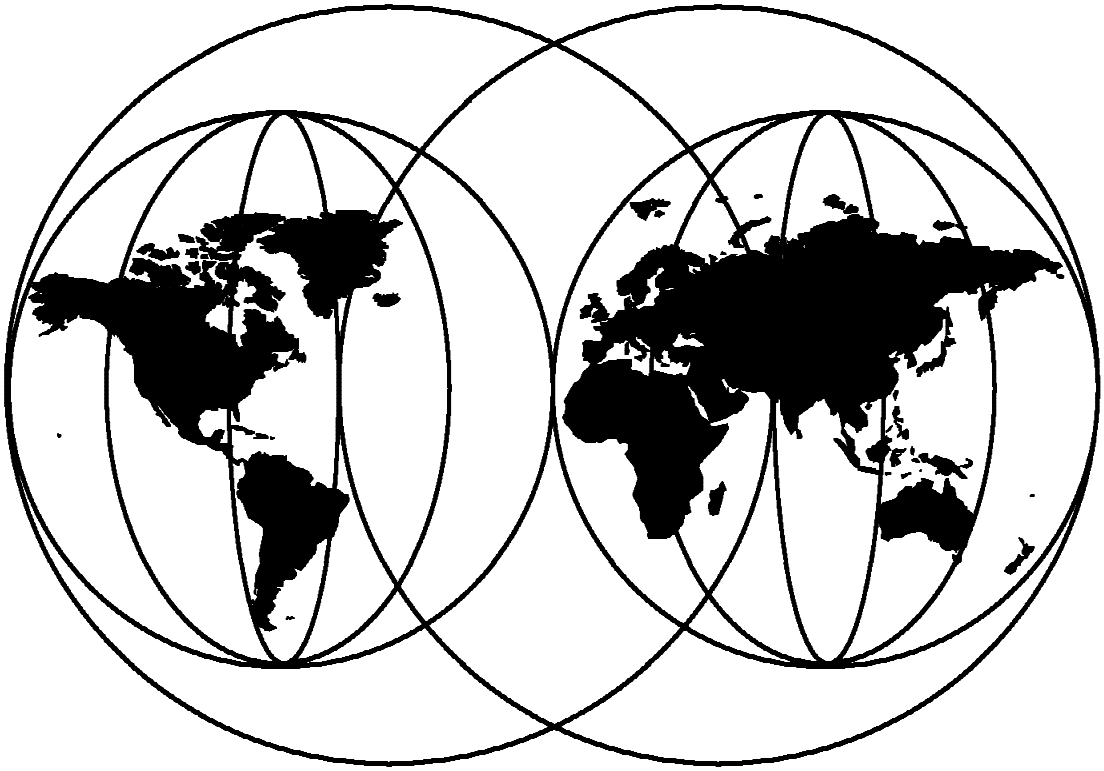
IBM

Parallel Sysplex Continuous Availability Case Studies

Frank Kyne, Paola Bari, Mary Petras



International Technical Support Organization

http://www.redbooks.ibm.com

SG24-5346-00

International Technical Support Organization

SG24-5346-00

Parallel Sysplex Continuous Availability Case Studies

June 1999

Take Note!

Before using this information and the product it supports, be sure to read the general information in Appendix B, “Special Notices” on page 89.

First Edition (June 1999)

This edition applies to

Program Name, Program Number	Version, Release Number
CICS TS for OS/390, 5655-147	1.2
DB2 for MVS/ESA V4, 5695-DB2	4.1
DB2 for OS/390 V5, 5655-DB2	5.1
DFSMS/MVS, 5695-DF1	1.4

for use with the:

Program Name	Version, Release Number
OS/390	1.3

Comments may be addressed to:
IBM Corporation, International Technical Support Organization
Dept. HYJ Mail Station P099
522 South Road
Poughkeepsie, New York 12601-5400

When you send information to IBM, you grant IBM a non-exclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Contents

Figures

Tables

Preface

This redbook discusses the topic of continuous availability in a Parallel Sysplex environment. It provides advice about how to create a business case for making an application continuously available, and presents two case studies of actual customer experiences of implementing Parallel Sysplex as part of their program for improved availability.

This redbook will be of assistance to anyone contemplating a continuous availability project. It discusses the various aspects of availability and provides advice about how to get the best value for your money in terms of availability.

The two case studies cover DB2 data sharing and VSAM Record Level Sharing and provide specific technical tips that may assist other customers in similar environments.

The Team That Wrote This Redbook

This redbook was produced by a team of specialists from around the world working at the International Technical Support Organization Poughkeepsie Center.

Frank Kyne is a Senior I/T Specialist at the International Technical Support Organization, Poughkeepsie Center. He has been an author of a number of other Parallel Sysplex Redbooks. Before joining the ITSO one year ago, Frank worked in IBM Global Services in Ireland as an MVS Systems Programmer.

Paola Bari works in the S/390 New Technology Center in Poughkeepsie. She is responsible for assisting customers migrate to Parallel Sysplex in both the MVS and VSAM/RLS areas. Prior to moving to the New Technology Center, Paola worked in the service organization of IBM Italy for 15 years. During her 17 years working with S/390 hardware and operating systems, she has worked in the product defect area, the service organization, and in the education division. Paola is an IBM Certified Services Specialist.

Mary Petras is a DB2 Specialist for the IBM Santa Teresa Laboratory responsible for assisting DB2 customers in the USA with special projects. She has recently presented at the DB2 Technical Conference and also at local DB2 regional user groups. She has more than 20 years experience in data processing and worked as a software developer, MVS systems programmer, technical trainer and application programmer before joining IBM in 1996. Mary holds a B.S. and an M.S. in Mathematics from Pratt Institute, School of Engineering and Science, New York. Her areas of expertise include DB2 performance and data sharing.

Thanks to the following people for their invaluable contributions to this project:

Jeff Josten
IBM Santa Teresa

Gary King
IBM Poughkeepsie

Gopal Krishnan
IBM Santa Teresa

Peggy Alkinc
UPS Mahwah

Chris Bolcato
UPS Mahwah

Rose Brown
UPS Mahwah

Diane Chan
UPS Mahwah

Cheryl Doberman
UPS Mahwah

Dave Feigenbaum
UPS Mahwah

William Grib
UPS Mahwah

Rose Jung
UPS Mahwah

Dave Koa
UPS Mahwah

Patty Loud
UPS Mahwah

Andrew MacDonald
UPS Mahwah

Tom O'Neill
UPS Mahwah

Kanu Patel
UPS Mahwah

Jim Pawlak
UPS Mahwah

Maurizio Petracca
SSB

Alex Rutman
UPS Mahwah

Benjamin Teixeira
UPS Mahwah

Leslie Vella
UPS Mahwah

Peter Tolosi
UPS Mahwah

Sharon Williams
UPS Mahwah

Cathy Wilson-Daly
UPS Mahwah

Theresa Young
UPS Mahwah

Comments Welcome

Your comments are important to us!

We want our redbooks to be as helpful as possible. Please send us your comments about this or other redbooks in one of the following ways:

- Fax the evaluation form found in “ITSO Redbook Evaluation” on page 97 to the fax number shown on the form.
- Use the online evaluation form found at http://www.redbooks.ibm.com/
- Send your comments in an Internet note to redbook@us.ibm.com

Chapter 1. Introduction

Continuous availability. This is a term that has received a lot of renewed attention over the last year or two. But exactly what is it, and why has it suddenly become such an issue?

What it is depends on who you ask. If you are the MVS systems programmer responsible for an OS/390 system, continuous availability may mean that your MVS system never has an unplanned outage. If you are a CICS systems programmer, it might mean that you have CICS set up with MRO and VTAM Generic Resources over a number of MVS images so that the CICS application will still be available even if one MVS image is shut down.

But what about if you are the end user? To the end user (which today is often the customer), continuous availability means that you can execute the transaction in full capability mode any time, day or night, any day of the week. You do not care about whether an outage is planned or unplanned. And you are not interested in the fact that the data must be backed up or reorganized. And it is not your concern that the Internet Service Provider that you use is recycling his DNS server. As far as you are concerned, you are trying to access your bank account to pay some bills and you cannot get at it.

It is obvious that continuous availability means different things to different people. IBM has defined three terms that help clarify things to some extent:

High Availability This has to do with keeping an application running during the planned service hours. It involves redundancy of components, to ensure there is always an alternative available if something breaks, and thorough testing to ensure that any potential problems are detected before they can affect the production environment.

Continuous Operations This means that a system can provide service to its users at all times without any outages, planned or otherwise. This is not that difficult to achieve, and there are many examples of specialized systems, such as a Communication Management Configuration, that run for many months without any sort of outage. However, the prerequisite is that few or no changes can be made to the system, and this is a very unlikely scenario in a normal production system.

Continuous Availability This is a combination of high availability and continuous operations. It means that the application service will remain available across planned and unplanned system outages.

In reality, when people say they need continuous availability, in most cases they mean that they want the application to be available all the time during the agreed service hours, regardless of problems with, or changes to the underlying hardware or software. What makes this more stringent than high availability is that the service hours are getting longer and longer, to the point that there is no time left for making changes to any of the system components.

The quest for higher and higher availability has to be tempered by what the business actually *needs* and what is possible with the available technology. In reality, most applications can still withstand some planned outage, either for batch work, backups and reorgs, or to effect application changes. Another example of reality tempering expectations is that in most cases, the application is expected to be available at the host, or maybe on the network that is owned and controlled by the organization, but continuous availability in the environment outside the control of the organization (that is, on the Internet) is not normally included in the business case.

The Parallel Sysplex is IBM's mechanism to help customers provide this definition of continuous application availability. It is designed to spread a workload over a number of images (thus benefiting from the redundancy inherent in high availability), so that changes can be made to one of the underlying images without affecting the availability of the application.

In this redbook, we provide two examples of how customers have used Parallel Sysplex to help improve the availability of their business-critical applications. These customers have used Parallel Sysplex to provide the level of continuous availability currently required by their businesses and to build a base on which further availability improvements can be made.

We had also hoped to provide guidelines for building a business case, based on actual customer examples. Unfortunately, customers that have an effective, accurate methodology for sizing the cost of an outage and the value of availability generally consider that methodology to be a business asset and are understandably loathe to share it. Therefore, we have instead provided an Appendix that simply provides a list of most of the costs associated with providing continuous availability, and identifies costs associated with unavailability, to help you build a business case of your own.

For further discussions about continuous availability, consult the redbooks *Continuous Availability Systems Design Guide*, SG24-2085, and *Continuous Availability S/390 Technology Guide*, SG24-2086. There are also other availability publications listed in the bibliography.

1.1 Determining Your Availability Requirements

Most people, when first asked how much availability they require, will reply that they want continuous availability. However, when shown the cost of truly continuous availability, the requirement often reduces to something between "What I have now is fine!" to "As much as I can afford.."

Not many applications can cost-justify 100% availability. The cost of availability increases dramatically as you get closer to 100%. Moving from 90% to 97%, for example, probably costs nothing more than better processes and practices, and very little in terms of additional hardware or software. Moving from 97% to 99.9% will require investing in current hardware technology, implementing very good processes and practices, and committing to staying current on software levels and maintenance. 99.9% is the availability claimed for a non-sysplex S/390 environment in a recent independent study. 99.9% availability equates to 8.9 hours downtime a year. The same study found that a Parallel Sysplex environment achieved availability of 99.998%, or just 10 minutes unplanned downtime a year. Removing that last 10 minutes of downtime is likely to be more costly than moving from 99.9% to 99.998%. Appendix A, "The Cost

Components of a Business Case" on page 75 will help you decide if the value of availability to the application justifies the expense.

What would probably be far more beneficial, and far less expensive, would be to address the planned outages. In an IBM study, planned outages account for over 90% of all outages. And of the planned ones, about 40% are for hardware, software, network, or application changes. All of these can be addressed, at least to some extent, with Parallel Sysplex. If you are already down to 10 minutes unplanned downtime a year, it is likely that you are already using Parallel Sysplex. If this is the case, then 36% of *all* your downtime can be addressed by exploiting the functionality provided by technology that you already have. This will buy you much more customer satisfaction at a much smaller price than trying to eliminate those last 10 minutes of unplanned downtime.

1.2 Case Studies

The customers whose experiences are documented in this book are excellent examples of how customers are striving to improve availability while operating in the real world of finite budgets and tight schedules.

Both of the customers that worked with us on this book have implemented Parallel Sysplex as part of their efforts to provide greater availability for their critical applications. Neither customer has an application that never has an outage, but at this time, neither business has that requirement. Both customers have wisely implemented the new technology in a staged manner, improving availability while gaining experience and still having scope for further exploitation and still higher availability.

The two case study chapters provide detailed information about the Parallel Sysplex implementation projects of the two customers and highlights the continuous availability aspects of both projects. It is hoped that the experiences of these customers will prove useful to other customers that are also planning to implement Parallel Sysplex as part of a plan for higher availability.

The appendix contains a discussion of the cost aspects of providing higher availability, as well as some suggestions about how to place a monetary value on the cost of unavailability. The information in this appendix should help you build one part of the business case for a continuous availability project.

Chapter 2. DB2 Case Study

Because so many customers use DB2, and many of those are considering making their DB2 data available via the Internet, it was critical that we present a case study of a DB2 application. The customer chosen for our DB2 case study had to meet the following criteria:

- The customer should be well known.
- The customer's technical staff should be very skilled, so as to be able to provide us with as much information as possible and answer any questions that might arise.
- The application should be both sizable and critical to the business.
- The application should use components and products that are in common usage.

We were very fortunate to find a customer that met and exceeded all of these criteria and was enthusiastic about taking part in this project. The customer that is the base for this case study is United Parcel Service (UPS) of the United States.

We would like to take this opportunity to thank the staff at UPS that worked with us on this project. It would not have been possible to provide so much useful information without their support and enthusiasm.

This chapter describes the data sharing migration project at UPS. It details the steps that were taken to move their most critical application to data sharing mode, and relates valuable lessons learned in the process.

At the time of writing, the application in question is not “continuously” available in absolute terms. However, it is providing the levels of availability required by the business at this time. We will discuss some of the inhibitors to truly continuous availability and UPS' plans to address those issues.

2.1 Company Description

UPS is the world's largest package distribution company, transporting more than 3 billion parcels and documents annually to more than 200 countries and territories worldwide. The brown UPS vans are a familiar sight; UPS has become a symbol for reliability in a changing world where the convenient overnight package delivery business is becoming more and more prevalent.

The world headquarters is in Atlanta, Georgia. Worldwide, UPS employs over 325,000 people. 1998 revenue was $24.8 billion. The delivery fleet consists of 157,000 vehicles, 224 privately-owned aircraft, and over 300 chartered planes, with over 1,500 daily flights serving 600 airports worldwide. A little-known fact is that as of December 1997, UPS was the tenth largest airline in the world.

UPS has data centers in Atlanta, Georgia and Mahwah, New Jersey. At the time of writing, there are a total of 8,013 mainframe MIPS spread across 14 mainframes with 55 TB of DASD storage. In addition, there are 868 midrange computers servicing 218,000 PCs connected by 3,500 LANs serving over 600,000 users.

In 1998, UPS delivered 3.14 billion packages, averaging over 12 million packages and documents each day. The delivery of packages in a convenient and timely fashion and reliable, efficient and accurate tracking of all packages is accomplished via technology that uses DB2 data sharing and Parallel Sysplex.

2.2 Background to Parallel Sysplex Project

The Package Level Detail Repository (PLDR) is the set of UPS applications that controls all aspects of processing all stages in the life of a parcel from collection from the shipper, through transportation within UPS, and on to final delivery or collection by the customer.

The PLDR applications were developed over a number of years, in separate stages. They all are DB2-based and are spread over the two UPS sites. PLDR consists of five major DB2 subsystems, each using 3 to 5 TB of DB2 data, and each requiring nearly 24 x 7 x 365 availability. In the summer of 1996, each of these five subsystems ran on its own IBM 9021-9X2 and all were approaching the capacity limits of those processors. The largest application, Delivery Information Automated Lookup System (DIALS), was expected to run out of capacity by the end of 1996.

There were also application enhancements in development that were expected to increase processor usage. And finally, the ability of customers to query the status of their packages via the Internet was expected to increase transaction rates for DIALS, leading to greater processor utilization.

The impending capacity issues caused UPS to initiate a project to look at the options available to them. Although capacity was an immediate concern, and the initial driving force behind the project, UPS also felt it was important to take this opportunity to try to eliminate any single points of failure and reduce the amount of time for any planned or unplanned outages.

Given this scenario, it was decided to investigate the option of migrating DIALS to a data sharing environment, using Parallel Sysplex. The scalable architecture of Parallel Sysplex and data sharing would support the growth of their business, including developing new applications, such as a new UPS initiative, Guaranteed Ground. The technology would meet the current business needs, future processing growth, and competitive demands of the business.

While an upgrade to a larger processor would address the immediate capacity concerns, UPS senior management made a strategic decision to move to Parallel Sysplex data sharing. When successfully implemented, this would remove any capacity constraints and also provide improved application availability.

2.2.1 Initial IT Infrastructure

When the data sharing project began, DIALS ran on a 9021-9X2 at nearly 100% capacity during peak daily periods. The processor ran in basic mode and DIALS was the only application running on it. ESCON Directors were used to provide any-to-any connectivity between the processors and I/O devices. The 9021-9X2 was configured as shown in Table 1 on page 7.

Table 1. DIALS Processor at Start of Project				
CPU Type	**Central Storage**	**Expanded Storage**	**MIPS (as rated by UPS)**	**Number of CPs**
IBM 9021-9X2	2,048	4,096	464	10

The DIALS 9X2 was in a sysplex together with four other 9X2s that ran other applications. The production Parallel Sysplex used two 9674-C04 Coupling Facilities. There was also a development Parallel Sysplex running on two 9672-C02 Coupling Facilities.

The following software products were installed on the DIALS system at that time. The products were at the required levels for resource sharing, but not for data sharing.

- MVS/JES2 5.2.2
- DB2 Version 3.1
- CICS/ESA Version 4.1.0
- IMS Fast Path Version 4.1
- Omegamon for CICS V300
- Omegamon for MVS V350
- Omegamon for DB2 V300
- BMC Unload Plus 3.1
- BMC Reorg Plus 4.2.02
- BMC Recover Plus 2.3.0
- BMC Load Plus 3.1.00
- BMC Copy Plus 5.2.0

2.3 Application Description

PLDR is comprised of a total of 50,000 IMS and DB2 data sets amassing over 16.8 TB of data with 324 billion rows. It was cited by the Wintercorp Corporation as the world's largest DB2 database repository in 1996 and 1997. The largest single table contains 64 GB and 3.5 billion rows.

2.3.1 Delivery Information Automated Lookup System (DIALS)

The first business application chosen to participate in DB2 data sharing was the Delivery Information Automated Lookup System (DIALS), which tracks each package and is considered by many to be the most mission-critical application at UPS. At the heart of the DIALS application is a DB2 database that contains all pertinent information regarding the instance of a package delivery event. A DIALS record is only produced when the package is delivered or attempted to be delivered to the customer. The other PLDR databases contain information about the earlier stages in the life of the parcel.

Multiple online processes are used to provide package delivery information on the demand of external UPS client requests, as well as requests from internal organizations responsible for managing and auditing the package delivery process.

DIALS contains two types of information:

- Package delivery information derived from other systems involved in the package shipment and tracking process (for example, shipper name and address, charge information, delivery center, and so on).

- Information that is directly related to the delivery event (for example, name and address of the person who received the package, receiver signature in bit map form, time and date of the delivery, driver information, money amounts collected for COD, and so on).

Using a hand-held computer device called a Delivery Information Acquisition Device (DIAD), the UPS delivery driver electronically captures information about each package. Overnight, package delivery information such as package shipper name and address, class of service, charges, ship to name and address, driver route and stop information for the current day is pre-loaded into the DIAD from systems outside the DIALS environment. The driver collects any remaining package delivery information as part of the delivery event; for example, delivery location, time of delivery, and the signature of the person receiving the package.

For guaranteed delivery packages, the minimal information necessary to track a package is transmitted from the DIAD via cellular telephone directly from the package delivery van and uploaded to the DIALS database. For all packages, complete information is uploaded from the DIAD when the delivery van returns to base each evening. In both cases, the information is passed to a CICS transaction which interacts with IMS DBCTL to load the information into an IMS Fast Path database. IMS BMPs then extract the information, load it into the DIALS DB2 databases, and delete it from the Fast Path database after it is successfully loaded to DB2.

The information is then available for customers to track their packages or to verify proof of delivery. For guaranteed delivery packages, the delivery information is generally available to customers 20 minutes after delivery. The uploading of all information when the delivery van returns to base causes very high levels of DB2 insert activity (up to 7 M inserts per hour) in the late evening and early night.

A diagram showing the flow of data in DIALS is contained in Figure 1 on page 9.

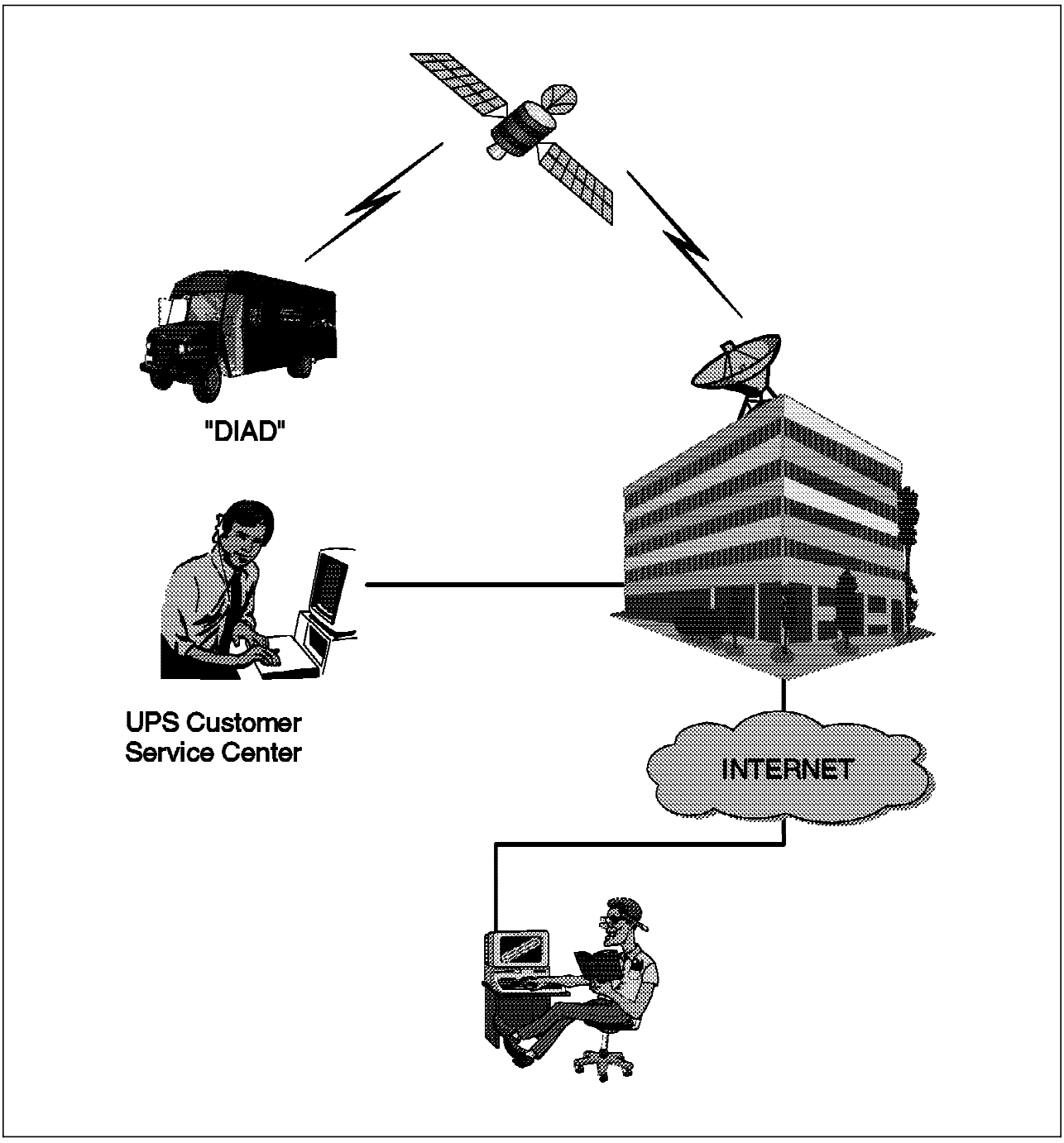


Figure 1. DIALS Data Flow

2.3.2 Online Processes

Online synchronous and asynchronous query processes search the database to provide package delivery status information. The synchronous queries are typical CICS 3270 queries used if the user has certain required package identification information. The longer-running asynchronous queries are PC-driven queries used for longer-running searches where the user only has a subset of the required information.

2.3.2.1 DIALS CICS Transaction Mix

The delivery information is accessed by several online CICS transactions in order to satisfy requests for confirmation of actual delivery. Table 2 contains a summary of some of the typical types of transactions and their daily volumes before migrating to data sharing.

Table 2. DIALS CICS Transaction Volumes - Before Data Sharing

Transaction	Description	Daily CICS transaction rate	Daily DB2 execution rate
OLI	OnLine Interface is the key online query process used by internal UPS customer service personnel to verify telephone inquires related to a package.	700,000	395,000
ADI	Automatic Delivery Information system is used internally within UPS to monitor and audit the package delivery process.	88,000	25,000
TIPS	Tracer Information Process System allows customer service to initiate a search to locate a package when the UPS client has incomplete information about a package. This request generates a broader search inquiry against the DIALS database.	500,000	122,000

2.3.3 Batch Processes

A started task, frequently referred to as a *propagator*, is used to transfer information into the DIALS DB2 database environment from IMS. The propagation workload consists of a suite of IMS BMP jobs that are used to select updated or inserted records from the IMS Fast Path database and apply those inserts and updates to the DIALS DB2 database.

2.3.4 DIALS Database Design

DIALS contained over 16,000 DB2 data sets and 100 IMS data sets amassing 4.2 TB of DASD and used 426 MIPS (using UPS' rating of processor MIPS) at the start of the data sharing project.

To support very large tables, DIALS employs logical tables. Each of 11 logical tables are further subdivided into 22 physical tables called *instances*. Each instance contains 28 partitions. There are also 35 global (or non-instanced) tables. Records are assigned to one of the instances using an algorithm designed to provide an even spread across all instances. The global tables act as an index to the instances. Each instance contains data for 56 weeks. Each partition contains up to two week's worth of data. The instances contain all information that is common to every package.

Access to DIALS data is index-driven; indexes are scanned in order to locate qualifying rows containing package information. DIALS does massive insert activity into both instance and global tables and each behaves quite differently:

- Inserts to the instance tables and global tables go at the end of the tables.

