MANAGEMENT BRIEF

e-TRANSACTIONS IN FINANCIAL SERVICES

Impact of Volume Growth on IT Infrastructures and Costs

International Technology Group

885 North San Antonio Road, Suite C
Los Altos, California 94022-1305
Telephone: (650) 949-8410
Facsimile: (650) 949-8415
Email: itg1@pacbell.net

TABLE OF CONTENTS

List of Figures

Bottom Line

By 2005, transaction workloads in financial services will be more than four times larger than today. By 2010, they will be more than 10 times larger. More frequent and complex interactions with customers, mergers, acquisitions and economic growth are already pushing transaction volumes to unprecedented levels. The Web impact will be even greater.

How will this growth be handled? In particular, how will extremely large, volatile workloads be supported while maintaining de facto 100% availability, fast response time and security? These remain central requirements for conventional business-critical systems. They are even more important on the Web, where customers are only a few clicks away from competitors.

E-business strategies must be launched or relaunched, new systems must be deployed, legacy applications must be replaced or upgraded. Mainframe systems can play a major role in meeting these challenges. Key arguments include:

- ***Volume experience***. Volume and service quality are closely related. Supporting large, volatile workloads with high service quality is a great deal more difficult than in lower-volume, more predictable environments. The critical question may thus not be simply whether any given platform can deliver “99.9 percent” or “7 x 24” availability — it is whether this level can be sustained under heavy workloads and business-critical operating conditions.

  The largest mainframe-based core banking system in the United States currently handles transaction volumes 78 times greater than the largest system on any other platform. In brokerage and insurance, mainframe systems are 35 and 29 times larger respectively. Availability is significantly higher than for UNIX and NT servers handling smaller, less demanding workloads.

- ***Cost picture***. Results from companies surveyed for this document indicate that in conventional electronic funds transfer (EFT) applications five-year hardware, software, maintenance and personnel costs for fault-tolerant servers are 1.7 to 2.5 times higher than for S/390 systems, for comparable levels of availability, response time and security. Disparities are consistent with results for other industries and types of application.

  Costs of ownership are, however, only part of the picture. Outages also result in lost revenues and customer lifetime value (CLV). In a retail brokerage example, the five-year cost of outages for UNIX servers (in this case, Sun Microsystems E10000s) is close to $438 million in lost revenues and CLV. Comparable mainframe losses are less than $4 million. Five-year costs of ownership for mainframe- and UNIX-based systems are $108 million and $230 million respectively.

It is often assumed that legacy systems are an obstacle to success in e-business. But the reverse may be the case. If existing mainframe-based infrastructures are leveraged effectively, the strengths of this architecture will represent a critical edge against less well-equipped competitors.

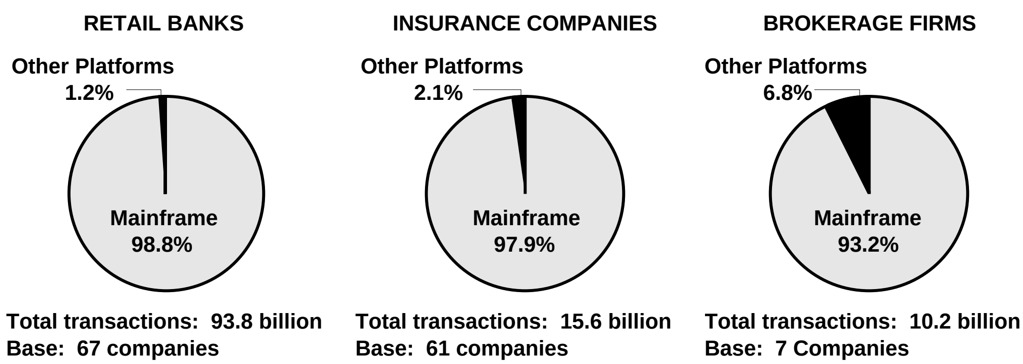
INFRASTRUCTURES

As of yearend 1999, financial services companies employed more than 459 billion instructions per second (BIPS) of mainframe processor capacity, and ran core business processes with more than 4.8 billion lines of mainframe code. Mainframe databases housed more than 435 terabytes (TB) of data, and a further 10.8 petabytes were maintained in archival form.

