Redbooks
ibm.com/redbooks

IBM Power System E850C
Technical Overview and Introduction

Alexandre Bicas Caldeira

Volker Haug

Cloud

Power Systems

IBM

Redpaper

International Technical Support Organization

IBM Power System E850C: Technical Overview and Introduction

October 2016

REDP-5412-00

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

First Edition (October 2016)

This edition applies to the IBM Power System E850C (8408-44E) server.

This document was created or updated on July 12, 2017.

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™	IBM Spectrum™	PowerPC®
AIX®	IBM z™	PowerSC™
Bluemix®	POWER®	PowerVM®
DB2®	POWER Hypervisor™	PowerVP™
DS8000®	Power Systems™	Redbooks®
Easy Tier®	Power Systems Software™	Redpaper™
Electronic Service Agent™	POWER6®	Redbooks (logo) ®
EnergyScale™	POWER6+™	Storwize®
GPFS™	POWER7®	System Storage®
IBM®	POWER7+™	SystemMirror®
IBM API Connect™	POWER8®	WebSphere®
IBM FlashSystem®	PowerHA®	XIV®

The following terms are trademarks of other companies:

SoftLayer, and SoftLayer device are trademarks or registered trademarks of SoftLayer, Inc., an IBM Company.

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.

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

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

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™ E850C (8408-44E) server that supports IBM AIX®, and Linux operating systems. The objective of this paper is to introduce the major innovative Power E850C offerings and their relevant functions.

The Power E850C server (8408-44E) is the latest enhancement to the Power Systems portfolio. It offers an improved 4-socket 4U system that delivers faster IBM POWER8® processors up to 4.22 GHz, with up to 4 TB of DDR4 memory, built-in IBM PowerVM® virtualization, and capacity on demand. It also integrates cloud management to help clients deploy scalable, mission-critical business applications in virtualized, private cloud infrastructures.

Like its predecessor Power E850 server, which was launched in 2015, the new Power E850C server uses 8-core, 10-core, or 12-core POWER8 processor modules. However, the Power E850C cores are 13%-20% faster and deliver a system with up to 32 cores at 4.22 GHz, up to 40 cores at 3.95 GHz, or up to 48 cores at 3.65 GHz, and use DDR4 memory. A minimum of two processor modules must be installed in each system, with a minimum quantity of one processor module's cores activated.

Cloud computing, in its many forms (public, private, or hybrid), is quickly becoming both the delivery and consumption models for IT. However, finding the correct mix between traditional IT, private cloud, and public cloud can be a challenge. The new Power E850C server and IBM Cloud PowerVC manager can enable clients to accelerate the transformation of their IT infrastructure for cloud while providing tremendous flexibility during the transition.

IBM Cloud PowerVC Manager provides OpenStack-based cloud management to accelerate and simplify cloud deployment by providing fast and automated VM deployments, prebuilt image templates, and self-service capabilities all with an intuitive interface. PowerVC management upwardly integrates into various third-party hybrid cloud orchestration products, including IBM Cloud Orchestrator, VMware vRealize, and others. Clients can simply manage both their private cloud VMs and their public cloud VMs from a single, integrated management tool.

IBM Power Systems is designed to provide the highest levels of reliability, availability, flexibility, and performance to bring you a world-class enterprise private and hybrid cloud infrastructure. Through enterprise-class security, efficient built-in virtualization that drives industry-leading workload density, and dynamic resource allocation and management, the server consistently delivers the highest levels of service across hundreds of virtual workloads on a single system.

The Power E850C server includes the cloud management software and services to assist with clients' move to the cloud, both private and hybrid. Those additional capabilities include the following items:

- Private cloud management with IBM Cloud PowerVC Manager, Cloud-based HMC Apps as a service, and Open source cloud automation and configuration tooling for AIX
- Hybrid cloud support
- Hybrid infrastructure management tools
- Securely connect system of record workloads and data to cloud native applications
- IBM Cloud Starter Pack

- Flexible capacity on demand
- Power to Cloud Services

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 E850C system.

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

Authors

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

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 has worked as an IBM Business Partner on Power Systems hardware, 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 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, AIX, and 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:

George Ahrens
Joanna Bartz
Thomas Bosworth
Petra Buehrer
David Bruce
Sertak Cakici
Joseph W Cropper
Timothy Damron

Kirk Dietzman
Daniel Henderson
Dan Hurliman
Roxette Johnson
Ann Lund
Edgar Michel Garcia
Bob Kovacs
Hans Mozes
Michael J Mueller
Mark Olson
Ravi A Shankar
Bill Starke
Dawn C Stokes
Jeff Stuecheli
Doug Szerdi
Joel Sebastian Torres De la Torre
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

General description

The Power E850C server (8408-44E) is the latest enhancement to the Power System portfolio. It offers an improved 4-socket system that delivers faster POWER8 processors up to 4.22 GHz, with up to 4 TB of DDR4 memory, built-in IBM PowerVM virtualization, capacity on demand. It also integrates cloud management to help clients deploy scalable, mission-critical business applications in virtualized, private cloud infrastructures.

The Power E850C is optimized for running AIX and Linux workloads. The Power E850C server is four EIA-units tall (4U).

The Power E850C server supports the IBM Solution Editions for the SAP HANA and Solution Edition for Healthcare offerings.

This chapter includes the following sections:

- Systems overview
- Operating environment
- Physical package
- System features
- Disk and media features
- I/O drawers
- Build to order
- IBM Power Systems Solution Editions
- Management options
- System racks

1.1 Systems overview

The following sections provide detailed information about the Power E850C system.

1.1.1 Power E850C server

The Power System E850C server offers new levels of performance, price/performance, and function with POWER8 technology using a 4-socket server. The Power E850C server provides these features:

- Up to 48 POWER8 cores
- Up to 4 TB of DDR4 memory on a 4-socket server
- Up to 51 Gen3 PCIe slots
- Integrated SAS controllers for integrated disk/SSD/DVD bays
- Up to 1,536 SAS bays for disk/SSD with EXP24S I/O drawers

The Power E850C server is a single enclosure server that uses IBM POWER8 processor technology. It can be configured with two, three, or four processor modules with one machine type-model (8408-44E).

The minimum processor configuration is two processor modules. However, with this configuration, only 7 of the 11 internal PCIe slots function. With three processor modules configured, 9 of the 11 PCIe slots function, and with four processor modules configured, all 11 PCIe slots are available for use.

The processor and memory subsystem in the Power E850C server is contained on a single planar. This CPU planar contains the four processor sockets and 32 memory Custom DIMM (CDIMM) slots. The minimum system configuration that is supported is two processor modules and eight CDIMMs. The rules covering the general offering is a minimum of four CDIMMs per processor socket populated.

The I/O subsystem in Power E850C consists of an IO planar, storage controller card, and a storage backplane. The IO planar supports 10 general-purpose PCIe Gen3 slots for full high/half length adapters. An additional (11th) slot is defaulted to the LAN adapter. Eight of the 11 slots are Coherent Accelerator Processor Interface (CAPI)-capable. The storage controller card contains two storage controllers that drive four SFF (2.5 in.) and eight 1.8 in. bays.

Three versions of the storage controller card are available:

- High performance dual controller with write cache
- Dual controller without write cache
- Split disk backplane

Up to 11 PCIe Gen3 slots are provided in the system unit. The number of PCIe slots depends on the number of processor modules:

- Two processor modules provide seven PCIe slots
- Three processor modules provide nine PCIe slots
- Four processor modules provide 11 PCIe slots

One of the slots must contain an Ethernet LAN adapter. Up to 40 additional PCIe Gen3 slots can be added by using PCIe Gen3 I/O Expansion Drawers:

- A two-processor module system can attach up to two PCIe Gen3 I/O drawers and provide up to 27 PCIe slots.
- A three-processor module system can attach up to three PCIe Gen3 I/O drawers and provide up to 39 PCIe slots.
- A four-processor module system can attach up to four PCIe Gen3 I/O drawers and provide up to 51 PCIe slots.

In addition to extensive hardware configuration flexibility, the Power E850C server offers elastic Capacity on Demand for both cores and memory, IBM Active Memory™ Expansion, and Active Memory Mirroring for Hypervisor. The Power E850C supports AIX and Linux operating systems, and PowerVM Standard and Enterprise editions.

The Power E850C provides strong reliability, availability, and serviceability characteristics:

- POWER8 chip capabilities
- Memory protection
- Multiple SAS storage protection options
- Hot plug SAS bays
- Hot plug PCIe slots
- Redundant and hot plug power supplies and cooling fans
- Redundant and spare internal cooling fans
- Hot plug Time of Day battery
- Highly resilient architecture for power regulators

The Power E850C server supports the IBM Solution Editions for SAP HANA and IBM Solution Edition for Healthcare offerings.

Figure 1-1 shows the Power E850C front view.

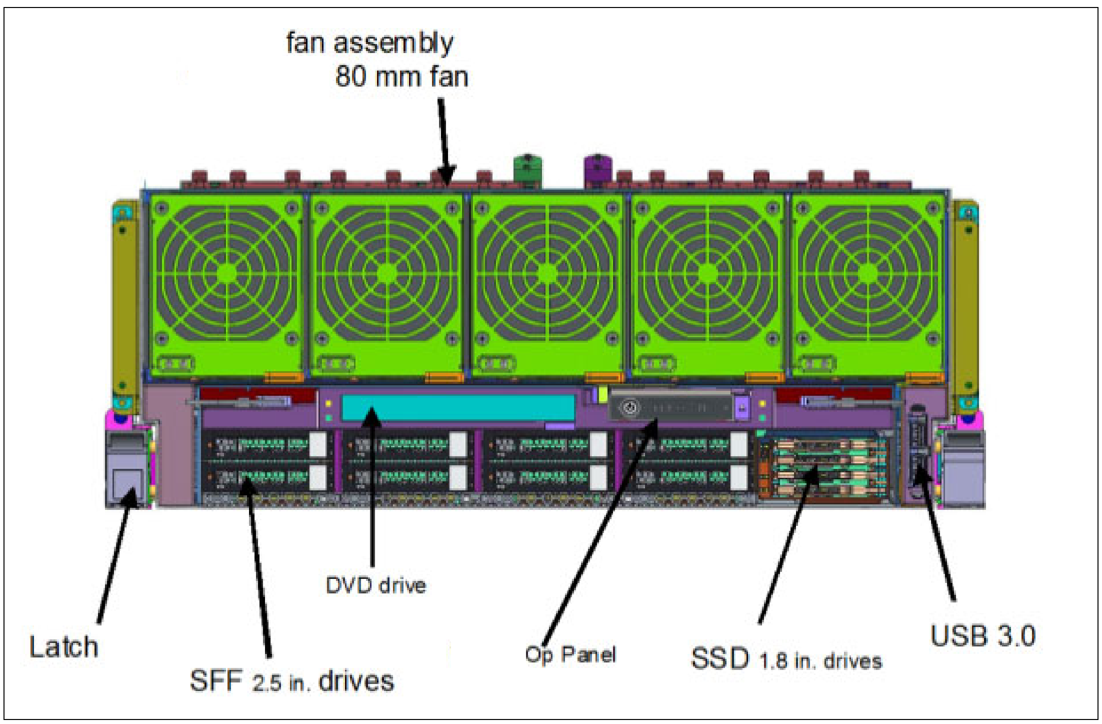

Figure 1-1 Front view of Power E850C server

Figure 1-2 shows the Power E850C rear view.

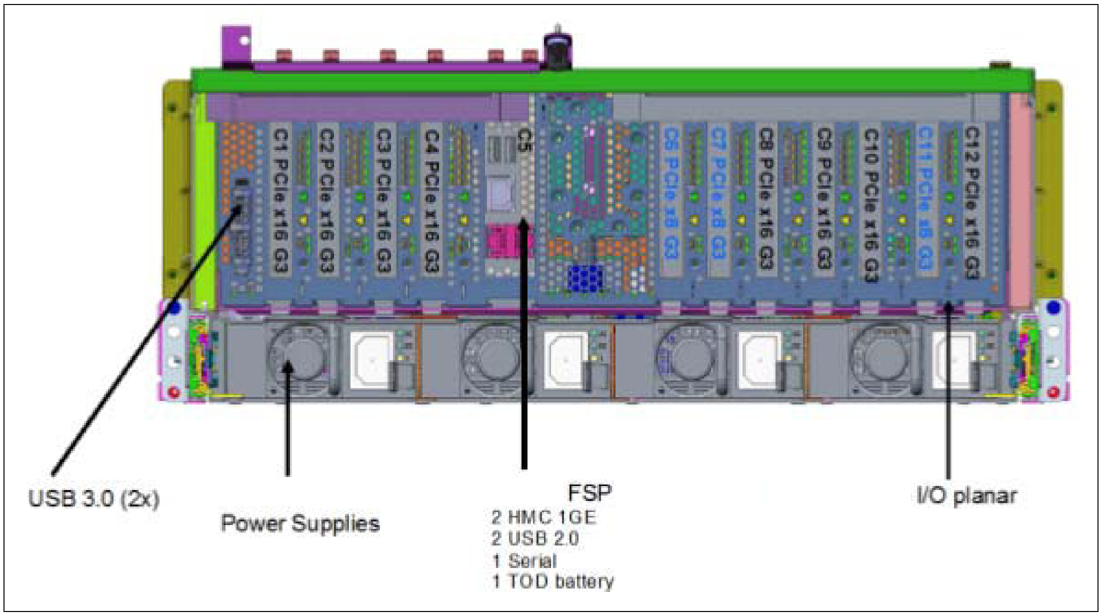

Figure 1-2 Rear view of Power E850C server

1.2 Operating environment

Table 1-1 details the operating environment for the Power E850C server.

Table 1-1 Operating environment for Power E850C

Power E850C operating environment		
System	**Power E850C**	
	Operating	**Non-operating**
Temperature	Recommended 18 - 27° C (64 - 80° F)	1 - 60° C (34 - 140° F)
	Allowable 5 - 40° C (41 - 104° F)	
Relative humidity	8 - 80%	5 - 80%
Maximum dew point	24° C (75° F)	27° C (80° F)
Operating voltage	200 - 240 V ac 180 - 400 V dc	N/A
Operating frequency	50 - 60 Hz +/- 3 Hz ac	N/A
Maximum power consumption	3,850 Watts	N/A
Maximum power source loading	3.90 kVA	N/A
Maximum thermal output	13,140 BTU/hour	N/A
Maximum altitude	3,050 m (10,000 ft.)	N/A

Environmental assessment: The IBM Systems Energy Estimator tool can provide more accurate information about power consumption and thermal output of systems based on a specific configuration, including adapters and I/O expansion drawers. The IBM Systems Energy Estimator tool can be accessed online at:

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

1.3 Physical package

Table 1-2 lists the physical dimensions of individual system nodes and of the PCIe I/O expansion drawer.

The system node requires 4U, and the PCIe I/O expansion drawer requires 4U. Thus, a single-enclosure system with one PCIe I/O expansion drawer requires 8U.

Table 1-2 Physical dimensions of the Power E850C and the PCIe I/O expansion drawer

Dimension	Power E850C system node	PCIe I/O expansion drawer
Width	448 mm (17.6 in.)	482 mm (19 in.)
Depth	776 mm (30.6 in.)	802 mm (31.6 in.)
Height	175 mm (6.9 in.) 4EIA units	175 mm (6.9 in.) 4 EIA units
Weight	69 kg (152 lb)	54.4 kg (120 lb)

To assure installability and serviceability in non-IBM industry-standard racks, review the installation planning information for any product-specific installation requirements.

Figure 1-3 shows a picture of a Power E850C system node from the front.

Figure 1-3 Front view of the Power 850C system node

1.4 System features

The Power E850C system nodes contain two to four processor modules with 512 KB L2 cache and 8 MB L3 cache per core, and L4 cache with DDR4 memory.

1.4.1 Power E850C system features

The following features are available on the Power E850:

- The Power E850C server supports 16 - 48 processor cores with two, three, or four POWER8 processor modules:
  - 8-core 4.22 GHz Processor Module (#EPW4).
  - 10-core 3.95 GHz Processor Module (#EPW5).
  - 12-core 3.65 GHz Processor Module (#EPW6).
- 128 GB to 4 TB high-performance DDR4 memory with L4 cache:
  - 16 GB CDIMM Memory (#EM8P).
  - 32 GB CDIMM Memory (#EM8Q).
  - 64 GB CDIMM Memory (#EM8R).
  - 128 GB CDIMM Memory (#EM8S).
- Optional Active Memory Expansion (#4798).
- Choice of three storage backplane features with different integrated SAS RAID controller options. All backplane options have eight SFF-3 SAS bays, four 1.8-inch SSD bays, and one DVD bay. All backplane options offer IBM Easy Tier® function for mixed HDD/SSD arrays:
  - Dual SAS Controller Backplane, without write cache (#EPVP).
  - Dual SAS Controller Backplane, with write cache (#EPVN).
  - Split Disk Backplane (two single SAS controllers), without write cache (#EPVQ).
- Up to 11 hot-swap PCIe Gen3 slots in the system unit:
  - Three x8 Gen3 full height, half length adapter slots.
  - Four, six, or eight x16 Gen3 full-height, half-length adapter slots.
  - With two processor modules, there are seven PCIe slots; with three modules, there are nine PCIe slots; and with four modules, there are 11 PCIe slots in the system unit.
  - One x8 PCIe slot is used for a LAN adapter.
- The PCIe Gen3 I/O Expansion Drawer (#EMX0) expands the number of full-high, hot-swap Gen3 slots:
  - Up to two PCIe3 drawers with two processor modules (maximum 27 slots on the server).
  - Up to three PCIe3 drawers with three processor modules (maximum 39 slots on the server).
  - Up to four PCIe3 drawers with four processor modules (maximum 51 slots on the server).
- Up to 64 EXP24S SFF Gen2-bay Drawers (#5887) can be attached, providing up to 1536 SAS bays for disk or SSD.

- System unit I/O (integrated I/O):
  - HMC ports: Two 1 GbE RJ45.
  - USB ports: Four USB 3.0 (two front and two rear) for general use and USB 2.0 (rear) for limited use.
  - System (serial) port: One RJ45.
- Hot-plug, redundant power supplies:
  - 4-AC Power Supply 2000 W (200 - 240 V) (#EB3M).
- The Power E850C provides these strong reliability, availability, and serviceability characteristics:
  - POWER8 chip capabilities.
  - Memory protection.
  - Multiple SAS storage protection options.
  - Hot-plug SAS bays.
  - Hot-plug PCIe slots.
  - Redundant and hot-plug power supplies and cooling fans.
  - Highly resilient architecture for power regulators.
- Other features:
  - Second-generation service processor.
  - Hot-plug SAS bays and PCIe slots.
  - Redundant and hot-plug power supplies and fans.
  - Optional Active Memory Mirroring for Hypervisor (#EM81).
  - Highly resilient power regulator architecture.
- System unit only 4U in a 19-inch rack-mount hardware.
- Primary operating systems:
  - AIX (#2146) (small tier licensing).
  - Linux (#2147): RHEL, SUSE Linux Enterprise Server, and Ubuntu.

1.4.2 Minimum configuration

The minimum Power E850C initial order must include two processor modules, 128 GB of memory, a storage backplane, one HDD or SSD DASD device, a PCIe2 4-port 1 GbE adapter, four power supplies and power cords, an operating system indicator, and a Language Group Specify.

Note:

- A machine type/model rack, if wanted, should be ordered as the primary rack.
- Permanent Capacity on Demand (CoD) processor core activations are required for the first processor module in the configuration and are optional for the second, third, and fourth modules. Specifically:
  - Two, three, or four 8-core #EPW4 require eight processor core activations (#EPW7) minimum.
  - Two, three, or four 10-core #EPW5 require ten processor core activations (#EPW8) minimum.
  - Two, three, or four 12-core #EPW4 require twelve processor core activations (#EPW9) minimum.
- A minimum of four memory features per processor module is required.
- At least 50% of available memory or 128 GB of activations, whichever is larger, must be activated through a combination of features #EMAA and #EMAB.
- The language group is auto-selected based on geographic rules.

1.4.3 Power supply features

The following are the key power supply features:

- (#EB3M) AC Power Supply - 2000 W for Server (200 - 240 VAC):
  - One 200 - 240 V, 2000-watt ac power supply.
  - The power supply is configured in a one-plus-one or two-plus-two configuration to provide redundancy. Supported in rack models only.
  - To be operational, a minimum of four power supplies in the server are required. During a power supply failure, any of the power supplies can be exchanged without interrupting the operation of the system.
- (#EMXA) AC Power Supply Conduit for PCIe3 Expansion Drawer:
  - Provides two 320-C14 inlet electrical connections for two separately ordered ac power cords with C13 connector plugs. Conduit provides electrical power connection between two power supplies in the front of a PCIe Gen3 I/O Expansion Drawer (#EMX0) and two power cords that connect on the rear of the PCIe Gen3 I/O Expansion Drawer.
  - A maximum of two per I/O Expansion Drawer #EMX0.

1.5 Disk and media features

This section describes the key disk and media features.

1.5.1 SAS bays and storage backplane options

Clients have a choice of three storage features with eight SFF disk bays, four x 1.8-inch disk bays, and one DVD bay:

- The Dual Controller Disk Backplane with write cache supports RAID 0, 1, 5, 6, 10, 5T2, 6T2, and 10T2 (#EPVN). The pair of controllers handles all 12 integrated SAS bays and the DVD bay.

- The Dual Controller Disk Backplane without write cache supports RAID 0, 1, 5, 6, 10, 5T2, 6T2, and 10T2 (#EPVP). The pair of controllers handles all 12 integrated SAS bays and the DVD bay.
- The Split Disk Backplane (two single controllers without write cache supports RAID 0, 1, 5, 6, 10, and 10T2 (#EPVQ). This is the default configuration in e-config. Each of the two controllers handles four SFF-3 and two 1.8-inch SAS bays. One of the controllers handles the DVD bay.

Each of the three backplane options provides SFF-3 SAS bays in the system unit. These 2.5-inch or small form factor (SFF) SAS bays can contain SAS drives (HDD or SSD) mounted on a Gen3 tray or carrier. Thus, drives that are designated for SFF-1, or SFF-2 bays do not fit in an SFF-3 bay. All SFF-3 bays support concurrent maintenance or hot-plug capability. All three of these backplane options support HDDs or SSDs or a mixture of HDDs and SSDs in the SFF-3 bays. If you mix HDDs and SSDs, they must be in separate arrays (unless you are using the Easy Tier function).

For more information about disk controller options, see 1.6, “I/O drawers” on page 9.

1.5.2 Storage Backplane Integrated Easy Tier function

The Easy Tier function is provided with all three storage backplanes (#EPVN, #EPVP, #EPVQ). Conceptually, this function is like the Easy Tier function found in IBM Storage products such as the IBM System Storage DS8000®, IBM Storwize® V7000, and SAN Volume Controller. However, it is implemented just within the integrated Power Systems SAS controllers and the integrated SAS bays. Hot data is automatically moved to SSDs, and cold data is automatically moved to disk (HDD) in an AIX, Linux, or VIOS environment. No user application coding is required.

Clients commonly have this hot/cold characteristic for their data. It is typical for 10% - 20% of the data to be accessed 80% - 90% of the time. This is called the *hot data*. If you can get the hot data onto SSDs, it can dramatically improve the performance of I/O-bound applications. And by keeping the cold data on HDDs, the total cost per gigabyte of the solution can be minimized. You can end up with high I/O performance at a reasonable price. By avoiding the need for lots of HDD arms for performance, you can reduce the number of I/O drawers, maintenance, rack/floor space, and energy.

For more information about Easy Tier, see 1.6, “I/O drawers” on page 9.

1.5.3 DVD and boot devices

A slimline media bay that can hold a SATA DVD-RAM is included in the feature EPVN, EPVP, and EPVQ backplanes. The DVD drive is run by one of the integrated SAS controllers in the storage backplane, and a separate PCIe controller is not required.

1.6 I/O drawers

The following sections describe the available I/O drawer options for this server.

1.6.1 PCIe Gen3 I/O expansion drawer

PCIe Gen3 I/O expansion drawers (#EMX0) can be attached to the Power E850C to expand the number of full-high, hot-swap Gen3 slots available to the server. The maximum number of PCIe Gen3 I/O drawers depends on the number of processor modules physically installed. The maximum is independent of the number of processor core activations:

- Up to two PCIe3 drawers with two processor modules
- Up to three PCIe3 drawers with three processor modules
- Up to four PCIe3 drawers with four processor modules

Each PCIe Gen3 I/O expansion drawer adds a total of 10 PCIe Gen3 adapter slots to the server’s capabilities.

Figure 1-4 shows the front of the PCIe Gen3 I/O expansion drawer.

Figure 1-4 Front of the PCIe Gen3 I/O expansion drawer

Figure 1-5 shows the rear of the PCIe Gen3 I/O expansion drawer.

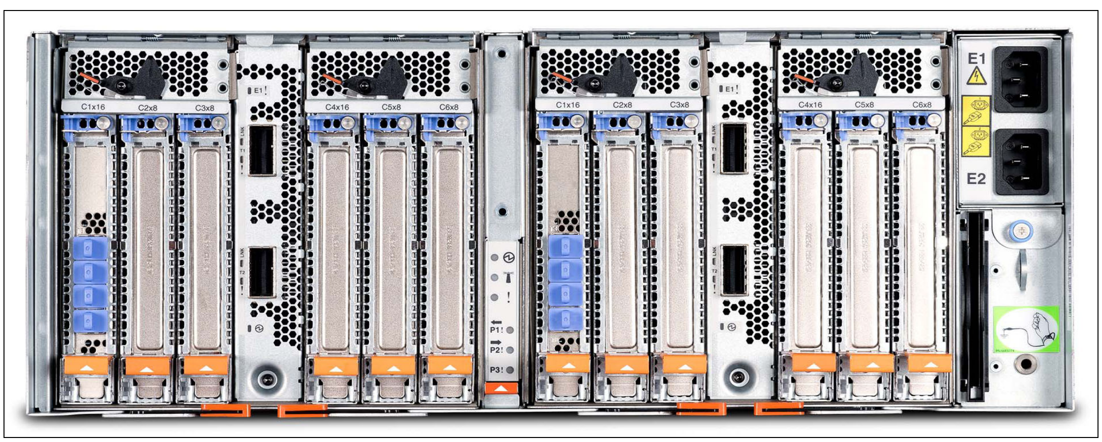

Figure 1-5 Rear of the PCIe Gen3 I/O expansion drawer

Table 1-3 shows PCIe adapters that are supported by the Power E850, either internally or in a PCIe Gen3 I/O expansion drawer.

Table 1-3 Feature of PCIe adapters supported internally or in a PCIe Gen3 I/O expansion drawer

Description	Feature
PCIe2 4-port 1 GbE Adapter	5899

Description	Feature
PCIe2 4-port 10 Gb + 1 GbE) SR+RJ45 Adapter	EN0S
PCIe2 4-port 10 Gb + 1 GbE) Copper SFP+RJ45 Adapter	EN0U
PCIe2 2-port 10/1 GbE BaseT RJ45 Adapter	EN0W
PCIe3 4-port 10 GbE SR Adapter	EN15
PCIe3 4-port 10 GbE SFP+ Copper Adapter	EN17
PCIe3 2-port 10 GbE NIC&RoCE SR Adapter	EC2N
PCIe3 2-port 10 GbE NIC&RoCE SFP+ Copper Adapter	EC38
PCIe3 2-port 40 GbE NIC RoCE QSFP+ Adapter	EC3B
PCIe3 Optical Cable Adapter for 4U CEC	EJ08
PCIe2 16 Gb 2-port Fibre Channel Adapter	EN0A
PCIe2 8 Gb 2-Port Fibre Channel Adapter	EN0G
PCIe2 8 Gb 4-port Fibre Channel Adapter	5729
PCIe2 4-port (10 Gb FCoE & 1 GbE) SR&RJ45	EN0H
PCIe2 4-port (10 Gb FCoE & 1 GbE) SFP+Copper&RJ45	EN0K
PCIe2 4-port (10 Gb + 1 GbE) SR+RJ45 Adapter	EN0S
PCIe2 4-port (10 Gb + 1 GbE) Copper SFP+RJ45 Adapter	EN0U
PCIe2 2-port 10 GbE SFN6122F Adapters	EC2J
PCIe2 2-port 10 GbE RoCE SR Adapter	EC30
PCIe2 2-port 10 GbE RoCE SFP+ Adapter	EC28
PCIe2 2-port 10 GbE SR Adapter	5287 (10 Gb FCoE)
PCIe3 12 GB Cache RAID SAS Adapter Quad-port 6 Gb x8	EJ0L
PCIe3 RAID SAS Adapter Quad-port 6 Gb x8	EJ0J
PCIe3 SAS Tape/DVD Adapter Quad-port 6 Gb x8	EJ10
PCIe1 SAS Tape/DVD Dual-port 3Gb x8 Adapter	EJ1P
PCIe3 12GB Cache RAID PLUS SAS Adapter Quad-port 6 Gb x8	57B1
PCIe 380 MB Cache Dual - x4 3 Gb SAS RAID Adapter	5805
PCIe Dual-x4 SAS Adapter	5901
8 Gigabit PCI Express Dual Port Fibre Channel Adapter	5735
4 Gigabit PCI Express Dual Port Fibre Channel Adapter	5774
PCIe3 Crypto Coprocessor no BSC 4767	EJ32
PCIe3 Crypto Coprocessor BSC-Gen3 4767	EJ33
4-Port 10/100/1000 Base-TX PCI Express Adapter	5717
2-Port 10/100/1000 Base-TX Ethernet PCI Express Adapter	5767
2 Port Async EIA-232 PCIe Adapter	EN27

Description	Feature
4 Port Async EIA-232 PCIe Adapter	5785
PCIe2 3D Graphics Adapter x1	EC42
IBM POWER® GXT145 PCI Express Graphics Accelerator	5748
PCIe2 4-port USB 3.0 Adapter	EC46
Note: ► Features #EC38 and #EC2N are not available in the following countries: Abu Dhabi, Algeria, Bahrain, Comoros, Djibouti, Dubai, Iraq, Kuwait, Lebanon, Libya, Malaysia, Morocco, Oman, Pakistan, Qatar, Saudi Arabia, Somalia, Tunisia, the United Arab Emirates, and Yemen.	

For more information about connecting PCIe Gen3 I/O expansion drawers to the Power E850C servers, see 2.11.1, “PCIe Gen3 I/O expansion drawer” on page 62.

1.6.2 EXP24S SFF Gen2-bay Drawer

The EXP24S SFF Gen2-bay Drawer (#5887) is an expansion drawer with 24 2.5-inch SFF SAS bays. The EXP24S supports up to 24 hot-swap SFF-2 SAS hard disk drives (HDDs) or solid-state drives (SSDs). It uses 2 EIA of space in a 19-inch rack. The EXP24S includes redundant ac power supplies and uses two power cords.

With AIX, Linux, and VIOS, you can order the EXP24S with four sets of six bays, two sets of 12 bays, or one set of 24 bays (mode 4, 2, or 1). Mode setting is done by IBM Manufacturing. If you need to change the mode after installation, ask your IBM support representative to see the following site:

http://w3.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS5121

The EXP24S SAS ports are attached to a SAS PCIe adapter or pair of adapters by using SAS YO or X cables.

Figure 1-6 shows the EXP24S SFF Gen2-bay drawer.

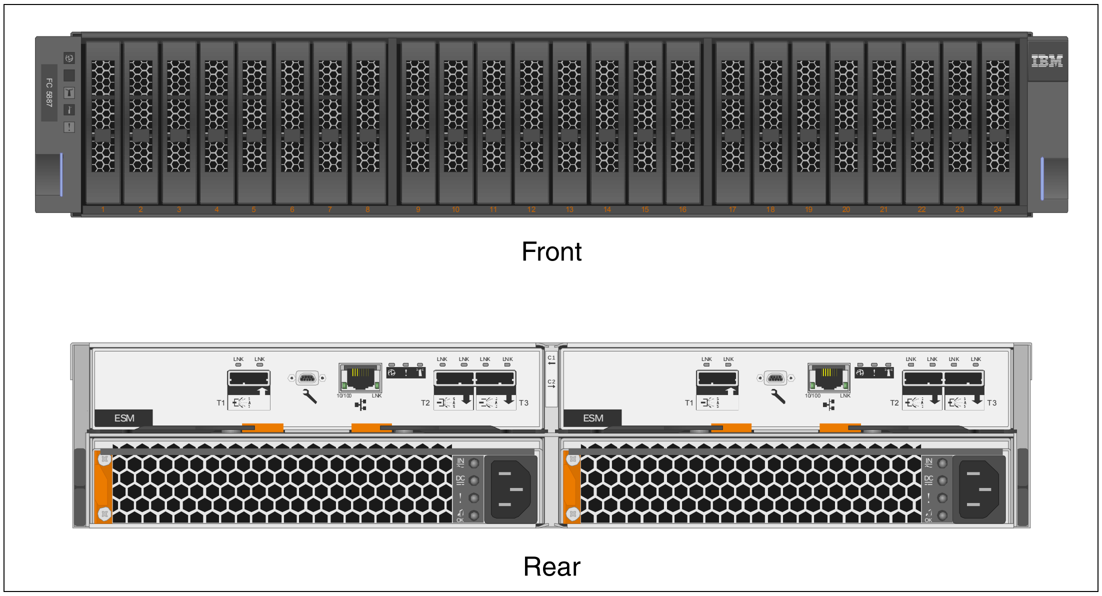

Figure 1-6 EXP24S SFF Gen2-bay drawer

For more information about connecting the EXP24S Gen2-bay drawer to the Power E850C server, see 2.12.1, “EXP24S SFF Gen2-bay Drawer” on page 69.

1.7 Build to order

You can do a *build to order* (also called *a la carte*) configuration by using the IBM Configurator for e-business (e-config). With it, you specify each configuration feature that you want on the system.

This method is the only configuration method for the Power E850C servers.

1.8 IBM Power Systems Solution Editions

The available IBM Solution Editions are described in the next sections.

1.8.1 IBM Power Systems Solution Editions for SAP HANA

IBM POWER8 servers are the ideal platform for running your SAP S/4HANA environment whether on premises or in a hybrid cloud environment. Power Systems Solution Edition for SAP HANA offers a competitively priced POWER8 server that is preconfigured to support the deployment of SAP HANA, SAP's in-memory database. This configuration acts as an optimized starter configuration that can be further customized for specific uses or anticipated growth and virtualization. The high performance of the POWER8 system, including 8-way multi-threading and high memory bandwidth, is ideal for fast in-memory processing such as with SAP HANA. This system results in strong price/performance results, with low total cost of ownership (TCO) and less complexity than competitive platforms. NVMe PCIe flash adapter for HANA on Power is also being introduced on the E850C and provides IT departments with faster data load, reduced maintenance downtime, and enhanced service levels.

This product represents a continuation of IBM's long-standing partnership with SAP. Power Systems Solution Edition enables clients to start fast and, to use SAP's phrase, “run simple.” It also provides the capacity to manage growth and the flexibility to integrate new functions, which helps minimize risks that are associated with adopting new business technologies.

The Solution Edition is available now for the Power 850C server with POWER8 processors.

Eligibility requirements for Power Systems Solutions Edition for SAP HANA:

- The client must possess new or existing software licenses for SAP HANA to be run on this server.
- Proof of purchase of the software (for example, a copy of the invoice) must be provided to IBM on request.
- SAP HANA software and Business Warehouse Solution software are available through an SAP authorized sales representative.

Power Systems Solution Editions for SAP HANA provide starting configurations that allow customers to quickly deploy HANA and grow with their business needs while meeting SAP design criteria. Power Systems servers are ideally suited to provide excellent SAP HANA performance and system reliability. The SAP HANA code that runs on the POWER architecture uses the exceptional POWER8 memory bandwidth, SIMD parallelization, and simultaneous multithreading (SMT).

Table 1-4 lists the solution edition offerings that are available by server.

Table 1-4 Solution edition offerings for SAP HANA

Server	Processor	Usage
E824	24 cores at 3.52 GHz	Suitable for HANA database that grow up to 0.5 TB
E850C	48 cores at 3.65 GHz	Suitable for HANA database that grow up to 2 TB
E870	40 cores at 4.19 GHz	Suitable for HANA database starting at 0.5 TB with expected growth beyond 1 TB
E870	80 cores at 4.19 GHz	Suitable for HANA database starting at 1 TB with expected growth beyond 2 TB

Note: Effective SAP database size (containing business data) is typically half the memory capacity of the server.

Power E850C system configuration

The Power E850C initial order for SAP HANA Solution Edition with 48 cores at 3.65 GHz processor and 2 TB memory must include a minimum of the items in Table 1-5.

Table 1-5 Features required for an initial order for SAP HANA

Feature number	Description
EPVR x 4	Solution edition for SAP HANA 3.65 GHz processor + 12 Activations
EPVS x 48	Power Processor Activation for #EPVR
EMBD x 2	1 TB DIMM Bundle for HANNA (8 x 128 GB DDR4 CDIMMs)
EMAJ x 8	256 GB Memory Activation for #EPVR and #EPVL
EPVN x 1	Dual Controller Disk Backplane with write cache

Feature number	Description
EN0S x 2	PCIe2 4-Port (10 Gb + 1 GbE) SR+RJ45 Adapter
5729 x 2	8 GB 4-port Fibre Channel Adapter
0837 x 1	SAN Load Source Specify
EB3M x 4	AC Power Supply - 2000W for Server (200-240 V ac)
6665 x 4	Power Cord 2.8 m (9.2 ft.), Drawer to Wall/IBM PDU (250V/10A)
9300/97xx	Language Group Specify
2147 x 1	Primary Operating System Indicator

Note:

- The minimum requirement of SUSE Linux Enterprise Server 11 (5639-S11) is 4-socket based licenses.
- The required SUSE Linux Enterprise Server 11 minimum level is SP3, plus the kernel update for SAP HANA on Power.
- SAN boot is required for SAP HANA solutions running on SUSE Linux Enterprise Server 11 SP3.
- Five units of STG Lab Services (6911-300) (#0003 x 5) are defaulted as applicable.
- Order configurations can be increased over this minimum configuration to meet customer requirements.

1.8.2 IBM Solution Edition for Healthcare

The Power E850C Solution Edition for Healthcare provides a cost-effective 48-core and 512 GB processor and memory activation feature package for eligible healthcare industry clients running approved ISV applications.

The Solution Edition for Healthcare minimum requirement is a 4-socket server (12-core 3.65 GHz E850C with 48 cores active), 512 GB memory (all active).

Note: Additional hardware components can be added as wanted following normal supported configuration rules.

For eligibility rules and registration of the Power Solution Edition for Healthcare by the sales channel, see the IBM Power Solution Editions website at:

http://www.ibm.com/systems/power/hardware/solutioneditions/eligibility.html

1.9 Management options

This section discusses the supported management interfaces for the servers.

A Hardware Management Console (HMC) is required for managing the IBM Power E850C. It has a set of functions that are necessary to manage the system:

- Creating and maintaining a multiple-partition environment

- Displaying a virtual operating system terminal session for each partition
- Displaying a virtual operator panel of contents for each partition
- Detecting, reporting, and storing changes in hardware conditions
- Powering managed systems on and off
- Acting as a service focal point for service representatives to determine an appropriate service strategy

The HMC could either be a hardware or a virtual appliance. Figure 1-7 shows the available configurations for a Power Systems HMCs infrastructure.

Note: Virtualization on the IBM Power E850C is also supported by the Integrated Virtualization Manager (IVM).

Multiple Power Systems servers can be managed by a single HMC. Each server can be connected to multiple HMC consoles to build extra resiliency into the management platform.

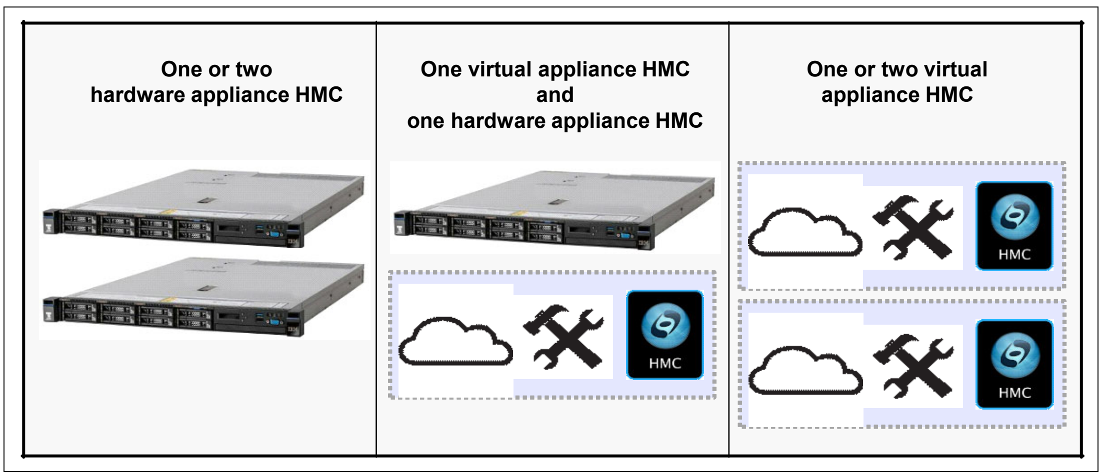

Figure 1-7 HMC configurations

Several hardware appliance HMC models are supported to manage POWER8 based systems. The 7042-CR9 is available for ordering at the time of writing, but you can also use one of the withdrawn models listed in Table 1-6.

Table 1-6 HMC models that support POWER8 processor technology-based servers

Type-model	Availability	Description
7042-C08	Withdrawn	IBM 7042 Model C08 Deskside Hardware Management Console
7042-CR5	Withdrawn	IBM 7042 Model CR5 Rack-Mounted Hardware Management Console
7042-CR6	Withdrawn	IBM 7042 Model CR6 Rack mounted Hardware Management Console
7042-CR7	Withdrawn	IBM 7042 Model CR7 Rack mounted Hardware Management Console
7042-CR8	Withdrawn	IBM 7042 Model CR8 Rack mounted Hardware Management Console
7042-CR9	Available	IBM 7042 Model CR9 Rack mounted Hardware Management Console

Existing HMC models 7042 can be upgraded to Licensed Machine Code Version 8 to support environments that might include IBM POWER6®, IBM POWER6+™, IBM POWER7®, IBM POWER7+™ and POWER8 processor-based servers.

If you want to support more than 254 partitions in total, the HMC requires a memory upgrade to a minimum of 4 GB.

