FUJITSU

White paper
PRIMEQUEST | THE BEST COST-EFFICIENCY IN MISSION CRITICAL OPERATION

Business continuity and cost-efficiency have become essential demands on IT platforms. Offering the best-blend of standard and high availability technologies, PRIMEQUEST is an open enterprise system platform that fully maximizes uptime and greatly lowers TCO.
This whitepaper explains the the features of the PRIMEQUEST that make it the best choice for mission critical operations, and demonstrates how the PRIMEQUEST total cost of ownership is significantly lower than the HP Superdome 2.

Content

Introduction

PRIMEQUEST is a mission-critical server that supports up to eight Intel® Xeon® CPU chips and maximum 80 cores. By combining the cost efficiency of x86 servers and high availability of UNIX server, customers can build their solid business platform and achieve a high return on investment with PRIMEQUEST.
The intention of this whitepaper is to convince the reader that PRIMEQUEST can help you reduce operational costs without sacrificing the business demands for high availability and performance scalability. First, PRIMEQUEST can minimize planned and unplanned downtime because almost all the components are fully redundant and hot replaceable. Plus, the heart of the server – the CPUs and memory – is protected from failure by multi-level data protection mechanisms. Second, PRIMEQUEST has very good performance scalability up to a maximum 80 cores and 160 threads. Third, PRIMEQUEST can also help you reduce your server operation and administration costs with a range of measures from Green IT to well-developed IT system support. PRIMEQUEST is unique breed in mission critical servers in its combination of openness, high-quality engineering, scalability and high availability delivers best-in-class cost efficiency and operational stability, comparable to high-end UNIX servers. This whitepaper compares HP Superdome 2 with PRIMEQUEST looking at operational costs, Oracle DB support charges, and the costs of ensuring high availability.

Matched high availability in mission critical systems

Maximizing business uptime is an axiom of mission critical server. However, to sustain server operation, even during a system failure, all components must be redundant. Especially the essential parts of the server such as CPU, memory, and system bus, must be especially assured as a failure of one of those components has the capability to cause an entire system shutdown. In addition maintenance operations such as component replacement, patch application, and testing, must be able to be executed while business applications continue to run, without interruption, or with the very minimum of downtime.

In-built high availability can slash mission critical server costs

Building mission critical systems could sacrifice cost efficiency. This is valid if the server being used lacks overall reliability. While clustering is a possible and practical solution with such servers, the need to double or triple the server count will bring extra costs – worse there are hidden costs that are easily overlooked.

In a multi-server cluster, maintenance costs will be more than doubled because administrators must apply the same patches to all servers in the cluster. They must also switch the cluster nodes before and after maintenance.

License and support charges for some software may also double.

Even clustered systems are not totally safe from server failure. So, business losses by such downtime must be contemplated. In a High Availability (HA) cluster, for example, if failed nodes are switched to a stand-by node on failure, the cluster switching process requires a number of minutes to restart applications. If cluster parallelism such as Oracle Real Application Clusters (RAC) is used, performance deterioration will be unavoidable. For instance here, with a dual node Oracle RAC system, the performance would be halved during the time the failed server was offline.

PRIMEQUEST on the other hand embeds the equivalent of high-end UNIX server availability in every unit; this means the operational costs related to high availability are as low as those of UNIX servers.

CPU protection

Xeon 7500 processor series and E7 processor family are designed to handle recoverable and unrecoverable errors.

- Recoverable errors
  Both data and tag fields in cache levels 1/2/3 can detect and correct bit errors. The data protection features of level 3 cache are described below.
  - Data array
    Up to three-bit errors can be detected and retried. Up to two-bit errors can be corrected.
  - Tag array, core valid array, and LRU (Least Recently Used)
    Up to two-bit errors can be detected and retried. One-bit errors can be corrected.
  - Registers, ALUs (Arithmetical and Logical Units), and TLBs (Translation Look-aside Buffer)
    One-bit errors are handled by each processor's circuits. They can detect and correct such errors.
- Unrecoverable errors
  If the above retry operations are successful, the application and operating system are not notified of the error. Only if the recovery is unsuccessful the application is stopped.

