May 2007

MANAGEMENT BRIEF

IMPACT OF IBM SYSTEM p SERVER VIRTUALIZATION

Transforming the IT Value Equation with POWER6 Architecture

International Technology Group

4546 El Camino Real, Suite 230
Los Altos, California 94022-1069
Telephone: (650) 949-8410
Facsimile: (650) 949-8415
Email: info-itg@pacbell.net

TABLE OF CONTENTS

List of Figures

EXECUTIVE SUMMARY

What Strategy?

Virtualization is no longer an “emerging” technology. Its potential to increase the efficiency of IT infrastructures has become widely recognized. Organization-wide improvements in overall capacity utilization and reductions in the costs of acquiring, managing and supporting server and storage platforms have become realistic opportunities.

Yet, in most organizations, use of virtualization is still at any early stage. The focus is on small-scale projects. One reason for this is that, while many organizations recognize the potential of virtualization in general terms, they are unclear as to what the scope and goals of virtualization strategy should be.

Key questions are raised. Should virtualization be implemented only for new projects, or applied to existing infrastructures? Should opportunities be pursued on a case-by-case basis, or targeted as a broader goal of IT strategy? What level of investment should be made, for what return? What priority should be given to virtualization initiatives relative to the many other claims on IT resources?

Before these questions can be answered, a broader understanding of the costs and benefits of virtualization at the enterprise level is needed. This report marks a first step in this process.

Economic Benefits

Overview

This report focuses on the potential economic benefits of large-scale virtualization for two subsets of the large organization IT environment – UNIX and Linux servers.

Specifically, it deals with the potential benefits of strategies that effectively exploit the virtualization strengths of the IBM System p platform. It addresses established System p strengths, as well as the new capabilities announced by IBM with the introduction of its POWER6 processor-based p570 servers in May 2007.

Potential benefits are illustrated by three composite profiles of large financial services, manufacturing and retail companies. For each profile, two sets of scenarios are presented: (1) conventional scenarios built around diverse, multivendor bases of UNIX and Intel-based Linux servers; and (2) virtualized scenarios in which the same applications run on System p servers leveraging the full potential of virtualization.

For each profile and scenario, costs are then compared. Results are summarized below.

UNIX Server Costs

Five-year server operating costs for System p virtualized scenarios including POWER6 processor-based p570 and POWER5+ processor-based System p5 servers averaged 72 percent less than those for conventional scenarios.

Operating costs, in this context, include hardware maintenance; update subscriptions and support for systems and database software; personnel for system administration-related functions; and facilities costs for data center occupancy, power and cooling.

Even allowing for the costs of acquiring new System p hardware and systems software required to realize these savings, overall costs were still significantly lower. If acquisition costs are included in virtualized scenarios, five-year costs for these averaged 66 percent less than those for conventional scenarios.

Figures 1 and 2 summarize results.

Figure 1
UNIX Server Comparisons: Five-year Operating Costs

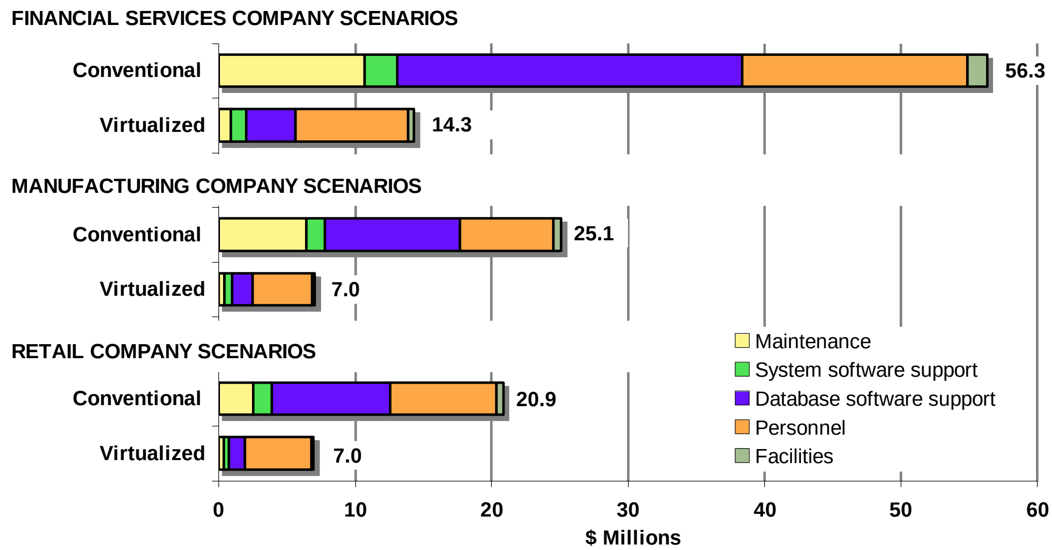
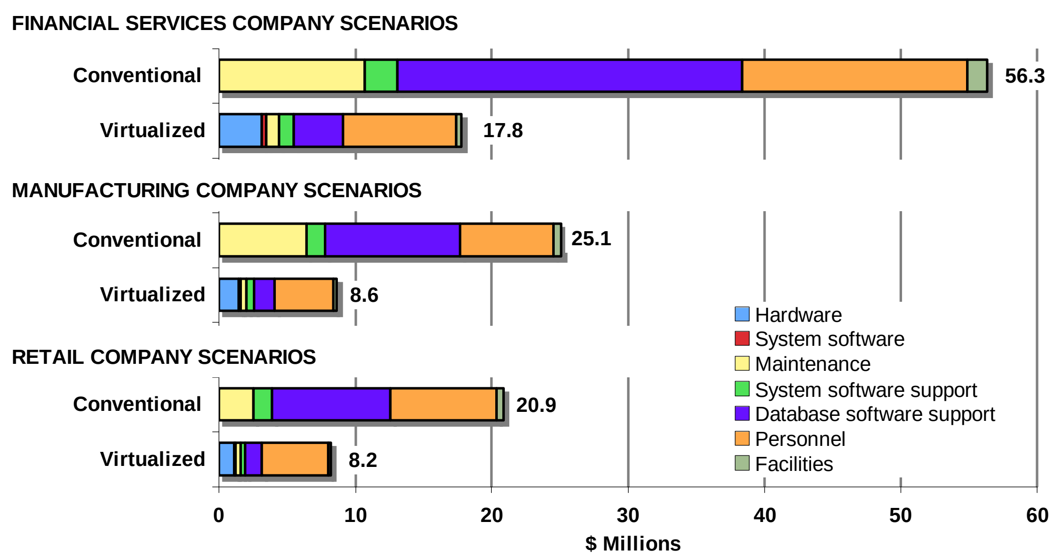


Figure 2
UNIX Server Comparisons: Five-year Overall Costs
(Virtualized Scenarios include Operating and Acquisition Costs)



Detailed breakdowns of costs, along with explanations of variations between scenarios for these comparisons are provided in figures 17 and 18 respectively in the Cost Picture section of this report.

Cost reductions were enabled through multiple System p virtualization capabilities. Dynamic logical partitions (LPARs), along with Workload Partitions (WPARs) combine with effective system and workload management facilities to enable high levels of server consolidation and capacity utilization.

The comparisons in figures 1 and 2 are between conventional scenarios and virtualized scenarios, which include p570 as well as System p5 servers.

The new POWER6 processor-based System p 570 servers, however, offer significant improvements in performance and virtualization capability. If costs of p570 models in virtualized scenarios are compared with those of equivalent servers in conventional scenarios, disparities widen. Five-year operating costs for p570 servers averaged 79 percent less than those for conventional equivalents.

If acquisition costs for p570 servers are included in comparisons, five-year costs averaged 72 percent less than those for conventional equivalents. Figures 3 and 4 summarize results.