It is thus not necessary to build new infrastructures to meet the challenges of e-business. They are already in place, and optimized for high-volume, business-critical operations. They may be expanded and enhanced without the costs, delays and disruptions of radical technology change.

Among U.S.-based banks, brokerages and insurance companies with more than $5 billion in assets, the vast majority of all retail transactions during 1999 — including more than 90% of transactions occurring via the Web — were, as figure 1 shows, processed by mainframe-based systems.

Figure 1
Business Transaction Volumes by Platform:
Companies with Assets of More Than $5 Billion



Source: *Strategies for Volume* Management Brief, International Technology Group

Among companies with more than $50 billion in assets, mainframe-based systems handled close to 100 percent of core transactions in retail banking, 94.6 percent in brokerage, and 98.6 percent in insurance. Numbers are for business transactions, including core retail banking transactions; policy issue and updates, claims, billing and payments in insurance; and retail brokerage transactions.

In most companies, UNIX and NT servers are employed in a variety of distributed, departmental and specialized roles. Tandem and equivalent fault-tolerant platforms also act as front-end switches. But even when multiple-tier architectures are employed, core business logic, data and back-end processing remain mainframe-based.

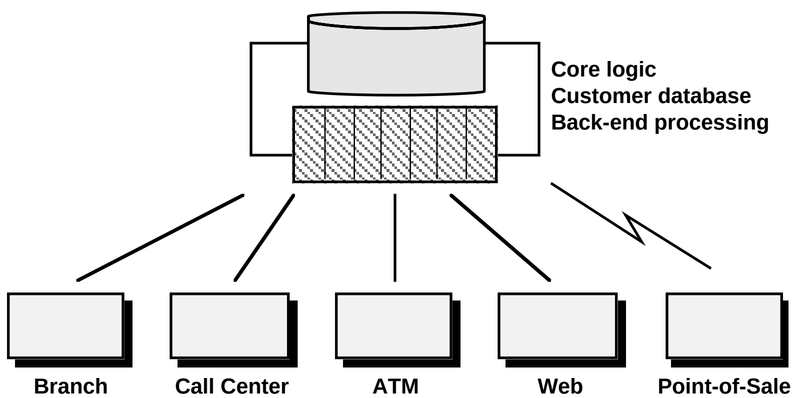
Each transaction at the customer interface typically generates at least 20 to 50 back-end transactions. Accounts must be updated, reconciliations and settlements performed, claims processed, bills and statements generated, payments posted, and so on. Back-end workloads are thus a great deal heavier than their front-end counterparts.

Which means that, in practice, mainframe-based infrastructures are the focal point of e-business architectures. Even if it were technically feasible to replace core mainframe-based systems — which is far from clear — it would be a prohibitively difficult and expensive process. Nor, from a business point view, would it make sense to do this.

Most companies will not operate uniquely via the Web. Call centers and financial advisors will remain critical, along with branches, ATMs and point-of-sale locations in banking, and agents and brokers in insurance. There is growing evidence that the ability to leverage customer relationships across multiple channels will be a critical competitive edge.

This has major implications for IT strategy. Core transaction-processing systems must scale across all major channels, data must be consolidated, common back-end processes must be executed, and strategies and solutions for customer relationship management (CRM) must address all points of contact. Figure 2 illustrates a retail banking example of this effect.

Figure 2
E-business Architecture: Retail Banking Example



For most of the 1990s, companies focused on distributed, departmental and specialized applications while neglecting back-end infrastructures that were, and remain, predominantly mainframe-based. This may have been justified to meet what were — at the time — higher competitive priorities. But e-business changes the picture.

It is not sufficient merely to have a presence on the Web. Core processes of product and service delivery must remain competitive. The key variables of competitiveness will be the combined performance and functionality of all systems throughout the enterprise, and the effectiveness with which they are integrated. Success or failure in e-business will be determined as much, if not more, by back-end systems as by the Web interface.

Under any scenario, mainframe-based infrastructures will play a central role in e-business strategies. There is the option to make that role a proactive source of competitive advantage.

“Mainframe,” in this context, refers to S/390 architecture-compatible systems.

