Redbooks
ibm.com/redbooks

IBM Power System S821LC
Technical Overview and Introduction

David Barron
Alexandre Bicas Caldeira
Volker Haug

Cloud

Power Systems

IBM

Redpaper

International Technical Support Organization

IBM Power System S821LC: Technical Overview and Introduction

December 2016

REDP-5406-00

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

First Edition (December 2016)

This edition applies to the IBM Power System S821LC (8001-12C) server.

Contents

Notices

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

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

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

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

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

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

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

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

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

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

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

COPYRIGHT LICENSE:

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

Trademarks

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

Active Memory™
AIX®
DS8000®
Easy Tier®
IBM®
IBM FlashSystem®
IBM Spectrum™
IBM Spectrum Accelerate™
IBM z Systems™
POWER®
Power Systems™
POWER8®
PowerVM®
Redbooks®
Redpaper™
Redbooks (logo)  ®
Storwize®
System Storage®
XIV®

The following terms are trademarks of other companies:

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

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.

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

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

Preface

This IBM® Redpaper™ publication is a comprehensive guide that covers the IBM Power System S821LC (8001-12C) server that uses the latest IBM POWER8® processor technology and supports the Linux operating system (OS).

The Power S821LC server is designed to maximize data center floor space with its dense 1U server design, which helps to reduce infrastructure cost. The Power S821LC server delivers superior performance and exceptional throughput for data center and cloud workloads that require dense virtualization, open source database deployment, and high-performance computing applications.

The Power S821LC server supports up to two processor sockets, offering 16-core 2.328 GHz (3.026 GHz turbo) or 20-core 2.095 GHz (2.827 GHz turbo) POWER8 configurations in a 19-inch rack-mount, 1U (EIA units) drawer configuration. All the cores are activated.

The objective of this paper is to introduce the Power S821LC offering and its relevant functions, including:

- Two POWER8 processors in a 1U form factor
- Dense virtualization and dense database deployment capability-providing more value per server footprint than 1U x86-based alternatives
- Leadership data throughput that is enabled by POWER8 multithreading with up to 4X more threads than x86 designs
- Superior application performance due to 2x per core performance advantage over x86-based systems
- Acceleration of a broad range of workloads with GPUs and superior I/O bandwidth with Coherent Accelerator Processor Interface (CAPI)

This publication is for professionals who want to acquire a better understanding of IBM Power Systems™ products. The intended audience includes the following roles:

- Clients
- Sales and marketing professionals
- Technical support professionals
- IBM Business Partners
- Independent software vendors

This paper expands the current set of IBM Power Systems documentation by providing a desktop reference that offers a detailed technical description of the Power S821LC system.

Authors

This paper was produced by a team of specialists from around the world working at the International Technical Support Organization, Austin Center.

David Barron is a lead engineer in IBM Power Systems Hardware Development. His current focus is on the development of the mechanical, thermal, and power subsystems of scale-out servers that are based on the IBM POWER® processor, and supporting OpenPOWER IBM Business Partners design IBM POWER based servers. He holds a degree in Mechanical Engineering from The University of Texas.

Alexandre Bicas Caldeira is a Certified IT Specialist and is the 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 16 years of experience working on IBM Systems Solutions and worked also as an IBM Business Partner on Power Systems hardware, IBM AIX®, and IBM PowerVM® virtualization products.

Volker Haug is an Executive IT Specialist & Open Group Distinguished IT Specialist within IBM Systems in Germany supporting Power Systems clients and IBM Business Partners. He holds a Diploma degree in Business Management from the University of Applied Studies in Stuttgart. His career includes more than 29 years of experience with Power Systems, IBM AIX, and IBM PowerVM virtualization. He has written several IBM Redbooks® publications about Power Systems and PowerVM. Volker is an IBM POWER8 Champion and a member of the German Technical Expert Council, which is an affiliate of the IBM Academy of Technology.

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:

Adrian Barrera, Scott Carroll, Ray Laning, Ann Lund, Ben Mashak, Michael J Mueller, Padma, Rakesh Sharma, Justin Thaler
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

Comments welcome

Your comments are important to us!

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

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

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, 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

1

Architecture and technical description

Today, the number of sources generating data is leading to an exponential growth in the data volume. Making sense of this data and doing it faster than the competition can leads to an unprecedented opportunity to gain valuable insights and apply these insights at the best point of impact to improve your business results.

The IBM Power System S821LC (8001-12C) server is designed to maximize data center floor space with its dense 1U server design, which helps to reduce infrastructure cost. The Power S821LC server delivers superior performance and exceptional throughput for data center and cloud workloads that require dense virtualization, open source database deployment, and high-performance computing applications.

The Power S821LC server supports up to two processor sockets offering 8-core, 10-core, 16-core, and 20-core POWER8 configurations and up to 512 GB DDR4 memory, based on a flexible design to address cloud and data center workloads ranging from database to high-performance computing applications.

The Power S821LC hardware advantages lead to superior application performance:

- Up to two POWER8 processors in a 1U form factor
- Leadership data throughput that is enabled by POWER8 multithreading with up to 4X more threads than x86 designs
- Superior application performance that is designed to provide 2x per core performance advantage over x86 based systems
- Support for the new NVMe drives

The frequency for both processor options can also be boosted dynamically as required.

The following sections provide detailed information about the Power S821LC server models.

1.1 Power S821LC system hardware overview

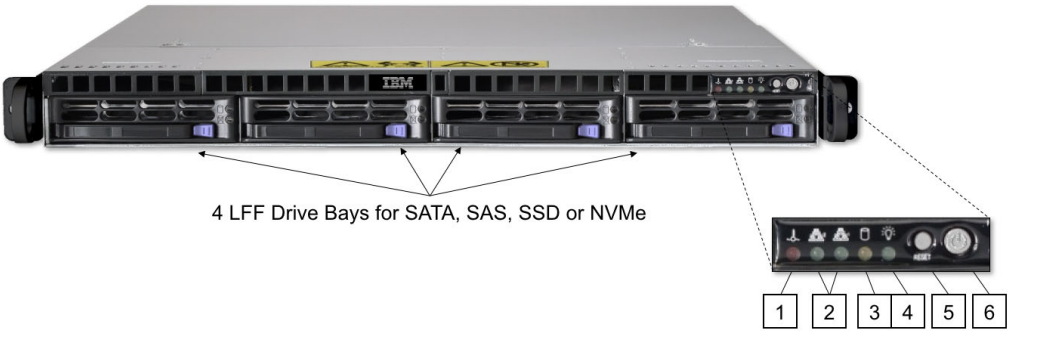
The Power S821LC server is a powerful, one-socket or two-socket server that offers 8, 10, 16, or 20 fully activated cores and the I/O configuration flexibility to meet today's growth and tomorrow's processing needs. The server features are designed for dense deployments with a 2U server performance in a 1U form factor.

By incorporating OpenPOWER Foundation community innovations, the Power S821LC server has a low acquisition cost through system optimization (industry-standard memory, focused configurations, focused I/O and expansion, and industry-standard warranty), which makes it ideal for clients that want the advantages of running their applications on a platform that is designed and optimized for data and Linux.

This section provides an overview of the system hardware. The new Power S821LC model offers:

- Powerful IBM POWER8 Single Chip Module (SCM) processors that offer 2.328 GHz or 2.095 GHz performance
- Up to 512 GB of DDR4 memory
- Four PCIe slots, of which three are Coherent Accelerator Processor Interface (CAPI)-enabled (a maximum of two CAPI devices can be used concurrently)
- One NVIDIA K80 or P100 GPU supported
- Four SATA/SAS/NVMe drives bays, allowing for a maximum of 32 TB hard disk drive (HDD) storage or 15.2 TB of solid-state drive (SSD) storage
- Two 64 GB or 128 GB SATA DOM flash modules (local internal boot devices)
- Two hot-swap, redundant 100-127 or 200-240 V AC power supplies. If GPU is configured 200-240 V AC power supplies must be used.
- 19-inch rack-mount hardware (1U)

Figure 1-1 shows the front view and a description of the major LEDs.



1	Information LED
2	Network Port Activity LED
3	Disk Activity LED
4	Power On LED
5	Reset Button
6	Power Button

Figure 1-1 Power S821LC server front view

The rear view, including PCIe slot identification and native ports, is shown in Figure 1-2.

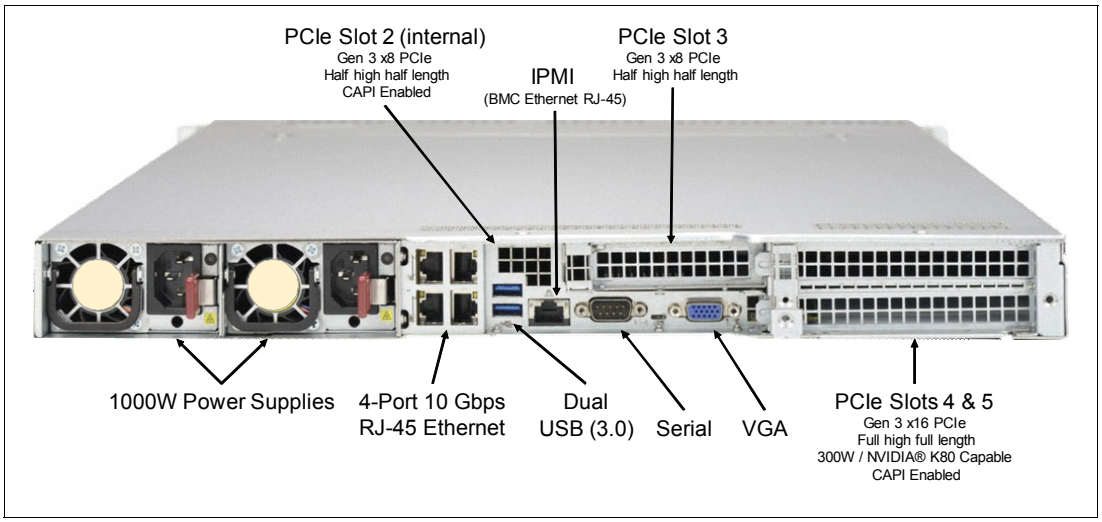


Figure 1-2 Power S821LC server rear view

Note: The four native Ethernet ports support 10Gb, 1Gb, or 100Mb speeds. Setting each Ethernet port to its own unique speed is supported.

The Power S821LC server is extremely dense, so its PCIe cards must be installed by using PCIe risers cards. There might be up to three PCIe riser cards in the system, configured according to the number of processor features:

- PCIe riser 1: Configured for systems with one or two processors that are installed, it allows for a 4-port 10Gbps Ethernet Adapter and an additional internal-only PCIe x8 slot that is used mostly by the internal storage adapters.
- PCIe riser 2: Configured for systems with one or two processors that are installed, it allows for one external PCIe x8 slot.
- PCIe riser 3: Configured only for systems with two processors that are installed, it allows for two external PCIe x16 slots that are CAPI-capable. If an NVIDIA GPU is configured, it must be installed on this PCIe riser card.

A diagram of the PCIe riser cards and their corresponding external slots is shown in Figure 1-3.

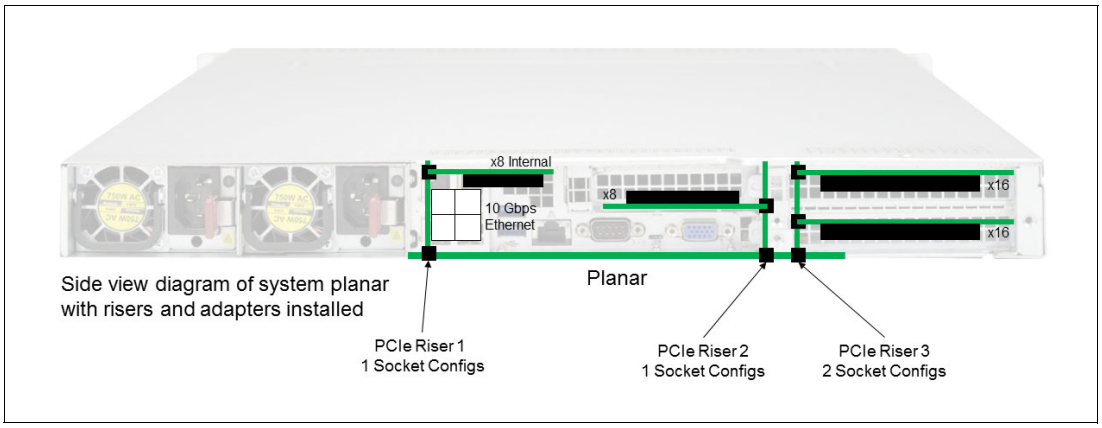


Figure 1-3 Rear-view diagram of system planar and PCIe riser cards

Figure 1-4 show the top view of a Power S821LC server.

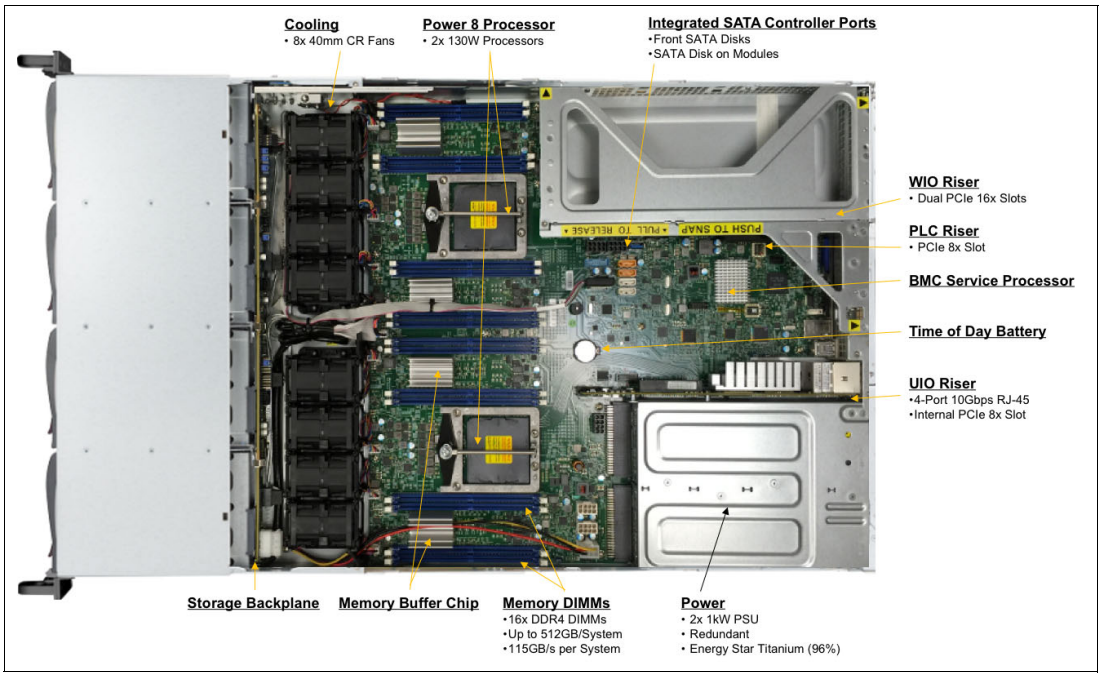


Figure 1-4 Power S821LC server top view

1.2 System architecture

The system is designed to maximize processor performance in a dense 1U form factor to maximize compute density, but also allow for configurability in terms of storage capacity, memory capacity, memory bandwidth, and PCIe adapter allowance. Figure 1-5 on page 5 illustrates the overall architecture.

Bandwidths that are provided throughout the section are theoretical maximums that are used for reference.

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.

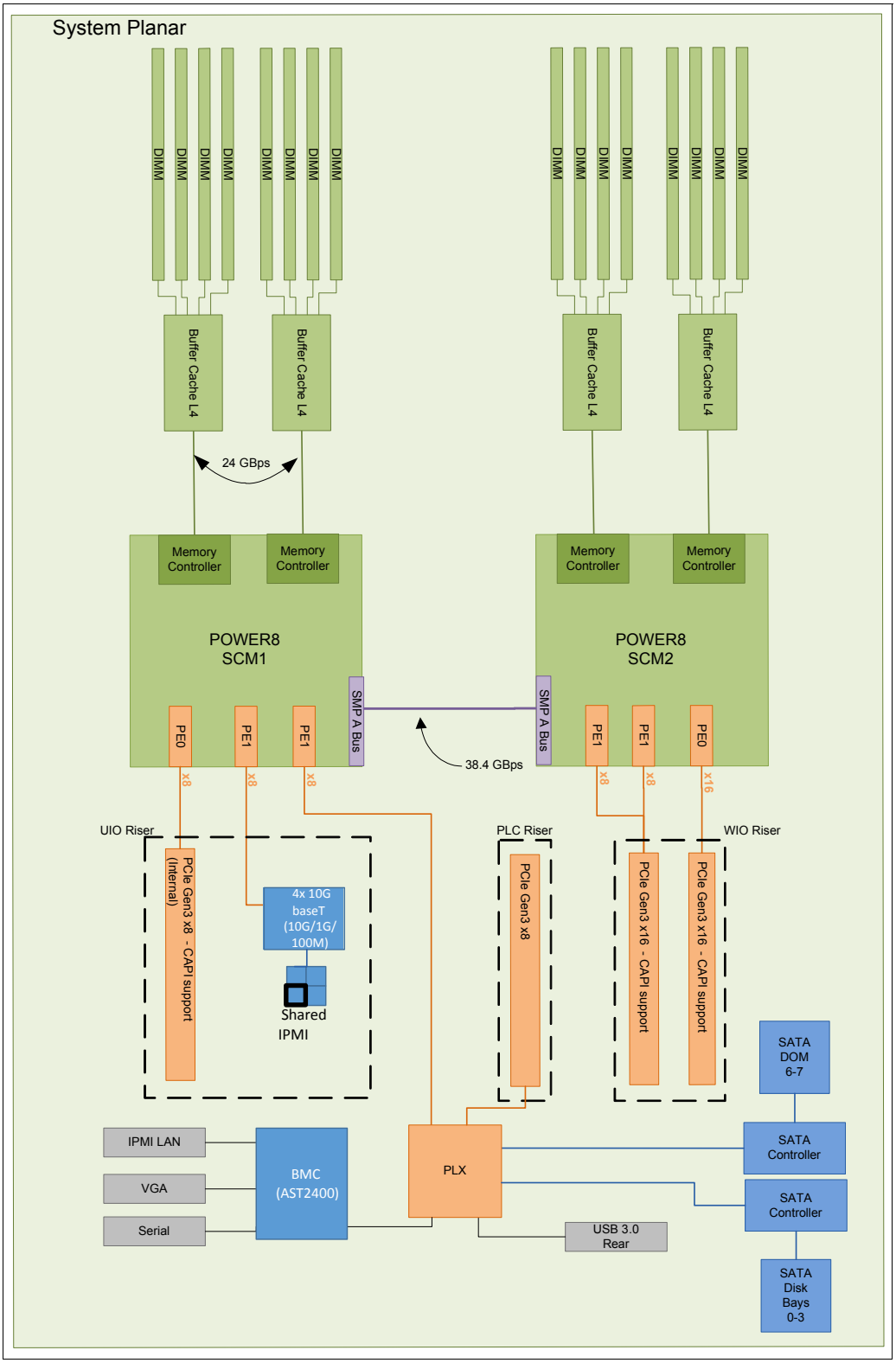


Figure 1-5 System architecture logical diagram

Some PCIe slots are bound to a specific processor. The overall processor to PCIe slot mapping and major component identification is provided in Figure 1-6 as a top-level depiction of the main system planar.