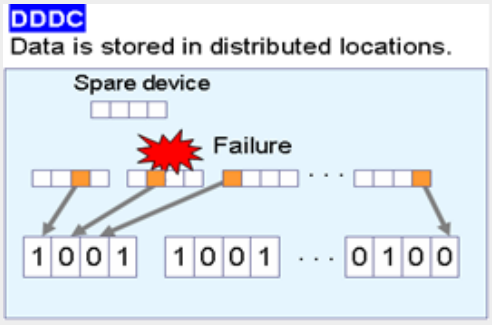
Memory protection

Memory chips and their interfaces to CPUs also have to be protected from errors. This is because memory is one of the most error-prone parts of the server and memory failures have the ability to cause an entire server stoppage.

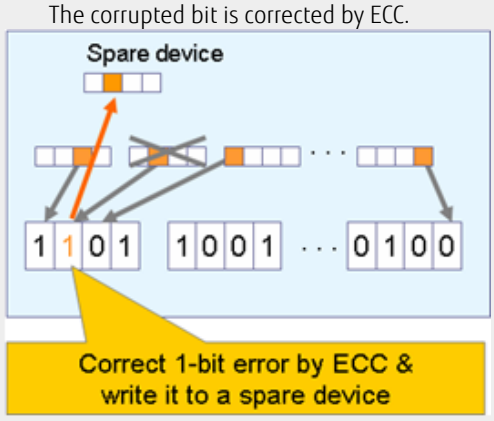
- Multi-bit error recovery
  Even with an error occurring in a DRAM module, the application can continue operating while the error is corrected.
  In DRAM 4-bit or 8-bit data chunks are typically assigned an additional DRAM bit. ECC (Error Check and Correct) uses this information to correct read errors so that CPU memory access can continue when an inconsistency is found.
  PRIMEQUEST 1800E with Xeon 7500 processor series is able to recover single DRAM failures using Single Device Data Correction (SDDC). PRIMEQUEST 1800E2 with Xeon E7 product family processors is able to recover dual DRAM failures using Dual Data Device Correction (DDDC) (Figure 1).
- Memory Mirroring
  Memory Mirroring is a memory redundancy function that allows each CPU to write to and read from a memory pair. This means CPU-memory access can continue even if a whole DRAM module fails, as the other available DRAM module still contains the correct data.
  Memory Mirroring is capable of handling concurrent failures such as a memory chip and memory buffer failure. Both PRIMEQUEST 1800E and 1800E2 successfully continue memory access even when a memory chip or memory buffer failure occurs.
  While some UNIX vendors argue their servers support memory mirroring, their memory buffers are not always a multiplex. This means their UNIX servers despite appearing to have memory mirroring may fail on a memory buffer failure.
- Guaranteed read/write operations
  PRIMEQUEST 1800E and 1800E2 detect and correct one-bit errors, detect two-bit errors and then performs retry operations using ECC. If an error occurs on one SMI (Scalable Memory Interconnect) lane, which is an interface between processor and memory, memory access is able to continue using a spare lane.
- Memory Scrubbing
  Memory Scrubbing detects a malfunctioning memory chip before it is used. This is designed to ensure early detection and correction of memory errors using ECC. This includes Demand Scrubbing error checking at memory read time, and periodic error checking by Patrol Scrubbing.

Figure 1. Memory error correction by DDDC

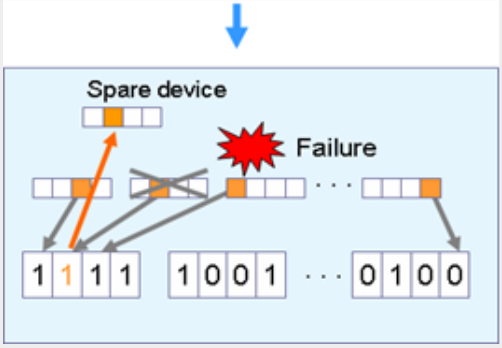
(1) First DRAM failure
Due to one DRAM failure, one-bit is corrupted.



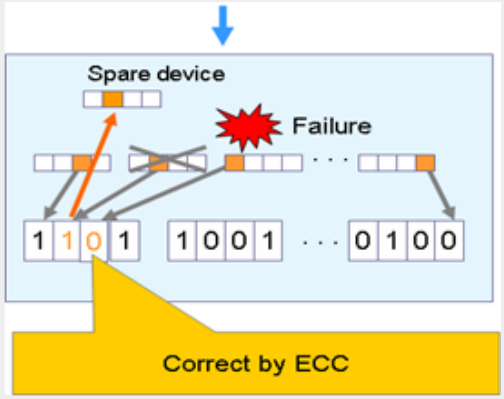
(2) Recovery of first DRAM failure
The corrupted bit is corrected by ECC.



