B77

IMS V9 HALDB Online Reorganization

Rich Lewis

IMS
Technical Conference

Sept. 27-30, 2004

Orlando, FL

RETURN TO INDEX

HALDB Online Reorganization

- HALDB Online Reorganization (OLR) is a standard part of IMS V9 DB
  - Not a feature, product, tool, etc.
- Benefits
  - PHDAM and PHIDAM databases are reorganized
  - 100% availability of database during reorganization
    - Zero outages
    - Applications are unaffected
      - They never get data unavailable conditions
  - Full integrity and recoverability are maintained
  - Eliminates database outages for reorganizations

HALDB Online Reorganization Overview

- Environments
  - Runs in TM/DB or DBCTL system
    - Executes in DLISAS address space
  - Concurrent online and data sharing updates are allowed
  - XRF and RSR are supported
- Recoverability
  - System, IMS, or media failures
  - DBRC support, standard recovery utilities, and DRF
- Performance
  - External parameter for pacing

HALDB Online Reorganization Overview

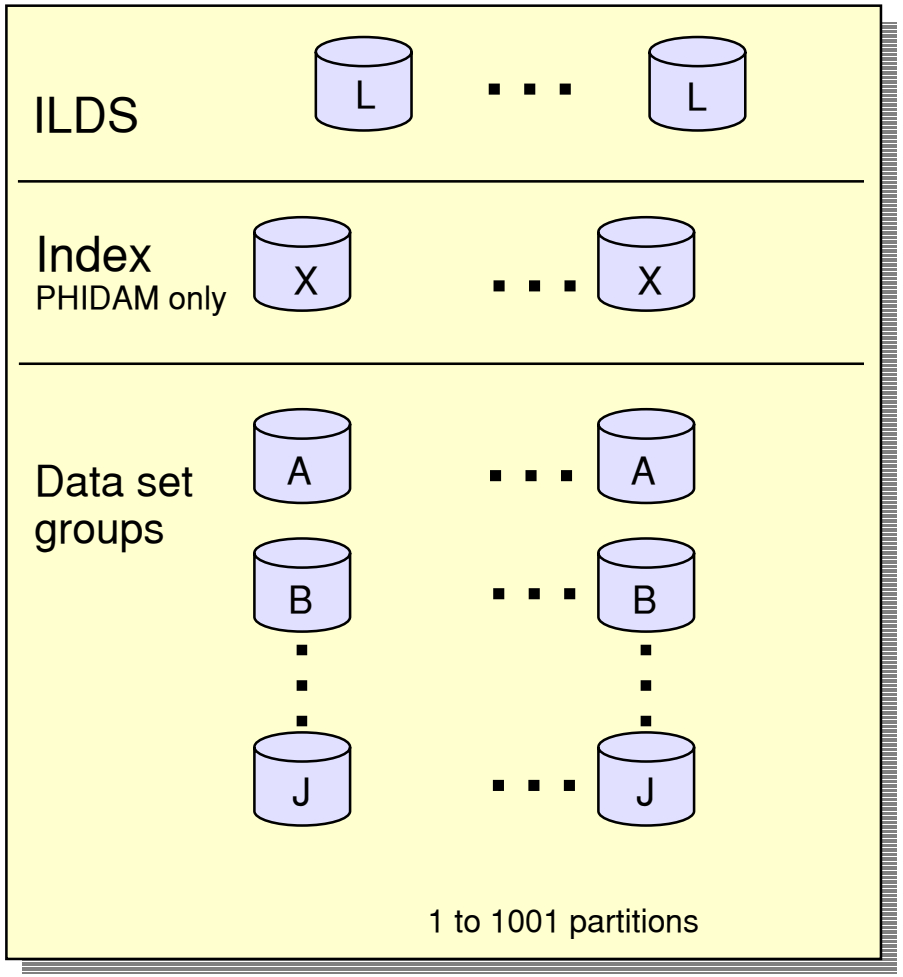
- HALDB PHDAM and PHIDAM only
  - Reorganize by partition
    - PHDAM data component
    - PHIDAM data component and primary index
- Secondary indexes and logical relationships
  - Database with secondary indexes can be reorganized
    - But secondary index (PSINDEX) CANNOT be reorganized
  - Database with logical relationships can be reorganized
  - ILDS (ILEs) updated with new target RBAs
- Restrictions
  - No DBD changes (DBDS space allocation changes are OK)

