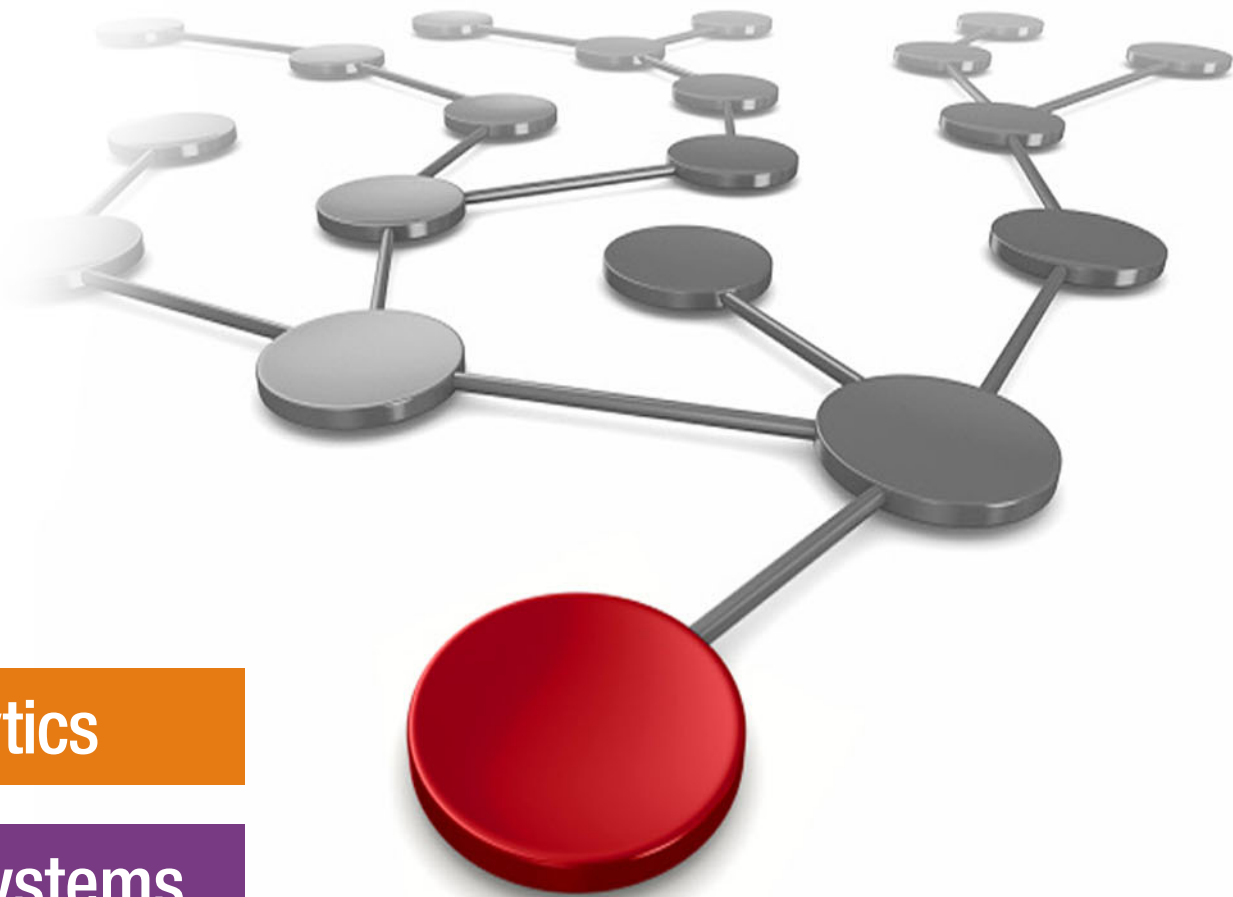


# IBM Power System AC922

## Introduction and Technical Overview

Alexandre Bicas Caldeira



 Analytics

Power Systems





International Technical Support Organization

**IBM Power System AC922 Introduction and Technical Overview**

March 2018

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

**First Edition (March 2018)**

This edition applies to the IBM Power System AC922 server models 8335-GTG and 8335-GTW.

**© Copyright International Business Machines Corporation 2018. All rights reserved.**

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

# Contents

<b>Notices</b> .....	v
Trademarks .....	vi
<b>Preface</b> .....	vii
Authors .....	vii
Now you can become a published author, too! .....	viii
Comments welcome .....	viii
Stay connected to IBM Redbooks .....	viii
<b>Chapter 1. Product summary</b> .....	1
1.1 Key server features .....	2
1.2 Server models .....	4
1.2.1 Power AC922 server model 8335-GTG .....	4
1.2.2 Power AC922 server model 8335-GTW .....	5
1.2.3 Minimum features .....	7
<b>Chapter 2. System architecture</b> .....	9
2.1 System architecture .....	10
2.2 Processor subsystem .....	15
2.2.1 POWER9 processor overview .....	16
2.2.2 Processor feature codes .....	18
2.3 Memory subsystem .....	18
2.3.1 Memory features and placement rules .....	18
2.3.2 Memory bandwidth .....	19
2.4 I/O subsystem .....	20
2.4.1 PCIe .....	20
2.4.2 IBM CAPI2 .....	21
2.4.3 OpenCAPI .....	22
2.4.4 The NVIDIA Tesla V100 .....	25
2.4.5 NVLINK 2.0 .....	28
2.5 PCI adapters .....	29
2.5.1 Slot configuration .....	29
2.5.2 Local area network adapters .....	30
2.5.3 Fibre Channel adapters .....	31
2.5.4 CAPI-enabled InfiniBand adapters .....	31
2.5.5 Compute-intensive accelerators .....	31
2.5.6 Flash storage adapters .....	32
2.6 System ports .....	32
2.7 Internal storage .....	32
2.7.1 Disk and media features .....	33
2.8 External I/O subsystems .....	33
2.9 Location codes .....	34
2.10 IBM System Storage .....	34
2.11 Operating system support .....	35
2.11.1 Ubuntu .....	35
2.11.2 Red Hat Enterprise Linux .....	35
2.11.3 More information .....	36
2.12 Java .....	36

<b>Chapter 3. Physical infrastructure</b> .....	37
3.1 Operating environment .....	38
3.1.1 Leak detection .....	40
3.1.2 Water pressure .....	41
3.2 Physical package .....	41
3.3 System power .....	41
3.4 System cooling .....	42
3.5 Rack specifications .....	46
3.5.1 IBM Enterprise Slim Rack 7965-S42 .....	46
3.5.2 AC power distribution units .....	51
3.5.3 Rack-mounting rules .....	53
3.5.4 OEM racks .....	53
<b>Related publications</b> .....	59
IBM Redbooks .....	59
Online resources .....	59
Help from IBM .....	59

# 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 <http://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®  
DS8000®  
Easy Tier®  
EnergyScale™  
IBM®  
IBM FlashSystem®  
OpenCAPI™

POWER®  
Power Systems™  
POWER9™  
PowerHA®  
PowerLinux™  
PowerVM®  
Real-time Compression™

Redbooks®  
Redpaper™  
Redbooks (logo) ®  
Storwize®  
System Storage®  
XIV®

The following terms are trademarks of other companies:

Linux is a trademark of Linus Torvalds 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.

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



# Preface

This IBM® Redpaper™ publication is a comprehensive guide that covers the IBM Power System AC922 server (8335-GTG and 8335-GTW models). The Power AC922 server is the next generation of the IBM Power processor-based systems, which are designed for deep learning and artificial intelligence (AI), high-performance analytics, and high-performance computing (HPC).

This paper introduces the major innovative Power AC922 server features and their relevant functions:

- ▶ Powerful IBM POWER9™ processors that offer 16 cores at 2.6 GHz with 3.09 GHz turbo performance or 20 cores at 2.0 GHz with 2.87 GHz turbo for the 8335-GTG
- ▶ Eighteen cores at 2.98 GHz with 3.26 GHz turbo performance or 22 at 2.78 GHz cores with 3.07 GHz turbo for the 8335-GTW
- ▶ IBM Coherent Accelerator Processor Interface (CAPI) 2.0, IBM OpenCAPI™, and second-generation NVIDIA NVLink technology for exceptional processor-to-accelerator intercommunication
- ▶ Up to six dedicated NVIDIA Tesla V100 GPUs

This publication is for professionals who want to acquire a better understanding of IBM Power Systems™ products and is intended for the following audiences:

- ▶ Clients
- ▶ Sales and marketing professionals
- ▶ Technical support professionals
- ▶ IBM Business Partners
- ▶ Independent software vendors (ISVs)

This paper expands the set of IBM Power Systems documentation by providing a desktop reference that offers a detailed technical description of the Power AC922 server.

This paper does not replace the current marketing materials and configuration tools. It is intended as an extra source of information that, together with existing sources, can be used to enhance your knowledge of IBM server solutions.

## Authors

This paper was produced by a specialist working at the International Technical Support Organization, Austin Center.

**Alexandre Bicas Caldeira** is a Certified IT Specialist and a former Product Manager for Power Systems Latin America. He holds a degree in computer science from the Universidade Estadual Paulista (UNESP) and an MBA in marketing. His major areas of focus are competition, sales, marketing, and technical sales support. Alexandre has more than 20 years of experience working on IBM Systems Solutions and has worked also as an IBM Business Partner on Power Systems hardware, IBM AIX®, and IBM PowerVM® virtualization products.

The project that produced this publication was managed by:

Scott Vetter  
**Executive Project Manager, PMP**

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

Adel El-Hallak, Volker Haug, Ann Lund, Cesar Diniz Maciel, Chris Mann, Scott Soutter, Jeff Stuecheli  
**IBM**

## 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 ITSO 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](http://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](http://ibm.com/redbooks)

- Send your comments in an email to:

[redbooks@us.ibm.com](mailto:redbooks@us.ibm.com)

- Mail your comments to:

IBM Corporation, International Technical Support Organization  
Dept. HYTD Mail Station P099  
2455 South Road  
Poughkeepsie, NY 12601-5400

## Stay connected to IBM Redbooks

- Find us on Facebook:

<http://www.facebook.com/IBMRedbooks>

- Follow us on Twitter:

<http://twitter.com/ibmredbooks>

- ▶ Look for us on LinkedIn:  
<http://www.linkedin.com/groups?home=&gid=2130806>
- ▶ Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:  
<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>
- ▶ Stay current on recent Redbooks publications with RSS Feeds:  
<http://www.redbooks.ibm.com/rss.html>





## Product summary

The IBM Power System AC922 is the next generation of the IBM POWER® processor-based systems, which are designed for deep learning and artificial intelligence (AI), high-performance analytics, and high-performance computing (HPC).

The system is a co-designed with OpenPOWER Foundation members and will be deployed at the most powerful supercomputer on the planet with a partnership between IBM, NVIDIA, Mellanox, and others. It provides the current technologies that are available for HPC, improving even more the movement of data from memory to GPU accelerator cards and back, enabling faster and lower latency data processing.

Among the new technologies the system provides, the most significant are the following ones:

- ▶ Two POWER9 processors with up to 40 cores (8335-GTG) or 44 cores (8335-GTW) and improved buses
- ▶ 1 TB of DDR4 memory (8335-GTG) with improved speed or 1 TB for the 8335-GTW
- ▶ Up to six NVIDIA Tesla V100 (Volta) GPUs, delivering up to 100 teraflops (TFLOPS) each, which is a 5x improvement compared to the previous generation
- ▶ Second-generation NVLINK with 2x throughput compared to the first generation

Because the massive computing capacity is packed into just 2Us of rack space, special cooling systems support the largest configurations. Therefore, to accommodate distinct data center infrastructure requirements, the system is available in two distinct models:

- ▶ 8335-GTG: Up to four GPUs and air-cooled
- ▶ 8335-GTW: Up to six GPUS and water-cooled

Figure 1-1 shows the front and rear views of a Power AC922 server.

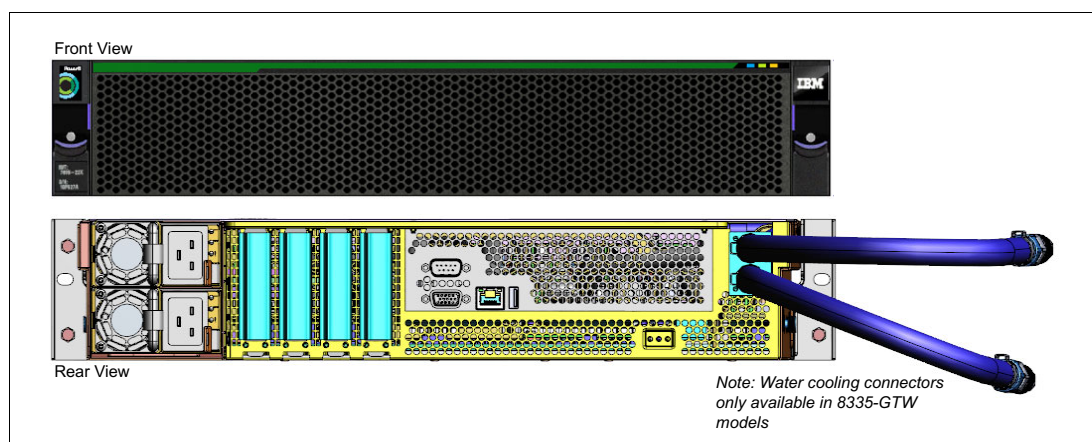


Figure 1-1 Front and rear views of the Power AC922 server

## 1.1 Key server features

The Power AC922 server addresses the demanding needs of deep learning and AI, high-performance analytics, and HPC.

An updated list of ported HPC applications that can use the IBM POWER technology is available at [IBM Power Systems HPC Applications Summary](#).

The system includes several features to improve performance:

- ▶ POWER9 processors:
  - Each POWER9 processor module has either 16 or 20 cores that are based on a 64-bit architecture:
    - Clock speeds for 16-core chip of 2.6 GHz (3.09 GHz turbo (8335-GTG))
    - Clock speeds for 20-core chip of 2.0 GHz (2.87 GHz turbo (8335-GTG))
    - Clock speeds for 18-core chip of 2.98 GHz (3.26 GHz turbo (8335-GTW))
    - Clock speeds for 18-core chip of 2.78 GHz (3.076 GHz turbo (8335-GTW))
  - 512 KB of L2 cache per core, and up to 120 MB of L3 cache per chip
  - Up to four threads per core
  - 120 GBps memory bandwidth per chip
  - 64 GBps SMP interconnect between POWER9 chips
- ▶ DDR4 memory:
  - Sixteen dual inline memory module (DIMM) memory slots
  - Maximum of 1024 GB DDR4 system memory (8335-GTG)
  - Improved clock from 1333 MHz to 2666 MHz for reduced latency
- ▶ NVIDIA Tesla V100 GPUs:
  - Up to six NVIDIA Tesla V100 GPUs, based on the NVIDIA SXM2 form factor connectors
  - 7.8 TFLOPs per GPU for double precision

- 15.7 TFLOPs per GPU for single precision
- 125 TFLOPs per GPU for deep learning, with new 640 Tensor Cores per GPU, which are designed for deep learning
- 16 GB HBM2 internal memory with 900 GBps bandwidth, 1.5x the bandwidth compared to Pascal P100
- Liquid cooling for six GPUs configurations to improve compute density
- NVLink 2.0:
  - Twice the throughput compared to the previous generation of NVLink
  - Up to 200 GBps of bidirectional bandwidth between GPUs
  - Up to 300 GBps of bidirectional bandwidth per POWER9 chip and GPUs, compared to 32 GBps of traditional Peripheral Component Interconnect Express (PCIe) Gen3
- OpenCAPI 3.0:
  - Open protocol bus to enable connections between the processor system bus in a high speed and cache coherent manner with OpenCAPI compatible devices, such as accelerators, network controllers, storage controllers, and advanced memory technologies
  - Up to 100 GBps of bidirectional bandwidth between CPUs and OpenCAPI devices
- Four PCIe Gen4 slots with up to 64 GBps bandwidth per slot, twice the throughput from PCIe Gen3, with three Coherent Accelerator Processor Interface (CAPI) 2.0 capable slots

Figure 1-2 shows the physical locations of the main server components.

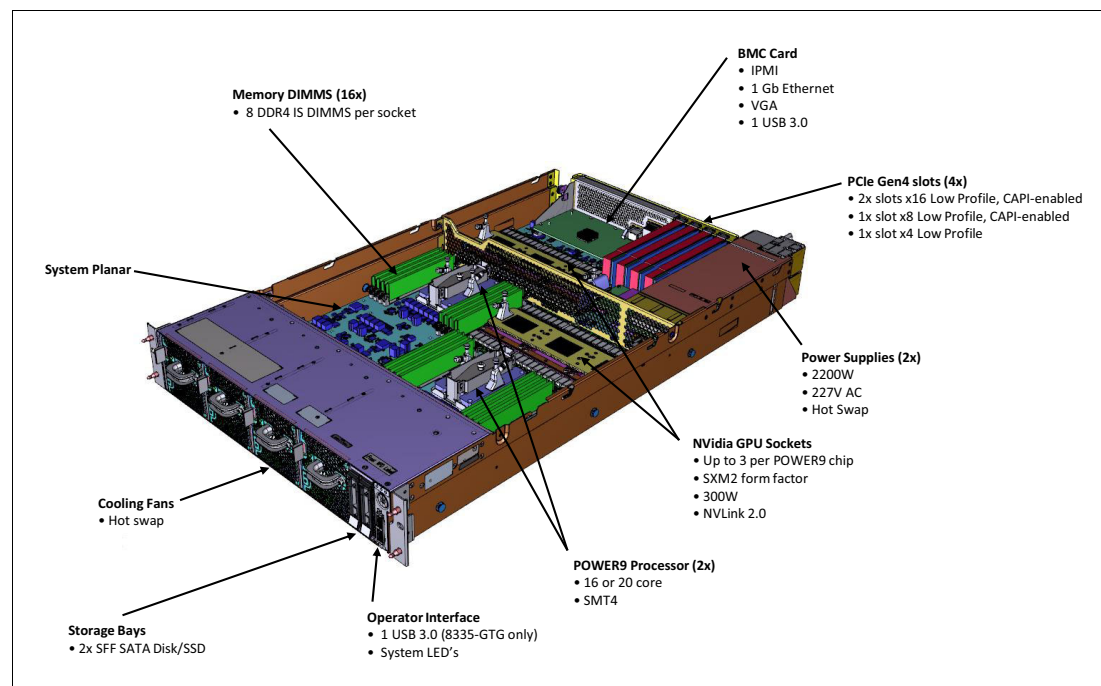


Figure 1-2 Location of server main components

## 1.2 Server models

The Power AC922 is manufactured as two distinct models, as shown in Table 1-1.

*Table 1-1 Summary of Power AC922 server available models*

Server model	POWER9 chips	Maximum memory	Maximum GPU cards	Cooling
8335-GTG	2	1 TB	4	Air-cooled
8335-GTW	2	1 TB	6	Water-cooled

### 1.2.1 Power AC922 server model 8335-GTG

This summary describes the standard features of the Power AC922 model 8335-GTG:

- ▶ 19-inch rack-mount (2U) chassis
- ▶ Two POWER9 processor modules:
  - 16-core 2.6 GHz processor module
  - 20-core 2.0 GHz processor module
  - Up to 1024 GB of 2666 MHz DDR4 error correction code (ECC) memory
- ▶ Two small form factor (SFF) bays for hard disk drives (HDDs) or solid-state drives (SSDs) that support:
  - Two 1 TB 7200 RPM NL SATA disk drives (#ELD0)
  - Two 2 TB 7200 RPM NL SATA disk drives (#ES6A)
  - Two 960 GB SATA SSDs (#ELU4)
  - Two 1.92 TB SATA SSDs (#ELU5)
  - Two 3.84 TB SATA SSDs (#ELU6)
- ▶ Integrated SATA controller
- ▶ Four PCIe Gen4 slots:
  - Two PCIe x16 Gen4 Low Profile slots, CAPI-enabled
  - One PCIe x8 Gen4 Low Profile slot, CAPI-enabled
  - One PCIe x4 Gen4 Low Profile slot
- ▶ Two or four NVIDIA Tesla V100 GPUs (#EC4J), based on the NVIDIA SXM2 form factor connectors air-cooled
- ▶ Integrated features:
  - IBM EnergyScale™ technology
  - Hot-swap and redundant cooling
  - Two 1 Gb RJ45 ports
  - One front USB 3.0 port for general use
  - One rear USB 3.0 port for general use
  - One system port with RJ45 connector
- ▶ Two power supplies (both are required)



The internal view of the fully populated server with four GPUs is shown in Figure 1-3. In this figure, the air baffles were removed to better show the major components.

