E 41

IMS Disaster Recovery

pete_sadler@uk.ibm.com

Anaheim, California October 23 - 27, 2000

RETURN TO INDEX

Trademarks

IBM Trademarks or registered Trademarks

- IMS/ESA

Other trademarks

- ENET - ENET Corporation

Topics

IMS Remote Site Recovery Overview
ENET RRDF
XRC hardware overview
PPRC hardware overview
GDPS implementation
IMS recommendations

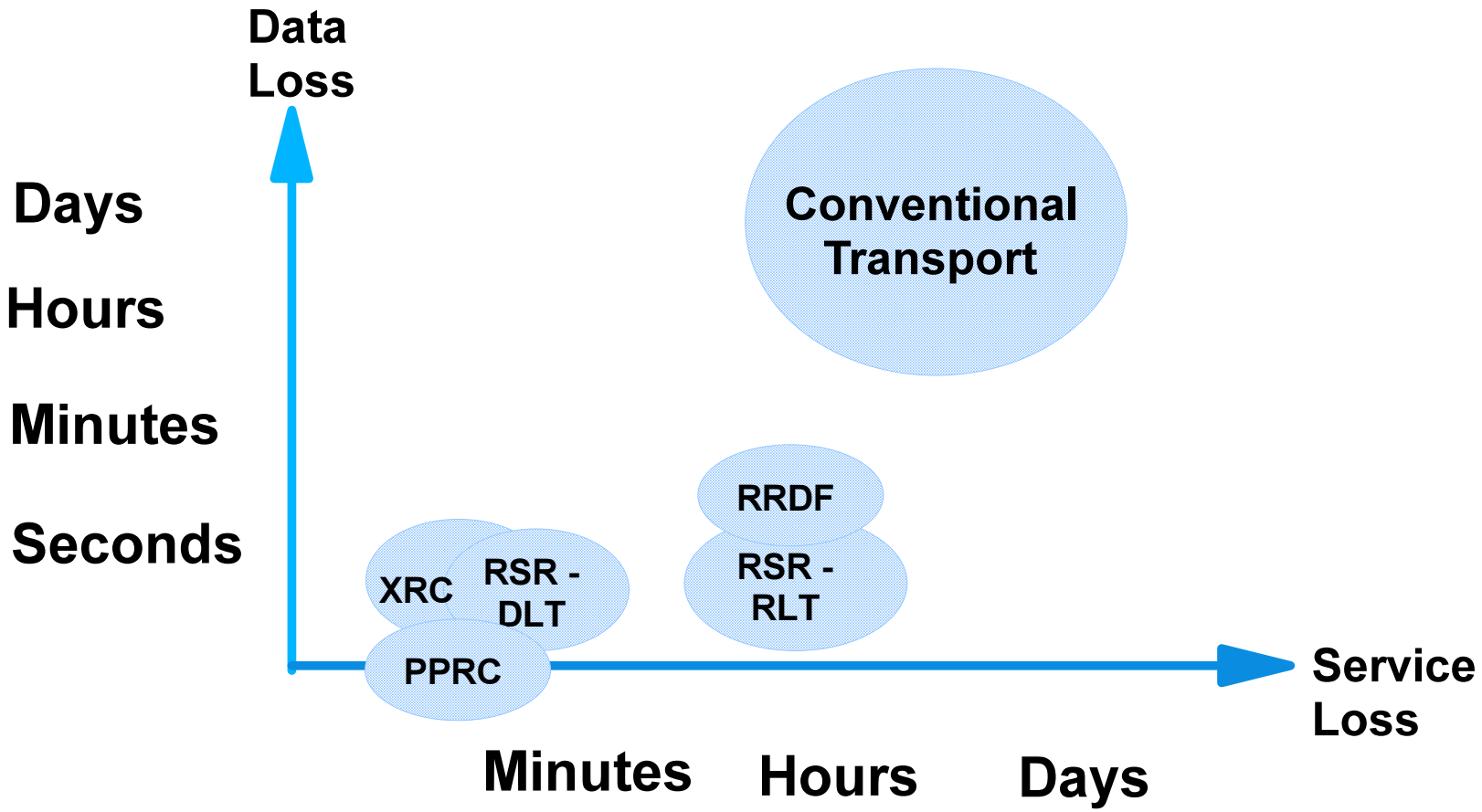
Topics

Notes

This presentation will cover IMS disaster recovery options including the IMS V6 RSR feature, ENET's RRDF, and Hardware redundancy solutions like XRC and PPRC.

We'll also discuss combination solutions and solutions that work with both IMS and DB2.

Data Loss and Takeover Speed



Data Loss and Takeover Speed

Notes

- RSR offers two options:
  - RLT where log data is transported and RECON states are maintained
  - DLT where DBs are shadowed in real time
- RRDF from ENET offers RLT type functionality
- XRC is a hardware asynchronous copy methodology
- PPRC is a hardware synchronous copy methodology
- Traditional methods send copies of logs and ICs to the remote site by over land transport.
- ALL asynchronous methods risk the loss of some in flight data

SHARE Defined Tiers

Tier 0 - No Disaster Recovery Plan

Tier 1 - Pickup Truck Access Method

Tier 2 - PTAM and Hot Site

Tier 3 - Electronic Vaulting

Tier 4 - Electronic Remote Journalling/Logging

Tier 5 - Two Site Two Phase Commit

Tier 6 - Zero Data Loss (Synchronous Remote Copy)

SHARE Defined Tiers

Notes

Tier 0 - No Disaster Recovery Plan

All data lost, no recovery possible.

Tier 1 - Pickup Truck Access Method

Secured backup but no D/R site. New data after backup is lost.

Tier 2 - PTAM and Hot Site

As Tier 1 but with D/R site. Recovery is 24-48 hours.

Tier 3 - Electronic Vaulting

Remote tape library. Can reduce data loss and recovery window.

Tier 4 - Electronic Remote Journalling/Logging

Small data loss. Applying updates minimises recovery time.

