Maximizing IMS Database Availability

Rich Lewis
IBM

August 3, 2010
Session 7853

SHARE in Boston

Agenda

- Why are databases unavailable
  - We will discuss the reasons
- What can we do about it
  - We will see how we can eliminate or minimize these outages

Why are IMS databases unavailable?

- System unavailability
  - Addressed by Parallel Sysplex
- Application design
  - Applications which cannot run with concurrent updates
  - Data processed outside of IMS

Not covered in this presentation

- Reorganizations
  - For database performance
  - For database restructure
- Application design
  - Applications copy data; data updates must be quiesced
- Database recoveries
  - Preparation
    - Image copies
    - Disaster recovery preparation
  - Recovery execution
    - Full recoveries
    - Timestamp recoveries

We'll talk about these reasons and how to minimize these database outages

Reorganizations

Reorganizations

- For database performance – typical reorg
  - "Optimizes" data placement in data sets
    - Puts segments near segments from which they are chained
  - Reduces I/Os required to process the database
    - Reduces CPU utilization
- For database restructure – less typical reorg
  - Adds, deletes, moves segment types in the database
  - Modifies segment definitions
  - Changes physical characteristics
    - Block sizes
    - Randomization parameters
    - Compression
    - …

Reorganizations

- Three categories of reorganizations
  - Offline
    - Database is not available during the reorganization
  - Online – almost
    - Database is available during almost all of the reorganization
      - Short outage required
  - Online – truly
    - Database is available during all of the reorganization
      - Absolutely no outage whatsoever

Offline

Database is not available during the reorganization

Full Function (non-HALDB) Offline Reorganizations

- HD Unload, HD Reload, HISAM Unload, HISAM Reload, Prereorganization, Database Scan, Prefix Resolution, and Prefix Update

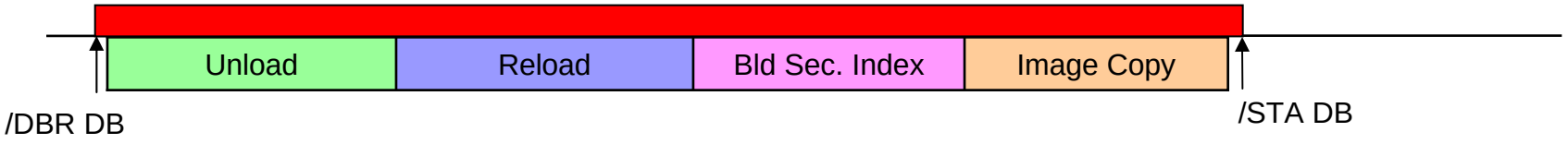
  - These utilities are supplied in the IMS product
    - They perform the reorganization functions
      - Unload, reload, update logical relationships, rebuild secondary indexes
    - They are not designed for optimum performance

  - Databases are unavailable during the reorganization process

  - These utilities may be tuned for better performance
    - More database buffers
    - More unload/reload data set and work file buffers

Full Function (non-HALDB) Offline Reorganizations

- IMS Tools: HP Unload, HP Load, Index Builder, HP Prefix Resolution, and IMS Parallel Reorganization (Renamed: IMS Database Reorganization Expert)
- These tools are not part of the IMS product
  - *They shorten the reorganization process*
  - *They simplify the management of reorganizations*
  - HP Unload, HP Load, Index Builder, HP Prefix Resolution are “high speed” replacements for standard utilities



  - IMS Parallel Reorganization does the unload, reload, etc. in parallel

HALDB Offline Reorganizations

- HALDB has advantages for offline reorganizations
  - Parallelism
    - Partitions may be reorganized in parallel
    - Smaller partitions provide shorter elapsed times
  - Elimination of much of the work
    - Secondary indexes and logical relationships do not have to be updated during the reorganization
      - They are dynamically updated when they are first used
      - Self-healing pointers
    - A subset of the partitions may be reorganized
      - Only those that are disorganized
  - Performance
    - *Much shorter outages for reorganizations*
      - *Could reduce the elapsed time by 90% or 95%!*