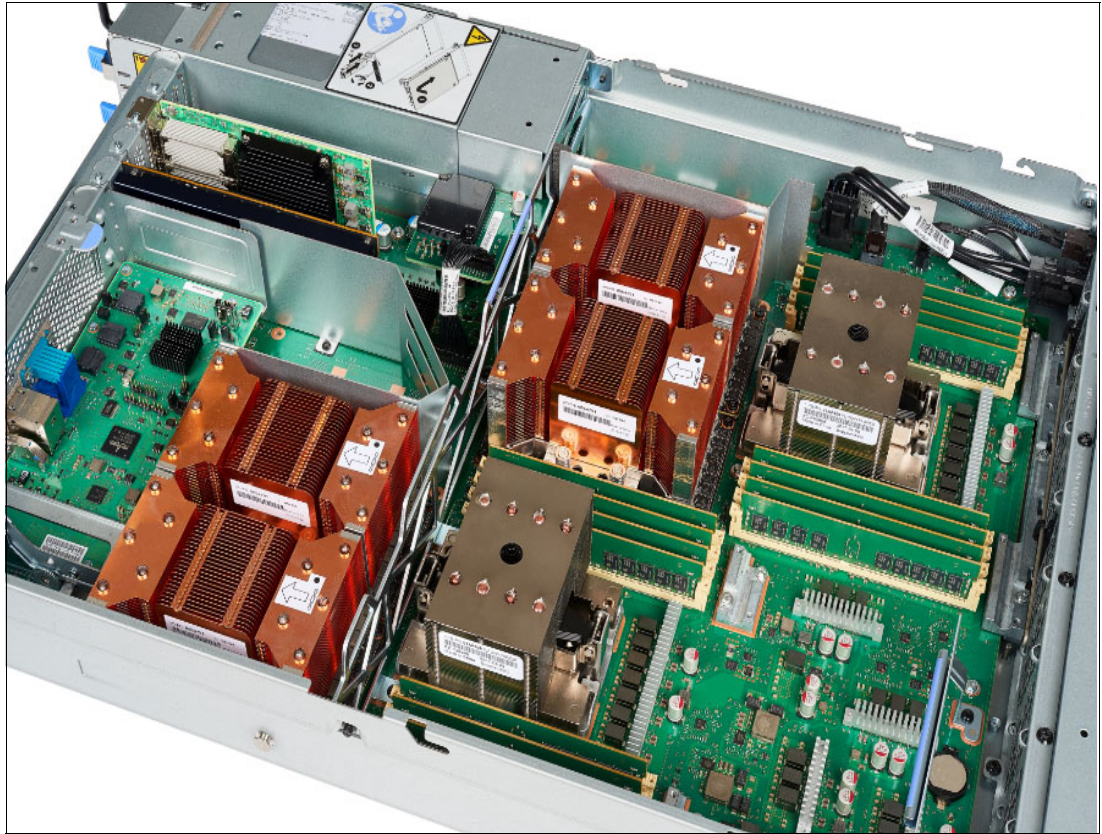


Figure 1-3 Power AC922 server model 8335-GTG fully populated with four GPUs

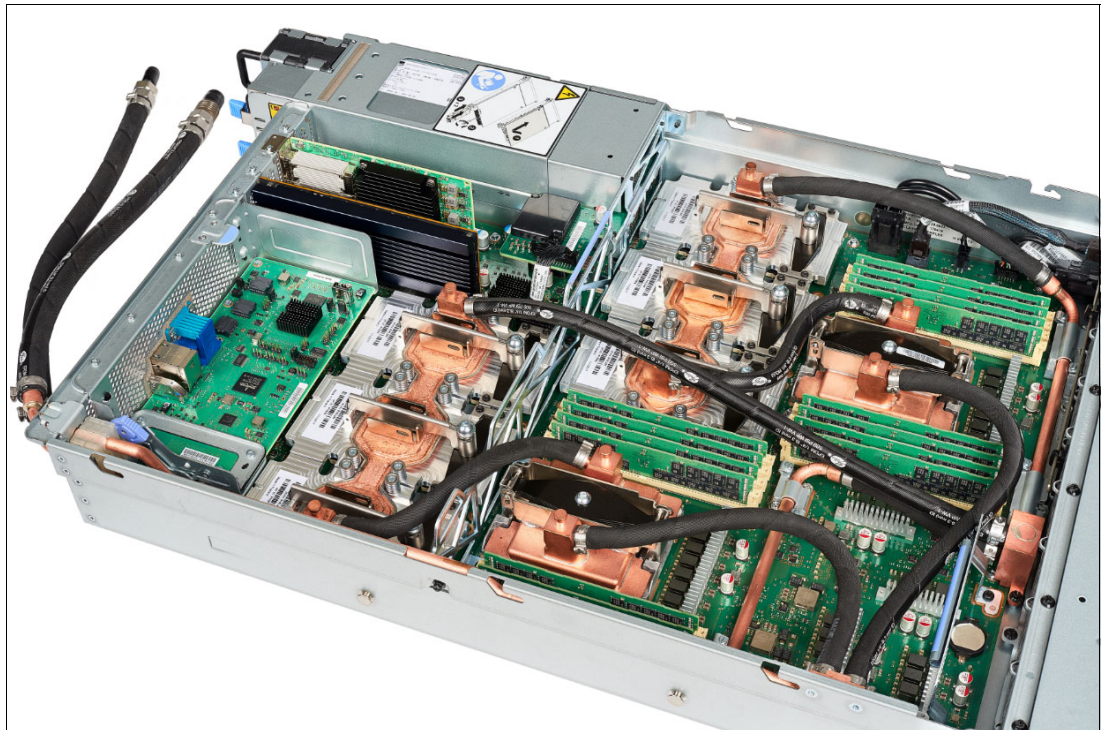
## 1.2.2 Power AC922 server model 8335-GTW

This summary describes the standard features of the Power AC922 model 8335-GTW:

- ▶ 19-inch rack-mount (2U) chassis
- ▶ Two POWER9 processor modules:
  - 18-core 2.98 GHz processor module
  - 22-core 2.78 GHz processor module
  - Up to 2048 GB of 2666 MHz DDR4 ECC memory
- ▶ Two SFF bays for HDDs or SSD that support:
  - Two 1 TB 7200 RPM NL SATA disk drives (#ELD0)
  - Two 2 TB 7200 RPM NL SATA disk drives (#ES6A)
  - Two 960 GB SATA SSDs (#ELU4)
  - Two 1.92 TB SATA SSDs (#ELU5)
  - Two 3.84 TB SATA SSDs (#ELU6)
- ▶ Integrated SATA controller

- ▶ Four PCIe Gen4 slots:
  - Two PCIe x16 Gen4 Low Profile slots, CAPI-enabled
  - One PCIe x8 Gen4 Low Profile slot, CAPI-enabled
  - One PCIe x4 Gen4 Low Profile slot
- ▶ Four or six NVIDIA Tesla V100 GPUs (#EC4J), based on the NVIDIA SXM2 form factor connectors water-cooled
- ▶ Integrated features:
  - IBM EnergyScale technology
  - Hot-swap and redundant cooling
  - Two 1 Gb RJ45 ports
  - One rear USB 3.0 port for general use
  - One system port with RJ45 connector
- ▶ Two power supplies (both are required)

The internal view of the fully populated server with six GPUs and the water-cooling system installed is shown in Figure 1-4. In this figure, the air baffles were removed to better show the major components.



*Figure 1-4 Power AC922 server model 8335-GTW fully populated with six GPUs*

**Note:** A Hardware Management Console (HMC) is not supported on the Power AC922 server.

### 1.2.3 Minimum features

The minimum initial order for the Power AC922 model 8355-GTG must include the following minimum features:

- ▶ Two processor modules with at least 16 cores each
- ▶ 256 GB of memory (sixteen 16 GB memory DIMMs)
- ▶ Two HDDs or SSDs
- ▶ Two #EC3L PCIe 2-port 100 Gbps Ethernet adapters
- ▶ Two #EC4J compute-intensive accelerators (NVIDIA V100 GPUs)
- ▶ Two power supplies and power cords (both are required)
- ▶ An Linux operating system (OS) indicator
- ▶ A rack integration indicator
- ▶ A Language Group Specify





# System architecture

This chapter describes the overall system architecture for the IBM Power System AC922 server. The bandwidths that are provided throughout the section are theoretical maximums that are used for reference.

**Note:** The speeds that are shown are at an individual component level. Multiple components and application implementation are key to achieving the preferred performance. Always do the performance sizing at the application-workload environment level and evaluate performance by using real-world performance measurements and production workloads.

## 2.1 System architecture

The Power AC922 server is a two single-chip module (SCM) system. Each SCM is attached to eight memory RDIMM slots. The server has a maximum capacity of 16 memory dual inline memory modules (DIMMs) which enables up to 1024 GB of memory.

The system board has sockets for four or six GPUs depending on the model, each of which is 300 watts capable. Additionally, the server has a total of four Peripheral Component Interconnect Express (PCIe) Gen3 slots, and three of these slots are Coherent Accelerator Processor Interface (CAPI)-capable.

Figure 2-1 shows the location of the processors, memory DIMMs, GPUs, and PCIe slots.

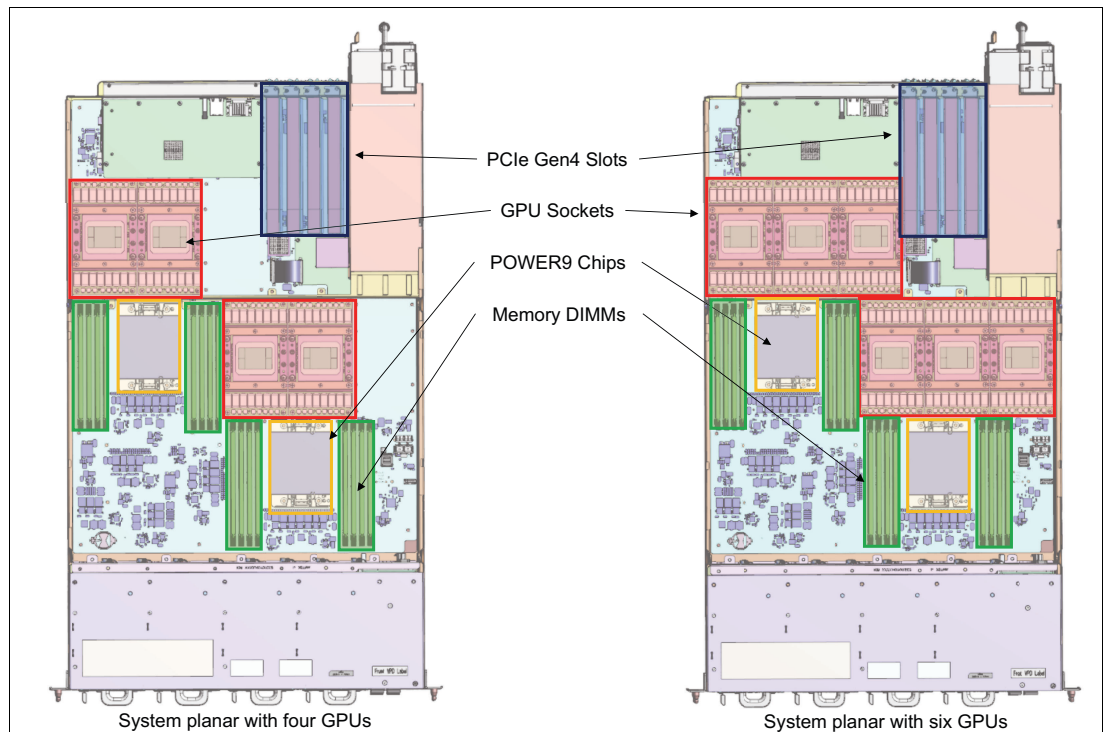


Figure 2-1 Component location for four GPU and six GPU system board



An integrated SATA controller is routed through a dedicated PCI bus on the main system board and enables up to two SATA hard disk drives (HDDs) or solid-state drives (SSDs) to be installed. Figure 2-2 shows the location of the integrated SATA connector. This bus also drives the integrated Ethernet and USB ports.

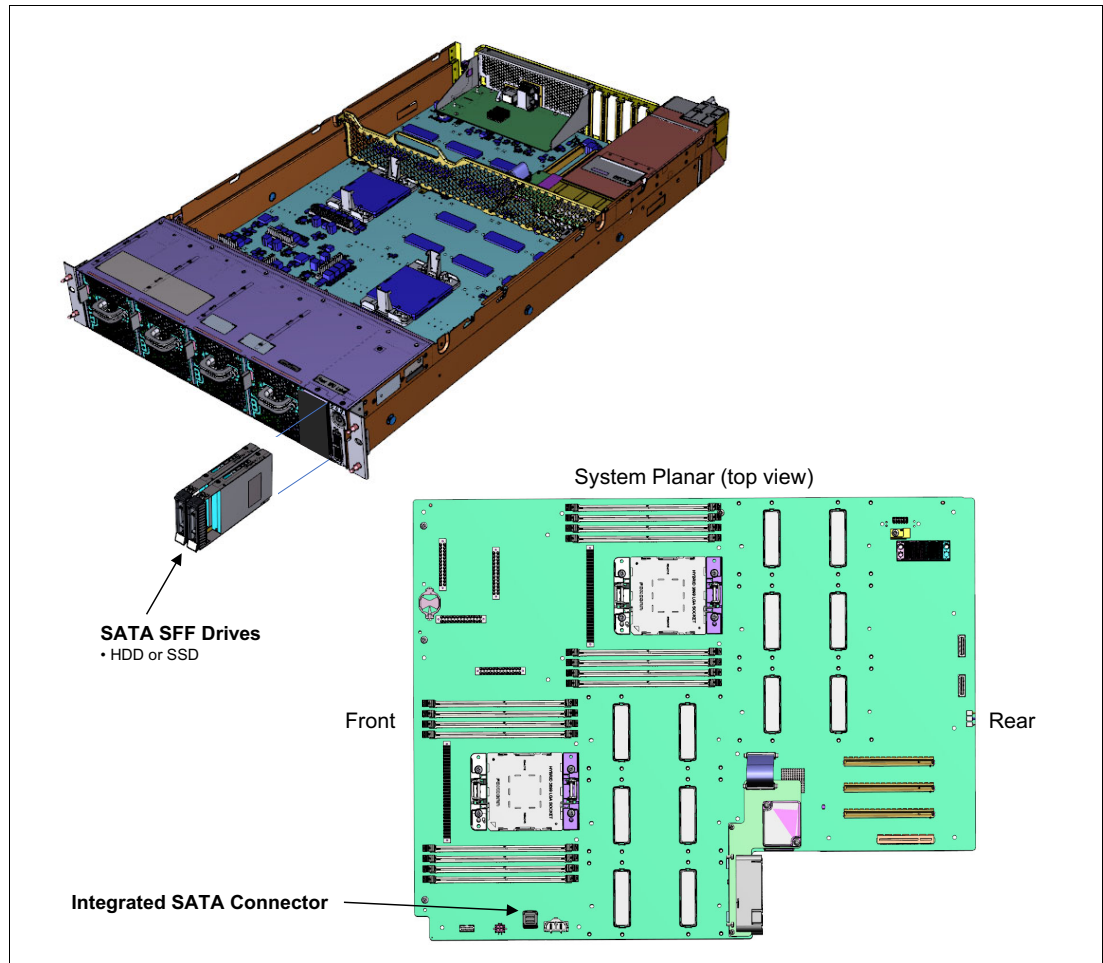


Figure 2-2 Integrated SATA connector

The POWER9 processor brings enhanced memory and I/O connectivity, improved chip to chip communication, and a new bus called NVLINK 2.0. Figure 2-3 shows the external processor connectivity.

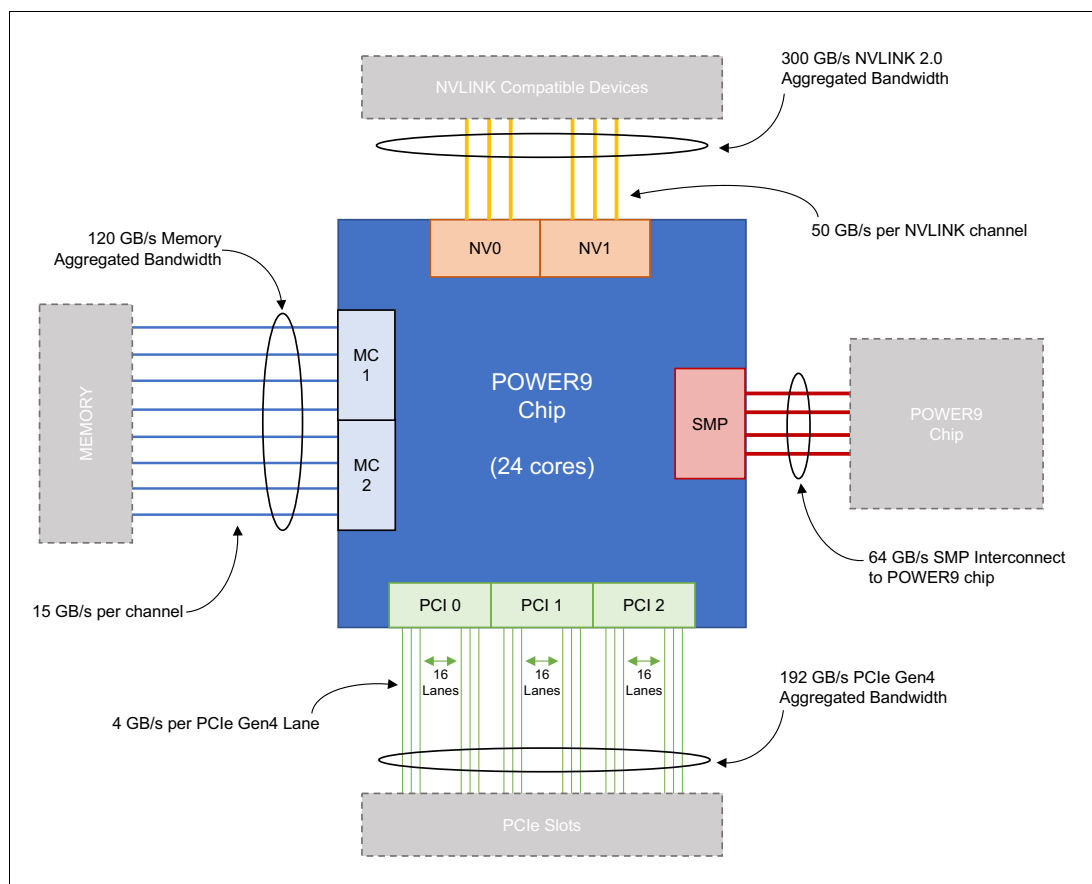


Figure 2-3 POWER9 chip external connectivity

Faster DDR4 memory DIMMs at 2666 MHz are connected to two memory controllers through eight channels with a total bandwidth of 120 GBps. Symmetric Multiprocessing chip-to-chip interconnect is done through a four channel SMP bus with 64 GBps bidirectional bandwidth.

The current PCIe Gen4 interconnect doubles the channel bandwidth of the previous PCIe Gen3 generation, enabling the 48 PCIe channels with a total of 192 GBps bidirectional bandwidth between the I/O adapters and the POWER9 chip.

The connection between GPUs and between CPUs and GPUs is done through a link that is called NVLINK 2.0, which was developed by IBM, NVIDIA and the OpenPOWER Foundation. This link provides up to 5x more communication bandwidth between CPUs and GPUs (compared to traditional PCIe Gen3) and enables faster data transfer between memory and GPUs and between GPUs. Complex and data-hungry algorithms such the ones that are used in machine learning can benefit from having these enlarged pipelines for data transfer after the amount of data that must be processed is many times larger than the GPU internal memory. For more information about NVLINK 2.0, see 2.4.5, “NVLINK 2.0” on page 28.

Each POWER9 CPU and each GPU have six NVLINK channels, called Bricks, with each one delivering up to 50 GBps bidirectional bandwidth. These channels can be aggregated to enable more bandwidth or more peer-to-peer connections.



Figure 2-4 compares the POWER9 implementation of NVLINK 2.0 with traditional processor chips that use PCIe and NVLINK.

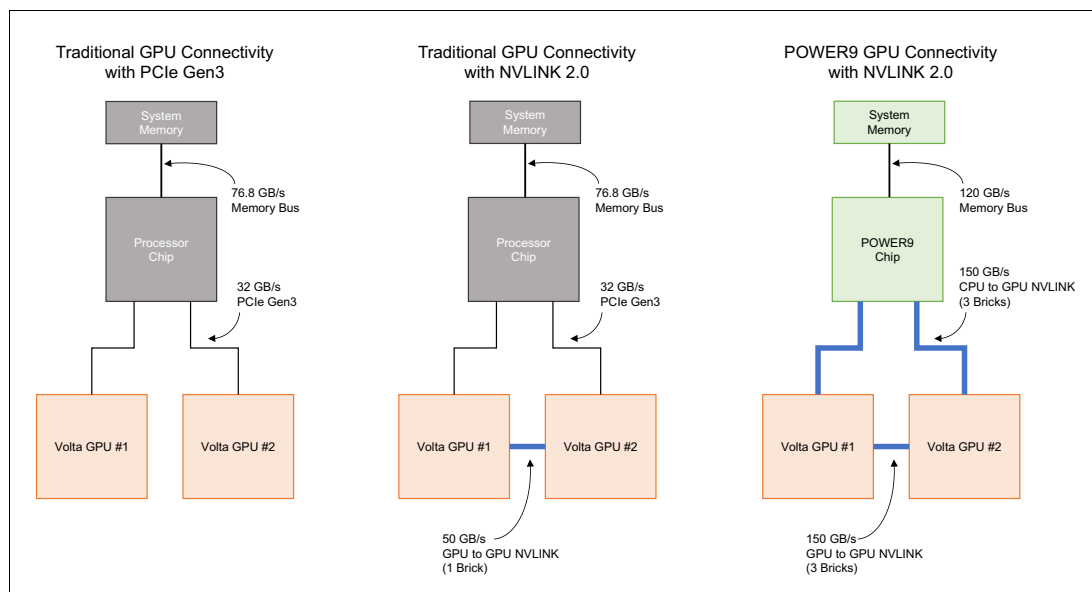


Figure 2-4 NVLINK 2.0 POWER9 implementation versus traditional architectures

On traditional processors, communication is done through PCIe Gen3 buses. When the processor must handle all the GPU-to-GPU communication and GPU-to-system memory communication, having more than two GPUs per processor potentially created a bottleneck on the data flow from system memory to GPUs.

To reduce this impact on the GPU-to-GPU communication, NVLINK provides a 50 GBps direct link between GPUs, reducing the dependency on the PCIe bus to exchange data between GPUs, but still depending on PCIe Gen3 to GPU to system memory communications.

The NVLINK 2.0 implementation on POWER9 goes beyond the traditional implementation by implementing 1.5x more memory bandwidth and aggregating NVLINK Bricks to enable 3x faster communication between GPUs and system memory to GPU, reducing potential bottlenecks throughout the system. The goal is to move data from system memory to the GPU 16 GB internal memory as fast as possible so that GPU processing does not have to stop and wait for data to be moved to continue.

After NVLINK Bricks are combined differently depending on the server having four or six GPUs (with two POWER9 processors) to maximize bandwidth, there are two distinct logical diagrams, depending on the numbers of maximum GPUs that are supported per system.

Figure 2-5 shows the logical system diagram for the Power AC922 server (8335-GTG) with four GPUs, where the six NVLINK Bricks are divided into groups of three, which enables 150 GBps buses between GPUs.