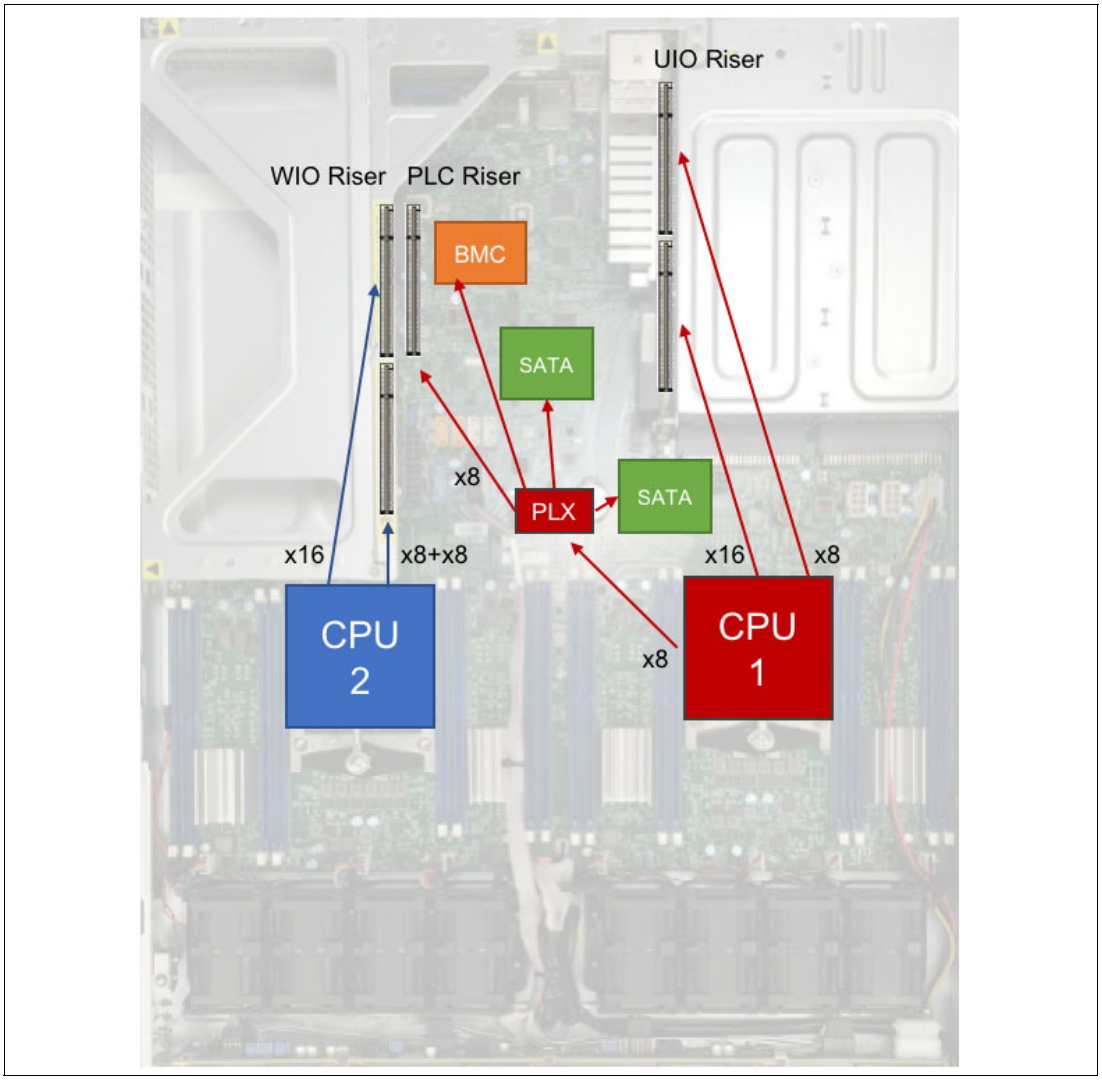


Figure 1-6 System planar overview with PCIe to CPU identification

1.3 Physical package

The Power S821LC server is offered exclusively as a rack mount 1U server. The width, depth, height, and weight of the server are shown in Table 1-1.

Table 1-1 Physical dimensions for the Power S821LC server

Dimension	Power S821LC server
Width	449 mm (17.7 in)
Depth	705.3 mm (27.76 in)
Height	43 mm (1.7 in)
Weight (maximum configuration)	18.59 kg (41 lbs)

1.4 Operating environment

The Power S821LC server is designed to operate at nominal processor frequencies within the ASHRAE A2 envelope. ASHRAE is a global society providing standards for environmentals.

For more information about ASHRAE A2, see the following website:

https://www.ashrae.org/standards-research--technology/standards--guidelines

1.5 Leveraging innovations of the OpenPOWER Foundation

This system is designed to incorporate a plethora of innovative technology, which is optimized to function with POWER processors through deep partnerships within the OpenPOWER Foundation. Details of each OpenPOWER Foundation member and its main contribution are shown in Table 1-2.

Table 1-2 Contribution of OpenPOWER Foundation members

OpenPOWER Foundation member	Contribution
NVIDIA	Tesla GPU accelerator
Ubuntu by Canonical	Ubuntu 16.04 Linux OS
Mellanox	InfiniBand/Ethernet connectivity in and out of server
HGST	Optional NVMe adapters
Alpha Data with Xilinx FPGA	Optional CAPI accelerator
Broadcom	Optional PCIe adapters
QLogic	Optional Fibre Channel PCIe
Samsung	SSDs & NVMe
Hynix, Samsung, and Micron	DDR4 memory
IBM	POWER8 processors

1.6 Base system and standard features

The server chassis of the Power S821LC server contains one or two processor modules. Each POWER8 processor module is either 8-core or 10-core and has a 64-bit architecture, up to 512 KB of L2 cache per core, and up to 8 MB of L3 cache per core. Processor options are 8-core POWER8 2.328 GHz (#EKP1) or 10-core POWER8 2.095 GHz (#EKP2).

The Power S821LC server provides 16 DIMM memory slots, of which eight, 12, or 16 can be populated. Memory features that are supported are 4 GB, 8 GB, 16 GB, and 32 GB, allowing for a maximum system memory of 512 GB.

The Power S821LC server provides the following features:

- Power Systems server that is built with POWER8 processor modules:
  - 8-core, 2.328 GHz
  - 10-core, 2.095 GHz
- High-performance DDR4 memory:
  - 4 GB (#EKM0), 8 GB (#EKM1), 16 GB (#EKM2), and 32 GB (#EKM3) memory features
  - Up to 512 GB memory
- Storage bays: Four LFF/SFF bays for four SAS/SATA/SSD drives or four NVMe drives
- PCIe Gen3 slots:
  - Two PCIe x16 G3 FH slots, CAPI-enabled
  - One PCIe x8 G3 LP slot, CAPI-enabled (internal slot)
  - One PCIe x8 G3 LP slot
- Two rear USB 3.0 ports
- Two hot-swap redundant 100-127 or 200-240 V AC power supplies (no power redundancy with GPUs installed. If GPU is installed 200-249 V AC power supples must be used)
- 19-inch rack-mount hardware (1U)

The Power S821LC server is composed of a base system, which determines whether the system accepts one or two processors.

Note: A one-socket system does not support the WIO PCIe slots. All Power S821LC base systems include an NVMe-enabled midplane allowing all four drive slots to be compatible with SATA, SAS, and NVMe drives. An additional storage controller might be required.

The two base system feature codes and descriptions are listed in Table 1-3.

Table 1-3 Available base systems and descriptions

Feature code	Description
EKB0	One-socket base system with LFF high-function drive midplane (NVMe drives are supported.)
EKB4	Two-socket base system with LFF high-function drive midplane (NVMe drives are supported.)

In addition to base system selection, a minimum of eight DIMMs and one processor are required to create a minimally orderable valid system.

1.6.1 Power S821LC server minimum configuration

The minimum Power S821LC server initial order must include one processor module, eight 4 GB or larger memory DIMMs, two power supplies, two line cords, rack-mounting hardware, a system software indicator, a rack integrator specify, and a Language Group Specify.

Linux is the supported OS.

The server supports bare-metal Linux installations only.

Note: Neither PowerVM or PowerKVM virtualization is supported on the Power S821LC server. KVM virtualization is available on this server through the appropriate Linux distribution.

1.7 Available features

The required and optional available features are described in the following sections.

1.7.1 IBM POWER8 processor

This section introduces the available POWER8 processors for the Power S821LC server and describes the main characteristics and general features of the processor.

Processor availability in the Power S821LC server

The number of processors in the system is determined by the base system that is selected:

- One-socket Fab Assembly with NVMe Backplane (#EKB0)
- Two-socket Fab Assembly with NVMe Backplane (#EKB4)

Table 1-4 shows the available processor features that are available for the Power S821LC server. Additional information about the POWER8 processors, including details about the core architecture, multithreading, memory access, and CAPI, can be found in the following sections.

Table 1-4 Processor features and description

Feature code	Description
EKP1	8-core 2.32 GHz POWER8 processor
EKP2	10-core 2.09 GHz POWER8 processor

POWER8 processor overview

The POWER8 processor is manufactured by using the IBM 22 nm Silicon-On-Insulator (SOI) technology. Each chip is 649 mm^2 and contains 4.2 billion transistors. As shown in Figure 1-7, the chip contains up to 12 cores, two memory controllers, Peripheral Component Interconnect Express (PCIe) Gen3 I/O controllers, and an interconnection system that connects all components within the chip. Each core has 512 KB of L2 cache, and all cores share 96 MB of L3 embedded DRAM (eDRAM). The interconnect also extends through module and system board technology to other POWER8 processors in addition to DDR4 memory and various I/O devices.

POWER8 processor-based systems use memory buffer chips to interface between the POWER8 processor and DDR4 memory. Each buffer chip also includes an L4 cache to reduce the latency of local memory accesses.

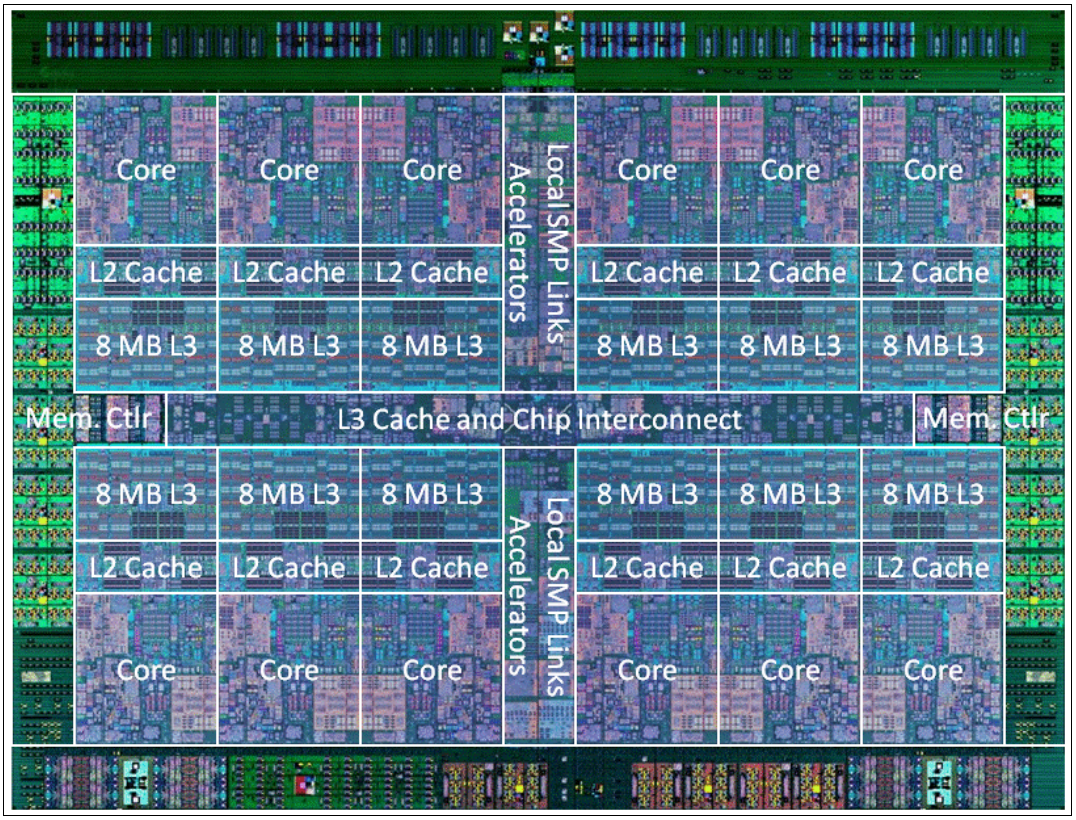


Figure 1-7 The POWER8 processor chip

Note: The processor that is shown is a 12-core. This system uses processors with 8 or 10 cores.

The POWER8 processor is available in system offerings from single-socket servers to multi-socket enterprise servers. It incorporates a triple-scope broadcast coherence protocol over local and global SMP links to provide superior scaling attributes. Multiple-scope coherence protocols reduce the amount of SMP link bandwidth that is required by attempting operations on a limited scope (single chip or multi-chip group) when possible. If the operation cannot complete coherently, the operation is reissued by using a larger scope to complete the operation.

Here are additional features that can augment the performance of the POWER8 processor:

- Support for DDR4 memory through memory buffer chips that offload the memory support from the POWER8 memory controller.
- An L4 cache within the memory buffer chip that reduces the memory latency for local access to memory behind the buffer chip; the operation of the L4 cache is not apparent to applications running on the POWER8 processor. Up to 128 MB of L4 cache can be available for each POWER8 processor.
- Hardware transactional memory.
- On-chip accelerators, including on-chip encryption, compression, and random number generation accelerators.
- CAPI, which allows accelerators that are plugged into a PCIe slot to access the processor bus by using a low latency, high-speed protocol interface.
- Adaptive power management.

Table 1-5 summarizes the technology characteristics of the POWER8 processor.

Table 1-5 Summary of POWER8 processor technology

Technology	POWER8 processor
Die size	649 mm^2
Fabrication technology	► 22 nm lithography ► Copper interconnect ► SOI ► eDRAM
Maximum processor cores	12
Maximum execution threads core/chip	8/96
Maximum L2 cache core/chip	512 KB/6 MB
Maximum On-chip L3 cache core/chip	8 MB/96 MB
Maximum L4 cache per chip	128 MB
Maximum memory controllers	2
SMP design-point	16 sockets with POWER8 processors
Compatibility	With prior generation of IBM POWER processors

POWER8 processor core

The POWER8 processor core is a 64-bit implementation of the IBM Power Instruction Set Architecture (ISA) Version 2.07 and has the following features:

- Multi-threaded design, which is capable of up to eight-way simultaneous multithreading (SMT)
- 32 KB, eight-way set-associative L1 instruction cache
- 64 KB, eight-way set-associative L1 data cache
- Enhanced prefetch, with instruction speculation awareness and data prefetch depth awareness
- Enhanced branch prediction, which uses both local and global prediction tables with a selector table to choose the preferred predictor
- Improved out-of-order execution

- Two symmetric fixed-point execution units
- Two symmetric load/store units and two load units, all four of which can also run simple fixed-point instructions
- An integrated, multi-pipeline vector-scalar floating point unit for running both scalar and Single Instruction Multiple Data (SIMD) type instructions, including the Vector Multimedia eXtension (VMX) instruction set and the improved Vector Scalar eXtension (VSX) instruction set, and capable of up to sixteen floating point operations per cycle (eight double precision or sixteen single precision)
- In-core Advanced Encryption Standard (AES) encryption capability
- Hardware data prefetching with 16 independent data streams and software control
- Hardware decimal floating point (DFP) capability.

More information about Power ISA Version 2.07 can be found at the following website:

https://www.power.org/wp-content/uploads/2013/05/PowerISA_V2.07_PUBLIC.pdf

Figure 1-8 shows a picture of the POWER8 core, with some of the functional units highlighted.

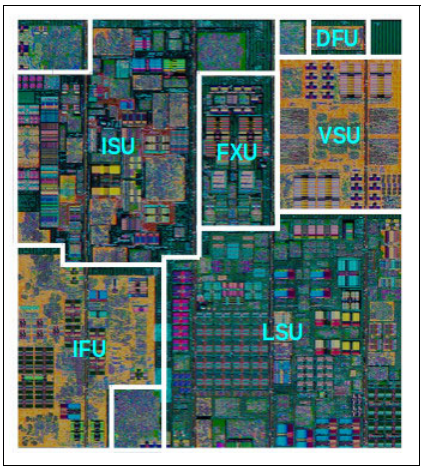


Figure 1-8 POWER8 processor core

Simultaneous multithreading

POWER8 processor advancements in multi-core and multi-thread scaling are remarkable. A significant performance opportunity comes from parallelizing workloads to enable the full potential of the microprocessor, and the large memory bandwidth. Application scaling is influenced by both multi-core and multi-thread technology.

SMT allows a single physical processor core to dispatch simultaneously instructions from more than one hardware thread context. With SMT, each POWER8 core can present eight hardware threads. Because there are multiple hardware threads per physical processor core, additional instructions can run at the same time. SMT is primarily beneficial in commercial environments where the speed of an individual transaction is not as critical as the total number of transactions that are performed. SMT typically increases the throughput of workloads with large or frequently changing working sets, such as database servers and web servers.

Table 1-6 shows a comparison between the different POWER processors options for the Power S821LC server and the number of threads that are supported by each SMT mode.

Table 1-6 SMT levels that are supported by a Power S821LC server

Cores per system	SMT mode	Hardware threads per system
8 (One-Socket System)	Single Thread (ST)	8
8 (One-Socket System)	SMT2	16
8 (One-Socket System)	SMT4	32
8 (One-Socket System)	SMT8	64
10 (One-Socket System)	Single Thread (ST)	10
10 (One-Socket System)	SMT2	20
10 (One-Socket System)	SMT4	40
10 (One-Socket System)	SMT8	80
16 (Two Socket System)	Single Thread (ST)	16
16 (Two-Socket System)	SMT2	32
16 (Two-Socket System)	SMT4	64
16 (Two-Socket System)	SMT8	128
20 (Two-Socket System)	Single Thread (ST)	20
20 (Two-Socket System)	SMT2	40
20 (Two-Socket System)	SMT4	80
20 (Two-Socket System)	SMT8	160

The architecture of the POWER8 processor, with its larger caches, larger cache bandwidth, and faster memory, allows threads to have faster access to memory resources, which translates to a more efficient usage of threads. So, POWER8 allows more threads per core to run concurrently, increasing the total throughput of the processor and of the system.