Figure 3
UNIX Server Comparisons: Five-year Operating Costs for p570 and Equivalent Conventional Servers

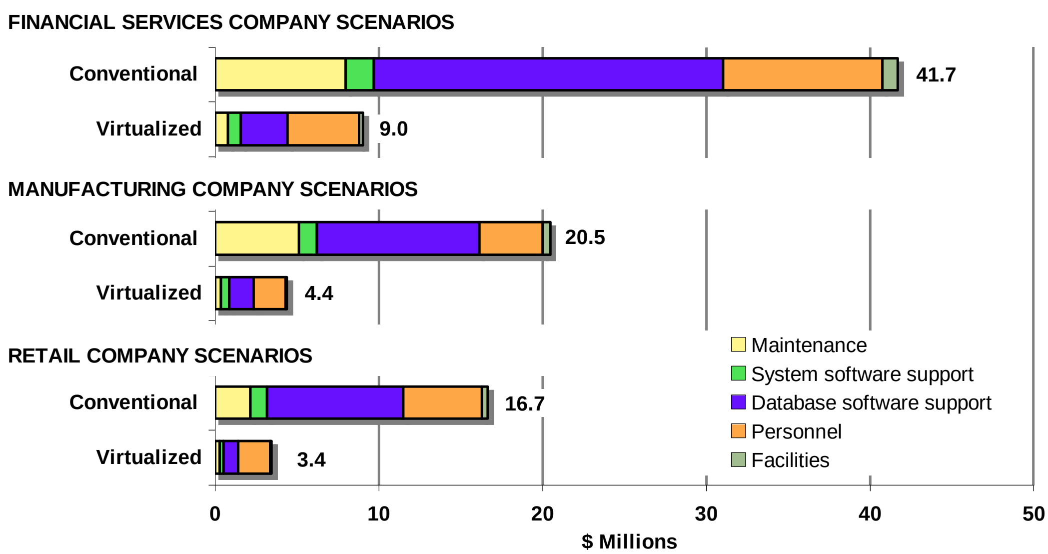
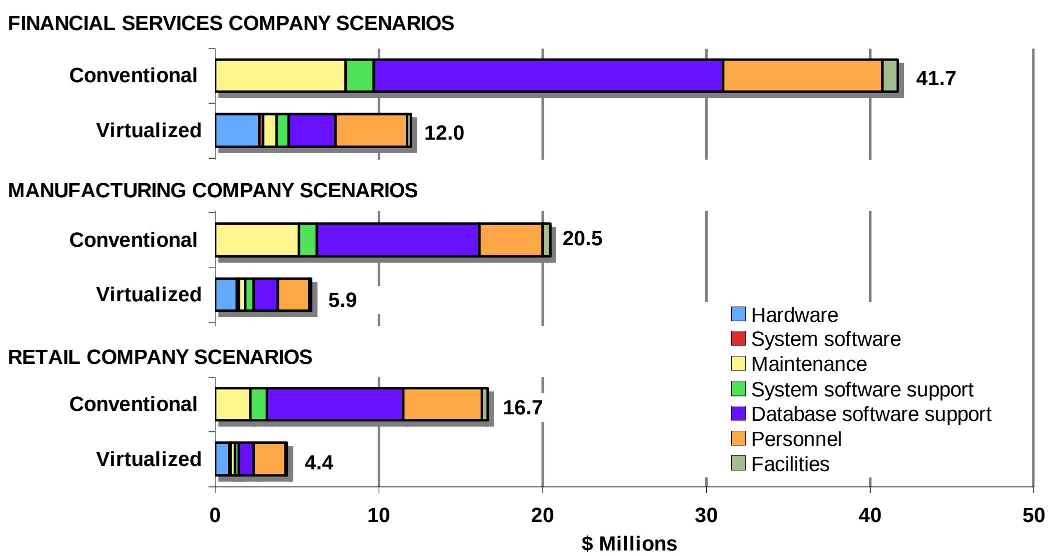


Figure 4
UNIX Server Comparisons: Five-year Overall Costs for p570 and Equivalent Conventional Servers (Virtualized Scenarios include Operating and Acquisition Costs)



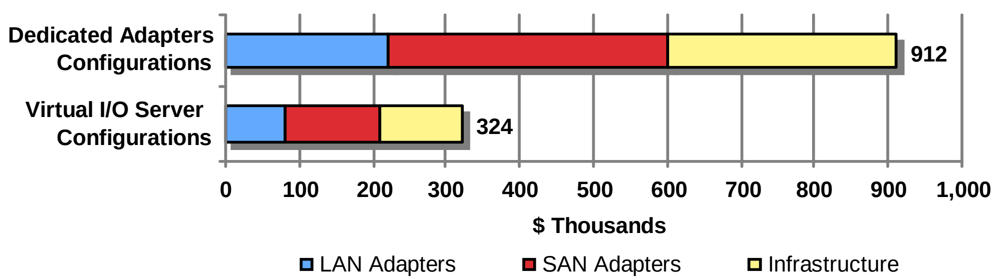
Detailed breakdowns of costs, along with explanations of variations between scenarios for these comparisons are provided in figures 17 and 19 respectively in the Cost Picture section of this report.

Lower costs for p570 servers are due to virtualization capabilities supported on System p5 servers, as well as to higher levels of performance and new POWER6 processor-specific capabilities such as Shared Dedicated Capacity, which enable further improvements in system-level capacity utilization.

For p570 as well as System p5 servers, significant economies were realized in I/O-related costs, including local area network (LAN) and storage area network (SAN) adapters, as well physical infrastructures supporting these through use of System p Virtual I/O Server technology. Physical infrastructures include transceivers, structured cabling and switches.

In virtualized scenarios, use of Virtual I/O Servers resulted in I/O-related costs for servers employing LPARs or both that averaged 65 percent less than would have been the case if dedicated adapters were employed. Figure 5 summarizes results.

Figure 5
Five-Year I/O-related Costs for Partitioned IBM System p Servers: All Virtualized Scenarios



For both sets of comparisons, full time equivalent (FTE) staffing levels for system administration-related functions such as asset, capacity, change, configuration and performance management are significantly lower for virtualized scenarios. Personnel costs for these functions are correspondingly less.

This is due to reductions in numbers of physical servers; reduced diversity of hardware and software platforms; replacement of older-technology servers and systems software with latest-generation System p hardware and software platforms; improved management tools and practices, including increased automation; and use of System p virtualization capabilities.

Linux Server Costs

In addition to the AIX operating system, System p servers may also run the major Linux distributions in native mode. The new IBM System p Application Virtual Environment (System p AVE) allows x86 Linux applications to run on the System p platform without modification. Many of the same virtualization capabilities are supported as for AIX, making this platform a candidate for Linux server consolidation.

To address this potential, comparisons were made for different sets of Linux-based applications for all three composite profiles. Comparisons are based on mixes of business applications as well as Web and intranet serving, file serving, software development and other functions.

Two sets of scenarios were compared. In conventional scenarios, applications are deployed on one- to four-way Intel-based servers. In virtualized scenarios, System p 570 and System p5 servers are employed.

Figures 6 and 7 summarize results.

Figure 6
Linux Server Comparisons: Five-year Operating Costs

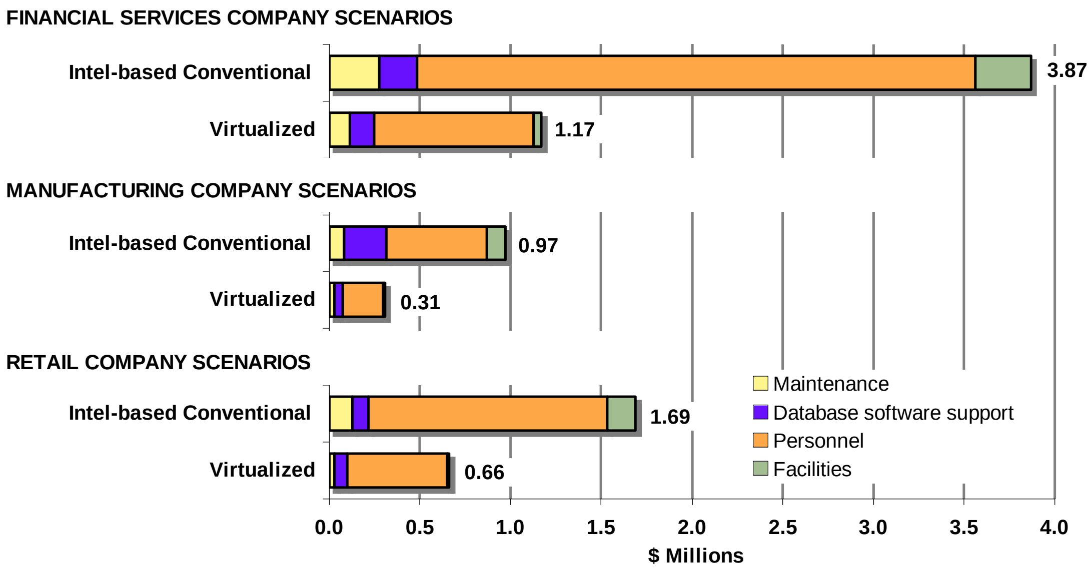
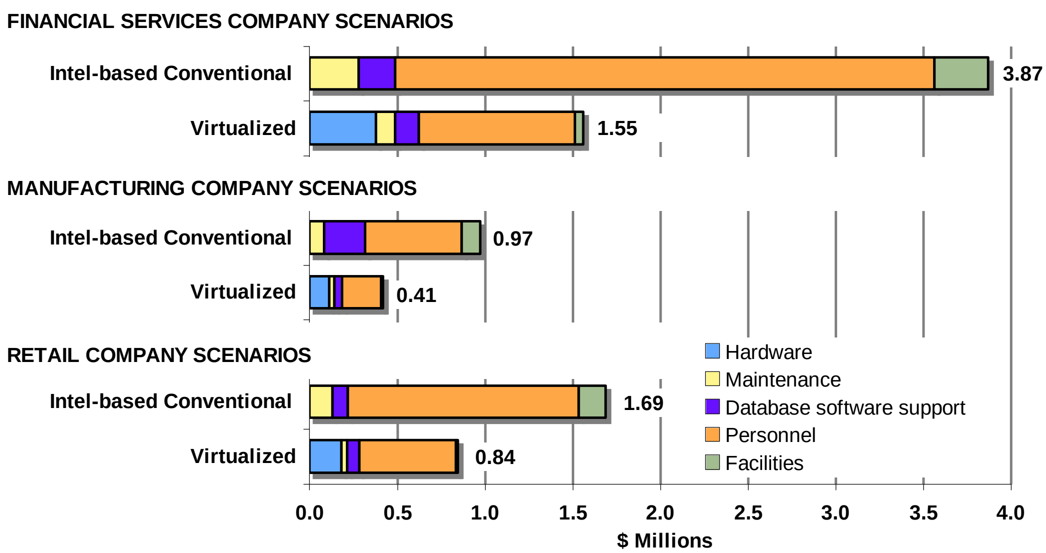