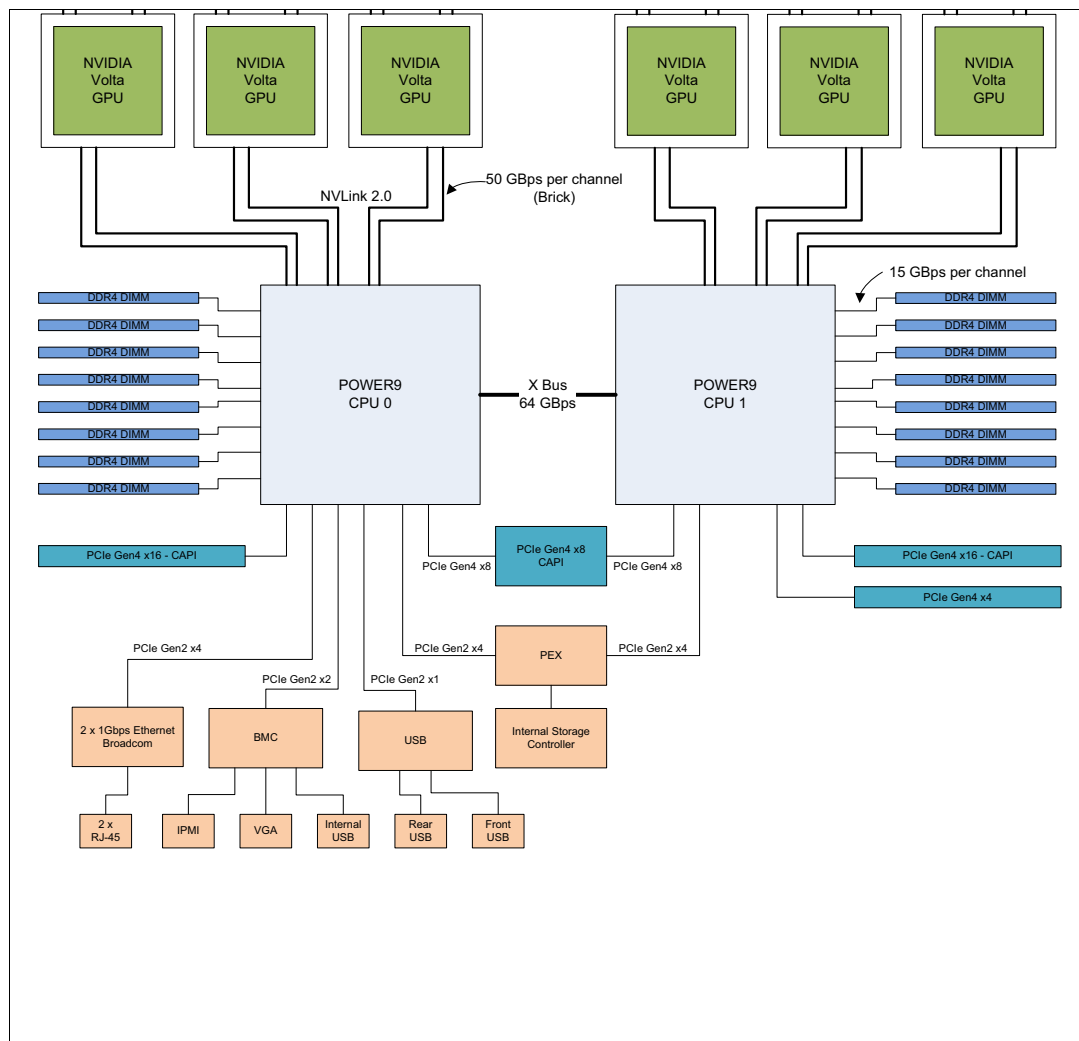


Figure 2-5 The Power AC922 server model 8335-GTG logical system diagram

Figure 2-6 shows the logical system diagram for the Power AC922 server (8335-GTW) with six connected GPUs, where the six NVLINK Bricks are divided into three groups of two Bricks, enabling 100 GBps buses between GPUs, but enabling more connected GPUs.

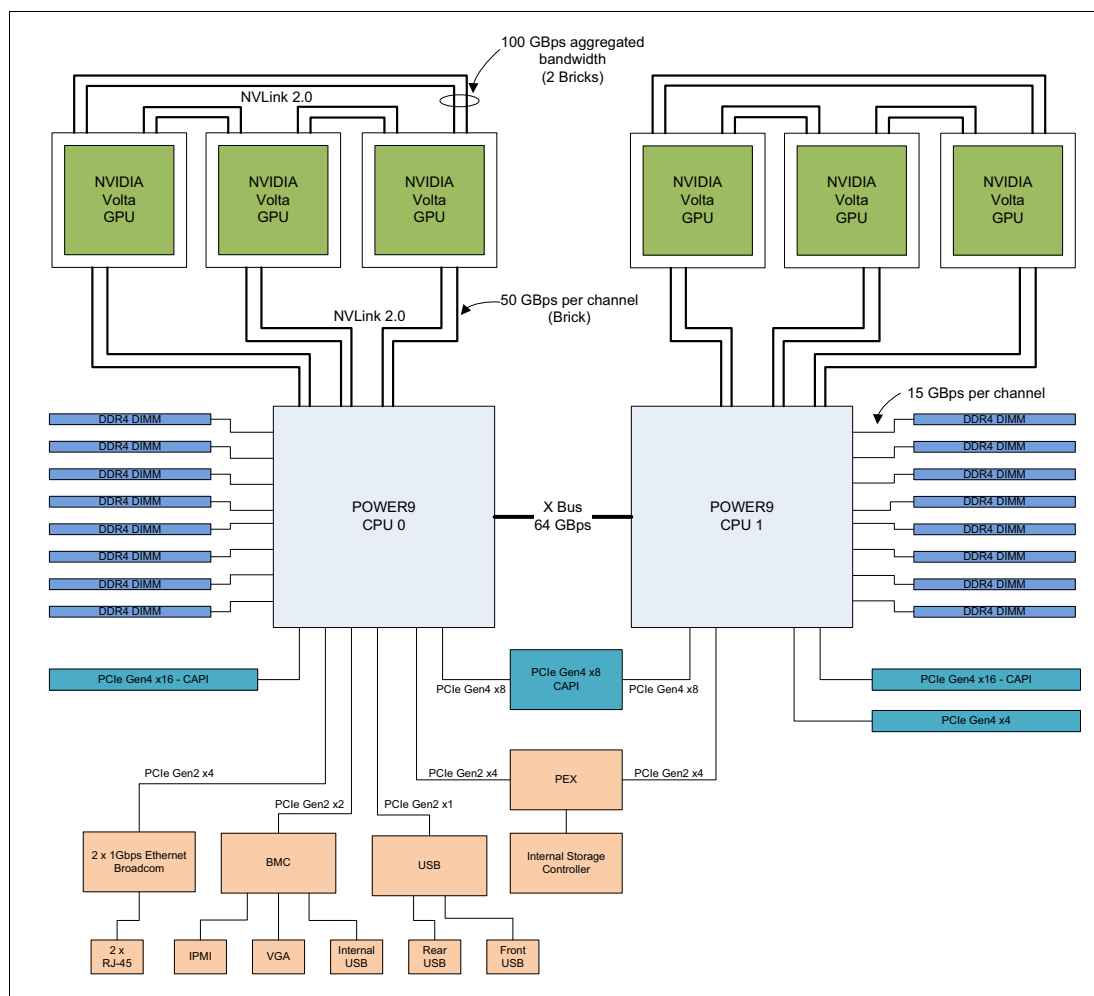


Figure 2-6 The Power AC922 server model 8335-GTW logical system diagram

## 2.2 Processor subsystem

This section introduces the current processor in the Power Systems product family and describes its main characteristics and features in general.

The POWER9 processor in the Power AC922 server is the current generation of the POWER processor family. Based on the 14 nm FinFET Silicon-On-Insulator (SOI) architecture, the chip size is 685 mm x 685 mm and contains eight billion transistors.

## 2.2.1 POWER9 processor overview

POWER9 chips have four variations, depending on the server scalability and whether they are being used on servers on a Linux infrastructure or servers on a PowerVM infrastructure. The main differences are in the scalability, maximum core count, SMT capability, and memory connection.

Table 2-1 compares the chip variations.

Table 2-1 POWER9 chip variations

Variation	Maximum cores	SMP connections	Maximum SMT	Memory connection	Memory bandwidth
Scale-Out Linux	24	2 sockets	SMT4	Direct	120 GBps
Scale-Out PowerVM	24	2 sockets	SMT4	Direct	120 GBps
Scale-Up Linux	12	16 sockets	SMT8	Memory buffer	230 GBps
Scale-Up PowerVM	12	16 sockets	SMT8	Memory buffer	230 GBps

The main reason for this differentiation between Linux and PowerVM infrastructures is that Linux infrastructures need more granularity and core counts per chip, and PowerVM infrastructures need stronger threads and higher per core performance for better efficiency on licensing.

Figure 2-7 shows the main differences.

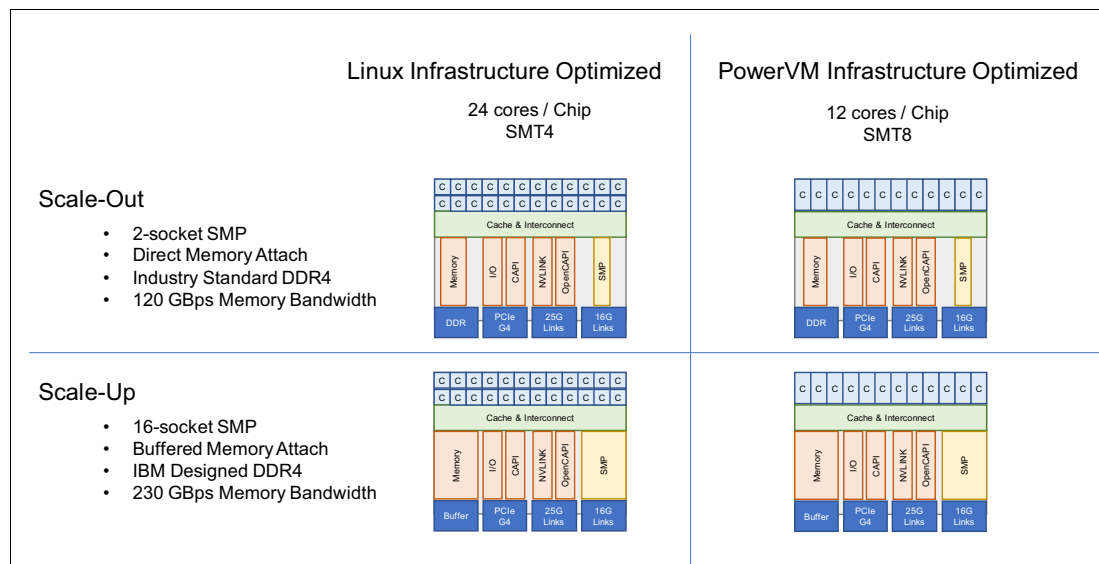


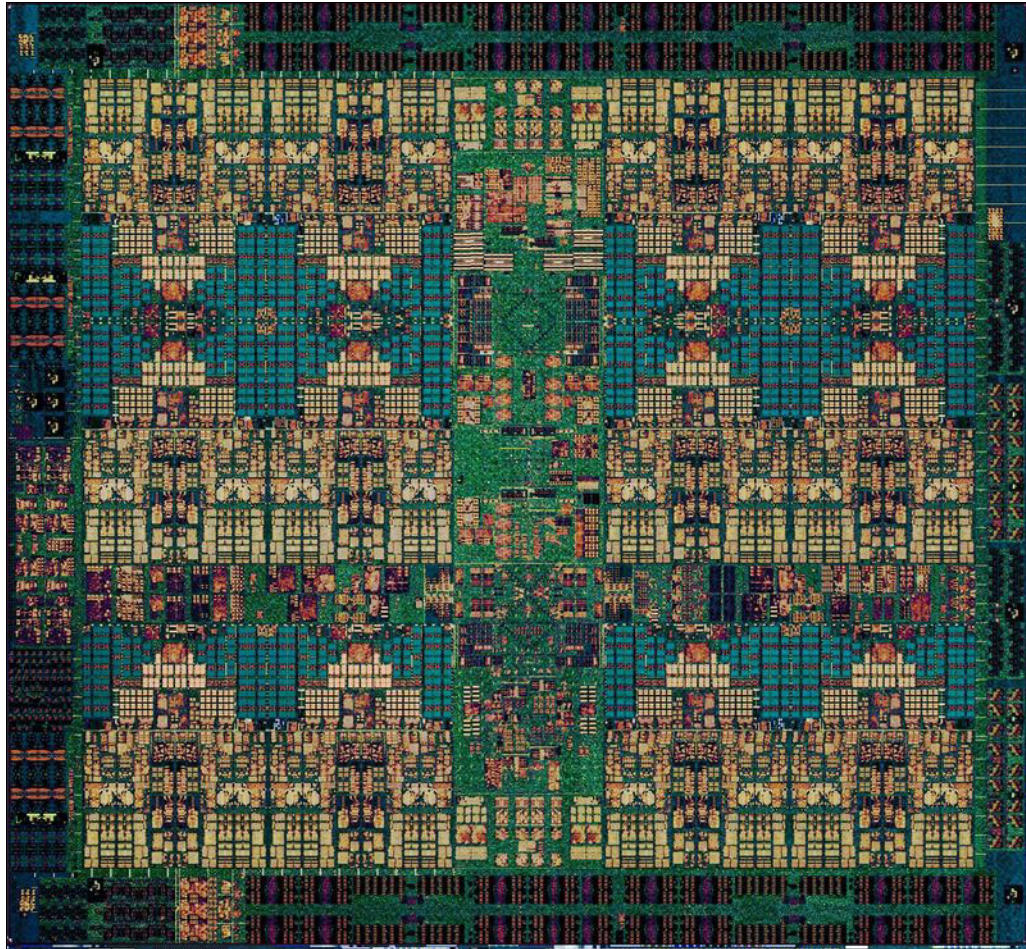
Figure 2-7 POWER9 chip variations

The Power AC922 server uses the scale-out Linux version, with up to 24 cores and SMT4.

The POWER9 chip contains two memory controllers, PCIe Gen4 I/O controllers, and an interconnection system that connects all components within the chip at 7 TBps. Each core has 256 KB of L2 cache, and all cores share 120 MB of L3 embedded DRAM (eDRAM). The interconnect also extends through module and system board technology to other POWER9 processors in addition to DDR4 memory and various I/O devices.

Although scale-out variations have direct memory connections, scale-up POWER9 processor-based systems use memory buffer chips to interface between the POWER9 processor and DDR4 memory. Each buffer chip also includes an L4 cache to reduce the latency of local memory accesses.

Figure 2-8 shows the POWER9 processor with 24 cores.



*Figure 2-8 The 24-core POWER9 processor*

Each POWER9 processor has eight memory channels that address up to 512 GB of memory. Theoretically, a two-socket server can address up to 8 TB of memory, and a 16-socket system can address up to 64 TB of memory. Because of the current state of memory DIMM technology, the largest available DIMM is 64 GB, so the largest amount of memory that is supported on the Power AC922 is 1024 GB (64 GB x 8 channels x 2 processors).

## 2.2.2 Processor feature codes

The Power AC922 (8335-GTG) server supports two processor configurations only, as shown in Table 2-2. Processor features must be in quantities of two and cannot be mixed.

Table 2-2 POWER9 processor features supported

Feature code	Description	Min/Max	Operating system (OS) support
EP0K	POWER9 16-core 2.60 GHz (3.09 GHz Turbo) - 190W	2/2	Linux
EP0M	POWER9 20-core 2.00 GHz (2.87 GHz Turbo) - 190W	2/2	Linux

## 2.3 Memory subsystem

The Power AC922 server is a two-socket system that supports two POWER9 SCM processor modules. The server supports a maximum of 16 DDR4 RDIMMs slots in the main system board directly connected to the POWER9 processor.

Memory features equate to a single memory DIMM. All memory DIMMs must be populated, and mixing of different memory feature codes is not supported. Memory feature codes that are supported are as follows:

- ▶ 16 GB DDR4
- ▶ 32 GB DDR4
- ▶ 64 GB DDR4

Plans for future memory growth needs should be accounted for when you decide which memory feature size to use at the time of initial system order because an upgrade requires a full replacement of the installed DIMMs.

### 2.3.1 Memory features and placement rules

Each feature code equates to a single memory DIMM. Table 2-3 shows the available memory feature codes for ordering.

Table 2-3 Memory features that are supported

Feature code	Description	Min/Max	OS support
EM61	16 GB DDR4 2666 MHz DDR4 RDIMM	16/16	Linux
EM63	32 GB DDR4 2666 MHz DDR4 RDIMM	16/16	Linux
EM64	64 GB DDR4 2666 MHz DDR4 RDIMM	16/16	Linux

The supported maximum memory is 1024 GB by installing 16 #EM64 memory DIMMs. For the Power AC922 server (models 8335-GTG and 8335-GTW), the following requirements apply:

- ▶ All the memory DIMMs must be populated.
- ▶ Memory features cannot be mixed.
- ▶ The base memory is 256 GB with sixteen 16 GB, 2666 MHz DDR4 memory modules (#EM61).

Table 2-4 shows the total memory and how it can be achieved by using certain quantities of each memory feature code.

Table 2-4 Supported memory feature codes for the Power AC922 server

Memory features	Total installed memory		
	256 GB	512 GB	1024 GB
16 GB (#EM61)	16		
32 GB (#EM63)		16	
64 GB (#EM64)			16

## 2.3.2 Memory bandwidth

The POWER9 processor has exceptional cache, memory, and interconnect bandwidths. Table 2-5 shows the maximum bandwidth estimates for a single core on the server.

Table 2-5 The Power AC922 server single-core bandwidth estimates

Single core	8335-GTG and 8335-GTW	
	2.860 GHz	3.259 GHz
L1 (data) cache	137.28 GBps	156.43 GBps
L2 cache	137.28 GBps	156.43 GBps
L3 cache	183.04 GBps	208.57 GBps

The bandwidth figures for the caches are calculated as follows:

- ▶ L1 cache: In one clock cycle, two 16-byte load operations and one 16-byte store operation can be accomplished. The value varies depending on the clock of the core, and the formulas are as follows:
  - 2.860 GHz Core:  $(2 \times 16 \text{ B} + 1 \times 16 \text{ B}) \times 2.860 \text{ GHz} = 137.28 \text{ GBps}$
  - 3.259 GHz Core:  $(2 \times 16 \text{ B} + 1 \times 16 \text{ B}) \times 3.259 \text{ GHz} = 156.43 \text{ GBps}$
- ▶ L2 cache: In one clock cycle, one 32-byte load operation and one 16-byte store operation can be accomplished. The value varies depending on the clock of the core, and the formula is as follows:
  - 2.860 GHz Core:  $(1 \times 32 \text{ B} + 1 \times 16 \text{ B}) \times 2.860 \text{ GHz} = 137.28 \text{ GBps}$
  - 3.259 GHz Core:  $(1 \times 32 \text{ B} + 1 \times 16 \text{ B}) \times 3.259 \text{ GHz} = 156.43 \text{ GBps}$
- ▶ L3 cache: One 32-byte load operation and one 32-byte store operation can be accomplished at half-clock speed, and the formula is as follows:
  - 2.860 GHz Core:  $(1 \times 32 \text{ B} + 1 \times 32 \text{ B}) \times 2.860 \text{ GHz} = 183.04 \text{ GBps}$
  - 3.259 GHz Core:  $(1 \times 32 \text{ B} + 1 \times 32 \text{ B}) \times 3.259 \text{ GHz} = 208.57 \text{ GBps}$



Table 2-6 shows the overall bandwidths for the entire Power AC922 server that is populated with the two processor modules.

Table 2-6 The Power AC922 server total bandwidth estimates

Total bandwidths	8335-GTG	
	20 cores @ 2.860 GHz	16 cores @ 3.259 GHz
L1 (data) cache	2746 GBps	2503 GBps
L2 cache	2746 GBps	2503 GBps
L3 cache	3661 GBps	3337 GBps
Total memory	240 GBps	240 GBps
SMP interconnect	64 GBps	64 GBps
PCIe interconnect	272 GBps	272 GBps

Where:

- ▶ Total memory bandwidth: Each POWER9 processor has eight memory channels running at 15 GBps. The bandwidth formula is calculated as follows:  
 $8 \text{ channels} \times 15 \text{ GBps} = 120 \text{ GBps per processor module}$
- ▶ SMP interconnect: The POWER9 processors are connected by using an X-bus. The bandwidth formula is calculated as follows:  
 $1 \text{ X bus} \times 4 \text{ bytes} \times 16 \text{ GHz} = 64 \text{ GBps}$
- ▶ PCIe interconnect: Each POWER9 processor has 34 PCIe lanes running at 16 Gbps full-duplex. The bandwidth formula is calculated as follows:  
 $34 \text{ lanes} \times 2 \text{ processors} \times 16 \text{ Gbps} \times 2 = 272 \text{ GBps}$

## 2.4 I/O subsystem

The key components of the I/O subsystem are described in this section.

### 2.4.1 PCIe

PCIe uses a serial interface and enables point-to-point interconnections between devices by using a directly wired interface between these connection points. A single PCIe serial link is a dual-simplex connection that uses two pairs of wires, one pair for transmit and one pair for receive, and can transmit only one bit per cycle. These two pairs of wires are called a *lane*. A PCIe link can consist of multiple lanes. In these configurations, the connection is labeled as x1, x2, x8, x12, x16, or x32, where the number is effectively the number of lanes.

The Power AC922 server supports the new PCIe Gen4, which are capable of 32 GBps simplex (64 GBps duplex) on a single x16 interface. PCIe Gen4 slots also support previous generation (Gen3 and Gen2) adapters, which operate at lower speeds according to the following rules:

- ▶ Place x1, x4, x8, and x16 speed adapters in the same size connector slots first before mixing adapter speed with connector slot size.



- Adapters with lower speeds are allowed in larger sized PCIe connectors, but larger speed adapters are not compatible in smaller connector sizes (that is, a x16 adapter cannot go in an x8 PCIe slot connector).

PCIe adapters use a different type of slot than PCI adapters. If you attempt to force an adapter into the wrong type of slot, you might damage the adapter or the slot.

POWER9 processor-based servers support PCIe low profile (LP) cards because of the restricted height of the server.

Before adding or rearranging adapters, use the [System Planning Tool](#) to validate the new adapter configuration.

If you are installing a new feature, ensure that you have the software that is required to support the new feature and determine whether there are existing update prerequisites to install. To obtain this information, see [Power Systems Prerequisites](#).

The following sections describe other I/O technologies that enhance or replace the PCIe interface.

## 2.4.2 IBM CAPI2

IBM CAPI2 is the evolution of CAPI and defines a coherent accelerator interface structure for attaching special processing devices to the POWER9 processor bus. As with the original CAPI, CAPI2 can attach accelerators that have coherent shared memory access with the processors in the server and share full virtual address translation with these processors by using standard PCIe Gen4 buses with twice the bandwidth compared to the previous generation.

