors real-time performance and health of cloud resources	Ensures continuous operational compliance by detecting performance and security issues.

Tool	Primary Use	Compliance Role
IBM Cloud Security and Compliance Center	Manages security policies and checks against compliance standards	Automates compliance checks and generates regulatory reports for auditing purposes.
IBM QRadar (SIEM)	Analyzes security events and detects potential incidents	Provides audit trails for security violations and helps respond to incidents in real-time.
IBM Spectrum® Protect	Automates backup and data protection	Ensures compliance with data retention policies and encryption standards.
IBM Guardium	Monitors database activity and protects sensitive data	Supports compliance by logging database access and ensuring data security standards are met.
IBM Log Analysis	Centralized logging and analysis	Provides audit trails and logs for compliance auditing and investigation purposes.
IBM PowerVC	Manages virtualized Power systems	Monitors resource usage and system configurations, ensuring infrastructure meets compliance standards.

Case Study: Multinational Compliance in Action

A global pharmaceutical company adopting a hybrid cloud model must comply with GDPR in Europe, HIPAA in the U.S., and local data laws in Asia. By using Power Virtual Server Private Cloud in each region, the company ensures that sensitive healthcare data remains compliant with regional regulations, while benefiting from the cloud's scalability and flexibility.

Summary

These tools ensure that businesses using Power Virtual Server Private Cloud can automate compliance checks, generate audit reports, and continuously monitor their environment to maintain compliance with industry regulations. Regular use of these tools enables businesses to proactively address compliance challenges and maintain robust security.

Navigating government regulations is a critical aspect of cloud adoption. By using IBM Power Virtual Server Private Cloud, organizations can ensure compliance with data sovereignty, security, and industry-specific regulations, all while taking advantage of the cloud's flexibility. Whether managing healthcare, financial, or government data, Power Virtual Server offers a comprehensive solution to meet regulatory needs.

1.5.1 Black or disconnected sites

For an enterprise or organization which has locations with security requirements that do not allow any outside network connectivity, then those locations are not a good fit for the Power Virtual Server Private Cloud offering. The IBM Power Virtual Server Private Cloud requires a management network connection to the IBM Cloud Management Console (CMC) for remote provisioning and management of the infrastructure. The CMC does not enable access to any client data or applications.

1.6 Use cases

There are many ways to get started on moving workloads to the cloud. The following sections provide some use cases for workloads that are good choices to implement in a hybrid cloud environment.

1.6.1 Burst to cloud

Burst to cloud refers to the capability of extending an organization's on-premises infrastructure by temporarily utilizing cloud resources during periods of high demand. This hybrid cloud strategy allows businesses to “burst” workloads into the cloud when local resources are insufficient, providing additional computing power, storage, and other resources without the need for permanent hardware investments. IBM Power Systems, especially when used with IBM Power Virtual Server, enables businesses to burst workloads to IBM Cloud seamlessly, offering both flexibility and cost savings.

Key features of burst to cloud

By leveraging burst to cloud, organizations can optimize their IT infrastructure, reduce costs, and improve their ability to handle unpredictable workloads. Figure 1-13 provides an illustration of burst to cloud.

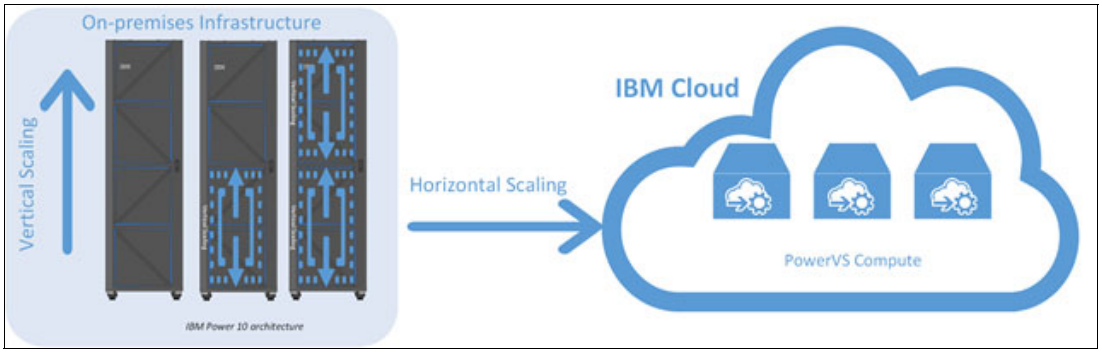


Figure 1-13 Burst to cloud

1. Dynamic Resource Expansion

   When the demand on an organization's on-premises infrastructure exceeds capacity, workloads can be automatically redirected to cloud environments to maintain performance. This is ideal for businesses experiencing seasonal traffic spikes, sudden increases in customer demand, or data processing bursts.

   Example: A retailer could use burst-to-cloud during Black Friday or holiday shopping seasons to accommodate spikes in web traffic.

2. Cost Efficiency

   Rather than investing in additional on-premises servers that may only be needed occasionally, businesses can leverage pay-as-you-go cloud resources. This avoids over-provisioning and helps manage costs more efficiently.

   Example: A media company streaming a large live event can temporarily use cloud resources to handle the spike in viewer traffic without purchasing new hardware.

3. Hybrid Cloud Integration

   IBM Power Virtual Server allows for seamless integration between on-premises environments and IBM Cloud. Workloads can be shifted between the two environments with minimal disruption, giving organizations flexibility in where they run their applications.

   This model is perfect for companies that require a balance between data residency and cloud scalability.

4. Disaster Recovery and Flexibility

   Cloud bursting can also be used as part of a disaster recovery strategy, where critical workloads can shift to the cloud in case of on-premises failures. This ensures business continuity without investing heavily in secondary infrastructure.

Implementing burst to cloud

Implementing a burst to cloud requires some planning and preparation. To be successful you need to implement a hybrid cloud architecture that integrates your private infrastructure and public clouds. You need to be able to utilize load balancing techniques to distribute workloads between on-premises and cloud resources.

When deciding to use a burst to cloud strategy, be sure to consider data transfer costs and latency when workloads are moved between environments and ensure that your cloud provider meets your security and compliance requirements.

Here are some additional considerations:

1. Monitor Usage Patterns

   Regularly monitor resource usage to predict when cloud bursting might be necessary. Automating the scaling process based on performance metrics can help avoid bottlenecks and downtime.

2. Optimize Workload Placement

   Identify which workloads are suitable for bursting to the cloud. Applications with variable demand or batch processing tasks are ideal candidates for cloud bursting.

3. Leverage Auto-Scaling Tools

   Use IBM Cloud Auto-Scaling to automatically adjust resources based on preset thresholds, ensuring workloads remain responsive while minimizing unnecessary cloud usage.

   Example Use Case: A financial services firm experiences fluctuating demand for its data analysis services. During normal operation, its on-premises Power Systems handle the workload. However, at quarter-end, when demand for financial reports surges, the firm bursts workloads into IBM Cloud. This strategy allows the firm to avoid hardware over-provisioning while ensuring timely report generation.

For more information see the following documents:

- IBM Power Systems for Hybrid Cloud
- Power Virtual Server Burst to Cloud Guide
- IBM Auto-Scaling Overview

1.6.2 AI on Power

IBM Power is designed for AI and advanced workloads, positioning enterprises to inference and deploy AI algorithms on sensitive data and transactions that reside on Power systems. More specifically, IBM Power Virtual Server can help enterprises by providing a flexible, scalable, and secure platform to run mission-critical workloads, including AI, which extends on-premises environments to the cloud.

At the same time, clients want to be agile and drive stronger business outcomes with simpler experiences. This includes ramping up quickly, spending less time maintaining infrastructure, and paying as they go. With a flexible as a service offering, we're helping clients with frictionless workload migration and modernization from on-premises to the cloud. This is possible because of architectural parity between IBM Power and IBM Cloud that helps our clients achieve growth while adopting innovative AI and hybrid cloud technologies. Power Virtual Servers delivers a virtual machine as a service on Power architecture, engineered to run your mission critical applications. With the same architecture on premises and on cloud, workload migration does not require re-platforming to move it to other clouds.

We are continuing to make improvements to our IBM Power Virtual Server offering on IBM Cloud, including enhancing the user experience and capabilities and expanding the infrastructure itself, so clients can focus on business outcomes.

IBM Power clients have valuable data residing on their servers, giving them an opportunity with enterprise AI to derive insights from this data to help address business challenges such as providing a more seamless and secure customer experience. IBM Power10 has on-chip acceleration for production-ready AI at the point of data to enable faster time to insight and lower latency.

Technology services provider OpenTech_utilizes *OpenXAI* platform, a private document sourced knowledge-based generative AI chat application that generates content for users based on their own private data. Choosing the right infrastructure to run *OpenXAI* was critical for OpenTech as the company's chat application was being trained on sensitive and confidential documentation. Furthermore, OpenTech needed to stay on-premises within the country to comply with local regulations regarding data privacy and sovereignty.

IBM and OpenTech joint clients use IBM Power10 processor-based servers to run generative AI chatbots that leverage private documentation to keep confidential data secure.

"Putting our OpenXAI platform on Power10 gives us the best of both worlds. Our open-source technology stack makes development and expansion easy to handle as we add new features constantly," said Moatasim Al-Masri, CEO, OpenTech. "In addition, clustering multiple Power10 servers together with IBM Storage solutions gives our clients a unique and robust generative AI solution that takes the worry out of having to find hardware to match the software. This solution is powerful for today and into the future and can be a closed on-premises private solution or a private corporate cloud solution hosted on-prem, keeping confidential information secure to meet data privacy and sovereignty requirements."[10]

Statement of direction

IBM intends to incorporate the IBM Spyre accelerator in future Power offerings to provide additional AI compute capabilities. Working together, IBM Power processors and IBM Spyre accelerator will enable the next generation infrastructure to scale demanding AI workloads for businesses.[11]

[10] Delivering an On-Prem Generative AI Chatbot with IBM Power

[11] https://www.ibm.com/docs/en/announcements/statement-direction-spyre-accelerator-power-platform

1.6.3 Database support

IBM Power is an excellent platform for databases due to its high availability design, its performance and scalability, and its integrated security capabilities. The data managed by the database is the life blood of an enterprise's business and if it is unavailable for any reason, business operations are crippled. This can lead to a financial loss due to lost business and also can lead to a loss of reputation as customers are unable to interface with the enterprise.

IBM Power is the choice to host databases for many enterprises and IBM works closely with database vendors to ensure that our customers can get the best value when running their databases on IBM Power.

As IBM Power Virtual Server is based on the same IBM Power architecture as is run on-premises, database vendors do not need to specifically provide support for databases on Power, if it runs on-premises, then it will run on IBM Power Virtual Server. This section focuses on database support for IBM Power Virtual Server including IBM Db2, Oracle Database, and other databases that run on IBM Power.

IBM Db2

IBM Db2 is a relational database product that allows users to store, manage, and retrieve data in a structured format. Db2 offers features like data security, scalability, and support for various data types. It can also handle complex queries and transactions.

IBM Db2 is the cloud-native database built to power low–latency transactions and real-time analytics at scale, providing customers with both self-managed and fully managed (SaaS) options on the cloud. It provides a single engine for DBAs, enterprise architects and developers to keep critical applications running. It also stores and queries anything and powers faster decision-making across organizations.

For the past three decades, Db2 has provided stability and dependability for customers' data management solutions. Its robust architecture and proven performance have given businesses uninterrupted access to critical data while powering their enterprise-level applications. Db2 has been recognized as a leading data management product as shown in Figure 1-14.

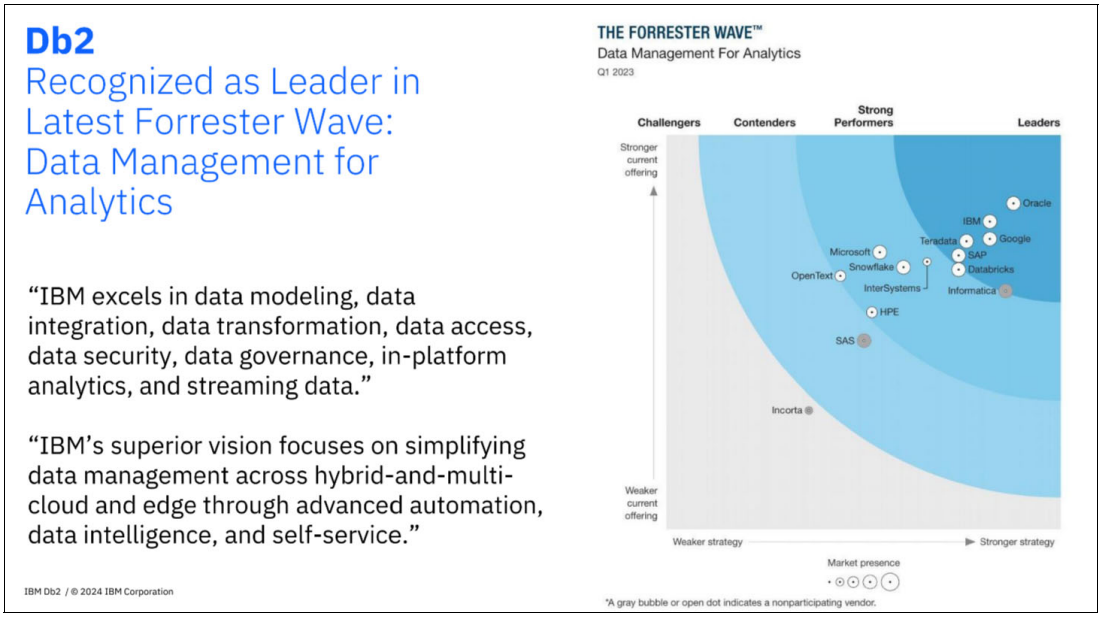


Figure 1-14 Db2 recognized as a leader

IBM Db2 first started as a mainframe database over 40 years ago. After its success on that platform, it was ported to run on Linux, Unix (for example AIX), and Windows hence the name Db2 LUW. Since then, it has grown into a highly utilized cloud-ready database engine for use both on-premises and in the cloud. IBM Db2 is integrated into IBM i as an integral component in the IBM i ecosystem.

Many customers run Db2 on IBM Power, running in AIX or Linux, and do all IBM i users. It is known for its integrated features for performance and data replication, serving data for business-critical applications across many industries.

IBM Power Virtual Server provides an excellent infrastructure and environment to relocate your Power on AIX assets to high performance virtual computing infrastructure. With IBM Db2 on Power Virtual Server you get the option to leverage Db2 backup/restore procedures and Db2 HADR database synchronization commands. You can create an off-premises backup of your Db2 database for disaster recovery or fully migrate your Db2 workload from on-premises to a Power Virtual Server environment in a safe and secure IBM datacenter, quickly and easily using common Db2 tools.

Power Virtual Server can offer significant price performance advantages over comparable x86-based platforms. Consider the following:

- IBM Power9®-based Power Virtual Server processors, available now, offer competitive price performance.
- Since Db2 LUW licensing is per core based, the per core performance advantage of Power Virtual Server over x86-based compute provides significant TCO advantages.
- Power10 processors, which are becoming generally available in more datacenters, offer > 2x per core performance advantage Power9.

In addition, modern AIX on Power Virtual Server infrastructure can provide significant operational benefits. For example, an IBM Aspera® client/server infrastructure can be installed to facilitate extremely fast data transfers across networks.

The future of IBM Db2® is bright, and performance, modernization and TCO benefits can be realized by migrating Db2 on AIX from Power servers to IBM Power Virtual Server.

Oracle

Clients worldwide, across multiple industries, are finding an edge over competitors with Oracle workloads running on IBM Power. Running on IBM Power provides better performance than x86 based options and takes advantage of the high availability of the IBM Power platform. The ability to run more workload per core on IBM Power also allows those clients to reduce licensing costs, providing a high performance and cost-effective choice for running business critical workloads based on Oracle. Figure 1-15 on page 31 lists five key motivations for running your Oracle workloads on IBM Power.

Utilizing IBM Power Virtual Server, clients can build a hybrid cloud environment to run Oracle on an IBM Power stack – either in IM data centers or client locations – identical to their existing on-premises Power-based infrastructure for Oracle workloads. Running Oracle on Power Virtual Server Private Cloud allows clients to leverage the 1:20 entitled core-to-virtual processor ratio (just like traditional on-premises implementations) to maximize Oracle licensing TCO. Power Virtual Server uses a fully Oracle certified stack with PowerVM, AIX, IBM i, SAN based storage, and network adapters which enables full Oracle support. IBM and Oracle work closely together to identify and fix any issues that might arise.

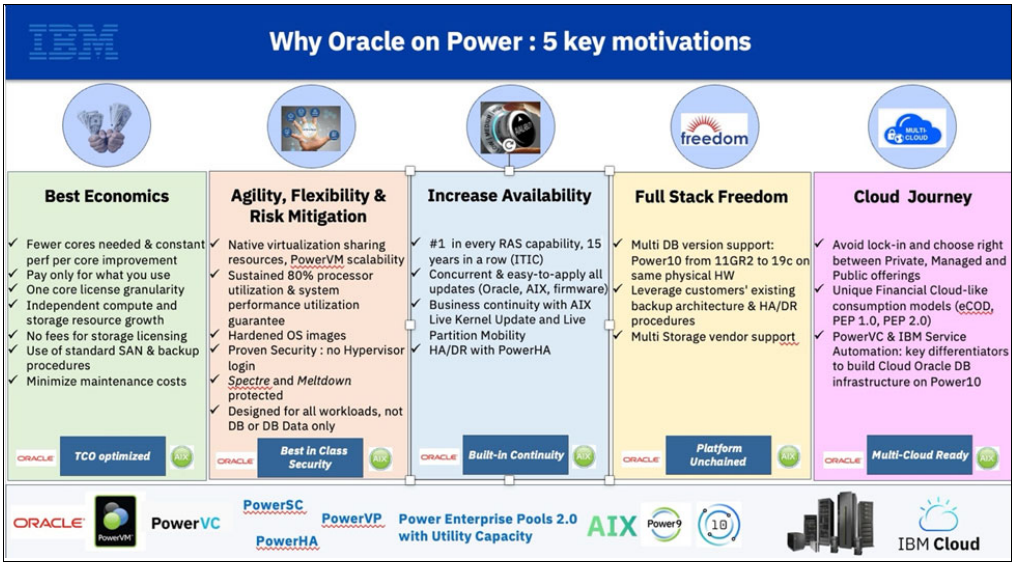


Figure 1-15 Five motivations for running Oracle on IBM Power

Importantly, running on IBM Power Virtual Server meets Oracle's hard partitioning requirements which can reduce the number of licenses required for your solution. This provides:

- Sub-capacity licensing with the ability to license only the CPUs available to Oracle Database and middleware. This allows a company to optimize Oracle licensing to provide a reduced TCO, lower than any hyperscale cloud environment.
- Full support for Oracle Database, RAC, Fusion Middleware, and Applications. Any IBM Power supported Oracle solutions are supported on Power Virtual Server.
- Highest off-premises scalable Oracle DB Single Instance capability.
- Superior storage resiliency with an enterprise class platform with all flash NVMe enabled SAN storage.

Figure 1-16 summarizes the reasons that clients choose IBM Power to run their Oracle business workloads.

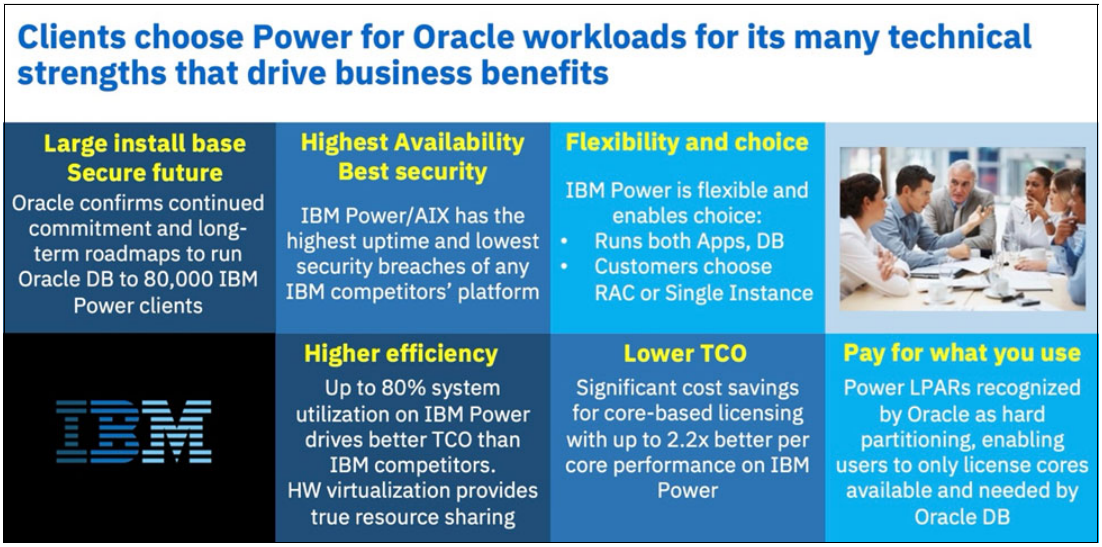


Figure 1-16 Summary of why clients choose IBM Power for Oracle workloads

Other databases

There are many other databases, either SQL based or noSQL available on IBM Power. As such, they can also be easily installed in your IBM Power Virtual Server servers. Running these databases on Power Virtual Server can provide a significant TCO advantage compared to running them on x86 servers while providing enterprise level support.

Figure 1-17 shows a number of databases that are supported on IBM Power.

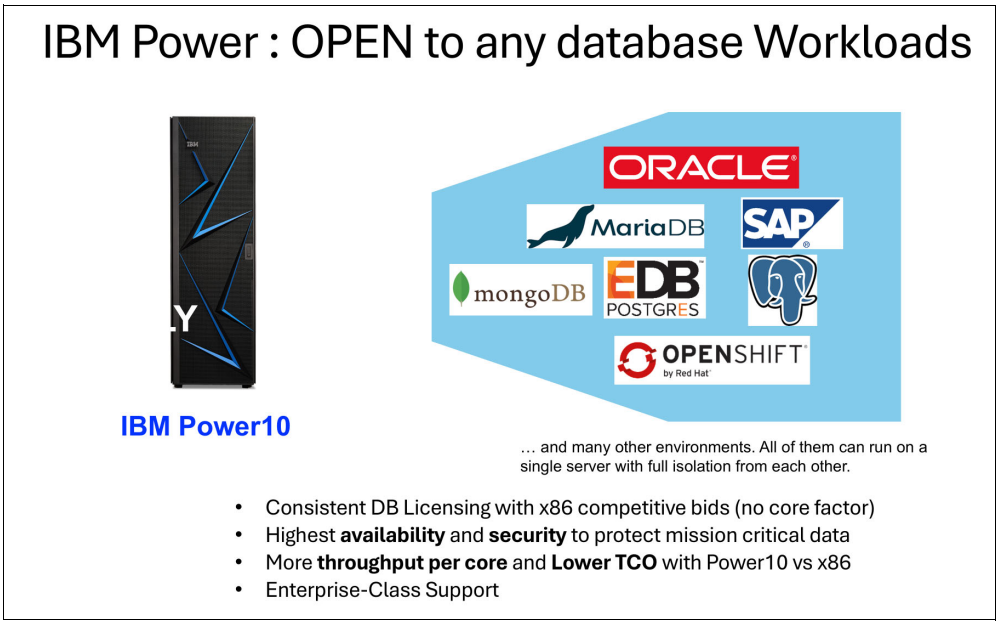


Figure 1-17 Other database choices for IBM Power Virtual Server

You can run these databases on Power Virtual Server by bringing your own licenses.

1.6.4 Development and test environment

For an organization to build, test, and validate an application before going live into production, development and test environments are necessitated. That is where the developers make new features, test their performance, and debug problems without affecting the live production environment.

Creating a development and test environment in an on-premises location can be a difficult and time-consuming process. Developers need to work with the infrastructure team to find resources for their environment. In some cases, this requires that new servers be purchased, requiring time to acquire funding, order and install the equipment and then when the equipment is no longer needed, find another project that can utilize that equipment.

Moving development and test to the cloud can provide a significant productivity benefit for the development staff as well as reduced expenses for the business. IBM Power Virtual Server is ideal for creating a temporary sandbox environment for testing, as a step before deploying production applications. For example:

- You need a remote environment to test software or hardware updates. You have temporary needs for some system resources.
- You need to assess, plan, or test next-generation hardware or OS versions.
- You are interested in an isolated infrastructure, and you need it temporarily just for testing applications.

- You want to test hardware before a hardware refresh.

Using IBM Power Virtual Server for your development environment provides:

- Dynamic Resource Allocation

  Cloud platforms allow you to quickly scale up or down computing resources based on project demands. This ensures that you have the necessary capacity for peak workloads without over investing in hardware.

- Flexible Environments

  You can easily create and configure different development and testing environments with varying configurations (e.g., different operating systems, databases, or software versions) to accommodate diverse project requirements.

- Pay-as-You-Go Model

  Cloud services typically follow a pay-as-you-go pricing model, meaning you only pay for the resources you use. This can significantly reduce upfront costs and eliminate the need for long-term hardware commitments.

- Reduced Overhead

  Cloud providers manage infrastructure maintenance, updates, and security, reducing your operational overhead and freeing up resources for core development activities.

- Rapid Provisioning

  Cloud environments can be provisioned in minutes, allowing you to quickly set up new development or testing environments. This accelerates development cycles and enables faster time-to-market.

- Simplified Management

  Cloud platforms often provide user-friendly interfaces and automation tools, simplifying the management of development and testing environments.

- Remote Access

  Cloud-based environments can be accessed from anywhere with an internet connection, enabling seamless collaboration among development teams, regardless of their physical location.

- Centralized Repository

  Cloud-based version control systems and artifact repositories can provide a centralized location for storing and managing code, ensuring that everyone has access to the latest versions.

- Data Backup and Recovery

  Cloud providers often offer robust data backup and recovery services, helping to protect your valuable assets and minimize downtime in the event of a disaster.

- High Availability

  Cloud infrastructure is typically designed for high availability, ensuring that your development and testing environments remain accessible even in the face of hardware failures or disruptions.

In summary, the cloud offers a flexible, scalable, and cost-effective solution for development and testing environments. By leveraging cloud-based resources, organizations can accelerate development cycles, improve collaboration, and enhance their overall agility and competitiveness.

Modernizing your applications

The advantages of using IBM Power Virtual Server as a development environment are also applicable to clients that are developing Cloud Ready applications. IBM Power Virtual Server provides an excellent platform to modernize your applications utilizing Red Hat OpenShift.

Clients can begin their modernization journey by deploying Red Hat OpenShift on IBM Power Virtual Server utilizing the simplified Installer Provisioned Infrastructure (IPI). IPI is designed to simplify and automate the deployment of a Red Hat OpenShift cluster on Power. For more information on using Red Hat OpenShift on Power see Deploy Red Hat OpenShift.

Power Virtual Server allows integration with on-premises environments to create a hybrid cloud development environment where part of the environment is on-premises while others are on the cloud. This will be helpful in cases when organizations require an on-premises infrastructure due to compliance or performance reasons but would like to have the flexibility of the cloud for testing new features.

CI/CD Pipeline Support

Power Virtual Server environments can support the integration into a Continuous Integration/Continuous Deployment (CI/CD) pipeline. Developers will be able, using their own scripting and automation, to automate the deployment of test environments and run automated tests and promote builds to production seamlessly and in an automated workflow.

Example of Development and Test Environment in Power Virtual Server

A software company development leverages the IBM Power Virtual Server Private Cloud to establish multiple test environments. Each environment emulates a different development life stage, all the way from simple unit testing to preproduction.

In the cloud, the development team provisions an isolated VM for different application modules. They run their tests on IBM AIX VMs and scale up the resources with growing test data complexity.

The company automates the deployment of new builds, execution of unit tests, and reporting by integrating the Power Virtual Server test environments into their CI/CD pipeline. This increase in release cycles ensures that only high-quality code reaches production.

Conclusion

IBM Power Virtual Server provides agile development and testing in an economical, highly scalable environment. The use of Power Virtual Server for development accelerates development cycles through on-demand provisioning, seamlessly integrated hybrid cloud, and automation in workflows while taking ultimate control to the business by assuring maximum utilization of budget spent on resources. Utilizing Power Virtual Server provides a test environment for new applications or to validate updates, and provides the flexibility required by modern development teams.

2

IBM Power Virtual Server Private Cloud

Enterprises adopting innovative hybrid cloud and AI technologies need flexible infrastructures that ensure performance, security, and interoperability to meet evolving IT demands. To stay competitive, reduce technology costs, and respond to changing customer needs, organizations require platforms that enable quick adjustments in workload consumption, whether on the cloud or on-site, prioritizing business outcomes over infrastructure management. Generative AI applications hold significant potential for driving business transformation, and enterprises taking a strategic approach to hybrid cloud may maximize its benefits.

IBM's Hybrid Cloud strategy prioritizes client choice and flexibility. IBM recently expanded the capabilities of Power Virtual Server to allow deployment within client data centers. This approach combines the benefits of cloud consumption and management with on-premises data storage, addressing regional compliance and governance requirements. Power Virtual Server Private Cloud is designed to meet these goals.

Clients have the choice of fully managed off-premises services or private on-premises capabilities in their data center, providing:

- Predefined server, storage, and network architecture in semi-configurable options.
- No re-platforming required.
- Consistent cloud user experience.
- Highly reliable infrastructure with leading business continuity and disaster recovery.
- Superior IBM Power security and comprehensive Cloud compliance.
- Flexible consumption, fully owned and operated by IBM, delivering industry-leading service level agreements (SLAs) using in-country resources.
- Fully metered consumption model with no upfront cost, pay-as-you-use billing, and a minimum monthly commitment.
- Intentional workload placement, integrating with IBM Cloud or on-premises, and benefiting from a strong ecosystem for IBM i, AIX, and Linux.

The following topics are covered in this chapter:

- 2.1, “Power Virtual Server Private Cloud Design” on page 37
- 2.2, “Technology components” on page 43
- 2.3, “Pricing concepts” on page 65

2.1 Power Virtual Server Private Cloud Design

IBM Power Virtual Server is an IBM offering providing cloud services for IBM Power workloads in IBM Cloud data centers. Power Virtual Server was originally offered in 2019 and has grown to 21 data centers world-wide and over 650 customers.

The new on-premises extension of IBM Power Virtual Server was created to provide intentional workload placement while maintaining a consistent management experience. IBM Power Virtual Server Private Cloud benefits clients that:

- Have data sovereignty requirements in a country with no IBM Public Cloud.
- Have regulated or sensitive data and workloads that need to remain on-premises.
- Have workloads with ultra short latency requirements to other on-premises infrastructure.

The newer 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 require some data and workloads to stay in the 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 your 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.

2.1.1 Same management interface as IBM Power Virtual Server

IBM Power Virtual Server Private Cloud is designed to be managed with the same interface that is currently used in the public or off-premises version of the offering. This allows easy integration and migration of workloads between the offering options.

Provisioning a Power Virtual Server environment starts at the IBM Power Virtual Server page in IBM Cloud. Before you can provision any cloud resources you need to have set up an IBM Cloud account with a valid username and password. Figure 2-1 shows the initial screen for provisioning Power Virtual Server.

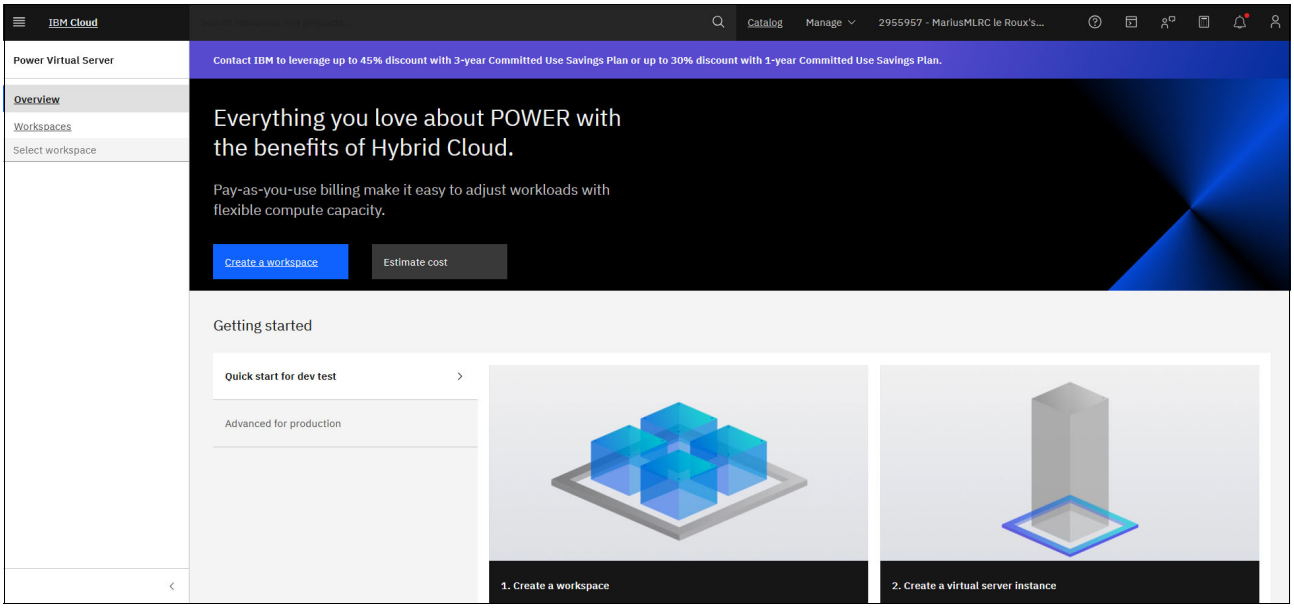


Figure 2-1 Power Virtual Server provisioning page

The first thing that you need to select, depending on your requirements and how you are planning to use these resources, is the type of instance to be created. There are two options available:

1. `Quick start for dev / test`
2. `Advanced for production`

Choosing *`Quick start for dev/test`* provides a simplified process allowing an immediate creation of Power Virtual Server resources intended for dev/test work in the cloud. In contrast choosing *`Advanced for production`* displays additional panels to enter options with an emphasis on network and security configuration.