HMC code and system firmware requirements:

- HMC base Licensed Machine Code Version 860, or later is required to support the Power E850C (8408-44E).
- System firmware level 860.10, or later.
- When using the POWER E850C server to access the cloud-based HMC Apps as a Service, the HMC must be at a minimum level of V8.60 with PTF MH01655.

You can download or order the latest HMC code from the Fix Central website:

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

For further information about managing the Power E850C serves from an HMC, see 2.13, “Hardware Management Console” on page 80.

1.10 System racks

The Power E850C server and its I/O drawers are designed to fit a standard 19-inch rack such as IBM features 7014-T00, 7014-T42, 7014-B42, 7965-94Y, #0551, and #0553.

The Power E850C IBM manufacturing rack integration can be ordered with a 36U (7014-T00) or a 42U (7014-T42) rack. A same serial-number model upgrade MES is placed in either feature #0551 or #0553 rack. This is done to ease and speed client installation, provide a more complete and higher-quality environment for IBM Manufacturing system assembly and testing, and provide a more complete shipping package. IBM Manufacturing does not support the use of other racks with the Power E850C initial system order or model upgrade. The 7014-T00 or feature #0551 is a 1.8-meter enterprise rack that provides 36U or 36 EIA of space. The 7014-T42 or feature #0553 is a 2-meter enterprise rack that provides 42U or 42 EIA of space.

Notes: 7014-B42 (42U) support only, cannot be ordered from IBM and 7965-94Y (42U) support only by field merge.

Compared to the existing Power E850, the E850C server draws additional power. Using previously provided IBM power distribution unit (PDU) features #7188, #7109, and #7196 reduces the number of E850C servers and other equipment that can be held most efficiently in a rack. The new high-function PDUs provide more electrical power per PDU and thus offer better “PDU footprint” efficiency. In addition, they are intelligent PDUs that provide insight to actual power usage by receptacle, and also provide remote power on/off capability for easier support by individual receptacle.

The new PDUs are features #EPTJ, #EPTL, #EPTN, and #EPTQ available on the 7014-T00, 7014-T42, and 7965-94Y racks. For more information, see 1.10.6, “The ac power distribution unit and rack content” on page 19.

Installing in non-IBM racks: The client is responsible for ensuring that the installation of the server in the preferred rack or cabinet results in a configuration that is stable, serviceable, safe, and compatible with the server requirements for power, cooling, cable management, weight, and rail security.

1.10.1 IBM 7014 model T00 rack

The 1.8-meter (71-inch) model T00 is compatible with past and present IBM Power Systems. The T00 rack provides these features:

- 36U (EIA units) of usable space.
- Optional removable side panels.
- Optional highly perforated front door.
- Optional side-to-side mounting hardware for joining multiple racks.
- Standard business black or optional white color in OEM format.
- Increased power distribution and weight capacity.
- Supports both ac and dc configurations.
- The rack height is increased to 1926 mm (75.8 in.) if a power distribution panel is fixed to the top of the rack.
- The #6068 feature provides a cost effective plain front door.
- Weights are as follows:
  - T00 base empty rack: 244 kg (535 lb.).
  - T00 full rack: 816 kg (1795 lb.).
  - Maximum Weight of Drawers is 572 kg (1260 lb.).
  - Maximum Weight of Drawers in a zone 4 earthquake environment is 490 kg (1080 lb.). This equates to 13.6 kg (30 lb.)/EIA.

1.10.2 IBM 7014 model T42 rack

The 2.0-meter (79.3-inch) Model T42 addresses the requirement for a tall enclosure to house the maximum amount of equipment in the smallest possible floor space. The following features differ in the model T42 rack from the model T00:

- The T42 rack has 42U (EIA units) of usable space (6U of additional space).
- The model T42 supports ac power only.
- Weights are as follows:
  - T42 base empty rack: 261 kg (575 lb.).
  - T42 full rack: 930 kg (2045 lb.).
- The feature #ERG7 provides an attractive black full-height rack door. The door is steel, with a perforated flat front surface. The perforation pattern extends from the bottom to the top of the door to enhance ventilation and provide some visibility into the rack.
- The feature #6069 provides a cost-effective plain front door.
- The feature #6249 provides a special acoustic door.

Notes: The Power E850C server with 7014-T42 rack includes one standard PDU.

1.10.3 IBM 7965 model 94Y rack

The 2.0-meter (79-inch) model 7965-94Y is compatible with past and present IBM 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 fit easily on standard 24-inch floor tiles.

The IBM 42U Slim Rack has a lockable perforated front steel door, providing 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 located horizontally, but they each use 1U of space in this position.

A minimum of one of the following features is required:

- Feature #ER2B allows you to reserve 2U of space at the bottom of the rack.
- Feature #ER2T allows you to reserve 2U of space at the top of the rack.

1.10.4 Feature code #0551 rack

The 1.8-meter rack (#0551) is a 36U (EIA units) rack. The rack that is delivered as #0551 is the same rack that is delivered when you order the 7014-T00 rack. The included features might differ. Several features that are delivered as part of the 7014-T00 must be ordered separately with feature code #0551. The #0551 is available with an initial order of a Power E850C server.

1.10.5 Feature code #0553 rack

The 2.0-meter rack (#0553) is a 42U (EIA units) rack. The rack that is delivered as #0553 is the same rack that is delivered when you order the 7014-T42. The included features might differ. Several features that are delivered as part of the 7014-T42 or must be ordered separately with the #0553. The #0553 is available with an initial order of a Power E850C server.

1.10.6 The ac power distribution unit and rack content

For rack models T00, T42, and the slim 94Y, 12-outlet PDUs are available. The following PDUs are available:

- PDUs Universal UTG0247 Connector (#7188)
- Intelligent PDU+ Universal UTG0247 Connector (#7109)
- High Function 9xC19 PDU: Switched, Monitoring (#EPTJ)
- High Function 9xC19 PDU 3-Phase: Switched, Monitoring (#EPTL)
- High Function 12xC13 PDU: Switched, Monitoring (#EPTN)
- High Function 12xC13 PDU 3-Phase: Switched, Monitoring (#EPTQ)

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

Feature #7109

Intelligent PDU with Universal UTG0247 Connector is for an intelligent ac power distribution unit (PDU+) that allows the user to monitor the amount of power being used by the devices that are plugged in to this PDU+. This ac power distribution unit provides 12 C13 power outlets. It receives power through a UTG0247 connector. It can be used for many different countries and applications by varying the PDU to Wall Power Cord, which must be ordered separately. Each PDU requires one PDU to Wall Power Cord. Supported power cords include the following features: 6489, 6491, 6492, 6653, 6654, 6655, 6656, 6657, and 6658.

Feature #7188

Power Distribution Unit mounts in a 19-inch rack and provides 12 C13 power outlets. Feature 7188 has six 16A circuit breakers, with two power outlets per circuit breaker. System units and expansion units must use a power cord with a C14 plug to connect to the feature 7188. One of the following power cords must be used to distribute power from a wall outlet to the feature 7188: feature 6489, 6491, 6492, 6653, 6654, 6655, 6656, 6657, or 6658.

Feature #EPTJ

This feature is an intelligent, switched 200 - 240 volt ac PDU with nine C19 receptacles on the front of the PDU. The PDU is mounted on the rear of the rack, making the nine C19 receptacles easily accessible. Each receptacle has a 20 amp circuit breaker. Depending on country wiring standards, the PDU is single-phase or three-phase wye. See three-phase feature #EPTK or #EPTL for countries that do not use wye wiring. The PDU can be mounted vertically in rack side pockets, or it can be mounted horizontally. If mounted horizontally, it uses 1 EIA (1U) of rack space. See feature #EPTH for horizontal mounting hardware. Device power cords with a C20 plug connect to C19 PDU receptacles and are ordered separately. One country-specific wall power cord is also ordered separately and attaches to a UTG524-7 connector on the front of the PDU.

Supported power cords include features #6489, #6491, #6492, #6653, #6654, #6655, #6656, #6657, #6658, and #6667.

Feature #EPTL

This feature is an intelligent, switched 208 volt 3-phase ac PDU with nine C19 receptacles on the front of the PDU. The PDU is mounted on the rear of the rack, making the nine C19 receptacles easily accessible. Each receptacle has a 20 amp circuit breaker. The PDU can be mounted vertically in rack side pockets, or it can be mounted horizontally. If mounted horizontally, it uses 1 EIA (1U) of rack space. See feature #EPTH for horizontal mounting hardware. Device power cords with a C20 plug connect to C19 PDU receptacles and are ordered separately. One wall power cord is provided with the PDU (no separate feature number) and has an IEC60309 60A plug (3P+G). The PDU supports up to 48 amps. Two RJ45 ports on the front of the PDU enable the client to monitor each receptacle's electrical power usage and to remotely switch any receptacle on or off. The PDU is shipped with a generic PDU password. IBM strongly urges clients to change it upon installation.

Feature #EPTN

This feature is an intelligent, switched 200 - 240 volt ac PDU with twelve C13 receptacles on the front of the PDU. The PDU is mounted on the rear of the rack, making the twelve C13 receptacles easily accessible. Each receptacle has a 20 amp circuit breaker. Depending on country wiring standards, the PDU is single-phase or three-phase wye. See three-phase feature #EPTK or #EPTL for countries that do not use wye wiring. The PDU can be mounted vertically in rack side pockets, or it can be mounted horizontally. If mounted horizontally, it uses 1 EIA (1U) of rack space. See feature #EPTH for horizontal mounting hardware. Device power cords with a C14 plug connect to C13 PDU receptacles and are ordered separately.

One country-specific wall power cord is also ordered separately and attaches to a UTG524-7 connector on the front of the PDU.

Supported power cords include features #6489, #6491, #6492, #6653, #6654, #6655, #6656, #6657, #6658, and #6667. Two RJ45 ports on the front of the PDU enable the client to monitor each receptacle's electrical power usage and to remotely switch any receptacle on or off. The PDU is shipped with a generic PDU password. IBM strongly urges clients to change it upon installation.

Feature #EPTQ

This feature is an intelligent, switched 208 volt 3-phase ac PDU with twelve C13 receptacles on the front of the PDU. The PDU is mounted on the rear of the rack, making the twelve C13 receptacles easily accessible. Each receptacle has a 20 amp circuit breaker. The PDU can be mounted vertically in rack side pockets, or it can be mounted horizontally. If mounted horizontally, it uses 1 EIA (1U) of rack space. See feature EPTH for horizontal mounting hardware.

Device power cords with a C14 plug connect to C13 PDU receptacles and are ordered separately. One wall power cord is provided with the PDU (no separate feature number) and has an IEC60309 60A plug (3P+G). The PDU supports up to 48 amps. Two RJ45 ports on the front of the PDU enable the client to monitor each receptacle's electrical power usage and to remotely switch any receptacle on or off. The PDU is shipped with a generic PDU password. IBM strongly urges clients to change it upon installation.

For detailed power cord requirements and power cord feature codes, see the IBM Power Systems Hardware IBM Knowledge Center website:

https://www.ibm.com/support/knowledgecenter/TI0002C/p8had/p8had_85x_86x_kickoff.htm

> **Power cord:** Ensure that the appropriate power cord feature is configured to support the power being supplied.

1.10.7 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.

Three to four service personnel are required to manually remove or insert a system unit into a rack, depending on its dimensions, weight, and content. To avoid the need for this many people to assemble at a client site for a service action, a lift tool can be useful. Similarly, if you have chosen to install this customer setup (CSU) system, similar lifting considerations apply. The Power E850C has a maximum weight of 69 kg (152 lbs). However, by temporarily removing the power supplies, fans, and RAID assembly, the weight is easily reduced to 55 kg (121 lbs).

When lowering the Power E850C onto its rails in the rack, the server must be tilted on one end about 15 degrees so that the pins on the server enclosure fit onto the rails. This process equates to lifting one end of the server about 4 cm (1.6 in.). This can be done using a tip plate on a lift tool or manually adjusting the load on a lift tool or tilting during the manual lift. Consider the optional feature #EB2Z Lift Tool.

2

Architecture and technical overview

This chapter describes the overall system architecture for the IBM Power System E850C (8408-44E). The 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 best performance.

Always do the performance sizing at the application workload environment level and evaluate performance by using real-world performance measurements and production workloads.

This chapter includes the following sections:

- Physical system design
- Logical system design
- The IBM POWER8 processor
- Memory subsystem
- Memory Custom DIMMs
- Capacity on Demand
- System buses
- Internal I/O connections
- PCI adapters
- Flash storage adapters
- External I/O subsystems
- External disk subsystems
- Hardware Management Console
- Operating system support
- Energy management

2.1 Physical system design

The Power E850C is configured as a 4U (4 EIA) rack-mount server. It is designed to fit into an industry standard 19-inch rack. The server capabilities can be extended by adding I/O expansion drawers to the system. These include the 4U (4 EIA) PCIe Gen3 I/O expansion drawer and the 2U (2 EIA) EXP24S I/O drawer for added local storage capacity.

Table 2-1 lists the dimensions of the system components for installation planning.

Table 2-1 Dimensions of the system components for installation planning

Dimension	Power E850C system node	PCIe Gen3 I/O expansion drawer
Width	448 mm (17.6 in.)	482 mm (19 in.)
Depth	776 mm (30.6 in.)	802 mm (31.6 in.)
Height	175 mm (6.9 in.), 4 EIA units	175 mm (6.9 in.), 4 EIA units
Weight	69.0 kg (152 lb)	54.4 kg (120 lb)

A Power E850C server should be installed into a rack that has been certified and tested to support the system, such as the IBM 7014-T00 rack.

The front of the system node contains the front fans (five) for the system. It also provides access to the internal storage bays, the operator panel, and the optional DVD-RAM drive. Figure 2-1 shows a front view of the Power E850C server.

Figure 2-1 Front view of the Power E850C server

The rear of the system provides access to the internally installed PCIe cards, the service processor connections, and the power supply units. Figure 2-2 shows a rear view of the Power E850C server showing power supply detail.

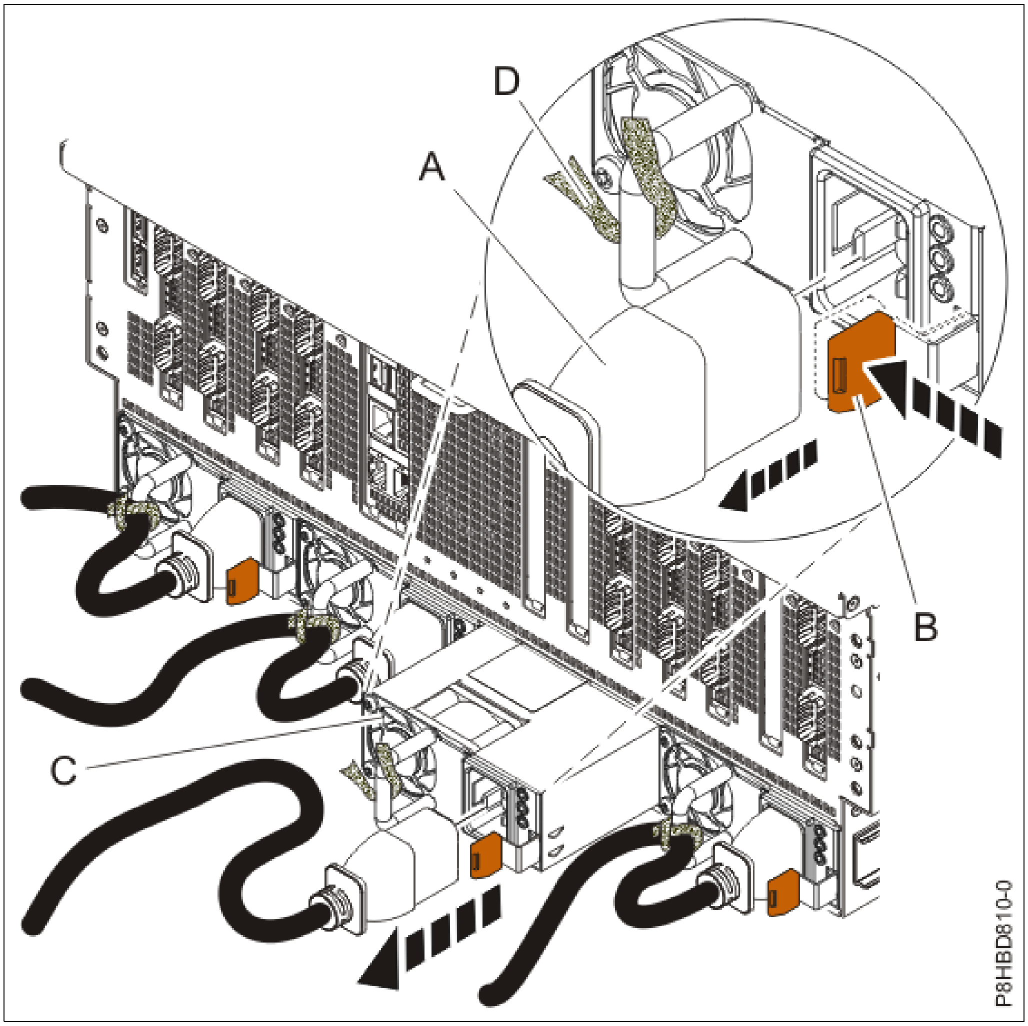

Figure 2-2 Rear view of the Power E850C server

Figure 2-3 shows the internal layout of major components of the Power E850C server within the 4U chassis.

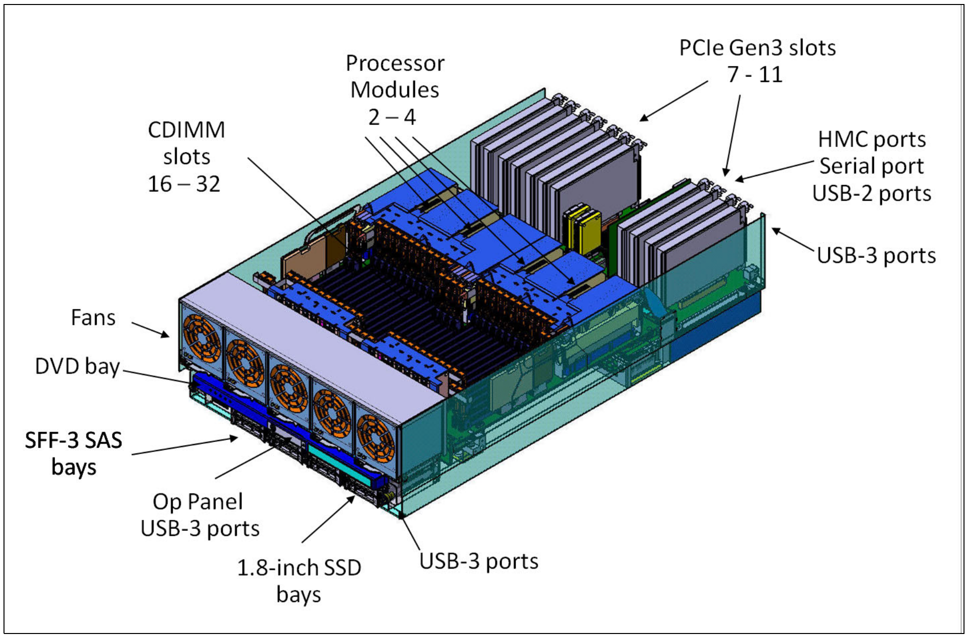

Figure 2-3 Layout view of the Power E850C showing component placement

2.2 Logical system design

This section contains logical diagrams of the Power E850C server in different configurations. It also covers some factors to consider when configuring a new or upgraded Power E850C server.

The Power E850C can be configured with two, three, or four processor modules installed. The number of memory CDIMM slots and the number of PCIe adapter slots that can be used varies based on the number of processor modules installed. Table 2-2 shows the available options.

Table 2-2 Usable memory CDIMM slots and PCIe adapter slots in the Power E850C

Number of processor modules	Memory CDIMM slots	PCIe adapter slots in system node
Two	16	Four x16 and three x8 (seven total)
Three	24	Six x16 and three x8 (nine total)
Four	32	Eight x16 and three x8 (11 total)

The number of available PCIe adapters supported can be increased by adding PCIe Gen3 I/O expansion drawers to the server. The Power E850C can support up to one PCIe I/O expansion drawer per processor module.

The following factors can influence the placement of PCIe cards in the Power E850C server:

- All PCIe slots in the system node are SR-IOV capable.

- All PCIe x16 slots in the system node have dedicated bandwidth (16 lanes of PCIe Gen 3 bandwidth).
- One of the PCIe x8 slots (C11) has dedicated bandwidth (eight lanes of PCIe Gen3 bandwidth). The other two PCIe x8 slots (C6 and C7) share bandwidth (by using the PEX unit, and also shared with the USB 3.0 controller).
- All PCIe x16 slots in the system node are CAPI capable.
- At least one PCIe Ethernet adapter is required on the server by IBM to ensure proper manufacture and test of the server. One of the x8 PCIe slots is used for this required adapter, identified as the C11 slot.

Figure 2-4 shows the logical design of the Power E850C server with two processor modules installed.

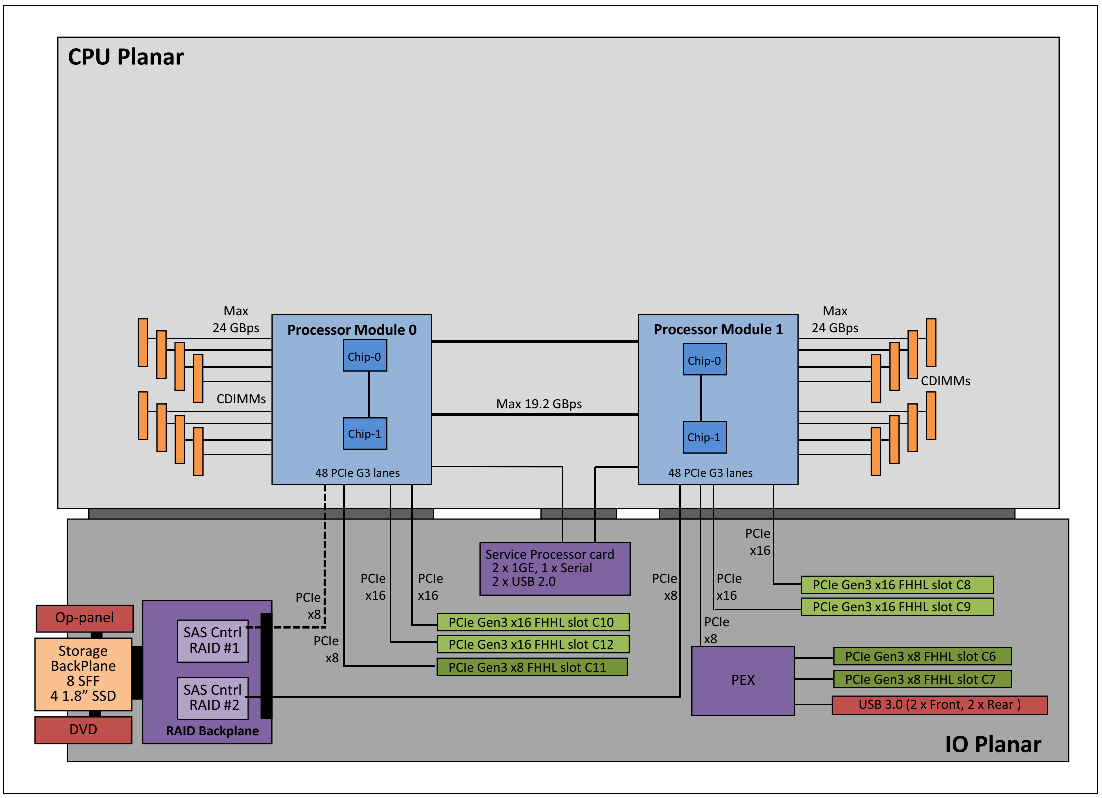

Figure 2-4 Logical diagram for a Power E850C with two processor modules installed

Figure 2-5 shows the logical design of the Power E850C server with three processor modules installed.

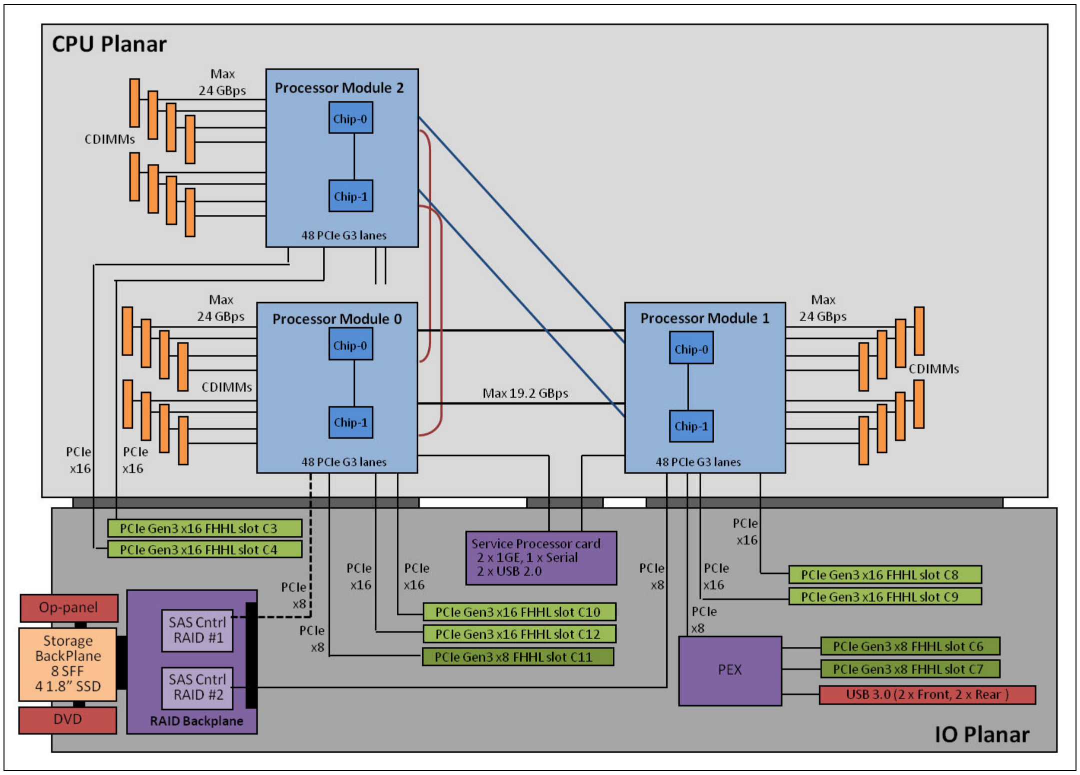

Figure 2-5 Logical diagram for a Power E850C with three processor modules installed

Figure 2-6 shows the logical design of the Power E850C server with four processor modules installed.

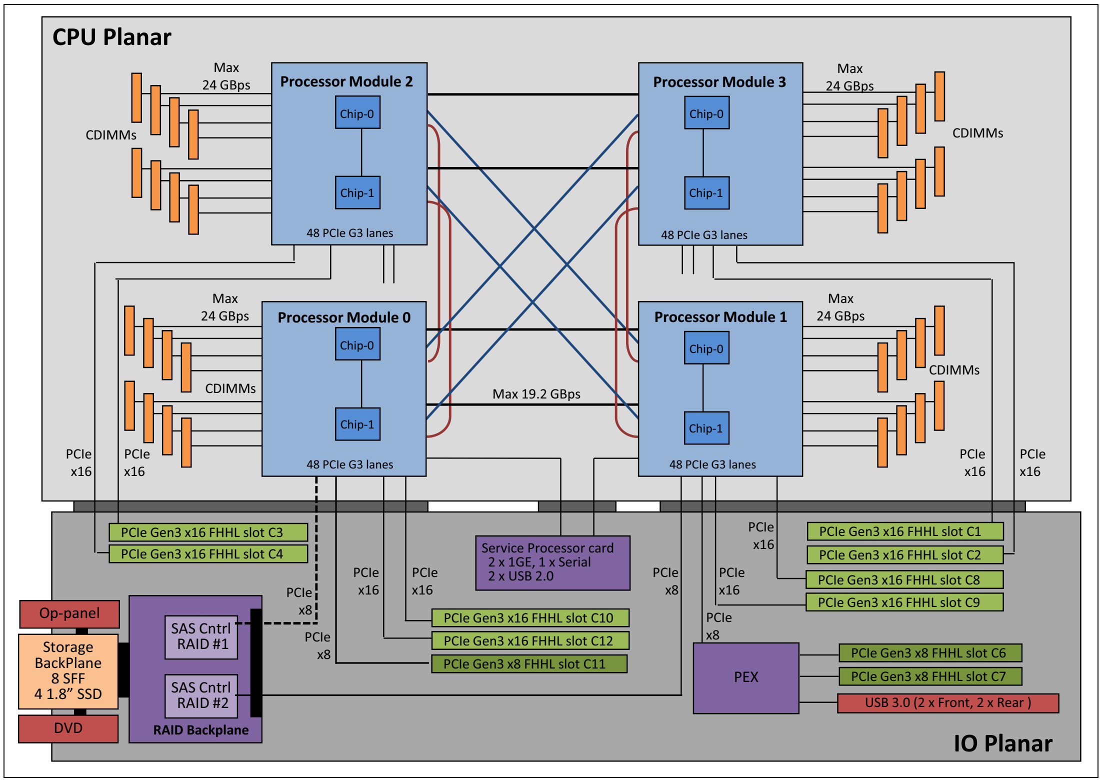

Figure 2-6 Logical diagram for a Power E850C server with four processor modules installed

2.3 The IBM POWER8 processor

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. The chip contains 12 cores, two memory controllers, and an interconnection system that connects all components within the chip. On some systems, only 6, 8, 10, or 12 cores per processor might be available to the server. Each core has 512 KB of L2 cache, and all cores share 96 MB of L3 embedded dynamic random access memory (eDRAM). The interconnect also extends through module and board technology to other POWER8 processors in addition to DDR3 memory and various I/O devices.

POWER8 systems use memory buffer chips to interface between the POWER8 processor and DDR3 or DDR4 memory. Each buffer chip also includes an L4 cache to reduce the latency of local memory accesses.

For an in-depth discussion on the POWER8 processor, see the *IBM Power System E850 Technical Overview and Introduction*, REDP-5222 at:

http://www.redbooks.ibm.com/abstracts/redp5222.html

2.4 Memory subsystem

The Power E850C can have two, three, or four processor modules installed per server. Each processor module enables eight DDR4 CDIMM slots capable of supporting 16 GB, 32 GB, 64 GB, and 128 GB CDIMMs. A server with four processor modules can support up to 4 TB of installed memory, a server with three processor modules installed can support up to 3 TB of installed memory, and a server with two processor modules installed can support up to 2 TB of installed memory.

The memory on the systems is Capacity on Demand-capable, allowing for the purchase of additional memory capacity and dynamically activate it when needed. It is required that at least 50% of the installed memory capacity is active.

The Power E850C server supports an optional feature called *Active Memory Expansion* (#4798). This features allows the effective maximum memory capacity to be much larger than the true physical memory maximum. Sophisticated compression and decompression of memory content by using the POWER8 processor along with a dedicated coprocessor present on each POWER8 processor can provide memory expansion up to 100% or more. This ratio depends on the workload type and its memory usage. As an example, a server with 256 GB of RAM physically installed can effectively be expanded over 512 GB of RAM. This approach can enhance virtualization and server consolidation by allowing a partition to do more work with the same physical amount of memory or allowing a server to run more partitions and do more work with the same physical amount of memory. The processor resource used to expand memory is part of the processor entitlement assigned to the partition enabling Active Memory Expansion.

Active Memory Expansion is compatible with all AIX partitions. Each individual AIX partition can have Active Memory Expansion enabled or disabled. Control parameters set the amount of expansion wanted in each partition to find a balance of memory expansion and processor utilization. A partition needs to be rebooted to turn on Active Memory Expansion.

A planning tool is included within AIX that allows you to sample actual workloads and estimate the level of expansion and processor usage expected. This tool can be run on any Power Systems server running PowerVM as a hypervisor. A one-time, 60-day trial of Active Memory Expansion is available on each server to confirm the estimated results. You can request the trial activation code on the IBM Power Systems Capacity on Demand website:

http://www.ibm.com/systems/power/hardware/cod

To activate Active Memory Expansion on a Power E850C server, the chargeable feature code #4798 must be ordered, either as part of the initial system order or as an MES upgrade. A software key is provided which is applied to the server. There is no need to reboot the system.

2.5 Memory Custom DIMMs

Custom DIMMs (CDIMMs) are innovative memory DIMMs that house industry-standard DRAM memory chips and a set of components that allow for higher bandwidth, lower latency communications, and increased availability. CDIMMs include these components:

- Memory Scheduler
- Memory Management (RAS Decisions and Energy Management)
- Memory Buffer

By adopting this architecture for the memory DIMMs, several decisions and processes regarding memory optimizations are run internally into the CDIMM. This technique saves bandwidth and allows for faster processor-to-memory communications. It also allows for a more robust RAS. For more information, see Chapter 4, “Reliability, availability, and serviceability” on page 101.

2.5.1 CDIMM design

The CDIMMs exist in two different form factors:

- A 152 SDRAM design that is named the *Tall CDIMM*
- An 80 SDRAM design that is named the *Short CDIMM*

Each design is composed of multiple 4 GB SDRAM devices depending on its total capacity. The CDIMMs for the Power E850C server are short CDIMMs. Tall CDIMMs from other Enterprise Systems such as the Power E870 and Power E880 are not compatible with the Power E850C server.

The Power E850C supports CDIMMs in 16 GB, 32 GB, 64 GB, and 128 GB capacities. Each CDIMM incorporates a 16 MB Memory Buffer, also known as *L4 cache*. The L4 cache is built on eDRAM technology (same as the L3 cache), which has a lower latency than regular SRAM. Each CDIMM has 16 MB of L4 cache and a fully populated Power E850C server with four processor modules has 512 MB of L4 Cache. The L4 Cache performs several functions that have direct impact on performance and bring a series of benefits for the Power E850C 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 being 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.

2.5.2 Memory placement rules

For the Power E850C, each memory feature code provides a single CDIMM. These memory features must be ordered in pairs of the same memory feature. Both CDIMMs of a CDIMM pair must be installed in CDIMM slots supporting one processor. Different size pairs can be mixed on the same processor. However, for optimal performance ensure that all CDIMM pairs that are connected to a processor are of the same capacity. Also, ensure that the number of memory CDIMM pairs is the same on each processor module of a system.

System performance improves as more CDIMM pairs match. System performance also improves as more CDIMM slots are filled, because this configuration increases the memory bandwidth available. Therefore, if 256 GB of memory is required, using sixteen 16 GB CDIMMs would offer better performance than using eight 32 GB CDIMMs. This configuration allows memory access in a consistent manner and typically results in the best possible performance for your system. You should account for any plans for future memory upgrades when you decide which memory feature size to use at the time of the initial system order.

A minimum of four CDIMM slots must be populated for each installed processor module. For a Power E850C with two processor modules, eight CDIMM slots must be populated. For a Power E850C with three processor modules installed, 12 CDIMM slots must be populated, and for a Power E850C with four processor modules installed, 16 CDIMM slots must be populated.

All the memory CDIMMs are capable of capacity upgrade on demand and must have a minimum of 50% of its physical capacity activated. For example, if a Power E850C has 512 GB of memory installed, a minimum of 256 GB would need to be activated. A minimum of 128 GB of memory must be activated on each system. For more information about Capacity on Demand and activation requirements, see 2.6, “Capacity on Demand” on page 40.

For the Power E850C, the following memory options are orderable:

- 16 GB CDIMM, 1600 MHz DDR4 DRAM (#EM8P)
- 32 GB CDIMM, 1600 MHz DDR4 DRAM (#EM8Q)
- 64 GB CDIMM, 1600 MHz DDR4 DRAM (#EM8R)
- 128 GB CDIMM, 1600 MHz DDR4 DRAM (#EM8S)

Note: DDR4 DRAMs provide the same memory throughput as DDR3.

Table 2-3 summarizes the minimum and maximum CDIMM and memory requirements for the Power E850C server.

Table 2-3 Minimum and maximum memory requirements

Number of processor modules installed	Minimum number of CDIMMs	Minimum memory capacity (16 GB CDIMMs)	Maximum number of CDIMMs	Maximum memory capacity (128 GB CDIMMs)
Two	8	128 GB	16	2048 GB (2 TB)
Three	12	192 GB	24	3072 GB (3 TB)
Four	16	256 GB	32	4096 GB (4 TB)

The following are the basic rules for memory placement:

- Each feature code provides a single memory CDIMM.
- Memory CDIMMs must be ordered and installed in matching pairs.
- For each installed processor, there must be at least four CDIMMs populated.
- There is a maximum of eight memory CDIMMs per installed processor.
- At least 50% of the installed memory must be activated by using memory activation features.
- DDR4 memory modules cannot be mixed with DDR3. If a DDR4 feature is used, all the memory on the system must be DDR4.

2.5.3 CDIMM plugging order

Each processor module has two memory controllers. Each of these memory controllers connects to four CDIMM slots within the system. The physical connections and location codes of the memory slots are shown in Figure 2-7.

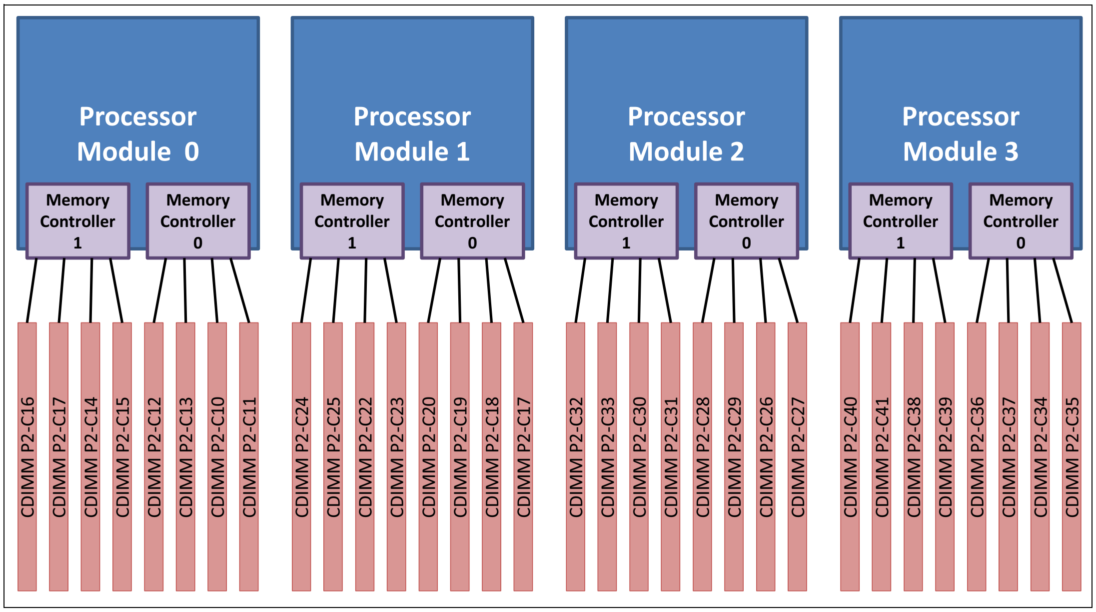

Figure 2-7 Memory slot physical connections and location codes for the Power E850C

Table 2-4 shows the recommended plugging order for CDIMMs in a Power E850C server with two processor modules installed to ensure optimal performance and CDIMM size flexibility. Each number represents two matching capacity CDIMMs in two adjacent CDIMM slots.

Table 2-4 Memory plugging order for Power E850C with two processor modules

Processor Module 0				**Processor Module 1**			
Memory Controller 1		Memory Controller 0		Memory Controller 1		Memory Controller 0	
C16 C17	C14 C15	C12 C13	C10 C11	C24 C25	C22 C23	C20 C21	C18 C19
1	5	3	7	2	6	4	8

Table 2-4 shows the following pairs:

- First CDIMM pair is identical and plugged into P2-C16 and P2-C17
- Second CDIMM pair is identical and plugged into P2-C24 and P2-C25
- Third CDIMM pair is identical and plugged into P2-C12 and P2-C13
- Fourth CDIMM pair is identical and plugged into P2-C20 and P2-C21
- Fifth CDIMM pair is identical and plugged into P2-C14 and P2-C15
- Sixth CDIMM pair is identical and plugged into P2-C22 and P2-C23
- Seventh CDIMM pair is identical and plugged into P2-C10 and P2-C11
- Eighth CDIMM pair is identical and plugged into P2-C18 and P2-C19

Table 2-5 shows the recommended plugging order for CDIMMs in a Power E850C server with three processor modules installed to ensure optimal performance and CDIMM size flexibility. Each number represents two matching capacity CDIMMs in two adjacent CDIMM slots.

Table 2-5 Memory plugging order for Power E850C with three processor modules

Processor Module 0				**Processor Module 1**			
Memory Controller 1		Memory Controller 0		Memory Controller 1		Memory Controller 0	
C16 C17	C14 C15	C12 C13	C10 C11	C24 C25	C22 C23	C20 C21	C18 C19
1	7	4	10	2	8	5	11

Processor Module 0			
Memory Controller 1		Memory Controller 0	
C32 C33	C30 C31	C28 C29	C26 C27
3	9	6	12

Table 2-5 shows the following pairs:

- First CDIMM pair is identical and plugged into P2-C16 and P2-C17
- Second CDIMM pair is identical and plugged into P2-C24 and P2-C25
- Third CDIMM pair is identical and plugged into P2-C32 and P2-C33
- Fourth CDIMM pair is identical and plugged into P2-C12 and P2-C13
- Fifth CDIMM pair is identical and plugged into P2-C20 and P2-C21
- Sixth CDIMM pair is identical and plugged into P2-C28 and P2-C29
- Seventh CDIMM pair is identical and plugged into P2-C14 and P2-C15
- Eighth CDIMM pair is identical and plugged into P2-C22 and P2-C23
- Ninth CDIMM pair is identical and plugged into P2-C30 and P2-C31
- Tenth CDIMM pair is identical and plugged into P2-C10 and P2-C11
- Eleventh CDIMM pair is identical and plugged into P2-C18 and P2-C19
- Twelfth CDIMM pair is identical and plugged into P2-C26 and P2-C27

Table 2-6 shows the recommended plugging order for CDIMMs in a Power E850C server with four processor modules installed to ensure optimal performance and CDIMM size flexibility. Each number represents two matching capacity CDIMMs in two adjacent CDIMM slots.