Applications can have customized functions in Field Programmable Gate Arrays (FPGAs) and queue work requests directly in shared memory queues to the FPGA. Applications can also have customized functions by using the same effective addresses (pointers) they use for any threads running on a host processor. From a practical perspective, CAPI enables a specialized hardware accelerator to be seen as an extra processor in the system with access to the main system memory and coherent communication with other processors in the system.

Figure 2-9 shows a comparison of the traditional model, where the accelerator must go through the processor to access memory with CAPI.

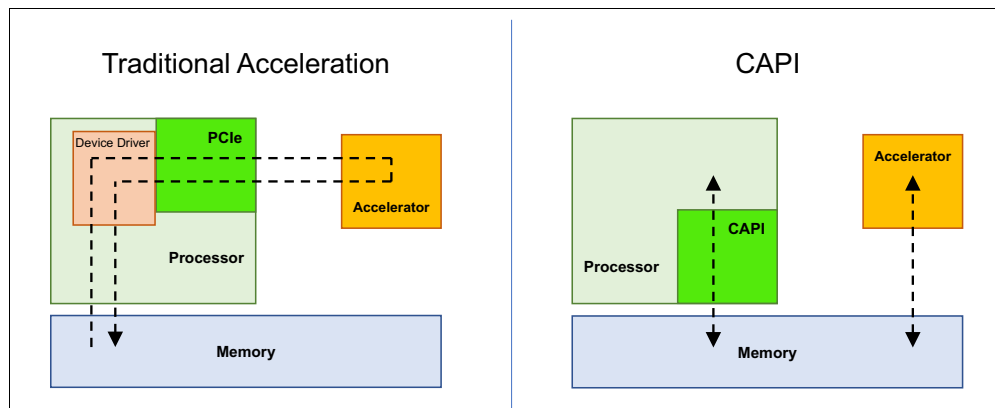


Figure 2-9 CAPI accelerator that is attached to the POWER9 processor

The benefits of using CAPI include the ability to access shared memory blocks directly from the accelerator, perform memory transfers directly between the accelerator and processor cache, and reduce the code path length between the adapter and the processors. This reduction in the code path length might occur because the adapter is not operating as a traditional I/O device, and there is no device driver layer to perform processing. CAPI also presents a simpler programming model.

The accelerator adapter implements the POWER Service Layer (PSL), which provides address translation and system memory cache for the accelerator functions. The custom processors on the system board, consisting of an FPGA or an ASIC, use this layer to access shared memory regions, and cache areas as though they were a processor in the system. This ability enhances the performance of the data access for the device and simplifies the programming effort to use the device. Instead of treating the hardware accelerator as an I/O device, it is treated as a processor, which eliminates the requirement of a device driver to perform communication. It also eliminates the need for direct memory access that requires system calls to the OS kernel. By removing these layers, the data transfer operation requires fewer clock cycles in the processor, improving the I/O performance.

The implementation of CAPI on the POWER9 processor enables hardware companies to develop solutions for specific application demands. Companies use the performance of the POWER9 processor for general applications and the custom acceleration of specific functions by using a hardware accelerator with a simplified programming model and efficient communication with the processor and memory resources.

For a list of supported CAPI adapters, see 2.5.4, “CAPI-enabled InfiniBand adapters” on page 31.

### 2.4.3 OpenCAPI

Although CAPI is a technology that is present in IBM POWER processors and depends on IBM intellectual property (the PSL), several industry solutions might benefit from having a mechanism of connecting different devices to the processor, with low latency, including memory attachment. The PCIe standard is pervasive to every processor technology, but its design characteristics and latency, do not enable the attachment of memory for load/store operations.

Therefore, the OpenCAPI Consortium was created, with the goal of defining a device attachment interface to open the CAPI interface to other hardware developers and extending its capabilities. OpenCAPI aims to enable memory, accelerators, network, storage, and other devices to connect to the processor through a high bandwidth, low latency interface, becoming the interface of choice for connecting high-performance devices.

By providing a high-bandwidth, low latency connection to devices, OpenCAPI enables several applications to improve networking, use FPGA accelerators, use expanded memory beyond server internal capacity, and reduce latency to storage devices. Some of these use cases and examples are shown in Figure 2-10.

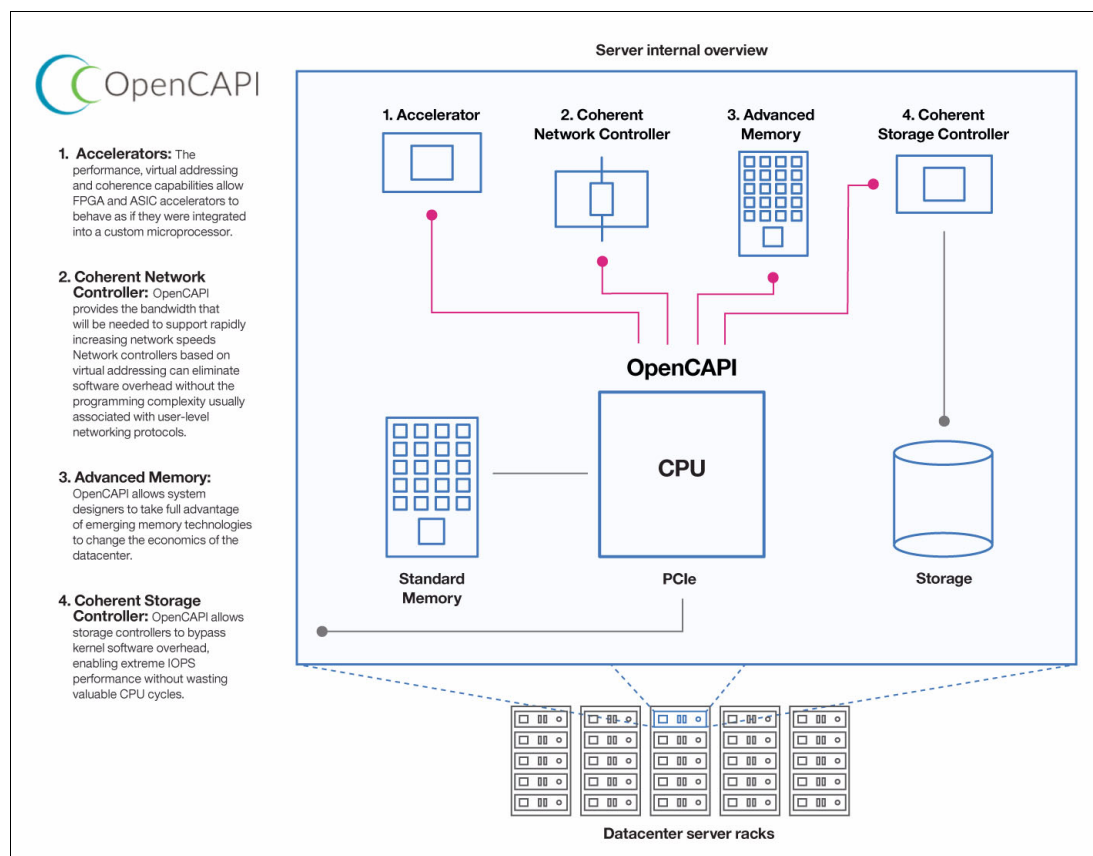


Figure 2-10 OpenCAPI use cases

The design of OpenCAPI enables low latency in accessing attached devices (in the same range of DDR memory access, that is, 10 ns), which enables memory to be connected through OpenCAPI and serve as main memory for load/store operations. In contrast, PCIe latency is 10 times larger (around 100 ns). Therefore, OpenCAPI has a significant enhancement compared to traditional PCIe interconnects.

OpenCAPI is neutral regarding processor architecture, so the electrical interface is not defined by the OpenCAPI consortium or any of its workgroups. On the POWER9 processor, the electrical interface is based on the design from the 25G workgroup within the OpenPower Foundation, which encompasses a 25 Gbps signaling and protocol that is built to enable a low latency interface on CPU and attached devices. Future capabilities include increased speeds of up to 32 Gbps and 56 Gbps signaling.

The current design for the adapter is based on a PCIe card that draws power from the PCIe slot while connecting to the OpenCAPI port on the system board through a 25 GBps cable, as shown in Figure 2-11.

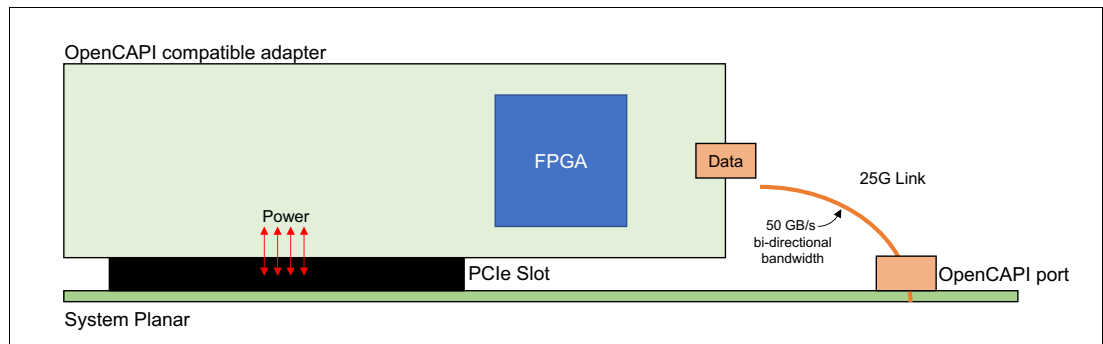


Figure 2-11 OpenCAPI compatible adapter and 25G link

The OpenCAPI interface uses the same electrical interconnect as NVLink 2.0. Systems can be designed to have an NVLink-attached GPU, an OpenCAPI-attached device, or both. The use of OpenCAPI adapters limits the number of NVLINK ports that are available for GPU communication. Each POWER9 chip has six NVLINK ports, four of which can be used for OpenCAPI adapters, as shown in Figure 2-12.

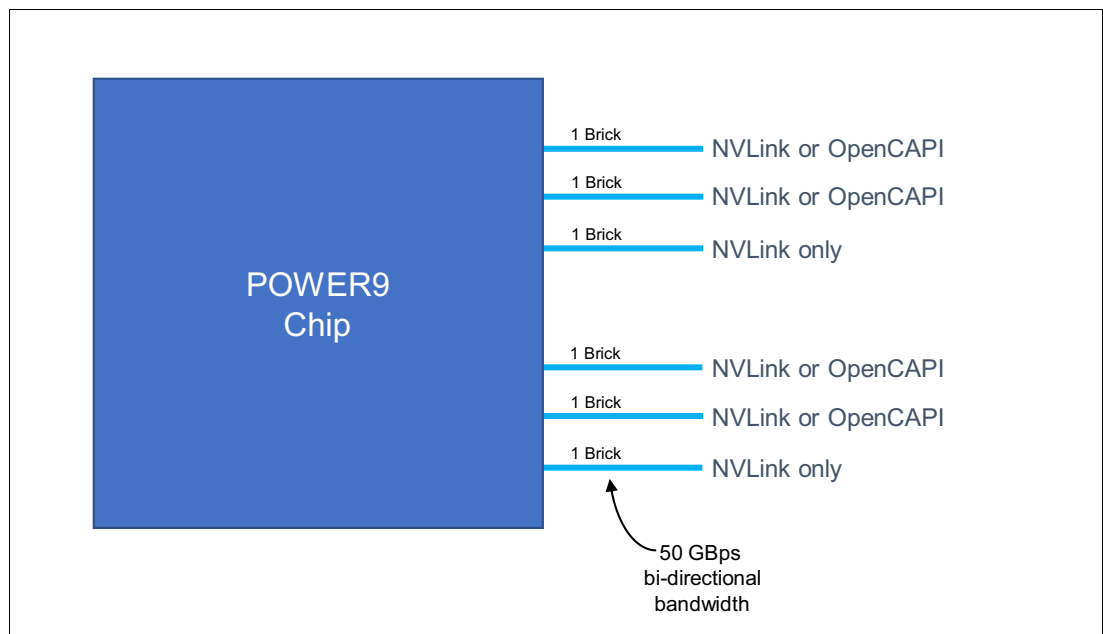


Figure 2-12 OpenCAPI and NVLINK shared ports on the POWER9 chip

## 2.4.4 The NVIDIA Tesla V100

The new NVIDIA Tesla V100 accelerator, codenamed Volta, takes GPU computing to the next level. This section describes the Tesla V100 accelerator, which is shown in Figure 2-13.

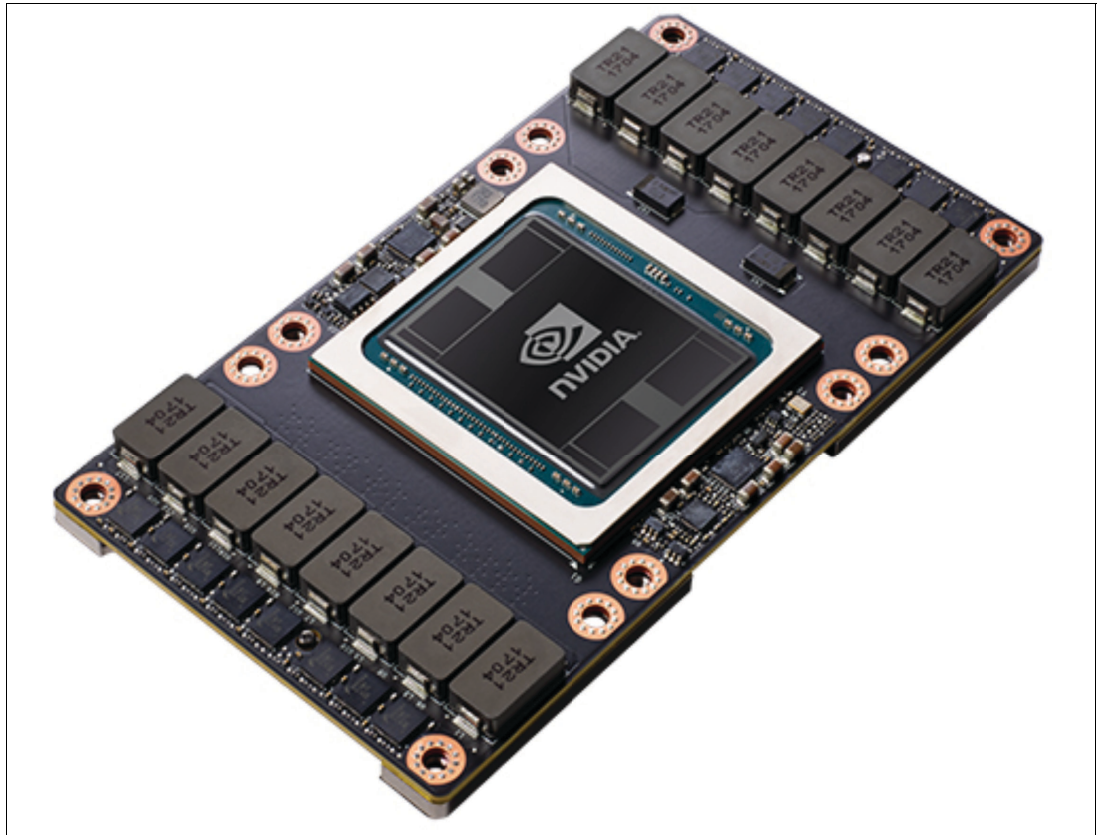


Figure 2-13 NVIDIA Tesla V100 for NVLINK accelerator

NVIDIA Tesla V100 is the most advanced data center GPU built to accelerate artificial intelligence (AI), high-performance computing (HPC), and graphics. Powered by NVIDIA Volta, the current GPU architecture, Tesla V100 offers the performance of 100 CPUs in a single GPU, which enables data scientists, researchers, and engineers to tackle challenges that were once impossible.

The Tesla V100 includes the following key features:

- Volta architecture

By pairing CUDA cores and Tensor cores within a unified architecture, a single server with Tesla V100 GPUs can replace hundreds of commodity CPU servers for traditional HPC and deep learning.

- Tensor core

Equipped with 640 Tensor cores, Tesla V100 delivers 125 teraflops (TFLOPS) of deep learning performance, that is, 12x Tensor TFLOPS for DL training, and 6x Tensor TFLOPS for DL inference compared to NVIDIA Pascal GPUs.

- Next generation NVLINK

NVIDIA NVLink in Tesla V100 delivers 2x higher throughput compared to the previous generation. Up to eight Tesla V100 accelerators can be interconnected at up to 300 GBps to unleash the highest application performance possible on a single server.

- ▶ Maximum efficiency mode

The new maximum efficiency mode enables data centers to achieve up to 40% higher compute capacity per rack within the existing power budget. In this mode, Tesla V100 runs at peak processing efficiency, providing up to 80% of the performance at half the power consumption.

- ▶ HBM2

With a combination of improved raw bandwidth of 900 GBps and higher DRAM usage efficiency at 95%, Tesla V100 delivers 1.5x higher memory bandwidth over Pascal GPUs, as measured on STREAM.

- ▶ Programmability

The Tesla V100 is designed to simplify programmability. Its new independent thread scheduling enables finer-grain synchronization and improves GPU usage by sharing resources among small jobs.

The Tesla V100 delivers exceptional performance for the most demanding compute applications. It delivers the following performance benefits:

- ▶ 7.8 TFLOPS of double-precision floating point (FP64) performance
- ▶ 15.7 TFLOPS of single-precision (FP32) performance
- ▶ 125 Tensor TFLOPs of mixed-precision

With 640 Tensor cores, Tesla V100 is the first GPU to break the 100 TFLOPS barrier of deep learning performance. The next generation of NVIDIA NVLink connects multiple V100 GPUs at up to 300 GBps to create the most powerful computing servers. AI models that would use weeks of computing resources on previous systems can now be trained in a few days. With this dramatic reduction in training time, many problems are now solvable with AI.

Multiple GPUs are common in workstations, as are the nodes of HPC clusters and deep-learning training systems. A powerful interconnect is valuable in multiprocessing systems. The NVIDIA Tesla V100 does not rely on traditional PCIe for data transfers, but instead uses the new NVLINK 2.0 bus that creates an interconnect for GPUs that offer higher bandwidth than PCI Express Gen3 (PCIe). The GPUs are compatible with the GPU ISA to support shared memory multiprocessing workloads.

Once PCIe buses are not used for data transfer, the GPU cards do not need to comply with the traditional PCIe card format. To improve density in the Power AC922 server, the GPUs have a different form factor called *SXM2*. This form factor enables the GPU to be connected directly on the system board. Figure 2-14 on page 27 shows the SXM2 GPU module top and bottom views and the connectors that are used for the GPU modules.

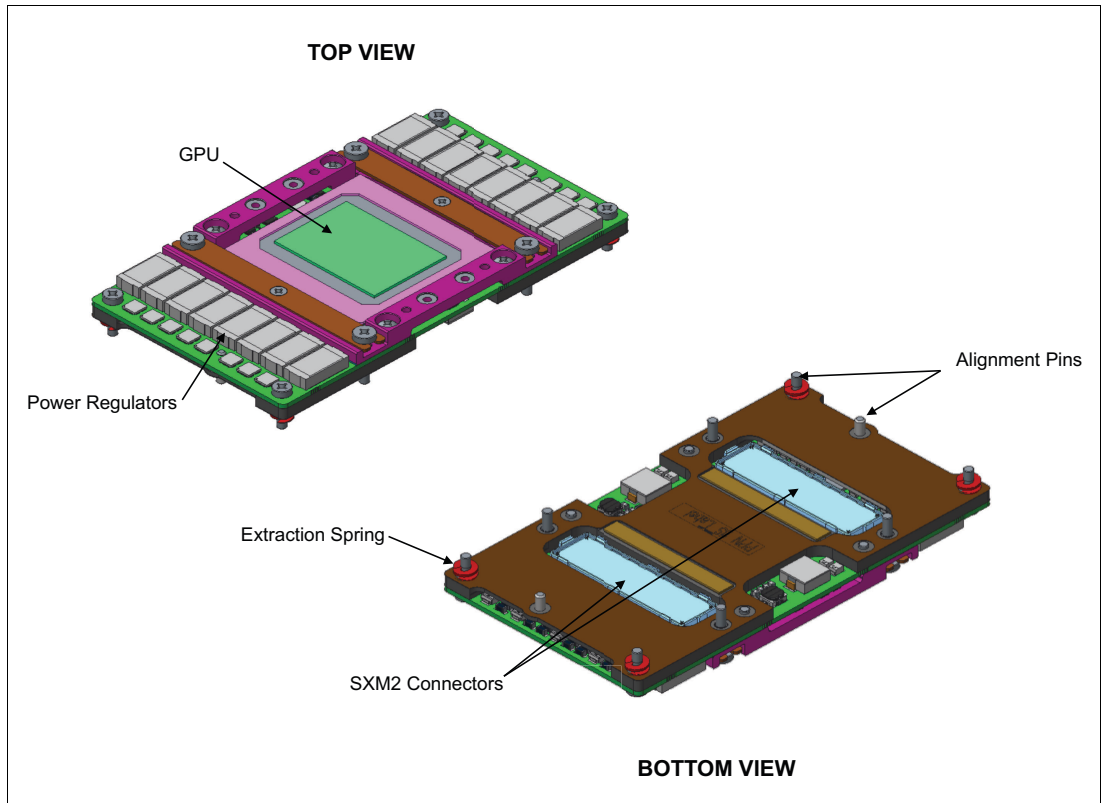


Figure 2-14 SXM2 GPU module views

Figure 2-15 shows the location of the GPUs on the Power AC922 system board.