Next note that there are two panels to be chosen from.

1. `Create a workspace.`
2. `Create a virtual server instance.`

A workspace is a logical container or cloud instance that is used to create and deploy virtual server instances. A workspace is tied to a specific Power Virtual Server cloud location – either off-premises or on-premises. An active workspace has to be configured with the correct region specified and allows for resource grouping and utilization.

A virtual server instance (VSI) is a single server image or logical partition (LPAR) defined in Power Virtual Server. The VSI definition includes defining processors, memory, networking and storage resources that are defined for the specific LPAR. Multiple VSI can be defined

within a workspace. If you have more than one physical location or region, you will have multiple workspaces defined.

Figure 2-2 shows a sample list of workspaces defined within an account. This list is accessed by selecting the `Workspaces` tab on the left of your screen.

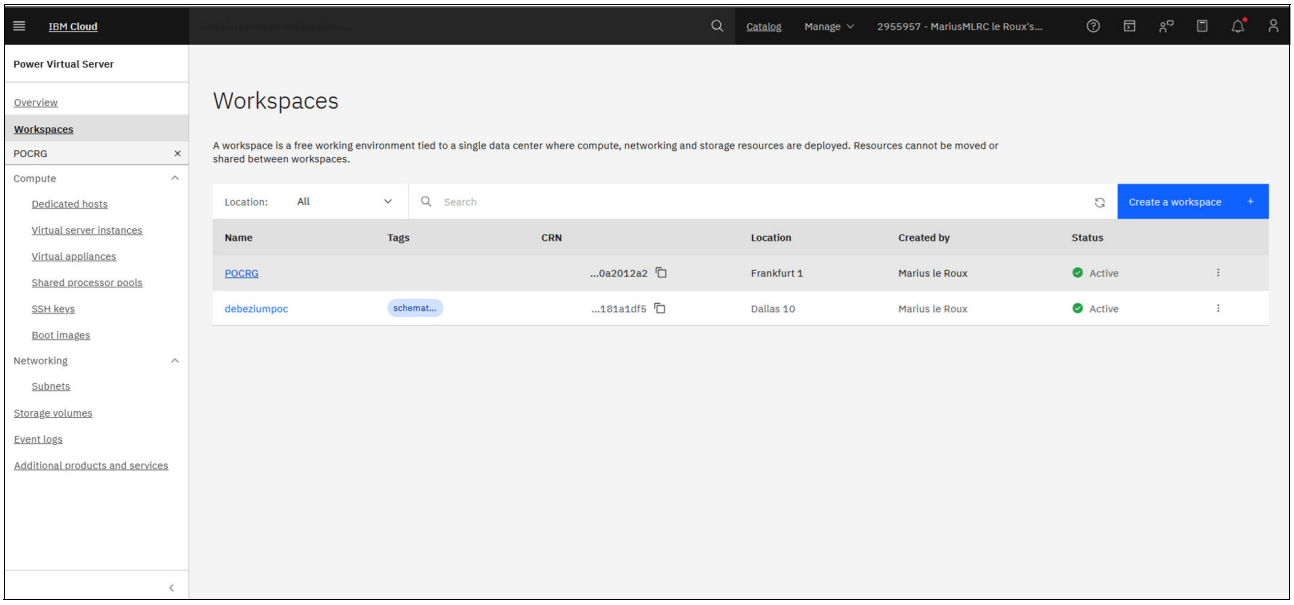


Figure 2-2 List of defined workspaces

> **Important:** A workspace must be defined before you can create a VSI. At least one workspace needs to be defined in each location that you want to run workloads within Power Virtual Server.

If you do not have a workspace defined in the location you wish to define Power Virtual Server LPARs, then you need to select Create a Workspace from the initial Power Virtual Server provisioning screen shown in Figure 2-1 on page 38. The process of defining a workspace is detailed in section 5.2, “Setting up your workspace” on page 115. This is the point where you differentiate between defining resources in the public cloud or in your own private cloud infrastructure.

Once your workspace is defined, you can define your Virtual Server Instances within your workspace. Defining your VSI is detailed in 5.3, “Setting up a virtual server instance” on page 117.

2.1.2 Security and data governance

Power Virtual Server Private Cloud implements foundational security measures, covering Security and Privacy by Design, security controls, data management and encryption, data privacy and jurisdiction considerations, and incident management.

Power Virtual Server Private Cloud prioritizes adherence to privacy regulations, particularly those governing application data placement and location, for customers evaluating the service. To ensure data jurisdiction, Power Virtual Server Private Cloud implements robust security principles and controls, keeping customer data exclusively within their data center or premises.

Power Virtual Server Private Cloud's architecture segregates the tenant and management networks robustly. The tenant networks, which handle customer workload data and VM communication, use both Ethernet and SAN Fiber Channel storage. In contrast, the management network handles control plane traffic, as discussed in the Data management and encryption section. This segregation starts at the physical layer, where each server has at least two Ethernet network interface controllers (NICs), one dedicated solely to management/control traffic. Only Power Virtual Server site reliability engineering (SRE) personnel and service accounts can access the tenant network, with no access to the management network. The tenant network also remains isolated from the internet, ensuring customer workload data on the SAN network and storage systems does not traverse the management network.

Customer data remains in the client data center and is not transferred to the control or management plane. All customer data stored in IBM FlashSystems is encrypted while at rest.

IBM Security and Privacy by Design

Power Virtual Server Private Cloud follows IBM Security® and Privacy by Design[1] (SPbD) practices, which focus on threat modeling, privacy assessment, security testing, static and dynamic code scans, and vulnerability management based on IBM Product Security Incident Response Team (PSIRT) guidelines. The team maintains and reviews the threat model annually or upon significant architectural changes, as required by IBM SPbD.

Security controls

IBM and clients divide security responsibilities according to the joint responsibility model shown in Figure 2-3.

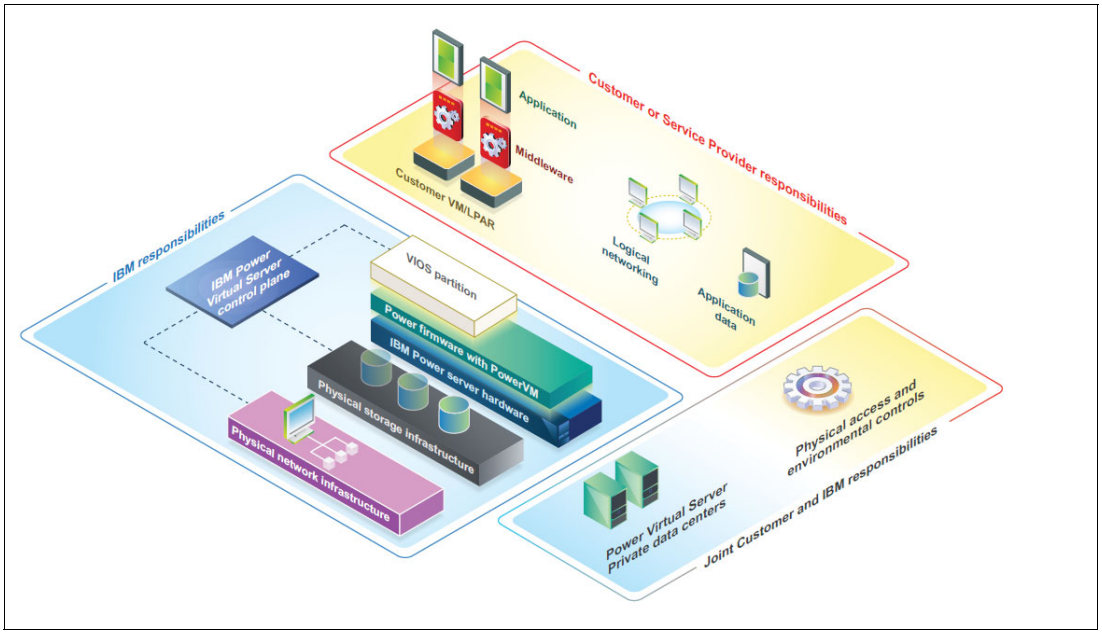


Figure 2-3 Joint responsibility model

In this model:

- IBM handles logical controls within the blue boundary, including Power hardware and firmware, Virtual I/O Servers (VIOS), Network Installation Management servers, IBM PowerVC, NovaLink, Hardware Management Console (HMC), physical network

[1] https://www.ibm.com/support/pages/ibm-security-and-privacy-design

infrastructure, and storage infrastructure, including storage area network (SAN) switches, storage controllers, and storage arrays.

- Clients handle logical controls within the red boundary, covering operating systems, middleware, applications, logical networking, and application data.
- Data center physical security controls, including physical access and environmental controls, are a joint responsibility between customers and Power Virtual Server.

For more information see IBM Power Virtual Server Private Cloud: Security overview.

Data management and encryption

Power Virtual Server Private Cloud minimizes data flow out of the customer data center to the IBM Cloud data center virtual private cloud instance, limiting it to infrastructure device configuration, control/management information, monitoring status, telemetry, and security event data. This data transmission is crucial for the product's optimal functionality and does not include customer workload data or sensitive personally identifiable information, unless explicitly authorized by the customer. All communications between the customer data center and the IBM Cloud VPC instance adhere to stringent security protocols, employing Transport Layer Security (TLS) 1.3 or TLS 1.2 protocols along with National Institute of Standards and Technology (NIST) approved ciphers and algorithms.

Virtual machine lifecycle events, such as VM creation and deletion, are communicated over a dedicated management plane, strictly adhering to TLS 1.2 and above protocols for HTTPS traffic. Connections to infrastructure devices for configuration management, patching, and problem resolution are secured using Secure Shell (SSH).

2.1.3 Pay as you go model

The IBM Power Virtual Server Private Cloud commercial terms are:

- Fully metered flexible capacity consumption (usage) with minimum commit at pod level
  - Metering of actual usage on hourly basis [core, memory (GB), storage (GB)]
  - Monthly minimum commit with monthly true-up if monthly usage < monthly minimum commit
- Contract terms 3 or 5-year with 1-year renewal
  - Option to terminate after 12 months with 60-day notice and pay only 50% of remaining commit
- Single monthly bill one month in arrears
  - Capacity consumption details available on IBM Cloud Portal (UI)
- Advanced support default w/ Premium upgrade option
  - SLA’s and SLO’s aligned to IBM Cloud
- Technology & upgrades:
  - Add Building blocks of Compute & Storage
  - Expand capacity into a pod that is not fully populated
  - Increase the memory capacity only by adding new server(s)
  - Removing Servers and/or Storage from pod not supported
  - Pod upgrades to latest technology only if deemed by IBM
- Countries planned for Limited & General availability:
  - USA, Canada, Germany, Belgium, Netherlands, Luxembourg, Switzerland, Austria, Denmark, UK, Australia

2.1.4 Dynamic provisioning

Power Virtual Server offers Infrastructure as a Service (IaaS) in the public cloud or in a customer's designated data center, known as Power Virtual Server Private Cloud. IaaS solutions are designed to quickly and efficiently allocate compute and storage resources for client workloads. You can create a new Virtual Server Instance, or logical partition (LPAR), through a simple GUI interface, API calls, or automation scripts in just minutes.

IBM PowerVM provides the underlying virtualization function for IBM Power servers, efficiently sharing server resources across multiple LPARs. You can dynamically modify each LPAR configuration by adding or removing processors and memory while it runs. This dynamic provisioning and deprovisioning also applies to storage configuration, allowing you to add storage capacity to an operational LPAR. In some cases, you can even remove storage capacity if the application is not accessing data on those volumes.

This flexibility enables you to scale your infrastructure up or down to meet business requirements. For example, you can quickly increase processing capability during peak workload times and reduce it when the peak subsides, resulting in lower computing expenses.

2.1.5 Scalable

IBM Power Virtual Server Private Cloud is an as-a-service offering that includes a prescriptive set of physical infrastructure which includes compute, network, and storage. The physical infrastructure, also known as, the pod, is deployed in the customer data center. The pod is maintained and operated by IBM site reliability engineering (SREs) and managed through the IBM Cloud platform. Each pod is associated with an IBM Cloud Satellite™ location that is owned by customer IBM Cloud account.

The architecture is set up to allow clients to scale their existing pods by adding additional compute and storage within the existing racks. For medium pods with two racks, the architecture supports the expansion to four racks. It is also possible to add additional pods to scale the environment horizontally to meet additional workload demands.

2.1.6 Ease of implementation and migration

IBM Power Virtual Server Private Cloud, an as-a-service offering is designed to be easy to implement, giving customers cloud experience with infrastructure in their data center. Here are some of the features that make it easy to use:

- Quick deployment
- Automation interfaces

  GUI, CLI, API, or Terraform interfaces makes the management of resources very easy
- Bring your own image

  Customers have flexibility to bring their own custom IBM AIX, Linux, or IBM i images
- Dynamic resource adjustment

  Configure and customize number of cores, amount of memory and storage volume size on the virtual server.
- Private virtual server access

  Customer can use a VPN to access the virtual server.

- Pay-as-you-use-billing

  Based on specific needs, customers can choose to scale up or down resources dynamically based on need.

2.2 Technology components

IBM Power Virtual Server Private Cloud is an as-a-service offering with highly prescriptive infrastructure (compute, storage, and network) residing in the customer's location (on-premises). The management will be performed via IBM Cloud interfaces (GUI, API, CLI, Terraform). The various IBM Cloud regions (i.e., Dallas, Washington DC, London, Frankfurt, Sao Paulo, etc.) will host the Power Virtual Server Private Cloud control plane software (Power Service Broker) and the client locations will be configured to connect to the nearest IBM Cloud region.

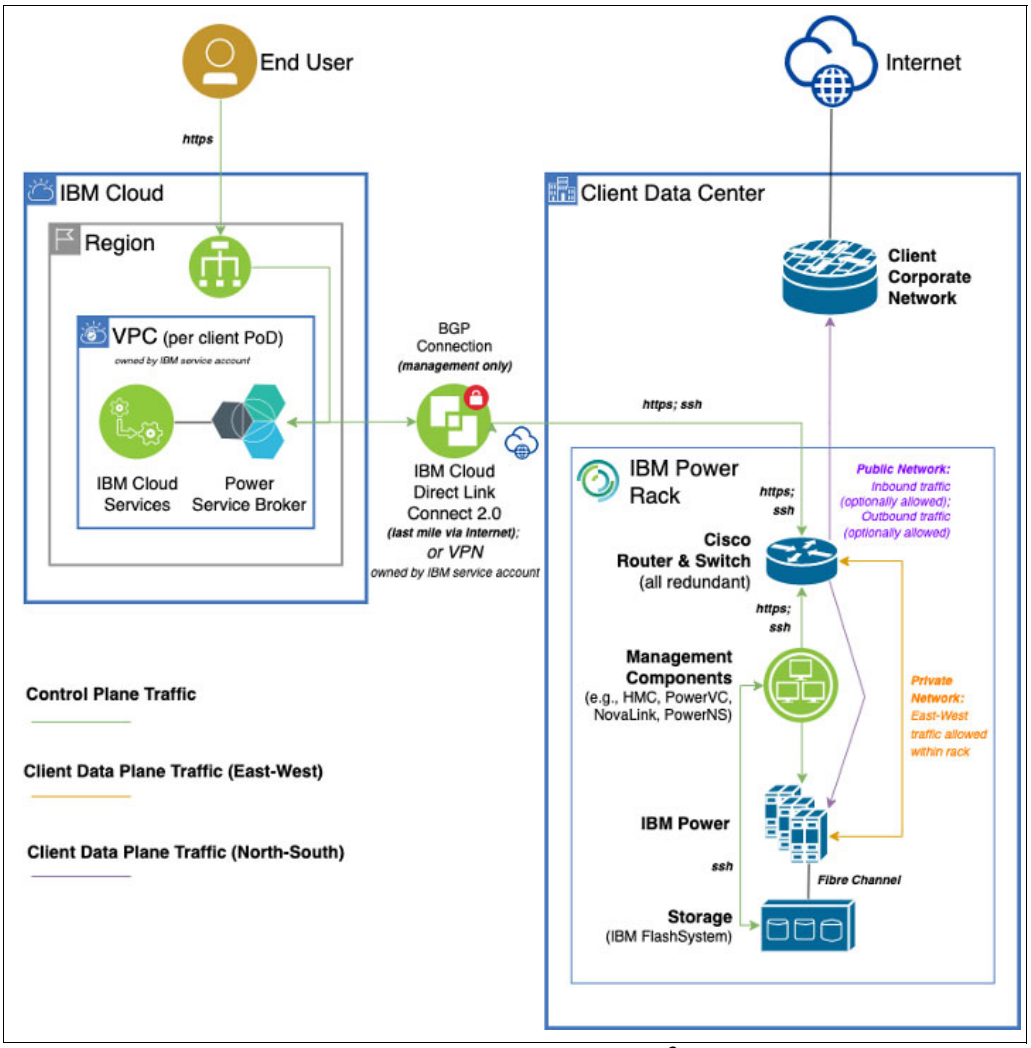


Figure 2-4 Power Virtual Server Private Cloud high level view[2]

[2] https://cloud.ibm.com/docs/power-iaas?topic=power-iaas-private-cloud-architecture

2.2.1 Pods

A point of delivery (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.

As of the time writing this book, the following pod sizes are available:

- Small: 1 rack of IBM Power10 (S1022 and E1050) processors, storage, networking and management components
- Medium: Either 2 or 4 racks of IBM Power10 (S1022, E1050, or E1080) processors. storage, networking and management components.

Small pods

Small pods are designed as an entry configuration that support either the Power S1022 processor or the Power E1050. When using the Power S1022, the small pod supports up to 297 customer available cores and 36 TB of memory. With the Power E1050, a small pod supports up to 340 customer available cores and 32 TB of memory. A small pod provides either 438 TB or 876 TB of usable storage. Table 2-1 shows the different capabilities of the small pod.

Table 2-1 Small pod attributes

Attribute	Value
Client-usable hosts S1022 2 TB	Min: 6 Max: 9
Client-usable hosts S1022 4 TB	Min: 5 Max: 9
Client-usable hosts E1050	Min: 2 Max: 4
Client-usable cores per Pod S1022 (2U): 40 total; 33 usable E1050 (4U): 96 Total; 85 usable	 Min: 198 – S1022 (2 TB) Min: 165 – S1022 (2 TB) Max: 297 – S1022 Min: 170 – E1050 Max: 340 – E1050
Client-usable memory per Pod S1022 (2U) 2/4 TB options E1050 (4U) 4/8 TB options	 Min: 12 TB –S1022 (2 TB) Max: 36 TB–S1022 Min: 8 TB – E1050 Max: 32 TB – E1050
Racks per Pod	1
Redundant HMC	No
Client usable storage per Pod Assumes 2x compression 1– 2 FS5200 controller 12 19.2 TB drives per controller	 Min: 438 TB Max: 876 TB
Max power consumption **Estimated max weight**	Rack 1: 24.7 kVA Rack 1: 1748 lbs

Figure 2-5 is a diagram showing the small pod setup.

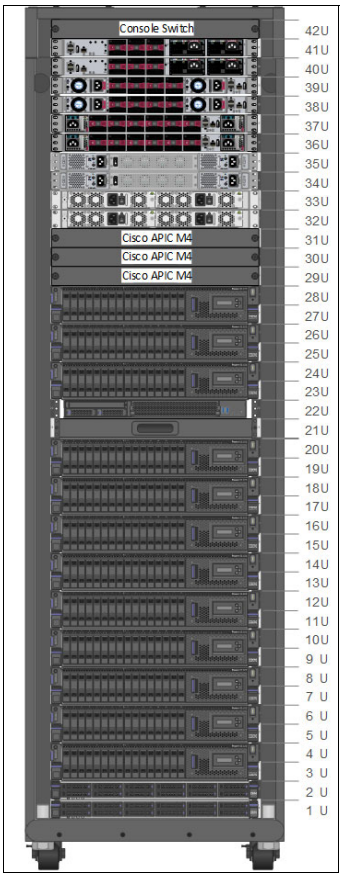


Figure 2-5 Small Pod

Medium pods

The medium pods support any of the Power10 models, the Power E1080, the Power E1050 or the Power S1022, providing significant flexibility. Depending on the configuration, the medium pod consists of either 2 racks or 4 racks. A medium pod can provide up to 1287 usable cores using the Power S1022, up to 1615 usable cores using the Power E1050, or up to 535 usable cores using the Power E1080. A medium pod provides up to 160 TB of memory (using the Power E1080).

Up to 3.5 PB of storage is available with 2 racks and up to 7 PB is available with 4 racks in the medium pod. Additionally, the medium pod provides a redundant HMC. Table 2-2 shows the attributes of the medium pod.

Table 2-2 .Medium pod details

Attribute	Value
Client-usable hosts S1022	Min: 12 Max:15 (2-rack) Max: 39 (4-rack)
Client-usable hosts E1050	Min: 5 (2-rack) Min: 8 (4-rack) Max:7 (2-rack) Max: 19 (4-rack)

Attribute	Value
Client-usable hosts E1080	Min: 2 (4-rack) Max: 5 (4-rack)4
Client-usable cores per pod S1022 (2U): 40 total; 33 usable E1050 (4U): 96 Total; 85 usable E1080 (10U): 120 Total; 107 usable	 Min: 396 – S1022 Max: 495 – S1022 (2 Rack) Max: 1287 – S1022 (4 Rack) Min: 425 – E1050 (2 Rack) Min: 680 – E1050 (4 Rack) Max: 595 – E1050 (2 Rack) Max: 1615 – E1050 (4 Rack) Min: 214 – E1080 (4 Rack) Max: 535 – E1080 (4 Rack)
Client-usable memory per pod S1022 (2U) 2/4 TB options E1050 (4U) 4/8 TB options E1080 (10U) 8/16/32 TB options	 Min: 24 TB –S1022 (2 TB) Max: 60 TB–S1022 (4 TB) (2 Rack) Max: 156 TB–S1022 (4 TB) (4 Rack) Min: 20 TB – E1050 (4 TB) Max: 56 TB – E1050 (8 TB) (2 Rack) Max: 152 TB – E1050 (8 TB) (4 Rack) Min: 16 TB – E1080 (8 TB) Max: 160 TB – E1080 (32 TB) (4 Rack)
Racks per pod	2 or 4 (must add in pairs)
Redundant HMC	Yes
Client usable storage per pod Assumes 2x compression 2 FS9500 controllers (2 Rack) 1 – 4 FS9500 controllers (4 Rack) 28 or 48 19.2 TB drives per controller	 Min: 1.75 PB (24 drives per controller; 2 Rack) Max: 3.5 PB (48 drives per controller; 2 Rack) Max: 7.0 PB (48 drives per controller;4 Rack)
Max power consumption	Rack 1: 22.3 kVA Rack 2: 22.3 kVA Rack 3: 24.6 kVA Rack 4: 24.6 kVA
Estimated max weight	Rack 1: 1581 lbs Rack 2: 1581 lbs Rack 3: 1593 lbs Rack 4: 1593 lbs

Figure 2-6 shows the configuration of the medium pod with four racks.

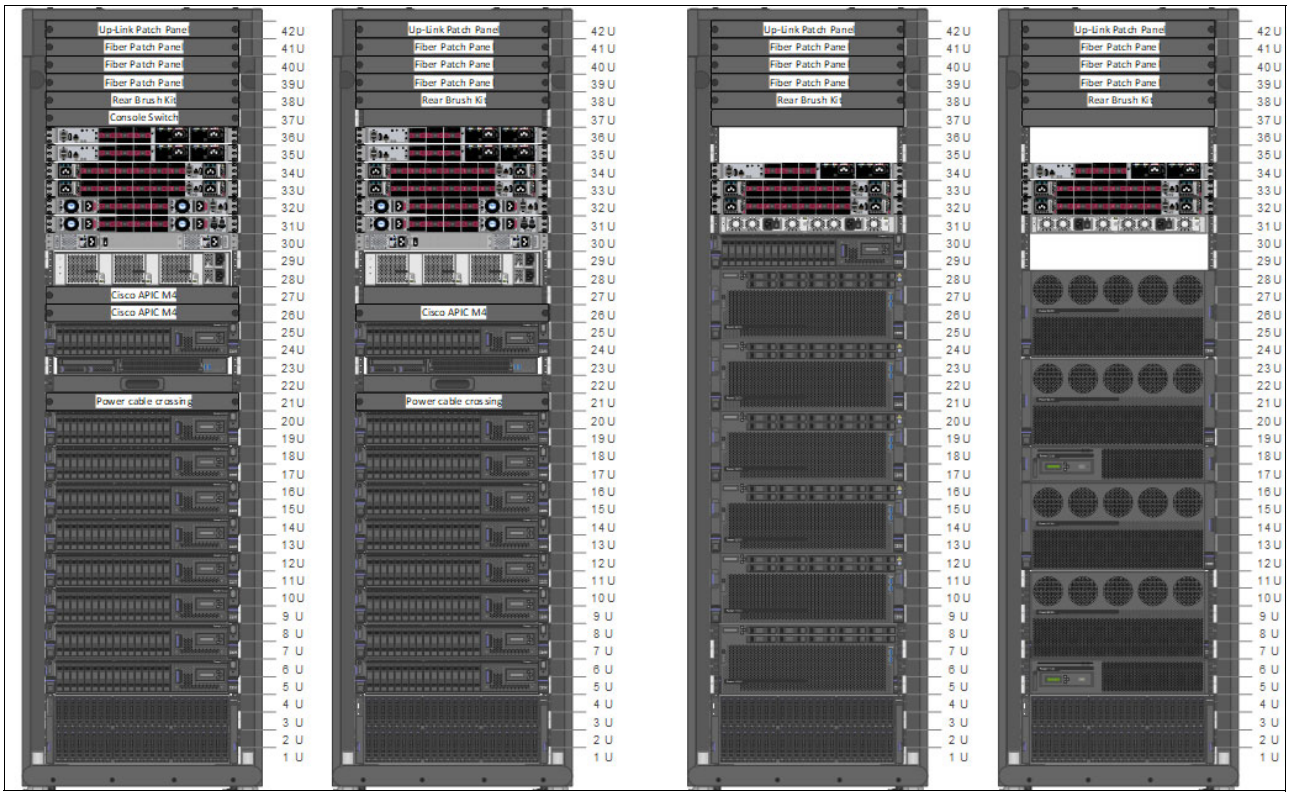


Figure 2-6 Four rack medium pod diagram

2.2.2 Compute

The compute nodes used in the IBM Power Virtual Server Private Cloud all utilize IBM Power10 based processors. The IBM Power10 processor is the newest generation of IBM Power processors, providing highly available, high-performance systems specifically built with security in mind.

The IBM Power10 processor was introduced to the public on 17 August 2020 at the 32nd HOT CHIPS1 semiconductor conference. At that meeting, the new capabilities and features of the latest IBM POWER® processor microarchitecture and the Power Instruction Set Architecture (ISA) v3.1B were revealed and categorized according to the following Power10 processor design priority focus areas:

- Data plane bandwidth

  Terabyte per second signaling bandwidth on processor functional interfaces, petabyte system memory capacities, 16-socket symmetric multiprocessing (SMP) scalability, and memory clustering and memory inception capability.

- Powerful enterprise core

  New core micro-architecture, flexibility, larger caches, and reduced latencies.

- End-to-end security

  Hardware-enabled security features that are co-optimized with PowerVM hypervisor support.

- Energy-efficiency

  Up to threefold energy-efficiency improvement in comparison to IBM Power9 processor technology.

- Artificial intelligence (AI)-infused core

  A 10 - 20x matrix-math performance improvement per socket compared to the IBM Power9 processor technology capability.

The IBM Power10 Processor session material that was presented at the 32nd HOT CHIPS conference is available through the HC32 conference proceedings archive at this web page.

IBM S1022 Server (9105-22A)

The Power S1022 (9105-22A) 2U rack mount server is a powerful one- or two-socket server that is available with one or two processors per system, with the following options:

- One or two 12-core Power10 processors running at a typical 2.90 - 4.0 GHz (maximum)
- Two 16-core Power10 processors running at a typical 2.75 - 4.0 GHz (maximum)
- Two 20 core Power10 processor running at a typical 2.45 - 3.90 GHz (maximum)

All processor cores can run up to eight simultaneous threads to deliver greater throughput. When two sockets are populated, both must be the same processor model.

The Power S1022 server is available in small and medium Pods. Within the Power Virtual Cloud Private configuration, it comes with the features shown in Table 2-3.

Table 2-3 Resource specifications for IBM Power S1022

Attribute	Value
Applicable pod sizes	small; medium
Rack space	2U
Total cores	40
Total usable cores	33
Memory options	2 TB 4 TB
H/W-based transparent memory encryption	Yes
VIOS configuration	2 LPARs 3 cores/VIOS 16 GB memory/VIOS
NovaLink configuration	1 core 16 GB memory
Network adapters	Small: 25GbE (2X) Medium: 100GbE (2X)
Fibre Channel adapters	64 Gbps; 2-port (2X)
Max kVA	1.513
Amps	7.57
Watts	1,468
BTU per hour	5010
Weight	71 lbs

Figure 2-7 shows the S1022 server from the front and the back.

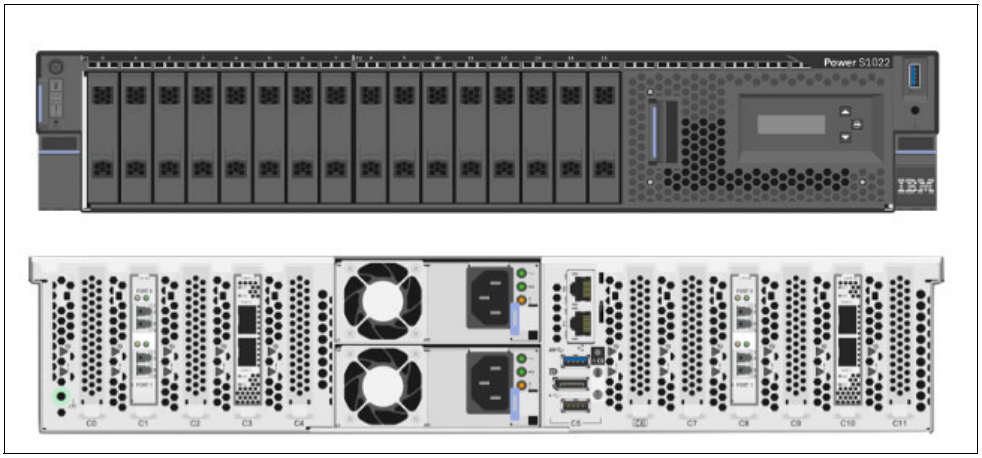


Figure 2-7 S1022 server

For more information regarding the IBM S1022, see *IBM Power 10 Scale Out Servers Technical Overview S1012, S1014, S1022s, S1022 and S1024*, REDP-5675.

IBM E1050 Server (9043-MRX)

The Power E1050 server is ideal for cloud deployments because of its built-in virtualization, flexible capacity, and high usage. The machine type model number of the Power E1050 server is 9043-MRX. It is a single enclosure server that is four EIA units tall (4U). It can be configured with two, three, or four dual-chip modules (DCMs). Three processor options are available:

- Twelve cores running at a typical 3.35 - 4.00 GHz (max) frequency range
- Eighteen cores running at a typical 3.20 - 4.00 GHz (max) frequency range
- Twenty-four cores running at a typical 2.95 - 3.90 GHz (max) frequency range

The Power E1050 server is available in small and medium Pods. Within the Power Virtual Cloud Private configuration, it comes with the features shown in Table 2-4.