Table 2-6 Memory plugging order for Power E850C with four processor modules

Processor Module 0				**Processor Module 1**			
Memory Controller 1		Memory Controller 0		Memory Controller 1		Memory Controller 0	
C16 C17	C14 C15	C12 C13	C10 C11	C24 C25	C22 C23	C20 C21	C18 C19
1	9	5	13	2	10	6	14

Processor Module 0				**Processor Module 1**			
Memory Controller 1		Memory Controller 0		Memory Controller 1		Memory Controller 0	
C32 C33	C30 C31	C28 C29	C26 C27	C40 C41	C38 C39	C36 C37	C34 C35
3	11	7	15	4	12	8	16

Table 2-6 on page 34 shows the following pairs:

- First CDIMM pair is identical and plugged into P2-C16 and P2-C17
- Second CDIMM pair is identical and plugged into P2-C24 and P2-C25
- Third CDIMM pair is identical and plugged into P2-C32 and P2-C33
- Fourth CDIMM pair is identical and plugged into P2-C40 and P2-C41
- Fifth CDIMM pair is identical and plugged into P2-C12 and P2-C13
- Sixth CDIMM pair is identical and plugged into P2-C20 and P2-C21
- Seventh CDIMM pair is identical and plugged into P2-C28 and P2-C29
- Eighth CDIMM pair is identical and plugged into P2-C36 and P2-C37
- Ninth CDIMM pair is identical and plugged into P2-C14 and P2-C15
- Tenth CDIMM pair is identical and plugged into P2-C22 and P2-C23
- Eleventh CDIMM pair is identical and plugged into P2-C30 and P2-C31
- Twelfth CDIMM pair is identical and plugged into P2-C38 and P2-C39
- Thirteenth CDIMM pair is identical and plugged into P2-C10 and P2-C11
- Fourteenth CDIMM pair is identical and plugged into P2-C18 and P2-C19
- Fifteenth CDIMM pair is identical and plugged into P2-C26 and P2-C27
- Sixteenth CDIMM pair is identical and plugged into P2-C34 and P2-C35

2.5.4 Memory activations

Several types of Capacity on Demand capability are available for processors and memory on the Power E850C server. All the memory CDIMMs in a Power E850C system are capable of capacity upgrade on demand, and must have a minimum of 50% of their physical capacity activated. A minimum of 128 GB of memory must be activated on each system. The remaining installed memory capacity can be activated as Capacity on Demand, either permanently or temporarily.

Initial activations of memory resources must meet the minimum requirement of 50% of all installed memory, or 128 GB, whichever is higher. The maximum initial order for activations is the entire installed memory capacity.

Initial memory activations can be ordered in quantities of 1 GB (#EMAA) and 100 GB (#EMAB). This memory capacity remains permanently active.

Any memory resources that are installed but not activated as part of the initial order are available for future use of the Capacity on Demand features of the Power E850C.

For more information about Capacity on Demand and activation requirements, see 2.6, “Capacity on Demand” on page 40.

2.5.5 Memory throughput

The peak memory and I/O bandwidths per system node have increased over 300% compared to the previous generation POWER7 processor-based servers, providing the next generation of data-intensive applications with a platform capable of handling the needed amount of data.

> **Warning:** All bandwidth figures listed in the section are theoretical maximums, and are based on the nominal clock speeds of the processors listed. Because clock speeds can vary depending on power-saving mode and server load, these bandwidths are provided for information only.

Cache bandwidths

Table 2-7 shows the maximum bandwidth estimates for a single core on the Power E850C system.

Table 2-7 Power E850C single core bandwidth estimates

Single core	Power E850C	Power E850C	Power E850C
	One core @ 3.65 GHz	One core @ 3.95 GHz	One core @ 4.22 GHz
L1 (data) cache	175.58 GBps	189.94 GBps	202.70 GBps
L2 cache	175.58 GBps	189.94 GBps	202.70 GBps
L3 cache	234.11 GBps	253.25 GBps	270.27 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 formula is as follows:
  - 3.65 GHz Core: (2 x 16 B + 1 x 16 B) x 3.658 GHz = 175.58 GBps
  - 3.95 GHz Core: (2 x 16 B + 1 x 16 B) x 3.957 GHz = 189.94 GBps
  - 4.22 GHz Core: (2 x 16 B + 1 x 16 B) x 4.223 GHz = 202.70 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:
  - 3.65 GHz Core: (1 x 32 B + 1 x 16 B) x 3.658 GHz = 175.58 GBps
  - 3.95 GHz Core: (1 x 32 B + 1 x 16 B) x 3.957 GHz = 189.94 GBps
  - 4.22 GHz Core: (1 x 32 B + 1 x 16 B) x 4.223 GHz = 202.70 GBps
- L3 cache: In one clock cycle, one 32-byte load operation and one 32-byte store operation can be accomplished. The value varies depending on the clock of the core and the formula is as follows:
  - 3.65 GHz Core: (1 x 32 B + 1 x 32 B) x 3.658 GHz = 234.11 GBps
  - 3.95 GHz Core: (1 x 32 B + 1 x 32 B) x 3.957 GHz = 253.25 GBps
  - 4.22 GHz Core: (1 x 32 B + 1 x 32 B) x 4.223 GHz = 270.27 GBps

Memory bandwidths

Each processor module in the Power E850C server has two memory controllers, each of which controls four memory channels, each of which can be connected to a CDIMM. These high-speed memory channels run at 8 GHz, and can support 2 byte read operations and 1 byte write operation concurrently. This support is independent of the processor clock speed.

So a single processor module can support a memory bandwidth of 8 GHz x 8 channels x (2 x read byte + 1 x write byte) = 192 GBps. This calculation assumes that all memory CDIMM slots are populated. If fewer memory CDIMM slots are populated, this figure is lower.

The theoretical maximum memory bandwidth of a server depends on the number of processor modules installed. These figures are listed in Table 2-8.

Table 2-8 Maximum theoretical memory bandwidths

Processor modules installed	Two processor modules installed	Three processor modules installed	Four processor modules installed
Maximum theoretical memory bandwidth	384 GBps	576 GBps	768 GBps

Server summaries

The following tables summarize the cache and memory bandwidths for a number of different Power E850C server configurations.

Table 2-9 shows the theoretical maximum bandwidths for two, three, or four installed 3.65 GHz (12 core) processor modules.

Table 2-9 Theoretical maximum bandwidths for 3.65 GHz system configurations

System bandwidth	Power E850C	Power E850C	Power E850C
	2 processor modules @ 3.65 GHz (24 cores)	3 processor modules @ 3.65 GHz (36 cores)	4 processor modules @ 3.65 GHz (48 cores)
L1 (data) cache	4213.92 GBps	6320.88 GBps	8427.84 GBps
L2 cache	4213.92 GBps	6320.88 GBps	8427.84 GBps
L3 cache	5618.64 GBps	8427.96 GBps	11,237.28 GBps
L4 cache/memory	384 GBps	576 GBps	768 GBps

Table 2-10 shows the theoretical maximum bandwidths for two, three, or four installed 3.95 GHz (10 core) processor modules.

Table 2-10 Theoretical maximum bandwidths for 3.95 GHz system configurations

System bandwidth	Power E850C	Power E850C	Power E850C
	2 processor modules @ 3.95 GHz (20 cores)	3 processor modules @ 3.95 GHz (30 cores)	4 processor modules @ 3.95 GHz (40 cores)
L1 (data) cache	3798.8 GBps	5698.2 GBps	7597.6 GBps
L2 cache	3798.8 GBps	5698.2 GBps	7597.6 GBps
L3 cache	5065.0 GBps	7597.5 GBps	10,130.0 GBps
L4 cache/memory	384 GBps	576 GBps	768 GBps

Table 2-11 shows the theoretical maximum bandwidths for two, three, or four installed 4.22 GHz (8 core) processor modules.

Table 2-11 Theoretical maximum bandwidths for 4.22 GHz system configurations

System bandwidth	Power E850C	Power E850C	Power E850C
	2 processor modules @ 4.22 GHz (16 cores)	3 processor modules @ 4.22 GHz (24 cores)	4 processor modules @ 4.22 GHz (32 cores)
L1 (data) cache	3243.2 GBps	4864.8 GBps	6484.4 GBps
L2 cache	3243.2 GBps	4864.8 GBps	6484.4 GBps
L3 cache	4324.3 GBps	6486.5 GBps	8648.7 GBps
L4 cache/memory	384 GBps	576 GBps	768 GBps

2.5.6 Active Memory Mirroring

The Power E850C server can provide mirroring of the hypervisor code across multiple memory CDIMMs. If a CDIMM that contains the hypervisor code develops an uncorrectable error, its mirrored partner enables the system to continue to operate uninterrupted.

Active Memory Mirroring (AMM) is a chargeable feature on the Power E850C. It can be ordered as part of the initial order, or as an MES upgrade later by using feature code #EM81. After it is licensed, it can be enabled, disabled, or reenabled depending on the user's requirements.

The hypervisor code logical memory blocks are mirrored on distinct CDIMMs to allow for more usable memory. No specific CDIMM hosts the hypervisor memory blocks, so the mirroring is done at the logical memory block level, not at the CDIMM level. To enable the AMM feature, the server must have enough free memory to accommodate the mirrored memory blocks.

Besides the hypervisor code itself, other components that are vital to the server operation are also mirrored:

- Hardware page tables (HPTs), which are responsible for tracking the state of the memory pages assigned to partitions
- Translation control entities (TCEs), which are responsible for providing I/O buffers for the partition's communications
- Memory used by the hypervisor to maintain partition configuration, I/O states, virtual I/O information, and partition state

It is possible to check whether the Active Memory Mirroring option is enabled and change its status through Hardware Management Console (HMC), under the Advanced tab on the System Properties window (Figure 2-8).

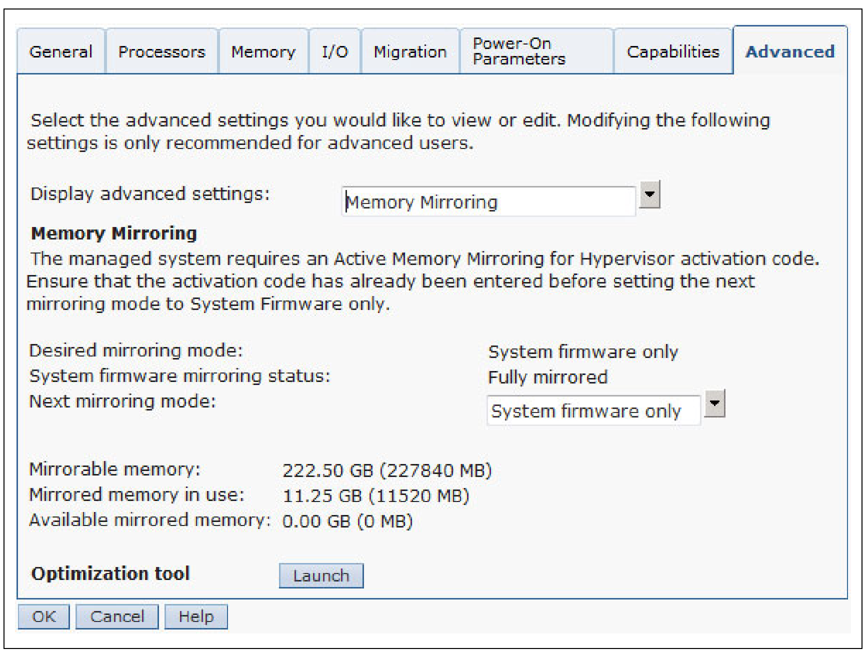

Figure 2-8 System Properties panel on an HMC

After a failure on one of the CDIMMs containing hypervisor data occurs, all the server operations remain active and the flexible service processor (FSP) isolates the failing CDIMMs. Systems stay in the partially mirrored state until the failing CDIMM is replaced.

Some components are not mirrored because they are not vital to the regular server operations and require a larger amount of memory to accommodate their data:

- Advanced Memory Sharing Pool
- Memory used to hold the contents of platform dumps

> **Partition data:** Active Memory Mirroring will *not* mirror partition data. It was designed to mirror only the hypervisor code and its components, allowing this data to be protected against a CDIMM failure.

With AMM, uncorrectable errors in data that is owned by a partition or application are handled by the existing Special Uncorrectable Error handling methods in the hardware, firmware, and operating system.

2.5.7 Memory Error Correction and Recovery

Many features within the Power E850C memory subsystem are designed to reduce the risk of errors, or to minimize the impact of any errors that do occur. These features help ensure that those errors do not affect critical enterprise data.

Each memory chip has error detection and correction circuitry built in. This circuitry is designed so that the failure of any one specific memory module within an ECC word can be corrected without any other fault.

In addition, a spare DRAM per rank on each memory CDIMM provides for dynamic DRAM device replacement during runtime operation. Also, dynamic lane sparing on the memory link allows for replacement of a faulty data lane without affecting performance or throughput.

Other memory protection features include retry capabilities for certain faults detected at both the memory controller and the memory buffer.

Memory is also periodically scrubbed to allow for soft errors to be corrected and for solid single-cell errors reported to the hypervisor. This process supports operating system deallocation of a page that is associated with a hard single-cell fault.

For more information about Memory RAS, see 4.3.10, “Memory protection” on page 110.

2.5.8 Special Uncorrectable Error handling

Special Uncorrectable Error (SUE) handling prevents an uncorrectable error in memory or cache from immediately causing the system to terminate. Rather, the system tags the data and determines whether it will ever be used again. If the error is irrelevant, it does not force a checkstop. If the data is used, termination can be limited to the program/kernel or hypervisor that owns the data, or can freeze the I/O adapters controlled by an I/O hub controller if data is to be transferred to an I/O device.

2.6 Capacity on Demand

Several types of Capacity on Demand (CoD) offerings are optionally available on the Power E850C server to help meet changing resource requirements in an on-demand environment. CoD uses resources that are installed on the system but that are not activated.

> **Hardware Management Consoles:** The Power E850C does not require an HMC for management. However, most CoD capabilities require an HMC to manage them. Capacity Upgrade on Demand (CUoD) does not require an HMC because permanent activations can be enabled through the Advanced System Management Interface (ASMI) menus.

2.6.1 Capacity Upgrade on Demand

A Power E850C server includes a number of active processor cores and memory units. It can also include inactive processor cores and memory units. Active processor cores or memory units are processor cores or memory units that are already available for use on your server when it comes from the manufacturer. Inactive processor cores or memory units are processor cores or memory units that are installed in your server, but not available for use until you activate them. Inactive processor cores and memory units can be permanently activated by purchasing an activation feature called *Capacity Upgrade on Demand* and entering the provided activation code on your server.

With the CUoD offering, you can purchase additional processor or memory capacity in advance at a low cost, and dynamically activate them when needed. Activation does not require that you restart your server or interrupt your business. All the processor or memory activations are restricted to the single server they are licensed to, and cannot be transferred to another system.

Capacity Upgrade on Demand can have several applications to allow for a more flexible environment. CUoD allows a company to reduce their initial investment in a system. Traditional projects that use other technologies require that the system is acquired with all the resources available to support the whole lifecycle of the project. This setup can incur costs that are only necessary in later stages of the project, which affects software licensing costs and software maintenance.

By using Capacity Upgrade on Demand, the company could start with a system with enough installed resources to support the whole project lifecycle but only with enough active resources necessary for the initial project phases. More resources could be activated as the project continues, adjusting the hardware platform to meet the changing project needs. This process allows the company to reduce the initial investment in hardware and only acquire software licenses that are needed on each project phase, reducing the Total Cost of Ownership and Total Cost of Acquisition of the solution.

Figure 2-9 shows a comparison between two scenarios: A fully activated system versus a system with CUoD resources being activated along with the project timeline.

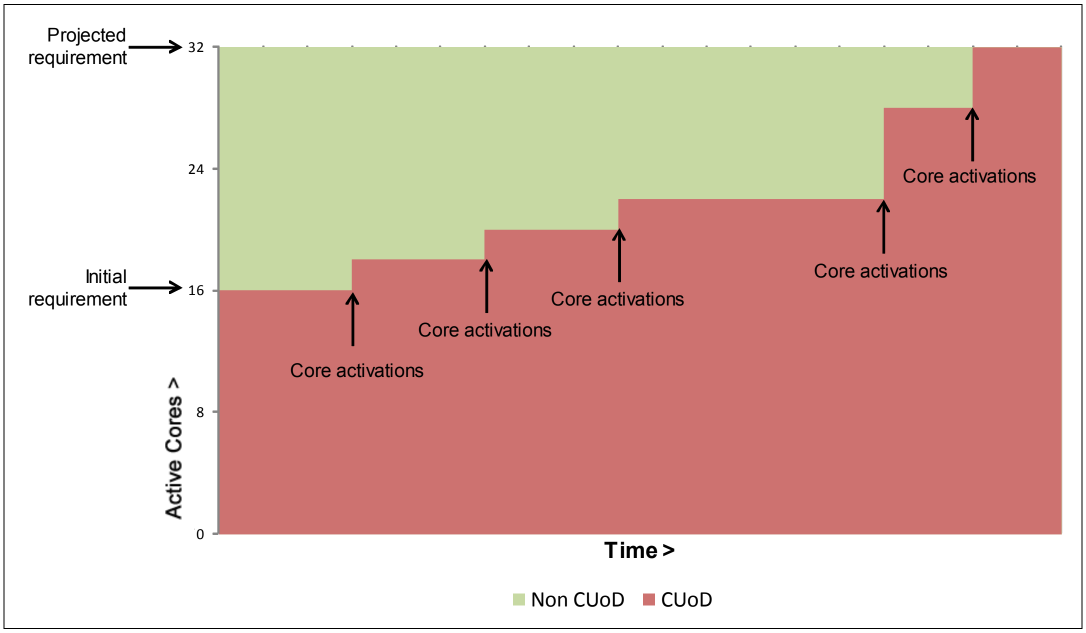

Figure 2-9 An example of Capacity Upgrade on Demand usage

Table 2-12 lists the processor activation features for the Power E850C server. Each feature code activates a single processor core. You cannot activate more processor cores than are physically installed in the server.

Table 2-12 Permanent processor activation codes for Power E850C

Processor module	Processor module feature code	Permanent activation feature code
12-core 3.65 GHz	#EPW6	#EPW9
10-core 3.95 GHz	#EPW5	#EPW8
8-core 4.22 GHz	#EPW4	#EPW7

Permanent activations for memory features in the Power E850C can be ordered by using the following feature codes:

- #EMAA for 1 GB activation of memory
- #EMAB for 100 GB activation of memory

These feature codes are the same regardless of the capacity of the memory CDIMMs installed in the system. You cannot activate more memory capacity than is physically installed in the server.

2.6.2 Elastic Capacity on Demand

With the Elastic CoD offering, you can temporarily activate and deactivate processor cores and memory units to help meet the demands of business peaks such as seasonal activity, period-end, or special promotions. Elastic CoD enables processors or memory to be temporarily activated in one day increments as needed. These activations cover a period of 24 hours from the time the activation is enabled on the system. When you order an Elastic CoD feature, you receive an enablement code that allows a system operator to make requests for additional processor and memory capacity in increments of one processor day or 1 GB memory day. The system monitors the amount and duration of the activations. Both prepaid and postpay options are available.

> **Note:** Some websites or documents still refer to Elastic Capacity on Demand (Elastic CoD) as *On/Off Capacity on Demand*.

Charges are based on usage reporting that is collected monthly. Processors and memory can be activated and turned off an unlimited number of times, whenever additional processing resources are needed.

This offering provides the system administrator with an interface at the HMC to manage the activation and deactivation of resources. A monitor that resides on the server records the usage activity. This usage data must be sent to IBM monthly. A bill is then generated based on the total amount of processor and memory resources used, in increments of processor and memory (1 GB) days.

Before using temporary capacity on your server, you must enable your server. To enable it, an enablement feature (MES only) must be ordered and the required contracts must be in place. The feature codes are #EP9T for processor enablement and #EM9T for memory enablement.

The Elastic CoD process consists of three steps: Enablement, activation, and billing.

- Enablement

  Before requesting temporary capacity on a server, you must enable it for Elastic CoD. To do this, order an enablement feature and sign the required contracts. IBM generates an enablement code, mails it to you, and posts it on the web for you to retrieve and enter on the target server.

  A *processor enablement* code (#EP9T) allows you to request up to 90 processor days of temporary capacity for each inactive processor core within the server. For instance, if you have 32 processor cores installed, and 16 are permanently activated, you would have 16 inactive processor cores. Therefore, you would receive enablement for (16 x 90) = 1440 processor days of elastic CoD. If the 90 processor-day limit is reached, place an order for another processor enablement code to reset the number of days that you can request back to 90 per inactive processor core.

  A *memory enablement* code (#EM9T) lets you request up to 90 memory days of temporary capacity for each GB of inactive memory within the server. For instance, if you have 256 GB of memory installed in the system, and 156 GB is permanently activated, you would have 100 GB of inactive memory. Therefore, you would receive an enablement code for (100 x 90) = 9000 GB days of elastic CoD. If you reach the limit of 90 memory days, place an order for another memory enablement code to reset the number of allowable days you can request back to 90.

- Activation requests

  When Elastic CoD temporary capacity is needed, use the HMC menu for On/Off CoD. Specify how many inactive processors or gigabytes of memory are required to be temporarily activated for some number of days. You are billed for the days that are requested, whether the capacity is assigned to partitions or remains in the shared processor pool.

  At the end of the temporary period (days that were requested), ensure that the temporarily activated capacity is available to be reclaimed by the server (not assigned to partitions), or you will be billed for any unreturned processor days.

- Billing

  The contract, signed by the client before receiving the enablement code, requires the Elastic CoD user to report billing data at least once a month (whether or not activity occurs). This data is used to determine the proper amount to bill at the end of each billing period (calendar quarter). Failure to report billing data for use of temporary processor or memory capacity during a billing quarter can result in default billing equivalent to 90 processor days of temporary capacity for each inactive processor core.

For more information about registration, enablement, and usage of Elastic CoD, visit the following website:

http://www.ibm.com/systems/power/hardware/cod

HMC requirement: Elastic Capacity on Demand requires that an HMC is used for management of the Power E850C server.

2.6.3 Utility Capacity on Demand

Utility Capacity on Demand (Utility CoD) automatically provides additional processor performance on a temporary basis within the shared processor pool.

With Utility CoD, you can place a quantity of inactive processors into the server's shared processor pool, which then becomes available to the pool's resource manager. When the server recognizes that the combined processor utilization within the shared processor pool exceeds 100% of the level of base (permanently activated) processors that are assigned across uncapped partitions, then a Utility CoD processor minute is charged. This level of performance is then available for the next minute of use.

If additional workload requires a higher level of performance, the system automatically allows the additional Utility CoD processors to be used. The system automatically and continuously monitors and charges for the performance needed above the base (permanently activated) level.

Registration and usage reporting for Utility CoD is done using a public website, and payment is based on reported usage. Utility CoD requires PowerVM Enterprise Edition to be active on the Power E850C system.

For more information about registration, enablement, and use of Utility CoD, visit the following location:

http://www.ibm.com/systems/support/planning/capacity/index.html

HMC requirement: Utility Capacity on Demand requires that an HMC is used for management of the Power E850C server.

2.6.4 Trial Capacity on Demand

A *standard request* for Trial Capacity on Demand (Trial CoD) requires you to complete a form that includes contact information and vital product data (VPD) from your Power E850C server with inactive CoD resources.

A standard request activates eight processors or 64 GB of memory (or both eight processors and 64 GB of memory) for 30 days. Subsequent standard requests can be made after each purchase of a permanent processor activation. An HMC is required to manage Trial CoD activations.

An *exception request* for Trial CoD requires you to complete a form that includes contact information and VPD from your Power E850C server with inactive CoD resources. An exception request activates all inactive processors or all inactive memory (or all inactive processor and memory) for 30 days. An exception request can be made only one time over the life of the machine. An HMC is required to manage Trial CoD activations.

To request either a Standard or an Exception Trial, visit the following location:

https://www-912.ibm.com/tcod_reg.nsf/TrialCod?OpenForm

2.6.5 Software licensing and Capacity on Demand

Although Capacity on Demand orders are placed by using a hardware feature code, some CoD offerings include entitlement to certain IBM Systems Software and operating systems for the temporarily enabled capacity. Any other software products, including IBM applications and middleware, are not included in this cost. Check with your software vendor to determine how they charge for temporarily activated resources.

Capacity Upgrade on Demand permanently activates processor cores on the server. As such, this is seen as new permanent capacity, and licenses need to be purchased for any operating system and systems software running on this capacity. The activation code that is provided is for the hardware resources only.

Elastic CoD, Utility CoD, and Trial CoD activations include incremental licensing for the following IBM Systems Software and operating systems:

- AIX
- PowerVM
- PowerVC
- IBM PowerVP™
- IBM Cloud Manager with OpenStack
- IBM Cloud PowerVC Manager
- IBM PowerHA®
- IBM PowerSC™
- Cluster Systems Management (CSM)
- IBM General Parallel File System (GPFS™)

Other IBM Systems Software or operating systems might also be included in the activations. Linux operating systems licenses might already cover the additional capacity, dependent on the licensing metric used. Check with your distributor for more details.

> **Note:** CoD does not ship any software or provide the base licensing entitlement. The software has to be initially installed and licensed on the server before temporary CoD provides the incremental licensing to cover the additional processor cores that are temporarily activated.

For more information about software licensing considerations with the various CoD offerings, see the most recent revision of the *Power Systems Capacity on Demand User's Guide*:

http://www.ibm.com/systems/power/hardware/cod

2.6.6 Integrated Facility for Linux

When running Linux workloads on a Power E850C server, it is possible to reduce the overall cost of processor and memory activations by ordering the Integrated Facility for Linux (IFL) package (#ELJN). This chargeable option provides four processor activations, 32 GB of memory activations, and four licenses of PowerVM for Linux. This package is at a lower overall cost than ordering the same number of processor, memory, and PowerVM activations separately. However, the activated resources can only be used to run Linux server-based and VIO server-based workloads.

Any combination of Linux distributions supported by PowerVM can be used in logical partitions that use these resources. Any partitions that are running AIX cannot use these resources, and shared processor pools need to be set up on the system to prevent this. The Linux partitions can use any other resources that are permanently activated on the system, or that are temporarily activated by using Elastic CoD or Utility CoD.

IFL packages can be ordered as part of the initial system order, or as an MES upgrade. These activations are permanent, and count towards the minimum activations that are required for the system.

It is possible to use only IFL packages for permanent activations, in which case the entire system is restricted to running only Linux workloads.

When ordering an Integrated Facility for Linux package (#ELNN), the following no-cost features are added to the system:

- 1 x four processor activations (#ELNK, #ELNL, or #ELNM depending on processor module)
- 1 x 32 GB memory activation (#ELNP)
- 1 x four PowerVM for Linux entitlements (#ELNQ)

2.7 System buses

This section provides additional information related to the internal buses.

2.7.1 PCI Express Generation 3

The internal I/O subsystem in the Power E850C is connected directly to the PCIe controllers on the POWER8 processor modules. Each POWER8 processor module has four buses, two of which have 16 lanes (x16) and two of which have eight lanes (x8). This configuration provides a total of 48 lanes of PCIe connectivity per processor node. Each lane runs at 1 GBps in each direction, giving a total maximum PCIe bandwidth per processor module of 96 GBps.

The first and second processor modules in a system use all of their available PCIe connectivity. The third and fourth processor modules use only the two x16 connections currently. The maximum PCIe bandwidth of a Power E850C server with four processor modules installed is therefore 320 GBps.

2.7.2 PCIe logical connectivity

Figure 2-10 shows how the internal PCIe Gen3 adapter slots are connected logically to the processor modules in the system. Each processor module supports two PCIe Gen3 x16 slots. Processor module 0 also supports a single PCIe Gen3 x8 slot that is used for the default LAN adapter. The other x8 connection is used by the RAID adapter in the system. Processor module 1 supports two x8 slots connected through a PCIe Gen3 switch, along with the internal USB 3.0 host adapter for the four USB ports on the server (two rear, two front). Therefore, these adapters share the bandwidth of the x8 connection from the processor module.

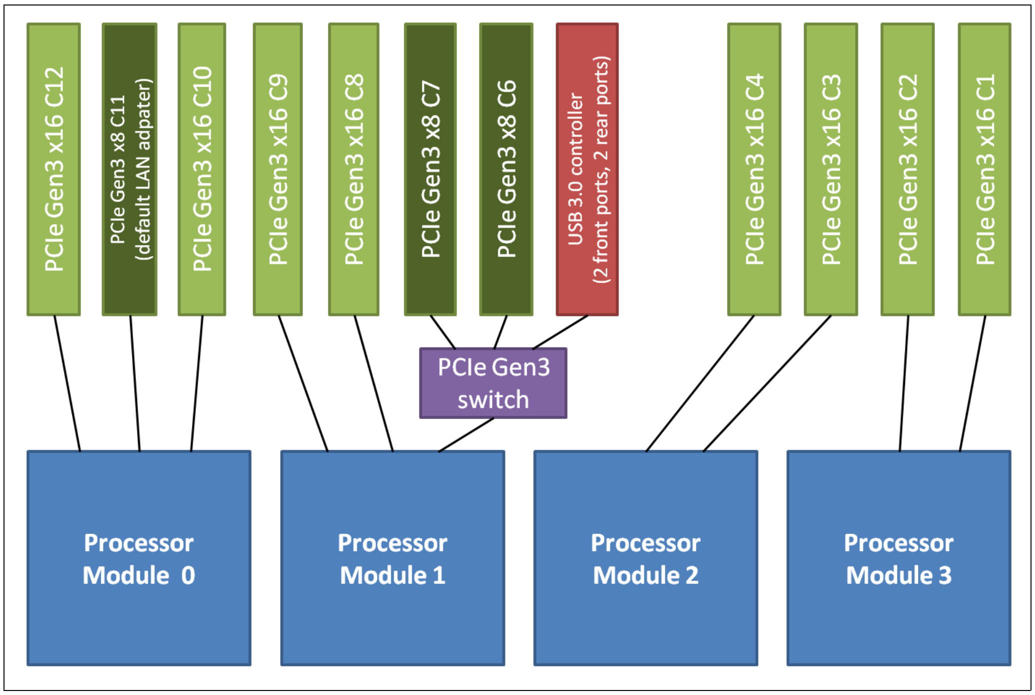

Figure 2-10 PCIe connectivity for the Power E850C server

A Power E850C server with fewer than four processor modules installed will not support all of the PCIe slots inside the system enclosure. The third and fourth processor modules each support two of the PCIe Gen3 x16 slots. Therefore, the maximum number of adapters supported and the total PCIe bandwidth available varies by configuration.

All PCIe Gen3 slots in the server are hot-plug compatible for concurrent maintenance and repair. These procedures should be initiated through the HMC or ASMI menus.

Where the number of PCIe adapters in a server is important, the Power E850C supports the external PCIe Gen3 I/O expansion drawer (#EMX0). This drawer contains two Fan-Out modules, each of which supports six PCIe adapter slots (four x8 slots and two x16 slots). These Fan-Out modules connect through optical cables to a PCIe Optical Cable Adapter (#EJ08) card, which is placed in a PCIe Gen3 x16 adapter slot in the server.

Each processor module in the Power E850C server supports up to two PCIe Optical Cable Adapters, and therefore up to two Fan-Out modules. This configuration means that each processor module can support up to one full PCIe Gen3 I/O expansion drawer, giving a total of 12 PCIe Gen3 slots. All of the adapters that are connected to a Fan-Out module share the bandwidth of the single x16 slot in the server. For more information about the PCIe Gen3 I/O expansion drawer, see 2.11.1, “PCIe Gen3 I/O expansion drawer” on page 62.

Table 2-13 summarizes the maximum numbers of adapters and bandwidths for different configurations of the Power E850C server.

Table 2-13 Maximum PCIe adapters and bandwidth supported on the Power E850C

Processor Modules	Maximum PCIe bandwidth	PCIe slots in server enclosure	Maximum I/O expansion drawers	Maximum PCIe slots supported
Two	192 GBps	7	2	27
Three	256 GBps	9	3	39
Four	320 GBps	11	4	51

2.8 Internal I/O connections

The internal I/O subsystem resides on the I/O planar, which supports all of the PCIe Gen3 x16 and x8 slots. All PCIe slots are hot-pluggable and enabled with Enhanced Error Handling (EEH). In the unlikely event of a problem, EEH-enabled adapters respond to a special data packet that is generated from the affected PCIe slot hardware by calling system firmware. The firmware examines the affected bus, allows the device driver to reset it, and continues without a system reboot. For more information about RAS on the I/O buses, see 4.3.11, “I/O subsystem availability and Enhanced Error Handling” on page 111.

All of the PCIe slots within the system enclosure support full-height half-length PCIe cards. These can be PCIe Gen1, Gen2, or Gen3 adapters. The server also supports full-height full-length audiotapes in the I/O expansion drawer. For more information, see 2.11.1, “PCIe Gen3 I/O expansion drawer” on page 62.

Table 2-14 lists the slot configuration of the Power E850C server.

Table 2-14 Slot configuration and capabilities

Slot	Location code	Slot type	CAPI capable	SRIOV capable
Slot 1[a]	P1-C1	PCIe Gen3 x16	Yes	Yes
Slot 2[ab]	P1-C2	PCIe Gen3 x16	Yes	Yes
Slot 3[c]	P1-C3	PCIe Gen3 x16	Yes	Yes
Slot 4[cb]	P1-C4	PCIe Gen3 x16	Yes	Yes
FSP slot[d]	N/A[d]	N/A[d]	N/A[d]	N/A[d]
Slot 5	P1-C6	PCIe Gen3 x8	No	Yes
Slot 6	P1-C7	PCIe Gen3 x8	No	Yes
Slot 7[b]	P1-C8	PCIe Gen3 x16	Yes	Yes
Slot 8	P1-C9	PCIe Gen3 x16	Yes	Yes
Slot 9[b]	P1-C10	PCIe Gen3 x16	Yes	Yes
Slot 10[e]	P1-C11	PCIe Gen3 x8	No	Yes
Slot 11	P1-C12	PCIe Gen3 x16	Yes	Yes

a. Slots 1 and 2 are only active when the fourth processor module is installed.
b. This slot is capable of supporting 75 W adapters.
c. Slots 3 and 4 are only active when the third processor module is installed.

d. The space for slot P1-C5 is used for the FSP card and connections, so it is not a usable PCIe slot.
e. Slot 10 (P1-C11) is used by the default LAN adapter, which is required for manufacturing and testing.

Figure 2-10 on page 46 shows the physical and logical placement of the slots.

Table 2-15 shows the priorities for the PCIe adapter slots in the Power E850C.

Table 2-15 Adapter slot priorities for the Power E850C

Configuration	Two processor modules installed	Three processor modules installed	Four processor modules installed
PCIe slot priority	10, 9, 7, 11, 8, 5, 6	10, 9, 7, 4, 11, 8, 3, 5, 6	10, 9, 7, 4, 2, 11, 8, 3, 1, 5, 6

2.8.1 System ports

The Power E850C server has a number of system ports, which are used for management of the server. These are not accessible from the host operating system. Instead, they connect to the FSP. The system ports sit within the space allocated for C5 on the backplane of the system enclosure. The following system ports can be found:

- 2 x USB 2.0 ports

  These ports can be used to connect uninterruptible power supplies (UPSs), and for FSP installation and code updates if necessary. These USB 2.0 ports are not available to host operating systems. They are only accessible by the service processor.

- 2 x RJ-45 Ethernet ports

  These ports are used for management connectivity to the system processor. They can be used to connect to the ASMI menus through a web interface, or can be used for connection to an HMC if applicable. Each Ethernet port has a unique MAC address, and both can be assigned different Ethernet addresses on two distinct subnets. One Ethernet port is used as the primary connection, and the second Ethernet port can be used for redundancy.

- 1 x serial port

  This RJ-45 port can be converted to a standard serial connection by using an optional system serial port converter cable (#3930). This addition allows management of the system through a standard serial connection.

2.9 PCI adapters

This section covers the types and functions of the PCI cards that are supported by the Power E850C server.

2.9.1 PCI Express

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 such 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 supported on this server are PCIe Gen3, 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 connector size slots first, before mixing adapter speed with connector slot size.
- Adapters with smaller 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).

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

- PCIe low profile (LP) cards, which are not used with the Power E850C server.
- PCIe full height cards, which are used in the Power E850C server and the PCIe Gen3 I/O expansion drawer (#EMX0).

Low-profile PCIe adapters are supported only in low-profile PCIe slots, and full-height cards are supported only in full-height slots.

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

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

The following sections describe the supported adapters and provide tables of orderable feature numbers. The tables indicate operating system support (AIX and Linux) for each of the adapters.

Note: PCIe full height and full high cards are used in the Power E850C server and any attached PCIe Gen3 I/O expansion drawer (#EMX0).

2.9.2 LAN adapters

To connect the Power E850C server to a local area network (LAN), you can use the LAN adapters that are supported in the PCIe slots of the system. Table 2-16 lists the available LAN adapters. Information about Fibre Channel over Ethernet (FCoE) adapters can be found in Table 2-20 on page 52.

Table 2-16 Available LAN adapters

Feature code	CCIN	Description	Max per system	OS support
5767	5767	2-Port 10/100/1000 Base-TX Ethernet PCI Express Adapter	50	AIX, Linux
5899	576F	PCIe2 4-port 1 GbE Adapter	50	AIX, Linux

Feature code	CCIN	Description	Max per system	OS support
EC2N		PCIe3 2-port 10 GbE NIC&RoCE SR Adapter	50	AIX, Linux
EC38		PCIe3 2-port 10 GbE NIC&RoCE SFP+ Copper Adapter	50	AIX, Linux
EC3B	57B6	PCIe3 2-Port 40 GbE NIC RoCE QSFP+ Adapter	50	AIX, Linux
EC3M	2CEC	PCIe3 2-port 100GbE (NIC&RoCE) QSFP28 Adapter x16	50	AIX, Linux
EN0S	2CC3	PCIe2 4-Port (10 Gb +1 GbE) SR+RJ45 Adapter	50	AIX, Linux
EN0U	2CC3	PCIe2 4-port (10 Gb +1 GbE) Copper SFP+RJ45 Adapter	50	AIX, Linux
EN0W	2CC4	PCIe2 2-port 10/1 GbE BaseT RJ45 Adapter	50	AIX, Linux
EN15	2CE3	PCIe3 4-port 10 GbE SR Adapter	50	AIX, Linux
EN17	2CE4	PCIe3 4-port 10 GbE SFP+ Copper Adapter	50	AIX, Linux

2.9.3 Graphics accelerator adapters

Table 2-17 lists the available graphics accelerator adapters. An adapter can be configured to operate in either 8-bit or 24-bit color modes. The adapter supports both analog and digital monitors.

Table 2-17 Available graphics accelerator adapters

Feature Code	Description	Max per system	OS support
5748	POWER GXT145 PCI Express Graphics Accelerator	9	AIX, Linux
EC42	PCIe2 3D Graphics Adapter x1	10	AIX, Linux

2.9.4 SAS adapters

Table 2-18 lists the SAS adapters that are available for the Power E850C server.

Table 2-18 Available SAS adapters

Feature code	CCIN	Description	Max per system	OS support
5901	57B3	PCIe Dual-x4 SAS Adapter	50	AIX, Linux
EJ0J	57B4	PCIe3 RAID SAS Adapter Quad-port 6 Gb x8	34	AIX, Linux
EJ0L	57CE	PCIe3 12 GB Cache RAID SAS Adapter Quad-port 6 Gb x8	34	AIX, Linux
EJ10	57B4	PCIe3 SAS Tape/DVD Adapter Quad-port 6 Gb x8	34	AIX, Linux
EJ1P	57B3	PCIe1 SAS Tape/DVD Dual-port 3Gb x8 Adapter	8	AIX, Linux
EJ14	57B1	PCIe3 12 GB Cache RAID PLUS SAS Adapter Quad-port 6 Gb x8	24	AIX, Linux

2.9.5 Fibre Channel adapter

The Power E850C supports direct or SAN connection to devices that use Fibre Channel adapters. Table 2-19 summarizes the available Fibre Channel adapters, which all have LC connectors.

If you are attaching a device or switch with an SC type fibre connector, an LC-SC 50 Micron Fibre Converter Cable (#2456) or an LC-SC 62.5 Micron Fibre Converter Cable (#2459) is required.

Table 2-19 Available Fibre Channel adapters

Feature Code	CCIN	Description	Max per system	OS support
5729		PCIe2 8 Gb 4-port Fibre Channel Adapter	50	AIX, Linux
5735	577D	8 Gigabit PCI Express Dual Port Fibre Channel Adapter	50	AIX, Linux
EN12		PCIe2 8 Gb 4-port Fibre Channel Adapter		AIX, Linux
EN0A	577F	PCIe2 16 Gb 2-port Fibre Channel Adapter	50	AIX, Linux
EN0G		PCIe2 8 Gb 2-Port Fibre Channel Adapter	50	AIX, Linux

2.9.6 Fibre Channel over Ethernet

FCoE allows for the convergence of Fibre Channel (FC) and Ethernet traffic onto a single adapter and a converged fabric.

Figure 2-11 compares existing Fibre Channel and network connections and FCoE connections.

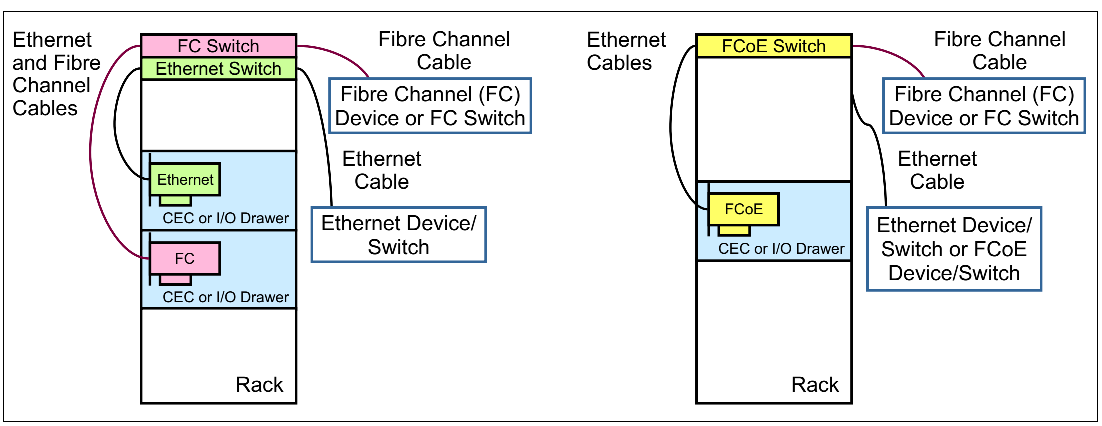

Figure 2-11 Comparison between existing FC and network connections and FCoE connections

Table 2-20 lists the available FCoE adapters. They are high-performance Converged Network Adapters (CNAs) that use SR optics. Each port can simultaneously provide network interface card (NIC) traffic and Fibre Channel functions.