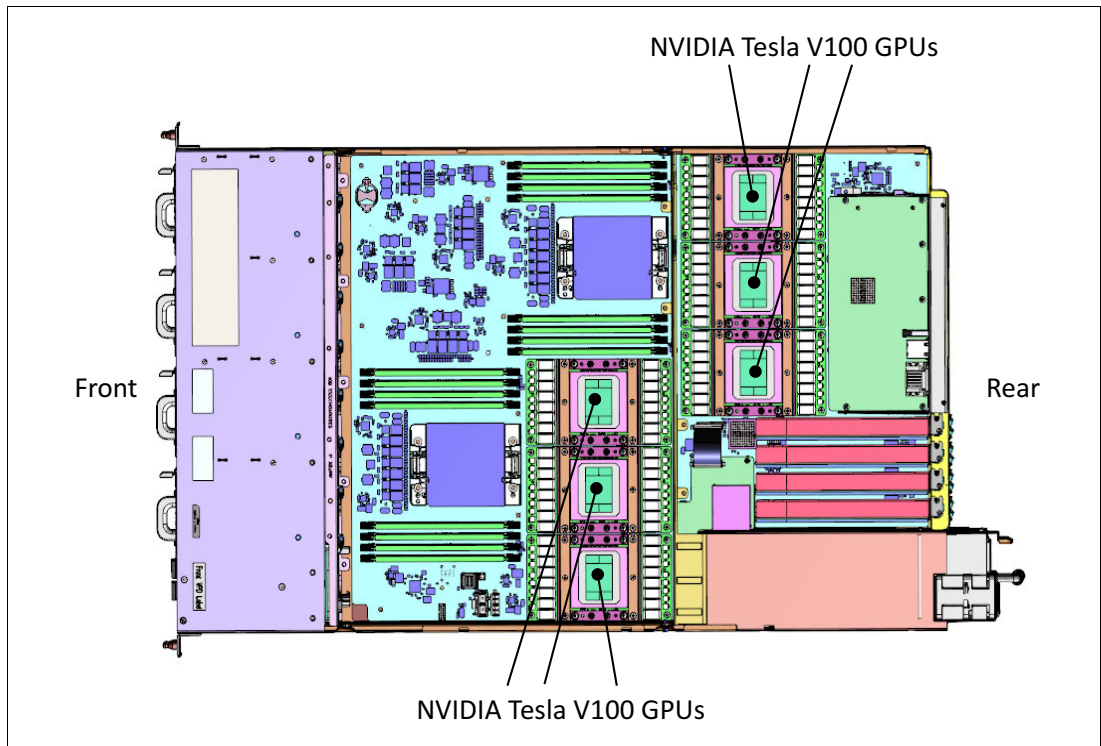


Figure 2-15 GPU location in a six-GPU configuration

Cooling for four-GPU configurations (Power AC922 model 8335-GTG) is done by air cooling, and cooling for six-GPU configurations (model 8335-GTW) is done by water cooling. For more information about server water cooling, see Chapter 3, “Physical infrastructure” on page 37.

For more information about the Tesla V100, see [Inside Volta Parallel for All](#).

## 2.4.5 NVLINK 2.0

NVLINK 2.0 is the NVIDIA new generation high-speed interconnect technology for GPU-accelerated computing. Supported on SXM2-based Tesla V100 accelerator system boards, NVLink increases performance for both GPU-to-GPU communications and for GPU access to system memory.

Support for the GPU ISA enables programs running on NVLINK-connected GPUs to run directly on data in the memory of another GPU and on local memory. GPUs can also perform atomic memory operations on remote GPU memory addresses, enabling much tighter data sharing and improved application scaling.

NVLINK 2.0 uses the NVIDIA High-Speed Signaling interconnect (NVHS). NVHS transmits data over a link that is called Brick that connects two processors (GPU-to-GPU or GPU-to-CPU). A single Brick supports up to 50 GBps of bidirectional bandwidth between the endpoints. Multiple links can be combined to form *Gangs* for even higher-bandwidth connectivity between processors. The NVLINK implementation in Tesla V100 supports up to six links, enabling a gang with an aggregate maximum theoretical bandwidth of 300 GBps bidirectional bandwidth.

Although traditional NVLINK implementation primarily focuses on interconnecting multiple NVIDIA Tesla V100 GPUs together, under POWER9 it also connects Tesla V100 GPUs with IBM POWER9 CPUs, enabling direct system memory access and providing GPUs with an extended memory orders of magnitude larger than the internal 16 GB memory.

On a Power Systems implementation, Bricks are always combined to provide the highest bandwidth possible. Figure 2-16 compares the bandwidth of the POWER9 processor that is connected with two GPUs and three GPUs.

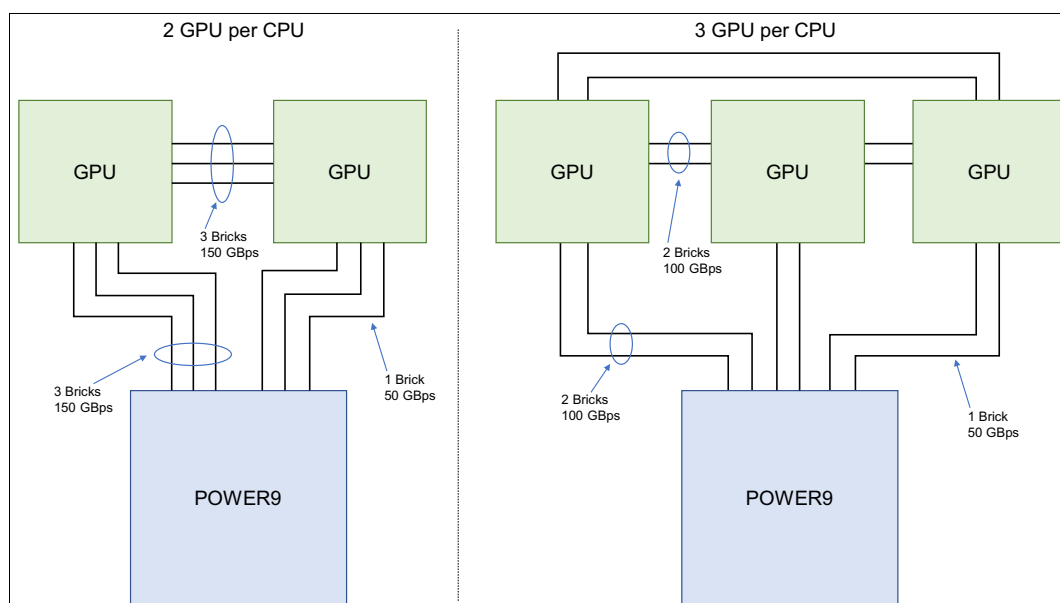


Figure 2-16 CPU to GPU and GPU to GPU interconnect that uses NVLink 2.0



All the initialization of the GPU is through the PCIe interface. The PCIe interface also contains the side-band communication for status, power management, and so on. After the GPU is running, all data communication uses the NVLink.

## 2.5 PCI adapters

This section describes the types and functions of the PCI adapters that are supported by the Power AC922 server.

The Power AC922 server uses the current PCIe Gen4 technology, enabling 32 GBps unidirectional and 64 GBps bidirectional bandwidth.

**Note:** PCIe adapters on the Power AC922 server are not hot-pluggable.

### 2.5.1 Slot configuration

The Power AC922 server has four PCIe Gen4 slots. Figure 2-17 shows a rear-view diagram of the Power AC922 server with its PCIe slots.

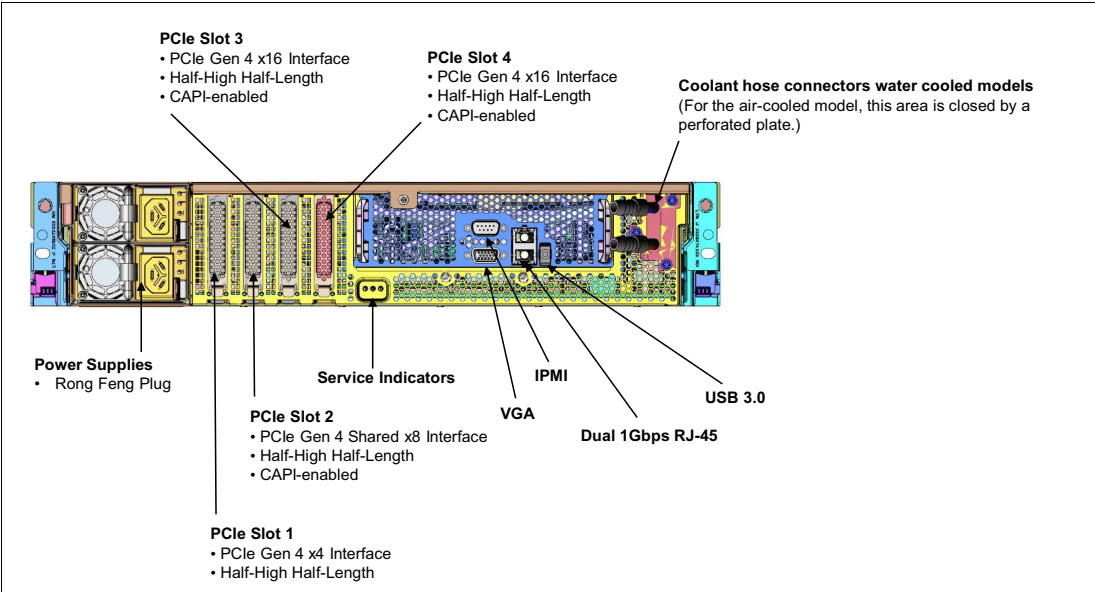


Figure 2-17 Rear-view PCIe slots and main components

Table 2-7 provides the PCIe Gen4 slot properties.

Table 2-7 The Power AC922 server PCIe Gen4 slot properties

Slot	Description	Card size	CAPI-capable
Slot 1	PCIe Gen4 x4	Half-height, half-length	No
Slot 2	PCIe Gen4 x8 Shared	Half-height, half-length	Yes

Slot	Description	Card size	CAPI-capable
Slot 3	PCIe Gen4 x16	Half-height, half-length	Yes
Slot 4	PCIe Gen4 x16	Half-height, half-length	Yes

Slot 2 has a shared connection between the two POWER9 CPUs. When you use a dual-channel Mellanox InfiniBand ConnectX5 (IB-EDR) Network Interface Card (#EC64), it enables each CPU to have direct access to the InfiniBand card. If the #EC64 card is not installed, the shared slot operates as a single x8 PCIe Gen4 slot that is attached to processor 0.

Figure 2-18 shows the logical diagram of the slot 2 that is connected to the two POWER9 processors.

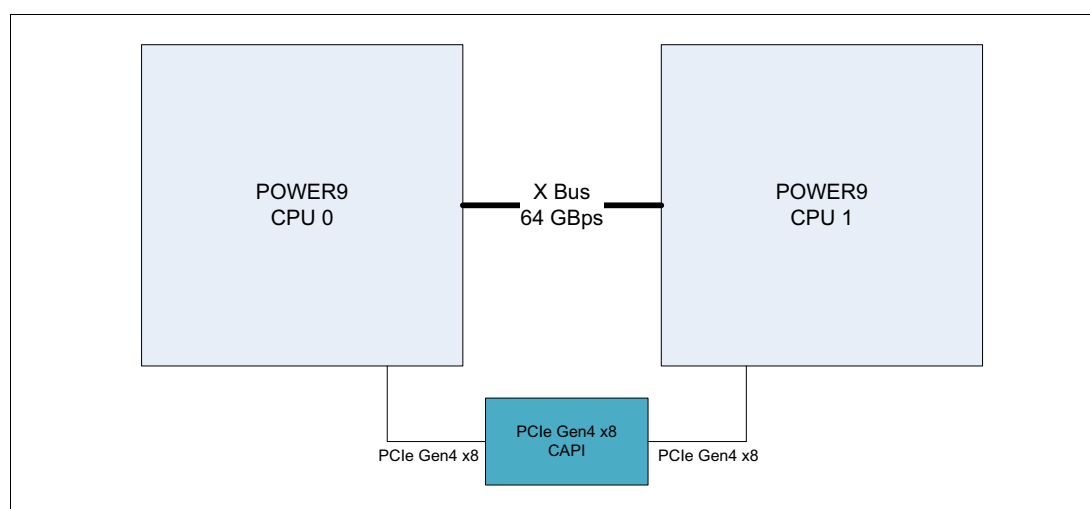


Figure 2-18 Shared PCIe slot 2 logical diagram

Only LP adapters can be placed in LP slots. A x8 adapter can be placed in a x16 slot, but a x16 adapter cannot be placed in a x8 slot.

## 2.5.2 Local area network adapters

To connect the Power AC922 server to a local area network (LAN), you can use the LAN adapters that are supported in the PCIe slots of the system unit. Table 2-8 lists the supported LAN adapters for the server.

Table 2-8 Supported LAN adapters

Feature code	Description	Max	OS support
EC2R	PCIe3 LP 2-port 10GbE (NIC& RoCE) SFP28 Adapter x8	3	Linux
EC2T	PCIe3 LP 2-port 25/10GbE (NIC& RoCE) SFP28 Adapter x8	2	Linux
EC3L	PCIe3 LP 2-port 100GbE (NIC& RoCE) QSFP28 Adapter x16	2	Linux
EL3Z	PCIe2 LP 2-port 10/1GbE BaseT RJ45 Adapter x8	3	Linux

Feature code	Description	Max	OS support
EL4M	PCIe2 LP 4-port 1GbE Adapter	4	Linux
EN0T	PCIe2 LP 4-Port (10Gb+1GbE) SR+RJ45 Adapter x8	3	Linux
EN0V	PCIe2 LP 4-port (10Gb+1GbE) Copper SFP+RJ45 Adapter x8	2	Linux

### 2.5.3 Fibre Channel adapters

The Power AC922 server supports direct or SAN connection to devices that use Fibre Channel adapters. Table 2-9 summarizes the available Fibre Channel adapters, which all have LC connectors.

If you are attaching a device or switch with an SC-type fiber connector, an LC-SC 50 micron fiber converter cable (#2456) or an LC-SC 62.5 micron fiber converter cable (#2459) is required.

Table 2-9 Fibre Channel adapters that are supported

Feature code	Description	Max	OS support
EL43	PCIe3 LP 16 Gb 2-port Fibre Channel Adapter x8	3	Linux
EL5V	PCIe3 LP 32 Gb 2-port Fibre Channel Adapter x8	3	Linux

### 2.5.4 CAPI-enabled InfiniBand adapters

Table 2-10 shows the available CAPI adapters.

Table 2-10 Available CAPI adapters

Feature code	Description	Max	OS support
EC64	PCIe4 LP 2-port 100 Gb EDR InfiniBand Adapter x16	3	Linux
EC62	PCIe4 LP 1-port 100 Gb EDR InfiniBand Adapter x16	3	Linux

### 2.5.5 Compute-intensive accelerators

Compute intensive accelerators are GPUs that are developed by NVIDIA. With NVIDIA GPUs, the server can offload processor-intensive operations to a GPU accelerator and boost performance. The Power AC922 server aims to deliver a new class of technology that maximizes performance and efficiency for all types of scientific, machine learning, deep learning, AI, engineering, Java, big data analytics, and other technical computing workloads.

Table 2-11 lists the available compute-intensive accelerators.

Table 2-11 Graphics processing units adapters that are supported

Feature code	Description	Max	OS support
EC4J	One air-cooled NVIDIA Tesla V100 GPU 16 GB (model 8335-GTG only)	4	Linux
EC4H	One water-cooled NVIDIA Tesla V100 GPU 16 GB (model 8335-GTW only)	6	Linux

## 2.5.6 Flash storage adapters

The available flash storage adapter is shown in Table 2-12.

Table 2-12 Available flash storage adapter

Feature code	Description	Max	OS support
EC5A	PCIe3 1.6 TB NVMe Flash Adapter	3	Linux

## 2.6 System ports

The system board has two 1 Gbps Ethernet ports, one Intelligent Platform Management Interface (IPMI) port, one rear USB 3.0 port (the model 8335-GTG also has one front USB 3.0 port), and a VGA port, as shown in Figure 2-17 on page 29.

The integrated system ports are supported for modem and asynchronous terminal connections with Linux. Any other application that uses serial ports requires a serial port adapter to be installed in a PCI slot. The integrated system ports do not support IBM PowerHA® configurations. The VGA port does not support cable lengths that exceed 3 meters.

## 2.7 Internal storage

The internal storage on the Power AC922 server contains the following features:

- ▶ A storage backplane for two 2.5-inch small form factor (SFF) Gen4 SATA HDDs or SSDs.

**Limitation:** The disks use an SFF-4 carrier. Disks that are used in other Power Systems servers usually have an SFF-3 or SFF-2 carrier and are not compatible with this system.

- ▶ One integrated SATA disk controller (non-RAID).
- ▶ The storage split backplane feature is not supported.

Table 2-13 presents a summarized view of these features.

*Table 2-13 Summary of features for the integrated SATA disk controller*

Option	Integrated SATA disk controller
Supported RAID types	None - JBOD
Disk bays	Two SFF Gen4 (HDDs/SDDs)
SATA controllers	Single
IBM Easy Tier® capable controllers	No
External SAS ports	No
Split backplane	No

The 2.5 inch or SFF SAS bays can contain SATA drives (HDD or SSD) that are mounted on a Gen4 tray or carrier (also known as SFF-4). SFF-2 or SFF-3 drives do not fit in an SFF-4 bay. All SFF-4 bays support concurrent maintenance or hot-plug capability.

## 2.7.1 Disk and media features

The server supports the attachment of up to two SATA storage devices. Table 2-14 lists the supported devices that can be installed. Disk features cannot be mixed.

*Table 2-14 Supported storage devices*

Feature code	Description	Max	OS support
ELD0	1 TB 7.2k RPM SATA SFF-4 disk drive	2	Linux
ELU4	960 GB SSD SFF-4 disk drive	2	Linux
ELU5	1.92 TB SSD SFF-4 disk drive	2	Linux
ELU6	3.84 TB SSD SFF-4 disk drive	2	Linux
ES6A	2 TB 7.2k RPM 5xx SATA SFF-4 disk drive	2	Linux

The Power AC922 server is designed for network installation or USB media installation. It does not support an internal DVD drive.

## 2.8 External I/O subsystems

The Power AC922 server does not support external PCIe Gen3 I/O expansion drawers or EXP24S, EXP12X, and EXP24SX storage drawers.

## 2.9 Location codes

Figure 2-19 shows the location codes for the server main components.

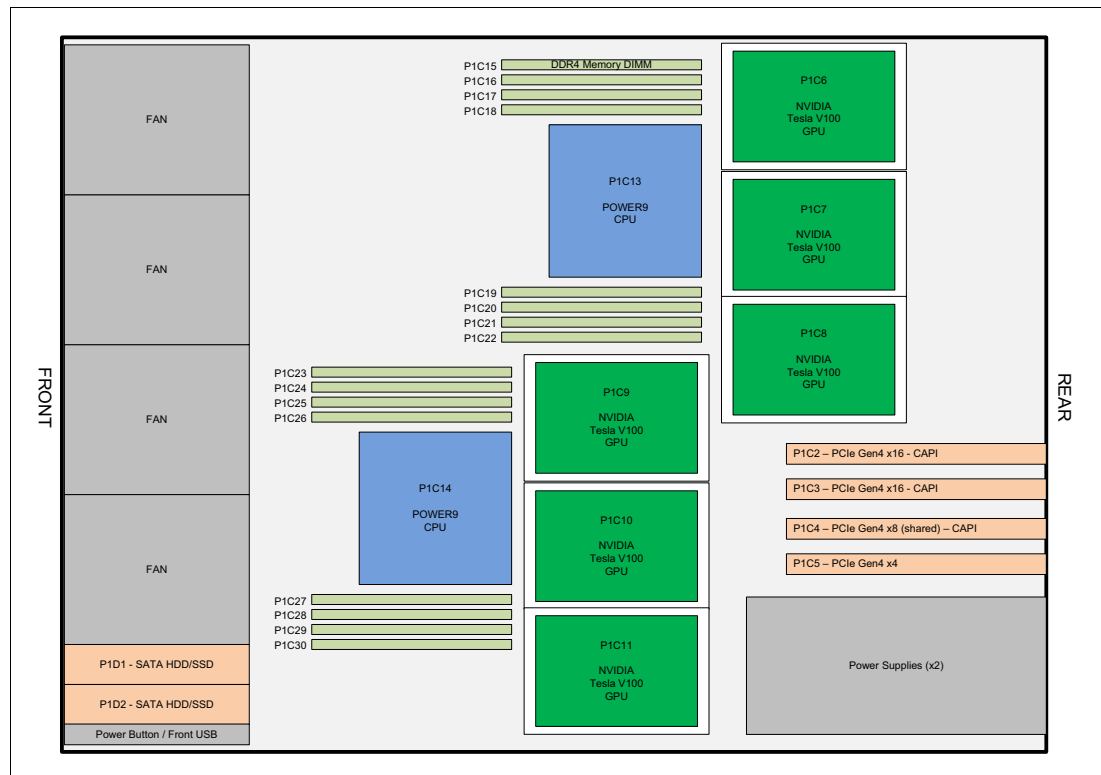


Figure 2-19 Power AC922 main components location codes

## 2.10 IBM System Storage

The IBM System Storage® disk systems products and offerings provide compelling storage solutions with superior value for all levels of business, from entry-level to high-end storage systems. For more information about the various offerings, see [IBM Disk Storage](#).

The following sections describe sample System Storage offerings.

## IBM Network Attached Storage

IBM Network Attached Storage (NAS) products provide a wide-range of network attachment capabilities to a broad range of host and client systems, such as IBM Scale Out Network Attached Storage and the IBM System Storage N series. For more information about the hardware and software, see [IBM Spectrum Scale](#).

## IBM Storwize family

The IBM Storwize® family is the ideal solution to optimize the data architecture for business flexibility and data storage efficiency. Different models, such as the IBM Storwize V3700, IBM Storwize V5000, and IBM Storwize V7000, offer storage virtualization, IBM Real-time Compression™, Easy Tier, and many more functions. For more information, see [IBM Storwize](#).

## **IBM FlashSystem family**

The IBM FlashSystem® family delivers performance to derive measurable economic value across the data architecture (servers, software, applications, and storage). IBM offers a comprehensive flash portfolio with the IBM FlashSystem family. For more information, see [IBM FlashSystem](#).

## **IBM XIV Storage System**

The IBM XIV® Storage System is a high-end disk storage system that helps thousands of enterprises meet the challenge of data growth with hotspot-free performance and ease of use. Simple scaling, high service levels for dynamic, heterogeneous workloads, and tight integration with hypervisors and the OpenStack platform enable optimal storage agility for cloud environments.

XIV Storage Systems extend ease of use with integrated management for large and multi-site XIV deployments, reducing operational complexity and enhancing capacity planning. For more information, see [IBM XIV Storage System](#).

## **IBM System Storage DS8000**

The IBM System Storage DS8800 storage system is a high-performance, high-capacity, and secure storage system that delivers the highest levels of performance, flexibility, scalability, resiliency, and total overall value for the most demanding, heterogeneous storage environments. The storage system can manage a broad scope of storage workloads that exist in today's complex data center, doing it effectively and efficiently.

Additionally, the IBM System Storage DS8000® storage system includes a range of features that automate performance optimization and application quality of service, and also provide the highest levels of reliability and system uptime. For more information, see [IBM System Storage DS8800](#).