HALDB Online Reorganization Technique

- Online reorganization (OLR) is into new “partner” data sets
  - A-J and X data sets alternate with M-V and Y data sets
  - Only one ILDS (L) per partition
- Both sets of data sets are used during OLR
- At end of OLR, old data sets may be discarded
- 100% availability of database during the reorganization
  - No outages
  - No data set renames

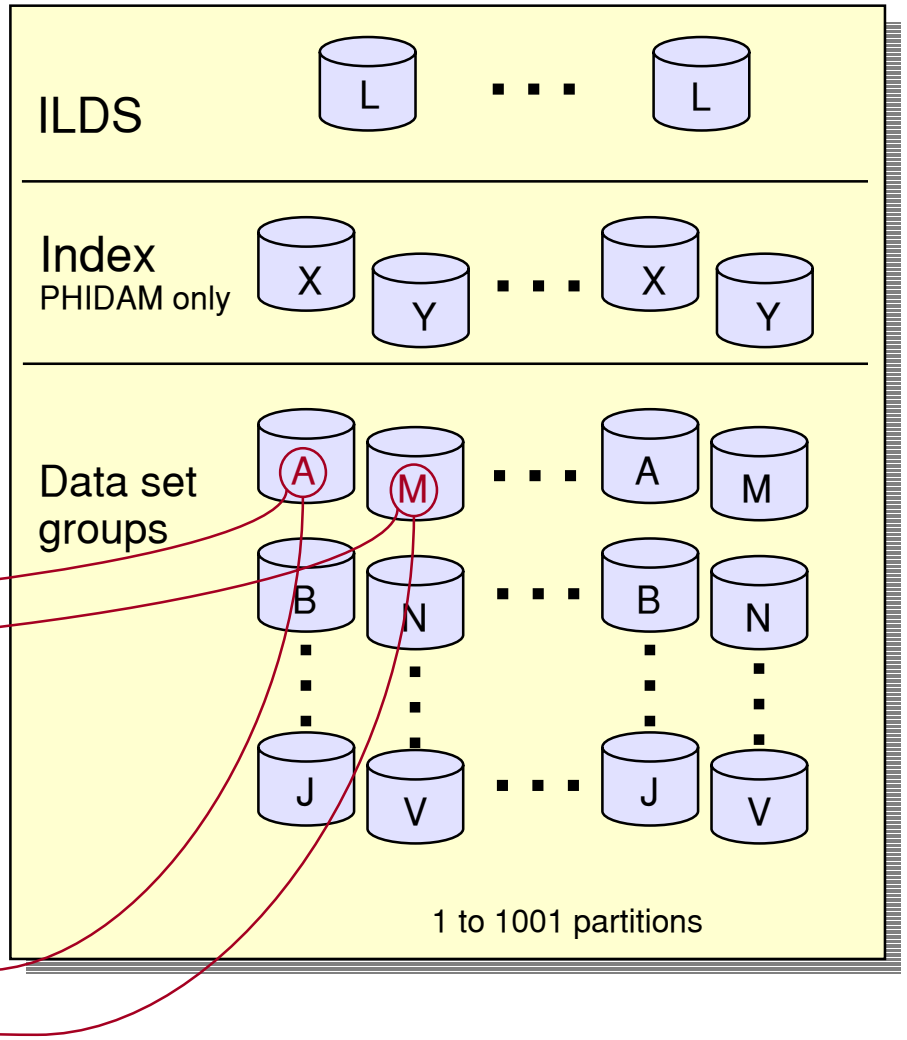
HALDB Naming Conventions

- DDNAMEs
  - Partition name and data set letter
    - Partition name: DJXK21
    - DDNAMEs:
      - DJXK21L, DJXK21X, DJXK21A, DJXK21B,…
- Data set names
  - Data set name prefix, data set letter, and partition id
    - DSN prefix: IMSP.DB.DJXAB
    - Partition id: 00001
    - Data set names:
      - IMSP.DB.DJXAB.L00001
      - IMSP.DB.DJXAB.X00001
      - IMSP.DB.DJXAB.A00001
      - IMSP.DB.DJXAB.B00001
      - …