Tier 5 - Two Site Two Phase Commit

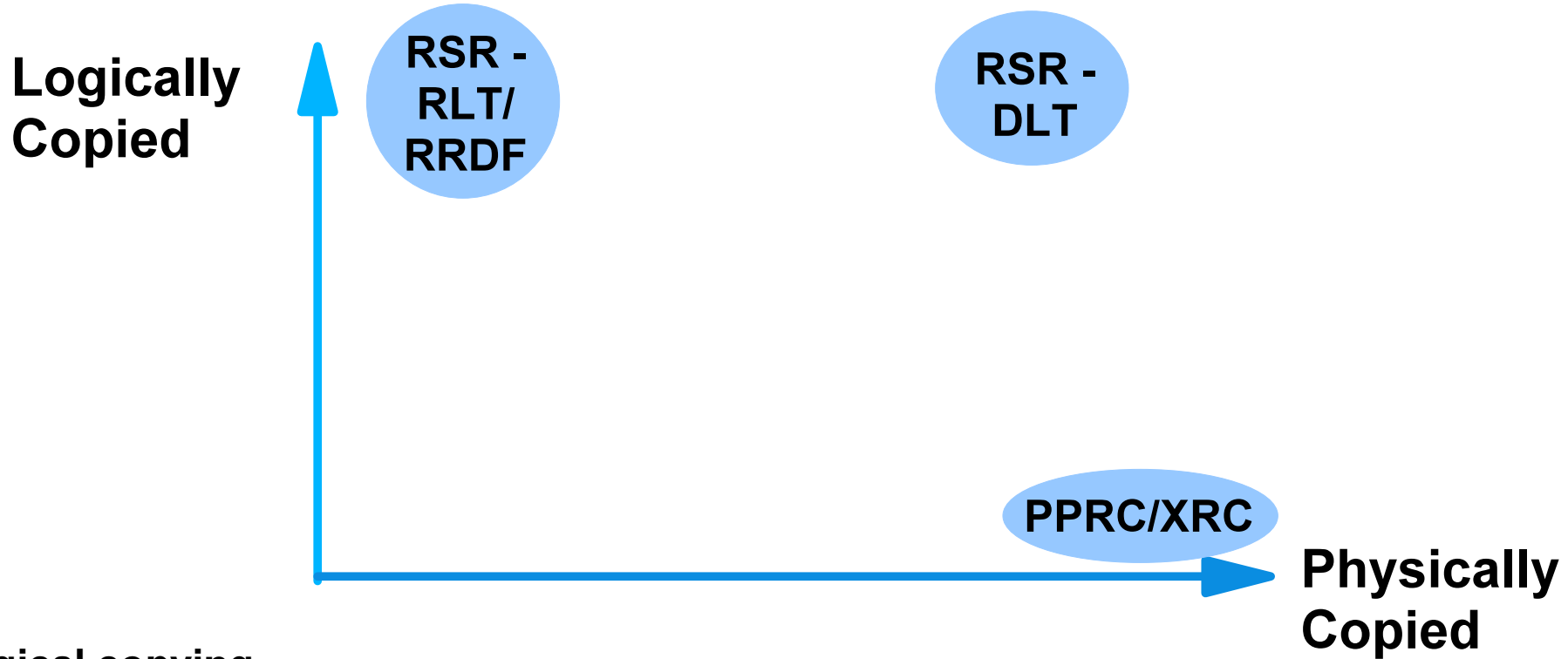
Application controlled. Secondary outage affects Primary.

Tier 6 - Zero Data Loss (Synchronous Remote Copy)

Zero data loss if Synchronous, seconds if Asynchronous

Rolling failure means inconsistent data at Secondary site

Logical Recovery Vs Physical Recovery



Logical copying

all data needed for recovery is available

Physical copying

all data is physically available

RSR is built to ensure that the remote site is logically correct.

Logical Recovery Vs Physical Recovery

Notes

It is important that the state datasets for IMS be logically correct at the remote site.

By that I mean that the RECON at the remote site must accurately reflect the DB states at the remote site, NOT at the active site. The same is true for OLDS datasets.
For this reason hardware copy methods must be configured such that all the data or none of the data is transferred.

Archived SLDSs on tape are not hardware copied.
Theoretically, customers could manually track and adjust for pack by pack differences, but the risk of error is probably too high for any to attempt it.

IMS RSR Design Points

- Asynchronous Tracking
- Efficient Tracking
- Parallel Sysplex support
- Non-disruptive resynch
- Disaster Relief
- Site switching in under 60 minutes
- Complement XRF, PPRC, XRC
- Allow warm start of system

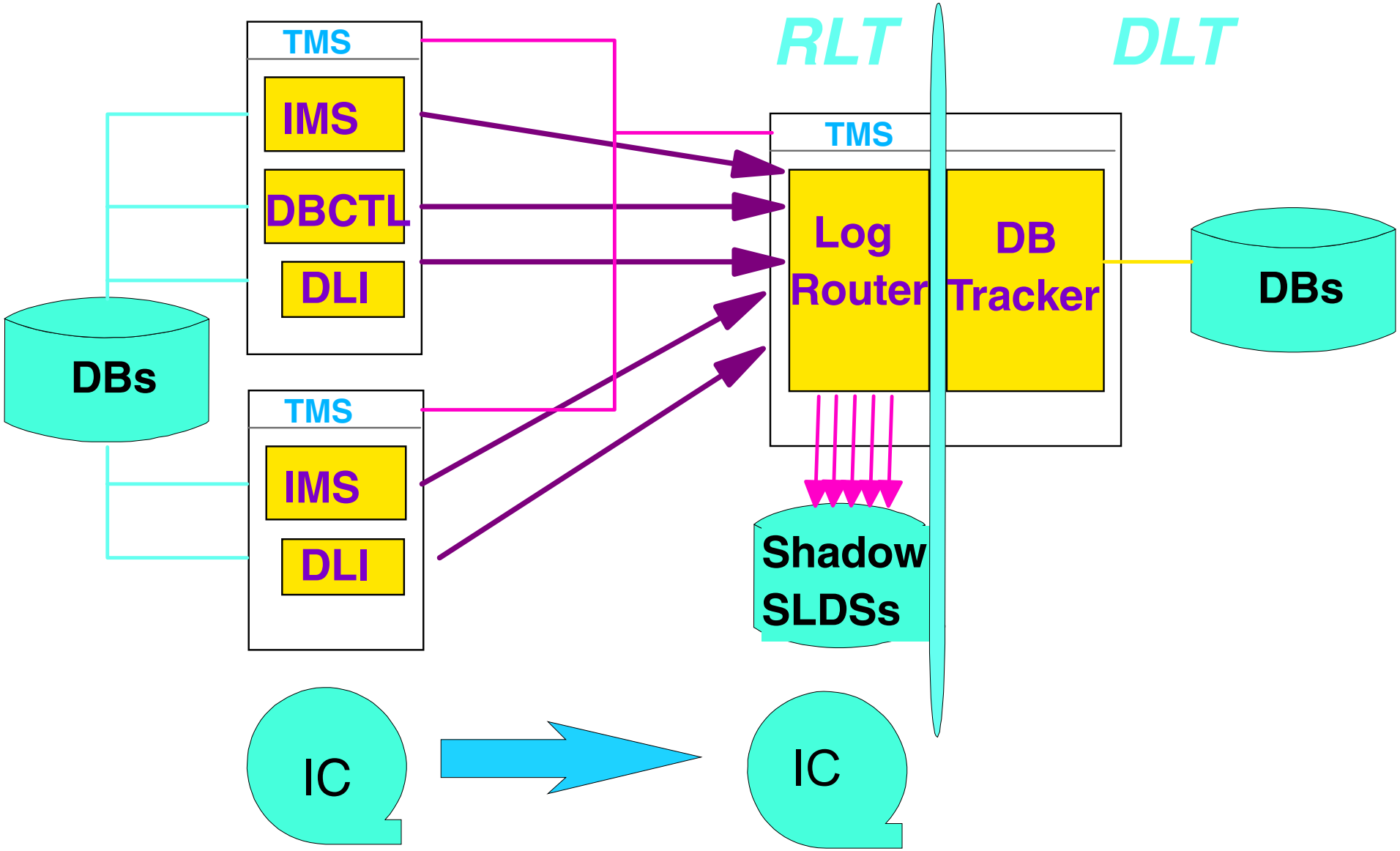
IMS RSR Design Points

Notes

RSR was meant to complement XRF, not replace it. It was intended for infrequent use so it was deemed important that