(3) Second DRAM failure
Due to another DRAM failure, one-bit is corrupted.



(4) Recovery of second DRAM failure
The corrupted bit is corrected by ECC.

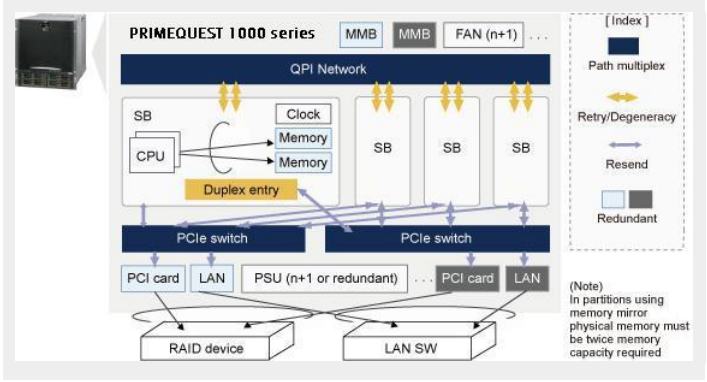


Component Redundancy

The figure below shows that almost every component is redundant or can be used in multiplex configuration.

- Redundant components
  Memory, PCI cards, standard LAN ports, fans, HDDs, service processors (Management Board, called MMB)
- Path multiplex
  Interconnections between System boards and PCI switches, CPUs and other System board components.
  In addition to CPU protection system boards can be protected from failure by use of a Reserved System Board. This can be used for automatic replacement of the original board on failure.

Figure 2. PRIMEQUEST component diagram



Hot Replacement

All main components are hot-replaceable.

- Power supplies, fans, disk drives, PCI cards, service processors, and DVD drive

Comparison with HP Superdome 2

Table 1 below compares HP Superdome 2 with PRIMEQUEST in terms of redundancy and hot replacement.

- Redundancy
  - PRIMEQUEST memory can be made redundant. But, HP Superdome 2 has no equivalent redundancy.
  - PRIMEQUEST and HP Superdome 2 have the same level of redundancy for system interconnect and other components
- Hot replacement
  - PRIMEQUEST standard LAN ports, PCI cards, and a DVD drive can be replaced during operation, but no corresponding statement can be found regarding HP Superdome 2

Table 1. Redundancy and hot replacement comparison of PRIMEQUEST with HP Superdome 2

	PRIMEQUEST 1800E/1800E2	HP Superdome 2*1
Redundancy		
System board (including CPU and memory)	Memory can be redundant by memory mirroring	No redundant components
System interconnect	Multiplex by Intel Quick Path Interconnect	Multiplex by Intel Quick Path Interconnect
Other components	Components below are redundant: - PCI cards - Standard LAN ports - FAN - PSU - Service processor (MMB)	Components below are redundant: - PCI cards - FAN - PSU - Service processor
Hot replacement		
System board (including CPU and memory)	Cannot be hot replaced (system board part of a running partition) Can be hot replaced (system board not part of a running partition)	Cannot be replaced during operation
System interconnect	Cannot be replaced during operation	Cannot be replaced during operation
Other components	Components below can be replaced during operation: - PCI cards - Standard LAN ports - FAN - PSU - Service processor (MMB) - DVD drive	Components below can be replaced during operation: - FAN - PSU - Service processor (MMB) - clock

*1 For HP Superdome 2 specification, see documents below.
http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA1-1678ENW.pdf

Table 2 looks at processor and memory protection. As explained previously in this section, PRIMEQUEST 1800E and 1800E2 have multi-level protection for processor and memory. On the other hand, HP Superdome 2 does not support memory redundancy. For other memory and processor protection mechanisms, PRIMEQUEST and HP Superdome 2 have comparable functions.