- Inserts to global indexes are random.
- Inserts into instance indexes are in skip sequential mode.

Inherent in the design of DIALS is a level of physical separation of resources. Should a DASD volume fail and make one of the databases unavailable, only a fraction of the data would be affected, and the records are distributed such that no one area or office would lose access to all their information. While this is not as much of an issue now that RAID devices are the norm, it demonstrates the importance that UPS places on the availability of this application. Also, the division of the data into so many partitions provides great flexibility to stop just a small portion of the data to do a database reorg while keeping the majority of the application available to users.

2.3.4.1 DIALS Bufferpool Definitions

DIALS local bufferpools are configured using BP0 for the DB2 catalog and directory databases, BP1 for tablespaces, BP2 for indexes, BP7 for work file tablespaces and BP8 for vendor product tables. The size of the various buffer and Hiperpools at the start of the data sharing project is shown in Table 3.

Table 3. DIALS DB2 Bufferpool Definitions

Bufferpool	Use	VP	HP
BP0	Catalog/Directory	2,000	2,000
BP1	Tables	58,000	116,000
BP2	Indexes	150,000	150,000
BP7	Sort	3,500	0
BP8	Vendor Tables	500	0

2.3.4.2 DIALS DB2 SQL Activity

At the start of the data sharing project, the daily DB2 activity against the DIALS databases was as shown in Table 4.

Table 4. DIALS DB2 SQL Activity - 1997 (Before Data Sharing)

Type of DDL	Daily occurrence
SELECT	13,600,000
INSERT	36,000,000
UPDATE	805,000
DELETE	154,000

2.4 Data Sharing Project

While DIALS is a well-designed, robust system, its enormous size and business criticality dictate a more detailed risk assessment than any other application at UPS. Before it could be moved to data sharing mode, UPS needed to be absolutely sure that there would be no problems following the migration. Given the size of the DIALS databases, and the level of activity against the databases, reverting to non-data sharing mode was not an option.

To provide the required level of confidence, it was decided to stress test the DIALS application. Specifically, the stress test was to ensure that the application in data sharing mode would:

- Meet acceptable levels of performance
- Handle throughput expected during the peak season
- Have code stability
- Scale at minimum cost (overhead)

2.4.1 First Stress Test

Because of the size of its applications and the volumes of data it processes, UPS has a history of encountering product failures that other customers may never encounter. Successful completion of the stress test would give UPS the necessary preparation and confidence for a low-risk implementation of data sharing for the DIALS application.

A test Parallel Sysplex was already running in-house, and was used for basic DIALS testing to ensure there were no basic data sharing issues. However, UPS did not have sufficient spare CPU and DASD capacity to handle the size and volume the desired stress test would require.

UPS decided to use the System Verification Services facility at the IBM Washington Systems Center (WSC) in Gaithersburg, Maryland for the test. UPS selected a data sharing team representing personnel from all pertinent departments including operations, capacity planning, performance, DB2 and IMS DBAs, MVS, VTAM, CICS, IMS and DB2 systems programmers, applications and storage management.

Because of the size and complexity of the DIALS environment, only a subset of the data was transported to the WSC. Since the majority of inserts and queries are against the last two partitions, this subset was chosen. The key transactions (OLI, ADI, and TIPS) and data and business function dependencies were identified. It was impossible to test every possible transaction, so tests with transactions that were representative of the workload were selected. Also, as FVT and OCAS transactions have basically the same DB2 functionality as OLI, they were not included in the stress test.

Teleprocessing Network Simulator (TPNS) scripts were used to simulate online query workloads. Input transactions for each workload were captured from the production environment prior to the test. TPNS scripts and input had to be fine-tuned until calibration-run statistics matched the transaction profiles in production. Transactions that accessed parts of the database that were not going to be brought to the WSC had to be removed. There were difficulties calibrating the DIALS environment to run consistently in a non-data sharing environment for a baseline measurement. Detailed transaction profiles were created whose attributes included data manipulation language (DML) mix, activity, lock rates, BP activity, suspension, getpages, and so on. Some transactions had to be deleted and modified until the right mix was obtained. Different processes and transactions were captured at different times, and getting the correct corresponding DB2 and system data copied to take along was a challenge.

On the production system, the propagators were stopped long enough to allow 24 million inserts to build up. The IMS system and its databases were then

backed up. For the test, the propagator tasks would use this data to simulate the database package delivery information insert jobs. Peak production conditions projected for the 1997 holiday season were simulated for both workloads.

At this time, DIALS was still running on DB2 Version 3, so this had to be migrated to Version 4 to provide data sharing support. Type 1 indexes were converted to type 2. The baseline measurements for the stress test were then taken.

Enabling the DB2 subsystem to 1-way and then to 2-way data sharing was done for the first time and some operational errors occurred. DB2 bugs, unique to UPS' workload characteristics, were detected, diagnosed and quickly fixed. OEM vendor utilities were upgraded and tested and some bugs were also detected.

There were issues encountered during recovery testing that were not resolved to UPS' satisfaction. Throughput was within their expectation but response time and overhead was higher than expected.

The test indicated that more work was required before DIALS could run acceptably in a production data sharing environment. The issues that arose strongly suggested to UPS management that they were not yet prepared to place their most mission-critical application in data sharing. Based on the results of the test, UPS decided not to move forward with the DIALS DB2 data sharing project at that time, especially in light of projected forecasts that they would be able to get through their 1997 peak within the capacity of an IBM 9021-9X2 for each of their major applications. In addition, UPS felt that their operational procedures for backup and recovery, as well as for planned and unplanned outages, needed to be documented, more thoroughly tested, and understood by their operations and technical support staff before a move to data sharing could be considered.

2.4.2 Another UPS Application: InfoCenter

It was decided to move a less critical application to data sharing first, in order to give UPS a higher level of comfort and experience with this new technology.

A small application, InfoCenter, was chosen as the first candidate for production DB2 data sharing. This application had a much lower profile, but as it was also DB2-based, it would give UPS practical experience with all aspects of DB2 data sharing.

InfoCenter is basically a read-only application using QMF with a lot of dynamic SQL. Although it is not similar in functionality to DIALS (it is TSO- rather than CICS- and batch-based), it was felt that it would give the operations staff familiarity with the new commands and procedures and could be used as a training ground in preparation for the real thing for the entire staff. Everyone involved learned many new concepts and became familiar with data sharing naming conventions, activating policies, fine-tuning production migration procedures, and procedures for moving Coupling Facility structures for planned outages, among other tasks.

UPS used this time to gain more experience in preparation for DIALS data sharing. A detailed migration checklist was continually updated with additional tasks so that when DIALS would be migrated to data sharing, the precise steps needed would be thoroughly documented.

2.4.2.1 InfoCenter Results

The objective throughout was to get to data sharing with a high degree of confidence and a minimal amount of error. With InfoCenter, UPS would get the needed reassurance of code stability and reliability, assuming all the following:

- There were no software failures.
- There were no extended outages to both members at the same time.
- There was minimal impact to the UPS production environment and overall business.

The 2-way data sharing implementation of InfoCenter was completed in November 1997 and was judged a success. This provided UPS with a proof of concept, real-life experience with data sharing, and paved the way for another attempt to move DIALS to data sharing mode.

2.4.3 Second Stress Test

Another trip to IBM's Washington Systems Center (WSC) was scheduled for early 1998. The objective was to apply all the experience gained in the InfoCenter migration to DIALS to check UPS' processes and procedures and also check the performance of DIALS in a data sharing environment. This test was not to be merely a performance stress test, but was also to ensure that the UPS technical staff would be ready for the operational changes needed in a data sharing environment. This test was named the Systems Support Readiness Test (SSRT).

2.4.3.1 Test Setup

The 2-way environment from the previous stress test had been saved to tape. This saved considerable setup time and allowed testing to begin almost immediately after the data was restored. Maintenance to the DB2 subsystem brought it up to a current level and OS/390 R1.3 was also installed. The test was run with DB2 Version 4 since this was going to be used in the production environment. However, it was understood that further performance benefits could be expected when UPS moved to DB2 Version 5.

Analysis of the performance reports from the SMF data gathered from the first stress test indicated that the DASD configuration had to better match the DIALS production environment in order to eliminate the I/O bottlenecks encountered in the first test. That first stress test used only 300 DASD volumes; this test would require over 800 volumes. The DIALS and system datasets were redistributed and spread across 800 volumes and control units similar to the UPS production environment. RMF and CMF performance reports on a volume level were scrutinized until an optimal DASD configuration was found with virtually no I/O bottlenecks. Full volume backups were then taken which would be used to restore the environment prior to subsequent test runs. If datasets were moved after these backups were taken, new backups would be generated. Volume backups required 2,500 tape volumes and subsequent restores took 6 hours using 64 tape drives!

Compared to the first stress test, the Group Buffer Pools were reduced in size. All of the inserts and updates in DIALS are for the very latest transactions, whereas most of the queries are against data that is less current. This means that most of the data that is written to the Group Buffer Pools at commit time is not referenced again until after it has been cast out to DASD. For an application of this type, there is little or no benefit in having very large GBPs. Having very large GBPs increases the processing involved any time they need to be

scanned, and increases the amount of data to be moved at castout time. On the other hand, the GBPs must be sized large enough to be able to handle the peak insert rate without filling up faster than the changed records can be cast out. UPS felt that the current GBP sizes were the best balance between these two requirements.

Another change implemented for the second stress test was to reduce the GBP thresholds to spread the castout activity more evenly. Also, the checkpoints were defined to minimize the number of times when both systems would be taking checkpoints at the same time. The service levels for CICS, IMS, and DB2 were all higher, and MVS 5.2.2 was replaced with OS/390 1.3 for the second test. It was felt that some of the performance improvement was due to changes introduced by the intervening service. Since moving to data sharing, UPS has attempted to keep the DIALS DB2 software at as current a service level as is reasonable, to minimize rediscovery of problems and also to be sure of picking up any performance enhancements that are introduced in the service stream.

It is important to note that the IMS databases used for the tests would only support at most 2 hours of inserts. When the data was exhausted, or if there were any problems encountered, a full restore had to occur to ensure test repeatability. Knowing the restore was a lengthy process, it was generally run at the end of the day when all testing was complete. The schedule was tight and when a successful run was made, the remaining data was used to do recovery testing so as not to waste precious time. Each test generated 100 tape volumes of SMF data to be copied and analyzed for a review meeting the following morning.

Various hardware failures were simulated, such as varying links to the Coupling Facilities off-line, deactivating each of the Coupling Facilities, and disabling critical DASD during periods of peak activity, to test and validate operational and recovery procedures. The MVS systems programmers crashed MVS and left it to the operators to control movement of CICS regions from the failing MVS to the other. Operations personnel were left to recover tasks using the existing documented procedures. Based on the results of the recovery tests, the operational procedures were modified.

2.4.3.2 Run Descriptions

Several tests were run to measure overall response time as well as the DB2 and CICS response times for specific transactions (OLI, ADI, TIPS), throughput and overhead. *Throughput* is the number of CICS transactions/second, as well as DB2 plan executions per second along with the number of DB2 inserts per hour. *Overhead* calculations are based on the number of processor MIPS required to run the application in each environment. Since the test system was not exactly like the production environment, the production overhead would be extrapolated based on December 1997 production measurements.

The first test was run in 2-way data sharing mode using a 9021-9X2 and an LPAR on a second 9021-9X2. This test demonstrated the ability to run the projected DIALS 1998 workload in a data sharing environment.

Since the projected workload would exceed the capacity of a single 9021-9X2, it was not possible to run this workload in data sharing and non-data sharing mode for a direct comparison. However, response times and the total MIPS used were measured and compared to the 1997 base measurements to extrapolate the data sharing overhead.

Two runs were made using the projected DIALS 1998 workload.

Availability Run: The CICS transactions for the applications OLI, ADI and TIPS were each split across the two members of the data sharing group. All the propagators (IMS BMPs) ran on one member. This configuration provides for continuous availability of the CICS transactions. If one member fails, the CICS transactions would still be available on the remaining member. This run was called the *high availability* configuration.

Capacity Run: In this run, the ADI and TIPS transactions were run on one member with the propagators and OLI on the other. This method of splitting the workload was predicted to cause less data sharing overhead than the availability configuration, because of reduced inter-DB2 read/write interest. This approach was considered minimum risk since the new data sharing code was exercised less frequently due to the smaller number of GBP-dependent pagesets.

In this configuration, if a member fails, the function on that member would not be available until the failing member is recovered or the function is moved to the other image. This configuration was envisaged as a way of moving the CPU-intensive transactions to a second processor, thus protecting the response times of those transactions as the CPU got busier, and also as a way of providing more capacity for the enhanced MMS application which was expected to require significant processor resources on the processor running IMS. This run was named the *capacity* configuration.

2.4.3.3 Overhead Calculation Methodology

For all the runs, the data sharing overhead was calculated using a method developed by Gary King from the IBM S/390 Performance Evaluation Group. This has been shown to be accurate within 2% to 3%. This methodology is documented in the redbook *Batch Processing in a Parallel Sysplex*, SG24-5329, and uses a fixed software and hardware cost based on the Coupling Facility activity and processor speed.

A set of runs was made to measure the processor overhead and validate Gary King's methodology. This set of runs used a scaled-down DIALS workload that fit on a single 9021-9X2 with rates similar to the then-current production rate. This workload was run on a single 9021-9X2 in 1-way data sharing. It was then split across two data sharing members on the same 9021-9X2 and run again to match, as close as possible, the same volume of the 1-way run. The response time and MIPS used was measured and a valid comparison of the data sharing impact could then be made with confidence.

Having compared a number of runs using actual measurements, and comparing those results with the output of Gary King's calculation, UPS had confidence in the accuracy of this formula.

2.4.3.4 Test Results

These four sets of runs helped to provide UPS with an accurate assessment of how DB2 data sharing would perform in their high-volume production environment. It also proved the capability of supporting this new architecture in their environment and gave the technical and operations staff at UPS the high level of confidence needed to continue the project in their production environment.

UPS rates an IBM 9021-9X2 at 464 MIPS. Given this rating, a true 9021-982 would be rated at 385 MIPS. However, the processor that member 2 ran on was actually an 8-way dedicated LPAR of a 9021-9X2, with TPNS running on the other two CPs. Thus, this system ran a little slower than a true 9021-982. The best estimate taking into account LPAR effects is that it would run at about 378 MIPS.

Normalizing the Results: The production workload measured in December 1997 was during the peak query production hour on the peak query day. In Table 5 on page 19, one might notice that the number of DB2 inserts/hour is only 240,000 on the production system. Ordinarily the peak query activity is during the *day* and the peak insert activity is in the *early evening*. The workload run during the stress test did not actually represent the workload characteristics in production. The aim during the test was to simulate the worst case scenario where both query and batch would peak at the *same time*. This worst case scenario can happen in DIALS if there are system problems or the propagators back up, leading to a burst of activity when things return to normal.

Also, enhancements to an existing application, Mobile Message Switch (MMS), were expected to go into production, and this would change the current workload mix due to its doing realtime inserts to DIALS during the day. It was expected that the MMS changes would increase the day time insert rate tenfold from previous levels up to about 2,000,000 per hour.

Another apparent anomaly was in the response times for the TIPS transactions. The TIPS transactions all use dedicated threads. However, the accounting record for the DB2 thread associated with the TIPS transaction in production is cut once, when the thread terminates, since the CICS RCT contains the parameter TOKENE set to NO. In the test environment, the run was made with TOKENE=YES which specifies that the accounting record is to be cut upon termination of the program. As a result the response times for TIPS in the test environment appear to be shorter than those for the production environment. To be sure of the validity of the figures, UPS changed one of the production TIPS CICS regions to also use TOKENE=YES, and the subsequent response times were in line with those found during the test. As a final reality check, UPS checked the DB2 processing associated with the TIPS transactions and found that the measurements from the test were in fact very similar to those from the production system.

To translate the results to reflect the actual production environment, the percent of MIPS used by DB2 in 1-way data sharing was compared to the percent of MIPS used by DB2 in production on December 1, 1997. In 1-way data sharing, DB2 consumed 50.7% of the MIPS, while in the production measurement, DB2 consumed 31.7% of the MIPS. From this, the ratio 31.7/50.7 = .625 was calculated. Thus, the estimated production overhead is .625 * the benchmark overhead.

Availability Configuration Results: Even with a transaction rate of 130.5 Transactions Per Second, and an insert rate of 5.985 million inserts/hour, the response times for TIPS, OLI and ADI were well under the success criteria. The Gary King methodology was used to calculate the additional MIPS required. The results estimated an increase of 14.7% MIPS, which translates to a 9.2% increase in production MIPS usage.

Even more important to UPS was the consistency of the CICS and DB2 performance. Acceptable average response times were not sufficient. UPS had to be confident that there were no instances of transactions getting very poor

response times, but being masked by a large number of transactions with good response times. To satisfy themselves, UPS looked at every plan execution to make sure there were no such instances. Of 90,000 plan executions, there was not one that was over 1 second. This not only gave the UPS technical staff the confidence to proceed, but also to be content that any performance issues that arose after the migration were not caused by data sharing. And in fact, this has proved to be the case.

Capacity Configuration Results: In this test, the response time results were similar to the availability configuration, resulting in an overall production overhead of 8.8%. This also proved that moving the CICS regions around did not adversely affect the performance characteristics or have a significant impact on the data sharing overhead.

In a separate test designed specifically to measure the maximum achievable insert rate, just over 18 M inserts per hour were achieved. It was vital to UPS that the data sharing configuration could handle this level of activity. In case of a prolonged component outage, it is possible for this level of activity to build up, and UPS had to be confident that implementing data sharing would not impede their ability to catch up in a very short time. The results of this test gave UPS complete confidence that data sharing would not adversely affect the throughput capacity of the DIALS DB2 subsystem.

1-way to 2-way Comparison: For the 1-way to 2-way comparison, the following information was collected from the Coupling Facility activity and system reports to determine the extra MIPS used in the 2-way test:

```
1-way:  83.2% CPU busy * 464 MIPS = 386.0 MIPS used
2-way:  65.6% CPU busy on member 1 * 464 MIPS = 305.78 MIPS used
        38.4% CPU busy on member 2 * 378 MIPS = 145.15 MIPS used
```

In the 2-way run, the transaction and insert rates match very closely to the 1-way run, totaling 450.93 MIPS used. To account for the slightly lower query and insert rates in the 2-way, the total MIPS were uplifted by 0.7% or 3.16 MIPS. This yields a new adjusted total for the 2-way of 454.1 MIPS, representing a growth of 17.6% over the 1-way configuration.

The results are summarized in Table 5 on page 19.

Table 5. Stress Test Results								
	CICS txns/sec	**DB2 ins/hour**	**TIPS resp. time**	**OLI resp. time**	**DD31 resp. time**	**ADI resp. time**	**Benchmark overhead**	**Production overhead**
Success criteria	89.0	5.800 M	10 sec	1.0 sec	1.5 sec	5.0 sec	25%	15%
Availability configuration	130.5	5.985 M	1.55	0.36	0.44	1.68	14.7%[1]	9.2%[1]
Capacity configuration	129.4	6.251 M	1.94	0.39	0.51	1.22	14.0%[1]	8.8%[1]
1-way comparison configuration	106.4	4.850 M	1.36	0.33	0.40	1.31	N/A	N/A
2-way comparison configuration	106.0	4.800 M	1.32	0.36	0.45	1.08	17.6%/ 16.9%[1]	11.0%/ 10.6%[1]
Production 12/97	124.0	240,000	7.09	0.42	0.53	2.06		

2.4.3.5 Test Benefits

The contribution and involvement of the various technical support personnel from UPS in each step along the way provided the required hands-on experience in implementing, operating, monitoring and tuning this new environment. The skills transfer from participation in such a test prepared the UPS technical staff for the implementation of DIALS data sharing in an actual production environment.

2.4.3.6 Summary

UPS deemed the test a success. Recovery was tested in every scenario possible, the software ran reliably, and performance exceeded the stated goals. In one test, known as the *maximum stress run*, close to 190 transactions per second was achieved, exceeding all expectations. In addition, the UPS technical support and operations staff received valuable training and experience.

The performance and recovery testing done in Gaithersburg proved that DIALS could perform acceptably in the data sharing environment. The bottom line was that even with a 15% overhead, UPS did not see an increase in response time. In addition, UPS gained the confidence that if the major types of transactions are evenly split across both members of the data sharing group, performance would not be adversely affected. This arrangement provides for maximum availability and gives UPS the most flexibility in running their workload, especially if the need arises to run without one member of the data sharing group.

[1] These results were computed using Gary King's methodology

2.4.4 Project Timetable

The elapsed time from the start of the data sharing project to full implementation was about 1.5 years. This may seem a long time; however, consider that in this time, UPS moved from having no data sharing to a faultless implementation of data sharing for their most critical application.

The most important factor of the project was that availability of this key application was maintained or improved. Fallback to non-data sharing mode was not an option, and any application outages caused by the migration would not be acceptable. The timetable for the project is shown in Table 6.

Table 6. DIALS Data Sharing Implementation Timetable

Date	Milestone
02/05/97	Test subsystem migration to 1-way in development
02/21/97	Test subsystem migration to 2-way in development
03/97 - 04/97	First stress test
11/23/97	InfoCenter migration to 1-way
12/08/97	InfoCenter migration to 2-way
01/98 - 02/98	Second stress test
04/19/98	DIALS migration to 1-way
04/22/98	DIALS migration to 2-way

2.4.5 Current IT Infrastructure

Compared to the environment at the beginning of this project, there have been changes to both hardware and software in support of data sharing. The DIALS processors and Coupling Facilities have been upgraded. And the software is now at the levels required to support DB2 data sharing. The details are provided in the following sections.

At the time of writing, UPS is just about to install a third DIALS member, initially only in standby mode. This third member will be ready to be activated should additional capacity be required over the peak holiday period.

2.4.5.1 DIALS Data Sharing Group

Figure 2 on page 21 provides a diagrammatic representation of the hardware involved in the DIALS data sharing group.

Table 7 on page 22 provides more detail about the processors involved in the DIALS data sharing group at the time of writing. As UPS are constantly upgrading their hardware, the configuration will more than likely be different by the time this book is published.

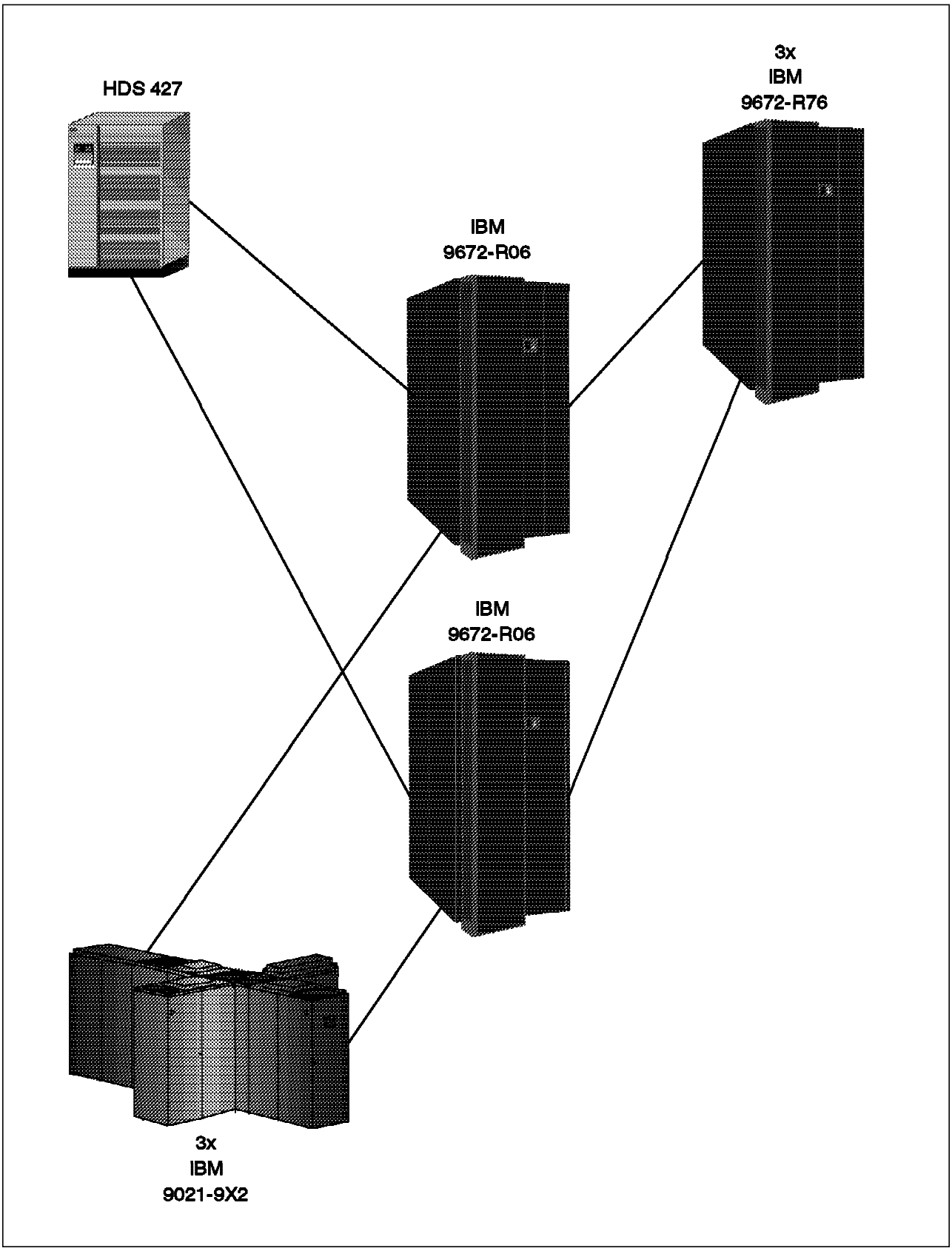


Figure 2. Current DIALS Hardware Configuration

Table 7. DIALS Data Sharing Complex				
CPU Type	**Central Storage**	**Expanded Storage**	**MIPS (as rated by UPS)**	**Number of CPs**
IBM 9021-9X2	1,392	2,784	464	6 (LPAR)
IBM 9672-R76	2,016	6,144	713	7
HDS 427	2,048	4,096	539	4
IBM 9672-R06	2,048	2,048	N/A - CF	3
IBM 9672-R06	2,048	2,048	N/A - CF	3

2.4.5.2 Current Software Levels

UPS planned to upgrade to DB2 Version 5, IMS Version 6, and OS/390 Version 2.6 in the first quarter of 1999. At the time of writing, the software levels installed on the DIALS systems are as follows:

- OS/390 JES2 Version 1.3
- DB2 Version 4.1
- CICS Transaction Server 1.2
- IMS/ESA Version 4.1
- Omegamon for CICS V400
- Omegamon for MVS V400
- Omegamon for DB2 V400
- BMC Unload Plus 3.1
- BMC Reorg Plus 4.2.02
- BMC Recover Plus 2.3.0
- BMC Load Plus 3.1.00
- BMC Copy Plus 5.2.0
- Responsive Systems Buffer Pool Tool 4.3

2.4.5.3 New Releases That Were Required

UPS uses many vendor product tools and utilities for DB2. It was necessary to test each vendor product in a data sharing environment. To avoid any possible problems, UPS ordered and installed the latest service or release available for each product prior to the test.