- Performance of the active system must NOT be impacted; therefore tracking is asynchronous
- The tracking site should use minimal computer resources.
  The design of the tracker avoids IRLM locking, bulks the updates, and allows DB recoveries in a single pass
- Service availability, in a Disaster vs Cost equation, indicates 60 minutes is a reasonable trade off
- Warm start to handle available messages was desired and the ability to permit /ERE BLDQ
- BLDS in a parallel sysplex is supported
- Batch DL/I jobs are included
- Non-disruptive resynchronisation is possible because
  DB Recovery is possible at the remote site in case of media failure
  log gaps are filled automatically
  Online Forward Recovery is automatic for DBs in DLT

RSR Overview



RSR Overview

Notes

- RSR comprises
  - the Transport Manager address spaces in each MVS
  - the Tracker system at the remote site
  - appropriate log data sets and RECONs
- The TMS assists the IMS subsystem loggers establish APPC conversations on which log buffers are sent at OLDS write time
- The log buffers are filtered to drop log records that are not of interest
  - for DBs not covered, trace records and so on
- Log Router produces shadow SLDSs at the Tracking Site for RLT
  - separate log stream for each active subsystem
- Router passes DB log records to Database trackers for DLT
- Image Copies are transported to the Tracking site outside RSR
- Tracking site RECONs maintain recovery status information for the Tracker

Tracking Subsystem Components

Transport Manager - Helps Conversation setup between actives and tracker
Isolated Log Send - to catch up
Log Router

- **Manage Shadowing of active site Logs**
- **Manage Automatic Log Archive of shadowed logs**
- **Manage Online Forward Recovery of shadowed DBs**
- **Manage Filling of Log Gaps in shadowed logs**
- **Manage Datasharing Stream Merge**
- **Route Data to Database Trackers**
- **Synchronize with XRC**

Database Tracker - Apply data to shadow DBs

- **Separate DLI and FP trackers**

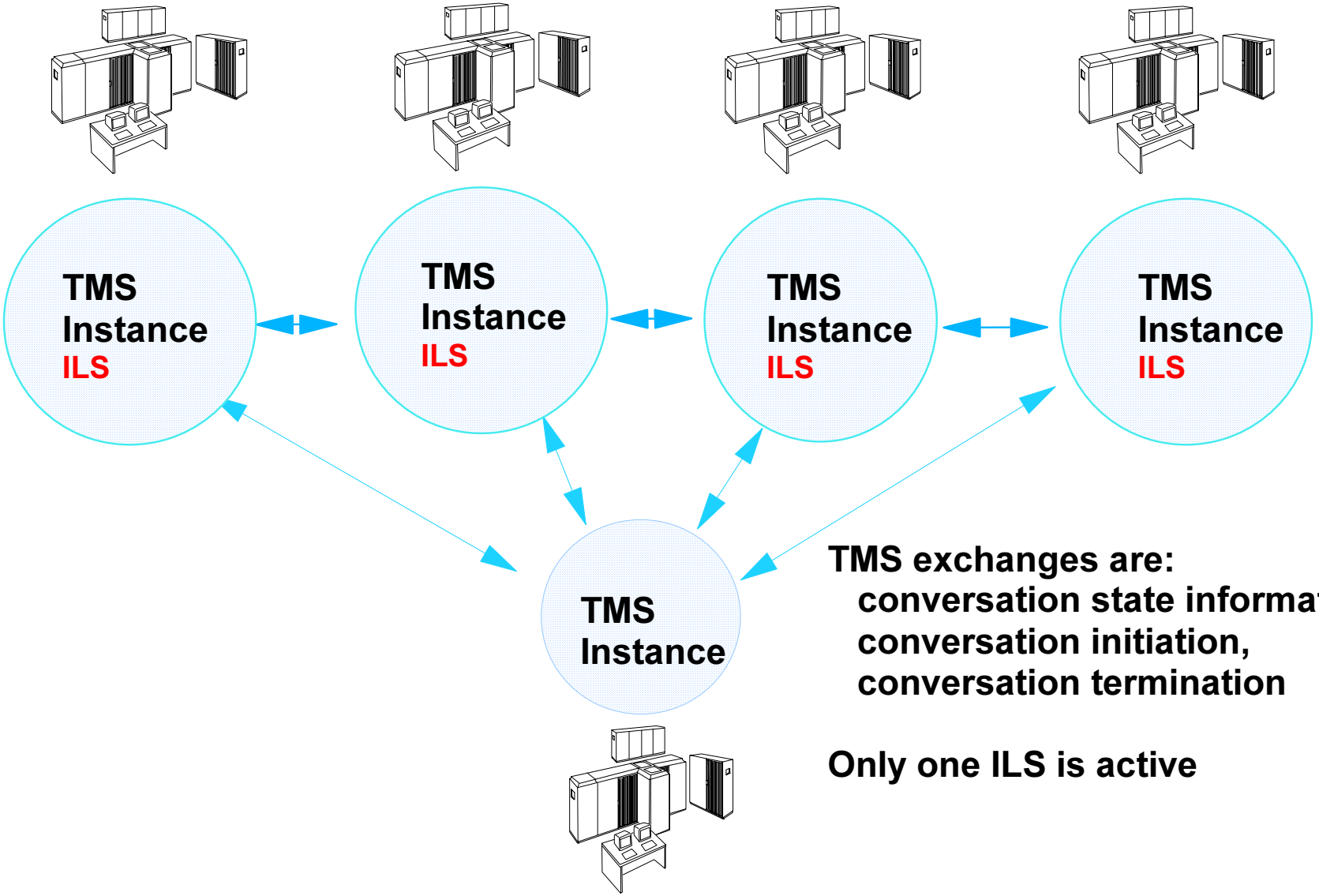
DBRC - Knows everything about Logs and DBs

Tracking Subsystem Components

Notes

- Transport Manager - separate new address space to facilitate APPC conversations between active site images and tracker
- Log Router - listed functions, but XRC support only available with V6. The log router runs as a major component in the tracking IMS.
- The tracking IMS uses the underlying IMS systems engine, much of the FF DB and DEDB engines, the logger, and DBRC. It does not support concurrent transaction work.