SERVICE QUALITY

Availability

Basics

From a business perspective, availability may be defined as efficient access to system resources at all times when required by users, or by business-critical application processes running without direct user intervention. In practice, it means not only minimizing risks of outages, but also shielding their effects in such a way that no interruption in service, or degradation of service quality is experienced.

Two types of system outage must be addressed:

1. ***Unplanned outages***. Application errors, hardware and systems software failures are typical causes of unplanned outages. Also contributing to the problem are mistakes by IT personnel, including operators, system and database administrators and (indirectly) application programmers.

   Another, increasingly common cause is inability to handle growth. Workload surges are, for example, the most frequent reason for major Web site outages. Systems become overloaded and crash, or must be taken off-line for capacity upgrades, or both. Outages thus tend to occur at times when very large numbers of users are online — i.e., when the business impact is most severe.

   The critical requirement is thus not necessarily for generic “7 x 24” or “99.9 percent” availability. It may be to maintain de facto 100 percent availability under exceptionally heavy workload conditions.

   Failure to process customer transactions in a timely manner may result not only in lost revenues and customers, but also in exposure to lawsuits and regulatory action. Brokerages are already vulnerable to these, and it is likely that this will increasingly be the case for banks, insurers and other financial services companies operating via the Web.

2. ***Planned outages***. Preventative maintenance, hardware and software upgrades, and other tasks may require systems to be shut down. In most companies, however, 95 percent or more of off-line time is spent running batch application processes, along with backups, replication and other data movement processes in batch mode.

   Batch processing thus becomes a key component of overall availability. Delays in completing overnight jobs may impact the next day’s online operations. In large companies, high levels of batch throughput will be required, along with ability to move large volumes of data between processors, storage and networks.

Maintaining high levels of availability is not a simple or inexpensive process. There is an industry rule of thumb that moving from, say, 99.9 to 99.99 percent availability is a task ten times more difficult than moving from 99 to 99.9 percent. Costs and difficulties increase with even higher levels. While actual ratios will vary, the principle applies to any business-critical transaction-processing system.

Workload Characteristics

Availability cannot be measured in isolation. Costs and difficulties are heavily dependent on transaction characteristics and volumes, along with database utilization, response time and other variables that impact system loading.

It is thus not useful to know that a platform is capable of, say, 99.999 percent availability without details of the applications, workloads and service quality for which this level is realized. A system that functions well in a stable, low-volume environment may not perform as well with larger, more demanding workloads.

For planning purposes, effective quantification of all variables that will affect availability thus becomes essential. Figure 3 illustrates this approach using data from a representative financial services company.

Figure 3
Key Availability Planning Variables

CURRENT	Normal Workload	Peak Workload (Predictable)	Peak Workload (Unpredictable)
Transaction complexity *	1 business = 5 computer	1 business = 5 computer	1 business = 5 computer
Transactions per second	70	150	300 (x2)
Database size	80 GB	80 GB	80 GB
Availability target	100%	100%	100%
Acceptable fallover time	Subsecond	Subsecond	Subsecond
Response time at host	Subsecond	Subsecond	Subsecond
IN FIVE YEARS	**Normal Workload**	**Peak Workload (Predictable)**	**Peak Workload (Unpredictable)**
Transaction complexity *	1 business = 10 computer	1 business = 10 computer	1 business = 10 computer
Transactions per second	200	450	2,250 (x5)
Database size	300 GB	300 GB	300 GB
Availability target	100%	100%	100%
Acceptable fallover time	Subsecond	Subsecond	Subsecond
Response time at host	Subsecond	Subsecond	Subsecond

* The number of computer transactions generated by each business transaction.

In this example, transaction complexity doubles over a five-year period, reflecting changes in the company's portfolio of services, and in the frequency and sophistication of its interactions with customers. By 2005, the number of computer transactions generated by each business transaction has doubled.

Peak transactions per second and response time are based on sustained throughput during five-minute windows at times of maximum loading. Workload spikes within this window would also normally be quantified, and allowance would be made for the speed at which unpredictable peaks developed; e.g., a 50% increase might occur over an eight-hour period, or workloads might double within a few hours.

The impact of changes in variables is cumulative. In this example, the challenges of maintaining availability would be 10 to 20 times greater by the end of 2005 than in 2000.