Table 2-4 Resource specification for IBM Power E1050

Attribute	Value
Applicable pod sizes	small; medium
Rack space	4U
Total cores	96
Total usable cores	85
Memory options	4 TB 8 TB
H/W-based transparent memory encryption	Yes
VIOS configuration	2 LPARs 5 cores/VIOS 32 GB memory/VIOS
NovaLink configuration	1 core 16 GB memory

Attribute	Value
Network adapters	Small: 25 GbE (4X) Medium: 100 GbE (4X)
Fibre Channel adapters	64 Gbps; 2-port (4X)
Max kVA	3,109
Amps	15.55
Watts	3,016
BTU per hour	10,292
Weight	153 lbs

Figure 2-8 shows the Power E1050.

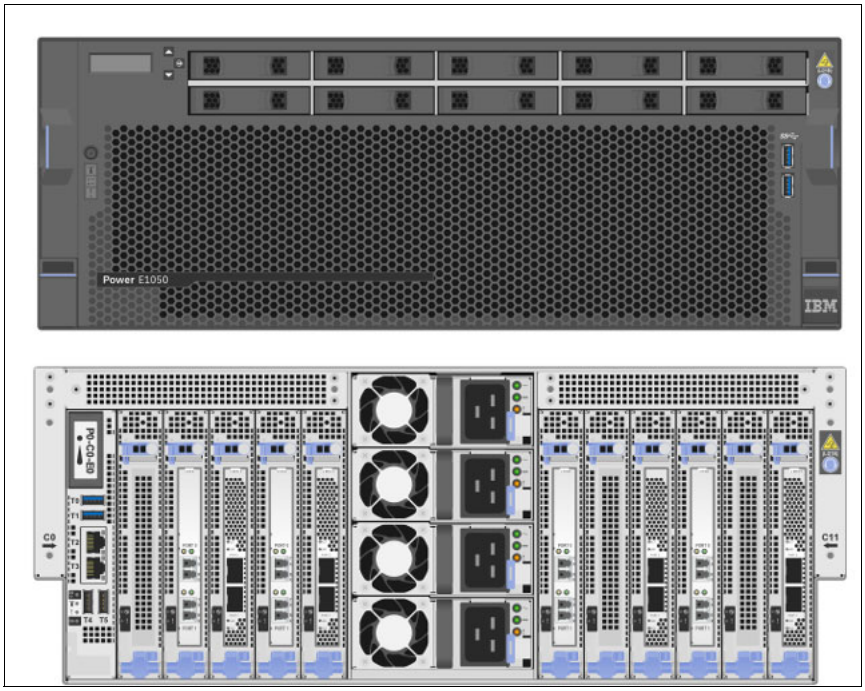


Figure 2-8 IBM Power E1050

For more information on the IBM Power E1050 see *IBM Power E1050 IBM Power E1050 Technical Overview and Introduction*, REDP-5684

IBM E1080 Server (9080-HEX)

The Power E1080, also referred to by its 9080-HEX machine type-model designation, represents the most powerful and scalable server in the IBM Power portfolio. It is comprised of a combination of CEC enclosures that are called nodes (or system nodes) along with other drawers to support additional i/o adapters or internal disk capacity. The Power E1080 is designed with reliability, scalability, and security built in and has fully integrated encryption capability as well as advanced AI capabilities.

The Power E1080 server provides the following hardware components and characteristics:

- 10-, 12-, or 15-core Power10 processor chips that are packaged in a single chip module per socket.
- One, two, three, or four system nodes with four Power10 processor sockets each.
- Redundant clocking in each system node.
- Up to 60 Power10 processor cores per system node and up to 240 per system

In Power Virtual Server Private Cloud, the IBM Power E1080 Server is only available in Medium Pods. The configuration supplied in Power Virtual Server Private Cloud includes two system nodes, each with four 15-core 3.55 - 4.00 GHz (max) processor for a total of 60 cores in each system node. This provides a total of 120 processors per system – of which 107 are available for customer deployments. There are three memory configurations available – 8 TB, 16 TB and 32 TB – for configuration flexibility. More details about the configuration are shown in Table 2-5.

Table 2-5 .Resource specification for IBM Power E1080

Attribute	**Value**
Applicable pod sizes	medium
Rack space	10U
Total cores	120
Total usable cores	107
Memory options	8 TB 16 TB 32 TB
H/W-based transparent memory encryption	Yes
VIOS configuration	2 LPARs 10 cores/VIOS 48 GB memory/VIOS
NovaLink configuration	1 core 16 GB memory
Network adapters	100 GbE (4X)
Fibre Channel adapters	64 Gbps; 2-port (4X)
Max kVA	7.621
Amps	38.10
Watts	7,392
BTU per hour	25,229
Weight	410 lbs

Figure 2-9 shows an IBM Power E1080 system node. A Power E1080 built for Power Virtual Server Private Cloud will have two of these system nodes and a System Control Unit which contains additional components for reliability and serviceability.

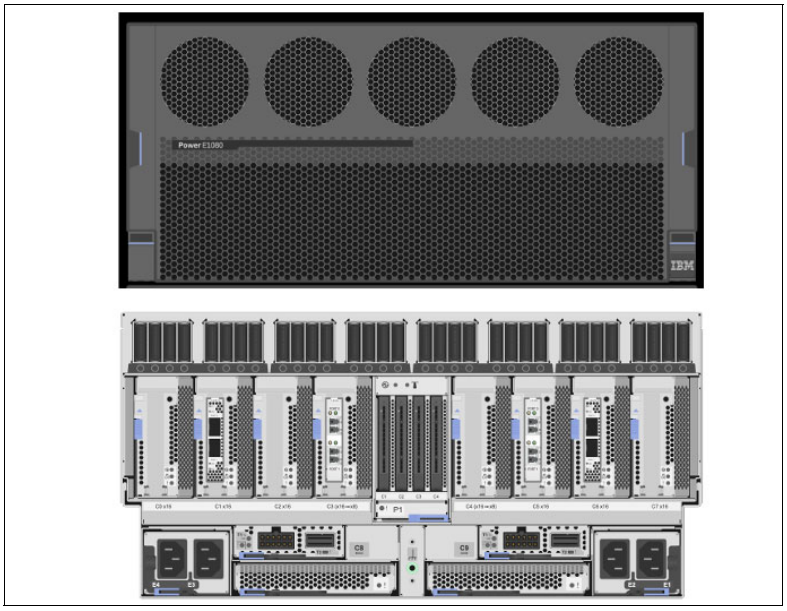


Figure 2-9 IBM Power E1080

For more information regarding the IBM E1080 Server, see *IBM Power E1080 Technical Overview and Introduction*, REDP-5649

Additional Compute components

In addition to the compute components listed above, the pod contains an IBM Hardware Management Console and display and an IBM System Management node (9105-22A).

2.2.3 Storage

Storage within the pod is provided by IBM FlashSystem® controllers which are connected using SAN switches to the compute nodes to provide storage. Small Pods have options for either one or two IBM FlashSystem 5200 storage controllers with a capacity of 435 TB per controller. Medium Pods have options for either two or four FlashSystem 9500 storage controllers with a capacity of 1.75 PB per controller.

The underlying Brocade-based SAN fabric connects the compute nodes at 64 Gbps and the storage nodes connect at 32 Gbps with redundancy built in at both the port and switch levels.

IBM FlashSystem 5200

For small Pods, storage is provided by SAN attached FlashSystem 5200 storage controllers. Each controller enclosure has dual redundant storage controllers that connect to the SAN utilizing Fibre Channel connections. The control enclosure supports 12 FlashCore Modules (FCMs). The FlashCore Module design utilizes the NVMe protocol, a PCIe Gen4 interface, and high-speed NAND memory to provide high throughput and IOPS with consistent and predictable latency. The FCM supports hardware-based data compression and self-encryption.

At least one FlashSystem 5200 controller is required in each small Pod, with the ability to choose two. This allows support for up to 876 TB of allocatable storage. Figure 2-10 on

page 53 shows the IBM FlashSystem 5200 (4662-6H2). The FlashSystem 5200 is always populated with 12 19.2 TB FCMs.

Figure 2-10 shows the IBM FlashSystem 5200.

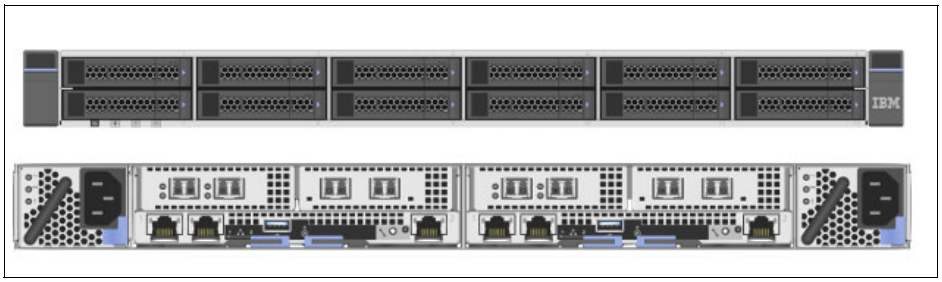


Figure 2-10 IBM FlashSystem 5200

Table 2-6 shows the attributes of the FlashSystem 5200 as configured in the pod.

Table 2-6 FlashSystem 5200 attributes

Attribute	Value
Applicable pod sizes	small
Rack space	1U
Total capacity per controller	438 TB
Drives per controller	12
Network management connections	1 GbE (2X)
Fibre Channel connections	32 Gbps (8X)
Max kVA	0.928
Amps	4.64
Watts	900
BTU per hour	3,072
Weight	43 lbs

For more information regarding the IBM Storage FlashSystem 5200 see *IBM Storage FlashSystem 5200 Product Guide for IBM Storage Virtualize 8.6*, REDP-5617.

IBM FlashSystem 9500

Storage for medium Pods is provided by the FlashSystem 9500. The FlashSystem 9500 is a storage controller with dual redundant controllers and supports up to 48 storage devices. The storage controllers attach via SAN connections to the SAN switches built into the pod. For Power Virtual Server Private Cloud Pods, each controller is configured with either 24 or 48 19.2 TB FCMs providing either 876 TB or 1.75 PB per controller. A medium pod must have at least one FlashSystem 9500 and can have up to four for a total maximum storage capacity of 7 PB.

Figure 2-11 shows the IBM Flash System 9500 (4666-AH8) which is used in the medium Pods.

Figure 2-11 FlashSystem 9500 used in medium Pods

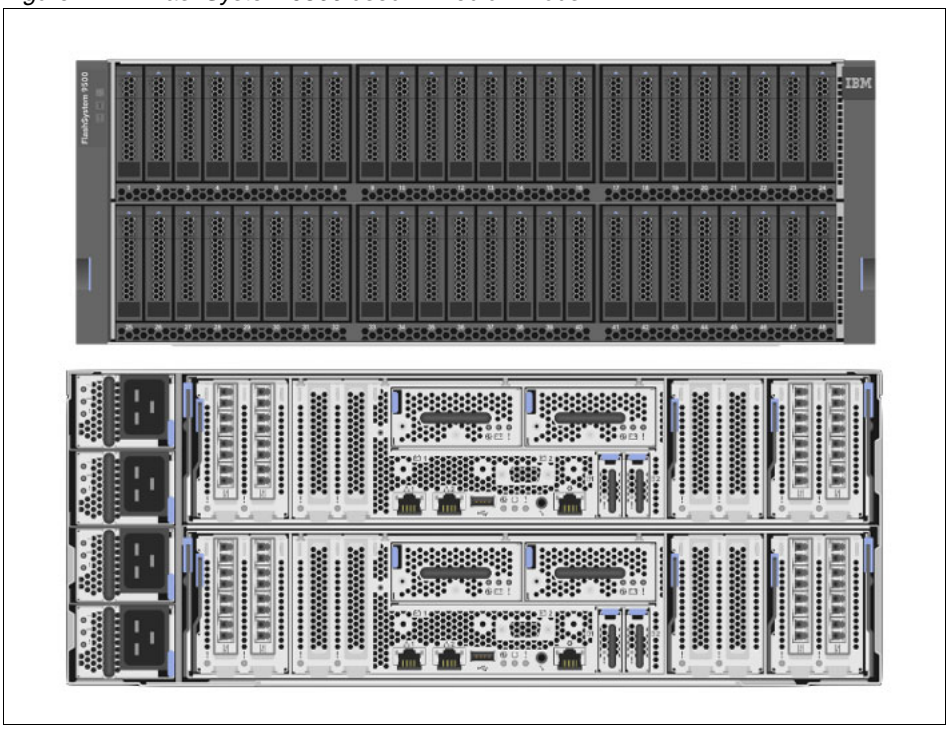


Table 2-7 details the attributes of the FlashSystem 9500 in the medium Pods.

Table 2-7 FlashSystem 9500 attributes

Attribute	Value
Applicable pod sizes	medium
Rack space	4U
Total capacity per controller	876 TB (24 drives) 1.75 PB (48 drives)
Drives per controller	24 48
Network management connections	1 GbE (2X)
Fibre Channel connections	32 Gbps (32X)
Max kVA	3.403
Amps	17.01
Watts	3,300
BTU per hour	11,263
Weight	155 lbs

For more information on the FlashSystem 9500, see *IBM Storage FlashSystem 9500 Product Guide: Updated for IBM Storage Virtualize 8.7*, REDP-5742.

Storage tiers

IBM Power Virtual Server provides a tier-less storage offering that removes the notion of a disk type and replaces it with a storage pool. Each storage pool can support multiple storage tiers. The storage tiers are based on different IOPS levels.

Storage tiers are designed to provide different levels of storage pricing. The client is charged more for higher performance and less for lower performance or inactive data. Storage tier pricing is based on I/O operations per second (IOPS).

Tier0, Tier1, and Tier3 performance is based on IOPS per GB, meaning that the performance of your storage volumes is connected to the size of the volume. This works well for many workloads, but there are some workloads that have a smaller amount of data and are hampered by the IOPS/GB calculation. For these workloads, Power Virtual Server provides the Fixed IOPS tier which provides 5000 IOPS per data volume, regardless of the size.

Note: The Fixed IOPS tier is only available for volumes with a size of 200 GB or less. Above 200 GB using Tier0 provides higher performance (200 GB @ 25 IOPS/GB = 5000 IOPS).

Table 2-8 shows the supported storage tiers with corresponding IOPS.

Table 2-8 Storage tiers within Power Virtual Server

Tier level	IOPS	Performance
Tier 0	25 IOPS/GB	A 100-GB volume receives 2500 IOPS. This is 2.5x faster than tier 1 and 8.3x faster than tier 3.
Tier 1	10 IOPS/GB	A 100-GB volume receives 1000 IOPS. This is 3.3x faster than tier 3.
Tier 3	3 IOPS/GB	A 100-GB volume receives 300 IOPS.
Fixed IOPS[a]	5000 IOPS regardless of size	A 100-GB volume receives 5000 IOPS.

a. The use of fixed IOPS is limited to volumes with a size of 200 GB or less, which is the break-even size with Tier 0 (200 GB @ 25 IOPS/GB = 5000 IOPS).

Data resiliency

Power Virtual Server provides a highly available storage environment with redundant SAN connections to the storage controllers and RAID protection to protect from storage media failures. In addition, Power Virtual Server supports two methods of data replication to avoid application outages due to site failures such as power failures and other issues such as fires or floods that prevent access to the data stored in the pod; application or system-based data replication, or asynchronous storage replication.

Application or system data replication

System and application replication replicates the data over the network to a Power Virtual Server pod in another physical location or even to a Power Virtual Server instance in an IBM datacenter. This can be done using:

- Cluster tools such as IBM PowerHA® SystemMirror® for AIX or IBM i, and high availability clustering solutions utilizing PaceMaker on Linux.
- Other system replication tools such as MIMIX or Maxava.
- Database replication technologies such as IBM DB2® Mirror for i.

When utilizing system-based or application-based tools, data is copied to a server in a remote location and in case of an outage, the tools restart the application on the remote server if a failure occurs (often described as a role swap). When the site is recovered, the data is migrated back to the primary site and the applications are then restarted in the primary site.

Asynchronous Storage Replication with Global Replication Services

Global Replication Services (GRS) is the Power Virtual Server offering that provides asynchronous storage replication. Replication at the storage layer is where the primary storage controller located in the pod in your Power Virtual Server Private Cloud instance is connected over the network to a storage controller in a remote location - either in another Power Virtual Server Private Cloud pod or in a Power Virtual Server instance in an IBM datacenter.

In case of a site failure, the applications are restarted on a server in the remote location and production can be resumed. When the site is recovered the storage replication is restarted to copy the data from the remote site back to the primary site. When the data is fully replicated the applications can be restarted in the primary location.

The primary advantage of storage replication is that the data replication does not take any additional compute resources from your application or database servers which can either reduce the consumption costs of your Power Virtual Server Private Cloud or provide additional compute resources for other applications and services. GRS replicated volumes are charged as twice the rate of equivalent tier volumes.

Additional details on GRS can be found in “Storage Replication using Global Replication Services” on page 96.

SAN Fabric

A Storage Area Network is built into the design of the pod. It is not customer accessible. The design of the SAN is done using best practices and provides switch level redundancy within the pod. In addition, each component of the SAN is designed with port level redundancy to eliminate any single point of failure within the SAN.

Figure 2-12 shows the SAN setup for the small Pods.

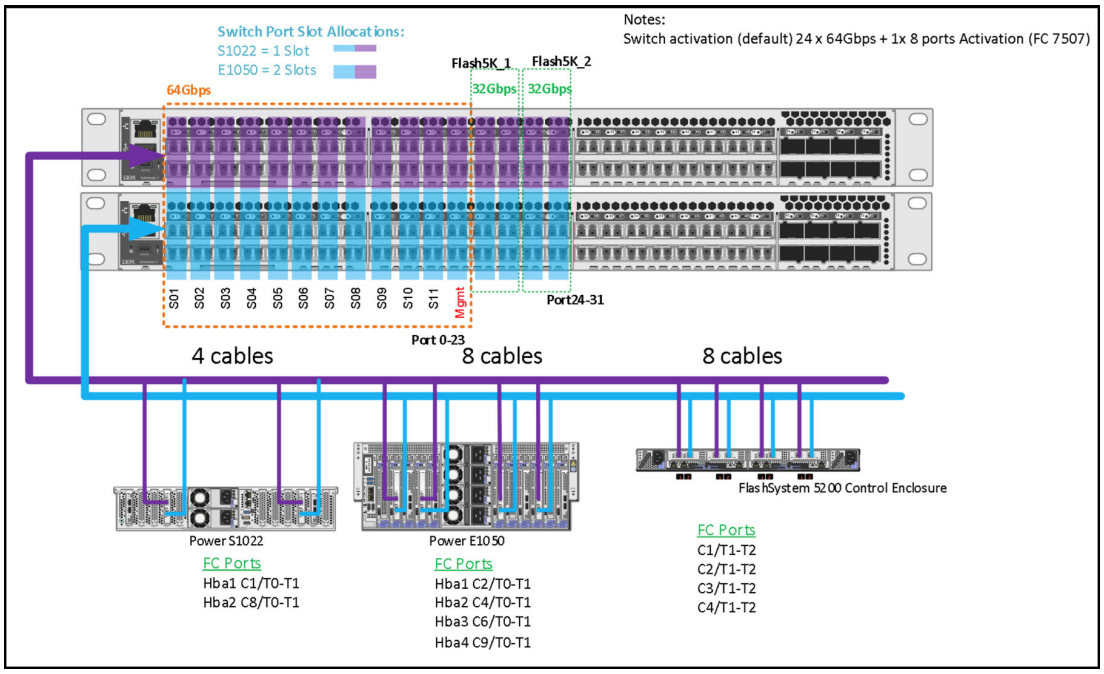


Figure 2-12 Small pod SAN design

Similarly, Figure 2-13 shows the SAN design for the medium pod. When only two racks are configured in a medium Pod, all the servers, the storage controllers and the SAN switches are installed in those two racks. When racks three and four are added to the medium Pod, those racks contain additional servers and the third and fourth storage controllers (if configured). Racks three and four also contain additional SAN switches which are used to connect the servers in those racks. These switches are connected via inter-switch links (ISL) to the switches in racks 1 and 2. All of the storage controllers are connected to switches in rack 1 and rack 2.

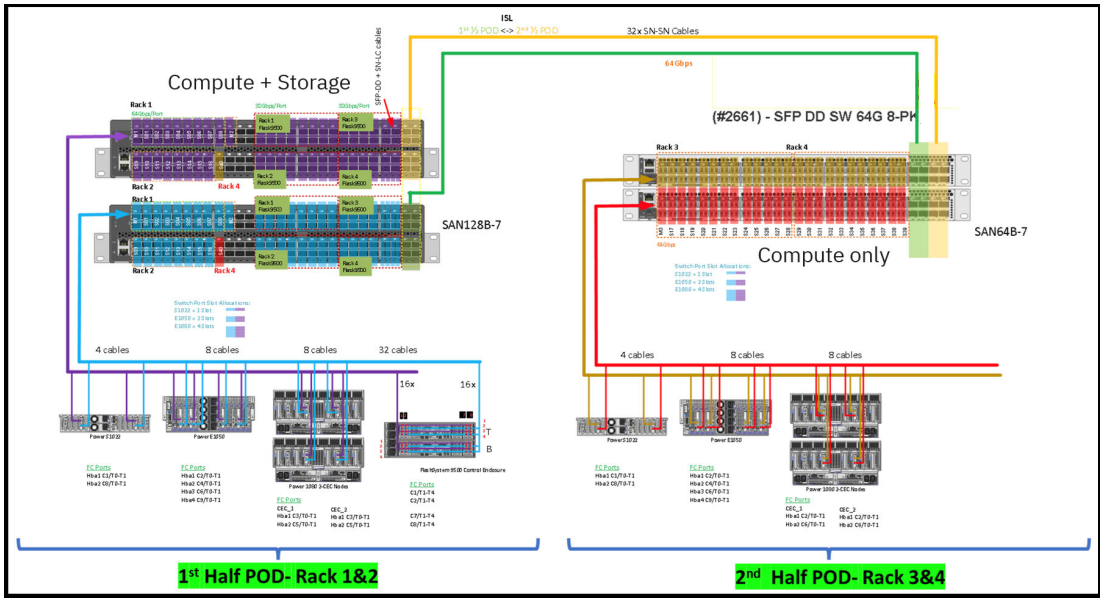


Figure 2-13 Medium pod SAN design

2.2.4 Network

The Power Virtual Server Private Cloud pod comes with all of the networking components for:

- Interconnections for data traffic between the virtual servers running in the pod.
- Connections for data traffic between the virtual servers running in the pod and the client data center network.
- Connections to the IBM Cloud network for control and management.
- Connections to the servers for management and control.
- Connections to the storage controllers for management and control.
- Connectivity to the hardware management consoles for management and control.

Other than the connections to the client data center and the IBM Cloud control points, the pod is completely self-contained, and no other connectivity is supported.

Network components

Depending on which pod configuration is being installed, the actual components being used are slightly different. The following is a list of the components that are used in the pod.

- Cisco ASR1001 Router
- Cisco APIC M4 Server
- Cisco 400G Spine Switch (N9K-C93600CD-GX) (medium pod only)
- Cisco Leaf 25G Switch (N9K-C93180YC-FX)
- Cisco Leaf 100G Switch (N9K-C93600CD-GX) (Used as Spine on small pod)
- Cisco Leaf 1G Switch (N9K-C9348GC-FXP)
- Cisco 1G Switch (N9K-92348GC-X), OOBM (medium pod only)
- Vertiv Avocent ACS Serial Console

Network Internal Design

The networking topology in the pod is a leaf-spine design. Leaf-spine is a two-layer network topology composed of spine and leaf switches. A spine-leaf architecture helps data center networks reduce network latency and hop count and improve network efficiency. Servers and storage connect to leaf switches which aggregate traffic from servers and connect directly to the spine.

In the small Pod, the connections between the compute nodes and the ACI fabric use the 25gig leaf switches. The 25 gig leaf switches connect into 100gig spines. The customer connections would be between either the ASR1k or the 25 gig border leafs.

On the medium Pod, the connections between the compute nodes and the ACI fabric use the 100gig leaf switches. All leafs in the ACI fabric connect up to 400gig spines. The customer connections would also use either the ASR1k or the 25gig border leafs.

Figure 2-14 on page 59 shows the networking components and connectivity options for the customer data center switches. The customer data center connection can be either Layer 2 or Layer 3 and we support 10G, 25G or 100G.

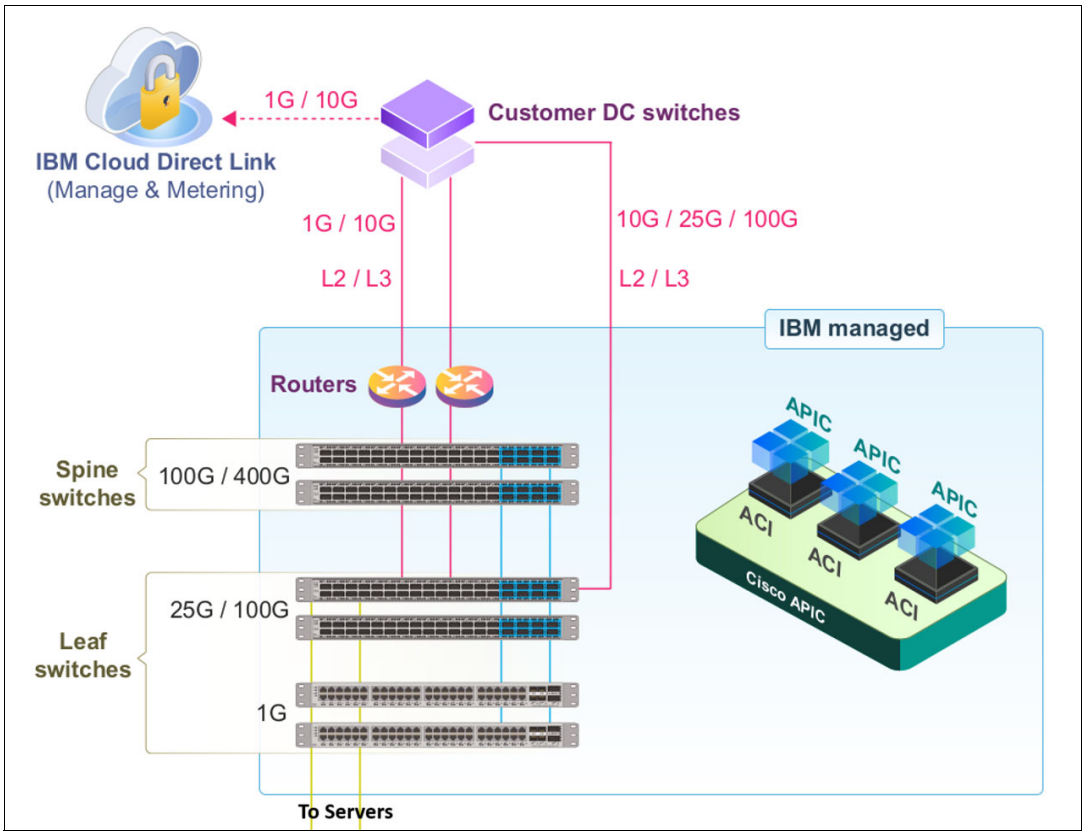


Figure 2-14 External connections for networking

Network External Design

Based on what is shown in Figure 2-14, customers would need to provide IBM with either an IBM Cloud Direct Link or internet connectivity with guaranteed 1gig connection and less than 200ms latency. This connection would be utilized for the control plane that IBM would utilize to maintain the pod.

For connections into the pod for customers we can set up either an L2 or L3 connection connecting to the ASR1k or directly into the ACI Fabric.

2.2.5 OS licensing

For AIX, IBM i, and Linux, OS licenses are included in the cost of the processor cores. IBM Power Virtual Server On-premises provides a complete Red Hat Enterprise Linux (RHEL) offering experience with RHEL stock images. The offering includes support from IBM and access to RHEL bug fixes from Satellite servers that are hosted in IBM Cloud. Currently, you must bring your own licenses for all the other operating system images. For more flexibility, you can always bring your own custom Linux image that is tested and deployed. The AIX stock images are supported on the Power10 with AIX operating system.

IBM i

When you select IBM i, in addition to the IBM i operating system, the following LPP, and IBM i operating system features are included in the IBM Power Virtual Server offering:

- 5770-SS1 IBM i processor and unlimited users
- 5770-SS1: IBM i Option 23, OptiConnect
- 5770-SS1: IBM i Option 44, Encrypted Backup Enablement

- 5770-SS1: IBM i Option 45, Encrypted ASP Enablement
- 5770-SS1 IBM i Option 18 Media & Storage Extensions
- 5770-SS1 IBM i Option 26 Db2 Symmetric Multiprocessing
- 5770-SS1 IBM i Option 27 Db2 Multisystem
- 5770-SS1 IBM i Option 38 PSF for IBM i Any Speed Printer Support
- 5770-SS1 IBM i Option 41 HA Switchable Resources
- 5770-SS1 IBM i Option 42 HA Journal Performance
- 5770-SS1 IBM i Options: all other IBM i Options that are included with IBM i at no additional charge. Examples:
  - Option 8, AFP Compatibility Fonts
  - Option 12, Host Servers
  - Option 13, System Openness Includes
- 5770-DG1: HTTP Server for i
- 5770-JV1: Developer Kit for Java
- 5770-NAE: Network Authentication Enablement for i
- 5733-SC1: Portable Utilities for i
- 5770-TC1: TCP/IP
- 5770-TS1: Transform Services for i
- 5770-UME: Universal Manageability Enablement for i
- 5770-XE1: IBM i Access for Windows operating system
- 5733-ACS: IBM i Access Client Solutions
- 5733-ARE: IBM Administration Runtime Expert
- 5798-FAX: IBM Facsimile Support for i
- 5770-SM1: IBM System Manager for i
- 5770-DFH: IBM CICS® Transaction Server for i
- 5770-MG1: IBM Managed System Services for i
- 5770-AF1: Advanced Function Printing for i
- 5761-AMT: IBM Rational® Application Management Toolset
- 5761-AP1: Advanced DBCS Printer Support
- 5733-B45: AFP Font Collection for i
- 5770-BR1: Backup, Recovery, and Media Services
- 5770-BR2: Backup, Recovery, and Media Services
- 5761-DB1: System/38 Utilities
- 5761-CM1: Communications Utilities
- 5761-DS2: Business Graphics Utility
- 5648-E77: InfoPrint Fonts
- 5769-FN1: AFP DBCS Fonts
- 5769-FNT: AFP Fonts
- 5722-IP1: Infoprint Server for i
- 5770-JS1: Advanced Job Scheduler for i
- 5770-PT1: Performance Tools
- 5770-QU1: Query for i
- 5770-ST1: IBM Db2 Query Manager and SQL Dev Kit for i
- 5733-XT2: XML Toolkit
- 5770-XH2: IBM i Access for web
- 5770-XW1: IBM i Access Family

Each LPP in the package contains all the features, which includes the optional features. For example, the 5770-BR1 solution includes the network feature and the advanced features in addition to the base product.

2.2.6 IBM Cloud catalog

The IBM Cloud Catalog serves as a comprehensive and user-friendly interface that allows businesses to explore, select, and deploy cloud-based services and products. This marketplace simplifies the cloud journey by offering a centralized location where users can quickly find and deploy relevant IBM cloud solutions, including IBM Power Virtual Server Private Cloud. Whether users are looking for compute, storage, or networking solutions, the IBM Cloud Catalog provides access to both IBM-native and partner services.

Figure 2-15 shows a screen shot of the catalog.

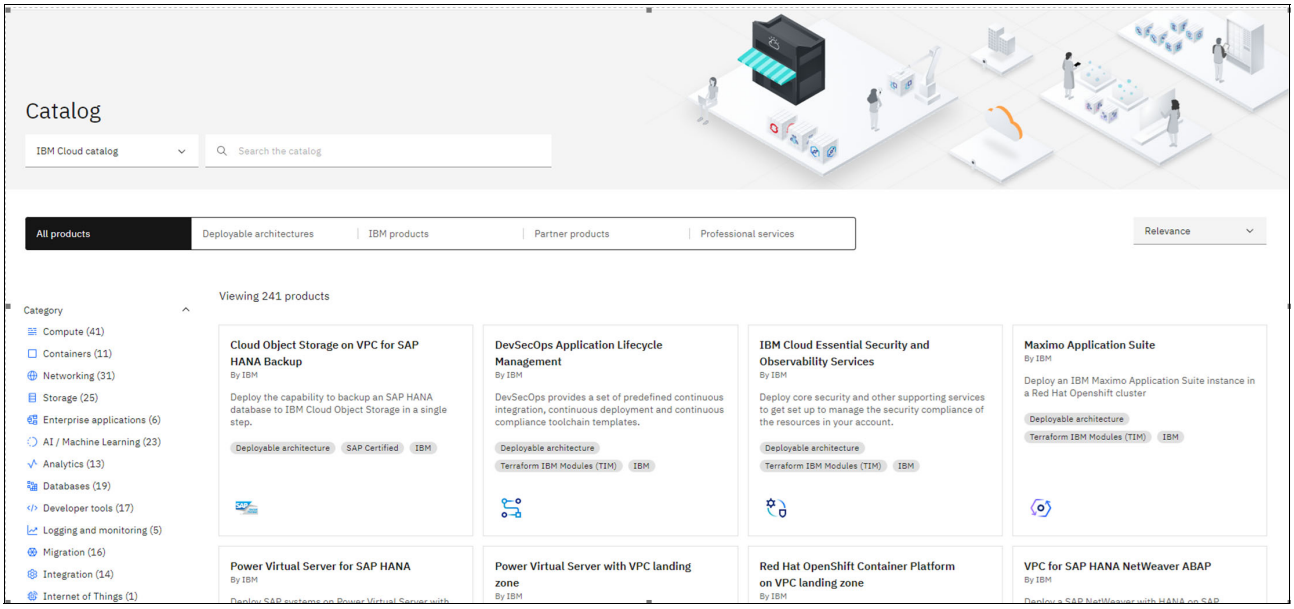


Figure 2-15 IBM Cloud catalog

The catalog also supports a seamless integration of preconfigured templates and custom configurations for cloud services, ensuring that businesses can easily meet their technical and operational needs.

