Care and Feeding of Oracle Rdb Hot Standby

Paul Mead / Norman Lastovica
Oracle New England Development Center

Copyright 2001, 2003 Oracle Corporation
Overview

- What Hot Standby provides
- Basic architecture of Hot Standby components
- Considerations and trade-offs for implementing Hot Standby
Hot Standby Provides Disaster Tolerance

- You can implement an alternate (standby) database providing disaster tolerance for your application
- Standby database is physically equivalent to original (master) database, including journals
- Replication is automatic and efficient -- low overhead
- Standby database can be used for read access
  - Read Committed, i.e., no snapshots
  - DDL changes not visible
  - Contention between LRS and query processes
  - Logminer is better
Components

- Hot Standby is essentially continuous RMU/RECOVER
- Log Catch-up Server (LCS) does initial synchronization then monitors replication
- AIJ Log Servers (ALS) transfer journal contents to standby Log Roll-forward Server (LRS) which applies journal entries to standby database
Basic Server Model Components

**Master Database**
- ALS
- LCS
- .AIJ

**Standby Database**
- LRS
- .AIJ
- .RDA
Catch-up Phase

- When replication is started the standby database must “catch up” with the master
- The longer it has been since standby was setup (or replication previously terminated) the longer it takes to catch up
Basic Server Model
Catch-up Phase

Master Database
- ALS
- LCS
- .AIJ

Standby Database
- LRS
- .AIJ
- .RDA
Basic Server Model
Normal Operation

Master Database

ALS

LCS

.AIJ

Standby Database

LRS

.AIJ

.RDA
Basic Server Model
Additional Processes

Master Database
- ALS
- LCS
- .AIJ
- DBR

Standby Database
- LRS
- .AIJ
- .RDA
- .AIJSERVER
- AIJSERVER
- AIJSERVER
- AIJSERVER
Setting Up Hot Standby

- Restore database on standby and start replication
- Standby database cannot be modified
- Journal disks must have same cluster size or **AIJSIGNATURE** error will be returned at startup
- Can use TCP/IP or DECnet as transport
- Make sure network objects configured
- Grant WORLD and GROUP privileges to RDMAIJnn network object account
Setting Up Hot Standby

- Use concealed logical for disk devices if you might add storage areas or journals to the master
- Allow for storage area extensions
- RMU/REPLICATE AFTER START qualifiers ignored if you don’t specify /MASTER or /STANDBY
- Row cache must be disabled on standby RMU/OPEN/ROW=DISABLED
- Use circular journals and automatic ALS
- Use manual open databases
- More details in documentation
Setting Up…
Simple Example

$ RMU /OPEN MFP.RDB
$ RMU /SERVER BACKUP_JOURNAL SUSPEND MFP.RDB
$ RMU /SHOW AFTER MFP.RDB /OUTPUT=OPT.DAT
$ RMU /BACKUP /ONLINE /QUIET_POINT MFP.RDB MFS.RBF

$ RMU /RESTORE /AIJ_OPT=OPT.DAT /ROOT=MFS.RDB MFS.RBF
$ RMU /OPEN MFS.RDB
$ RMU /REPLICATE AFTER_JOURNAL START MFS.RDB -
   /MASTER_ROOT=MFP.RDB /ONLINE /WAIT

$ RMU /REPLICATE AFTER_JOURNAL START MFP.RDB -
   /STANDBY_ROOT=MFS.RDB /WAIT
…On the Master

$ RMU /SHOW SYSTEM

... 
database DKA400:[RDBDEMO]MFP.RDB;1
- current after-image journal file is
  DISK$DEMO1:[RDBDEMO]J3.AIJ;1
- hot standby "Log Catch-Up Server" is active
- hot standby "Log Shipping Server" is active
- AIJ Log Server is active
- 2 active database servers on this node
  - 20A0097D:1 - AIJ Log server
    - image ...RDMALS711.EXE
  - 20A00980:1 - AIJ Log Catch-Up server
    - image ...RDMLOG711.EXE
- standby database
  "DKA400:[RDBDEMO_HS]MFS.RDB"
...On the Standby

$ RMU /SHOW SYSTEM

... 
database DKA300:[RDBDEMO_HS]MFS.RDB;1
  * database is opened by an operator
  - current after-image journal file is
    DISK$DEMO3:[RDBDEMO_HS]J3.AIJ;1
  - hot standby "Log Roll-Forward Server" is active
  - 1 active database server on this node
    - 20A0097F:1 - AIJ Log Roll-Forward server
      - image ...RDMLRS711.EXE
    - master database "DKA400:[RDBDEMO]MFP.RDB"
Standby Configuration

• Standby system is intended to be used for fail-over
  – Should have similar performance characteristics as master

• Under-configured standby system can affect performance of master database
Synchronization Modes

• Set based on business need
• Modes define when ALS process may proceed (and thus when committing process may proceed)
  – Cold: send and don’t wait for confirmation of receipt (default)
  – Warm: wait for confirmation of receipt of message
  – Hot: wait for confirmation of write to standby journal
  – Commit: wait for confirmation transaction(s) applied on standby
• “Higher” modes mean
  – Less chance updates lost if both master and standby go down simultaneously
  – More potential delays on master
  – Less possible “lag” between master and standby
Synchronization Modes
Fail Over Scenarios

Master Database

ALS

Standby Database

Commit

LRS

Ack

.AIJ

.RDA
Checkpointing

- Larger checkpoint intervals reduces write I/O in LRS
- Checkpointing is most critical during AIJ switchover, and is done automatically when switchover occurs
- Make large, but not so large that Hot Standby server checkpoints are “too old”
- Master database should be open on “active” nodes only
Governor