Back-end Workloads

In most companies, UNIX- and NT-based servers handle workloads that are a fraction of their back-end counterparts. Although the overall volume of, say, a branch automation network may be large, the workload handled by any single server will be a small fraction of this. It is, moreover, comparatively easy to add new servers to meet branch or departmental growth. The real challenge will be how well back-end systems and databases scale.

Fault-tolerant servers typically handle large transaction volumes. But workloads are still less demanding. Servers are used to switch rather than process transactions, database sizes are relatively small, and batch is executed on mainframe systems. High levels of front-end availability may thus not be representative of what would occur under heavier workloads.

Batch routinely represents 60 to 80 percent of back-end workloads. Use of platforms whose batch performance is generally weaker than that of mainframes would tend to increase the duration of planned outages. More limited capabilities for workload management and movement also impact overall availability levels.

Fallover Efficiency

While risks of unplanned outages may be minimized, it is unlikely that they will be eliminated altogether. Use of clustering to shift workloads to standby capacity thus represents a critical part of the availability equation.

Effective fallover has, however, not proved to be an easy process. Major Web site outages have routinely occurred when two or even three layers of servers were in place. This was, again, most common when systems were experiencing rapid workload growth.

The fallover process is sensitive to volume. A cluster may, for example, be able to transfer a five transactions per second (tps) workload effectively. But at, say, 50 or more tps, the process is a great deal more demanding. If the system is not properly equipped, risks of disruption will be significantly larger.

Response Time

It is not sufficient simply to keep a system online. If response time is inadequate, or key application features do not function effectively, business damage may be even greater than if an outage occurs. This will particularly be the case in that, unlike outages, the effects are often not immediately visible.

One popular Web site, for example, experienced more than 73 hours of outages during a 12-month period in 1998 and 1999. But over this period, users also experienced at least 400 hours when response time was unacceptably slow, or certain system functions were not available, or both. Competitors were entering the market on a large scale during this period, and customers inevitably turned to them.

Consumer patience erodes when Web site delays extend into the five- to ten-second range, and risks that customers will go elsewhere increase exponentially in proportion to wait time. If the experience is repeated, odds of customer loss increase even further. There may also, as with outages, be significant legal and regulatory implications.

In contrast, the normal consumer expectation for response time at an automated teller machine or point-of-sale location is two to five seconds, and this range will probably become the Web norm. Widespread use of DSL and cable modems will reduce network latencies, making server-level response time an increasingly important competitive variable.

As a general principle, if an application is properly optimized, halving response time means doubling configuration size. The corollary is that significantly more capacity is required to maintain response time under peak workloads. Failure to allow for this effect will result in slowdowns, and may (because server capacity is saturated more rapidly) cause servers to crash. Costs will also be higher.

Security

Security is becoming an increasingly important issue in most areas of e-business. It is inevitably critical in financial services. Less generally realized, however, is that provision of high levels of security has a major impact on workloads.

At the most basic level, authorization of credit card, ATM and debit/credit transactions requires significant processor capacity. This will particularly be the case if positive verifications, or large databases or both are employed. Workloads will also tend to be more I/O-intensive; i.e., similar to conventional mainframe transaction-processing systems.

The impact of other techniques may be even greater. Software-based encryption, for example, routinely consumes four or five times more server capacity than the application it supports. Again, these effects have major implications for performance, availability and costs of e-business systems.

Conclusions

In terms of transaction volumes, peak loading and availability, response time and security requirements, the profile of the large-scale e-business system corresponds more closely to the characteristics of S/390 architecture than to any other platform.

Transaction-processing workloads are inherently I/O-intensive. Compared to UNIX and NT servers, the greater mainframe I/O bandwidth results in higher levels of performance under high-volume workloads. Response time can also be maintained more effectively, and data transfer processes occur more rapidly, reducing the duration of planned outages.

S/390 systems also employ more efficient hardware-based encryption techniques. Because these are integrated into the core system, the workload and cost impact is significantly less than for outboard devices.

The S/390 environment is also equipped with workload management capabilities which are generally recognized to be superior to those of any other platform. These enable capacity to be switched transparently to high-priority workloads in the event of growth surges or accidents. This is particularly the case where Parallel Sysplex technology is employed.

UNIX or NT servers might eventually handle mainframe-class workloads with equivalent service quality. But their ability to do so is, at best, unproven. Early experiences have not been encouraging, even with today’s transaction volumes.

