Enterprise Server Division

High Availability in an IMS Environment

IMS XRF and Parallel Sysplex:
A Positioning Paper

January 2001

Author: David Raften

Contact: David Raften
raften@us.ibm.com

Introduction

IMS is the fastest, most reliable database computing system in the world, plain and simple. When immediate access to mission-critical information is imperative, over 95% of the world's major corporations rely on IMS to provide a continuous link to data that is accurate, up-to-date, and quickly accessed by many end users. Customers rely on IMS systems to process billions of vital transactions a day. Any time you make an airline reservation, rent a car, get cash from an ATM, or pick up a prescription from the pharmacy, chances are you've used IMS.

IMS provide many features to provide availability and recovery of the IMS systems. There are four features and/or products in particular that can be used to provide a high availability environment for IMS systems and the data in IMS databases. They rely on duplicating IMS subsystems and possibly data on another OS/390 or z/OS system.

Extended Recovery Facility (XRF)

In 1987, IMS was the first major database manager to provide a high availability, fault tolerant solution. This was done with the extended recovery facility (XRF). XRF is delivered as an integral part of the IMS program product. Availability is improved by using additional resources to minimize the impact of certain events that disrupt service to the end users. The time that end users cannot access the system is reduced, for some users down to seconds, and their involvement in the recovery process is simplified.

Remote Site Recovery (RSR)

IMS continued its long history of being the first to provide new technology in a commercial transaction processing product in 1994 with RSR, extending IMS recovery to remote sites for disaster recovery. This further eliminates single points of failure that can disrupt end user service. Changes to an active IMS system's resources are tracked at a remote site that can then takeover the IMS workload should an extended outage of the active system occur that effectively disables the active computing site (such as a planned power shutdown, fire, flood or earthquake). RSR provides for remote takeover with minimum or no data loss without reducing availability of the active site during normal operations. Customers requiring very high availability can use RSR to reduce recovery time from days or weeks to hours. IMS TM/DB, CICS/IMS DB, IMS DB batch and IMS TM are supported as are XRF and data sharing.

Parallel Sysplex Data Sharing

Also in 1994, IMS was the first data base to announce support of Parallel Sysplex n-way data sharing. Although IMS has supported the 2-way data sharing environment since the mid-1980's, that solution required the use of IRLM message passing with VTAM to manage locking requests.

Due to this overhead, only a 2-way solution was practical. IMS Parallel Sysplex customers can benefit from reduced computing costs, improved performance, and incremental growth up to 255 IMS subsystems and 32 operating system images (nodes) running in a S/390 Parallel Sysplex configuration. This parallel processing support is available for a wide range of data bases including Full Function, and Fast Path DEDBs. DEDBs with the SDEPs and VSO features were supported in IMS V6.

Also with IMS V6, IMS delivered enhancements in message processing for clustered systems that provides increased capacity, incremental growth, automatic workload balancing (within a Sysplex), enhanced reliability, and increased availability. This new capability, IMS Shared Message Queues (SMQ), utilizes the Sysplex Coupling Facility for shared queues that service multiple IMS Transaction Managers. A message can be shared and processed by any IMS sharing the queues. Shared queues can be used to distribute work across the Sysplex. For IMS Fast Path users, the new Expedited Message Handler provides the same function with similar benefits to users. In addition, VTAM Generic Resource support provides a single system image to end users wishing to log on to the data sharing group.

Geographically Dispersed Parallel Sysplex (GDPS)

Developed independently of IMS is the Geographically Dispersed Parallel Sysplex disaster recovery solution. GDPS is a multisite management facility that is a combination of system code and automation that utilizes the capabilities of Parallel Sysplex technology and storage subsystem mirroring to manage processors, storage, and network resources. GDPS is an integrated D/R readiness solution and handles all the complexity of switching the network, the systems and the DASD subsystems.

The main focus of GDPS is to make sure that, whatever happens in the primary site, the image of all data in the surviving location is time consistent. **Time consistent** means the secondary disks contain all updates until a specific point in time, without anything missing, and no updates beyond that point. GDPS will actually go one step further: it will also prevent logical contamination of secondary data so that the plethora of errors that should be expected during a rolling disaster will not be copied forward to the surviving site, either fully or in part.

Overall Availability Strategy

The solutions described above concentrate on just one of the approaches to improving system availability. They should be used in combination with other techniques as part of an overall system availability improvement strategy.

A good place to start in developing a plan to improve system availability is to understand the extent, impact, and root causes of recent outages. Once the specific types of outages that have been occurring in the environment have been understood, action plans for dealing with each can

be developed. General approaches for improving system availability include:

Reducing the FREQUENCY of failure: The outage analysis will often identify some types of failures for which corrective action can prevent the problems from recurring. For example, improved testing of system and application changes and improving change control disciplines often help reduce the frequency of failure.

MASKING the failure: The occurrences of some types of failures, such as I/O errors, can be masked so that they do not cause a visible application outage to the end user. Examples include RAMAC devices, IMS's I/O Toleration buffer, TCP/IP's VIPA Takeover, and Parallel Sysplex.

Reducing the DURATION of outages: The duration of system outages can often be significantly reduced by automating the recovery process, reducing the dependence on humans to detect the problem and execute recovery actions. XRF, RSR, and GDPS all work to effectively reduce the duration of the outages.

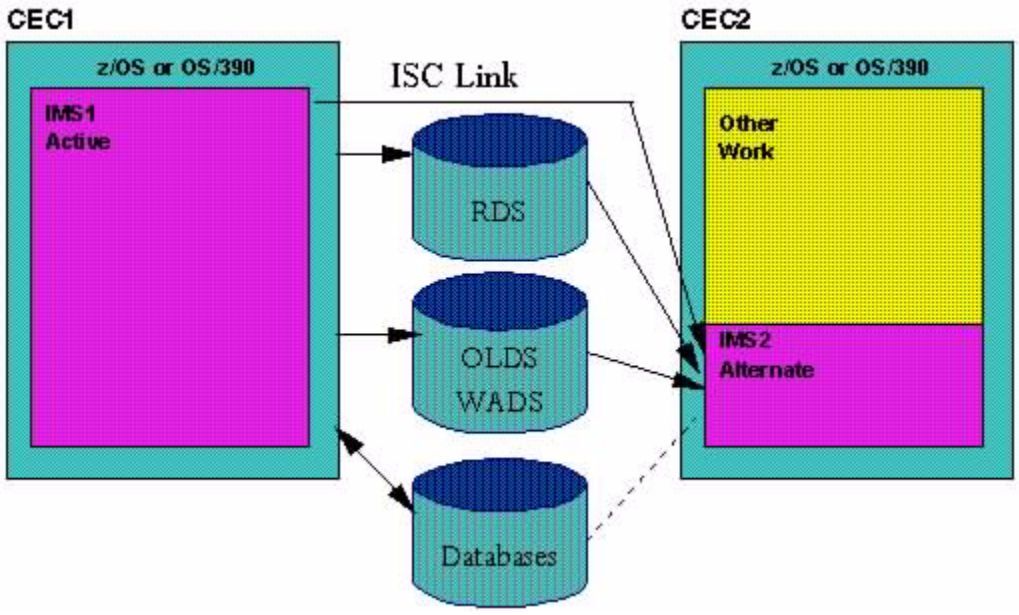
Extended Recovery Facility (XRF)

XRF works by having a second IMS system running. This alternative IMS system runs on a separate OS/390 or z/OS image, which should be on a physically separate machine. The alternate tracks the work on the active IMS system via the IMS log data sets. If certain disruptions occur, the alternate IMS, with the help of network support, takes over the workload of the active IMS system. It is the installation's option whether the takeover is initiated automatically, or is under operator control.

XRF is delivered as an integral part of the IMS program product. It is intended to provide increased availability for IMS subsystems. There is an overhead, both in machine usage and support, in using XRF. However, if you have an application that can only tolerate minimal outages, then you may wish to consider XRF.

The alternate system tracks the active system by reading the IMS system logs. The alternate checks these records, updating its own control blocks to keep its control blocks and shadow copy of the data base synchronized with the active IMS's system. In this way, the alternate IMS remains ready to take over at any time.

An IMS configuration could look like this:



In addition to tracking the log, the alternate is also constantly tracking the health of the active sys-

tem. This heartbeat detection is done in three ways:

The active IMS sends messages over the ISC link between the active and the alternate IMS subsystems.

The active IMS places a time stamp in the RDS.

The active IMS continues to add new records to the IMS system log.

The active IMS records certain failures on the system log, including

- IMS control region ABENDs,
- VTAM failures that lead to TPEND exits in IMS,
- and IRLM failures that lead to STATUS exits in IMS.

In addition, absence of IMS log updates can indicate

- System failures,
- System loops or wait states,
- or CEC failures.