# **2.11 Operating system support**

The Power AC922 server supports Linux, which provides a UNIX-like implementation across many computer architectures.

For more information about the software that is available on Power Systems servers, see [Linux on Power Systems](#).

The Linux OS is an open source, cross-platform OS. It is supported on every Power Systems server IBM sells. Linux on Power Systems is the only Linux infrastructure that offers both scale-out and scale-up choices.

## **2.11.1 Ubuntu**

Ubuntu Server 16.04.03 LTS and any subsequent updates are supported. For more information, see [Ubuntu for POWER9](#).

## **2.11.2 Red Hat Enterprise Linux**

Red Hat Enterprise Linux 7.4 for Power LE (POWER9) and any subsequent updates are supported.

Starting with Red Hat Enterprise Linux 7.1, Red Hat provides separate builds and licenses for big endian and little endian versions for Power Systems servers. For more information about RHEL for POWER9, see [Red Hat](#).

### 2.11.3 More information

For more information about the IBM PowerLinux™ Community, see the [Linux on Power Community](#).

For more information about the features and external devices that are supported by Linux, see [Linux on IBM Power Systems](#).

## 2.12 Java

When running Java applications on the POWER9 processor, the prepackaged Java that is part of a Linux distribution is designed to meet the most common requirements. If you require a different level of Java, there are several resources that are available.

For more information about IBM Java and tested Linux distributions, see [IBM developerWorks](#).

For more information about the OpenJDK port for Linux on PPC64 LE and pregenerated builds, see the [Java OpenJDK website](#).

Launchpad.net has resources for Ubuntu builds. For more information, see [OpenJDK 9](#), [OpenJDK 8](#), and [OpenJDK 7](#).





# Physical infrastructure

The objective of this section is to summarize all the physical infrastructure requirements regarding the IBM Power System AC922 servers.

For more information, see [IBM Knowledge Center for the Power AC922 server](#).

## 3.1 Operating environment

Table 3-1 and Table 3-2 on page 39 provide the operating environment specifications for the Power AC922 server.

Table 3-1 Operating environment for the 4-GPU 8335-GTG Power AC922 server

Server operating environment			
Description	Recommended operating	Allowable operating	Non-operating
Temperature <sup>ab</sup>	18 - 27 °C (64 - 80.6°F)	5 - 40°C (41 - 104°F)	1 - 60°C (34 - 140°F)
Humidity range	5.5°C (42°F) dew point (DP) to 60% relative humidity (RH) and 15°C (59°F) dew point	-12°C *19.4°F) DP and 8% - 80% RH	8 - 80% RH
Maximum dew point		24°C (75° F)	27°C (80°F)
Maximum operating altitude		3050 m (10,000 ft.)	
Operating voltage		200 - 240 V AC	N/A
Operating frequency		50 - 60 Hz +/- 3 Hz	N/A
Power consumption		2300 watts maximum	N/A
Power source loading		2.6 kVA maximum	N/A
Thermal output		8872 BTU/hr maximum	N/A
Noise level and sound power		7.6/6.7 bels operating/idling	N/A

- a. Derate maximum allowable dry-bulb temperature 1°C (1.8°F) per 175 m above 950 m. IBM recommends a temperature range of 18°C - 27°C (64°F - 80.6°F).
- b. For model 8335-GTG, heavy workloads might see performance degradation above 30°C (86°F), 900 m (2953 ft.), or both if internal temperatures result in a central processing unit (CPU) or graphics processing unit (GPU) clock reduction.

Table 3-2 Operating environment for the 4-GPU 8335-GTW Power AC922 server

Server operating environment			
Description	Recommended operating	Allowable operating	Non-operating
Temperature <sup>a</sup>	18 - 27 °C (64 - 80.6°F)	5 - 40°C (41 - 104°F)	1 - 60°C (34 - 140°F)
Humidity range	5.5°C (42°F) dew point (DP) to 60% relative humidity (RH) and 15°C (59°F) dew point	-12°C *19.4°F) DP and 8% - 80% RH	8 - 80% RH
Maximum dew point		24°C (75° F)	27°C (80°F)
Maximum operating altitude		3050 m (10,000 ft.)	
Operating voltage		200 - 240 V AC	N/A
Operating frequency		50 - 60 Hz +/- 3 Hz	N/A
Power consumption		2300 watts maximum	N/A
Power source loading		2.6 kVA maximum	N/A
Thermal output		8872 BTU/hr maximum	N/A
Noise level and sound power		7.6/6.7 bels operating/idling	N/A

a. Derate maximum allowable dry-bulb temperature 1°C (1.8°F) per 175 m above 950 m. IBM recommends a temperature range of 18°C - 27°C (64°F - 80.6°F).

Figure 3-1 shows the flow rate of water that is required based on the inlet temperature of the water to the rack for a single system.

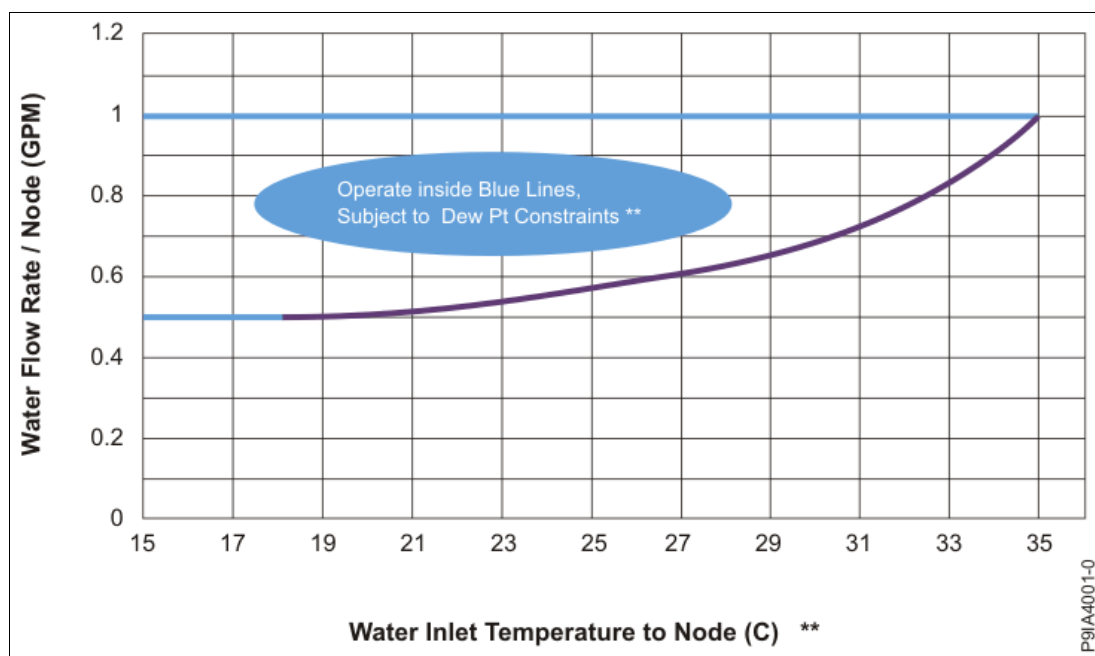


Figure 3-1 Water flow rate versus temperature

**Note:** The customer must constantly monitor the dew point and adjust the water temperature if necessary. The water temperature must *always* be above the dew point so that there is never a chance of condensation.

Figure 3-2 provides data about the water flow versus pressure drop as a function of the number of systems in a rack. The facility rack-level pressure drop includes the following pressure drops:

- ▶ Supply-side Eaton ball valve quick-connect pair
- ▶ Supply-side 1-in. ID, 6-ft. long hose going to the supply manifold
- ▶ Supply-side manifold
- ▶ 8335-GTW nodes
- ▶ Return-side manifold
- ▶ Return-side 1-in. ID, 6-ft. long hose leaving the return manifold
- ▶ Return-side Eaton ball valve quick-connect pair

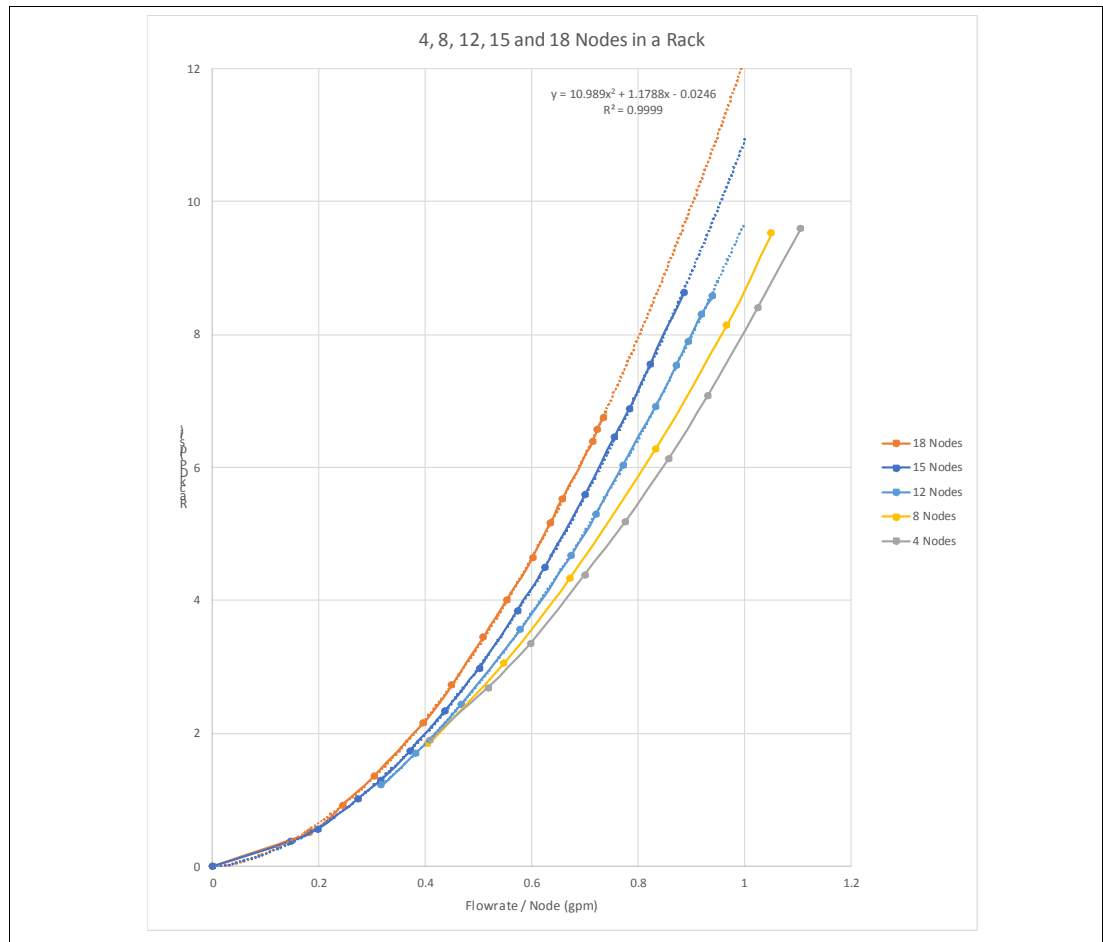


Figure 3-2 Water flow rate versus pressure drop

### 3.1.1 Leak detection

No leak detection is included in the system. You should have facility-level leak sensors or detectors as a preventive measure.

### 3.1.2 Water pressure

The rack manifold cannot exceed 40 PSI water pressure upon entrance to the rack during normal operating conditions. In a worst case, single-fault condition, the maximum pressure cannot exceed 55 PSI.

## 3.2 Physical package

Table 3-3 shows the physical dimensions of the chassis. The server is available only in a rack-mounted form factor and requires 2U (2 EIA units) of rack space.

*Table 3-3 Physical dimensions for the Power AC922 server*

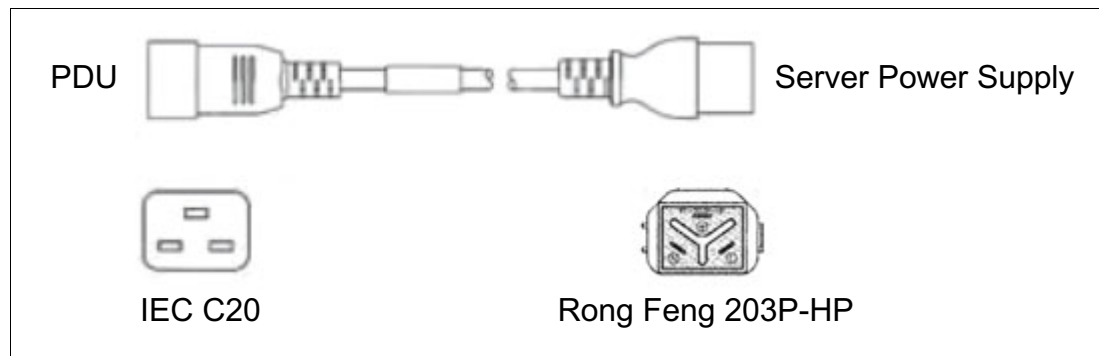
Dimension	Power AC922 server models 8335-GTG and 8335-GTW
Width	441.5 mm (17.4 in.)
Depth	845.0 mm (33.3 in.)
Height	86.0 mm (3.4 in.)
Weight (maximum configuration)	30 kg (65 lbs.)

## 3.3 System power

The Power AC922 server is powered by two 2200-W power supplies at the rear of the unit.

The power supplies provide redundancy if a power supply failure occurs. If GPUs are the largest power-consuming devices in the server, depending on the configuration and utilization, throttling might occur if a power supply failure occurs when six GPUs are installed. In this case, the system remains operational, but might experience reduced performance until the power supply is replaced.

The power supplies on the server use a Rong Feng 203P-HP connector. A new power cable to connect the power supplies to the power distribution units (PDUs) in the rack is required, rendering the reuse of existing power cables not viable. The PDU connector type (IEC C20 or IEC C19) depends on the selected rack PDU. Figure 3-3 shows an example of the power cable with its connectors.



*Figure 3-3 Power AC922 power cables with the Rong Feng connector*

Both 1-phase and 3-phase PDUs are supported. For more information, see 3.5.2, “AC power distribution units” on page 51.

When opting for 3-phase 60A PDUs, a total of 4 PDUs are required to support a full rack with 18 Power AC922 servers that are configured with four GPUs. If 1-phase PDUs are selected, a minimum of five PDUs are required to support a full rack of 18 Power AC922 servers with a four-GPU configuration. If the 1-phase PDUs are limited to 48A, no more than four Power AC922 servers can be connected to a single PDU.

## 3.4 System cooling

Air or water cooling depends on the model of the server and the GPUs feature codes that are selected. For a list of available GPUs, see 2.5.5, “Compute-intensive accelerators” on page 31.

**Rack requirement:** The IBM 7965-S42 rack with feature #ECR3 or #ECR4 installed supports the water-cooling option for the Power AC922 server (see “Optional water cooling” on page 47).

When using water-cooled systems, the customer is responsible for providing the system that supplies the chilled conditioned water to the rack. Water condensation can occur in certain combinations of temperature and relative humidity, which define the dew point. The system that supplies the cooling water must be able to measure the room dew point and automatically adjust the water temperature several degrees above dew point. Otherwise, the water temperature must be above the maximum dew point for that data center installation. Typical primary chilled water is too cold for use in this application because building-chilled water can be as cold as 4°C - 6°C (39°F - 43°F).

In air-cooled systems (8335-GTG), all components are air-cooled, including processors and GPUs that use heat sinks. Figure 3-4 shows the internal view of the server and the two processors and four GPUs heat sinks that are installed.

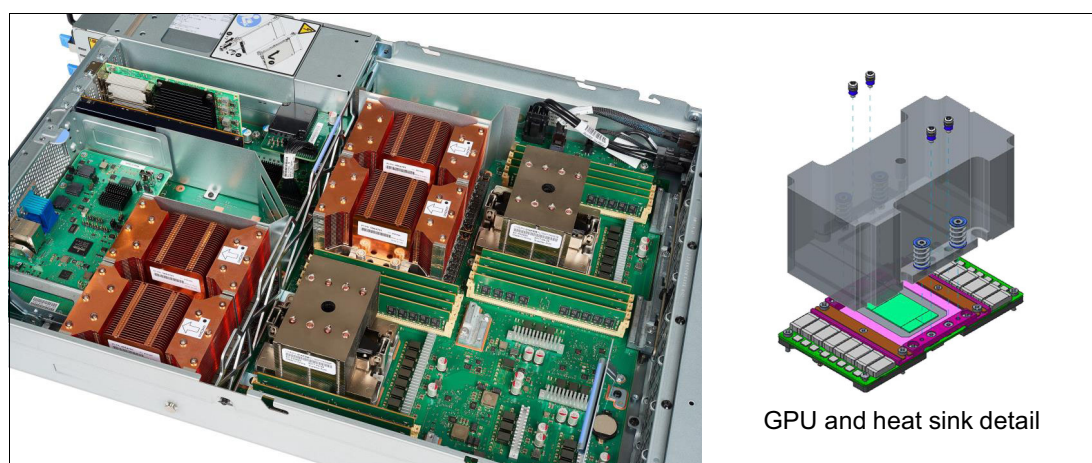


Figure 3-4 Power AC922 air-cooled model internal view

In water-cooled systems (8335-GTW), the processors and GPUs are cooled by using water, and other components, such as memories, Peripheral Component Interconnect Express (PCIe) adapters, and power supplies are cooled by using traditional air-cooling systems. Coldplates to cool two processor modules and up to six GPUs are included. Water lines carrying cool water in and warm water out are also included. This feature is installed in the system unit when the server is manufactured and is not installed in the field.

When ordering the Power AC922 model 8335-GTW, a cooling kit is required. It contains the pipes, coldplates, and splitters that are required to cool the system. Feature code #EJ31 provides the internal cooling system of the server for a 4-GPU configuration, #EJ34 is for a 6-GPU configuration as shown in Figure 3-5.

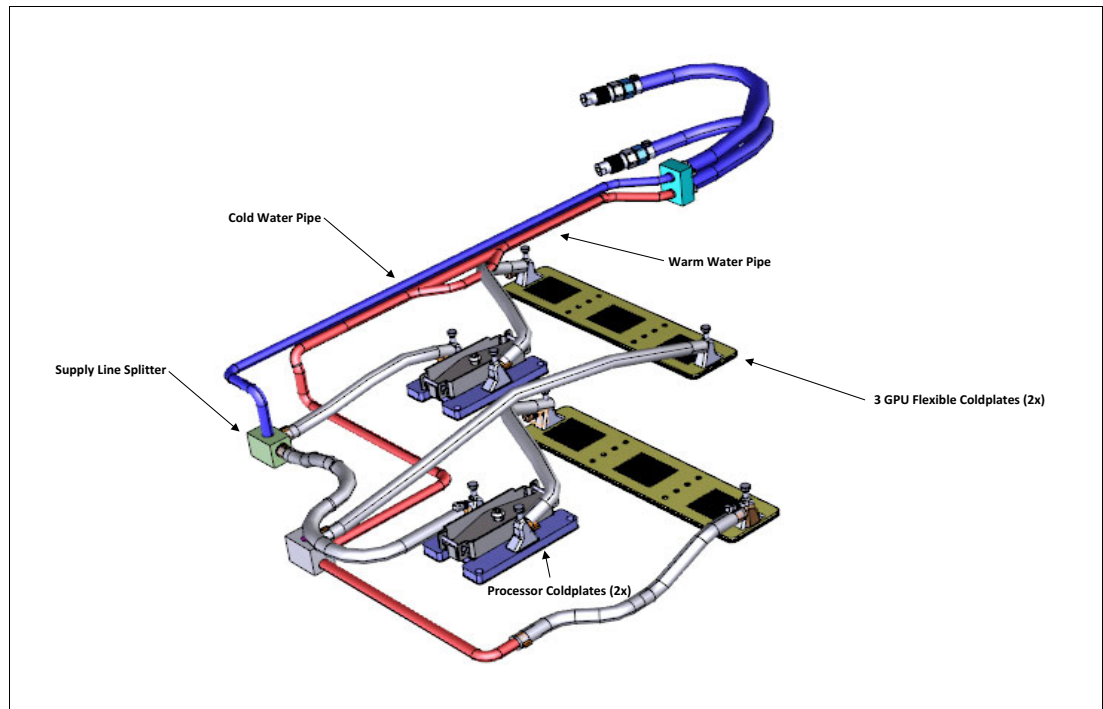


Figure 3-5 Internal cooling components for the 8335-GTW server



Figure 3-6 shows a view of the cooling system that is installed in the server.

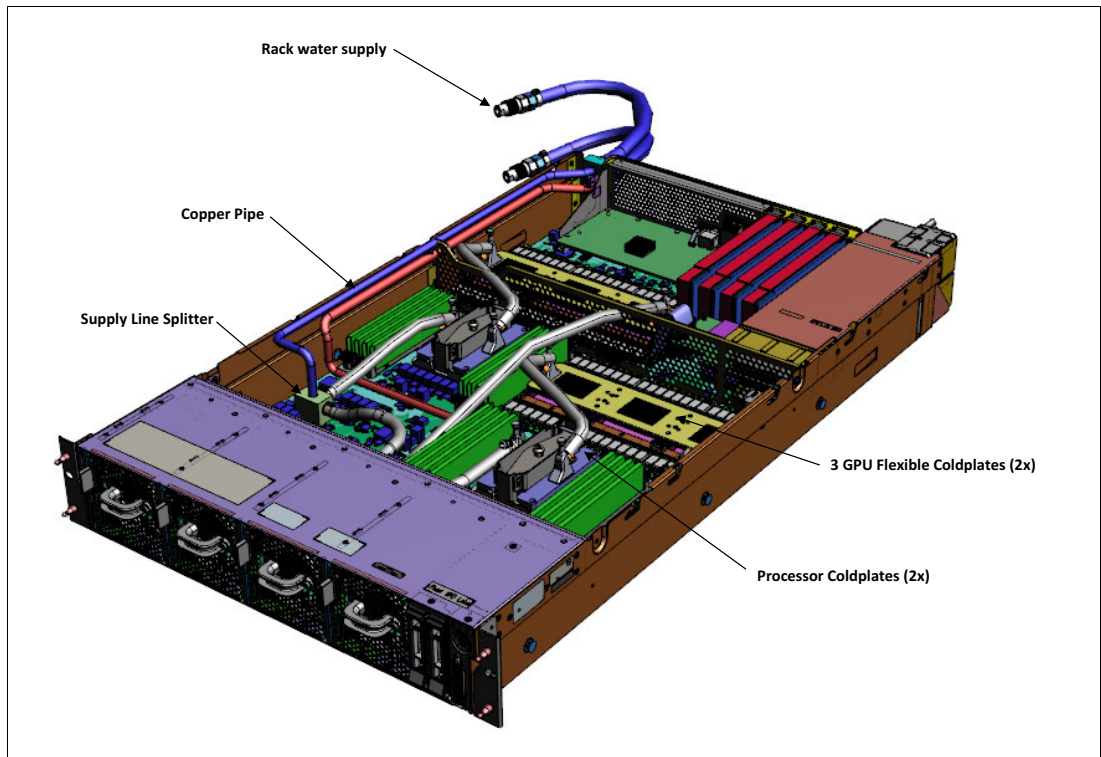


Figure 3-6 Internal cooling that is installed in an 8335-GTW model server

Figure 3-7 shows a detailed view of a processor and three GPUs cooling system.