COST PICTURE

Infrastructure Example

Platform Parameters

What does it take to run the transaction-processing infrastructure of a large financial services company — handling not only conventional business-critical systems, but also the new, high-volume workloads that will be generated by transaction processing via the Web?

Mainframe parameters are known. In large banks, brokerages and insurance companies, and in many smaller organizations, core transaction-processing systems remain mainframe-based. Other platforms have typically been employed in distributed, departmental and specialized roles, or as front-end servers.

Which raises questions. For example, should cost comparisons be based only on front-end applications, or on the entire process of service delivery of which back-end workloads are a major component? While calculations based on the former may be useful, organizational costs will be determined by systems supporting the overall process structure. Comparisons may thus reasonably be made at this level.

No large financial services companies run core infrastructures on UNIX, NT or Tandem and equivalent servers. Any attempt to project costs for these must thus be speculative. There is, however, some experience with UNIX and (to a lesser extent) NT servers in transaction-processing applications in small financial services companies, and in departments and divisions of their larger counterparts. General extrapolations are thus legitimate.

Costs of Ownership

Figure 4 shows the business profile, workload characteristics and quality of service parameters of a $10 billion brokerage company, along with configuration, staffing and costs of ownership data for infrastructures based on IBM S/390 and Sun E10000 servers. Data is for all transaction-processing operations.

Results are based on detailed inputs from seven major brokerages, including five predominantly mainframe users, one company (with under $5 billion in assets) employing only Sun servers, and one employing both platforms. The profile is a composite of these, and calculations are based on their experiences. The assumption of 50 percent online customers is higher than the current norm, but reflects the growth trend for this channel.

Brokerage is probably the most advanced model for the competitive impact of e-business. Workloads are subject to heavy peaks, including sudden surges which may be 50 percent or higher than average volumes. High levels of availability during trading hours are required for business, legal and regulatory reasons. Applications are highly time-sensitive. And frustrated customers can move easily to competitors.

These comparisons are for underlying database and server platforms only, and do not include applications software or application development and support personnel.

Using any configuration assumption, high levels of proliferation occur in the Sun scenario. At a minimum, this would involve 36 systems, and at least the same number of copies of database, transaction monitor and system management software. In practice, E10000 servers are normally configured with multiple partitions. There would thus probably be at least 100 to 500 separate software stacks.

Figure 4
Costs of Ownership Calculation: Brokerage Example

BUSINESS PROFILE			
FULL SERVICE BROKERAGE			
Revenues:	$10 billion	Peak window: 50 hours/week	
Assets:	$75 billion	— 75% online, 25% batch	
Trades/day:	1 million+	Off-peak workloads	
Customers:	8 million	— 70% batch, 30% online	
Online customers:	50%	5 TB production data	
2 data centers, 7 x 24 operation		Subsecond response time at host	
9,000 internal users			
S/390 SCENARIO		**E10000 SCENARIO**	
Hardware		**Hardware**	
6 x Z77		36 x E10000	
52 GB RAM each		64 CPUs - 400 MHz each	
30 TB disk		48 GB RAM each	
Parallel Sysplex (2 x 3)		50 TB disk	
		SPARCcluster (9 x 4)	
Software		**Software**	
OS/390, DB2, CICS/ESA, VTAM, TCP/IP, IBM & Candle system management tools, COBOL, Assembler, REXX, C++, Java		Solaris, Oracle 8i, Tuxedo, Veritas, TCP/IP, CA-Unicenter & Sun system management tools, C++, Java	
Personnel		**Personnel**	
System administration	27	System administration	66
Database administration	21	Database administration	49
Operations	25	Operations	39
TOTAL	**73**	**TOTAL**	**154**
FIVE-YEAR COSTS ($000)		**FIVE-YEAR COSTS ($000)**	
Hardware	26,265	Hardware	96,456
Maintenance	11,136	Maintenance	40,372
Software	38,265	Software	31,460
Personnel	31,852	Personnel	61,481
TOTAL	**107,518**	**TOTAL**	**229,769**

This situation would, obviously, have major implications for the maintenance of availability, and would create high overheads for system and database administration. It would also be extremely difficult to consolidate data for decision support, operational, reporting and customer relationship management (CRM) applications.