Table 2. Processor and memory protection comparison of PRIMEQUEST and HP Superdome 2

	PRIMEQUEST 1800E/1800E2	HP Superdome 2*1
Processor		
Cache protection	Error detection and correction mechanisms for data and tag arrays	Error detection and correction mechanisms for data and tag arrays
Error check circuits	Register, ALU, and TLB equip with error detection and correction mechanisms	Soft error hardened latches embedded
Memory		
Multi-bit error recovery	Dual Data Device Correction supported*2	Not supported
Memory mirroring	Supported	Not supported
Guaranteed memory read/write operation	Multiplex CPU-Memory lanes	Multiplex CPU-Memory lanes
Memory scrubbing	Supported	Supported (presumption)

*1 Documents below were used for this comparison.
HP Superdome 2 QuickSpec :
http://h18004.www1.hp.com/products/quickspecs/13682_div/13682_div.HTML
HP Superdome 2 datasheet :
http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA1-1678ENW.pdf
*2 This function is available for PRIMEQUEST 1800E2. PRIMEQUEST 1800E with Xeon 7500 series processors only supports Single Data Device Correction

Operational cost comparison with UNIX server

In this section we discover the excellent cost-efficiency of PRIMEQUEST in mission critical operation. Further detailed analysis will be explained in the following sections.

Operational cost comparison with HP Superdome 2

PRIMEQUEST 1800E2 has much lower operational costs than HP Superdome 2. The operational cost per year for a PRIMEQUEST 1800E2 with Windows server is more than $50,000 lower than that of HP Superdome 2. Configured as a Linux server, PRIMEQUEST 1800E2 has operational cost around $43,000 lower than that of HP Superdome 2.

But when it comes to Oracle database support charges, those charges for PRIMEQUEST 1800E2 are just half those of HP Superdome 2.

Assumptions in making these comparisons are as follows:

(1) Operational costs compared : administrative costs, server power consumption, and air conditioning costs

(2) PRIMEQUEST 1800E2 and HP Superdome 2 with the same numbers of cores have comparable performance.
PRIMEQUEST 1800E2 : Intel Xeon E7-8830 (8 cores) 8 CPU chips
HP Superdome 2 : Intel Itanium 9350 4c (4 cores) 16 CPU chips

(3) PRIMEQUEST 1800E2 and HP Superdome 2 have the same level of high availability.

(4) 24 hours, 365 days a year operation

(5) Operating in the US

Figure 3 and Table 3 show operational costs per year excluding Oracle database support fees.

- Operational costs for PRIMEQUEST 1800E2 Windows server is $20,000 lower than that of a Linux server PRIMEQUEST 1800E2 configuration.
- Operational costs for PRIMEQUEST 1800E2 Windows server is $51,474 lower than that of HP Superdome 2.
- Operational costs for PRIMEQUEST 1800E2 Linux server is $31,474 lower than that of HP Superdome 2.

Figure 3. TCO comparison between PRIMEQUEST and HP Superdome 2

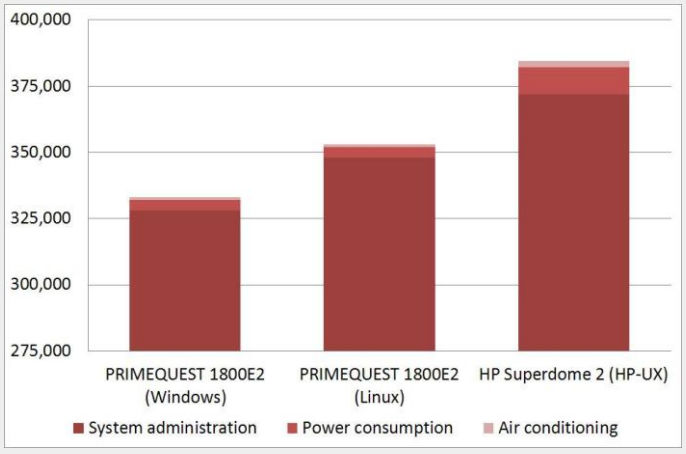


Table 3. TCO comparison between PRIMEQUEST and HP Superdome 2

US$

	PRIMEQUEST 1800E2 (Windows)	PRIMEQUEST 1800E2 (Linux)	HP Superdome 2 (HP-UX)
System administration	328,000	348,000	372,000
Power consumption	4,058	4,058	10,067
Air conditioning	989	989	2,454
Total	333,047	353,047	384,521