Table 2-20 Available FCoE adapters

Feature Code	CCIN	Description	Max per system	OS support
EN0H	2B93	PCIe2 4-port (10 Gb FCoE & 1 GbE) SR & RJ45	50	AIX, Linux
EN0K	2CC1	PCIe2 4-port (10 Gb FCoE & 1 GbE) SFP+Copper & RJ45	50	AIX, Linux

For more information about FCoE, see *An Introduction to Fibre Channel over Ethernet, and Fibre Channel over Convergence Enhanced Ethernet*, REDP-4493.

Note: Adapters #EN0H, and #EN0K, support SR-IOV when minimum firmware and software levels are met.

2.9.7 InfiniBand Host Channel adapter

The InfiniBand Architecture (IBA) is an industry-standard architecture for server I/O and inter-server communication. It was developed by the InfiniBand Trade Association (IBTA) to provide the levels of reliability, availability, performance, and scalability that are necessary for present and future server systems with levels better than can be achieved by using bus-oriented I/O structures.

InfiniBand is an open set of interconnect standards and specifications. The main InfiniBand specification is published by the IBTA and is available at the following website:

http://www.infinibandta.org

InfiniBand is based on a switched fabric architecture of serial point-to-point links, where these InfiniBand links can be connected to either host channel adapters (HCAs), which are used primarily in servers, or target channel adapters (TCAs), which are used primarily in storage subsystems.

The InfiniBand physical connection consists of multiple byte lanes. Each individual byte lane is a four-wire, 2.5, 5.0, or 10.0 Gbps bidirectional connection. Combinations of link width and byte lane speed allow for overall link speeds of 2.5 - 120 Gbps. The architecture defines a layered hardware protocol and a software layer to manage initialization and the communication between devices. Each link can support multiple transport services for reliability and multiple prioritized virtual communication channels.

For more information about InfiniBand, see *HPC Clusters Using InfiniBand on IBM Power Systems Servers*, SG24-7767.

A connection to supported InfiniBand switches is accomplished by using the QDR optical cables #3290 and #3293.

Table 2-21 lists the available InfiniBand adapter.

Table 2-21 Available InfiniBand adapter

Feature code	CCIN	Description	Max per system	OS support
EC3F[a]	2CEA	PCIe3 2-port 100 Gb EDR InfiniBand Adapter x16	3	Linux
EC3U[a]	2CEC	PCIe3 1-port 100Gb EDR IB Adapter x16	3	Linux

a. This adapter is CAPI capable

2.9.8 Asynchronous and USB adapters

Asynchronous PCIe adapters provide the connection of asynchronous EIA-232 or RS-422 devices. If you have a cluster configuration or high-availability configuration and plan to connect the IBM Power Systems using a serial connection, you can use the features that are listed in Table 2-22.

Table 2-22 Available asynchronous and USB adapters

Feature code	CCIN	Description	Max per system	OS support
5785	57D2	4 Port Async EIA-232 PCIe Adapter	9	AIX, Linux
EN27		PCIe 2-Port Async EIA-232 adapter	10	AIX, Linux
EC46		PCIe2 LP 4-Port USB 3.0 Adapter	30	AIX, Linux

2.9.9 Cryptographic coprocessor

The cryptographic coprocessor cards that are supported for the Power E850C are shown in Table 2-23.

Table 2-23 Available cryptographic coprocessor

Feature code	Description	Max per system	OS support
EJ32	PCIe3 Crypto Coprocessor no BSC 4767	10	AIX
EJ33	PCIe3 Crypto Coprocessor BSC-Gen3 4767	10	AIX

2.9.10 Flash storage adapters

The available flash storage adapters are shown in Table 2-24.

Table 2-24 Available flash storage adapters

Feature code	CCIN	Description	Max	OS support
EC55	58CB	PCIe3 1.6 TB NVMe Flash Adapter	7	Linux
EC57	58CC	PCIe3 3.2 TB NVMe Flash Adapter	7	Linux

2.10 Internal storage

There are three storage controller options for the Power E850C server. All of these options connect to the storage backplane that has front mounted storage bays, which include eight hot-plug Small Form Factor (SFF) disk bays and four 1.8-inch disk bays and one DVD bay. Figure 2-12 shows the positions of the different bays on the front of the Power E850C server.

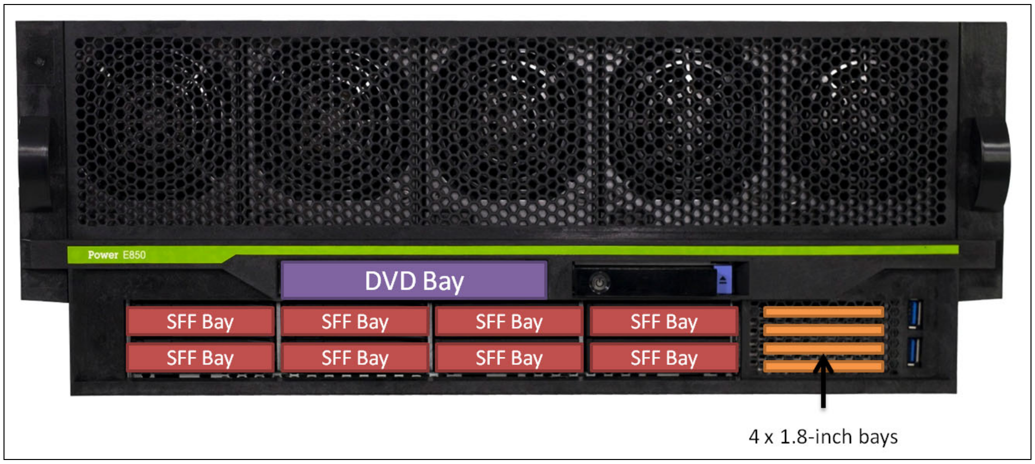

Figure 2-12 Disk bay positions on the front of the Power E850C

2.10.1 Storage controller options

There are three storage controller options to choose from when configuring the Power E850C server:

- Dual controller disk backplane with write cache (#EPVN)

  The pair of controllers handles all 12 integrated disk bays and the DVD bay.
- Dual controller disk backplane without write cache (#EPVP)

  The pair of controllers handles all 12 integrated disk bays and the DVD bay.
- Split disk backplane (two single controllers) without write cache (#EPVQ)

  Each one of the two controllers handles four SFF disk bays and two 1.8-inch disk bays. One of the controllers handles the DVD bay.

None of the controller options provide any external SAS connections for further expansion. If you want to expand the storage capability of the server, you must add an EXP24S expansion drawer that uses a PCIe SAS adapter. This expansion drawer handles further SFF hard disk drives (HDDs) and solid-state devices (SSDs). For more information about the EXP24S expansion drawer, see 2.12.1, “EXP24S SFF Gen2-bay Drawer” on page 69.

Dual controller disk backplane options

The dual controller disk backplane options provide both performance and protection advantages. Patented Active-Active capabilities enhance performance when there is more than one array configured. Each of the dual controllers has access to all the disk bays and can back up the other controller if there was a problem with the other controller. For the dual controller backplane with write cache, each controller mirrors the other’s write cache, providing redundancy protection.

The write-cache capability increases the speed of write operations by committing them to the write-cache flash memory first, and then writing to the disks attached to the controller. This configuration allows write activities to be committed in a shorter time, reducing the latency of the write operation. When the write cache is full, write operation latency reverts to the speed of writes on the attached disks. The write cache has a raw capacity of 1.8 GB, but uses advanced compression techniques to store up to 7.2 GB of writes at a time. Integrated flash memory for the write-cache content provides protection against electrical power loss to the server and avoids the need for write cache battery protection and battery maintenance.

Clients with I/O performance-sensitive workloads with a large percentage of writes should consider using the dual controllers with write cache, or use PCIe SAS controllers with write cache to connect external storage. This consideration is especially relevant for HDDs. Note also that RAID 5 and RAID 6 protection levels result in more drive write activity than mirroring or unprotected drives.

The dual controllers should be treated as a single resource, so both should be assigned to the same partition or VIOS. They both have access to all of the internal disks. If multiple arrays are configured on the internal disks, the controllers split primary responsibility for handling the arrays. If one of the dual controllers fails, the remaining controller takes over all work.

Split disk backplane option

The split disk backplane option has two independent controllers. One of these handles the top four SFF disk bays, the top two 1.8-inch disk bays, and the DVD bay. The other controller handles the bottom four SFF disk bays and the bottom two 1.8-inch disk bays. Figure 2-13 shows how the bays are split across the two controllers.

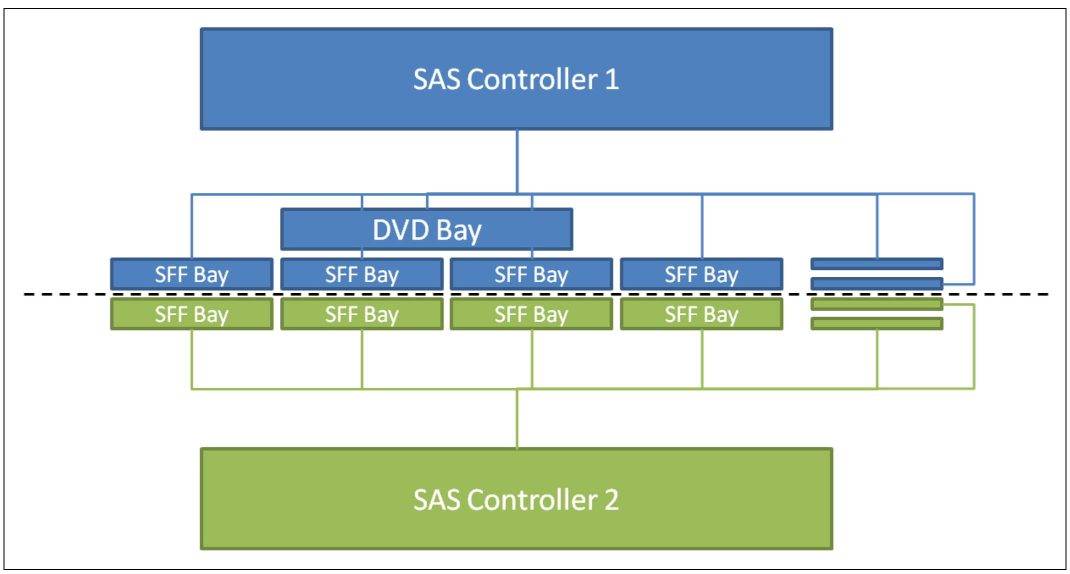

Figure 2-13 Controller to disk bay connections for split disk backplane option on Power E850C

The independent split disk controllers should each be treated as a single resource. Each controller only has access to the six disk bays that are assigned to it, so it cannot take over control of any disks that are assigned to the other controller. During a failure of a disk controller, you might lose all access to a set of disks, and potentially to the DVD drive. Consider this limitation during the system planning phase. For example, you might want to assign each controller to a VIOS partition, and then mirror the two VIOS for protection. Or you might want to assign each to the same partition and then mirror the two sets of drives.

2.10.2 RAID protection for internal disks

There are multiple protection options for HDD/SSD drives in the Power E850C server, whether they are contained in the internal SFF or 1.8-inch bays in the system unit or in disk-only I/O drawers like the EXP24S. Although protecting drives is always recommended, AIX and Linux users can choose to leave a few or all drives unprotected at their own risk. IBM supports these configurations.

HDD/SSD drive protection can be provided by AIX and Linux, or by the available hardware controllers.

All three of these controller options can offer different drive protection options: RAID 0, RAID 5, RAID 6, or RAID 10. RAID 5 requires a minimum of three drives of the same capacity. RAID 6 requires a minimum of four drives of the same capacity. RAID 10 requires a minimum of two drives. Hot spare capability is supported by RAID 5 or RAID 6.

All three controller options offer Easy Tier functionality, which is also called *RAID 5T2* (2-tiered RAID 5), *RAID 6T2* (2-tiered RAID 6), and *RAID 10T2* (2-tiered RAID 10). The split disk backplane option supports RAID 10T2, but does not support RAID 5T2 or 6T2.

Table 2-25 details the drive protection options that are available with each storage controller option available with the Power E850C server.

Table 2-25 Drive protection capabilities of the Power E850C storage controller options

Protection type	Dual controller with write cache (#EPVN)	Dual controller without write cache (#EPVP)	Split disk backplane (#EPVQ)
JBOD	No	No	Yes
RAID 0/1	Yes	Yes	Yes
RAID 5/6/10	Yes	Yes	Yes
RAID 5T2 (Easy Tier)	Yes	Yes	No
RAID 6T2 (Easy Tier)	Yes	Yes	No
RAID 10T2 (Easy Tier)	Yes	Yes	Yes
Split backplane	No	No	Yes

2.10.3 Drive protection levels

The following are the different available levels of drive protection:

- Just a bunch of disks (JBOD) provides no drive protection.

  JBOD presents the drives as just a bunch of disks to the system. The failure of a single drive results in the loss of all data on that disk. Any data protection must be provided through the operating system or software.

- RAID 0 provides striping for performance, but does not offer any fault tolerance.

  The failure of a single drive results in the loss of all data on the array. This version of RAID increases I/O bandwidth by simultaneously accessing multiple data paths.

- RAID 1 provides mirroring for fault tolerance, but halves total drive capacity.

  The failure of a single drive results in no loss of data on the array. This version of RAID provides simple mirroring, requiring two copies of all data. However, this option has the highest cost because only half of the installed capacity is usable.

- RAID 5 uses block-level data striping with distributed parity.

  RAID 5 stripes both data and parity information across three or more drives. Fault tolerance is maintained by ensuring that the parity information for each block of data is placed on a drive that is separate from the ones that are used to store the data itself. This version of RAID provides data resiliency if a single drive fails in a RAID 5 array.

- RAID 6 uses block-level data striping with dual distributed parity.

  RAID 6 is the same as RAID 5 except that it uses a second level of independently calculated and distributed parity information for extra fault tolerance. A RAID 6 configuration requires N+2 drives to accommodate the additional parity data, making it less cost-effective than RAID 5 for equivalent storage capacity. This version of RAID provides data resiliency if one or two drives fail in a RAID 6 array. When you work with large capacity disks, RAID 6 allows you to sustain data parity during the rebuild process.

- RAID 10 is a striped set of mirrored arrays.

  RAID 10 is a combination of RAID 0 and RAID 1. A RAID 0 stripe set of the data is created across a two-disk array for performance benefits. A duplicate of the first stripe set is then mirrored on another two-disk array for fault tolerance. This version of RAID provides data resiliency if a single drive fails, and can provide resiliency for multiple drive failures.

RAID 5T2, RAID 6T2, and RAID 10T2 are RAID levels with EasyTier enabled. It requires that both types of disks exist on the system under the same controller (HDDs and SSDs) and that both are configured under the same RAID type.

If Easy Tier functionality is not being used, an array can consist of only one drive type. Therefore, you can have HDD only arrays and SDD only arrays.

It is possible to mix drive capacities in an array. However, if an array has multiple capacity points, the used capacity might be lower. For example, with a mix of 300 GB HDDs and 600 GB HDDs, only 300 GB of the larger 600 GB HDDs will be used. Similarly, if an array has both 387 GB SSDs and 775 GB SSDs, only 387 GB of the 775 GB SSDs will be used.

Note: The block size of the drives in an array must match, covering both HDDs and SSDs. So either all drives must be formatted with a 4 k block size, or all must be formatted with a 5 *xx* block size.

2.10.4 Easy Tier

The Power E850C server supports both HDDs and SSDs in the SFF bays, which are connected to the storage backplane and accessible from the front of the server. The 1.8-inch bays support only SSDs. It is possible to create multiple arrays using any of the storage controller options. However, you cannot mix HDDs and SSDs in a standard (non-Easy Tier) array. Instead, you would need to create separate arrays for HDDs and SSDs.

All of the storage controller options also support Easy Tier functionality, which allows them to support mixed arrays of HDDs and SSDs. When the SSDs and HDDs are under the same array, the adapter can automatically move the most accessed data to faster storage (SSDs) and less accessed data to slower storage (HDDs).

No coding or software intervention is needed after the RAID is configured correctly. Statistics on block accesses are gathered every minute. When the adapter detects that some portion of the data is being frequently requested, it moves this data to faster devices. The data is moved in chunks of 1 MB or 2 MB called *bands*.

By moving hot data onto faster SSDs, Easy Tier can dramatically improve the performance of applications that are limited by disk I/O operations. Cold data blocks are moved onto more cost effective HDDs, reducing the overall cost. Combining this process with the dual controller storage backplane with write cache can provide even higher levels of overall storage performance.

From the operating system point-of-view, there is just a regular array disk. From the SAS controller point-of-view, there are two arrays with parts of the data being serviced by one tier of disks and parts by another tier of disks.

Figure 2-14 shows a representation of an Easy Tier array.

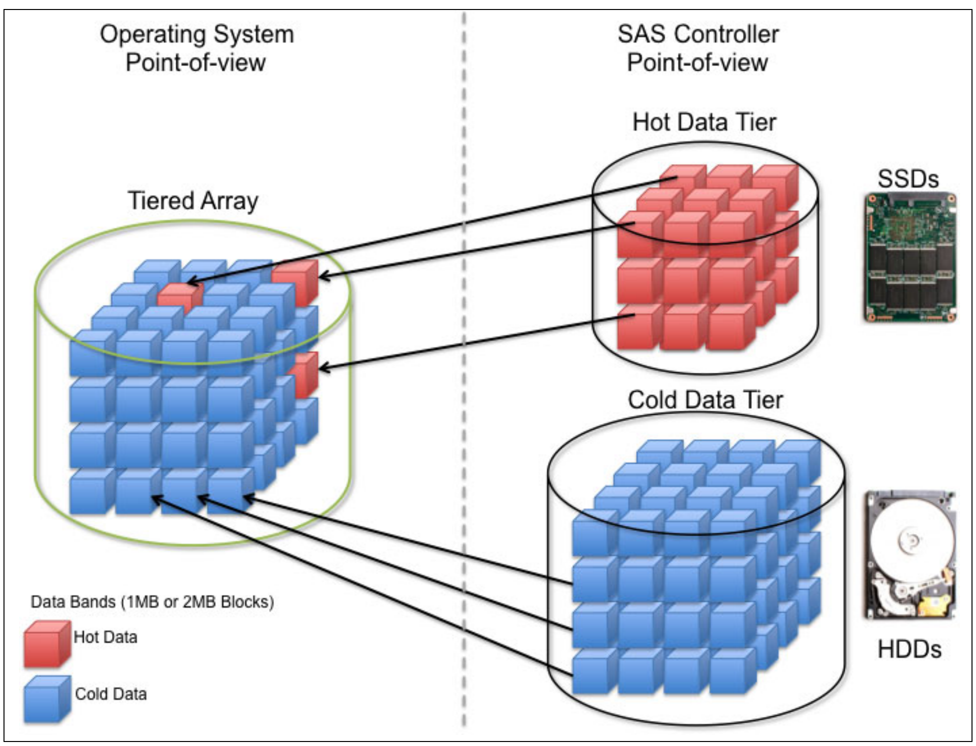

Figure 2-14 An Easy Tier array

The Easy Tier configuration is accomplished through a standard operating system SAS adapter configuration utility. Figure 2-15 shows an example of tiered array creation for AIX.

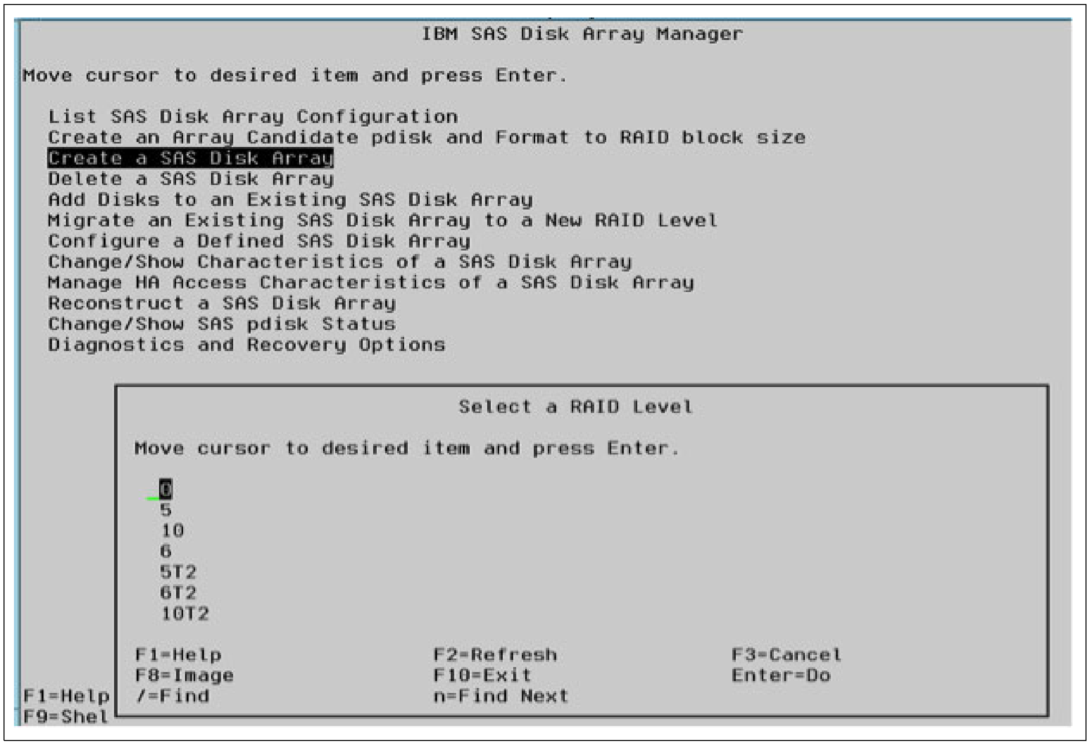

Figure 2-15 Array type selection panel on AIX RAID Manager

Figure 2-16 shows a created tiered array for AIX.

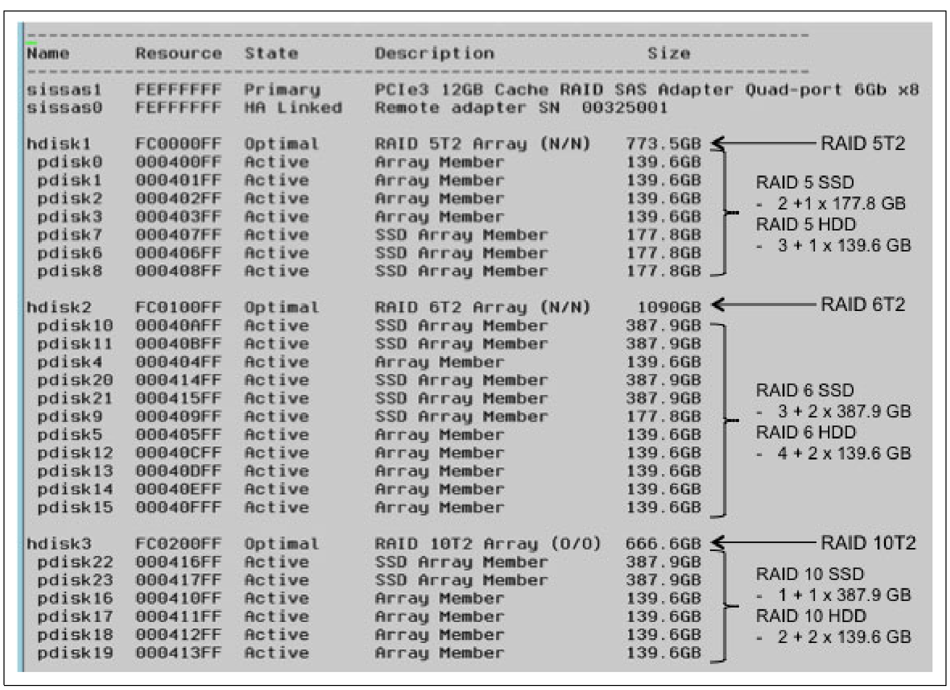

Figure 2-16 Tiered arrays (RAID 5T2, RAID 6T2, and RAID 10T2) example on AIX RAID Manager

Each Easy Tier array is made up of two individual arrays of the same type, one populated with HDDs and one with SSDs. So a RAID 5T2 Array is formed of a RAID 5 array of HDDs and a RAID 5 array of SSDs. As such, the following minimum device quantities required:

- RAID 5T2 requires at least three HDDs and three SSDs
- RAID 6T2 requires at least four HDDs and four SSDs
- RAID 10T2 requires at least two HDDs and two SSDs

The HDD and SSD can be different capacities in an Easy Tier array. However, if either half of the array has multiple capacity points, the used capacity might be lower. For example, with a mix of 300 GB HDDs and 600 GB HDDs, only 300 GB of the larger 600 GB HDDs are used. Similarly, if an array has both 387 GB SSDs and 775 GB SSDs, only 387 GB of the 775 GB SSDs are used. A combination of 600 GB HDDs and 387 GB SSDs allows you to use the full capacity of all devices.

Note: The block size of the drives in an array must match, covering both HDDs and SSDs. So either all drives must be formatted with a 4 k block size, or all must be formatted with a 5 *xx* block size.

2.10.5 Internal disk options

Each of the controller options that are available provides a backplane with eight SFF-3 disk bays in the server. These 2.5-inch (SFF) SAS bays can support both HDDs and SSDs mounted in a Gen3 carrier. Previous generation SFF disk types (SFF-1 and SFF-2) do not fit in these bays. All SFF-3 bays support concurrent maintenance or hot-plug capability. All three of the controller options support HDDs or SSDs or a mixture of HDDs and SSDs in the SFF-3 bays. If you mix HDDs and SSDs, they must be in separate arrays (unless using the Easy Tier function).

The storage backplane also has four 1.8-inch storage bays. These bays can hold 1.8-inch SSDs to provide high-performance storage for Easy Tier arrays.

Table 2-26 lists the drives that are available for the internal bays of the Power E850C server.

Table 2-26 Drives available for internal storage bays on the Power E850C

Feature Code	Capacity	Type	Placement
ES0L	387 GB	SSD	SSF-3 bay
ES0N	775 GB	SSD	SSF-3 bay
ES0U	387 GB	4k block SSD	SFF-3 bay
ES0W	775 GB	4k block SSD	SFF-3 bay
ES16	387 GB	SSD	1.8-inch bay
ES1C	387 GB	5xx SSD eMLC4	1.8-inch bay
ES2V	387 GB	4k SSD eMLC4	1.8-inch bay
ES2X	775 GB	5xx SSD eMLC4	1.8-inch bay
ES4K	775 GB	4k SSD eMLC4	1.8-inch bay
ES7K	387 GB	SSD 5xx eMLC4	SFF-3 bay
ES7P	775 GB	SSD 5xx eMLC4	SFF-3 bay

Feature Code	Capacity	Type	Placement
ES8J	1.9 TB	4k block Read Intensive SSD	SFF-3 bay
ES8N	387 GB	SSD 4k eMLC4	SFF-3 bay
ES8Q	775 GB	SSD 4k eMLC4	SFF-3 bay
ES8V	1.55 TB	SSD 4k eMLC4	SFF-3 bay
ESD5	600 GB	10k RPM HDD	SFF-3 bay
ESD9	1.2 TB	10k RPM HDD	SFF-3 bay
ESDB	300 GB	15k RPM HDD	SFF-3 bay
ESDF	600 GB	15k RPM HDD	SFF-3 bay
ESDR	300 GB	10k RPM HDD	SFF-3 bay
ESDT	146 GB	15k RPM HDD	SFF-3 bay
ESF5	600 GB	4k block 10k RPM HDD	SFF-3 bay
ESF9	1.2 TB	4k block 10k RPM HDD	SFF-3 bay
ESFB	300 GB	4k block 15k RPM HDD	SFF-3 bay
ESFF	600 GB	4k block 15k RPM HDD	SFF-3 bay
ESFV	1.8 TB	4k block 10k RPM HDD	SFF-3 bay

If you want to expand the storage capability of the server, you need to add an EXP24S expansion drawer using a PCIe SAS adapter. This expansion drawer handles further SFF HDDs and SSDs. For more information about the EXP24S expansion drawer, see 2.12.1, "EXP24S SFF Gen2-bay Drawer" on page 69.

2.10.6 Media drives

A slimline media bay is included with all of the controller options on the Power E850C server. This bay can hold the optional DVD-RAM drive (#5771). If using one of the two dual controller options, both integrated controllers can access the DVD device if fitted. If using the split disk backplane option, only one controller can access the DVD. This is the controller that has access to the top disks in the system.

A DVD drive can be included in the system. It is then available to perform operating system installation, maintenance, problem determination, and service actions, such as maintaining system firmware and I/O microcode at their latest levels. Alternatively, the system must be attached to a network with software such as AIX Network Installation Management (NIM) or Linux Install Manager to perform these functions.

The Power E850C supports the RDX USB External Docking Station for Removable Disk Cartridge (#EUA4). The USB External Docking Station accommodates RDX removable disk cartridge of any capacity. The disks are in a protective rugged cartridge enclosure that plug into the docking station. The docking station holds one removable rugged disk drive/cartridge at a time. The rugged removable disk cartridge and docking station backs up similar to tape drive. This can be an excellent alternative to DAT72, DAT160, 8 mm, and VXA-2 and VXA-320 tapes.

Table 2-27 shows the available media device feature codes for the Power E850C server.

Table 2-27 Media device feature code descriptions for Power E850C

Feature code	Description
5771	SATA Slimline DVD-RAM Drive
EUA4	RDX USB External Docking Station for Removable Disk Cartridge

2.11 External I/O subsystems

This section describes the PCIe Gen3 I/O expansion drawer that can be attached to the Power E850C server.

2.11.1 PCIe Gen3 I/O expansion drawer

PCIe Gen3 I/O Expansion drawers (#EMX0) can be attached to the system unit to expand the number of full-high, hot-swap Gen3 slots available to the server. The maximum number of PCIe Gen3 I/O drawers depends on the number of processor modules physically installed. The maximum is independent of the number of processor core activations.

The PCIe Gen3 I/O expansion drawer is a 4U high, PCI Gen3-based and 19-inch rack mountable I/O drawer. It offers two PCIe Fan Out Modules (#EMXF), each of which provides six full-length, full-height PCIe slots that are labeled C1 through C6. Slots C1 and C4 are x16 slots, and C2, C3, C5, and C6 are x8 slots. Slots C1 and C4 of the 6-slot fanout module in a PCIe Gen3 I/O drawer are SR-IOV enabled.

An #EMX0 drawer can be configured with one or two EMXF fanout modules. Adding a second fanout module is not a hot-plug operation and requires scheduled downtime.

The physical dimensions of the drawer are 444.5 mm (17.5 in.) wide by 177.8 mm (7.0 in.) high by 736.6 mm (29.0 in.) deep for use in a 19-inch rack.

A PCIe x16 to Optical CXP converter adapter (#EJ08) and two 3.0 m (#ECC7), or two 10.0 m (#ECC8) CXP 16X Active Optical cables (AOC) connect the system node to a PCIe Fan Out module in the I/O expansion drawer. One feature #ECC7, or one #ECC8 ships two AOC cables.

Concurrent repair and add or removal of PCIe adapters is done by HMC guided menus or by operating system support utilities.

A blind swap cassette (BSC) is used to house the full high adapters that go into these slots. The BSC is the same BSC as used with the previous generation server's #5802/5803/5877/5873 12X attached I/O drawers.

A maximum of four PCIe Gen3 I/O drawers can be attached to the Power E850C server, if equipped with four processor modules. The maximum number of PCIe Gen3 I/O drawers depends on the number of installed processor modules. Table 2-28 on page 63 lists the maximum PCIe Gen3 I/O drawer configurations for the Power E850C server.

Table 2-28 Maximum PCIe Gen3 I/O drawer configurations

Power E850C configuration	Maximum number of attached PCIe Gen3 I/O drawers	Maximum number of PCIe slots on the server
Power E850C with two processor modules	2	27
Power E850C with three processor modules	3	39
Power E850C with four processor modules	4	51

Figure 2-17 shows the back view of the PCIe Gen3 I/O expansion drawer.

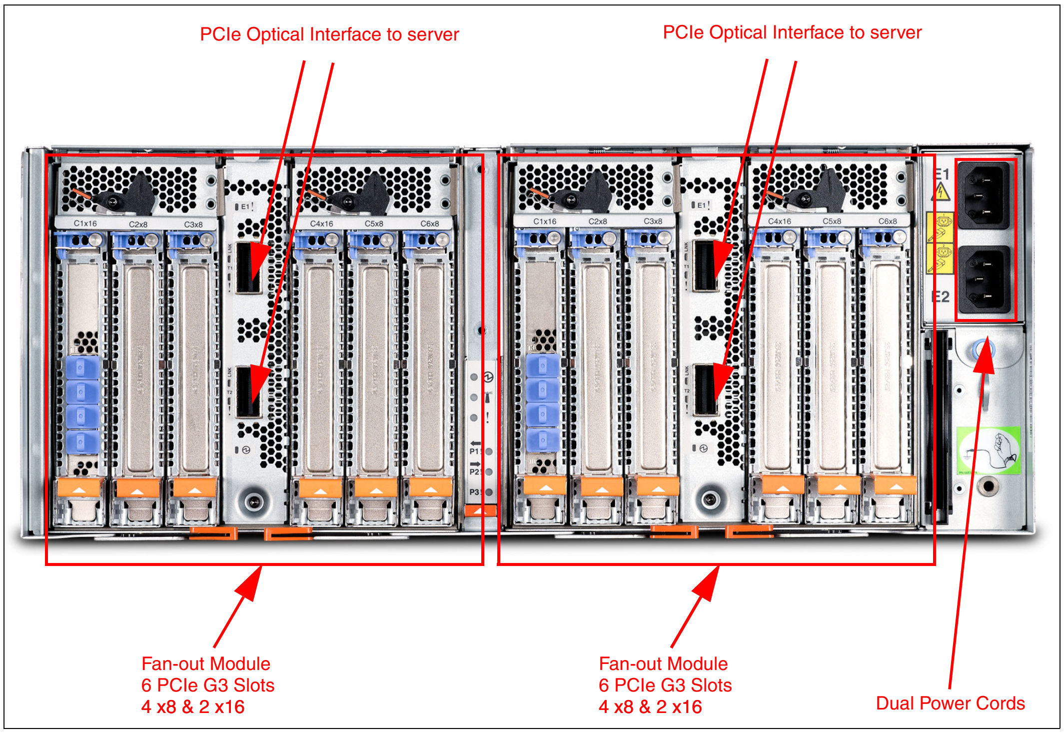

Figure 2-17 Rear view of the PCIe Gen3 I/O expansion drawer

PCIe Gen1, Gen2, and Gen3 full-high adapters are supported. The set of full-high PCIe adapters that are supported is found in the Sales Manual, which is identified by feature number. See the PCI Adapter Placement manual for the 8408-E8E and 44E for details and rules associated with specific adapters supported and their supported placement in x8 or x16 slots, found at:

http://www.ibm.com/support/knowledgecenter/8408-E8E/p8eab/p8eab_85x_slot_details.htm

2.11.2 PCIe Gen3 I/O expansion drawer optical cabling

I/O drawers are connected to the adapters in the server with data transfer cables:

- 3.0 m Optical Cable Pair for PCIe3 Expansion Drawer (#ECC7)
- 10.0 m Optical Cable Pair for PCIe3 Expansion Drawer (#ECC8)

Recommendation: Locate any attached PCIe Gen3 I/O Expansion drawer in the same rack as the POWER8 server for ease of service. However, the drawers can be installed in separate racks if the application or other rack content requires it. Generally, use 3 m cables for PCIe drawers in the same rack as the system unit and 10 m cables for drawers that are in a different rack.

A minimum of one PCIe3 Optical Cable Adapter for PCIe3 Expansion Drawer (#EJ08) is required to connect to the PCIe3 6-slot Fan Out module in the I/O expansion drawer. The top port of the fan-out module must be cabled to the top port of the #EJ08 port. Likewise, the bottom two ports must be cabled together:

1. Connect an active optical cable to connector T1 on the PCIe3 optical cable adapter in your server.
2. Connect the other end of the optical cable to connector T1 on one of the PCIe3 6-slot Fan Out modules in your expansion drawer.
3. Connect another cable to connector T2 on the PCIe3 optical cable adapter in your server.
4. Connect the other end of the cable to connector T2 on the PCIe3 6-slot Fan Out module in your expansion drawer.
5. Repeat the preceding four steps for the other PCIe3 6-slot Fan Out module in the expansion drawer, if required.

Figure 2-18 shows connector locations for the PCIe Gen3 I/O expansion drawer.

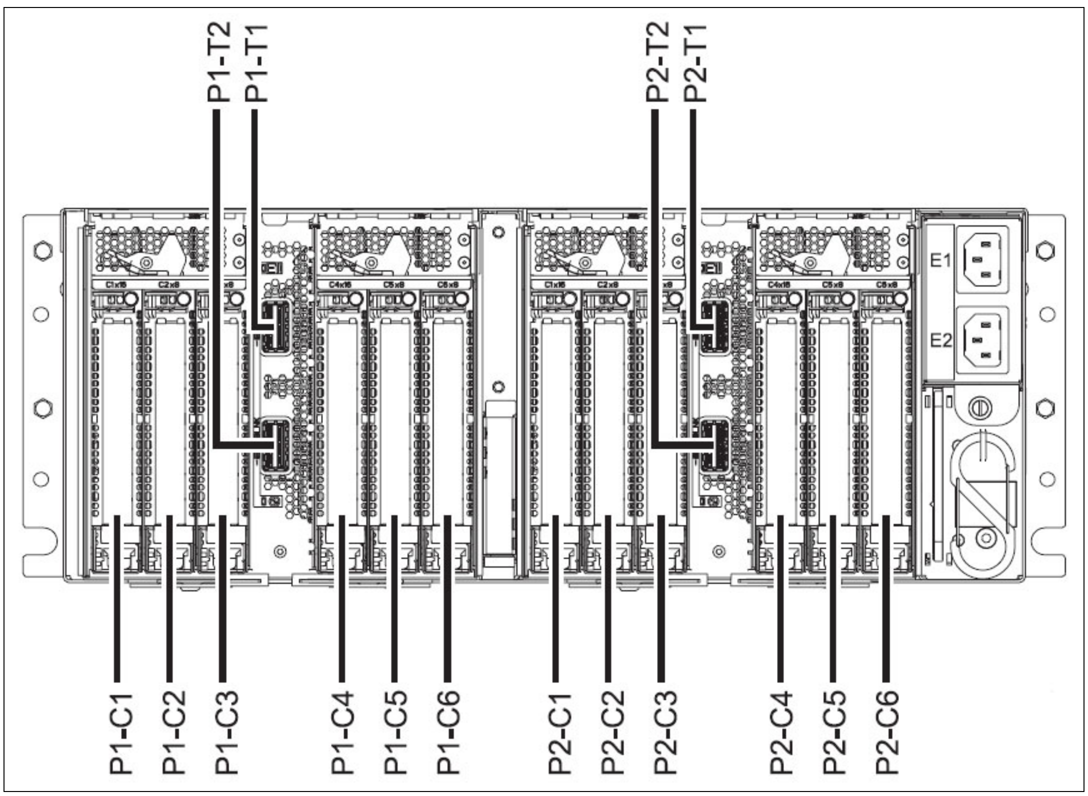

Figure 2-18 Connector locations for the PCIe Gen3 I/O expansion drawer

Figure 2-19 shows typical optical cable connections.

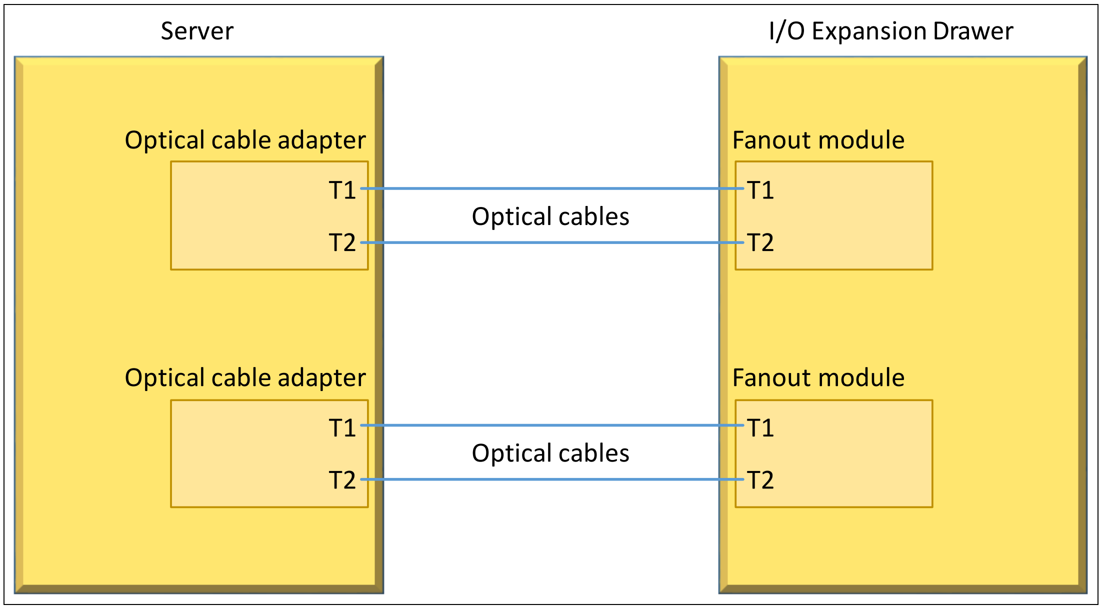

Figure 2-19 Typical optical cable connection

General rules for the PCI Gen3 I/O expansion drawer configuration

The PCIe3 optical cable adapter can be in any of the x16 PCIe Gen3 adapter slots in the Power E850C system node. However, generally use the PCIe adapter slot priority information while selecting slots for installing PCIe3 Optical Cable Adapter (#EJ08).

Table 2-29 shows PCIe adapter slot priorities in the Power E850C server.

Table 2-29 PCIe adapter slot priorities

Feature code	Description	Slot priorities		
		Two processor modules	Three processor modules	Four processor modules
EJ08	PCIe3 Optical Cable Adapter for PCIe3 Expansion Drawer	9, 7, 11, 8[a]	9, 7, 4, 11, 8, 3[a]	9, 7, 4, 2, 11, 8, 3, 1[a]

a. For information about how the slot numbers listed relate to physical location codes, see Table 2-14 on page 47.

