overview of the new Itanium® 2-based hp servers rx2600 and rx5670: how hp is putting Intel® Itanium 2 processors to work

an executive white paper from hp

july 2002

table of contents

executive summary

For many types of computing, systems based on the Intel® Itanium® 2 processor offer markedly superior performance over other systems that utilize RISC or IA-32 processors, while still maintaining a very competitive price. As the first servers in the computer industry to offer the awesome compute power of the Itanium 2 processor, the HP Servers rx2600 and rx5670 signal a dramatic change in the 64-bit server market.

The Intel Itanium 2 processor is the first volume microprocessor in the Intel Itanium processor family architecture, which was co-developed by HP and Intel. Both of the new HP systems also utilize the exclusive HP zx1 Chipset, which supports this new 64-bit processor with industry-leading high bandwidth and low latency. In the rx2600 and rx5670, HP is proving the Intel Itanium architecture concept of highly-parallel, 64-bit performance with a pair of servers that are full featured, have a unique high-bandwidth system architecture, and use industry-leading commodity parts to ensure compatibility and economical pricing.

This white paper provides details about these new HP servers, shows their features and benefits, discusses their electrical architectures, and provides a breakdown of their unique high-availability features. What's more, this paper illustrates how HP makes the transition from RISC and IA-32 servers easy, while preserving our customers' investments in existing software. Finally, the paper provides an appendix with a technical discussion of the Itanium 2 microarchitecture.

introducing the Intel Itanium 2 processor

The Intel Itanium 2 processor is a milestone in the continuing evolution of microprocessors because it will be the first enterprise-class 64-bit processor that has the power to become pervasive. Let's take a closer look at what this means.

Today, 32-bit servers or proprietary 64-bit RISC servers are the norm. They have respectable price/performance ratios but are either fundamentally limited in performance scalability or are exceedingly expensive.

Servers based on the Intel-architecture 32-bit (IA-32) processors, for example, are unable to address large amounts of memory. Meanwhile, 64-bit RISC architectures have the necessary performance and addressing, but they are more expensive and generally lock the customer into a proprietary operating environment and a single computer vendor. End users and IT professionals alike are clamoring for high performance and large addressing at an economical price. And they also are asking for pervasiveness so that they do not have to deal with multiple architectures and complexity.

HP and Intel co-developed the Intel Itanium 2 microarchitecture. The result is a high-performance, parallel 64-bit architecture that has the performance headroom to grow in the future and can be priced at a level to ensure its widespread adoption. The Itanium 2 microarchitecture fulfills both of these promises and is likely to become pervasive very quickly. The Itanium 2 processor is the fundamental building block of HP's new entry-level servers, the rx2600 and the rx5670.

Itanium 2-based rx2600 and rx5670 servers from hp

HP's Itanium 2-based servers are targeted at performance-hungry markets such as technical and scientific computing, Secure Sockets Layer (SSL) Web serving, application serving, and database applications. What's more, these systems are highly affordable, making them extremely attractive to software developers. Read on to learn about the system features that fulfill these promises.

hp server rx2600 at a glance

The HP Server rx2600 is the industry's first 2-way Itanium 2-based server. It has a sleek 2U footprint, but it can be equipped with up to two 1-GHz Intel Itanium 2 processors loaded with 3 MB of on-chip L3 cache and as much as 12 GB of RAM. Soon after the initial release of this server, larger memory modules will increase RAM scalability to 24 GB. This means that it has extraordinary compute density. Fitting 20 servers into a 2-meter rack delivers an astounding 160 gigaflops of potential power.

The HP Server rx2600 also has extensive availability and management features, which make it ideal for deployments in mission-critical data centers or compute-intensive server farms. Features such as hot-swap redundant components, memory chip spare, an integrated management processor, and high availability clustering support make the rx2600 the clear leader among 2-way Itanium 2-based servers.

The HP rx2600 is flexible, too, with a choice of 64-bit operating systems—HP-UX, Linux, or Microsoft® Windows®—to suit any need. It can be installed in a rack or in a standalone, vertical tower configuration. And there's a full range of HP storage peripherals and I/O adapters to complete the package.

figure 1. the hp server rx2600 can be installed in a rack or as a standalone unit

hp server rx2600 details

hp server rx2600 product specifications

- central processor
  - 1 or 2 Intel Itanium 2 processors at 900 MHz or 1 GHz
- cache (all on-chip)
  - 16 KB instruction and 16 KB data level 1 cache
  - 256 KB level 2 cache
  - 1.5 MB level 3 cache (900 MHz)
  - 3.0 MB level 3 cache (1 GHz)
- main memory
  - 1 GB minimum to 12 GB maximum (24 GB expected in 2003) PC2100 parity-protected ECC chip spare DDR CL2 memory in 12 DIMM slots (DIMMs must be installed in groups of four)
  - 8.5 GB/s memory bandwidth
- chipset
  - HP zx1 Chipset
  - 80 nanosecond memory latency
  - 6.4 GB/s system bus bandwidth
  - 4.0 GB/s aggregate I/O bandwidth
- expansion slots
  - 1 PCI-X, 1 GB/s sustained, 64-bit 133 MHz
  - 3 PCI-X, 0.5 GB/s sustained, 64-bit 133 MHz
  - each slot is full-length and has an independent bus
- hot-plug disk drives (3 bays for 1-in high 3.5-in disks)
  - 219 GB maximum internal storage
  - integrated dual-channel SCSI controller
  - disk sizes available: 36 GB 10,000 rpm; 36 GB 15,000 rpm; and 73 GB 10,000 rpm
- removable media
  - one slimline media bay for optional IDE optical drives
  - choice of 16X DVD-ROM or 16X/10X/40X CD-RW (also reads DVD)
- core I/O interconnect ports
  - Gigabit-TX LAN with RJ-45 connector (10/100/1000BT autosensing)
  - 10/100BT LAN with RJ-45 connector and Wake-on-LAN capability
  - 2 general-purpose RS-232 serial ports
  - VGA
  - 4 USB Series A 2.0 (480 Mb/s) ports
  - management processor interconnect
    - 10/100BT management LAN with Web console access
    - RS-232 local console
    - RS-232 remote/modem console
    - RS-232 general purpose
- power and cooling
  - one 650W hot-swap power supply standard
  - optional second 650W hot-swap power supply for N+1 redundancy
  - 4 cooling fans with N+1 redundancy
- power requirements
  - input current: 100–120V 7.2A/200–240V 3.2A (autoranging)
  - line frequency: 50 Hz to 60 Hz
  - maximum power input: 714W

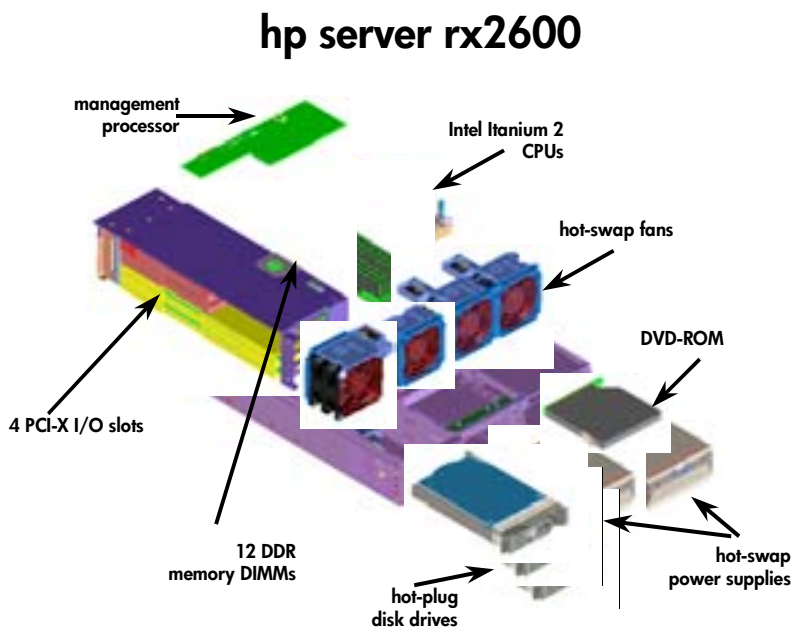
physical and environmental specifications

- environmental specifications
  - altitude
    - operating: 3000 m (10,000 ft) max
    - storage: 4600 m (15,000 ft) max
  - temperature
    - operating: +05°C to +35°C
    - non-operating: –40°C to +70°C
  - humidity
    - operating: 15% to 80% (relative)
- physical dimensions in rack orientation
  - height: 8.6 cm (3.4 in)
  - width: 48.2 cm (19 in)
  - depth: 68.0 cm (26.8 in)
- physical dimensions in standalone, vertical orientation
  - height: 49.5 cm (19.5 in)
  - width: 29.7 cm (11.7 in)
  - depth: 67.3 cm (26.5 in)
- net weight
  - minimum standalone configuration: 22.4 kg (49.5 lb)
  - maximum standalone configuration: 25.5 kg (56.3 lb)
  - minimum rack configuration: 17.5 kg (38.6 lb)
  - maximum rack configuration: 22.2 kg (49.0 lb)

mechanical design and packaging of the hp server rx2600

The exploded view reveals the location of major components as well as the mechanical and architectural features of the rx2600. The server is partitioned into three electrical partitions—the system board, including CPUs, memory, and core I/O; the I/O backplane, including four PCI-X I/O slots; and the management processor board.

figure 2. major components of the hp server rx2600



Two hot-swap power supply bays are located in the lower right corner of the server (when viewed from the front). Just above the power supplies is a slimline optical media drive bay, supporting either a DVD or a DVD/CD-RW combo drive. To the left of the unit's front are three bays for hot-plug hard disk drives. Directly behind the power supplies and peripheral bays are four hot-swap cooling fans.