Transport Manager



TMS exchanges are:
- **conversation state information,**
- **conversation initiation,**
- **conversation termination**

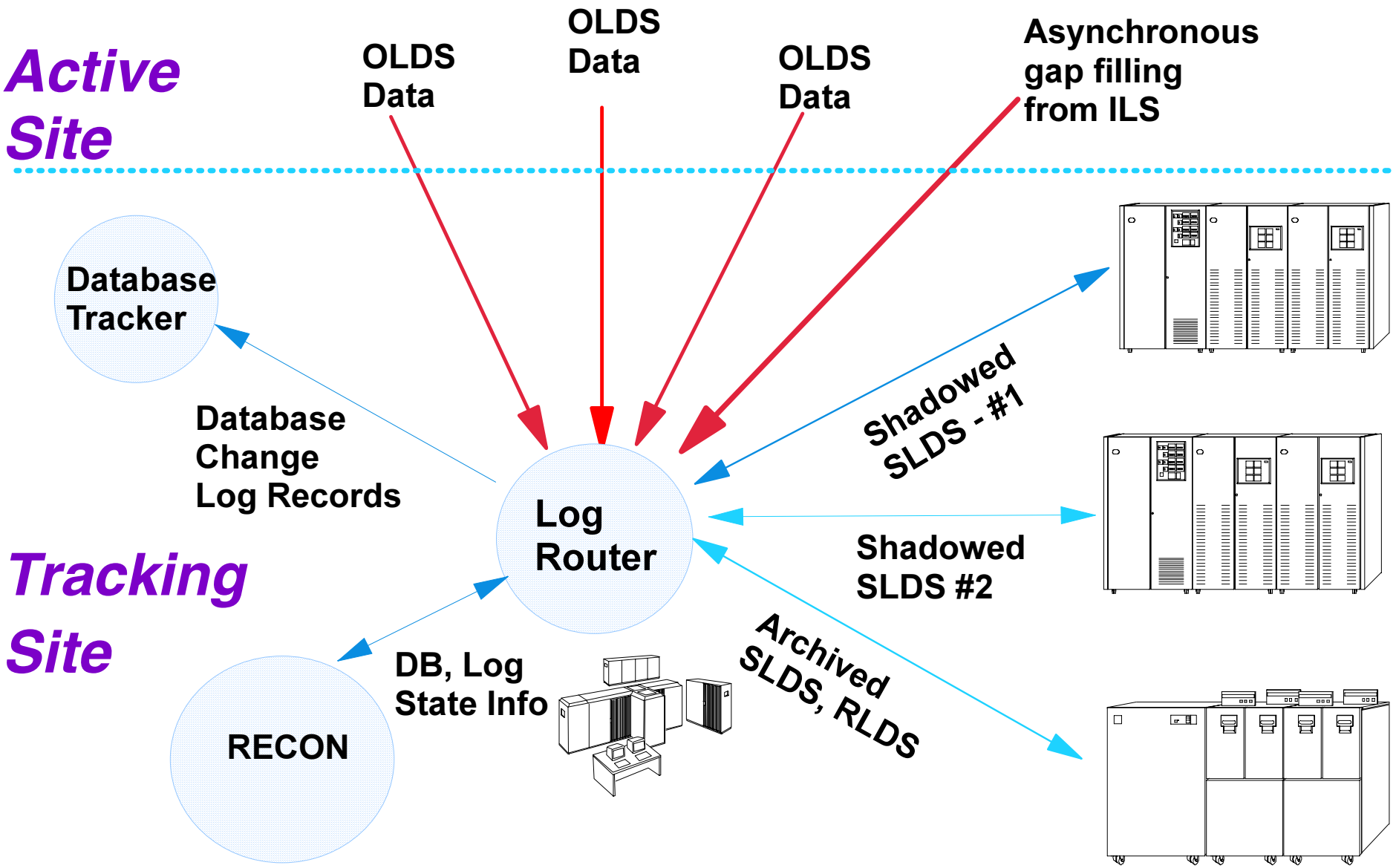
Only one ILS is active

Transport Manager

Notes

- There needs to be a Transport Manager instance (TMS) for every MVS image that has IMS image(s) creating IMS log data
- All TMS instances are in communication with each other
- Only one Isolated Log Send (ILS) instance will be employed at the active site, but it must be able to access all active site log data
- The TMS initiates and facilitates conversations, but then IMS does all the work of sending the log blocks
- TMS is a new and separate IMS address space and started task

The Tracking Subsystem - Log Router

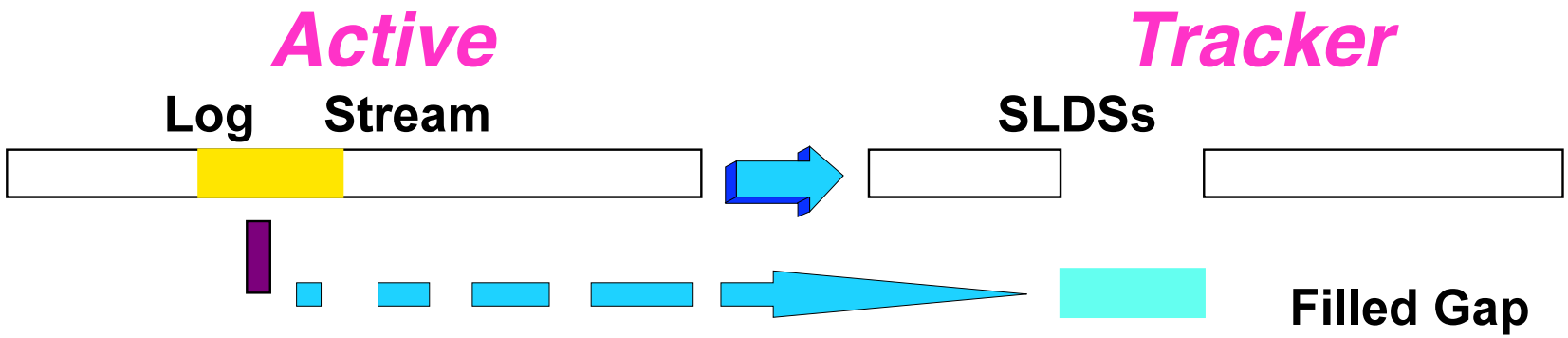


The Tracking Subsystem - Log Router

Notes

- ★ Dynamic allocation of all shadow log datasets.
- ★ Automated management of log archiving.
- ★ Single control point management of all active site log streams.
- ★ Automation of log write error handling by copying to new log dataset.
- ★ Each active site log stream is kept separate at the remote site (IMS EREs require their own logs)
- ★ Blocksizes for SLDS can differ from active.
- ★ Active and Tracker SLDS boundaries will not match
- ★ Log Router
  - internally handles data sharing merge for DB trackers
  - Coordinates data processing by DBRC, System, and DB tracker components to ensure proper logical copy of active site state.
  - Controls parallel, multi DB database recoveries with OFR

Log Router - Isolated Log Transport



A combination of active site ILS function and tracking site log router function

Manages the filling of gaps in received log streams. Initiates and controls the sending of missing log segments.

Maintains conversation with Active Site Isolated Log Send Component

Supports multiple parallel ILS conversations - up to a user specified maximum or the number of datasets being transmitted by isolated log send.

Log Router - Isolated Log Transport

Notes

- There are many reasons why data can not be sent from an active site IMS to the tracking system real time
  - When this occurs there are two options:
    - Send it later
    - Pause service until you can send it
- IMS customers can rarely afford to reduce their active site availability in order to guarantee data is hardened at the tracking site. RSR provides a separate log transmit facility that acts as a log transmit server - driven by requests made by the tracking IMS system.
- Only 1 Isolated Log Send (ILS) component is active per active site and it is a sub-component of one of the Transport Managers.
  - An ILS must be able to access all active site log data. The user may control the number of transmit pipes; the maximum is one per OLDS or SLDS of which part is to be transmitted. If you start multiple ILS images only one will connect - others will wait in case of the active TMS/ILS failing
- Each gap will cause its own SLDS to be created.