Costs of Outages

In addition to potential legal and regulatory damage, discussed earlier, the business costs of outages can be divided into two main categories:

1. ***Revenue loss***. Customer trades may be postponed, abandoned or conducted through competitors if outages occur. Even if business is later recouped, there may be delays in posting income, and many customers will have been exposed to competitors.

2. ***Customer loss***. The potential impact of lost customers is even greater, particularly if measured in terms of CLV. Defection rates will be particularly high if outages occur during periods of heavy activity — which, experience has shown, is when poorly optimized systems are most likely to crash. They will be even higher if customers are exposed to multiple outages.

   It may, again, be possible to recoup business. But this will be an expensive process. As a general principle, it costs four to five times more in marketing expenditures to acquire a customer than to retain one, and the ratio may be significantly higher.

Figures 5 and 6 illustrate the cost implications of revenue and customer loss for the same composite brokerage firm. The values of individual companies will vary in practice, but the general approach is clearly valid.

Figure 5
Costs of Outages Calculation: Revenue Loss

VARIABLES	S/390 SCENARIO	E10000 SCENARIO
Revenues per year	$10 billion	$10 billion
Trading days per year	250	250
Revenues per trading day	$40 million	$40 million
Trading hours per day	9	9
Revenues per hour	$4,444,000	$4,444,000
Revenues per hour per cluster	$2,222,000 (2 clusters)	$494,000 (9 clusters)
Availability during trading day	99.998%	99.75%
Downtime per cluster per year	0.05 hours	5.63 hours
Lost revenues per cluster per year	$111,000	$2,781,000
TOTAL LOST REVENUES PER YEAR	$222,000 (x2 clusters)	$25,031,000 (x9 clusters)
TOTAL LOST REVENUES — FIVE YEARS	**$1.11 million**	**$125.16 million**

Overall, the costs of employing E10000 servers in this example are significantly higher than for the S/390 scenario. Over a five-year period, combined costs of lost revenues and CLV would be $437.91 million for the Sun scenario and $3.86 million for the S/390 scenario.

Figure 6
Costs of Outages Calculation: Customer Loss

VARIABLES	S/390 SCENARIO	E10000 SCENARIO
Online customers	4 million	4 million
Customers active per day	400,000 (10%)	400,000 (10%)
Customers active per hour	44,444	44,444
Customers active per hour per cluster	22,222 (2 clusters)	4,938 (9 clusters)
Availability during trading day	99.998%	99.75%
Downtime per cluster per year	0.05 hours	5.63 hours
Lost customers per hour of outage	222 (1% of active)	49 (1% of active)
Total lost customers per cluster per year	11	278
Average customer lifetime value (CLV)	$25,000	$25,000
Lost CLV per cluster per year	$275,000	$6,950,000
TOTAL LOST CLV PER YEAR	$550,000 (2 clusters)	$62,550,000 (9 clusters)
TOTAL LOST CLV — FIVE YEARS	**$2.75 million**	**$312.75 million**

Basis of Calculations

Costs of ownership comparisons are based on fully utilized production systems, and do not include additional capacity to handle exceptional workload peaks, idle fallover capacity, or test and development systems. Configurations, staffing levels and costs for both scenarios would be commensurately larger if allowance were made for these.

S/390 hardware, software and personnel costs are for two Parallel Sysplex clusters with three systems each, employing full data-sharing, located in two separate data centers. E10000 calculations are for nine SPARCclusters, also divided between two data centers. Workload distribution, data management and fallover capabilities are weaker than those of the Parallel Sysplex environment.

The values employed here are favorable to the Sun platform. It is assumed, for example, that four-way clusters are employed (the current norm is two); and that multiple partitions are employed in each 64-way configuration (i.e. that there is no N-way performance degradation, which is unlikely to occur in practice).

Similarly, a value of 55 percent peak processor utilization is used for E10000s, which is the norm for smaller applications on this platform. It is unclear whether it could be sustained with the volumes, response times and availability levels that characterize this installation.

E10000 batch performance is also probably overstated, and personnel costs are understated. Since there is no real experience with UNIX server configurations of this size in financial services — or indeed any other industry — there is no direct basis for comparisons. S/390 configurations and staffing are based on data supplied by users.