Figure 7
Linux Server Comparisons: Five-year Overall Costs (Virtualized Scenarios include Acquisition Costs)



Five-year operating costs for System p Linux server scenarios averaged 67 percent less than those for conventional scenarios. If acquisition costs are included in System p scenarios, five-year costs for these averaged 57 percent less than those for conventional scenarios.

In this presentation, operating costs again include hardware maintenance; update subscriptions and support for systems and database software; personnel for system administration-related functions; and facilities costs for data center occupancy, power and cooling. Database costs are for Oracle 10g.

Detailed breakdowns of costs, along with explanations of variations between scenarios for these comparisons are provided in figures 21 and 22 in the Cost Picture section of this report.

Factors causing lower costs for virtualized scenarios are generally similar to those for UNIX server comparisons. The same caveats as to definitions of system administration-related functions apply.

Costs of Downtime

Because of higher levels of availability are enabled by virtualized environments, there are also significant differences in costs of downtime – meaning costs incurred by organizations due to disruptions affecting critical business processes – between conventional and virtualized scenarios.

For UNIX serving, five-year costs of downtime for System p virtualized scenarios ranged from 41 percent to 59 percent less, and averaged 48 percent less than those for conventional scenarios.

For Linux serving, System p virtualized scenarios ranged from 83 percent to 89 percent less, and averaged 85 percent less than those for Intel-based conventional scenarios. Figures 8 and 9 summarize results.

Figure 8

UNIX Server Comparisons: Five-year Costs of Downtime

COMPANY	FINANCIAL SERVICES	MANUFACTURING	RETAIL
Applications	Online banking & brokerage, CRM financial services, various	ERP, SCM, CRM, BI, CFM, SEM, PLM, procurement	SCM, logistics, procurement, eCommerce, CRM
CONVENTIONAL SCENARIOS			
Availability levels	99.3% – 99.92%	99.75% – 99.9%	98.25% – 99.9%
Five-year costs ($000)	**25,649**	**36,950**	**13,419**
VIRTUALIZED SCENARIOS			
Availability levels	99.7% – 99.95%	99.9% – 99.96%	98.5% – 99.95%
Five-year costs ($000)	**10,505**	**21,444**	**7,882**

Figure 9

Linux Server Comparisons: Five-Year Costs of Downtime

COMPANY	FINANCIAL SERVICES	MANUFACTURING	RETAIL
CONVENTIONAL SCENARIOS			
Availability levels	99.32% – 99.83%	98.05% – 99.61%	99.14% – 99.77%
Five-year costs ($000)	**1,858**	**3,299**	**2,588**
VIRTUALIZED SCENARIOS			
Availability levels	99.90% – 99.98%	99.85% – 99.95%	99.73% – 99.98%
Five-year costs ($000)	**213**	**557**	**420**

In these figures, availability percentages reflect hours of system-level outages relative to annual hours of operation of companies, or business areas within companies that are supported by specific applications. The basis of these calculations is detailed in the Cost Picture section of this report.

Costs of downtime for UNIX server scenarios represent lost operating profit (for the financial services profile) or gross profit (for the manufacturing and retail companies). For Linux server scenarios, costs of downtime represent lost productivity. Both sets of calculations are based on industry- and organization-specific values for the effects of outages. The basis of these calculations is detailed in the Cost Picture section of this report.

Lower costs of downtime for virtualized scenarios are due to multiple factors that reduce the frequency and duration of planned as well as unplanned outages. These include reliability, availability and serviceability (RAS) features of System p hardware and the AIX operating system, and the IBM High Availability Cluster Multi Processing (HACMP) failover solution.

LPARs also assist in avoiding planned outages by allowing users to upgrade or modify software without taking systems offline. Two new capabilities – Live Partition Mobility and Live Application Mobility – make it possible to transfer partitions between servers if these need to be shut down for hardware-related reasons.

Live Partition Mobility is supported for AIX and Linux on p570 servers, and enables transfers to occur with no application downtime. It is designed to support applications requiring continuous availability. Live Application Mobility involves a brief interruption of service. It is supported on POWER6 processor-based as well as earlier POWER4 and POWER5 processor-based models running AIX 6.

Effective system and workload management capabilities further reduce risks of bottlenecks and outages caused by workload spikes and operational errors.

Conclusions

Realization of the gains described in this report would require investments in a number of areas, such as migration of applications and workloads, organizational change and staff retraining, which are not addressed by this report. Actual user savings may vary widely in practice.

The principle is nevertheless demonstrated: efficient virtualized server infrastructures built around the System p offer the potential for major improvements in IT cost structures and availability levels. Even if organizations do not realize large-scale, short-term turnover of server bases, the phased creation of such infrastructures over time represents a new, highly attractive opportunity to increase IT cost-effectiveness.

There is a particularly significant opportunity to support large-scale Linux deployment. Use of the System p platform could enable organizations to put in place efficient, scalable and robust server infrastructures at an early stage of deployment. It would become possible to avoid many of the challenges of infrastructure fragmentation, problematic availability and management overhead that, all too often, have been characteristic of x86 server bases.

Linux and virtualization have both been described as “disruptive” or “transformative” technologies. Less well recognized, however, is that the combination of these may represent a force for change that is broader and more powerful than either individually.

Most industry observers believe that the potential of Linux virtualization will not be realized for years. But that is the case for the x86 world. The System p server offers that potential now. Which means, for many organizations, a new field of opportunities may be realized sooner rather than later.

STRATEGIC ROLE

Beyond the quantifiable cost and benefit issues that are the focus of this report, a larger question is raised: what role might virtualization play in meeting the central challenges that IT strategy must address in the 21st century?

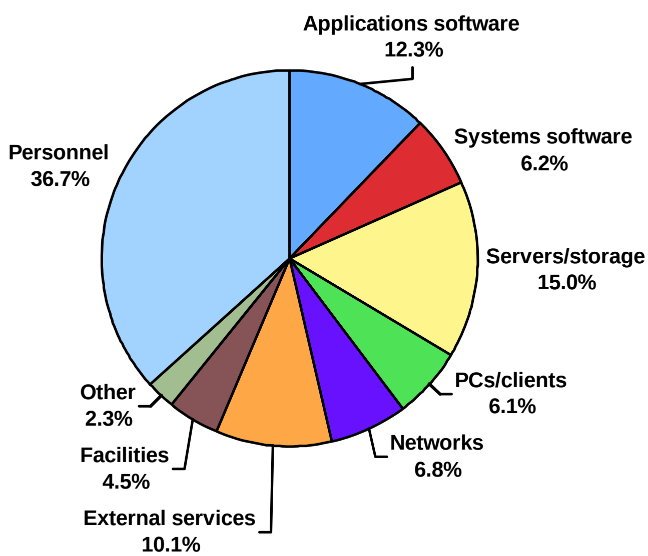
The large organization IT community is in a conservative mood. After aggressive deployment of new applications such as enterprise resource planning (ERP), customer relationship management (CRM) and eBusiness systems in the 1990s, the focus has shifted to solutions that offer more incremental forms of business advantage. Integration of existing environments, rather than delivery of new functionality, dominates many corporate IT agendas.

Structural factors have contributed to this situation. Past deployments have created layers of complexity that undermine overall IT effectiveness. Resources have increasingly been diverted into managing and maintaining underlying platforms, rather than into the higher value-added processes of application delivery and business transformation.

The extent to which this has occurred has been obscured, all too often, by the way in which organizations account for IT expenditure. Conventional categories such as hardware, systems software, middleware, applications, networks, external services, personnel and facilities may be useful for budgetary purposes. They do not necessarily provide insight into underlying cost structures.

The implications may be simply illustrated. Figure 10 shows a conventional breakdown of IT expenditures by Fortune 1000 companies during 2006.

