

# Description Paper

We make sure



## **PRIMEQUEST Usage Patterns Usage template database consolidation**

Issue October 2007

Pages 7

### Contents

<b>Introduction</b>	<b>2</b>
<b>Definition and Target Market</b>	<b>2</b>
<b>Description</b>	<b>2</b>
Customer Problem/Challenges	2
PRIMEQUEST Business Value	2
Critical Server Attributes	3
<b>Solution Description and Requirements</b>	<b>4</b>
<b>Competitive Analysis</b>	<b>5</b>
Performance	5
Reliability	5
Flexibility	6
<b>Summary</b>	<b>6</b>
<b>Appendix A: Detailed Summary of PRIMEQUEST Attributes</b>	<b>7</b>

## Introduction

This paper describes the features of the Fujitsu Siemens Computers PRIMEQUEST family of servers and how they combine to provide an extremely powerful, reliable, and flexible platform for database consolidation.

It will be shown that the features of PRIMEQUEST servers, coupled with the Linux or Windows operating system, provide an easy to manage server environment offering robust performance and scalability.

## Definition and Target Market

Workload consolidation is a popular model in IT nowadays. The goal is to simplify administrative duties and associated costs by consolidating multiple workloads onto the same physical platform, or better yet, consolidating multiple similar applications into a single large application.

However, not all application instances are amenable to platform or application consolidation. For example, Microsoft Exchange Server is designed for active-active multi-server environments, and consolidating the images onto a larger system often introduces complexity and overheads that it was not designed for.

Other applications, however, are excellent candidates for consolidation, databases being a prime example. Both Microsoft SQL Server 2005 (2008 planned) and Oracle 10g/11g are very capable of handling multiple database instances or views, and can host multiple concurrent transactions routed to those diverse instances. Customers can consolidate their databases such that multiple 2 or 4-way servers can be consolidated onto a single 8 or 12-way server hosting a unified database service. While this does not directly reduce the database administration task, it does simplify server and operating system management and maintenance for real cost savings.

## Description

The following sections describe the business and operational challenges of a database consolidation host.

### Customer Problems / Challenges

Server, and database, proliferation can quickly become an administrative and maintenance nightmare. Over the last decade IT shops have seen their administrative budgets expand from 25 to 75 percent of total budget, sometimes more, as ever increasing numbers of servers are added to their IT infrastructure.

One way to address the challenge is server consolidation. Moving all the servers into a common room, called "physical consolidation" by Gartner, is the first and easiest step. The next step is called logical consolidation, whereby multiple servers are moved on to partitions on larger single machine. The final, and the most challenging yet rewarding, is application consolidation, or rational consolidation. It is this process where the number of instances of the hardware, operating system, and application are reduced, simplifying the administrative loads of the IT staff.

The issue is that rational consolidation cannot be hosted by just any server platform. It must be robust to handle the additive loads, it must be reliable to insure availability of merged application services, and it must be able to dynamically accommodate changes in application service demands.

Each of these criteria is well addressed by the Fujitsu Siemens Computers PRIMEQUEST server.

### PRIMEQUEST Business Value

The Fujitsu Siemens Computers PRIMEQUEST server family represents best-in-class SMP (Symmetrical Multi-Processor) capabilities for Itanium-based systems. From the ground up, the PRIMEQUEST server is designed to support robust and demanding database workloads, providing excellent performance and reliability.

Examples of this are found in:

- High performance memory sub-system that provides the optimum host platform for demanding memory-intensive database workloads. These attributes are critical to the robust hosting of aggregate loads without degradations in transaction response times.
- High-bandwidth I/O sub-system that provides optimal access to multiple database resources.
- Unparalleled hardware reliability with mirror-mode operations. For the most demanding business environments PRIMEQUEST can provide best-in-class availability.

Independent of the PRIMEQUEST hardware design, Fujitsu has worked with Microsoft and the Linux vendors, notably Red Hat and SuSE, to improve the reliability, availability, and robustness of the respective operating systems.

Specific to database consolidation, PRIMEQUEST servers provide near linear scaling as the hardware configuration is grown. This allows the server to host a larger aggregate database image without degrading performance. Furthermore, Automatic Reconfiguration can be used to grow the database domain or partition quickly, accommodating business growth or further consolidation as needed.

## Critical Server Attributes

The following sections describe the specific features of PRIMEQUEST servers in the context of critical server attributes required for database consolidation.