Figure 7 summarizes the basis of these comparisons. Personnel costs are based on prevailing 1999 United States salary levels for system programmers, system administrators, database administrators and operations staff, plus 36.2 percent for benefits, travel, training, overhead and related items.

Figure 7
Basis of Costs of Ownership Calculation: Brokerage Example

S/390 SCENARIO	
Hardware Configurations	
Base:	6,107 MIPS + 10% (Sysplex hardware & overhead) = 6,718 MIPS
	6 x Z77 (1,137 MIPS) = 6,822 MIPS
Peak Utilization	
CPU:	80%
Disk:	85%
Hardware Costs	
CPU:	$3.12K/MIPS loaded configuration
Disk:	$150/GB
Maintenance Costs	
10.6% of hardware acquisition cost/year x 4 years (First year warranty)	
Software Configurations	
OS/390 includes DFSMS, RACF, DFSORT & SDSF)	
Software Pricing	
PSLC-D (56.3% of norm) x 5 years	

E10000 SCENARIO	
Hardware Configurations	
Base:	2,326 CPUs (400 MHz equivalent)
	36 x E10000 64 CPUs = 2,304 CPUs
	Multiple partitions/system (assume no MP slowdown)
Peak Utilization	
CPU:	55%
Disk:	60%
Hardware Costs	
CPU:	List prices discounted 50%
Disk:	$144/GB
Maintenance Costs	
CPU:	9.0% of hardware acquisition cost/year x 4 years (First year warranty)
Disk:	12.3% of hardware acquisition cost/year x 4 years (First year warranty)
Software Configurations	
Equivalent functionality to OS/390 environment	
Software Pricing	
List prices discounted 50%, includes initial license & 4 years support fees (First year warranty)	

Outage calculations are also favorable to the E10000 scenario. It is assumed, for example, that this platform is capable of achieving 99.75 percent availability during periods of peak utilization, compared to an industry norm in the 95 to 97 percent range. Equally, the value of 5.6 hours downtime per cluster per year is generous, given the limited nature of SPARCcluster technology. The norm is over six hours.

Front-end Example

EFT Application

This example compares costs of ownership for front-end electronic funds transfer workloads. It is based on inputs from banks and service bureaus employing functionally equivalent applications on S/390 systems (five cases), Tandem servers (four) and both platforms (four). Extremely high levels of 7 x 24 availability were required.

Details of six representative installations are presented in figure 8. Results were similar for companies that were not included in this comparison.

Figure 8
Costs of Ownership Calculation: EFT Example

COMPANY	A	B	C	D	E	F
Peak transactions per second	18	53	70	150	265	375
Transactions per month (millions)	4	28	40	75	184	363
Number of cardholders (millions)	0.3	2.5	4	6.5	6	5
S/390 CONFIGURATION						
Hardware						
MIPS	16	37	60	76	185	372
RAM	1 GB	2 GB	3 GB	4 GB	6 GB	10 GB
Disk	14 GB	65 GB	120 GB	100 GB	90 GB	70 GB
Software	OS/390, VTAM, X.25, Assembler, C++					
Personnel (FTEs)	1	1	1.5	1.5	1.5	2
TANDEM CONFIGURATION						
Number of nodes	2	6	8	10	18	44
Personnel (FTEs)	3	2	2	2.5	4	5
S/390 FIVE-YEAR COSTS ($000)						
Hardware	52	125	205	252	591	1,184
Software	99	231	329	428	1,054	2,108
Maintenance	22	53	87	107	250	502
Subtotal	**173**	**409**	**621**	**787**	**1,895**	**3,794**
Subtotal x 2	346	818	1,242	1,574	3,790	7,588
Personnel	421	421	632	632	632	843
TOTAL FIVE-YEAR COSTS ($000)	**767**	**1,239**	**1,874**	**2,206**	**4,422**	**8,431**
TANDEM FIVE-YEAR COSTS ($000)						
Hardware/software	478	1,435	1,913	2,392	4,305	10,523
Maintenance	143	430	574	717	1,291	3,157
Personnel	1,291	861	861	1,076	1,722	2,152
TOTAL FIVE-YEAR COSTS ($000)	**1,912**	**2,726**	**3,348**	**4,185**	**7,318**	**15,832**

The application switches transactions from ATMs, point-of-sale devices and teller terminals, and provides authorization. Key workload variables include overall transaction volume, peak loading in tps, and volume of calls to cardholder databases for verification purposes. Peak tps numbers are for sustained volume over a five-minute period.