Figure 10
IT Expenditure by Fortune 1000 Companies in 2006: Conventional Breakdown

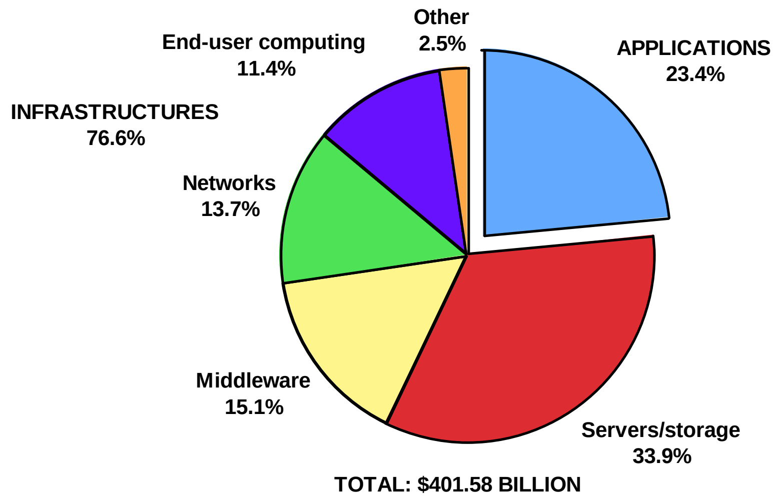


TOTAL: $401.58 BILLION

Alternatively, IT expenditure may be divided into two main categories: applications and infrastructures. If Fortune 1000 outlays during 2006 are broken down in this manner, the results are as shown in figure 11.

Figure 11

IT Expenditure by Fortune 1000 Companies in 2006: Alternative Breakdown



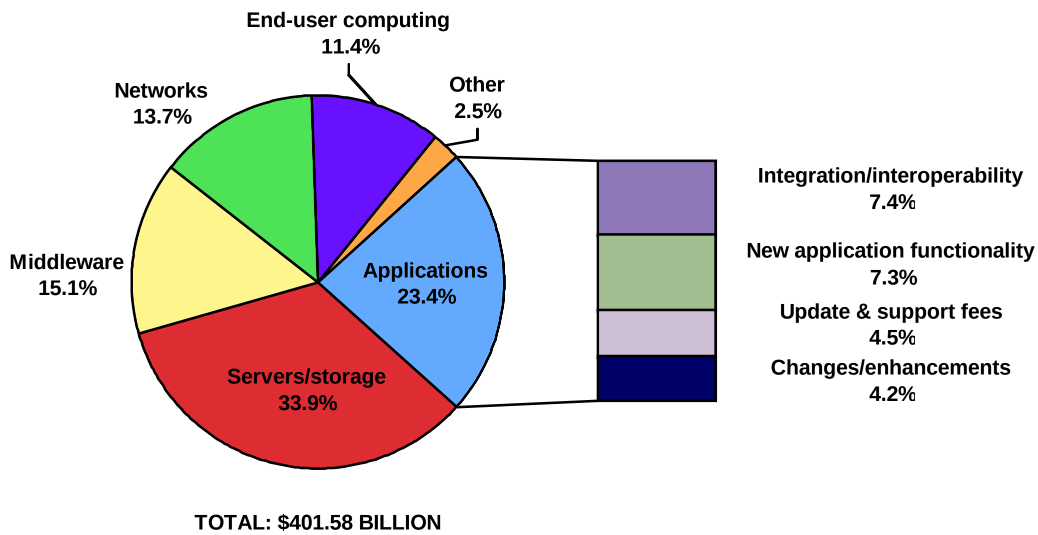
In this presentation, personnel, services and facilities costs, as well as those for hardware and software, are assigned to functional categories. For example, "applications" costs include applications software license and support fees, along with costs of in-house personnel and external services for application development, deployment and maintenance.

A striking conclusion emerges: less than a quarter of overall Fortune 1000 IT expenditures in 2006 were for applications. More than three-quarters went into infrastructures.

Moreover, the majority of applications expenditure is accounted for by update and support fees for installed applications, changes and enhancements to these, and tools and programming for systems integration and interoperability. As figure 12 shows, an average of only approximately seven percent of overall IT expenditure went into the delivery of new application functionality.

Figure 12

Applications Expenditure by Fortune 1000 Companies in 2006



A key principle should be highlighted. Applications are the direct source of business value delivered by IT. Users interact with, and business processes are enabled by applications. Underlying IT resources such as databases, middleware, servers, storage, PCs and networks are simply the delivery mechanisms for these. Yet, the majority of resources are expended on delivery mechanisms.

The impact of infrastructure fragmentation, however, extends beyond simple diversion of resources. Complexity of underlying infrastructures has materially increased the cost and risk involved in major new application deployments, and has encouraged a focus on maintenance and incremental enhancement of existing applications.

This situation has inevitably discouraged large-scale application initiatives. It has also helped to create a perception among many corporate executives that IT investment is a comparatively high-risk, low-reward proposition, and that organizational resources should be focused elsewhere.

It is from this perspective that the strategic significance of virtualization should be viewed. Even small reductions in overall infrastructure complexity could enable organizations to significantly increase the resources devoted to new application initiatives, and to reduce costs, risks and delivery times for these.

Virtualization initiatives may thus act as a catalyst not only for major shifts in IT cost structures, but also for broader, transformational increases in IT effectiveness.

Organizations are often wary of “quantum leaps” in technology. But a quantum leap in the organization and operation of server infrastructures has nevertheless become possible.

IBM SYSTEM p SERVER

Overview

The virtualization capabilities of the IBM System p server form part of a larger system design that is also characterized by high levels of system-level performance, advanced virtualization capabilities, optimization for availability as well as other variables of service quality, and extensive autonomic capability – meaning the application of artificial intelligence technologies to IT administration tasks.

The potential of the System p server for IT cost reduction, which is illustrated in this report, is a function of all of these capabilities. The central strength of the platform is that these capabilities are designed into the core system architecture, and integrated and optimized in a mutually reinforcing manner. The overall impact in terms of operating efficiency is such that "the whole is more than the sum of the parts."

Virtualization Capabilities

Partitioning and Management

System p capabilities in this area include the following:

- ***Partitioning***. The System p platform supports three complementary forms of partitioning. Hardware-based ***LPARs*** are implemented through a firmware-based ***hypervisor***, and enable processor, memory and Ethernet resources to be shared between instances of AIX and Linux. Up to 64 LPARs with one or more cores each are supported.

  Advanced POWER Virtualization (APV), an optional System p feature, allows I/O resources to be shared and dynamically re-allocated between instances of AIX 5.3 and later, and Linux. Partitions may be configured in increments as small as one tenth of a core. This capability is commonly employed to improve load balancing for large, complex workloads.

  WPARs, which are supported for AIX 6, provide a further level of capability. WPARs allow users to create multiple software-based partitions on top of a single AIX instance. This approach enables high levels of flexibility and capacity utilization for applications executing heterogeneous workloads, and simplifies patching as well as other operating system maintenance tasks.

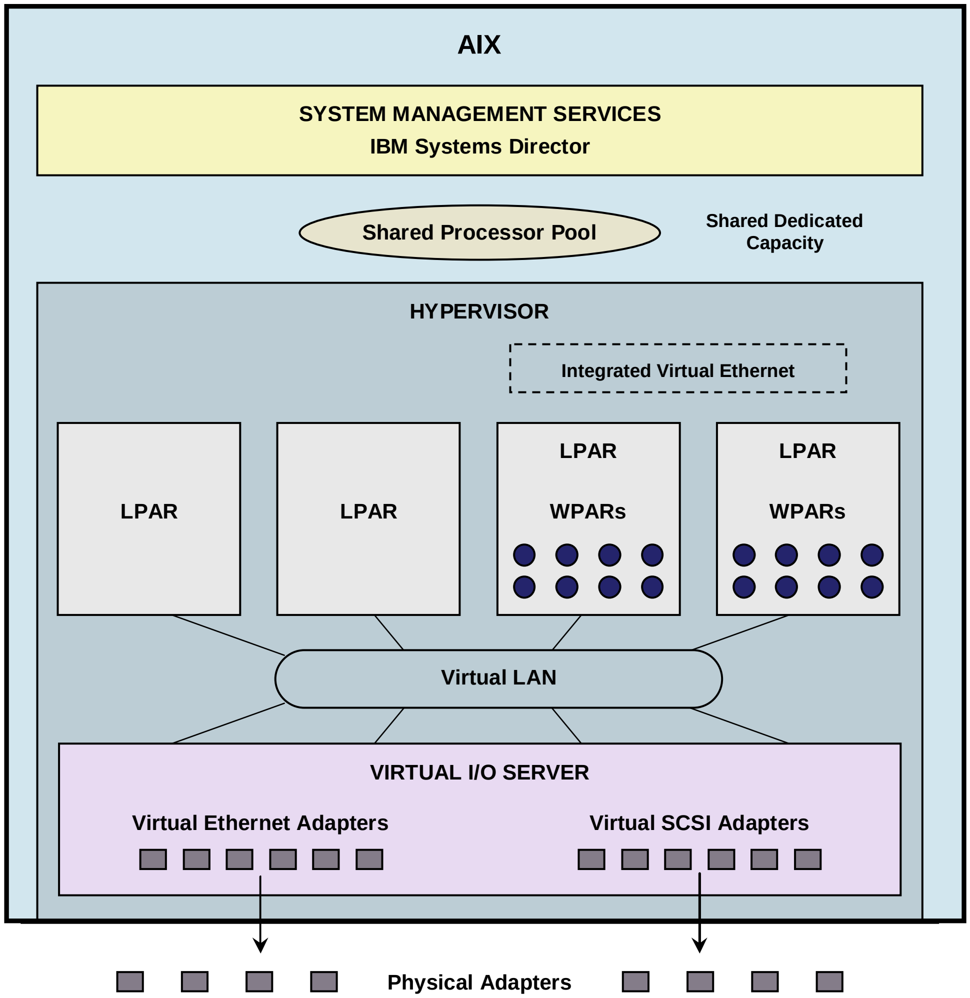
  These technologies, separately and – to an even greater extent – combined, allow organizations to consolidate multiple applications, as well as multiple components of the same application (e.g., database, application and Web serving) that would otherwise require dedicated servers.