Key Features of IBM Cloud Catalog

The following are some key features and capabilities of using the IBM Cloud Catalog:

- Efficient Search and Discovery

  At the top of the interface, the Search Bar allows users to quickly search for cloud services by entering keywords or product names such as “AI,” “Containers,” or “Storage,” or “Db2” as shown in Figure 2-16.

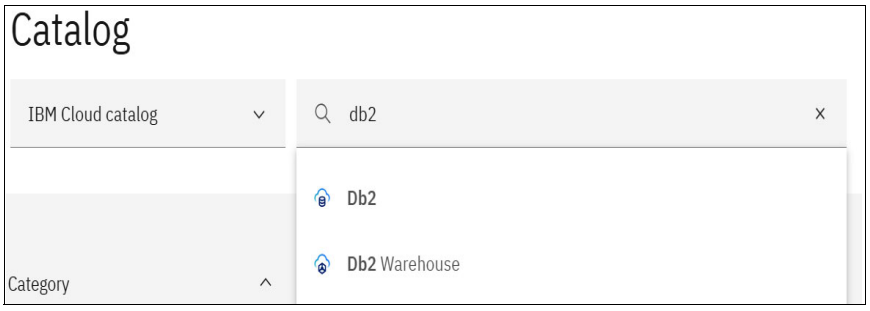


Figure 2-16 Catalog search

The search function supports fast, intuitive access to products, ensuring that users can easily locate relevant services.

- Comprehensive Category Filters

  The Category Sidebar is located on the left-hand side of the interface and offers users the ability to filter their search based on product categories. Users can select categories such as:

  - Compute
  - Networking
  - Storage
  - Enterprise Applications
  - AI and Analytics

  This is shown in Figure 2-17.

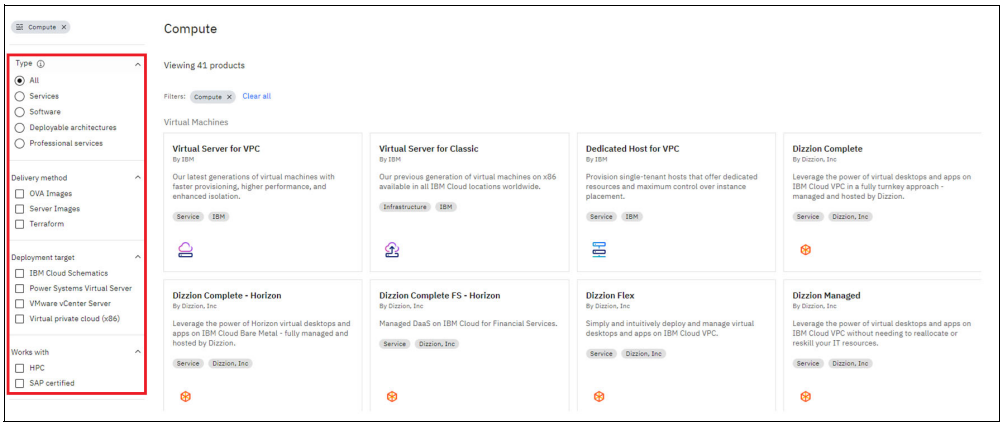


Figure 2-17 Catalog filters

By filtering by category, users can quickly navigate to services that fit their specific business needs.

- Product Display and Detailed Information

  In the center of the interface, a Product Grid is displayed where each cloud offering is displayed as a tile with key details as shown in Figure 2-18.

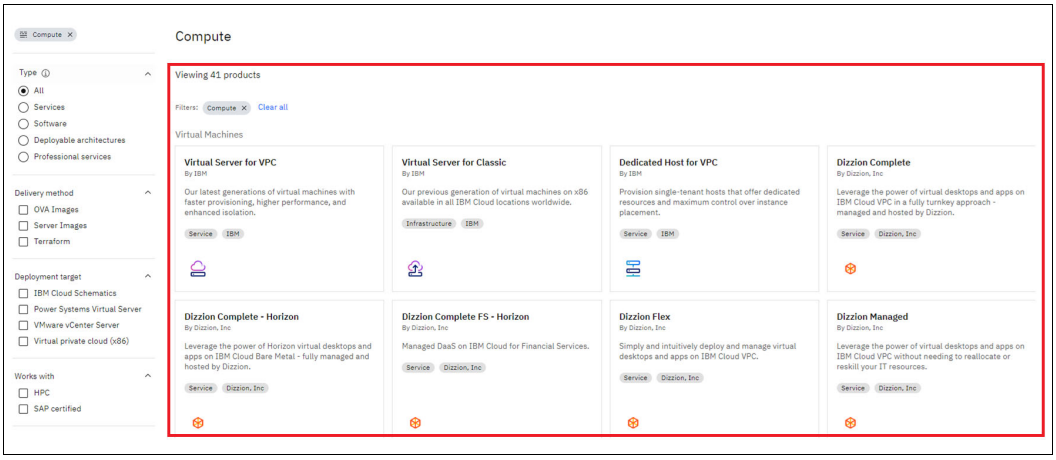


Figure 2-18 Product grid area

Each catalog product is displayed as a tile with information such as:

- Product Name.
- Provider (e.g., IBM or a third-party partner).
- Description: A brief overview of the product's purpose.
- Tags indicating deployment options, certifications and other essential product features.

- Powerful Filtering and Sorting

  The filtering and sorting features allow users to refine their searches further as shown in Figure 2-19.

Figure 2-19 Filter and sort options in the catalog

Choosing a category helps reduce the number of options found and then those options can be sorted based on characteristics such as:

- Price
- Popularity
- Newest releases
- Industry
- Compliance

This ensures that users can focus on products that align with their specific technical requirements or financial constraints.

- Responsive and Flexible Interface

  The IBM Cloud Catalog interface is designed to be responsive, meaning it adapts to different screen sizes and devices. Whether accessing the catalog from a desktop, tablet, or smart phone, users can easily browse, configure, and deploy services.

  This ensures flexibility for teams working remotely or on-site, ensuring cloud resources are available wherever they are needed.

Navigating the IBM Cloud Catalog for Power Virtual Server Services

Figure 2-20 on page 64 illustrates the process to select and implement Power Virtual Server services using the IBM Cloud Catalog.

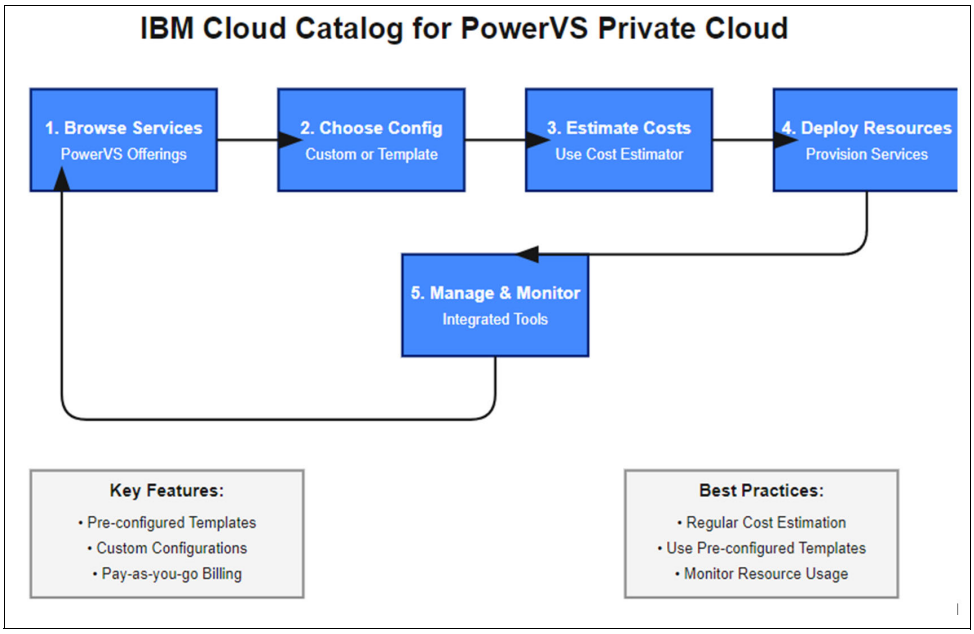


Figure 2-20 Using Cloud Catalog

To use the IBM Cloud Catalog effectively for IBM Power Virtual Server Private Cloud deployment, follow these steps:

1. Browse Services

   Start by selecting Power Virtual Server Private Cloud from the available services in the catalog. This includes products related to compute, storage, and networking. Users can either browse the predefined Power Virtual Server service templates or configure their own.

2. Select Configuration

   Choose from a preconfigured template or build a custom configuration. This flexibility ensures that the services align with the organization's operational, and scalability needs.

3. Estimate Costs

   Use the Cost Estimator to calculate the expenses associated with different configurations. This step is critical for managing cloud budgets and aligning with financial goals.

4. Deploy Services

   Once the configuration and cost estimation are completed, the next step is to deploy the services. The deployment can be automated or manual, depending on the organization's requirements and chosen configurations. Manage and Monitor

   After deployment, IBM Cloud provides tools to monitor resource usage, manage services, and ensure that everything is running smoothly. This includes real-time monitoring dashboards and alert systems to ensure optimal performance.

Recommendations for Using IBM Cloud Catalog

When using cloud services through the IBM Cloud Catalog consider the following:

- Regularly Review Service Costs: Frequently use the Cost Estimator to ensure you're staying within budget.
- Leverage Preconfigured Templates: If you're new to IBM Power Virtual Server, start with a preconfigured template to simplify the deployment process.

- Monitor Resources: Use the monitoring tools provided in IBM Cloud to track your resource usage and performance in real-time.

The IBM Cloud Catalog is a powerful tool that simplifies cloud service selection, configuration, and deployment. For businesses using IBM Power Virtual Server, the catalog offers a seamless interface for managing hybrid cloud infrastructures, ensuring that organizations can scale efficiently while maintaining full control over their environments.

2.3 Pricing concepts

There are two main concepts involved with the pricing of the IBM Power Virtual Server Private Cloud:

- Metered consumption charges

  Consumption charges are based on:

  - Number of processors allocated to virtual servers
    - Pricing varies by processor type (S1022, E1050, E1080)
  - Amount of memory allocated
    - Pricing varies by processor class (scale out - S1022 or scale out- E1050 or E1080)
  - Amount of storage (
    - Pricing based on storage tier allocated.
  - OS licenses

  Consumption billing is based on allocated resources on a monthly basis. This is prorated based on an hourly rate for virtual servers whose configuration is changed during the month.

- Minimum monthly commitment

  The minimum monthly commitment is determined by the configuration installed on your data center floor and the term commitment.

  - Term commitment can be either three-years or five-years. Renewals are available for a one-year term. The minimum monthly commitment is less when choosing the five-year term.
  - Option to terminate after 12 months (with 60 days notice) and pay only 50% of the remaining commitment.

The amount charged per month is the actual metered consumption charges for your defined virtual servers unless that amount is less than the minimum monthly commitment, in which case you are billed the minimum monthly commitment.

- Single monthly bill in arrears.
- Capacity consumption details are available on the IBM Cloud Portal.
- Each client's billing will be different. Use the pricing estimator to model your expected usage.

2.3.1 Total cost of ownership versus cost of acquisition

Most clients have had the building blocks used to create a solution like IBM Power Virtual Server since 2019 but have not invested the development resources and time to create a self-provisioning portal that enables cloud like service delivery. This is millions of dollars of development and integration testing to create, enhance and maintain a "cloud service delivery" capability in addition to the cost of the servers, storage, network, development pipelines and management software.

This requires a large, dedicated team of experts at IBM to design and maintain this IaaS solution that needs to evolve as the technology components and services evolve (follow-on generations of IBM Power, storage evolution, faster networks, quantum-safe encryption, continuous service delivery, and other enhancements).

A recent study by Precision IT[3] examined the total cost of ownership (TCO) for migrating a customer's existing IBM Power servers to an on-premises native cloud solution. The study compared IBM Power Virtual Server Private Cloud to several other vendor solutions for a five-year TCO and found that the IBM Power Virtual Server Private Cloud was the best option.

The study concluded that:

- IBM Power Virtual Server Private Cloud offers a "compelling solution" for clients looking for an on-premises hybrid cloud option offering cloud benefits such as cloud management, OPEX pricing, and scalability with the security and control of an on-site solution.
- IBM Power Virtual Server Private Cloud offers superior scaling to x86 based alternatives while maintaining the advantage of private cloud governance and security.

IBM's new Power Virtual Server Private Cloud requires a 3-5-year commitment to a minimum monthly charge with no upfront capital costs. This approach enables clients to configure the target solution they want and implement a private cloud rapidly with just a monthly amount (the greater of the minimum commitment or actual usage) billed in arrears with a typical 3 year cross-over vs purchasing the elements and developing the cloud services. This allows clients to start small and expand as required by the business, enabling same day delivery of new services compared to months that would be required if using equipment that needs to be ordered and deployed.

2.3.2 Pricing tools

IBM provides an estimating tool which is integrated in the assessment and ordering process. The tool is available to estimate pricing for Power Virtual Server, both off-premises and on-premises. Doing a pricing estimate is the first step in the customer journey to implementing IBM Power Virtual Server Private Cloud as shown in Figure 2-21.

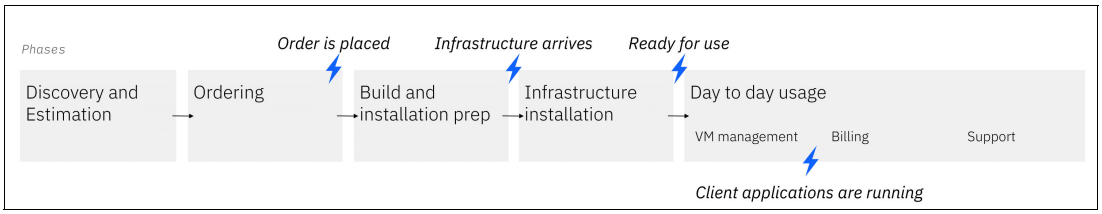


Figure 2-21 Customer journey

The estimator can be found at https://cloud.ibm.com/power/estimate which is shown in Figure 2-22 on page 67.

[3] https://fk2816.p3cdn1.secureserver.net/wp-content/uploads/2024/11/Precision-IT-White-Paper-v10.pdf

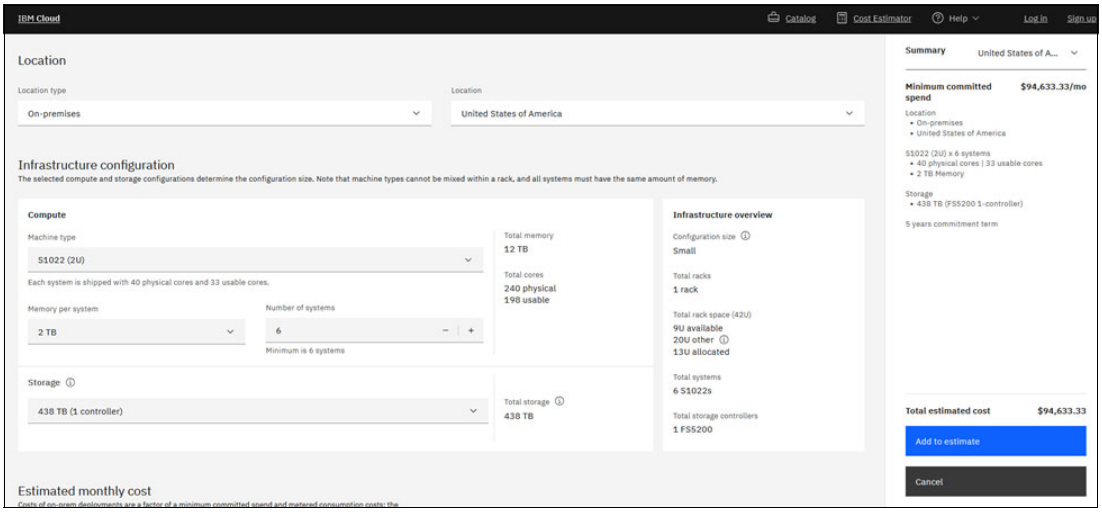


Figure 2-22 IBM Power Virtual Server estimator

When you arrive at the estimator, first choose **"**`On-premises`**"** in the **Location Type** field at the top left of the screen. Then you need to choose the cloud data center that will be the connection point for your on-premises pod from the **Location** drop down menu in the top middle of the screen. Note that the right panel shows the estimated price and the configuration details. This will change as you change the selections in the pricing tool.

Note: When you select **"**`On-premises`**"** you will likely see an error as shown in Figure 2-23. `This` warning is to avoid losing any information you might have entered for an estimate prior to selecting **"**`On-premises`**"**. If you have already saved all previous entries, you may safely select **"**`Change location type`**"**.

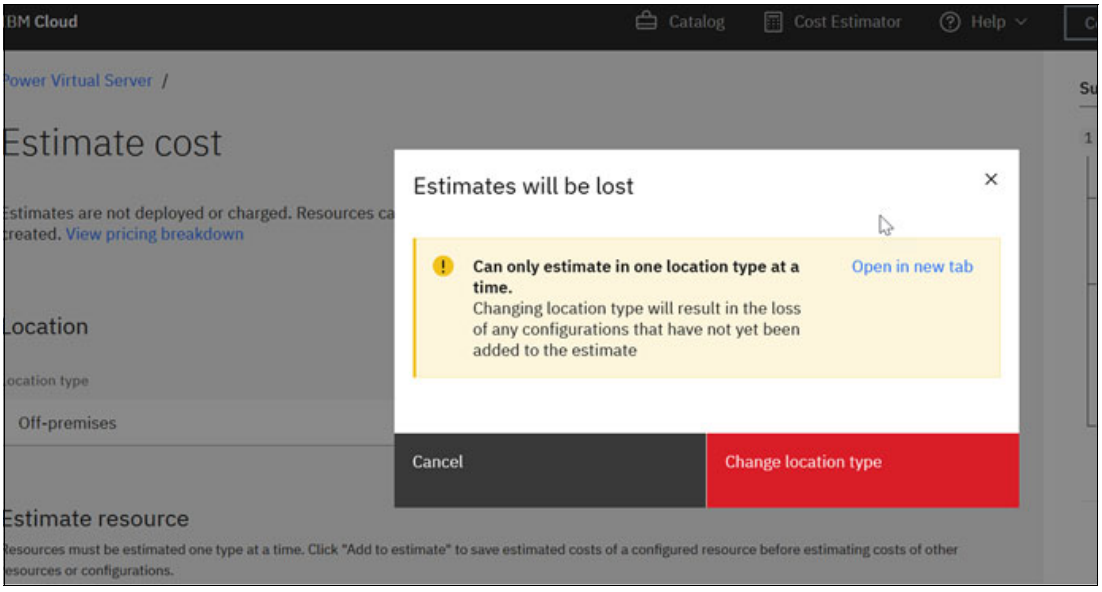


Figure 2-23 Error screen for estimator

Choosing your compute and storage resources

The next step in the process is to choose the compute components you want included in your pod as shown in Figure 2-24 on page 68.

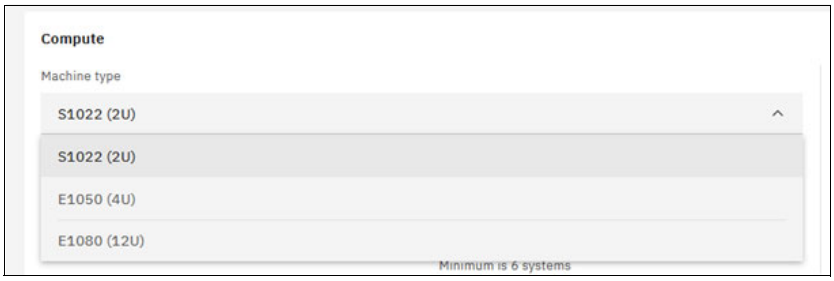


Figure 2-24 System selection

The three choices are the ones described in section 2.2.2, “Compute” on page 47. Choosing an E1080 will automatically start creating a medium pod as the E1080 is only supported in a medium pod.

Once you have chosen the system type, you will need to choose additional details such as the memory and the number of systems you want to include. The tool will identify the minimum number of systems that need to be included based on these selections. If you choose more systems than can be supported in a small Pod, the tool will switch to a medium pod instead. The pricing estimate will automatically be updated as you make different selections.

In addition to the compute resources, the panel allows you to define the storage options you want to configure and will show you the storage capacities available. If you choose the FlashSystem 9500, the tool will automatically change you to a medium Pod, while if you choose the FlashSystem 5200 you will be kept in a small pod configuration. If you select the FlashSystem 5200 when the number of systems included is larger than what is supported in a small Pod, the tool will show you an error and tell you that you need to reduce the number of systems.

Selecting the commitment term

The next panel on the screen is used to define whether you want to price the system with either a 5-year or a 3-year commitment. as shown in Figure 2-25. You can toggle between the two options to see the pricing differences between these options. The commitment term and associated terms are discussed in section 2.3, “Pricing concepts” on page 65.

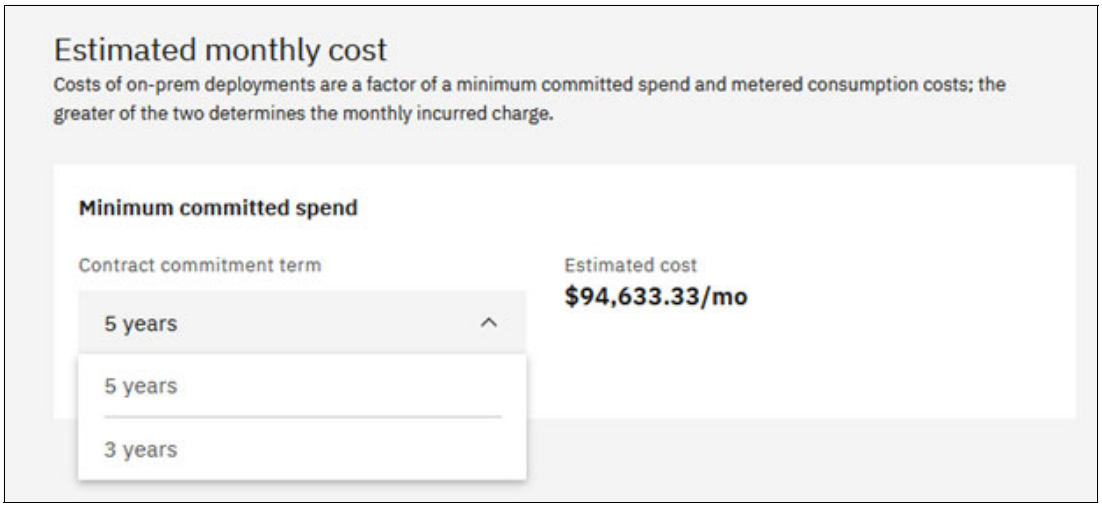


Figure 2-25 Commitment term selection

Metered consumption cost

Pricing for Power Virtual Server Private Cloud is based on consumption with a minimum monthly commitment component. The usage of resources is metered and reported. The client is only charged for consumption of resources that exceeds the minimum committed spend. The pricing tool has a section to allow you to model consumption pricing to help you estimate your billing. Figure 2-26 show the Metered consumption cost estimator which becomes available after you have successfully defined your desired configuration.

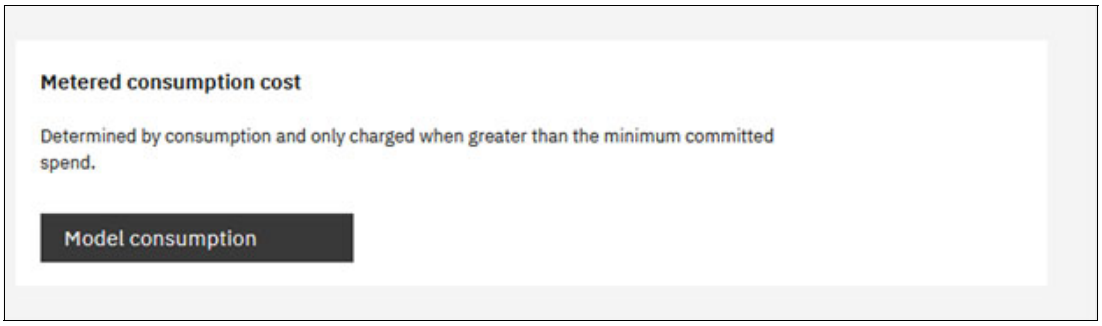


Figure 2-26 Model consumption

When you select the **Model consumption** button, you are presented with another screen which is shown in Figure 2-27. In this screen you can input the expected usage of compute (cores), memory (TB) and storage (TB across the different storage tiers). The screen then shows you the minimum monthly commitment and the expected metered usage that would be billed based on your defined configuration.

Figure 2-27 Modeled metered consumption panel

3

Planning Considerations

A successful implementation of Power Virtual Server Private Cloud on your premises requires careful planning. This chapter provides an insight into some of the considerations that you need to plan for as you decide to implement an on-premises Power Virtual Server pod. Your IBM team will be working with you through each of these considerations in order to make the implementation successful and ensure that the solution meets your requirements.

The following topics are discussed.

3.1 Planning overview

Planning to utilize IBM Power Virtual Server for your power workloads requires some planning whether you are planning on using Power Virtual Server off-premises or Power Virtual Server Private Cloud (on-premises). Some of the same tasks need to be considered in both choices, for example:

- Determining network connectivity.
- Deciding which applications need to be migrated.
- Sizing your Power Virtual Server virtual servers to match the workload.
- Choosing how to migrate the workloads.
- Determining backup and recovery, and disaster recovery requirements.

The major differentiation between the two options is that the on-premises option requires a more precise sizing to ensure that you choose the right starting configuration to provide the best ROI for your environment. Choosing a pod configuration that is too large will mean that you are paying more for the solution than required. Choosing a pod that is too small, will limit your expansion

Figure 3-1 provides an overview of the planning process that will be used to:

- Size your workload.
- Choose your configuration.
- Determine the cost of the solution.
- Validate that you have the appropriate data center resources to install the solution.
- Have the appropriate network connectivity to support the solution.

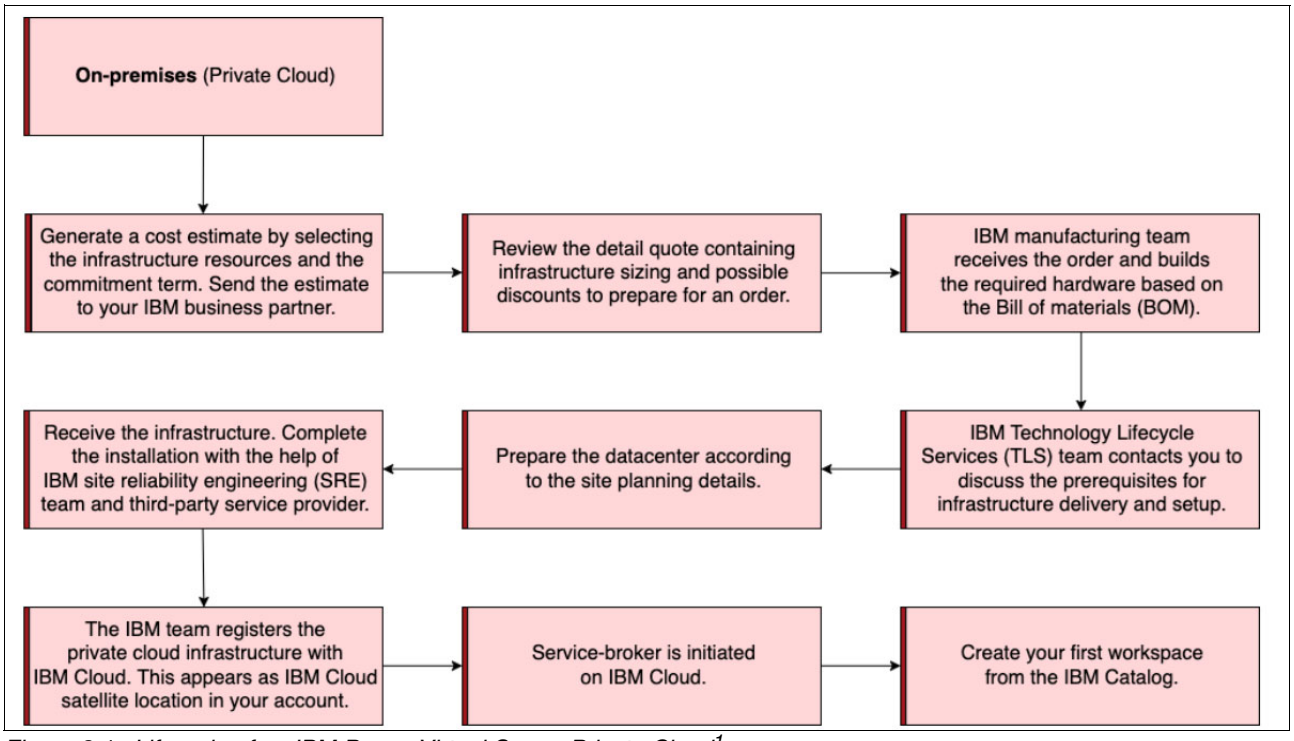


Figure 3-1 Lifecycle of an IBM Power Virtual Server Private Cloud[1]

IBM will work side by side with your team to ensure that your implementation is successful.

[1] https://cloud.ibm.com/docs/allowlist/power-iaas?topic=power-iaas-getting-started

3.2 Prerequisites

Before choosing to implement Power Virtual Server Private Cloud in your data center, ensure that the following prerequisites are met.

3.2.1 Basic Prerequisites

Power Virtual Server Private Cloud is an IBM owned infrastructure installed and operated by IBM in a client data center.

- The client is responsible for maintaining the datacenter
  - Floor space
  - HVAC
  - Access security
- The client needs to provide access to IBM service technicians to install and maintain the servers and storage.
- The client needs to provide network connectivity:
  - Connection to the IBM Cloud data center for control and management.
  - Connection between the pod networking to the client networking for customer data and workflow.

  For more information on the networking connections see section 2.2.4, “Network” on page 58.

3.2.2 Minimum size

Ensure that the IBM Power Virtual Server Private Cloud configuration is not oversized for the workloads that are moving to the offering. If you have multiple smaller workloads, you might want to consolidate them into a single pod to provide a better ROI. Consider the minimum configuration for a pod as shown below.

Small pod

The minimum compute configuration of a small pod is:

- 6 S1022(40c) with 2 TB memory each, or
- 5 S1022(40c) with 4 TB of memory each, or
- 2 E1050(96c) with 8 TB of memory each

Medium pod

The minimum compute configuration of a medium pod is:

- 12 S1022(40c) with 2 TB or 4 TB memory each, or
- 5 E1050(96c) with 4TB or 8TB each, or
- 2 E1080(2CEC-120c) with 8TB or 16TB or 32TB each

3.2.3 Supported Workloads and Operating Systems

IBM Power Virtual Server Private Cloud utilizes IBM Power10 servers. The workloads you want to move to Power Virtual Server Private Cloud must be able to run on IBM Power10 servers and be hosted by a supported operating system.

IBM Power Virtual Server Private Cloud is an excellent choice for enterprise applications such as Oracle and other databases.

The following operating systems are supported by the IBM Power Virtual Server Private Cloud offering. If you are running on previous versions of these operating systems, you need to upgrade to supported versions.

- AIX

  The minimum version of AIX supported on the supported servers in the IBM Power Virtual Server Private Cloud is:

  - S1022 - 7.2 or later
  - E1050 - 7.2 or later
  - E1080 - 7.2 or later

  The following stock images are available when you create a virtual machine:

  - AIX 7.3 TL1 SP2
  - AIX 7.2 TL5 SP6
  - AIX 7.2 TL5 SP8
  - AIX 7.3 TL2 SP1

- IBM i

  The IBM Power Virtual Server Private Cloud (On-premises) supports IBM i 7.3, or later.