HALDB Offline Reorganizations

= Database Outage

- Non-HALDB reorganization

- HALDB reorganization
  - Entire database

  - Reorganize partitions in parallel

HALDB Offline Reorganizations

- IMS product utilities
  - HD Unload and HD Reload may be used
- IMS Tools
  - HP Unload and HP Load may be used
  - IMS Parallel Reorganization may be used (Renamed: IMS Database Reorganization Expert)
  - Index Builder may be used
    - Eliminates the need to heal pointers after the reorganization
  - Performance
    - *The IMS Tools further shorten the reorg process*

Online - Almost

Database is available during almost all of the reorganization

Online Reorganization Facility (ORF)

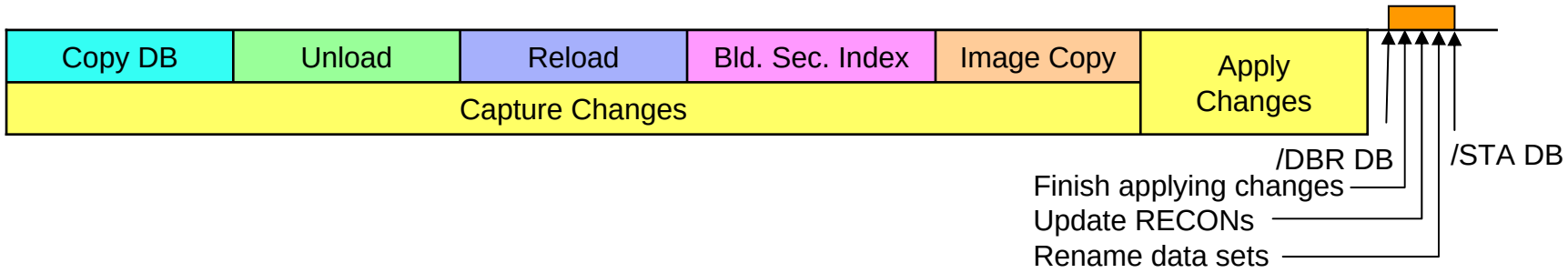
- Online Reorganization Facility (ORF) is an IMS tool
  - Reorganizes full function databases, including HALDB
  - Allows several types of restructures
- Technique
  - Uses shadow data sets (reorganizes and renames them)
- Availability
  - Allows concurrent updates from IMS online systems
  - A very short outage is required
- Restrictions
  - No external logical relationships
  - Concurrent batch (DLI or DBB) not allowed

Online Reorganization Facility (ORF) …

- Allows several types of restructures
  - Can convert non-HALDB to HALDB
  - Can modify HALDB partitioning
  - Can make many changes to DBD (add segments, modify segments,…)
- BMPs and message regions do not have to be quiesced
  - Special processing during the switch of data sets
    - BMPs “fail” after the next sync point
      - They are restarted after ORF completes
        - No operator action
    - Message regions are “held” after next sync point

Online Reorganization Facility (ORF) …

- Reorganization process with ORF
  - Database is copied
  - Copy is reorganized
  - Updates to production database are captured
  - Captured updates are applied to the copy at the end of the reorganization
  - Production database is “/DBRed”
    - RECONs are updated
    - Names of data sets are changed
  - Production database is “/STARTed”



Online - Truly

Database is available during all of the reorganization

HALDB Online Reorganization

- HALDB Online Reorganization
  - Standard part of IMS
- Absolutely no outages
- Reorganizes partition into a new set of data sets
- Supports:
  - Secondary indexes, logical relationships, data sharing, …
- Changes allowed:
  - Data set sizes may be changed
    - Recovery from an out-of-space abend (U0844) without a database outage!
  - Data set placement may be changed
  - Restrictions
    - Cannot make DBD changes
    - Cannot make partition selection changes
- IMS 11 makes significant performance improvements
  - Reduces logging, reduces CPU and shortens elapsed time