- Oracle Database Enterprise Edition 11g support fee
  Figure 4 and Table 4 show Oracle support fees.
  PRIMEQUEST 1800E2 (64 cores): $304,000/year
  HP Superdome 2 (64 cores) : $608,000 /year

Figure 4. Oracle databa0se support fees (one year) comparison

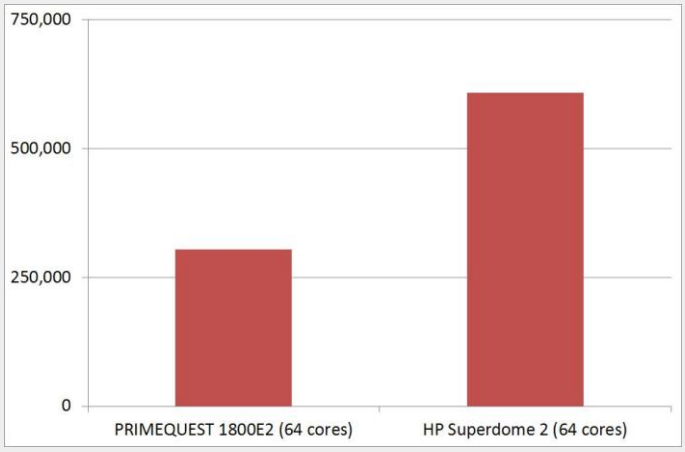


Table 4. Oracle database support fee

US$

	PRIMEQUEST 1800E2 (64 cores)	HP Superdome 2 (64 cores)
Support charge for five years *1	1,520,000	3,040,000
Charge for one year	304,000	608,000
Formula	64 (number of cores) x $47,500 (fee per Oracle processor license) x 0.5 (Oracle multiplier)	64 (number of cores) x $47,500 (fee per Oracle processor license) x 1.0 (Oracle multiplier)

*1 Relevant service available for five years
http://www.oracle.com/us/corporate/pricing/technology-price-list-070617.pdf

System administration cost comparison

This section explains why the selection of a standard OS such as Windows or Linux can slash administrative costs due to their wider familiarity. In this section, system administration costs are defined as labor costs for administrative work such as server monitoring for problem detection, OS configuration changes, and OS patch application.
In this comparison the following is assumed:

- PRIMEQUEST and HP Superdome 2 will require the same administrative effort. The equivalent of four full time administrators are required to provide 365 days a year and 24 hours a day operation for either type of server.
- PRIMEQUEST has web interface comparable to HP Superdome 2 Online Administrator. PRIMEQUEST also has an intuitive administration interface called ServerView Operation Manager.

Figure 5 and Table 5 show 24x7 operational administrator costs per year.

- Costs for Windows server is $44,000 lower than that for HP-UX
- Costs for Linux server is $24,000 lower than that for HP-UX

Figure 5. Server administration TCO comparison

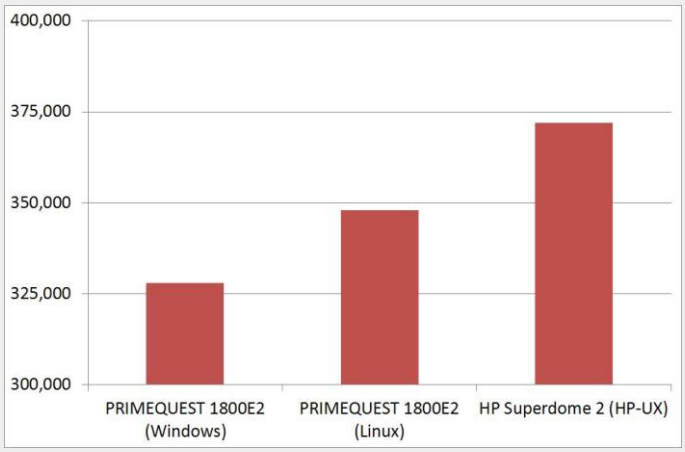


Table 5. Server administration TCO comparison

	PRIMEQUEST 1800E2 (Windows)	PRIMEQUEST 1800E2 (Linux)	HP Superdome 2 (HP-UX)
Cost per administrator per year *1	82,000	87,000	93,000
Annual cost for four persons for 24-hour operation	328,000	348,000	372,000