  IBM i stock images currently available when you create a VM are:

  - IBM i 7.5 TR4
  - IBM i 7.5 TR3
  - IBM i 7.4 TR10
  - IBM i 7.4 TR9
  - IBM i 7.3 TR13

- Linux

  The IBM Power Virtual Server (On-premises) supports Red Hat Enterprise Linux (RHEL) with RHEL stock images that includes support from IBM and access to RHEL bug fixes from Satellite servers hosted on IBM Cloud. This capability is referred to as the Full Linux Subscription (FLS) model. For more information, see Full Linux subscription for IBM Power Virtual Server (On-premises).

  FLS provides access to RHEL OS fixes and updates through activation keys for Power servers, which are hosted on an IBM satellite server within the IBM Cloud environment. To register for FLS, select one of the stock (RHEL OS) images that are provided by IBM.

  The following list is an example of the FLS offerings:

  - Stock images: RHEL 8.4, RHEL 8.6, RHEL 9.2
  - Support: You pay IBM for support
  - Patches: You receive keys for satellite servers to obtain Linux patches from Linux distribution (Linux distros)

3.2.4 Proximity to IBM Cloud data center

The client data center used for IBM Power Virtual Server Private Cloud (see list of supported data centers below) needs to be close enough to the IBM Cloud data centers to provide a low latency connection between the data centers. The network latency between the client data center and IBM needs to be less than 200 ms round trip delay.

The following IBM Cloud regions can host connections from the Pods for IBM Power Virtual Server (On-premises) in your data center. Selection of the IBM Cloud region is one of the factors for computing pricing. Select the IBM Cloud region that is closest to the physical location of your data center. Choose from:

- Dallas
- Frankfurt
- London
- Madrid
- Osaka
- Sao Paulo
- Sydney
- Tokyo
- Toronto
- Washington, DC

Important: The network latency between your data center and the selected IBM Cloud region must maintain a network round-trip time (RTT) of less than or equal to 200 milliseconds.

3.2.5 Countries supported

At the time of publication, IBM Power Virtual Server private cloud is only available in the following countries.

- United States
- Canada
- Netherlands
- Denmark
- Austria
- Germany
- Belgium
- Luxembourg
- Switzerland
- United Kingdom
- Australia

3.3 Application assessment

Prior to moving workloads to IBM Power Virtual Server Private Cloud, you should do an assessment of the application environment to ensure that you can successfully migrate those workloads. some of the things that need to be evaluated are:

- Is the workload currently running on IBM Power?

  If the workloads are currently running on IBM Power, then the process of migration is greatly simplified. While it is possible to migrate workloads running on x86 platforms to IBM Power, it does take additional planning.

- Is the workload running on a supported operating system?

  Migrating a workload from an operating system that is older and not supported on IBM Power Virtual Server requires that you update to a supported operating system version. In some cases, this also means updating to newer versions of middleware products such as the database manager.

- Is the workload a standalone application or is it split between different systems that require interconnections and data flow?

  If your application requires connections to servers that are not running on IBM Power, you need to plan for the continued connectivity to those resources after the migration.

- What are the availability and disaster recover requirements for the workload?

  IBM Power Virtual Server provides many capabilities for workloads that require high availability, such as automatic failover and data replications. You need to plan appropriately to support these requirements.

  For example, if you need disaster recovery, you may want to utilize Power Virtual Server Private Cloud in a second data center, or consider utilizing Power Virtual Server off-premises for recovering the workloads. The key is that these solutions require special planning to be successful.

These are just some examples of the things you need to consider. IBM has several offerings available to help you build a migration plan. For more information see IBM Technology Expert Labs.

3.4 Sizing

Sizing your workload to determine the configuration of your Power Virtual Server Virtual Server Instance consists of two distinct components: sizing the compute requirements (CPU and memory) and sizing the storage requirements.

3.4.1 Compute sizing

Collecting data from your existing IBM Power servers will provide you with the information required to correctly size your Power Virtual Server environment. Data can be gathered using tools such as NMON and TOP or you can use scan reports from your existing Hardware Management Controllers (HMCs) to get the information on the existing CPU utilization and memory allocation.

After choosing the size of your LPAR, monitor the performance and adjust as necessary. It is possible to dynamically Scale LPAR cores and memory on the fly from 0.25x-8x.

CPU Cores

When sizing for CPU, size for typical LPAR processor utilization, not the maximum allocation. Remember to not include any Virtual I/O Server (VIOS) LPARs in your sizing. Sizing on Power Virtual Server often saves from 25%-50% of CPU compared to your existing on-premises utilization as Power Virtual Server doesn't need the control plane and capacity headroom.

If you are not currently running on Power10 based servers, then you will need to convert the CPU numbers on your existing IBM Power technology to the IBM Power10. You can use a loose estimate.

Optimize Cores by LPAR loosely based on the numbers in Table 3-1.

Table 3-1 Estimated core conversion factors

Existing technology	Power10 technology
1 IBM Power7 core	0.3 Power10 core
1 IBM Power8® core	0.5 Power10 core
1 IBM Power9 core	0.75 Power10 core

Processor type

Processors can be defined as shared uncapped, shared capped or dedicated. For most LPARs, shared uncapped cores are the appropriate choice. Use shared capped cores for compliance or ISV license purposes, and use dedicated cores only as required based on usage in your private on-premises environment.

Memory sizing

Memory is charged on allocated memory at the LPAR, not on the memory utilization shown by the operating system. In general, the memory size for your Power Virtual Server LPAR will be the same as your current implementation. Reminder that if you allocate more than 64 GB of memory per core, then there is an additional charge for the memory. Avoid this 1.5x memory premium if possible.

Choosing the processor model

Consider using the Power S1022 unless you need additional cores, memory or performance of the Power E1050 or Power E1080. IBM i on the Power S1022 only supports up to four cores. For partitions larger than four cores, use the Power E1080.

3.4.2 Storage sizing

There are two components in sizing your storage configuration: amount of storage and the storage performance requirement or tier.

Storage Tier Selection Guidance

Consider starting with tier 3 storage. All of the storage is IBM FlashSystem, so the difference between the tiers is the number of IOPs supported. Storage tiers are discussed in “Storage tiers” on page 55. Tier 0 supports the most IOPS/GB, followed by Tier 1 and then Tier 3.

- Use Tier 3 storage for non-production LPARS, application servers, web servers, and other low IOPS workloads.
- Use Tier 1 storage for production, mission-critical servers, or databases.
- Use Tier 0 for high-performance databases.
- Fixed 5K IOPS is for databases with high-performance requirements that are <200 GB.

Storage tiers can be mixed as appropriate. The storage tier of a LUN can be dynamically changed.

Storage sizing

Right-size storage volume sizes based on utilization. Aim for 70-75% utilization of allocated storage on your volumes. This means that if you have a volume allocated at 20 TB with only 30% utilization, then it can be optimized to 10TB.

Cloud Object Storage

Cloud Object Storage (COS) is not a primary storage for Power Virtual Server. Instead, use COS for:

- Archive
- Long-term storage and Backup repository
- Capturing and exporting a virtual machine (VM)
- Importing a boot image

3.4.3 Scalability and flexibility

Scaling is the ability to dynamically adjust computing resources such as CPU, memory, storage, and network capacity in response to changing workloads. With IBM Power Virtual Server, either off-premises in the public cloud or on-premises in Private Cloud, businesses can scale their infrastructure up or down based on demand, ensuring that they maintain optimal performance while avoiding unnecessary costs.

IBM Power Virtual Server offers both vertical scaling where resources within an existing virtual machine are scaled up or down, and horizontal scaling where virtual machines are added to distribute workloads and later removed when the workload requirements go down. This flexibility allows organizations to handle fluctuating workloads efficiently without needing to over-provision resources.

Vertical Scaling

Vertical scaling in Power Virtual Server involves increasing or decreasing the resources allocated to an individual Logical Partition (LPAR). This can include scaling the number of processors, the amount of memory, or even increasing storage capacity to handle larger workloads.

Vertical scaling is useful when a single application or LPAR requires more resources to process higher workloads, such as during peak usage periods for transactional systems or data-intensive applications.

For example, a retail company running an inventory management system on an AIX LPAR might increase the processing units and memory allocation during the holiday season when transactional data is significantly higher, ensuring smooth and fast operations without downtime.

Horizontal Scaling

Horizontal scaling in Power Virtual Server involves creating additional LPARs to distribute the load across multiple instances. Instead of increasing resources on a single LPAR, horizontal scaling adds more LPARs to share the workload, improving redundancy and fault tolerance.

Horizontal scaling is ideal for clustered applications, distributed databases, and other systems where workloads can be spread across multiple LPARs for better performance and reliability.

For example, a financial services company may create multiple LPARs running instances of their trading application in Power Virtual Server to handle increased market activity during high-volume trading periods. By adding more LPARs, they can distribute the processing load across these instances, improving overall system performance and preventing any single point of failure.

Scaling in Power Virtual Server

In Power Virtual Server, resources are allocated to LPARs from a shared pool, allowing for flexible and dynamic scaling. Note that:

- The hypervisor (PowerVM) manages the allocation of physical resources to LPARs, enabling efficient use of the underlying hardware.
- Power Virtual Server supports both dedicated and shared processor LPARs, offering flexibility in how processing resources are allocated.
- Memory can be dynamically added to LPARs in Power Virtual Server, allowing for on-the-fly vertical scaling of memory resources.

Some of the key benefits of Power Virtual Server Private Cloud are:

- Dynamic Provisioning

  Dynamic provisioning allows businesses to add or reduce resources in real-time without needing to shut down virtual machines. This flexibility helps companies respond to sudden workload spikes without causing downtime or disruption to services.

  Power Virtual Server can automatically scale workloads using preconfigured policies. For example, resources can be provisioned automatically when CPU utilization reaches a certain threshold, ensuring workloads always have the necessary resources.

- Pay-as-You-Go Flexibility

  IBM Power Virtual Server Private Cloud offers a pay-as-you-go model, which ensures that businesses only pay for the resources they use. Scaling resources dynamically allows organizations to avoid the costs associated with over-provisioning hardware for worst-case scenarios.

  This model is particularly valuable for seasonal businesses or those with variable workloads, such as retail during holiday seasons or financial services during tax season.

- Resource Monitoring and Management

  IBM Power Virtual Server includes integrated monitoring tools that allow businesses to track resource usage across their VMs. These tools provide insights into CPU, memory, and storage utilization, helping IT teams make informed decisions about scaling.

  IBM Cloud Monitoring integrates seamlessly with Power Virtual Server environments, providing real-time metrics and enabling automated alerts when resources need to be scaled.

- Hybrid Cloud Scaling

  For organizations using a hybrid cloud strategy, Power Virtual Server supports seamless scaling between on-premises and cloud environments. Workloads can be moved to the cloud during periods of high demand and scaled back to on-premises infrastructure when demand decreases.

  This approach is particularly useful for businesses that need to meet regulatory requirements for data residency while still enjoying the flexibility of cloud-based scaling.

Scaling Strategies in Power Virtual Server Private Cloud

There are multiple strategies for scaling infrastructure in Power Virtual Server to adjust to expanding workloads.

Proactive scaling involves planning for predictable increases in demand. Businesses can schedule increases in resources based on known events, such as planned marketing campaigns, product launches, or seasonal spikes. An example of proactive scaling is when an e-commerce company increases VM capacity ahead of a major sales event to ensure the site can handle the expected surge in traffic without performance degradation.

Reactive scaling allows businesses to respond to unexpected spikes in workload. Power Virtual Server can automatically provision additional resources based on preset thresholds, such as CPU or memory utilization exceeding 80%. As an example of reactive scaling consider a media company which is streaming a live sports event. The company might experience an unexpected surge in viewers. Power Virtual Server can automatically add more VMs to handle the load, ensuring uninterrupted service for viewers.

Auto-scaling in Power Virtual Server enables automatic adjustment of resources based on real-time demand. This means that workloads can be dynamically scaled without manual intervention, ensuring optimal performance and cost-efficiency. A cloud-based gaming company uses auto-scaling to manage fluctuating player numbers, automatically increasing server capacity during peak gaming hours and reducing it during off-peak times.

Effectively managing resources

As described previously, Power Virtual Server provides the ability to dynamically adjust the resources assigned to an application to enable a business to scale up resources during high demand events and later scale back down to maintain a cost-effective solution. To be effective at managing resource utilization to meet business requirements, an enterprise needs to:

- Monitor Resource Utilization

  Use Power Virtual Server monitoring tools to regularly assess resource utilization and performance metrics. Set up alerts and automated scaling triggers to ensure that you are not over-provisioning or under-provisioning resources.

- Implement Scaling Policies

  Define clear scaling policies that align with your business objectives. For example, set thresholds for when to scale up or down, and determine whether to prioritize vertical or horizontal scaling based on the type of workload.

- Test Scaling Scenarios

  Regularly test your scaling strategies by simulating high-load conditions. This will help you identify potential bottlenecks or issues with resource allocation, allowing you to fine-tune your scaling policies.

- Cost Management

  Continuously review the costs associated with scaling. Make use of the IBM Cloud Cost Estimator to track resource usage and forecast expenses based on your scaling needs.

Conclusion

Scaling in IBM Power Virtual Server Private Cloud offers businesses the flexibility to dynamically adjust resources based on workload demands, ensuring that applications run efficiently while maintaining cost control. Whether scaling vertically to enhance the performance of individual VMs or horizontally to distribute workloads across multiple instances, Power Virtual Server provides the tools needed for seamless scalability. By leveraging dynamic provisioning, proactive and reactive scaling strategies, and auto-scaling, businesses can optimize their cloud environments for performance, reliability, and cost-efficiency.

3.5 Network connectivity

Establishing a connectivity between the pod in your data center and IBM Cloud requires planning, site preparation, and understanding the network architecture. You also need to ensure that you have met the prerequisites in order to be successful.

3.5.1 Network prerequisites

To facilitate the IBM Power Virtual Server pod infrastructure connectivity, you must evaluate the following network requirements:

- The data center site must provide network cables to connect the IBM Power Virtual Server (On-premises) network infrastructure and the data network at the site.
- The site must provide two uplink cables to connect the IBM Power Virtual Server (On-premises) network infrastructure to the IBM Cloud region through IBM Direct Link connections or through VPN connections.
- Contract with a service provider to:
  - Provide redundant connections to the IBM Direct Link connection or VPN connection.
  - Provide the last mile connection from the point-of-presence (PoP) of your service provider to the customer data center.

3.5.2 Network architecture

Figure 3-2 provides an overview of the network architecture used to connect to the client data center and your Power Virtual Server pod.

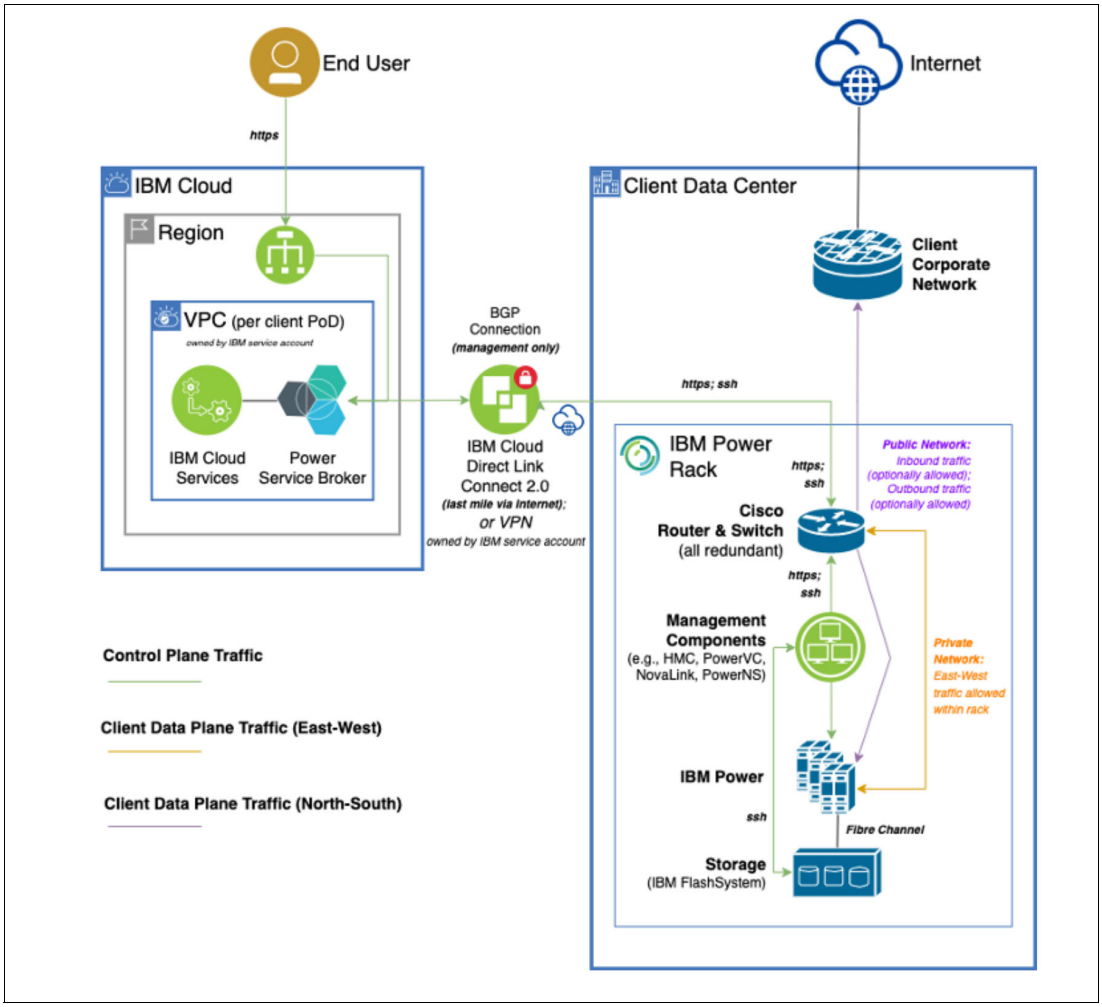


Figure 3-2 Network architecture

Note that there are two networks of interest:

1. Control plane network

   The control plane network is used for communication between IBM Cloud and the pod that is located in your data center. This is used only for setting up and controlling the operations of the virtual servers running in the pod. No user data is transmitted over this network.

2. Data plane network

   The data plane network is used when clients are accessing the applications running on the virtual servers or for server-to-server communication between the different servers and services running in the pod and in your business network.

Control plane network

The control plane network is crucial for communication between your local IBM Power Virtual Server data center (the Pod) and the IBM Cloud. It enables the Service Broker to manage virtual machines effectively by starting APIs on PowerVC. To ensure reliable connectivity, this network includes multiple paths and can be established through either an IBM Cloud Direct Link 2.0 or a Virtual Private Network (VPN).

Before installing the Pod, make sure to provide the necessary network-specific information to facilitate the setup of either the IBM Cloud Direct Link 2.0 or VPN connection.

IBM Cloud Direct Link 2.0 Connect

IBM Cloud Direct Link 2.0 Connect is a technology designed to connect remote data centers with the IBM Cloud. It facilitates the linking of a Virtual Private Cloud (VPC) on IBM Cloud to a router connected to your pod in your data center. Using IPsec over VPN, also known as last-mile connectivity, this solution ensures a secure connection to the Pod's router while minimizing data center environment exposure.

The control plane network is configured for redundancy, providing multiple pathways to maintain reliable connectivity between the Pods and IBM Cloud regions. IBM facilitates the setup of the IBM Cloud Direct Link 2.0 Connect using details provided for third-party network providers (like Megaport or Equinix). Attributes such as service keys are collected during the ordering process, enabling third-party providers to establish remote connectivity between the IBM Cloud and your data center.

IBM Cloud Direct Link 2.0 Connect serves as an alternative to traditional site-to-site VPN solutions, offering enhanced security, privacy, and consistent high-throughput connectivity between your remote network and IBM Cloud environments. For more information, refer to the Getting Started with IBM Cloud Direct Link guide.

The connectivity between IBM Cloud and the pod consists of two main components:

1. Shared Connectivity

   This involves the integration of IBM Cloud network infrastructure with the third-party network provider infrastructure.

2. Internet-Based Connectivity

   Often referred to as “last-mile connectivity,” this connection typically utilizes an IPsec VPN. You'll need to contract a service provider to establish this last-mile link.

Establishing this connectivity requires collaboration between you, IBM, and the third-party network provider.

Virtual Private Network (VPN)

VPN connections between IBM Cloud and the pod can be established in multiple ways:

- Site-to-site VPN connectivity

  A VPN gateway can be created on the VPC housing the service broker, which will provide an internet-facing IP address. You can also set up a VPN client in your data center with its own internet-facing IP address. This creates an IPsec VPN tunnel between the two endpoints, allowing the VPN client to extend connectivity to the IBM Power Virtual Server (on-premises) routers.

- VPN connectivity using IBM Cloud Classic Environment

  You can deploy a Juniper vSRX Virtual Firewall in the classic environment of IBM Cloud. This firewall can connect to the VPC containing the service broker via a transit gateway. Alternatively, a similar VPN setup can be configured in your data center, linking through an IPsec VPN tunnel to the IBM Power Virtual Server (on-premises) routers. It's advisable to keep your data center infrastructure's connection to the pod routers minimal. You may configure one or more virtual routers in your environment and establish a Border Gateway Protocol (BGP) deployment using the IBM Power Virtual Server (on-premises) routers.

Data plane network

The data plane network becomes active when the Pods at your data center are linked to your local IBM Power Virtual Server network. This allows you to access the virtual servers inside the pod directly through your own network, not via IBM Cloud. Each pod includes essential software components like HMC and PowerVC, as well as necessary storage and network devices such as ACI and routers, enabling it to connect with IBM Cloud.

Important: When you create a network within a Pod, ensure that the network does not overlap with other existing networks in the same workspace within the pod. If you create an overlapping network, an error message is displayed.

As part of the network planning, you can review the following use cases and identify the use cases that are applicable to you. You can communicate about such requirements before the installation so that you do not have to open separate support tickets to implement the use-cases and configurations.

1. Private network within a Pod

   With this use case, you can establish a private network within a pod that allows communication between the applications that are located in the pod. You can establish a private network within the pod by using the IP address allocation method, Classless Inter-Domain Routing (CIDR). CIDR allows network routers to route data packets to the respective device based on the indicated subnet.

   You can deploy virtual machines in a pod that has a default configuration by using one of the following patterns:

   - Affinity: In this pattern, virtual machines are deployed on the same physical host. Therefore, the virtual machines can communicate with each other on the same host through the attached Ethernet switch.
   - Anti-affinity: In this pattern, virtual machines are deployed on different physical hosts. A custom configuration is required on the externally connected Ethernet switch to enable communication between virtual machines that are deployed on different physical hosts.

   As an example, if you have a database server and a web server that need to communicate exclusively with each other. You can connect both servers to the same private network to enable communication between them.

Figure 3-3 describes the private network within a pod type of network setup.

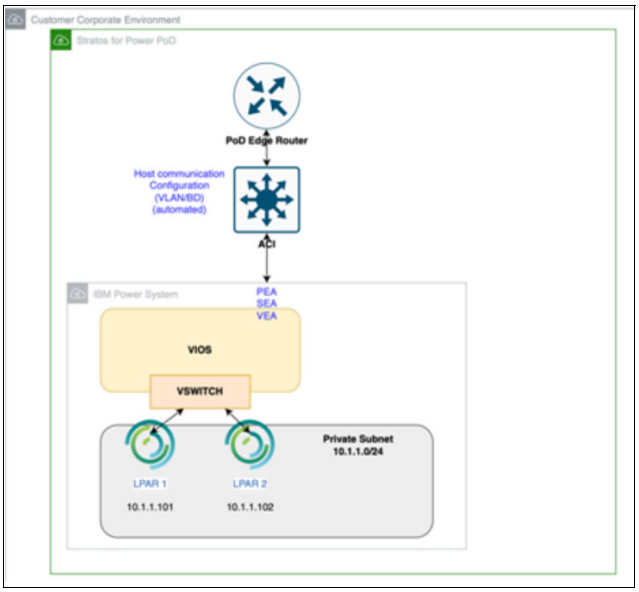


Figure 3-3 Private network within a cloud use case

2. Outbound-only

   With this use case, you can establish a private network that allows communication between applications within the pod and with external destination points. However, the applications within the pod are not accessible from the destination points on the external network. You can establish an outbound-only network connectivity through dynamic Network Address Translation (NAT) gateway configuration, resembling a network established by using routers. Figure 3-4 shows this use case.

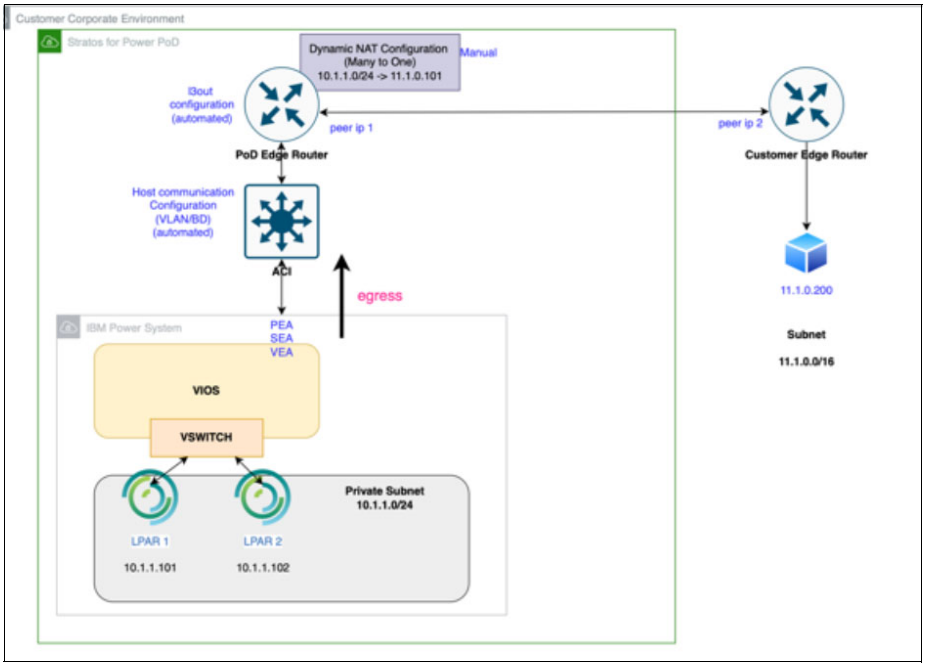


Figure 3-4 Outbound-only use case

3. Bidirectional external connectivity through BGP

   With this use case, you can establish a network that allows communication between the applications within the pod and with the destination points on the external network. By setting up the Layer 3 Firewall rules, you can allow both inbound and outbound connections. Configure the Border Gateway Protocol (BGP) manually between the pod router and the corporate network. By using the BGP configuration, establish a connection between the private network and the corporate network. To configure BGP manually, contact the Support Center.

   As an example of this use case, there is a database server that is running inside the pod that needs to connect to a database server from another application that resides outside the Pod, but within your corporate network. Using Layer 3 inbound access, you can route the traffic between the servers securely utilizing corporate firewalls or routing rules as the corporate network can access the pod subnets through a BGP connection. Figure 3-5 describes the bidirectional external connectivity using BGP.

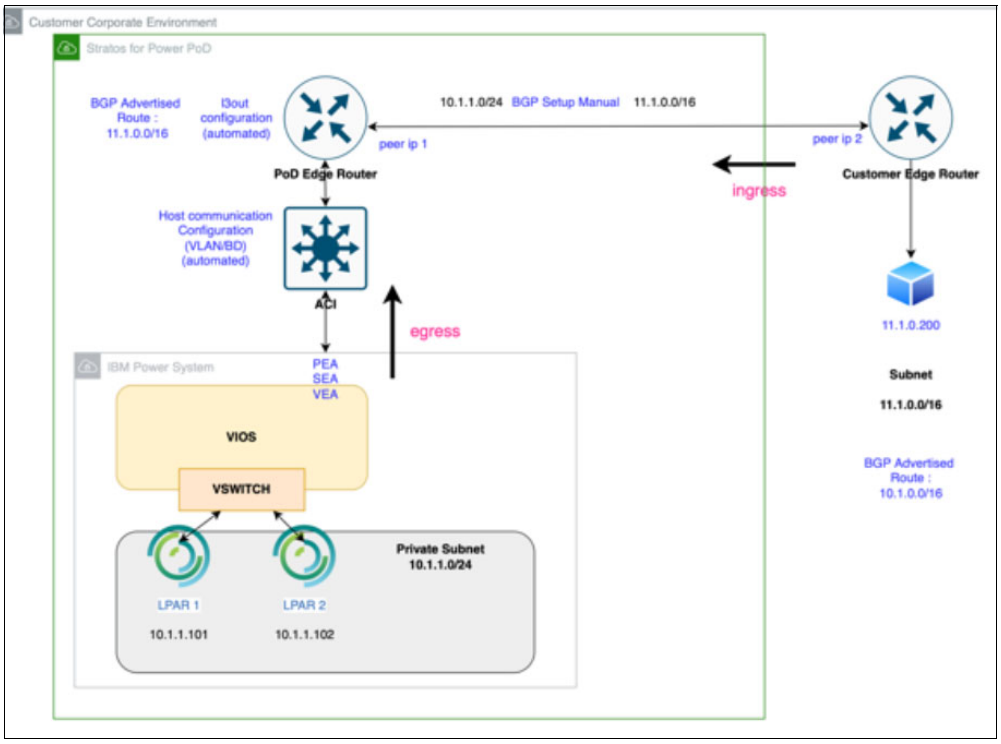


Figure 3-5 Bi-directional access with BGP protocol

4. Bidirectional external connectivity through static routes

   This is a variation of option 3 “Bidirectional external connectivity through BGP” on page 85 but this case uses static routes instead of BGP. Routing still occurs at Layer 3, but a static route is defined between the edge routers, one in the pod and the other in the corporate network. The static route establishes a connection between the pod subnet and the corporate network.

Figure 3-6 describes the bidirectional external connectivity with a static route type of network setup.

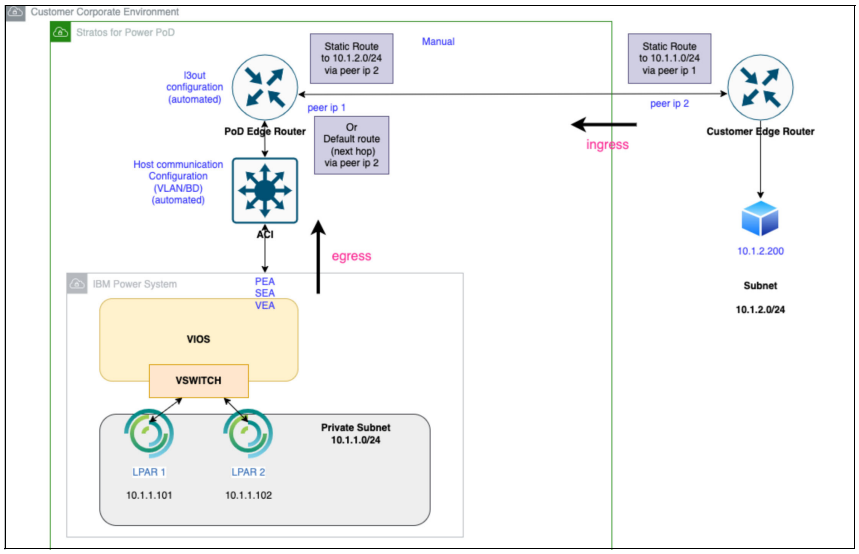


Figure 3-6 Bi-directional connectivity through static routes

5. Bidirectional external connectivity - ACI Layer 2

   This use case uses a Layer 2 network connection between applications within the pod and with destination points on the external network. By connecting to one of your existing corporate networks in this case you bypass the router and connect to the Cisco Application Centric Infrastructure (Cisco ACI). You can establish this type of connectivity when you want the same IP address space on both internal and external networks. All other external network use cases involve two distinct subnets. Figure 3-7 describes the bidirectional external connectivity by using ACI Layer 2 firewall type of network setup.