The left side of the system houses the I/O backplane and I/O card bay. There are four PCI-X slots in the I/O card bay.

The right rear of the server contains the main system board. The system board contains two Intel Itanium 2 CPU sockets, 12 memory DIMM slots, and the core I/O controllers. The management processor sits on an independent circuit board that attaches to the rear of the main system board.

racking density

The rx2600 is designed to provide unprecedented performance density. At only 2 EIA units (one EIA unit = 1.75 inches) per server, up to 20 systems can be installed into a single 2-meter HP cabinet.

The rx2600 is supported in HP's A490xA cabinets. The rx2600 is also supported in a variety of third-party, non-HP racks and cabinets.

Note: Dimensions for rack configuration are as follows: H = 3.4 inches (8.6 cm), D = 26.8 inches (68.0 cm), W = 19 inches (48.2 cm).

cabinet spacing requirements

For proper ventilation, the rx2600 requires a minimum of 24 inches (61 cm) of free space at both the front and rear of the cabinet. During product installation and servicing, a total of 32 inches (82 cm) of free space is needed at the front of the cabinet. The depth of HP A490xA cabinet is 39 inches (99 cm); therefore, a minimum of 87 inches (221 cm) of total space is needed for each cabinet during normal operation. An additional 8 inches (21 cm) is needed during installation and servicing.

standalone pedestal configuration

The rx2600 is also available in a standalone configuration when a cabinet is not desired. The standalone system is ideal for an office environment, under a desk, or on a shelf. The standalone configuration uses a stylish tower mounting shell, with the system simply placed inside this shell.

Note: Dimensions for standalone/pedestal configuration: H = 19.5 inches (48.5 cm), D = 26.5 inches (67.2 cm), W = 11.7 inches (29.7 cm).

workstation or server? here's how to choose

Hewlett-Packard recently introduced the HP Workstation zx6000, which also uses the Intel Itanium 2 processor. This section describes the technical differences between the HP Workstation zx6000 and the HP Server rx2600.

The HP Server rx2600 and HP Workstation zx6000 are both 1- or 2-way Itanium 2-based systems. Both utilize HP's high-performance zx1 Chipset. Beyond that, however, the server is a very different class of system.

The HP Server rx2600 is optimized for high-performance server I/O with four 133-MHz PCI-X slots, including one full-bandwidth 1-GB/s slot. The HP Workstation zx6000 is optimized for high-performance AGP graphics support. Moreover, the HP Server rx2600 includes many high-availability features that are critical in commercial and many technical server environments; these high-availability features such as chip spare memory and high-availability clustering support are unique to the HP Server rx2600. The rx2600 is the best choice for compute node clusters and mission-critical commercial applications because of its advanced connectivity, remote manageability, and superior availability.

The systems also support different versions of HP's strategic operating systems. The HP Server rx2600 supports the commercial and technical operating environments of HP-UX as well as Microsoft's Advanced Server operating system. The HP Workstation zx6000 supports the client and technical versions of HP-UX as well as the Microsoft Windows XP client operating system. For Linux, the workstation and server support different distributions, focused on either client or server computing.

The following table summarizes the differences between the HP Server rx2600 and HP Workstation zx6000.

comparing the hp workstation zx6000 with the hp server rx2600

	hp workstation zx6000	hp server rx2600
category	Optimized for high-performance AGP graphics support for the 1- or 2-way technical computing workstation markets	Optimized for high-performance PCI-X I/O for the 2-way server and performance cluster markets
target usage model	Computer aided engineering, scientific research, life sciences, digital content creation rendering, mechanical CAD, workstation software development, graphics clusters	Computer aided engineering clusters, scientific research clusters, secure Web serving, application serving, server software development
Intel Itanium 2 processors	1 or 2 900-MHz with 1.5 MB L3 on-chip cache, or 1 or 2 1-GHz with 3 MB L3 on-chip cache	1 or 2 900-MHz with 1.5 MB L3 on-chip cache, or 1 or 2 1-GHz with 3 MB L3 on-chip cache
memory capacity	12 GB (24 GB when 2 GB DIMMs are available) Installed in pairs	12 GB (24 GB when 2 GB DIMMs are available) Installed in quads
system bus bandwidth	6.4 GB/s	6.4 GB/s
system memory bandwidth	8.5 GB/s	8.5 GB/s
aggregate I/O bandwidth (at hp zx1 chipset channels)	4.0 GB/s	4.0 GB/s
I/O slots	1 AGP-4X (1 GB/s) 66 MHz 3 PCI-X (0.5 GB/s) 133 MHz	1 PCI-X (1 GB/s) 133 MHz 3 PCI-X (0.5 GB/s) 133 MHz
operating systems supported	HP-UX client and technical versions Microsoft Windows (client version) 64-bit Linux (workstation distributions)	HP-UX 11i v 1.6 OE HP-UX 11I v 1.6 technical OE Microsoft Windows Advanced Server 64-bit Linux
graphics support	3D and 2D AGP	2D via integrated VGA port
audio support	Yes—optional PCI card	Not available
internal storage	3 3.5-in bays for hot-plug Ultra160 or Ultra320 SCSI devices	3 3.5-in bays for hot-plug Ultra160 or Ultra320 SCSI devices
management processor card	Optional	Standard
hot-plug disks	Yes	Yes
memory chip spare	No	Yes
redundant hot-swap power supplies	Yes	Yes
redundant hot-swap fans	Yes	Yes
mc/serviceguard clustering support	No	Yes
tower configuration	Yes (with quieter fans for deskside use)	Yes (with redundant fans)
rack-optimized configuration	Yes—2U Maximum 20 systems in a 2-meter rack	Yes—2U Maximum 20 systems in a 2-meter rack

hp server rx5670 at a glance

The HP Server rx5670 is the industry's most expandable 4-way Itanium 2-based server. The rx5670 can be equipped with up to four 1-GHz Intel Itanium 2 processors loaded with 3 MB of on-chip L3 cache, as much as 48 GB of RAM, and ten PCI-X I/O expansion slots. Simply put, the rx5670 is the highest performing, most expandable 4-way Itanium 2-based server on the market.

The rx5670 also has extensive availability and management features, which make it ideal for deployment in mission-critical data centers or compute-intensive server farms. Features such as hot-swap redundant components, memory chip spare, an integrated management processor, and high-availability clustering support make the rx5670 ideal for nearly any computing environment.

Furthermore, the rx5670 was designed for investment protection. It is the only Itanium 2-based server that can be "built" by a board-swap, in-chassis upgrade from an existing RISC server. In fact, *any HP rp5400, rp5430, rp5450, rp5470, or HP 9000 L-Class server can be upgraded to an rx5670!* This allows HP's PA-RISC customers to upgrade their investment to the Intel Itanium 2 microarchitecture.

The HP rx5670 is flexible, too. It offers a choice of 64-bit operating systems, including HP-UX, Linux, or Microsoft Windows. And it can be installed in a rack or in a standalone, pedestal configuration along with a host of choices from the full range of HP storage peripherals and I/O adapters.

figure 3. the hp server rx5670 is suitable for rack-mount or standalone duty

hp server rx5670 details

hp server 5670 product specifications

- central processor
  - 1, 2, 3, or 4 Intel Itanium 2 processors at 900 MHz or 1 GHz
- cache (all on-chip)
  - 16 KB instruction and 16 KB data level 1 cache
  - 256 KB level 2 cache
  - 1.5 MB level 3 cache (900 MHz)
  - 3.0 MB level 3 cache (1 GHz)
- main memory
  - 1 GB minimum to 48 GB maximum PC2100 parity-protected ECC chip spare DDR CL2 memory
  - 1 or 2 memory carrier boards with 24 DIMM slots each
  - 12.8 GB/s memory bandwidth
- chipset
  - HP zx1 Chipset
  - 105 ns memory latency
  - 6.4 GB/s system bus bandwidth
  - 4.0 GB/s aggregate I/O bandwidth
- expansion slots
  - 3 PCI-X on independent buses, 64-bit 133 MHz
  - 6 PCI-X on 3 shared buses, 64-bit 66 MHz
  - 1 PCI for graphics/USB, 64-bit 33 MHz
- hot-plug disk drives (4 bays)
  - 272 GB maximum internal storage
  - 2 integrated dual-channel Ultra160 SCSI controllers
  - disk sizes available: 36 GB 10,000 rpm; 36 GB 15,000 rpm; and 73 GB 10,000 rpm
- removable media
  - one open media bay for optional SCSI optical drives
  - choice of DVD-ROM or DDS-3
- core I/O interconnect ports
  - Gigabit-TX LAN with RJ-45 connector (10/100/1000BT autosensing)
  - Ultra160 SCSI port for external peripherals
  - management processor interconnect
    - 10/100BT management LAN with Web console access
    - RS-232 local console
    - RS-232 remote/model console
    - RS-232 general purpose
  - 2 USB Series A 2.0 (480 Mb/s) ports (optional)
  - VGA (optional)
- power and cooling
  - two 930W hot-swap power supplies standard
  - optional third 930W hot-swap power supply for N+1 redundancy
  - 8 cooling fans with N+1 redundancy
- power requirements
  - input current: 100–120V 10A/200–240V 5A (autoranging)
  - line frequency: 50 Hz to 60 Hz
  - typical power requirements (max configuration): 950W
  - theoretical maximum power input: 2089W