Memory access

On the Power S821LC server, each POWER8 module has two memory controllers, each connected to one memory channel. Each memory channel operates at 1600 MHz and connects to a memory buffer that is responsible for many functions that were previously on the memory controller, such as scheduling logic and energy management. The memory buffer also has 16 MB of L4 cache. Each memory buffer connects to four industry standard DDR4 DIMMs, as shown in Figure 1-9.

With four memory channels that are populated per memory buffer (two per socket) and four DIMMs per buffer at 32 GB per DIMM, the system can address up to 512 GB of total memory.

Note: In a one-socket configuration, the number of populated memory buffers is reduced to two, so the maximum memory capacity for a one-socket system is 256 GB.

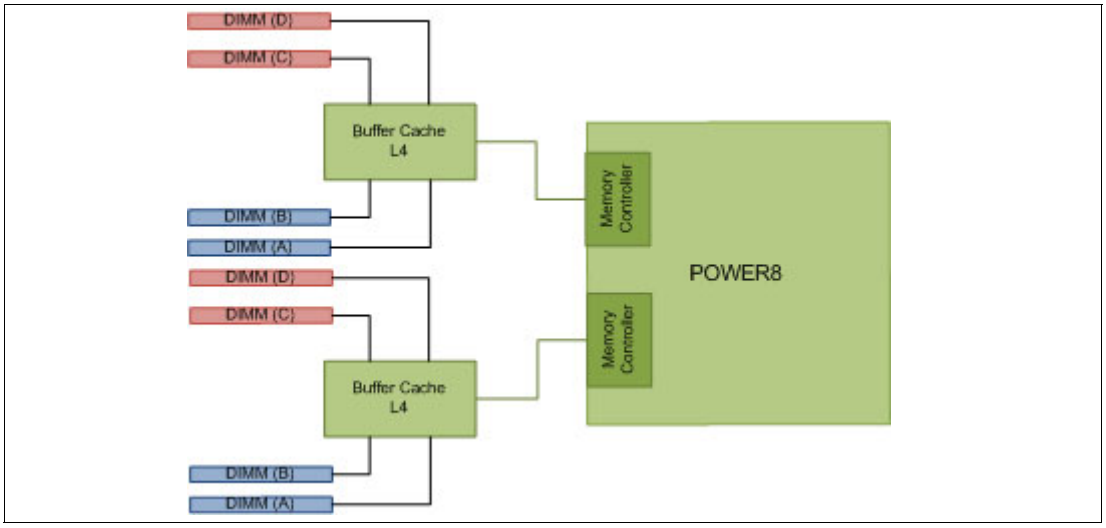


Figure 1-9 Power S821LC server memory logical diagram and DIMM pairs

On-chip L3 cache Innovation and Intelligent Cache

The POWER8 processor uses a breakthrough in material engineering and microprocessor fabrication to implement the L3 cache in eDRAM and place it on the processor die. L3 cache is critical to a balanced design, as is the ability to provide good signaling between the L3 cache and other elements of the hierarchy, such as the L2 cache or SMP interconnect.

The on-chip L3 cache is organized into separate areas with differing latency characteristics. Each processor core is associated with a fast 8 MB local region of L3 cache (FLR-L3), but also has access to other L3 cache regions as a shared L3 cache. Additionally, each core can negotiate to use the FLR-L3 cache that is associated with another core, depending on reference patterns. Data can also be cloned to be stored in more than one core's FLR-L3 cache, again depending on reference patterns. This Intelligent Cache management enables the POWER8 processor to optimize the access to L3 cache lines and minimize overall cache latencies.

Figure 1-7 on page 10 shows the on-chip L3 cache, and highlights the fast 8 MB L3 region that is closest to a processor core.

The innovation of using eDRAM on the POWER8 processor die is significant for several reasons:

- Latency improvement

  A six-to-one latency improvement occurs by moving the L3 cache on-chip compared to L3 accesses on an external (on-ceramic) Application Specific Integrated Circuit (ASIC).

- Bandwidth improvement

  A 2x bandwidth improvement occurs with on-chip interconnect. Frequency and bus sizes are increased to and from each core.

- No off-chip driver or receivers

  Removing drivers or receivers from the L3 access path lowers interface requirements, conserves energy, and lowers latency.

- Small physical footprint

  The performance of eDRAM when implemented on-chip is similar to conventional SRAM but requires far less physical space. IBM on-chip eDRAM uses only a third of the components that conventional SRAM uses, which has a minimum of six transistors to implement a 1-bit memory cell.

- Low energy consumption

  The on-chip eDRAM uses only 20% of the standby power of SRAM.

L4 cache and memory buffer

POWER8 processor-based systems introduce an additional level in memory hierarchy. The L4 cache is implemented together with the memory buffer. Each memory buffer contains 16 MB of L4 cache. On a Power S821LC server, you can have up to 64 MB of L4 cache by using all of the four memory buffers.

Figure 1-10 shows a picture of the memory buffer, where you can see the 16 MB L4 cache and processor links and memory interfaces.

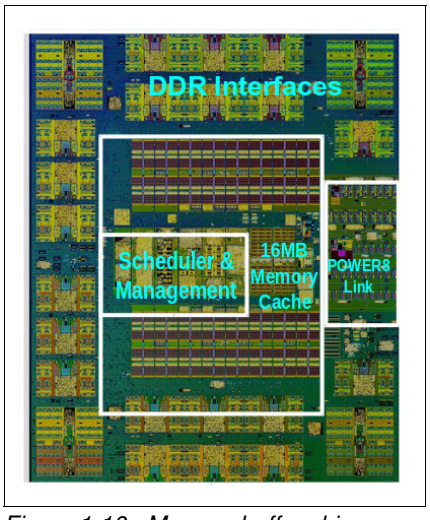


Figure 1-10 Memory buffer chip

Table 1-7 shows a comparison of the different levels of cache in the POWER8 processor.

Table 1-7 POWER8 cache hierarchy

Cache	POWER8
L1 instruction cache: Capacity/associativity	32 KB, 8-way
L1 data cache: Capacity/associativity bandwidth	64 KB, 8-way Four 16 B reads or one 16 B writes per cycle
L2 cache: Capacity/associativity bandwidth	512 KB, 8-way Private 64 B reads and 16 B writes per cycle
L3 cache: Capacity/associativity bandwidth	On-Chip 8 MB/core, 8-way 32 B reads and 32 B writes per cycle
L4 cache: Capacity/associativity bandwidth	Off-Chip 16 MB/buffer chip, 16-way Up to 8 buffer chips per socket

Hardware transactional memory

Transactional memory is an alternative to lock-based synchronization. It attempts to simplify parallel programming by grouping read and write operations and running them as a single operation. Transactional memory is like database transactions, where all shared memory accesses and their effects are either committed all together or discarded as a group. All threads can enter the critical region simultaneously. If there are conflicts in accessing the shared memory data, threads try accessing the shared memory data again or are stopped without updating the shared memory data. Therefore, transactional memory is also called a *lock-free synchronization*. Transactional memory can be a competitive alternative to lock-based synchronization.

Transactional memory provides a programming model that makes parallel programming easier. A programmer delimits regions of code that access shared data and the hardware runs these regions atomically and in isolation, buffering the results of individual instructions, and trying execution again if isolation is violated. Generally, transactional memory allows programs to use a programming style that is close to coarse-grained locking to achieve performance that is close to fine-grained locking.

Most implementations of transactional memory are based on software. The POWER8 processor-based systems provide a hardware-based implementation of transactional memory that is more efficient than the software implementations and requires no interaction with the processor core, therefore allowing the system to operate in maximum performance.

Coherent Accelerator Processor Interface

CAPI defines a coherent accelerator interface structure for attaching special processing devices to the POWER8 processor bus.

The CAPI can attach accelerators that have coherent shared memory access with the processors in the server and share full virtual address translation with these processors, which use a standard PCIe Gen3 bus.

Applications can have customized functions in FPGAs and enqueue work requests directly in shared memory queues to the FPGA, and by using the same effective addresses (pointers) it uses for any of its threads running on a host processor. From a practical perspective, CAPI allows a specialized hardware accelerator to be seen as an additional processor in the system, with access to the main system memory, and coherent communication with other processors in the system.

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 is possible because the adapter is not operating as a traditional I/O device, and there is no device driver layer to perform processing. It also presents a simpler programming model.

Figure 1-11 shows a high-level view of how an accelerator communicates with the POWER8 processor through CAPI. The POWER8 processor provides a Coherent Attached Processor Proxy (CAPP), which is responsible for extending the coherence in the processor communications to an external device. The coherency protocol is tunneled over standard PCIe Gen3, effectively making the accelerator part of the coherency domain.

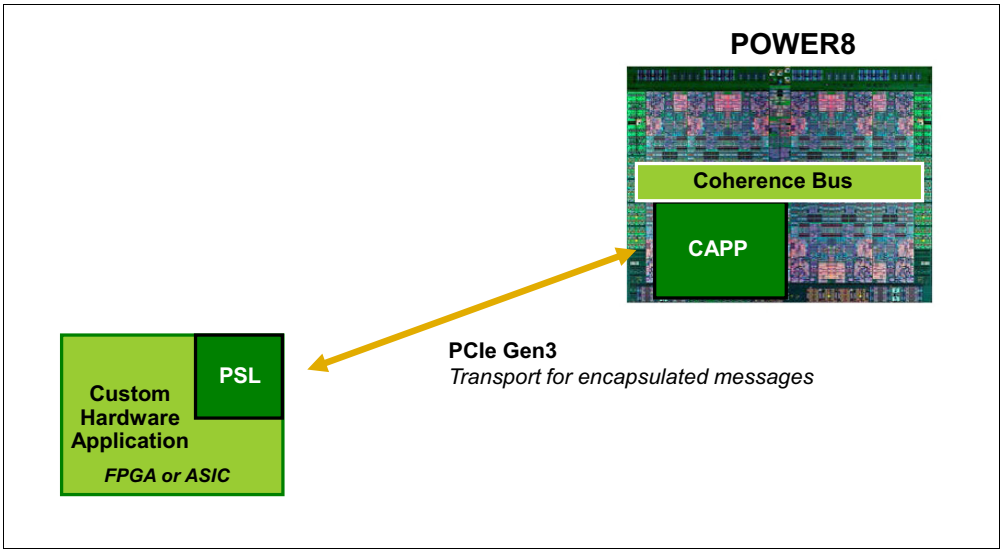


Figure 1-11 CAPI accelerator that is attached to the POWER8 processor

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, and the need for Direct Memory Access that requires system calls to the operating system (OS) kernel. By removing these layers, the data transfer operation requires much fewer clock cycles in the processor, improving the I/O performance.

The implementation of CAPI on the POWER8 processor allows hardware companies to develop solutions for specific application demands and use the performance of the POWER8 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.

Note: There is no MES upgrade option between 8-core and 10-core processors.

1.7.2 Memory

The available memory options and bandwidth are described in this section.

Memory availability in the Power S821LC server

The Power S821LC server is a one-socket or two-socket system that supports POWER8 SCM processor modules. It supports a maximum of 16 DDR4 DIMMs directly plugged into the main system board. The maximum number of DIMMs is allowed only in a two-socket system. Therefore, a one-socket system is limited to eight DIMMs.

The memory features equate to one DDR4 memory DIMM. The sizes and feature codes are described in Table 1-8.

Table 1-8 Memory features and description

Feature code	Description
EKM0	4 GB DDR4 Memory DIMM
EKM1	8 GB DDR4 Memory DIMM
EKM2	16 GB DDR4 Memory DIMM
EKM3	32 GB DDR4 Memory DIMM

The maximum supported memory in a two-socket system is 512 GB by installing a quantity of 16 feature #EKM3, and the maximum supported memory in a one-socket system is 256 GB by installing a quantity of eight feature #EKM3s.

Memory placement rules

For the Power S821LC server, the following rules apply to memory:

- A minimum of eight DIMMs is required (both a one-socket and s two-socket server).
- A maximum of 16 DIMMs are allowed.
- Memory must be plugged in quads and populating in quads maximizes memory bandwidth.
- Memory features cannot be mixed.
- Memory upgrades must be of the same capacity as the initial memory.
- Valid quantities for memory features in a one-socket system are 8.
- Valid quantities for memory features in a two-socket system are 8, 12, or 16.

If memory upgrades must be of the same capacity as the initial memory, account for any plans for future memory upgrades when you decide the number of processors and which memory feature size to use at the time of the initial system order.

Table 1-9 shows the number of features codes that are needed for each possible memory capacity.

Table 1-9 Number of memory feature codes that are required to achieve memory capacity

	Total installed memory								
Memory feature	**32 GB**	**48 GB**	**64 GB**	**96 GB**	**128 GB**	**192 GB**	**256 GB**	**384 GB**	**512 GB**
4 GB (#EKM0)	8	12[a]	16[a]						
8 GB (#EKM1)			8	12[a]	16[a]				
16 GB (#EKM2)					8	12[a]	16[a]		
32 GB (#EKM3)							8	12[a]	16[a]

a. Quantities are only available on two-socket systems.

The required approach is to install memory evenly across all processors in the system. Balancing memory across the installed processors allows memory access in a consistent manner and typically results in the best possible performance for your configuration.

One-socket systems always have all memory slots (P1) fully populated.

Two-socket systems have a minimum of eight DIMMs and maximum of 16 DIMMs per system.

Figure 1-12 shows the memory plugging order for two-socket systems.

To ease the memory installation, A and B slots are indicated on the system planar by black DDR4 DIMM connectors and C and D slots are blue DDR4 DIMM connectors.

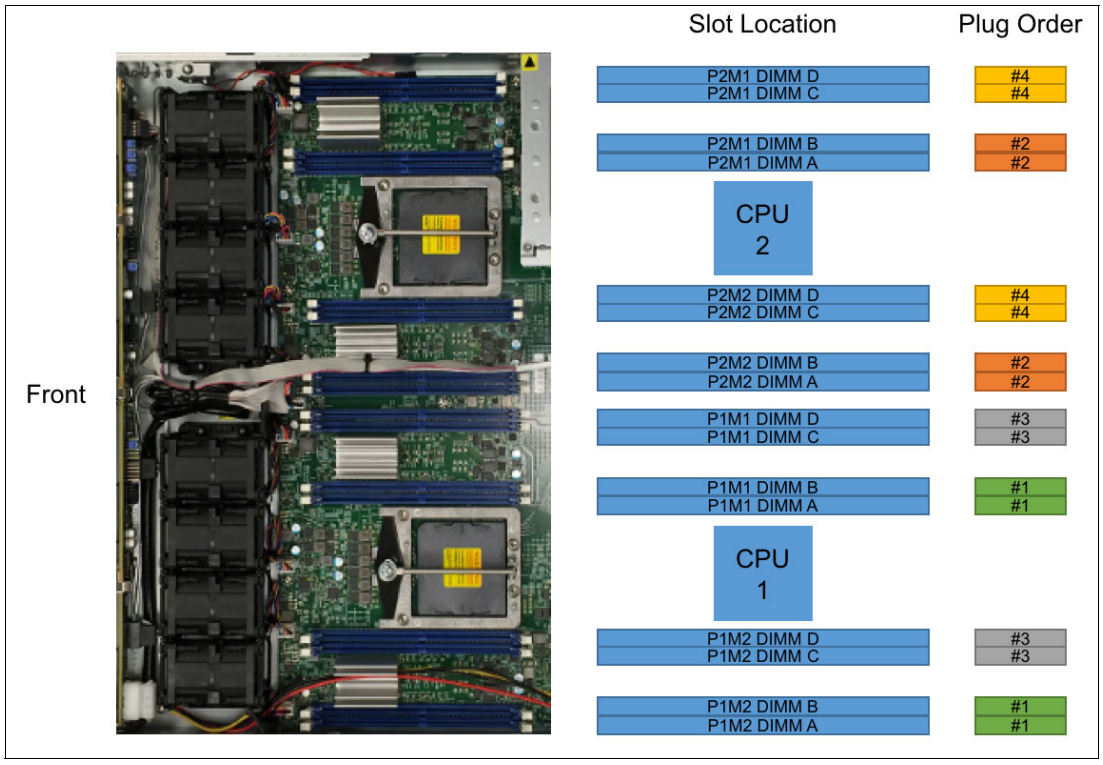


Figure 1-12 Memory slot numbering and plugging order for two-socket systems

The DIMM plug sequence for two-socket systems is also shown in Table 1-10.

Table 1-10 Memory slot placement rules

Slot location	Slot	DIMM quantity	Plug sequence	Notes
P1M2	A and B	2	1	Minimum required memory in a two-socket system
P1M1	A and B	2	1	
P2M2	A and B	2	2	
P2M1	A and B	2	2	
P1M2	C and D	2	3	Memory bandwidth maximized
P1M1	C and D	2	3	
P2M2	C and D	2	4	
P2M1	C and D	2	4	

Memory buffer chips

Memory buffer chips can connect to up to four industry-standard DRAM memory DIMMs and include a set of components that allow for higher bandwidth and lower latency communications:

- Memory Scheduler
- Memory Management (RAS Decisions & Energy Management)
- Buffer Cache

By adopting this architecture, several decisions and processes regarding memory optimizations are run outside the processor, saving bandwidth and allowing for faster processor to memory communications. It also allows for more robust reliability, availability, and serviceability (RAS). For more information about RAS features, see Chapter 3, “Reliability, availability, and serviceability” on page 39.

A detailed diagram of the memory buffer chip that is available for the Power S821LC server and its location on the server are shown in Figure 1-13.