- ***System and workload management***. Partitioning techniques enable consolidation, creating the potential for significantly higher levels of overall capacity utilization than may be realized with dedicated servers.

  The extent to which this potential will be realized in practice, however, depends heavily on the mechanisms that allocate system resources between, and monitor and control workload execution processes across partitions. If these are ineffective, a high proportion of system capacity may be idle at any given time. Surges in workloads running in individual partitions may also create bottlenecks if additional capacity is not available in a timely manner.

  One of the core strengths of the System p platform is that highly effective management facilities are implemented at multiple levels, and closely integrated and optimized within the overall system environment. This is illustrated in figure 13.

Figure 13
Key IBM System p Virtualization Capabilities



For p570 servers, an additional capability has been announced by IBM called ***Shared Dedicated Capacity***, which provides options to donate excess cycles from dedicated processors to a shared pool

Higher-level management capabilities are provided by the ***IBM Systems Director*** family of tools and by IBM Tivoli enterprise management solutions.

System p management facilities also address a set of requirements that have proved important in encouraging large-scale adoption of virtualization: the ability to track usage of shared system resources in a manner that enables organizations to implement effective chargeback procedures.

AIX contains an extensive suite of system-level accounting features that enable administrators to collect statistics on usage of resources such as processors, memory, disks and adapters by application, process, transaction or other variables, or combinations of these.

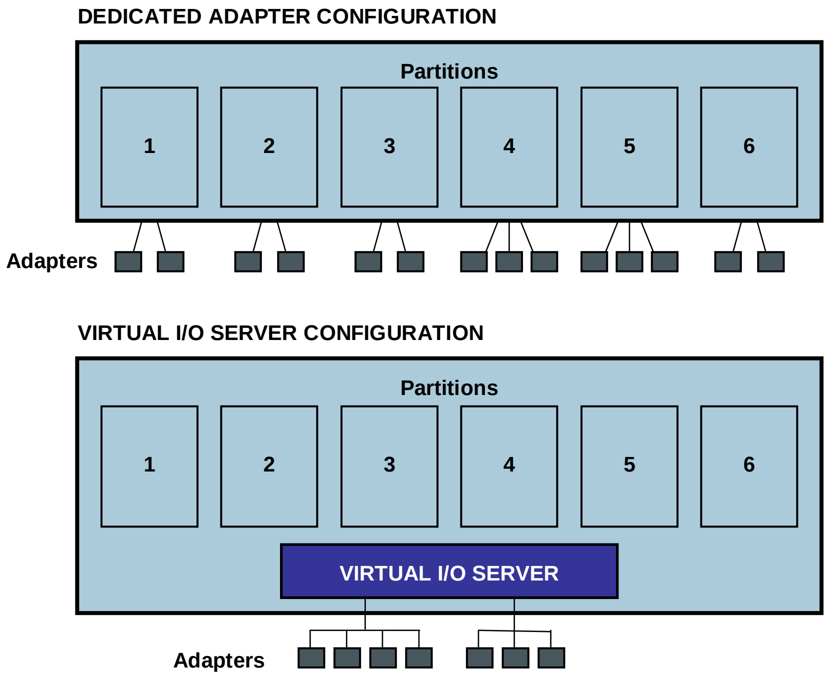
In many organizations, one of the reasons for continuing to employ dedicated servers is that this approach enables the business units that control the applications that run on these servers to track costs in a simple and non-controversial manner. If, however, costs of shared server platforms can be transparently and credibly determined, and allocated accordingly, a significant organizational obstacle to large-scale server consolidation may be removed.

Virtual I/O Server

A second key set of System p virtualization capabilities is implemented in the ***Virtual I/O Server***, an LPAR-based appliance that allows for the creation of virtual Ethernet and SCSI adapters.

The principal benefit of the Virtual I/O Server is that it allows operating system instances running in multiple partitions to share a common pool of LAN adapters, as well as Fiber Channel, SCSI and RAID devices. It is not necessary to dedicate adapters to individual partitions. As figure 14 illustrates, the number of physical adapters may be significantly reduced.

Figure 14
Dedicated Adapter and Virtual I/O Server Configurations: Examples



This approach may not be appropriate for all partitions. For most workloads, however, it offers the potential for significant savings in adapter and related LAN and SAN infrastructure costs.

Virtual I/O Servers interface to a virtual LAN or multiple virtual LANs that provide high-speed interconnection between LPARs within the System p environment. Virtual LAN capability further reduces network complexity and vulnerability, and may significantly reduce throughput times for interaction between LPAR-based systems, as well as for replication and other data movement processes.

A more basic form of I/O virtualization is provided for p570 servers through ***Integrated Virtual Ethernet***. This hypervisor-based facility enables sharing of Ethernet adapters without use of a Virtual I/O Server. It is designed for comparatively light workloads that do not require high levels of I/O throughput, such as those generated by development and test systems.

System-level Performance

The System p platform benefits from high levels of performance delivered by the IBM POWER RISC design. Current POWER5+ processors include dual- and quad-core units with clock speeds of 1.5 GHz to 2.3 GHz, while p570 servers employ dual-core 3.5 GHz, 4.2 GHz and 4.7 GHz processors.

Processor performance, however, is only part of the picture. System-level performance potential has been optimized at all levels of design and implementation – including microelectronics, module- and subsystem-level components, internal communications, I/O, and system-level hardware and software.

Key System p capabilities include highly effective compiler- and operating system-level performance acceleration, including chip symmetric multithreading; low levels of symmetric multiprocessing (SMP) overhead; and extensive system-level integration and optimization of all performance-related features.

Availability Optimization

Core RAS Features

The System p server benefits from a wide range of hardware- and software-based reliability, availability and serviceability features. These are designed to reduce the potential for unplanned outages, as well as to limit the frequency and duration of planned outages for tasks such as configuration upgrades and scheduled maintenance.

Core RAS features include the following:

- ***Basic capabilities*** include high levels of component reliability and redundancy, along with pervasive monitoring, diagnostic, and fault isolation and resolution facilities. These are built into processors, main memory, cache and packaging modules, as well as into all major hardware components. In many cases, multiple layers of protection and self-test are implemented.

  Key functionality is delivered by IBM-developed Chipkill and First Failure Data Capture (FFDC) technologies. Chipkill, which performs error checking for memory devices, is regarded as more reliable than conventional error correction code (ECC) techniques. FFDC employs thousands of embedded sensors that identify and report failures to a separately powered Service Processor, which also monitors environmental conditions.

  FFDC architecture forms the basis for predictive failure analysis functions that identify potential as well as actual failures throughout the system. In both cases, the Service Processor can automatically notify system administrators or contact an IBM Support Center directly ("call home" service) to report events requiring service intervention.

- ***Failure masking capabilities*** prevent outages in case failures do occur. For example, processors may be automatically disabled if they begin to malfunction, and standby processors may be activated without interrupting operations. Dynamic LPARs facilitate this process.

Concurrent maintenance ("hot plugging") functions and dynamic also reduce requirements for planned outages. LPARs reduce them further. Systems software and application upgrades, for example, may be performed in one LPAR while the original system continues operating in another. Software may be copied to and modified in LPARs. Backups may be executed concurrently with online processing.

System p RAS features are implemented at multiple levels, including LPARs (availability-related functions are built into the System p hypervisor), the Virtual I/O Server (for adapters) and system-level management facilities. These capabilities draw extensively on mainframe high availability design concepts and technologies.

Higher-level Capabilities

Higher-level capabilities include the following:

- ***Live Partition Mobility*** and ***Live Application Mobility*** are new capabilities that enable users to move partitions between systems with no application downtime, or limited downtime respectively.