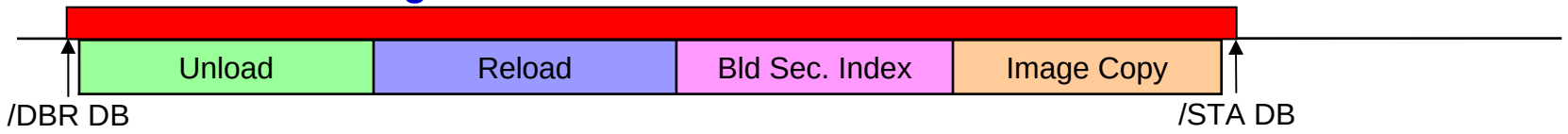
Fast Path DEDB Reorganizations

- Fast Path Data Entry Database (DEDB)
  - Database is architected for online reorganizations
    - Designed to facilitate online reorganization
  - IMS product includes online reorganization capability
    - High-Speed DEDB Direct Reorganization utility
      - Absolutely no outage
      - Reorganizes area (partition) in the same data set
        - Reads and rewrites segments within units of work (sets of CIs)
      - Supports concurrent updates
      - Supports data sharing
      - Restriction:
        - Restructures not allowed

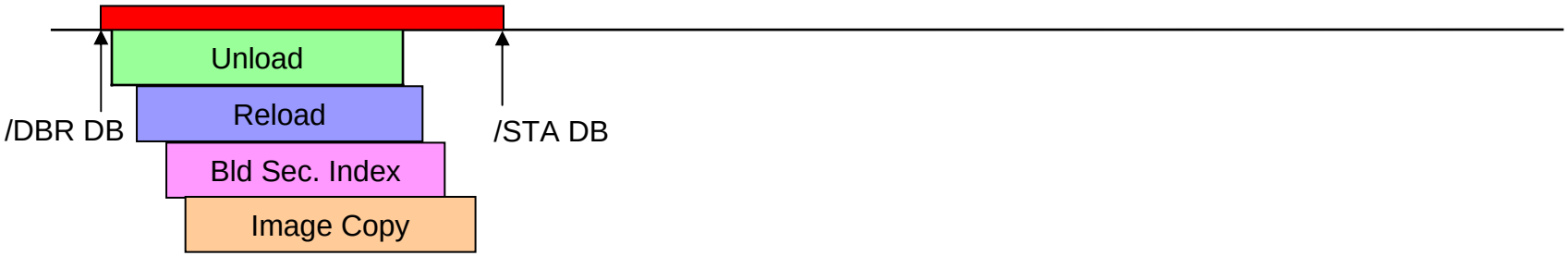
Comparison of Reorgs

= Database Outage

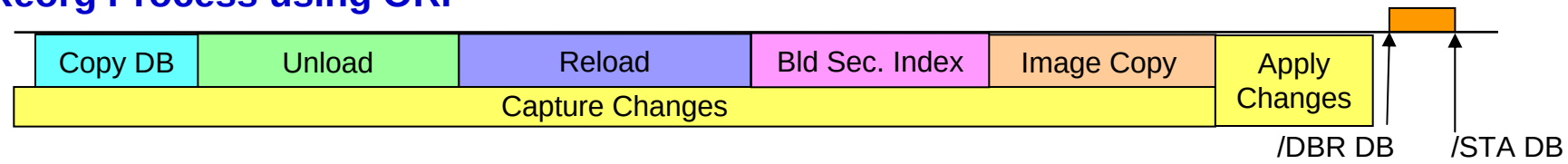
Standard Offline Reorg Process



Offline Reorg Process using Parallel Reorg (Renamed: IMS Database Reorganization Expert)



Reorg Process using ORF



True Online Reorg Process

No outage!