A takeover can occur when there is:

- IMS ABEND
- A surveillance-detectable IMS failure
- A surveillance-detectable OS/390 failure, loop, or wait state
- A central processor complex (CPC) failure
- A VTAM failure that results in a TPEND exit
- An IRLM failure that results in a STATUS exit
- XRF terminal switching is limited to SNA terminals

End user recovery time is based upon the terminal class.

Class 1: When the active IMS establishes or terminates a terminal session with Class 1 terminals, the alternate IMS establishes or terminates a backup session. When the active IMS fails, the alternate IMS takes over the terminal session without losing control from the viewpoint of the LU by changing the mode of the pre-established backup session from BACKUP to ACTIVE. When the session is switched, the NCP sends the alternate IMS its view of the terminal's status. IMS compares this with its own record of status to decide what, recovery action to take, if any. When the REVERIFY operand is used with RACF, the user must sign on again, otherwise this is not needed. Class 1 terminals have the shortest recovery time, in the order of half a minute once the outage is detected. Examples of Class 1 terminals would include those SNA terminals with human end users.

Supported Class 1 terminals are those that:

- Use SNA protocol
- Are controlled by a VTAM and NCP that support XRF
- Are connected to a 37x5 Communication Controller
- Can be defined on the UNITYPE keyword on the TYPE macro in the IMS System gener-

ation as one of the following:

- SLUTYPE1, SLUTYPE2, SLUTYPE4, SLUTYPEP, FINANCE

Class 2: Class 2 terminals do not have backup session support on the alternate IMS, but are restarted automatically after takeover. The interface to the terminal, including recovery procedures after session reestablishment on the alternate IMS, is exactly the same as in the prior version of IMS, except for the automatic session re-establishment at system takeover.

When a Class 2 terminal session is established on the active IMS, the alternate IMS tracks the session initiation or termination using the log records. When the active IMS terminates abnormally, the alternate IMS tries to establish a new session with the network resources that were active before the failure.

Class 2 terminals receive good support from XRF. In fact, at takeover, IMS might be able to reestablish service on some of your Class 2 terminals before all of your Class 1 terminal sessions are switched. Using a BACKUP= parameter on system definition macros or ETO logon descriptors, you can set the priority that controls the order in which IMS reestablishes service to Class 2 terminals at takeover. Class 2 terminal recovery time depends upon how fast it takes VTAM to reestablish the network session and how quickly the terminals can then reenter the "Signon" command.

Terminals that qualify as Class 2 include:

- Multiple Systems Coupling (MSC) and ISC subsystems that communicate with IMS XRF through VTAM or bisynchronous lines
- Terminals on leased lines controlled by BTAM and connected to a 37x5 Communication Controller with multi-system line access (MSLA)
- Spool line groups on shared DASD, locally attached to both CPCs
- Non-SNA 3270 terminals controlled by VTAM
- Devices controlled by the Network Terminal Option (NTO) licensed program
- Locally attached devices
- Terminals that do not use the SNA protocol
- Terminals that qualify as Class 1 terminals, except:
  - The terminals are not controlled by a 37x5 network communication controller
  - The terminals are not controlled by an NCP or VTAM that supports XRF or defined BACKUP=(n,NO) on the system definition macro or ETO logon descriptor

Class 3: Class 3 terminals do not have backup session support on the alternate IMS. Their terminal sessions must be restarted manually on the new active IMS after takeover, either by the MTO or by user logon. Class 3 terminals have the longest recovery time as there is the most manual intervention. Terminals are required to enter the "Logon" command to reestablish a session, followed by the "Signon" to IMS. As with Class 2, recovery time for Class 3 terminals depend upon

how quickly VTAM can reestablish the network sessions. Examples of Class 3 terminals would include those managed by TCP/IP and intelligent workstations such as APPC (LU 6.2) type bank Automatic Teller Machines (ATMs).

Terminals that are eligible for Class 3 include

- All the terminals not eligible for Class 1 or 2 service
  - OTMA terminals
  - APPC connections
  - TCP/IP clients
- Class 1 and Class 2 terminals that have indicated no switching should be done. This is done by specifying BACKUP=NO on the system definition macro or ETO logon descriptor.

The principal drawbacks of XRF are:

- It will not, in itself, protect against network outages. You will have to plan for this separately.
- XRF does not support DB2 or VSAM files. However, if you are designing an application of this sort, an alternative would be to use IMS databases, particularly the Data Entry Database (DEDB). The DEDB has provisions for performing most database maintenance with the databases remaining available. It will also automatically maintain multiple copies of the data sets containing the data to guard against media failure.
- Some maintenance to the IMS software may need to be applied to both the active and standby IMS systems at the same time.
- XRF requires an investment in the processor capacity used by the alternate IMS when tracking. This is in the order of 12% of the active IMS's requirements.
- IMS failures will take a longer time to recover as the failing IMS control region must shut down completely before the terminals could be switched.

So, while XRF can prevent most unplanned and planned outages, it cannot keep the IMS system available indefinitely. You will eventually have to have plan outages for software maintenance and upgrades, and some changes to the IMS configuration.

Remote Site Recovery (RSR)

When your computing system is disabled, you need to recover quickly and ensure that your database information is accurate. Interruption of computer service can be either planned or unplanned. When interruption on the primary computing system occurs, you need to resume online operations with minimal delay and minimal data loss.

The remote site recovery feature (RSR) allows quick recovery from an interruption of computer services at an active (primary) site. RSR supports recovery of IMS DB full-function databases, IMS DB Fast Path DEDBs, IMS TM message queues, and the IMS TM telecommunications network.

IMS database and online transaction information is continuously transmitted to a tracking (remote, or secondary) site. The tracking site is continually ready to take over the work of the active site in the event of service interruption at the active site.

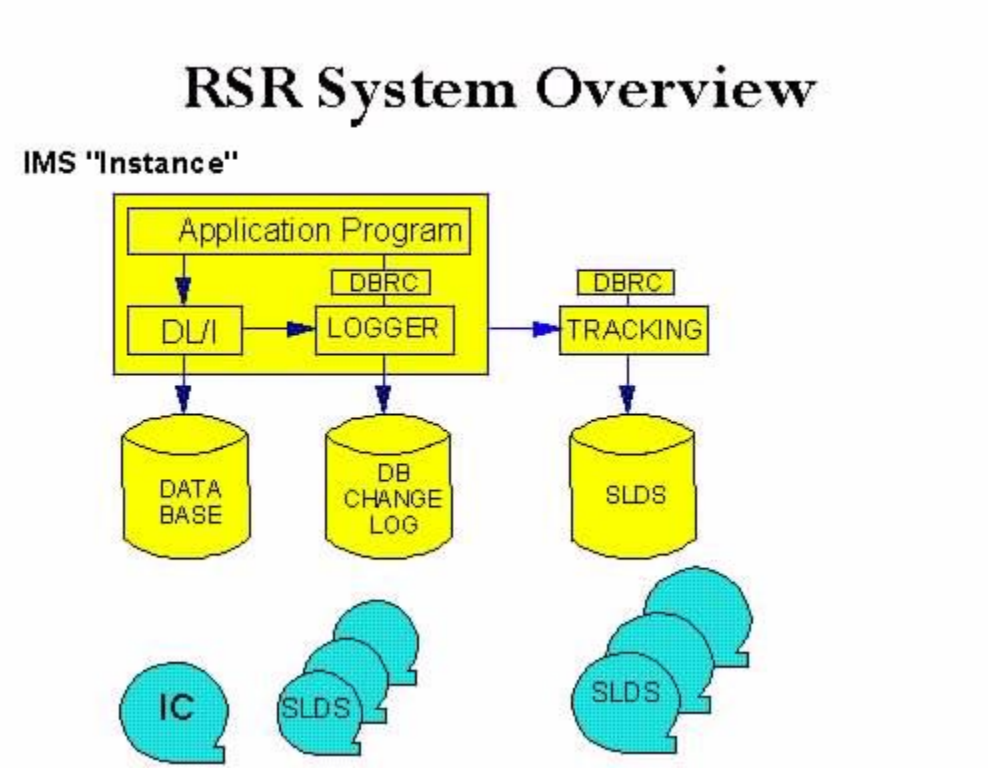
Because IMS needs to be able to resume online operations at a remote tracking site in the event of an extended outage (either planned or unplanned) at the active site, RSR does the following:

- Provides a remote copy of the necessary IMS DB and IMS TM log records for database and message queue recovery at the tracking site.
- Reduces the time required to resume computer service to approximately an hour.
- Lets you select and filter out the log records that are not needed to support the defined critical environment.
- Continues to operate when the active or tracking sites or the RSR transmission facility become temporarily unavailable, and provides a way to resynchronize the sites as soon as possible.
- Provides transaction consistency between the active and tracking sites.
- Supports IMS DB and IMS DBCTL. Supports full-function databases and Fast Path DEDBs.
- Supports both online IMS DB and DBCTL workloads, as well as batch workloads, at the active site.
- Supports data sharing at the active site.
- Coexists with XRF
- Recognizes that DBRC is operating at the active site and, separately, at the tracking site.
- Supports standard ACF/VTAM communication protocols, so that new technology is not required for data transmission.