*1 These figures refer to job search site, www.indeed.com
http://www.indeed.com/salary?q1=Windows+Server+Administrator&l1=
http://www.indeed.com/salary?q1=Linux+Server+Administrator&l1=
http://www.indeed.com/salary?q1=HP-UX+Server+Administrator&l1=

IT knowledge shift to Windows and Linux

Nobody is likely to argue against the tendency that server engineers are choosing Windows and Linux. While such tendency seems obvious, it is also worthwhile to confirm the growing dominance of these OS products and how UNIX demographics are changing. According to the National Salary Tendency statistics from www.indeed.com, HP-UX server administrator salaries in the US tend to be higher than those for Windows and Linux in 2009 and 2010 (Figure 6). It is therefore realistic to conclude that administrative costs for Windows and Linux servers will remain at a lower level than HP-UX in the coming years. On the transition to a Windows and Linux server administrator population, Figure 7 shows the percentage job growth in the US market from January 2005. These graphs also refer to www.indeed.com.

- Windows and Linux administrator job percentages grew until the end of January 2008, and then decreased during 2009. They started to grow again during 2010.
- While HP-UX job percentages grew from 2005 to 2008, with ups and downs during the period. But from the beginning of 2009, the trend has been towards decline.

Figure 6. Salary transition for server administration in the US

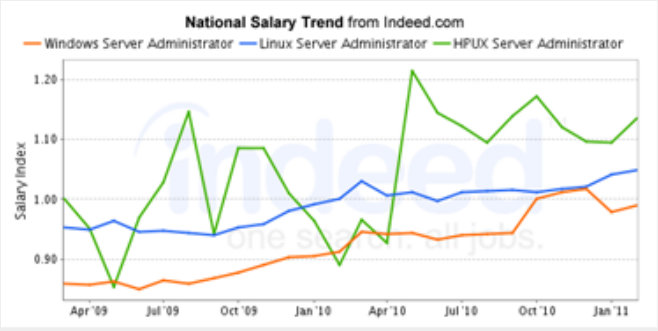


Figure 7. Server administration job percentage transition in the US for Windows/Linux/HP-UX

(1) Windows server job trend

(2) Linux server job trend

(3) HP-UX server job trend

There is an outstanding difference in the programmer job market. Job demands for Linux programmers grew by more than 50% from 2005 to 2010. On the other hand, job demand for HP-UX programmers declined by 75% from 2005 to 2010 (Figure 8).
The graphs below also refer to www.indeed.com

- Linux programmer job percentage grew steadily from 2005 to 2010.
- Windows programmer job percentages are flat overall during 2005 to 2007 while up and down on an annual basis, there was a decrease from 2008 until the end of 2009, with growth again in 2010.
- HP-UX programmer job percentages sharply decreased during 2005. Then in 2009, it decreased again.

As the programming population declines, companies using HP Superdome 2 severs will likely have to pay higher costs for application development. Worse still, application enhancement may become major challenge with new application development becoming almost impossible in the future.

Figure 8. Programmer server administration job percentage transition in the US

(1) Windows server job trend

(2) Linux Server Job trend

(3) HP-UX Server Job trend

Green IT cost comparison

If you are managing a datacenter, slimming down server costs would be an urgent issue. Choosing unmatched servers may require facility reconstruction for additional power supply as well and additional floor space.

But a cost-efficient server can reduce the risk of datacenter reconstruction. PRIMEQUEST consumes about 40% of the power used by HP Superdome 2. (Figure 9.) The comparison is based on maximum power consumption. But server power consumption and cooling costs are included in this comparison.

The maximum power consumption used in this comparison is because HP only discloses the maximum value for HP Superdome 2. While this maximum value represents the peak power consumption for the maximum configuration, such a large value is likely larger than you will observe in your datacenter.