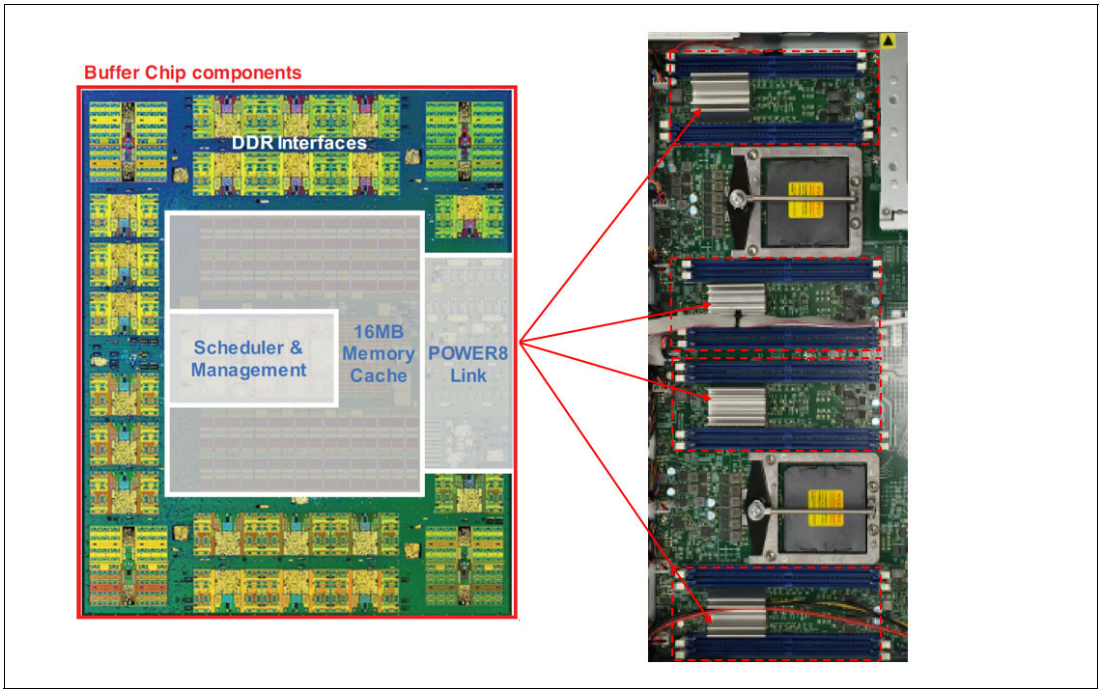


Figure 1-13 Detail of the memory buffer chip and location on the system board

The buffer cache is a L4 cache and is built on eDRAM technology (same as the L3 cache), which has lower latency than regular SRAM. Each buffer chip on the system board has 16 MB of L4 cache, and a fully populated Power S821LC server has 64 MB of L4 cache. The L4 cache performs several functions that have a direct impact on performance and provides a series of benefits for the Power S821LC server:

- Reduces energy consumption by reducing the number of memory requests.
- Increases memory write performance by acting as a cache and by grouping several random writes into larger transactions.
- Partial write operations that target the same cache block are “gathered” within the L4 cache before they are written to memory, becoming a single write operation.
- Reduces latency on memory access. Memory access for cached blocks has up to 55% lower latency than non-cached blocks.

Memory bandwidth

The POWER8 processor has exceptional cache, memory, and interconnect bandwidths. Table 1-11 shows the maximum bandwidth estimates for a single core on the Power S821LC server.

Table 1-11 Power S821LC server single-core bandwidth estimates

Single core	Power S821LC server (8001-12C)	
	8-core 2.328 GHz processor	**10-core 2.095 GHz processor**
L1 (data) cache	111.74 GBps	100.56 GBps
L2 cache	111.74 GBps	100.56 GBps
L3 cache	148.99 GBps	134.08 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:
  - 2.095 GHz Core: (2 * 16 B + 1 * 16 B) * 2.095 GHz = 100.56 GBps
  - 2.328 GHz Core: (2 * 16 B + 1 * 16 B) * 2.328 GHz = 111.74 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:
  - 2.095 GHz Core: (1 * 32 B + 1 * 16 B) * 2.095 GHz = 100.56 GBps
  - 2.328 GHz Core: (1 * 32 B + 1 * 16 B) * 2.328 GHz = 111.74 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:
  - 2.095 GHz Core: (1 * 32 B + 1 * 32 B) * 2.095 GHz = 134.08 GBps
  - 2.328 GHz Core: (1 * 32 B + 1 * 32 B) * 2.328 GHz = 148.99 GBps
- Total memory bandwidth: Each POWER8 processor has two memory channels running at 8.0 GBps and is capable of writing 2 bytes and reading 1 byte at a time. The bandwidth formula is calculated as follows:
  - 2 channels per CPU * 1 CPU per server * 8.0 GBps * 3 bytes = 48.0 GBps per one-socket server
  - 2 channels per CPU * 2 CPU per server * 8.0 GBps * 3 bytes = 96.0 GBps per two-socket server

On a system level basis, for both one-socket and two-socket Power S821LC servers that are configured with either 8 or 10 core processors, overall memory bandwidths are shown in Table 1-12.

Table 1-12 Power S821LC server: total bandwidth estimation

Total bandwidth	Power S821LC server (8001-12C)			
	8-cores @ 2.328 GHz	**10-cores @ 2.095 GHz**	**16-cores @ 2.328 GHz**	**20-cores @ 2.095 GHz**
L1 (data) cache	893.92 GBps	1,005.6 GBps	1,781 GBps	2,011.2 GBps
L2 cache	893.92 GBps	1,005.6 GBps	1,781 GBps	2,011.2 GBps
L3 cache	1,191.9 GBps	1,340.8 GBps	2,383.8 GBps	2,681.6 GBps
Total memory	48.0 GBps	48.0 GBps	96.0 GBps	96.0 GBps

1.7.3 Internal storage subsystems

The Power S821LC server supports several internal storage subsystems, including SATA and SAS HDDs, SATA SSDs, SATA Disk on Modules (DOM) and NVMe. Depending on the selection, a different PCIe storage controller might be necessary.

Note: The default configuration contains one #EKAB storage controller. Other controllers are available to suit different configuration needs. For more information, see “Storage adapters” on page 28.

Additional information about the PCIe storage controllers can be found in “Storage adapters” on page 28. This section details drive features, plugging rules, and general data about the support of drive features.

Additional information about the SATA DOM can be found in “SATA Drive on Modules” on page 24.

System-level drive-slot numbering and rules

The general slot numbering is presented in Figure 1-14. There is one connector on the interior side of the system backplane and the numbers represent the logical mapping within the connector.

The Power S821LC server has a standard midplane that allows for SATA, SAS, and NVMe drives.

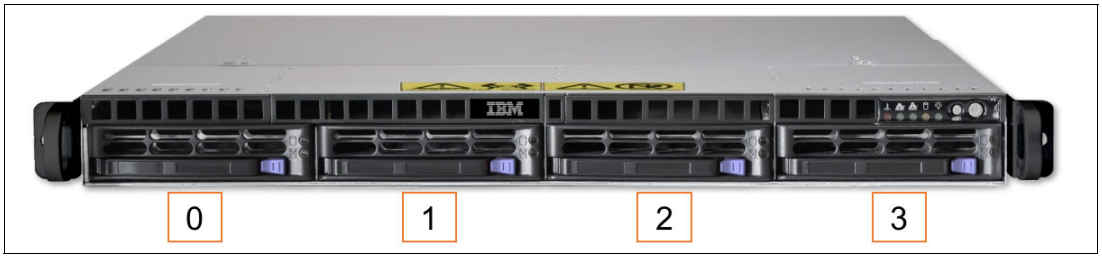


Figure 1-14 Drive slot mapping

Depending on the storage selection, different PCIe storage controllers might be necessary:

- Sata Only: The SATA controllers on the main planar can support a full population of four SATA drives in the front and two SATA DOM features plugged into the main planar. For more information about SATA DOM, see “SATA Drive on Modules” on page 24.
- SAS and SATA: For SAS (or SATA drives), the storage controller features available are #EKAA and #EKAB. Only one SAS/SATA storage adapter should ever be required.
- NVMe-Only: For NVMe, one feature #EKAE is required for every two NVMe Drives, up to four drives yielding the need for up to two #EKAE adapters.

Figure 1-15 summarizes the possibilities for internal storage and its storage controller requirements by using either the integrated controller or featured controllers.

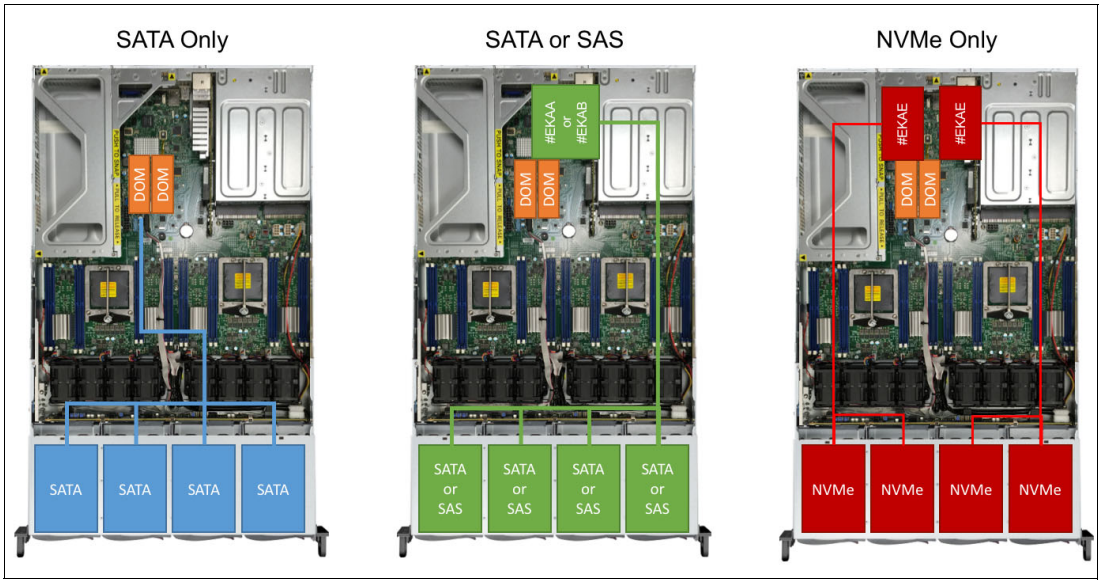


Figure 1-15 Storage options and storage controllers

The quantitative rules for plugging SATA, SAS, NVMe drives, and SATA DOM are presented in Table 1-13.

Table 1-13 Quantity of SATA DOM features

Quantity of SATA DOM features	Without storage adapter #EKEA or #EKEB		With storage adapter #EKAA or #EKAB	
	SATA	SAS	SATA	SAS
0	4	0	4	4
1	4	0	4	4
2	4	0	4	4

Additional drive restrictions:

- RAID is limited to 0, 1, and 10 for drives that are supported by the main planar SATA controllers. Additional RAID options are enabled by storage controllers.
- NVMe devices are not hot pluggable. All other drives are hot pluggable.
- NVMe devices are not bootable, so use SATA DOM for the internal boot.

SATA Drive on Modules

SATA DOM is a flash drive with a SATA interface, intended to be plugged directly into the motherboard and used as a computer HDD. The parts look like a USB thumb drive, but have a male SATA connector instead of a male USB connector and plug directly into the main planar.

By using less power, SATA DOM allows for an easy and more efficient way of storing the boot data internally. It is preferable for systems that have access to external storage or do not need much internal storage for their workload.

After they are plugged directly on the motherboard, SATA DOMs are not hot-pluggable.

A diagram of a SATA DOM drive and its location on the system planar is shown in Figure 1-16.

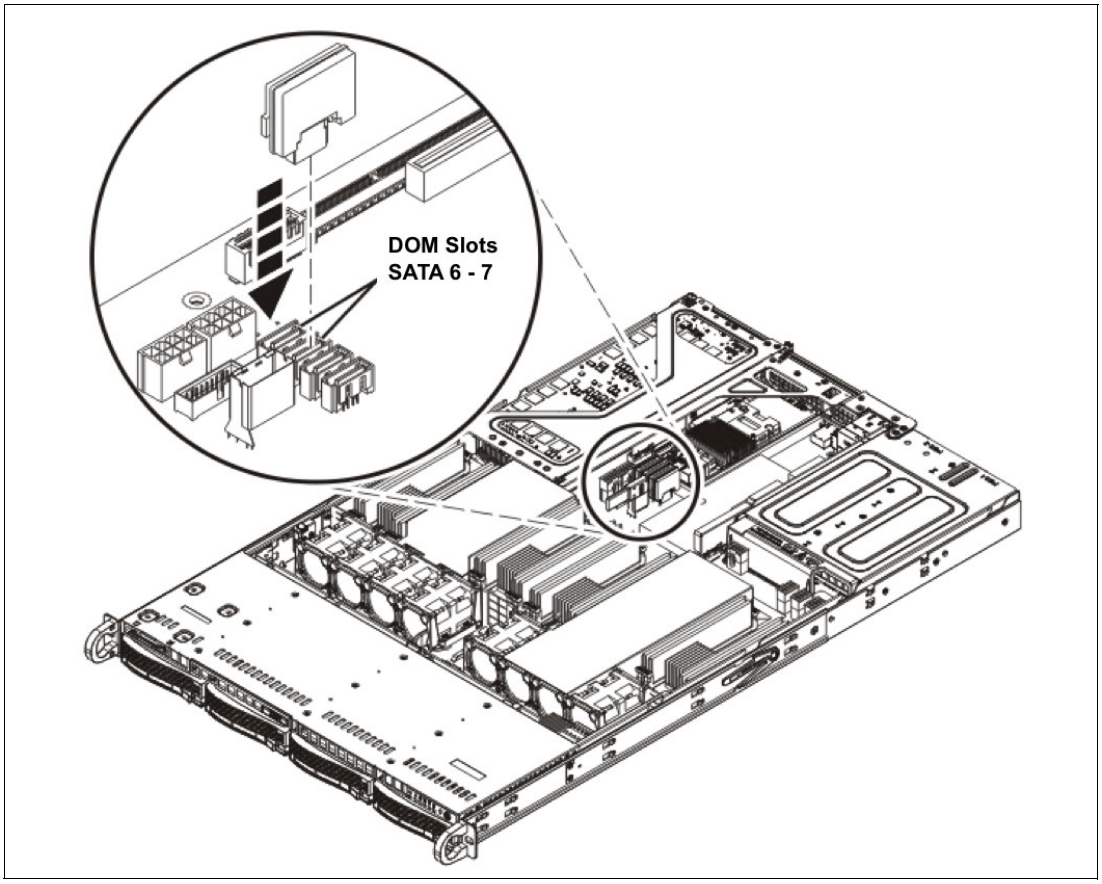


Figure 1-16 SATA DOM and its location on the system planar

Drive features and descriptions

Table 1-14 lists all the available drive features for the Power S821LC server.

Table 1-14 Available drives for the Power S821LC server

Disk type	Feature code	Description
SATA HDD	EKDA	2 TB 3.5" SATA HDD
	EKDB	4 TB 3.5" SATA HDD
	EKDC	6 TB 3.5" SATA HDD
	EKDD	8 TB 3.5" SATA HDD

Disk type	Feature code	Description
SAS HDD	EKD1	2 TB 3.5" SAS HDD
	EKD2	4 TB 3.5" SAS HDD
	EKD3	6 TB 3.5" SAS HDD
	EKD4	8 TB 3.5" SAS HDD
	EKD5	4 TB 3.5" Self Encrypting SAS HDD
	EKD6	8 TB 3.5" Self Encrypting SAS HDD
SATA DOM	EKSK	128 GB SATA Disk on Module SuperDOM
	EKSL	64 GB SATA Disk on Module SuperDOM
SATA SSD	EKS1	240 GB, SFF SATA SSD; 1.2 Disk Writes Per Day (DWPD) Kit
	EKS2	160 GB, SFF SATA SSD; 0.3 DWPD Kit
	EKS3	960 GB, SFF SATA SSD; 0.6 DWPD Kit
	EKS5	1.9 TB, SFF SATA SSD; 1.2 DWPD Kit
	EKS4	3.8 TB, SFF SATA SSD; 1.2 DWPD Kit
	EKS7	1.9 TB, SAS 2.5-inch SSD 1-DWPD
	EKS6	3.8 TB, SAS 2.5-inch SSD 1-DWPD
SDWPD NVMe	EKNA	800 GB, SFF NVMe; 3 DWPD Kit
	EKNB	1.2 TB, SFF NVMe; 3 DWPD Kit
	EKNC	1.6 TB, SFF NVMe; 3 DWPD Kit
	EKND	2.0 TB, SFF NVMe; 3 DWPD Kit
SDWPD NVMe	EKNJ	800 GB, SFF NVMe; 5 DWPD Kit
	EKNN	3.2 TB, SFF NVMe; 5 DWPD Kit

Note: SSDs and NVMe drives include an SFF to LFF converter tray.

1.7.4 PCI adapters

For a listing of PCIe slots and type, and a graphical rear view of the system and table with slot capability, see Figure 1-2 on page 3.

This section provides an overview of PCI Express and bus speed and feature listings, segregated by function, for the supported PCIe adapters in the Power S821LC server.

PCI Express

PCIe uses a serial interface and allows for 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 PCIe interfaces that are supported on this server are PCIe Gen3, which are capable of 16 GBps simplex (32 GBps duplex) on a single x16 interface. PCIe Gen3 slots also support previous generation (Gen2 and Gen1) 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.

POWER8 based servers can support two different form factors of PCIe adapters:

- PCIe low profile (LP) cards
- PCIe full height and full high cards

Before adding or rearranging adapters, use the System Planning Tool to validate the new adapter configuration. For more information, see the System Planning Tool website:

http://www.ibm.com/systems/support/tools/systemplanningtool/

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 any existing update prerequisites to install. To obtain this information, use the IBM prerequisite website:

https://www-912.ibm.com/e_dir/eServerPreReq.nsf

Each POWER8 processor has 32 PCIe lanes running at 8 Gbps full-duplex. The bandwidth formula is calculated as follows:

Thirty-two lanes * 2 processors * 8 Gbps * 2 = 128 GBps

As seen in the PCIe bus to CPU mapping in Figure 1-5 on page 5, the 32 lanes feed various PCIe slots and adapter slots. In general, PCIe lanes coming direct from the processor and not through a switch are CAPI-enabled.

Storage adapters

Storage adapters are required to enable full function of the drive features in the front of the system. The first drive adapters features support the SAS and SATA protocols (#EKAA, #EKAB, and EKAD). Feature #EKAA includes a battery back-up for cache protection and feature #EKAE is the NVMe host bus adapter that is required to support NVMe drives. One of these adapters is required for every two NVMe devices, up to the system limit of four.

All of the storage adapters are kitted with system-specific internal cables to optimize serviceability. Available adapters are provided in Table 1-15.

Table 1-15 Storage adapters

Feature code	Type	Description
EKAA	x8	PCIe3 SAS RAID Controller w/cable for 1U server, based on LSI MegaRAID 9361-8i
EKAB	x8	PCIe3 SAS RAID Controller w/cable for 1U server, based on LSI 3008L
EKAD	x8	Storage adapter - SAS-3, 3008 Chipset, 8 Ports, external for 1U server
EKAE	x8	PCIe3 2-port NVMe Adapter w/cable for 1U server, based on PLX PEX8718
EKAG	x8	4-port NVMe HBA w/cable for 1U server, based on PLX PEX9733
EKAH	x8	PCIe3 SAS RAID Controller w/cable for 1U server, based on LSI MegaRAID 9361-8i 2 GB
EKN2	x8	PCIe3 1.6 TB NVMe Flash Adapter 3-DWPD

LAN adapters

To connect the Power S821LC server to a local area network (LAN), you can use the LAN adapters that are supported in the PCIe slots of the system, as listed in Table 1-16, in addition to the standard 4-port 10 Gb BaseT Ethernet that is present in every system.