Note that RSR does not support shared queue. Only IMS/TM local message queues are supported.

RSR is a separately priced component available with IMS. It provides similar facilities to XRF,

but with some differences. RSR can track details of IMS full function databases, Fast path DEDB's, IMS/TM message queues and the current IMS/TM telecommunication network on an alternate machine. This machine is connected the machine with the active systems on by a network connection using the VTAM APPC protocol. The VTAM connection is between separate transport manager subsystems (TMS) on the active and tracking machines.



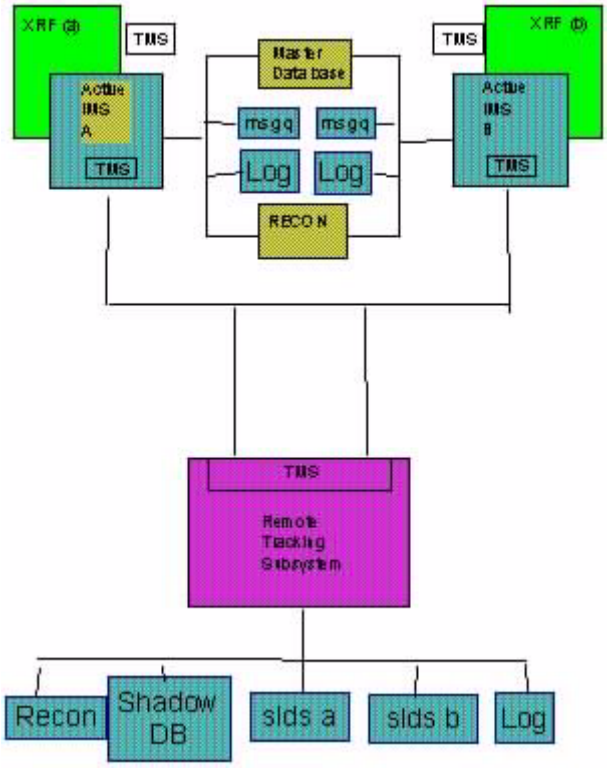
The log router manages the tracking end of communication between the active and tracking sites. In the tracking subsystem it receives data from the active subsystems, stores the log data in tracking log data sets, and routes log records to individual tracking subcomponents, called trackers. The log router is unique to tracking subsystems; it is not found in active subsystems.

From the tracking data set, IMS on the tracking system processes the data and logs it using normal IMS logging. Depending on what level of tracking has been requested, the IMS region may also apply the updates to the shadow IMS databases at the tracker site. These updates are done by the DL/I database tracker.

If there are any interruptions to the network connection, RSR will note the gaps in the logging and perform catch up processing when the link is re-established. The IMS system on the tracking machine normally can only process input from the TMS. It only becomes a fully functioning system if it has to take over.

With RSR, not all databases need to be tracked. You define the databases that are to be tracked by specifying this when you define them to DBRC.

IMS RSR and XRC can work together, as shown in the following diagram:



Here, IMS logs are sent from the active subsystems as they are created. The active site could have two active subsystems, A and B, which are configured with XRF and data sharing. One tracking subsystem is used for storing log data received from the active subsystems.

If one of the active site subsystems is disrupted, you could use its XRF alternate subsystem to take over its processing. If both subsystems at the active site are disrupted, this setup is designed on the assumption that you are willing to recover both subsystems, A and B, on the tracking subsystem.

When updates are made to an active subsystem, this subsystem writes log data to an online log data set (OLDS), or, for batch, to a system log data set (SLDS). At the same time as the disk write, the active subsystem also sends this log data to the tracking subsystem. The tracking subsystem stores this data in SLDS.

At the tracking site, DBRC maintains log and database recovery information for the tracking subsystem. The tracker's RECON data set is not a mirror of the RECON data set at the active site. DBRC notes the log data received from both active subsystems and maintains database recovery information for the tracker in the RECON data set of the tracking subsystem.

RSR does not support active application processing against the shadow databases until a remote takeover of active processing occurs. At that point, the tracking site becomes the new active site and you can restart the active subsystems at that site. If the active subsystem participates in data sharing, you need to switch all sharing subsystems (the entire service group) to the tracking site on an RSR takeover.

The following table gives a comparison of the features of XRF and RSR.

XRF	**RSR**
Uses same physical log data sets and database data sets for active and tracking system, thus yielding a single point of failure	Uses completely separate log data sets and database data sets, so a site-wide physical problem can be overcome
Supports DB/DC and DCCTL configurations	Supports DB/DC, DBCTL, DCCTL, and batch DL/I configurations
Performs takeovers on a subsystem by subsystem basis	Remote takeover includes all subsystems that share data bases
No exposure to marooned data	Unplanned takeovers have exposures to marooned data.
Active and tracking system must be within channel attach distance of each other	Active and tracking systems are connected by network, only limit on separation is network response
Active and tracking systems must use IMS/TM	Active systems can be any system updating IMS resources DB/DB, TM only, DB only, or batch. The IMS tracker must be DB/DC.
One-to-one relationship between active and tracking system.	One Tracker tracks all members of an data sharing group
All committed updates recorded on tracking system	Possible for gap in data at tracking system after unplanned takeover
Switching to/from alternative comparatively simple.	Planned takeovers, switching back to the original site is more complex than XRF After unplanned takeovers, switching back is very difficult and requires a planned takeover
Switches over to alternate in order of one minute.	Switch to alternate can take an hour or more.

Requires some subsystems management	Requires more subsystems management due to need to replicate descriptors, programs, and other resources at the second site

Both features rely on having another IMS subsystem, situated on another OS/390 system, that tracks the update activity of the primary IMS subsystem (only one for XRF, one or more for RSR) to provide a backup.

RSR has these recovery features:

- It provides a remote copy of the necessary IMS DB and IMS TM log records for database and message queue recovery at the tracking site.
- It supports these DL/I database access methods: HDAM, HIDAM, HISAM, and SHISAM, and supports fast path DEDBs.
- It supports IMS DB, IMS DBCTL, and batch workloads.
- It maintains remote copies of full-function databases and fast path DEDBs.
- It recognizes that IMS database recovery control (DBRC) is operating at the active site and, separately, at the tracking site.
- It supports data sharing at the active site.
- It coexists with the IMS extended recovery facility (XRF).
- It lets you filter out log records that are not needed to support the defined critical environment. An added benefit of this is reduced line traffic.
- It continues to operate when the active or tracking sites, or the RSR transmission facility, become temporarily unavailable, and provides a way to resynchronize the sites.
- It provides transaction consistency between the active and tracking sites.
- It supports standard VTAM communication protocols, so new technology is not required for data transmission.

To summarize:

- XRF is suitable for situations where you have a single IMS DB/DC system that requires very high system availability (greater that 99.5%). However the second OS/390 must be channel attached to the OS/390 system the first IMS is running on.

- RSR is suitable for situations where you have one or more IMS applications, which may run in a number of address spaces, and where you wish to minimize data loss in a failure situation, but can tolerate outages of around an hour. RSR uses network connections between the two OS/390 systems, so there are no restrictions on the distance separating them.

Parallel Sysplex Data Sharing

The Parallel Sysplex cluster contains multisystem data sharing technology, allowing direct, concurrent read/write access to shared data from all processing nodes in the parallel sysplex configuration. This is done without sacrificing performance or data integrity. Through this technology, the power of multiple OS/390 and z/OS systems can be harnessed to work in concert on common workloads to improve levels of price/performance, scalable growth, and continuous availability.

Just as work can be dynamically distributed across the individual processors within a single S/390 SMP server, so too can work be dynamically directed to and balanced between any node in a Parallel Sysplex cluster having available capacity. This avoids the need to partition data or applications among individual nodes in the cluster or to replicate databases across multiple servers.

Through data sharing and dynamic workload balancing, continuous availability and continuous operations characteristics are significantly improved for the clustered system, as servers can be dynamically removed or added to the cluster in a non-disruptive manner. If the processing demands grow and exceed the capacity of the existing server systems, it is possible to add an additional system to the Parallel Sysplex cluster and grow the application workload transparently. This can be accomplished without splitting applications or databases across multiple servers. Together with Communication Server support, the entire Parallel Sysplex cluster can be viewed as a single logical resource to end users and business applications.