Figure 9. Server related power cost comparison

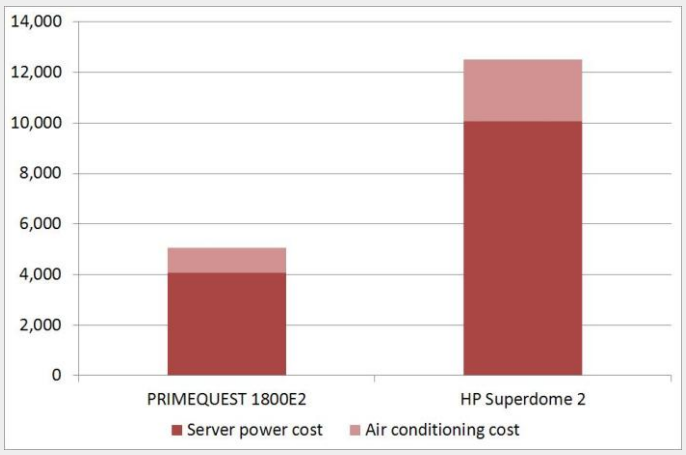


Table 6. Server related power cost comparison

US$

	PRIMEQUEST 1800E2	HP Superdome 2
Total power cost	5,047	12,522
Server power cost *1	4,058	10,067
Air conditioning cost	989	2,455

*1 for calculation details, please see Tables 7 and 8.

Table 7. Power consumption cost calculation

	PRIMEQUEST 1800E2	HP Superdome 2 *1	Formula
Max. Power consumption (kW) [A]	4.0	9.924	
Annual Power consumption (kWh)	35,040	86,934.24	=A x 8,760 (365 days x 24 hours)
Power consumption cost (US$)	4,058	10.067	=B x 11.58 *2 (cents/kWh)

*1 One HP Superdome 2 with 64 cores and two IO Expansion (IOX) cabinets.

This configuration is selected because Superdome 2 does not have PCI slots inside its cabinet – for making 16 PCI Express slots available, two IOX must be mounted. PRIMEQUEST 1800E2 has 16 PCI Express slots in the cabinet.

Respective power consumption is as follows.

- 8.82 kW for Superdome 2 cabinet
- 0.525KW for IOX

For the detail, see HP Superdome 2 QuickSpec:
http://h18004.www1.hp.com/products/quickspecs/13682_div/13682_div.HTML

*2 Cost per kWatt : 11.58 cents/kWh according to US Energy Information Administration
http://www.eia.doe.gov/emeu/steo/pub/contents.html

Table 8. Air conditioning cost comparison

	PRIMEQUEST 1800E2	HP Superdome 2	Formula
Server power consumption (kWh) [A]	35,040	86,934.24	
Server power consumption (BTU) [B]	119,626,560	296,793,495	=A x (3,414BTU/kWh) *1
Air conditioning power consumption (Wh) [C]	8,544,754	21,199,535	=B/14 *2
Air conditioning power consumption cost (US$)	989	2,455	=C/1000 x 11.58 (Cents/kWh)/100 *3

*1 Conversion ratio from kWh to BTU is 3,414

*2 Seasonal Energy Efficiency Ratio (SEER): 14 is set as ENERGY STAR qualified Central Air Conditioners must have a SEER of least 14.

*3 Cost per kWatt : 11.58 cents/kWh according to US Energy Information Administration
http://www.eia.doe.gov/emeu/steo/pub/contents.html

Green technology inside PRIMEQUEST

PRIMEQUEST's energy efficiency is sustained by Fujitsu and Intel technologies.

- Intel Intelligent Power Technologies
  Intel's Intelligent Power Technology reduces power supply inline with CPU/core utilization and memory use.
  - The Integrated Power Gate can reduce power to idle cores to near-zero
  - Low power state automation means processors and memory can be placed in their lowest available power state
- Fujitsu smart power management technologies
  - Multi-step fan speed (Figure 10)
    Fan rotation changes smoothly matched to temperature changes. Efficient cooling means cool air is concentrated on hot spots like CPUs and memory. Heat radiation boards called heat sink are used with high heat generating components parts like CPUs. The large surface area ensures more efficient airflow use.
  - System Board cooling (Figure 11)
    System Board design to concentrate cool air onto hot spots like CPUs and memory.
  - Power Supply Unit (PSU) Efficiency (Figure 12)
    PRIMEQUEST PSU are highly energy efficient when converting AC-DC and DC-DC current.