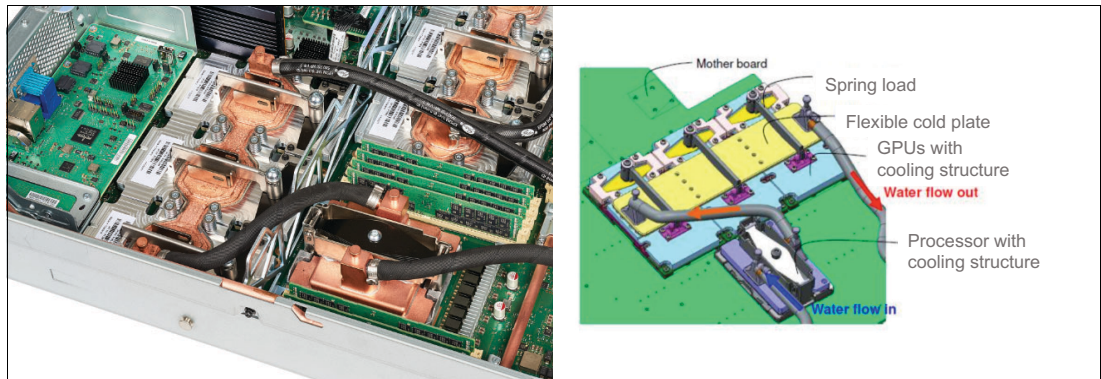


Figure 3-7 Processor and GPU water cooling details

Water enters the system and passes through a splitter block, where the water goes to two different flowpaths. In each flowpath, the water flows first through the CPU coldplate and then through the GPU coldplate. Then, the warm water enters a return line splitter block and goes out of the server.



Figure 3-8 shows the cold water in blue and warm water in red.

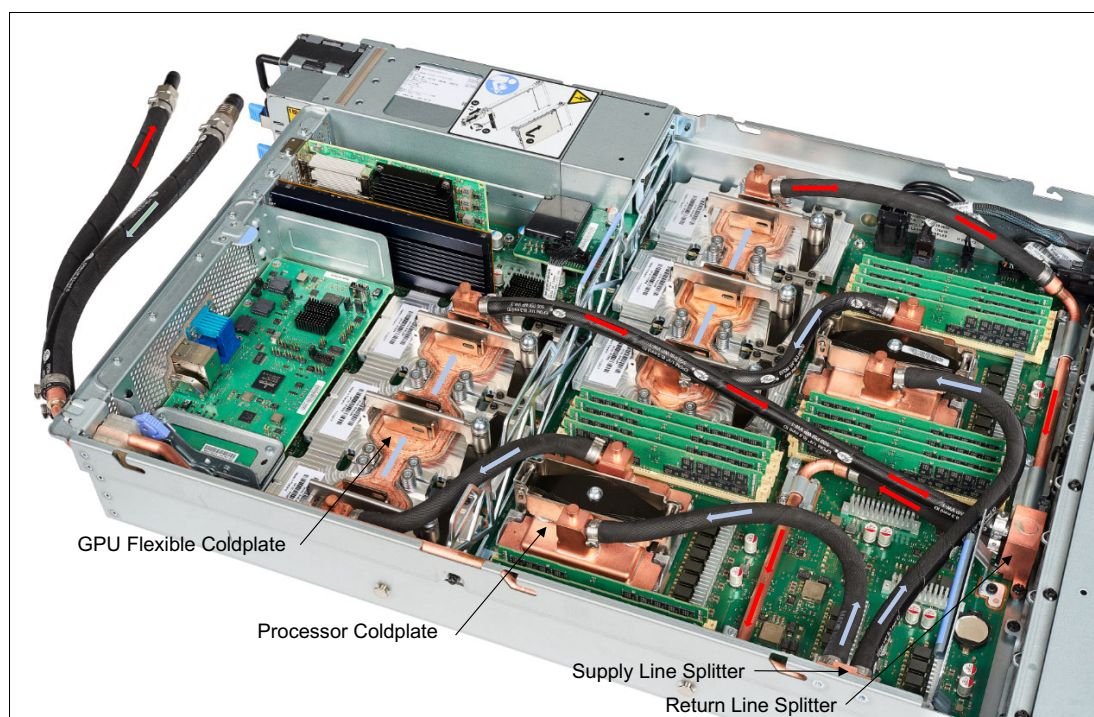


Figure 3-8 Cold and warm water flow through the Power AC922 system

When shipped from IBM, an air-cooled server cannot be changed into a water-cooled server, and a water-cooled server cannot be changed into an air-cooled server.

Customer setup is not supported for water-cooled systems.

**Important:** You must provide a 1-inch ID barb fitting to attach your facility to the hose kit for each hose. Only clean, filtered, and chemically treated water must be used, *not generic building water*.

The GPU air-cooled and water-cooled servers have the following ordering differences:

- ▶ With an air-cooled server (8335-GTG), an initial order can be ordered with two GPUs or four GPUs by using feature code #EC4J.
- ▶ With a water-cooled server (8335-GTW), a quantity of four or six feature code #EC4H GPUs must be ordered.

**Note:** The Power AC922 model 8335-GTW server offers only the fixed rail kit option. Ordering this model with slide rails is not supported. Maintenance of components other than power supplies and fans must be done on a bench with the server unplugged from the cooling system.

For more information about the water-cooling option, see [IBM Knowledge Center](#).

## 3.5 Rack specifications

Depending on the model that is chosen for the Power AC922 server, there are different racks that are supported. Although the air-cooled model (8335-GTG) supports many racks, water-cooled options have just one rack that is supported, as shown in Table 3-4.

Table 3-4 Supported racks by model

Type/Model	Description	Supported by 8335-GTG	Supported by 8335-GTW
7014-T00	IBM 7014 Rack Model T00	Yes	No
7014-T42	IBM 7014 Rack Model T42	Yes	No
7014-S25	IBM Entry Rack Cabinet Model S25	Yes	No
7965-94Y	IBM 42U Slim Rack	Yes	No
7965-S42	IBM Enterprise Slim Rack	Yes	Yes
N/A	OEM 19" Rack	See 3.5.4, "OEM racks" on page 53.	No

**Note:** Because of the water-cooling system, the model 8335-GTW server mounts only in the 42U IBM Enterprise Slim Rack (7965-S42).

These racks are built to the 19-inch EIA 310D standard.

**Order information:** The Power AC922 server cannot be integrated into these racks during the manufacturing process and are not ordered together with servers. If the server and any of the supported IBM racks are ordered together, they are shipped at the same time in the same shipment, but in separate packing material. IBM does not offer integration of the server into the rack before shipping.

If a system is installed in a rack or cabinet that is not an IBM rack, ensure that the rack meets the requirements that are described in 3.5.4, "OEM racks" on page 53.

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

### 3.5.1 IBM Enterprise Slim Rack 7965-S42

The new 2.0-meter (79-inch) Model 7965-S42 is compatible with past and present Power Systems servers and provides an excellent 19-inch rack enclosure for your data center.

This is a 19-inch rack cabinet that provides 42U of rack space for use with rack-mounted, non-blade servers, and I/O drawers. Its 600 mm (23.6-in.) width combined with its 1070 mm (42.1-in.) depth plus its 42 EIA enclosure capacity provides great footprint efficiency for your systems, and enables it to be easily placed on standard 24-inch floor tiles, enabling better thermal and cable management capabilities.

Another difference between the 7965-S42 model rack and the 7014-T42 model rack is that the “top hat” is on the 40U and 41U boundary instead of the 36U and 37U boundary in the 7014-T42 model.

The IBM PDUs are mounted vertically in four side bays, two on each side. After the side bays are filled, PDUs can be mounted horizontally at the rear of the rack. For more information about IBM PDUs, see 3.5.2, “AC power distribution units” on page 51.

To enable maximum airflow through the data center and the rack cabinets, filler panels are mounted at the front of the rack in empty EIA locations, and the rack offers perforated front and rear door designs.

Figure 3-9 shows the front view of the 7965-S42 rack.



*Figure 3-9 IBM 7965-S42 racks front view*

Ballasts for more stability are available, so the 7965-S42 racks should not require the depopulate rules above the 32 EIA location, as required with 7014-T42 rack models.

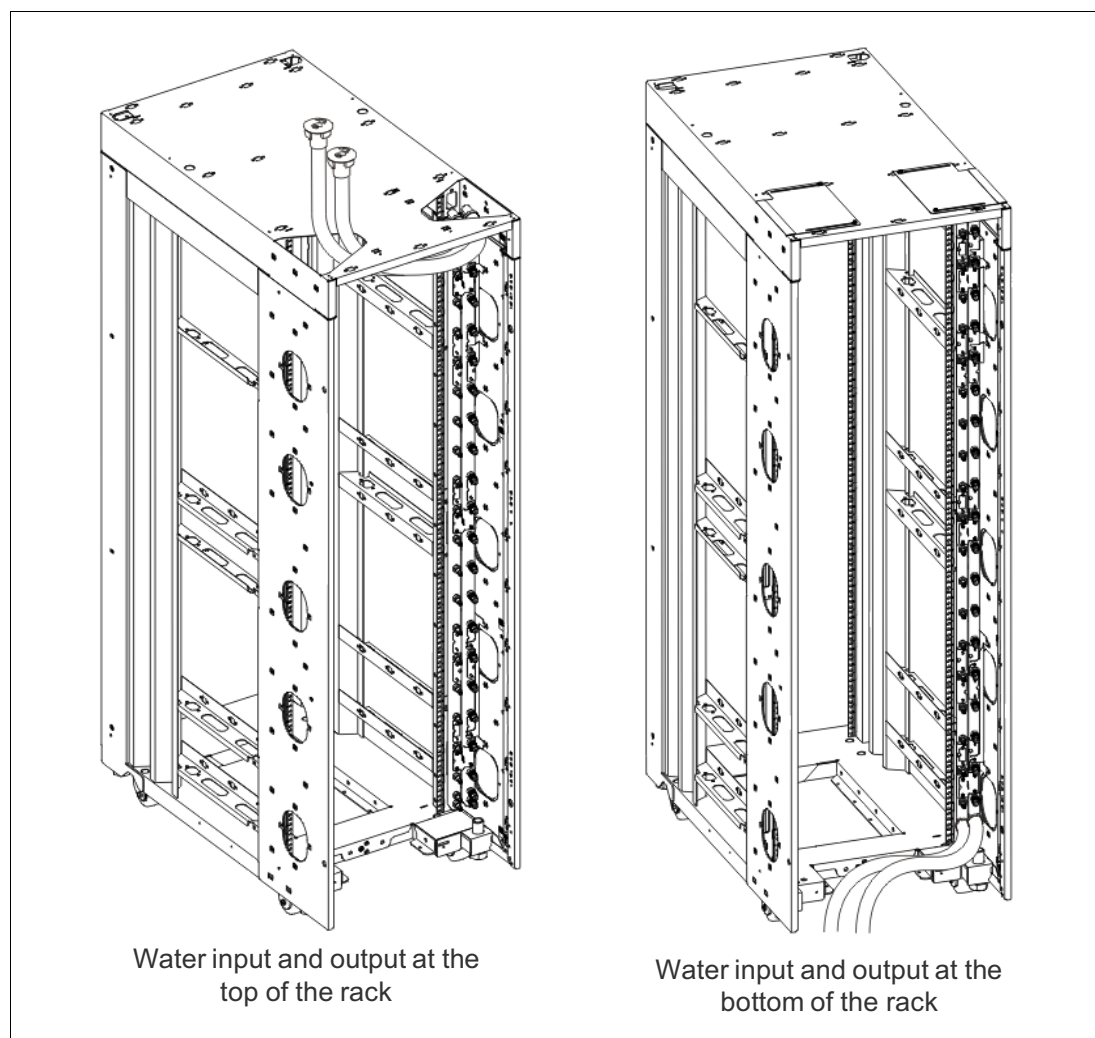
### **Optional water cooling**

When opting for model 8335-GTW of the Power AC922 server, water cooling is mandatory, so the 7965-S42 rack must be ordered with the water-cooling option (feature code #ECR3 or #ECR4). There is no miscellaneous equipment specification (MES) for these features in the field.

These features represent a manifold for water cooling, and provide a water supply and water return for 1 - 20 servers that are mounted in a 7965-S42 Enterprise Slim Rack.

Feature code #ECR3 indicates the manifold with water input and output at the top of the rack. Feature code #ECR4 can be used to order the manifold with water input and output at the bottom of the rack. Because the hose exits might require some space inside the rack, leave a 2U space vacant on the top or bottom of the rack, depending on the location of the hoses that you choose.

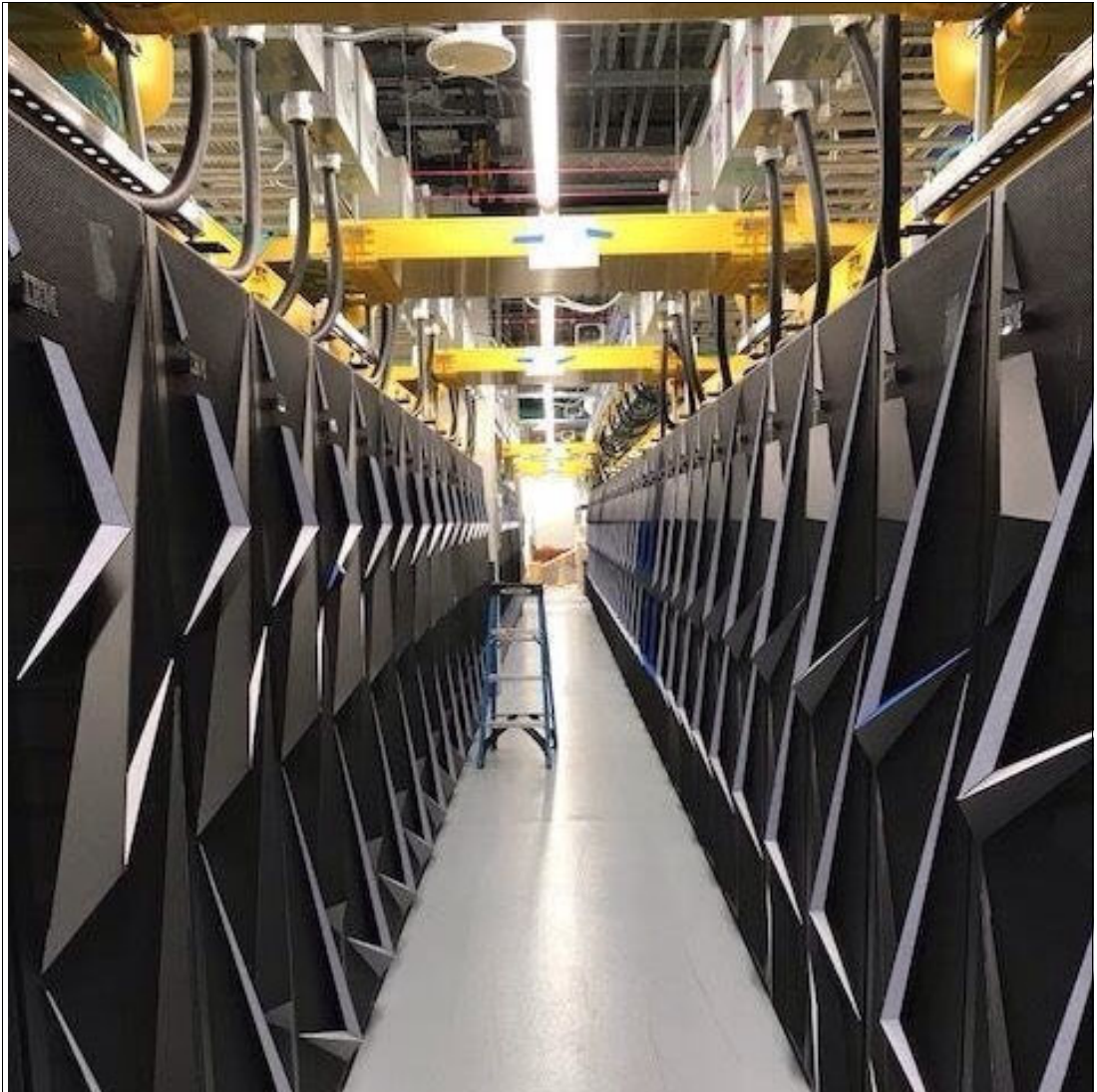
Figure 3-10 shows both options of the water input and output.



*Figure 3-10 Top and bottom water input and output for the 7965-S42 rack*



Figure 3-11 shows a data center rack row with the 7965-S42 racks with water input and output at the top of the rack.



*Figure 3-11 Datacenter rack row with water input and output at the top of racks*

The manifold is mounted on the right side of the rack as viewed from the rear and extends for 40U. The manifold does not interfere with the placement of servers or other I/O drawers. Quick connect fittings are located every 2U on the manifold for water supply and return, which provides 20 pairs of fittings.

Figure 3-12 shows a manifold for the 7965-S42 rack.

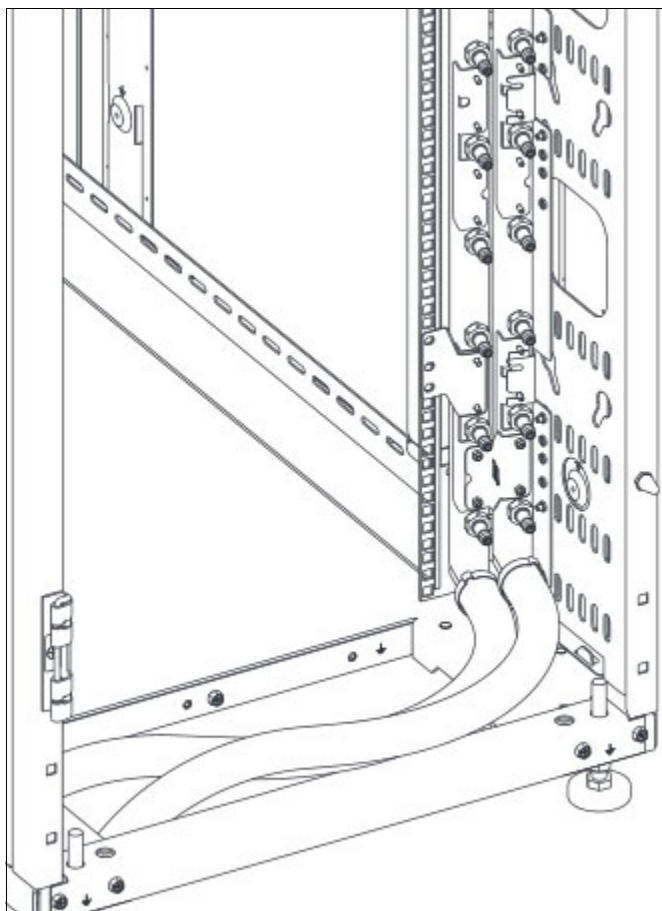


Figure 3-12 Manifold for the 7965-S42 rack

The servers are connected to the manifold through quick-connects. Supply and return hoses from the manifold to the server are provided as part the server cooling feature.

The manifold has one cold water inlet that leads to the rack and one warm water outlet. Two 4.25 m (14-foot) hose kits are provided with the manifold to connect water supply and return. The outer diameter of the hoses is approximately 34.5 mm (1.36 in.).

You must provide a 1-inch ID barb fitting to attach your facility to the hose kit for each hose. Only clean, filtered, and chemically treated water must be used, *not generic building water*.

For more information, see [IBM Knowledge Center](#).

**Important:** Avoid vertically mounted PDUs on the right side as viewed from the rear of the rack. The manifold makes access to PDU impossible. Use either horizontally mounted PDUs, or use vertically mounted PDUs on the left side of the rack.

### 3.5.2 AC power distribution units

AC power distribution is fulfilled by PDUs, which include the AC PDU #7188 and the AC Intelligent PDU+ #7109. The Intelligent PDU+ is identical to #7188 PDUs, but it is equipped with one Ethernet port, one console serial port, and one RS232 serial port for power monitoring.

The PDUs have 12 client-usable IEC 320-C13 outlets. Six groups of two outlets are fed by six circuit breakers. Each outlet is rated up to 10 amps, but each group of two outlets is fed from one 15 amp circuit breaker. Four PDUs can be mounted vertically in the back of the T00 and T42 racks.

Figure 3-13 shows the placement of the four vertically mounted PDUs.