The DB2 PM monitor, the DB2 DISPLAY GROUP and GROUPBUFFERPOOL commands and Omegamon for DB2 provide useful information for monitoring and tuning DB2 data sharing and collecting historical data. The DB2 DISPLAY DB LOCKS command is also very useful, but should used judiciously as it can have a negative impact on DB2 performance while it is running. UPS also uses Boole and Babbage's CMF and Strobe and Candle's Omegamon for monitoring system-wide performance.

The Candle Command Center feature of Omegamon will provide UPS with a sysplex-wide view of performance and GRS enqueue contention. This product is in the process of being tested at the time of writing. The DB2PLEX feature of the Candle Command Center is still in test and is expected to be migrated to production shortly.

2.4.5.4 Current CICS Transaction Mix

The current transaction levels are shown in Table 8. The numbers of transactions have changed both because of growth in business volumes, and because of the introduction of the ability for customers to query the status of their parcel through the Internet.

Table 8. DIALS CICS Transaction Volumes - 1998 (After Data Sharing)

Transaction	Daily CICS transaction rate	Daily DB2 execution rate
OLI	526,000	307,000
ADI	392,000	111,000
TIPS	200,000	51,000
OCAS (Internal)	2,053,000	565,000
OCAS (External/Internet)	1,325,000	760,000
FVT	2,525,000	528,000

2.4.5.5 Current DIALS DB2 SQL Activity

The application has changed significantly since the project began; propagators run during the day and transactions check to see if the data is already there, performing a delete followed by an insert.

Therefore, the figures shown in Table 9 reflect both business growth and variances caused by application changes.

Table 9. DIALS DB2 SQL Activity - 1998 (After Data Sharing)

Type of DDL	Daily occurrence
SELECT	19,800,000
INSERT	48,000,000
UPDATE	805,000
DELETE	11,990,000

2.4.5.6 Current Bufferpool Definitions

The current Bufferpool specifications are shown in Table 10. Note that the same definitions are used on each of the DIALS subsystems.

Table 10. Current DIALS Bufferpool Specifications

Bufferpool	Use	VP	HP
BP0	Catalog/Directory	2,000	2,000
BP1	Tables	57,000	170,000
BP2	Indexes	149,000	150,000
BP7	Sort	3,500	0
BP8	Vendor tables	500	0

2.4.5.7 Coupling Facility Structures

UPS has updated its Coupling Facilities to take advantage of the performance and availability benefits of the latest level of IBM Coupling Facility technology. UPS uses the Coupling Facilities for the full range of resource sharing exploiters, as well as DB2 data sharing.

The other current exploiters of the Coupling Facilities are CICS Transaction Services Logging, XCF signaling, IEFAUTOS (tape sharing), JES2 checkpoint, Operlog, LOGREC, VTAM Generic Resources, GRS Star and RACF.

The Group Buffer Pool sizes for the DIALS subsystem, together with other key DB2 values, are shown in Table 11.

Table 11. DB2 Structure Sizes

Structure	INITSIZE	SIZE	DIR RATIO	CLASST	GBPOOLT	CHECKPOINT
GBP0	10 M	15 M				
GBP1	143 M	160 M	18	1	10	5
GBP2	300 M	450 M	13	1	10	3
GBP8	3 M	4 M				
LOCK	64 M	128 M				
SCA	16 M	48 M				

2.4.6 Potential Impediments

At the start of the project, UPS and IBM looked at the full set of products and facilities used by DIALS and identified all areas that could limit UPS' ability to fully exploit the benefits of Parallel Sysplex. A number of items were identified, and each was studied to see if it was significant enough to negate the value of data sharing. In the end, it was decided that while these other issues would reduce the full value, there was still sufficient benefit in using DB2 data sharing to justify the project.

The use of IMS by the DIALS application was a potential issue given that IMS was (and at the time of writing, still is) at a level that did not support Fast Path data sharing.

Also, both of the DIALS DB2 subsystems would be using the same set of DB2 system libraries. This would affect the ability to provide continuous availability as both DB2 subsystems would need to be stopped together to apply maintenance or new releases of DB2.

UPS currently shuts down DB2 to take weekly disaster recovery dumps of the DB2 system datasets. This requires a complete shutdown of the DIALS data sharing group.

The DIALS CICS regions do not use MRO, and affinities have been identified that have to be addressed before MRO is implemented.

APPN would not be fully implemented in time to provide VTAM Generic Resources (VTAM GR) support at the same time as DB2 data sharing went into production.

It was not known if the session manager used by UPS supports VTAM GR. As all users logon to DIALS through the session manager, its support of VTAM GR would be required in order to get the full benefit of Generic Resources.

All of these potential impediments are discussed in greater detail in the following sections.

2.4.6.1 IMS Data Sharing

The lack of IMS data sharing forces the propagators, which are responsible for all updates to the DIALS database, to remain on one member. This does not affect the availability of DIALS as such. However, it can have a serious impact in that the database will not reflect the latest deliveries if the propagators are not running. Any problem with IMS in the current environment would cause the propagators not to run until corrected. In case there are problems with IMS, up to 3 days of updates can be stored locally, and will be fed through to the IMS Fast Path database as soon as IMS becomes available again. Similarly, the propagators will start updating DB2 again as soon as IMS becomes available. During the stress test in the WSC, the IMS regions were successfully moved from one member to the other. The success of this test proved that the whole IMS subsystem could be moved between MVS images in case of a hardware failure.

New functions in the propagators, Mobile Message Switch (MMS) and System Notification Exception System (SNES), do real-time inserts to the DIALS database. These new functions were due to be rolled out around the same time that DIALS data sharing was enabled. UPS was concerned that as MMS and SNES grow, the IMS workload may exceed the storage and CPU constraints of the system it currently runs on.

IMS Version 6 has support for shared Fast Path DEDBs and this will alleviate the current limitation by allowing the propagators with the new MMS and SNES functions to run on all systems in the data sharing group.

Given that a version of IMS that supports data sharing was available from IBM, and could be implemented whenever UPS required it, it was decided that this was not a sufficient impediment to stop the DIALS data sharing project.

2.4.6.2 Maintenance Issues

UPS has two issues in relation to software maintenance. The first of these is the challenge of remaining at the "right" maintenance level. UPS participated in conference calls with several other customers to discuss how data sharing affects their maintenance methodology. Based on their own experiences and recommendations from IBM, UPS has been staying very current on maintenance for the data sharing environment. Traditionally, when a new system release is required or system software changes are installed, the UPS strategy has been to roll it out to all other DB2 subsystems before it is applied to the DIALS system. Now, new service is applied to the DIALS systems much earlier in the process. In addition, new PTFs are received in-house and maintained in separate SMP/E zones in case it becomes necessary to apply service at short notice. To date, this methodology has been very effective and UPS has suffered no unplanned outages of DB2 since moving to data sharing mode.

The other issue relates to the need to stop all DB2 subsystems concurrently to apply new maintenance. This is because all the DIALS DB2 subsystems run from the same set of DB2 target libraries. Only one set of libraries is used because the DB2 library names are coded into hundreds of UPS batch jobs and

procedures. When the second DB2 subsystem was initially set up, the easiest way to do this was to have it use the same set of libraries - changing all the JCL would have been a nontrivial task.

UPS is currently investigating the new Enhanced Alias Facility in DFSMS 1.5 to see if this will provide relief. This is discussed in more detail in 2.5.12, "DFSMS Enhanced Catalog Alias Support" on page 48.

At the time of writing, UPS is still investigating various methods of applying service, but has not yet settled on a methodology that will meet all their requirements.

While this issue is an inconvenience, it is not an overriding issue at the moment as the other PLDR applications contain sufficient information to support the user community while DIALS is down. Obviously, DIALS and the other applications would not be stopped at the same time. To ensure continuity, the other applications are not in the same data center as DIALS. Also, the DIALS DB2 stoppage is timed to coincide with other changes, such as NCP reloads or OS/390 IPLs. However, UPS is aware that this situation will change as availability requirements get more stringent, and they are currently investigating different methods of maintaining DB2 that will allow them to stop and service just one member at a time.

2.4.6.3 DB2 System Backups

Currently, all members of the data sharing group are shut down once a week so that full pack dumps of the DB2 subsystem volumes can be taken for fast recovery when needed, and also for disaster recovery. All the DB2s are stopped in an orderly manner to ensure there are no in-flight transactions. This permits a restart of DB2 from the restored datasets without the requirement for any of the DB2 archive logs.

As DASD and tape technology have improved, the duration of this outage has gotten shorter and shorter. UPS is presently looking at the latest DASD subsystem capabilities, such as SnapShot, to try to reduce this outage still further. Longer term, they are looking at alternatives so that the group-wide shutdown can be eliminated completely.

As stated for the maintenance issue, a short weekly interruption for DIALS is currently acceptable and therefore was not considered to be a reason to stop the data sharing project.

2.4.6.4 CICS Affinities

When UPS was in the process of implementing MRO on their various CICS systems, the DIALS CICS regions were investigated, but it was felt that implementing MRO was not technically desirable for DIALS at that time. Given the way UPS structures their session managers and the way they route CICS logons, every UPS Customer Service Center will continue to have access to DIALS even if a CICS region abends. This means it can provide high availability for the CICS users without having to implement MRO.

Since the original decision not to implement MRO for DIALS, IBM has made a number of enhancements to CICS that remove some of the affinities that have inhibited the use of MRO and dynamic transaction routing in other Parallel Sysplex customers. For example, it is now possible to have shared data tables and temporary storage queues resident in the Coupling Facility. UPS intended to

look at the DIALS CICS regions again in the future, and determine if MRO would be more suitable to their environment at that time.

The bottom line is that even though there are some affinities at the moment, UPS was still able to go ahead with the data sharing implementation and provide improved application availability without having to address all the affinities first.

2.4.6.5 VTAM Generic Resources (VTAM GR)

The full benefit of VTAM GR would only be achieved if the session manager supported this facility. As all DIALS CICS users logon through the session manager, it would need to support VTAM GR in order to allow users to logon to any of the session manager instances if the instance they are currently using were to fail.

However, as stated previously, it was felt that availability provided by the current implementation was acceptable at the moment. While VTAM GR would enable improved load balancing and faster re-logon following a failure, this is something that can be implemented at a later stage. The absence of VTAM GR when DIALS moved to data sharing mode was not considered to be an impediment.

2.4.6.6 Session Manager Support of VTAM GR

It was not known if the session manager used by UPS supported VTAM GR. However, there were two reasons that this was not considered to be an issue that had to be addressed at the time. The first reason, as stated previously, is that the availability provided to users at the time was considered to be acceptable. The other reason is that full implementation of VTAM GR and CICS MRO would require a redesign of the method used to assign users to CICS regions. Currently, the session manager will sign a given user on to a selected CICS region based on their user ID. And specific screens in each location are automatically logged on to a specific session manager instance. Thus every site has users spread over a number of session manager instances and CICS regions. This guarantees that every site will still have access to DIALS even if a CICS region or a session manager instance has a planned or unplanned outage.

The use of VTAM GR and CICS MRO, together with a redesign of the session manager signon method, would improve availability in case of an outage. But outages of the session manager or a CICS region were so rare, that an improvement in this area was not required at this time. Therefore, it was felt that this function could be investigated in the future and should not hold up the implementation of DIALS data sharing.

2.4.6.7 Impediments Summary

All of the potential impediments affect continuous availability. However, none of them would affect the implementation of DB2 data sharing. UPS decided that the benefits of data sharing were sufficient on their own to justify its implementation. The resolution of these issues will provide for even better availability, and will be addressed as business requirements for higher and higher availability increase.

2.4.7 Application Design Issues

DIALS is a well-behaved and well-tuned application that conforms to “good programming” guidelines. There were no issues regarding the design of the application from a DB2 standpoint that would prevent it from taking part in data sharing. Even though IMS Version 4 does not support data sharing, it was

decided that this constraint would not deter the project, even though IMS is confined to run on one member.

2.4.7.1 Required Changes

Very few changes to the application were necessary to prepare it for a data sharing environment. Naturally, it was important to ensure that the application as well as the data sharing configuration was well-tuned in an attempt to minimize the data sharing costs associated with requests to the Coupling Facility for cache and lock operations.

Most importantly to UPS was the fact that no application code changes were required. Some environmental and BIND option changes were made, and these are discussed in the following sections.

2.4.7.2 New CICS Regions

New CICS regions had to be created on the system with the new member in the data sharing group. All DIALS CICS regions are clones of each other. These regions are spread across the Parallel Sysplex, providing each service center with availability in case one member is down.

2.4.7.3 Type 2 Index Conversion

Most of the overhead associated with data sharing has to do with locking costs, so it was important to use type 2 indexes in order to reduce the number of locks. Type 1 indexes were in use for DIALS and conversion to type 2 was critical. All conversions to type 2 indexes including the DB2 catalog and directory occurred early in the project.

2.4.7.4 Reduce Parent L-Locks

The bind parameters RELEASE(DEALLOCATE) and RELEASE(COMMIT) control when the parent l-locks are released. With RELEASE(COMMIT), all parent l-locks are released at COMMIT and are reacquired when the next transaction begins. With RELEASE(DEALLOCATE), parent l-locks are released when the thread is deallocated.

Trying to design for more thread reuse, the bind parameter, RELEASE(DEALLOCATE) was used for the most frequently run transactions - DD31, KA47, and DDPI. Since these transactions were heavily used, they benefited the most from RELEASE(DEALLOCATE).

Other transactions continued to use RELEASE(COMMIT) since they were not as frequently used. Using RELEASE(DEALLOCATE) can increase usage of the EDM Pool as the life of a thread is much longer. Therefore RELEASE(DEALLOCATE) was not considered for short, infrequently used packages. One of the general recommendations for data sharing is to reuse threads whenever possible and to bind with the option RELEASE(DEALLOCATE). Depending on how much your threads get reused, this bind option can mean more EDM Pool storage is necessary for storing objects used by the plan or package. Remember to consider increasing the size of the EDM Pool when using the bind parameter RELEASE(DEALLOCATE) and thread reuse.

More protected threads for the same three CICS transactions also proved to be beneficial - the CICS RCT parameter THRDA was increased from 5 to 8 for each transaction. This change implementing more thread reuse, coupled with a rebind of its associated package with the RELEASE(DEALLOCATE) option, significantly reduced the number of physical locks (p-locks) taken. UPS still uses

RELEASE(COMMIT) for most plans and packages due to the impact on the EDM Pool caused by rebinding their collections with RELEASE(DEALLOCATE). The issue of EDM Pool usage was especially critical for UPS because of the fact that each package is replicated up to 42 times. So for a package that takes up only 8 KB, 336 KB must be allowed in the EDM Pool. Other customers with less data may not find the EDM Pool size to be as critical.

2.4.7.5 Vendor Utilities

One of the ISV utilities used type 1 indexes for its internal tables and a new release had to be installed to convert to type 2. In addition, one tool used to reorg tablespaces that spanned multiple volumes had some difficulties converting to type 2 indexes. Until this issue was addressed, these tables were isolated in their own bufferpool (BP8).

2.4.8 Implementation Issues

A number of implementation issues arose during the project and are discussed in the following sections.

2.4.8.1 Compatibility with Non-IBM Hardware

DIALS was moved to data sharing mode around the same time as it was moved to a new non-IBM processor. The measured overhead and CPU usage on this new processor was significantly more in this configuration than during the stress test in the Washington Systems Center, which had all IBM hardware in the sysplex. This led to some initial concerns about the performance of the data sharing configuration; however, these concerns were allayed when UPS more fully understood the reasons for the different performance profile.

It is important to understand that there is a much closer relationship between the systems in a Parallel Sysplex than there is in a traditional multi-system environment, so changes on one system can have a more significant effect on other systems in the sysplex than would have been the case in the past.

2.4.8.2 Shutdown of the DB2 Data Sharing Group

During periods of low activity, shutdown of the DIALS data sharing group takes approximately three minutes. During peak periods of high update activity, shutdown takes approximately 45 minutes. The shutdown time is an issue for UPS since, in an abnormal situation, the DB2 data sharing group may need to be recycled during problem determination and the increased shutdown time would lengthen any outage.

During the initial stress test, UPS was surprised to find that there was activity to the DB2 Coupling Facility structures during shutdown, even if they took a checkpoint just before the shutdown. Further investigation showed that this activity was caused by Coupling Facility DeleteName processing. The duration of this processing is directly related to the size of the GBP structures. As UPS initially had very large GBPs, DeleteName processing during shutdown was taking a significant amount of time. The DeleteName processing was for the pageset or partitions going out of GBP-dependency during physical close.

When a DB2 subsystem that is a member of a data sharing group is shut down, any structures that are owned by that member must have the ownership transferred to another surviving member. The amount of time that this transfer of ownership takes is directly related to the size of the structures. UPS found that the fastest shutdown could be achieved by leaving the member that is the

structure owner for their large group buffer pools to be the *last* subsystem to shutdown. If that member was stopped first, then DB2 would transfer structure ownership to the other member before shutdown, and this process elongated the shutdown of the data sharing group.

This issue will be resolved in a future level of Coupling Facility Control Code (CFCC) and a future release of OS/390, together with a Small Programming Enhancement (SPE) to DB2 Version 5 and Version 6.

2.4.8.3 BIND Considerations

When using different processors in a data sharing group, you need to decide on which member to run the BIND. The recommendation is to bind a plan or package on the processor where that plan is expected to run.

Different processors, as well as different DB2 configurations and customization values, may affect access path selection. It may be necessary to compare Explain output to make comparisons, which can be time-consuming. Constant monitoring and tuning help to ensure that efficient access paths are chosen. Reorgs should be run frequently to keep the data in clustering sequence, thus optimizing the decisions of the DB2 optimizer.

2.4.8.4 Archive Logs

An important issue for DB2 logging in a data sharing environment is to avoid reading archive log tapes in case a recovery is required. In a data sharing environment, IBM recommends keeping on DASD all the logs required for recovery from the last image copy. UPS had to size the DASD pool for the archive logs based on peak activity. The full impact of this recommendation, in terms of DASD space required, was not fully realized until quite late in the project.

Image copies are taken daily since UPS guarantees its users the fastest possible recovery from any failure. In case of a failure, the image copy will be restored, and forward recovery will typically be run using both the active and archive DB2 logs. In a 24-hour peak period, DIALS can generate 31 logs of 800 cylinders each. Each member in the data sharing group is sized for the same number of logs, for availability reasons. UPS archives logcopy 1 to DASD and logcopy 2 to tape. An SMS-managed pool was defined with 100 GB of 3390-3 volumes for the archive logs, with both members using the same pool. DFHSM manages the movement of ARCHIVE logs from DASD to tape and back, if required.

As DFHSM space management is based on allocation information contained in the VTOC (which does not go to a level of granularity lower than Julian date), the “primary days” DFHSM value had to be set to 2. If “primary days” were set to 1, when migration kicked off after midnight, any dataset created “yesterday,” which might actually be just minutes old, is eligible to be migrated. Thus UPS ended up having between 24 and 48 hours of archive logs available on DASD, effectively doubling the amount of space required (which had already been doubled because there are now two DB2 subsystems)!

There was also contingency allowed in case of a HSM failure, which could cause archive logs to remain on DASD for more than 48 hours. While all this contingency came with a cost, the ability to recover DIALS in the shortest possible time following a failure was the overriding concern.

Archiving to DASD can be a significant cost when you are dealing with the amount of logging in an application that is as insert-intensive as DIALS - few sites log as much as UPS. As DIALS is the only DB2 application in data sharing mode, it is currently the only DB2 subsystem at UPS that is archiving to DASD. Since the size of the DASD pool had to be large enough to accommodate up to 48 hours during a peak period, archiving to DASD was a costly issue.

Although IMS runs on only one system, if that system were to fail, IMS may be moved to the other system. Therefore, both DB2 subsystems' logs had to be sized to be able to cope with the full updating workload.

The reasoning for archiving to DASD is explained in the following extract from *DB2 Data Sharing: Planning and Administration*, SC26-3269.

"The RECOVER job needs at least one tape unit for each DB2 member whose archived log records are to be merged. (More might be needed if you run more than one recover at the same time for different partitions of a partitioned table space.) Therefore, do not archive logs from more than one system to the same tape.

Archiving to tape is not recommended because there can be negative consequences to not having enough tape units allocated: If there are not enough tape units to do the recovery, DB2 can possibly deadlock. If this happens, use the command SET ARCHIVE to increase the number of tape units that can be used.

If you must archive to tape, make sure the value for READ TAPE UNITS on installation panel DSNTIPA for each member is high enough to handle anticipated recovery work. For example, if you have 8 members, each member should specify at least 8 tape drives. You will need more if you run more than one recovery job at the same time on a given member, or if multiple members run recovery jobs at the same time.

Also, make sure you specify 0 for the DEALLOC PERIOD field on installation DSNTIPA to avoid making an archive tape inaccessible to other members of the data sharing group. (If you intend to run all RECOVER jobs from a single DB2, this suggestion does not apply.)"

2.4.8.5 Fallback to Non-Data Sharing

Once you decide to enable data sharing, fallback to non-data sharing is not recommended and it is unusual to have to do so. Although the fallback procedures are simple and well-documented, it does require a cold start. You must choose a surviving member since the logs used during data sharing are now unusable. Full image copies or non-DB2 backups of all data including the DB2 catalog and directory are required to establish a new recovery point. It is recommended that this step be performed as soon as possible after data sharing is disabled. Few customers have had to fall back and it is extremely rare to have to do so in production.

Fallback to non-data sharing was estimated to take 24 to 36 hours at UPS because of all the image copies required for DIALS. For availability reasons, fallback to non-data sharing at UPS was *not* an option.

2.4.8.6 Fallback from 2-way to 1-way

There was a general misconception that bringing down a member in a 2-way data sharing group would be equivalent to running in a non-data sharing environment. Assumptions were made that GBPs would not be used in a 1-way environment and that all GBPs would be de-allocated when the second member was quiesced. In DB2 Version 4, if there are GBP-dependent pagesets in the remaining member and that member is holding update locks against that resource, GBP-dependency does not go away. In a 1-way data sharing environment, it is still possible to see activity against the GBP structures in the Coupling Facility.

In DB2 Version 5, the new dynamic inter-DB2 R/W tracking enhancement helps to remove the GBP-dependency sooner.

2.4.8.7 Unique LRSN Timestamp

During time changes to reset the clock in the fall, operations staff keeps everything down for 61 minutes to prevent duplicate timestamps values in the DB2 logs. Prior to data sharing, the time-of-day value was not used in the DB2 logs, so this was not an issue. If UPS sets their Sysplex Timers TOD clock to GMT and then offset for the local time, this would not be an issue as DB2 uses the time from the TOD clock.

2.4.8.8 Optical DASD Devices

DIALS uses IBM 3995 optical DASD devices to store data older than 13 months, due to the lower cost per MB of those devices. TIPS is the only transaction that accesses the infrequently used data stored on these optical devices. The 3995 devices have a limited number of interfaces. Because of the hardware limitation of four channel interfaces, access from more than 2 CPUs is not possible. If UPS decides to use 3 members in the DIALS data sharing group, TIPS will only have the flexibility to run on one or both of the two processors connected to the 3995. While this is a restriction, it should still be possible to provide continuous availability for TIPS as at least one of those two systems should always be available.

2.4.8.9 MAXSPACE

In the event of a problem, it is important to gather complete documentation for every address space pertinent to the problem. Of course, ensuring that you have sufficient space for the dumps depends upon many factors including working set sizes for the IRLM, DBM1, MSTR, DDF and SPAS address spaces.

In the informational APAR II06335, IBM recommends 2500 MB for the MAXSPACE parameter of the MVS SDUMP command, which is significantly higher than the default of 500 MB. The value chosen should be tested to ensure that it is adequate. In addition, remember that dump services do not use ESTOR but do require auxiliary storage (see informational APAR II06471). As a result, UPS added auxiliary storage packs to provide the 3500 MB space needed in their environment.

For UPS, it is critical to ensure that any problems that arise can be diagnosed and fixed based on the information gathered on the first occurrence. It is not acceptable to have to suffer a reoccurrence and possibly an outage just to provide diagnostic information. If all the information can be collected on the first occurrence, it should be possible in most cases to install a fix before the problem is encountered again.

2.4.8.10 Restarting Failed Jobs

There used to be a restriction whereby failed DB2 utilities had to be restarted on the same member. APARs PN81211, PN81212, and PN81213 address this restriction and introduced the ability for utilities to use the DB2 group attachment name. If the PTFs for these APARs are applied, the utility can be restarted on any system if it was the utility that failed. However, if it was the DB2 subsystem that failed then the utility cannot be restarted until the failed DB2 member has been restarted.

2.4.8.11 CICS Support of Group Attachment Name

Each CICS region that communicates with DB2 must have the DB2 subsystem name specified in the Resource Control Table (RCT). CICS currently does not support the DB2 Group Attachment Name facility. This limits somewhat the ability to start a CICS region on any member of the data sharing group.

The attachment facility that is shipped with CICS 4.1 and subsequent releases has some advantages for use with data sharing. The old attachment facility has the subsystem name defined in the RCT, so if you want to attach a CICS region to two DB2 subsystems, you have to assemble two RCTs with different suffixes. The new attachment facility allows you to override the subsystem name on startup and with the INITPARM. This eliminates the need for a second instance of the RCT if you wish to connect to a different DB2 subsystem.

As UPS is using CICS Transaction Server 1.2, they have the new attachment facility and thus the ability to select the desired DB2 subsystem name at CICS startup without having to maintain two RCTs.

2.4.9 Implementation Lessons

The least understood and most difficult DB2 concepts to grasp in a data sharing environment are *global locking* and *castout processing*. While the underlying concepts appear straightforward, only actual experience will lead to a thorough understanding of how they work in practice.

Using firsthand experience in setting up the right environment for DIALS, many things were learned along the way by trial and error. DB2 data sharing tuning and performance is still something of an “art.” Experimentation with data sharing and advice from the DB2 development and performance experts at Santa Teresa Laboratory during the stress test proved very beneficial.

2.4.9.1 Local Bufferpool Sizes

UPS was pleased with the existing performance of the bufferpools and decided to stay with this winning combination. The bufferpool configuration on the originating DB2 member had been sized to achieve good performance in a non-data sharing environment, and the recommendation was made to use the same configuration for each member of the data sharing group. Therefore, local buffer pool sizes and thresholds in both the originating member and cloned member were defined to be equivalent to those in the non-data sharing DIALS subsystem. Although it appeared that this resulted in essentially doubling the amount of storage in the bufferpools, it was not yet known what workload would be routed to each member, so this sizing gave UPS the most flexibility.