Reorganizations Summary

- Full function non-HALDB
  - IMS tools can significantly reduce outage time
  - ORF tool reduces outage to a very short time
  - Conversion to HALDB may provide significant benefits
- HALDB
  - Partitioning and self-healing pointers significantly reduce offline reorganization times
  - IMS tools can significantly reduce outage time for offline reorgs
  - IMS HALDB Online Reorganization has no outage
    - Part of the IMS product
- Fast Path DEDB
  - Real online reorganization with no outage
    - Part of the IMS product

Reorganizations Summary

- Database restructuring considerations
  - HALDB OLR has very limited restructuring capability
    - Only allows changes of data set properties
  - Fast Path High Speed Direct Reorg does not allow restructuring
  - ORF allows many types of restructuring
  - Some restructuring requires offline process
  - Applications may not tolerate online restructuring
    - Can they handle two forms of the data simultaneously?

Image Copies

Image Copies

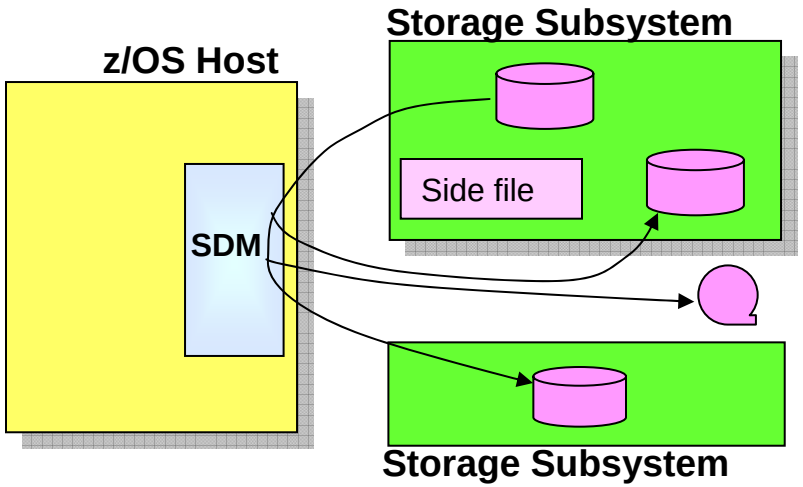
- Clean IC vs. Fuzzy IC
  - Clean IC
    - All segment images are from the same time
      - Recovery may be done without logs
    - Database is not available for update for some time
  - Fuzzy IC
    - Segment images are from different times
      - Recovery requires logs
    - Database is available for update during the IC process

Copy Technologies

- Standard copy
  - z/OS reads and writes each block
  - Clean copy does not allow any updates before the copy completes
- Concurrent copy
  - Logical copy is very short
    - Updates are not allowed during the logical copy
  - Physical copy is done by z/OS System Data Mover
    - Updates are allowed during the physical copy
    - Uses z/OS resources
- Flash copy
  - Copy time is very short
  - Copy is done entirely by the storage subsystem
    - No z/OS resources are used to copy the data sets

Concurrent Copy

- Concurrent Copy uses storage subsystem and the System Data Mover (SDM)
  - Copy may be to another subsystem, including tape or disk
  - Two phase copy
    - Logical copy – sets up side file for updates during physical copy (very quick)
      - Updates are not allowed
    - Physical copy – writes data using SDM and host resources
      - Updates are allowed



FlashCopy

- FlashCopy is a storage subsystem capability to copy volumes or data sets
  - Copy is created in the same subsystem
  - Copy is in same format (identical copy)
  - Almost instantaneous
    - Done be creating another “map” to the data
  - Subsequent updates are written to different locations for the two data sets or volumes
    - New and old “maps” are used