Isolated Log Data Transport Flow

Log Router recognizes log data missing

Log Router sends request to ILS to fill gap

ILS queries DBRC at active for log DS info for gap

ILS selects OLDS/SLDS

ILS transmits gap data to Log Router

Log Router writes data and updates Recon

Log Router resumes tracking functions with provided data

Isolated Log Data Transport Flow

Notes

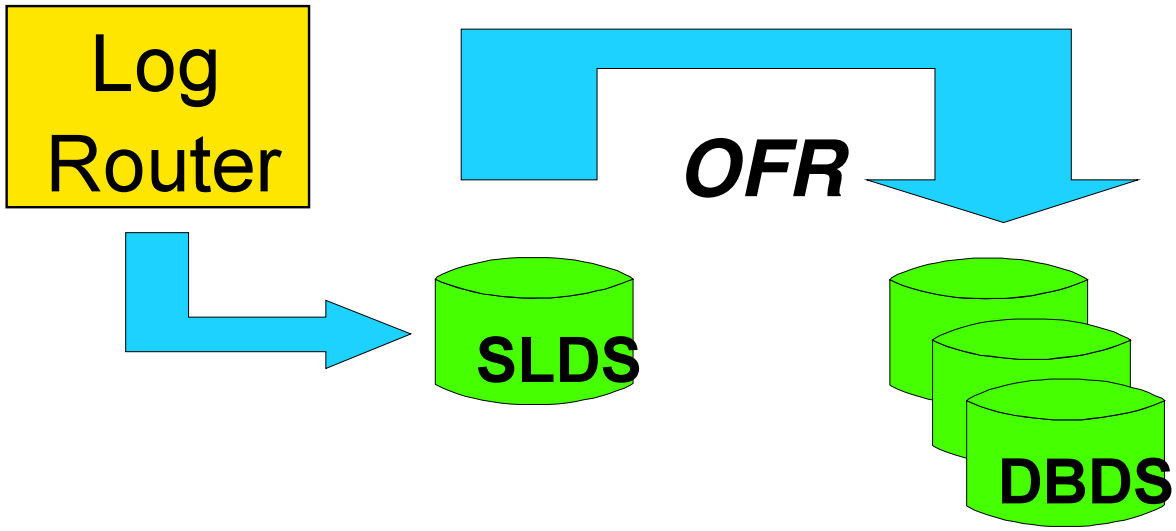
The number of APPC sessions for ILS can be controlled by the user, but one per OLDS or SLDS to be transmitted is possible for RSR.

ILS transmissions compete for transmission bandwidth with active site IMS log data.

It can be argued which is more important, but overrunning your bandwidth can cause numerous small gaps in the active IMS transmissions.

These may be operationally disruptive at the remote site since each becomes a separate SLDS.

Log Router - Online Forward Recovery



Provides catchup for DLT
Multiple Areas/DBs in a single log pass
Initiated by /START DB or /START AREA
OFR restarts at tracker ERE
Supports sharing merge automatically
Supports smooth transition into normal Tracking

Log Router - Online Forward Recovery

Notes

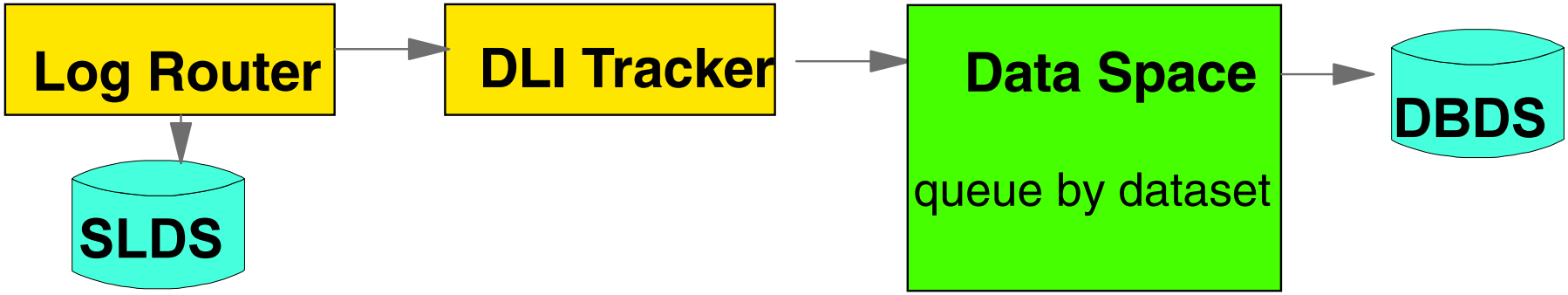
RSR as an environment can involve lots of DB recovery activity.

As a result it is important that the recovery is automatic and efficient.

In addition the target is constantly moving and the solution must be able to blend into normal remote site DB tracking.

OFR allows all DB datasets to be recovered from different points in a single pass of the necessary log data under control of the tracking IMS system.

DL/I DB Tracker Main Points



- **No locking required**
- **Workload unit is Block Number / CI Number, not transaction**
- **Databases are updated in REDO (Forward Recovery) manner**
- **Data is buffered in Dataspaces**
- **Databases are not usable prior to service transfer**
  - not guaranteed to be transaction consistent while in active tracking

DL/I DB Tracker Main Points

Notes

DLI DB tracking is done in the SAS

DB work is subtasked in a different way to that at the active site.

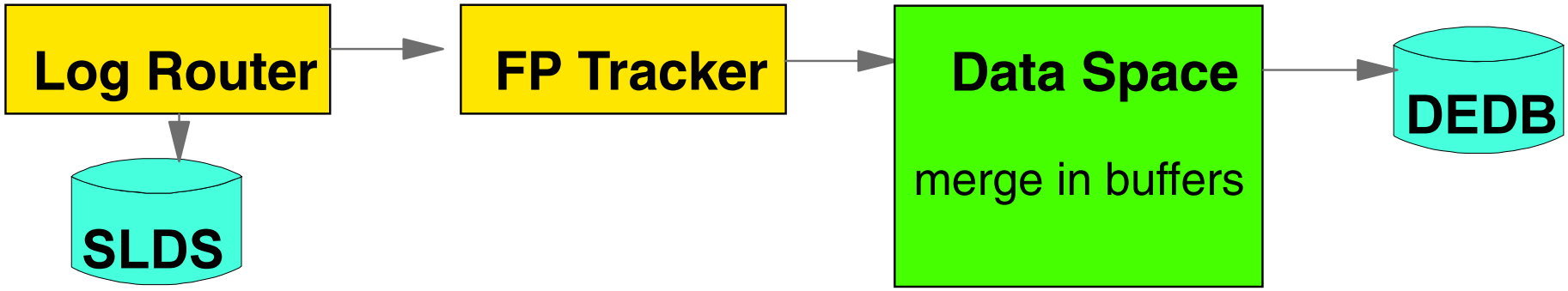
For DLI data it is not even handled in transactional units.

This is done to avoid locking and to get maximum efficiency for the tracker.

If you think of an XRF system as being in perpetual restart, you should think of an RSR system as being in perpetual forward recovery.