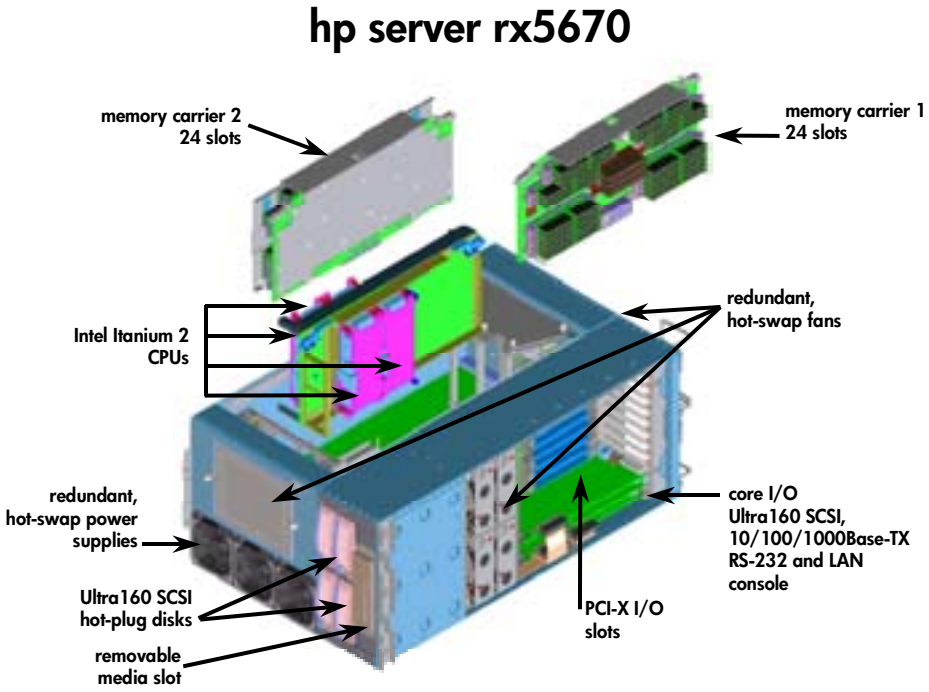
physical and environmental specifications

- environmental specifications
  - altitude
    - operating: 3000 m (10,000 ft) max
    - storage: 4600 m (15,000 ft) max
  - temperature
    - operating: +05°C to +35°C
    - non-operating: –40°C to +70°C
  - humidity
    - operating: 15% to 80% (relative)
- physical dimensions in rack orientation
  - height: 31.1 cm (12.25 in)
  - width: 48.2 cm (19 in)
  - depth: 77.4 cm (30.5 in)
- net weight
  - maximum configuration: 73 kg (160 lb)

mechanical design and packaging of the hp server rx5670

The exploded view illustrates the location of major components as well as the mechanical and architectural features of the rx5670. The server is partitioned into two main electrical partitions—the system partition, including baseboard, CPU board, and memory carrier boards; and the I/O partition, consisting of PCI-X I/O slots, core I/O, and the management processor.

figure 4. major components of the hp server rx5670



Three hot-swap power supply bays are located in the lower left corner of the server (when viewed from the front). To the right of the front panel, a peripheral bay provides space for four hot-plug disks and one removable media device (either DVD-ROM or DDS-3). Directly above the power supply bays is the first of eight hot-swap cooling fans.

The right side of the system houses the I/O backplane and I/O card bay. There are twelve PCI-X/PCI slots in the I/O card bay. Two or three of these slots are factory-loaded with core I/O cards, depending on the choice of operating system. The remaining nine or 10 slots are available for a wide variety of optional I/O adapter cards. Two pairs of fans provide cooling for the I/O bay as well as the peripheral bay.

An opening at the top of the server provides access to the CPU board, memory carriers, and system baseboard. In aggregate, these boards provide support for up to four CPUs and 48 dual inline memory modules (DIMMs).

racking density

The rx5670 is designed to provide performance density that easily adapts to different environments. At 7 EIA units per server, up to five servers can be installed into a single 2-meter HP cabinet.

The rx5670 is supported in HP's A490xA and A189xA cabinets. When using the high-availability slider rails, bolt-on anti-tip feet are required. In addition, when using the slider in A189xA cabinets, a ballast is required. (See the *HP Server Configuration Guide* for details.)

The rx5670 is also supported in a variety of third-party, non-HP racks and cabinets. Refer to the *HP Server Configuration Guide* for the latest list of qualified third-party racks.

Note: Dimensions for rack configuration: H = 12.25 inches (31.1 cm), D = 30.5 inches (77.5 cm), W = 19 inches (48.2 cm).

high-availability slider rails

There are two rail options, static or slider, available for racking the rx5670 into an HP cabinet. The high-availability (HA) slider rails were designed to allow easy service access to the system, as well as to enable hot-swapping of the four fans in the side cavity.

With the HA slider rails the unit can be completely serviced without removing it from the rack, thus allowing side-by-side racks of systems to be completely supported without sacrificing floor space for side access to the system. For this reason, the high-availability slider rails are highly recommended.

Note: The slider mechanism occupies 1 EIA unit of rack space. When used with the rx5670, the combination will occupy 8 EIA units of rack space. Static rails do not consume EIA space within the cabinet, leaving more EIA space for peripherals. However, using static rails prohibits hot-swapping of the I/O bay fans.

cabinet spacing requirements

The rx5670 requires a minimum of 24 inches (61 cm) of free space in both the front and rear of the cabinet for proper ventilation. During product installation and servicing, a total of 32 inches (82 cm) of free space is needed at the front of the cabinet. The depth of HP A490xA cabinets is 39 inches (99 cm). Therefore, a minimum of 87 inches (221 cm) of total space is needed for each cabinet during normal operation. An additional 8 inches (21 cm) is needed during installation and servicing.

standalone/deskside configuration

When a cabinet is not desired, the rx5670 is also available in a standalone configuration. The standalone system is ideal for an office environment, under a desk, or on a shelf. The standalone configuration utilizes the same internal chassis and front plastic bezel as the racked version. However, a sheet metal cover, base, and casters are added for functionality and aesthetics. Casters can be removed when not desired.

Note: Dimensions for standalone/deskside configuration: H = 14.5 inches (36.8 cm), D = 30.5 inches (77.5 cm), W = 19 inches (48.2 cm).

in-box upgrade from PA-RISC servers

Any PA-RISC-based HP Server rp5400, rp5430, rp5450, rp5470, or HP 9000 L-Class server is field-upgradable to an HP Server rx5670. The rx5670 chassis, power, and cooling infrastructure are identical to legacy rp54xx and L-Class systems. Customers simply order the appropriate upgrade product numbers, and an HP customer engineer will arrive to swap the components. Please see the *HP Server rx5670 Ordering Guide* for more information.

Itanium 2-based hp server architecture

This section discusses the electrical architecture of the HP Servers rx2600 and rx5670. Topics covered include HP's zx1 Chipset, block diagrams, and I/O layout.

overview of the hp zx1 chipset

HP develops chipsets to meet the needs of enterprise and technical customers. In a world where every company has access to the same 64-bit processors, HP's strength is to develop and tune systems to deliver the kind of performance and reliability that IT, engineering, and research professionals demand.

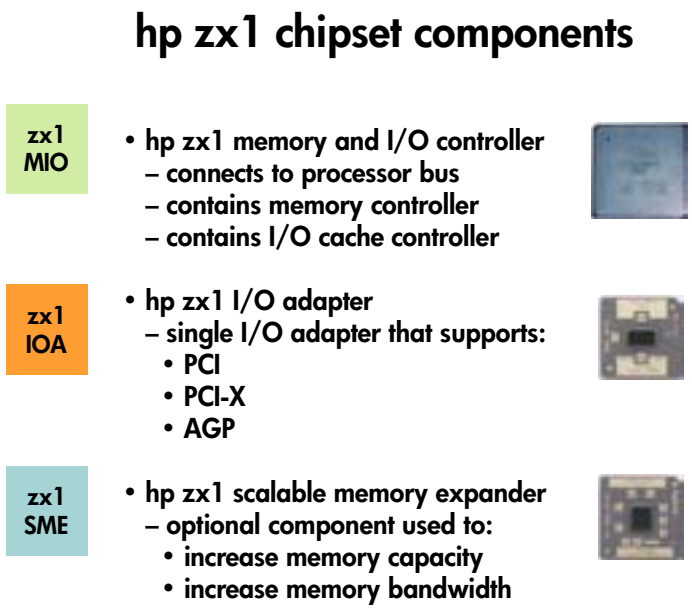
The HP zx1 Chipset is the central building block of both the HP Server rx2600 and the HP Server rx5670. The HP zx1 Chipset is a modular three-chip solution designed for cost-effective, high-bandwidth, low-latency 1- to 4-way symmetrical multiprocessing (SMP) servers and workstations. Invented entirely by HP, the zx1 is an exclusive value-add in the standards-based world of Itanium 2-based computing.