Companies reported savings of two to four times in moving from Tandem to S/390 systems. A key factor was that, in all companies, mainframes were already employed for back-end processing. At least some of the online EFT workload could be handled by systems which otherwise ran batch during off-peak hours.

S/390 capacity was thus in some cases available to handle EFT workloads during the day. Where this was the case, the incremental costs of processor and storage resources were minimal. Companies were also typically able to handle EFT workloads with their existing S/390 data center staffs, or with one or two additional full time equivalent (FTE) persons.

The experiences of companies employing S/390 and Tandem servers highlight the effects of duplication. It was necessary to support two very different system architectures, requiring separate sets of skills. This affected cost structures in all of the companies surveyed. Its impact was particularly severe in smaller banks, where dedicated Tandem specialists had to be employed to support even low-end configurations.

The central issue, as one respondent noted, was that Tandem servers could not realistically replace mainframe-based back-end systems. The latter could, however, replace front-end servers while maintaining equivalent performance and availability, and offering significantly lower costs. Even Tandem users recognized this.

Higher integration levels also characterize OS/390. It is thus not necessary to support certain functions directly, as is the case in the Tandem environment. This is reflected in higher Tandem staffing levels.

Basis of Calculations

Key values are summarized in figure 9.

Figure 9
Basis of Costs of Ownership Calculation: EFT Example

S/390 Hardware:	$3.12K per MIPS (loaded configuration) + $150 per GB disk storage
Tandem Hardware/Software:	S 70000-400 MHz node with 128 MB RAM, 16 GB disk & NSK operating system. $239,150 per node
S/390 Software:	PSLC-C pricing x 5 years
Maintenance:	10.6% (S/390) & 7.5% (Tandem) per year of hardware costs x 4 years (assumes first year warranty)
Personnel:	$61,867 (S/390) & $63,189 (Tandem) average salary per FTE + 36.22% (benefits, travel, training, overhead) x 5 years

S/390 hardware, software and maintenance costs are doubled to provide the same level of fault tolerance as the Tandem architecture; i.e., standby capacity equivalent to 100 percent of the production system is available within Parallel Sysplex clusters.

Calculations are, again, for underlying servers and staff handling system administration, operations and other tasks required for these. They do not include applications software and related personnel costs.

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 information services and management consulting firm specializing in timely analyses of current developments and future trends in information systems technologies and markets.

Particular attention is given to analyzing trends in emerging technologies and markets. ITG's staff and supporting consortium of international analysts are well experienced and skilled in creative assessment techniques and research methodologies.

Services provided to clients are designed to provide factual data and reliable documentation to assist in the decision-making process. The information provided establishes the basis for developing tactical and strategic plans. Important developments are analyzed and practical guidance is offered on the most profitable ways to respond to the changes influencing and shaping complex IT deployment.

A broad range of services is offered, furnishing clients with the information necessary to complement their internal capabilities and resources. Customized programs are designed to provide a continuous source of technology, product and market information involving delivery of various combinations of the following services:

Status Reports	In-depth studies analyzing important topics and future industry trends
Flash Reports	Timely documents covering latest developments and significant market activity
Management Briefings	Periodic interactive planning meetings with management to address tactical issues
Telephone Consultation	Immediate response to clarify conclusions and review current informational needs
Executive Presentations	Scheduled interactive strategic meetings with decision-makers to examine future directions

The firm's **Enterprise Information Systems** service is particularly noteworthy. This service covers various architectures, technologies and techniques that may be effectively deployed in order to achieve the goals of enterprise-wide information systems. Scenarios being proposed by leading system suppliers and software vendors are objectively evaluated. Clients include systems suppliers, software vendors, service firms, financial institutions, government agencies, trade associations and end users.

International Technology Group
885 North San Antonio Road, Suite C
Los Altos, California 94022-1305
Telephone: (650) 949-8410
Facsimile: (650) 949-8415
Email: itg1@pacbell.net