Partner Data Sets

- Index and data set group data sets have partners
  - Each X data set has a Y partner
  - Each A data set has an M partner
  - Each B data set has an N partner
  - …
- DDNAMEs differ by the letter
  - For example:
    - DJXK21A
    - DJXK21M
- Data set names differ by the letter
  - For example:
    - IMSP.DB.DJXAB.A00001
    - IMSP.DB.DJXAB.M00001

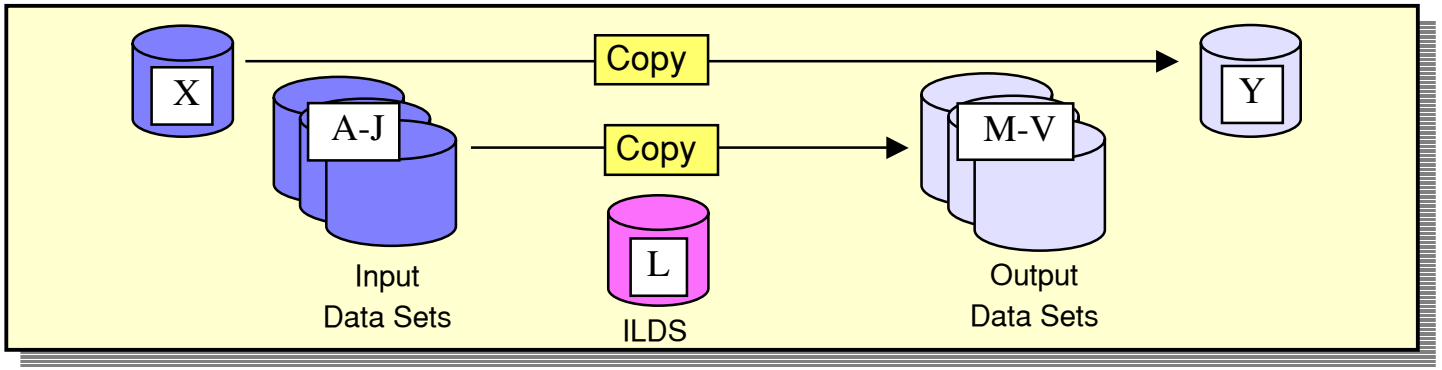


Terminology

- Before or after reorganization
  - Active data sets (either A-J, X or M-V, Y))
    - Data sets being accessed by applications
  - Inactive data sets
    - Data sets not being accessed by applications
- During reorganization
  - All data sets (A-J, X and M-V, Y) are active data sets
  - Input data set: Contains unreorganized data
    - Includes both active and inactive data
  - Output data set: Contains reorganized data
  - Cursor
    - Dividing line between active data and inactive data
    - Only used while reorganization in progress or suspended