The HP zx1 Chipset consist of three modular components:

- The HP zx1 Chipset memory and I/O controller connects to the processor bus and contains dual memory controllers and the I/O cache controller. It interfaces to the Itanium 2 processor bus and provides a low-latency connection to DDR memory, either directly or through zx1 scalable memory expanders. The controller can connect up to 12 zx1 memory expanders for quadruple the base memory capacity. It can also connect up to eight zx1 I/O adapters, capable of sustaining 4.0 GB/s of I/O bandwidth.
- The HP zx1 Chipset I/O adapter chip is a scalable solution designed to support PCI-X, PCI, and AGP bus architectures. It provides a scalable I/O implementation for a wide variety of systems. The rx2600 and rx5670 do not deploy AGP graphics bus technology. AGP is available in the HP Workstations zx2000 and zx6000, which also employ the HP zx1 Chipset.
- The HP zx1 Chipset scalable memory expander is an optional component used to increase memory capacity and increase memory bandwidth. Acting as a memory hub, it decreases the number of signal loads on the memory bus, thereby allowing the system to increase its memory transfer rate.

Memory expanders are not used in the rx2600. The rx5670, however, deploys 12 memory expanders, resulting in exceptional memory capacity (48 DIMM slots) and bandwidth (12.8 GB/s).

figure 5. the hp zx1 chipset consists of three components



The HP zx1 Chipset was designed with several goals in mind:

- *provide the best performance*—for demanding applications that don't fit within the processor cache, the memory system design is the key to performance. The HP zx1 Chipset's memory bandwidth has been optimized with dual memory controllers to provide from 8.5 to 12.6 GB/s of memory bandwidth with 80 to 105 nanoseconds of open page latency.
- *provide the right functionality*—including both memory capability and PCI-X support.
  - at 12 GB and 48 GB, respectively, the rx2600 and rx5670 provide enough memory for the most demanding tasks.
  - both the rx2600 and rx5670 support 133 MHz PCI-X buses capable of handling the latest generation of high-speed I/O adapters.
- *enable a family of systems via a modular, multi-chip design*—designers can choose the chipset components they need and select the number of these components to meet system cost and design requirements. For example, the more expandable HP Server rx5670 deploys HP zx1 Chipset scalable memory expanders for greater memory capacity. The rx2600, on the other hand, does not deploy zx1 scalable memory expanders, resulting in lower system costs.

The HP zx1 Chipset is ideal for use with the Intel Itanium 2 processor because it complements the processor's price/performance advantages. Moreover, the new HP chipset leverages HP's co-developer knowledge of the CPU itself. Indeed, the HP zx1 Chipset was the turn-on vehicle for the Intel Itanium 2 processor in February 2001. At that time, Itanium 2-based systems with the HP zx1 Chipset were already running HP-UX, Linux, and Microsoft Windows.

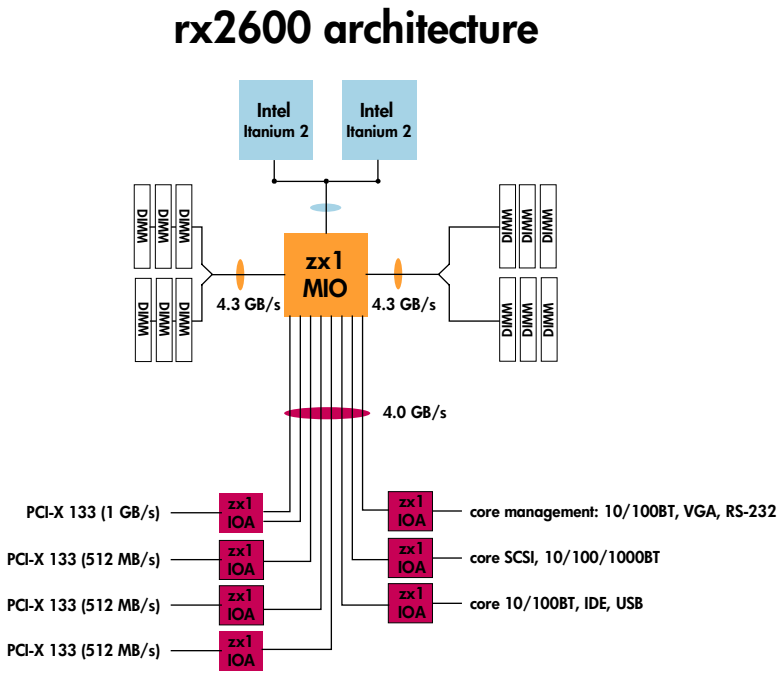
features and benefits of the hp zx1 chipset

feature	user benefit
high memory bandwidth, low memory latency	top application performance, faster time to solution
high memory capacity	optimum performance for large models/databases
133 MHz PCI-X	highest performance I/O adapters
modularity	family of Itanium-processor-based servers and workstations, each optimized for the right level of cost and scalability

architectural overview of the hp server rx2600

The HP Server rx2600 supports either one or two Intel Itanium 2 processors linked to the HP zx1 Chipset memory and I/O controller through a 200-MHz, 128-bit front-side system bus. Total bandwidth on the system bus is 6.4 GB/s.

figure 6. the hp server rx2600 architecture features Itanium 2 processors and hp's zx1 chipset



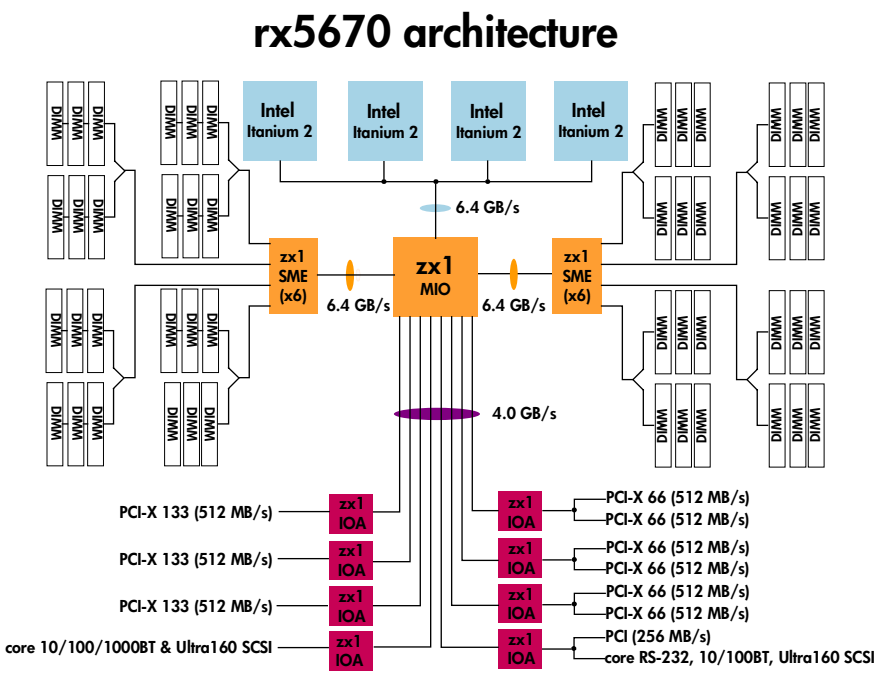
Memory DIMMs are attached directly to two 266-MHz, 4.3-GB/s memory buses. Combined memory bandwidth across both buses is 8.5 GB/s. Each bus links up to six double-data-rate (DDR) SynchDRAM memory DIMMs. Total system memory capacity is 12 GB, via twelve 1-GB DIMMs. When 2-GB DIMMs become available, total memory capacity will increase to 24 GB.

The I/O architecture consists of eight 0.5-GB/s channels allocated among seven zx1 I/O adapters. Each of these seven adapters provides a PCI-X or PCI bus to the available I/O slots and core I/O devices. The first two channels connect to a single 133-MHz PCI-X I/O slot, providing 1 GB/s of sustained throughput. This slot is ideal for high-bandwidth I/O adapters such as high-performance clustering interconnects. The next three I/O channels link to three, independent 133-MHz PCI-X I/O slots, each with 0.5 GB/s of sustained throughput. The remaining three I/O channels link to three PCI buses, which in turn link to the core LAN, SCSI, IDE, and USB interfaces, and to the management processor.

architectural overview of the hp server rx5670

The HP Server rx5670 supports one, two, three, or four Itanium 2 processors linked to the zx1 memory and I/O controller through a 200-MHz, 128-bit system bus. Total bandwidth on the system bus is 6.4 GB/s.

figure 7. the hp server rx5670 supports up to four Itanium 2 processors linked to the zx1 chipset



The zx1 memory controller links to two independent 200-MHz, 6.4-GB/s memory buses. Each bus connects to six zx1 scalable memory expanders, which in turn allocate bandwidth to the double data rate (DDR) SynchDRAM memory DIMMs. Total DIMM capacity is 48 units, distributed over two 24 DIMM memory carrier boards, with one board on each memory bus.