IMS Exploitation of Parallel Sysplex: Since IMS V5.1 when IMS initially supported n-way data sharing, there has been continuous enhancements to the Parallel Sysplex support. A summary of the major support items includes the following:

Release	Support Item	Value
IMS V5.1	•Full Function Data Sharing •(basic) DEDB Data Sharing	Availability and Capacity

IMS V6.1	•Shared Message Queues •VTAM Generic Resources •Fast Data Base Recovery •DEDB / VSO Data Sharing •DEDB / SDEP Data Sharing •OSAM Caching •Command Reference Character Enhancements •Asynchronous APPC/OTMA support	Dynamic Workload Balancing Single system image to end users Faster backouts of retained locks Usability and functionality enhancements Performance Operational support System Management Functional enhancements
IMS V6.1 Functional APARs	•Generic IMSID •VTAM G/R Enhancements	Usability and functionality enhancements
IMS V7.1	•Rapid Network Reconnect •Online Recovery Services •IMS Monitor Enhancements	Faster recovery for terminals Faster recovery for data bases System management
Follow-on releases	•More enhancements coming	Faster recovery for terminals and data bases

While much can (and have) been written about each of these items, of partitular note to this document is Fast Data Base Recovery (FDBR), and also IMS's interaction with OS/390's Automatic Restart Manager (ARM).

FDBR, although incompatible with XRF, uses many of XRF's techniques to track a production IMS subsystem's health. This is done through monitoring the logs (as XRF does) and by XCF monitoring for a heartbeat (instead of using the Availability Monitor used by XRF). In the event of the active system failing, the FDBR code will dynamically back out in-flight full function database updates, invoke DEDB redo processing, and purge retained locks from IRLM. The FDBR region will then end. This speeds up the IMS restart process, speeds up the backing out of in-doubt locks, and improves availability to the shared data bases.

Using a different philosophy, OS/390 (and z/OS) increase availability by providing ARM, a policy based technique to automate the restart of subsystems in the event of a failure. When a failure is detected, ARM invokes the restart of registered subsystems in the correct order, either in place (subsystem failure) or on a surviving partition in the JESPlex / Sysplex. With the Workload Manager Goal Mode active, it would be restarted on the partition with the most available capacity to handle the work. The subsystems would then use their normal emergency restart process to recover.

A comparison of the recovery options is described in the following table:

Recovery Feature	XRF	FDBR	ARM
Data Base recovery	Begins its database recovery processing sooner than an IMS restart initiated by ARM. XRF is tracking the active IMS by reading its log. This eliminates most of the log reading that is required in an ARM-initiated restart.	Begins its backout processing sooner than an IMS restart initiated by ARM. FDBR is tracking the active IMS by reading its log. This eliminates most of the log reading that is required in an ARM-initiated restart	ARM-initiated restarts begin the database recovery processes later than FDBR or XRF would
in-doubt threads	XRF can resolve in-doubt threads with CICS and DB2. This requires that the CICS and DB2 subsystems be restarted on the MVS where the XRF alternate executes.	FDBR does not resolve in-doubt threads with CICS and DB2	ARM-initiated restarts can resolve in-doubt threads with CICS and DB2 ARM automatically moves IMS, CICS, and DB2 in a group to the same MVS.
Message Queues	XRF handles the recovery of IMS message queues.	Does not recover IMS message queues	Restarts recover the IMS message queues.
MSDBs	Recovers MSDBs	Does not recover MSDBs	ARM-initiated restarts recover MSDBs.
New Work	XRF accepts new work on the new active system.	FDBR does not restart failed IMS	Restarts accept new work on the new active system.
DBRC Authorizations	The alternate assumes the authorizations of the failed active	Does not release DBRC database authorizations	Restarts releases DBRC database authorizations during emergency restart processing.

Performance	XRF requires more CPU and storage than FDBR while in tracking phase	FDBR uses less CPU and storage than XRF during the tracking phase.	ARM requires no resources for tracking.
Operations	Takeover on another CPU requires operator intervention to indicate that IO prevention is complete.	Requires no operator intervention when IMS runs as a started task. XCF monitoring is used to inform FDBR that IO prevention is complete.	Does not require operator intervention

Geographically Dispersed Parallel Sysplex (GDPS)

Over the past 15 years much technology has been made available to create a High Availability environment in a single site. Examples include:

- Concurrent hardware Install
- Concurrent Maintenance
- Sysplex (not only workload balancing but also High Availability)
- Fault tolerant equipment
  RAID is an example and has been accepted such that non-RAID equipment is no longer installed.
- N, N+1 coexistence which allows for the upgrading of software levels without ever having to take applications down (OS/390 provides even N, N+3 coexistence).

GDPS extends these technologies to provide Disaster Recovery capability.

GDPS provides switching capability from one site to another site, for planned and unplanned events. This covers the need to free up a site for maintenance purposes, as well as the type of unforeseen event that closes down a site unexpectedly. In addition, it solves the problem of routine PPRC and XRC configuration management by providing a high level, functional interface that handles virtually all of the technical details and thus protects the user from making accidental errors.

GDPS was originally built on Sysplex and PPRC technologies, bringing together two advanced technologies to create an integrated D/R readiness solution. Now GDPS also supports IBM's Extended Remote Copy technology, providing the same advanced D/R solution over unlimited distance. Also included with the GDPS solution are the automation products Tivoli Netview for OS/390 and System Automation for OS/390. Using automation makes good sense for this type of solution because of the need to interface to many software and system components: OS/390, RACF, JES2, Console log, ERP, Disk Subsystems, etc.

GDPS provides *near*-continuous availability: in the process of making the switch from one site to another it will be necessary to restart all applications.

The diagram below shows a high-level view of a GDPS configuration. Key ingredients are

- A Parallel Sysplex with, when PPRC is used, components in two sites.
- Data replication (synchronously through PPRC or asynchronously through XRC)
- The configuration as a whole is managed through GDPS automation.

Although GDPS started as a tool to manage unplanned outages, it was soon obvious that its framework and code could be used equally well to manage planned exception conditions such as

the recycle of systems for software maintenance or the shutdown of a site for planned site maintenance.

The main focus of the automation is to make sure that, whatever happens in the primary site (site 1 in the diagram), the image of all data in the surviving location (site 2) is time consistent. **Time consistent** means: the secondary disks contain all updates until a specific point in time, without anything missing, and no updates beyond that point. GDPS will actually go one step further: it will also prevent logical contamination of secondary data so that the plethora of errors that should be expected during a rolling disaster will not be copied forward to the surviving site, either fully or in part. Such errors would leave the transaction and data base managers, and potentially also the operations staff, in a confused state, preventing a fast and efficient application restart.

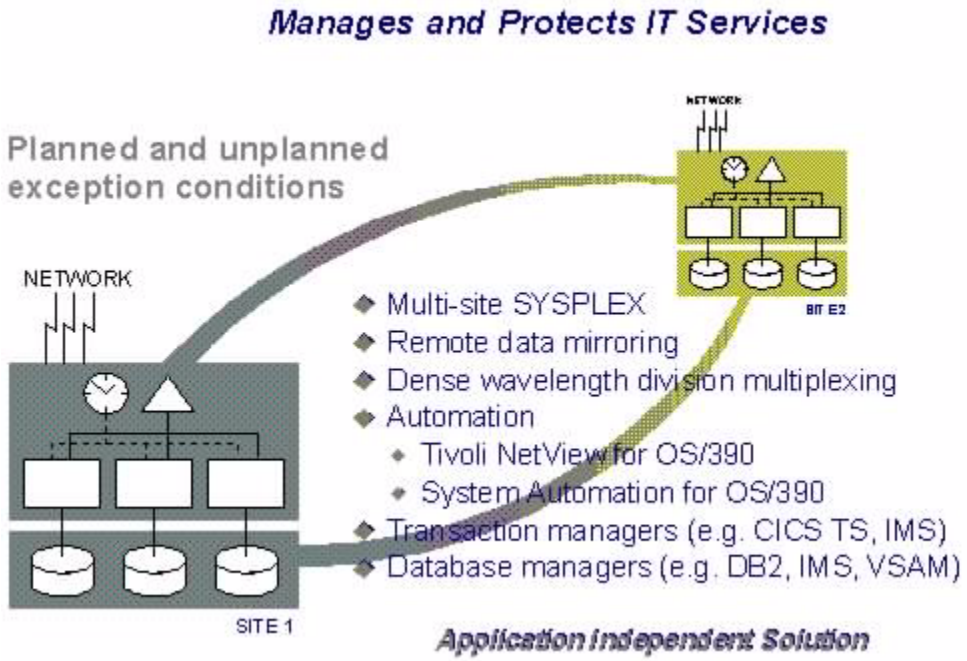
The fact that the secondary data image is time consistent and without logical contamination means that applications can be emergency-restarted in the secondary location, without having to go through a lengthy and time-consuming data recovery process. This should allow an installation to be up and running within an hour, even when the primary site has been rendered totally unusable.

This qualifies the relation between GDPS and the Transaction/Database Managers (or any other type of application environment): GDPS guarantees time consistent data in the surviving site, leaving applications in the same situation as if there had been a sudden and complete power failure. No error analysis is needed (other than what is required to make the decision to move to the alternate data processing location) and applications can be emergency-restarted. This solution is therefore application independent.

Perhaps the most unique GDPS capability is that it will always leave the secondary disks in a time consistent state, no matter the nature of the problem in the primary site, and irrespective of the number of primary/secondary Control Units in the PPRC configuration.