Both DLI and FP store in-flight data in a dataspace.

Fast Path Database Tracker



- **Work is done under Tracker's Control Region Fast Path TCB**
- **Work is buffered in dataspaces and changes are rolled up**
- **Work is hardened at regular intervals**
- **DEDB Areas are not usable prior to service transfer**
  - Areas are normally unit of recovery consistent (except during I/O processing).

Fast Path Database Tracker

Notes

FP DB work is done under the Control Region address space (FP TCB).

Since FP doesn't log undo records, the FP tracker only applies data that has reached a commit point.

The data is written out at intermittent intervals and in large chunks. As a result the transactional consistency can not be depended on until the tracking system as been gracefully shut down.

DBRC at the Tracker

DBRC is Required

- **Covered DBs registered with DBRC in Global Service Group**

Tracker owns resources exclusively

DBRC tracks OFR, DB, Log state info

DBRC tracks takeover state info

DBRC at the Tracker

Notes

DBRC is required.

It owns all DB resources exclusively - it will not allow subsystem authorisations at the tracking site until a service transfer is performed.

It tracks the system state, the shadows of the various active logs, and the logical IC/DB state at the tracking location.

Normal Tracking Data Flow

Active IMS Logger sends data at OLDS write

Log Router receives data and writes to SLDS for that stream

DBRC Tracker component processes log records that pass RECON state information

System Tracker component processes log records that require system state tracking

Logs for shared DBs are merged in timestamp order

Merged logs are passed to DB trackers

DLI & FP Trackers apply data to DBs

Missing log records trigger Isolated Log Send

Normal Tracking Data Flow

Notes

Because some WADS activity can be skipped in the event of an ERE the tracking system can not harden anything that has not hardened on the active site OLDS.

The Log Router not only ensures the various log streams are written out and reflected in DBRC, but it also makes sure all log data is received,

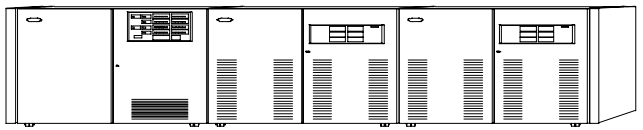
- that DBRC information passed in special DBRC information log records is processed,

- that some system information reflected in IMS log records gets reflected in DBRC,

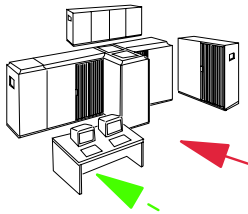
- that DB sharing data is properly merged and then passed to DLI and FP DB trackers,

- and with V6 it makes sure that IMS data stays in time consistent synch with XRC data like DB2 DB updates.

Four Way XRF / RSR Site Overview

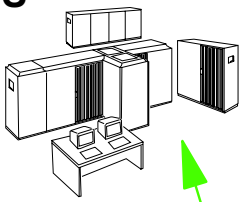


1 RECON, OLDS, SLDS,
Shared DBs, Unshared DBs

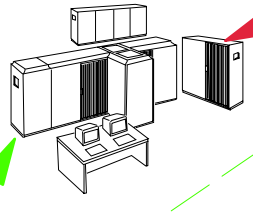


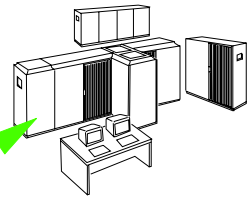
IMS XRF Alt
IRLM
TMS

IMS XRFActive
IRLM
TMS



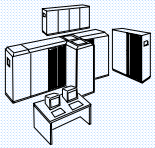
IMS XRF Active
IRLM
TMS



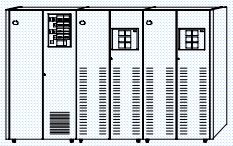


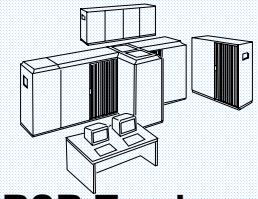
IMS XRF Alt
IRLM
TMS
ILS Instance

1 RECON
Tracker's OLDS
No Local DBs



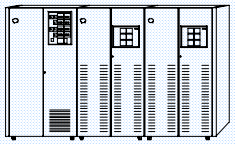
Dormant

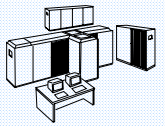


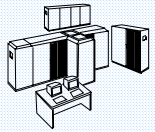


RSR Tracker
TMS

Shadowed Logs
Shadowed DBs







Dormant

Four Way XRF / RSR Site Overview

Notes

A complex 4 way XRF data sharing RSR overview.

The tracking site uses only one CPU to track the active site IMS systems.

Only XRF actives are in active communication with the tracker because only actives are creating log data.

Isolated Log Send (ILS) could have been started on any one of the active site machines .

Only one RECON at each site.

Active site could also have batch jobs and utilities.

Planned Takeover Flow

- Initiated at the Active site ONLY
  - No more normal signons (abnormal OK)
  - All IMS subsystems must terminate normally
- Recognized by Tracker in passed Log Stream
- Tracker receives and processes all log data
- OFRs need not complete
- Tracking subsystem shuts down and RECON roles changed
- New Actives are restarted (ERE or COLD)
- Switch is hardened by DBRC processing at first signon

Planned Takeover Flow

Notes

This type of takeover can also be used for routine site maintenance because the "old active" is left at a consistent and known state that is in synch with the remote site.

It can immediately take on the role of tracker.

Unplanned Takeover Flow

Initiated at the Tracking Site ONLY

User must terminate Active Site subsystems

- If not true disaster

Conversations with Active Site are terminated

- If not already broken

Tracker processes all received log data

OFRs need not complete

Tracking subsystem shuts down and RECON role changed

New Actives are restarted (ERE or COLD)

Switch is hardened by DBRC processing at first signon

Old Active Site REQUIRES tracking site rebuild

Unplanned Takeover Flow

Notes

After a real unplanned takeover from the active site, the DBs and RECON no longer reflect the true state of IMS.

The NEW TRACKER DBs and RECON must be replaced by refresh copies from the NEW ACTIVE site.

Once you enter the command at the remote site it will stop taking data from the active - don't do it until you've squeezed what you want/can from the active site.

Remember that in a sharing environment the remote site MUST truncate to the least common denominator (the oldest timestamped log end point).

DO NOT USE THIS FOR PLANNED OUTAGES.

Tracker Performance

Recovery Level Tracking - 5% Active Site CPU.
Exact numbers determined by log volume and VTAM RU size definitions.

Database Level Tracking - 10-15% Active Site CPU
Exact numbers depend on activity and DLT choices

Transmission and Processing capacity needs to be at least 2x normal so that the tracking subsystem can catchup from gaps, transmission drops, etc.

Tracker Performance

Notes

The exact cost is a function of many things:

- VTAM RU size is best for IMS when it is large (OLDS block size), but that is not efficient for transaction traffic. Customers will need to tune.

-DB overhead is a function of application DB calls vs application logic

-Higher transmission bandwidth allows you to plug holes faster which translates to reduced data loss exposure.