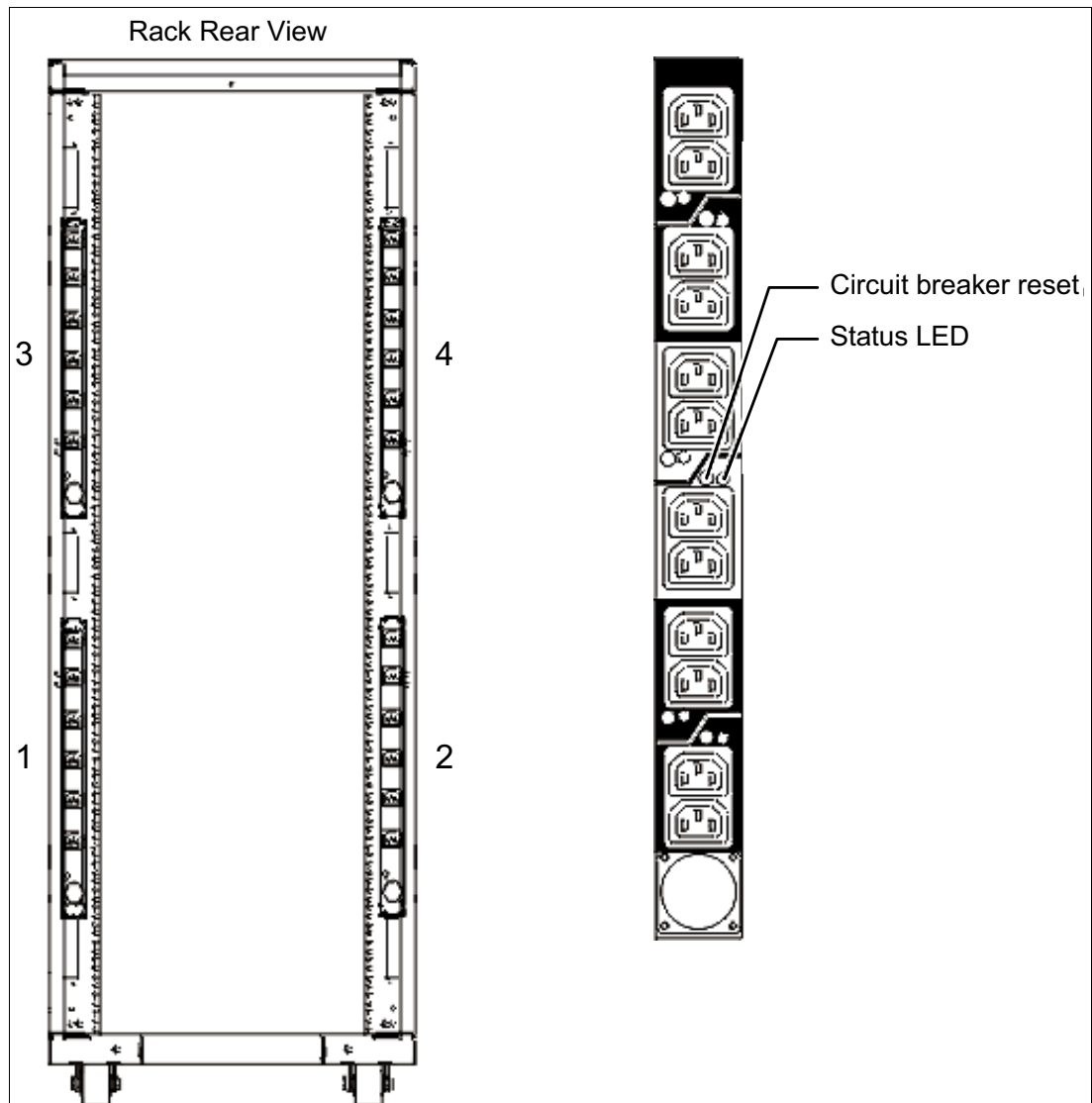


Figure 3-13 PDU placement and PDU view

In the rear of the rack, two more PDUs can be installed horizontally in the T00 rack and three in the T42 rack. The four vertical mounting locations are filled first in the T00 and T42 racks. Mounting PDUs horizontally uses 1U per PDU and reduces the space that is available for other racked components. When mounting PDUs horizontally, the preferred practice is to use fillers in the EIA units that are occupied by these PDUs to facilitate the correct airflow and ventilation in the rack.

The PDU receives power through a UTG0247 power-line connector. Each PDU requires one PDU-to-wall power cord. Various power cord features are available for various countries and applications by varying the PDU-to-wall power cord, which must be ordered separately. Each power cord provides the unique design characteristics for the specific power requirements. To match new power requirements and save previous investments, these power cords can be requested with an initial order of the rack or with a later upgrade of the rack features.

Table 3-5 shows the available wall power cord options for the PDU and iPDU features, which must be ordered separately.

*Table 3-5 Wall power cord options for the PDU and iPDU features*

Feature code	Wall plug	Rated voltage (VAC)	Phase	Rated amperage	Geography
6653	IEC 309, 3P+N+G, 16A	230	3	16 amps/phase	Internationally available
6489	IEC309, 3P+N+G, 32A	230	3	32 amps/phase	EMEA
6654	NEMA L6-30	200 - 208, 240	1	24 amps	US, Canada, LA, and Japan
6655	RS 3750DP (watertight)	200 - 208, 240	1	24 amps	US, Canada, LA, and Japan
6656	IEC 309, P+N+G, 32A	230	1	24 amps	EMEA
6657	PDL	230 - 240	1	32 amps	Australia and New Zealand
6658	Korean plug	220	1	30 amps	North and South Korea
6492	IEC 309, 2P+G, 60A	200 - 208, 240	1	48 amps	US, Canada, LA, and Japan
6491	IEC 309, P+N+G, 63A	230	1	63 amps	EMEA

**Notes:** Ensure that the correct power cord feature is configured to support the power that is being supplied. Based on the power cord that is used, the PDU can supply 4.8 - 19.2 kVA. The power of all of the drawers that are plugged into the PDU must not exceed the power cord limitation.

The Universal PDUs are compatible with previous models.

To better enable electrical redundancy, each server has two power supplies that must be connected to separate PDUs, which are not included in the base order.

For maximum availability, a preferred approach is to connect power cords from the same system to two separate PDUs in the rack, and to connect each PDU to independent power sources.



For detailed power requirements and power cord details about the 7014 racks, see [IBM Knowledge Center](#).

For detailed power requirements and power cord details about the 7965-94Y rack, see [IBM Knowledge Center](#).

### 3.5.3 Rack-mounting rules

Consider the following primary rules when you mount the system into a rack:

- ▶ The system can be placed at any location in the rack. For rack stability, start filling a rack from the bottom.
- ▶ Any remaining space in the rack can be used to install other systems or peripheral devices if the maximum permissible weight of the rack is not exceeded and the installation rules for these devices are followed.
- ▶ Before placing the system into the service position, be sure to follow the rack manufacturer's safety instructions regarding rack stability.

### 3.5.4 OEM racks

The system can be installed in a suitable OEM rack if that the rack conforms to the EIA-310-D standard for 19-inch racks. This standard is published by the Electrical Industries Alliance. For more information, see [IBM Knowledge Center](#).

The website mentions the following key points:

- The front rack opening must be 450 mm wide  $\pm 0.75$  mm (17.72 in.  $\pm 0.03$  in.).

Figure 3-14 is a top view that shows the specification dimensions.

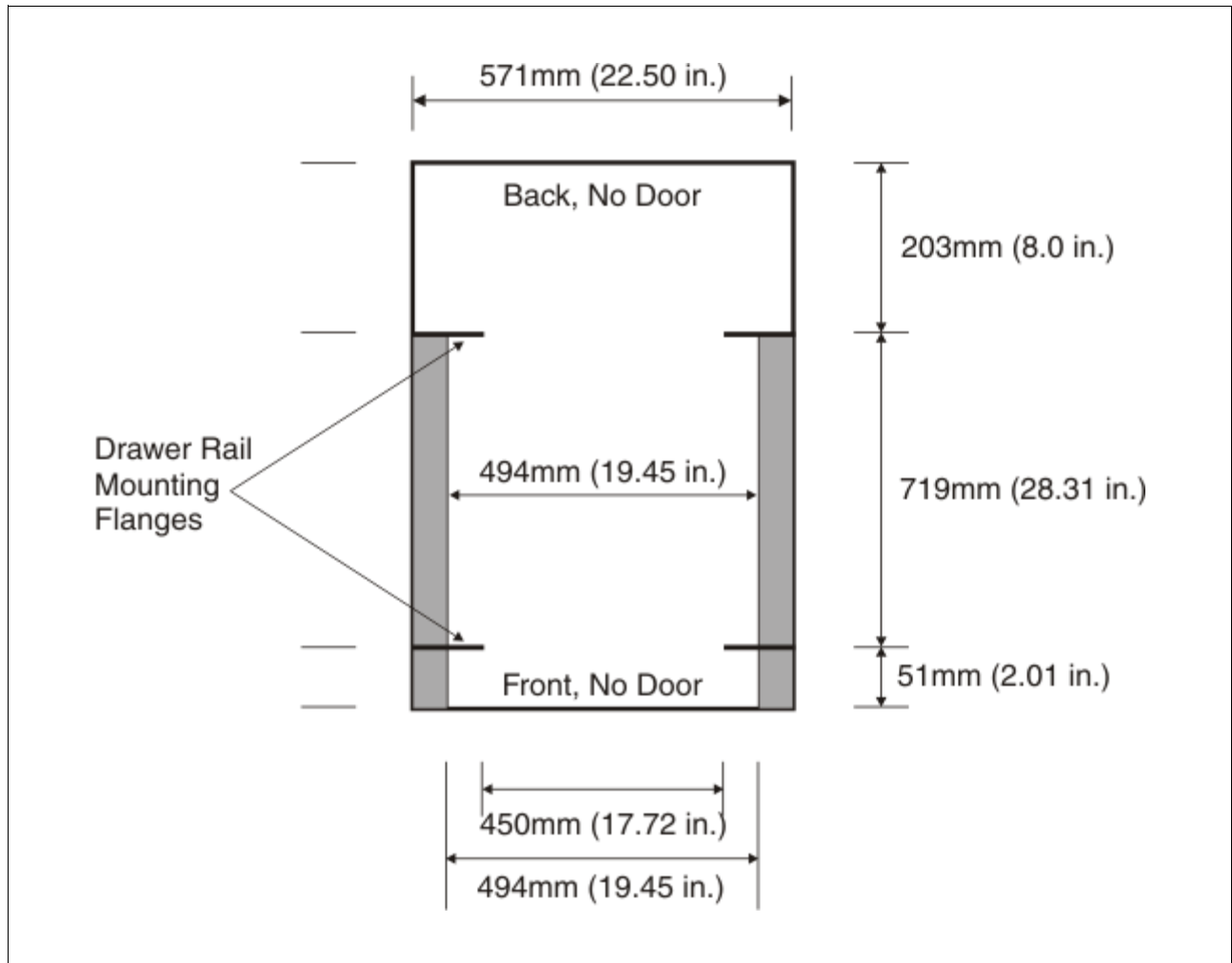


Figure 3-14 Top view of rack specification dimensions (not specific to IBM)

- The rail-mounting holes must be  $465 \text{ mm} \pm 0.8 \text{ mm}$  ( $18.3 \text{ in.} \pm 0.03 \text{ in.}$ ) apart on-center (horizontal width between the vertical columns of holes on the two front-mounting flanges and on the two rear-mounting flanges), as shown in Figure 3-15.

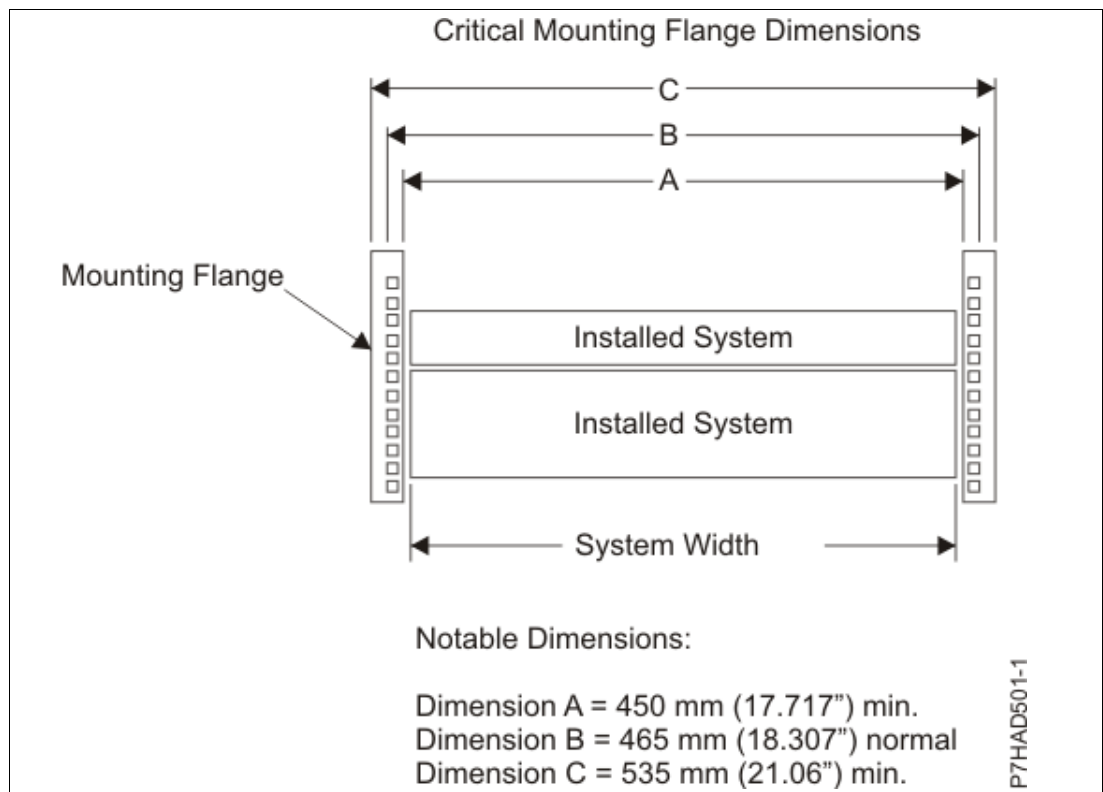


Figure 3-15 Mounting flange dimensions

- The vertical distance between the mounting holes must consist of sets of three holes spaced (from bottom to top) 15.9 mm (0.625 in.), 15.9 mm (0.625 in.), and 12.67 mm (0.5 in.) on-center, which makes each three-hole set of vertical hole spacing 44.45 mm (1.75 in.) apart on center. Rail-mounting holes must be 7.1 mm  $\pm$  0.1 mm (0.28 in.  $\pm$  0.004 in.) in diameter.

Figure 3-16 shows the top front specification dimensions.

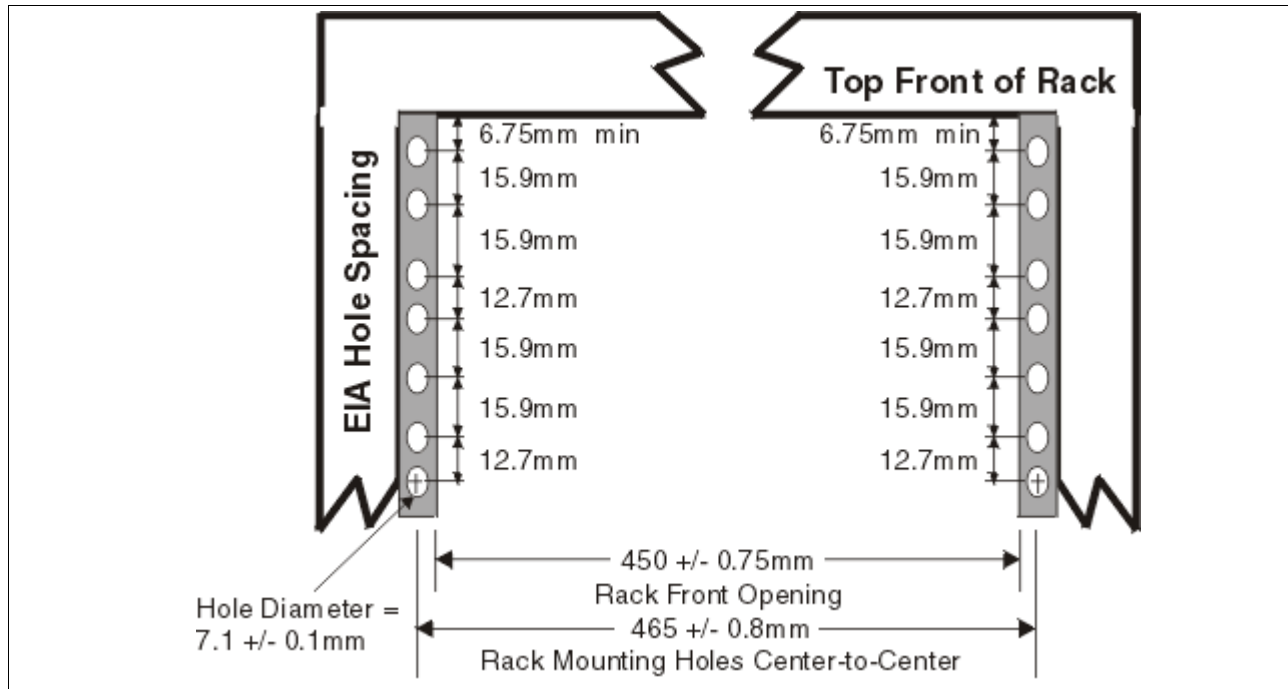


Figure 3-16 Rack specification dimensions top front view

- A minimum rack opening width of 500 mm (19.69 in.) for a depth of 330 mm (12.99 in.) is needed behind the installed system for maintenance, service, and cable management. The recommended depth is at least 254 mm (10 in.) within the rack from the rear rack mount flange to the frame line, as shown in Figure 3-17.

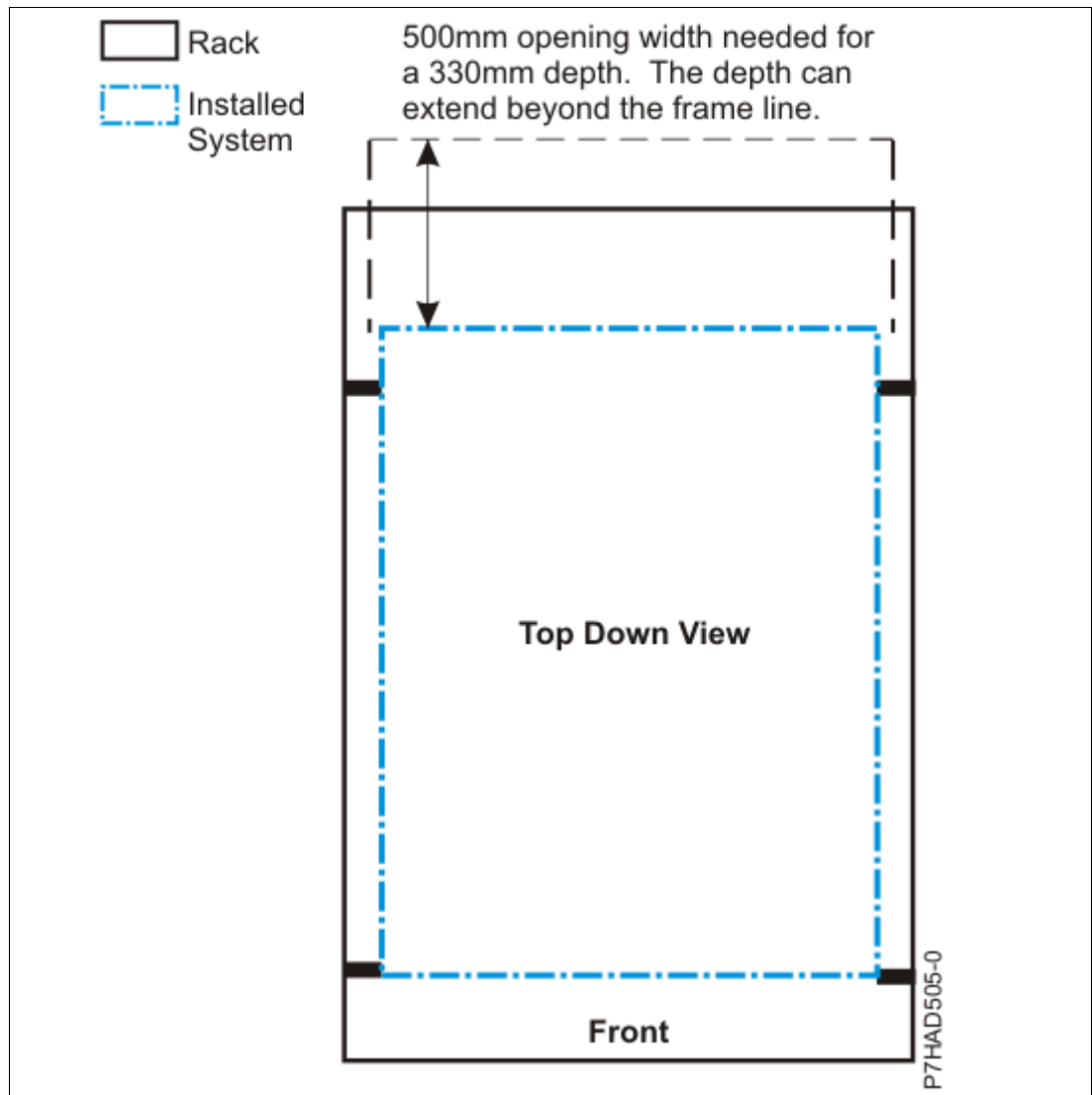


Figure 3-17 OEM rack opening depth



# Related publications

The publications that are listed in this section are considered suitable for a more detailed description of the topics that are covered in this paper.

## IBM Redbooks

The following IBM Redbooks publications provide more information about the topic in this document. Some publications that are referenced in this list might be available in softcopy only.

- ▶ *IBM PowerAI: Deep Learning Unleashed on IBM Power Systems*, SG24-8409
- ▶ *IBM Power System S822LC for High Performance Computing Introduction and Technical Overview*, REDP-5405

You can search for, view, download, or order these documents and other Redbooks publications, Redpapers, web docs, drafts, and additional materials, at the following website:

[ibm.com/redbooks](https://ibm.com/redbooks)

## Online resources

These websites are also relevant as further information sources:

- ▶ IBM Portal for OpenPOWER - POWER9 Monza Module  
[https://www.ibm.com/systems/power/openpower/tgcmDocumentRepository.xhtml?aliasId=POWER9\\_Monza](https://www.ibm.com/systems/power/openpower/tgcmDocumentRepository.xhtml?aliasId=POWER9_Monza)
- ▶ NVIDIA Tesla V100  
<https://www.nvidia.com/en-us/data-center/tesla-v100/>
- ▶ NVIDIA Tesla V100 Performance Guide  
<http://images.nvidia.com/content/pdf/volta-marketing-v100-performance-guide-us-r6-web.pdf>
- ▶ OpenCAPI  
<http://opencapi.org/technical/use-cases/>
- ▶ OpenPOWER Foundation  
<https://openpowerfoundation.org/>

## Help from IBM

IBM Support and downloads

[ibm.com/support](https://ibm.com/support)

IBM Global Services

[ibm.com/services](https://ibm.com/services)









REDP-5472-00

ISBN 0738456705

Printed in U.S.A.

Get connected