Lastly: GDPS does not take away any of the Sysplex High Availability capabilities. With a GDPS the installation still takes the full benefit of a Sysplex, while inter site High Availability is managed through GDPS.

The diagram shows a GDPS configuration with Sysplex components in each site. Note that this is no longer a must, now that GDPS also supports XRC as a remote copy technology.



Usability comparison of RSR and GDPS

RSR	GDPS
Uses completely separate log data sets and database data sets, so a site-wide physical problem can be overcome	Same.
IMS-only solution. Supports DB/DC, DBCTL, DCCTL, and batch DL/I configurations RSR does not support shared queues Only IMS/TM local message queues are supported.	Application independent. Supports any environment, any data base, any configuration.
Remote takeover includes all subsystems that share data bases	Remote takeover includes all subsystems that you wish to recover
Unplanned takeovers have exposures to marooned data.	•Key to GDPS is full point-in-time Data Consistancy. (all GDPS solutions) •No marooned data (with PPRC) •Full data integrity. This is true even when using multiple data bases such as IMS together with DB2.

Active and tracking systems are connected by network, only limit on separation is network response	Tracking systems are <40 km apart (PPRC) or unlimited (XRC)
Active systems can be any system updating IMS resources DB/DB, TM only, DB only, or batch. The IMS tracker must be DB/DC.	Application independent
One tracking system tracks many active systems	GDPS tracker tracks the remote site with any number of systems
Inconsistant data with some data loss after unplanned takeover.	Full point in time recovery. Some data loss.
Planned takeovers, switching back to the original site is more complex than XRF After unplanned takeovers, switching back is very difficult and requires a planned takeover	Planned or unplanned takeovers are automated for easy management. Recovery back to original site is also fully automated.
Switch to alternate can take an hour or more.	Takeover in under an hour, time needed to re-IPL other system.
Requires more subsystems management due to need to replicate descriptors, programs, and other resources at the second site	As with RSR, GDPS requires a remote disaster recovery site. Many of the system parameters can be "Cloned".

Functional comparison of XRF and Parallel Sysplex

Both XRF and Parallel Sysplex address availability, but go about it in a different way.

Terminal Recovery

The power of XRF lies in the fact that Class 1 terminals can be switched from the active IMS to the alternate very quickly. Measurements done in 1987 with 2500 Class 1 terminals on a 3084-QX (approximately 22 "MIPS" on a 4-way processor) showed all terminals switched in under 35 seconds after the failure, with the first transaction resuming in uner 40 seconds after the failure. A takeover is invisible to a Class 1 terminal. The takeover appears as a period of time where there is long response time. Class 2 terminals have XRF switch the network and the terminal have to just re-signon to IMS. Class 3 terminals have no XRF support and they need to re-logon to the network, then resignon to IMS. XRF can not, however, recover from many kinds of failures. Anecdotal customer experience is that XRF provides a 75% chance of recovery.

With IMS V7 and Multi-Node Persistent Session (MNPS) support, called IMS Rapid Network Reconnect, or RNR, SNA terminals have their sessions reestablished on the alternate site. As with XRF, this saves network costs. When IMS with the transaction manager fails, the failing IMS needs to completely stop, then the IMS can be restarted again either in place or on another LPAR. This is similar to what XRF does as XRF also has to wait until IMS control region shutdown before switching sessions. There are two differences, though:

RNR support currently requires the failing IMS be restarted somewhere in the Parallel Sysplex. This would elongate recovery.

Assuming a 2-way data sharing environment, only half of these users would be affected. The other half would continue to run on the data sharing partner. This would make recovery a non-issue for a percentage of the users.

With Shared Message Queues, there would be no lost messages, even if the user logs on to a different IMS.

VTAM Generic Resources support is another option that can be used to recover the users. It is compatible with RNR, so both can be optionally done together. VTAM Generic Resources presents a single IMS node name to the end users. Instead of having to issue a "LOGON IMS1", or "LOGON IMS2", or "LOGON IMS3", the user just issues "LOGON IMS", for example. VTAM knows which IMSs are active and directs the logon to the IMS that has the least number of sessions at that time. This balances the network load. With Workload Manager Goal Mode, WLM recommendations are honored, based upon available capacity. If a user is on an IMS that failed, the user can either wait for RNR to reconnect the terminal, or just re-enter the "LOGON IMS" request again. In this case, a new session will then be established with one of the surviving IMSs. This has the potential to be much faster than RNR, although there is additional CPU cost in the network manager to reestablish the session. VTAM APPN is required for Generic Resource sup-

port. VTAM Generic Resources is incompatible with XRF and Remote Site Recovery.

IMS does not have direct TCP/IP support with in it. If a user wants to use TCP/IP support for its network, the user either uses IMS Connect (the TCP/IP gateway) or the user writes their own TCP/IP socket support (chose not to use T/N3270) and uses the OTMA callable interface to get to IMS transactions. Be that as it may, the percentage of TCP/IP in a typical network is growing. Support for this environment within the Parallel Sysplex is growing as well. Communication Server support such as "VIPA Takeover" establishes the session to a backup IP address. If IMS comes up with a predefined IP address, then when IMS is restarted either in place or on a different LPAR, its clients can reconnect using the same address. Alternatively, with DNS support, the clients can use the same DNS name immediately. This would get resolved to one of the surviving IMSs with currently the most available capacity. Although the clients would need to reconnect to the network, this would only apply to a percentage of the users, depending upon how many IMSs are in the data sharing group. The other clients would continue running unaffected. Current TCP/IP support, then, is similar to what is available with VTAM Generic Resources.

The future direction of IBM is have support similar to XRF's "Class-1" for TCP/IP connections.

Supported Data Bases

XRF is an **IMS-only** solution. If an application has any DB2, VSAM files, or other data base managers updating data, then XRF by itself **WILL NOT WORK**. (XRF supports VSAM data bases managed by IMS, but not direct VSAM calls such as those done by a CICS FOR). Other recovery techniques must be put into place to support non-IMS data. This increases the complexity of managing the data and the time to recover the data.

Parallel Sysplex technology is exploited by many data base managers. This includes:

- IBM products
  - IMS
  - DB2
  - VSAM
- Non-IBM products
  - Datacom DB
  - ADAbas
  - Oracle for MVS

All of these data bases can together be used by a single application.

Again, if there are any non-IMS data base calls by an application, then XRF by itself will not work without extensive recovery for the other data bases.

In addition to data sharing, the Parallel Sysplex enables a host of other technologies. These "Resource Sharing" exploitations were designed to simplify the process of managing multiple LPARs, each with resources, reduce the total number of resources required, improve performance,

or combinations of the above. One recent example is "Intelligent Resource Director, or IRD. This function can be used to manage the resources used by multiple LPARs within a single (zSeries) server. IRD includes the functions of:

- LPAR CPU Management: Dynamically changes the LPAR weights and number of logical CPUs so the most important work can meet its goals
- Dynamic Channel Path Management: Dynamically adjusts the bandwidth distribution from the z/OS LPARs to each of the control units so the business critical work gets access to the data.
- Channel Subsystem Priority Queueing: Gives the most important work priority access to the channel subsystem in the server so it can meet its business goals.

A more full explanation of IRD and other exploitations can be found in various white papers off of the Parallel Sysplex home page at: www.ibm.com/s390/pso/

Steady-state Operations Performance Implications

XRF relies on a hot-standby IMS system that is tracking the production workload. Due to heart-beat processing, the cost on the Active system was measured in 1987 at 3% ITR impact (Internal Transaction Rate, or Transactions per CPU-Second). The CPU cost on the Alternate IMS was measured at 12% of the production's system's CPU. This includes IMS and all other subsystems needed to support this workload such as the network server, OS/390 processing, etc. On the average, then, there would be a total impact of approx. 7.5% ITR cost on the two systems.

The cost of migrating to a two-way IMS data sharing environment from a single IMS system was measured at 3% for migrating to a Parallel Sysplex. This includes overhead for managing a JES2 MAS, shared DASD, and sysplex management. If some of these were already in place, then this 3% cost would be lower. In addition, with current technology there is a data sharing cost of approximately 9% of the data sharing workload. If the data sharing workload takes up, for example, 50% of the LPAR's capacity while running at 80% busy, then the final total CPU % increase would be (.50 * .80) * .09 = .4 * .09 = .036, or a 3.6% increase bringing the utilization to 83.6%. On the average, then, there would be a total impact of 12%, assuming the 100% of the production processor was running the data sharing workload. Since this is never the case (TSO, monitors, other applications are also running), customer experiences shows the cost to be under 10%. This is comparable to XRF's cost.

Data Sharing has other performance features that are not available with XRF. Since production work can be dynamically routed to multiple systems, one can split the workload, resulting in lower peak CPU utilizations and eliminating any possible CPU bottlenecks that may exist. As well as being the only way to manage growth, a performance benefit of lower CPU utilization is less queueing for CPU and other resources, reducing batch elapsed times and transaction response times.