2.4.9.2 Sizing Coupling Facility Structures

Although some guidelines are provided in the redbook *DB2 for MVS/ESA Version 4 Data Sharing Implementation*, SG24-4791, the accuracy of your estimate will only be known when the application is actually up and running. However, it is very important not to just apply the provided formulas to your numbers and blindly use the result. The formulas are based on an average workload. If your workload is different to the norm, then the results provided may not be appropriate for your installation. This in fact was the case for UPS, as described previously in 2.4.3.1, “Test Setup” on page 14. The provided formulas give a good starting point, but you must then apply a reality check as well.

The redbook *DB2 for OS/390 Capacity Planning*, SG24-2244 contains additional information for sizing the Coupling Facility structures.

Lock Structure Sizing: The lock structure was sized using the current 32 MB ECSA size in use for the DIALS environment. As a result, an INITSIZE of 64 MB and a SIZE of 128 MB seemed sufficient based on the methodology provided in *DB2 Data Sharing: Planning and Administration*, SC26-3269.

In the event of a Coupling Facility failure, the lock and SCA structures are rebuilt using information that is held in each member of the data sharing group. For availability reasons, it is strongly recommended to keep the SCA and Lock structures in a Coupling Facility that is failure isolated from any members of the data sharing group. UPS has adhered to this recommendation and has all its DB2 structures in external Coupling Facilities.

SCA Structure Sizing: The Shared Communications Area (SCA) list structure was sized based on the number of objects in DIALS according to the methodology provided in *DB2 Data Sharing: Planning and Administration*, SC26-3269. The number of unique objects not including views was approximately 2000, so an INITSIZE of 16 MB and a SIZE of 32 MB was used. Running out of space in this structure will cause DB2 to come down, so it is better to err on the side of caution and specify more space than needed. Because much of the space in the SCA is taken up with exception information, space is reclaimed by correcting the database exception conditions.

Group Buffer Pool Sizing: Coupling Facility Structures are defined using the IXCMIAPU program in terms of 1 KB increments, whereas DB2 bufferpool sizes are allocated in terms of 4 KB pages. Keep this difference in mind when deciding on cache structure sizes. Also, if changes are made to the local buffer pool configuration, be sure to re-evaluate its effect on the group buffer pool sizes.

The approach taken when sizing GBPs for DIALS is to have a directory entry for each unique page in each local bufferpool, including Hiperpools and pages in the GBP. The number of pages needed in the GBP was based on the current number of pages written, which was obtained from a DB2 PM statistics report. As an example, if there are 200 pages/sec written for BP1 and the time needed to keep those changed pages cached is 60 seconds, then 200*60 pages would be needed. This would require 12,000 4 KB pages in the cache structure for each member. Assuming 100% data sharing, this yields a total of 24,000 4 KB pages.

It is also important to ensure there are enough directory entries for the pages in the local buffer pool and Hiperpool and in the corresponding group buffer pool. The number of directory entries needed for a GBP is:

```
number of members * (Virtual Buffer Pool + Hiperpool) + number of GBP cache pages
```

For GBP1, this equates to 2*(58,000 + 116,000) + 24,000, or 372,000. Each directory entry is 208 bytes, so the number of bytes for directory entries is 75,563 KB. Add to that 24,000 4 KB pages and this yields 171,563 KB of storage. Structures are allocated in the Coupling Facility in increments of 256 KB, so rounding up to next 256 KB multiple gives an INITSIZE of 171776. The directory entry ratio is 372,000/24,000 or 16 to 1. Using this methodology sizes the group buffer pools according to their corresponding local buffer pool's use and need.

Having used the preceding value initially, UPS found that performance could actually be improved by *reducing* the sizes of the two largest GBPs. The reasoning behind this is discussed in 2.4.3.1, “Test Setup” on page 14.

2.4.9.3 INITSIZE Vs. SIZE

To give yourself flexibility in case you picked a size that is too small, it is a good idea to specify the maximum structure size (SIZE) of the structure somewhat larger than the initialize allocation size (INITSIZE), so that you can dynamically expand it, if required. However, specifying a maximum size *significantly* larger than its initial size is not a good idea, even if it gives you more flexibility.

The Coupling Facility structures contain some control structures that are static in nature, and at the time the structure is initially allocated, these static structures are allocated large enough to accommodate the biggest the structure could ever be. If the maximum size is very much larger than the INITSIZE, it is possible that most of the INITSIZE is used up for these controls, leaving you little in the way of entries and elements. In the extreme case, the static controls might need to be *larger* than the INITSIZE, in which case the structure may not be allocated at all!

Tests performed at Santa Teresa Laboratory to see the impact of different INITSIZE/SIZE combinations on directory/data element allocation revealed some interesting results. In one experiment for the SCA list structure with INITSIZE 1024 (1 MB) and SIZE 10240 (10 MB), DB2 failed to start due to an “SCA full” condition. Why? Too much of the 1 MB allocation is taken up by Coupling Facility static control structures, leaving insufficient space for the list elements used by DB2.

The results of lab tests for different INITSIZE/SIZE combinations for GBP cache structures are shown in Table 12. It appears that the higher the SIZE-to-INITSIZE ratio, the more control structure overhead that is consumed in the initial structure allocation, and the less space that is left in the structure for DB2's use.

Table 12. Effect of INITSIZE/SIZE Combinations

Structure	INITSIZE	SIZE	Allocated size	# dir entries	# data elements
GBP0	1024	1024	1024	926	185
GBP1	1024	2048	1024	926	185
GBP11	1024	4096	1024	926	178
GBP22	1024	8182	1024	865	169
GBP33	1024	16384	1024	763	151
GBP44	1024	32768	1024	497	99

To conclude, it is a good idea to specify a maximum size that is larger than the INITSIZE to give yourself flexibility. However, do not overspecify the Coupling Facility structure maximum size. In most cases, a 50% cushion is probably sufficient. For more information, refer to *PR/SM Planning Guide*, GA22-7236.

For DIALS, the structure SIZEs were initially specified as twice the INITSIZE, and then tuned based on actual experiences.

2.4.9.4 Coupling Facility Structures Placement

Structures are defined in a Coupling Facility Resource Manager (CFRM) policy, and placement of the structures is specified using the PREFLIST parameter. At UPS, it was decided to place the index group buffer pools in one Coupling Facility and the data group buffer pools, lock structure, and SCA structure in the other. All of the structures for the production DIALS system are in external Coupling Facilities and so have failure-independence.

Be sure that each Coupling Facility is large enough to hold *all* the structures, in case you need to move all the structures into a single Coupling Facility for planned or unplanned outages.

UPS also used a REBUILDPERCENT of 1. This ensures automatic rebuild of the LOCK and SCA structures in case any member of the DB2 data sharing group loses access to either of the structures.

2.4.9.5 Group Buffer Pool Monitoring and Tuning

For DB2 Version 4, the two factors that most influence the GBP recovery time are GBP checkpoint frequency and DB2 system checkpoint frequency. Monitoring and tuning group buffer pools is very important for good performance of the data sharing group.

Group Buffer Pool Thresholds: In an effort to ensure that changed pages are written out to DASD more evenly over time and avoid performance spikes due to intermittent floods of activity to the I/O subsystem, the local bufferpool deferred write (DWQT) and vertical deferred write (VDWQT) thresholds were used for the corresponding GBP thresholds, GBPOOLT and CLASST.

Group Buffer Pool Checkpoint: All changed pages in the GBP are cast out to DASD during each GBP checkpoint. The structure owner is responsible for supervising the castout process. The structure owner is the first DB2 that allocated the GBP due to accessing a pageset for the first time causing GBP-dependency. Between GBP checkpoints, pages will also be cast out if the GBPOOLT threshold is reached.

At GBP checkpoint, three activities take place in the Coupling Facility:

1. The Log Record Sequence Number (LRSN) of the oldest changed page needs to be found.
2. All changed pages need to be marked for castout processing.
3. The changed pages get transferred to the structure owner, who then writes them to DASD.

In the first stress test, the GBPOOLT threshold was set to 50, meaning that if 50% of the GBP pages were changed, enough pages would be cast out to bring the number of changed pages back down to 40%. It was found that this threshold was being reached up to 8 times between each GBP checkpoint. At

GBP checkpoint, significant delays were occurring. It was felt that these delays were largely caused by the amount of activity occurring in the Coupling Facility at checkpoint time.

Based on these experiences, UPS reduced the checkpoint frequency from the default of eight minutes, and changed them to 5 minutes for GBP1 and 3 minutes for GBP2. This would minimize the number of times that checkpoint processing for these two largest GBPs was overlapped. In addition, the GBPOOLT was reduced to 10%. As there were no performance problems during the first stress test when GBPOOLT was reached (thus causing 10% of the pages to be cast out), it was felt that this value should avoid performance problems at checkpoint time. Finally, the size of the GBPs was reduced, again reducing the amount of processing that was required at GBP checkpoint. As well as reducing the activity in the Coupling Facility, these changes also eliminated the I/O bottlenecks that were occurring in the first test at GBP checkpoint time.

UPS experimented with creating dynamic SQL to force each member to become the structure owner for different group buffer pools in an effort to evenly spread the GBP checkpoint and castout activity across the group. However, test results did not show that performance was enhanced measurably, so this approach was subsequently rejected in the production environment.

Group Buffer Pool Directory Entry Ratio: The group buffer pool contains information necessary for maintaining cache coherency. Pages of GBP-dependent page sets are registered in the group buffer pool. When a changed page is written to the group buffer pool, all DB2 subsystems that have this page cached in their buffer pools are notified that the page has been invalidated (this notification does not cause a processing interruption on those systems). This is called cross-invalidation (XI).

XIs also occur if there are insufficient directory entries in the GBP for all the pages in the local and group buffer pools. If a buffer has been cross-invalidated, that page must be read from DASD if it is needed again, rather than being found in the local buffer pool. If the number of XIs gets too high, this can have a negative impact on performance.

Initially, there was a significant number of XIs. While this did not cause a noticeable performance problem at the time, it was felt that the situation should be addressed before it got worse. Investigation showed that the XIs were due to a shortage of directory entries. Increasing the directory ratio from 11 to 13 reduced the number of directory reclaims due to XIs to zero.

2.4.9.6 DB2 Subsystem Parameters

Each member in a data sharing group has its own DSNZPARM settings and some of these parameters can be different for each member. Carefully evaluate the values of all the subsystem parameters in DSNZPARM for each member depending on the current activity. It was of particular interest to UPS to reduce the number of times datasets were going in and out of GBP dependency. The LOGLOAD, PCLOSET, and PCLOSEN parameters control this activity.

LOGLOAD: The DB2 customization parameter LOGLOAD controls the frequency with which DB2 bufferpool checkpoints are taken based on the number of log records. At each checkpoint, DB2 tries to externalize all changed data from its buffer pools to DASD. The DB2 performance recommendation is to take a

checkpoint approximately every 10 to 15 minutes in order to ensure quick DB2 restart time in case of failure.

Before the first stress test, the LOGLOAD parameter for DIALS was set to 3,000,000, causing a checkpoint to occur infrequently. The LOGLOAD for the DIALS subsystem was reduced to 300,000 which gives a checkpoint every 7 to 8 minutes during periods of peak insert activity. This value was found to provide the best balance between performance and maximum availability in case of a DB2 outage.

The LOGLOAD for both members can be different since each have a different update frequency. The updating member's LOGLOAD was set to 300,000, whereas the other member which does mostly query has it set to 60,000.

Prior to DB2 V6, a restart of DB2 is required to change this parameter. In DB2 V6, it can be changed dynamically while DB2 is running. This provides the flexibility to have a lower value on systems with lower levels of updates, while still being able to react to unplanned changes in the workload profile, without a DB2 outage.

PCLOSEN and PCLOSET: PCLOSEN and PCLOSET are two thresholds that control when page sets or partitions are converted from read-write to read-only state. PCLOSEN represents the number of consecutive DB2 checkpoints since a page set or partition was last updated. PCLOSET represents the amount of elapsed time since a page set or partition was last updated. The default for PCLOSEN is 5 checkpoints, and the default for PCLOSET is 10 minutes. If either of these thresholds is exceeded, the data set is converted to read-only status.

The pageset/partition P-lock is released when it is physically closed. Typically, once a pageset or partition is physically opened, it remains open until 99% of the maximum number of open datasets (DSMAX), is hit. In data sharing, there is an added consideration for those pageset/partition(s) that are GBP-dependent. The R/O DB2 member will physically close it to try to remove the inter-DB2 R/W interest assuming none of the following:

- It has not been referenced in PCLOSET minutes.
- It has not been referenced during the last PCLOSEN DB2 system checkpoints, and it is defined with CLOSE YES.

The recommendation is to keep PCLOSEN high and to use PCLOSET to control when data sets are physically and pseudo closed. However, it is also important not to have datasets going in and out of R/W to R/O switching too frequently. It is sometimes difficult to balance this activity. Monitor the R/W-to-R/O counter in DB2 PM statistics to ensure it remains no higher than 10 per minute.

DIALS encountered some high transaction response times after initial start-up of DB2, with DB2 class 2 elapsed time as long as 7.5 seconds instead of the expected 0.6 seconds. Analysis of these long-running transactions found most of the time was spent in Service Task Switch (STS) time. A trace of IFCIDs 170 and 171 indicated that the STS time was due to the physical allocation of data sets. As DB2 is normally only restarted at the weekend, and the subsequent level of activity is quite low, it was felt that they could live with this small number of long running transactions. However, to avoid having to reallocate datasets during normal processing, the PCLOSET value was increased to 30 minutes, thus reducing the likelihood of any datasets getting de-allocated.

To change PCLOSEN or PCLOSET, update the values in the DSNTIJUZ installation job, run DSNTIJUZ, and restart DB2.

2.4.9.7 GBP Recovery Process

Much was learned during the many recovery scenarios tested and the most important findings are described in the following sections.

Recovery of Tablespaces in GRECP Status: With an active workload running, a Coupling Facility containing the GBP cache structures was deactivated in order to test recovery procedures. Many tablespaces went into group recovery pending (GRECP) status and over 1000 pages were recorded in the logical page list (LPL). The DB2 members remained up and functioning in spite of the Coupling Facility failure. During this simulated failure, users accessing pagesets or partitions that were not affected could still function. Any transaction accessing GRECP objects received an SQLCODE -904 indicating that the resource is not available. In this scenario, it is important to get the objects out of GRECP status as quickly as possible.

The procedure in DB2 Version 4 to recover these objects is to issue appropriate START DATABASE commands for those database and tablespace names in GRECP status. As soon as GBP-dependent objects begin to be accessed, the GBP structures now required would automatically be allocated in the alternate Coupling Facility. UPS experimented with various ways to issue the START commands and timed each approach.

The time taken for the conventional approach to issue START DATABASE(*) SPACENAM(*) took almost 2 hours to recover. Recovery using the wildcard for both database and tablespace results in the log being accessed repeatedly for each object in each database from the last checkpoint.

Subsequent recovery tests used the output of a DISPLAY DATABASE(*) RESTRICT command to identify the databases that were in GRECP status. Using this information, the fastest recovery took less than 18 minutes using the START DATABASE(dbname) SPACENAM(*) command, which allowed for parallel I/O during recovery processing.

Remember to recover the DB2 catalog and directory databases (DSNDB01 and DSNDB06) first. It is also wise to recover the most important databases first, so prioritize the list of databases accordingly.

In DB2 V5, recovery of tables in GRECP status is automatic, so the question of how to identify the tables in GRECP and issue the START commands is irrelevant. As UPS plans to move to DB2 V5 in the near future, they have not done any further work to automate or speed up this process.

2.4.9.8 Any-to-Any Connectivity

At UPS, the general philosophy regarding storage management is to ensure that every I/O device, including all DASD, tape drives, and channels, is available to every processor or MVS image. The cost of providing this connectivity can be an issue. However, this is outweighed by the flexibility to move a changing workload to a more appropriate processor and to provide fallback in case of planned or unplanned outages. This provides UPS with the ability to bring any OS/390 image up on any processor in the sysplex. The full connectivity is complemented by UPS' use of HCD and ESCON Manager to provide centralized administration and management of the I/O configuration.

UPS feels that the flexibility provided by any-to-any connectivity is even more important in a data sharing environment than it was prior to data sharing.

2.4.9.9 Automation Changes

Around the same time as the start of the data sharing project, UPS was converting their automated operations product from Legent's Automate/MVS to CA OPS/MVS II. Given the impending move to data sharing mode, they used the opportunity to design the automation to handle a Parallel Sysplex environment, where tasks are less tied to a particular image than was the case in the past. The operations group designed the automation around a master list of all (17) systems and all tasks. The master list contains information about which system a task is running on. They also developed some in-house move routines to make it easier to move a task from one image to another. Now, if they wish to move a task from system A to system B, they issue the move command which stops the task on system A, updates the database to say that this task should now be on system B, and starts the task on system B.

All of the automation routines are generic and common to all systems, so that a task can be moved to any system and automation will handle it in the same manner.

At the moment, UPS does not use Automatic Restart Management (ARM) in production. However, it is in test on the development systems. The next release of automation will provide specific support for ARM, so that if ARM restarts a task on the same or a different system, the automation will trap this event and update its database accordingly.

Beyond these design changes to make automation easier to maintain and to give it more flexibility in a Parallel Sysplex environment, minimal changes had to be made to automation. New data sharing-related messages were initially trapped and just forwarded to an expert who then decided what, if any, automation was required. Very little new code had to be added to handle new messages.

Managing those functions that control recycling all the members is handled by maintaining separate batch processing streams and different names for each member of the data sharing group. The batch processing is dependent upon all of the different tasks where strict naming conventions are employed, and individual names for each started task makes automation as simple as managing a single task.

2.4.9.10 Training and Education

In a technical environment, the right training and education is important, and Parallel Sysplex and data sharing training and education is no exception. The cost of training and education is small when compared to the cost of the hardware and software, yet the availability of the system depends on a proper understanding by the people controlling the system. There are many classes in data sharing and Parallel Sysplex offered for a wide variety of audiences. It is important to get a basic understanding of this new environment; however, this is just the foundation. In spite of excellent training, the real test is *using* these concepts: what is involved in activating and deactivating a policy, moving structures from one Coupling Facility to another, and so on.

Although the UPS IT staff thought they fully understood DB2 data sharing, they found they were challenged during recovery testing during sessions with the IBM S/390 EdPlex Center in Poughkeepsie and the stress testing at the Washington

Systems Center. Here, the UPS staff could "play" with and test recovery procedures in an environment that allowed them to fully test their understanding of these concepts.

An integral and important part of the training experience was having the UPS MVS systems programmers work together with the UPS operations staff during the tests to help them better understand the differences in their environment. Having to deal with operational procedures in a hands-on environment that actually mimicked the production environment was invaluable to the success of the project.

2.4.9.11 Project Management

Project management and detailed project planning is key to the success of any complex project. The project plan drawn up by UPS extended to over 30 pages and included input from all affected areas. With so many areas involved, it was imperative to have a strong project manager who understood of all the key components. In this and other projects that IBM has been involved in, strong project management is a recurring theme in successful projects.

To manage the project plan, UPS used Microsoft Project to assign tasks, establish successors and predecessors to all tasks, make assignments of particular tasks, and set timelines for scheduling completion of tasks. Some type of project management tool is key to monitoring, managing, and reporting on the timeline for all tasks necessary to complete within the guidelines established.

2.4.9.12 Networking

If you are contemplating a continuous availability project, it is important to make contact with other users to share their experiences. UPS established a network of resources using the Internet, and contacts at conferences such as Guide, Share, the DB2 Technical Conference, and the International DB2 User's Group, IDUG. Most countries also have local DB2 user groups. If you still have questions, your IBM representative can arrange conference calls with other customers who have experience with migrating to Parallel Sysplex and data sharing.

2.4.9.13 Availability of IBM Resources

UPS identified the partnership that developed between IBM and UPS as one of the unexpected benefits of the data sharing project. IBM identified critical resources to assist with the project, and several experts worked hand-in-hand with UPS at critical steps of the data sharing effort. Many IBMers were involved, including experts from the Washington and Dallas Systems Centers, the DB2 development lab and S/390 development, to assist with questions throughout the project.

2.4.10 Opportunities for Improvement

Although UPS was very pleased with the support they received from IBM, they felt that there was room for improvement in the documentation relating to Parallel Sysplex implementation.

2.4.10.1 Documentation

UPS felt that while all the information they needed was available somewhere, there was no *single* document they could refer to that would lead them through the whole data sharing project. UPS felt that a checklist should be developed for data sharing projects so that potential data sharing implementers can be aware of all the issues that such a project may entail. Also, a list of sources of information should be created pointing to such resources as redbooks, Web sites, product publications, and so on.

The three books that were used the most by the DB2 staff were *DB2 Data Sharing Planning and Administration*, SC26-3269, and the redbooks *DB2 for MVS/ESA Version 4 Data Sharing Implementation*, SG24-4791, and *DB2 for MVS/ESA Version 4 Data Sharing Performance Topics*, SG24-4611.

Since the project was completed, a Web site has been created that provides a task list for implementing Parallel Sysplex, complete with Hiperlinks to the appropriate sections of the relevant manuals. At this time, the task list mainly addresses Parallel Sysplex at the MVS level. It does not go into great detail about how to move any of the subsystems (CICS, IMS, DB2) to Parallel Sysplex. However, the list is continually being extended to make it more useful and user-friendly. The Web site address is:
`http://www.s390.ibm.com/pso`

There is also a site that provides hints and tips for DB2 data sharing implementation, based on experience gathered from customer projects. The DB2 data sharing hints and tips Web site is accessible at:
`http://www.software.ibm.com/data/db2/os390/dshints`

2.4.11 Results That Were Achieved

The rollout of DIALS in a data sharing environment was transparent to the user. Since the migration, there have been no unplanned DB2 outages. From time to time, performance issues have arisen, but none have turned out to be related to data sharing. The DB2 data sharing implementation and the availability of the system ever since the cutover have exceeded the expectations of the project team.

2.5 Future Plans

At the time this book was written, the Christmas holiday season was approaching. This is traditionally the busiest period for DIALS, so further changes or developments would have to wait until after this time. Following their experience with DIALS data sharing over this critical period, further exploitation of DB2 data sharing is being considered for the remaining PLDR applications. Following the announcement of the IBM G5 and G6 processors and the capacity they provide, further implementation of data sharing will be driven by availability rather than capacity concerns.

In addition, the following enhancements will be considered for DIALS and the other PLDR applications in the future.

2.5.1 VTAM Generic Resources

At the time of writing, the VTAM Generic Resource support is being tested on the UPS development system. Some changes are required to get APPN routing working correctly, after which VTAM GR will be available.

UPS is currently contacting the session manager vendor to establish if that product will support VTAM GR, and to identify general Parallel Sysplex exploitation plans for that product. For example, will it store information in the Coupling Facility that would allow a user to reconnect to a different session manager instance in case of an unplanned outage?

2.5.2 CICSPLEX/System Manager (CP/SM)

UPS investigated CP/SM when it was first introduced by IBM. At the time, it was felt that the product was not appropriate to the UPS CICS topology. Since then, however, a number of enhancements have been made to the product, and UPS will review it again at some time in the future.

CP/SM, together with CICS TS support for Temporary Storage queues in the Coupling Facility, may help UPS address the affinity issues in the DIALS CICS regions.

2.5.3 IMS Version 6 Data Sharing

UPS planned to migrate to IMS Version 6 early in 1999. This version of IMS was to be rolled out to all UPS IMS subsystems by the end of the first quarter.

Once the rollout of IMS V6 is complete, UPS will then start investigating the steps necessary to move the DIALS IMS system to data sharing mode. Given the limited time between the implementation date and the pre-year 2000 change freeze, the actual move to data sharing mode for the DIALS IMS system is likely to be held off until early 2000.

2.5.4 Higher Performance Coupling Facility Links

Currently 50 MB links are installed on the processors running DIALS, as opposed to the 100 MB links used in the testing at the Washington Systems Center, (WSC). The difference in service time between Hiperlinks with 100 MB links and ISC links with 50 MB links depends on the amount of data being sent to and from the Coupling Facility. For a lock operation, the difference might be 15%, whereas for a cache operation where the amount of data transferred is larger, the difference would be greater, depending on the size of the transfer.

In the DIALS application, there is very little reuse of data from the GBPs by the CICS query regions, so most of the Coupling Facility requests generated as a result of CICS queries are lock and registration requests. On the other hand, every page created by the propagators will be written to the GBPs and later cast out to DASD. Based on this profile, it was estimated that the use of 50 MB ISC links would increase overhead by 1.5% to 2% and make no discernible difference in response time. At the time of the migration to data sharing, it was felt that the performance with the installed 50 MB links would be acceptable.

At the time of writing, UPS is implementing 100 MB Hiperlinks as new processors are added to the configuration, and all 50 MB links will be phased out over time. In addition, UPS is monitoring the developments in Coupling Facility link technology, such as Internal Coupling Bus (ICB) and Internal Coupling (IC)

channels, and will continue to implement the level of technology that is most appropriate to their environment.

2.5.5 3-Way Data Sharing Group

Depending on capacity requirements for new initiatives, and steady growth in the current production environment, UPS may consider 3-way data sharing for DIALS. A third member is being installed in the data sharing group and will be placed in standby mode. This member will be ready to take over some of the DIALS workload should the holiday peak load require additional capacity.

2.5.6 DB2 Compression

UPS has implemented DB2 hardware compression on a number of the other PLDR databases and is very pleased with the results. Space savings are in the range of 35% to 55%. CPU time for the DB2 address space is up slightly, but overall utilization is down marginally. I/O times have improved (presumably due to a reduced load on the I/O subsystem), and batch elapsed times have reduced.

UPS planned to extend the use of DB2 compression to the DIALS databases in the first quarter of 1999.

2.5.7 DB2 Online Image Copy

UPS currently takes weekly Sharelevel Reference image copies. These require stopping all update activity against the databases while the backups are in progress. There are also Sharelevel Change image copies taken on a nightly basis. These image copies do not require update activity to be quiesced, however, they do require the logs to be available to recover the databases.

Due to the mix of DASD technologies installed in UPS (IBM and non-IBM), UPS is unable to take advantage of the Concurrent Copy or Virtual Concurrent Copy features of the image copy utility.

2.5.8 DB2 Online Database Reorg

UPS currently does twice-weekly reorganizations of the index tables for the DIALS databases. These reorgs require all activity against the tables in question to be stopped for the duration of the process.

UPS has the BMC Sharelevel Reference online reorg product installed, and uses it for reorging some databases, but not the DIALS databases. Due to lack of DASD space for the temporary copy, they have not investigated full online reorg yet. After DB2 compression is implemented for the DIALS databases, freeing up some DASD space, they will start to investigate the use of online reorg to allow them to reorg their databases while only causing a brief outage of the databases.