### Scalability

Scalability is very important to the database consolidation environment. The server ability to host the additive transaction loads of multiple consolidated databases directly influences the transaction response times and how efficiently they run. This is true whether the workload demands are computational or I/O-intensive: both must scale to meet the demands of the consolidated database services.

PRIMEQUEST servers offer unmatched scalability for Itanium-based servers. The combination of processor power, tied together into a true SMP system with the Fujitsu-designed memory and I/O sub-systems provide a robustly scalable platform for hosting large database operations. Furthermore, the PRIMEQUEST design is such that as the system grows in processor count, memory, or I/O resources, the system performance does not degrade, remaining near linear as the system expands.

### Performance

Much like scalability, outright performance is critical to the database consolidation environment. The speed at which the increased number of transactions executes is heavily dependent upon the overall system performance, ranging from processor speed to memory sub-system latency and throughput.

The PRIMEQUEST server is an ideal blend of processor performance, courtesy of the Intel Itanium2 design and implementation, and the Fujitsu-designed memory sub-system. Individual transactions will demonstrate shorter execution times relative to other competing platforms, and the excellent scalability will provide an ideal host platform for the concurrent users and transactions from consolidated databases.

I/O performance is very important to consolidated database operations, as the aggregate loads of multiple concurrent users operating against different database views stresses the latency, bandwidth, and flexibility of the host I/O sub-system. The PRIMEQUEST server incorporates a PCI-Express fabric for handling the memory-to-I/O transfers, and by virtue of being separate from the processor system boards, can offer as much connectivity as the workload requires. PRIMEQUEST supports extensive PCIe and PCI-X external interfaces for connecting to external resources.

### Big Memory

Databases thrive on memory. In many instances certain tables or indices will be permanently fixed in memory to avoid time-consuming I/O operations, and there is a direct relationship between the amount of memory, the number of fixed tables or indices, and database performance. For the database consolidation server, the total amount of memory is important to efficiently support the multiple fixed tables and their memory requirements.

One of the critical factors is the amount of memory that can be hosted per system board, and therefore per CPU. The PRIMEQUEST design hosts 256GB per system board at the outset, or 32GB per CPU core. This allows a generous pairing of memory with processors, even for systems that do not require the full complement of system boards.

### High I/O Bandwidth

Consolidated database servers may host a mixture of OLTP (online transaction processing) and data warehouse or business intelligence workloads.

Normally, OLTP workloads do not present any bandwidth requirements. However, business intelligence, data warehousing, or decision support workloads can present significant I/O bandwidth requirements due to the mostly sequential processing (table scans) of data.

PRIMEQUEST servers easily satisfy these requirements, providing up to 128 PCI-X 100/133 (or up to 64 PCIe as well as a mix of PCI-X and PCIe) external connections distributed across (up to) eight I/O Boards. Each board is capable of providing a sustained 6.25GB/sec bandwidth, bi-directional, that can satisfy the most demanding database access streams.

### Flexible I/O

The ability to float I/O resources within partitions, either as the result of system component failure or system growth, is a critical requirement for database consolidation platforms. Storage resources, and access to them, change or grow with the business, and insuring access to those resources for all components of the database server is paramount. For example, the customer can add system boards to an existing partition, and using the PRIMEQUEST Flexible I/O facility, provide direct access to the I/O resources for the new board. Similarly, new I/O resources can be added to an existing partition, and be made available to all system board components.

### Dynamic Reconfiguration<sup>1</sup>

Dynamic Reconfiguration is a desirable function for the host database consolidation platform. This is because DR provides a means to:

1. Perform concurrent maintenance, where resources can be taken in and out of service to perform upgrades without taking the application service and database down, and
2. To expand the system configuration with little or no service intervention.

Both of these are important to a 24\*7 business environment.

---

<sup>1</sup> Dynamic Reconfiguration is planned to be supported from 1<sup>st</sup> half of 2008

## Operational Efficiency

Database consolidation servers must be easy to administer, if only to help the database administrators deal with the complexities of a composite database environment.

PRIMEQUEST has a number of integrated management facilities that simplify host server maintenance and administration, foremost of which is the web-based management interface for service and diagnostics, the integrated Gbit network connections and associated VLAN assignments, and integrated KVM facilities for controlling the host OS. These points of control allow the system administrator to control system resources with assurance, and they can quickly and confidently control the PRIMEQUEST host platform.