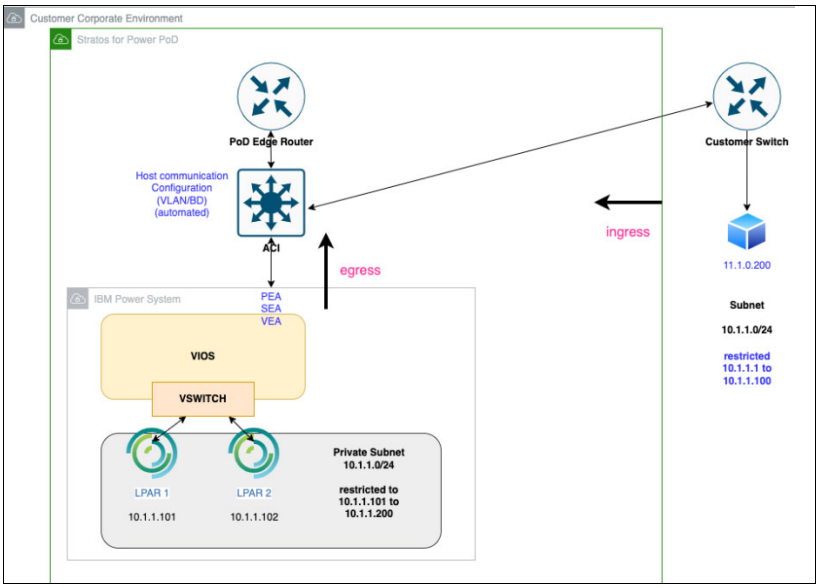


Figure 3-7 Bi-directional connectivity using ACI Layer 2

6. Network connectivity for full Linux subscription

   With this use case, you can establish a network between a virtual machine in the pod and the Red Hat Satellite server on IBM Cloud. The virtual machine has the stock image of Red Hat Enterprise Linux (RHEL) with full Linux subscription. Connect the virtual machine in the pod to a proxy network in the corporate network environment and connect the proxy network to the Red Hat Satellite server on IBM Cloud by using either Direct Link or VPN connection. The virtual machine in the pod can access the Linux satellite server to retrieve software fixes and other artifacts. Figure 3-8describes the network connectivity between a virtual machine and a Red Hat Satellite server on IBM Cloud setup.

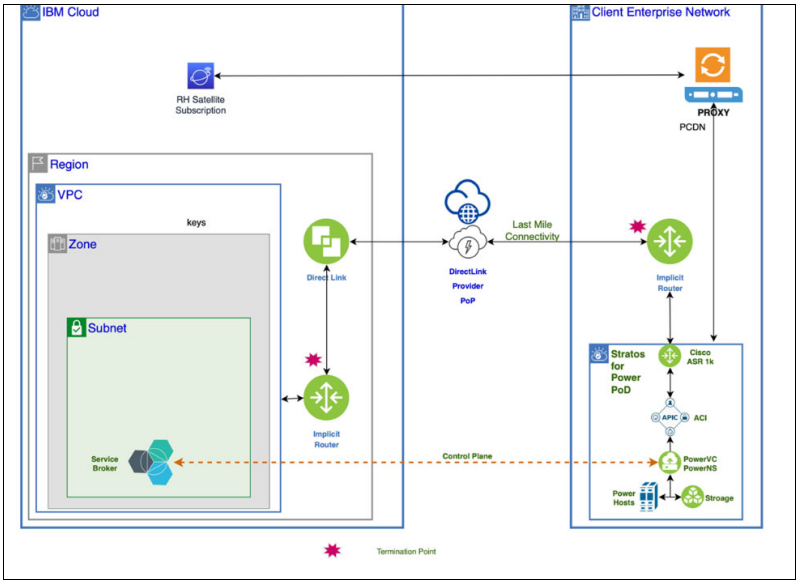


Figure 3-8 Connection to Red Hat Satellite server in IBM Cloud

7. DHCP network inside the Pod

   This use case provides a Dynamic Host Configuration Protocol (DHCP) protocol server within the pod to dynamically assign an IP address to a virtual machine.

   Restriction: The presence of the DHCP network within the pod is mandatory when you are using the OpenShift Container Platform on the IBM Power Virtual Server (On-premises) environment. The DHCP network is intended for use only in the OpenShift Container Platform.

   IBM Power Virtual Server Pods can be configured to include a private and hardware-based DHCP network. The edge router within the pod is configured with the DHCP pool and gateway. You can deploy virtual machines in the DHCP network. The virtual machines are assigned IP addresses from the DHCP server. You can attach only one DHCP network interface card (NIC) to a virtual machine. If you attach more than one DHCP NIC to a virtual machine, only one NIC acquires the IP address from the DHCP server that is assigned to the virtual machine.

   Important: When creating a DHCP network, note that the first four IP addresses are reserved. You must configure a network that has more than four IP addresses. For example, if the subnet mask is 255.255.255.248, the total number of IP addresses is eight. You cannot create a network with a subnet mask beyond 255.255.255.248 as it has less than or equal to five IP addresses.

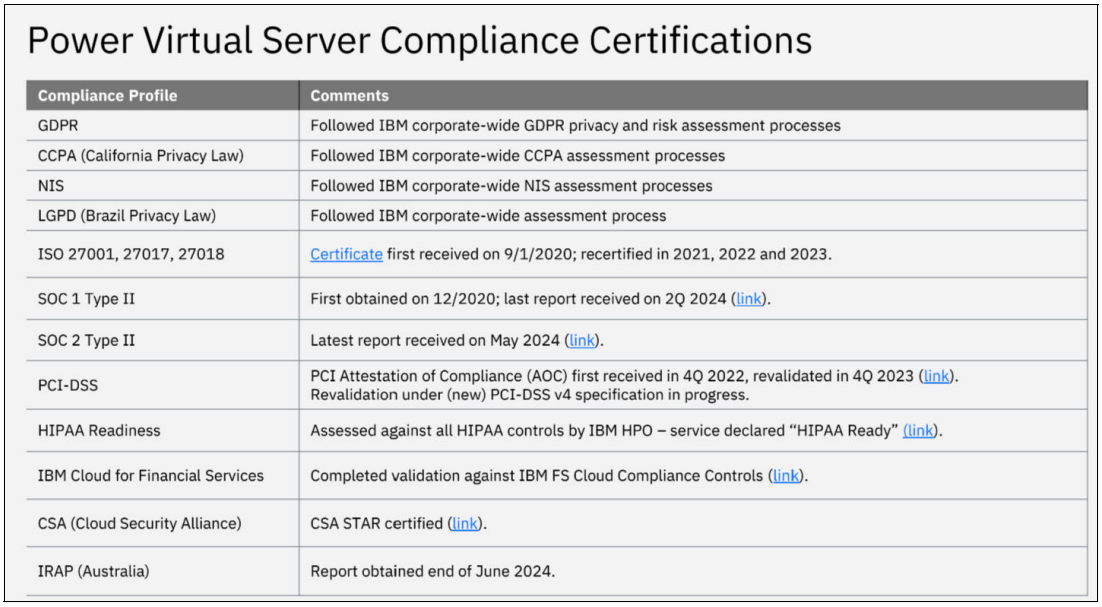
All of these network use cases can be set up initially when the pod is set up and defined. If you need to add or modify one of these use case implementations, then you can do so at any time by opening a service ticket.

For more information about setting up networking for your Power Virtual Server pod see Networking overview.

3.6 Compliance

For customers evaluating Power Virtual Server Private Cloud, adherence to privacy regulations, including those governing the placement and location of application data, is paramount. Acknowledging the significance of data jurisdiction, Power Virtual Server Private Cloud implements robust security principles and controls to guarantee that customer data remains exclusively within their data center or premises.

Power Virtual Server Private Cloud infrastructure meets the minimum regulatory compliance profiles necessary for operations across various regions and aligns with the IBM Secure Engineering and SPbD processes. These standards encompass GDPR, NIS, LGPD (Brazil privacy regulation), and CCPA (US State of California privacy law). If customers are interested in seeking compliance with additional profiles applicable to Power Virtual Server (public) infrastructure, such as SOC2, PCI-DSS, HIPAA, or IBM Cloud for financial services, they can reach out to their representative to discuss the required steps. It is important to note that all these regulatory profiles incorporate physical controls, thus relying on the implementation of controls within the customer's data center. Figure 3-9 shows a current list of Power Virtual Server certifications.



Power Virtual Server Compliance Certifications

Compliance Profile	Comments
GDPR	Followed IBM corporate-wide GDPR privacy and risk assessment processes
CCPA (California Privacy Law)	Followed IBM corporate-wide CCPA assessment processes
NIS	Followed IBM corporate-wide NIS assessment processes
LGPD (Brazil Privacy Law)	Followed IBM corporate-wide assessment process
ISO 27001, 27017, 27018	Certificate first received on 9/1/2020; recertified in 2021, 2022 and 2023.
SOC 1 Type II	First obtained on 12/2020; last report received on 2Q 2024 (link).
SOC 2 Type II	Latest report received on May 2024 (link).
PCI-DSS	PCI Attestation of Compliance (AOC) first received in 4Q 2022, revalidated in 4Q 2023 (link). Revalidation under (new) PCI-DSS v4 specification in progress.
HIPAA Readiness	Assessed against all HIPAA controls by IBM HPO – service declared "HIPAA Ready" (link).
IBM Cloud for Financial Services	Completed validation against IBM FS Cloud Compliance Controls (link).
CSA (Cloud Security Alliance)	CSA STAR certified (link).
IRAP (Australia)	Report obtained end of June 2024.

Figure 3-9 Power Virtual Server Compliance certifications

3.7 Planning for integration

Integrating IBM Power Virtual Server Private Cloud into an organization's existing IT infrastructure requires careful planning and coordination. Effective integration ensures that workloads, data, and applications can function seamlessly between on-premises systems and cloud environments, while maintaining performance, security, and compliance standards. Planning for integration involves considerations related to networking, data migration, security, management, and application compatibility.

3.7.1 Network Connectivity

For businesses adopting a hybrid cloud strategy, ensuring seamless network connectivity between on-premises infrastructure and Power Virtual Server Private Cloud is critical. Power Virtual Server supports IBM Cloud Direct Link, a high-speed, low-latency connection that provides secure, reliable access between on-premises environments and the IBM Cloud.

Power Virtual Server also supports integration with Virtual Private Networks (VPNs) and Virtual Private Clouds (VPCs), enabling secure communication between cloud-based and on-premises resources. This ensures that sensitive data can be transferred securely across locations without exposing it to the public internet.

Plan for bandwidth requirements, latency tolerance, and redundancy to ensure consistent performance across hybrid environments. Additionally, organizations should ensure that their internal network security policies align with cloud networking configurations.

3.7.2 Data Migration

Data migration is a key part of the integration process, particularly for businesses moving large workloads to the cloud. Planning involves evaluating the size, type, and sensitivity of the data to be migrated, as well as the bandwidth and time required for migration.

Data will need to be migrated to your Power Virtual Server instances using a network-based transfer tool similar to what would be used for any on-premises migration. There are many data migration strategies to choose from:

- For file-based migration, Cloud Object Storage can be used as an intermediary location to store files from your on-premises environment. You can retrieve and send your files to the Power Virtual Server environment from this location. You must create Cloud Object Storage buckets to transfer data over the public internet or privately secured links. For more information, see IBM Cloud Object Storage: FAQ.
- Using your current backup tools can create a backup copy of your data to Cloud Object Storage. This backup is then used to build your Power Virtual Server instance.
- Power Virtual Server supports FalconStor Virtual Tape Library which can be used to backup and restore your data.
- Databases can be migrated utilizing built-in replication capabilities provided by the database vendor.
- IBM PowerHA can be used to mirror systems (AIX and IBM i) to Power Virtual Server. Red Hat and SUSE provide a high availability clustering option utilizing Pacemaker. For IBM i, there are multiple third-party solutions such as Mimix and Maxava.
- IBM Aspera is supported and provides fast, secure, and efficient data transfer to the cloud which can reduce the amount of time required for your data migration.

Plan for potential challenges like downtime, data consistency, and application compatibility during migration. Test the migration process in a sandbox environment to ensure smooth execution.

As an example, a healthcare organization planning to migrate patient records from its on-premises IBM i system to Power Virtual Server Private Cloud uses IBM Aspera to securely transfer data, ensuring that the migration is completed quickly with no loss of data integrity.

3.7.3 Backup

In addition to being able to backup your virtual machines via the network into your existing locally hosted backup solution, IBM Power Virtual Server Private Cloud also offers a comprehensive suite of tools and services for backing up data to IBM Cloud Object Storage (COS). These solutions are designed to extend existing backup strategies seamlessly into IBM cloud, providing an efficient and secure method of protecting critical business data.

For Cloud backups, utilizing your existing backup solution in Power Virtual Server is possible assuming that the solution can backup to network attached solutions such as IBM Cloud Object Storage. In addition, IBM Power Virtual Server supports the virtual tape library solution from FalconStor which emulates tapes to the system for backups, while providing capabilities for writing the data over the network to IBM COS.

Provided below are some solutions for backing up to IBM Cloud Object Storage based on the type of operating system hosting the workloads.

- IBM i workloads:

  For IBM i, Backup Recovery and Media Service (BRMS) in conjunction with IBM Cloud Storage Solutions for i (ICC) can be leveraged for backing up data to the cloud (to IBM Cloud Object Storage). This method could be adopted for smaller LPARs/VMs based on bandwidth and other factors. However, a FalconStor StorSafe VTL solution (introduced in below section) in conjunction with BRMS is recommended when there is moderate to large amount of data to be backed up, many LPARs to be backed up, or when backup/restore performance matters most. In addition, a VTL solution is an essential capability for hybrid deployments.

- AIX and Linux workloads:

  For AIX and Linux workloads, backup software such as Storage Protect can be configured to have it's backup storage pool send and store backups to IBM Cloud Object Storage.

Appropriate network connectivity, preferably Direct Links with bandwidth that is adequately sized, will be an essential factor to consider based on your backup and restore window requirements.

Recommended Practices for Implementing Backup in IBM Power Virtual Server

Consider the following guidelines for planning an efficient backup solution.

1. Define Retention Policies

   Establish data retention policies that align with regulatory requirements and business needs, ensuring compliance while managing storage costs effectively.

2. Optimize Backup Scheduling

   Schedule backups during non-peak hours to avoid affecting production workloads. Implement incremental backups to reduce bandwidth usage and optimize cloud storage.

3. Regular Testing of Backup and Restore Processes

   Conduct regular tests of the backup and restore processes to ensure that data recovery is swift and reliable, verifying that the backups work as expected under different scenarios.

4. Implement Encryption and Compliance Standards

   Utilize IBM's encryption capabilities to protect sensitive data and ensure compliance with relevant regulations such as GDPR, HIPAA, and PCI DSS.

5. Optimal connectivity for transferring backups to IBM Cloud Object Storage

   Ensure that the network connectivity between the on-premises location hosting the workloads to IBM cloud is sized and configured with adequate bandwidth for ensuring backups can be sent and received within the expected backup/restore window.

Backup media options

There are three primary media options available for use in backing up your data.

- Block disk provided by the IBM FlashSystem storage in your pod.
- Cloud Object Storage which is network attached object storage.
- Virtual tape library solution that emulates a tape library which is network attached using iSCSI connections

Local attached disk

Locally attached disk can be a good solution for backing up relatively small amounts of data. If you backup up data on high IOP tiers to lower IOP tiers within your POD, this might be cost effective for your requirements. However, for most environments, this is not an effective solution.

Cloud Object Storage

IBM Cloud Object Storage is a highly available, durable, and secure platform for storing unstructured data. Unstructured data (sometimes called binary or "blob" data) refers to data that is not highly structured in the manner of a database. Object storage is the most efficient way to store PDFs, media files, database backups, disk images, or even large structured datasets.

The files that are uploaded into IBM Cloud Object Storage are called objects. Objects can be anywhere from very small (a few bytes) [to very large] (up to 10TB). They are organized into buckets that serve as containers for objects, and which can be configured independently from one another in terms of locations, resiliency, billing rates, security, and object lifecycle. Objects themselves have their own metadata in the form of user-defined tags, legal holds, or archive status. All data stored in IBM Cloud Object Storage is encrypted, erasure-coded, and dispersed across three locations (with the distance between locations ranging from within a single data center, across a Multi-Zone Region or MZR, or even across multiple MZRs). This geographic range of dispersal contributes to a bucket's resiliency. All requests and responses are made over HTTPS. Figure 3-10 illustrates backing up your data to IBM Cloud.

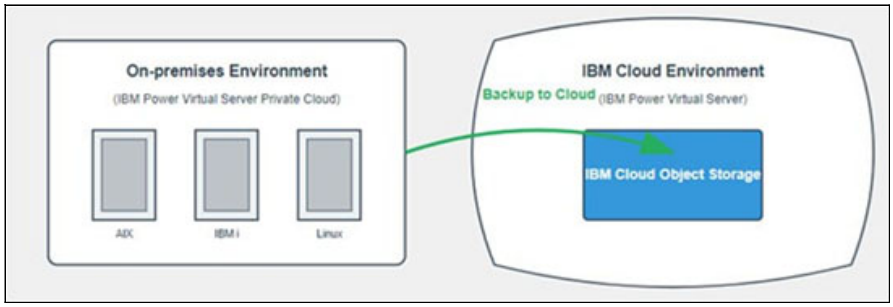


Figure 3-10 Backing up to IBM Cloud

FalconStor Virtual Tape Library

FalconStor Virtual Tape Library (VTL) is an optimized backup and deduplication solution that provides tape library emulation, high-speed backup or restore, data archival to IBM Cloud Object Storage for long-term storage, global data deduplication, enterprise-wide replication, and long-term cloud-based container archive, without requiring changes to the existing environment.

Here are some of the StorSafe VTL benefits

- Up to 95% data reduction. FalconStor reduces the storage capacity required by backup data in the client-managed environment and in the cloud by up to 95%.
- Ransomware Protection. Recover data from any point in time with air-gapped, immutable backups.
- Off-site Protection. For off-site protection, StorSafe VTL exports virtual tapes as physical tapes, and keeps track of their location with a management dashboard. Virtual tapes can remain in the library for fast restores or can be stubbed when exported to local object storage.
- High Performance. Improve performance for both backup and recovery.
- Modernize. FalconStor is compatible with existing backup software, hardware, and operational procedures.

Additional information on FalconStor can be found at the FalconStor StorSafe VTL site.

FalconStor VTL is offered through the IBM Cloud Catalog for both Power Virtual Server, both IBM data center and client location. For more information see the FalconStor Deployment Guide.

General Backup Strategies

Virtual machines hosted in IBM Power Virtual Server private cloud could also be integrated seamlessly with existing backup solutions such as IBM Storage Protect, allowing organizations to extend their current backup strategies into Power Virtual Server Private Cloud without the need for entirely new tools / backup softwares.

Backup strategies for IBM i

There are many products that can be used to backup your data in an IBM i environment. However, IBM i has a built-in backup solution (Backup Recovery and Media Services or BRMS) which is the choice of many clients.

Using IBM Backup, Recovery and Media Services

A common IBM i backup strategy is to use IBM Backup, Recovery, and Media Services (BRMS) and IBM Cloud Storage Solutions (ICC). Together, these products automatically back up your LPARs to IBM Cloud Object Storage. The ICC product can be integrated with BRMS to move and retrieve objects from remote locations, including Cloud Object Storage. In most cases, this process involves backing up to virtual tapes and image catalogs.

Note: You need extra storage for the LPAR to host the image catalogs until they are moved to Cloud Object Storage.

The typical IBM i customer uses the following flow to back up LPARs and objects:

1. Use the 5733-ICC product to connect to Cloud Object Storage (~2 times the disk capacity to hold the backup images).

2. Connect to IBM Cloud Object Storage by following the steps that are mentioned in Using Cloud Object Storage.
3. Complete the back up to Cloud Object Storage by choosing the speed and resiliency that is required.

Backup strategies for AIX and Linux

Your current backup software can continue to be used in Power Virtual Server Private Cloud, to either perform backups locally over the LAN or alternatively, if supported, write the backup data to IBM Cloud Object Storage. Another option is to use the FalconStor VTL solution as a backup target for your existing backup solution. The FalconStor VTL solution provides tape emulation for your backups.

Solution for Hybrid Cloud Backup Strategy

IBM Power Virtual Server facilitates a hybrid cloud backup strategy, where on-premises systems could be backed up to the cloud or a copy of the backups replicated to IBM Cloud leveraging Virtual tape library (VTL) replication or Backup replication technologies (E.g. IBM Storage protect replication). This adds an extra layer of protection beyond local backups, safeguarding data from physical disasters, theft, or hardware failure.

FalconStor StorSafe VTL Replication

By leveraging FalconStor StorSafe VTL, backup solutions can direct backups into the FalconStor StorSafe VTL that is hosted in Power Virtual Server Private Cloud. Additionally, a corresponding FalconStor StorSafe VTL can also be provisioned in Power Virtual Server Public Cloud (in IBM cloud) with VTL replication configured between the two VTLs ensuring a copy of all backups taken on-premises are replicated into the FalconStor StorSafe VTL in Power Virtual server public cloud (IBM cloud). In addition to the below benefits, this method provides you with the capability to render hybrid cloud deployment:

- Cost effective cold-site disaster recovery.
- Alternative solution to off-site physical tape vaulting.

Figure 3-11 shows a sample of a hybrid deployment leveraging FalconStor StorSafe VTL.

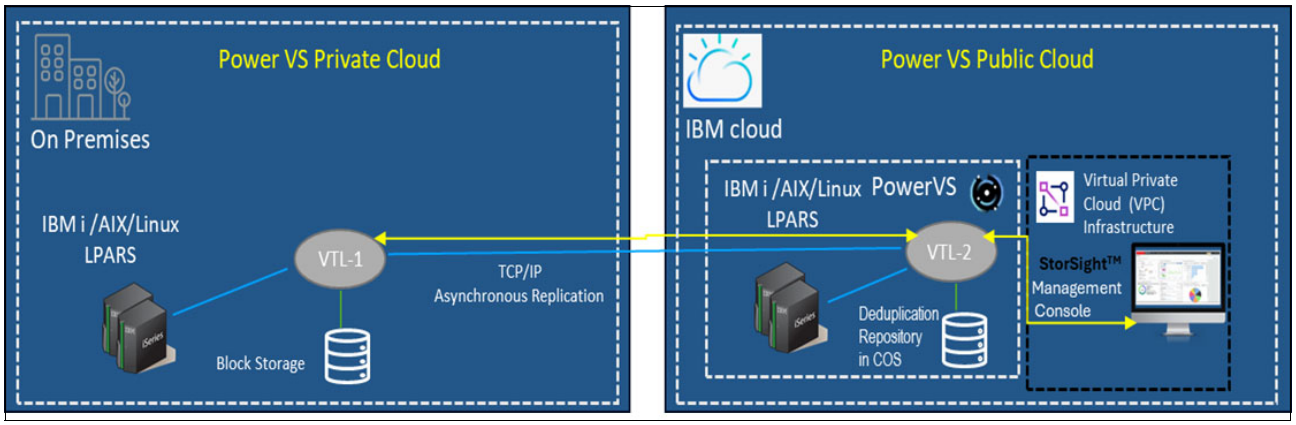


Figure 3-11

Conclusion

By leveraging IBM Power Virtual Server Private Cloud, businesses can take advantage of a scalable, secure, and flexible backup solution that aligns with modern IT requirements. IBM Power Virtual Server ensures that data remains protected and readily recoverable, allowing

organizations to focus on their operations without the risks associated with data loss or system failures.

3.7.4 Disaster Recovery

Business Continuity is of utmost importance for enterprises that run mission-critical workloads, and their resiliency depends on how quickly enterprises can recover (RTO) from a disaster event and how much data they can recover (RPO) after the event. Disaster events can be hardware, software failures or catastrophic incidents such as fires, storms, earthquakes, etc. and a highly reliable, secured disaster recovery solution is a must.

Disaster recovery (DR) is a critical component of business continuity planning, focusing on restoring services and minimizing downtime following an unexpected event such as system failure, cyberattacks, or natural disaster. Disaster recovery addresses the loss of the ability to run your business at its primary site due to any event or situation that causes the primary site to become unavailable.

There are varying levels of disaster recovery with each offering different levels of recovery. Advanced configurations involving continuous data replication can provide quick recovery with a near-zero recovery point objective (RPO). RPO represents the amount of data (generally expressed in a measure of time) that cannot be recovered and must be reentered or recreated. Less advanced options that involve saving point in time copies which are then restored. This provides an RPO equal to the time since the last save operation and recovery also takes longer. In general, the cost of managing disaster recovery options is inversely related to the RPO of the solution. In other words, the smaller the RPO, the greater the cost.

As you design your solution, you need to consider your availability requirements and how to respond to potential interruptions in the service. One option is to use Power Virtual Server instances in an IBM datacenter as disaster recovery location. Another option is to use a second Power Virtual Server Private Cloud pod in a different location.

The disaster recovery solution must provide for replication of your appliction data to the recovery location and allow the applications to be restarted in that location. You can restart your workload on a different host in the cloud if a hardware failure occurs. The use of Power Virtual Server, either on-premises or off-premises, for your DR location can reduce your recovery costs as the cloud resources do not require an initial capital investment and you only pay for the resources when they are used.

Recommended Practices for Implementing DR with IBM Power Virtual Server

It is recommended that you take the following into account as you plan and implement your Disaster Recovery solution for Power Virtual Server.

1. Regular Testing

   Regularly test the failover and failback processes to ensure the DR setup works as expected. IBM Power Virtual Server provides the tools to simulate disaster scenarios and verify recovery procedures.

2. RPO and RTO Considerations

   Define Recovery Point Objectives (RPO) and Recovery Time Objectives (RTO) based on business needs. IBM Power Virtual Server can help organizations achieve low RPO and RTO targets through continuous replication and fast failover mechanisms.

3. Compliance and Data Sovereignty

Ensure that the chosen DR region complies with regulatory requirements, particularly for industries like healthcare and finance. Power Virtual Server Private Cloud allows businesses to select DR locations that meet GDPR, HIPAA, or other regional data protection regulations.

Data Replication

Data replication is a critical component of your disaster recovery plan. There are multiple methods of ensuring that you have a copy of your application data.

For applications with minimal data change it is possible that you cound use backup tools such as IBM Storage Protect or BRMS to back up the data to either Cloud Object Storage or to a virtual tape library solution such as FalconStor. However, the RPO and RTO of a backup and restore solution is unlikely to meet the disaster recovery requirements of most systems.

There are two classes of data replication solutions that can be considered for continuous replication of your data to the disaster recovery location:

- Software replication
- Storage replication

Software replication

Software replication solutions utilize either system level tools (generally described as clustering tools) or application or database level replication tools.Power Virtual Server supports many software replication tools:

- PowerHA SystemMirror for AIX

  PowerHA SystemMirror for AIX can be used to manage disaster recovery solutions for AIX instances utilizing Global Logical Volume Mirroring (GLVM). GLVM replicates or mirrors data between different AIX instances at the logical volume layer. As data is written to a mirrored logical volume the data is also sent over the network to the remote system and GLVM maintains the data consistency. PowerHA is responsible for moving the applications to the remote AIX instance in case of failures in the primary. Using PowerHA your system administrators can synchronize your data back to the primary site when service is restored and eventually move the applications back to the primary location. Reference Using PowerHA System Mirror using GLVM on Power Virtual Server for more information.

- PowerHA SystemMirror for IBM i

  Similar to the AIX solution, PowerHA SystemMirror provides support for IBM i replication using Geo Mirroring. IBM PowerHA System Mirror for i provides end-to-end integrated clustering solutions for high availability and disaster recovery. PowerHA is an integrated extension of the IBM i operating system and offers environmental, application, and data resiliency solutions for managing access and storage in the event of planned or unplanned outages. Geographic Mirroring is integrated directly with the storage management component in the IBM i operating system. As the system writes data from main memory to disk, the same data is sent over TCP/IP to the target system to be written to disk, ensuring identical data at the byte level replicated between both systems. Reference IBM i Disaster Recovery with IBM Power Virtual Server.

- Application-specific replication

  Applications might have replication mechanisms that can sync multiple environments. These options are commonly used for application-specific replication:

  - Db2 HADR
  - Rocket iCluster HA/DR
  - Maxava HA
  - Migrate 23

- MIMIX
- Oracle Data Guard
- Oracle GoldenGate

Work with the vendor for implementation guidance on Power Virtual Server for these products.

Storage Replication using Global Replication Services

The GRS enablement in IBM Power Virtual Server Private Cloud enables asynchronous replication of data between the primary location infrastructure and the secondary location infrastructure. The two infrastructure locations have the identical set of capabilities that IBM Power Virtual Server in IBM data center provides.

The infrastructure in the IBM Power Virtual Server Private Cloud in which your workspace is located, has the primary volumes of the replication pairs. The infrastructure in the secondary location has the auxiliary volumes. IBM Cloud infrastructure internally uses IBM FlashSystems Global Mirror Change Volume (GMCV) as storage technology that provides asynchronous replication.

Each time a replicated volume is created, four copies of volumes are created across the infrastructure:

- Primary volume in the primary infrastructure
- Primary change volume in the primary infrastructure to store the delta changes
- Auxiliary volume in the secondary infrastructure
- Auxiliary change volume in the secondary infrastructure to update the delta changes

You must provide the required network configuration between the primary location and the secondary location for replication, which includes the following prerequisites:

- The network bandwidth must be greater than or equal to 10 Gbps.
- The network latency must be less than or equal to 200 ms.

During the first sync, the entire data from primary volumes is copied to the auxiliary volumes. For subsequent syncs, only the delta changes are copied. The effective Recovery Point Objective (RPO) depends on the capability of the underlying network throughput and the application characteristics. If the network throughput is insufficient to reach the defined RPO, then the time duration between the synchronization increases.

3.7.5 Application Compatibility

Power Virtual Server supports a wide range of operating systems, including AIX, IBM i, and Linux, making it easier to migrate legacy applications to the cloud without requiring refactoring. However, it is crucial to evaluate each application's compatibility with the Power Virtual Server environment before migration.

For cloud-native or containerized applications, Power Virtual Server integrates seamlessly with DevOps tools and CI/CD pipelines, enabling smooth deployment and testing in hybrid environments.

Businesses must assess application dependencies, performance requirements, and any potential adjustments required for cloud deployment. Additionally, ensure that licensing for any proprietary software is aligned with cloud deployment models.

For example, a manufacturing company moves its legacy AIX-based supply chain management system to Power Virtual Server Private Cloud, ensuring that the application’s

dependencies are compatible with the cloud environment and that the performance remains consistent post-migration.

3.7.6 Security and Compliance

Security is paramount when integrating cloud environments with on-premises systems. Power Virtual Server provides robust security features, including data encryption (at rest and in transit), access controls, and firewall configurations to protect sensitive data.

Businesses in regulated industries such as healthcare, finance, and government must ensure that their cloud integration complies with industry standards like GDPR, HIPAA, and PCI DSS. Power Virtual Server Private Cloud supports compliance by providing tools to manage data residency, audit trails, and secure data handling practices.

Power Virtual Server integrates with IBM Cloud Identity and Access Management to provide granular control over who can access specific resources. Role-based access controls (RBAC) ensure that users only have access to the data and systems they are authorized to interact with.

A financial institution integrates its core banking applications with Power Virtual Server while ensuring compliance with SOX and PCI DSS regulations. Data encryption and secure access controls are implemented to protect sensitive customer information throughout the integration process.

3.7.7 Management and Monitoring

Unified Management Power Virtual Server Private Cloud offers tools like IBM Cloud Monitoring, and IBM Cloud Automation Manager for managing cloud resources and monitoring performance. These tools help businesses maintain visibility into their cloud environments, optimize resource usage, and automate routine management tasks.

By centralizing monitoring and management, businesses can track resource utilization, performance metrics, and security incidents across both on-premises and cloud environments. This ensures that any issues can be quickly identified and addressed, minimizing disruptions.

Ensure that IT teams are trained in using cloud management tools and that clear processes are in place for monitoring, troubleshooting, and scaling cloud resources as needed.

A retail business integrates its Power Virtual Server cloud infrastructure with its on-premises ERP system. Using IBM Cloud Monitoring, the IT team monitors both environments from a single dashboard, ensuring real-time insights into system performance and resource allocation.

3.7.8 Steps for Effective Integration Planning

The following steps should be taken to ensure an effective integration.

1. Assess Current Infrastructure

   Conduct a thorough assessment of the existing on-premises infrastructure, including applications, workloads, network configurations, and security policies. Identify areas that may need to be adjusted or updated to ensure compatibility with the cloud environment.

2. Define Integration Objectives

Clearly define the objectives for the integration, such as improving performance, enhancing scalability, or optimizing costs. These objectives will help guide decisions around cloud configurations, data migration strategies, and workload allocation between on-premises and cloud environments.

3. Develop a Migration and Integration Plan

   Create a detailed migration plan that outlines the steps, tools, and timelines for moving data and workloads to the cloud. Include contingency plans for potential issues like downtime or data migration failures.

4. Test the Integration

   Before fully deploying workloads to the cloud, test the integration in a pilot environment. This allows businesses to identify and resolve any issues related to network connectivity, data migration, or application compatibility before going live.

5. Monitor and Optimize Post-Integration

   After the integration is complete, continuously monitor the cloud environment to ensure that performance, security, and compliance objectives are being met. Use monitoring tools to track resource utilization and make adjustments as needed to optimize costs and performance.

Example of Integration Planning in Action

A multinational corporation is planning to integrate its existing on-premises ERP system with Power Virtual Server Private Cloud to improve scalability and performance. The IT team begins by assessing the current infrastructure and identifying key applications that can be migrated to the cloud. After defining the goals of the integration, they establish secure network connectivity and migrate their data. The team tests the integration in a pilot environment before migrating the full workload, ensuring that the process runs smoothly and without disruption to business operations.