Table 1-16 LAN adapter features and descriptions

Feature code	Type	Description
EKA0	x8	PCIe3 2-port 10 GbE BaseT RJ45 Adapter, based on Intel X550-A
EKA1	x8	PCIe3 4-port 10 GbE SFP+ Adapter, based on Broadcom BCM57840
EKA2	x8	PCIe3 2-port 10 GbE SFP+ Adapter, based on Intel XL710
EKA3	x4	PCIe2 2-port 1 GbE Adapter, based on Intel 82575EB
EKAL	x16	PCIe3 1-port 100 GbE QSFP28 x16, based on Mellanox ConnectX-4
EKAM	x16	PCIe3 2-port 100 GbE QSFP28 x16, based on Mellanox ConnectX-4
EKAU	x16	PCIe3 2-port 10/25 GbE (NIC&RoCE) Adapter, based on Mellanox ConnectxX-4 Lx

Fibre Channel adapters

The servers support direct or SAN connection to devices that use Fibre Channel adapters. Table 1-17 summarizes the available Fibre Channel adapters (all have LC connectors). The infrastructure that is used with these adapters determines the need to procure LC fiber converter cables.

Table 1-17 Fibre Channel adapter features and descriptions

Feature code	Type	Description
EKAP	x8	PCIe 2-port 8 Gb Fibre Channel, based on QLogic QLE2562
EKAQ	x8	PCIe 2-port 16 Gb Fibre Channel, based on QLogic QLE2692SR
EKAF	x8	PCIe 2-port 16 Gb Fibre Channel, based on Emulex LPE16002B-M6

CAPI adapters

The CAPI Field Programmable Gate Array (FPGA) adapter in Table 1-18 acts as a co-processor for the POWER8 processor chip by handling specialized, repetitive functions extremely efficiently.

Table 1-18 CAPI adapter features and descriptions

Feature code	Type	Description
EKAT	x8	PCIe3 CAPI adapter, Alpha-Data ADM-PCIE-KU

Compute-intensive accelerator adapters

Compute-intensive accelerators are GPUs that are developed by NVIDIA. They are shown in Table 1-19. With NVIDIA GPUs, the Power S821LC server can offload processor-intensive operations to a GPU accelerator and boost performance. This adapter requires two x16 slots.

Note: If GPUs are configured, 200-240 V AC power supplies must be used.

Table 1-19 Compute intensive accelerator adapter features and description

Feature code	Type	Description
EKAJ	x16	NVIDIA Tesla K80 24 GB GPU Accelerator
EKAZ	x16	NVIDIA Tesla P100 16 GB GPU Accelerator

NVIDIA Tesla GPUs are massively parallel accelerators that are based on the NVIDIA Compute Unified Device Architecture (CUDA) parallel computing platform and programming model. Tesla GPUs are designed from the ground up for power-efficient, high-performance computing, computational science, supercomputing, big data analytics, and machine learning applications, delivering dramatically higher acceleration than a CPU-only approach.

These NVIDIA Tesla GPU Accelerators are based on the NVIDIA Kepler Architecture and designed to run the most demanding scientific models faster and more efficiently. Table 1-20 shows a summary of NVIDIA Tesla characteristics.

Table 1-20 NVIDIA Tesla specifications

Features	Tesla K80	Tesla P100
Number and type of GPUs	2 GK210 GPUs	1 GP100 GPU
Peak double precision floating point performance	1.87 Tflops	4.7 Tflops
Peak single precision floating point performance	5.60 Tflops	9.3 Tflops
Memory bandwidth (error correction code, off)	480 GBps	732 GBps
Memory size	24 GB (GDDR5)	16 GB (HBM2)
CUDA cores	4,992	3,584

For more information about the NVIDIA Tesla GPU, see the NVIDIA Tesla data sheet, found at:

http://www.nvidia.com/object/tesla-servers.html

NVIDIA CUDA is a parallel computing platform and programming model that enables dramatic increases in computing performance by harnessing the power of the GPU. Today, the CUDA infrastructure is growing rapidly as more companies provide world-class tools, services, and solutions. If you want to start harnessing the performance of GPUs, the CUDA Toolkit provides a comprehensive development environment for C and C++ developers.

The easiest way to start is to use the plug-in scientific and math libraries that are available in the CUDA Toolkit to quickly accelerate common linear algebra, signal and image processing, and other common operations, such as random number generation and sorting. If you want to write your own code, the Toolkit includes a compiler, and debugging and profiling tools. You also find code samples, programming guides, user manuals, API references, and other documentation to help you get started.

The CUDA Toolkit is available at no charge. Learning to use CUDA is convenient, with comprehensive online training available, and other resources, such as webinars and books. Over 400 universities and colleges teach CUDA programming, including dozens of CUDA Centers of Excellence and CUDA Research and Training Centers. Solutions for Fortran, C#, Python, and other languages are available.

To learn more, explore the GPU Computing Ecosystem on CUDA Zone at the following website:

https://developer.nvidia.com/cuda-tools-ecosystem

The production release of CUDA V7.5 for POWER8 (and any subsequent release) is available for download at the following website:

https://developer.nvidia.com/cuda-downloads

1.8 Operating system

The Power S821LC 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, see the Linux on Power Systems website:

http://www.ibm.com/systems/power/software/linux/index.html

1.8.1 Ubuntu

Ubuntu Server 14.04.5 LTS and Ubuntu Server 16.04.1 LTS for IBM POWER8 are supported on the Power S821LC server.

For more information about Ubuntu Server for Ubuntu for POWER8, see the following website:

http://www.ubuntu.com/download/server/power8

1.8.2 Red Hat Enterprise Linux

Red Hat Enterprise Linux (ppc64le) Version 7.2 is supported on the Power S821LC server.

For additional questions about this release and supported Power Systems servers, consult the Red Hat Hardware Catalog, found at the following website:

https://hardware.redhat.com

1.9 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. This system connects to storage by using an options Fibre Channel adapter (#EKAP or #EKAQ).

For more information about the various offerings, see the following website:

http://www.ibm.com/systems/storage/disk

The following sections highlight a few of the offerings.

1.9.1 IBM Storwize family

The IBM Storwize® family is part of the IBM Spectrum™ Virtualize offering and 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 (RtC), Easy Tier®, and many more functions. For more information, see the following website:

http://www.ibm.com/systems/storage/storwize

1.9.2 IBM FlashSystem family

The IBM FlashSystem® family delivers extreme 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 the following website:

http://www.ibm.com/systems/storage/flash

1.9.3 IBM XIV Storage System

The IBM XIV® Storage System is part of IBM Spectrum Accelerate™ and is a high-end disk storage system, helping 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 the following website:

http://www.ibm.com/systems/storage/disk/xiv/index.html

1.9.4 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 (QoS), and also provide the highest levels of reliability and system uptime. For more information, see the following website:

http://www.ibm.com/systems/storage/disk/ds8000/index.html

1.10 Java

When running Java applications on the POWER8 processor, the pre-packaged 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 available.

Current information about IBM Java and tested Linux distributions are available at the following website:

https://www.ibm.com/developerworks/java/jdk/linux/tested.html

Additional information about the OpenJDK port for Linux on PPC64 LE and some pre-generated builds can be found at the following website:

http://cr.openjdk.java.net/~simonis/ppc-aix-port/

Launchpad.net has resources for Ubuntu builds. You can discover more about them at the following websites:

https://launchpad.net/ubuntu/+source/openjdk-9
https://launchpad.net/ubuntu/+source/openjdk-8
https://launchpad.net/ubuntu/+source/openjdk-7

2

Management and virtualization

As you look for ways to maximize the return on your IT infrastructure investments, virtualizing workloads becomes an attractive proposition.

The IBM Power System S821LC server is excellent for clients that want the advantages of running their big data, Java, open source, and industry applications on a platform designed and optimized for data and Linux.

This chapter attempts to identify and clarify the tools that are available for managing Linux on Power Systems servers.

2.1 Main management components overview

Figure 2-1 shows the logical management flow of a Linux on Power Systems server.

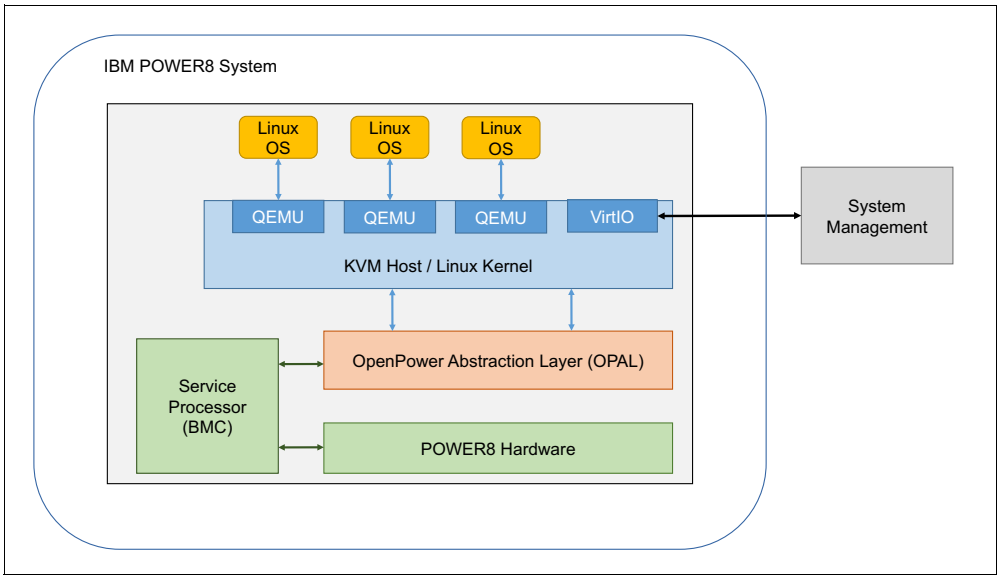


Figure 2-1 Logical diagram of a Linux on Power Systems server

The service processor, or baseboard management controller (BMC), uses KVM for virtual machines (VMs) and provides robust error detection and self-healing functions that are built into the IBM POWER8 processor and memory buffer modules.

Open Power Abstraction Layer (OPAL) is the system firmware in the stack of POWER8 processor-based Linux on Power Systems servers.

The KVM virtualization technology offers key capabilities that can help you consolidate and simplify your IT environment. QEMU is a generic and open source machine emulator and virtualizer that hosts the VMs on a KVM hypervisor. It is the software that manages and monitors the VMs.

KVM hosts can be managed by open source Linux tools that use the libvirt API, such as the Kimchi point-to-point administration tool and IBM Cloud PowerVC Manager.

IBM Cloud PowerVC Manager delivers easy-to-use advanced virtualization management capabilities that are virtualized by KVM. IBM Cloud PowerVC Manager manages KVM VMs within a resource pool and enables the capture, deployment, and inventory of VM images.

2.2 Service processor

The service processor, or BMC, is the primary control for autonomous sensor monitoring and event logging features on the Power S821LC server.

BMC supports the Intelligent Platform Management Interface (IPMI V2.0) and Data Center Management Interface (DCMI V1.5) for system monitoring and management.

BMC monitors the operation of the firmware during the boot process and also monitors the hypervisor for termination. The firmware code update is supported through the BMC and IPMI interfaces.

2.2.1 Open Power Abstraction Layer

On Linux on Power systems, the OPAL firmware provides a hypervisor interface to the underlying hardware. The OPAL firmware allows Linux hosts to use the VirtIO API. The VirtIO API specifies an independent interface between VMs and the service processor.

The VirtIO API is a high-performance API that para-virtualized devices use to gain speed and efficiency. VirtIO para-virtualized devices are especially useful for guest operating systems (OSs) that run I/O heavy tasks and applications.

For the Power S821LC server, OPAL bare metal is the only available option for the system firmware in the stack of POWER8 processor-based servers.

2.2.2 Intelligent Platform Management Interface

The IPMI is an open standard for monitoring, logging, recovery, inventory, and control of hardware that is implemented independent of the main CPU, BIOS, and OS. It is the default console to use when you configure a KVM host. The Power S821LC server provides one 10M/100M baseT IPMI port.

The *ipmitool* is a utility for managing and configuring devices that support IPMI. It provides a simple command-line interface (CLI) to the service processor. You can install the ipmitool from the Linux distribution packages in your workstation or another server (preferably on the same network as the installed server). For example, in Ubuntu, run the following command:

```
$ sudo apt-get install ipmitool
```

To connect to your system with IPMI, you must know the IP address of the server and have a valid password. To power on the server with ipmitool, complete the following steps:

1. Open a terminal program.
2. Power on your server by running the following command:

   ```
   ipmitool -I lanplus -H fsp_ip_address -P ipmi_password power on
   ```

3. Activate your IPMI console by running the following command:

   ```
   ipmitool -I lanplus -H fsp_ip_address -P ipmi_password sol activate
   ```

For more help with configuring Linux on a Power Systems server, see the following website:

https://www.ibm.com/support/knowledgecenter/linuxonibm/liabp/liabpusingipmi.htm

2.2.3 Petitboot bootloader

Petitboot is a kexec-based bootloader that is used by POWER8 processor-based systems that are configured with Linux.

After the POWER8 processor-based system powers on, the petitboot bootloader scans local boot devices and network interfaces to find boot options that are available to the system. Petitboot returns a list of boot options that are available to the system.

If you are using a static IP or if you did not provide boot arguments in your network boot server, you must provide the details to petitboot. You can configure petitboot to find your boot server by following the instructions that are found at the following website:

https://www.ibm.com/support/knowledgecenter/linuxonibm/liabp/liabppetitbootadvanced.htm

You can edit petitboot configuration options, change the amount of time before petitboot automatically boots, and so on, by following the instructions found at the following website:

https://www.ibm.com/support/knowledgecenter/linuxonibm/liabp/liabppetitbootconfig.htm

After you select to start the Linux installer, the installer wizard walks you through the steps to set up disk options, your root password, time zones, and so on.

You can read more about the petitboot bootloader program at the following website:

https://www.kernel.org/pub/linux/kernel/people/geoff/petitboot/petitboot.html

2.3 IBM Cloud PowerVC Manager

IBM Cloud PowerVC Manager (5765-VCS) is an advanced enterprise virtualization management offering for Power Systems based on the OpenStack technology. OpenStack is an open source software that controls large pools of server, storage, and networking resources throughout a data center. The latest IBM Cloud PowerVC Manager Version 1.3.1 was announced in April 2016 and is built on OpenStack (liberty). This comprehensive virtualization management offering enables VM setup and management.

2.3.1 Benefits

IBM Cloud PowerVC Manager includes the following features and benefits:

- Self-service portal to allow provisioning of new VMs in a PowerVM-based private cloud with an approval process
- Cloud management policies, including approvals and expirations
- Deploy templates for cloud deployments
- Metering data that can be used for chargeback
- Enhanced security isolation by project with integrated, role-based security controls

- An improvement to the Dynamic Resource Optimizer that allows schedules for monitoring policies, NovaLink-based mobile core CoD optimization, and improved Dynamic Resource Optimizer policy controls
- Enhanced storage support, which includes the following items:
  - Standard Edition storage management enhancements that include support for Hitachi Data Systems Unified Storage Disk arrays and virtual Storage Platform storage arrays for NPIV storage
  - Improved data deletion policies
- VM deployment with IBM Active Memory™ Expansion (AME) enabled Integrated management of storage, network, and compute resources

For more information about hardware and OS support for IBM Cloud PowerVC Manager hosts, see IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/SSXK2N_1.3.1/com.ibm.powervc.standard.help.doc/powervc_hwandsw_reqs_hmc.html

2.3.2 Lifecycle

With the introduction of IBM Cloud PowerVC Manager V1.3.0, the end of service date for PowerVC V1.2 for standard support is April 2017. For more information about the PowerVC lifecycle, see the following website:

http://www.ibm.com/systems/power/software/virtualization-management/lifecycle.html

3

Reliability, availability, and serviceability

This chapter provides information about IBM Power Systems reliability, availability, and serviceability (RAS) design and features.

The elements of RAS can be described as follows:

Reliability	Indicates how infrequently a defect or fault in a server occurs.
Availability	Indicates how infrequently the functioning of a system or application is impacted by a fault or defect.
Serviceability	Indicates how well faults and their effects are communicated to system managers and how efficiently and nondisruptively the faults are repaired.

3.1 Introduction

The IBM Power System S821LC server is bringing POWER8 processor and memory RAS functions into a highly competitive cloud data center with open source Linux technology as an operating system (OS) and virtualization.

The Open Power Abstraction Layer (OPAL) firmware provides a hypervisor and OS independent layer that uses the robust error-detection and self-healing functions that are built into the POWER8 processor and memory buffer modules.

The processor address-paths and data-paths are protected with parity or error-correcting codes (ECCs); the control logic, state machines, and computational units have sophisticated error detection. The processor core soft errors or intermittent errors are recovered with processor instruction retry. Unrecoverable errors are reported as machine check (MC) errors. Errors that affect the integrity of data lead to system checkstop.

3.1.1 RAS enhancements of POWER8 processor-based scale-out servers

The Power S821LC server, in addition to being built on advanced RAS characteristics of the POWER8 processor, offer reliability and availability features that often are not seen in such scale-out servers.

Here is a brief summary of these features:

- Processor enhancements integration

  POWER8 processor chips are implemented by using 22 nm technology, and are integrated on SOI modules.

  The processor design supports a spare data lane on each fabric bus, which is used to communicate between processor modules. A spare data lane can be substituted for a failing one dynamically during system operation.

  A POWER8 processor module has improved performance, including support of a maximum of 12 cores because doing more work with less hardware in a system supports greater reliability. The Power S821LC server offers two processor socket offerings with 8-core and 10-core processor configurations. So, there are 16-core and 20-core configurations that are available.

  The On Chip Controller (OCC) monitors various temperature sensors in the processor module, memory modules, and environmental temperature sensors, and steers the throttling of processor cores and memory channels if the temperature rises over thresholds that are defined by the design. The power supplies have their own independent thermal sensors and monitoring.

  Power supplies and voltage regulator modules monitor Over-Voltage, Under-Voltage, and Over-Current conditions. They report to a “power good” tree that is monitored by the service processor.

- I/O subsystem

  The PCIe controllers are integrated into the POWER8 processor. All the PCIe slots are directly driven by the PCIe controllers.

- Memory subsystem

  The memory subsystem has proactive memory scrubbing to prevent accumulation of multiple single-bit errors. The ECC scheme can correct the complete failure of any one memory module within an ECC word. After marking the module as unusable, the ECC logic can still correct single-symbol (two adjacent bit) errors. An uncorrectable error (UE) of data of any layer of cache up to the main memory is marked to prevent usage of fault data. The processor's memory controller and the memory buffer have retry capabilities for certain fetch and store faults.

3.2 IBM terminology versus x86 terminology

The different components and descriptions in the boot process have similar functions, but have different terms for POWER8 processor-based and x86-based scale-out servers. Table 3-1 shows a quick overview of the terminology.

Table 3-1 IBM Terminology versus x86 terminology

IBM	x86	Description
SBE	Undisclosed	Self-Boot Engine: Starts the boot process.
Host Boot	BIOS	Core, Powerbus (SMP), and memory initialization.
OPAL	BIOS/ VT-d / UEFI	KVM hardware abstraction, PCIe RC, IODA2 (VT-d), and open firmware.
OCC	PCU, off chip microprocessors	Performs real-time functions, such as power management.
HBRT	N/A	Correctable error monitoring and OCC monitoring.