## Environmental Advantages

PRIMEQUEST is the penultimate database consolidation platform:

- It has a small footprint relative to its capabilities, and is in fact the smallest server on the market.
- It has superior energy dissipation characteristics.
- It has a number of integrated resources, one of which is an integrated KVM switch that precludes the maintenance of a complex jungle of jumper cables and external switches.

All of these PRIMEQUEST attributes contribute to a best in class physical, logical, and rational database consolidation platform

## Fault Immunity

A single database may have a relaxed availability requirement, where it must only be operational during normal business hours. Another may be accessed only at night during batch processing. Yet another may be needed only for week-end processing. Combining them all onto a consolidated database server, though, creates an environment that needs to be accessible everyday, all the time.

PRIMEQUEST servers provide the utmost reliability for hosting these demanding environments. The PRIMEQUEST servers are the only Itanium-based servers to provide mirror-mode operations, thereby providing extra assurance that if something did indeed happen to the hardware that its effect would be completely nullified. The level of redundancy in the PRIMEQUEST design is based upon years of experience with mission critical mainframe and large UNIX servers, and provides a highly reliable host platform.

## Mission-Critical OS

Similar to the hardware reliability requirements, the host operating system for a 24\*7 database consolidation environment needs to be resilient and robust.

Microsoft Windows Server, particularly Enterprise Edition and Datacenter Edition, has been proven to be a very suitable operating system for hosting the SQL Server database, which in turn is a robust and flexible database management product.

Much of the work that Fujitsu has put into the Linux open source community has been for operating system "hardening". Examples are device driver enhancements, improved dump and diagnostic capabilities, error recovery and retry. These capabilities enhance the mission-critical aspects of the Linux operating system and how it will operate on the PRIMEQUEST server.

## Multiple OS Support

The essence of the database workload environment is determined by the choice of database vendor. In the context that not all database products are available on all operating systems, the ability of PRIMEQUEST to support multiple operating systems is important.

The issue is whether or not these operating systems need to be concurrently supported on a database consolidation platform. In the typical deployment, this is not the case: the database(s) and host OS are typically the sole workload on the platform.

## Solution Description and Requirements

Hosting a mission-critical 24\*7 consolidated database environment requires a number of key contributions:

1. Optimal hardware platform, represented by the Fujitsu Siemens Computers PRIMEQUEST server.
2. Database software with proven capabilities  
Fujitsu Siemens Computers recommends Oracle 10g as the class-leading relational database management system for Linux, which offers proven scalability, reliability, and manageability. Microsoft SQL Server is also a robust database management system, and is the recommended solution for all Windows Server deployments.
3. Proven service and support  
Fujitsu Siemens Computers and Oracle Professional Services can accommodate the most demanding customer application requirements, and with Fujitsu Siemens Computers' service organization, provide excellent ongoing support to guard against service outages.

The combination of top-flight hardware, software, and services and support provides the customer with the best solution for their database consolidation requirements.

## Competitive Analysis

There are a number of vendors that offer products to compete against PRIMEQUEST, based both on Itanium and RISC processors.

Competing Itanium vendor in EMEA mainly is HP, offering Superdome Integrity servers. Competing RISC system vendor is IBM, offering systems based upon the Power technology.

PRIMEQUEST either matches or surpasses the performance of these respective platforms using several criteria: performance, reliability, and flexibility.

### Performance

#### CPU Performance

At the heart of PRIMEQUEST servers are the Itanium 2 processors. The current incarnation, the "Montecito" design, incorporates a 24MB 3<sup>rd</sup>-level cache running on a 1.6 GHz processor. These capacities, coupled with the EPIC (Explicitly Parallel Instruction Computing) architecture, allow Itanium-based servers to out-perform servers using competitive vendor processors, such as IBM and AMD.

For reference, the Standard Performance Evaluation Corporation (SPEC) has a number of benchmark results that highlight the Itanium capabilities. See <http://www.spec.org>. By almost any measure, the Itanium-powered servers provide class-leading performance.