3.7.9 Conclusion

Planning for integration with IBM Power Virtual Server Private Cloud is essential for ensuring a smooth transition from on-premises infrastructure to a hybrid cloud environment. By addressing key areas such as network connectivity, data migration, application compatibility, security, and management, businesses can fully leverage the benefits of cloud computing while maintaining control over their IT operations. With proper planning and execution, Power Virtual Server Private Cloud can provide the scalability, flexibility, and security needed to meet the demands of modern enterprises.

3.8 Management

With Power Virtual Server, IBM manages the complex setup of servers, storage, and networking, adhering to best practices for a reliable, high-performance infrastructure. You retain control of your applications and data, while IBM and our partners continue to introduce enhanced options to increase value.

IBM also ensures the firmware and platform (from the virtualization layer down) are kept up to date, providing timely security and resilience updates, along with integrated Live Partition Mobility (LPM) technology to minimize downtime.

Figure 3-12 shows the responsibility assignment matrix for Power Virtual Server.

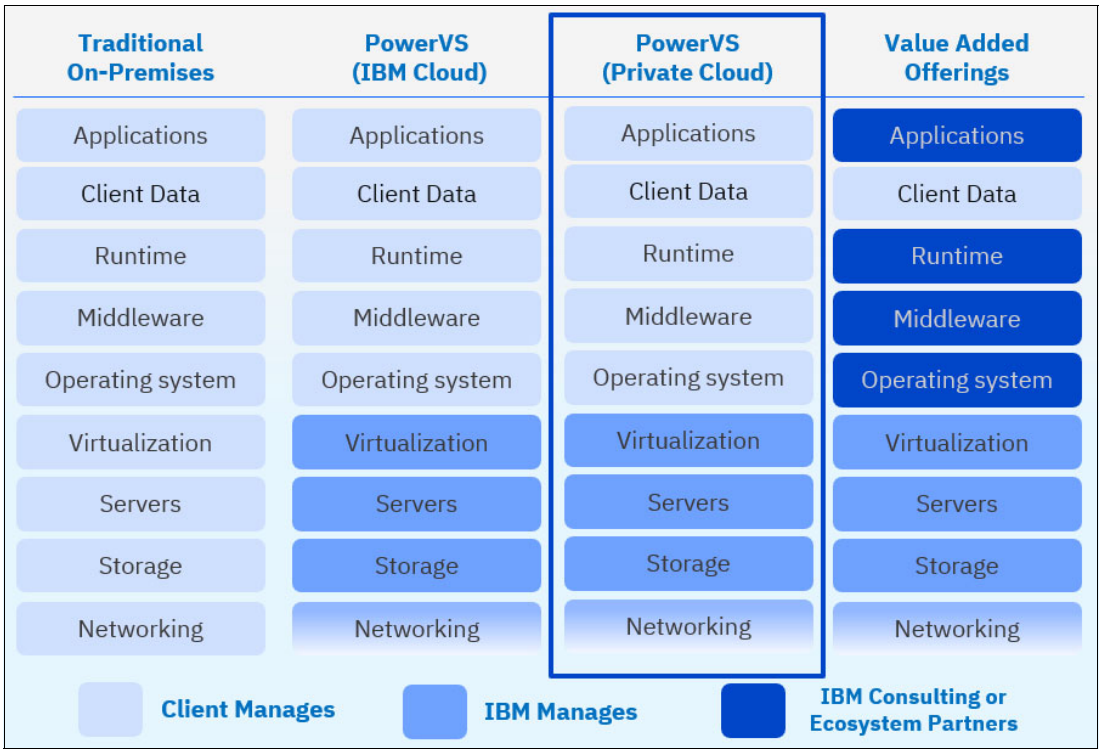


Figure 3-12 Responsibility matrix for Power Virtual Server

IBM site reliability engineering (SREs) focus on several key areas to ensure a robust and optimized Power Virtual Server experience, including:

- Monitoring and Alerting
- Incident Management
- Analysis (Improvement Cycle)
- Firmware and supporting code currency
- Security
- Compliance
- Deployment and Provisioning

Availability and serviceability

Call Home is enabled for the Power server HMCs and FlashSystems storage. While software subsystem and infrastructure logs remain within the pod, diagnostic data is shared with IBM support as needed. All data flows are encrypted across the restricted Power Virtual Server Private Cloud Control Plane Network (PCCN).

Note: Client data is never accessible or transmitted over this network.

Client operations

Table 3-2 lists optional 'Day 2' operational technologies that enhance the IBM Power Virtual Server Private Cloud solution.

Table 3-2 Client Operational Capability

Client Operational Capability	Reference Technology
Backup and Restore	► IBM Storage Protect (formerly IBM Spectrum Protect) ► IBM i: BRMS -> COS (recommended <3TB); BYO Falconstor VTL over iSCSI
High Availability	► AIX: PowerHA for AIX Enterprise Edition ► IBM i: PowerHA Geographic Mirroring; all 3rd party logical replication technologies such as MIMIX, iCluster, and others ► RHEL: Red Hat Enterprise Linux High Availability ► SUSE: SUSE Linux Enterprise High Availability ► IaaS: Automated remote restart provided by the underlying IaaS management layer
Disaster Recovery	► AIX: PowerHA for AIX Enterprise Edition with GLVM mirroring ► IBM i: PowerHA Geographic Mirroring; all 3rd party logical replication technologies such as MIMIX, iCluster, and others ► RHEL: Red Hat Enterprise Linux High Availability ► SUSE: SUSE Linux Enterprise Clustered Disaster Recovery
Observability and Monitoring (for both virtual machines and containers)	IBM Instana™ Observability: ► Provides operating system agents for AIX, Linux and IBM i ► Provides ability to monitor Red Hat OpenShift clusters
Security and Compliance	IBM PowerSC: ► Provides security and compliance capabilities for AIX, Linux and IBM i

3.8.1 IBM services

Should you need assistance with any aspect of your IBM Power Virtual Server Private Cloud installation or operation, IBM Cloud Support team members are available to assist you in resolving the issues.

IBM Cloud Support plans

The IBM Cloud Services team is responsible for working with clients to resolve any issues that they have in accessing or using IBM Power Virtual Server, either off-premises or on-premises. The support team is available 24 x 7 and can be accessed in a number of ways. Clients can choose to open cases online, through online chats, or via the phone.

Clients can choose an Advanced or Premium support plan to customize your IBM Cloud support experience for your business needs. The Advanced plan is the default. The level of support that you select determines the severity that you can assign to support cases and your level of access to the tools available in the Support Center. Table 3-3 differentiates the two different offerings.

Table 3-3 IBM Cloud support plan offerings

	Advanced	Premium
Description	For environments with a limited number of business-critical applications.	For mission-critical environments with a strategic dependency on IBM Cloud.

	Advanced	Premium
Case management	Prioritized case handling and support experience that is aligned with your business needs.	Client engagement that is aligned with your business outcomes to accelerate time-to-value
Availability	24 x 7 access to the IBM Cloud technical support team through cases, phone, and chat	24 x 7 access to the IBM Cloud technical support team through cases, phone, and chat
Initial response time objectives	Severity 1 — 1 hour Severity 2 — 2 hours Severity 3 — 4 hours Severity 4 — 8 hours	Severity 1 — 15 minutes Severity 2 — 1 hour Severity 3 — 2 hours Severity 4 — 4 hours
Additional Support	N/A	Technical Account Manager

IBM Technical Account Manager

Premium clients are assigned a Technical Account Manager that provides a personalized experience, advising on processes, policies, and operation related to on-premises services. The Technical Account Manager provides the following services to enhance your IBM Power Virtual Server Private Cloud experience:

- Personalized Client Advocacy for private cloud services
  - The Technical Account Manager conducts a Welcome Call tailored for the clients' services.
  - Delivers a customized IBM Support Plan ensuring IBM Support engagement processes are understood.
  - Key interface into Support, SRE Operations & Experts.
- Acceleration Issue Resolution
  - Technical Account Manager coordinates across the IBM SRE Operations & IBM Support teams to reduce case resolution time.
- Personal Onboarding Assistance
  - Helps with onboarding services and serves as client advocate, aligning IBM resources to meet client needs.
  - Ensures that roles and responsibilities are understood.
- Advocate for on-premises Infrastructure Maintenance
  - Coordinates with IBM SRE to provide personalize client communication about planned maintenance events.
  - Ensures that any maintenance activity is known and understood.
  - Highlights key dates and times for actions.
- Expert Coordination
  - Can help coordinate experts for client-initiated events, including architectural support on maintenance events, to review required critical functions & services.
- Quarterly Review Meetings
  - Hosts periodic reviews of on-premises services, focused on business outcomes and future planning activities.
  - Reviews case activity and planned maintenance.

Opening a support case

To open a support case for Power Systems Virtual Server Private Cloud, complete the following steps:

1. Log in to the IBM Cloud with your IBM Cloud account credentials as shown in Figure 3-13.

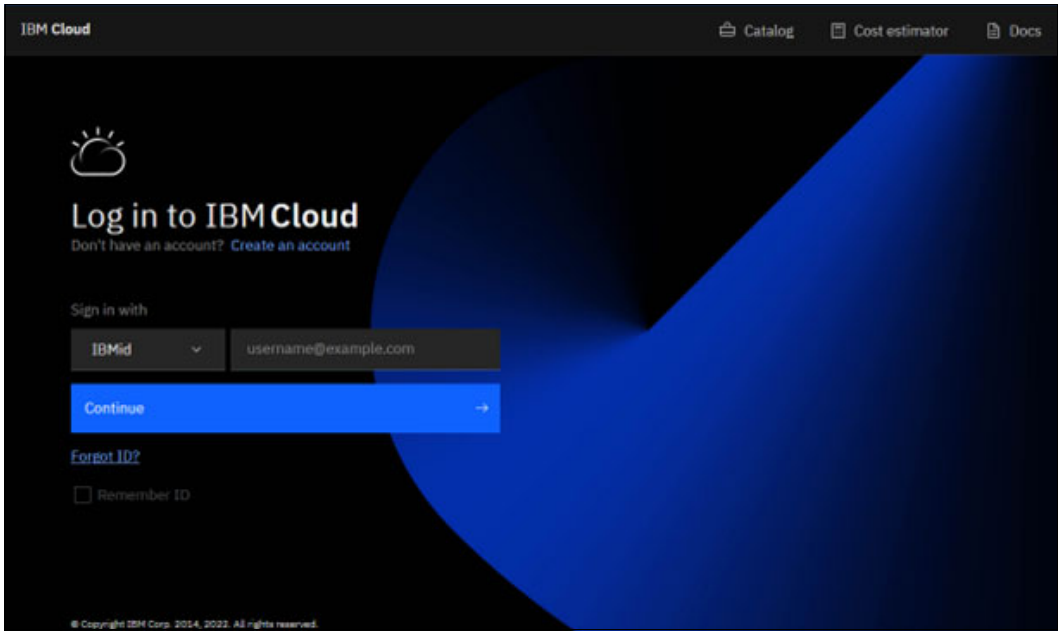


Figure 3-13 Login screen for IBM Cloud

2. In the menu bar, click the question mark with a circle, and select **Support Center** as shown in Figure 3-14.

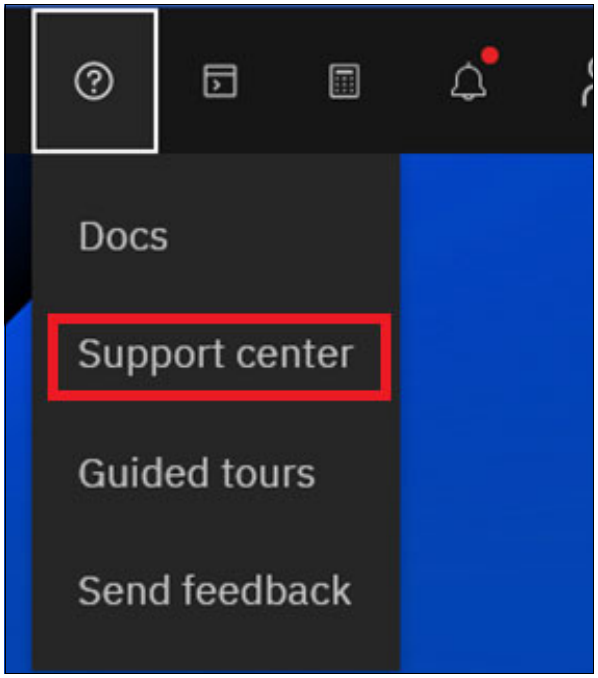


Figure 3-14 Select support center

3. From the Contact support section, click **Create a case** as shown in Figure 3-15.

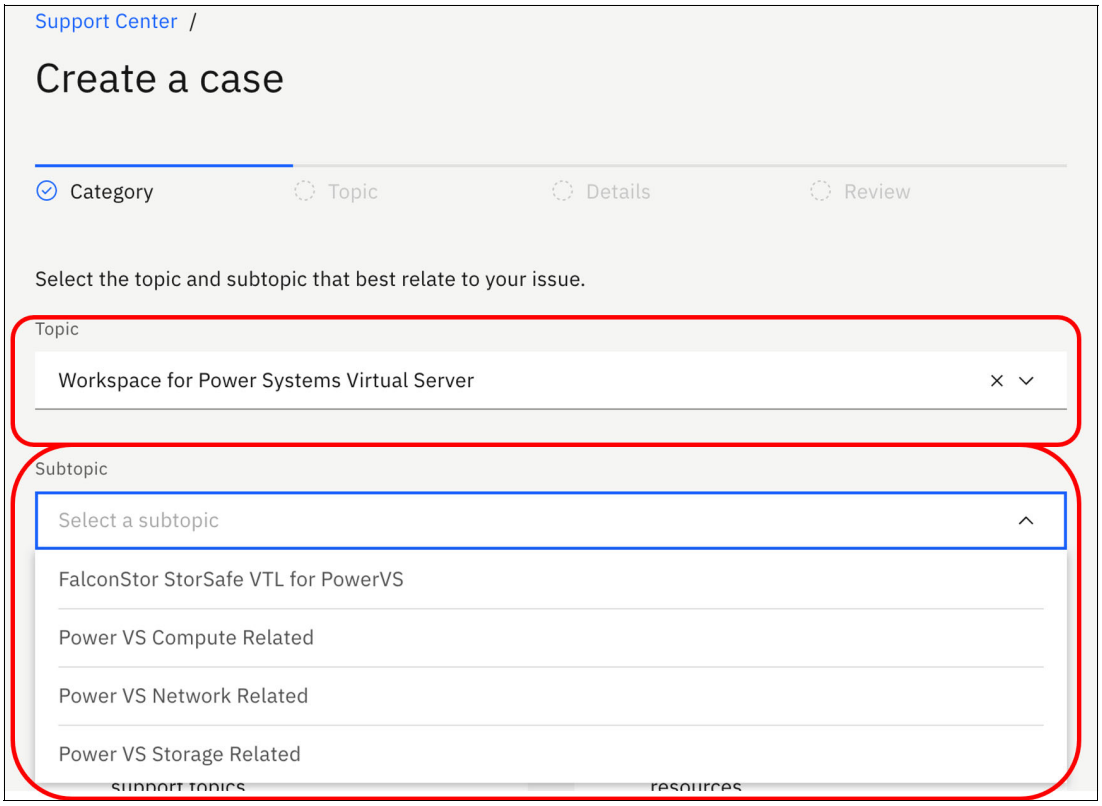


Figure 3-15 Create a case

4. In the category section, select the topic “Workspace for Power Systems Virtual Server”.
5. Select the subtopic that is most closely related to your issue. E.g.: Power VS On Premises Related.
6. Complete the description details and other required fields.
7. Optional steps:
   - Attach files and resources to provide more details about the issue you are experiencing.
   - If you would like a user in your account to be updated about the case, add them by using the Contact watchlist.
   - Select **Email me updates about this case** to receive support case notifications.
8. Click **Next**, review your case summary, and click **Submit case**. After you receive email verification for the case, follow the instructions for further communication on the issue.

Important: To maintain security, do not include any personal information, sensitive data, or device or service credentials in case responses. For example, do not include passwords, API keys, secrets, or credit card information.

For more details see Using the Support Center. After your support case is created, you can view its progress on the Manage Cases Page.

3.8.2 Unplanned network disconnection of management control plane

In case there is an unplanned network outage for the management network connecting the IBM Cloud-resident Service Broker instance and the pod infrastructure, the VMs will continue to run within the pod.

Table 3-4 describes the implications of a pod that is running in an unexpected, disconnected mode which is due to an unplanned network outage where both the primary and secondary management connections (Direct Link or site-to-site VPN) to IBM are lost.

Table 3-4 Impact of disconnected mode

Capability	Impact of disconnected mode	Comment
Your workload and data	No impact	Client workload remains fully operational, and data remains fully available.
GUI / API (for read operations)	Minimal impact	GUI remains operational and leverages last-known cached data. Incoming updates for data, such as storage consumption, remains fixed until control plane connectivity is re-established
GUI / API (for write operations)	Unavailable	For example, VM or volume creation. Resource write operations are unavailable until control plane connectivity is re-established.
Command-line interface (CLI)	Minimal impact	Read operations remain operational and write operations are unavailable until connectivity is re-established
Billing and metering	No impact	Metering uses last-known cached data (if the pod gets disconnected, no write operations can occur in the interim)
Telemetry	Unavailable	In-pod telemetry data is unavailable until control plane connectivity is reestablished (one exception is that IBM Storage Insights caches information for a selected period)
DHCP service (for client data networks)	No impact	DHCP services are provided by the pod-resident network infrastructure and does not require a connection to IBM Cloud.
IBM remote support	Unavailable	IBM Operations staff would not be able to remotely connect to the pod until communication is re-established.

4

Ordering and Installation

Once you have determined that IBM Power Virtual Server Private Cloud is the appropriate solution for your enterprise infrastructure and have completed the planning process to determine the appropriate Pod configuration for your requirements, the next step is to work with IBM or your IBM Business Partner to order the Pod and have IBM install it in your data center location. This chapter describes the ordering and installation process for IBM Power Virtual Server Private Cloud.

The following topics are included in this chapter:

4.1 Process overview

IBM Power Virtual Server Private Cloud is listed in the IBM Cloud Catalog, it is available under IBM Power Virtual Server tile. Clients can use the Estimator[1] located in the GUI to define the desired configuration and see an estimated monthly list price.

IBM is responsible for offering content, price and billing – metered hourly and billed monthly. The offering is designed as a Business Partner friendly go-to-market offering. The services Infrastructure will reside in clients' data centers. Whereas the client is responsible for physical site location (e.g., space, energy and network connectivity). IBM is responsible for installation and configuration of the infrastructure. IBM will own and operate the Power infrastructure. The client (or partner) provisions the virtual machines. IBM operates the infrastructure up through the hypervisor layer, including all hardware and software maintenance operations. Client (or partner) is responsible for the OS, middleware and applications. IBM will provide support (for all IBM managed components).

Power Virtual Server Private Cloud requires a specific ordering and installation process which we describe in this section.

Figure 4-1 provides an overview of the planning and installation lifecycle for a Power Virtual Server Private Cloud instance.

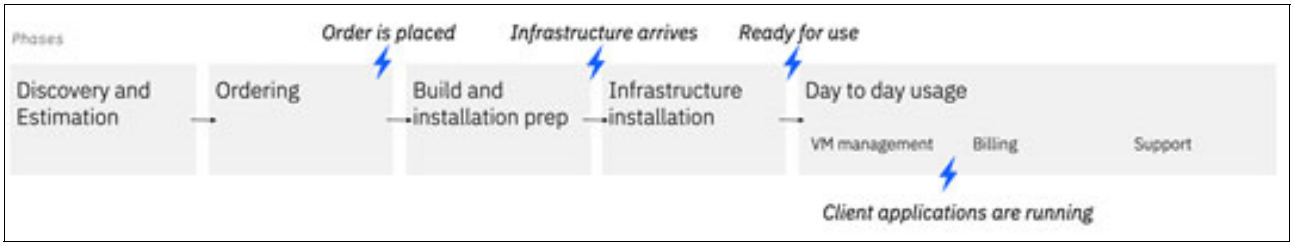
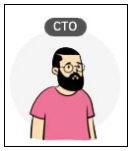
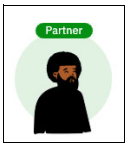


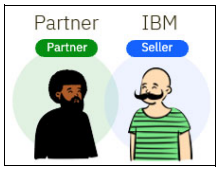
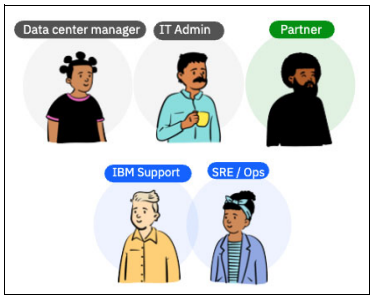
Figure 4-1 Power Virtual Server Private Cloud Planning and installation lifecycle

Table 4-1 presents sample scenario to illustrate the process and the people that would be involved in the planning, installation and operation of a Power Virtual Server Private Cloud instance.

Table 4-1 Sample installation planning scenario

Persons involved	Activity
	Wade, the CTO at ABC Company, is looking to optimize IT operational cost and improve efficiency for the next year. He's looking to maintain data locally and to offload the operations to vendors, so they can focus on innovation.
	Bruno, their business partner, recommends Power Virtual Server Private Cloud as a solution that can reduce costs, address IT skills, improve IT operations and allow them to focus on their applications. Additionally, Bruno shares that despite being on-premises, IBM ships and configures the necessary hardware and is responsible for maintaining it in a full as-a-service model.

[1] See the IBM Cloud estimator website. https://cloud.ibm.com/estimator

Persons involved	Activity
	Bruno connects with his IBM seller to get a more detailed quote that includes any possible discounts and to prepare for an order.
	IBM manufacturing team receives the BOM with all the details to assemble ABC Company's Power Virtual Server Private Cloud.
	IBM TLS connects with ABC Company to discuss floorspace planning, networking planning, and other items in preparation for the Power Virtual Server Private Cloud delivery and setup.
	TLS SREs setup and configure the infrastructure and then register the Power Virtual Server Private Cloud infrastructure with IBM Cloud. The Power Virtual Server Private Cloud instance appears as a new IBM Cloud Satellite Location in the client's account and is ready toy to run client applications

4.2 Finalizing configuration

During the initial planning as discussed in Chapter 3, “Planning Considerations” on page 71, you have done an analysis of the workloads that you plan on deploying to IBM Power Virtual Server Private Cloud and have determined the appropriate Pod configuration to support that workload. Working with your IBM representative, you will have received initial pricing, based on your estimated configuration. Now it is time to finalize the configuration and submit the order for the Pods that IBM will install in your data center.

To ensure that your experience with IBM Power Virtual Server Private Cloud is successful, IBM – and your IBM business partner, if one in involved – will work with you to validate the order and will schedule a technical review. Once the technical delivery assessment is completed and your configuration is finalized, the configuration will be turned over to IBM manufacturing to build and ship your IBM Power Virtual Server Pod to your location.

4.2.1 Technical delivery assessment

As part of the process to ensure that the proper configuration is being proposed and delivered, IBM (and any associated IBM business partner) work with the client to do a technical delivery assessment (TDA). The IBM sales team and the client meet with a group of

IBM Cloud experts to review the client's requirements and goals and assess whether the proposed solution will meet expectations. The TDA process looks at the client's current environment:

- What existing IBM Power hardware is being replaced by the solution?
  - How many LPARs?
  - What are the processor and memory requirements?
  - What are the storage requirements?
- What applications will be migrated to the new Power Virtual Server Private Cloud Pod?
  - What Operating Systems are currently being used?
  - What databases are involved?
  - Are there any high availability or disaster recovery requirements?
  - Are there integration requirements with other applications and environments?
- What are the network requirements?

The TDA then ensures that specific prerequisites for the installation and management of the Pod are met. Such as:

- Base prerequisites
  - Is the proposed installation in one of our currently supported countries?
  - Is the client location able to have continuous connectivity to IBM Cloud for management, provisioning, Call Home, and support? (Direct Link Connect or Site-to-Site VPN)
  - Does the client REQUIRE external direct SAN Access to local storage or can the storage be accessed via the TCP/IP network?
  - Is the management network latency able to meet the 200 msec maximum delay?
- Physical Planning
  - Can the facility support the 60 amp single or three phase power drops required for the Racks
  - Is the facility able to support the cooling requirements of the racks?
  - Are sufficient network drops available to connect to the racks and can TCP/IP Network addresses be allocated?
- Sizing
  - Does the Client expect their workload on Power to Grow?
  - Does the proposed sizing of the configuration provide room for anticipated growth?
  - Does the proposed sizing meet the sizing for minimum commit?
  - Is the Power Virtual Server Private storage being proposed sufficient for client?
- Installation and service
  - Can IBM get access to the facility for maintenance and support of hardware?
  - Is the client location able to accept delivery of full IBM 42U racks of equipment? (for example: loading docks, elevators)?
- Compliance and Security
  - Are there any compliance standards required (for example:FFIEC, PCI, HIPAA, PII, ISO)
  - Are there any data residency or data protection regulation (GDPR) requirements for the Power servers and Storage?

After any issues found in the TDA are addressed, the system can be ordered and scheduled for delivery to the customer data center. Once the Pods are delivered, IBM is responsible for setting them up and connecting them to the IBM Cloud to enable management of the environment as discussed in section 4.4, "IBM installation and decommission services" on page 112.

4.3 Customer responsibilities

This section describes the different tasks that need to be completed throughout the ordering and installation process and assigns the responsibility for those steps to the customer or to IBM.

4.3.1 Prerequisites for installing the Pod

There are several prerequisites that must be met in order to complete the installation of an IBM Power Virtual Server Private Cloud pod. IBM will work with the client to ensure that all requirements have been satisfied.

Preorder Steps

This group of requirements need to be satisfied prior to the order being accepted.

Customer Responsibilities

1. Identify the closest IBM Cloud region to minimize network latency. To be successful, the round-trip time (RTT) for a data packet needs to be less than 200 milliseconds. For more information on determining the RTT from your location to the chosen IBM Cloud region see Power IaaS Network Latency.
2. Confirm that site and environmental conditions are suitable for Pod installation. Validate that you have floor space available, have the power capacity required, have the appropriated heat and air conditioning capacity, and have the appropriate security controls in place. For more information work with your IBM representative and consult the Power IaaS Preinstallation Checklist.

IBM Responsibilities

1. Provide information and resources to assist in site planning and environmental requirements.

Order and Preinstallation

After the order has been placed an prior to the installation, the following tasks need to be completed.

Customer Responsibilities

1. Complete preinstallation checklist provided by IBM.
2. Prepare data center site according to the checklist requirements, including space for 4/2U racks, floor load capacity, HVAC setup, and power source readiness
3. Ensure security measures and access control are in place.
4. Route power and network cables to the installation site.

IBM Responsibilities

1. Provide a preinstallation checklist.
2. Conduct an installation readiness review

Networking Setup

Prior to installation, the following network setup tasks need to be completed.

Customer Responsibilities

1. Provide network-specific information such as ASN and service key. For more information see Power IaaS Network Requirements.
2. Establish connections using IBM Direct Link or VPN.
3. Work with a service provider for last mile connectivity.

IBM Responsibilities

1. Assist with network use case identification and requirements communication.

Installation and Activation

These tasks need to be completed during the installation and activation stage, after the Pod has arrived at the customer data center.

Customer Responsibilities

1. Work with IBM SRE team for physical cabling and initial configuration of the data plane network.

IBM Responsibilities

1. Install, upgrade, and update hardware and software for the Pod infrastructure.
2. Configure the network for validation testing.
3. Conduct provisioning tests over the control plane.
4. Activate the account post-testing, marking the start of billing.
5. Ensure visibility of the Pod Satellite location in the customer's IBM Power Virtual Server (On-premises) account.

Post-Installation Testing

After the installation, the following tasks and tests must be completed.

Customer Responsibilities:

1. Perform provisioning tests using the service broker to ensure functionality (VM provisioning, IP address assignment, basic command operations).

IBM Responsibilities:

1. Support through IBM Support Center for any connectivity or installation issues post-setup.

4.3.2 Additional customer requirements

This section addresses some additional areas that need to be considered before the IBM Power Virtual Server Private Cloud Pod can be installed.

Site Readiness

- Ensure that the IBM Power Virtual Server Pod is protected with restricted access that is consistent with your company data protection and physical access control policies.
- Ensure that electric power and communication facilities are available in adequate quantities for operation. If these facilities are inadequate, contact the utility company to determine whether additional services can be made available.

Site Access Requirements

- Define an access route from your loading dock to your data processing area before delivery of your server. A preliminary check of the building is required to determine if adequate access for the normal delivery of supplies and servers exists. A small alley, a narrow door opening, or limited access to the delivery area can become inhibitive to installation. The loading dock, passageways, and elevators must be able to accommodate heavy, oversized data processing support equipment such as air conditioning equipment. For more information see Planning Access.

Power requirements

- Your On-premises site for IBM Power Virtual Server must be provisioned with A-side and B-side power redundancy that meets the Power Virtual Server rack connector and load requirements. See Power Requirements to determine the rack connector and load requirements.

Network Requirements

- The data center site must provide network cables to connect the IBM Power Virtual Server (On-premises) network infrastructure and the data network at the site.
- The site must provide two uplink cables to connect the IBM Power Virtual Server (On-premises) network infrastructure to the IBM Cloud region through IBM Direct Link connections or through VPN connections.
- Contract with a service provider to:
  - Provide redundant connections to the IBM Direct Link connection or VPN connection.
  - Provide the last mile connection from the point-of-presence (PoP) of your service provider to the customer data center.

IBM provides networking connection between the components within the Pod, but the client needs to provide connections between the Pod and their internal data network at the site. In addition, the customer is responsible for providing connections from the Pod to the IBM Cloud region. This is shown in Figure 4-2.

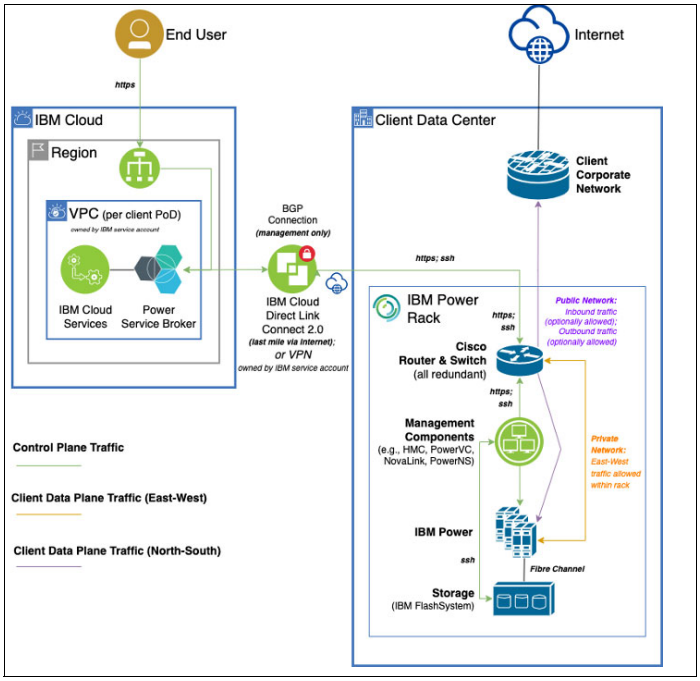


Figure 4-2 Networking architecture

Networking requirements are discussed in section 3.5, “Network connectivity” on page 80. For additional information see Network Requirements.

4.4 IBM installation and decommission services

IBM is responsible for deploying and managing IBM Power Virtual Server Private Cloud installations and is responsible for removing the equipment when the contract expires.

4.4.1 Deployment services

Deployment services include:

- Physical infrastructure installation
- Configuration into IBM Cloud
- Service activation.

The client is responsible for providing and maintaining the physical facilities, including the floor space, power, and HVAC components to allow the equipment to operate. The client is also responsible to ensure that the equipment racks can be delivered and ensure that the racks can be moved to and placed properly in the data center. This includes ensuring that the racks have a clear delivery path from the dock to the computer room floor and providing any ramps and floor support as necessary.

When the equipment is in place, IBM service technicians will come to the client data center and will complete the final installation steps to ensure that the racks are in the correct location and if there are multiple racks, will install the interconnecting cables.

After the racks are installed and the equipment is powered on, the IBM technicians will work with the IBM Cloud support team to connect the Pod to the cloud and activate the IBM Cloud services.

4.4.2 End of services

At the end of the contract, IBM is responsible for deactivating the IBM Cloud connections and remove the equipment from the client data center. IBM is responsible for:

- IBM data disposal,
- Services disconnect
- Hardware tear down and packaging
- Return to IBM

The client is responsible for migrating all client data from the Pod storage and deactivating all applications, prior to IBM coming to deactivate the pod.

