

July 2004



**IBM @server zSeries:
Premier Business Resiliency
for the On Demand World**

Contents

- 3 Introduction**
- 5 *The built-in availability features of zSeries servers: it's all about the application!***
- 7 *Single-Site / Multi-server environment that can deliver 99.999% availability***
- 8 *Multi-Site Availability: The zSeries Solution with GDPS***
- 9 Conclusion**

Introduction

April 7, 2004 marked the 40th anniversary of the System/360™. What began as a computing revolution in 1964 has evolved with relentless innovation, continuous refinement, time-tested core competencies and proven technologies – with the result being today’s IBM @server® zSeries® advantage: world-class servers that have been designed from the start to address critical business requirements today, defining the gold standard of computing in the server marketplace.

High on that list of critical business requirements today is the need for IT infrastructures to better provide continuity of business operations in the event of planned or unplanned disruptions. Mission-critical application availability directly correlates to successful business operations. In today’s on demand business environment, “downtime” whether planned or unplanned, is not only unwelcome – it’s costly. Downtime statistics are staggering and range from tens of thousands of dollars to multiple millions of dollars per hour of downtime. And while the financial impact is significant, the damage can extend well beyond the financial realm into key areas of customer loyalty, market competitiveness and regulatory compliance.

Downtime costs will quite likely continue to rise and become more visible as enterprises increase their reliance upon IT solutions. Your customers often access products via the Internet, so the amount of time and effort it takes for them to switch loyalties to another supplier is minimal, to say the least. If your site isn’t up or responsive when your customers want it, they’re more likely to go somewhere else in the time it takes to click a mouse. The fact is – frequent system outages in today’s highly competitive marketplace can negatively impact your business.

Companies that have neither designed adequate IT infrastructures to accommodate continuous availability nor developed business continuity plans in general often make the assumption that back-office and manual processes will keep things running until computer systems are again available. Characteristics of these recovery models may allow critical applications to recover within 24 to 48 hours, with data loss potentially exceeding 24 hours, and full business

Highlights

recovery taking days or weeks. As companies transform to compete in today’s on demand world, those older processes will most likely not be good enough. Server platform decisions, business continuity strategies, and application availability requirements must be re-evaluated to ensure they’re based on today’s business requirements.

Brokerage Retail	\$6.5 million
Credit Card Sales Authorization	2.6 million
Airline Reservation Centers	90,000
Package Shipping Services	28,250
Manufacturing Industry	26,761
Banking Industry	17,093
Transportation Industry	9,435

Figure 1: Financial Impact of Downtime Per Hour (by various Industries)
Source: Contingency Planning Research & Strategic Research Corp.

A resilient on demand business must be highly available despite planned and unplanned outages – as zSeries servers are.

In the on demand world, two of the most stringent demands for survival are continuous availability and near-transparent disaster recovery. Systems that are designed to deliver continuous availability combine the characteristics of high availability and continuous operations to deliver high levels of service (24x7). High availability is an attribute of a system that provides service at agreed-upon levels and masks unplanned outages from end users and critical business processes. Continuous operations, on the other hand, typify systems that can continuously operate and mask planned outages from end users and critical business processes.

To attain levels of continuous availability and near-transparent disaster recover, the ideal solution must be based on a shared view of data, geographical clusters and data mirroring. And these are precisely the key technologies which IBM @server zSeries servers, Parallel Sysplex® data sharing and GDPS® (Geographically Dispersed Parallel Sysplex™) architectures can deliver. Most importantly, these zSeries-based technologies have evolved over time, and the integrated hardware, firmware, operating system and middleware elements have been designed to tightly work together providing an application environment with world-class levels of high availability and disaster recovery. zSeries servers approach application availability with an integrated and cohesive strategy that encompasses single-server, multi-server and multi-site environments.

The built-in availability features of zSeries servers: it's all about the application!

It's important to note that disasters are not the only cause of downtime. After all, some companies must shut down their systems to make scheduled updates, perform maintenance or upgrade their servers. The built-in availability features on the zSeries platform can empower you to avoid both scheduled and unscheduled outages, as well as position your data center for world-class disaster recovery.

The zSeries product line offers layer upon layer of fault tolerance and error checking. If a failure occurs, the built-in redundancy on zSeries servers can re-route work from failing components to operational ones in an effort to prevent the end-user service from being interrupted. The failed components may be removed and replaced while the application is still active, allowing service to continue. The spec sheet of IBM's flagship zSeries server, the IBM @server zSeries 990 (z990), highlights key availability features integrated into the overall zSeries system design, and the list is largely representative of those availability features also included on the newest member of the zSeries family, the IBM @server zSeries 890 (z890) server.

Some of the features are designed to directly support key middleware such as CICS® – enabling greater efficiency and higher levels of availability for transaction execution. Other features are designed to work directly with operating systems such as z/OS®, z/VM® and even Linux for zSeries – enabling greater flexibility when dynamically reconfiguring zSeries servers if workloads need to be re-routed in the event of an outage.