#### Memory Performance

In commercial systems hosting databases, memory performance is the key to robust and sustainable **system** performance. PRIMEQUEST couples the Itanium 2 processors together using a custom-designed memory sub-system that offers three benefits:

1. Low latency for local (on-board) and remote (off-board) access. Memory latency is the defining attribute for many database applications, and consistent memory latency is the hallmark of a robust memory design.
2. Symmetrical multi-processor (SMP) memory design. Because of the low ratio of remote/local memory access, the PRIMEQUEST server offers an SMP environment for the hosted operating system and applications.
3. High bandwidth. Memory latency determines how fast a processor can get the first data element. Bandwidth determines how fast the processor can get the last. PRIMEQUEST offers class-leading memory bandwidth, both peak and sustained. Of the two values (see Appendix A for details), sustained is the more realistic, in that it reflects how the memory sub-system supports the aggregate configuration of processors and I/O components.

These attributes combine to provide class-leading memory-system performance in the PRIMEQUEST servers. Competitive designs suffer from ccNUMA (cache coherent Non-Uniform Memory Architecture) effects, where memory performance degrades as the system gets physically larger, i.e., more processors are added to the server.

#### I/O Performance

PRIMEQUEST servers are well-balanced systems, with robust Itanium 2 processors coupled together with an SMP memory sub-system to provide excellent computational and database processing power. The PRIMEQUEST server extends these capabilities by providing a flexible, robust, and powerful I/O sub-system.

Key to the PRIMEQUEST I/O capabilities is the Fujitsu-designed "Southbridge" implementation. It combines a custom ASIC that interfaces directly to the PRIMEQUEST memory cross-bars, that in turn interfaces to a flexible high-performance PCI Express fabric. These PCIe interfaces, four-lanes wide for an aggregate bi-directional bandwidth of 2GB/sec per link, are externally extended to both PCIe and PCI-X host bus adapters and network interface cards.

With 128 PCI-X (or 64 PCIe as well as a mix of PCI-X and PCIe) interfaces the connectivity to PRIMEQUEST servers is excellent. Adding to this capability is the large number of integrated services, where PRIMEQUEST servers offer up to 32 Gbit interfaces, 32 Ultra320 SCSI drives, and integrated Management LANs and KVM interfaces., none of which requires the customer to consume valuable PCI slots.

### Reliability

PRIMEQUEST servers offer unparalleled reliability. Not only are all major components redundant and hot-pluggable, PRIMEQUEST servers are the only systems to offer System Mirror mode.

System Mirror mode takes the two Global Address crossbars (GAC) and runs them in lock-step. It also takes the four Global Data crossbars (GXD) and mirrors them into two steps. Finally, System Mirror mode mirrors banks of memory on systems boards.

The benefit of System Mirror mode is that a failure of any of the components will not take the system down. This stands in unique contrast to systems that, despite offering significant redundancy, will "bounce" if one of their redundant components fails. PRIMEQUEST servers offer that last bit of insurance for application availability, by providing a platform that is, on all intents and purposes, immune to faults.

## Flexibility

PRIMEQUEST servers today support up to sixteen partitions, each hosting an independent operating system instance. While even this is higher than competitive products such as the HP Superdome, PRIMEQUEST flexibility goes a step further by allowing customers to separately configure processor/memory resources from I/O resources.

Fujitsu Siemens Computers calls this Flexible I/O, and it offers several benefits. For one, partition configurations are determined by the application requirements. Both System Boards and I/O Boards can each be physically split into two extended partitions (XPAR), or can be combined into larger SMP partitions hosting larger applications. Thus, if the application requires a high processor count with relatively little I/O, then multiple System Boards can be configured with a single I/O Board. Inversely, if the application requires a high degree of connectivity but not too many processors, multiple I/O Boards can be combined with a single System Board running XPAR.

If there is a System Board failure, a hot-standby (reserve) System Board can be deployed to take its place. Using Flexible I/O, all of the resources available to the previous System Board become transparently available to the replacement system board.

## Summary

The family of PRIMEQUEST servers offers “game changing” platforms for hosting advanced Linux and Windows applications. Engine for engine, they offer class-leading performance and scalability, unmatched reliability, and superior flexibility.

The robustness of the Itanium 2 processors, coupled with the Fujitsu-designed memory and I/O sub-systems, offers industry-leading performance for database platforms. The System Mirror mode places PRIMEQUEST servers into the rarified plateau of fault-tolerant servers, where component failures are not even noticed by the hosted operating systems. Finally, PRIMEQUEST XPAR capabilities, coupled with Flexible I/O, offer a high degree of freedom for configuring physical components to best match application requirements.