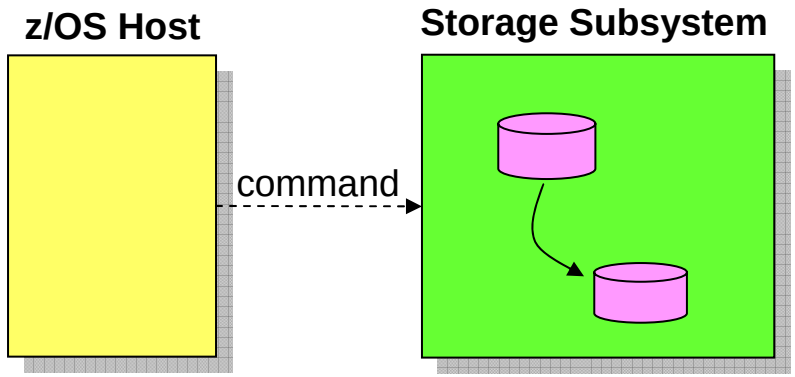


Image Copies

- IMS product provides four image copy utilities
  - Image Copy
  - Image Copy 2
  - Online Image copy
  - Fast Path DEDB HSSP image copy
- IMS tools
  - High Performance Image Copy (HPIC)

Image Copies in IMS Product

- Database Image Copy (DFSUDMP0)
  - Clean and fuzzy image copies
    - Fuzzy IC not valid for KSDSs
      - KSDS: (S)HISAM, HIDAM indexes and secondary indexes
  - Output is in “image copy” format

- Database Image Copy 2 (DFSUDMT0)
  - Can use concurrent copy capability of the storage subsystem
    - Creates output in “dump” format
  - Can use FlashCopy capability of the storage subsystem
  - Clean and fuzzy image copies
    - Valid for all types of data sets: OSAM, ESDS, and KSDS
    - Outage for clean copies is much shorter
      - Database unavailable only during the logical copy for concurrent copy or “equivalent” copy time for FlashCopy

Image Copies in IMS Product

- Online Database Image Copy (DFSUICP0)
  - Runs in online system – special BMP
    - Uses online buffer pools – performance implications
  - Does not support Fast Path databases
  - Valid for all data set types: OSAM, ESDS, KSDS
  - Output is in “image copy” format
  - Fuzzy image copies
    - Updates allowed only in the same online system
      - Limited data sharing support

Image Copies in IMS Product

- Fast Path DEDB High-Speed Sequential Processing (HSSP)
  - Fast Path DEDBs only
  - HSSP is an option for BMP application programs
    - High speed process to read the database
      - Anticipatory reads eliminate many read waits
  - HSSP has an image copy option
    - Produces an image copy while application processes the area
    - Fuzzy image copy
      - Concurrent updates are allowed in any data sharing system

Image Copy Tool

- High Performance Image Copy (HPIC)
  - Includes Concurrent Copy, FlashCopy, and SnapShot support
    - Concurrent Copy
      - Can make fuzzy image copies of all data sets, including KSDSs
        - KSDS image copies are not identical copies (they have header records, etc.)
    - FlashCopy and SnapShot Copy
      - Can make fuzzy image copies of data sets except KSDSs
      - Faster copies of database data sets

Image Copies Summary

- Clean image copies
  - Available with
    - Image Copy, Image Copy 2, and HPIC
      - Image Copy 2 and HPIC can minimize the outage

- Fuzzy image copies
  - Available for OSAM and ESDSs with
    - Image Copy, Image Copy 2, and HPIC
  - Available for KSDSs with
    - Image Copy 2 and HPIC
  - Available for DEDBs with
    - Image Copy, Image Copy 2, HSSP, and HPIC

Database Recoveries

Database Recoveries

- Database recoveries are done for three reasons
  - Full recovery
    - Due to DASD failure
    - Puts database back to its last state
    - RAID technology has eliminated the need for most of these
  - Timestamp recovery (to a previous state)
    - Usually due to an application processing error
    - Related databases must be recovered to the same time
    - Database must be recovered to a *recovery point*
      - Time when there were no uncommitted updates
        - No transactions in-flight
      - Exception for DRF
  - Disaster recovery
    - May be either full recovery or timestamp recovery
  - Most recoveries today are timestamp recoveries

Database Recoveries

- Preparing for timestamp recoveries
  - Creating recovery points
    - Database must be quiesced
      - Typically done with /DBR command
      - For data sharing, database must be quiesced on all systems at the same time
      - DBRC enforces these rules
        - Database data set cannot have an ALLOC record which spans the time
    - IMS 11 contains a database quiesce function
      - Eliminates the need to /DBR the databases
  - Creating recovery points is a significant cause of database unavailability
    - Many installations /DBR their databases once every day for this purpose
    - Outages are not caused by failures
      - They are caused to prepare for potential failures