So with zSeries... the availability design-point focuses on “the application.” Applications don't just rely on server hardware. They require an integrated environment in which hardware, firmware, operating systems and middleware work together to provide application and data availability. MTBF for zSeries servers is now measured in decades; but the zSeries availability story doesn't stop there. Approximately one third of the z/OS code base provides rich RAS functionality delivering reliability, availability and serviceability – often resulting in outage events being completely masked from applications, and in severe cases resulting in graceful degradation rather than complete failure. And concurrent maintenance capabilities – supported by both the hardware as well as operating systems – help mask planned outage events from the application as well.

Furthermore, focusing on application availability means introducing design-points that address as many of the potential causes of downtime as possible:

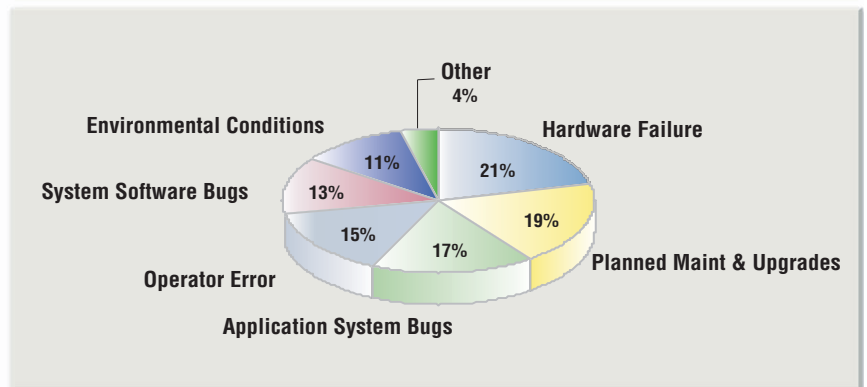


Figure 2: Common Causes for “Application Downtime”
Source: Standish Group Research

As noted in the above figure, “operator error” is attributed to 15% of data center outages in a recent survey. Today’s data center environments are admittedly complex, so any technologies that help reduce operator intervention are a welcome addition to the IT operating environment. In the zSeries world, z/OS raises operator automation and operator assistance to state-of-the-art levels with msys for Operations and a z/OS Healthchecker – both of which complement a robust suite of Tivoli® systems automation products.

Advanced virtualization technologies and heterogeneous workload management are also a critical part of the business continuity strategy for zSeries. Indeed, as a growing number of elements in today’s IT infrastructure can be virtualized on zSeries – including server images, I/O connectivity, and now even networking elements – the computing platform is ideally positioned to handle everyday IT infrastructure dynamics as well as the occasional disaster recovery events that may occur. Combine zSeries advanced virtualization technologies, workload management capabilities, On/Off Capacity on Demand, Capacity BackUp (CBU), and the mature disaster recovery techniques in place at many state-of-the-art data centers today, and you have an outstanding environment that can deliver the ultimate in business resiliency.

A resilient on demand business depends upon a uniform view of data for all servers and applications – as zSeries servers provide through the ability to support very large databases.

Single-Site / Multi-server environment that can deliver 99.999% availability

Introduced in 1994, the zSeries Parallel Sysplex architecture allows clustered zSeries servers to provide resource sharing, workload balancing and data sharing capabilities for on demand data centers –delivering ultimate flexibility when supporting different application topologies. While zSeries hardware, operating systems and middleware have long supported multiple applications on a single server, Parallel Sysplex clustering allows multiple applications to communicate across servers – and even supports the concept of a large, single application spanning multiple servers, resulting in optimal availability characteristics for that application.

With Parallel Sysplex clustering and its ability to support data sharing across servers, IT architects can design and develop applications that have one, integrated view of a shared data store – effectively eliminating the need to partition databases that so often occurs when UNIX® and Intel® processor-based computing platforms are deployed as the database server. Data sharing with Parallel Sysplex has the unique advantage of allowing nondisruptive database growth with automatic load re-balancing. zSeries shared database architectures also provide inherent locking services for the ultimate in data integrity – thereby avoiding application-level locking and the need to frequently partition and re-partition applications as is so often the case with UNIX and Intel processor-based server architectures.

Application and/or database growth in some non-zSeries partitioned database environments would typically create workload skews requiring lengthy and disruptive database re-partitioning which so often means downtime for the application. And allowing for data integrity with some non-zSeries partitioned databases often requires application-level locking – which in high-volume transaction environments could lead to service level agreements not being met... again, not optimal for high-profile, high-availability applications.

So Parallel Sysplex data sharing capabilities can avoid these obstacles to near-continuous availability encountered with partitioned database architectures, plus it can deliver an added bonus: single-view database simplicity helps remove unnecessary complication in the IT infrastructure. And simpler IT infrastructures help reduce the likelihood of unexpected outages, while at the same time allowing planned outages to have a smaller impact across the overall application space.

A resilient on demand business depends upon multi-site application availability – as zSeries servers provide through IBM's Geographically Dispersed Parallel Sysplex.