5

Setup and operation examples for Power Virtual Server Private Cloud

This chapter is designed to show examples of setting up and operating the IBM Power Virtual Server Private Cloud. As described earlier, managing IBM Power Virtual Server whether in the public cloud or in your data center is, by design, done using the same interfaces. This synergy and ease of management allows you to choose the appropriate location for your IBM Power Virtual Server workloads and easily migrate workloads between on-premises and public cloud locations.

The following topics are included in this chapter:

5.1 Introduction

With IBM Power Virtual Server you can run IBM Power based workloads in the cloud or on your data center floor with the same management tools. The only difference between running on IBM Power Virtual Server or IBM Power Virtual Server Private Cloud is the amount of time it takes to provision your first resources as you need to work with IBM to order and install the infrastructure for your IBM Power Virtual Server Private Cloud pod.

IBM Power Virtual Server Private Cloud provides businesses with a flexible, scalable cloud infrastructure tailored specifically for running IBM Power Systems workloads. Power Virtual Server Private allows enterprises to leverage cloud computing benefits while maintaining control over critical workloads that require high security, performance, or compliance with regulatory requirements. It is designed to run mission-critical workloads, such as AIX, IBM i, and Linux environments, enabling businesses to extend or replace their on-premises infrastructure with cloud-based resources.

Power Virtual Server Private offers robust security features, including data encryption (both at rest and in transit), firewalls, and role-based access controls. It also complies with a range of industry regulations such as HIPAA, PCI DSS, and GDPR, making it suitable for industries like healthcare, finance, and government. By hosting sensitive workloads in a dedicated environment, businesses can ensure compliance with stringent data protection laws and maintain control over where their data is stored and processed.

Power Virtual Server Private Cloud provides on-demand scalability, allowing businesses to provision additional resources such as compute, memory, and storage when needed. This ensures that organizations can easily scale their infrastructure to meet changing business demands without the need for additional capital investment. Resources can be scaled up or down depending on workload requirements, making it highly adaptable to both predictable and unexpected changes in demand.

Power Virtual Server Private Cloud is designed to support high availability for mission-critical applications. It provides disaster recovery options, allowing businesses to replicate workloads and data across multiple data centers for redundancy. This ensures minimal downtime and business continuity in the event of failures or disasters.

Power Virtual Server Private Cloud is ideal for businesses pursuing a hybrid cloud strategy. It integrates seamlessly with public clouds, allowing businesses to extend their infrastructure to the cloud while keeping critical workloads on-premises or in a private environment. Workloads can be moved easily between on-premises and cloud environments, providing the flexibility to manage workloads where they are most efficient and cost-effective.

By adopting a pay-as-you-go model, Power Virtual Server Private allows businesses to manage costs effectively by only paying for the resources they consume. This avoids the need for over-provisioning resources and ensures that IT budgets align with actual infrastructure usage. Power Virtual Server Private Cloud reduces the need for upfront capital expenditures and allows organizations to scale resources based on business needs.

This section introduces how businesses can get started with Power Virtual Server Private Cloud, including the features, capabilities, and core benefits it offers. By leveraging Power Virtual Server Private, organizations can enjoy cloud-like flexibility without compromising on control, security, or performance, making it an ideal solution for hybrid cloud strategies.

5.2 Setting up your workspace

A workspace in IBM Cloud is a logical container to group a set of cloud resources. It can be used to group resources for specific projects and to simplify management of your cloud resources.

All of the resources defined in a workspace are located in a single cloud datacenter or location. The location is specified when you create the workspace. You can define multiple workspaces within a single datacenter or you can specify multiple workspaces across different data centers. Resources cannot be moved or shared across workspaces.

With the availability of Power Virtual Server Private cloud, the Create Workspace has a new parameter to define the location, the Location type field allows you to select:

- IBM data center to select from one of the IBM datacenter locations
- Client location which allow you to select a satellite location registered to your enterprise.

Figure 5-1 shows the Create Workspace screen with Client location selected.

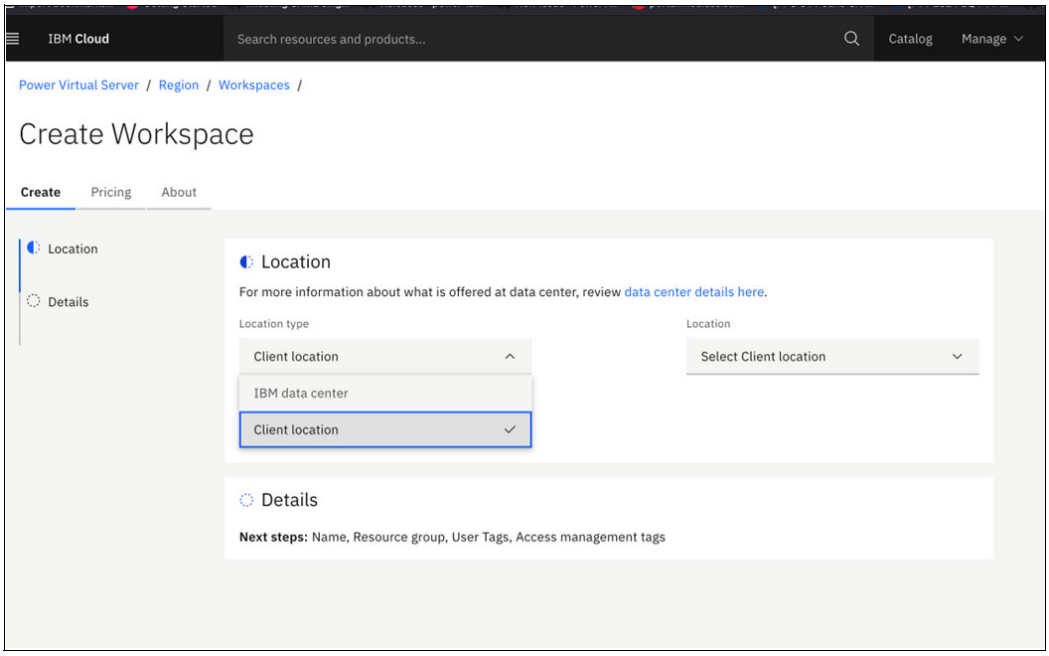


Figure 5-1 Selecting IBM data center or Client location

After choosing from IBM data center or Client location, the Location field on the page is populated with the options available for your IBM Cloud account. Figure 5-2 on page 116 shows the client locations list. Choose the location that contains the resources that you want to include in your workspace.

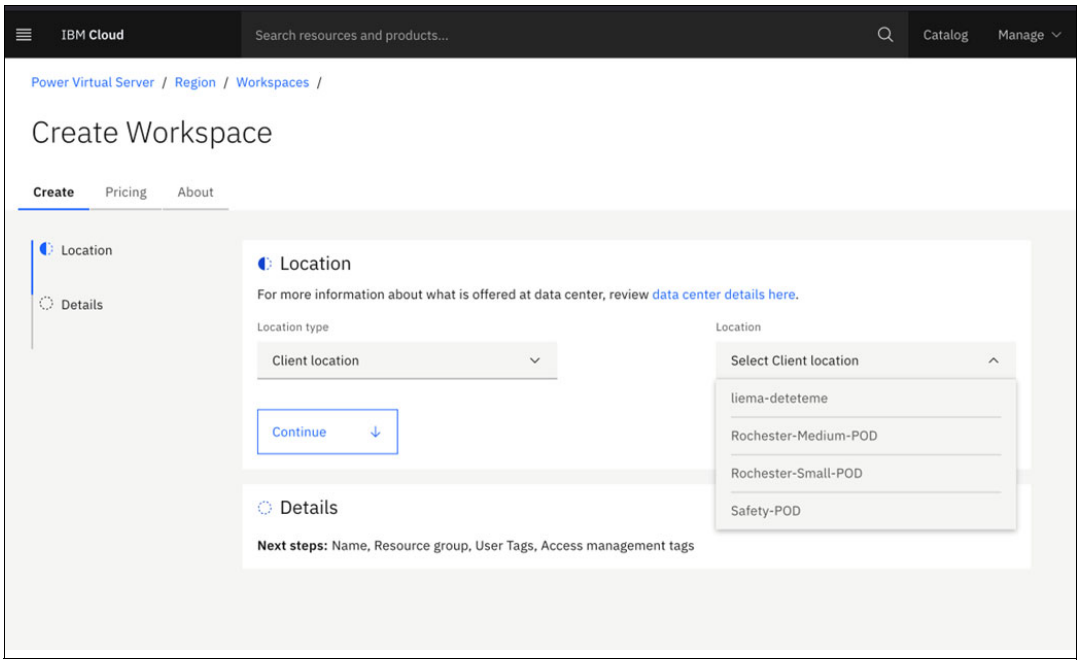


Figure 5-2 Choosing a client location

Once you select your location type satellite location, select Continue to fill in additional data about your workspace as shown in Figure 5-3.

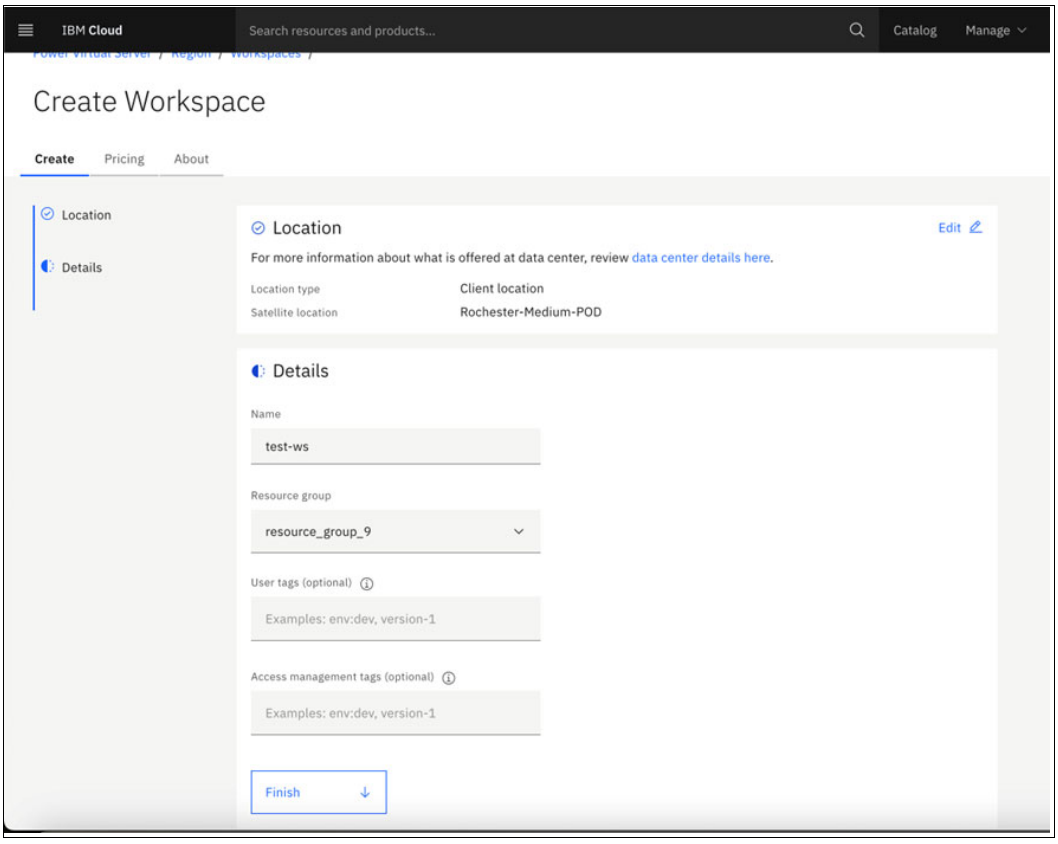


Figure 5-3 Additional information for workspace definition

Required fields are the Name and Resource group fields. The name is defined by you and should be used to differentiate this workspace for others you have in your environment. The User tags and Access management tags fields are designed to help you manage your IBM Cloud resources.

Authorized users can add tags to resources or service IDs in the account. By tagging your resources, you can better organize, track, and manage costs associated with related assets. Using a consistent tagging schema to link resources to specific teams allows you to group and filter by those tags when analyzing costs in your exported usage report.

Access management tags help control access to resources. These tags can be predefined for use in access policies that grant permissions to the resources they are attached to. Only the account administrator has the ability to create these tags and can delete them only if they are not linked to any resources.

The last option displayed as you create your workspace is whether you want to enable IBM Cloud Monitoring for your workspace.

5.3 Setting up a virtual server instance

The next step in setting up your Power Virtual Server Private Cloud resources is to define a virtual server instance – this is machine LPAR which will be used to run your business applications or databases. This process is exactly the same for resources in your private Pod or on the public cloud. We have documented many of the steps involved here. For additional information, see Creating an IBM Power Virtual Server.

Note: Before you can create a Virtual Server Instance, you need to select the workspace that will contain the VSI. Select the workspace from the left screen panel by selecting "**Select workspace**" and then selecting the workspace you want to contain your new VSI.

Figure 5-4 shows the initial screen used to define a virtual server instance. On this screen you provide a name for your instance and define how many you want to create with these specifications (Number of instances). You also select the boot image to be used for the instance, choosing from standard boot images provided by IBM or from boot images you have uploaded to the IBM Cloud.

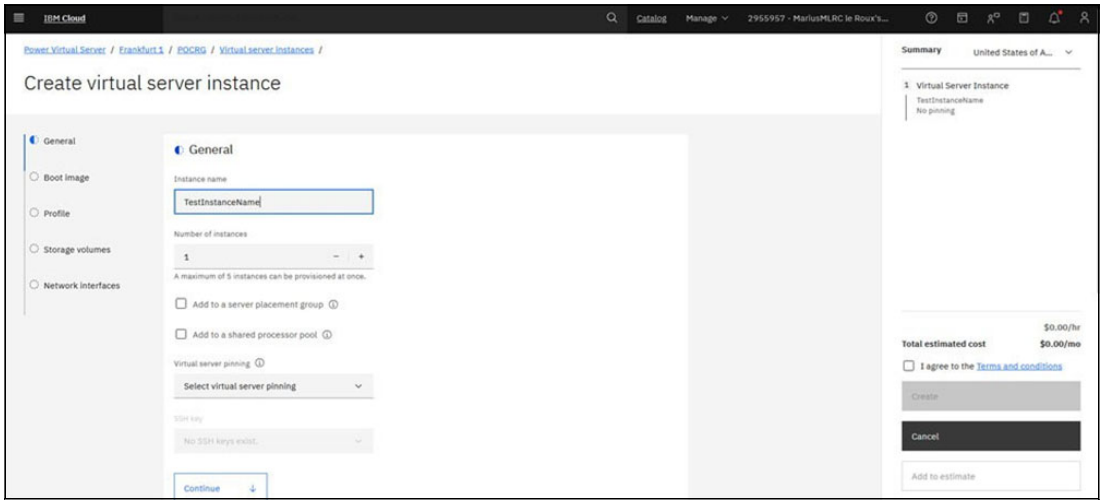


Figure 5-4 Creating a virtual server instance

Additional screens are presented as you continue, allowing you to define the compute, memory, storage, and network resources to be used by your instance.

In the General tab, the required fields are Instance name and Number of instances. The name is a user defined label that should help you identify the instance and its workloads. Only alphanumeric characters, underscores and dashes are allowed in the name. This would be the equivalent to the LPAR name in your existing Power VM environment. If you want to create multiple VSI instances with the same parameters, select the number you want to create.

Additional options available for selection are:

- Add to a server placement group.

  Server placement groups provide you control over the host or server on which a new virtual machine (VM) is placed. By using server placement groups, you can build high availability within a data center. You can apply an affinity or anti-affinity policy to each VM within a server placement group. After you create a placement group, you can provision a new VM in the placement group. When you set a placement group with an affinity policy, all VMs in that placement group are provisioned on the same server. When you set a placement group with an anti-affinity policy, all VMs in that placement group are provisioned on different servers. More information can be found in Managing server placement groups. When the check box is checked, a selection panel drops down to select the server placement group.

- Add to a shared processor pool.

  Shared processor pools within an IBM PowerVM environment allow more efficient sharing of processor cores across multiple LPARs. This can provide better efficiency and reduce software licensing charges. For more information see Managing the shared processor pool. When the check box is checked, a selection panel drops down to select the shared processor pool.

- Virtual server pinning

  VM pinning allows you to control the movement of VMs during disasters and other restart events. Hard pinning is recommended for applications with serial number-based licenses.

  You can choose a pinning policy: soft pin or hard pin, to pin a VM to the host where it is running. When you soft pin a VM for high availability, PowerVC automatically migrates the VM to the original host. The PowerVC is migrated when the host is back to its operating state. When you hard pin a VM, the movement of the VM is restricted if the VM has a licensing restriction with the host. The VM movement is restricted during remote restart, automated remote restart, DRO, and live partition migration. The default pinning policy is none.

- Select SSH key

  Select an SSH key to use to access the VSI after provisioning. For more details see Generating and using an SSH key.

When you are finished with this section, press **Continue.** You can come back and make changes by selecting **Edit** in the upper right corner of the General tab. Figure 5-5 on page 119 is an example of the next screen displayed. This screen allows you to enter information about the operating system to be run in the VSI.

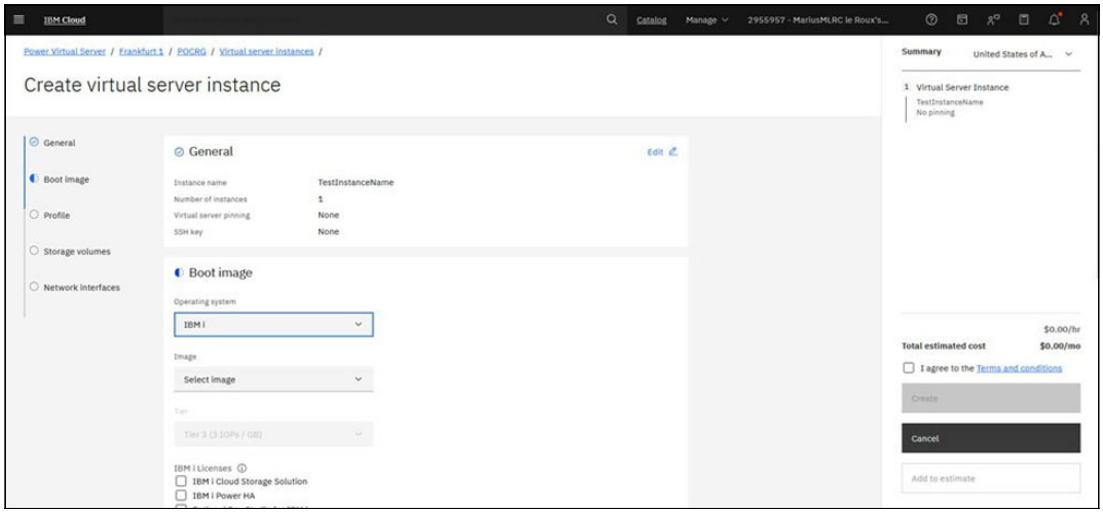


Figure 5-5 Boot image selection screen

In this example we have selected IBM i as the operating system. Other available options are: AIX, and Linux.

There are also options to provide your own license for Linux implementations. If you bring your own image, you must obtain the subscription for Linux directly from the vendor. After you deploy your Linux VM, you must log in to the VM and register it with the Linux vendor's satellite server. To reach the Linux vendor satellite servers (where you can register and obtain packages and fixes), you must attach a public network to your VM. To learn more about the registration process, see Registering and subscribing to SLES or Registering and subscribing to RHEL.

Once you have selected your operating system, you must select an image. You need to choose from a range of IBM supplied options, or if you chose to bring your own license, an image with a BYOL suffix.

There are additional options to choose to help customize your installation depending on the operating system.

Figure 5-6 shows the Boot image section filled in for our IBM i example.

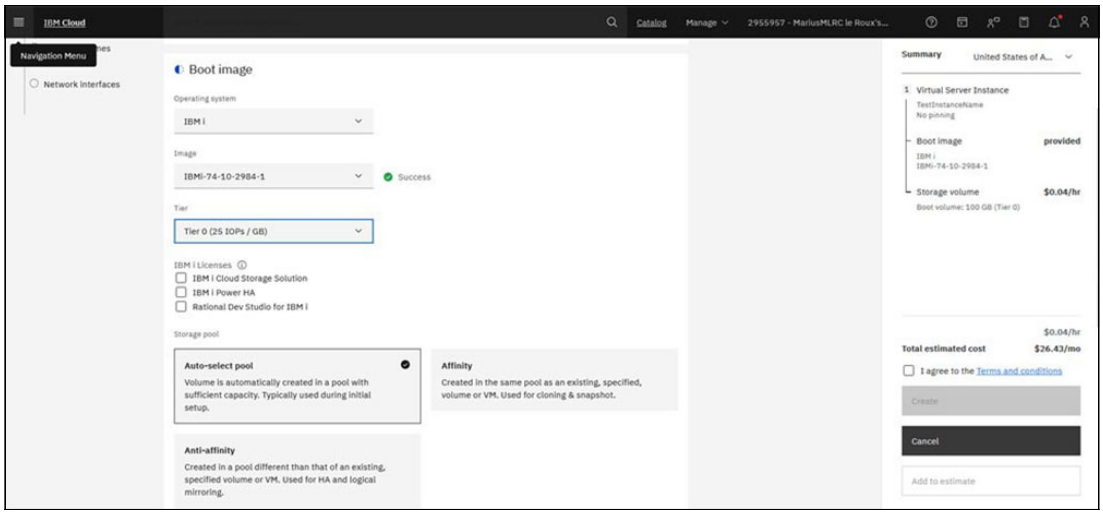


Figure 5-6 Boot image selection screen with entries

After selecting your image, you are then provided options for the storage tier that will be provisioned for your boot volume. The storage tier options were discussed in “Storage tiers” on page 55. For each operating system, there may be additional selections, for example in our IBM i instance, we can add additional IBM product licenses.

At the bottom of the screen, as shown in Figure 5-7 are some advanced configuration options that can be selected.

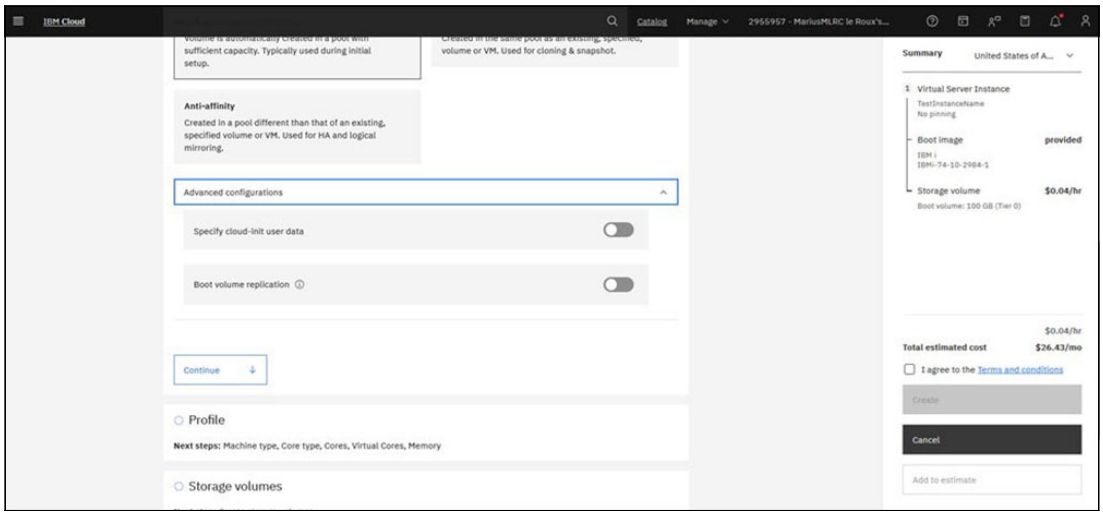


Figure 5-7 Additional configuration options

The Specify cloud init user data allows you to specify a script to be run at initialization to further customize your image. The Boot volume replication allows you to set up replication of your boot image for high availability.

Pressing **Continue** takes you to the next section where you define the hardware profile of your virtual server. This is shown in Figure 5-8.

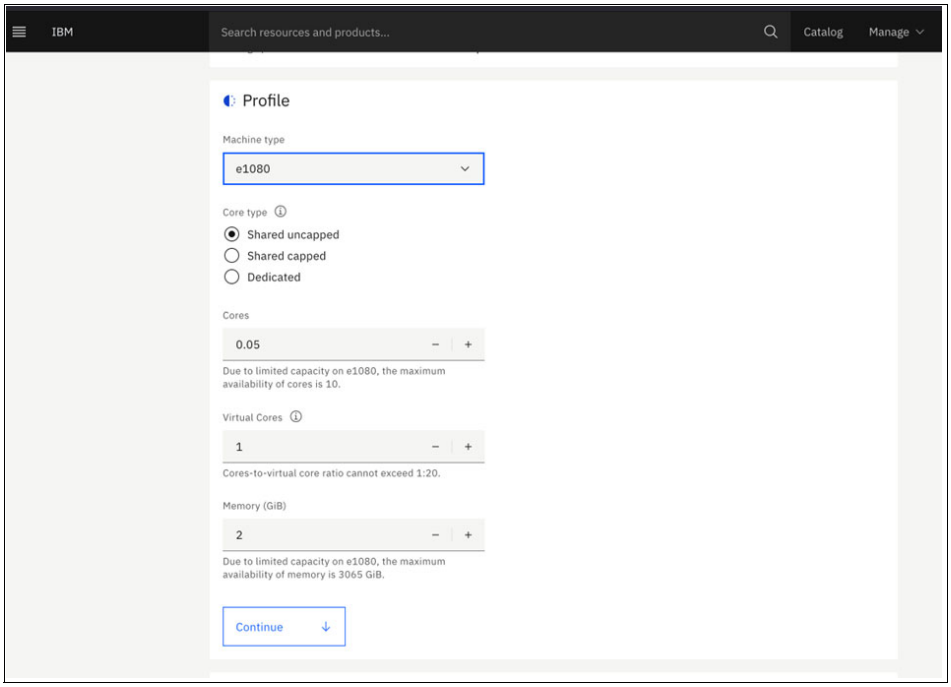


Figure 5-8 Machine profile selection screen

For most operating systems, you are free to customize your machine configuration as needed.

Hitting **Continue** takes you to the next section where you define storage for your instance. This is shown in Figure 5-9. If you do not want to define the storage configuration at this time, you can come back later and add the appropriate storage resources to the virtual server.

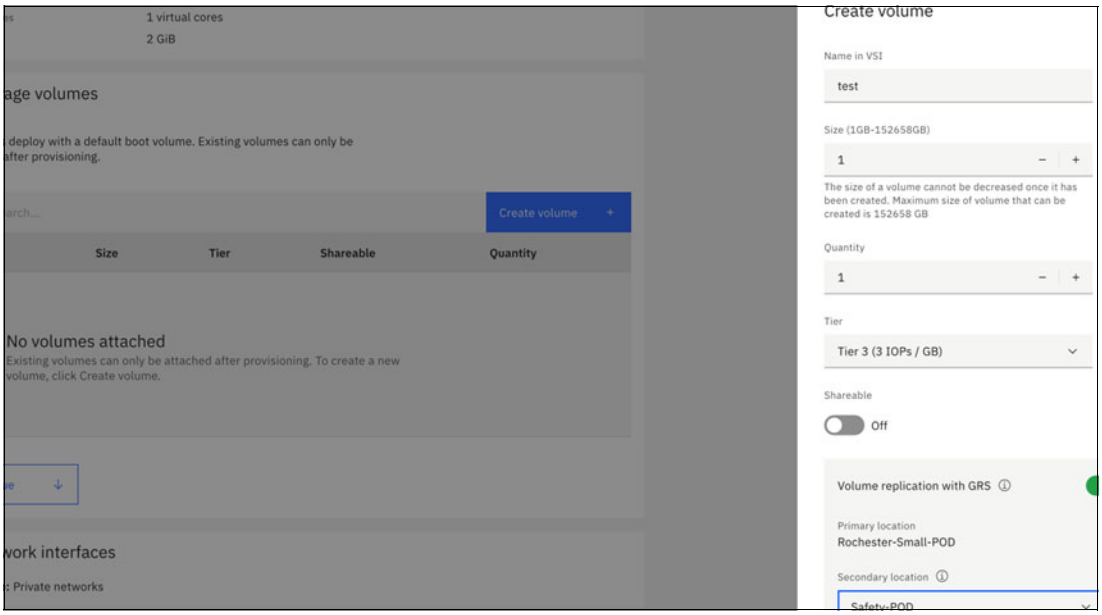


Figure 5-9 Define storage for virtual server

The last section allows you to define the network resources that are available to your virtual server. This is shown in Figure 5-10. Just like the storage configuration, you can initially turn on the Public networks option and then later come back to add additional network interfaces as needed.

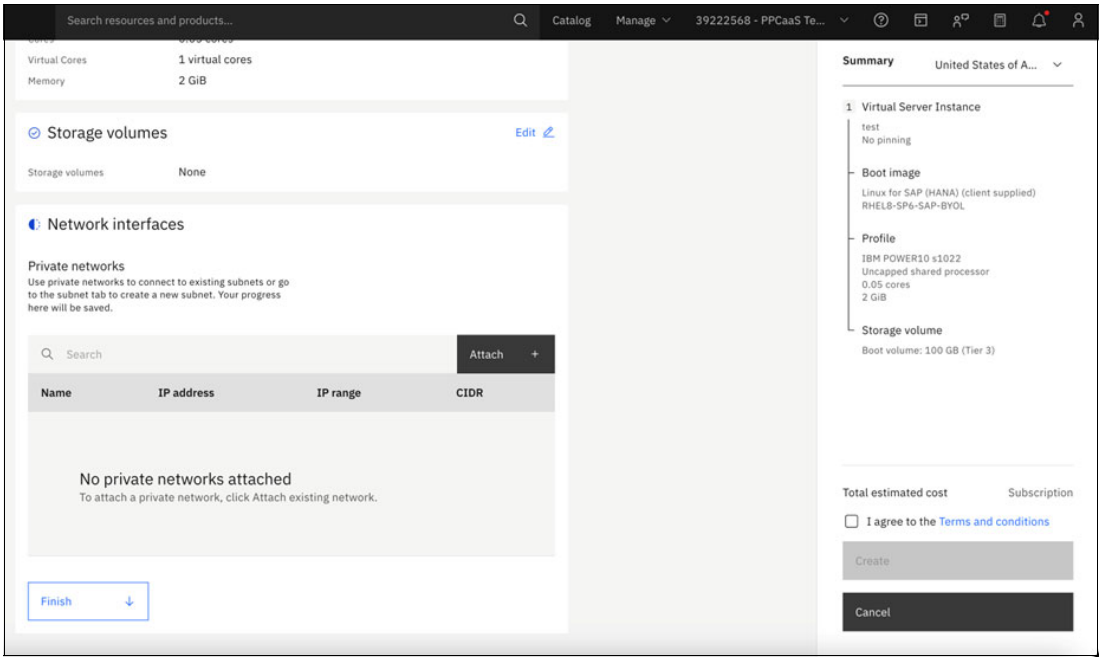


Figure 5-10 Define networking for your virtual server

If you would like to make any changes before initiating the provisioning of the virtual server you can select **Edit** on any of the sections. Once the virtual server instance is defined, you can select **Create** in the lower right corner. The screen also provides you a pricing summary for the virtual server as it is configured. It may take several minutes for the instance to be provisioned.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this paper.

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.

- *IBM Power Systems Virtual Server Guide for IBM i*, SG24-8513
- *IBM Power Virtual Server Guide for IBM AIX and Linux*, SG24-8512
- *IBM Power Systems Cloud Security Guide: Protect IT Infrastructure In All Layers*, REDP-5659
- *SAP HANA on IBM Power Systems Virtual Servers: Hybrid Cloud Solution*, REDP-5693
- *IBM Power Systems Private Cloud with Shared Utility Capacity: Featuring Power Enterprise Pools 2.0*, SG24-8478
- *IBM Power E1080 Technical Overview and Introduction*, REDP-5649
- *IBM Power E1050 IBM Power E1050 Technical Overview and Introduction*, REDP-5684
- *IBM Power 10 Scale Out Servers Technical Overview S1012, S1014, S1022s, S1022 and S1024*, REDP-5675
- *IBM Storage FlashSystem 9500 Product Guide: Updated for IBM Storage Virtualize 8.7*, REDP-5742
- *IBM Storage FlashSystem 5200 Product Guide for IBM Storage Virtualize 8.6*, REDP-5617

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:

- IBM Power Virtual Server Private Cloud: Solution overview

  https://www.ibm.com/downloads/cas/NNP2JPDA
- IBM Power Virtual Server Private Cloud: Security overview

  https://www.ibm.com/downloads/cas/6RP3MPL9

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

REDP-5745-00

ISBN

Printed in U.S.A.