2.5.9 DB2 Version 5

There are many enhancements in Version 5 that affect DB2 data sharing and improve availability, performance and usability. The migration to Version 5 was scheduled for the first quarter of 1999. Some of the key features of interest to UPS are as follows:

Dynamic Inter-DB2 R/W Interest Tracking Enhancement

The usefulness of CLOSE YES to remove inter-DB2 R/W interest is limited in Version 4 because a pageset remains GBP-dependent even

after all the R/O DB2 members have physically closed it. Refer to the redbook *DB2 for MVS/ESA Version 4 Data Sharing Implementation*, SG24-4791, for more information.

The Version 5 inter-DB2 R/W interest tracking enhancement allows the updating DB2 to convert the pageset to become non-GBP-dependent soon after the inter-DB2 R/W interest has gone away. An internal process is initiated soon after pagesets are physically closed to try to convert pagesets to non-GBP-dependent without having to wait for the last updater to perform a pseudo close.

CLOSE YES does help to remove the inter-DB2 R/W interest sooner, but Version 4 does not take full advantage of this. With this Version 5 enhancement, DB2 takes advantage of the disappearance of inter-DB2 R/W interest to stop using GBP protocols for the pageset sooner. If you are planning to migrate to Version 5, evaluate using CLOSE YES before the migration.

Group Buffer Pool Recovery

There are enhancements in Version 5 that will speed up the GBP recovery process. Automatic GBP recovery will allow for automatic recovery during Coupling Facility hardware failures.

Manual GBP rebuild support will allow for non-disruptive Coupling Facility maintenance procedures. Before this support became available, a scheduled outage was required to the data sharing environment in order to do maintenance upgrades to the Coupling Facility containing the group buffer pools.

Duplex Group Buffer Pools

GBP duplexing was originally announced for Version 6 but retrofit to Version 5 (APAR PQ17797). This enhancement will help by allowing the GBPs to be duplexed across two Coupling Facilities, thus minimizing any service disruption caused by a structure or Coupling Facility hardware failure or 100% loss of connectivity to a Coupling Facility. It also provides significant benefits even if only one system loses connectivity to the Coupling Facility containing the GBPs. Without Duplex GBPs, all GBP structures in the affected Coupling Facility must be rebuilt in the alternate Coupling Facility. Duplex GBPs remove this requirement to rebuild the structures, and thus provide performance benefits when there is a less-than-100% loss of connectivity.

UPS is particularly interested in this enhancement, due to the great impact this has on availability in case of a Coupling Facility or CF link failure. As each Coupling Facility must have enough storage to hold *all* the DB2 structures (to cater for an outage of one of the Coupling Facilities), the introduction of Duplex Group Buffer Pools will not require any additional Coupling Facility storage compared to the current situation.

DB2 PM Enhancements

Many requests pertaining to the Group Buffer Pool are recorded in “OTHER REQUESTS” in the DB2 PM statistics report. DB2 Version 5 breaks down this information by categorizing it into new counters such as UNLOCK CASTOUT, READ CASTOUT CLASS, READ CASTOUT STATISTICS, READ DIRECTORY INFO, READ STORAGE

STATISTICS, REGISTER PAGE, UNREGISTER PAGE, DELETE NAME, PARTICIPATION IN GBP REBUILD, and GBP CHECKPOINTS TRIGGERED. This more detailed information will ease the task of monitoring and tuning DB2's use of the Coupling Facility.

LARGE Partitioned Table Spaces

With the introduction of large partitioned table spaces in DB2 Version 5, a table can hold up to a terabyte of data in compressed or uncompressed format.

In prior releases, table space storage was restricted to 64 partitions of 1 GB each. With large partitioned table spaces, this limit increases to 254 partitions of 4 GB each. The size of each index partition of the partitioned index also increases to 4 GB.

The size of a data set for a nonpartitioned index on a LARGE partitioned table expands to 4 GB. With a limit of 128 data sets, this increases the maximum size of a nonpartitioned index from 64 GB to 512 GB.

Selective Partition Locking

Prior to DB2 Version 5, it was not possible to lock individual partitions of a partitioned table space. Even when SQL activity occurs on only a single partition, the entire table space is locked. For data sharing, the inability to lock individual partitions means that all child locks for a table space are propagated to the Coupling Facility, even when you do the following:

- Route work to specific members of the data sharing group to create an affinity between a specific table space partition and a data sharing member.
- Run batch jobs on separate partitions to avoid locking conflicts.

This is because all child locks on all partitions have the same lock parent in the locking hierarchy, the *partitioned table space lock*.

For some types of work loads, even those in a non-data-sharing environment, you can improve performance by specifying a single gross lock on a partition rather than using individual page or row locks. It was not possible to do this before DB2 Version 5.

Now, by defining a table space with the LOCKPART YES option, you tell DB2 that you want individual partitions locked only as they are accessed. For those special cases in which you are purposefully creating an affinity between data partitions and DB2 members, each locked partition is a lock parent in the explicit locking hierarchy. DB2 and IRLM can detect when no inter-DB2 read/write interest exists on that partition, and therefore they do not propagate child locks unnecessarily.

Although selective partition locking can benefit certain data sharing applications, there can also be benefits for non-data sharing applications that use partitioned table spaces. For these applications, it might be desirable to acquire gross locks (S, U, or X) on partitions to avoid numerous lower-level locks but still maintain concurrency. When locks escalate, and the table space is defined with LOCKPART YES, applications that access different partitions of the same table space are not affected by update activity. The LOCK TABLE

statement is extended to allow applications to explicitly lock certain partitions for those table spaces defined with LOCKPART YES.

TCP/IP

Transmission Control Protocol/Internet Protocol (TCP/IP) is a standard communication protocol for network communication. Previous versions of DB2 supported TCP/IP requesters, although additional software and configuration was required. Native TCP/IP support eliminates these requirements, allowing gateway-less connectivity to DB2 for systems running MVS OpenEdition. Now DB2 can be accessed from a SNA network, a TCP/IP network, or a mixed network.

UPS is considering developing applications using TCP/IP, so this support is important to them.

Preformatting Table Spaces and Partitions

You can now preformat table spaces and partitions when you use the LOAD or REORG utility.

Prior to DB2 Version 5, during insert processing DB2 must preformat pages before they can be used. Preformatting occurs as needed, and can delay insert processing, or make the insert processing time unpredictable.

For applications that have heavy insert activity into new table spaces, you can have DB2 preformat pages when you load or reorganize the table space or partition, instead of during insert processing in Version 5. After data has been loaded or reorganized, DB2 preformats the remaining pages up to the high allocated RBA in the table space or partition, and in the associated index spaces.

If you can preallocate the entire table space before using it, use the PREFORMAT option of LOAD or REORG when preformatting is causing measurable delays in your insert processing, or if you must have predictable elapsed times for insert processing.

Shutdown of the DB2 Data Sharing Group

The elongated length of time during shutdown of the data sharing group is partly due to DeleteName processing for the pageset or partitions going out of GBP-dependency during physical close. A solution is planned to be delivered with a future CFLEVEL and DB2 and OS/390 APARs.

GBP Checkpoint Enhancements

When DB2 takes a GBP checkpoint, it scans the GBP cache structures to find the Log Record Sequence Number (LRSN) of the oldest changed page. This LRSN is used in case GBP recovery is required. This scan can cause spikes in Coupling Facility utilization, especially noticeable for large GBPs, adversely affecting Coupling Facility response times.

CFLEVEL=5 introduced User Defined Field (UDF) order queue support, which maintains a queue in order of time of update. Thus the oldest changed page can now be found easily at the top of the queue.

This support requires CFLEVEL=5 or higher, OS/390 2.6 or OS/390 Releases 3 through 5 with APAR OW28460, and DB2 APAR PQ17650 (since superseded by APAR PQ19714).

2.5.10 DB2 Version 6

There are many enhancements in DB2 Version 6, in addition to the ones that have been rolled back to DB2 Version 5. UPS have not had an opportunity yet to evaluate the impact of Version 6 on them. However, one enhancement that may benefit them, due to the manner with which they use their Group Buffer Pools, is the ability to specify that updated pages are not written to the CF. GBPCACHE NONE can be used to specify that only cross-invalidation information is kept in the CF, while updated pages are written directly to DASD. Given the low reuse of pages from the GBPs in UPS, this new ability might provide some performance benefits for them.

2.5.11 OS/390 Structure Rebuild Enhancements

After the Coupling Facility failure was tested, a SETXCF START,REBUILD,CF=cfname,LOC=NORMAL command was issued to move structures back to their original Coupling Facility after the failed Coupling Facility came back online. This command caused an unnecessary rebuild of the Lock and SCA structures which delayed the GBP recovery. The new POPULATECF function of the SETXCF REBUILD command in OS/390 V2.6 will eliminate this unnecessary rebuild.

The new function will give the user a way to repopulate a Coupling Facility after it comes back online without having to do unnecessary rebuilds. As a result, the operational complexity is lessened by issuing just one command while at the same time avoiding all the unnecessary rebuilds in the current release.

2.5.12 DFSMS Enhanced Catalog Alias Support

A new feature in DFSMS 1.5 (available with OS/390 V2.7) will help customers in data sharing mode apply maintenance while still having some members of the data sharing group up and running.

Some customers, including UPS, currently have a single set of DB2 libraries shared across a number of DB2 subsystems. This is often done to avoid the nontrivial task of mass JCL changes. Starting with DFSMS 1.5, DFSMS/MVS will provide a mechanism to allow an alias in a shared master catalog to be interpreted differently on each system in the sysplex based on the MVS symbolics defined on each system.

Using this new facility, you can define an alias in the master catalog, with a new subparameter, for example,

```
DEFINE ALIAS(NAME(DB2.DSNLOAD)
             SYMBOLICRELATE('DB2.&SYSNAME..DSNLOAD'))
```

On system SYSA, if you have the symbolic SYSNAME set to SYSA, a STEPLIB pointing to DB2.DSNLOAD would end up using DB2.SYSA.DSNLOAD. If SYSNAME is set to SYSB on system SYSB, the same job would end up using dataset DB2.SYSB.DSNLOAD when run on SYSB.

This enhancement may help those customers applying maintenance in a Parallel Sysplex environment for all products including DB2. Having two sets of DB2 Target libraries will allow the installation to shut down one DB2 subsystem and apply maintenance, while another member of the same data sharing group continues to run on another image.

2.6 Summary

As can be seen, UPS has invested a great deal of time and effort in achieving a high performance, high availability data sharing solution. They moved an extremely large, business-critical application into data sharing mode, improving its availability without having any negative impact on its user population. They are now positioned with the technical and management skills, as well as the hardware and software infrastructure, to provide these benefits to additional applications with minimal incremental cost or effort.

The DIALS application is currently not a continuously available application in the strictest sense of the phrase. However, the important thing is that it is currently providing the availability required by its users. And, as can be seen in the previous sections, there are additional facilities in DB2 and the other products that will provide even higher availability without much more effort or cost. As factors such as business growth and Internet access drive the requirement for fewer and fewer outages, UPS now has the skills and the infrastructure in place to respond to that requirement in DIALS and other applications.

The benefits that UPS has obtained--improved availability and more operational flexibility--are also available to other customers that implement Parallel Sysplex data sharing and workload balancing. An important lesson from this project is that many benefits can be achieved from data sharing, *even if only a part of the application exploits Parallel Sysplex*. Parallel Sysplex has helped UPS provide the level of availability required by a very demanding internal and external user community. And Parallel Sysplex data sharing gives UPS the capability to extend DIALS capability to users throughout the world. As the Internet and globalization of business continue to grow, this level of availability will become a prerequisite for large businesses that wish to remain competitive in the global marketplace.

Late breaking news

Just as this book was going to print, UPS completed the migration of another major application to DB2 data sharing. Based on the knowledge and experience gained during the DIALS migration, UPS completed the migration of this 1000+ MIPS application in less than four months, with a flawless cutover.

Chapter 3. VSAM Record Level Sharing Case Study

Traditionally, companies in the financial services industry have had more stringent reliability and availability requirements than most other large computer users. For our second case study, we wanted to work with a customer in this industry. If possible, we also wanted to work with a customer that was doing data sharing using a data manager other than DB2.

The customer that worked with us on this study was Societa' per i Servizi Bancari (SSB) in Italy. SSB provides services to a number of Italian banks and therefore has even higher than normal availability requirements. Additionally, SSB is doing VSAM Record Level Sharing (RLS). As this is the most recent exploiter of Parallel Sysplex data sharing among the IBM data managers, it does not have as many users as IMS or DB2 data sharing yet, so we were very fortunate to get a customer that was both in the financial services industry *and* using VSAM RLS.

We would like to take this opportunity to thank the staff at SSB for their assistance and enthusiasm during this project.

3.1 Company Description

The rapid evolution of information technology is radically changing our daily habits. Even money is being transformed, becoming virtual data, pure electronic information. This transformation involves not only cash but also all the services connected to it. These dramatic changes are driving a need for more and more advanced systems. Systems supporting modern-day banking institutions must be fast, secure, integrated, and able to satisfy all specific needs, from those of multinational companies to small businesses, through to the individual. This is the business environment in which SSB operates.

SSB has a specific mission--to provide services for a constantly evolving financial environment using the most technologically advanced facilities. Its activities are laid out along two lines:

- Studying and developing new information services for the banking sector
- Creating and applying innovative solutions to facilitate interfaces among banks and between banks and their clients

In this sense, SSB can be seen as a technological link between the existing systems, and those that will be required to meet the demands of tomorrow.

SSB's stockholder portfolio consists of over 200 Italian banks. Its customers include all the major credit institutions, who collectively manage over 70% of the activities of the entire banking system in Italy. SSB has up to 650 customers for some of its applications.

SSB works in close collaboration with the Bank of Italy (the Italian Central Bank), Convenzione Interbancaria per i Problemi dell'Automazione--Interbank Convention on Automation Issues (CIPA) and Associazione Bancaria Italiana--Italian Banking Association (API).

SSB manages the authorization processing for Bancomat and PagoBancomat operations, and Visa, Europay and Mastercard transactions from Italy and from

abroad. SSB also manages the authorization and the switching of the JCB, American Express and Diners cards, with roughly 55 million transactions per year.

In accordance with the European BEST Agreement, SSB works together with companies providing similar services in other countries, in order to harmonize and integrate European services for banking customers. SSB holds a share of Europay, the company that promotes Eurocheque, Eurocards, Cirrus, and Maestro cards internationally. Furthermore, SSB is a Principal Member of Visa. It can thus develop its services in the best way, and, at the same time, take part directly on strategic committees of Europay and Visa in order to be involved from the start in the creation and design of all new international payment services.

SSB's 1997 net revenue was 103 billion lira (roughly 65 million US Dollars). In the same year, SSB managed 2.5 billion interbanking operations. SSB's data center is located in Milano, Italy, with about 1600 mainframe MIPS, and 2.2 TB of DASD. SSB currently employes about 200 people.

More information on SSB can obtained from the SSB Web site:
`http:\\www.ssb.net`

3.2 Background to Parallel Sysplex Project

One of the services provided by SSB is a Point Of Sale (POS) application. The POS application supports three types of electronic transactions: ATM transactions using debit cards via the Bancomat application, debit card transactions in shops, post offices, and railway terminals via the PagoBancomat application, and credit card transactions. The authorization can be granted by SSB directly, by the issuing bank to which SSB routes the authorization requests, or by the relevant credit card company.

The POS application is one of SSB's most critical applications due to the high number of transactions, the stringent availability requirements, and its high growth rate.

Table 13. Transaction Volumes

Transaction type	**01/98-09/98**	**01/97-12/97**	**01/96-12/96**
Bancomat ATM	65,000,000	75,000,000	65,000,000
POS Payments (PagoBancomat)	109,000,000	97,000,000	50,000,000
Eurocheque	22,000,000	18,000,000	13,000,000
Fastpay (PagoBancomat)	19,000,000	21,000,000	10,000,000
POS Siteba (PagoBancomat)	51,000,000	50,000,000	27,000,000
Bank Transfer	70,000,000	87,000,000	77,000,000
RIBA & RID	91,000,000	109,000,000	97,000,000
Cheque Transactions	133,000,000	169,000,000	171,000,000

Starting in 1994, SSB knew they were going to have to make major changes to their IT strategy, mainly due to the rapid growth of their workload. SSB needed a system which would be highly available and able to handle peak volumes that could be as high as one million authorizations per day, or 50 to 70 (or more) authorization transactions per second.

In addition to the projected growth in the number of transactions, SSB was also looking for ways to provider even higher availability for the POS application. They needed fault tolerance and a way to maintain the application availability across planned or unplanned outages.

Table 13 on page 52 shows the yearly transaction volumes for SSB. Note that the figures only cover nine months for 1998. The table includes the POS transactions as well as the more traditional transactions, and displays the trend for rapid growth of the POS transactions while the traditional transactions are steady or declining.

Figure 3 contains information about the workload profile for all the transactions in Table 13 on page 52 for a typical month.

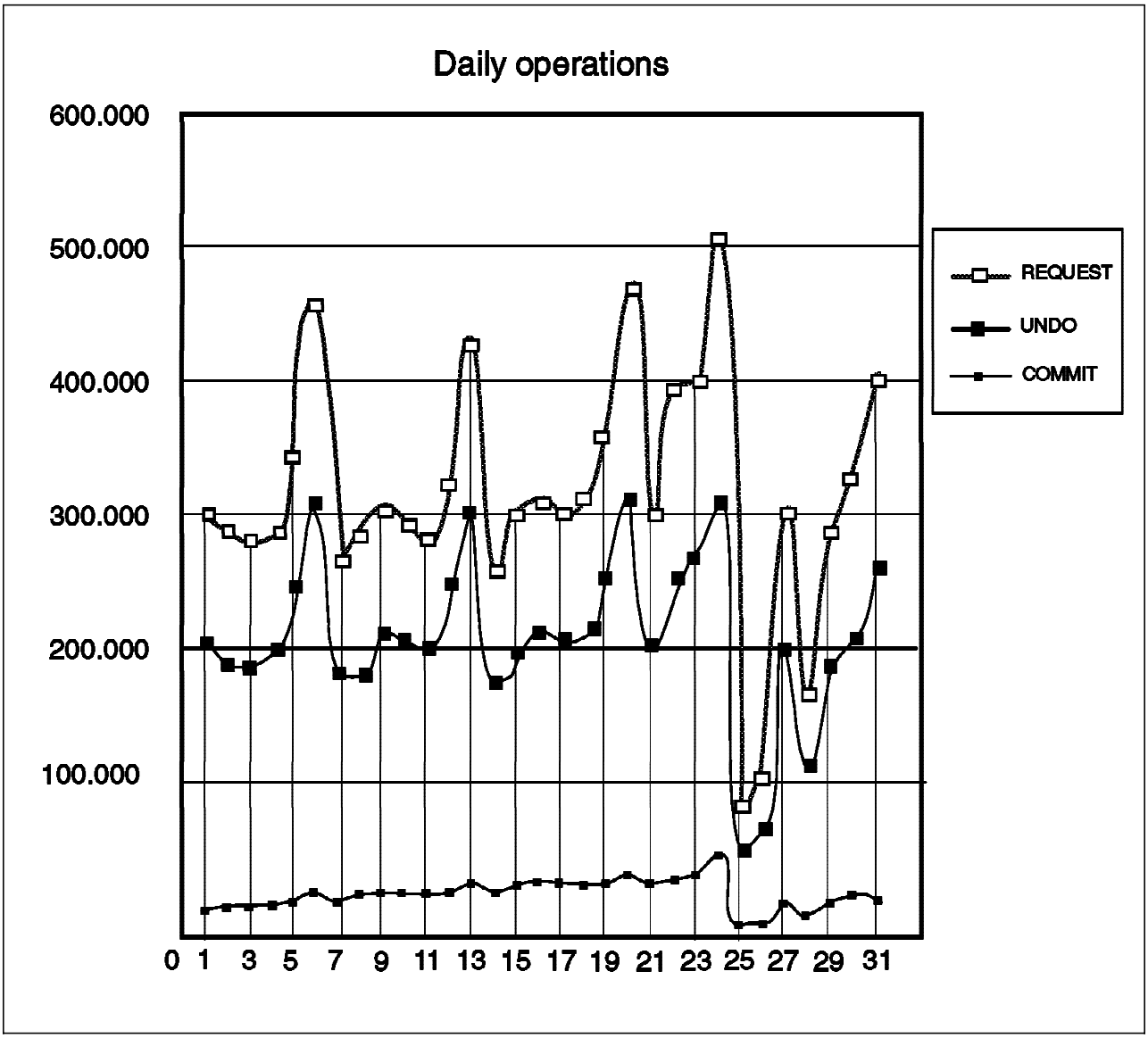


Figure 3. SSB Typical Month Transaction Volumes

3.3 Application Description

At the start of the project, the POS application consisted of a single CICS V3.3 region with local KSDS VSAM work files and a DL/I cards database (under DBCTL control). A CICS V3.3 file owning region was used for access to the shared VSAM KSDS files. The journal and forward recovery management was provided by the JMP and RPCV products from BMC Software.

POS was run on an IBM 9021-952 under MVS/ESA V5.2.2. POS is written in OS/VS COBOL and consists of about 1000 programs. There were about 50 KSDS VSAM work files, and the daily batch window consisted of more than 100 jobs.

There were four POS environments (four CICS regions) to support four different networks. This separation was driven by business requirements rather than technical issues. At the time, there were about 50,000 POS terminals and the peak load in each region was about 13 authorization transactions per second.

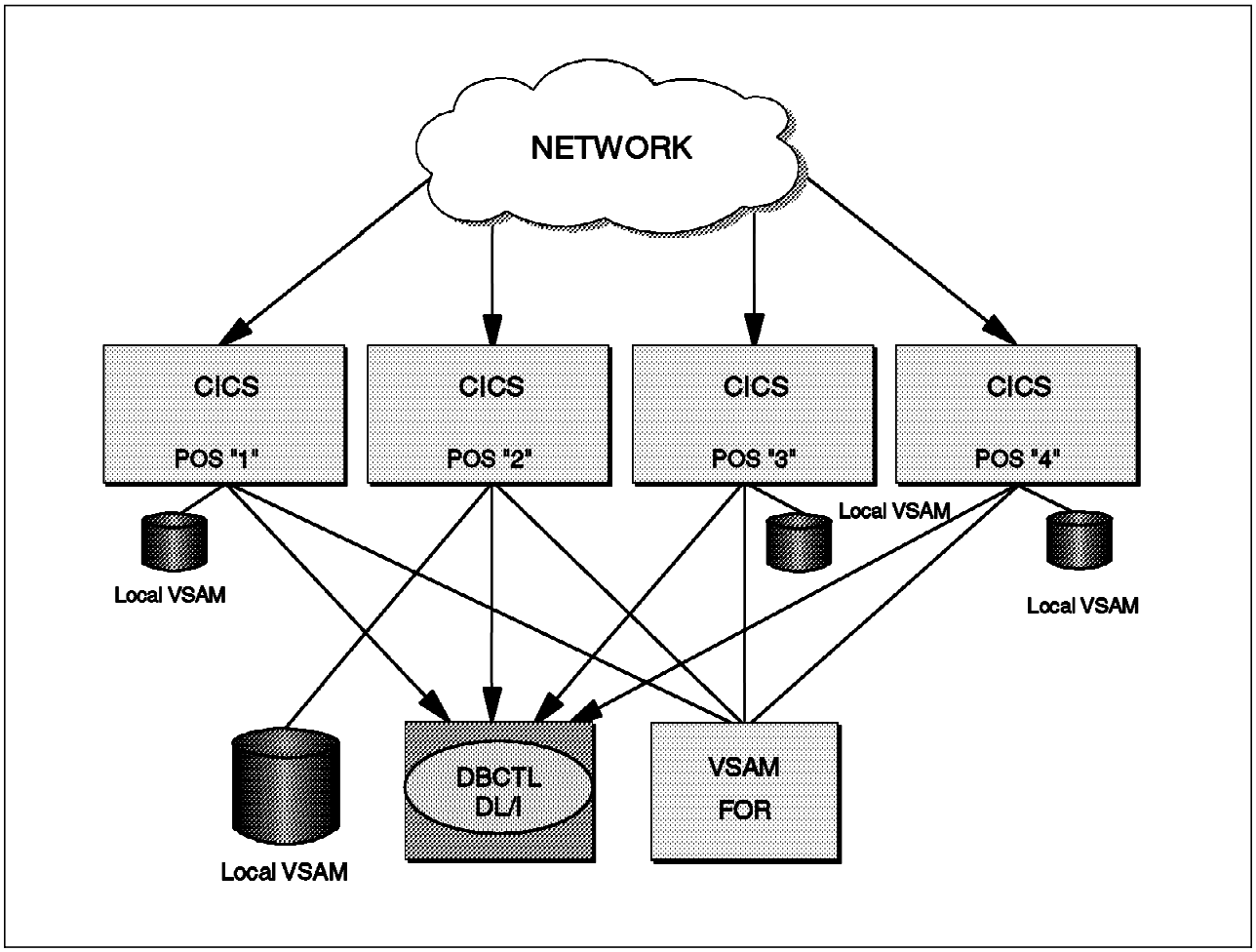


Figure 4. SSB Original Production POS Configuration

Figure 4 displays the configuration used by the POS application at the start of this project.

3.4 Data Sharing Project

Faced with a growing workload, at least one application that required near-continuous availability, and the requirement to provide a scalable environment, SSB decided to instigate a data sharing project. This project was tasked with investigating data sharing and, if it was deemed to be appropriate, implementing Parallel Sysplex and data sharing.

3.4.1 Continuous Availability Requirements

In 1994, the POS CICS region was available 21 to 22 hours a day. The remaining time was need for reorgs of the files and databases, batch jobs, and scheduled maintenance windows.

One of the most common causes of planned outages was application changes. Even a minor application change could affect the whole service.

In addition to these daily planned outages, the POS application had a single point of failure--the FOR used for accessing the VSAM files. Problems in the CICS region or the FOR or an MVS failure would cause a service outage for the entire POS application.

If Parallel Sysplex was to be the appropriate solution for SSB, it would have to address the following issues:

- Removing the single point of failure
- Providing the ability to do maintenance without affecting POS availability
- Providing for controlled release of changes to the CICS application
- Scalability
- Fault tolerance through redundant elements
- Exploitation of hardware and software functions

3.4.2 Business Case

Based on the business drivers and the associated availability requirements, SSB decided to buy a small processor to test the implementation of Parallel Sysplex. SSB purchased an IBM 9672-E01 with two CPs and built a two-way Parallel Sysplex, with one test MVS image on the 9672 and the other on the existing 9021. The Coupling Facility was run in an LPAR on the 9021.

The POS application was analyzed in order to be able to scale it to a smaller environment and fit on the smaller 9672 engines available at the time. The objective of this study was to understand the feasibility of the solution, and the overhead due to the application changes and to using data sharing.