The following figures show several examples of supported configurations. For simplicity, these figures do not show every possible combination of the I/O expansion drawer to server attachments.

Figure 2-20 shows an example of a Power E850C with two processor modules and a maximum of two PCIe Gen3 I/O expansion drawers.

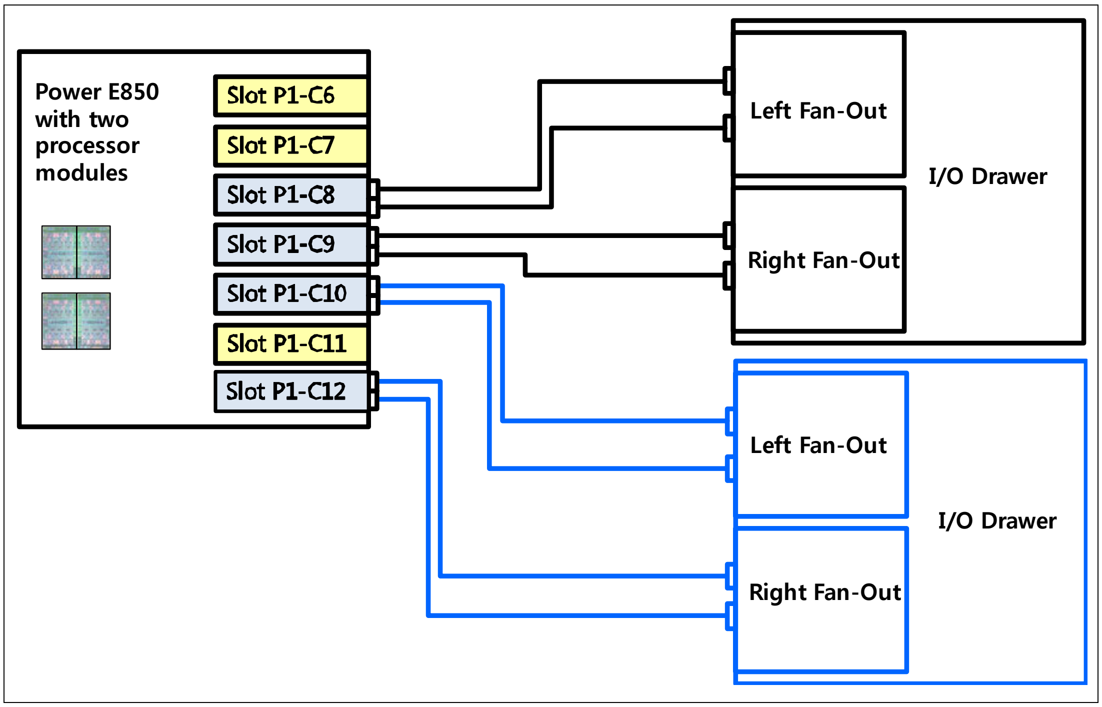

Figure 2-20   Example of a Power E850C and a maximum of two I/O drawers

Figure 2-21 shows an example of a Power E850C with three processor modules and a maximum of three PCIe Gen3 I/O expansion drawers.

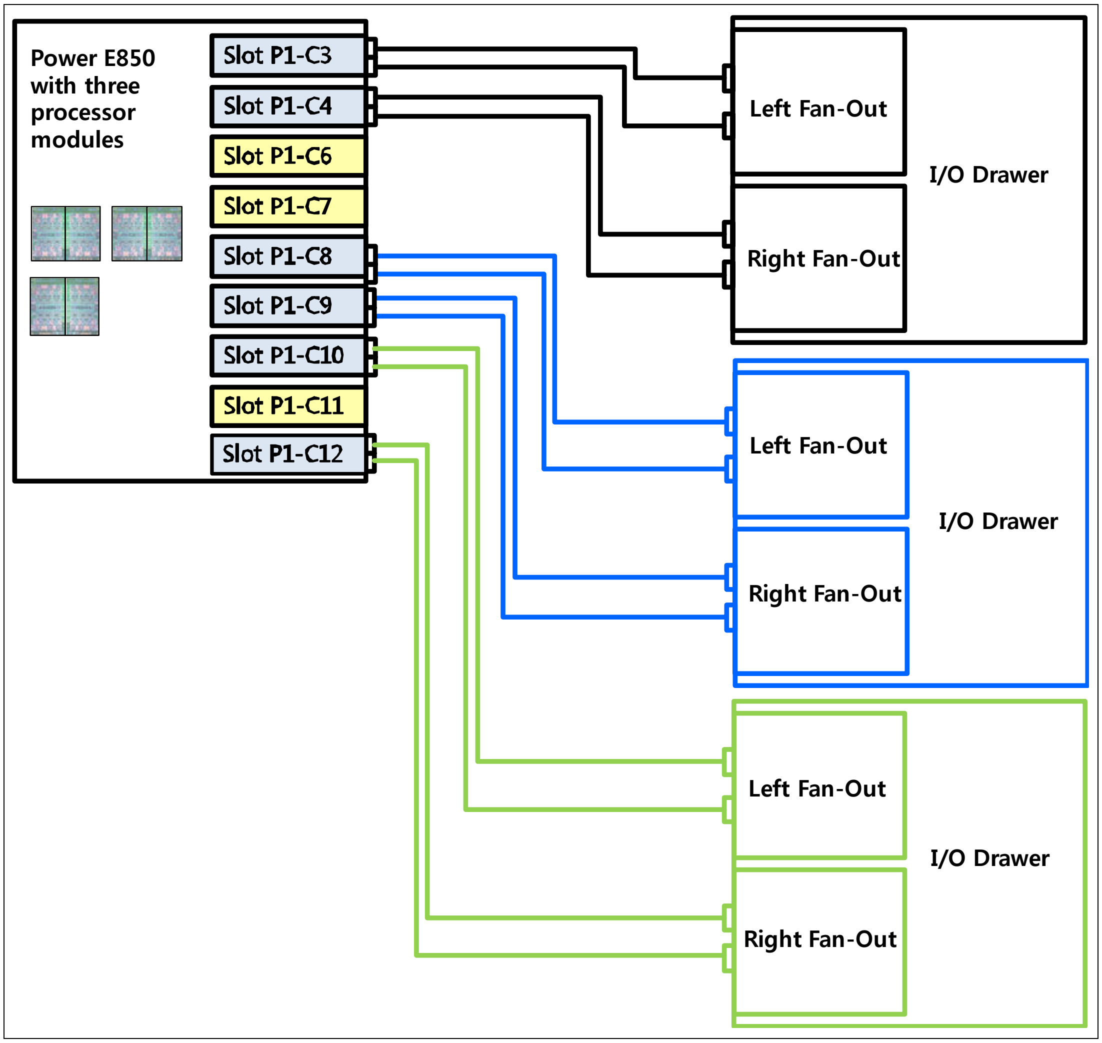

Figure 2-21 Example of a Power E850C and a maximum of three PCI Gen3 I/O expansion drawers

Figure 2-22 shows an example of Power E850C with four processor modules and a maximum of four PCIe Gen3 I/O expansion drawers.

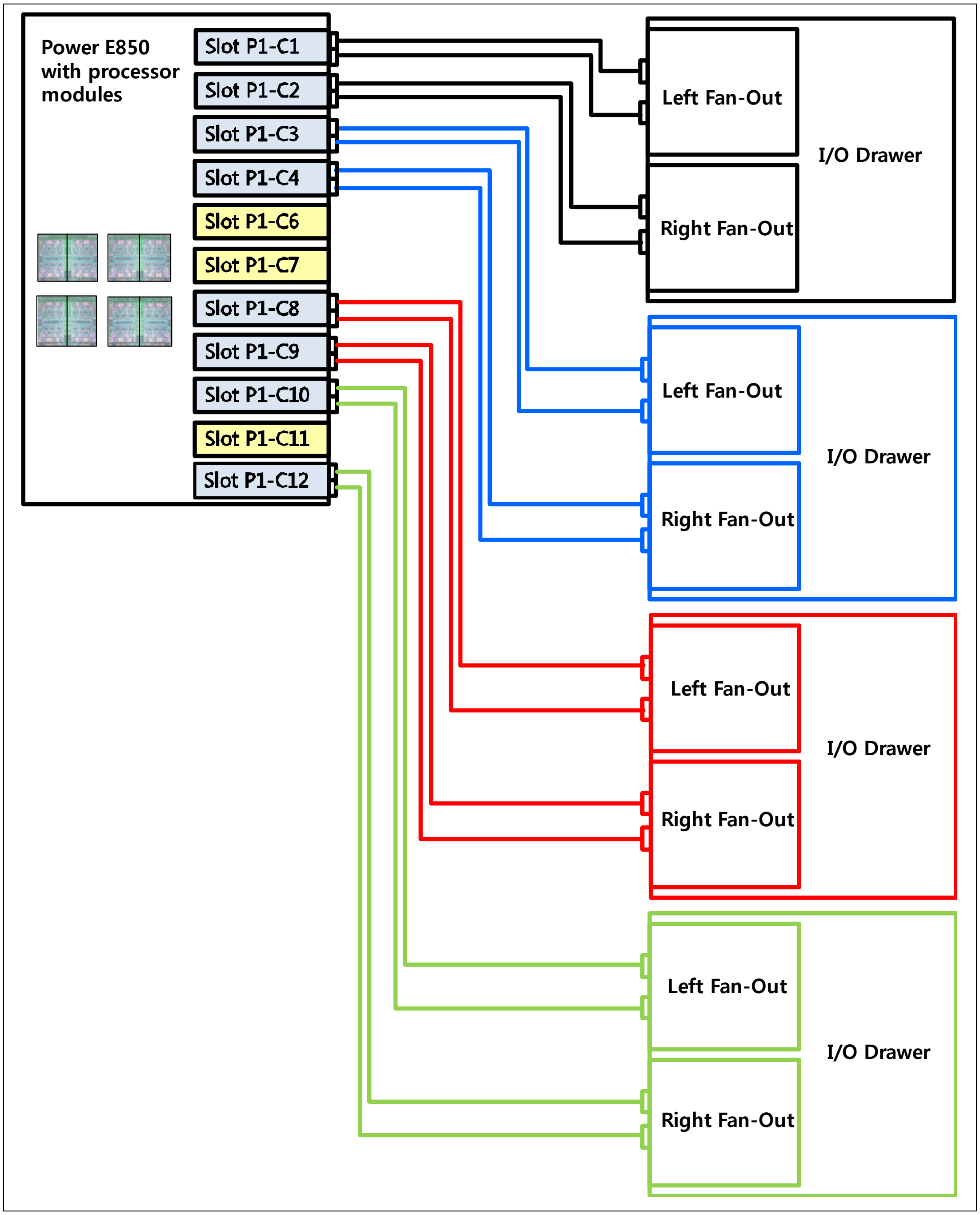

Figure 2-22 Example of a Power E850C and a maximum of four I/O drawers

2.11.3 PCIe Gen3 I/O expansion drawer SPCN cabling

There is no system power control network (SPCN) used to control and monitor the status of power and cooling within the I/O drawer. SPCN capabilities are integrated in the optical cables.

2.12 External disk subsystems

This section describes the following external disk subsystems that can be attached to the Power E850C system:

- EXP24S SFF Gen2-bay Drawer for high-density storage (#5887)
- IBM System Storage®

Note:

- The EXP30 Ultra SSD Drawer (#EDR1 or #5888), the EXP12S SAS Disk Drawer (#5886), and the EXP24 SCSI Disk Drawer (#5786) are not supported on the Power E850C server.
- IBM offers a 1U multimedia drawer that can hold one or more DVD drives, tape drives, or RDX docking stations. The 7226-1U3 is the most current offering. The earlier 7216-1U2 and 7214-1U2 are also supported. Up to six of these multimedia drawers can be attached by using a PCIe SAS adapter.

2.12.1 EXP24S SFF Gen2-bay Drawer

The EXP24S SFF Gen2-bay Drawer (#5887) is an expansion drawer with twenty-four 2.5-inch small form-factor SAS bays. The EXP24S supports up to 24 hot-swap SFF-2 SAS HDDs or SSDs. It uses only 2 U (2 EIA units) of space in a 19-inch rack. The EXP24S includes redundant ac power supplies and uses two power cords.

Note: A maximum of 64 EXP24S drawers can be attached to the Power E850C server, providing an extra quantity of 1536 disks.

To further reduce possible single points of failure, EXP24S configuration rules consistent with previous Power Systems are used. All Power operating system environments that are using SAS adapters with write cache require the cache to be protected by using pairs of adapters.

With AIX, Linux, and VIOS, you can order the EXP24S with four sets of six bays, two sets of 12 bays, or one set of 24 bays (mode 4, 2, or 1). Figure 2-23 shows the front of the unit and the groups of disks on each mode.

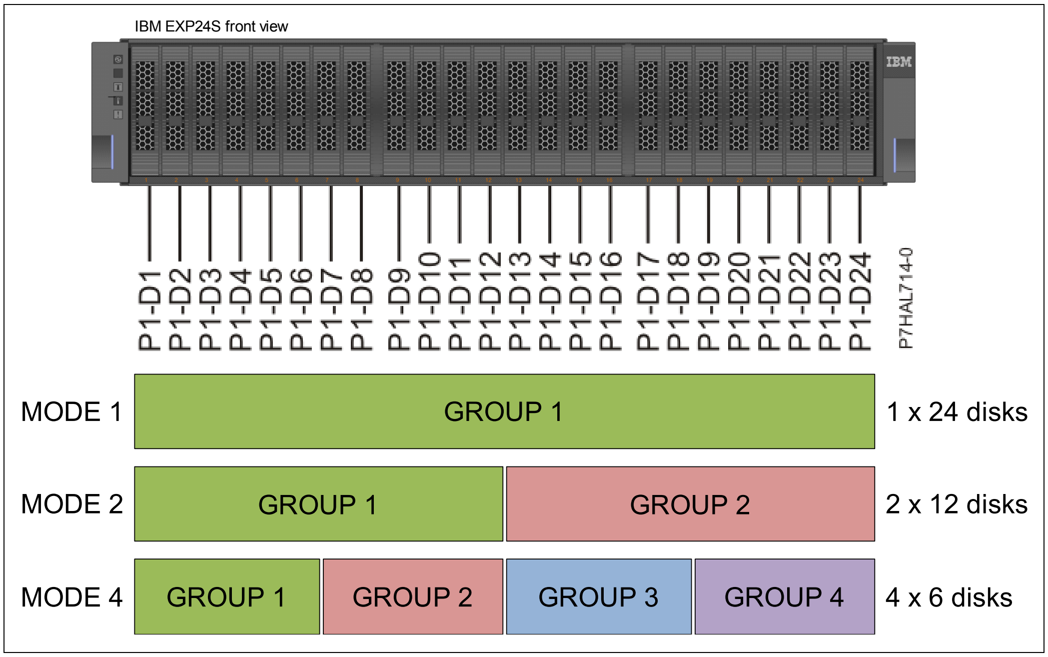

Figure 2-23 EXP24S front view with location codes and disk groups depending on its mode of operation

Mode setting is done by IBM manufacturing. If you need to change the mode after installation, ask your IBM support representative to refer to the following site:

http://w3.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS5121

The stickers indicate whether the enclosure is set to mode 1, mode 2, or mode 4. They are attached to the lower-left shelf of the chassis (A) and the center support between the enclosure services manager modules (B).

Figure 2-24 shows the mode stickers.

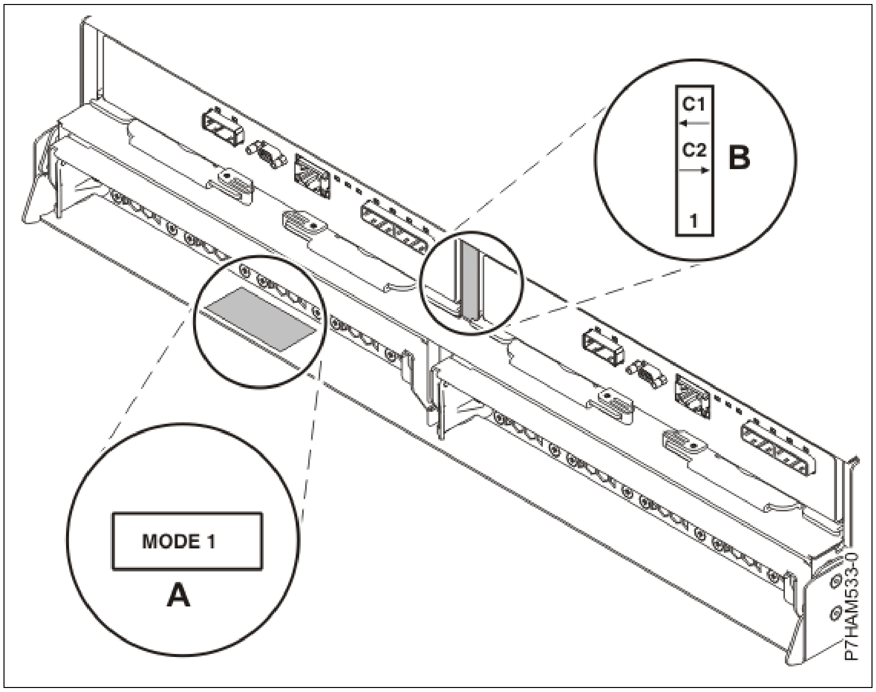

Figure 2-24 Mode sticker locations at the rear of the 5887 disk drive enclosure

The EXP24S SAS ports are attached to a SAS PCIe adapter or pair of adapters by using SAS YO or X cables. Cable length varies depending on the feature code. Calculate the proper length considering routing for proper airflow and ease of handling. A diagram of both types of SAS cables can be seen in Figure 2-25.

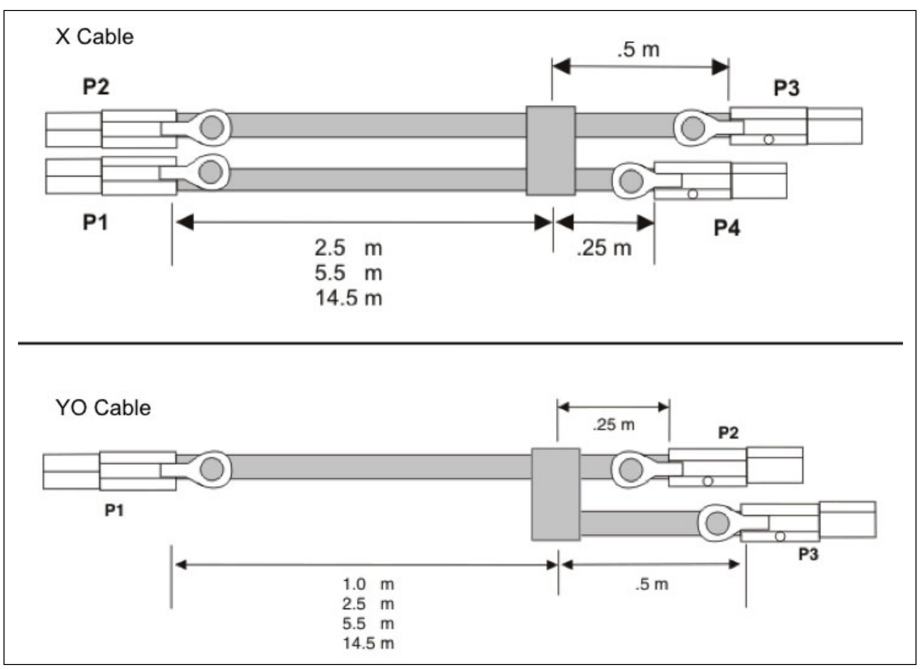

Figure 2-25 Diagram of SAS cable types X and YO

The following SAS adapters support the EXP24S:

- PCIe Dual-x4 SAS Adapter (#5901)
- PCIe3 RAID SAS Adapter Quad-port 6 Gb x8 (#EJ0J)
- PCIe3 12GB Cache RAID SAS Adapter Quad-port 6 Gb x8 (#EJ0L)
- PCIe3 SAS Tape/DVD Adapter Quad-port 6 Gb x8 (#EJ10)

The EXP24S drawer can support up to 24 SAS SFF Gen-2 disks. Table 2-30 lists the available disk options.

Table 2-30 Available disks for the EXP24S

Feature code	Description	Max per server	OS support
ES0G	775 GB SFF-2 SSD for AIX/Linux	768	AIX, Linux
ES0Q	387 GB SFF-2 4 K SSD for AIX/Linux	768	AIX, Linux
ES0S	775 GB SFF-2 4 K SSD for AIX/Linux	768	AIX, Linux
ES19	387 GB SFF-2 SSD for AIX/Linux	768	AIX, Linux
1917	146 GB 15 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	1536	AIX, Linux
1925	300 GB 10 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	1536	AIX, Linux
1953	300 GB 15 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	1536	AIX, Linux
1964	600 GB 10 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	1536	AIX, Linux
ES0Q	387 GB SFF-2 SSD 5xx eMLC4 for AIX/Linux	768	AIX, Linux
ES78	775 GB SFF-2 SSD 5xx eMLC4 for AIX/Linux	768	AIX, Linux
ES7E	1.9 TB Read Intensive SAS 4k SFF-2 SSD for AIX/Linux	768	AIX, Linux
ES80	387 GB SFF-2 SSD 4k eMLC4 for AIX/Linux	768	AIX, Linux
ES85	775 GB SFF-2 SSD 4k eMLC4 for AIX/Linux	768	AIX, Linux
ES8C	1.55 TB SFF-2 SSD 4k eMLC4 for AIX/Linux	768	AIX, Linux
ESD3	1.2 TB 10 K RPM SAS SFF-2 Disk Drive (AIX/Linux)	1536	AIX, Linux
ESDP	600 GB 15 K RPM SAS SFF-2 Disk Drive - 5xx Block (AIX/Linux)	1536	AIX, Linux
ESEV	600 GB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4096	1536	AIX, Linux
ESEZ	300 GB 15 K RPM SAS SFF-2 4 K Block - 4096 Disk Drive	1536	AIX, Linux
ESF3	1.2 TB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4096	1536	AIX, Linux
ESFP	600 GB 15 K RPM SAS SFF-2 4 K Block - 4096 Disk Drive	1536	AIX, Linux
ESFT	1.8 TB 10 K RPM SAS SFF-2 Disk Drive 4 K Block - 4096	1536	AIX, Linux

There are six SAS connectors on the rear of the EXP24S drawer to which two SAS adapters or controllers are attached. They are labeled T1, T2, and T3. There are two T1, two T2, and two T3 connectors. While configuring the drawer, special configuration feature codes indicate for the plant the mode of operation in which the disks and ports are split:

- In mode 1, two or four of the six ports are used. Two T2 ports are used for a single SAS adapter, and two T2 and two T3 ports are used with a paired set of two adapters or dual adapters configuration.
- In mode 2 or mode 4, four ports are used, two T2 and two T3 to access all SAS bays.

Figure 2-26 shows the rear connectors of the EXP24S drawer, how they relate to the modes of operation, and disk grouping.

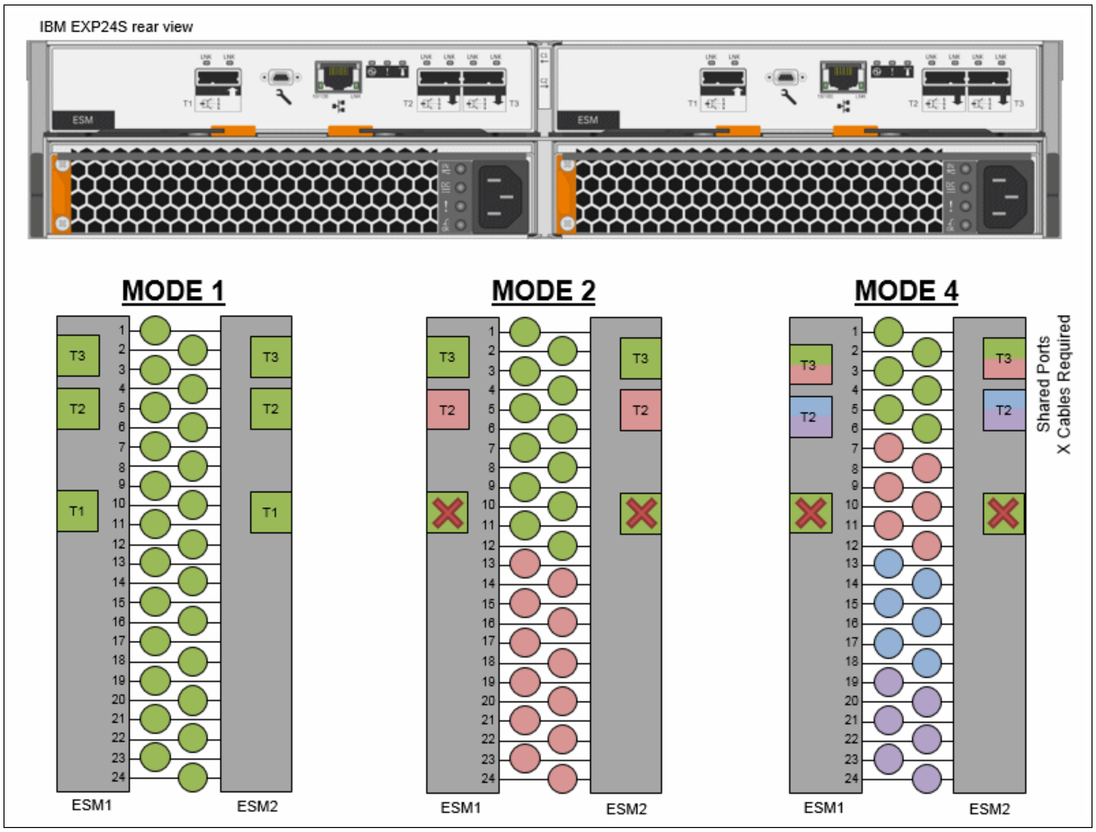

Figure 2-26 Rear view of EXP24S with the 3 modes of operation and the disks assigned to each port

An EXP24S drawer in mode 4 can be attached to two or four SAS controllers and provide high configuration flexibility. An EXP24S in mode 2 has similar flexibility. Up to 24 HDDs can be supported by any of the supported SAS adapters or controllers.

The most common configurations for EXP24S with Power Systems are detailed in 2.12.2, “EXP24S common usage scenarios” on page 74. Not all possible scenarios are included. For more information about SAS cabling and cabling configurations, search for “Planning for serial-attached SCSI cables” in the IBM Knowledge Center, which can be accessed at:

http://www.ibm.com/support/knowledgecenter/POWER8/p8hdx/POWER8welcome.htm

2.12.2 EXP24S common usage scenarios

The EXP24S drawer is versatile in the ways that it can be attached to Power Systems. This section describes the most common usage scenarios for EXP24S and Virtual I/O Servers, using standard PCIe SAS adapters (#5901).

Note: Not all possible scenarios are included. See the “Planning for serial-attached SCSI cables” guide in the IBM Knowledge Center to see more supported scenarios.

Scenario 1: Basic non-redundant connection

This scenario assumes a single Virtual I/O Server with a single PCIe SAS adapter #5901 and an EXP24S set on mode 1, allowing for up to 24 disks to be attached to the server. Figure 2-27 shows the connection diagram and components of the solution.

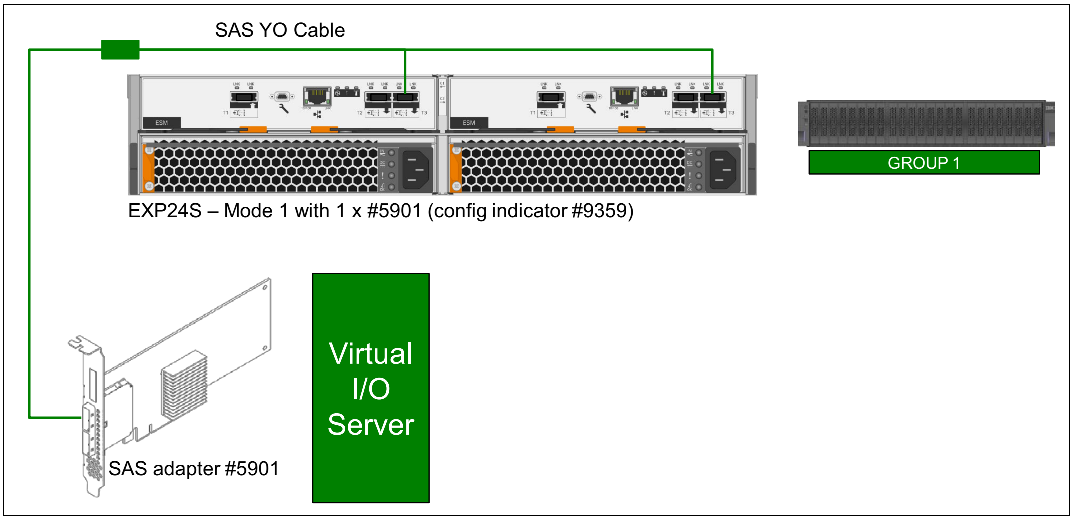

Figure 2-27 Scenario 1: Basic non-redundant connection

The following are the required feature codes for this scenario:

- One EXP24S drawer #5887 with indicator feature #9359 (mode 1 with single #5901)
- One PCIe SAS adapter #5901
- One SAS YO cable 3 Gbps with proper length

Scenario 2: Basic redundant connection

This scenario assumes a single Virtual I/O Server with two PCIe SAS adapters #5901 and an EXP24S set on mode 1. This configuration allows for up to 24 disks to be attached to the server. Figure 2-28 shows the connection diagram and components of the solution.

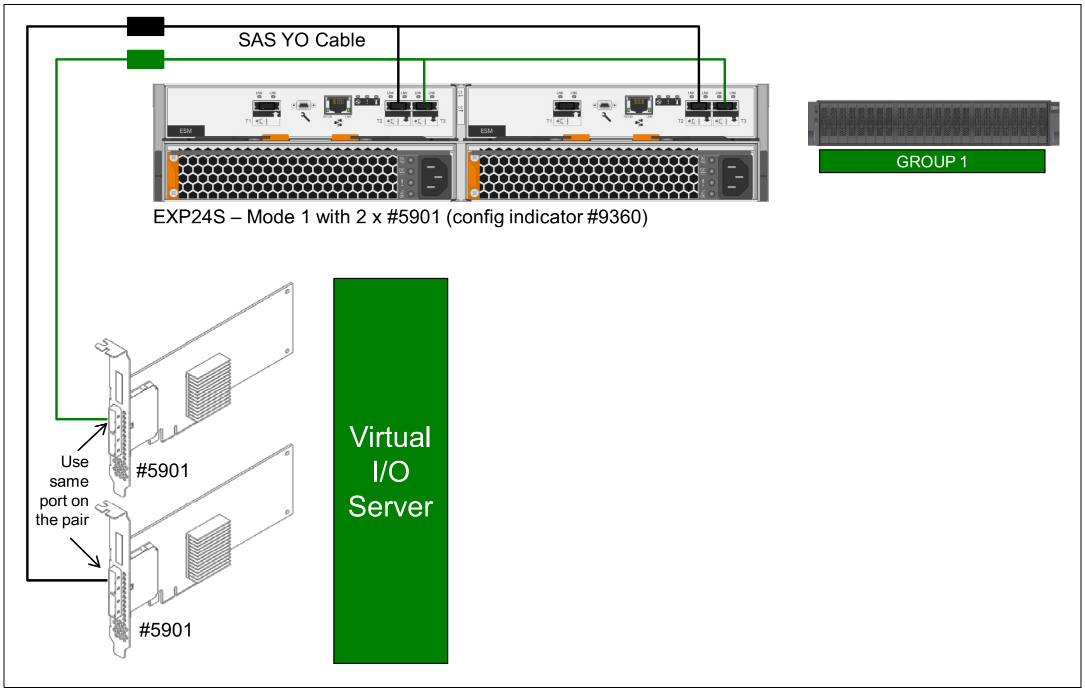

Figure 2-28 Scenario 2: Basic redundant connection

The following are the required feature codes for this scenario:

- One EXP24S drawer #5887 with indicator feature #9360 (mode 1 with dual #5901)
- Two PCIe SAS adapter #5901
- Two SAS YO cables 3 Gbps with proper length

The ports that are used on the SAS adapters must be the same for both adapters of the pair. This scenario provides no SSD support.

Scenario 3: Dual Virtual I/O Servers sharing a single EXP24S

This scenario assumes a dual Virtual I/O Server with two PCIe SAS adapters #5901 each and an EXP24S set on mode 2. This configuration allows for up to 12 disks to be attached to each Virtual I/O Server. Figure 2-29 shows the connection diagram and components of the solution.

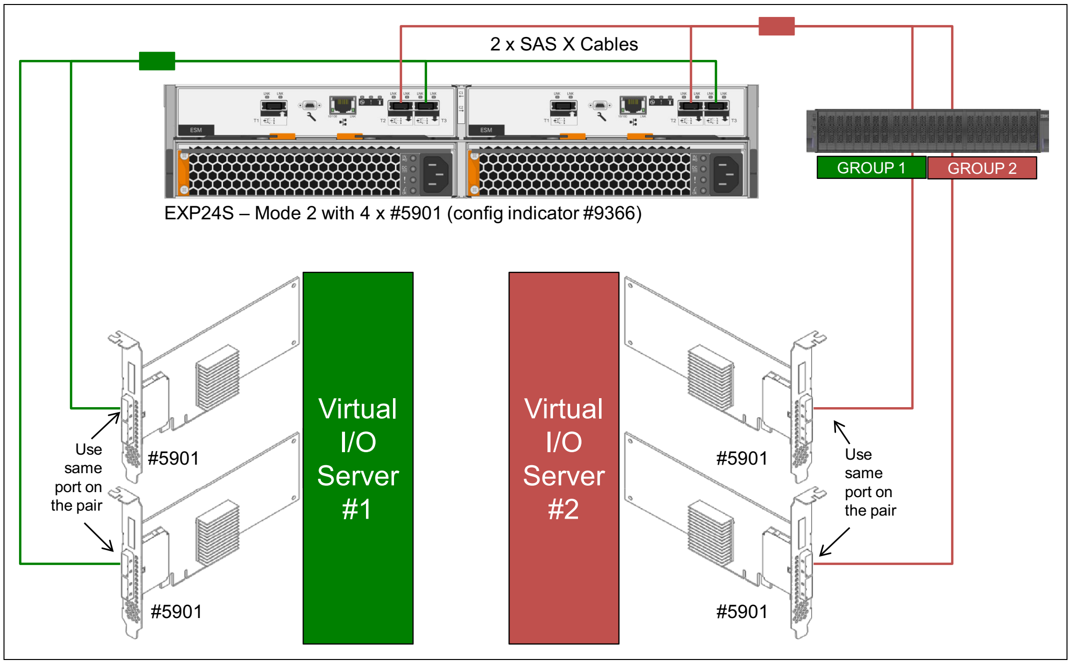

Figure 2-29 Dual Virtual I/O Servers sharing a single EXP24S

The following are the required feature codes for this scenario:

- One EXP24S drawer #5887 with indicator feature #9366 (mode 2 with quad #5901)
- Four PCIe SAS adapter #5901
- Two SAS X cables 3 Gbps with proper length

The ports that are used on the SAS adapters must be the same for both adapters of the pair. This scenario provides no SSD support.

Scenario 4: Dual Virtual I/O Servers sharing two EXP24S

This scenario assumes a dual Virtual I/O Server with two PCIe SAS adapters #5901 each and two EXP24S set on mode 2. This configuration allows for up to 24 disks to be attached to each Virtual I/O Server (2 per drawer). If compared to scenario 3, this scenario has the benefit of allowing disks from different EXP24S drawers to be mirrored. This mirroring allows for hot maintenance of the whole EXP24S drawers if all data is properly mirrored. Figure 2-30 shows the connection diagram and components of the solution.

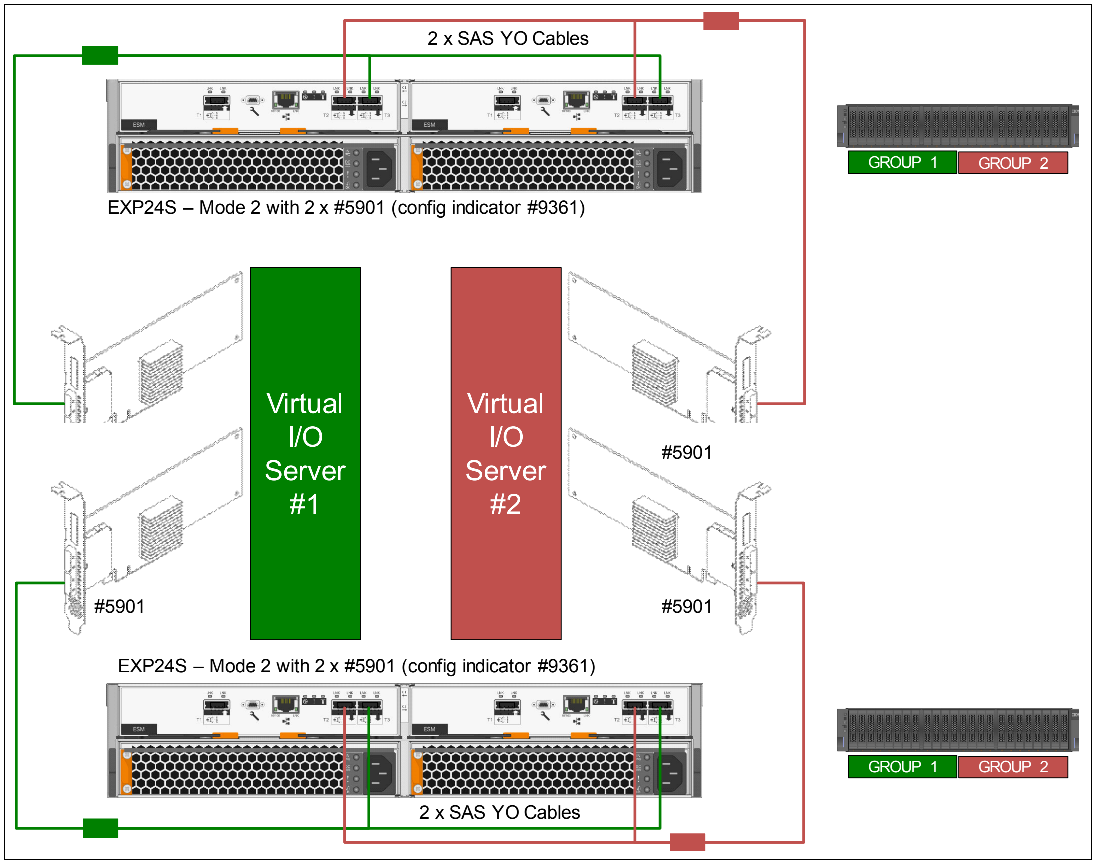

Figure 2-30 Dual Virtual I/O Servers sharing two EXP24S

The following are the required feature codes for this scenario:

- Two EXP24S drawers #5887 with indicator feature #9361 (mode 2 with dual #5901)
- Four PCIe SAS adapter #5901
- Four SAS YO cables 3 Gbps with proper length

This scenario provides no SSD support.

Scenario 5: Four Virtual I/O Servers sharing two EXP24S

This scenario assumes four Virtual I/O Servers with two PCIe SAS adapters #5901 each and two EXP24S set on mode 4. This configuration allows for up to 12 disks to be attached to each Virtual I/O Server (6 per drawer). This scenario allows disks from different EXP24S drawers to be mirrored, allowing for hot maintenance of the whole EXP24S drawers if all data is properly mirrored. Figure 2-31 shows the connection diagram and components of the solution.

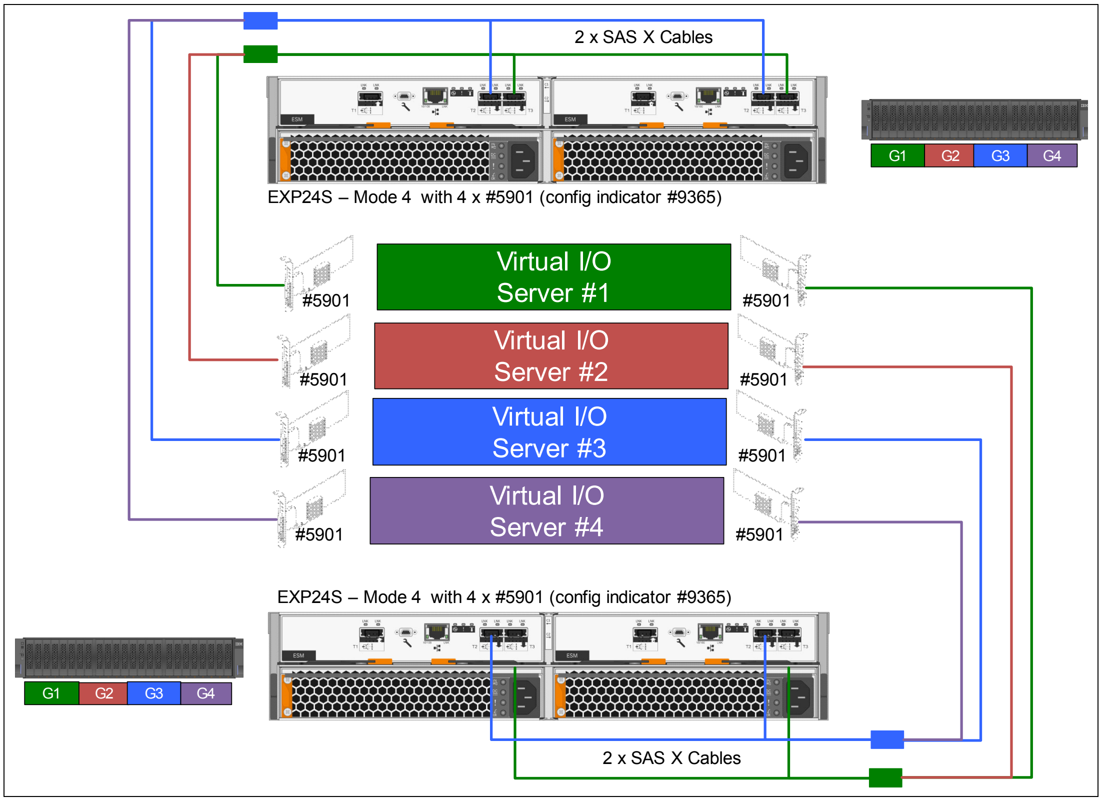

Figure 2-31 Four Virtual I/O Servers sharing two EXP24S

The following are the required feature codes for this scenario:

- Two EXP24S drawers #5887 with indicator feature #9365 (mode 4 with four #5901)
- Eight PCIe SAS adapter #5901
- Four SAS X cables 3 Gbps with proper length

This scenario provides no SSD support.