3.3 Error handling

This section describes how the Power S821LC server handles different errors and recovery functions. It provides some general information and helps you understand some techniques.

3.3.1 Processor core/cache correctable error handling

The OPAL firmware provides a hypervisor and OS-independent layer that uses the robust error-detection and self-healing functions that are built into the POWER8 processor and memory buffer modules.

The processor address-paths and data-paths are protected with parity or ECC. The control logic, state machines, and computational units have sophisticated error detection. The processor core soft errors or intermittent errors are recovered with processor instruction retry. Unrecoverable errors are reported as an MC. Errors that affect the integrity of data lead to system checkstop.

The Level 1 (L1) data and instruction caches in each processor core are parity-protected, and data is stored through to L2 immediately. L1 caches have a retry capability for intermittent errors and a cache set delete mechanism for handling solid failures.

The L2 and L3 caches in the POWER8 processor and L4 cache in the memory buffer chip are protected with double-bit detect, single-bit correct ECC.

Special Uncorrectable Error handling

Special Uncorrectable Error (SUE) handling prevents an UE in memory or cache from immediately causing an MC with an UE. The system marks the data such that if the data ever is read again, it generates an MC with an UE. Termination may be limited to the program / partition or hypervisor owning the data. If the data is referenced by an I/O adapter, it freezes if data is transferred to an I/O device.

3.3.2 Processor Instruction Retry and other try again techniques

Within the processor core, soft-error events might occur that interfere with the various computation units. When such an event can be detected before a failing instruction is completed, the processor hardware might try the operation again by using the advanced RAS feature that is known as *Processor Instruction Retry*.

Processor Instruction Retry allows the system to recover from soft faults that otherwise result in outages of applications or the entire server. Try-again techniques are used in other parts of the system as well. Faults that are detected on the memory bus that connects processor memory controllers to DIMMs can be tried again. In POWER8 processor-based systems, the memory controller is designed with a replay buffer that allows memory transactions to be tried again after certain faults internal to the memory controller faults are detected. This function complements the try-again abilities of the memory buffer module.

3.3.3 Other processor chip functions

Within a processor chip, there are other functions besides just processor cores.

POWER8 processors have built-in accelerators that can be used as application resources to handle such functions as random number generation. POWER8 also introduces a controller for attaching cache-coherent adapters that are external to the processor module. The POWER8 design contains a function to “freeze” the function that is associated with some of these elements without taking a system-wide checkstop. Depending on the code that uses these features, a “freeze” event might be handled without an application or partition outage.

As indicated elsewhere, single-bit errors, even solid faults, within internal or external processor *fabric buses*, are corrected by the ECC that is used. POWER8 processor-to-processor module fabric buses also use a spare data lane so that a single failure can be repaired without calling for the replacement of hardware.

3.4 Serviceability

The server is designed for system installation and setup, feature installation and removal, proactive maintenance, and corrective repair that is performed by the client:

- Customer Install and Setup (CSU)
- Customer Feature Install (CFI)
- Customer Repairable Units (CRU)

Warranty service upgrades are offered for an onsite repair (OSR) by an IBM System Services Representative (SSR), or an authorized warranty service provider.

3.4.1 Detection introduction

The first and most crucial component of a solid serviceability strategy is the ability to detect accurately and effectively errors when they occur.

Although not all errors are a guaranteed threat to system availability, those errors that go undetected can cause problems because the system has no opportunity to evaluate and act if necessary. POWER processor-based systems employ IBM z Systems™ server-inspired error detection mechanisms, extending from processor cores and memory to power supplies and hard disk drives (HDDs).

3.4.2 Error checkers and fault isolation registers

POWER processor-based systems contain specialized hardware detection circuitry that is used to detect erroneous hardware operations. Error-checking hardware ranges from parity error detection that is coupled with Processor Instruction Retry and bus try again, to ECC correction on caches and system buses.

Within the processor/memory subsystem error checker, error-checker signals are captured and stored in hardware fault isolation registers (FIRs). The associated logic circuitry is used to limit the domain of an error to the first checker that encounters the error. In this way, runtime error diagnostic tests can be deterministic so that for every check station, the unique error domain for that checker is defined and mapped to CRUs that can be repaired when necessary.

3.4.3 Service processor

The service processor supports the Intelligent Platform Management Interface (IPMI 2.0) and Data Center Management Interface (DCMI 1.5) for system monitoring and management. The service processor provides the following platform system functions:

- Power on/off
- Power sequencing
- Power fault monitoring
- Power reporting
- Fan/thermal control
- Fault monitoring
- VPD inventory collection
- Serial over LAN (SOL)
- Service Indicator LED management
- Code update
- Event reporting through System Event Logs (SELs)

All SELs can be retrieved either directly from the service processor or from the host OS (Linux). The service processor monitors the operation of the firmware during the boot process.

The firmware code update is supported through the service processor and IPMI interface. Multiple firmware images exist in the system and the backup copy is used if the primary image is corrupted and unusable.

3.4.4 Diagnosing

General diagnostic objectives are to detect and identify problems so that they can be resolved quickly.

Using the extensive network of advanced and complementary error detection logic that is built directly into hardware, firmware, and OSs, Power Systems servers can perform considerable self-diagnosis.

Host Boot IPL

In POWER8, the initialization process during IPL changed. The service processor is no longer the only instance that initializes and runs the boot process. With POWER8, the service processor initializes the boot processes, but on the POWER8 processor itself, one part of the firmware is running and performing the central electrical complex (CEC) chip initialization. A new component that is called the *PNOR chip* stores the Host Boot firmware. The SBE is an internal part of the POWER8 chip itself and is used to start the chip.

Device drivers

In certain cases, diagnostic tests are preferably performed by OS-specific drivers, most notably adapters or I/O devices that are owned directly by a logical partition. In these cases, the OS device driver often works with I/O device Licensed Internal Code to isolate and recover from problems. Potential problems are reported to an OS device driver, which logs the error.

3.4.5 General problem determination

Accessing the Advanced System Management GUI interface provides a general overview of sensor information and possible errors.

Using an event sensor display as a primary interface for problem determination

This function has the following aspects:

- Covers 90% of typical failures
- Does not handle transient failure scenarios

Using SEL logs or operating system syslog records for remainder

This function has the following aspects:

- Sensors can be enabled/disabled by a client.
- The “Get Sensor Event Enable” IPMI command is available.

SEL events: Platform-related events

The following platform-related events are available under the SEL events:

- SELs link to extended Service Event Logs (eSELs)
- eSEL represents a service action required event:
  - SELs linked to the eSEL represent “service action required” and a part to be replaced.
  - You may have multiple SELs that are linked to the eSEL.
  - SELs not linked to eSEL may not represent a service action required event.
  - Without an eSEL Event, the System Attention LED does not turn on.

For an SEL event that is associated with an eSEL event, see Example 3-1. In this case, events 63 and 64 are the SEL events and event 62 is the associated eSEL event.

Example 3-1 SEL and eSEL events

```
60 | 09/04/2015 | 15:12:27 | Power Supply #0xcd | Presence detected | Asserted
61 | 09/04/2015 | 15:12:27 | Power Supply #0xce | Presence detected | Asserted
62 | 09/04/2015 | 15:12:35 | OEM record df | 040020 | 0c2207aaaaaa
63 | 09/04/2015 | 15:12:35 | Memory #0x22 | Transition to Non-recoverable | Asserted
64 | 09/04/2015 | 15:12:36 | Memory #0x23 | Transition to Non-recoverable | Asserted
65 | 09/04/2015 | 15:12:54 | System Firmware Progress #0x05 | Memory initialization | Asserted
```

OEM vendor SELs: Platform-related events

The following platform-related events are available under the OEM vendor SELs:

- SELs are developed to provide specific OEM information in the error record.
- Not interpretable by IPMI.
- No corresponding IPMI SEL events.

Generic system event SELs

Here are some of the generic system event SELs:

- Firmware
- Isolates and symbolics as highest priority FRUs

Syslog events: OS-detected events

PCI adapters and devices are OS-detected events.

3.4.6 Error handling and reporting

If there is a system hardware or environmentally induced failure, the system error capture capability systematically analyzes the hardware error signature to determine the cause of failure.

The CEC recoverable errors are handled through the CEC diagnostic capability in a Linux application and generate a SEL. There is also an eSEL that contains extra First Failure Data Capture (FFDC) from the Host Boot, OCC, and OPAL subsystems that are associated with each SEL. For system checkstop errors, OCC collects FIR data for PNOR, and Host Boot CEC diagnostic tests creates a SEL based on the FIR data in PNOR.

When the system can be successfully restarted either manually or automatically, or if the system continues to operate, the host Linux OS can monitor the SELs on the service processor through the IPMI tool. Hardware and software failures are recorded in the SELs and can be retrieved through the IPMI interface. There is a plan to report SELs in the system log of the OS.

The system can report errors that are associated with PCIe adapters/devices.

For some example SEL events, see Example 3-2.

Example 3-2 Example of SEL events

```
31 | 09/04/2015 | 15:11:40 | Power Unit #0x1c | Power off/down | Asserted
32 | 09/04/2015 | 15:11:40 | Power Supply #0xcd | Presence detected | Deasserted
33 | 09/04/2015 | 15:11:40 | Power Supply #0xce | Presence detected | Deasserted
34 | 09/04/2015 | 15:11:43 | Power Supply #0xcd | Presence detected | Asserted
35 | 09/04/2015 | 15:11:43 | Power Supply #0xce | Presence detected | Asserted
36 | 09/04/2015 | 15:11:47 | System Firmware Progress #0x05 | Motherboard initialization | Asserted
```

```
37 | 09/04/2015 | 15:12:11 | Fan #0xd4 | Upper Non-critical going high | Asserted
38 | 09/04/2015 | 15:12:11 | Fan #0xd4 | Upper Critical going high | Asserted
39 | 09/04/2015 | 15:12:11 | Fan #0xd4 | Upper Non-recoverable going high | Asserted
3a | 09/04/2015 | 15:12:12 | Fan #0xd5 | Upper Non-critical going high | Asserted
3b | 09/04/2015 | 15:12:12 | Fan #0xd5 | Upper Critical going high | Asserted
3c | 09/04/2015 | 15:12:12 | Fan #0xd5 | Upper Non-recoverable going high | Asserted
3d | 09/04/2015 | 15:12:12 | Fan #0xd6 | Upper Non-critical going high | Asserted
3e | 09/04/2015 | 15:12:13 | Fan #0xd6 | Upper Critical going high | Asserted
3f | 09/04/2015 | 15:12:13 | Fan #0xd6 | Upper Non-recoverable going high | Asserted
40 | 09/04/2015 | 15:12:13 | Fan #0xd7 | Upper Non-critical going high | Asserted
41 | 09/04/2015 | 15:12:13 | Fan #0xd7 | Upper Critical going high | Asserted
42 | 09/04/2015 | 15:12:13 | Fan #0xd7 | Upper Non-recoverable going high | Asserted
43 | 09/04/2015 | 15:12:13 | Fan #0xd4 | Upper Non-recoverable going high | Deasserted
44 | 09/04/2015 | 15:12:13 | Fan #0xd4 | Upper Critical going high | Deasserted
45 | 09/04/2015 | 15:12:13 | Fan #0xd4 | Upper Non-critical going high | Deasserted
46 | 09/04/2015 | 15:12:13 | Fan #0xd5 | Upper Non-recoverable going high | Deasserted
47 | 09/04/2015 | 15:12:13 | Fan #0xd5 | Upper Critical going high | Deasserted
48 | 09/04/2015 | 15:12:14 | Fan #0xd5 | Upper Non-critical going high | Deasserted
```

To service a Linux system end to end, Linux service and productivity tools must be installed. You can find them at the following website:

http://www.ibm.com/support/customercare/sas/f/lopdiags/home.html

The tools are automatically loaded if IBM manufacturing installs the Linux image or IBM Installation Toolkit. PowerPack is the preferred way to install required service packages from the website. The Linux call home feature is also supported in a stand-alone system configuration to report serviceable events.

3.4.7 Locating and servicing

The final component of a comprehensive design for serviceability is the ability to locate and replace effectively parts requiring service. POWER processor-based systems use a combination of visual cues and guided maintenance procedures to ensure that the identified part is replaced correctly every time.

Packaging for service

The following service enhancements are included in the physical packaging of the systems to facilitate service:

- Color coding (touch points)

  Terracotta-colored touch points indicate that a component (FRU or CRU) can be concurrently maintained.

  Blue-colored touch points delineate components that may not be concurrently maintained (they might require that the system is turned off for removal or repair).

- Positive retention

  Positive retention mechanisms help ensure proper connections between hardware components, such as from cables to connectors, and between two adapters that attach to each other. Without positive retention, hardware components risk becoming loose during shipping or installation, which prevents a good electrical connection. Positive retention mechanisms such as latches, levers, thumb-screws, pop Nylatches (U-clips), and cables are included to help prevent loose connections and aid in installing (seating) parts correctly. These positive retention items do not require tools.

Service Indicator LED function

The Service Indicator LED function is for scale-out systems, including Power Systems such as the Power S821LC server, that can be repaired by clients. In the Service Indicator LED implementation, when a fault condition is detected on the POWER8 processor-based system, an amber FRU fault LED is illuminated (turned on solid), which is then rolled up to the system fault LED.

When the ID LED button on the front panel is pressed, the blue LED on the front panel and the blue ID LED on the rear panel light up. The technical personnel can easily locate the system on the rack, disconnect cables from the system, and remove it from the rack for later repair.

The Service Indicator operator panel contains the following items:

- Power On LED (Green LED: Front)
  - Off: Enclosure is off.
  - On Solid: Enclosure is powered on.
  - On Blink: Enclosure is in the standby-power state.
- Enclosure Identify LED (Blue LED: Front)
  - Off: Normal.
  - On Solid: Identify state.
  - On Blink: Reserved.
- System Information/Attention LED (Amber LED: Front)
  - Off: Normal.
  - On Solid: System Attention State.
- Enclosure Fault Roll-up LED (Amber LED: Front)
  - Off: Normal.
  - On Solid: Fault.
  - Power On/Off Switch.
  - Pin-hole Reset Switch.
  - USB Port.
  - Beeper.
  - Altitude Sensor with Ambient Thermal Sensor.
  - VPD Module.

Concurrent maintenance

The following components can be replaced without powering off the server:

- Drives in the front bay
- Power supplies
- Fans

The POWER8 processor-based systems are designed with the understanding that certain components have higher intrinsic failure rates than others. These components can include fans, power supplies, and physical storage devices. Other devices, such as I/O adapters, can wear from repeated plugging and unplugging. For these reasons, these devices are concurrently maintainable when properly configured. Concurrent maintenance is facilitated because of the redundant design for the power supplies and physical storage.

IBM Knowledge Center

IBM Knowledge Center provides you with a single place where you can access product documentation for IBM systems hardware, OSs, and server software.

The purpose of IBM Knowledge Center, in addition to providing client-related product information, is to provide softcopy information to diagnose and fix any problems that might occur with the system. Because the information is electronically maintained, changes because of updates or the addition of new capabilities can be used by SSRs immediately.

IBM Knowledge Center provides the following up-to-date documentation to effectively service the system:

- *Quick Install Guide*
- *User's Guide*
- *Trouble Shooting Guide*
- *Boot Configuration Guide*

The documentation can be downloaded in PDF format or used online through an internet connection.

IBM Knowledge Center can be found at the following website:

`http://www.ibm.com/support/knowledgecenter/`

Supporting information for the Power S821LC server is available online at IBM Knowledge Center:

`http://www.ibm.com/support/knowledgecenter/HW4L4/p8hdx/8335_gca_landing.htm`

Warranty and spare parts

The system comes with a 3-year warranty for parts. The replacement parts can be ordered through the Advanced Part Exchange Warranty Service, which can be found at the following website:

`http://www.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=877/ENUSZG15-0194&infotype=AN&subtype=CA&appname=skmwww`

3.5 Manageability

Several functions and tools help you can efficiently and effectively manage your system.

3.5.1 Service user interfaces

The service interface allows support personnel or the client to communicate with the service support applications in a server by using a console, interface, or terminal. Delivering a clear, concise view of available service applications, the service interface allows the support team to manage system resources and service information in an efficient and effective way. Applications that are available through the service interface are carefully configured and placed to give service providers access to important service functions.

Various service interfaces are used depending on the state of the system and its operating environment. Here are the primary service interfaces:

- Service Indicator LEDs (See “Service Indicator LED function” on page 47 and “Concurrent maintenance” on page 47.)
- Service processor

Service Interface

The service interface allows the client and the support personnel to communicate with the service support applications in a server by using a browser. It delivers a clear, concise view of available service applications. The service interface allows the support client to manage system resources and service information in an efficient and effective way. Different service interfaces are used depending on the state of the system, hypervisor, and operating environment. Here are the primary service interfaces:

- Service processor: Ethernet Service Network with IPMI Version 2.0
- Service Indicator LEDs: System attention and system identification (front and back)
- Host OS: Command-line interface (CLI)

The service processor is a controller that is running its own OS.

3.5.2 IBM Power Systems firmware maintenance

The IBM Power Systems Client-Managed Licensed Internal Code is a methodology that you can use to manage and install Licensed Internal Code updates on a Power Systems server and its associated I/O adapters.

Firmware updates

System firmware is delivered as a release level or a service pack. Release levels support the general availability (GA) of new functions or features, and new machine types or models. Upgrading to a higher release level is disruptive to customer operations. These release levels are supported by service packs. Service packs are intended to contain only firmware fixes and not introduce new functions. A *service pack* is an update to an existing release level.

IBM is increasing its clients’ opportunity to stay on a given release level for longer periods. Clients that want maximum stability can defer until there is a compelling reason to upgrade, such as the following reasons:

- A release level is approaching its end of service date (that is, it has been available for about a year, and soon service will not be supported).
- Move a system to a more standardized release level when there are multiple systems in an environment with similar hardware.
- A new release has a new function that is needed in the environment.
- A scheduled maintenance action causes a platform restart, which provides an opportunity to also upgrade to a new firmware release.

The updating and upgrading of system firmware depends on several factors, such as the current firmware that is installed, and what OSs are running on the system. These scenarios and the associated installation instructions are comprehensively outlined in the firmware section of Fix Central, found at the following website:

http://www.ibm.com/support/fixcentral/

3.5.3 Updating the system firmware with the ipmitool command

General firmware update steps for the Power S821LC server are managed by running the **ipmitool** command. Complete the following steps, but be sure that you always see the provided Firmware Release notes for the most current Installation instructions:

1. Power off the machine and install the code from the Standby Power state by running the following command:

   ```
   ipmitool -H <hostname> -I lan -U ADMIN -P admin chassis power off
   ```

2. Issue a BMC reset (establish a stable starting point) by running the following command:

   ```
   ipmitool -H <BMC IP> -I lan -U ADMIN  -P admin mc reset cold
   ```

   From a companion system, run the following commands to flash the BMC and firmware:

   - **ipmitool -H <BMC IP> -I lanplus -U ADMIN -P admin raw 0x32 0xba 0x18 0x00** (The command protects the BMC memory content so that you do not lose the network settings.)
   - **ipmitool -H <BMC IP> -U ADMIN -I lanplus -P admin hpm upgrade <xxxxx.hpm> -z 30000 force**

   Attention: If you experience a seg fault error during the code update, run the command again and change the block size from 30000 to 25000.