The environment created was as close as possible to the production environment. This was done for two reasons:

- To make the tests as realistic as possible
- To make the task of reflecting the changes to the production environment as easy as possible

3.4.3 Migration to Data Sharing

The testing and implementation of data sharing was to be done in three phases:

1. Phase 1, consisting of:
   - Performing transaction affinities analysis
   - Performing affinities resolution through application changes
   - Splitting the single CICS region into a TOR/AOR topology
   - Implementing the CP/SM product for dynamic transaction routing
   - AOR cloning using CICS FORs to share the VSAM files
2. Phase 2, consisting of:
   - Migrating the CICS TOR/AOR regions to CICS TS 1.1
   - Enabling new logging functions for the CICS regions
3. Phase 3, consisting of:
   - Migrating some of the VSAM files over to DB2 in order to allow batch update jobs to run concurrently with the online systems
   - Performing CICS/VR implementation to support forward recovery functions
   - Performing SMSVSAM implementation
   - Performing batch assessment and modification to meet the restrictions of the RLS environment
   - Migrating from the FOR to RLS on a file-by-file basis

3.4.4 Phased Implementation

SSB tried to keep the test and production environments as close as possible. Any change that was considered successful in the test environment was then applied to the production environment. This allowed the production system to benefit from each small improvement as it was proven. It also avoided the complexity involved in trying to maintain two increasingly different versions of one application, and eliminated any problems that might have been encountered if all the changes were applied to the production system at one time.

For example, the splitting of the single CICS region into TOR/single AOR configuration was immediately brought into the production environment to get benefits from a capacity point of view and also to avoid heavy change management activities for the application changes.

The following charts show the different configurations used during the three stages, with the response time measured during the stress tests.

3.4.5 Creating and Maintaining the Test Environment

The POS application logs all operations, including the input data, to an application file. SSB wrote a CICS application that can read this file and replay the transactions. SSB also restored a copy of all the production POS files onto the test system each day. Using this capability, SSB was able to rerun the previous day's POS transactions on the test system.

This allowed SSB to collect comparative performance data during the different migration steps to evaluate potential overhead and the performance impact of

each change. SSB's operational procedures were updated to keep the test databases up to date and to save each day's POS transaction data.

The original configuration consisted of a single CICS AOR/TOR region, a single DBCTL, and a single CICS FOR region, as depicted in Figure 5.

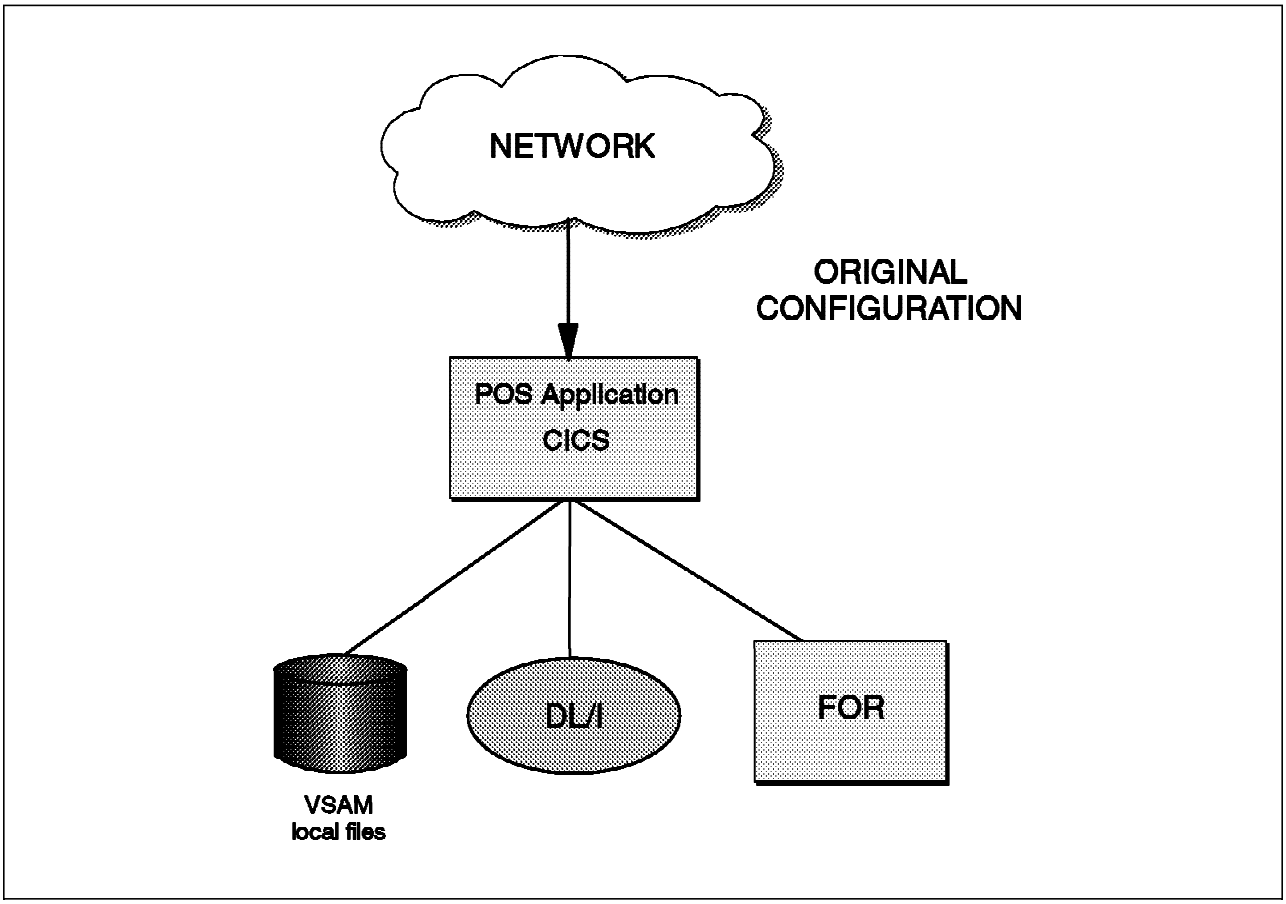


Figure 5. SSB Original POS Instance

In the first phase, a copy of the POS application was made and brought into the test Parallel Sysplex environment. No changes from the production copy were made to this version of the application. The objective of this step was to create a stable environment, testing of the application and obtaining a base measurement, against which future changes could be compared.

A verification test checked that the test application matched the logic of the production version. SSB scaled the workload from the production to this smaller test environment. 15 transactions/second had to be generated to reproduce the same relative system load on the smaller CMOS processors.

Table 14 contains the performance data collected at this stage of the project.

Table 14 (Page 1 of 2). Performance Data for Production System

Transaction type	Percentage of transactions	Response time in sec
Auth-00	70%	1.79
Auth-RM	23%	0.38

Table 14 (Page 2 of 2). Performance Data for Production System		
Transaction type	**Percentage of transactions**	**Response time in sec**
Auth-15	5%	1.57

3.4.6 Migration Phase 1

The configuration for this phase of the project consisted of one TOR with multiple AOR regions, a single DBCTL, and a single CICS FOR region, as depicted in Figure 6.

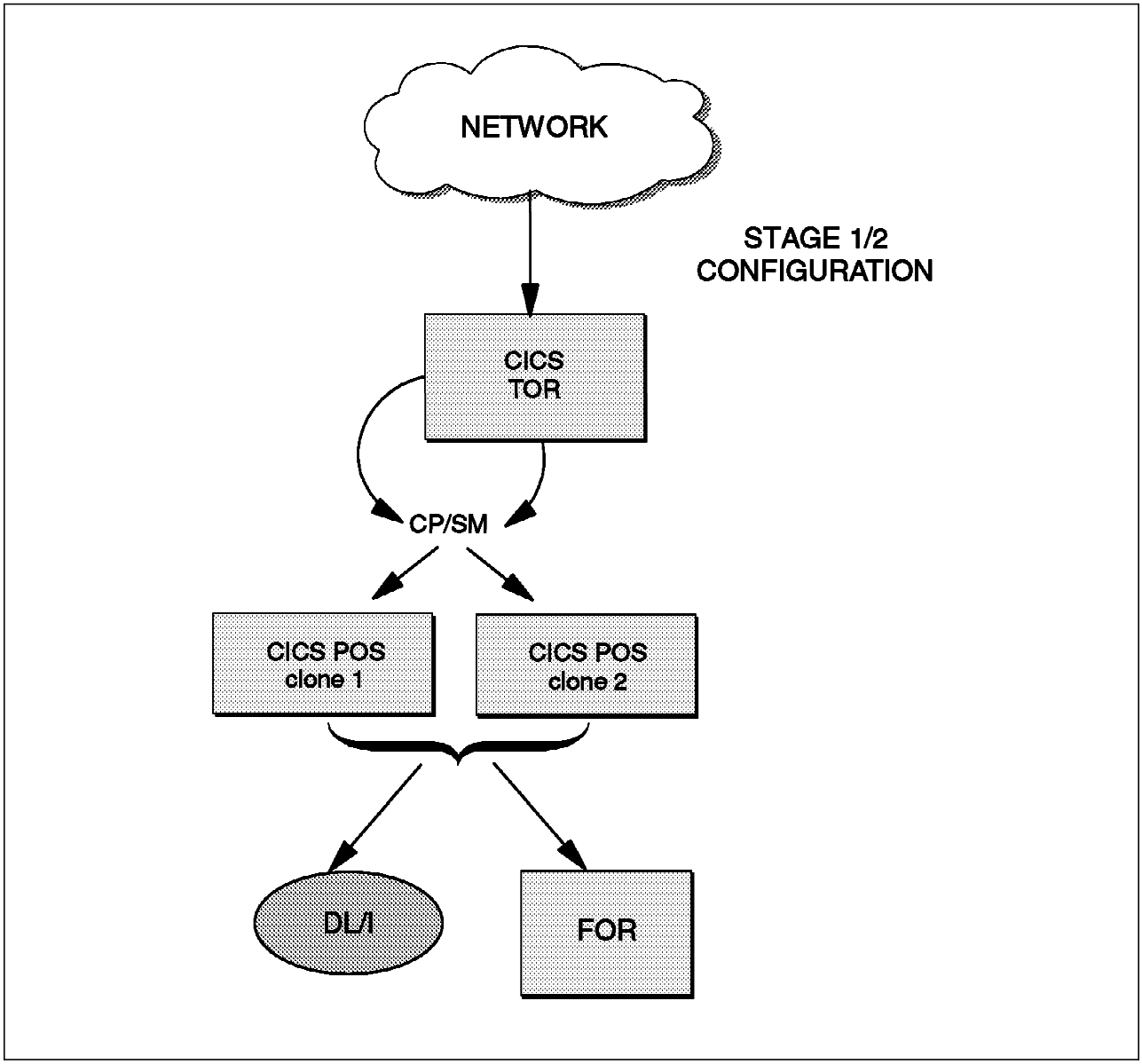


Figure 6. SSB Phase One POS Instance

In this configuration, the network was connected to a new front-end region (TOR).

VTAM Generic Resources (GR) could not be used due to the design of the application. Rather than the remote devices logging on a host application, the

host application actually initiates the logon. For the POS application, this precludes any of the benefits of VTAM GR.

However, even though VTAM GR could not be used, application availability could be maintained since there were now two routes from the TOR to the AORs. Work was routed from the front-end region to the clones using CICSPlex/SM (CP/SM).

The same workload as the original configuration was used in this phase, and the results are shown in Table 15.

Table 15. Performance Data for Phase One

Transaction type	Percentage of transactions	Response time in sec
Auth-00	60%	1.97
Auth-RM	37%	0.58
Auth-15	3%	1.82

3.4.7 Migration Phase 2

In this phase the TOR and AOR CICS regions were upgraded to use CICS Transaction Server 1.1. There was still a single DBCTL and just one CICS FOR region.

Initially, the CICS regions were activated without any logging functions. This was done in order to verify the application behavior under the new level of CICS, without the added complication of the logging changes introduced by CICS TS 1.1.

After successful testing of the application with CICS TS 1.1, CICS logging functions were activated. This step had to be approached carefully as it was found to be difficult to optimize the system logger environment to get optimal performance.

Once the new level of CICS was considered stable, the next step took place.

3.4.8 Migration Phase 3

This phase involved moving to a TOR/Multiple AOR configuration with DBCTL for IMS data sharing and VSAM/RLS for VSAM data sharing. This is depicted in Figure 7 on page 60.

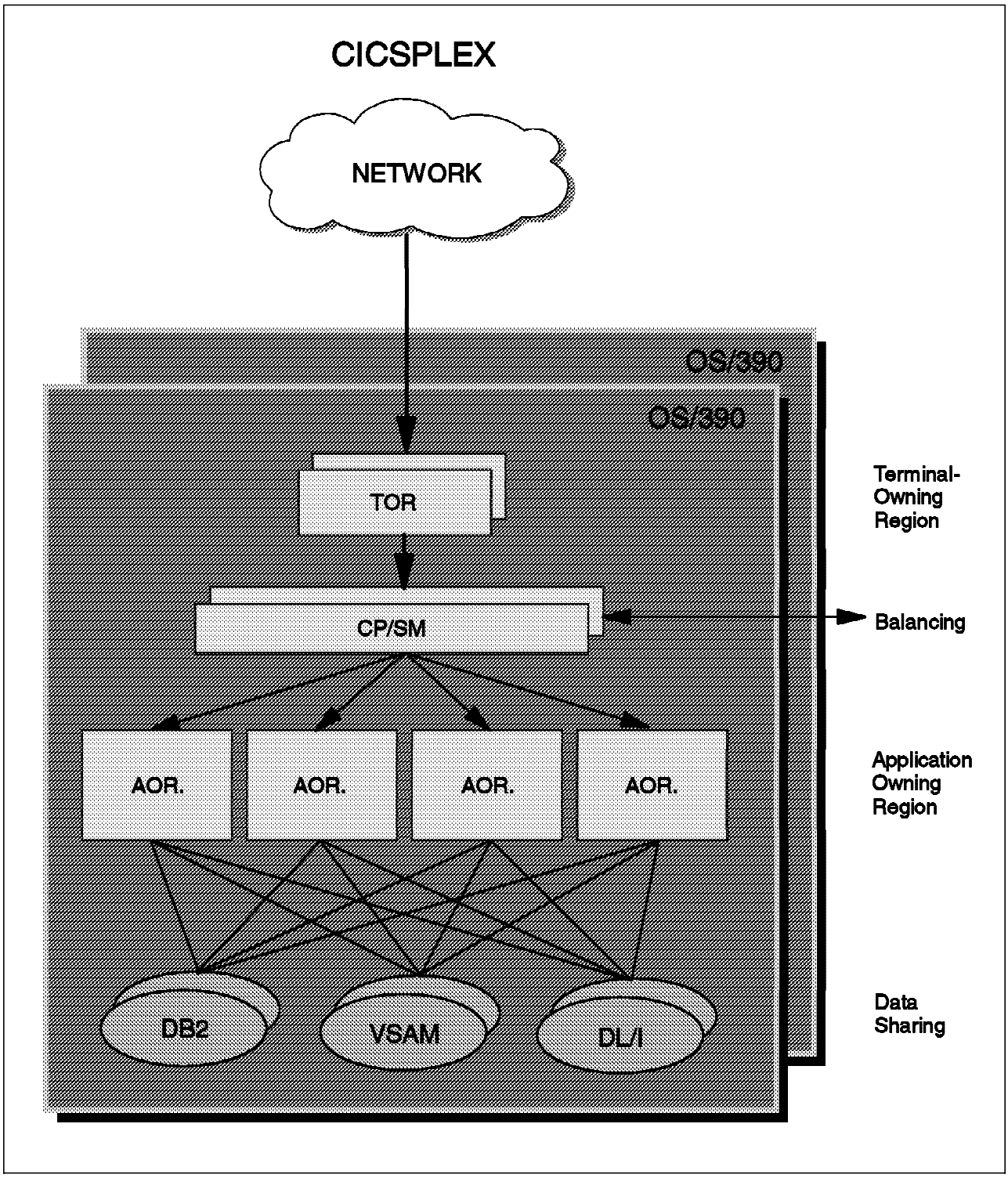


Figure 7. SSB Final POS Instance

This stage involved the replacement of the FOR with VSAM data sharing. Also in this phase, the DL/I databases were moved to sysplex data sharing mode.

The results that were obtained in this stage are contained in Table 16 on page 61.

Table 16. Performance Data for Phase Three		
Transaction type	**Percentage of transactions**	**Response time in sec**
Auth-00	61%	1.99
Auth-RM	29%	0.57
Auth-15	10%	1.95

3.4.9 Test Results

The pilot test was deemed to be a success.

The final performance data showed an increased response time due to the new application design, the use of RLS and the new products involved in the migration; however, the response time was still within the committed service levels.

Most importantly, SSB was also able to verify that the new solution provided the availability requirements that drove this project:

- Removal of critical single points of failure
- Fault tolerance
- Removal of application outages caused by maintenance/upgrade activities
- Reduced batch window
- Removal of application outages due to hardware failures

Based on these results, SSB decided to implement this architecture on the production environment for all their remaining critical applications.

3.4.10 Application Design Issues

SSB used this project to review the way their application is designed, and to identify which design objectives were not conducive to data sharing. This information could then be used when reviewing other applications for data sharing, and was also fed back into the design process for future application design.

3.4.10.1 Ease of Testing

As stated previously, the POS application writes a log of all operations. This information can then be used to drive test cases against a test version of the databases. Even though this log was originally created for audit purposes, SSB was able to use the same information to drive comprehensive tests in an automated manner.

3.4.10.2 Affinities

One of the largest inhibitors to better availability was the fact that the entire POS application resided within a single CICS region. In order to achieve the required availability, SSB needed to break the POS application into a TOR and multiple cloned AORs. This provided a structure with redundant elements in order to eliminate the single point of failure. Thus the first step was to split the POS CICS region into a TOR and cloned AORs configuration.

In the initial implementation, static routing was used to direct transactions from the TOR to the AOR. In order to implement CP/SM dynamic routing of the

transactions to cloned AORs, SSB had to find out if the transactions had affinities. If any existed, they would either have to be removed, or else they had to be honored by the routing rules in CP/SM.

The Transaction Affinity Utility product was used to analyze the application programs to be migrated. Both the batch and online analyzer/reporter functions were used to detect affinities. The result showed several affinities due to the use of temporary storage queues across multiple transactions, as well as a small number of ENQ/DEQ commands.

Luckily the application programs did not use any CWA or TWA or any other potential inhibitors. At that time (pre-GA CICS TS code) the Temporary Storage Queue Server was not yet available[2]. Similarly, there was no mechanism for handling ENQ/DEQ affinities. Most of the use of the Temporary Storage Queue was replaced by DB2 tables. ENQ/DEQ commands were removed, changing the application programs flow. There is still one TS Queue, which is loaded from a DB2 table at CICS start-up in each AOR region and used in read-only mode by the application programs.

One of the affinities that existed in the application was a function used to monitor the status and health of the application. No changes were made to remove affinities for this part of the application. These transactions were statically routed to a specific AOR.

3.4.10.3 COMMIT Changes

Another change was made to application logic. Programs were changed to keep the unit of work (UOW) as short as possible in order to avoid contention on recoverable resources. This was achieved by adding a sync point to the application logic in order to be able to commit (and consequently release) resources. Also, changes were made to make the application flow asynchronous, starting a new transaction to perform a specific function instead of the old synchronous single transaction process.

3.4.10.4 Change Management

Due to the volume of changes being applied, and the fact that there were now two “production” copies of the POS application, additional change management tools had to be used in order to track all the application changes.

As alluded to earlier, SSB was very conscious of not letting the production and test versions of POS get too far out of sync. This was to avoid having to make a large number of changes to the production application at one time, increasing the chance of introducing problems into the application. Therefore, as changes were tested and approved on the test version, they were moved to the production version of POS.

---

[2] The Temporary Storage Queue Server was provided by CICS TS 1.1 (GA code), and sysplex support for ENQ/DEQ was provided by CICS TS 1.3.

3.4.10.5 Project Duration

The task of splitting the single CICS region into a TOR/AORs structure was the heaviest and longest activity on the migration plan--the time frame for this activity was four to five months, including time spent for the implementation on the test system and then the migration over to the production version.

3.4.11 Implementation Issues

One of the new requirements in a data sharing environment is the availability of one or more Coupling Facilities. There are three types of Coupling Facility structures required in a CICS RLS environment, and SSB used the recommendations provided in the redbook *OS/390 MVS Parallel Sysplex Configuration, Volume 2: Cookbook*, SG24-2076, to help estimate the sizes and utilizations of these structures.

There were also some additional considerations in relation to the usage profile of the data sets that were going to be accessed in RLS mode. These issues are discussed in the following sections.

3.4.11.1 Lock Structure

SSB used the formula provided in the redbook *OS/390 MVS Parallel Sysplex Configuration, Volume 2: Cookbook*, SG24-2076, which recommends making the cache structure the same size as the total of all the LSR pools and hiperpools for the files that are going to be accessed in RLS mode. This proved to be successful in their environment. SSB uses the RMF post-processor Coupling Facility Report to monitor false contention. Their aim is to keep false contention below 1%. The IGWLOCK structure is 40 MB and the false contention is usually around 0.1%.

3.4.11.2 Cache

As there was only one Coupling Facility available in the test environment, SSB defined a single cache set with a single CF structure for all the files that were to be accessed in RLS mode. Later, when the production application was initially migrated to RLS mode, SSB kept the same Coupling Facility configuration: there was a single cache set associated with a single cache structure in just one Coupling Facility. As more files were moved to RLS mode, this configuration continued to be used, in the belief that isolating the cache structures would lead to better performance.

However, based on information in the SMF Type 42 records, SSB noticed that the CF requests were not balanced across the Coupling Facilities. Depending on the application workload, time of day, and so forth, SSB saw different CF requests rate on the two Coupling Facilities. Sometimes, one Coupling Facility would be overutilized while the other had only a small number of CF requests. They redefined the configuration and currently have *all* the RLS files, for multiple applications, related to a single cache set. In turn, that cache set is associated with two cache structures, one in each Coupling Facility.

SSB used the formula from the redbook *OS/390 MVS Parallel Sysplex Configuration, Volume 2: Cookbook*, SG24-2076, to determine an initial size for the cache structure. To avoid the overhead of resizing the structures every time more files were moved to RLS mode, SSB allocated structures that were larger than this initial estimate, then used SMF Type 42 records to check the effectiveness of the allocated structures.

At the time of writing, SSB has two structures, one with an INITSIZE and SIZE of 20 MB and the other one with an INITSIZE of 60 MB with a SIZE of 80 MB. The difference in size is a result of the initial configuration, where each cache set was associated with just one cache structure. Each cache structure was sized for the files that were going to be using that structure.

The first set of files to be moved to RLS mode were sized as requiring a structure of just 20 MB. The second application to be moved to RLS mode initially used a different structure, and these files were sized as requiring a 60 MB structure.

After the move to RLS, it was found that the use of multiple cache sets, with each cache set only being associated with one cache structure, was not an effective way of balancing the load across the Coupling Facilities. The files were changed to just use a single cache set, and that cache set was associated with both cache structures. SSB plans to change the cache structures in the future to have similarly sized structures in each Coupling Facility.

In addition to the cache structures in the Coupling Facility, the SMSVSAM address space also maintains a data space for all the buffers for files being accessed in RLS mode. It is very important to size this buffer so as to maintain the same hit ratio that was achieved with the LSR buffer pools. The size of the data space is determined by the RLS_MAX_BUFSIZE parameter in IGDSMSxx, and was determined by adding the size of the two defined cache structures. SSB currently uses an 80 MB data space, which the SMF Type 42 records show is adequate.

3.4.11.3 System Logger

In the release of MVS installed at the time, the System Logger did not provide the ability to control the ratio of data entries to data elements in the logger structures. SSB found that one of the most difficult tasks was in understanding the characteristics of each CICS TS logstream and trying to group ones of similar characteristics, in order to optimize the use of Coupling Facility storage and logger function, such as offloading.

SSB used the information from the DFHLSCU utility to set up the structures and logstreams and then group them by application. Some of the estimates from this utility were not sufficiently accurate, especially the AVGBUFSIZE definitions. SSB used the RMF Coupling Facility report and the IXGRPT1 sample program in SYS1.SAMPLIB (which prints the SMF Type 88 records) to analyze the logging configuration. Those reports indicated that the definitions were in some cases wasting storage in the Coupling Facility, and in other cases causing a large number of offloads. This was mainly due to the fact that log streams with differing characteristics were using the same logger structure in the Coupling Facility.

To achieve more efficient use of the logger structures, SSB now groups logstreams from cloned regions, assuming they have the same characteristics. For new logstreams, a test structure is used to understand the logstream characteristics and consequently choose the most suitable structure.

The current logger values used for logstreams are contained in Table 17 on page 65.

Table 17. SSB Logstream Values					
Logstream type	**Low Threshold**	**High Threshold**	**CF_structure**	**Logstream#**	**Offload size**
DFHSHUNT	50	80	6 up to 12 MB	20	50 cyls
DFHLOG	50	80	40 up to 50 MB	10	50-200 cyls
DFHJOURN	20	80	24 up to 32 MB	10	200-825 cyls

Non-active offload data sets are migrated after three days.

The POS transactions were heavy users of CICS logging functions. A side effect of using the System Logger services in the Coupling Facility was a reduction in the elapsed time of the POS transactions, due to the reduced response times for logging to the Coupling Facility compared to logging to DASD (CF response time is microseconds compared to a DASD response time of milliseconds).

3.4.11.4 VSAM File Considerations

During the migration of the VSAM files to RLS, SSB noticed degraded performance on a particular CICS transaction. Investigation showed that high response time was occurring on VSAM data sets allocated with a CI size of 512 bytes, and having a logical record length of 400 bytes. Since in a non-RLS data sharing environment locking occurs at the CI level, this CI size was chosen with the intention of limiting lock contention. However, in an RLS environment, with locks at the record level, it was not necessary to have such a small CI to avoid lock contention. Having such small CIs causes a split every time a record is inserted, generating additional overhead and fragmenting the file.

SSB also discovered another negative impact of the small CI size. From the SMF Type 42 records, SSB noticed that only CIs between 2 KB and 4 KB are written to the Coupling Facility cache structure. So, in the case of their 512 byte CI size data set, SSB was using a 2 KB entry in the cache structure to store a 512 byte CI, wasting Coupling Facility storage, and also having a low hit ratio in the CF. SSB increased the CI size to 4 KB, which provided better performance, higher hit ratios, and a reduction in the number of I/Os.

Without this change, it would not have been possible to migrate this application to RLS mode because of the unacceptable response time.