Other scenarios

For direct connection to logical partitions, different adapters, and cables, see "5887 disk drive enclosure" in the IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/POWER8/p8hdx/POWER8welcome.htm

2.12.3 IBM System Storage

The IBM System Storage Disk Systems products and offerings provide compelling storage solutions with superior value for all levels of business, from entry-level to high-end storage systems. For more information about the various offerings, see the following website:

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

The following section highlights a few of the offerings.

IBM Storwize family

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

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

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

IBM XIV Storage System

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

XIV Storage Systems extend ease of use with integrated management for large and multi-site XIV deployments, reducing operational complexity and enhancing capacity planning. For more information, see the following website:

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

IBM System Storage DS8000

The IBM System Storage DS8000 storage subsystem is a high-performance, high-capacity, and secure storage system. It is designed to deliver the highest levels of performance, flexibility, scalability, resiliency, and total overall value for the most demanding, heterogeneous storage environments. The system is designed to effectively and efficiently manage a broad scope of storage workloads that exist in today’s complex data center.

Additionally, the IBM System Storage DS8000 includes a range of features that automate performance optimization and application quality of service, and also provide the highest levels of reliability and system uptime. For more information, see the following website:

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

2.13 Hardware Management Console

The Power E850C platforms support two main service environments:

- One or more HMCs:

  This is a supported option by the system with PowerVM. This environment is the common configuration for servers supporting logical partitions with dedicated or virtual I/O. In this case, all servers have at least one logical partition.

- No HMC:

  There are two service strategies for non-HMC systems:

  – Full-system partition with PowerVM: A single partition owns all the server resources and only one operating system may be installed.

  – Partitioned system with PowerVM: In this configuration, the system can have up to three partitions and can be running more than one operating system. In this environment, partitions are managed by the Integrated Virtualization Manager (IVM), which provides some of the functions that are provided by the HMC.

    For more information about IVM, see *Integrated Virtualization Manager for IBM Power Systems Servers*, REDP-4061.

An HMC is a dedicated appliance that allows administrators to configure and manage system resources on IBM Power Systems servers that use IBM POWER6, POWER6+ POWER7, POWER7+, and POWER8 processors and the PowerVM Hypervisor. The HMC provides basic virtualization management support for configuring logical partitions (LPARs) and dynamic resource allocation, including processor and memory settings for selected Power Systems servers. The HMC also supports advanced service functions, including guided repair and verification, concurrent firmware updates for managed systems, and around-the-clock error reporting through IBM Electronic Service Agent™ for faster support.

The HMC management features help improve server usage, simplify systems management, and accelerate provisioning of server resources by using the PowerVM virtualization technology.

The HMC could either be a hardware or a virtual appliance, which are described in more detail in the next sections. You can use Power Systems HMCs in any of the following configurations:

- Only hardware appliance HMCs
- Only virtual appliance HMCs
- A combination of hardware appliance and virtual appliance HMCs

Requirements:

- When using any HMC with the Power E850C, the HMC code must be running at V860.10 level or later. The minimum firmware level for the Power E850C is V860.10 or later.
- When using the Power E850C server to access the cloud-based HMC Apps as a Service, the HMC code must be at a minimum level of V860.

2.13.1 Hardware appliance HMC

The 7042-CR9 hardware appliance HMC is a dedicated rack-mounted workstation that can be used to manage any of the systems that are supported by the version 8 HMC. It provides hardware, service, and basic virtualization management for your Power Systems servers.

HMC RAID 1 support

There is a high availability feature available for the hardware appliance HMC. The 7042-CR9, by default, includes two HDDs with RAID 1 configured. RAID 1 is also offered on the 7042-CR6, 7042-CR7, 7042-CR8, and 7042-CR9 models (if the feature was removed from the initial order) as an MES upgrade option.

RAID 1 uses data mirroring. Two physical drives are combined into an array, and the same data is written to both drives. This combination makes the drives mirror images of each other. If one of the drives experiences a failure, it is taken offline and the HMC continues operating with the other drive.

HMC models

To use an existing HMC to manage any POWER8 processor-based server, the HMC must be a model CR5 or later rack-mounted HMC, or model C08 or later deskside HMC. The latest HMC model is the 7042-CR9.

> **Note:** The 7042-CR9 ships with 16 GB of memory, and is expandable to 192 GB with an upgrade feature. 16 GB is advised for large environments or where external utilities, such as PowerVC and other third-party monitors, are to be implemented.

2.13.2 Virtual appliance HMC

The 5765-HMV virtual appliance HMC can be used to manage any of the systems that are supported by the version 8 HMC. It provides hardware, service, and basic virtualization management for your Power Systems servers. The virtual HMC runs as a virtual machine on an x86 server virtualized either by VMware ESXi or Red Hat KVM.

It offers the same functionality as the hardware appliance HMC. However, the virtual appliance HMC has the following differences from the hardware appliance HMC:

- An activation engine that allows providing unique configuration during initial deployment.
- Differences in the way the license acceptance dialog is presented.
- Support for multiple virtual disks for additional data storage.
- Formatting of physical media is not supported. This process is supported through a virtual device attached to the VM.

Hardware requirements

At the time of writing, the following hardware is required:

- x86 64-bit hardware with hardware virtualization assists (Intel VT-x or AMD-V)
- Resources for the HMC virtual appliance VM: Four CPUs, 8 GB of memory, 160 GB of disk space, and two network interfaces

Software requirements

At the time of writing, the following software is required:

- Either VMware ESXi V5 or later, or Red Hat Enterprise Linux 6.x with KVM

2.13.3 HMC code level

When using any HMC with the Power E850C servers, the HMC code must be running at V860.10 level or later. The minimum firmware level for the Power E850C is V860.10 or later.

When using the Power E850C server to access the cloud-based HMC Apps as a Service, the HMC code must be at a minimum level of V860.

If you are attaching an HMC to a new server or adding a function to an existing server that requires a firmware update, the HMC machine code might need to be updated to support the firmware level of the server. In a dual HMC configuration, both HMCs must be at the same version and release of the HMC code.

To determine the HMC machine code level that is required for the firmware level on any server, access the Fix Level Recommendation Tool (FLRT) on or after the planned availability date for this product:

https://www14.software.ibm.com/webapp/set2/flrt/home

FLRT identifies the correct HMC machine code for the selected system firmware level.

> **Note:** Access to firmware and machine code updates is conditional on entitlement and license validation in accordance with IBM policy and practice. IBM can verify entitlement through customer number, serial number electronic restrictions, or any other means or methods that are employed by IBM at its discretion.

2.13.4 HMC connectivity to the POWER8 processor-based systems

POWER8 processor-based servers, and their predecessor systems, that are managed by a hardware or virtual appliance HMC require Ethernet connectivity between the HMC and the server's service processor. In addition, if dynamic LPAR, Live Partition Mobility, or PowerVM Active Memory Sharing operations are required on the managed partitions, Ethernet connectivity is needed between these partitions and the HMC. A minimum of two Ethernet ports are needed on the HMC to provide such connectivity.

For the HMC to communicate properly with the managed server, eth0 of the HMC must be connected to either the HMC1 or HMC2 ports of the managed server, although other network configurations are possible. You can attach a second HMC to the remaining HMC port of the server for redundancy. The two HMC ports must be addressed by two separate subnets.

Figure 2-32 shows a simple network configuration to enable the connection from the HMC to the server and to allow for dynamic LPAR operations. For more information about HMC and the possible network connections, see *IBM Power Systems HMC Implementation and Usage Guide*, SG24-7491.

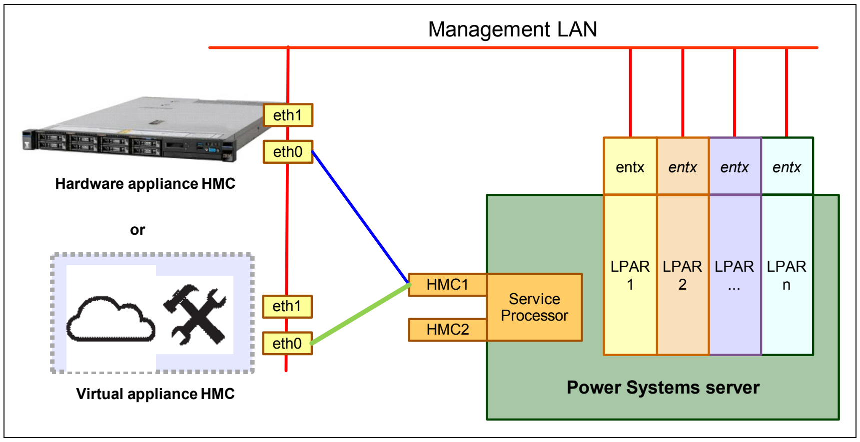

Figure 2-32 Network connections from the HMC to service processor and LPARs

By default, the service processor HMC ports are configured for dynamic IP address allocation. The HMC can be configured as a DHCP server, providing an IP address at the time that the managed server is powered on. In this case, the FSP is allocated an IP address from a set of address ranges that are predefined in the HMC software.

If the service processor of the managed server does not receive a DHCP reply before timeout, predefined IP addresses are set up on both ports. Static IP address allocation is also an option and can be configured by using the ASMI menus.

Notes: The two service processor HMC ports have the following features:

- Run at a speed of 1 Gbps
- Are visible only to the service processor and can be used to attach the server to an HMC or to access the ASMI options from a client directly from a client web browser
- Use the following network configuration if no IP addresses are set:
  - Service processor eth0 (HMC1 port): 169.254.2.147 with netmask 255.255.255.0
  - Service processor eth1 (HMC2 port): 169.254.3.147 with netmask 255.255.255.0

For more information about the service processor, see 4.5.3, “Service processor” on page 117.

2.13.5 High availability HMC configuration

The HMC is an important hardware component. Although Power Systems servers and their hosted partitions can continue to operate when the managing HMC becomes unavailable, certain operations, such as dynamic LPAR, partition migration using PowerVM Live Partition Mobility, or the creation of a new partition, cannot be performed without the HMC. To avoid such situations, consider installing a second HMC in a redundant configuration, to be available when the other is not (during maintenance, for example).

To achieve HMC redundancy for a POWER8 processor-based server, the server must be connected to two HMCs:

- The HMCs must be running the same level of HMC code.
- The HMCs must use different subnets to connect to the service processor.
- The HMCs must be able to communicate with the server's partitions over a public network to allow for full synchronization and functionality.

For more information about redundant HMCs, see *IBM Power Systems HMC Implementation and Usage Guide*, SG24-7491.

2.14 Operating system support

The IBM Power E850C systems support the following operating systems:

- AIX
- Linux

In addition, the Virtual I/O Server can be installed in special partitions that provide support to the other operating systems for using features such as virtualized I/O devices, PowerVM Live Partition Mobility, or PowerVM Active Memory Sharing.

For details about the software available on IBM Power Systems, visit the IBM Power Systems Software™ website at:

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

2.14.1 Virtual I/O Server

The minimum required level of Virtual I/O Server for Power E850C is VIOS 2.2.5.10 or later.

IBM regularly updates the Virtual I/O Server code. To find information about the latest updates, visit the Fix Central website at:

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

2.14.2 IBM AIX operating system

The following sections discuss the various levels of AIX operating system support.

IBM periodically releases maintenance packages (service packs or technology levels) for the AIX operating system. Information about these packages is on the Fix Central website at:

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

The Fix Central website also provides information about how to obtain the fixes that are included on CD-ROM.

The Service Update Management Assistant (SUMA) can help you to automate the task of checking and downloading operating system downloads. It is part of the base operating system. For more information about the **suma** command, go to the following website:

https://www.ibm.com/developerworks/aix/library/au-aixservicepacks/

Table 2-31 provides a list of minimum levels that are required for AIX.

Table 2-31 Supported AIX levels on the E850C

AIX Version	With Virtual I/O Server	Without Virtual I/O Server
6.1	► 6100-09 Technology Level and Service Pack 6 or later	► 6100-09 Technology Level Service Pack 8, and APAR IV88679 or later
7.1	► 7100-03 Technology Level and Service Pack 6 or later ► 7100-04 Technology Level and Service Pack 1 or later	► 7100-03 Technology Level Service Pack 8 or later (planned availability: January 27, 2017) ► 7100-04 Technology Level Service Pack 3, and APAR IV88680 or later
7.2	► 7200-00 Technology Level or later	► 7200-00 Technology Level and Service Pack 3 or later (planned availability January 27, 2017) ► 7200-01 Technology Level or later

2.14.3 Linux operating systems

Linux is an open source operating system that runs on numerous platforms from embedded systems to mainframe computers. It provides an implementation like UNIX across many computer architectures.

The supported versions of Linux on Power E850C are shown in Table 2-32.

Table 2-32 Supported Linux versions on the E850C

Linux vendor	Big endian	Little endian
SUSE	► SUSE Linux Enterprise Server 11 Service Pack 4 (IBM Power Systems Solution Editions for SAP HANA clients only)	► SUSE Linux Enterprise Server 12 Service Pack 1 or later
RedHat	► Red Hat Enterprise Linux 7.2 or later ► Red Hat Enterprise Linux 6.8 or later	► Red Hat Enterprise Linux 7.2 or later
Ubuntu	► n/a	► Ubuntu 16.04.1 or later

If you want to configure Linux partitions in virtualized Power Systems, be aware of the following conditions:

- Not all devices and features that are supported by the AIX operating system are supported in logical partitions that run the Linux operating system.
- Linux operating system licenses are ordered separately from the hardware. You can acquire Linux operating system licenses from IBM to be included with the POWER8 processor-based servers, or from other Linux distributors.

For information about features and external devices that are supported by Linux, see this site:

http://www.ibm.com/systems/p/os/linux/index.html

For information about SUSE Linux Enterprise Server, see this site:

http://www.novell.com/products/server

For information about Red Hat Enterprise Linux Advanced Server, see this site:

http://www.redhat.com/rhel/features

For information about Ubuntu Server, see this site:

http://www.ubuntu.com/server

2.14.4 Java versions that are supported

Java is supported on POWER8 servers. For best use of the performance capabilities and most recent improvements of POWER8 technology, upgrade Java based applications to Java 8, Java 7, or Java 6. For more information, visit the following website:

http://www.ibm.com/developerworks/java/jdk/aix/service.html

2.15 Energy management

The Power E850C systems have features to help clients become more energy efficient. IBM EnergyScale™ technology enables advanced energy management features to conserve power dramatically and dynamically, and further improve energy efficiency. Intelligent Energy optimization capabilities enable the POWER8 processor to operate at a higher frequency for increased performance and performance per watt, or dramatically reduce frequency to save energy.

2.15.1 IBM EnergyScale technology

IBM EnergyScale technology provides functions to help the user understand and dynamically optimize processor performance versus processor energy consumption, and system workload, to control IBM Power Systems power and cooling usage.

EnergyScale uses power and thermal information that is collected from the system to implement policies that can lead to better performance or better energy usage. IBM EnergyScale has the following features:

- Power trending

  EnergyScale provides continuous collection of real-time server energy consumption. It enables administrators to predict power consumption across their infrastructure and to react to business and processing needs. For example, administrators can use such information to predict data center energy consumption at various times of the day, week, or month.

- Power saver mode

  Power saver mode lowers the processor frequency and voltage on a fixed amount, reducing the energy consumption of the system while still delivering predictable performance. This percentage is predetermined to be within a safe operating limit and is not user configurable. The server is designed for a fixed frequency drop of almost 50% down from nominal frequency (the actual value depends on the server type and configuration).

  Power saver mode is not supported during system start, although it is a persistent condition that is sustained after the boot when the system starts running instructions.

- Dynamic power saver mode

  Dynamic power saver mode varies processor frequency and voltage based on the usage of the POWER8 processors. Processor frequency and usage are inversely proportional for most workloads, implying that as the frequency of a processor increases, its usage decreases, given a constant workload. Dynamic power saver mode takes advantage of this relationship to detect opportunities to save power, based on measured real-time system usage.

  When a system is idle, the system firmware lowers the frequency and voltage to power energy saver mode values. When fully used, the maximum frequency varies, depending on whether the user favors power savings or system performance. If an administrator prefers energy savings and a system is fully used, the system is designed to reduce the maximum frequency to about 95% of nominal values.

  Dynamic power saver mode is mutually exclusive with power saver mode. Only one of these modes can be enabled at a time.

- Power capping

  Power capping enforces a user-specified limit on power usage. Power capping is not a power-saving mechanism. It enforces power caps by throttling the processors in the system, degrading performance significantly. The idea of a power cap is to set a limit that must never be reached but that frees extra power that was never used in the data center. The *margined* power is this amount of extra power that is allocated to a server during its installation in a data center. It is based on the server environmental specifications that usually are never reached because server specifications are always based on maximum configurations and worst-case scenarios.

- Soft power capping

  There are two power ranges into which the power cap can be set: Power capping, as described previously, and soft power capping. Soft power capping extends the allowed energy capping range further, beyond a region that can be ensured in all configurations and conditions. If the energy management goal is to meet a particular consumption limit, soft power capping is the mechanism to use.

- Processor core nap mode

  IBM POWER8 processor uses a low-power mode that is called *nap* that stops processor execution when there is no work to do on that processor core. The latency of exiting nap mode is small, typically not generating any impact on applications running. Therefore, the IBM POWER Hypervisor™ can use nap mode as a general-purpose idle state. When the operating system detects that a processor thread is idle, it yields control of a hardware thread to the POWER Hypervisor. The POWER Hypervisor immediately puts the thread into nap mode. Nap mode allows the hardware to turn off the clock on most of the circuits in the processor core. Reducing active energy consumption by turning off the clocks allows the temperature to fall, which further reduces leakage (static) power of the circuits and causes a cumulative effect. Nap mode saves 10 - 15% of power consumption in the processor core.

- Processor core sleep mode

  To save even more energy, the POWER8 processor has an even lower power mode referred to as *sleep*. Before a core and its associated private L2 cache enter sleep mode, the cache is flushed, transition lookaside buffers (TLB) are invalidated, and the hardware clock is turned off in the core and in the cache. Voltage is reduced to minimize leakage current. Processor cores that are inactive in the system (such as CoD processor cores) are kept in sleep mode. Sleep mode saves about 80% power consumption in the processor core and its associated private L2 cache.

- Processor chip winkle mode

  The most amount of energy can be saved when a whole POWER8 chiplet enters the *winkle* mode. In this mode, the entire chiplet is turned off, including the L3 cache. This mode can save more than 95% power consumption.

- Fan control and altitude input

  System firmware dynamically adjusts fan speed based on energy consumption, altitude, ambient temperature, and energy savings modes. Power Systems are designed to operate in worst-case environments, in hot ambient temperatures, at high altitudes, and with high-power components. In a typical case, one or more of these constraints are not valid. When no power savings setting is enabled, fan speed is based on ambient temperature and assumes a high-altitude environment. When a power savings setting is enforced (either Power Energy Saver Mode or Dynamic Power Saver Mode), the fan speed varies based on power consumption and ambient temperature.

- Processor folding

  Processor folding is a consolidation technique that dynamically adjusts, over the short term, the number of processors that are available for dispatch to match the number of processors that are demanded by the workload. As the workload increases, the number of processors made available increases. As the workload decreases, the number of processors that are made available decreases. Processor folding increases energy savings during periods of low to moderate workload because unavailable processors remain in low-power idle states (nap or sleep) longer.

- EnergyScale for I/O

  IBM POWER8 processor-based systems automatically power off hot-pluggable PCI adapter slots that are empty or not being used. System firmware automatically scans all pluggable PCI slots at regular intervals, looking for those that meet the criteria for being not in use and powering them off. This support is available for all POWER8 processor-based servers and the expansion units that they support.

- Server power down

  If overall data center processor usage is low, workloads can be consolidated on fewer numbers of servers so that some servers can be turned off completely. Consolidation makes sense during long periods of low usage, such as weekends. Live Partition Mobility can be used to move workloads to consolidate partitions onto fewer systems, reducing the number of servers that are powered on and therefore reducing the power usage.

  On POWER8 processor-based systems, several EnergyScale technologies are embedded in the hardware and do not require an operating system or external management component. Fan control, environmental monitoring, and system energy management are controlled by the On Chip Controller (OCC) and associated components.

The power mode can also be set up without external tools, by using the ASMI interface, as shown in Figure 2-33.

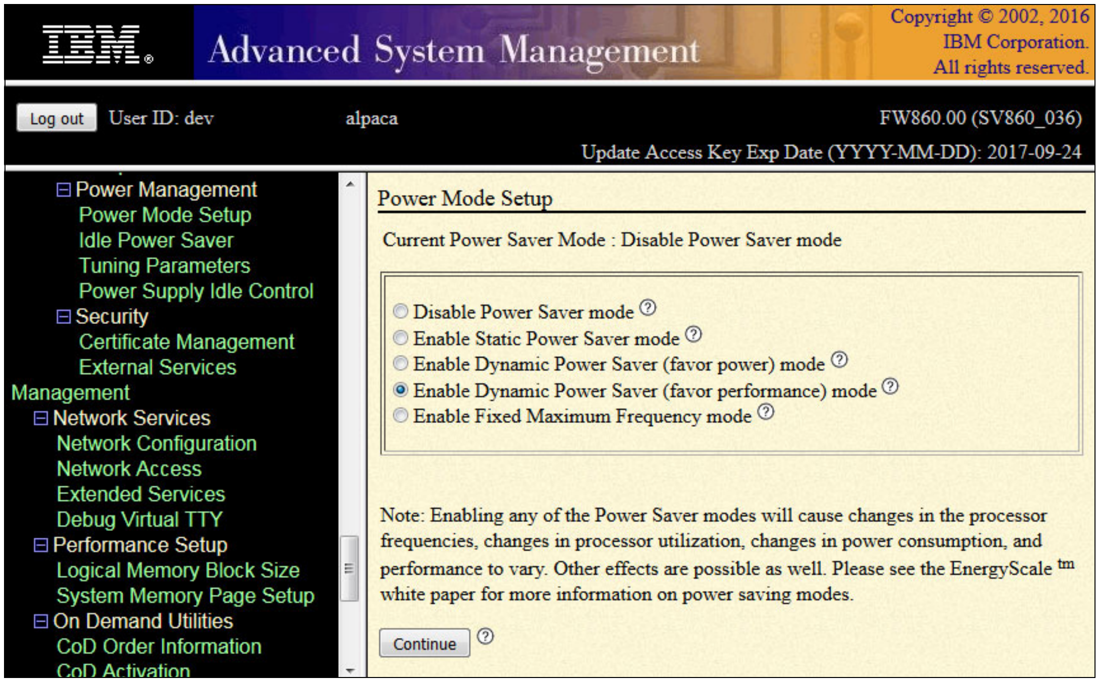

Figure 2-33 Setting the power mode in ASMI

2.15.2 On Chip Controller

To maintain the power dissipation of POWER7+ with its large increase in performance and bandwidth, POWER8 invested significantly in power management innovations. A new OCC that uses an embedded IBM PowerPC® core with 512 KB of SRAM runs real-time control firmware. It responds to workload variations by adjusting the per-core frequency and voltage based on data from activity, thermal, voltage, and current sensors.

The on-die nature of the OCC allows for approximately 100× increase in speed in response to workload changes over POWER7+. The capability enables reaction under the time scale of a typical OS time slice and allows for multi-socket, scalable systems to be supported. It also enables more granularity in controlling the energy parameters in the processor, and increases reliability in energy management by having one controller in each processor that can perform certain functions independently of the others.

POWER8 also includes an internal voltage regulation capability that enables each core to run at a different voltage. Optimizing both voltage and frequency for workload variation enables better increase in power savings versus optimizing frequency only.

2.15.3 Energy consumption estimation

Often, for Power Systems, 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 the IBM Knowledge Center, found at the following website:

  http://www.ibm.com/support/knowledgecenter/POWER8/p8hdx/POWER8welcome.htm

  Search for type and model number and "server specifications". For example, for the Power E850C system, search for "8408-44E server specifications".

- An estimation of the energy consumption for a certain configuration

  The calculation of the energy consumption for a certain configuration can be done in the IBM Systems Energy Estimator, found at the following website:

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

In that tool, select the type and model for the system, and enter some details about the configuration and estimated CPU usage. The tool will then provide the estimated energy consumption and the waste heat at the estimated usage and also at full usage.

3

Private and Hybrid Cloud Features

The Power E850C server includes cloud management software and services to assist with the move to the cloud, both private and hybrid. This combination allows for fast and automated VM deployments, based on prebuilt image templates and self-service capabilities, all with an intuitive interface.

This chapter provides an overview of these software and services.

This chapter includes the following sections:

- Private cloud software
- Hybrid cloud support
- Geographically Dispersed Resiliency for Power
- IBM Power to Cloud Rewards Program

3.1 Private cloud software

Companies are creating private clouds that join characteristics from both internal private environments and public cloud to create a more flexible, agile, secure, and customized infrastructure environment.

Having a single point to control this infrastructure is key to achieving these goals, which is where the private cloud software excels.

The following sections describe the software that is used to allow for the Power E850C to be part of a private cloud, and the tools available to ease the deployment and management of virtual machines in a private cloud.

3.1.1 IBM Cloud PowerVC Manager

Managing a private cloud requires software tools to help create a virtualized pool of compute resources, provide a self-service portal for users, and policies for resource allocation, control, security, and metering data for resource billing. Management tools for private clouds tend to be service driven, as opposed to resource driven, because cloud environments are typically highly virtualized and organized in terms of portable workloads.

The OpenStack-based IBM Cloud PowerVC Manager provides a self-service cloud portal for IBM Power Systems. This self-service portal enables users to quickly request cloud resources and reliably deploy virtual machines with approval policies to maintain control over provisioning of cloud resources.

Among its several capabilities, IBM Cloud PowerVC Manager provides the following features:

- Create VMs and resize the VMs CPU and memory
- Attach disk volumes to those VMs.
- Import existing VMs and volumes so that they can be managed by PowerVC
- Monitor the use of resources in your environment
- Migrate VMs while they are running (live migration between physical servers)
- Improve resource usage to reduce capital expense and power consumption
- Increase agility and execution to quickly respond to changing business requirements
- Increase IT productivity and responsiveness
- Simplify Power Systems virtualization management
- Accelerate repeatable, error-free virtualization deployments

IBM Cloud PowerVC Manager can manage AIX, IBM i, and Linux VMs running under PowerVM virtualization and Linux VMs running under PowerKVM virtualization.

Because it is OpenStack based, it provides upward integration with hybrid cloud orchestration products, allowing Power Systems to take part in a heterogeneous private cloud infrastructure. For more information, see 3.2.1, “Hybrid infrastructure management tools” on page 96.

IBM Cloud PowerVC Manager extends OpenStack with the additional components, so IBM Power Systems can be easily managed just like any other platform. Figure 3-1 shows the OpenStack components and IBM additions to it.

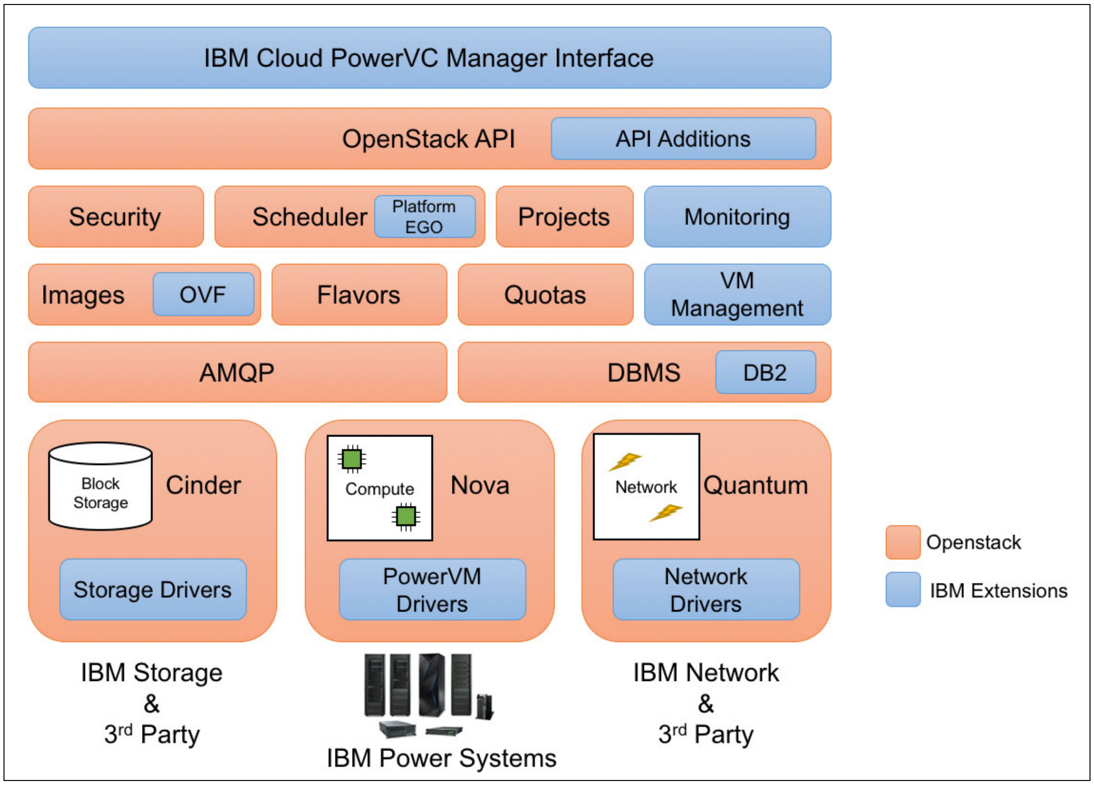

Figure 3-1 OpenStack components and IBM Cloud PowerVC Manager additions

Detailed information about IBM Cloud PowerVC Manager can be obtained in *IBM PowerVC Version 1.3.1 Introduction and Configuration Including IBM Cloud PowerVC Manager*, SG24-8199.

3.1.2 Cloud-based HMC Apps as a Service

The new HMC Apps as a Service provides powerful insights into your Power Systems infrastructure by using a set of hosted as-a-service apps, with no additional software or infrastructure setup. Through these new apps, clients gain the capability to aggregate Power Systems performance, auditing, and inventory data from across their enterprise, removing the burden of manual collection and collation. These IBM-developed apps are hosted in a secure, multi-tenant cloud. They provide health scores, search and filtering, and threshold-based alerts that can be accessed through a secure portal from desktops or mobile devices.

The applications gather data sent by the customer HMCs through the Internet by using the built-in cloud connector included with HMC V8R8.6.0 and NovaLink v1.0.0.4.

Figure 3-2 shows a logical diagram of a company with three data centers that consolidate information that is provided by HMC and Novalink into HMC Apps as a Service.

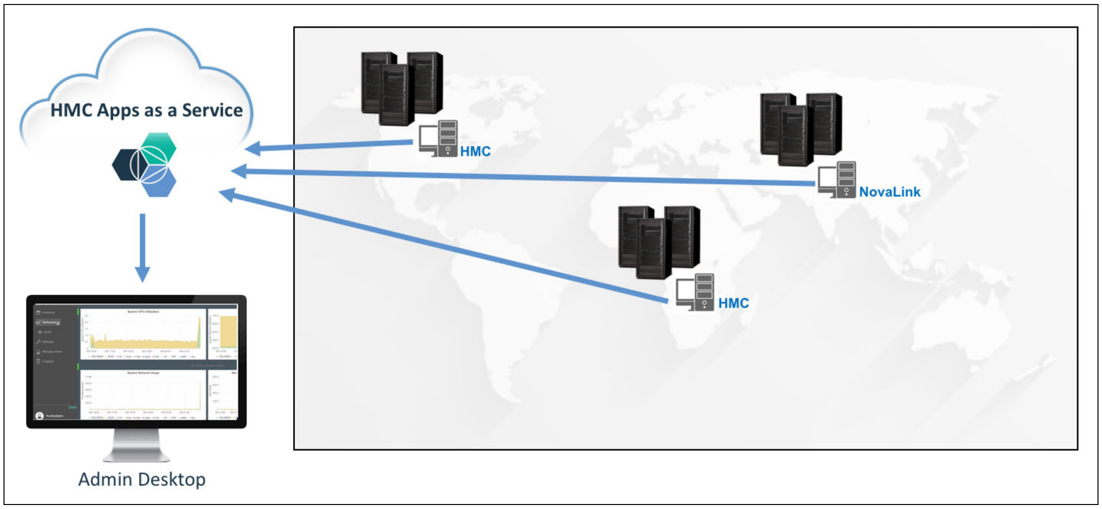

Figure 3-2 Logical diagram of HMC Apps as a Service consolidating information from several sites

After the data is gathered, it can be organized and analyzed to provide the administrators with a single point for these tasks:

- View all Power Systems, VIOS, HMCs, NovaLinks, and LPARs across entire enterprise
- Organize and show resources by tags
- See basic health and state
- Check hardware inventory
- Aggregated performance views across your Power enterprise
- Manage thresholds and alerts

When clients purchase a new Power E850C server, they are entitled to this new service offering for no additional charge.

> **Note:** At the time of writing, the performance and inventory applications are initially scheduled to be offered in a technology preview in 2016 and to be followed by a full GA offering with more applications in 2017.

3.1.3 Open source cloud automation and configuration tooling for AIX

IBM has expanded its commitment to keep key open source cloud management packages updated and to provide timely security fixes to enable clients to take advantage of open source skills. IBM Power Systems servers are well positioned to take advantage of key packages that were recently provided to enable cloud automation. This section lists several key packages.

Chef

Chef is an automation platform that provides automation for configuration, deployment, and management of virtual machines. It is based on a client/server architecture where the server stores the policies that are applied to nodes, called cookbooks, and their associated metadata. After a cookbook is created, it can be applied to a client virtual machine being deployed, configuring this VM according to the established policies automatically.

IBM is collaborating along with clients in this community to provide useful resources for using Chef with AIX systems. Chef-client for AIX is now enhanced with new recipes. The AIX cookbook can be reached at this URL:

https://supermarket.chef.io/cookbooks/aix/

Yum

Yum stands for “Yellowdog Updater, Modified” and allows for automatic updates, package, and dependency management on RPM-based distributions.

It is now available with repository access from both FTP and https protocols for AIX, and is also updated to enable automatic dependency discovery. For more information, see the AIX Toolbox for Linux Applications website at:

http://www.ibm.com/systems/power/software/aix/linux/toolbox/alpha.html

Cloud-init

Cloud-init is a tool created to help with the early initialization of a virtual machine being deployed. It uses OpenStack metadata information to, among other tasks, set a root password, grow file systems, set a default locale, set host names, generate SSH private keys, and handle temporary mount points. After it is script based, it can be easily extended to do other tasks specific to a customer environment.

Cloud-init and all its dependencies are now available, and includes support for licensed AIX users. For more information, see the AIX Toolbox for Linux Applications website at:

http://www.ibm.com/systems/power/software/aix/linux/toolbox/alpha.html

GitHub

GitHub is a version control system that manages and stores revisions of projects. GitHub provides a web-based interface, access control, and several collaboration features such as task management and wikis.

In the past, if you wanted to contribute to an open source project, you would need to download the source code, make the changes, take note of the changes, talk to the owner of the code to explain the changes, and get the owner’s approval to apply your changes to the official code. With GitHub, a user can fork an existing project, download and change only the needed files, pull the new code to the platform, and notify the owner who can then publish the changes automatically with version control.

Open source projects for AIX can be found at the following repository:

http://github.org/aixoss

Node.js

Node.js is a platform that is built on JavaScript for building fast and scalable applications. Node.js uses an event-driven, non-blocking I/O model that makes it lightweight and efficient. It is perfect for data-intensive and real-time applications that run across distributed devices.

Node.js is available for both Linux and AIX platforms and can be downloaded from:

https://nodejs.org/en/download/

3.2 Hybrid cloud support

Hybrid cloud is quickly becoming the de facto standard of IT. Two-thirds of organizations that blend traditional and cloud infrastructure together are already gaining advantage from their hybrid cloud. A hybrid cloud model enables the building and deploying of applications quickly with optimized utilization of resources and the lowering of costs. In addition, being able to centrally manage private, public, or dedicated cloud resources with a single management tool while securely connecting traditional workloads with cloud-native apps enables clients to respond to their dynamically changing business priorities in a more agile and timely fashion.

3.2.1 Hybrid infrastructure management tools

Power Systems OpenStack-based PowerVC management upwardly integrates into various third-party hybrid cloud orchestration products, including IBM Cloud Orchestrator, VMware vRealize, and others. Clients can simply manage both their private cloud VMs and their public cloud VMs from a single, integrated management tool.

VMware vRealize uses PowerVC OpenStack capabilities to interact with PowerVM and PowerKVM, allowing for management and integration of Power-based virtual machines into a single orchestration platform. A diagram of the OpenStack-based PowerVC upward integration with VMware vRealize can be seen in Figure 3-3.

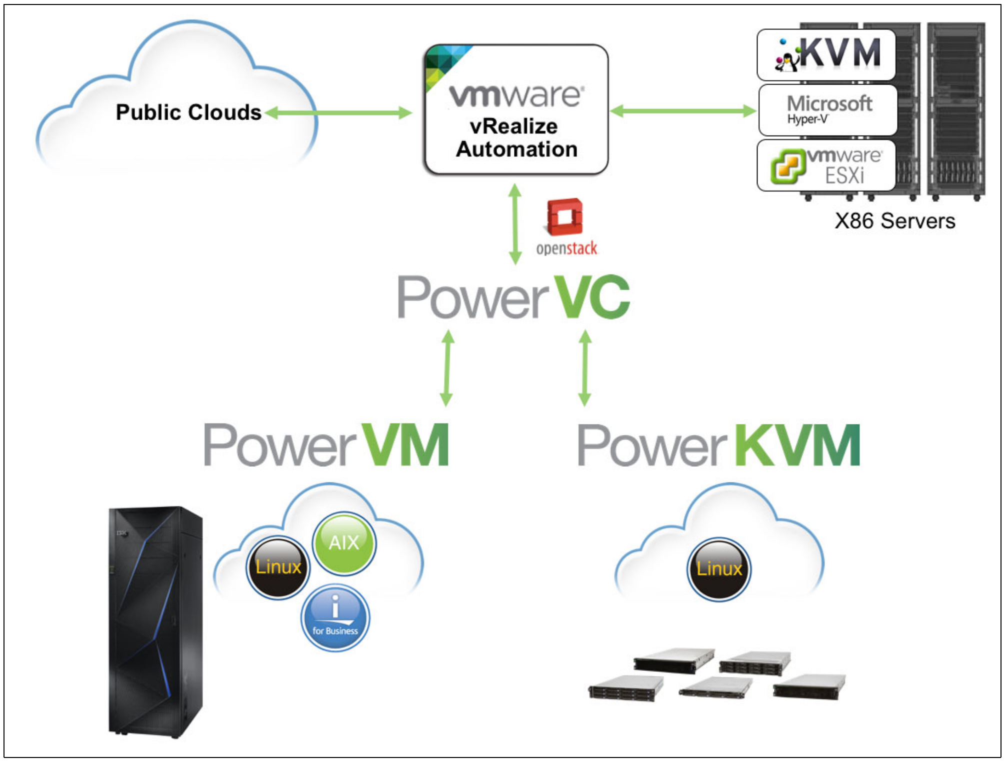

Figure 3-3 Diagram of the OpenStack-based PowerVC integration with VMware vRealize

3.2.2 Securely connect system of record workloads to cloud native applications

IBM API Connect™ and IBM WebSphere® Connect provide secure connectivity to cloud-based applications, giving clients the ability to rapidly develop new applications and services, accelerating their time to value. IBM Power to Cloud services can help clients get started with these solutions and to design new applications that use IBM Bluemix®. This application enables clients to rapidly build, deploy, and manage their cloud applications, while being enhanced by a growing ecosystem of available services and runtime frameworks.

3.2.3 IBM Cloud Starter Pack

To help clients get started with their hybrid cloud infrastructure, the Power E850C offering includes entitlement of a POWER8 Linux bare metal system in the IBM Cloud (SoftLayer®).

Customers are entitled to six months of a C812L-M POWER8 Linux bare metal with the following configuration:

- 10-core POWER8 processor 3.49 GHz
- 256 GB RAM
- Two 4 TB SATA HDD
- Ubuntu Linux

3.2.4 Flexible capacity on demand

With the purchase of a new Power E850C server, clients can now convert previously purchased capacity (Elastic COD Processor Days) to SoftLayer Linux on Power bare metal monthly server usage.

This application allows you to preserve investments after which a customer will be able to convert excess unused capacity into PowerLinux bare metal servers that run on public cloud.

This feature has the following possible uses:

- Offload workloads to the public cloud to make room for critical private cloud workloads
- Provide more processing capacity for applications during peaks
- Use the public cloud as a DR site for the Linux on Power virtual machines
- Move development virtual machines to public cloud

3.3 Geographically Dispersed Resiliency for Power

Geographically Dispersed Resiliency for Power (GDR for Power) is a tool that allows for the remote restart of failed virtual machines in a disaster recovery scenario.

Most of the disaster recovery scenarios for virtual environments are based on two different approaches, each one with its own benefits and points of attention: Clustered Virtual Machines and Virtual Machine Remote Restart. Figure 3-4 shows a diagram of both disaster recovery models.