An IMS extended recovery facility places heavy demands on the logging process. The alternative

IMS system reads all of the log records written by the active system; therefore, the I/O rate is effectively doubled. If the OLDS DASD response is continually delayed, the active system transaction response is delayed and the alternative system falls behind in its surveillance. This in turn causes an extended takeover time. Generally, the maximum logging rate of the XRF environment is one-half that of a non-XRF environment. The maximum logging rate for a single system is determined by the data rate available to the DASD device. Cached control units should mitigate this problem.

Growth Management

As mentioned above, a Parallel Sysplex can split workloads to multiple LPARs across CECs. Many workloads are growing faster than processor capacity. This is due to new applications, exploiting new technologies such as e-business, consolidation of computer centers, or corporate mergers. Once a single subsystem's workload exceeds the capacity of the server, the ONLY solution to managing growth is the Parallel Sysplex data sharing. XRF can not help with this environment. As well as being the premier clustering solution for functionality and usability, Parallel Sysplex is also the most scalable, with only a 0.5% cost for each additional node beyond the 2nd. This has been proven to scale all the way to 32 images.

Systems Management

XRF and Parallel Sysplex share some of the same system management requirements: One needs to manage a second IMS subsystem, both acessing the same data base, across multiple servers. Support for this, however, vary greatly.

Data sharing technology enables being able to manage the multiple IMSs and their operating system images as a single system. This is done with the resource sharing technology of sharing Consoles, Tapes, RACF, system logs, GRS resources, and signaling paths that are used by VTAM, XCF, and others.

Sysplex services include policy-based information to help automate recovery of subsystems (ARM), managing LPAR failures (SFM), and managing and tuning a mixed-workload environment (WLM). Recent Workload Manager enhancements that are built upon the Parallel Sysplex technology include dynamic management of LPAR CPU management, Channel Subsystem Priority Queueing, and Dynamic Channel Path Management. None of these features are available with XRF in a non-sysplex environment.

In addition, System Automation for OS/390 has many pre-coded routines to manage the applications that run on multiple systems. This removes operator and human error into making what could be a recoverable situation now unrecoverable.

The industry's premier disaster recovery product, GDPS, is based upon the sysplex technology.

IBM Single Site Recovery Support and Direction

XRF was introduced by IMS in 1987 with IMS/VS V2.1. Since then, there has been no new function or features to enhance XRF's capability either by IMS, VTAM (Communication Server), NCP, or OS/390. The customer base, while dedicated to the platform, is not very large.

Parallel Sysplex n-way Data Sharing support was introduced by IMS in 1994 with IMS/ESA V5.1. Since then, support has been added with IMS V6, IMS V7, new functionality between releases, and continued plans for further support to enhance usability and functionality in the future. By 4Q 1999 there were a total of over 1500 Parallel Sysplex customers with over 600 doing application data sharing, many doing IMS data sharing. These numbers are growing.

Since its introduction, **EVERY MVS, OS/390, or z/OS release** has had support to make Parallel Sysplex easier to set up, manage, perform, or have had new function added. This is impressing considering OS/390 releases come every 6 months! This trend is not stopping. There are planned enhancements in the hardware, z/OS, communication server, the data base managers, and other subsystems including MQSeries, to further this support.

At IMS user group conferences, one would need to go hunting to find XRF sites. On the other hand, to find a data sharing site just ask the person standing next to you. Chances are that person either has data sharing in production or soon will. It is easy to see that Parallel Sysplex coupling is the strategic direction for IBM.

The future direction is for IMS is to provide recovery equal to or better than XRF's in terms of speed and impact to the users. This will be in IMS follow-on releases.

A comparison of the usability and functionality enabled by XRF and Parallel Sysplex is shown in the following table:

Category	XRF	Parallel Sysplex
Single Point of Failure	Uses same physical log data sets and data-base data sets for active and tracking sys-tem, thus yielding a single point of failure	Each IMS has their own logs, although data base is shared
IMS Configurations	Supports DB/DC and DCCTL configura-tions. No support for non-IMS data bases	Supports all IMS configurations Supports DB2, VSAM, and many ISV DBMs
Data Bases Supported	•IMS FF and DEDBs	•IMS FF and DEDBs, •DB2 •VSAM •Datacom DB, Adabase, Oracle
Takeover restrictions	Performs takeovers on a subsystem by sub-system basis	No takeover needed as both systems ar "Active IMS"

Data Integrity exposures	No exposure to marooned data	No exposure to marooned data. Shared Message Queue logs all messages
Distance	Active and tracking system must be within channel attach distance of each other	Up to 32 LPARs within 40 km distanc
Transaction Manager restrictions	Active and tracking systems must use IMS/TM or CICS (CICS/XRF support)	No restrictions.
Capacity Considerations	Single LPAR/CEC solution	Supports up to 32 CECs
Number of subsystems	One-to-one relationship between active and tracking system.	Up to 255 IMS subsystems on 32 LPA
Installation	Intensive planning and considerations required for •Network setup •NCP •VTAM Uservars •IMS	Much easier to set up •msys for Setup and other wizards and to set up Parallel Sysplex environm •Relatively few, well documented procedures •Many redbooks and support available
Ease of Use	Operationally, many chances for human error	Recovery can be easily fully automated Products exist to do this.
Recovery considerations.	Recovers from many IMS and H/W failures	Recovers from **all** IMS and H/W failur Also recovers from some network failu
User Impact of Failure	Impact depends upon terminal class. •Class 1 - No action needed. •Class 2 - Re-signon to IMS •Class 3 - Re-logon to network followed by re-signon to IMS.	•Users on surviving IMSs are unaffecte •SNA Terminals switched (RNR) when ing IMS is restarted. Users need to signon to IMS •TCP/IP users need to re-logon to netw and re-signon to IMS •All users can use same Applid / IP Ad as before
Time for Recovery from an IMS failure	•IMS CTL region must fully come down before terminals can switch. Following this, •Class 1 terminals switches over to alternate in order of one minute •Class 2 and 3 terminals switches take a few minutes	•Users on surviving IMSs are unaffecte •IMS CTL region must fully come dow then IMS restarted before RNR ree lishes sessions •Any user (TCP/IP or SNA) can break sion and re-logon to surviving IMS
System Management	Requires additional management for multiple IMSs on multiple LPARs Each system resource managed separately	Many tools, functions, and automation ucts exist to simplify system managem Each LPAR appears as a Single System Image to •operations •system programmers •end users

Automation considerations	Care must be taken to avoid false takeovers. Many customized automation procedures must be hand	Existing products can be used to succe fully fully automate system and applica management (SA for OS/390, GDPS)
Application Considerations	•Conversational transaction •may have undelivered messaged •Many common connections found today, including TCP/IP, APPC and OTMA are not managed by XRF (Class 3)	•Conversational transactions supported Shared Message Queues •Asynchronous APPC/OTMA support
Network Considerations	No support for HPR or APPN	Full support for HPR and APPN netwo
Other Benefits	None	Resource Sharing exploiters •Automatic Tape Switching •GRS Star •XCF (VTAM links) •Consoles •RACF Data sets •Catalogs •System Logs •JES2 Checkpoint •MQ Series shared queues •Intelligent Resource Director •LPAR CPU Management •Dynamic Channel Path Management •Channel Subsystem Priority Queueing

Functional comparison of RSR and GDPS

Share '92 users defined the following disaster recovery tiers. It is based upon the type of recovery being performed, data loss, and time to recover. The percentage of each option used is on the right-most column.

As one moves down the table, the cost of each solution increases. Although almost all those surveyed have a D/R plan, is is also true that almost all the customers surveyed would see up to 24 hours of lost data, while taking about 24 hours to bring the D/R site on-line. For many industries, this is acceptable. For others, such as Finance and Banking, much tighter recovery is required to protect the business. Two solutions are IMS's Remote Site Recovery (RSR) and Geographically Dispersed Parallel Sysplex (GDPS).

IMS Remote Site Recovery (RSR) and Geographically Dispersed Parallel Sysplex (GDPS) share many characteristics.

- Both are remote site disaster recovery solutions
- Both require maintaining a "shadow" data base (optional with RSR)
- Both depend upon the transmission of data to reflect updates, keeping the recovery time down.

Here, the designs diverge. RSR relies upon the transmission of log data to the tracker site asynchronously through VTAM (CS/390) communications. The IMS tracker reads these changes and applies them to the shadow data base if the shadow data base exists. GDPS is based upon either PPRC or XRC remote copy technology together with automation, sysplex, and GDPS specific code. This combines to guarantee time-consistant data for all PPRC, and with XRC for almost all

circumstances. In addition, one can have no data lost with GDPS/PPRC. No other disaster recovery solution on the planet can do this. None.