SSB also discovered some other considerations relating to the CI size of the files being accessed in RLS mode. CIs greater than 4 KB are not (at the time of writing) written to the cache structure in the Coupling Facility, but they are written to the SMSVSAM data space. SSB used this characteristic to improve the performance of data sets that were only accessed in read mode and had a high locality of reference. These records were loaded to the SMSVSAM data space and never invalidated, thus providing a high read hit rate without wasting Coupling Facility storage. Again, SSB used the SMF Type 42 records to check performance and make adjustments.

The resulting response time for the file operations was slightly better than a file accessed by an FOR using cross memory for function shipping.

An issue that could potentially affect the complete availability of the POS application was the lack of support at the time for VSAM Extended Addressability

files to be accessed in RLS mode. SSB had a few files that were very large (bordering on 4 GB) and could not be migrated to RLS mode if the VSAM Extended Addressability support was used. SSB circumvented the problem by reorganizing the files on a frequent basis, thus containing the growth of the file. DFSMS/MVS 1.4 provided RLS support for these Extended Addressability files, so this is no longer an issue.

3.4.11.5 Forward Recovery Considerations

Prior to the migration to VSAM RLS, SSB used tools from BMC to perform the forward recovery activities against VSAM files. After SSB migrated to System Logger, they started using CICS/VR because that product was able to natively support the System Logger environment. This change necessitated the migration of all the recovery procedures, as well as the requirement to provide education for operators and application programmers.

3.4.11.6 Batch Considerations

Application programmers were heavy users of the FileAid product from Compuware, using the product to do online browsing of the VSAM files that were planned to be moved to RLS mode. However, at the time FileAid did not support VSAM RLS. SSB had to wait for the availability of a Small Program Enhancement PTF (APAR numbers UW40778 and UQ08023) to be able to run File-Aid and batch in Non-RLS mode against files that were concurrently being accessed in RLS mode by CICS. The PTF-enabled files open in RLS mode to also be opened for read processing in non-RLS mode without any JCL changes. To avail of this facility, the files had to be defined with a cross-region SHAREOPTION of 2.

3.4.12 Operational Changes

During the migration, SSB had to make some changes to its processes in order to integrate the new environment. The following is a description of the main changes made to resolve some differences found during the migration.

3.4.12.1 CICS Region Management

SSB needed to manage the new design of the CICS environment, and needed to be able to control the transaction routing and the new environment.

One of the main changes from an operational point of view is the use of CP/SM. All the CICS regions are managed through CP/SM. CP/SM provides the single point of control for the CICS environment. Also, the automation product benefits from this by using the CP/SM capability of consolidating some CICS events at a single point. This made the management of all the CICS regions an easy task.

3.4.12.2 CICS Logging Changes

CICS TS uses the System Logger services to log its data. SSB needed to completely replace their current procedures for collecting, archiving and retrieving log data.

A lot of operational changes were required as a result of the use of the System Logger. SSB had to update or change all the operational and recovery procedures they had in place to manage log data. One of the changes is the use of CICS/VR to be able to retrieve data from the logstream for the RLS shared files. For the system journal (DFHLOG and DFHSHUNT), SSB uses the tail management feature provided by CICS TS. One of the changes is the use of CICS/VR to be able to retrieve data from the logstream. SSB is using the

AUTODELETE option on the forward recovery logstream. It only accumulates log data to perform forward recovery functions on the RLS files.

SSB logs application data to user journals. At the end of the online day, batch jobs are run to retrieve application data from these journals. However, the move to CICS TS changed the log data format, requiring a change in the retrieving methodology.

CICS TS provides a compatibility interface to the new format of data written by the system logger. To use this interface, SSB needed to change the JCL and use the options COMPAT41 and COMPAT41V, which allowed their batch programs to retrieve data from the system logger without any changes. The use of these options translates the log data from the new format to the old one.

DFSMS/MVS 1.3 updates the last reference date field in the F1_DSCB when a data set is opened rather than when it is closed. SSB manages system logger offload data sets with SMS and allocates them with a Data Class, Storage Class and a Management Class. The Management Class was set to allow migration after three days of non-usage. Because the last referenced date is the date the file was opened rather than when it was closed, it was possible that the latest offload data set might be migrated when CICS is shut down, even though it is less than three days since the data set was referenced. This behavior created some operational problems in SSB. At CICS start up, it was sometimes found that the most recent offload data set had been migrated and therefore needed to be recalled. This was unacceptable because of the delay this caused in the startup of the region.

SSB had to implement a circumvention to force the recall before the CICS region was restarted. This problem has been solved in DFSMS/MVS 1.4, which now updates the last reference date field when the data set is closed. Ultimately, SSB would like to be able to manage those data sets in the same way that SMS manages GDG files.

3.4.12.3 Sharing Files with Batch

With the previous version of CICS, SSB used to close and deallocate the files to the CICS regions and then start the batch. With CICS TS, VSAM files in data sharing are open in RLS mode to the CICS regions. In order for batch jobs to be able to update these files, it is necessary to first QUIESCE the files to all the sharing CICS regions.

SSB wrote a program to be able to inquire on the status of the RLS files and change their status without external intervention, switching from RLS mode to non-RLS mode and vice versa. This program verifies if the file is open in RLS mode and if so, it will issue the CICS Quiesce command in order to make the file available for batch update access. At the time of writing, SSB only use RLS mode access for CICS - all batch access to VSAM files is still in non-RLS mode. Therefore, if a file is being accessed in RLS mode, SSB knows that it is CICS, and only CICS, that is accessing it. It the file was being accessed in RLS mode by a batch job when the QUIESCE command is issued, the QUIESCE would fail. The same interface is used to switch the files back to RLS mode at the end of the batch window.

SMSVSAM uses a locking mechanism to provide integrity for multiple update access from different CICS regions. It is possible that some locks are still held in the SMSVSAM lock structure in the Coupling Facility without this being

immediately obvious. This prevents access to the locked records to avoid potential integrity problems. If there are retained locks on a given file, any attempted NSR open for update against that file will be refused. However, a Delete/Define of the file will be allowed, almost certainly compromising the integrity of the file by losing the locks. SSB had a few cases where a batch job failed because one or multiple files still had retained locks.

This situation caused SSB to change some of their procedures to support the possibility of having retained locks on an RLS file. Since it was difficult to catch this error in the job chain, SSB wrote a program that processes the output from the SHCDS LIST command and, depending on the presence or absence of retained locks, either releases the suite of batch jobs or initiates recovery actions.

There is further information about the sharing of VSAM files between CICS and batch when RLS is being used, in the redbook *Batch Processing in a Parallel Sysplex*, SG24-5329.

3.4.13 Other Subsystems

For the DB2 portions of the POS data, SSB used the DB2 concurrent reorg capability in order to limit the time that data was unavailable to the application. For the POS data still residing in DL/I databases, SSB used the concurrent reorg product from BMC and the Concurrent Image Copy facility from IBM. They also decided to convert the batch DL/I jobs to BMPs. To be able to benefit from IMS data sharing capabilities, SSB added checkpoint functions to these jobs.

3.4.14 ISV Products

SSB had some key products that had to be available in the RLS environment at the time of migration, in order to be able to provide a complete solution and to continue normal operations. Table 18 contains a list of these products and the actions that were required for each.

Table 18. SSB Vendor Products

Products	Vendor	Usage	Notes
JMP	BMC	Journal Management	Replaced by CICS Logging function
RPCV	BMC	Forward/Recovery	Replaced by CICS/VR
OMEGAMON	Candle	Monitoring/Debugging	Updated
AF/OPERATOR	Candle	Automated Operations	Updated
DMS	Sterling Software	Storage Management	Updated
FileAid	Compuware	VSAM file management	Shareoption 2 PTF installed
SMARTEST	ViaSoft	Development tool	Updated
INTERTEST / SYMDUMP	CA	Online Debugging	Updated

3.4.15 Lessons Learned

The main reason for SSB moving to data sharing was to achieve continuous availability, and especially to protect its application from unplanned outages.

What SSB found is that it also achieved daily benefits such as scalability of the application, unlimited growth and no application outages during the maintenance window.

The POS workload is unpredictable depending on the time of the year, economy, and so on, and in the past it has always been difficult (and risky) to predict if the single region could sustain the future workload (for example, at Christmas or vacation time). This was due to the single region architecture and to the hardware engine capacity.

Following the move to data sharing, the reaction to a changing load is rapid and very easy to manage from an architectural point of view. During times of heavy loads, new AORs can be started to react to the increase of the load, thus maintaining the agreed service levels to the end user. In addition, the engine capacity is removed as a constraint for future growth of the application.

Also on a daily basis, SSB benefits from the capability of putting application changes into the production environment in a controlled way and without causing application outages. A new version of the application can now be activated in one of the AORs, tested with a limited number of transactions, using CP/SM to control the routing, and depending on the result, either remove the changes without affecting the whole application or, if successful, dynamically spread the changes to the remaining AORs. This avoids a lot of downtime for the POS application and provides a better service level to the end user.

3.4.16 Results That Were Achieved

As a result of the implementation of data sharing, the POS application now only has a scheduled downtime of a couple of hours a week. This downtime is used to back up and reorganize the VSAM files. The application is still performing comfortably within the Service Level Agreement, and the application can now continue to provide acceptable service through peak demand periods.

The current production environment is a two-way Parallel Sysplex running OS/390 R1.3 with CICS TS 1.1, CICS/VR 2.3, CICSPlex/SM 1.2, DBCTL 5.1, and DB2 5.1. Figure 8 on page 70 contains a diagram of the major software components.

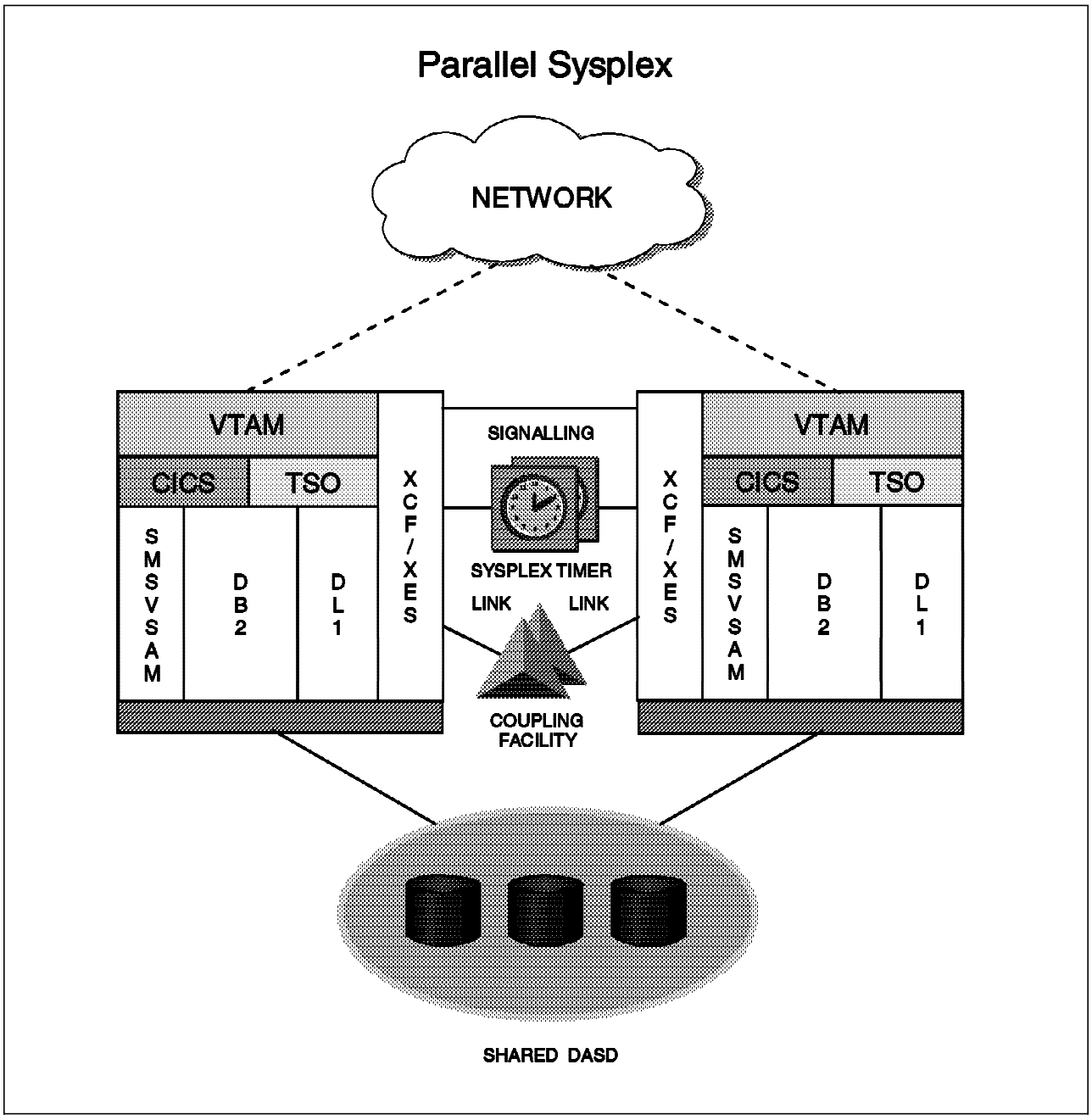


Figure 8. SSB Software Configuration

The current hardware is 2 x 9672-R86 and 2 x 9674-C05 Coupling Facilities with 2 GB of storage each, and is shown in Figure 9 on page 71.

DASD Storage is 2.2 TB of EMC and IBM RVA2 devices.

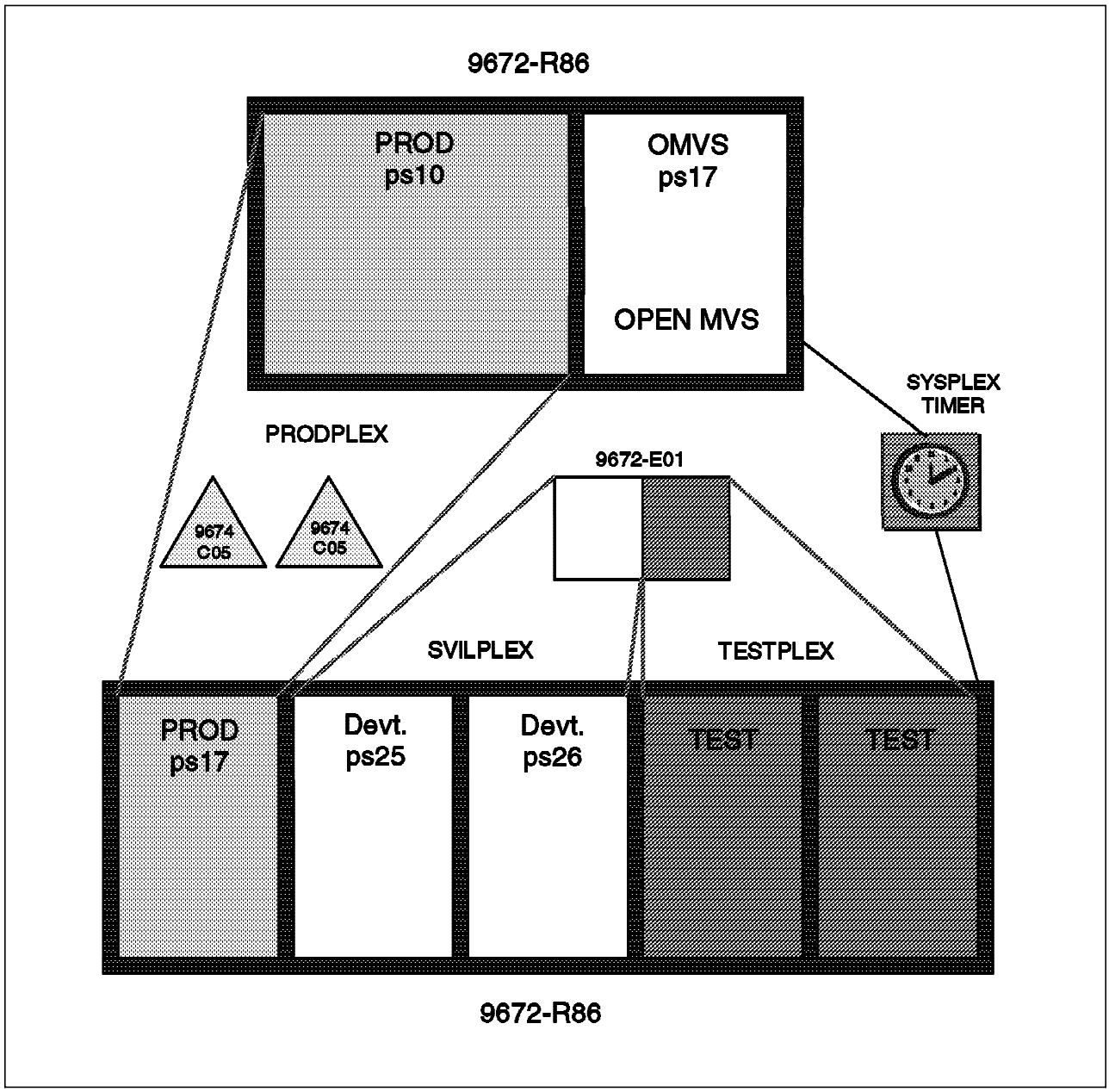


Figure 9. SSB Hardware Configuration

The two OS/390 systems are running in LPARs on the 9672s. On the same processors, SSB has two LPARs for the development two-way Parallel Sysplex. This complex is using a different Coupling Facility on a 9672-E01 (similar to a 9674-C01).

SSB has always had two totally separate environments for production and development. This is to isolate and protect the production environment from any errors that could potentially arise from the test environment.

Each instance of the POS application is based on a topology of a single TOR and one to four AORs. The single TOR is still a single point of failure. SSB could not use VTAM Generic Resources or VTAM Multi-Node Persistent Sessions (MNPS) to balance the logon over multiple TORs because of the way the network is designed in SSB.

As SSB cannot get the availability benefits of multiple TORs, they plan to activate Automatic Restart Manager (ARM) to provide quick restart of a failed TOR.

The current daily transaction rate varies from 800,000 up to 1,500,000 transactions (user and system) on each AOR. The peak transaction rate is up to 100 transactions per second on each AOR.

Transaction routing and balancing is controlled by the queue mode algorithm provided by CP/SM. Data sharing is used for VSAM, DB2, and DL/I. DL/I databases are still used, but a plan is in place to move these to DB2.

SSB currently has about 150 production VSAM files being accessed in RLS mode, with additional files being added as the related applications require the data sharing capability.

The current peak access to RLS files is about 1,000,000 application operations each day to a single file.

In addition to VSAM data sharing, SSB has also implemented the following Parallel Sysplex exploiters as part of the move to a Parallel Sysplex environment:

- XCF is now using both CTC connections and structures within the Coupling Facility.
- In addition to CICS TS log files, the system logger is also being used for Logrec and Operlog.
- IMS/DB V5 data sharing.
- DB2 V5 data sharing.
- Tape sharing with the IEFAUTOS facility.
- RACF data sharing. There is a RACF database in data sharing mode in each environment (production and development), and the RRSF facility is used to keep the two databases in sync.

3.5 Future Plans

Parallel Sysplex implementation is an ongoing project in SSB, and the POS application has not yet achieved fully continuous availability. There are still some single points of failure that SSB is studying and hopes to address in the near future.

One example is the network. One of the future plans is to isolate the network on a separate image in order to separate it from the application processes, thus avoiding application errors or malfunctions interfering with the whole network.

VTAM MNPS and Generic Resources do not help SSB due to their network configuration. To achieve better availability of the TOR function, SSB plans to implement ARM to quickly and automatically restart the failed region in case of a failure. The TORs are not affected by application changes and usually are stable and reliable. TOR unavailability is usually a consequence of a hardware or OS/390 outage.

SSB plans to implement High Performance Routing functions on top of their current APPN configuration to provide better availability and recovery for the NCP.

VSAM Online Reorg is another function that SSB has long required. It plans to implement the Backup_While_Open function to avoid having to stop the application to take file backups. However, there will still be a requirement to stop the application in order to reorg the files. At the time of writing, SSB has been unable to find any product that provides the ability to do online reorg on VSAM files, so this is one outage that cannot be avoided. The best that can be achieved is careful use of the VSAM FREESPACE parameter to minimize the number of CI and CA splits, and thus reduce the frequency of reorgs.

Another capability that SSB is looking forward to is to be able to update VSAM files concurrently from batch and CICS. This capability will completely eliminate the batch window, even though right now it is still something SSB can tolerate. IBM has issued a statement of direction stating that this is a function that will be provided in the future.

SSB also has a short-term plan to migrate other applications requiring high availability to data sharing mode. Other applications have already been migrated, but some of them do not have the capability of dynamically balancing their workload due to the usage of DPL and CICS EXEC START functions. For this reason, SSB is planning to install CICS TS 1.3 to be able to balance these workloads by utilizing the enhancements provided by this new level of CICS.

SSB also plans to expand the usage of WLM goal mode to the CP/SM balancing algorithm to ensure the service levels being delivered at the individual transaction level match those contained in the Service Level Agreement.

Another short-term plan is to move from a GRS ring configuration to GRS star, and to eliminate the use of CTCs for XCF communications, standardizing on CF structures for XCF.

3.6 Summary

SSB has invested a great deal of time and money into implementing Parallel Sysplex for their critical POS application. This investment has been rewarded by significant improvements in their application availability. The level of availability that they can deliver at the moment exceeds that required by their Service Level Agreements.

SSB has already started to leverage that investment by extending the use of data sharing to other critical applications. This allows SSB to improve the service delivered by these other applications for a relatively small incremental cost, and also increases the return on the investment that it has already made in Parallel Sysplex.

SSB sees Parallel Sysplex as a key component to its continuing ability to meet the ever-increasing requirements of its industry in a timely and cost-efficient manner.

Appendix A. The Cost Components of a Business Case

We had hoped to provide a sample business case in this book; however, it was difficult to find a customer that actually creates business cases. And those that have gone to the trouble of creating a formula for accurate business cases, generally feel that that formula is a valuable business asset and are not willing to share it.

The task of creating an accurate business case for IT applications is not trivial. Many of the benefits provided by new applications are intangible, and thus difficult to size. As a result of this, most customers do not actually create a business case when investigating new applications. Instead, the cost of developing the application is just included as one component of the product development costs.

However, if you wish to attempt to create a business case for improving the availability of an application, this appendix provides a list of costs--both costs for providing availability, and costs associated with outages--that can be used as *part* of that business case. We also provide some advice about where you might get some of this information from.

To provide some perspective, we include Table 22 on page 86, which shows outage costs for companies in a variety of industries. Without a detailed study, it is practically impossible to know whether an outage in your company would have similar impact; however, this table provides some idea of the scale of costs that you are looking at.

A very good description of the components contributing to continuous availability is available in the redbooks *Continuous Availability - Systems Design Guide*, SG24-2085 and *Continuous Availability S/390 Technology Guide*, SG24-2086.

A.1 Costs of Availability

Unfortunately for the person trying to justify an availability improvement, the *costs* of providing that improvement are very tangible, and immediately felt, whereas the *value* of availability is much harder to ascertain.

To make matters worse, while your project may have to bear the costs of the improvements, the results of your spending will more than likely benefit others as well. So, for example, if you are trying to justify a second processor, for availability and Parallel Sysplex implementation, it is unlikely that yours will be the only application using that resource once it is installed. The extent to which this is a problem for you varies from company to company. Some companies have very compartmentalized budgeting and planning, while other companies just place everyone's requirements into a pot and divide the result among the participants.

A.1.1 Identify Areas for Improvement

One of the first things that you must do when approaching an availability project is to study your current availability and understand exactly what it is that your customer is trying to achieve. It is conceivable that a single faulty component, such as a local terminal controller, has created the invalid perception that there is an availability problem with an application.

You should review all the sources of information available to you--problem management reports, system logs, operator logs, and so forth--to identify every outage over the past year. Verify this list with the application owner, to ensure you both have the same perception of the current availability.

Next, identify and categorize the root cause for every outage, both planned and unplanned. From this list, identify the items that are having the highest impact on availability.

It is only when you understand what is causing the application to be unavailable at the moment that you can effectively create a plan to improve that availability. You should also use this information to identify which causes of outages you have to address in order to gain the availability improvement that you are aiming for.

The plan is likely to include some change in processes, and may also include hardware and software changes. The following sections help you identify the cost components involved in improving application availability.

A.1.2 Hardware Costs

Early in the project you should carry out a Component Failure Impact Analysis (CFIA). Ideally, this should be done by someone outside your firm, or at least not affiliated with the application in question, to provide objectivity to your project. This analysis should identify both the single points of failure and those components whose loss would have a serious impact on the application availability. The only way to provide continuous operations is to have redundancy for all critical components.

The next step is to take the result of that study and discuss it with your capacity planners. It may well be that some of the identified components are about to be addressed as part of your normal upgrade process.

The remaining components will need to be sized and priced. This may at first appear difficult, given the rate at which hardware prices are currently dropping. However, depending on the scope of your project, you are likely to find that additional hardware only constitutes a relatively small part of the overall project cost, so pinpoint accuracy is not a requirement.

Table 19 on page 77 contains a list of hardware components that you should ensure are covered by the CFIA study. Depending on the results of the CFIA study, this may identify additional hardware costs that should be included in your project costs.

Table 19 (Page 1 of 2). Hardware Components for Availability

Component	Notes
Premises	Depending on the application, it is conceivable that you would have to consider splitting the application over two locations, to provide disaster recovery capability using a facility such as Geographically Dispersed Parallel Sysplex (GDPS).
Power Supplies	If your site is not equipped with Uninterruptable Power Supplies and generators, you may need to provide these, especially in locations that suffer a high incidence of power interruptions.
Processor	Check if sufficient spare capacity exists on any existing processors
Sysplex Timer	The 9037-001 Sysplex Timers had to reside in the same location due to a limited distance of 3 meters between the two processors. You may wish to replace these timers with model 2 timers which can be placed further apart--up to 3 kilometers, or 26 kilometers with repeaters. The loss of both timers in a sysplex environment will stop all processors in the sysplex.
Coupling Facility	If this is your first data sharing application, you may need to purchase a stand-alone Coupling Facility or an ICF on an existing processor. You should consult with your local Parallel Sysplex expert to ensure that your Coupling Facility configuration provides the level of availability and flexibility required, *taking into account the products that will be using the Coupling Facility*. If you have existing Coupling Facilities, they may require additional storage to support your workload. You should also ensure that there are at least two CF links between each Coupling Facility and every processor in the Parallel Sysplex.
ESCON Director	If already installed, you may only need additional ports. All critical devices should be attached to each processor via at least two ESCON directors. Ideally ESCON Manager should be used to provide maximum manageability of your directors.
DASD Subsystem	There are three reasons for expenditure on new DASD: • If you are implementing Parallel Sysplex for the first time. It is strongly recommended to keep DB2 archive logs on DASD rather than on tape in a DB2 data sharing environment. Depending on the volume of log data produced by your application, this could add up to a sizeable amount of DASD. • To provide RAID or mirroring capability. • To provide remote copy capability. If you are considering remote copy, be sure to include the cost of connecting the two DASD subsystems.
Tape Subsystem	You may consider a remote tape library, so that backups and recovery logs are immediately offsite in case of a disaster.
Network Controllers	Depending on where the bulk of the users of your application are, you may wish to provide additional network controllers to reduce the impact any time a controller has a planned or unplanned outage.