The server can operate with only one memory carrier board; however, this will only use one memory bus and, therefore, only half of the available memory bandwidth. For maximum performance, the rx5670 should be configured with both memory carrier boards, allowing full utilization of the rx5670's 12.8-GB/s total memory bandwidth.

The I/O architecture consists of eight 0.5-GB/s channels allocated among eight zx1 I/O adapters. Each of these eight adapters provides a PCI-X or PCI bus to the available I/O slots and core I/O devices. The first three I/O channels connect to three independent 133-MHz PCI-X I/O slots, each with 0.5 GB/s of sustained throughput. The next three I/O channels connect to three zx1 I/O adapters, each of which in turn connects to a pair of 66-MHz PCI-X I/O slots. Each slot-pair shares 0.5 GB/s of bandwidth.

The final two I/O channels connect to the core I/O. One channel provides 0.5 GB/s of bandwidth to the core 10/100/1000BT LAN as well as to one dual-channel Ultra160 SCSI controller. The other channel provides 0.5 GB/s of bandwidth to the core management LAN, RS-232 serial ports, another dual-channel Ultra160 SCSI controller, and a 33-MHz PCI slot. The 33-MHz slot is ideal for the optional VGA/USB adapter (HP Product Number A6869A). The VGA/USB card is automatically placed in this slot on systems that are factory-loaded with Microsoft Windows operating systems.

extensible firmware interface

The extensible firmware interface (EFI) is an interface between the HP-UX, Linux, and Windows operating systems and the Itanium 2-based platform firmware. The file system supported by the extensible firmware interface is based on the file allocation table (FAT) file system. EFI allows the use of FAT-32 for the system partition. (The system partition is required on a bootable disk for the Itanium 2-based platform.)

For a hard disk, the system partition is a contiguous grouping of sectors on the disk. The starting sector and size are defined by the EFI partition table residing on the second logical block of the hard disk, and/or by the master boot record (MBR), which resides on the first sector of the hard disk. The system partition can contain directories, data files, and EFI images. The EFI system firmware may search the \EFI directory of the EFI system partition, EFI volume, to find possible EFI images that can be loaded. (The HP-UX boot loader is one example of an EFI image.)

baseboard management controller

The baseboard management controller provides ease of system management. The baseboard management controller supports the industry-standard intelligent platform management Interface (IPMI) 1.0 specification. This specification describes the management features that have been built into the system; these features include diagnostics, configuration management, hardware management, and troubleshooting. The baseboard management controller interacts with the management processor to provide the highest level of system manageability and high-availability monitoring.

The baseboard management controller provides the following:

- 40-MHz ARM7TDMI RISC core, 1 MB flash ROM, 512 KB battery-backed RAM
- power and reset management
- system "health" management: fans, power supplies, temperatures, voltages
- event logging and reporting: system event log, forward progress log, diagnostic LEDs on status panel
- device inventory
- hardware and data protection: automatic clean OS shutdown on critical events, secure storage of system configuration parameters, protection of system flash ROM
- link to dedicated management processor (MP) via IPMB: enables remote management through the MP LAN or MP serial ports
- compliance with Intelligent Platform Management Interface 1.0

hp management processor

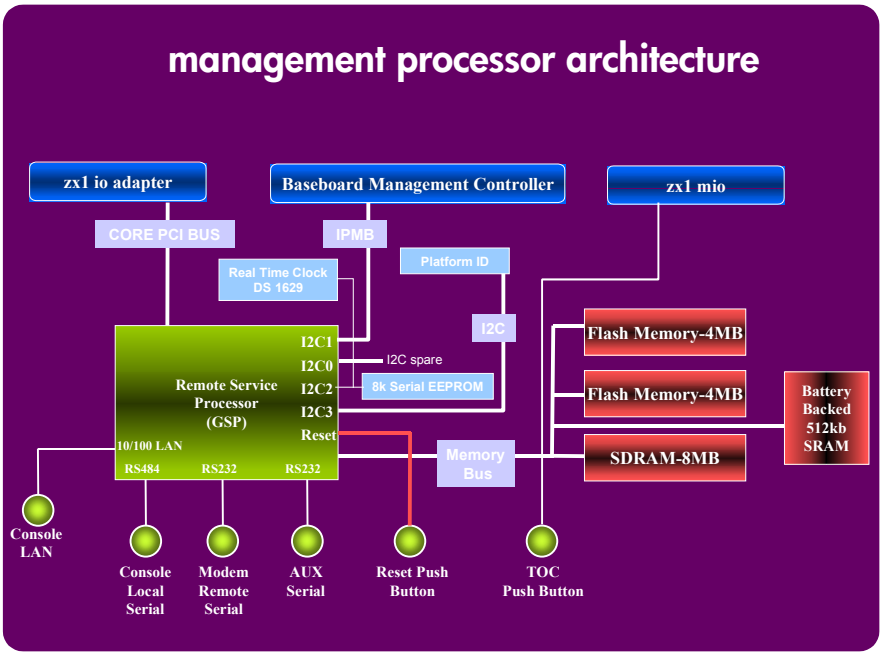
The management processor (MP) is included as a standard part of both the rx2600 and rx5670 servers. This processor provides a remote interface into the baseboard management controller to manage system resources, diagnose the health of the system, and facilitate system repair. Administrators can interact with the management processor on a dedicated, out-of-band (that is, independent of the main system data paths) communication link that can be accessed via RS-232 serial ports or a 10/100BT management LAN.

The management processor minimizes or eliminates the need for the system administrator to be physically at the system to perform tasks such as diagnostics, system management, or even hard resets. The management processor has its own battery backup, so it can be accessed even in the unlikely event that the main system power is out and the operating system has stopped functioning.

Here are some of the features enabled by the management processor:

- system management over the Internet or intranet (Web console)
- system console redirection
- console mirroring
- system configuration for automatic restart
- viewing history log of system events
- viewing history log of console activity
- setting MP inactivity timeout thresholds
- remote system control
- remote power cycle (except for MP housekeeping power)
- viewing system status
- event notification to system console, e-mail, pager, and/or HP Response Centers; e-mail and pager notification work in conjunction with HP's Event Monitoring System (EMS)
- automatic hardware protection of critical environmental problems
- access to management interface and consoles on WAN failure (modem required)
- automatic system restart
- forward progress indicator (via a virtual front panel)
- out-of-band manageability and system firmware update
- configuration of manageability and console security
- Secure Sockets Layer SSL encryption on Web console access

figure 8. high-level depiction of rx2600 and rx5670 management processor architecture



built for high availability

The HP Servers rx5670 and rx2600 have been designed to be an integral part of a mission-critical environment, delivering from 99.95 percent availability to close to 99.999 percent availability, depending on the specific solution configuration. Delivering these levels of uptime requires a strong base of single-system high availability (SSHA) in the hardware. The rx5670 and rx2600 servers have redundancy and resiliency built in from the ground up, starting with the chassis infrastructure, through the I/O, then continuing through the memory and processor subsystems.

The servers' strong SSHA is further bolstered by HP's fault event monitoring service (EMS). And for the maximum possible uptime, the rx5670 and rx2600 machines can be configured as an integral part of a high availability cluster, using clustering software such as HP's MC/Serviceguard.

high-availability chassis infrastructure (power and cooling)

Fans in the rx5670 and rx2600 provide excellent cooling, pulling cool air from the front of the unit, flowing the air back over internal system components, and then discharging heated air out the back of the server. All the fans in the rx5670 and rx2600 are easily accessible and provide N+1 redundancy.

These servers have high-availability power supplies, too. For instance, the rx5670 and rx2600 power subsystems provide high availability with N+1 redundant power options. The rx5670 comes standard with two hot-swap power supplies, and an optional third supply can be ordered for 2+1 redundancy. The rx2600 comes standard with a single hot-swap power supply; and an optional second supply gives this server 1+1 redundancy of power supplies. To further enhance availability, each power supply has its own dedicated power feed or line cord. Cords can be plugged into separate power grids for the maximum level of power protection.

hot-plug disk drives

The rx5670 supports up to four SCSI disks, while the rx2600 supports up to three. All disks are accessible from the front of the system and can be removed (or hot-plugged) while the server continues to run.

Two dual-channel SCSI controllers manage the four internal hot-plug disks in the rx5670. For added availability, disk pairs are on separate channels as well as separate SCSI controllers. This means that with disk mirroring, a SCSI controller, SCSI channel, or root disk could fail and the server would continue to run properly.

A single dual-channel SCSI controller manages the three disks on the rx2600. One channel links to two internal disks, while the second channel is connected to the third internal disk. This allows disk mirroring across separate SCSI channels, further enhancing availability.

multiple I/O channels

The multiple zx1 I/O channels in the rx5670 and rx2600 provide failover, load balancing, and failure isolation. In these servers, failures on one channel do not disrupt activities on other channels. Furthermore, both servers deploy fully independent PCI-X buses to isolate traffic on I/O adapters. If a problem occurs on one adapter, it will not interfere with traffic on another bus.

ECC and chip spare memory

The memory systems for both the HP Server rx5670 and the HP Server rx2600 utilize error correcting code to correct single-bit errors, as well as HP's chip spare technology to protect against multi-bit errors.