Packaging of RSR

- Separately Priced Features
- Both DLT & RLT variations are available
  - (RLT means no Online Forward Recovery)
- Feature Required on Actives plus you need a Tracker
- DLT requires RLT as ordering prereq

Packaging of RSR

Notes

You can order just the log tracking (RLT) or the log and database tracking (DLT) features.

For ordering, the DLT feature requires the RLT feature as a prereq. From an internals view the RLT function is a subset of the DLT function.

In addition RLT requires a control region (DB or TM) while DLT requires DB.

If you want to have DLT and messages to your MTO you need RLT+DLT+DB+TM which is what I recommend for the tracker.

RSR Pros & Cons

Tier 4

Pro

- Minimal bandwidth.
- VTAM distances.
- Tailorable by DB - DLT, RLT or not covered.
- Tolerates transmission outages.

Con

- Flexibility leads to complexity
- Only supports IMS data
- Operations involvement

Other Software Transport Products available

ENET

Tier 4

Provides a functionality similar to IMS RLT

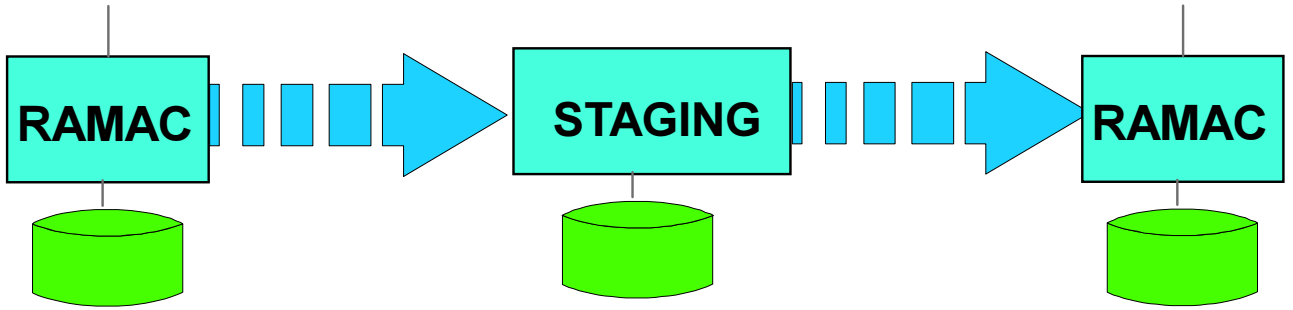
Supports both IMS and DB2

Maintains IMS RECON DB info

Internally works quite differently from RSR

ENET was marketed by IBM as RRDF

eXtended Remote Copy - XRC



Allows time stamp consistent copying of multiple DASD volumes to a remote site in real time.

IMS recommendation is copy OLDS, WADS, DBs, RECON - all or nothing.

IMS recommendation of ERRORLEVEL=SESSION to ensure both logical and physical views at the remote site.

Use multiple redundant ESCON

Enlarge the OLDs capacity to avoid remote site need for tape SLDS.

Make sure you do ICs after a disaster since you may not have enough log data to go back to the oldest valid IC without archived log data.

eXtended Remote Copy - XRC

Notes

XRC provides asynchronous, multiple DASD, transport to a remote site that will keep all packs at a time consistent state. Because the copying of the IMS state datasets (like RECON) must be also LOGICALLY valid it is the IMS recommendation that you copy all or nothing for IMS.

XRC works for DASD not tape, so Log Archiving is a problem.

Make sure you don't need archived logs for recovery after a disaster.

Enlarge your OLDS/SLDS DASD capacity.

In the case of a disaster you need to ensure you have current image copies of all DBs.

Remember that you only have the log data that is currently on DASD.

Any recoveries that need archived log data will have problems - the reason to have a large OLDS/SLDS capacity on DASD.

Staging can be done at a third site - ensure it is not on the primary site.

Consider the configuration when you have switched sites

XRC Pros & Cons

Tier 4

Pro

- Hardware only solution and works on any type of DASD content.
- Application independent

Con

- Requires more transmission bandwidth (10x)
- Does not handle IMS log archiving
- ESCON (Channel Extender) distances only
- You may have DBs that lack enough log data to be recovered until you take new image copies

XRC Pros & Cons

Notes

Because XRC will transmit entire altered blocks (not just the altered data) the bandwidth can be substantially larger than with RSR.

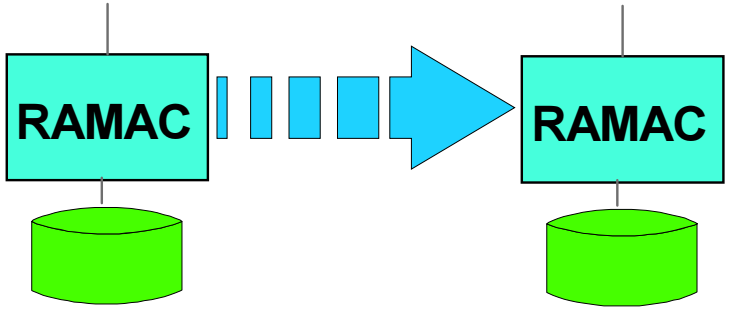
Log archives are a problem for both IMS and DB2.

- The IBM 3494 Magstar Vitual Tape Server duplexing function may help

Easier to operate than a software approach

It works for all DASD data

Peer to Peer Remote Copy - PPRC



Allows control unit level synchronous copying of DASD data to a remote site.

NO DATA LOSS in a disaster.

Because active site writes must wait for remote site duplication there is an unavoidable active site performance impact for both throughput and turnaround.

Remote Write failure results in Write failure reported to MVS

IMS recommendation is only to use if transaction characteristics demand zero data loss in a disaster

IMS recommendation is CRIT=Y to ensure an all or nothing copy mode.

Peer to Peer Remote Copy - PPRC

Notes

This approach impacts both active site availability and response times.

It is only viable when protection of the data is the highest priority.

An example would be an inter-bank system that handled billion dollar (choose your own currency) transfers.

PPRC Pros & Cons

Tier 6

Pro

- Only synchronous method and only method to ensure no data loss in a disaster
- Hardware only and minimal interaction with operators.

Con

- Impacts IMS throughput and turnaround
- Unable to handle logs archived to tape
- Requires more transmission bandwidth than RSR (10x)
- ESCON distances only

PPRC Pros & Cons

Notes

Guarantees no lost data.