Table 19 (Page 2 of 2). Hardware Components for Availability	
Component	**Notes**
Consoles	Because the console traffic in a sysplex gets routed to all systems in the 'plex, it is vital to ensure that there is always an alternate console available to take over in case of a failure of any console. While it has always been important to provide for console availability, it is even more important in a sysplex, as a console problem on one system can have a ripple effect on other systems in the sysplex.

A.1.3 Software

Just as redundant hardware is required to provide for continuous operations, similarly, redundant software is also required. It is probable that this will involve some additional expenditure. If additional processors are required by the project, you will need to provide the base software stack--OS/390 and other required base software--plus any other software required by the application on those additional processors.

If you are considering your first data sharing project, you may have to purchase an additional license for the database and/or transaction manager.

If you do not have an existing application change management product, such as the Software Configuration and Library Manager (SCLM) component of ISPF, we strongly recommend that you implement one. As SCLM is part of ISPF, there is no additional license cost; however, OEM equivalents will carry a cost.

Similarly, if you do not have a change and problem management system such as Info/Man, you should seriously consider purchasing one as part of the project. Poor change and problem management causes far more downtime than defective hardware or software. Having a good change and problem management product will not magically give you a good change and problem management process, but it is a key tool in the implementation and effectiveness of that process.

Finally, if you do not have one, you should seriously consider purchasing a product to assist with stress and regression testing. TPNS is an IBM product that many customers and IBM itself use for automated testing of new hardware, software, and applications.

Table 20 contains a list of software products that may be required to improve the availability of your application.

Table 20 (Page 1 of 2). Software Products for Availability	
Product	**Notes**
Software Upgrade Charges	If your plan entails upgrading existing processors, you will more than likely have to pay an upgrade charge for all the software installed on that processor.
OS/390 and related base software products	If your plan entails the procurement of an additional processor, you will have to budget for the cost of the base software stack for that machine.
Database and Transaction Managers	If you will require products on images that do not currently contain those products, you may need to purchase additional licences.

Table 20 (Page 2 of 2). Software Products for Availability	
Product	**Notes**
Application Change Control	In most companies, errors within applications are far more numerous than errors within system software. A powerful change control product is a key tool to help you get more structure and control over your application programs; however, such tools can be very expensive.
Change and Problem Management	You want a tool that is easy to use and tailor, maintains an historical database with powerful search capabilities, and has an API to facilitate automated problem recording.
Other Utilities	You should have a complete inventory of all the products used by the application in question. You must make sure that licences are held for every processor that the application might run on. You may also be able to get significant availability improvements by utilizing the features provided by the latest utilities. For example, file backup and reorganization are two of the largest causes of planned outages, yet these can be addressed, or at least reduced by the latest levels of the database utilities and database managers.
Automation Product	If you wish to keep an application available, there are some key requirements for your recovery actions: • They must be fast, so the problem can be addressed before it spreads. • They must be complete--you want to make sure that you do not miss any key messages. • They must be tested. You will want to create any possible failure and test that your response has the desired effect. • They must be consistent. If the response does not fix the problem, you want to be sure that you are aware of all the actions that were taken so that you can accurately diagnose what went wrong and fix your process. The best way of ensuring all of these is to automate as much as possible. Problem recording should happen automatically, to ensure that no problems are overlooked. And, as far as possible, the responses to problems should also be automated. In the heat of the moment, it is easy for people to take inappropriate or incorrect actions. If you have tested and automated your recovery procedures, the recovery from problems *should* be more successful overall.

A.1.4 Processes

The current industry trend is that hardware prices are dropping sharply, software costs are more or less stable, but labor costs are rapidly increasing. This being the case, the labor content of your project is likely to be the highest cost area.

Parallel Sysplex is a tool that IBM provides to help you achieve improved availability. However, like any tool, the end result depends on how effectively you use it. Also, it is only one part of the equation. Perhaps the most important part is your processes.

Your processes determine how you use the available technology. For Parallel Sysplex, your processes must be designed to exploit the capabilities that it delivers. For example, if you are not doing rolling restarts of your subsystems,

then your processes are holding you back from getting the maximum value from Parallel Sysplex.

Your processes also determine how you respond to problems. As stated in Table 20 on page 78, if you want to get the maximum benefit from automation, you must have processes in place to create comprehensive automation rules, *and* to continually test those rules to make sure they are still working and appropriate.

Another contributor to system availability is your service strategy. IBM recommends keeping within six months of the current service level for optimum availability and serviceability. Your processes should allow for updating the system on a regular basis, and monitoring HIPERs and PEs in the interim.

Perhaps the largest contributor to application and overall system availability is a comprehensive, automated, and *up-to-date* testing capability. The more thorough and representative your test cases, the more likely you are to catch any problems before they reach your production system. However, the investment required to create good test cases should not be underestimated. Not only is there a lot of work involved in creating the test cases and test environment in the first place, but even more important is that the test cases are kept up to date. As you implement changes to applications, the test case must be changed to test the new and the original function. As you create new applications, new test cases must be created, along with all the required databases, flat files, and so forth. And as you start to utilize new system functions, you need to create test cases to test those functions.

IBM performs a great deal of testing on both its products and on the service to those products, and this has led to improvements in the quality of code being delivered to customers. However, we cannot replicate all the applications of every customer, and the subtleties inherent to each application. Therefore, it is vital that your test cases are representative of the use that you make of your system.

Table 21 contains a list of some of the processes that you are going to have to create or modify as part of your project to improve application availability.

Table 21 (Page 1 of 3). Processes and Actions

Process/Action	**Notes**
Implement New Hardware	If you have identified a need for additional hardware, there will be a cost associated with implementing that hardware, and integrating it into the existing systems management processes.
Implement New Software	If you have identified a need for new or upgraded software, there will be a cost involved in installing and customizing that software. It is likely to be far more costly however to actually set up that software and integrate it into your existing processes.
Implement Parallel Sysplex Data Sharing	Depending on whether this is your first application to move to data sharing, the level or investment will vary. If you already have experience with data sharing, the incremental cost of implementing data sharing for another application should be much lower, assuming there are a manageable number of affinities. There are many other sources of detailed information about the cost of implementing data sharing.

Table 21 (Page 2 of 3). Processes and Actions

Process/Action	Notes
Implement/Improve Testing	This includes setting up and maintaining a set of test cases of your applications, application changes, and system facilities that you use. You also have to allow for the time to run the tests and check the output from those tests.
Implement/Improve Change Management	Your change management process should have the following attributes: • It should identify an owner for every change. • It should ensure that all interested parties are automatically notified. • It should be easy to use. If it is overly complex, people will not use it and it will lose its effectiveness. • It should be tied to the problem management system so that any problems caused by the change can be associated with the change. This provides valuable information for future problem diagnosis. If your current change management system does not meet these criteria, you should enhance or change it.
Ongoing Change Management	Having provided a sound change management infrastructure, you have to be prepared to make the investment in using it. This means allowing extra time for every change, to cover the administration work. Plus you must allow the time required for people to check the changes proposed by other people, to make sure there are no clashes or potential problems. And you have to have regular meetings *at least* once a week to review the proposed changes.
Implement/Improve Problem Management	Your problem management process should have the following attributes: • It should have an API to allow automatic problem reporting and notification. • It should provide a powerful search engine to enhance future problem diagnosis. • It should have a structure to allow easy searching and reporting. If your current problem management system does not meet these criteria, you should enhance or change it.
Ongoing Problem Management	Having made the investment in providing a powerful problem management system, you have to make an ongoing investment in using that system. You have to allow time for regular problem management meetings, somewhere between daily and weekly, depending on the size of your installation. You have to make sure that the problem resolutions are documented in the problem management system, to speed up future problem determination. You should have a mechanism in place to make sure that all problems are actually being recorded and tracked in the problem management system. And you must have regular analysis to identify any trends that can be addressed to reduce the number of future problems.

Table 21 (Page 3 of 3). Processes and Actions	
Process/Action	**Notes**
Ongoing Automation Support	It is vital that the maintenance of your automation procedures is an ongoing task. New messages frequently get added by new releases of products, or even by service. Also, as you start using new system facilities, new messages may start to appear. You have to ensure that your automation not only caters for the normal day-to-day messages, but most importantly, for those messages that only appear in a problem situation. There are various tools available to help you identify which messages are automated and which are not. These will generally help you trap all the common messages. But if you rely solely on these tools to identify new messages, you will not have automation in place to handle new problem scenarios until after the first occurrence. Therefore, you should use your system programmer test environment to develop and test your automation routines, to ensure that you have already addressed the new problem messages before they hit your production systems.

A.2 Identify Value of Availability

Unfortunately, identifying the value of availability, or the cost of unavailability, is much more difficult than arriving at the cost or providing that availability. For example, if you lose a network controller, what is the cost impact of that failure? Well, it depends on many things, for example:

- Which applications do the users on that controller use? If the users are all application developers accessing a test system, the cost will be lower than if the users are all customer service representatives accessing the order processing application.
- During which shift did the outage occur? It is likely that an outage of a network controller will have less impact during the night shift than the prime shift.
- If during the prime shift, what time did the outage occur? One controller might be busiest during the lunch hour, whereas another controller might be least busy at the same time.
- How long did it take to recover? The cost per minute for a one-minute outage is likely to be far less than the cost per minute outage for a four-hour outage.

To make matters worse, how do you report availability? There are so many components involved in delivering an application, and so many users of each application, what do you report on? If you report based on component availability, the reporting will be horrendously complicated, and not really all that informative. For example, if one of your tape drives is down, what is the impact of that? Just reporting the unavailability of the failed component does not in itself give you any idea of the cost to the business.

On the other hand, if you report based on application availability, how do you cater for the fact that a component failure (for example, a terminal controller) has only affected a small percentage of your application users?

There are no easy answers to these questions. Customers that we have spoken to have used component reporting, application reporting, and many combinations of the two. And most of them readily agree that there are weaknesses in their methodology, whatever it may be.

Unfortunately, the bottom line is that there is no easy method for arriving at a foolproof, all-encompassing formula for valuing availability.

However, there are still things that you can do to arrive at an initial cost. In the following sections, we list some of the cost factors and suggest sources of additional information.

A.2.1 Lost Business

One of the most obvious impacts of a loss of availability is business that may be lost as a result. The amount of business that is lost can vary from individual transactions, which is unfortunate, to loss of the customer, which can be serious if it happens often enough.

It may be possible to identify the cost of lost transactions. If the amount of business transacted by your application is consistent, it may be possible to compare the average value of business with the amount of business transacted on the day that an application outage is suffered. This would probably require a change in processes, at least, and possibly an application change to tally and record the amount of business transacted each day.

Depending on your business, it is likely to be more difficult to identify the loss of a customer, and tie that loss to the fact that the application was unavailable. If you have a relatively small number of high-value customers, you will probably have a close relationship with those customers, and they may make you aware of why they moved their business elsewhere.

On the other hand, if you are in the retail industry, it is unlikely that you will be able to produce definite figures for the number of customers lost as a direct result of application unavailability. The best you can possibly do is detect a trend in the amount of repeat business following a series of application outages. Again, this will require working with the application owner to obtain and record the required information.

It is likely that a single outage may result in lost transactions, whereas a series of outages may result in lost customers. So the cost of each outage is also impacted by the frequency of outages.

A.2.2 Image/Publicity

As business becomes more and more computerized, and electronic links between suppliers and customers become the norm rather than the exception, the availability of your applications becomes far more visible to people outside your company.

If you suffer from recurring application outages, this will quickly become known and you may find that this impacts your ability to win new contracts, or renew existing ones.

Similarly, with the huge growth in use of the Internet and the number of customers involved in e-commerce, the availability of your applications is far more public that in the past, and the potential customers more transient. If a

potential customer cannot transact the business he wants on your system, he can simply go to a competitor with a click of his mouse.

To make matters worse, these potential customers could be anywhere in the world. So your view of outage impact has to be modified to reflect this. You can no longer assume that an outage at 3:00 AM in your time zone is less important than an outage at 3:00 in the afternoon.

If your company becomes known for poor availability, that type of publicity is very difficult to rectify. Even worse, customers may begin to doubt the integrity of your applications, and move their business elsewhere as a result.

If it was difficult to identify the cost of lost customers, it is nearly impossible to assess the cost of poor publicity caused by poor availability. Perhaps your best chance of identifying the cost would be to speak to your company's public relations department. They should be aware of any existing negative publicity and should have some idea of the cost of improving public perception of the company.

A.2.3 Fines and Penalties

If bad publicity is difficult to cost, fines and penalties imposed as a result of application unavailability are not. In some industries, application availability is monitored by controlling bodies, and companies are expected to maintain a certain level of availability.

For example, some countries impose fines on airlines that do not use their allotted flight windows. If the system is down, causing a delay of a flight, this could lead to the imposition of significant financial penalties.

There are also moves within the financial world to encourage companies to maintain high levels of availability. Extended outages could lead to fines from the governing bodies.

A.2.4 Staff Costs

If you suffer a prolonged outage, or a number of smaller outages, there may be a significant staff cost both during the outage and afterwards.

Depending on the application that is affected, a prolonged outage may have a significant financial cost in lost productivity. For example, if the application controls a factory production line, there could conceivably be thousands of people sitting around with nothing to do, with a cost running into tens or even hundreds of thousands of dollars.

Worse still is the overtime that must be paid to catch up on where you would have been without the outage. Taking the factory example again, overtime may be paid at 1.5 times normal salary. So if 2000 workers, each paid 15 dollars an hour, are idle for four hours, the total cost would be 300,000 dollars--120,000 for the outage, and 180,000 for overtime to recover.

It should be possible to identify the users of a given application, roughly how many there are, and what the impact of an application outage is on those people. Multiply that by the average salary per hour and you have a rough idea of the cost in terms of lost productivity. If the lost productivity has to be made up by overtime, factor in the overtime rate, and you have a rough recovery cost.

On top of this, you must add in the cost of the IT staff involved in recovering from the outage, and supporting any additional availability hours that may be required by the end users to catch up with their work.

A.2.5 Impact on Business Decisions

This is another area that is difficult to size, particularly in advance. And it is very dependent on the type of application.

If the application is a Business Intelligence type of application, the cost may vary from very small, to very large, depending on how timely the information must be, and the type of data involved.

If the application is an order processing application for a large factory that works on a just-in-time basis, the impact could be significant. If items are not ordered in time, the whole factory could come to a halt due to the lack of one key component.

You need to work with the application owners to identify the impact of the application not being available for a given amount of time. Try to identify both a worst case scenario and a best case scenario, and a cost for each, then agree on a realistic average cost. In an actual outage situation, the real cost is likely to vary quite widely from your estimate.

A.2.6 Information Sources

There are a number of places where you may be able to obtain some data to help in your calculations.

A.2.6.1 Application Business Case

If a business case was created for the application when it was first developed, there may be some cost benefit estimates in that document that may be useful. However, you have to factor in the age of the document to determine the applicability of those figures. You may be fortunate and find that some recent enhancements to the application have an associated business case, which should contain updated figures.

A.2.6.2 Disaster Recovery Business Case

If your company has a disaster recovery agreement, it is likely that a business case had to be presented in relation to that expenditure. That business case should contain estimates of the impact of a system outage. It is possible that the business case will go down to an application or system level, in which case you can use those figures in your estimates. Even if the business case does not go down to that level of detail, you should be able to extrapolate some idea of the financial cost of the loss of a given application.

You should also speak to the developers of that business case to see where they obtained the financial impact information to help them build the case.

A.2.6.3 Transaction Values

Depending on the type of application, it may be relatively easy to find out the value of business transacted by that application in an average day. This information can be recorded and used to assess the impact of an outage of that application. Many applications print check sums as part of their audit trail, so it should be possible to extract that information.

You should also record the number of transactions per day for each application. If you have an outage, you can at least identify the change in the number of transactions. If you can agree on an average value per transaction, this allows you to fairly easily arrive at an estimate of the financial impact of the lost transactions.

A.2.6.4 Industry Surveys

Table 22 contains a list of sample outage costs. In fact, realistic outage costs are now probably higher, as these figures are extracted from a Datamation study in 1995.

Table 22. Sample Outage Costs

Business	Industry	Cost Range (per hour)	Average Cost (per hour)
Brokerage Operations	Finance	$5.6 to 7.3 Million	$6.45 Million
Credit Card/Sales Authorization	Finance	$2.2 to 3.1 Million	$2.6 Million
Pay Per View	Media	$67 to 233 Thousand	$150 Thousand
Home Shopping (TV)	Retail	$87 to 140 Thousand	$113 Thousand
Catalog Sales	Retail	$60 to 120 Thousand	$90 Thousand
Airline Reservations	Transportation	$67 to 112 Thousand	$89.5 Thousand
TeleTicket Sales	Media	$56 to 82 Thousand	$69 Thousand
Package Shipping	Transportation	$24.5 to 32 Thousand	$28 Thousand
ATM Fees	Finance	$12 to 17 Thousand	$14.5 Thousand

A.3 Summary

The items listed in this appendix are by no means all-encompassing. Every industry, and every company, will have unique costs and requirements.

And remember that tangible outage costs are only one part of the equation. You may be fortunate enough to suffer *no* unplanned outages at the moment, yet still require better application availability. The following lists some of the drivers for higher application availability:

- To maintain competitiveness. If all of your competitors are offering 24 x 7 online service, you may have no choice but to move in that direction, even though there is currently not enough business outside the normal business hours to justify the change by itself.
- Additional availability may give you access to a set of customers that you currently do not address--for example people on shift work may wish to carry out all their banking business in the middle of the night. If you are the only bank offering this capability, it certainly gives you a competitive advantage.
- The advent of Internet shopping introduces a completely new pattern of consumer behavior. Previously, once you had a customer in your shop, he

was more inclined to wait ten minutes for the system to come back than to get back in his car and drive 30 minutes to your competitor. Now, it takes the consumer ten seconds to move from your “shop” to your competitors “shop.” If you have better availability that your competitors, you have the opportunity to pick up customers from your competitor's site while his systems are down.

Another example is the credit authorization business. Most retail outlets will have the ability to use more than one credit authorization provider, but will tend to stick with the one provider until an interrupt occurs. At that point, they will switch to one of the other providers, and stay there until the next interrupt. Once again, if your systems have better availability, you have the opportunity to pick up your competitor's business when he suffers an outage.

- If your customer support systems are available 24 x 7, this gives you the flexibility to have fewer call center staff. Once your customers realize that they can get service at any time, there will be a trend towards fewer calls during the day, with more in the off-peak hours, thus allowing you to reduce the number of operators required to answer the volume of calls at a given time.
- Once again, if you can spread the workload associated with servicing your customers over a longer period of time, the peak processing power required to service that workload will decrease. Depending on how busy your current processor is, and your software licensing agreements, this could actually contribute to a significant savings, or at least a deferral of additional expense.
- Due to mergers or business growth, you may be required to support multiple time zones. As the number of zones that you have to support increases, the number and duration of acceptable outage times rapidly disappears.

A successful business case will include these considerations, plus many more that are specific to your company and circumstances. Most importantly, a successful availability project requires total commitment on the part of management and staff to work together towards a common objective.

Appendix B. Special Notices

This publication is intended to help systems programmers and availability specialists plan for improved availability through the use of Parallel Sysplex technology. The information in this publication is not intended as the specification of any programming interfaces that are provided by OS/390, CICS TS, IMS/ESA, or DB2. See the PUBLICATIONS section of the IBM Programming Announcement for OS/390, CICS TS, IMS/ESA, and DB2 for more information about what publications are considered to be product documentation.

References in this publication to IBM products, programs or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent program that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program or service.

Information in this book was developed in conjunction with use of the equipment specified, and is limited in application to those specific hardware and software products and levels.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to the IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact IBM Corporation, Dept. 600A, Mail Drop 1329, Somers, NY 10589 USA.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS. The information about non-IBM ("vendor") products in this manual has been supplied by the vendor and IBM assumes no responsibility for its accuracy or completeness. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

Any pointers in this publication to external Web sites are provided for convenience only and do not in any manner serve as an endorsement of these Web sites.

Any performance data contained in this document was determined in a controlled environment, and therefore, the results that may be obtained in other

operating environments may vary significantly. Users of this document should verify the applicable data for their specific environment.

This document contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples contain the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Reference to PTF numbers that have not been released through the normal distribution process does not imply general availability. The purpose of including these reference numbers is to alert IBM customers to specific information relative to the implementation of the PTF when it becomes available to each customer according to the normal IBM PTF distribution process.

The following terms are trademarks of the International Business Machines Corporation in the United States and/or other countries:

APPN
CICS
CICS/ESA
DB2
DFSMS
DFSMS/MVS
ESCON
IBM
IMS/ESA
MVS/ESA
OpenEdition
PR/SM
RACF
RMF
S/390
Sysplex Timer
VTAM

The following terms are trademarks of other companies:

C-bus is a trademark of Corollary, Inc.

Java and HotJava are trademarks of Sun Microsystems, Incorporated.

Microsoft, Windows, Windows NT, and the Windows 95 logo are trademarks or registered trademarks of Microsoft Corporation.

PC Direct is a trademark of Ziff Communications Company and is used by IBM Corporation under license.

Pentium, MMX, ProShare, LANDesk, and ActionMedia are trademarks or registered trademarks of Intel Corporation in the U.S. and other countries.

UNIX is a registered trademark in the United States and other countries licensed exclusively through X/Open Company Limited.

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

Appendix C. Related Publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this redbook.

C.1 International Technical Support Organization Publications

For information on ordering these ITSO publications see “How to Get ITSO Redbooks” on page 93.

- *Batch Processing in a Parallel Sysplex*, SG24-5329
- *Continuous Availability - Systems Design Guide*, SG24-2085
- *Continuous Availability S/390 Technology Guide*, SG24-2086
- *DB2 for MVS/ESA Version 4 Data Sharing Performance Topics*, SG24-4611
- *DB2 for MVS/ESA Version 4 Data Sharing Implementation*, SG24-4791
- *DB2 for OS/390 Capacity Planning*, SG24-2244
- *DB2 UDB for OS/390 Version 6 Performance Topics*, SG24-5351
- *High Availability Considerations: SAP R/3 on DB2 for OS/390*, SG24-2003
- *OS/390 MVS Parallel Sysplex Configuration, Volume 2: Cookbook*, SG24-2076

C.2 Redbooks on CD-ROMs

Redbooks are also available on the following CD-ROMs. Click the CD-ROMs button at http://www.redbooks.ibm.com/ for information about all the CD-ROMs offered, updates and formats.

CD-ROM Title	Collection Kit Number
System/390 Redbooks Collection	SK2T-2177
Networking and Systems Management Redbooks Collection	SK2T-6022
Transaction Processing and Data Management Redbooks Collection	SK2T-8038
Lotus Redbooks Collection	SK2T-8039
Tivoli Redbooks Collection	SK2T-8044
AS/400 Redbooks Collection	SK2T-2849
Netfinity Hardware and Software Redbooks Collection	SK2T-8046
RS/6000 Redbooks Collection (BkMgr Format)	SK2T-8040
RS/6000 Redbooks Collection (PDF Format)	SK2T-8043
Application Development Redbooks Collection	SK2T-8037

C.3 Other Publications

These publications are also relevant as further information sources:

- *DB2 Data Sharing: Planning and Administration*, SC26-3269
- *PR/SM Planning Guide*, GA22-7236

How to Get ITSO Redbooks

This section explains how both customers and IBM employees can find out about ITSO redbooks, redpieces, and CD-ROMs. A form for ordering books and CD-ROMs by fax or e-mail is also provided.

- **Redbooks Web Site** http://www.redbooks.ibm.com/

  Search for, view, download, or order hardcopy/CD-ROMs redbooks from the redbooks Web site. Also read redpieces and download additional materials (code samples or diskette/CD-ROM images) from this redbooks site.

  Redpieces are redbooks in progress; not all redbooks become redpieces and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

- **E-mail Orders**

  Send orders by e-mail including information from the redbook fax order form to:

In United States:	e-mail address: usib6fpl@ibmmail.com
Outside North America:	Contact information is in the "How to Order" section at this site: http://www.elink.ibmlink.ibm.com/pbl/pbl/

- **Telephone Orders**

United States (toll free)	1-800-879-2755
Canada (toll free)	1-800-IBM-4YOU
Outside North America	Country coordinator phone number is in the "How to Order" section at this site: http://www.elink.ibmlink.ibm.com/pbl/pbl/

- **Fax Orders**

United States (toll free)	1-800-445-9269
Canada	1-403-267-4455
Outside North America	Fax phone number is in the "How to Order" section at this site: http://www.elink.ibmlink.ibm.com/pbl/pbl/

This information was current at the time of publication, but is continually subject to change. The latest information may be found at the redbooks Web site.

IBM Intranet for Employees

IBM employees may register for information on workshops, residencies, and redbooks by accessing the IBM Intranet Web site at http://w3.itso.ibm.com/ and clicking the ITSO Mailing List button. Look in the Materials repository for workshops, presentations, papers, and Web pages developed and written by the ITSO technical professionals; click the Additional Materials button. Employees may access MyNews at http://w3.ibm.com/ for redbook, residency, and workshop announcements.

IBM Redbook Fax Order Form

Please send me the following:

Title	Order Number	Quantity

First name Last name

Company

Address

City Postal code Country

Telephone number Telefax number VAT number

□ Invoice to customer number

□ Credit card number

Credit card expiration date Card issued to Signature

We accept American Express, Diners, Eurocard, Master Card, and Visa. Payment by credit card not available in all countries. Signature mandatory for credit card payment.

Index

ITSO Redbook Evaluation

Parallel Sysplex Continuous Availability Case Studies
SG24-5346-00

Your feedback is very important to help us maintain the quality of ITSO redbooks. **Please complete this questionnaire and return it using one of the following methods:**

- Use the online evaluation form found at http://www.redbooks.ibm.com/
- Fax this form to: USA International Access Code + 1 914 432 8264
- Send your comments in an Internet note to redbook@us.ibm.com

Which of the following best describes you?
__Customer __Business Partner __Solution Developer __IBM employee
__None of the above

Please rate your overall satisfaction with this book using the scale:
(1 = very good, 2 = good, 3 = average, 4 = poor, 5 = very poor)

Overall Satisfaction _____

Please answer the following questions:

Was this redbook published in time for your needs? Yes____ No____

If no, please explain:

_____

_____

_____

_____

What other redbooks would you like to see published?

_____

_____

_____

Comments/Suggestions: (THANK YOU FOR YOUR FEEDBACK!)

_____

_____

_____

_____

_____

SG24-5346-00
Printed in the U.S.A.