Chip spare enables an entire SDRAM chip on a DIMM to be bypassed in the event that a multi-bit error is detected on that SDRAM. In order to use the chip spare functionality, identical sized DIMMs must be loaded in quads. Different DIMM sizes are supported as long as they are in different quads. For example, a quad of 512-MB DIMMs can be loaded along with a second quad of 1-GB DIMMs, and chip spare will be enabled on all the DIMMs.

Because of the chip spare feature, the rx5670 and the rx2600 servers are completely resilient to all SDRAM failures, regardless of the number of bits involved in the fault condition. This virtually eliminates memory failures as a source of system errors.

Some other vendors deal with multi-bit SDRAM failures by accepting the fact that they will occur. That is, they use a scheme that supports only failure *detection*, not failure *correction*. HP believes that this is unacceptable and is a dangerous choice for servers in business-critical environments. In fact, server systems that employ failure detection but not correction are at high risk to fail due to memory problem.

CPU error correction and dynamic processor resiliency

In the HP Servers rx2600 and rx5670, L1 and L2 caches both have full single-bit error checking and correcting, as well as double-bit error detection. Additionally, all the instruction and data paths also have single-bit error checking and correcting capabilities. What's more, the system processor bus has parity detection, and the data path is covered by error correction.

The rx2600 and rx5670 employ dynamic processor resiliency (DPR), too. With DPR, any CPU generating correctable cache errors at a rate deemed unacceptable is de-allocated from use by the system. This feature helps protect against a CPU degrading to the point where it may cause system crashes.

DPR works like this: When excessive errors are reported against a CPU, the CPU is deactivated (that is, the operating system will not schedule any new processes on it). The system firmware remembers the CPU's serial number and the time when this action was taken. From then on, at each poll interval the system monitor checks (by comparing the serial numbers) to see if the CPU has been replaced or not. If the processor has been replaced, its history is reset.

If the system is rebooted before the offending CPU has been replaced, the monitor generates a warning message and immediately de-allocates the CPU. (Such CPU de-allocation is only supported in the HP-UX operating system. It is not supported in Windows or Linux.)

comprehensive error logs

All system events are stored in the system event log (SEL) in nonvolatile memory. In addition, system firmware creates activity and forward progress logs (FPLs) in nonvolatile memory. In all but the most extreme situations—that is, in more than 95 percent of cases—this information will be sufficient to diagnose system failures to a single replaceable part. The SEL and FPL are available to both the management processor (and therefore are available remotely) and to system-level tools, leading to quick and accurate diagnosis.

fault management throughout the lifecycle

Fault management is HP's overall strategy and program to provide a complete value chain for detection, notification, and repair of system problems. Fault management starts right during the design phase, when hardware and OS designers include capabilities and instrumentation points that provide the ability to detect and isolate system anomalies. Monitors are created to poll for system health information or to asynchronously respond to instrumentation points that have been designed into the system to report problems or faults.

Fault management also involves implementing several methods for maintaining historical event information, allowing preservation of information for analysis or trending. Faults that generate errors and warnings are automatically logged to syslog, while notes and audit information are copied to an event log. Other options are available for preserving historical information as well.

Fault management provides immediate alerts of problems—and even potential problems—as soon as they are detected, so customers can take corrective action. In some cases fault monitors are actually smart enough to repair faults or prevent them from occurring.

capabilities of fault monitors

Fault management, coupled with the monitoring capabilities, keeps tabs on the health of system components and generates close to real-time events when problems develop. These events can trigger corrective action to enable the system to continue functioning, or they can trigger alerts to systems personnel to appropriately handle the situation before it becomes more severe.

Fault monitors are able to:

- poll the system for health information
- handle asynchronous events that have been designed into the hardware or software
- perform corrective action when possible
- de-allocate failing memory before it fails (Dynamic Memory Resiliency)
- de-allocate failing processors before they fail (Dynamic Processor Resiliency)
- de-configure failed processors from the working set before the next reboot
- shut down the system when power failure causes a switch to UPS
- manage events so that system performance is not hindered in the face of errors
- provide information on problem causes and what actions to take

notification and integrated enterprise management

Fault management currently uses the HP EMS (Event Monitoring Service) infrastructure for its notification methodology. EMS enables a wide variety of notification methods, including pager, e-mail, SNMP traps, system console, system log, text log file TCP/UDP, and OpenView Operationscenter (OPC) messaging. Fault management events can be viewed directly on the server, or through HP's Toptools management server, which can aggregate information from multiple systems in the data center.

Customers also have the option to integrate fault management events with enterprise management software from HP (OpenView) or from BMC, Tivoli, Computer Associates, or MicroMuse.

added options with hp support

For customers who purchase HP Support, fault management events can be forwarded to the HP Support Organization. In this case, HP can take responsibility for monitoring, filtering, and trending the events and taking action on items that need attention.

At the premium end of HP's Support offerings, customers can also sign up to receive services from HP's High Availability Observatory (HAO) organization. The HAO provides continuous and proactive monitoring of the customer's environment via a dedicated and private ISDN network. The ISDN link allows secure information flow between the customer site and HP's Support Organization and provides HP support engineers direct access to the customer's system. As part of the HAO implementation, HP installs a support node at the customer site, and this node is connected securely to the HP Support Organization.

proactive, not reactive

Fault management uses the philosophy of *proactive* (as opposed to *reactive*) management of problems. Fault management provides highly accurate fault diagnosis the first time, even as the problem occurs, and initiates or allows fast corrective action. Fault management results in a substantial decrease in unplanned downtime.

expected system availability

The tables following illustrate the expected system availability of the HP Servers rx2600 and rx5670. In these tables, each rx2600 is configured with 2 processors and 4 gigabytes of memory, while each rx5670 is configured with 4 processors and 8 gigabytes of memory.

The mean time between failures (MTBF) results reflect the hardware only. Cluster availability and downtime results refer to one or more servers taken together with a large HP xp512 data storage array. The two-server cluster is a two-node MC/Serviceguard cluster, where each node is a single server.

Cluster steady-state availability and expected annual downtime results are based on expected unplanned hardware and operating system interruptions, and include the associated database recovery time, application recovery time, and end-user re-login time. Fast database recovery and clustering software are assumed to be operating.

The results are based on detailed Markov chain models which are unique to each of the various servers and system configurations that have been modeled. Both the availability and downtime results are expected (or average) predictions.

The cluster results have been developed for systems where the servers are in active/active or active/hot standby mode. For example, in the last two rows of the rx2600 table below, two 2-way/4-GB servers in a two-node MC/Serviceguard cluster with one-node hot-standby has an expected system availability of 99.9964 percent and an expected annual downtime of 0.31 hours. These times may vary depending on the application, how the database is implemented, database activity, checkpoint frequency, and other factors. Any major variance in these values will affect the availability model results.

rx2600 availability modeling results

Single server	MTBF (hours)	123,554
	MTBF (years)	14.10
Two servers, two-node active/active MC/Serviceguard cluster	Availability	99.994%
	Downtime (hours)	0.53
Two servers, two-node active/hot-standby MC/Serviceguard cluster	Availability	99.9966%
	Downtime (hours)	0.29

rx5670 availability modeling results

Single server	MTBF (hours)	52,867
	MTBF (years)	6.04
Two servers, two-node active/active MC/Serviceguard cluster	Availability	99.9933%
	Downtime (hours)	0.59
Two servers, two-node active/hot-standby MC/Serviceguard cluster	Availability	99.9964%
	Downtime (hours)	0.31

an easy transition for RISC and IA-32 users

A major feature of the Intel Itanium 2 processor and HP's new servers is *backward compatibility of the Itanium 2 processor with IA-32 and PA-RISC processors*. Optimal performance can only be achieved when an application is compiled specifically for the Itanium 2 processor. However, most existing IA-32 and PA-RISC binaries are compatible. Backward compatibility is particularly useful for applications that are not performance-intensive or for system utilities and development tools that aid in porting and migration.

running 32-bit Windows applications

32-bit Microsoft Windows binaries can run on the 64-bit Windows operating system using the WOW64 software emulation that is already part of 64-bit Microsoft Windows. The OS automatically detects whether the application is 32-bit or 64-bit and handles it accordingly. This is fine for utility applications, but performance-sensitive applications will benefit from a recompile.

running 32-bit Linux applications

32-bit Linux binaries can be run on Itanium 2-based systems using the processor's built-in hardware translation feature. The Intel Value Engine (IVE) hardware is part of the Intel Itanium 2 processor and it gives a basic IA-32 functionality. As a rule of thumb, performance will be approximately equivalent to a 350-MHz Pentium® Pro processor. If an application is performance-critical, however, recompiling the application will allow it to take advantage of the strengths of the Itanium 2 microarchitecture.

running RISC applications

Users can run their PA-RISC binaries unchanged and completely transparently, thanks to the Aries dynamic code translator that is part of HP-UX 11i v1.6. This translator allows running 32-bit and 64-bit PA-RISC applications on the Itanium 2 microarchitecture without the need to recompile. A straightforward recompile of 32-bit and 64-bit PA-RISC applications yields native Itanium 2 processor binaries.

porting and migration services: transition help from hp

Thousands of programs run quite well in Itanium processor compatibility mode with minimal changes. However, porting these applications to the Intel Itanium 2 processor allows them to run even better because they can then take full advantage of the new processor's distinctive capabilities.