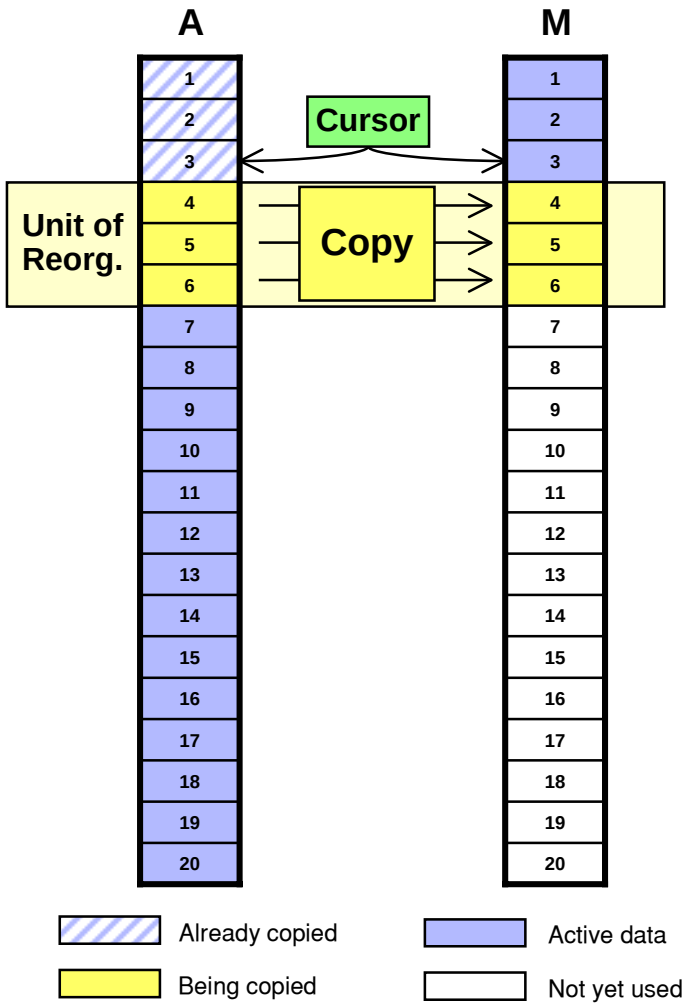
Reorganization

- Reorganize by copying segments
  - Read segments from one set of HALDB data sets (e.g. A-J, X)
  - Write (insert) segments to another set (e.g. M-V, Y)
    - Update ILDS for secondary index and logical relationship targets
  - Use locking protocols to provide concurrent access integrity
  - Log inserts for recoverability
  - Use cursor to identify which "set" to use to access a database record
    - Database records before cursor, use output data sets
    - Database records after cursor, use input data sets



Copying Records During Reorganization

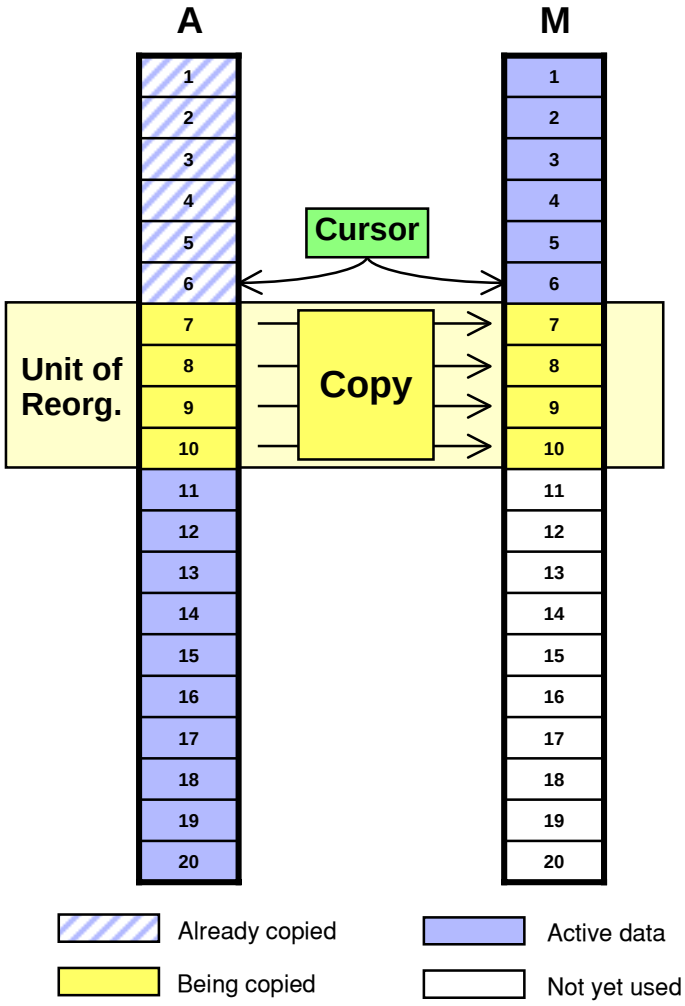
- Unit of Reorganization (UOR) is a set of database records
  - Records are copied from input to output data sets
  - Records in UOR are locked while being copied
  - At end of copy for UOR, the locks are released
  - Number of records in UOR is dynamically adjusted
    - Algorithm limits time taken, bytes copied, and locks held during copy