  Live Partition Mobility is designed primarily for organizations that need to shut down a System p server for maintenance, upgrades and other functions, but cannot afford to take business-critical production systems offline. These may be simply transferred to another server, and returned after the original server is restarted.

  The only interruption of service would be due to network latency. If sufficient bandwidth was available, a delay of – at most – a few seconds could typically be expected. Live Application Mobility involves longer delays (e.g., 20 seconds), and would normally be employed for less availability-sensitive applications.

- ***High Availability Cluster Multiprocessing*** is IBM's principal solution for System p failover clustering. Organizations employing HACMP have achieved mainframe-class availability levels even for highly demanding, complex workloads.

  If ***HACMP/Extended Distance (HACMP/XD)*** is employed, failover may occur at distances of up to 300 kilometers if sufficient wide area network bandwidth is available.

These solutions provide a spectrum of capability designed to meet a wide range of user needs. Live Partition Mobility and Live Application Mobility, for example, do not require use of HACMP or other cluster solutions that may reduce costs while avoiding disruptions of service. It can be expected that users will continue to employ HACMP to protect against unplanned outages, and for disaster recovery.

Autonomic Functions

The System p platform benefits from one of the most advanced implementations of autonomic capability within the IBM product line. Autonomic functions, which are grouped by IBM into four categories – self-configuring, self-optimizing, self-protecting and self-healing – are summarized in figure 15.

Figure 15
IBM System p Autonomic Functions

SELF-CONFIGURING	SELF-OPTIMIZING	SELF-PROTECTING	SELF-HEALING
Virtual IP address IP multipath routing Microcode discovery Hot-swappable disks Hot-swap PCI Wireless/pervasive configuration TCP explicit congestion notification	Static LPAR Dynamic LPAR Workload manager enhancement Extended memory allocator RSCT technology PSSP cluster management	Kerberos V5 Authentication Self-protecting kernel SecureWay LDAP directory integration SSL management Digital Certificates Encryption	Multiple default gateways Automatic system hang recovery Automatic dump analysis EtherChannel automatic failover Processor failure detection & failover First failure data capture Chipkill ECC Memory Dynamic bit steering Memory scrubbing Automatic dynamic deallocation Electronic Service Agent

In addition to augmenting the effectiveness of System p RAS capabilities, autonomic functions contribute to higher levels of system administration productivity by streamlining and automating tasks that would otherwise require extensive manual intervention.

COST PICTURE

UNIX Server Costs

Basis of Calculations

The UNIX server cost comparisons presented in this report are based on three composite profiles of large company installations employing a variety of Hewlett-Packard, IBM, Sun Microsystems and other servers running variants of the UNIX operating system other than Linux.

Profiles were constructed using data on application portfolios, server bases, configurations, utilization and service levels, staffing and other variables supplied by 16 companies in the same industries and approximate size ranges, with generally similar business profiles.

Two sets of scenarios were then developed:

1. ***Conventional scenarios*** are based on data reported by the 16 companies, and are built around diverse multivendor server bases that include different technology generations and systems software versions. Conventional management and operating practices are employed.
2. ***Virtualized scenarios*** are for the same applications and workloads as conventional scenarios. Configurations for these scenarios were developed on a case-by-case basis within major system groups. Where appropriate, multiple applications, application instances, or both were configured on the same physical servers using LPARs, WPARs, or combinations of these.

All profile scenarios include database, application and, where appropriate, Web and intranet servers; and production systems as well as non-production instances for such functions as development, test, quality assurance and training. Clustered failover configurations are employed in conventional and virtualized scenarios for systems requiring high levels of availability.

Financial services company scenarios include a variety of custom and packaged software solutions. Manufacturing company scenarios are built primarily around SAP AG and complementary third-party applications. Retail company scenarios include a mix of applications from EXE Technologies, i2 Technologies, JDA Software, Oracle (including PeopleSoft and Retek) and other vendors.

Profile installations are summarized in figure 16.

Figure 16
UNIX Server Profiles Summary

COMPANY	FINANCIAL SERVICES	MANUFACTURING	RETAIL
Business Profile	$325 billion assets $16 billion revenues 12 million customers 1,500 branches 50,000 employees	$27 billion revenues Consumer products 30 manufacturing plants 25 distribution centers 50,000 employees	$30 billion revenues Specialty chain 900+ outlets 10 distribution centers 125,000 employees
Focus of Comparisons	64 major applications System groups: • Corporate systems • Retail banking • Commercial banking • Financial services • Online systems • Intranet infrastructure	27 major applications System groups: • Core ERP systems • Supply chain management • Product management • Business intelligence • eProcurement • CRM	35 major applications System groups: • Corporate systems • Supply chain management • Logistics & transportation • Marketing, sales & service • Business intelligence • eCommerce
CONVENTIONAL SCENARIOS			
Servers	**Hewlett-Packard** Superdome, rx8640, rx8620, rx7620, rx6600, rx5670, rx4640, rx2620, rx2600, rp8420, rp8400, N4000, various Hewlett-Packard (Alpha) GS80, ES45, ES40 **IBM** p690, p680, p670, p660, p650, p640, p630, p615, p610, p5-570, p5-550, p5-510, p5-505, p5-185 **Sun Microsystems** E25K, E10K, E6800, E4900, E4800, E4500, E2900, E450, V890, V880, V490, V480, V440, V250, V240, V40Z, V20Z, X4600, X4200, X4100, various **Silicon Graphics** Altix 450 Total: 285 servers	**Hewlett-Packard** Superdome, rx7640, rx7620, rx6600, rx4640, rx2620, rp8420, rp5470, rp440 **IBM** p690, p670, p660, p650, p630, p615, p5-570, p5-550, p5-520, p5-510 **Sun Microsystems** E20K, E15K, E4800, V890, V880, V490, V480, V250, V40Z, V20Z, X4200, X4100, various Total: 85 servers	**Hewlett-Packard** rx8640, rx8620, rx7620, rx5670, rx4640, rx2620, rx2600, rx1600, rp8400 **IBM** p690, p680, p670, p650, p630, p615, p5-570, p5-550, p5-510, RS/6000 **Sun Microsystems** E15K, E6800, E4900, E4800, E4500, V890, V490, V480, V440, V250, V240, V40Z, V20Z, X4600, X4200, various Total: 112 servers
Personnel	34 FTEs	14 FTEs	16 FTEs
VIRTUALIZED SCENARIOS			
Servers	13 x POWER6 p570 4 x p5-550Q, 3 x p5-520Q 2 x p5-510Q, 24 x p5-505Q 22 x p5-505, 12 x p5-185 Totals: 80 servers 42 LPARs 82 WPARs	5 x POWER6 p570 5 x p5-550Q, 2 x p5-520Q 13 x p5-505Q, 5 x p5-505 Totals: 30 servers 28 LPARs 41 WPARs	7 x POWER6 p570 2 x p5-550Q, 3 x p5-520Q 14 x p5-505Q, 15 x p5-505 8 x p5-185 Totals: 49 servers 34 LPARs 56 WPARs
Personnel	17 FTEs	9 FTEs	10 FTEs

Cost Breakdowns

For conventional scenarios, calculations are for operating costs only. These include hardware maintenance, update and support subscriptions for systems and database software, along with personnel and facilities costs. For virtualized scenarios, calculations are for operating costs, which include the same components, along with costs of hardware and systems software acquisition for new System p servers.

Detailed breakdowns of calculations are shown in figure 17. Costs for p570 servers in virtualized scenarios and equivalent servers in conventional scenarios are shown separately.

Figure 17
Five-Year UNIX Server Costs Detail

COMPANY	FINANCIAL SERVICES	MANUFACTURING	RETAIL
CONVENTIONAL SCENARIOS – ALL SERVERS ($000)			
Maintenance	10,722.9	6,401.4	2,510.4
Systems software support	2,370.6	1,366.2	1,371.6
Database software support	25,264.2	9,922.3	8,714.3
Personnel	16,529.0	6,806.1	7,778.4
Facilities	1,393.9	584.3	498.1
TOTAL OPERATING COSTS	**56,280.6**	**25,080.3**	**20,872.8**
VIRTUALIZED SCENARIOS – ALL SERVERS ($000)			
Hardware	3,137.3	1,445.2	1,123.0
Systems Software	340.2	148.9	98.4
Acquisition costs (Subtotal)	***3,477.5***	***1,594.1***	***1,221.4***
Maintenance	931.2	401.7	362.6
Systems software support	1,094.3	585.0	348.6
Database software support	3,590.6	1,480.1	1,221.5
Personnel	8,264.5	4,375.3	4,861.5
Facilities	419.7	206.1	167.7
Operating costs (Subtotal)	***14,300.3***	***7,048.2***	***6,961.9***
TOTAL	**17,777.8**	**8,642.3**	**8,183.3**
CONVENTIONAL SCENARIOS – p570 EQUIVALENTS ($000)			
Maintenance	7,960.8	5,095.6	2,142.7
Systems software support	1,736.6	1,105.4	1,019.6
Database software support	21,334.3	9,922.3	8,293.3
Personnel	9,723.0	3,889.2	4,861.5
Facilities	920.4	465.2	340.6
TOTAL OPERATING COSTS	**41,675.1**	**20,477.7**	**16,657.7**
VIRTUALIZED SCENARIOS – p570 SERVERS ($000)			
Hardware	2,693.5	1,319.1	880.6
Systems software	248.7	140.2	76.4
Acquisition costs (Subtotal)	***2,942.2***	***1,459.3***	***957.0***
Maintenance	785.8	356.7	268.1
Systems software support	774.5	529.0	239.9
Database software support	2,861.4	1,480.1	888.0
Personnel	4,375.3	1,944.6	1,944.6
Facilities	216.1	83.6	84.2
Operating costs (Subtotal)	***9,013.1***	***4,394.0***	***3,424.8***
TOTAL	**11,955.3**	**5,853.3**	**4,381.8**

For all scenarios, database costs are for Oracle 10g. Calculations do not include initial license costs. It is assumed that these are covered by existing customer licensing arrangements.