To help customers make the transition to Intel Itanium processor family and get the most from this exciting new technology, HP offers a flexible set of services. Customers can select from the following:

- porting and migration workshop
- porting and migration guidance
- porting and migration detailed assessment
- porting and migration solution delivery
- online services

See http://www.hp.com/products1/itanium/services/porting_migration.html for all the details on these services.

conclusion

As the first processor based on the Intel Itanium architecture with the capability of becoming truly pervasive, the Itanium 2 processor is faster and more capable than its predecessor. Now, in its HP Servers rx2600 and the rx5670, Hewlett-Packard introduces a pair of powerful servers based on the Itanium 2 processor and featuring the new HP zx1 Chipset. The rx2600 server utilizes one or two Itanium 2 processors and up to 12 gigabytes of memory (with 24 GB expected in 2003); the rx5670 can be equipped with as many as 4 processors and as much as 48 GB of memory. These servers offer cutting-edge, 64-bit power along with excellent price/performance ratios.

The new servers are optimized for complex, floating-point-intensive computations, providing faster time to solution for demanding applications. They are especially suited to SSL Web serving, and they offer significant price/performance advantages over both IA-32- and RISC-based systems for Windows and database applications. They're also perfect for running SAP, Siebel, PeopleSoft, and SAS business application suites.

The Itanium 2 processor's ability to run IA-32 and RISC binaries without modification helps ensure protection for previous software investments, while HP's porting services can effect a complete transition that takes full advantage of the Itanium architecture. Another advantage is that customers with either of these servers can run any of three industry-leading operating systems—HP-UX, Linux, or Windows. This multi-OS capability overcomes the complexities and challenges associated with deploying and maintaining a heterogeneous operating environment.

Whether for technical computing or commercial IT, the HP Servers rx2600 and rx5670 offer superior power, scalability, and efficiency—with lower costs.

for more information

Looking for more information about the Intel Itanium processor family? Find out more about the architecture and how HP can help you make your transition by visiting:

http://www.hp.com/go/itanium

Or visit our Itanium-based servers and workstations site at:

http://www.hp.com/products1/itanium/servers_workstations/index.html

Or contact any of our worldwide sales offices or HP Channel Partners (in the U.S., call 1-800-637-7740).

HP product information and technical documentation is available online at:

http://www.hp.com/go/rx2600

http://www.hp.com/go/rx5670

appendix: technical details on the Itanium 2 microarchitecture

improved parallelism means better performance for Itanium 2 processor

The Intel Itanium 2 microarchitecture is based on the first new specification to be laid down in the Internet age. It was engineered from the beginning with the needs of commercial servers, technical computing, and e-commerce in mind.

The Intel Itanium 2 microarchitecture uses every practical technique to increase *parallelism*—the ability to execute multiple instructions during each machine cycle. Parallelism improves performance because it allows more instructions to be carried out at once. The Itanium architecture is designed to make certain the processor can execute as many instructions per cycle as possible and to ensure it also has the resources to sustain this high rate. In this regard, the Intel Itanium 2 processor is much more advanced than its predecessor—it actually has more resources available to ensure that higher execution rates are sustainable. And although all processors face penalties if an instruction is incorrectly executed or scheduled, these are minimized in the Itanium 2 processor, reducing their effect on parallelism.

A key to the high performance of Itanium chips is the design philosophy at the heart of the Itanium processor family: Explicitly Parallel Instruction Computing (EPIC). The EPIC philosophy is a big reason why Itanium processors are different from other 64-bit processors, providing much higher levels of instruction-level parallelism without unacceptable increases in hardware complexity.

How does EPIC achieve such excellent performance? By placing the burden of finding parallelism squarely on the compiler. Although processor hardware can extract a limited sort of parallelism, the best approach is to let the compiler, which can see the whole code stream, find the parallelism explicitly and make global optimizations. The compiler communicates this parallelism explicitly to the processor hardware by creating a three-instruction bundle that it issues to the hardware with directions on how the instructions should be executed. The hardware focuses almost entirely on executing the code as quickly as possible.[1]

The Intel Itanium 2 processor achieves greater performance than the original Itanium processor because its microarchitecture allows two of the three-instruction bundles to be issued for each clock cycle—a maximum of six instructions per cycle.

Greater performance is just one of the advantages of the Intel Itanium architecture and the Itanium 2 processor. We will now review some of the other important features of the Intel Itanium 2 processor.

[1] EPIC: Explicitly Parallel Instruction Computing. Schlansker & Rau, HP Labs; IEEE, February 2000.

the benefits of 64 bits

Itanium 2 processors deploy 64-bit addressing, giving them the inherent ability to address almost unlimited amounts of memory.

A few years ago, having more than 4 GB of memory in a single system was prohibitively expensive for all but customers of midrange and high-end servers. However, physical memory costs have fallen dramatically, and the density of physical memory continues to improve so that today, 4 GB of memory is reasonably priced for virtually any server user.

A 32-bit architecture is fundamentally unable to address more than 4 GB of memory. The architecture allows addressing of a maximum of 2^{32}, or 4,294,967,296 bytes. This equates to 4 GB of RAM since each byte needs a memory address. There are add-on hardware and software extensions that improve the memory scalability of 32-bit servers. However, these extensions are not part of the native architecture and, therefore, add performance-limiting overhead. The answer is 64-bit memory addressing, which allows a maximum of 2^{64} (18,446,744,073,709,551,616) bytes. This is 18 billion gigabytes of space, a limit that is unlikely to be reached anytime soon. Among other advantages, large memory space means that entire databases can be placed in memory, dramatically improving performance in data warehousing and customer relationship management (CRM) applications.

Another advantage of the 64-bit Intel Itanium 2 processor, with its 64-bit registers, is that it can work with incredibly large numbers without requiring additional operations that return results more slowly. This makes the Intel Itanium 2 processor ideal for floating-point-intensive calculations such as those used in technical and scientific computing.

the advantages of large register sets

In the Intel Itanium 2 processor, 128 floating-point registers combine both fixed and rotating registers (for software pipelining loops). This large register set allows the encoding of common algorithms without running out of register space—and without having to resort to complex tricks to move data between registers.[2]

The large register set provides a place to store intermediate results during complex calculations with significantly less latency than is involved in going back to the bus for a memory read. The registers act like a buffer to store data between the functional units and memory. Using registers in this way lessens the likelihood of stalls in the pipeline due to lack of data and ensures a higher rate of sustained parallel performance.

The rotating registers provide a tool that assists in paralleling loop execution by pipelining each loop. When combined with predication (see "predication and branch prediction" in a following section), this feature can prevent unnecessary code bloat that could hinder performance.

The 82-bit length of the floating-point register stack is also an advantage, as we will see in the next section.

[2] Scientific Computing on the Itanium Processor. Greer, Harrison, Henry, Li, Tang; Intel Corporation, SC2001 proceedings.

floating-point architecture in the Itanium 2 processor

Floating-point performance is key to many applications in scientific computing, life sciences, and some areas of business intelligence and e-commerce. The floating-point implementation in the Itanium 2 processor is a perfect example of the interplay that occurs between a number of EPIC concepts. Software and hardware features blend to make a unique floating-point architecture that has both high performance and excellent accuracy.

The 82-bit-wide register stack of 128 registers can support single, double, and double-extended IEEE-standard values. Broadly speaking, more registers provide more locations to conduct floating-point operations in parallel, while wider registers provide the necessary operating area for longer calculations. Floating-point-specific instructions like the Floating-point Multiply Add (fma) remove some intermediate rounding stages, speeding the execution of many floating-point calculations with no loss of accuracy. For example, some vector dot products and polynomials garner many benefits from the fma instruction.

For example, the Load/Pair instruction works in conjunction with the memory mapped file (MMF) instruction bundle template that encodes two memory instructions and one floating-point instruction in a single bundle to allow the construction of extremely efficient and tight inner loops. A combination of these two instructions allows the processor to serve up to four streaming floating-point operations, plus four loads and two stores per cycle. This is extremely efficient parallel floating-point performance.

The floating-point advantages of the Itanium 2 processor also work in conjunction with other benefits, such as software pipelining of loops, predication, speculation, and Single Instruction, Multiple Data streaming instructions.

speculative loads and prefetch

Removing branches from critical execution paths is important, but memory accesses are the real potential hindrance to parallelism because they have longer latencies that are often unpredictable. For example, in many cases the data required from a memory access is conditional on the outcome of a branch calculation.

Instead of having to wait for the outcome of a branch or store operation, the load can be made safely far in advance, then operated on or moved to the processor's registers in preparation for use. The Intel Itanium 2 processor has the ability to carry out such speculative loads and ensure they are implemented correctly, while minimizing the penalties if they are not. Speculative loading masks a great deal of access latency and ensures that more relevant data is available where it can be accessed quickly for better throughput.