Figure 10. PRIMEQUEST multi-step fan speed

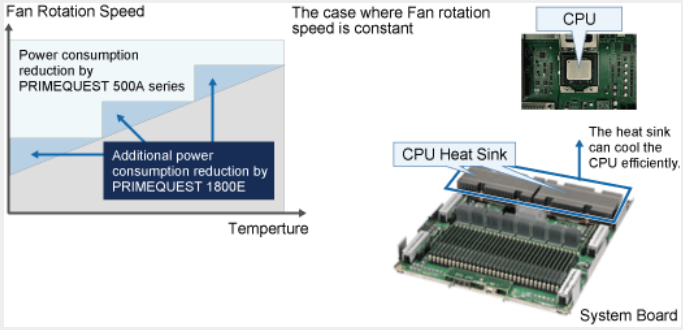


Figure 11. PRIMEQUEST system board cooling

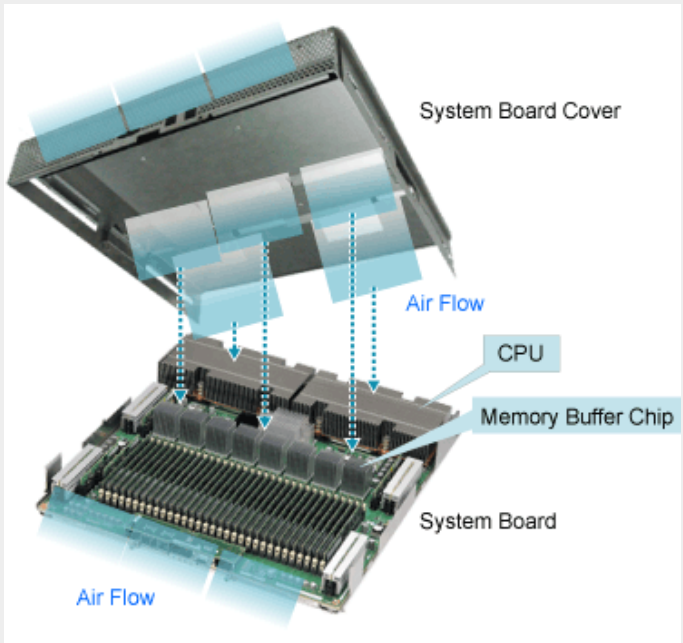
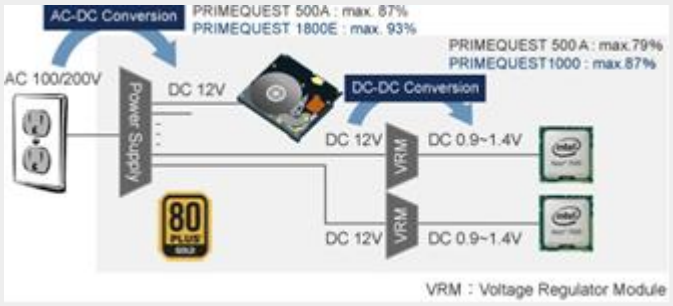


Figure 12. PRIMEQUEST power supply efficiency



Conclusion

In every aspect of server TCO, whether it is administrative costs, power consumption costs, or Oracle support licenses, PRIMEQUEST is more cost-efficient than the equivalent HP Superdome 2. Equally important is the high availability compared to the UNIX server, as a result of all the redundancy and data protection mechanisms embedded at the heart of PRIMEQUEST. This makes PRIMEQUEST the right choice for mission critical server customers looking to increase their Return on Investment (ROI).
Some may argue that the cost-efficiency shown in this whitepaper may be valid in the US, but may not be valid elsewhere. Such are the differences in labor and power costs between developed and developing countries.
But while such costs do vary country by country, overall the cost of using PRIMEQUEST will always be lower than HP Superdome 2.

- PRIMEQUEST power consumption is just 40% that of HP Superdome 2. So, power costs will always be much lower with PRIMEQUEST.
- A sensible assumption is that Windows and Linux administrator populations are tending to grow in most countries and this tendency will push down corresponding labor costs. On the other hand, we can see no factor pushing down labor costs related to HP-UX.
- Oracle Database support charges for PRIMEQUEST is half that of HP Superdome 2.
  - As for support charges per core, the ratio of Xeon and Itanium 9300 is 1 to 2. This ratio is common across all countries.

The obvious conclusion therefore is that PRIMEQUEST will be able to offer outstanding cost-efficiency and high availability to customers worldwide.

Contact
FUJITSU LIMITED
Website: www.fujitsu.com
2012-05-08-WW-EN