All costs, other than for personnel and facilities, are based on discounted vendor "street" prices for the products and services included in scenarios. Personnel costs are for the numbers of FTE UNIX system administrators shown in figure 16.

Facilities costs are for data center occupancy, power and cooling equipment, electricity and related operating costs, and are based on U.S. industry averages and norms. Additional detail on the basis of these calculations may be found in the Server Cost Calculations section this report.

Cost Factors

Lower five-year operating costs for virtualized scenarios relative to conventional scenarios in profile-based comparisons are due to the factors summarized in figures 18 (all servers) and 19 (p570 servers and conventional equivalents).

Figure 18

UNIX Server Comparisons: Factors Resulting in Lower Operating Costs for Virtualized Scenarios (All Servers)

Categories	Savings	Factors
Maintenance	86% - 94%	Fewer, newer servers reduce maintenance contract costs.
Software support	82% - 84%	Fewer software copies & CPUs result in lower license, update & support costs.
Personnel	36% - 50%	Fewer physical servers, reduced diversity & improved automation reduce system administration-related personnel costs.
Facilities	65% - 70%	Fewer physical servers, smaller footprints, & greater energy efficiency reduce data center occupancy, power & cooling costs.
Average	**72%**	

Figure 19

UNIX Server Comparisons: Factors Resulting in Lower Operating Costs for p570 Servers and Conventional Equivalents

Categories	Savings	Factors
Maintenance	87% - 93%	Fewer, newer servers reduce maintenance contract costs.
Software support	82% - 88%	Fewer software copies & CPUs result in lower license, update & support costs.
Personnel	50% - 60%	Fewer physical servers, reduced diversity & improved automation reduce system administration-related personnel costs.
Facilities	75% - 82%	Fewer physical servers, smaller footprints, & greater energy efficiency reduce data center occupancy, power & cooling costs.
Average	**79%**	

Linux Server Costs

Basis of Calculations

The Linux server cost comparisons presented in this report are based on the same company business profiles employed for UNIX server comparisons. Installation profiles, however, are different. Application portfolios, server bases and staffing levels were developed using data from 21 companies, including in some cases companies that did not supply UNIX server data.

Linux installation profiles are summarized in figure 20.

Figure 20
Linux Server Installation Profiles Summary

COMPANY	FINANCIAL SERVICES	MANUFACTURING	RETAIL
Applications	Market value & risk analytics Equities trading, market data Mortgage loans, antifraud Departmental applications Content management Software development & test Intranet applications Intranet infrastructure File serving, various	SAP xApps Computer aided design Departmental applications Software development & test File serving Web infrastructure, various	Internal portal, e-procurement Promotional e-mail Departmental applications Office applications POS software development File serving, fax serving Network management Web infrastructure, various
CONVENTIONAL SCENARIOS			
Servers	Dell, HP, IBM, various 165 servers	Fujitsu, HP, IBM 39 servers	Dell, HP, IBM, various 111 servers
Personnel	7 FTEs	1.25 FTEs	3 FTEs
VIRTUALIZED SCENARIOS			
Servers	POWER6 p570 8 x 3.5 GHz POWER6 p570 4 x 4.2 GHz POWER6 p570 4 x 3.5 GHz 2 x p5-550Q, 3 x p5-520Q 197 LPARs	POWER6 p570 4 x 4.2 GHz p5-520Q 44 LPARs	2 x POWER6 p570 4 x 3.5 GHz 2 x p5-550Q 122 LPARs
Personnel	2 FTEs	0.5 FTE	1.25 FTEs

As for UNIX server comparisons, two sets of scenarios were developed:

1. ***Conventional scenarios*** are based on data reported by the 21 companies, and are built around diverse bases consisting primarily of Dell, Hewlett-Packard, IBM and (for the manufacturing company) Fujitsu Intel-based servers.

   Bases include different hardware technologies – ranging from Pentium 4 processor-based servers installed in 2002 to current generation quad-core Xeon processor-based servers – and different Linux distributions and versions. Conventional Linux server management and operating practices are employed.

2. ***Virtualized scenarios*** are for the same applications and workloads as conventional scenarios. As for UNIX server comparisons, it is assumed that applications are deployed on System p servers in native mode in a manner that effectively exploits the potential of System p virtualization, and that management and operating practices are supportive of this potential.

   Configurations for these scenarios were developed on a case-by-case basis using LPAR-equipped System p servers to consolidate application and operating system instances.

As for UNIX server comparisons, database costs are for Oracle 10g, and calculations do not include initial license costs for copies of this. Linux systems software costs are not included.

Cost Breakdowns

Detailed breakdowns of Linux server costs are shown in figure 21. Cost categories are the same as those employed for UNIX server comparisons.

Figure 21
Five-Year Linux Server Costs Detail

COMPANY	FINANCIAL SERVICES	MANUFACTURING	RETAIL
CONVENTIONAL SCENARIOS ($000)			
Maintenance	276.6	82.2	126.5
Database software support	210.7	235.8	89.3
Personnel	3,073.5	548.8	1,317.2
Facilities	307.6	104.7	155.4
TOTAL OPERATING COSTS	**3,868.4**	**971.5**	**1,688.4**
VIRTUALIZED SCENARIOS ($000)			
Acquisition costs	**376.3**	**109.2**	**182.2**
Maintenance	112.2	32.3	31.9
Database software support	135.1	45.0	67.6
Personnel	878.1	219.5	548.8
Facilities	43.7	8.4	12.8
Operating costs (Subtotal)	***1,169.1***	***305.2***	***661.1***
TOTAL	**1,545.4**	**414.4**	**843.3**

Cost Factors

Lower five-year operating costs for virtualized relative to conventional scenarios are in most cases due to the same factors as for UNIX servers, and are summarized in figure 22.

Figure 22
Linux Server Comparisons: Factors Resulting in Lower Operating Costs for Virtualized Scenarios

Categories	Savings	Factors
Maintenance	59% - 75%	Fewer, newer servers reduce maintenance contract costs.
Database support	24% - 81%	Fewer software copies & CPUs result in lower license, update & support costs.
Personnel	58% - 71%	Fewer physical servers, reduced diversity & improved automation reduce system administration-related personnel costs.
Facilities	86% - 92%	Fewer physical servers, smaller footprints, & greater energy efficiency reduce data center occupancy, power & cooling costs.
Average	**67%**	

Costs of Downtime

Downtime Impacts

Although IT costs tend to dominate industry discussion, the bottom-line impact of outages may be equally, if not more significant.

For example, supply chain disruptions can result in inventory shortages, missed deliveries, production and distribution bottlenecks, and other impacts. Where supply chains are tightly integrated, even brief outages can generate "cascading" effects that spread rapidly across the entire organization, and whose effects may continue to be felt long after service is restored.

Experiences with Internet commerce have shown that online outages can result in lost sales. This is not only because customers cannot make purchases at a specific time, but also because they cannot obtain information on a company's products or services, and may look elsewhere. A customer who experiences an outage may never return, or may be more likely to divide their purchases between vendors.

Even where customer defections cannot be tied to a specific negative experience, outages contribute to overall levels of satisfaction or dissatisfaction, which affect attrition rates. These may have important bottom-line impact, particularly if lost sales or profit is measured in terms of customer lifetime value (CLV) and equivalents.

The impact on outages on productivity may also be significant. Internet and intranet applications, along with messaging and workflow systems, and a wide range of departmental and individual user tools have become deeply embedded into the business processes of many companies. Interruptions of service may be highly disruptive and, quite literally, can cause work to stop.

Effective virtualization of system resources can contribute to fewer outages. Not only risks of unplanned outages, but also planned downtime for preventative maintenance, configuration changes, software upgrades and other functions may be significantly reduced.