Easier to operate than software solutions

If the link between sites goes down - the active site is unavailable
(but see Geoplex later)

Log archives are a problem for both IMS and DB2.
The IBM 3494 Magstar Vitual Tape Server duplexing function may help

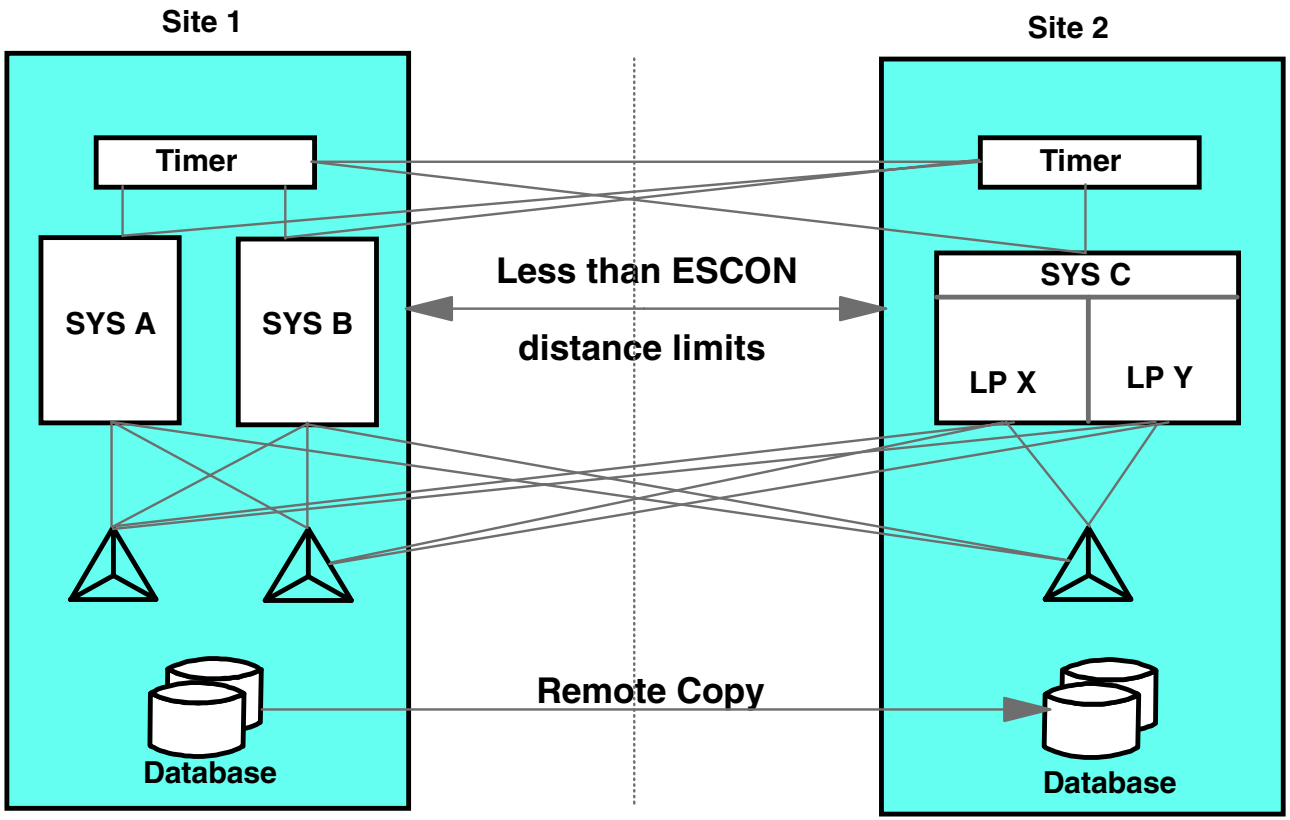
Like XRC you need more bandwidth than with RSR

Like XRC you need to ensure you have image copies after a disaster.

Geographically Dispersed Parallel Sysplex (GDPS)

Service offering based on enhancements with PPRC

Automation expects/tolerates rolling failures



Notes

Geogrphically Dispersed Parallel Sysplex (GDPS)

★ Additional PPRC operating characteristics when remote write failures occur

- FREEZE & GO - Tier 6
  - stops all secondary site writes but continues primary service
  - preserves consistency status at secondary site for later switch
  - backlevel data unless PPRC resync
- FREEZE & STOP UNCONDITIONAL - Tier 4
  - stops all writes (primary and secondary) and switches
  - interrupts service until switch complete but no missing data
- FREEZE & STOP CONDITIONAL - Tier 4 or Tier 6
  - stops all secondary writes if secondary site write error, otherwise switches
  - switches if problem is not with write into secondary control unit
  - preserves consistency status at secondary site for later switch
  - backlevel data unless PPRC resync

GDPS Pros & Cons

Tier 6 plus or Tier 4

Pro

- Based on PPRC but tolerates rolling failures
- Hardware only and minimal interaction with operators
- Application independent

Con

- Impacts IMS throughput and turnaround
- Unable to handle logs archived to tape
  - The IBM 3494 Magstar VTS duplexing function may help
- Requires more transmission bandwidth than RSR (10x)
- ESCON distances only
- Must restart on a Parallel Sysplex

GDPS Pros & Cons

Notes

Easier to operate than software solutions

Installation choices for failure management

Applies to all applications

If the link between sites goes down - the active site can continue
but Primary data may get marooned (like asynchronous)

Like XRC/PPRC you need more bandwidth than with RSR

Like XRC/PPRC you need to ensure you have ICs after a disaster.

Other Hardware Transport options

Non-IBM hardware offering

Both Synchronous and Asynchronous offerings

IMS recommendation - copy all or nothing

IMS recommendation - use asynchronous unless zero data loss is required.

Other Hardware Transport options

Notes

There are both synchronous and asynchronous methods available from non-IBM sources.

IMS recommends an all or nothing copy scheme for all hardware methods, plenty of transmission bandwidth, and increased OLDs sizes.

IMS Combination List

IMS RSR in DLT mode at the Version 6 level to support synchronization with XRC copied DB2 data

XRC to copy DB2 data, IMS program and system libraries

Electronic copying of 1st image copy after any reorg or DB load

Low cost transport of routine image copies

Enlarge DB2 log capacity to avoid need for archived data after a disaster

Don't penny pinch on data transmission bandwidth

IMS Combination List

Notes

RSR provides the best support for the IMS data.

You will need to use XRC for the DB2 or CICS data.

To use both you will want the IMS V6 enhancement to RSR that allows IMS to synchronize with the XRC data.

The greater your transmit bandwidth the quicker you can catch up from transmission holes. At minimum, allow for 2x your normal heavy transmit volumes.

Summary

The best solution is currently a combination software and hardware

Both software-only and hardware-only solutions are available

Archived logs are a problem for hardware only solutions

Summary

Notes

Don't forget your other required files.

Think about your ICs and archived logs

Understand that money is an unavoidable part of your decision.

Everyone would like to have zero data loss and zero availability loss.

Your actual solution will depend on a combination of your needs and your resources.

IMS Recovery Saver

Tool to enhance DB recovery processing

At local or remote site -

alternative to RSR for user-shipped logs

Conditions logs and RECONs

enables timestamp recovery to any point in time

Resolves DBRC information

Assists in co-ordinated IMS-DB2 recovery

Product number 5655-A68

RETURN TO INDEX

Bibliography

GG24-4210 Disaster Recovery Library : S/390 Technology Guide

Discusses products, communications and systems management

GG24-4211 Disaster Recovery Library : Design Concepts

Discusses concept, options, strategy and implementation

GG24-3993 Disaster Recovery Library : Database Recovery

Discusses DBMS data (IMS, CICS/VSAM and DB2)

GG24-3994 Disaster Recovery Library : Data Recovery

Discusses non-DBMS data (MVS and infrastructure data)

GG24-2595 Planning for IBM Remote Copy

Discusses 3990 PPRC and XRC

GF22-5114/GF22-5063 White Papers on GEOPLEX

SG24-5338 RAMAC Virtual Array: Implementing Peer-to-Peer Remote Copy