If a takeover occurs, the CPU cycles required by the D/R site will go up. IBM has an option called Capacity BackUp, or CBU. This enables one to configure a relatively small backup site and dynamically grow this to the desired capacity with a dynamic microcode change. GDPS can automate this procedure, removing human intervention from the picture and bringing the capacity online within seconds. This eliminates any CPU constraints in the D/R environment while keeping H/W and S/W charges down during steady-state operations. The RSR solution can also take advantage of CBU, but depending upon the CBU contract, it can take up to 2 hours for the capacity to become available.

Usability-wise, GDPS is fully automated. If a disaster occurs, the operator can respond to one WTOR and then sit back and watch everything take place. This includes

- Removing systems from Parallel Sysplex
- Perform Disk reconfiguration
- Perform CBU activation
- Perform CF reconfiguration
- Perform Couple Data Set reconfiguration
- Acquire processing resources and IPL systems into Parallel Sysplex
- Restart applications

A planned takeover to bring the configuration back to normal operations is just as easy. Both have been measured by multiple customers to complete in under one hour.

Functional Comparison

Functionally, there is no comparison. GDPS can do everything that RSR can do and more. They both are used to recovery from a server or site failure, but GDPS can do this recovery and a return to normal operations with much less skill requirements, much easier, entails much less planning, is easier, and on top of that can have fully time-consistant data and, with the PPRC option, zero data loss.

Supported Data Bases

RSR is an IMS-only disaster recovery solution. As with XRF, if there are any non-IMS data base calls, then further support is needed to recovery this data. GDPS is data base independent. It does not care what data bases are being used.

IBM Disaster Recovery Support and Direction

Since RSR's introduction, there has been a slow migration to this solution. There are still only a handful of RSR sites in production. IMS has added minimal enhancements to RSR since its release.

GDPS was introduced in 1998. While there are also a relatively small number of GDPS customers, the numbers are rapidly growing. In addition, there has been continued enhancements to the product. They include:

- Initial PPRC release
- Support for "Level 3 PPRC" to avoid resynchronizing data
- Automatic invocation of CBU
- XRC support
- Many further new functions planned for usability, functionality, and recovery time reductions.

GDPS is IBM's strategic solution for remote site disaster recovery.

Further information on GDPS can be found off of the Parallel Sysplex web site at:
ibm.com/s390/pso/

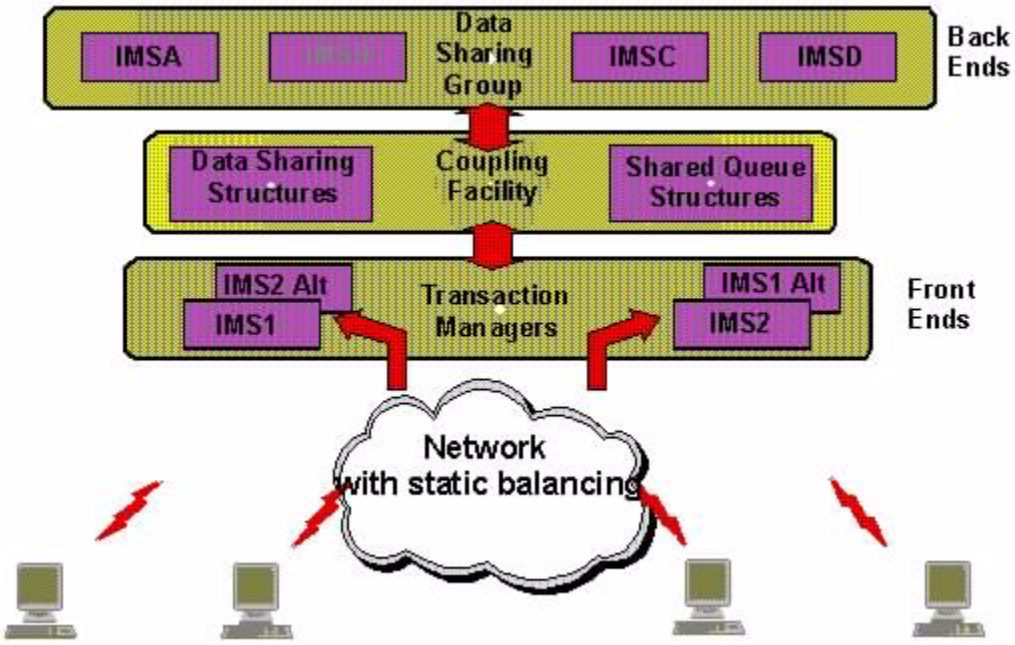
XRF and Parallel Sysplex Together

XRF and Parallel Sysplex are compatible. In fact, very good availability can be obtained by merging the best of each solution. One such environment could be an IMS fronto-end / back-end system, with the terminals managed by two IMS/TM subsystems, each with an XRF alternate. Users would sign on to IMS. The IMS/TMs would then pass all messages to an IMS Shared Message Queue, to be picked up and processed by multiple IMS/DBs.

Any DB2 data base or VSAM file requirements would be satisfied by DB2 or VSAM/RLS data sharing support.

In the event of an IMS/DB failure, the surviving IMS/DBs can load-balance the workload.
In the (unlikely) event of an IMS/TM failure, only half of the users would be affected. Of those on the failing IMS, existing Class 1 and Class 2 sessions would be routed to the XRF alternate while Class 3 and new sessions would log on there normally.

An example of this could look like the following:

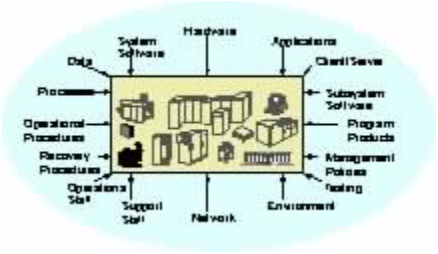


Note that VTAM Generic Resources and XRF are incompatible. Sessions can use "static" balancing; fixed routing based upon historic knowledge of how many users are in which group. Alternatively, one can create their own VTAM exits to do session balancing based on round robin or other techniques, with knowledge of which IMSs are active.

What Else Is Important?

Mindset:

A computing environment is composed of a variety of technologies, processes and people working together to deliver stable, reliable technology support to the business processes. What was discussed here are ways of improving the availability in the event of planned outages, or unplanned hardware, operating system, or subsystem failures. Through the techniques discussed, many of these outages can be masked from the end user or the outage time significantly reduced. This assumes that everything else that makes up the IT shop is working towards the same goal. In fact, the majority of all outages can be either prevented or their impact significantly reduced by creating and following good procedures, and following up on actions taken to prevent failures. Are the Operations staff trained to handle the situation? Are the recovery procedures will documented or automated? How skilled are system programmers? How robust is the application testing process? How often do the batch applications issue checkpoints/syncpoints? Where are the recovery procedures for failed batch jobs or transactions? How old are they? How would you handle media failures?



Some issues to think about when planning for availability include the following:

Issues / Topics	Tasks
•Configuration redundancy and isolation •Availability monitoring of configuration elements •Planned outages of configuration components •Availability impact of planned future configuration changes	•Understand IT Configuration Availability and Planned Outages

•Techniques for measuring the availability of systems and applications •Post processing management reports •Real-time monitoring	•Evaluate Availability Measurement Methods
•Technique for estimating costs	•Calculate the Cost of an Outage
•Client's process for collecting and analyzing outage data •Summarization of recent outage incident data •Client's process for quantifying impact to the business	•Analyze Outage Data
•Recovery processes and priorities •Human dependencies •Process for documenting, validating and maintaining currency of recovery processes	•Evaluate the Recovery Process
•Process to assess and manage the risk of change •System, application and stress test processes •Exceptions to test process and their impact •Test environment adequacy and isolation •Production migration process	•Understand Change Management, Testing and Migration Processes
•Application function and structure •Application data availability dependencies •Application data availability design characteristics •Other (non-data) availability design characteristics	•Determine Application Design Considerations

If you don't THINK continuous availability.....
you won't ACHIEVE continuous availability

Further Information

IMS/ESA Release Planning Guide V5 GC26-8031
MVS/ESA Setting Up a Sysplex GC28-1449
Continuous Availability S/390 Technology Guide SG24-2086
IMS/ESA Data Sharing in a Parallel Sysplex SG24-4303
IMS Sysplex Data Sharing: An Implementation Case Study SG24-4831
IMS/ESA Version 6 Shared Queues SG24-5088
IMS/ESA Shared Queues: A Planning Guide SG24-5257
IMS/ESA V6 Parallel Sysplex Migration Planning Guide SG24-5461
Using VTAM Generic Resources with IMS SG24-5487
IMS/ESA V6 Administration Guide: System SC26-8730

Appendix 1: Sample Migration from XRF to Parallel Sysplex

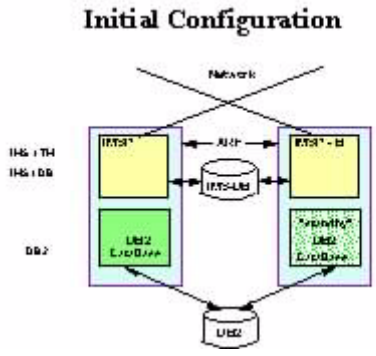
The following is an example of a migration from XRF to a data sharing environment.