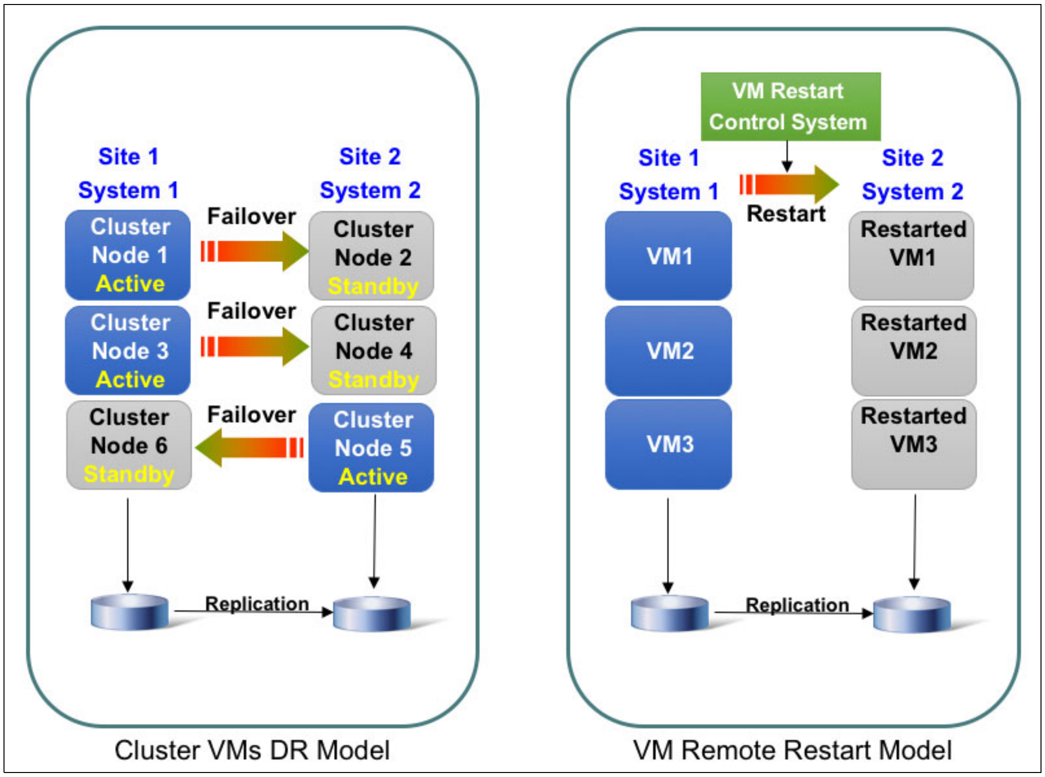

Figure 3-4 Disaster recovery models comparison for virtualized environments

Although critical applications can benefit from the clustered virtual machines model because of the faster RTO, its implementation can be more complex for many virtual machines. However, non-critical environments might benefit from the Virtual Machine Remote Restart model due to the simpler implementation and maintenance of the environment. Table 3-1 compares some aspects of both models.

Table 3-1 Disaster recovery models comparison for virtualized environments

	Clustered Virtual Machines	Virtual Machines Remote Restart
Complexity	Individual VM configuration	Host configuration
Failover Time	Faster (seconds to minutes)	Slower (minutes)
Cost	Higher due to extra software needs and labor	Lower
Maintenance	Each new VM requires a new cluster	After the host is configured, new VMs are automatically covered

3.4 IBM Power to Cloud Rewards Program

IBM Power to Cloud Rewards Program helps clients design, build, and deliver clouds platforms on IBM Power Systems servers. The IBM Power to Cloud Rewards Program transforms the successful IBM PowerCare program to a new points-based reward system and helps accelerate the transformation of IT infrastructure to private and hybrid cloud.

The IBM Power to Cloud Rewards Program enables clients to earn reward points on purchases of an IBM Power Systems server including the Power E850C.

IBM Power to Cloud Reward points can be used for a range of services that are focused on helping the transition from traditional IT platforms to private and hybrid cloud platforms. Power to Cloud Reward Program services offerings use the proven expertise of IBM Systems Lab Services consultants.

The following list is a sample of the services that are offered under the IBM Power to Cloud Rewards Program:

- Design for Cloud Provisioning and Automation
- Build Infrastructure as a Service for Private Cloud
- Build Cloud Capacity Pools across Data Centers
- Deliver with Automation for DevOps
- Design for Hybrid Cloud Workshop
- Deliver with Database as a Service
- Build and Provision for Mobility and Automation
- Design for Private Cloud Monitoring and Capacity Planning

For those clients looking for a hybrid cloud solution, workshops services are available to provide instruction on how to produce best in class applications that use API Connect and Bluemix with Power Systems.

To learn more about the full offering list, program details including points redemption value for specific services offers, and scope, contact your IBM Systems Lab Services representative.

4

Reliability, availability, and serviceability

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

RAS has the following :

Reliability Indicates how infrequently a defect or fault in a server occurs.

Availability Indicates how infrequently the functioning of a system or application is affected 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.

This chapter includes the following sections:

- Introduction
- Reliability
- Availability
- Availability impacts of a solution architecture
- Serviceability
- Manageability
- Selected POWER8 RAS capabilities by operating system

4.1 Introduction

The POWER8 processor-based servers are available in two different classes:

- Scale-out systems: For environments that consist of multiple systems that work in concert. In such environments, application availability is enhanced by the superior availability characteristics of each system.
- Enterprise systems: For environments that require systems with increased availability. In such environments, mission-critical applications can take full advantage of the scale-up characteristics, increased performance, flexibility to upgrade, and enterprise availability characteristics.

One key differentiator of the IBM POWER8 processor-based servers running PowerVM is that they use all the advanced RAS characteristics of the POWER8 processor design through the whole portfolio, offering reliability and availability features that are often not seen in other servers. Some of these features are improvements for POWER8 or features that were found previously only in higher-end Power Systems, which are now available across the entire range.

The POWER8 processor modules support an enterprise level of reliability and availability. The processor design has extensive error detection and fault isolation (ED/FI) capabilities to allow for a precise analysis of faults, whether they are hard faults or soft faults. They use advanced technology, including stacked latches and Silicon-On-Insulator (SOI) technology, to reduce susceptibility to soft errors. They also have advanced design features within the processor for correction or try again after soft error events. In addition, the design incorporates spare capacity that is integrated into many elements to tolerate certain faults without requiring an outage or parts replacement. Advanced availability techniques are used to mitigate the impact of other faults that are not directly correctable in the hardware.

Features within the processor and throughout systems are incorporated to support design verification. During the design and development process, subsystems go through rigorous verification and integration testing processes by using these features. During system manufacturing, systems go through a thorough testing process to help ensure high product quality levels, again taking advantage of the designed ED/FI capabilities.

Fault isolation and recovery of the POWER8 processor and memory subsystems are designed to use a dedicated service processor. They are meant to be largely independent of any operating system or application deployed.

The Power E850C server has processor and memory upgrade capabilities characteristic of enterprise systems, and is designed to support higher levels of RAS than scale-out systems.

4.1.1 RAS enhancements of POWER8 processor-based servers

Several features were included in the whole portfolio of the POWER8 processor-based servers. Some of these features are improvements for POWER8 or features that were found previously only in higher-end Power Systems, leveraging a higher RAS even for scale-out equipment.

The following is a brief summary of these features:

- Processor enhancements

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

The processor design now 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 compared to POWER7+, including support of a maximum of 12 cores compared to a maximum of eight cores in POWER7+. Doing more work with less hardware in a system provides greater reliability by concentrating the processing power and reducing the need for extra communication fabrics and components.

The memory controller within the processor is redesigned. From a RAS standpoint, the ability to use a replay buffer to recover from soft errors is added.

- I/O subsystem

  The POWER8 processor now integrates PCIe controllers. PCIe slots that are directly driven by PCIe controllers can be used to support I/O adapters directly in the systems, or can be used to attach external I/O drawers.

  For greater I/O capacity, the POWER8 processor-based Power E850C server also supports a PCIe switch to provide more integrated I/O capacity.

- Memory subsystem

  Custom DIMMs (CDIMMS) are used. These provide the ability to correct a single dynamic random access memory (DRAM) fault within an error-correction code (ECC) word (and then an additional bit fault) to avoid unplanned outages. They also contain a spare DRAM module per port (per nine DRAMs for x8 DIMMs), which can be used to avoid replacing memory.

  After all self-healing and other RAS-related features are implemented, the hypervisor might still detect that a DIMM has a substantial fault that, when combined with a future fault, could cause an outage. In such a case, the hypervisor attempts to migrate data from the failing memory to other available memory in the system, if any is available. This feature is intended to further reduce the chances of an unplanned outage, and can take advantage of any deallocated memory, including memory reserved for Capacity on Demand capabilities.

- Power distribution and temperature monitoring

  The processor module integrates a new On Chip Controller (OCC). This OCC is used to handle Power Management and Thermal Monitoring without the need for a separate controller, as was required in POWER7+. In addition, the OCC can also be programmed to run other RAS-related functions independent of any host processor.

4.1.2 RAS enhancements for enterprise servers

The following are RAS enhancements for enterprise servers:

- Memory Subsystem

  The Power E850C server has the option of mirroring the memory used by the hypervisor. This process reduces the risk of system outage linked to memory faults, as the hypervisor memory is stored in two distinct memory CDIMMs. The Active Memory Mirroring feature is only available on enterprise systems.

- Power Distribution and Temperature Monitoring

  All systems make use of voltage converters that transform the voltage level provided by the power supply to the voltage level needed for the various components within the system. The Power E850C server has redundant or spare voltage converters for each voltage level provided to any processor or memory CDIMM.

Converters that are used for processor voltage levels are configured for redundancy so that when one is detected as failing, it is called out for repair while the system continues to run with the redundant voltage converter.

The converters that are used for memory are configured with a form of sparing. In this sparing, when a converter fails, the system continues operation with another converter without generating a service event or the any sort of outage for repair.

As with the Power E870 and Power E880 servers, the Power E850C uses triple redundant ambient temperature sensors.

4.2 Reliability

The reliability of systems starts with components, devices, and subsystems that are highly reliable. On IBM POWER processor-based systems, this basic principle is expanded upon with a clear design for reliability architecture and methodology. A concentrated, systematic, and architecture-based approach is designed to improve overall system reliability with each successive generation of system offerings. Reliability can be improved in primarily three ways:

- Reducing the number of components
- Using higher reliability grade parts
- Reducing the stress on the components

In the POWER8 systems, elements of all three are used to improve system reliability.

During the design and development process, subsystems go through rigorous verification and integration testing processes. During system manufacturing, systems go through a thorough testing process to help ensure the highest level of product quality.

4.2.1 Designed for reliability

Systems that are designed with fewer components and interconnects have fewer opportunities to fail. Simple design choices, such as integrating processor cores on a single POWER chip, can reduce the opportunity for system failures. The POWER8 chip has more cores per processor module, and the I/O Hub Controller function is integrated in the processor module, which generates a PCIe BUS directly from the processor module. Parts selection also plays a critical role in overall system reliability.

IBM uses stringent design criteria to select server grade components that are extensively tested and qualified to meet and exceed a minimum design life of seven years. By selecting higher reliability grade components, the frequency of all failures is lowered, and wear-out is not expected within the operating system life. Component failure rates can be further improved by burning in select components or running the system before shipping it to the client. This period of high stress removes the weaker components with higher failure rates, that is, it cuts off the front end of the traditional failure rate bathtub curve (see Figure 4-1).

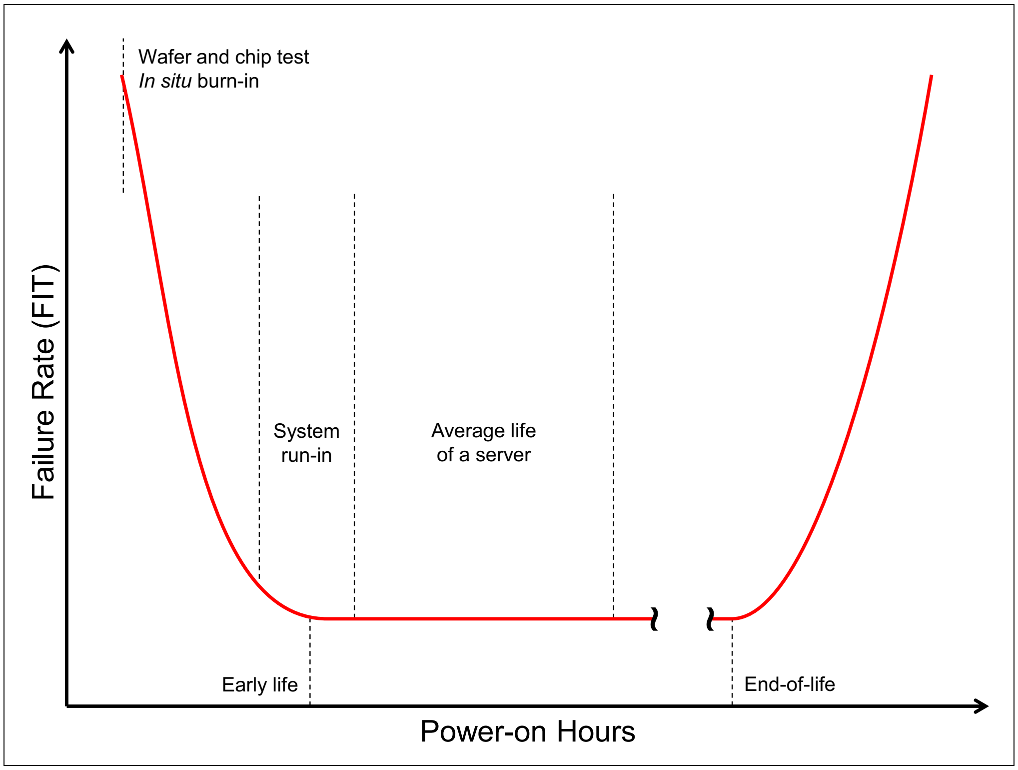

Figure 4-1 Failure rate bathtub curve

4.2.2 Placement of components

Packaging is designed to deliver both high performance and high reliability. For example, the reliability of electronic components is directly related to their thermal environment. Large decreases in component reliability are directly correlated to relatively small increases in temperature. All POWER processor-based systems are packaged to ensure adequate cooling. Critical system components, such as the POWER8 processor chips, are positioned on the system board so that they receive clear air flow during operation. POWER8 systems use a premium fan with an extended life to further reduce overall system failure rate and provide adequate cooling for the critical system components.

The Power E850C has two cooling channels. The front fans provide cooling for the upper part of the chassis, covering the memory cards, processors, and PCIe cards. These fans in this assembly provide redundancy, and support concurrent maintenance. The lower system fans, which are in the internal fan assembly, provide air movement for the lower part of the chassis, including the disk backplane and RAID controllers. The fans in the internal fan assembly provide redundancy, and also contain multiple integrated spares.

4.3 Availability

The more reliable a system or subsystem is, the more available it should be. Nevertheless, a lot of effort is made to design systems that can detect faults that do occur and take steps to minimize or eliminate the outages that are associated with them. These design capabilities extend availability beyond what can be obtained through the underlying reliability of the hardware.

This design for availability begins with implementing an architecture for ED/FI.

First-Failure data capture (FFDC) is the capability of IBM hardware and microcode to continuously monitor hardware functions. Within the processor and memory subsystem, detailed monitoring is done by circuits within the hardware components themselves. Fault information is gathered into fault isolation registers (FIRs) and reported to the appropriate components for handling.

Processor and memory errors that are recoverable in nature are typically reported to the dedicated service processor built into each system. The dedicated service processor then works with the hardware to determine the course of action to be taken for each fault.

4.3.1 Correctable error introduction

Intermittent or soft errors are typically tolerated within the hardware design by using error correction code or advanced techniques to try operations again after a fault.

Tolerating a correctable solid fault runs the risk that the fault aligns with a soft error and causes an uncorrectable error situation. A correctable error might also be predictive of a fault that continues to worsen over time, resulting in an uncorrectable error condition.

You can predictively deallocate a component to prevent correctable errors from aligning with soft errors or other hardware faults and causing uncorrectable errors. However, unconfiguring components, such as processor cores or entire caches in memory, can reduce the performance or capacity of a system. This process in turn typically requires that the failing hardware is replaced in the system. The resulting service action can also temporarily affect system availability.

To avoid such situations in solid faults in POWER8, processors or memory might be candidates for correction by using the “self-healing” features built into the hardware. It , for , take advantage of a spare DRAM module within a memory DIMM, a spare data lane on a processor or memory bus, or spare capacity within a cache module.

When such self-healing is successful, the need to replace any hardware for a solid correctable fault is avoided. The ability to predictively unconfigure a processor core is still available for faults that cannot be repaired by self-healing techniques or because the sparing or self-healing capacity is exhausted.

4.3.2 Uncorrectable error introduction

An uncorrectable error can be defined as a fault that can cause incorrect instruction execution within logic functions, or an uncorrectable error in data that is stored in caches, registers, or other data structures. In less sophisticated designs, a detected uncorrectable error nearly always results in the termination of an entire system. More advanced system designs in some cases might be able to terminate just the application that uses the hardware that failed. Such designs might require that uncorrectable errors are detected by the hardware and reported to software layers. The software layers must then be responsible for determining how to minimize the impact of faults.

The advanced RAS features that are built in to POWER8 processor-based systems handle certain “uncorrectable” errors in ways that minimize the impact of the faults. These features can even keep an entire system up and running after experiencing such a failure.

Depending on the fault, such recovery might use the virtualization capabilities of PowerVM in such a way that the operating system or any applications that are running in the system are not affected or required to participate in the recovery.

4.3.3 Processor Core/Cache correctable error handling

Layer 2 (L2) and Layer 3 (L3) caches and directories can correct single bit errors and detect double bit errors (SEC/DED ECC). Soft errors that are detected in the level 1 caches are also correctable by a try again operation that is handled by the hardware. Internal and external processor “fabric” busses have SEC/DED ECC protection as well.

SEC/DED capabilities are also included in other data arrays that are not directly visible to customers.

Beyond soft error correction, the intent of the POWER8 design is to manage a solid correctable error in an L2 or L3 cache by using techniques to delete a cache line with a persistent issue, or to repair a column of an L3 cache dynamically by using spare capability.

Information about column and row repair operations is stored persistently for processors. This process allows more permanent repairs to be made during processor reinitialization (during system reboot, or individual Core Power on Reset using the Power On Reset Engine).

4.3.4 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 is detected before a failing instruction is completed, the processor hardware might be able to 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 an outage 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 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 process complements the try again abilities of the memory buffer module.

4.3.5 Alternative processor recovery and Partition Availability Priority

If Processor Instruction Retry for a fault within a core occurs multiple times without success, the fault is considered to be a solid failure. In some instances, PowerVM can work with the processor hardware to migrate a workload running on the failing processor to a spare or alternative processor. This migration is accomplished by migrating the pertinent processor core state from one core to another with the new core taking over at the instruction that failed on the faulty core. Successful migration keeps the application running during the migration without needing to terminate the failing application.

Successful migration requires that enough spare capacity is available to reduce the overall processing capacity within the system by one processor core. Typically, in highly virtualized environments, the requirements of partitions can be reduced to accomplish this task without any further impact to running applications.

In systems without sufficient reserve capacity, it might be necessary to terminate at least one partition to free resources for the migration. In advance, PowerVM users can identify which partitions have the highest priority and which do not. When you use this Partition Priority feature of PowerVM, if a partition must be terminated for alternative processor recovery to complete, the system can terminate lower priority partitions to keep the higher priority partitions up and running. This prioritization can even occur when an unrecoverable error occurred on a core running the highest priority workload.

Partition Availability Priority is assigned to partitions by using a weight value or integer rating. The lowest priority partition is rated at zero and the highest priority partition is rated at 255. The default value is set to 127 for standard partitions and 192 for Virtual I/O Server (VIOS) partitions. Priorities can be modified through the Hardware Management Console (HMC).

4.3.6 Core Contained Checkstops and other PowerVM error recovery

PowerVM can handle certain other hardware faults without terminating applications, such as an error in certain data structures (faults in translation tables or lookaside buffers).

Other core hardware faults that alternative processor recovery or Processor Instruction Retry cannot contain might be handled in PowerVM by a technique called *Core Contained Checkstops*. This technique allows PowerVM to be signaled when such faults occur and terminate code in use by the failing processor core (typically just a partition, although potentially PowerVM itself if the failing instruction were in a critical area of PowerVM code).

Processor designs without Processor Instruction Retry typically must resort to such techniques for all faults that can be contained to an instruction in a processor core.

4.3.7 Cache uncorrectable error handling

If a fault within a cache occurs that cannot be corrected with SEC/DED ECC, the faulty cache element is unconfigured from the system. Typically, this process is done by purging and deleting a single cache line. Such purge and delete operations are contained within the hardware itself, and prevent a faulty cache line from being reused and causing multiple errors.

During the cache purge operation, the data that is stored in the cache line is corrected where possible. If correction is not possible, the associated cache line is marked with a special ECC code that indicates that the cache line itself has bad data.

Nothing within the system terminates just because such an event is encountered. Instead, the hardware monitors the usage of pages with marks. If such data is never used, hardware replacement is requested, but nothing terminates as a result of the operation. Software layers are not required to handle such faults.

Only when data is loaded to be processed by a processor core, or sent out to an I/O adapter, is any further action needed. In such cases, if the data is from a logical partition's host OS, the partition operating system might be responsible for terminating itself or just the program using the marked page. If data is owned by the hypervisor, the hypervisor might choose to terminate, resulting in a system-wide outage.

However, the exposure to such events is minimized because cache-lines can be deleted, which eliminates repetition of an uncorrectable fault that is in a particular cache-line.

4.3.8 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 busses”, are corrected by the error correction code that is used. POWER8 processor-to-processor module fabric busses also use a spare data-lane so that a single failure can be repaired without calling for the replacement of hardware.

4.3.9 Other fault error handling

Not all processor module faults can be corrected by these techniques. Therefore, a provision is still made for some faults that cause a system-wide outage. In such a “platform” checkstop event, the ED/FI capabilities that are built into the hardware and dedicated service processor work to isolate the root cause of the checkstop and unconfigure the faulty element where possible. This process allows the system to reboot with the failed component unconfigured from the system.

The auto-restart (reboot) option, when enabled, can reboot the system automatically following an unrecoverable firmware error, firmware hang, hardware failure, or environmentally induced (ac power) failure.

The auto-restart (reboot) option must be enabled from the Advanced System Management Interface (ASMI) or from the Control (Operator) Panel.

4.3.10 Memory protection

POWER8 processor-based systems have a three-part memory subsystem design. This design consists of two memory controllers in each processor module, which communicate to buffer modules on memory DIMMS through memory channels and access the DRAM memory modules on DIMMs, as shown in Figure 4-2.

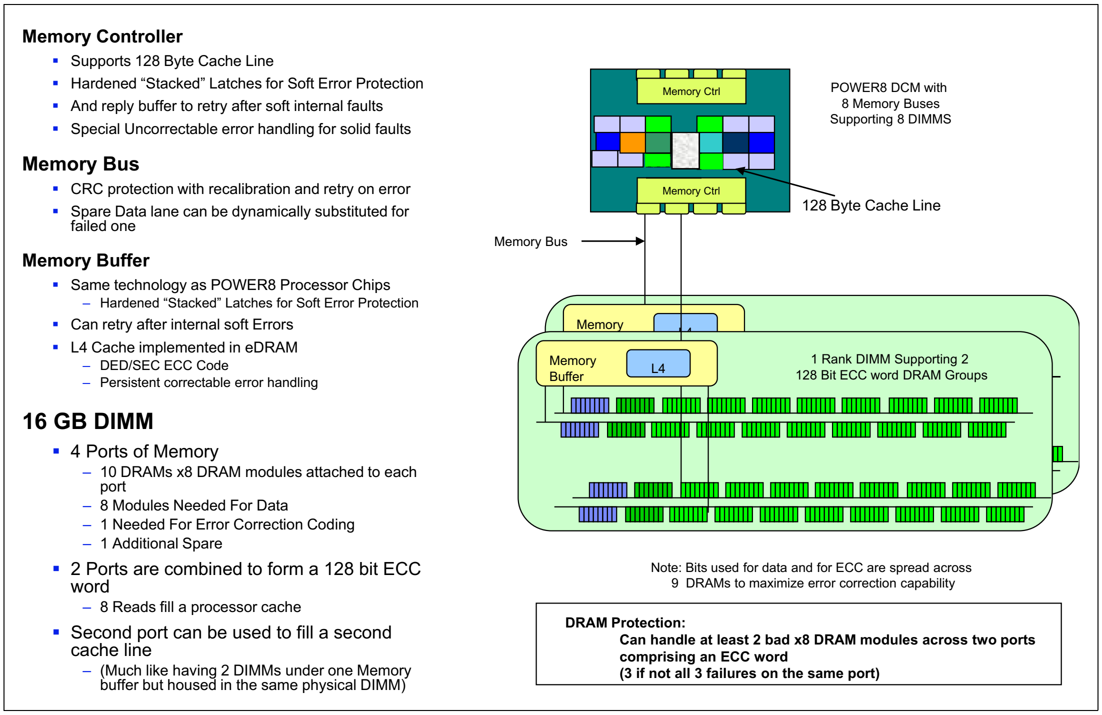

Figure 4-2 Memory protection features

The memory buffer chip is made by the same 22 nm technology that is used to make the POWER8 processor chip. The memory buffer chip incorporates the same features in the technology to avoid soft errors. It implements a try again process for many internally detected faults. This function complements a replay buffer in the memory controller within the processor, which also handles internally detected soft errors.

The bus between a processor memory controller and a DIMM uses CRC error detection that is coupled with the ability to try soft errors again. The bus features dynamic recalibration capabilities plus a spare data lane that can be substituted for a failing bus lane through the recalibration process.

The buffer module implements an integrated L4 cache using eDRAM technology (with soft error hardening) and persistent error handling features.

The memory buffer on each DIMM has four ports for communicating with DRAM modules. The 16 GB DIMM, for example, has one rank that is composed of four ports of x8 DRAM modules. Each of these ports contains 10 DRAM modules.

For each such port, there are eight DRAM modules worth of data (64 bits) plus another DRAM module’s worth of error correction and other such data. There is also a spare DRAM module for each port that can be substituted for a failing port.

Two ports are combined into an ECC word and supply 128 bits of data. The ECC that is deployed can correct the result of an entire DRAM module that is faulty. This is also known as *Chipkill* correction. Then, it can correct at least an additional bit within the ECC word.

The additional spare DRAM modules are used so that when a DIMM experiences a Chipkill event within the DRAM modules under a port, the spare DRAM module can be substituted for a failing module. This process avoids the need to replace the DIMM for a single Chipkill event.

Depending on how DRAM modules fail, it might be possible to tolerate up to four DRAM modules failing on a single DIMM without needing to replace the DIMM, and then still correct an additional DRAM module that is failing within the DIMM.

Other DIMMs are offered with these systems. A 32 GB DIMM has two ranks, where each rank is similar to the 16 GB DIMM with DRAM modules on four ports, and each port has 10 x8 DRAM modules.

In addition, a 64 GB DIMM is offered through x4 DRAM modules that are organized in four ranks.

In addition to the protection that is provided by the ECC and sparing capabilities, the memory subsystem also implements scrubbing of memory to identify and correct single bit soft errors. Hypervisors are informed of incidents of single-cell persistent (hard) faults for deallocation of associated pages. However, because of the ECC and sparing capabilities that are used, such memory page deallocation is not relied on for repair of faulty hardware.

If a more substantial fault persists after all the self-healing capabilities are used, the hypervisor also can dynamically moving logical memory blocks from faulty memory to unused memory blocks in other parts of the system. This feature can take advantage of memory that is otherwise reserved for Capacity on Demand capabilities.

Finally, if an uncorrectable error in data is encountered, the memory that is affected is marked with a special uncorrectable error code and handled as described for cache uncorrectable errors.

4.3.11 I/O subsystem availability and Enhanced Error Handling

Use multi-path I/O and VIOS for I/O adapters and RAID for storage devices to prevent application outages when I/O adapter faults occur.

To permit soft or intermittent faults to be recovered without failover to an alternative device or I/O path, Power Systems hardware supports Enhanced Error Handling (EEH) for I/O adapters and PCIe bus faults.

EEH allows EEH-aware device drivers to try again after certain non-fatal I/O events to avoid failover, especially in cases where a soft error is encountered. EEH also allows device drivers to terminate if an intermittent hard error or other unrecoverable errors occur, while protecting against reliance on data that cannot be corrected. This action typically is done by “freezing” access to the I/O subsystem with the fault. Freezing prevents data from flowing to and from an I/O adapter and causes the hardware/firmware to respond with a defined error signature whenever an attempt is made to access the device. If necessary, a special uncorrectable error code can be used to mark a section of data as bad when the freeze is started.

In POWER8 processor-based systems, the external I/O hub and bridge adapters were eliminated in favor of a topology that integrates PCIe Host Bridges into the processor module itself. PCIe busses that are generated directly from a host bridge can drive individual I/O slots or a PCIe switch. The integrated PCIe controller supports try again (end-point error recovery) and freezing.

IBM device drivers under AIX are fully EEH-capable. For Linux under PowerVM, EEH support extends to many frequently used devices. There might be third-party PCI devices that do not provide native EEH support.

4.3.12 Remote Restart capability

The Power E850C server supports logical partitions that are capable of remote restarting. This option can be enabled for each partition individually. If a host system fails, the workload of the host system can be automatically restarted on an operational target system. To recover from unexpected failures and to improve the availability of the systems, workloads are restarted automatically.

The HMC can restart an AIX or Linux partition remotely if the partition supports an attribute called *encapsulated state*. An encapsulated state partition is a partition in which the configuration information and the persistent data are stored external to the server on persistent storage.

This capability requires that an HMC manage the host system and the operational target system. The Remote Restart capability requires that the following requirements are met:

- The server supports the remote restart capability. This is the case for the Power E850C.
- The partition must not have physical I/O adapters that are assigned to the partition.
- The partition must not be a full system partition, or a Virtual I/O Server.
- The partition must not be an alternative error-logging partition.
- The partition must not have a barrier synchronization register (BSR).
- The partition must not have huge pages (applicable only if PowerVM Active Memory Sharing is enabled).
- The partition must not have its rootvg volume group on a logical volume or have any exported optical devices.

Remote Restart can be enabled for applicable partitions through the HMC interface.

4.4 Availability impacts of a solution architecture

A solution should not rely only on the hardware platform. Despite IBM Power Systems having superior RAS features to other comparable systems, design a redundant architecture for the application to allow for easier maintenance tasks and greater flexibility.

When running in a redundant architecture, some tasks that would otherwise require that a application be brought offline, can now be executed with the application running, allowing for even greater availability.

When determining a highly available architecture that fits your needs, consider the following topics:

- Will I need to move my workload from an entire server during service or planned outages?
- If I use a clustering solution to move the workload, how will the failover time affect my service?
- If I use a server evacuation solution to move the workload, how long will it take to migrate all the partitions with my current server configuration?

4.4.1 Clustering

IBM Power Systems running under PowerVM, AIX, and Linux support a number of different clustering solutions. These solutions are designed to meet requirements not only for application availability in regard to server outages, but also data center disaster management, reliable data backups, and so on. These offerings include distributed applications with IBM DB2® PureScale, HA solutions that use clustering technology with IBM PowerHA SystemMirror®, and disaster management across geographies with PowerHA SystemMirror Enterprise Edition.

For more information, see the following references:

- *Guide to IBM PowerHA SystemMirror for AIX Version 7.1.3*, SG24-8167

  http://www.redbooks.ibm.com/abstracts/sg248167.html
- *IBM PowerHA SystemMirror for AIX Cookbook*, SG24-7739

  http://www.redbooks.ibm.com/abstracts/sg247739.html

4.4.2 Virtual I/O redundancy configurations

Within each server, the partitions can be supported by a single VIOS. However, if a single VIOS is used and that VIOS terminates for any reason (hardware or software caused), all the partitions that use that VIOS will terminate.

Dual VIOS partitions can only be created through an optional HMC. A Power E850C server that is managed through the Integrated Virtualization Manager (IVM) is limited to a single VIOS partition.

Using redundant VIOS partitions mitigates that risk. Maintaining redundancy of adapters within each VIOS, in addition to having redundant VIOS, avoids most faults that keep a VIOS from running. Multiple paths to networks and SANs are therefore advised. Figure 4-3 shows a diagram of a partition accessing data from two distinct Virtual I/O Servers, each one with multiple network and SAN adapters to provide connectivity.

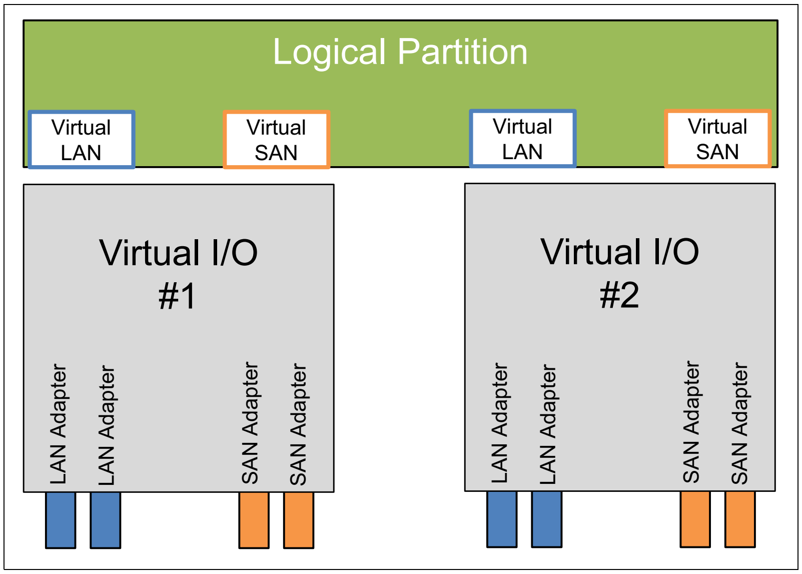

Figure 4-3 Partition using dual redundant virtual I/O servers for connectivity

Because each VIOS can largely be considered as an AIX based partition, each VIOS also needs the ability to access a boot image, have paging space, and other functions that require a root volume group or rootvg. The rootvg can be accessed through a SAN, through storage that is locally attached to a server, or through internal hard disks or solid-state devices. For best availability, the rootvg for each VIOS should use mirrored or RAID protected drives with redundant access to the devices.

4.4.3 PowerVM Live Partition Mobility

PowerVM Live Partition Mobility (LPM) allows you to move a running logical partition, including its operating system and running applications, from one system to another without any shutdown and without disrupting the operation of that logical partition. Inactive partition mobility allows you to move a powered-off logical partition from one system to another.

Live Partition Mobility provides systems management flexibility and improves system availability through the following functions:

- Avoid planned outages for hardware or firmware maintenance by moving logical partitions to another server and then performing the maintenance. Live Partition Mobility can help lead to zero downtime for maintenance because you can use it to work around scheduled maintenance activities.
- Avoid downtime for a server upgrade by moving logical partitions to another server and then performing the upgrade. This approach allows your users to continue their work without disruption.
- Avoid unplanned downtime. With preventive failure management, if a server indicates a potential failure, you can move its logical partitions to another server before the failure occurs. Partition mobility can help avoid unplanned downtime.

- Take advantage of server optimization:
  - Consolidation: You can consolidate workloads that run on several small, underused servers onto a single large server.
  - Optimized placement: You can move workloads from server to server to optimize resource use and workload performance within your computing environment. With live partition mobility, you can manage workloads with minimal downtime.

Live Partition Mobility can be completed between two systems that are managed by the same HMC. It is also possible to migrate partitions between two systems that are managed by different HMCs, which must be connected through a network.

Management through an HMC is optional for Power E850C servers. For systems managed by the IVM, partitions can only be migrated to other servers that are managed through an IVM, and these systems must be connected through a network.

Server Evacuation: This PowerVM function allows you to perform a Server Evacuation operation. Server Evacuation is used to move all migration-capable LPARs from one system to another if no active migrations are in progress on the source or the target servers.

With the Server Evacuation feature, multiple migrations can occur based on the concurrency setting of the HMC. Migrations are performed as sets, with the next set of migrations starting when the previous set completes. Any upgrade or maintenance operations can be performed after all the partitions are migrated and the source system is powered off.

You can migrate all the migration-capable AIX, IBM i, and Linux partitions from the source server to the destination server by running the following command from the HMC command line:

```
migrlpar -o m -m source_server -t target_server --all
```

Hardware and operating system requirements for Live Partition Mobility

Live Partition Mobility is supported by PowerVM Enterprise Edition in compliance with all operating systems that are compatible with POWER8 technology.

Logical partitions can only be relocated by using Live Partition Mobility if they are running in a fully virtualized environment, using external storage, which is accessible to both the exiting host server and the destination server.

VIOS partitions cannot be migrated.

For more information about Live Partition Mobility and how to implement it, see *IBM PowerVM Live Partition Mobility*, SG24-7460.

4.5 Serviceability

The purpose of serviceability is to repair or upgrade the system while attempting to minimize or eliminate service cost (within budget objectives) and maintaining application availability and high customer satisfaction. Serviceability includes system installation, miscellaneous equipment specification (MES) system upgrades and downgrades, and system maintenance or repair. Depending on the system and warranty contract, service might be performed by the customer, an IBM service support representative (SSR), or an authorized warranty service provider.

The serviceability features that are delivered in this system provide a highly efficient service environment by incorporating the following attributes:

- A design for customer setup (CSU), customer installable features (CIFs), and customer-replaceable units (CRUs)
- ED/FI incorporating FFDC
- Converged service approach across multiple IBM server platforms
- Concurrent Firmware Maintenance (CFM)

This section provides an overview of how these attributes contribute to efficient service in the progressive steps of error detection, analysis, reporting, notification, and repair found in all POWER processor-based systems.

4.5.1 Detecting errors

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

Although not all errors are a threat to system availability, those 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 storage devices.

4.5.2 Error checkers, fault isolation registers, and First-Failure Data Capture

IBM 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 and memory subsystem error-checkers, error-check signals are captured and stored in hardware 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 field-replaceable units (FRUs) that can be repaired when necessary.

Integral to the Power Systems design is the concept of FFDC. FFDC is a technique that involves sufficient error checking stations and coordination of fault reporting so that faults are detected and the root cause of the fault is isolated. FFDC also expects that necessary fault information can be collected at the time of failure without needing to re-create the problem or run an extended tracing or diagnostics program.

For the vast majority of faults, a good FFDC design means that the root cause is isolated at the time of the failure without intervention by a service representative. For all faults, good FFDC design still makes failure information available to the service representative. This information can be used to confirm the automatic diagnosis. More detailed information can be collected by a service representative for rare cases where the automatic diagnosis is not adequate for fault isolation.

4.5.3 Service processor

In POWER8 processor-based systems, the dedicated service processor is primarily responsible for fault analysis of processor and memory errors. The service processor is a microprocessor that is powered separately from the main instruction processing complex. In the Power E850C server, redundant connections to the service processor provide added reliability.

In addition to FFDC functions, the service processor performs many serviceability functions:

- Remote power control options
- Reset and boot features
- Environmental monitoring

  The service processor interfaces with the OCC function, which monitors the server's built-in temperature sensors and sends instructions to the system fans to increase rotational speed when the ambient temperature is above the normal operating range. By using an integrated operating system interface, the service processor notifies the operating system of potential environmental related problems so that the system administrator can take appropriate corrective actions before a critical failure threshold is reached. The service processor can also post a warning and start an orderly system shutdown in the following circumstances:

  - The operating temperature exceeds the critical level (for example, failure of air conditioning or air circulation around the system).
  - Internal component temperatures reach or exceed critical levels.
  - The system fan speed is out of operational specification (for example, because of multiple fan failures).
  - The server input voltages are out of operational specification.
- POWER Hypervisor (system firmware) and HMC connection surveillance

  The service processor monitors the operation of the firmware during the boot process, and also monitors the hypervisor for termination. The hypervisor monitors the service processor and can perform a reset and reload if it detects the loss of the service processor. If the reset or reload operation does not correct the problem with the service processor, the hypervisor notifies the operating system. The operating system can then take appropriate action, including calling for service. The service processor also monitors the connection to an HMC and can report loss of connectivity to the operating system partitions for system administrator notification.
- Uncorrectable error recovery

  The auto-restart (reboot) option, when enabled, can reboot the system automatically following an unrecoverable firmware error, firmware hang, hardware failure, or environmentally induced (ac power) failure.

  The auto-restart (reboot) option must be enabled from the ASMI menu or from the operator panel on the front of the server.

- Concurrent access to the service processor menu of the ASMI

  The ASMI provides management functionality for the server hardware through the service processor. Access to these menus allows nondisruptive changes to system default parameters, interrogation of service processor progress and error logs, and the ability to set and reset service indicators (Light Path). You can also access all service processor functions without having to power down the system to the standby state. The administrator or service representative can dynamically access the menu and functionality from any web browser-enabled console that is attached to the Ethernet support network, concurrently with normal system operation. Some options, such as changing the hypervisor type, do not take effect until the next boot.

- Management of the interfaces for connecting uninterruptible power source systems to the POWER processor-based systems and performing timed power-on (TPO) sequences.

4.5.4 Diagnosing

General diagnostic objectives are to detect and identify problems so that they can be resolved quickly. The IBM diagnostic strategy includes the following elements:

- Provide a common error code format that is equivalent to a system reference code, system reference number, checkpoint, or firmware error code.
- Provide fault detection and problem isolation procedures. Support a remote connection ability that is used by the IBM Remote Support Center or IBM Designated Service.
- Provide interactive intelligence within the diagnostic tests with detailed online failure information while connected to IBM back-end systems.

Using the extensive network of advanced and complementary error detection logic that is built directly into hardware, firmware, and operating systems, the Power E850C server can perform considerable self-diagnosis.

Because of the FFDC technology that is designed in to all Power Systems servers, re-creating diagnostic tests for failures or requiring user intervention is not necessary. Solid and intermittent errors are designed to be correctly detected and isolated at the time that the failure occurs. Runtime and boot time diagnostic tests fall into this category.

Boot time

When an IBM Power Systems server starts, the service processor initializes the system hardware. Boot-time diagnostic testing uses a multitier approach for system validation. It starts with managed low-level diagnostic tests that are supplemented with system firmware initialization and configuration of I/O hardware, followed by OS-initiated software test routines.

To minimize boot time, the system determines which of the diagnostic tests are required to be started to ensure correct operation. This determination is based on the way that the system was powered off, or on the boot-time selection menu.

Host Boot IPL

In POWER8, the initialization process during IPL changed. The flexible service processor (FSP) is no longer the only instance that initializes and runs the boot process. With POWER8, the FSP initializes the boot processes, but on the POWER8 processor itself, one part of the firmware is running and performing the Central Electronics Complex chip initialization. A new component called the PNOR chip stores the Host Boot firmware and the Self Boot Engine (SBE) is an internal part of the POWER8 chip itself that is used to boot the chip.

With this Host Boot initialization, new progress codes are available. An example of an FSP progress code is C1009003. During the Host Boot IPL, progress codes, such as CC009344, appear.

If a failure occurs during the Host Boot process, a new Host Boot System Dump is collected and stored. This type of memory dump includes Host Boot memory and is offloaded to the HMC when it is available.

Run time

All Power Systems servers can monitor critical system components during run time, and they can take corrective actions when recoverable faults occur. The hardware error-check architecture can report non-critical errors in the system in an *out-of-band* communications path to the service processor without affecting system performance.

A significant part of the runtime diagnostic capabilities originate with the service processor. Extensive diagnostic and fault analysis routines were developed and improved over many generations of POWER processor-based servers, and enable quick and accurate predefined responses to both actual and potential system problems.