## Appendix A: Detailed Summary of PRIMEQUEST Attributes

	PRIMEQUEST 580	PRIMEQUEST 540	PRIMEQUEST 520
Type	Floor-standing cabinet	Floor-standing cabinet	Rackmount Chassis: 12U
<b>CPU</b>			
Processor	Itanium® 2 Montecito dual-core Clock speed: 1.4 GHz (12MB L3 cache) or 1.6 GHz (24MB L3 cache)		
<b>System</b>			
Interconnect	Point-to-point crossbars: 2 Address, 4 Data		2 Address, 2 Data
	Bandwidth: 136.5 GB/sec sustained	68.2 GB/sec sustained	34.1 GB/sec sustained
	Latency: 208-276 ns	208 - 256 ns	208 - 256 ns
System Boards	8	4	2
Main Memory	2TB (256 x 8GB)	1TB (128 x 8GB)	256GB (64 x 4GB)
Internal HDD	Max 4.7TB (32 x 147GB)	Max. 2.35TB (16 x 147GB)	Max 584GB (8 x 73GB)
Expansion Slots	Maximum 128 PCI slots: IOU x 8 x 4 = 32 internal + PCI Box external: 8 x 12 PCI-X or 8 x 6 PCI-X plus 8 x 4 PCIe IOU: 64bit/133MHz Short x 16 64bit/100Mhz Short x 16  PCI Boxes max: PCI-X: 64bit/133MHz Long x 32 PCI-X: 64bit/100MHz Long x 64 PCI Express 4lane Short x 32	Maximum 64 PCI-slots: IOU x 4 x 4 = 16 internal + PCI Box external: 4 x 12 PCI-X or 4 x 6 PCI-X plus 4 x 4 PCIe IOU: 64bit/133MHz Short x 8 64bit/100Mhz Short x 8  PCI Boxes max: PCI-X: 64bit/133MHz Long x 16 PCI-X: 64bit/100MHz Long x 32 PCI Express 4lane Short x16	Maximum 16 PCI slots: IOU = 4 internal + PCI Box external: 1 x 12 PCI-X or 1 x 6 PCI-X plus 1 x 4 PCIe IOU: 64bit/133MHz Long x 2 64bit/100Mhz Long x 2 PCIe: 8lane Short x 4 PCI Boxes max: PCI-X: 64bit/133MHz Long x 4 PCI Express 4lane Short x 4
I/O Interfaces	10GBASE-LR: maximum 4 1000BASE-T: maximum 32 100BASE-Tx: 2 External SCSI Port: 16 Video (Analog RGB): 1 Serial (Dsub 9pin): 16 USB1.1: (KB, Mouse): 4 USB2.0: 32 DVD ROM: 1 Integrated KVM	10GBASE-LR: maximum 4 1000BASE-T: maximum 16 100BASE-Tx: 2 External SCSI Port: 8 Video (Analog RGB): 1 Serial (Dsub 9pin): 8 USB1.1: (KB, Mouse): 4 USB2.0: 16 DVD ROM: 1 Integrated KVM	1000BASE-T: 4 100BASE-Tx: 2 External SCSI Port: none Video (Analog RGB): 1 Serial (Dsub 9pin): 4 USB2.0: (KB, Mouse): 2  DVD ROM: 1 Integrated KVM
<b>Reliability</b>			
Redundancy	Disks, Power units, Fans, Cross-bars, Server Management processor, Gbit switch		Disks, Power units, Fans, Cross-bars
Hot Swap	PCI cards, Disks, Power units, Fans, Server Management processor		
Partitions	Maximum 16	Maximum 8	Maximum 4
<b>Installation Details</b>			
Dimensions	BASE Chassis: W x D x H 740 x 1,100 x 1800 mm (29.1 x 43.3 x 70.9 inches)	BASE Chassis: W x D x H 740 x 1,100 x 1800 mm (29.1 x 43.3 x 70.9 inches)	Rackmount Chassis: 12U 482 x 820 x 530 mm (17.3 x 32.3 x 20.9 inches)
Weight	Maximum 720 kg	Maximum 600 kg	Maximum 120 kg
Power Requirements	AC200-240 +/-10% Single Phase 50/60 Hz (+2/-4%)		
Power Consumption	10,900W	6,300W	3,320W
Operating Systems	Red Hat Enterprise Linux AS 4.0 , AP 5.0 SUSE Linux Enterprise Server 9, 10 Windows Server 2003, Enterprise Edition for Itanium-based Systems Windows Server 2003, Datacenter Edition for Itanium-based Systems		