Application Access During Online Reorganization

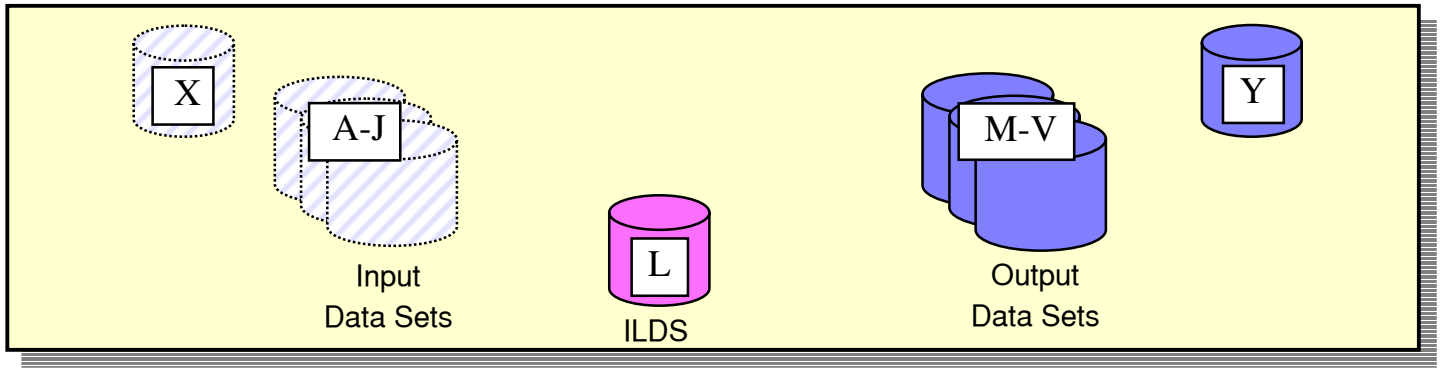
- Cursor points to last committed reorganized record
  - PHDAM RAP RBA
  - PHIDAM root key
- Data set used is based on cursor value
  - Cursor on record 6
  - Access Record 5:
    - Access from M data set
  - Access Record 14:
    - Access from A data set
  - Access Record 9:
    - Wait for lock,
      - then access from M data set

* Access includes gets and updates



Completion of Reorganization

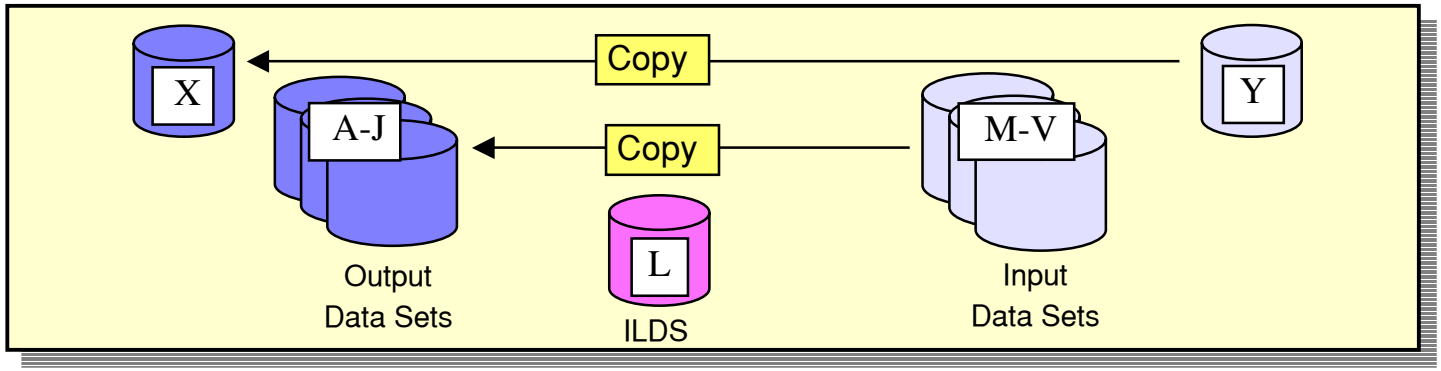
- When OLR completes
  - A-J,X becomes the "inactive" set - may be deleted
  - M-V,Y becomes the "active" set
- Cursor reset to inactive
- ILDS (ILEs) updated during reorganization



Next Reorganization

- Next reorganization
  - Reorganize from M-V,Y to A-J,X
  - A-J, X data sets may be reused

  Or
  - A-J, X data sets may be reallocated



Setting Up Online Reorganization

- DBRC is used to set online reorganization capability for a database

```
INIT.DB DBD(HALDB_master) OLRCAP|OLRNOCAP
CHANGE.DB DBD(HALDB_master) OLRCAP|OLRNOCAP
```