IMS is the transaction manage, with some applications access only IMS DL/I data, while others also access DB2 data.

Since DB2 does not support any non-data sharing means to have multiple DB2s access the same data bases (Shared Read-Only Data is no longer supported), DB2 is not active on the alternate (B system) during normal operations.

If there is a failure, the Class 1 and Class 2 terminals would be taken over by IMSP-B. The users can then continue processing any DLI-only applications. Class 2 users need to re-signon to IMS. DB2, however, is not active yet. Any applications that require DB2 data will fail with a -904 SQL return code: Resource Unavailable. Either sophisticated automation or operator intervention is required to possibly shut down DB2 on the active LPAR, and once DB2 finishes shutdown, restart it on the alternate system. Backouts of in-flight work will slow down DB2 restart. This amount of time depends upon how DB2-intensive the work was, and how often commit points take place. Eventually, after DB2 emergency restart finishes, any DB2-dependant transactions can now resume.

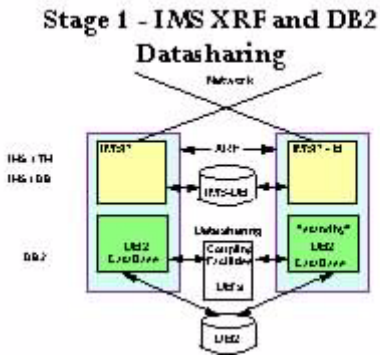
For those DB2-dependant transactions, there is an obvious availability impact, even with XRF, while having operations complexity.



The first step in the migration is to impliment a Parallel Sysplex environment and establish DB2 data sharing. Once done, a DB2 member of the data sharing group is started on both the Active and Alternate partitions. In the event of an IMS failure, the terminals would be switched as before with IMS and also DB2 data bases available. There would be some data unavailable as DB2 on the Active IMS's partition was holding locks for in-flights transactions. These transactions would have to be backed out for all the data to be available.

One point to consider is that if DB2-B has not expressed any interest in any of the data, DB2-A will acquire tablespace level "P-Locks" and "L-Locks" in the CF lock structure. If there was a system failure, DB2-B would not be able to access any of the data until DB2-A starts backing out these locks. To prevent that from happening, it is possible to kick off batch jobs on DB2-B every once in a while to keep inter-DB2 read/write interest.

Operationally, this is much easier to manage as there is no need to immediately restart DB2 on the alternate side in the event of an IMS failure. In the event of a partition or hardware failure, the restart process can be managed the OS/390's Automatic Restart Manager (ARM).

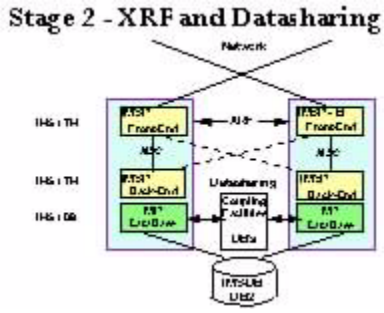


In Stage 2, IMS Block Level Data Sharing was set up for DL/I data. IMS XRF is still used by IMS/TM to manage the sessions. As transactions come in, static routing is done from the front end IMS/TM to the back end IMS/DBCTL regions. From known distributions of transaction rates, the workload is roughly balanced.

The customer now is able to use the full capacity of multiple servers and not be constrained by growth. This technique can similarly be used to expand beyond a 2-way, scalable to 255 IMS subsystems across 32 operating system images.

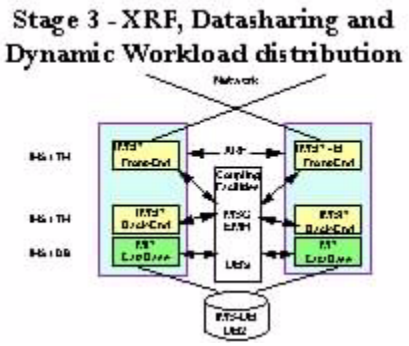
Availability is further improved over Stage 1 by separating the terminal and application management done by IMS. If a back-end IMS fails, MSC would just route all the tranactions to the surviving IMS. The IMS Workload Router tool is able to help in this by providing dyanmic support for MSC routing. If the front-end IMS fails, IMS-B will take over the terminals as before.

Since DB2 is active on both sides simultaneously, inter-DB2 read/write interest is always established, simplifying recovery issues.



Stage 3 removes MSC routing and replaces it with an IMS shared message queue. This provides dynamic workload balancing as the less-busy system will have more available Message Processing Regions to pull work off of the shared queue in the Coupling Facility. It is now much easier to dynamically add and remove IMSs from the configuration. Back-end IMS failures are totally invisible to users and operationally much simplier to manage.

Flexibility and simpler application recovery are also obtained by messages not being lost in the event of an IMS or system failure. Once on the message queue, the message will stay there and be sent to the terminal, even if the terminal moves to a different location.



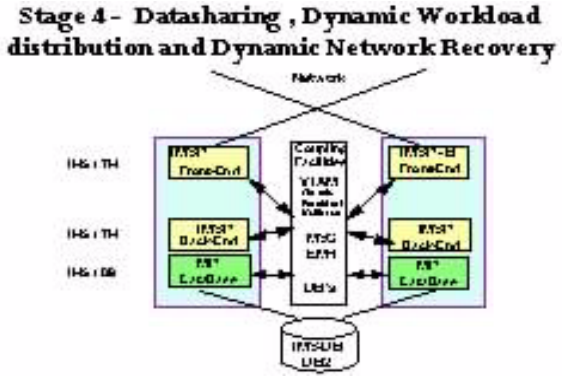
The final stage in the migration removes XRF and replaces it with two active IMS front ends. Terminal sessions are now dynamically balanced between the multiple IMS/TMs for SNA connections. TCP/IP connections can also be dynamically balanced through DNS support. With this support, multiple IMSs present a single system image to the end user.

As with Stage 3, a back-end IMS failure is invisible to the end-user. Messages would just get processed on the surviving IMSs.

An IMS/TM failure is handled differently now from Stage 3. Today, users are required to re-logon to IMS using VTAM G/R or DNS services, or else have Rapid Network Recovery recover the sessions when the failing IMS/TM is restarted (in place, or elsewhere). The users would then need to re-signon to IMS.

Operationally, without XRF, this is a much simpler environment to control and manage. There is no need for takeover actions, and much less chance for human errors.

The future direction of IMS and Communication Server development is to make IMS failures invisible to the end user.



Appendix 2: Migration Steps Moving to Parallel Sysplex Data Sharing

Migration to Parallel Sysplex Block Level Data Sharing

Naming Conventions
- Data Set High Level Qualifier
- IMS Group Name, Member Names, Proc Names
- IRLM Group Names, Member Names, Proc Names
- Plan IMS CF Structure Names
  - Lock, OSAM, VSAM

Plan IMS CF Structure Sizes
- ibm.com/s390/pso CF Structure Sizer

Define and activate CFRM Policy
- ibm.com/s390/pso CF Configuration Assistant

Review and update all database recovery procedures

Change Accumulation

Timestamp recoveries

Disaster Recovery

Review and update batch window procedures
- /DBR usage

Register databases with SHARECTL RECONs and SHARELVL(0) or (1)
- Implements authorization processing

Implement IRLM as lock manager (Mode = LOCAL)
- Eliminates segment locking
- Changes lock and deadlock processing

Allocate Group and Member level data sets

Register databases with SHARELVL(2) or (3)
- Databases may be registered in phases
- Invokes block locks for full function
- Eliminates FP lock manager use

Implement Lock & cache structures in CF
- Change IRLM to SCOPE=GLOBAL
- Adds "Read & Register" and lock processing overhead

Implement 2nd IMS subsystem
- Lock conflicts may occur
- Buffer invalidates may occur
- Additional overhead if lock contention

SMQ Migration

Plan CF Structure Names
Plan CF Structure Sizes
Update and activate new CFRM policy
Define logstream names to logger
Define RACF profies to protect CQS structures
Define data sets for CQS
Update IMS Start-up Procedures
Create CQS procs/parms
Update PROGxx and SCHEDxx for CQS

FDBR

Must be unique for each pair of IMS/FDBR
Enables FDBR for IMS (updates status field in CSA)
FDBR1 JCL
Use same parameters, proclib members, data sets, except
IMSID=FDRn
ARMRST=Y in DFSPBxxx
Disables ARM restart for active IMS
FDBR2 same as FDBR1 except IMSID

System Management

CMDMCS=Y enables command entry from MCS/EMCS console
CRC=@ for both IMSs allows one command to be routed to both IMSs
Generic start region capability in V6
    Could have used same JCL for dependent regions in IMS1 and IMS2
        /START REGION MPPDE LOCAL JOBNAME MPP1DE
    However, would have to start each dependent region one at a time
        /START REGION ALLDEPstarts all MPP regions
    Could use combination of generic start and TCO scripts