UNIX Server Comparisons

Costs of downtime for the financial services company represent lost operating profit, meaning net revenue after deduction of personnel, occupancy, equipment and other overheads. For the manufacturing and retail companies, the comparable metric is lost gross profit, meaning profit net of cost of goods sold (COGS), but before deduction of selling, general and administrative (SG&A) and other expenses.

Costs were determined using industry- and company-specific values assuming consistent levels of availability and cost structures over a five-year period.

For the financial services company, costs of downtime consists primarily of lost transaction fees and CLV for outages affecting online banking, brokerage and other customer self-service systems, along with CRM systems and Internet infrastructure servers supporting these.

Costs of downtime for the manufacturing company were calculated based primarily on outages affecting the company's core SAP ERP, supply chain management (SCM) and procurement systems. For the retail company, costs were calculated based on outages affecting the company's supply chain systems as well as Internet marketing, sales and customer service systems.

The impact of outages on the retail company's Internet systems includes the effects not only of lost online sales, but also of storefront sales lost because customers are unable to obtain information on products, promotions, pricing, availability, store locations and other subjects.

While online orders account for only 3 percent of total sales volume, more than 20 percent of storefront purchases are influenced by information obtained via the Internet.

Costs of downtime are greatest for the manufacturing company. This is because the company's bottom line depends heavily upon supply chain performance, and because all of its major business systems are deployed on UNIX servers. In the financial services and retail company profiles, certain business critical systems are deployed on mainframes and other platforms that are not addressed by this report.

Linux Server Comparisons

For all Linux server scenarios, costs are for employee productivity loss. This was measured in terms of (1) idle time, reduced productivity or both for periods when applications were not available to users during working hours and (2) reduced productivity following an outage; e.g., a one-hour outage causing a 60 percent productivity reduction might be followed by a two-hour period in which a 20 percent productivity loss occurs.

Productivity loss calculations were developed for each user community supported by Linux applications in the profile companies. Costs are based on industry median salaries and benefits for large U.S. financial services, manufacturing and retail companies for the occupational groups using applications.

For example, costs of downtime for the financial services company included calculations based on median salaries and benefits for securities traders, risk analysts, mortgage loan specialists, security professionals, internal auditors and others. Manufacturing company equivalents included multiple categories of managers, analysts, engineers, technicians, salespeople and other professionals, and in some cases administrative, production and logistics personnel.

Retail company equivalents included multiple categories of managers, analysts, buyers and other professionals, along with administrative, distribution and, in some cases, in-store personnel. Software developers were included in calculations for all three companies.

Productivity losses were quantified based on the number of individuals using each application who would be affected by outages, their average remuneration per hour, and percentage values for reductions in their productivity during and subsequent to outages.

Overall costs of downtime are lowest for the financial services company. This is because Linux applications were generally more specialized, and user communities were smaller than for the manufacturing and retail companies.

SERVER COST CALCULATIONS

General Approach

The cost and benefit implications of any IT initiative can be measured in a number of different ways. In most organizations, costs, benefits and – where this approach is employed – return on investment (ROI) values are determined on a project-by-project basis, usually over a three-year period.

While such techniques may be useful, they are inherently limited. For example, an application-specific project may deliver high levels of ROI while contributing to the fragmentation and complexity of an organization's overall IT infrastructure. Equally, inefficiencies whose effects are minor, or at least acceptable, for small-scale deployments may have significantly larger cumulative impacts at the enterprise level.

Short-term measurements may also obscure longer-term cost drivers. Growth in server bases has often been an incremental process whose implications have not been visible on a year-by-year basis. This has made it difficult to determine baselines for efficiency measurement. Even an organization that is more efficient than it was a year or two before may still fall far short of what it could achieve.

For these reasons, it was decided for this report to base comparisons on organization-wide application portfolios, workloads and server bases, and personnel supporting these; and to focus on five-year costs.

Configuration Sizing

In translating configurations employed in conventional scenarios into System p configurations employed in virtualized scenarios, the following approach was employed.

System p configurations delivering performance equivalent to conventional scenario servers were developed for each of the main applications in profile companies. Configurations were developed for production and non-production instances.

Further calculations were then undertaken to allow for the effects of virtualization. Nominal configuration requirements were first determined for groups of instances that were deployed on separate servers, where these represented realistic candidates for consolidation onto a single System p physical server.

An overall utilization value reflecting realistic potential consolidation efficiencies was then assigned to each group of instances, and allowance was made for other factors affecting System p capacity requirements. The resulting configuration was then rounded to next largest capacity increment offered by IBM; e.g., a nominal configuration of 3.3 x 4.2 GHz cores with 12.6 GB of RAM was rounded to a 4 x 4.2 GHz System p model with 16 GB of RAM. Other hardware components were configured similarly.

Other hardware components were configured similarly. Allowance was made for other System p virtualization capabilities including use of Virtual I/O Servers and Integrated Virtual Ethernet.

Cost Components

Maintenance, software update and support, and (for virtualized scenarios) hardware and software acquisition costs were calculated based on discounted vendor U.S. list prices. For database software, Oracle 10g multi-core pricing was employed where appropriate.

For UNIX server comparisons, systems software stacks for conventional scenarios include operating systems and, where appropriate, additional facilities providing functionality equivalent to that incorporated in AIX. Systems software stacks for virtualized scenarios include AIX and – for clustered configurations – HACMP facilities.

For Linux server comparisons, software costs are for Oracle 10g database support and update subscriptions only.

Personnel costs for UNIX and Linux system administrators were calculated using annual average salaries of U.S. $77,289 and $69,804 respectively, increased by 25.8 percent to allow for bonuses, benefits, training and other items.

Staffing levels are based on user-reported data. Definitions of system administration tasks and disciplines tend, however, to vary between organizations. The definitions employed in this report may not correspond to those employed in any individual user organization.

Facilities costs for conventional and virtualized scenarios for UNIX and Linux comparisons include costs for servers as well as power and cooling equipment.

Server facilities costs were calculated using vendor specifications for electricity consumption and footprints for server models and configurations included in scenarios. Costs include data center occupancy (calculations were based on industry standard rack mount units and service clearances for these, plus allowance for inactive areas) and electricity consumption.

Facilities costs also include hardware acquisition, maintenance, occupancy (including service clearances and inactive areas) for, and electricity consumption by power and cooling equipment. Configurations of this equipment are appropriate for overall electricity consumption and heat generation by server bases in scenarios. Cost calculations were based on U.S. specifications and discounted list prices for appropriate models from leading vendors.

Data center occupancy costs for servers and power and cooling equipment are based on a conservative assumption for annual average cost per square foot for existing facilities (i.e. costs do not include new facilities construction), while electricity costs for both are based on a conservative assumption for average price per kilowatt/hour. Both assumptions are for U.S. costs.

PSL03019-USEN-01

ABOUT THE INTERNATIONAL TECHNOLOGY GROUP

ITG sharpens your awareness of what's happening and your competitive edge
. . . this could affect your future growth and profit prospects

The International Technology Group (ITG), established in 1983, is an independent research and management consulting firm specializing in information technology (IT) investment strategy, cost/ benefit metrics, infrastructure studies, deployment tactics, business alignment and financial analysis.

ITG was an early innovator and pioneer in developing total cost of ownership (TCO) and return on investment (ROI) processes and methodologies. In 2004, the firm received a Decade of Education Award from the Information Technology Financial Management Association (ITFMA), the leading professional association dedicated to education and advancement of financial management practices in end-user IT organizations.

The firm has undertaken more than 100 major consulting projects, released approximately 160 management reports and white papers, and delivered nearly 1,800 briefs and presentations to individual clients, user groups, industry conferences and seminars throughout the world.

Client services are designed to provide factual data and reliable documentation to assist in the decision-making process. Information provided establishes the basis for developing tactical and strategic plans. Important developments are analyzed and practical guidance is offered on the most effective ways to respond to changes that may impact or shape complex IT deployment agendas.

A broad range of services is offered, furnishing clients with the information necessary to complement their internal capabilities and resources. Customized client programs involve various combinations of the following deliverables:

Status Reports	In-depth studies of important issues
Management Briefs	Detailed analysis of significant developments
Management Briefings	Periodic interactive meetings with management
Executive Presentations	Scheduled strategic presentations for decision-makers
Email Communications	Timely replies to informational requests
Telephone Consultation	Immediate response to informational needs

Clients include a cross section of IT end users in the private and public sectors representing multinational corporations, industrial companies, financial institutions, service organizations, educational institutions, federal and state government agencies as well as IT system suppliers, software vendors and service firms. Federal government clients have included agencies within the Department of Defense (e.g. DISA), Department of Transportation (e.g. FAA) and Department of Treasury (e.g. US Mint).

International Technology Group

ITG

4546 El Camino Real, Suite 230
Los Altos, California 94022-1069
Telephone: (650) 949-8410
Facsimile: (650) 949-8415
Email: info-itg@pacbell.net