Database Recoveries

- IMS 11 database quiesce
  - Created with UPD DB|AREA|DATAGRP … START(QUIESCE) command
  - Databases, areas and partition data sets are not closed and deallocated
  - Transactions and BMPs which access the databases do not have to be terminated
    - All uncommitted updates must be committed
    - After the commits, no updates are allowed until all other transactions and BMPs have committed their updates
      - Quiesce point is reached
  - Eliminates an “outage” and replaces it with a “delay”
    - During an “outage” applications fail if they attempt to access the database
      - Typically several minutes
    - During a “delay” applications wait for the database to become available
      - Typically a few seconds

Database Recoveries

- Eliminating recovery points
  - Database Recovery Facility (DRF) tool
    - Has Point-in-Time-Recovery (PITR) capability
      - Recovers databases to any time
        - Does not require a recovery point
        - Recovers only committed updates
        - Could recover to a DB2 determined time
    - Eliminates the need to create recovery points
      - Databases may remain available
        - Daily /DBRs or IMS 11 quiesces are not needed

Database Recoveries

- Database Recovery Facility (DRF) tool
  - Fast recovery capability
    - Recovery multiple database data sets in parallel
      - Reads inputs only once for multiple recoveries
    - Reads input logs in parallel
    - Reads image copies in parallel
    - Reads change accum data sets in parallel
    - Writes data sets sequentially after sorting log record updates
  - DRF provides a faster way to recover

Database Recoveries

- High Performance Change Accumulation (HPCA) tool
  - HPCA provides a faster way to create Change Accum data sets

- Index Builder tool
  - Creates or rebuilds indexes from indexed databases
    - Eliminates need for image copies and logs
  - Faster than database recovery for these indexes

Database Recoveries Summary

- Most outages are used to create recovery points
  - Outage may be converted to a ‘delay’ with IMS 11 DB quiesce
  - Outage may be eliminated with DRF

- Actual recoveries may be shortened with tools
  - DRF
  - HPCA
  - Index Builder

Disaster Recoveries

Disaster Recoveries

- Common techniques
  - Send clean copies to DR site
  - Send fuzzy image copies, logs, RECONs, etc. to DR site
  - Mirroring

Disaster Recoveries

- Send clean copies to DR site
  - Create common recovery point for all IMS databases
  - Probably requires common recovery point time with other data
    - DB2, VSAM, etc.
  - Regular outages to create clean copies 
  - Substantial data loss
    - Typically up to a full day 
  - Easy to recover at the DR site 
    - Just restore the image copies

Disaster Recoveries

- Send fuzzy image copies, logs, RECONs, etc. to DR site
  - Data is recovered to the end of the last log at the DR site
    - Less data loss than the “clean image copy” technique
  - Does not require outage if not data sharing ⇧
    - Recovery points are not used
  - Cannot be used with data sharing ⇩
  - More complex recovery at the DR site ⇩