If the BMC network settings are lost, it is possible to restore them by completing the following steps:

1. Set up a serial connection to the BMC by logging in and running the following commands to set up the network:
   - **/usr/local/bin/ipmitool -H 127.0.0.1 -I lan -U ADMIN -P admin lan set 1 ipsrc static**
   - **/usr/local/bin/ipmitool -H 127.0.0.1 -I lan -U ADMIN -P admin lan set 1 ipaddr x.x.x.x**
   - **/usr/local/bin/ipmitool -H 127.0.0.1 -I lan -U ADMIN -P admin lan set 1 netmask 255.255.x.x**
   - **/usr/local/bin/ipmitool -H 127.0.0.1 -I lan -U ADMIN -P admin lan set 1 defgw ipaddr x.x.x.x**
2. Power on and perform an IPL the machine by running the following command:

   ```
   ipmitool -H <hostname> -I lan -U ADMIN -P admin chassis power on
   ```

3.5.4 Updating the ipmitool on Ubuntu

The level of ipmitool on the Ubuntu 14.04.3 trusty archives (1.8.13-1ubuntu0.3) does not include all the fixes that are required for in-band code update support for Open Power systems. This section explains how to load, patch, and compile manually ipmitool on Ubuntu 14.04.3 to enable in-band code update support for the IBM S821LC server.

How to install ipmitool V1.8.15 and patches for an in-band code update

Open Power requires ipmitool level V1.8.15 (with patches) to run correctly on the OP810 firmware, especially the ipmitool code update function.

Note: All commands should be run as root or preceded by the **sudo** command.

Complete the following steps:

1. Remove ipmitool if it exists on your Ubuntu 14.04.3 installation by running the following command:

   ```
   apt-get remove ipmitool
   ```

2. Install the following packages by running the following command:

   ```
   apt-get install gcc make automake
   ```

3. Create a directory that is called `ipmitool_patch` and run **cd** to access it buy running the following commands:

   - **mkdir /ipmitool_patch**
   - **cd /ipmitool_patch**

4. Download the following files into the `/ipmitool_patch` directory by running the following commands:

   - **wget https://launchpad.net/ubuntu/+archive/primary/+files/ipmitool_1.8.15.orig.tar.bz2**
   - **wget https://launchpad.net/ubuntu/+archive/primary/+files/ipmitool_1.8.15-1ubuntu0.1.debian.tar.xz**

5. Decompress the files by running the following commands:

   - **bzip2 -d ipmitool_1.8.15.orig.tar.bz2**
   - **tar xvf ipmitool_1.8.15.orig.tar**
   - **tar xvf ipmitool_1.8.15-1ubuntu0.1.debian.tar.xz**

6. Copy the Debian patch files to the `ipmitool-1.8.15` directory by running the following command:

   ```
   cp debian/patches/*.patch ipmitool-1.8.15/
   ```

7. Change the directory to `ipmitool-1.8.15/` by running the following command:

   ```
   cd ipmitool-1.8.15/
   ```

8. Patch the source files by running the following commands:

   - **patch -p1 < usb_interface_support.patch**
   - **patch -p1 < memcpy_hpm_fix.patch**
   - **patch -p1 < 112_fix_CVE-2011-4339.patch**
   - **patch -p1 < 101_fix_buf_overflow.patch**
   - **patch -p1 < 098-manpage_typo.patch**
   - **patch -p1 < 096-manpage_longlines.patch**

9. Configure ipmitool for your system by running the following command:

   ```
   ./configure
   ```

Note: Be sure that are you in the `/ipmitool_patch/ipmitool-1.8.15/` directory!

10. Verify the command output. The last part of the output should look like Example 3-3. Note that the usb interface is `yes`.

Example 3-3 Verify output

```
ipmitool 1.8.15

Interfaces
  lan     : yes
  lanplus : no
  open    : yes
  free    : no
  imb     : yes
  bmc     : no
  usb     : yes
  lipmi   : no
  serial  : yes
  dummy   : no

Extra tools
  ipmievd   : yes
  ipmishell : no
```

11. Make the source files and install them by running the following commands:
    - **`make`**
    - **`make install`**
12. Log out of the system and log in to the system.
13. Verify what level of ipmitool is installed by running the following commands:
    - **`ipmitool -V`**
    - **`ipmitool version 1.8.15`**
14. Verify that the USB support is working by running the following command:

    ```
    ipmitool -I usb power status
    ```

    You should see the following output:

    ```
    Chassis Power is on
    ```

You should now be able to use this level of ipmitool for an in-band code update on Open Power systems.

For more information about this process, see the white papers that are found in IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/linuxonibm/liaai.ipmi/liaaiipmi.htm

3.5.5 Statement of direction: Updating the system firmware by using the Advanced System Management console

As a statement of direction, IBM plans to enhance the Advanced System Management console for firmware update activities. The most convenient method to update the system firmware on the Power S821LC server is to use the Advanced System Management GUI. It is comparable to the HMC GUI, and you can use it to simply go through the different windows and select and update the system firmware.

To update the system firmware by using the Advanced System Management console, complete the following steps:

1. Connect to the service processor interface. Use your browser and access the service processor by using the configured IP address. Log in by using the user name and password that are used in 2.2.2, “Intelligent Platform Management Interface” on page 35. Some browsers may not let you log in, but it is not a user name and password problem. If you cannot log in by using your browser, try to log in by using the Chrome browser.

   Figure 3-1 shows the Advanced System Management login window.

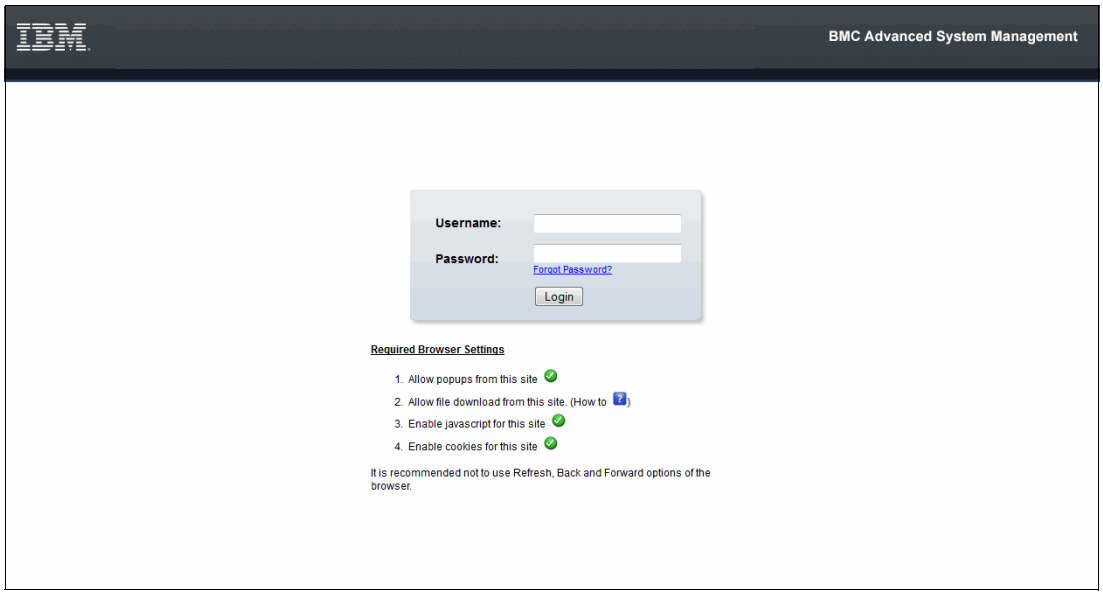


Figure 3-1 Advanced System Management GUI login window

After a successful login, the Advanced System Management Dashboard opens. It is the common window for multiple activities that can be performed, such as configuration, viewing FRU information, and performing firmware updates. General information about the current power consumption, sensor monitoring, and event logs is displayed.

Figure 3-2 shows the Dashboard window.

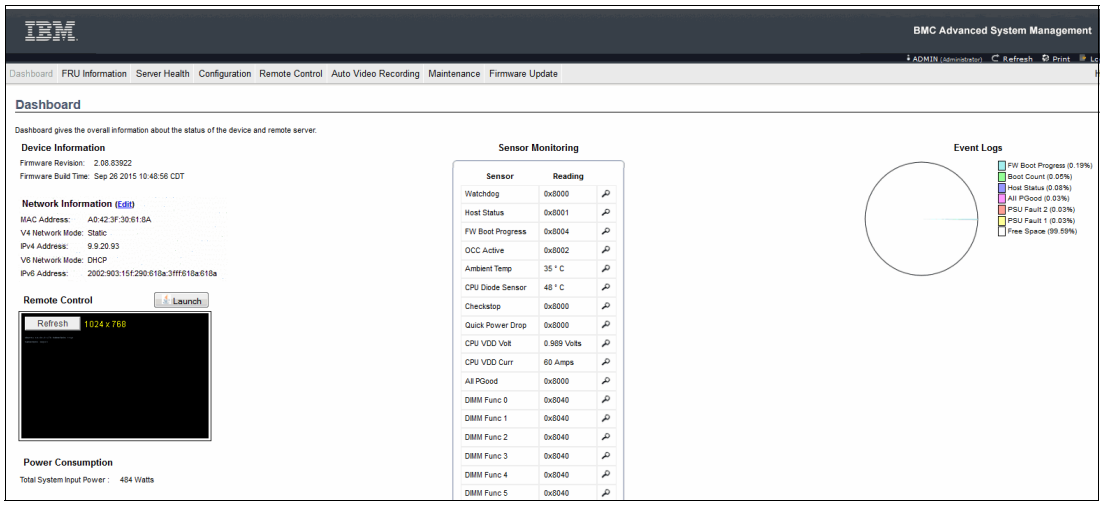


Figure 3-2 Advanced System Management Dashboard

2. Click **Firmware Update** → **Firmware Update**, as shown in Figure 3-3.

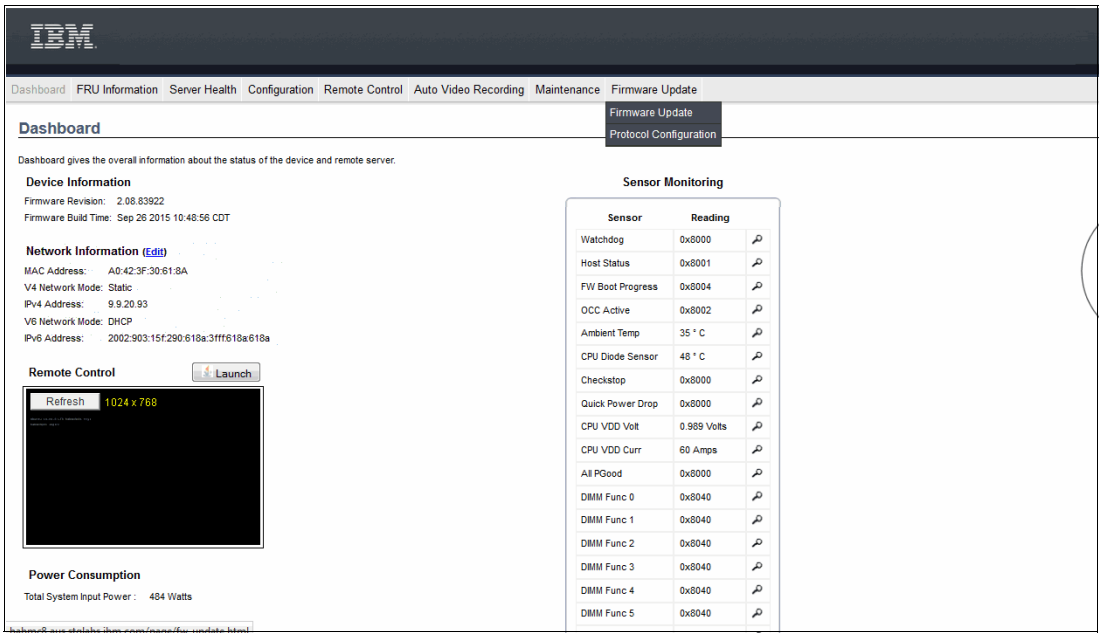


Figure 3-3 Dashboard Firmware Update menu

3. Select the correct firmware update image type. In this example, select **HPM**, which is the only type that is provided by the IBM Fix Central website, as shown in Figure 3-4.

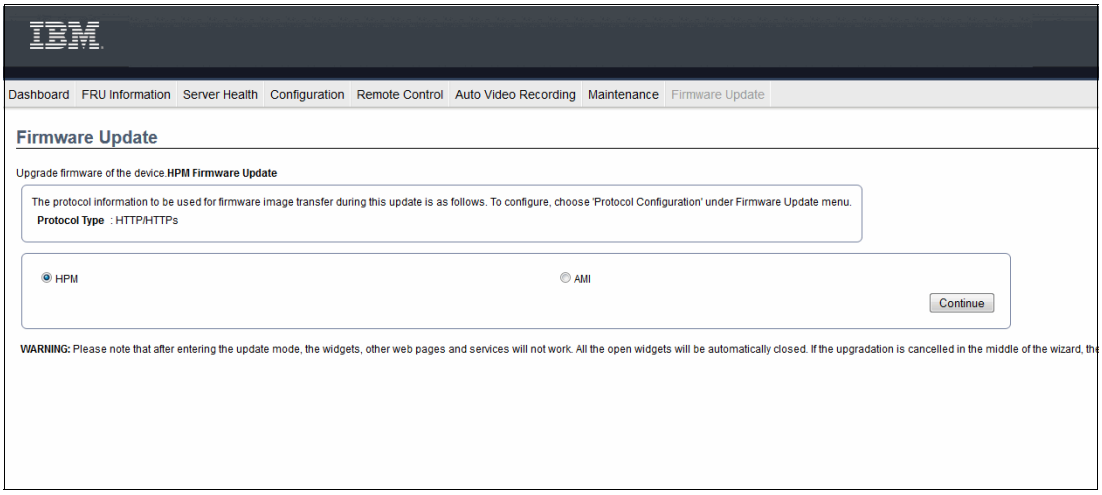


Figure 3-4 Select the firmware image type

4. Confirm that you want to update the HPM image by clicking **OK**, as shown in Figure 3-5.

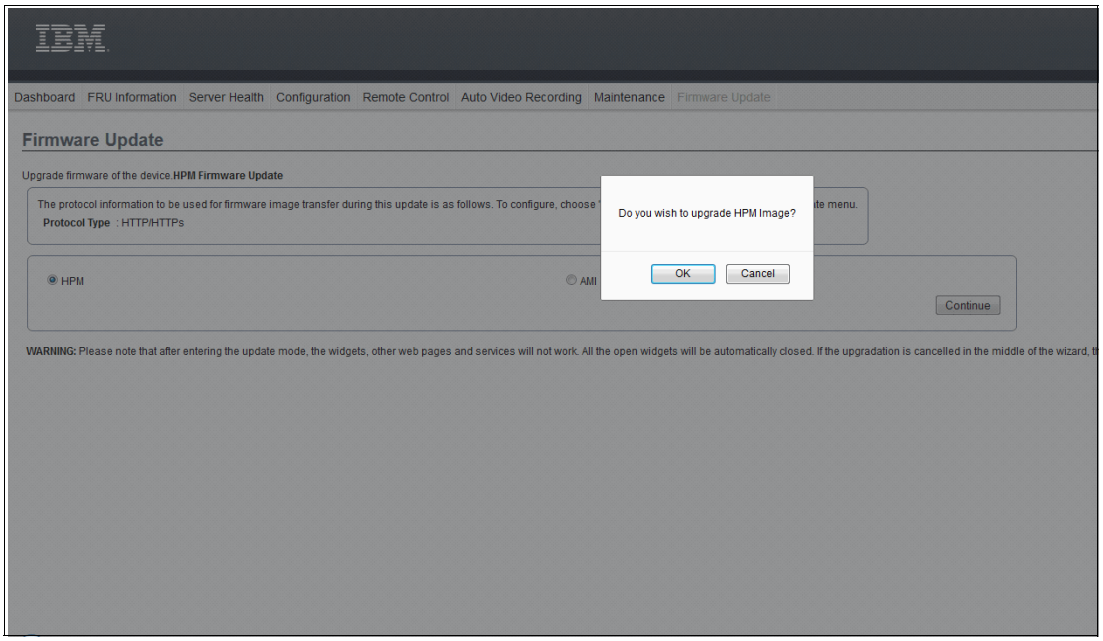


Figure 3-5 Confirm your update selection

A window opens that shows which components will be overwritten or preserved, as shown in Figure 3-6. For this example, the network settings are preserved.

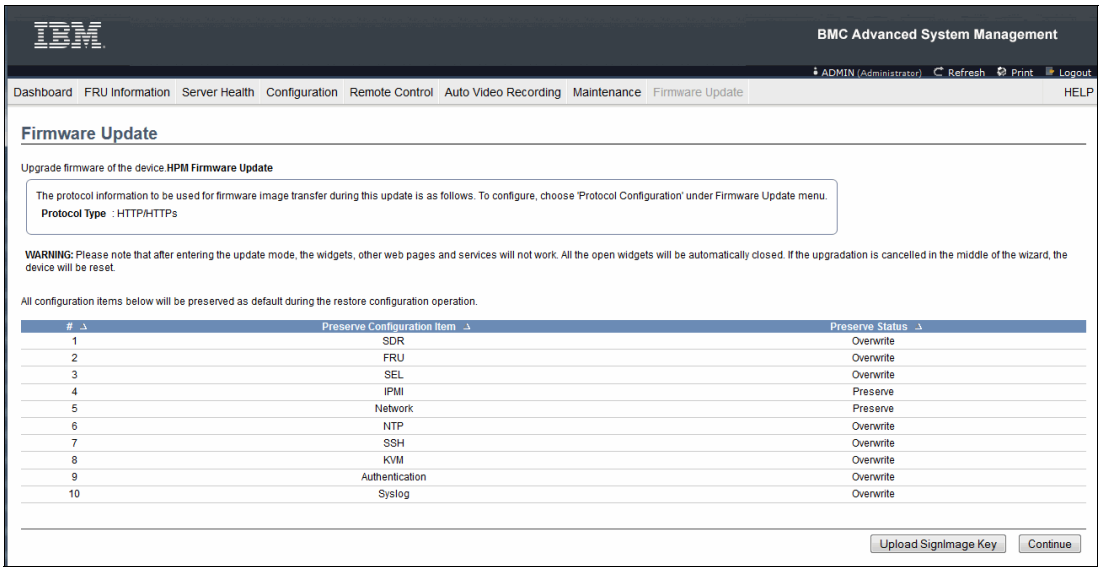


Figure 3-6 Firmware Update window

5. The next window prompts whether you want to continue to the update mode, as shown in Figure 3-7. Until the firmware update is completed, no other activities can be performed in the Advanced System Management Interface. If you want to proceed, click **OK**.