Multi-Site Availability: The zSeries Solution with GDPS

At the heart of zSeries multi-site availability is IBM's Geographically Dispersed Parallel Sysplex (GDPS) technology, which is positioned to provide a total business continuity solution for the z/OS environment. Based on geographical separation and advanced automation techniques, GDPS is a multi-site application availability solution that provides the capability to manage remote copy configuration and storage subsystems, automate Parallel Sysplex operation tasks, and perform failure recovery from a single point of control. In short, GDPS raises the "near-continuous availability" level of single site Parallel Sysplex to the "continuous availability" level – helping to further address data center concerns of external disruption to both applications and data.

GDPS extends the resource sharing, workload balancing, and continuous availability benefits of a Parallel Sysplex environment. It also significantly enhances the capability of an enterprise to recover from disasters and other failures, and to manage planned exception conditions – allowing business to achieve their own continuous availability and disaster recovery goals.

GDPS supports all popular zSeries transaction managers (i.e., CICS, IMS™, WebSphere®) and database managers (i.e., DB2®, IMS and VSAM) – and is enabled by a number of key IBM technologies and architectures.

Managed exclusively through IBM's GDPS is HyperSwap™, a new software technology that can substitute Peer-to-Peer Remote Copy (PPRC) secondary devices for PPRC primary devices. HyperSwap is designed to swap a large number of devices and to do it so fast, there is minimal impact to application availability (often measured in seconds instead of hours). The HyperSwap function – also available with single-site Parallel Sysplex environments – may be performed even if the primary disk subsystem is not operational, creating an opportunity to survive a primary disk subsystem – or even a complete site – failure without requiring systems to be recycled.

"Using the GDPS/PPRC HyperSwap technology is a significant step forward in achieving continuous availability," said Wolfgang Dungal, manager of Availability, Capacity and Performance management for iT-AUSTRIA. "The benefits in our GDPS environments are that a site switch by means of HyperSwap (i.e., dynamic

switching of the disk configuration) takes only a few minutes without application outage. Without HyperSwap, we had to shut down the GDPS/sysplex, switch the disk configuration, and restart systems and applications. The elapsed time was almost two hours for our largest GDPS with 11 systems and over 4,600 PPRC volume pairs.”

Prior to the implementation of GDPS technology, iT-AUSTRIA’s recovery time was in excess of 48 hours for only 2 Terabytes of DASD, whereas the current scenario exploits GDPS and HyperSwap – yielding a recovery time of a few minutes for approximately 76 Terabytes of DASD.

Additionally, IBM has already extended GPDS/PPRC technology to manage a heterogeneous environment of z/OS and Open Systems data. This has tremendous significance for enterprises seeking to simplify infrastructure complexities with combined zSeries and BladeServer platforms. If installations share their disk subsystems between the z/OS and Open System platforms, GDPS/PPRC can manage the PPRC and FlashCopy® for Open Systems storage. GDPS/PPRC is also designed to provide data consistency across both z/OS and Open Systems data.

Conclusion

The expanding scope of security threats, viruses and denial of service attempts so common in newspaper headlines today serve to remind us just how critical it is for enterprise IT infrastructures to be configured for maximum – yet cost-effective – availability and data integrity. Much work has been done, and much work continues to be done to analyze common, and not so common, causes for system outages. Results of these studies feed directly into zSeries server, operating system, middleware, Parallel Sysplex, and GDPS design and development teams. The resulting enhancements to these technologies – including increased simplification, advanced autonomics, and IBM @server cross-platform capabilities – will build upon the already-robust set of business resiliency capabilities that can best be delivered only when hardware, operating system and middleware components are designed in an integrated fashion to help deliver state-of-the-art business resiliency for today’s – and tomorrow’s – applications.

Today's zSeries leadership in delivering world-class availability levels and overall business resiliency has been 40 years in the making – and it's always been about the application and keeping the business running. No other computing platform has integrated the number of availability design-points throughout its entire architecture as long as – nor as well as – the mainframe. And no other company is in a better position than IBM to help bring the zSeries availability technologies to your entire IT infrastructure.

Choose an IT architecture based firmly on IBM @server zSeries products for your highly visible, mission-critical applications, and know that you're arming your enterprise with world-class business resiliency for the on demand era. The zSeries platform, along with related IBM products and service offerings, puts you in a position to take advantage of latest-generation technologies and evolve along with the demands of your business.



© Copyright IBM Corporation 2004

IBM Corporation
Software Communications
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
07-04
All Rights Reserved

The e-business logo, IBM, the IBM logo, IBM eServer, CICS, DB2, FlashCopy, GDPS, Geographically Dispersed Parallel Sysplex, HyperSwap, IMS, Parallel Sysplex, System/360, Tivoli, WebSphere, z/OS, z/VM, and zSeries are trademarks or registered trademarks of International Business Machines Corporation of the United States, other countries or both.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Intel is a trademark of Intel Corporation in the United States, other countries or both.

Linux is a registered trademark of Linus Torvalds.

Other company, product and service names may be trademarks or service marks of others.

Information concerning non-IBM products was obtained from the suppliers of their products or their published announcements. Questions on the capabilities of the non-IBM products should be addressed with the suppliers.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