  - OLRCAP allows online reorganization for partitions of the database

Output Data Set Creation

- Output data set allocation options
  - Preallocation by user
  - Automatic allocation by OLR
    - Invoked for each data set which is not cataloged
      - Invoked on data set by data set basis
- Why preallocate?
  - Want to allocate on specific volume
  - Change space allocation
    - Blocks/CIs
      - Primary and secondary allocations
    - For PHIDAM Primary Index
      - Free space percentage

Output Data Set Creation

- Automatic output data set creation
  - Space is equivalent to existing input data set
    - Requested as a number of OSAM blocks or VSAM records
  - SMS-managed
    - Same storage class as input data set
    - Same number of volumes as input data set
    - With guaranteed space attribute, primary space allocation is taken on all volumes
  - Non-SMS, OSAM
    - UNIT=SYSALLDA is used (storage or public volume)
    - If input is multivolume data set, output data set is not created
  - Non-SMS, VSAM
    - Data set is allocated on the same volume(s) as input data set

Starting Online Reorganization

- Command to initiate OLR
  - OM command (type-2 command):

    ```
    INIT OLREORG NAME(partname1, partname2,...)
    ```

  - Classic command (type-1 command):

    ```
    /INIT OLREORG NAME(partname1)
    ```

  - Command parameters:
    - Delete input data sets at completion of reorganization

      ```
      OPTION(DEL|NODEL)
      ```

    - Set rate of execution

      ```
      SET(RATE(100|nn)
      ```

Rate Parameter

- RATE parameter on INIT
  - RATE parameter determines how fast the reorganization runs
    - RATE(100) - runs at maximum speed
    - RATE(nn) - online reorganization waits after each commit so that average speed of reorganization is nn% of maximum speed
  - Examples:
    - If RATE(50), after each commit reorganization waits for the time that the last interval took
      - Possibly, run 1 second, wait 1 second, run 1 second, wait 1 second,...
    - If RATE(25), after each commit reorganization waits for 3 times as long as the last interval took
      - Possibly, run 1 second, wait 3 seconds, run 1 second, wait 3 seconds,...

Modifying Reorganization in Progress

- Command to modify OLR in progress
  - OM command (type-2 command):

    `UPD OLREORG NAME(*|partname1, partname2,...)`
  - Classic command (type-1 command):

    `/UPD OLREORG NAME(partname1)`
  - Command parameters:
    - Change delete option for input data sets

      `OPTION(DEL|NODEL)`
    - Change rate of execution

      `SET(RATE(100|nn)`

Commands to Show Status of Reorganization

- QRY command (type-2) example:

```
QRY OLREORG NAME(PVHDJ5A) SHOW(ALL)
```

  - Response:

```
Partition MbrName          CC   RATE   BYTES MOVED    STATUS
PVHDJ5A   IMS2            0    100         1256356   RUNNING
```

- /DIS command (type-1) example:

```
/DIS DB OLR
```

  - Response:

```
DATABASE   PART       RATE          BYTES   STATUS
DBHDOK01   PDHDOKA      50         2186776   RUNNING
```

Logging By Online Reorganization

- Log records written
  - Scheduling (x'08')
  - Termination (x'07')
  - UOR sync point (x'3730')
    - For each UOR
  - UOR statistics (x'2950')
    - For each UOR
  - Database change (x'50')
    - For all output data in the partition
      - This will be voluminous!

Logging By Online Reorganization

- UOR statistics log record (x'2950')
  - Written for each UOR
  - Data:
    - Total segments moved before this UOR
    - Total bytes moved before this UOR
    - Roots moved in UOR
    - Segments moved in UOR
    - Bytes moved in UOR
    - Locks held by UOR
    - Start time of UOR
    - Execution time (elapsed time) of UOR
    - Time interval waited before this UOR (due to RATE parameter)

Suspending and Restarting Online Reorganization

- Reorganization may be suspended
  - Commands:
    - TERM command (type-2) example:

      `TERM OLREORG NAME(PVHDJ5A)`
    - /TERM command (type-1) example:

      `/TERM OLREORG NAME(PVHDJ5A)`
  - Input and output data sets remain active
    - Cursor remains active