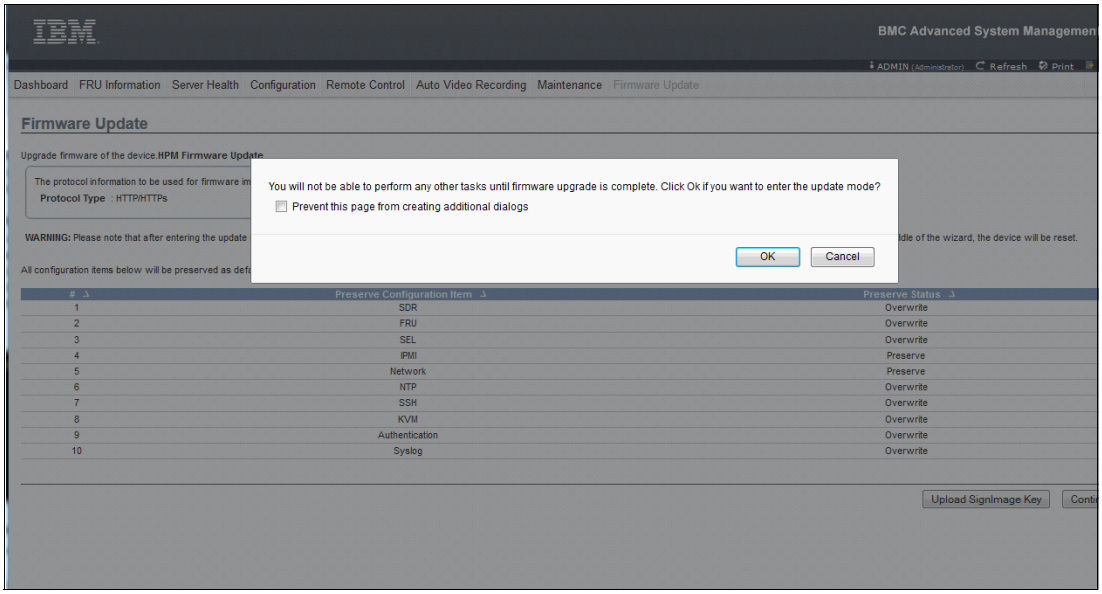


Figure 3-7 Confirm firmware update mode

6. Select the firmware update file from your local disk by selecting **Browse and Parse HPM Firmware Image**, clicking **Browse**, and selecting the file, as shown in Figure 3-8.

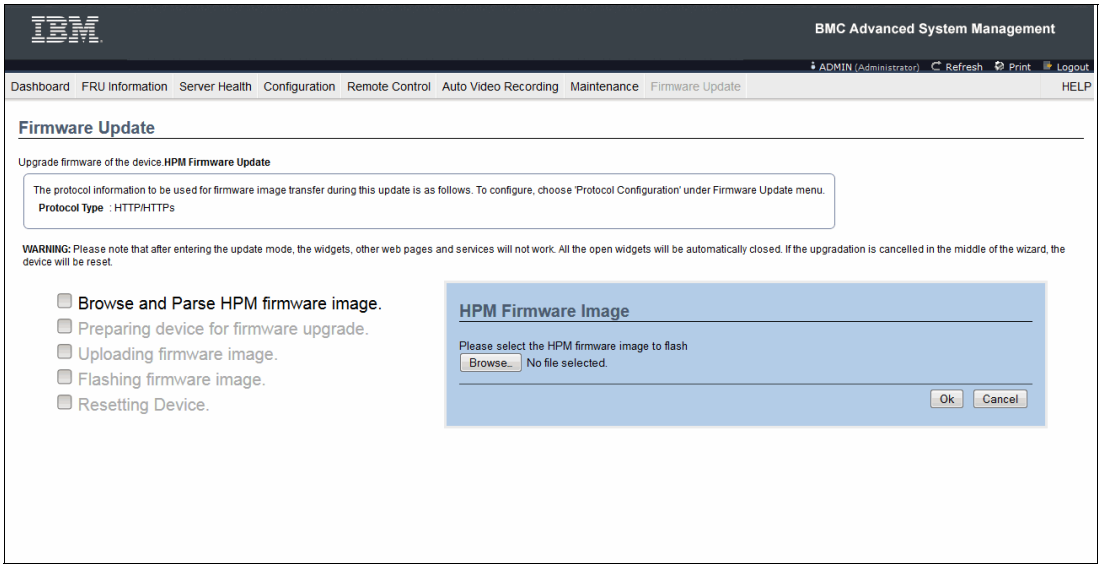


Figure 3-8 Select the firmware image

7. When the correct firmware image is selected, the GUI shows a list of components that will be updated, as shown in Figure 3-9. By default, all the components are selected. To update the firmware, click **Proceed**.

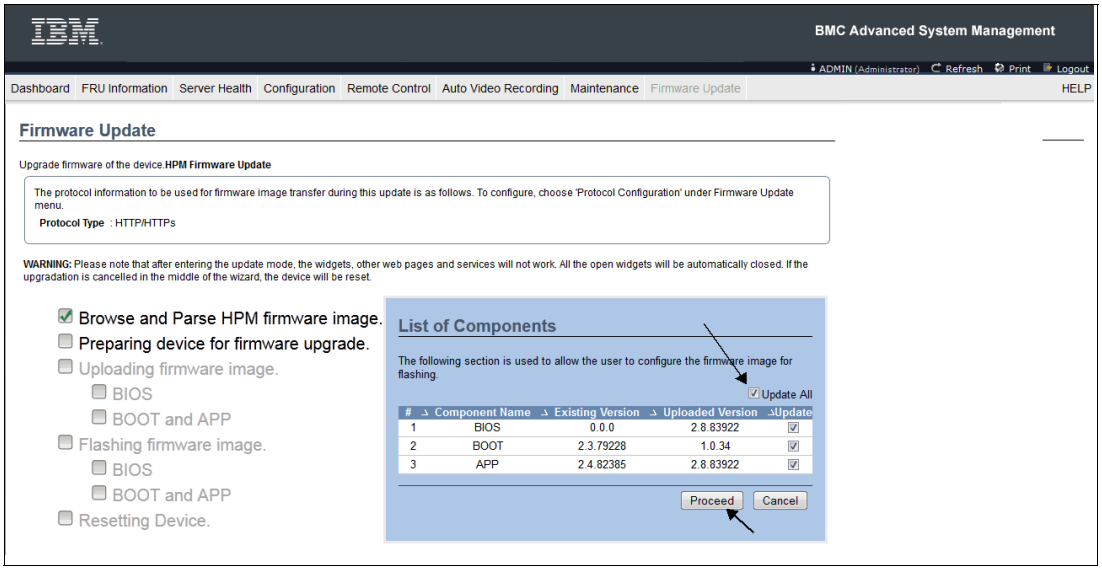


Figure 3-9 Start the firmware upgrade

8. After the firmware update is complete, the system restarts. After the restart, you can verify that the systems firmware was updated by opening the Advanced System Management Dashboard window.

A

Server racks and energy management

This appendix provides information about the racking options and energy management-related concepts that are available for the IBM Power System S821LC server.

IBM server racks

The Power S821LC server mounts in the IBM 42U Slim Rack (7965-94Y) along with square hole industry-standard racks. These racks are built to the 19-inch EIA 310D standard.

Order information: Power 821LC servers cannot be integrated into these racks during the manufacturing process, and are not orderable together with servers. If the Power 821LC 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 “OEM racks” on page 63.

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.

IBM 42U Slim Rack (7965-94Y)

The 2.0-meter (79-inch) Model 7965-94Y is compatible with past and present Power Systems servers and provides an excellent 19-inch rack enclosure for your data center. Its 600 mm (23.6 in.) width combined with its 1100 mm (43.3 in.) depth plus its 42 EIA enclosure capacity provides great footprint efficiency for your systems and allows it to be easily placed on standard 24-inch floor tiles.

The IBM 42U Slim Rack has a lockable perforated front steel door that provides ventilation, physical security, and visibility of indicator lights in the installed equipment within. In the rear, either a lockable perforated rear steel door (#EC02) or a lockable Rear Door Heat Exchanger (RDHX)(1164-95X) is used. Lockable optional side panels (#EC03) increase the rack’s aesthetics, help control airflow through the rack, and provide physical security. Multiple 42U Slim Racks can be bolted together to create a rack suite (indicate feature code #EC04).

Up to six optional 1U PDUs can be placed vertically in the sides of the rack. Additional PDUs can be placed horizontally, but they each use 1U of space in this position.

The AC power distribution unit and rack content

For the IBM 42U Slim Rack (7965-94Y) rack model, 12-outlet PDUs are available. These PDUs include the AC power distribution unit #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 7965-94Y rack. Figure A-1 shows the placement of the four vertically mounted PDUs. In the rear of the rack, two additional PDUs can be installed horizontally in the 7965-94Y rack. The four vertical mounting locations are filled first in the 7965-94Y slimline rack. Mounting PDUs horizontally consumes 1U per PDU and reduces the space that is available for other racked components. When mounting PDUs horizontally, the preferred approach is to use fillers in the EIA units that are occupied by these PDUs to facilitate the correct airflow and ventilation in the rack.

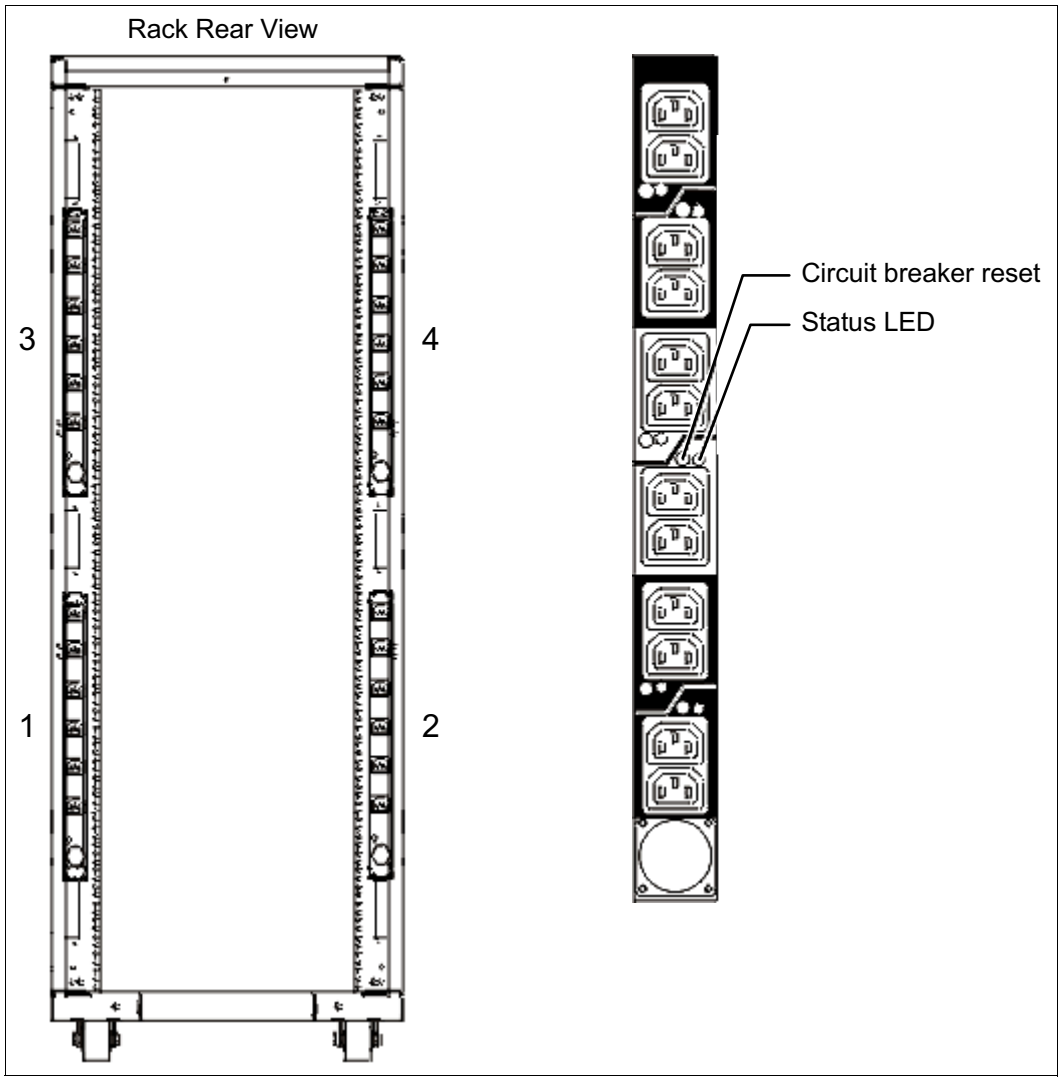


Figure A-1 PDU placement and PDU view

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 A-1 shows the available wall power cord options for the PDU and iPDU features, which must be ordered separately.

Table A-1 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 IBM 42U Slim Rack (7965-94Y) racks, see IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/api/redirect/powersys/v3r1m5/topic/p7had/p7hadrpower.htm

For detailed power requirements and power cord details about the 7965-94Y rack, see IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/api/redirect/powersys/v3r1m5/topic/p7had/p7hadkickoff795394x.htm

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.

OEM racks

The system can be installed in a suitable square hole 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:

http://www.ibm.com/support/knowledgecenter/api/redirect/systems/scope/hw/index.jsp

Energy consumption estimation

Often, for Power Systems servers, various energy-related values are important:

- Maximum power consumption and power source loading values

  These values are important for site planning and are described in IBM Knowledge Center:

  http://www.ibm.com/support/knowledgecenter/api/redirect/powersys/v3r1m5/index.jsp

  Search for type and model number and "server specifications". For example, for the Power S821LC server, search for "8001-12C".

- An estimation of the energy consumption for a certain configuration

  Calculate the energy consumption for a certain configuration in the IBM Systems Energy Estimator at the following website:

  http://www-912.ibm.com/see/EnergyEstimator

  In that tool, select the type and model for the system, and enter details about the configuration and CPU usage that you want. As a result, the tool shows the estimated energy consumption and the waste heat at the usage that you want and also at full usage.

Related publications

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

IBM Redbooks

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

- *IBM Power System E850 Technical Overview and Introduction*, REDP-5222
- *IBM Power System E850C Technical Overview and Introduction*, REDP-5412
- *IBM Power System S812LC Technical Overview and Introduction*, REDP-5284
- *IBM Power System S822 Technical Overview and Introduction*, REDP-5102
- *IBM Power System S822LC for Big Data Technical Overview and Introduction*, REDP-5407
- *IBM Power System S822LC for High Performance Computing Introduction and Technical Overview*, REDP-5405
- *IBM Power System S822LC Technical Overview and Introduction*, REDP-5283
- *IBM Power Systems E870 and E880 Technical Overview and Introduction*, REDP-5137
- *IBM Power Systems E870C and E880C Technical Overview and Introduction*, REDP-5413
- *IBM Power Systems HMC Implementation and Usage Guide*, SG24-7491
- *IBM Power Systems S812L and S822L Technical Overview and Introduction*, REDP-5098
- *IBM Power Systems S814 and S824 Technical Overview and Introduction*, REDP-5097
- *IBM Power Systems SR-IOV: Technical Overview and Introduction*, REDP-5065
- *IBM PowerVM Best Practices*, SG24-8062
- *IBM PowerVM Enhancements What is New in 2013*, SG24-8198
- *IBM PowerVC Version 1.3.1 Introduction and Configuration Including IBM Cloud PowerVC Manager*, SG24-8199
- *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- *Performance Optimization and Tuning Techniques for IBM Power Systems Processors Including IBM POWER8*, SG24-8171

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

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- IBM Fix Central website

  http://www.ibm.com/support/fixcentral/
- IBM Knowledge Center

  http://www.ibm.com/support/knowledgecenter/
- IBM POWER8 systems facts and features

  http://www.ibm.com/systems/power/hardware/reports/factsfeatures.html
- IBM Power Systems website

  http://www.ibm.com/systems/power/
- IBM Power Systems Hardware IBM Knowledge Center:

  http://www.ibm.com/support/knowledgecenter/api/redirect/powersys/v3r1m5/index.jsp
- IBM Power Systems S812L server specifications

  http://www.ibm.com/systems/power/hardware/s812l-s822l/index.html
- IBM Power Systems S814 server specifications

  http://www.ibm.com/systems/power/hardware/s814/index.html
- IBM Power System S821LC server specifications

  http://www.ibm.com/systems/power/hardware/s821lc/index.html
- IBM Power Systems S822 server specifications

  http://www.ibm.com/systems/power/hardware/s822/index.html
- IBM Power Systems S822L server specifications

  http://www.ibm.com/systems/power/hardware/s812l-s822l/index.html
- IBM Power Systems S822LC for Big Data server specifications

  http://www.ibm.com/systems/power/hardware/s822lc-big-data/index.html
- IBM Power System S822LC for Commercial Computing server specifications

  http://www.ibm.com/systems/power/hardware/s822lc-commercial/index.html
- IBM Power Systems S822LC for High-Performance Computing server specifications

  http://www.ibm.com/systems/power/hardware/s822lc-hpc/index.html
- IBM Power Systems S824 server specifications

  http://www.ibm.com/systems/power/hardware/s824/index.html
- IBM Power Systems S824L server specifications:

  http://www.ibm.com/systems/power/hardware/s824l/index.html
- IBM Power Systems E850 server specifications:

  http://www.ibm.com/systems/power/hardware/e850/index.html
- IBM Power Systems E870 server specifications:

  http://www.ibm.com/systems/power/hardware/e870/index.html
- IBM Power Systems E870C server specifications:

  http://www.ibm.com/systems/power/hardware/enterprise-cloud/index.html

- IBM Power Systems E880 server specifications:

  http://www.ibm.com/systems/power/hardware/e880/index.html
- IBM Power Systems E880C server specifications:

  http://www.ibm.com/systems/power/hardware/enterprise-cloud/index.html
- IBM Storage website

  http://www.ibm.com/systems/storage/
- IBM System Planning Tool website

  http://www.ibm.com/systems/support/tools/systemplanningtool/
- IBM Systems Energy Estimator

  https://www-912.ibm.com/see/EnergyEstimator
- *POWER8 Processor-Based Systems RAS - Introduction to Power Systems Reliability, Availability, and Serviceability*

  http://www.ibm.com/common/ssi/cgi-bin/ssialias?subtype=WH&infotype=SA&appname=STGE_PO_PO_USEN&htmlfid=POW03133USEN

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

REDP-5406-00

ISBN 0738455784

Printed in U.S.A.