(1)
- Fix up RECONs
- Full database recoveries
- Emergency restart from last log

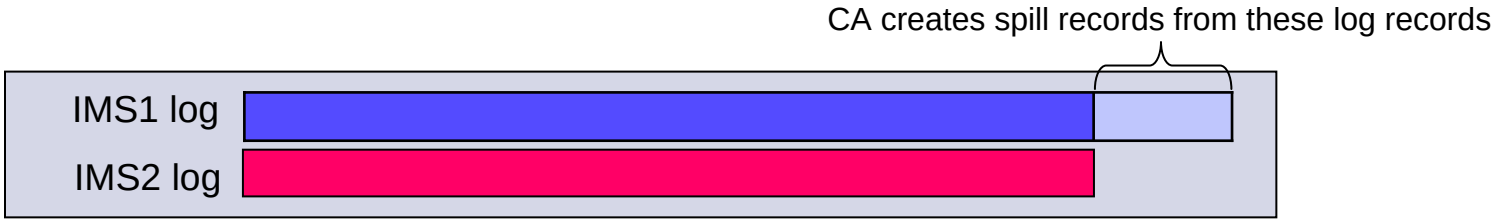
OR

(2)
- Fix up RECONs
- Timestamp recoveries with DRF
- Cold start IMS

  - May be difficult or easy to coordinate with other data
    - DB2, VSAM, etc.
      - Technique (1) recovers data to the “end” of the last IMS log ⇩
      - Technique (2) recovers to <u>any time</u> up to the “end” of the last IMS log ⇧

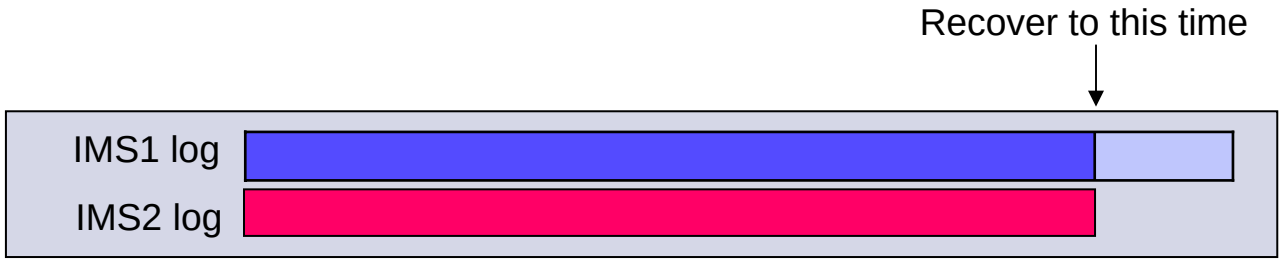
Disaster Recoveries with Data Sharing

- Timestamp recoveries without DRF require recovery points
  - Without IMS 11 database quiesce:
    - Concurrent outage required in all sharing systems
    - May extend outage for creating recovery points
  - With IMS 11 database quiesce:
    - Quiesce is coordinated across the IMSplex
- Full recoveries cannot be done
  - Logs from different systems have different end times
    - They cannot be merged with Change Accumulation
      - CA creates ‘spill records’ when one log extends past others



Disaster Recoveries with Data Sharing

- Timestamp recoveries with DRF
  - Does not require recovery point
  - May be done with fuzzy image copies
  - Benefits from parallel processing
  - May be done to any time for which all logs are available at remote site



Disaster Recoveries

- Mirroring
  - All databases and system data sets are mirrored
  - No outages for image copies 
  - Easy 
    - No recoveries at DR site
    - Works with data sharing
  - Almost no data loss 
  - Works with other subsystems too! 
    - All data is recovered to the same time
      - DB2, VSAM, …

Disaster Recoveries

- Mirroring resolves data sharing issues
  - Data at remote site matches production site at a recent time
    - Updates from all systems are reflected at the DR site
  - Eliminates need for recoveries

Application DB Copies

Application Database Copies

- Some installations copy databases from production to another system
  - Other system processes the data as it exists at the time of the copy
    - Could be ‘end of day’ data
- Typical technique
  - Image copy and restore
    - Using any utilities or tools
- FlashCopy for application database copies
  - Minimal outage for clean copy
  - One step
    - Copy is the same format as the original database

Application Database Copies

- IBM IMS Cloning Tool uses FlashCopy or other storage system facility
  - FlashCopy (IBM,EMC,HDS), TimeFinder/Snap(EMC), SnapShot (IBM,STK), DFSMSdss or FDR
  - Tool
    - Verifies that ACBLIB definitions in source and target are compatible
    - /DBRs source DBs
    - Copies databases
    - /STARTs source DBs
    - Adjusts RECON information in target
      - DELETE.DB to delete recovery information
      - INIT.DB, INIT.DBDS, INIT.PART and INIT.AREA to register/define databases

Summary

Maximizing Database Availability

- Practices to minimize database outages