- Suspended reorganization may be restarted
  - INIT and /INIT command will restart the reorganization
    - Restarts from the point of the cursor
  - Restart may be on the same IMS system or another IMS system

IMS Normal Termination and Restart with OLR Active

- If OLR is running when IMS is shutdown
  - /CHE FREEZE or /CHE DUMPQ
    - OLR is terminated at next commit
      - Cursor remains active
      - Ownership of OLR by this IMS is not relinquished
  - /CHE PURGE
    - Waits for OLR to complete
- When IMS is restarted after termination with OLR active
  - /NRE
    - Authorizes, allocates, and opens all input and output data sets
    - Resumes OLR automatically

Image Copies

- Active data set is copied
  - Control statement may specify A-J or M-V DDNAME
    - Image copy utilities determine which data set to copy
      - Copies active data set even when inactive partner is specified
  - Dynamic allocation allocates the active data set
- Any image copy utility may be used:
  - Image Copy
  - Image Copy 2
  - Online Image Copy
- Image Copy is not allowed while cursor is active
  - Online reorganization is active

  or
  - Online reorganization is suspended

Change Accumulation

- DBRC places partner data sets in the same change accum group
  - A and M data sets for a partition are in the same group
  - B and N data sets for a partition are in the same group
  - C and O data sets for a partition are in the same group
  - …
- GENJCL.CA treats start of OLR as purge time for output data sets
  - Log records from times before start of OLR are not used in later recoveries
- Change Accumulation utility is unchanged for online reorganization
  - Accumulates changes for all data sets specified on its control statements
    - GENJCL.CA generates the correct control statements

Database Recovery

- Database Recovery (DFSURDB0)
  - Recovers A-J or M-V data set
  - Full recovery
    - Allowed at any time
    - May be to time when OLR was active or suspended
  - Timestamp recovery
    - Not allowed to time when OLR was active or suspended
      - DRF tool may be used for this type of timestamp recovery
  - Recovery of data set reorganized by OLR (output data set)
    - Does not require image copy
      - Log only recovery is valid

Database Recovery Facility (DRF)

- Database Recovery Facility (IBM Tool)
  - Recovers active data sets
    - Understands which data sets to recover
  - Full recovery
    - Allowed at any time
  - Timestamp recovery
    - Allowed to any time
      - Includes PITR (point-in-time) capability

Performance Considerations

- OSAM sequential buffering may be used
  - Recommended
- Logging may affect performance
  - All data is logged when moved
  - A few additional log records
- Buffer pool contention
  - Partner data sets use the same buffer pool
    - Appropriate for times when reorganization is not running
    - Could cause buffer contention during reorganization

Performance Considerations

- Lock contention
  - Should be minimal
    - OLR has dynamic algorithm to limit the time that locks are held
  - OLR rarely causes a deadlock
    - Asks for database record locks conditionally
      - If lock is not available, the UOR is shortened
    - OLR is always the victim in its deadlocks
      - Application continues
      - OLR is dynamically backed out
        - Only the current UOR is backed out
      - OLR is automatically restarted at the current cursor position

Performance Considerations

- Online reorganization runs in DL/I address space
  - Each reorganization uses one of 10 database TCBs
    - Same TCBs that are used for allocation and open/close/EOV processing
- Online reorganization may run on any data sharing IMS system
  - Some installations may choose to dedicate an IMS to OLR
    - Buffer pool definitions may be tuned for OLR
    - Avoids buffer contention
    - Avoids logging contention
    - Limits the number of data sets with updates on the log
      - Logs are not required for change accum or recovery of other data sets

HALDB Online Reorganization Summary

- HALDB Online Reorganization is included in IMS V9 DB
  - Not a feature, product, tool, etc.

It's free!

- Benefits
  - Fast and efficient reorganizations
  - Full integrity and recoverability are maintained
  - Eliminates database outages for reorganizations

Full database availability during all of the reorganization process!

HALDB Online Reorganization

- Be ready for HALDB online reorganization when it becomes available

- Migrate your databases to HALDB now!
  - Any full function database may be migrated
    - No application changes required *

* Very small exception:
  Processing a secondary index as a database when using /SX and secondary index contains duplicate data

Easily handled

RETURN TO INDEX