- Adjusts synchronization mode if work backlog changes by factor of 10 since last checkpoint
- Unprocessed messages and buffered changes counted in backlog
- Checked every time ALS sends checkpoint request
- Enabled by default
- Gets invoked (even if disabled) if all LRS buffers filled
- Can increment up to “hot”
AIJ Switchover and Backups

- When a journal has been backed up and initialized on the master then it also needs to be initialized on the standby
- Spread AIJ files around disks and controllers to avoid contention when initialization is being done
- Journal initialization can be throttled by using the `RDM$BIND_AIJ_INITIALIZATION_IO_COUNT` and `RDM$BIND_AIJ_INITIALIZATION_IO_SIZE` logicals
ABS

- Documentation “recommends” ABS
- Not required
Performance Considerations

• Setting up LRS for fast and efficient I/O is critical
• Start with disk configuration and buffer count
  – Spread hot areas & journals over controllers/spindles
• CPU may ultimately be limiting factor in LRS’ ability to keep up with master database
• Avoid activities that cause long pauses in LRS processing of incoming messages
  – Bulk loads, CREATE INDEX, etc.
  – Storage area extensions
APF and ABW

- Most critical aspect of LRS performance
- The LRS sets its own APF and ABW parameters
- APF defaults to 50% of the buffer pool
- ABW defaults to 20% of the buffer pool, or 10 buffers, whichever is greater
- Can be manipulated from SHOW STATISTICS dashboard
Number of LRS Buffers

- More buffers means less I/O since buffers stay in buffer pool longer
- More buffers means it takes longer for the LRS to checkpoint since there are more modified buffers to write
- Long checkpoints could cause pauses on master and/or trigger governor to increase synchronization mode
Number of LRS Buffers

- Write I/Os
- CKPT time

Buffers

- 0
- 0.1
- 0.2
- 0.3
- 0.4
- 0.5
- 0.6
- 0.7
- 0.8

Write I/Os

- 0
- 1000
- 2000
- 3000
- 4000
- 5000
- 6000
- 7000
- 8000

Buffers

- 250
- 500
- 750
- 1000
- 1250
- 1500
- 1750
- 2000
Global or Local Buffers?

- Before 7.0.6.5 and 7.1.0.4 /BUFFERS ignored for local buffers
- Global Buffers provides mechanism to keep pages in memory while limiting the number of modified buffers held by LRS
- Global buffer pool acts as cache
- Use large enough “user limit” to allow LRS to keep that many modified buffers in its buffer pool
- Or, enable XFC (VMS 7.3-1 or later) and use local buffers
Transaction Sizes

- “Small” to “medium” sized transactions best
- Transaction data buffered by LRS until commit or rollback received
- LRS is not processing incoming data when applying transactions
- Large transactions may overflow to disk temporary storage
- LRS workfiles can be directed to location specified by `RDM$BIND_AIJ_WORK_FILE` logical
Adding Storage Areas or Journals Online

• Adding a storage area or journal can take a long time – LRS processing stops while doing these activities
• Best to do this type of activity during “quiet” times
Offline, Not-Journaled Operations

- Various database operations are not journaled and thus cannot be replicated in the standby database
- Expect to rebuild the standby database if those changes must be reflected in the standby database
- See list of operations listed in Hot Standby documentation
  - CREATE INDEX listed in docs but is OK.
Restarting Replication After Long Shutdown

- AIJ data extant at shutdown must still be available at restart
- Journals are normally backed up immediately after switchover
- HOTSEQBCK error returned if journal missing
- A manual RMU/RECOVER can be done to catchup the standby database
- If a large amount of journal data has been created since shutdown it may be faster to restore standby from newer backup
Suspend ABS When Hot Standby not Active

• **RMU/SERVER BACKUP_JOURNAL/LOG SUSPEND**

• Define **RDM$BIND_HOT_ABS_SUSPEND_SHUTDOWN** to inhibit ABS when Hot Standby is shutdown normally

• Define **RDM$BIND_HOT_ABS_SUSPEND** if you only want to suspend backups when there was a Hot Standby failure

• Be aware that if all the journals fill then master database will be shutdown
Failover and Failback

- Failover and failback are manual operations.
- Make sure LRS shutdown normally before failover.
- Backup journals before “swapping” databases (making the standby the primary db or “new master” and vice versa).
- Must backup original master after failback complete.
## Statistics

Rate: 3.00 Seconds           Hot Standby Statistics         Elapsed: 00:42:27.70
Page: 1 of 1 $2$DKA101:[RDB_RANDOM.V71M_RND_HOT_STANDBY]RNDB.RDB;1 Mode: Online

<table>
<thead>
<tr>
<th></th>
<th>rate.per.second</th>
<th>total</th>
<th>average</th>
<th>count</th>
<th>per.trans</th>
</tr>
</thead>
<tbody>
<tr>
<td>statistic............</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>name..................</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sync mode change</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>0.0</td>
</tr>
</tbody>
</table>
### Debug Logging

```bash
$ DEFINE /SYSTEM RDM$BIND_KODA_TRACE {ACHMY}
```

<table>
<thead>
<tr>
<th>Server Process</th>
<th>Define on…</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, AIJSERVER</td>
<td>Standby</td>
</tr>
<tr>
<td>C, LCS</td>
<td>Master</td>
</tr>
<tr>
<td>H, ALS</td>
<td>Master</td>
</tr>
<tr>
<td>M, ABS</td>
<td>Master or Standby</td>
</tr>
<tr>
<td>Y, LRS</td>
<td>Standby</td>
</tr>
</tbody>
</table>
For More Information

• Guide to Hot Standby Databases
• www.oracle.com/rdb
• metalink.oracle.com
• www.openvms.compaq.com
• Paul.Mead@oracle.com
QUESTIONS & ANSWERS