The service processor correlates and processes runtime error information by using logic that is derived from IBM engineering expertise. This logic is used to count recoverable errors (called *thresholding*) and predict when corrective actions must be automatically initiated by the system. These actions can include the following items:

- Requests for a part to be replaced
- Dynamic invocation of built-in redundancy for automatic replacement of a failing part
- Dynamic deallocation of failing components so that system availability is maintained

Device drivers

In certain cases, diagnostic tests are best performed by operating system-specific drivers, most notably adapters or I/O devices that are owned directly by a logical partition. In these cases, the operating system device driver often works with I/O device microcode to isolate and recover from problems. Potential problems are reported to an operating system device driver that logs the error. In non-HMC managed servers, the OS can start the Call Home application to report the service event to IBM. For HMC managed servers, the event is reported to the HMC, which can initiate the Call Home request to IBM. I/O devices can also include specific exercisers that can be started by the diagnostic facilities for problem re-creation (if required by service procedures).

4.5.5 Reporting

In the unlikely event that a system hardware or environmentally induced failure is diagnosed, IBM Power Systems servers report the error through various mechanisms. The analysis result is stored in system NVRAM. Error log analysis (ELA) can be used to display the failure cause and the physical location of the failing hardware.

Using the Call Home infrastructure, the system can automatically send an alert or call for service during a critical system failure. A hardware fault also illuminates the amber system fault LED, which is on the system unit, to alert the user of an internal hardware problem.

On POWER8 processor-based servers, hardware and software failures are recorded in the system log. When a management console is attached, an ELA routine analyzes the error. The routine then forwards the event to the Service Focal Point (SFP) application that runs on the management console, and can notify the system administrator that it isolated a likely cause of the system problem. The service processor event log also records unrecoverable checkstop conditions, forwards them to the Service Focal Point application, and notifies the system administrator.

After the information is logged in the Service Focal Point application, if the system is correctly configured, a Call Home service request is initiated and the pertinent failure data with service parts information and part locations is sent to the IBM service organization. This information also contains the client contact information as defined in the IBM Electronic Service Agent (ESA) guided setup wizard. With the new HMC V8R8.1.0, a Serviceable Event Manager is available to block problems from being automatically transferred to IBM. For more information, see “Service Event Manager” on page 135.

Error logging and analysis

When the root cause of an error is identified by a fault isolation component, an error log entry is created with basic data, such as the following examples:

- An error code that uniquely describes the error event
- The location of the failing component
- The part number of the component to be replaced, including pertinent data such as engineering and manufacturing levels
- Return codes
- Resource identifiers
- FFDC data

Data that contains information about the effect that the repair has on the system is also included. Error log routines in the operating system and FSP can then use this information and decide whether the fault is a Call Home candidate. If the fault requires support intervention, a call is placed with service and support, and a notification is sent to the contact that is defined in the ESA-guided setup wizard.

Remote support

The Remote Monitoring and Control (RMC) subsystem is delivered as part of the base operating system, including the operating system that runs on the HMC. RMC provides a secure transport mechanism across the LAN interface between the operating system and the optional HMC. It is used by the operating system diagnostic application for transmitting error information. It also performs several other functions, but they are not used for the service infrastructure.

Service Focal Point application for partitioned systems

A critical requirement in a logically partitioned environment is to ensure that errors are not lost before being reported for service. In addition, each error should be reported only once, regardless of how many logical partitions experience the effect of the error. The Service Focal Point application on the management console or in the IVM is responsible for aggregating duplicate error reports, and ensures that all errors are recorded for review and management. The Service Focal Point application provides other service-related functions, such as controlling service indicators, setting up Call Home, and providing guided maintenance.

When a local or globally reported service request is made to the operating system, the operating system diagnostic subsystem uses the RMC subsystem to relay error information to the optional HMC. For global events (platform unrecoverable errors, for example), the service processor also forwards error notification of these events to the HMC, providing a redundant error-reporting path in case of errors in the RMC subsystem network.

The first occurrence of each failure type is recorded in the Manage Serviceable Events task on the management console. This task then filters and maintains a history of duplicate reports from other logical partitions or from the service processor. It then looks at all active service event requests within a predefined time span, analyzes the failure to ascertain the root cause and, if enabled, initiates a Call Home for service. This methodology ensures that all platform errors are reported through at least one functional path, ultimately resulting in a single notification for a single problem. Similar service functionality is provided through the Service Focal Point application on the IVM for providing service functions and interfaces on non-HMC partitioned servers.

Extended error data

Extended error data (EED) is additional data that is collected either automatically at the time of a failure or manually later. The data that is collected depends on the invocation method. It includes information such as firmware levels, operating system levels, additional fault isolation register values, recoverable error threshold register values, system status, and any other pertinent data.

The data is formatted and prepared for transmission back to IBM either to assist the service support organization with preparing a service action plan for the service representative or for additional analysis.

System dump handling

In certain circumstances, an error might require a memory dump to be automatically or manually created. In this event, the memory dump can be offloaded to the optional HMC. Specific management console information is included as part of the information that optionally can be sent to IBM Support for analysis. If additional information that relates to the memory dump is required, or if viewing the memory dump remotely becomes necessary, the management console memory dump record notifies the IBM Support center of which management console the memory dump is on. If no management console is present, the memory dump might be either on the FSP or in the operating system. The location depends on the type of memory dump that was initiated and whether the operating system is operational.

4.5.6 Notifying

After a Power E850C server detects, diagnoses, and reports an error to an appropriate aggregation point, it then takes steps to notify the administrator and if necessary, the IBM Support organization. Depending on the assessed severity of the error and support agreement, this notification might range from a simple message to having field service personnel automatically dispatched to the client site with the correct replacement part.

Client Notify

When an event is important enough to report, but does not indicate the need for a repair action or the need to call home to IBM Support, it is classified as *Client Notify.* Clients are notified because these events might be of interest to an administrator. The event might be a symptom of an expected systemic change, such as a network reconfiguration or failover testing of redundant power or cooling systems. These events include the following examples:

- Network events, such as the loss of contact over a local area network (LAN)
- Environmental events, such as ambient temperature warnings
- Events that need further examination by the client (although these events do not necessarily require a part replacement or repair action)

Client Notify events are serviceable events because they indicate that something happened that requires client awareness, and the administrator might want to take further action. These events can be reported to IBM at the discretion of the administrator.

Call Home

Call Home refers to an automatic or manual call from a customer location to an IBM Support structure with error log data, server status, and other service-related information. The Call Home feature starts procedures within the service organization so that the appropriate service action can begin. Call Home can be done through the HMC if available, or directly from machines that are not managed by an HMC.

Although configuring a Call Home function is optional, clients are encouraged to implement this feature to obtain service enhancements, such as reduced time to problem determination and faster, more accurate transmission of error information. In general, using the Call Home feature can result in increased system availability. The ESA application can be configured for automated Call Home. For more information, see 4.6.4, “Electronic Services and Electronic Service Agent” on page 133.

Vital product data and inventory management

Power Systems store vital product data (VPD) internally, which keeps a record of how much memory is installed, how many processors are installed, the manufacturing level of the parts, and similar system data. These records provide valuable information that can be used by remote support and service representatives, enabling the service representatives to help keep the firmware and software current on the server.

IBM Service and Support Problem Management database

At the IBM Support center, historical problem data is entered into the IBM Service and Support Problem Management database. All of the information that is related to the error, along with any service actions that are taken by the service representative, are recorded for problem management by the support and development organizations. The problem is then tracked and monitored until the system fault is repaired.

4.5.7 Locating and servicing

The final component of a comprehensive design for serviceability is the ability to effectively locate and replace parts that require 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 on the part or a release lever indicates that the system might not be required to be powered off to perform service. This feature depends on system configuration, and preparatory steps might be required before the service action is taken on the system. For any concurrent maintenance procedures, care should be taken to follow the steps that are indicated by the HMC or maintenance menus in the correct order.

  Blue on the part or on a release lever, latch, or thumb-screw indicates that the procedure might require the unit or system to be shut down before servicing or replacing the part. Check your service procedure before attempting repair, and ensure that you fully understand the process that is required before starting work.

- Tool-less design

  Selected IBM systems support tool-less or simple tool designs. These designs require no tools, or require basic tools such as flathead screw drivers, to service the hardware components.

- Positive retention

  Positive retention mechanisms help ensure proper connections between hardware components, such as from cables to connectors, and between two cards 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.

Light Path

The Light Path LED function of the Power E850C helps with parts that can be replaced by the user. In the Light Path LED implementation, when a fault condition is detected on the Power E850C server, an amber fault LED is illuminated (turned on solid), which also illuminates the system fault LED. The Light Path system pinpoints the exact part by illuminating the amber fault LED that is associated with the part that needs to be replaced.

The service representative can clearly identify components for replacement by using specific component level identity LEDs. The system can also guide the service representative directly to the component by signaling (flashing) the component identity LED, and illuminating the blue enclosure identity LED to identify the server in a busy data center.

After the repair, the LEDs shut off automatically after the part is replaced. The Light Path LEDs are only visible while the system is powered on or has standby power connected.

IBM Knowledge Center

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

The latest version of the documentation is accessible through the Internet. In addition, a CD-ROM based version is available.

The purpose of the 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 due to updates or addition of new capabilities can be accessed by service representatives immediately.

The IBM Knowledge Center contains sections specific to each server model, and include detailed service procedures for a number of potential repair situations. The service procedure repository for a particular server model can be found in the “Troubleshooting, service and support” section.

The IBM Knowledge Center can be found online at the following link:

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

Service labels

Service representatives use these labels to assist with maintenance actions. Service labels are in various formats and positions, and are intended to transmit readily available information to the service representative during the repair process.

The following are some of these service labels and their purposes:

- *Location diagrams* are strategically positioned on the system hardware and provide information about the placement of hardware components. Location diagrams can include location codes, drawings of physical locations, concurrent maintenance status, or other data that is pertinent to a repair. Location diagrams are especially useful when multiple components are installed, such as DIMMs, sockets, processor cards, fans, adapters, LEDs, and power supplies.
- *Remove or replace procedure labels* provide systematic procedures, including diagrams, detailing how to remove and replace certain serviceable hardware components. They are often found on the cover of the system or in other locations that are accessible to the service representative.
- *Numbered arrows* are used to indicate the order of operation and serviceability direction of components. Various serviceable parts, such as latches, levers, and touch points, must be pulled or pushed in a certain direction and order so that the mechanical mechanisms can engage or disengage. Arrows generally improve the ease of serviceability.

QR code labels for servicing information

A label that lists a QR code can be found on the top service cover of the Power E850C. This label can be scanned with an appropriate app on a mobile device to link to a number of sources of information that simplify the servicing of the system.

From this quick access link you can find information about the following topics, among others:

- Installing and configuring the system
- Troubleshooting and problem analysis
- Reference code lookup tables
- Part location guides
- Removing and replacing field replaceable units
- Video guides for removal and installation of customer replaceable units
- Warranty and maintenance contracts
- Full product documentation

The operator panel

The operator panel on a POWER processor-based system is an LCD display (two rows of 16 characters) that is used to present boot progress codes. It indicates advancement through the system power-on and initialization processes. The operator panel is also used to display error and location codes when an error occurs that prevents the system from booting. It includes several buttons that enable a service representative or administrator to change various boot-time options, and for other limited service functions.

Concurrent maintenance

The IBM POWER8 processor-based systems are designed with the understanding that certain components have higher intrinsic failure rates than others. These components include fans, power supplies, and physical storage devices. Other devices, such as I/O adapters, can become worn from repeated plugging and unplugging. For these reasons, these devices are concurrently maintainable when properly configured. This feature allows parts to be replaced while the system is still running, without requiring any downtime to applications. Concurrent maintenance is facilitated by the redundant design for the power supplies, fans, and physical storage.

The following system parts allow for concurrent maintenance:

- Disk drives and solid-state devices (SSDs)
- DVD drive
- Front fans
- PCIe adapters (including PCIe optical cable adapter for I/O expansion drawer)
- Power supplies
- Time-of-day battery card
- Operator panel

Maintenance procedures: Concurrent maintenance functions need to be initiated through a service interface such as the HMC (if available) or the ASMI service functions menu. Some concurrent maintenance can be initiated through the host operating system. Attempting to replace parts without following the correct procedures can lead to further faults or system damage.

Concurrent maintenance of the PCIe optical cable adapter requires an HMC.

Repair and verify services

Repair and verify (R&V) services are automated service procedures that are used to guide a service representative, step-by-step, through the process of repairing a system and verifying that the problem was repaired. The steps are customized in an appropriate sequence for the particular repair for the specific system being serviced. The following scenarios are covered by R&V services:

- Replacing a defective FRU or a CRU
- Reattaching a loose or disconnected component
- Correcting a configuration error
- Removing or replacing an incompatible FRU
- Updating firmware, device drivers, operating systems, middleware components, and IBM applications after replacing a part

R&V procedures can be used by service representatives who are familiar with the task and those who are not. Education-on-demand content is placed in the procedure at the appropriate locations. Throughout the R&V procedure, repair history is collected and provided to the Service and Support Problem Management Database for storage with the serviceable event. This process helps ensure that the guided maintenance procedures are operating correctly.

Clients can subscribe through the subscription services on the IBM Support Portal to obtain notifications about the latest updates that are available for service-related documentation.

4.6 Manageability

Several functions and tools are available to help manage the Power E850C server. These items allow you to efficiently and effectively manage your system alongside other Power Systems servers and other machines.

4.6.1 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. The following primary service interfaces are used:

- Light Path (See “Light Path” on page 123 and “Service labels” on page 123)
- Service processor through the ASMI menus
- Operator panel (See “The operator panel” on page 124)
- Operating system service menu
- Service Focal Point through the HMC
- Service Focal Point Lite through the IVM

Service processor

The service processor is a micro processor-based controller that runs its own operating system. It is a component of the service interface card.

The service processor operating system has specific programs and device drivers for the service processor hardware. The host interface is a processor support interface that is connected to the POWER8 processor. The service processor is always running, regardless of the main system unit’s state. The system unit can be in the following states:

- Standby (power off)
- Operating, ready to start partitions
- Operating with running logical partitions

The service processor is used to monitor and manage the system hardware resources and devices. The service processor checks the system for errors, ensuring that the connection to the management console for manageability purposes and accepting ASMI Secure Sockets Layer (SSL) network connections. The service processor can view and manage the machine-wide settings by using the ASMI, and enables complete system and partition management from the HMC.

> **Analyzing a system that does not boot:** The FSP can analyze a system that does not boot. Reference codes and detailed data are available in the ASMI and are transferred to the HMC.

The service processor uses two Ethernet ports that run at 1 Gbps speed. Consider the following information:

- Both Ethernet ports are visible only to the service processor and can be used to attach the server to an HMC or to access the ASMI. The ASMI options can be accessed through an HTTP server that is integrated into the service processor operating environment.
- Both Ethernet ports support only auto-negotiation. Customer-selectable media speed and duplex settings are not available.
- Both Ethernet ports have the following default IP address:
  - Service processor eth0 (HMC1 port) is configured as 169.254.2.147.
  - Service processor eth1 (HMC2 port) is configured as 169.254.3.147.

The following functions are available through the service processor:

- Call Home
- ASMI
- Error information (error code, part number, and location codes) menu
- View of guarded components
- Limited repair procedures
- Generate dump
- LED Management menu
- Remote view of ASMI menus
- Firmware update through a USB key

Advanced System Management Interface

ASMI is the interface to the service processor that enables you to manage the operation of the server, such as auto-power restart. ASMI also allows you to view information about the server, such as the error log and VPD. Various repair procedures require that you have a connection to the ASMI.

The ASMI is accessible through the HMC if used to manage the server. It is also accessible by using a web browser on a system that is connected directly to the service processor (in this case, either a standard Ethernet cable or a crossed cable) or through an Ethernet network. ASMI can also be accessed from an ASCII terminal, but this option is available only while the system is in the platform powered-off mode.

Use the ASMI to change the service processor IP addresses or to apply certain security policies and prevent access from unwanted IP addresses or ranges.

You might be able to use the service processor's default settings. In that case, accessing the ASMI is not necessary. To access ASMI, use one of the following methods:

- Use an HMC.

  If configured to do so, the HMC connects directly to the ASMI for a selected system from this task.

  To connect to the ASMI from an HMC, complete the following steps:

  a. Open **Systems Management** from the navigation pane.

  b. From the work window, select one of the managed systems.

  c. From the System Management tasks list, click **Operations** → **Launch Advanced System Management (ASM)**.

- Use a web browser.

  At the time of writing, the supported web browsers are Microsoft Internet Explorer (Version 10.0.9200.16439), Mozilla Firefox ESR (Version 24), and Chrome (Version 30). Later versions of these browsers might work, but are not officially supported. The JavaScript language and cookies must be enabled, and TLS 1.2 might need to be enabled.

  The web interface is available during all phases of system operation, including the initial program load (IPL) and run time. However, several of the menu options in the web interface are unavailable during IPL or run time to prevent usage or ownership conflicts if the system resources are in use during that phase. The ASMI provides an SSL web connection to the service processor. To establish an SSL connection, open your browser by using the following address:

  `https://<ip_address_of_service_processor>`

> **Note:** To make the connection through Internet Explorer, click **Tools** → **Internet Options**. Clear the **Use TLS 1.0** check box, and click **OK**.

- Use an ASCII terminal.

  The ASMI on an ASCII terminal supports a subset of the functions that are provided by the web interface. It is available only when the system is in the platform powered-off mode. The ASMI on an ASCII console is not available during several phases of system operation, such as the IPL and run time.

- Command-line start of the ASMI.

  Either on the HMC itself or when properly configured on a remote system, it is possible to start ASMI web interface from the HMC command line. Open a terminal window on the HMC or access the HMC with a terminal emulation and run the following command:

  ```
  asmmenu --ip <ip address>
  ```

On the HMC itself, a browser window opens automatically with the ASMI window and, when configured properly, a browser window opens on a remote system when issued from there.

The operator panel

The service processor provides an interface to the operator panel, which is used to display system status and diagnostic information. The operator panel can be accessed in two ways:

- By using the normal operational front view
- By pulling it out to access the switches and view the LCD display

The operator panel has these features:

- A 2 x 16 character LCD display
- Reset, enter, power On/Off, increment, and decrement buttons
- Amber system information or attention LED, and a green Power LED
- Blue enclosure identify LED
- Altitude sensor
- USB port
- Speaker

The following functions are available through the operator panel:

- Error information
- Generate dump
- View machine type, model, and serial number
- View or change IP addresses of the service processor
- Limited set of repair functions

Operating system service menu

The system diagnostic tests consist of stand-alone diagnostic tests that are loaded from the DVD drive, and online diagnostic tests (available in AIX).

Online diagnostic tests, when installed, are a part of the AIX operating system on the server. They can be booted in single-user mode (service mode), run in maintenance mode, or run concurrently (concurrent mode) with other applications. They have access to the AIX error log and the AIX configuration data.

The following modes are available:

- Service mode

  This mode requires a service mode boot of the system and enables the checking of system devices and features. Service mode provides the most complete self-check of the system resources. All system resources, except the SCSI adapter and the disk drives that are used for paging, can be tested.

- Concurrent mode

  This mode enables the normal system functions to continue while selected resources are being checked. Because the system is running in normal operation, certain devices might require extra actions by the user or a diagnostic application before testing can be done.

- Maintenance mode

  This mode enables the checking of most system resources. Maintenance mode provides the same test coverage as service mode. The difference between the two modes is the way that they are started. Maintenance mode requires that all activity on the operating system is stopped. Run **`shutdown -m`** to stop all activity on the operating system and put the operating system into maintenance mode.

The system management services (SMS) error log is accessible on the SMS menus. This error log contains errors that are found by partition firmware when the system or partition is booting.

The service processor's error log can be accessed on the ASMI menus. You can also access the system diagnostics from a Network Installation Management (NIM) server.

> **Alternative method:** When you order a Power E850C, a DVD-RAM drive is available as an option. An alternative method for maintaining and servicing the system must be available if you do not order the DVD-RAM drive.

Depending on the operating system, the following service-level functions are what you typically see when you use the operating system service menus:

- Product activity log
- Trace Licensed Internal Code
- Work with communications trace
- Display/Alter/Dump
- Licensed Internal Code log
- Main storage memory dump manager
- Hardware service manager
- Call Home/Customer Notification
- Error information menu
- LED management menu
- Concurrent/Non-concurrent maintenance (within scope of the OS)
- Managing firmware levels:
  - Server
  - Adapter
- Remote support (access varies by OS)

Service Focal Point on the Hardware Management Console

Service strategies become more complicated in a partitioned environment. The Manage Serviceable Events task in the management console can help streamline this process.

Each logical partition reports errors that it detects and forwards the event to the Service Focal Point application that is running on the management console, without determining whether other logical partitions also detect and report the errors. For example, if one logical partition reports an error for a shared resource, such as a managed system power supply, other active logical partitions might report the same error.

By using the Manage Serviceable Events task in the management console, you can avoid long lists of repetitive Call Home information by recognizing that these are repeated errors and consolidating them into one error.

In addition, you can use the Manage Serviceable Events task to initiate service functions on systems and logical partitions, including the exchanging of parts, configuring connectivity, and managing memory dumps.

> **Optional HMC:** An HMC is not required to run and manage a Power E850C server. If you are running multiple Power Systems servers, use an HMC for management tasks to simplify the overall management structure.

4.6.2 IBM Power Systems Firmware maintenance

The IBM Power Systems Client-Managed Microcode is a methodology that enables you to manage and install microcode updates on Power Systems and its associated I/O adapters.

Firmware entitlement

With the HMC Version V8R8.1.0.0 and POWER8 processor-based servers, the firmware installations are restricted to entitled servers. The customer must be registered with IBM and entitled with a service contract. During the initial machine warranty period, the access key is already installed in the machine by manufacturing. The key is valid for the regular warranty period plus some additional time. The Power Systems Firmware is relocated from the public repository to the access control repository. The I/O firmware remains on the public repository, but the server must be entitled for installation. When the **`lslic`** command is run to display the firmware levels, a new value, **`update_access_key_exp_date`**, is added. The HMC GUI and the ASMI menu show the Update access key expiration date.

When the system is no longer entitled, the firmware updates fail. Some new System Reference Code (SRC) packages are available:

- E302FA06: Acquisition entitlement check failed
- E302FA08: Installation entitlement check failed

Any firmware release that was made available during the entitled time frame can still be installed. For example, if the entitlement period ends on 31 December 2015, and a new firmware release is released before the end of that entitlement period, it can still be installed. If that firmware is downloaded after 31 December 2015, but it was made available before the end of the entitlement period, it can still be installed. Any newer release requires a new update access key.

> **Note:** The update access key expiration date requires a valid entitlement of the system to perform firmware updates.

You can find an update access key at the IBM CoD Home website:

http://www-912.ibm.com/pod/pod

To access the IBM entitled Software Support page for further details, go to the following website:

http://www.ibm.com/servers/eserver/ess

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 can be disruptive to customer operations. IBM intends to introduce no more than two new release levels per year. These release levels will be supported by service packs. Service packs contain only firmware fixes and do not introduce new functions. A service pack is an update to an existing release level.

If the system is managed by a management console, use the management console to perform system firmware updates. By using the management console, you can take advantage of the CFM option when concurrent service packs are available. CFM allows the firmware update process to be partially or wholly concurrent (nondisruptive). With the introduction of CFM, IBM is increasing its clients' opportunity to stay on a release level for longer periods. Clients who 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).
- They want to 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 required in the environment.
- A scheduled maintenance action causes a platform reboot, which provides an opportunity to also upgrade to a new firmware release.

Any required security patches or firmware fixes are incorporated into service packs for the life of a release level. Customers are not required to upgrade to the latest release level to ensure security and stability of their systems.

Firmware can also be updated by using a running partition.

The updating and upgrading of system firmware depends on several factors, including the current firmware that is installed, and what operating systems 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

You might also want to review the preferred practice white papers that are found at the following website:

http://www14.software.ibm.com/webapp/set2/sas/f/best/home.html

Firmware update steps

The system firmware consists of service processor microcode, Open Firmware microcode, and system power control network (SPCN) microcode.

The firmware and microcode can be downloaded and installed either from the HMC, or from a running partition.

Power Systems has a permanent firmware boot side (A side) and a temporary firmware boot side (B side). New levels of firmware must be installed first on the temporary side to test the update's compatibility with existing applications. When the new level of firmware is approved, it can be copied to the permanent side.

For access to the initial websites that address this capability, see the POWER8 section on the IBM Support Portal:

https://www.ibm.com/support/entry/portal/product/power

For POWER8 based Power Systems, select the **POWER8 systems** link. Within this section, search for **Firmware and HMC updates** to find the resources for keeping your system's firmware current.

If you use an HMC to manage the server, that interface can be used to view the levels of server firmware and power subsystem firmware that are installed, and those that are available to download and install.

Each IBM Power Systems server has the following levels of server firmware and power subsystem firmware:

- Installed level

  This level of server firmware or power subsystem firmware is installed, and will be installed into memory after the managed system is powered off and then powered on. It is installed on the temporary side of system firmware.

- Activated level

  This level of server firmware or power subsystem firmware is active and running in memory.

- Accepted level

  This level is the backup level of server or power subsystem firmware. You can return to this level of server or power subsystem firmware if you decide to remove the installed level. It is installed on the permanent side of system firmware.

Figure 4-4 shows the different levels as shown in the HMC.

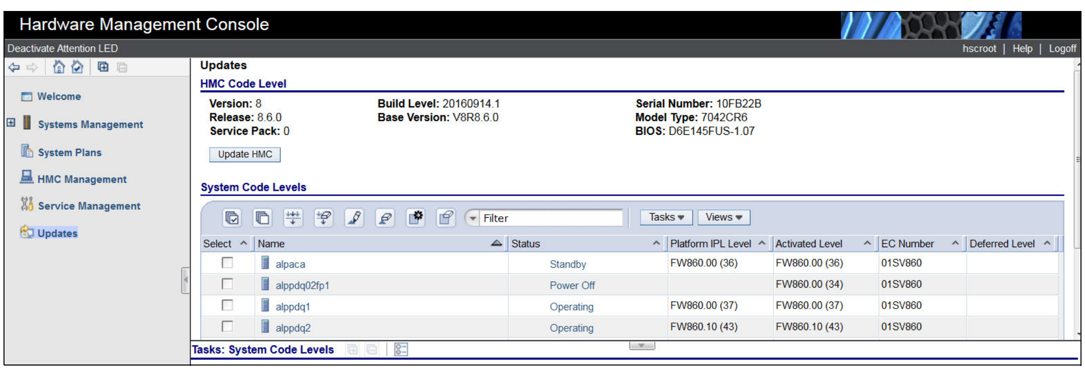

Figure 4-4 HMC and server firmware levels shown in the HMC

IBM provides the CFM function on the Power E850C model. This function supports applying nondisruptive system firmware service packs to the system concurrently (without requiring a reboot operation to activate changes).

The concurrent levels of system firmware can, on occasion, contain fixes that are known as *deferred*. These deferred fixes can be installed concurrently but are not activated until the next IPL. Deferred fixes, if any, are identified in the Firmware Update Descriptions table of the firmware document. For deferred fixes within a service pack, only the fixes in the service pack that cannot be concurrently activated are deferred.

Table 4-1 shows the file-naming convention for system firmware.

Table 4-1 Firmware naming convention

PPNNSSS_FFF_DDD			
PP	Package identifier	For example, 01	-
NN	Platform and class	For example, SV	Scale out systems
SSS	Release indicator		
FFF	Current fix pack		
DDD	Last disruptive fix pack		

The following example uses this convention: 01SV860_010_010 = Firmware for 8208-44E release 860 fix pack 10.

An installation is disruptive if the following statements are true:

- The release levels (SSS) of the currently installed and the new firmware differ.
- The service pack level (FFF) and the last disruptive service pack level (DDD) are equal in the new firmware.

Otherwise, an installation is concurrent if the service pack level (FFF) of the new firmware is higher than the service pack level that is installed on the system and the conditions for disruptive installation are not met.

4.6.3 Concurrent firmware maintenance improvements

Since POWER6, firmware service packs are concurrently applied and take effect immediately. Occasionally, a service pack is shipped where most of the features can be concurrently applied, but because changes to some server functions (for example, changing initialization values for chip controls) cannot occur during operation, a patch in this area required a system reboot for activation.

With the Power-On Reset Engine (PORE), the firmware can now dynamically power off processor components, change the registers, and reinitialize while the system is running, without discernible effect to any applications running on a processor. This feature allows concurrent firmware changes in POWER8, which in earlier designs required a reboot to take effect.

Activating new firmware functions requires installation of a higher firmware release level. This process is disruptive to server operations and requires a scheduled outage and full server reboot.

4.6.4 Electronic Services and Electronic Service Agent

IBM transformed its delivery of hardware and software support services to help you achieve higher system availability. Electronic Services is a web-enabled solution that offers an exclusive enhancement to the service and support that is available for IBM servers at no additional charge. These services provide the opportunity for greater system availability with faster problem resolution and preemptive monitoring. The Electronic Services solution consists of two separate, but complementary, elements:

- Electronic Services news page
- Electronic Service Agent

Electronic Services news page

The Electronic Services news page is a single Internet entry point that replaces the multiple entry points that are traditionally used to access IBM Internet services and support. With the news page, you can gain easier access to IBM resources for assistance in resolving technical problems.

Electronic Service Agent

The ESA is software that runs on the server and monitors events and transmits system inventory information to IBM on a periodic, client-defined timetable. The ESA automatically reports hardware problems to IBM.

Early knowledge about potential problems enables IBM to deliver proactive service that can result in higher system availability and performance. In addition, information that is collected through the Service Agent is made available to an IBM SSR when they help answer your questions or diagnose problems. Installation and use of ESA for problem reporting enables IBM to provide better support and service for your IBM server.

To learn how Electronic Services can work for you, see the following website (an IBM ID is required):

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

Electronic Services provides these benefits:

- Increased uptime

  The ESA tool enhances the warranty or maintenance agreement by providing faster hardware error reporting and uploading system information to IBM Support. This enhancement can translate to less time that is wasted monitoring the symptoms, diagnosing the error, and manually calling IBM Support to open a problem record.

  The tool's 24x7 monitoring and reporting mean no more dependence on human intervention or off-hours customer personnel when errors are encountered in the middle of the night.

- Security

  The ESA tool is secure in monitoring, reporting, and storing the data at IBM. The ESA tool securely transmits through the Internet (HTTPS or VPN), and can be configured to communicate securely through gateways to provide customers with a single point of exit from their site.

  Communication is one way. Activating ESA does not enable IBM to call into a customer's system. System inventory information is stored in a secure database that is protected behind IBM firewalls. It is viewable only by the customer and IBM. The customer's business applications or business data are never transmitted to IBM.

- More accurate reporting

  Because system information and error logs are automatically uploaded to the IBM Support center with the service request, customers are not required to find and send system information. This automated process decreases the risk of misreported or misdiagnosed errors.

  When inside IBM, problem error data is run through a data knowledge management system and knowledge articles are appended to the problem record.

- Customized support

  By using the IBM ID that you enter during activation, you can view system and support information by selecting **My Systems** at the Electronic Support website:

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

My Systems provides valuable reports of installed hardware and software by using information that is collected from the systems by ESA. Reports are available for any system that is associated with the customer's IBM ID. Premium Search combines the function of search and the value of ESA information, providing advanced search of the technical support knowledge base. Using Premium Search and the ESA information that was collected from your system, your clients can see search results that apply specifically to their systems.

For more information about how to use the power of IBM Electronic Services, contact your IBM SSR, or see the following website:

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

Service Event Manager

The Service Event Manager (SEM) allows the user to decide which of the Serviceable Events are called home with the ESA. It is possible to lock certain events. Some customers might not allow data to be transferred outside their company. However, after the SEM is enabled, the analysis of the possible problems might take longer.

The SEM can be enabled by running the following command:

```
chhmc -c sem -s enable
```

You can disable SEM mode and specify what state in which to leave the Call Home feature by running the following commands:

```
chhmc -c sem -s disable --callhome disable
chhmc -c sem -s disable --callhome enable
```

You can do the basic configuration of the SEM from the HMC GUI. After you select the Service Event Manager, as shown in Figure 4-5, you must add the HMC console.

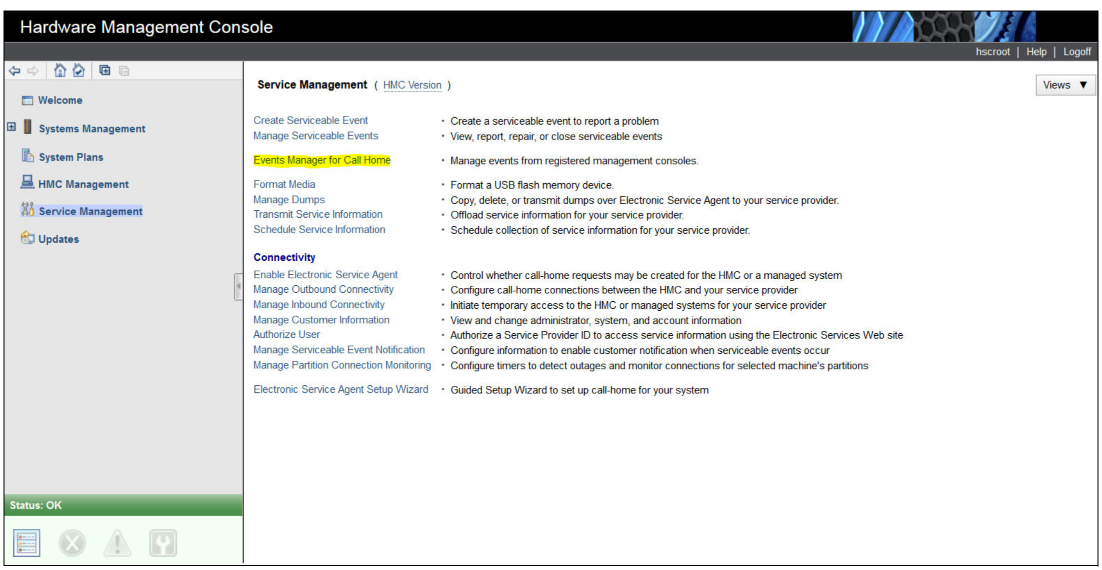

Figure 4-5 HMC selection for Service Event Manager

In the next window, you can configure the HMC that is used to manage the Serviceable Events and proceed with further configuration steps, as shown in Figure 4-6.

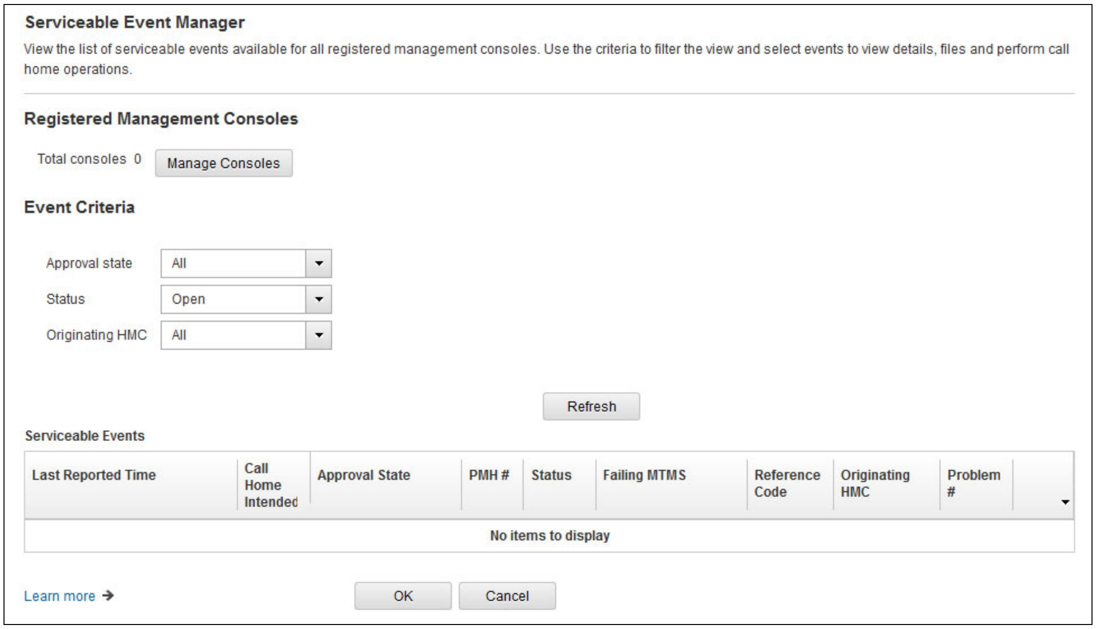

Figure 4-6 Initial SEM window

The following list provides detailed descriptions of the different configurable options:

- Registered Management Consoles

  "Total consoles" lists the number of consoles that are registered. Select **Manage Consoles** to manage the list of RMCs.

- Event Criteria

  Select the filters for filtering the list of serviceable events that is shown. After you make the selections, click **Refresh** to refresh the list based on the filter values.

- Approval state

  Select the value for approval state to filter the list.

- Status

  Select the value for the status to filter the list.

- Originating HMC

  Select a single registered console or **All consoles** to filter the list.

- Serviceable Events

  The Serviceable Events table shows the list of events based on the filters that are selected. To refresh the list, click **Refresh**.

The following menu options are available when you select an event in the table:

- **View Details**

  Shows the details of this event.

- **View Files**

  Shows the files that are associated with this event.

- **Approve Call Home**

  Approves the Call Home of this event. This option is available only if the event is not approved already.

The **Help/Learn more** function can be used to get more information about the other available windows for the Serviceable Event Manager.

4.7 Selected POWER8 RAS capabilities by operating system

Table 4-2 provides a list of the Power Systems RAS capabilities by operating system. The HMC is an optional feature on the Power E850C server.

Table 4-2 Selected RAS features by operating system

RAS feature	**AIX** **V7.2 TL1 SP1** **V7.1 TL3 SP5** **V6.1 TL9 SP5**	**Linux** **RHEL 6.8** **RHEL 7.2** **SUSE Linux Enterprise Server 12** **Ubuntu 16.04.01**
Processor		
FFDC for fault detection/error isolation	X	X
Processor instruction retry	X	X
Dynamic processor deallocation	X	X
Core error recovery		
► Alternative processor recovery	X	X
► Partition core contained checkstop	X	X
I/O subsystem		
PCI Express bus enhanced error detection	X	X
PCI Express bus enhanced error recovery	X	X
PCI Express card hot-swap	X	X
Cache availability		
Cache line removal	X	X
Dynamic bit-line sparing	X	X
Special uncorrectable error handling	X	X
Memory availability		
Memory page deallocation	X	X
Dynamic DRAM sparing	X	X
Periodic memory scrubbing	X	X
Special uncorrectable error handling	X	X

RAS feature	AIX V7.2 TL1 SP1 V7.1 TL3 SP5 V6.1 TL9 SP5	Linux RHEL 6.8 RHEL 7.2 SUSE Linux Enterprise Server 12 Ubuntu 16.04.01
Fault detection and isolation		
Storage Protection Keys	X	Not used by OS
Error log analysis	X	X
Serviceability		
Boot-time progress indicators	X	X
Firmware error codes	X	X
Operating system error codes	X	X
Inventory collection	X	X
Environmental and power warnings	X	X
Hot-swap DASD/media	X	X
Dual disk controllers/split backplane	X	X
Active-active dual disk controllers	optional	optional
EED collection	X	X
SP “Call Home” on non-HMC configurations	X	X
IO adapter/device stand-alone diagnostic tests with PowerVM	X	X
SP mutual surveillance with POWER Hypervisor	X	X
Concurrent firmware update with HMC	X	X
Service Agent Call Home Application	X	X
Service indicator LED support	X	X
System dump for memory, POWER Hypervisor, and SP	X	X
IBM Knowledge Center/IBM Systems Support Site service publications	X	X
System Support Site education	X	X
Operating system error reporting to HMC Service Focal Point application	X	X
RMC secure error transmission subsystem	X	X
Healthcheck scheduled operations with HMC	X	X
Operator panel (real or virtual)	X	X
Concurrent operator panel maintenance	X	X
Redundant HMCs	optional	optional
Automated server recovery/restart	optional	optional

RAS feature	AIX V7.2 TL1 SP1 V7.1 TL3 SP5 V6.1 TL9 SP5	Linux RHEL 6.8 RHEL 7.2 SUSE Linux Enterprise Server 12 Ubuntu 16.04.01
Repair and Verify guided maintenance with HMC	X	X
PowerVM Live Partition/Live Application Mobility with PowerVM Enterprise Edition	X	X
Emergency power-off warning (EPOW)		
EPOW errors handling	X	X

Related publications

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

IBM Redbooks

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

- *IBM Power System E850 Technical Overview and Introduction*, REDP-5222
- *IBM Power Systems E870 and E880 Technical Overview and Introduction*, REDP-5137
- *IBM Power System E850C Technical Overview and Introduction*, REDP-5412
- *IBM Power Systems HMC Implementation and Usage Guide*, SG24-7491
- *IBM Power Systems S812L and S822L Technical Overview and Introduction*, REDP-5098
- *IBM Power System S812LC Technical Overview and Introduction*, REDP-5284
- *IBM Power Systems S814 and S824 Technical Overview and Introduction*, REDP-5097
- *IBM Power System S821LC Technical Overview and Introduction*, REDP-5406
- *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 SR-IOV: Technical Overview and Introduction*, REDP-5065
- *IBM PowerVC Version 1.3.1 Introduction and Configuration Including IBM Cloud PowerVC Manager*, SG24-8199
- *IBM PowerVM Best Practices*, SG24-8062
- *IBM PowerVM Enhancements What is New in 2013*, SG24-8198
- *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- *Integrated Virtualization Manager for IBM Power Systems Servers*, REDP-4061
- *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:

- *Active Memory Expansion: Overview and Usage Guide* documentation:
  http://www.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=POW03037USEN
- *IBM EnergyScale for POWER8 Processor-Based Systems* white paper:
  http://public.dhe.ibm.com/common/ssi/ecm/po/en/pow03125usen/POW03125USEN.PDF
- IBM POWER8 systems facts and features
  http://www.ibm.com/systems/power/hardware/reports/factsfeatures.html
- 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 Systems 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 E850C server specifications:
  http://www.ibm.com/systems/power/hardware/e850c/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

- *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

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 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 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

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

REDP-5412-00

ISBN 0738455687

Printed in U.S.A.

Get connected

Redbooks®

ibm.com/redbooks