Prefetch is another advantage of the Intel Itanium 2 processor. With prefetch, the compiler directs the processor on how and when to prefetch instructions efficiently so that they are loaded into the cache.

predication and branch prediction

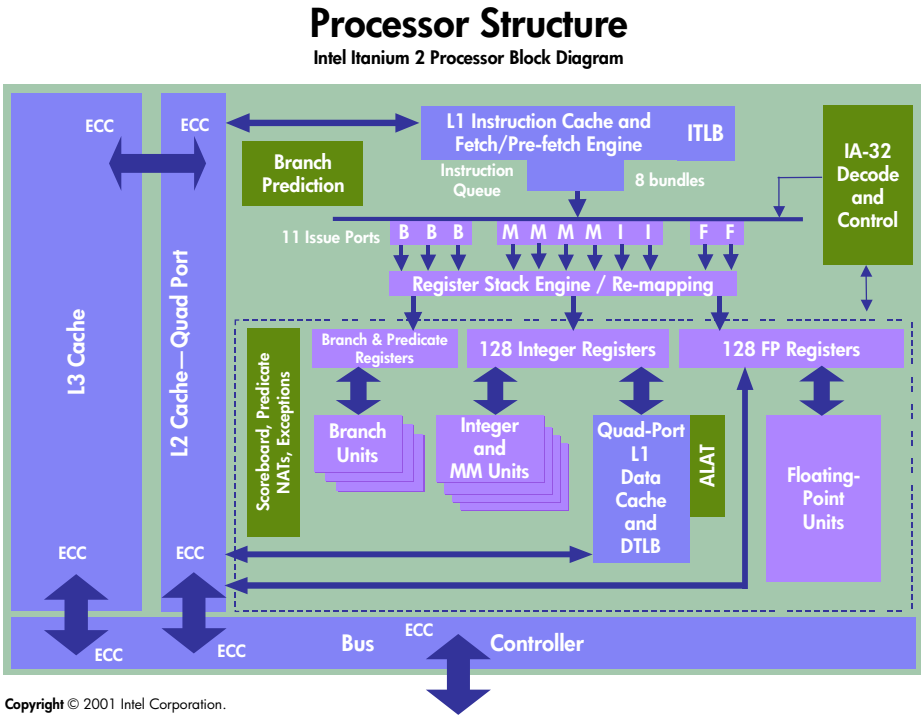
Predication provides a mechanism for smoothing out conditional branches that would otherwise interrupt instruction-level parallelism. It enables the removal of branches by taking both paths of execution and discarding the inaccurate outcome wherever possible. Predication is an important technique for maximizing parallelism.

This is far from being a comprehensive list of all the Itanium 2 processor features that enhance the performance of technical and commercial applications. But it does give an overview of some of the benefits that accrue from this remarkable new architecture.

a closer look at the Itanium 2 processor

For a closer look at the Itanium 2 processor, see the processor block diagram below.

figure 9. block diagram of the Intel Itanium 2 processor



Copyright © 2001 Intel Corporation.
Itanium is a trademark or registered trademark of Intel Corporation or its subsidiaries in the United States or other countries.

With 11 issue ports, 128 integer registers, and 128 floating-point registers, the Intel Itanium 2 processor has plenty of execution resources. The large L3 cache and the low-latency L1 and L2 caches work cohesively with these resources by staging data before it is needed.

comparing the Itanium 2 processor to other solutions

The Intel Itanium 2 processor has a number of major advantages over its IA-32 and RISC challengers. Of course, IA-32 is a fully mature technology at an attractive price, and it remains a great choice for workloads where 64-bit performance and scalability are not required, while RISC processors have been the vehicle for higher performance. And HP offers a full line of IA-32 and RISC servers.

However, the Itanium 2 processor has far higher floating-point performance and is more widely addressable than any IA-32 implementation. Moreover, the Intel Itanium architecture is still in the beginning of its lifecycle, and its performance is scaling much more sharply than that of IA-32 or RISC. The already excellent price/performance ratio of Itanium processors will only continue to improve.

Price/performance is where Itanium 2-based servers clearly beat RISC servers. The combination of a superlative microarchitecture, strong compilers, and industry-standard parts, along with a commodity processor manufacturing process for Itanium 2-based servers, means that low-volume, proprietary RISC servers are at the beginning of their decline.

The fundamental differences between the Intel Itanium 2 microarchitecture and that of two different dual-CPU microarchitectures can be compared using the following table.

comparing the Intel Itanium 2 processor with the Sun UltraSPARC III and an IA-32 implementation

feature	Intel Itanium 2 processor	Sun UltraSPARC III	IA-32 implementation
word size	64-bit	64-bit	32-bit
system bus bandwidth	6.4 GB/s	4.8 GB/s	2.1 GB/s
on-die cache	L1 = 16 KB + 16 KB L2 = 256 KB L3 = 1.5 MB or 3 MB	L1 = 96 KB* * 8 MB L2 external cache	L1 = 64 KB + 64 KB L2 = 256 KB
issue ports	11	4	9
on-die registers	328	96	26
execution units	6 integer, 3 branch, 2 FP, 1 SIMD, 2 load, 2 store	2 integer, 2 FP/VIS, 1 branch, 1 load/store	3 address generation, 3 integer, 3 x87 FP
pipeline stages	8-stage pipeline	14-stage pipeline	10-stage pipeline* * 15-stage FP
core frequency	900 MHz/1 GHz	900 MHz	1.4 GHz
instructions/clock	6 instructions/cycle	4 instructions/cycle	3 instructions/cycle

Comparing these three architectures is instructive. It is immediately clear that both the Sun and the IA-32 processor are bottlenecked by system buses that restrict how much data can flow to the processor each cycle. It is also clear that the Intel Itanium 2 processor has far more execution resources.

The address units on an IA-32 system manage the loads/stores from and to the caches. The three address generation units do nothing but perform address calculations for the IA-32's load store unit and therefore do not directly influence instruction throughput.

The IA-32 architecture is limited in two major ways: it is limited to a 32-bit address space, and it is effectively register-starved. Furthermore, it has only limited superscalar parallelism—it will break down the three instructions into a potential of nine micro operations, but it is still effectively only working on three x86 instructions at a time. Note also that the shorter eight-stage pipeline of the Intel Itanium 2 processor means that pipeline errors such as stalls have the smallest penalty on this architecture.

comparing the Itanium 2 processor to its Itanium processor predecessor

The Intel Itanium 2 processor is, as its name suggests, the second processor based on the Intel Itanium architecture. It has a number of microarchitectural advantages over its predecessor that ensure it is able to maintain a higher sustained rate of instruction issue per cycle, which means it can offer better parallelism than its predecessor or any other processor available today.

The Itanium 2 processor is manufactured on the 0.18 μm process. Its eight-stage basic execution pipeline is 20 percent shorter than the 10-stage pipeline of its Itanium processor predecessor, meaning there are smaller penalties from stalls in the pipeline, and there is a higher sustainable rate of performance. The Intel Itanium 2 clocks at 1 GHz.

execution resources

The Intel Itanium 2 processor's additional execution resources over those of its predecessor are as follows:

- 2 additional integer units
- 2 additional multimedia units
- 2 additional load/store memory ports

With enough issue control and data paths to sustain a wider combination of instruction in each instruction bundle, the Intel Itanium 2 processor is more likely to achieve its peak rate of six instructions per cycle and maintain a higher sustained rate. It has a much better chance of filling most, if not all, instruction issue slots each cycle.

cache system

The cache latencies of the Intel Itanium 2 processor are approximately half those of its predecessor. This means that the right data is more likely to be available when the processor needs it so that high throughput of instructions can be maintained.

The Intel Itanium 2 processor's cache system looks like this:

- L1 instruction cache: 4-way 16 KB with a 1-cycle latency
- L1 data cache: 4-way 16 KB with a 1-cycle latency
- L2 instruction + data cache: 8-way 256 KB with a 5- to 6-cycle latency
- L3 instruction + data cache: 12-way, 1.5 MB or 3 MB depending on the processor, with a 12-cycle latency

Through advanced memory design techniques, the designers of the Intel Itanium 2 processor were able to fit 3 MB of L3 cache on the processor, the largest on-chip cache in the industry. This provides much lower latency compared with the 4 MB of off-chip cache on the Intel Itanium processor. This large L3 cache complements some of the other design techniques examined earlier, including speculative loads and prefetching instructions.

front-side bus

The processor's system bus, or "front-side" bus, has a 3X improvement in data bandwidth over the first Intel Itanium processor. The data bus bandwidth has been increased from 2.1 GB/s to 6.4 GB/s, the data bus word size has been increased from 64 bits to 128 bits, and the frequency of the bus clock has been increased from 133 MHz to 200 MHz. This frequency is double-pumped, for a total of 400 mega-transfers per second of system bus data.

failover capabilities

The failure of a processor is possible but uncommon. The Intel Itanium 2 processor has highly available failover capabilities built in. Any L1 data or instruction cache error is correctable by the processor. The processor will also attempt to correct any random single-bit logic failure in the L2 and L3 caches or in data cycles on the bus. If the error is uncorrectable, the relevant data thread is poisoned, and, if necessary, the firmware effects a recovery while minimizing application damage.

Pentium is a U.S. registered trademark of Intel Corporation. Intel Itanium Processor Family is a trademark of Intel Corporation in the U.S. and other countries and is used under license. Microsoft and Windows NT are U.S. registered trademarks of Microsoft Corporation. UNIX is a registered trademark of The Open Group.

Technical information contained in this document is subject to change without notice.

© Copyright Hewlett-Packard Company 2002

07/02

5981-2240EN