High Availability
and
Oracle Parallel Fail Safe

James Viscusi
Oracle Corporation
The Challenges of High Availability Data Servers

• Many high availability products from Oracle
• To use them you must....
  – architect / implement a solution
  – perform extensive testing to verify...
    – components coexist
    – failure scenarios
  – maintain the solution across upgrades
• Customers typically deploy one of two solutions:
  – Single Instance Oracle with Cold Failover
  – Oracle Parallel Server
Single Instance Oracle with Cold Failover

Active

Oracle Instance 1

HA Cluster Manager

Server

Database for Oracle Instance 1

Application 2 Storage

Active with other app.

Application 2

HA Cluster Manager

Server

failover

Exclusive Access
Oracle Parallel Server Architecture

Active

Oracle8/Oracle8i Parallel Server Instance 1

Cluster Manager

Server

Shared Access

Shared Database for Oracle Instances 1 and 2

Active

Oracle8/Oracle8i Parallel Server Instance 2

Cluster Manager

Server

Shared Access
Oracle Parallel Fail Safe

• Let Oracle manage this complexity
  – Oracle Parallel Fail Safe is a pre-packaged, pre-configured, pre-tested high availability solution
  – Built upon Oracle8i Parallel Server
  – Integrated with high availability products from UNIX vendors
  – Compatible with off-the-shelf and custom applications
Oracle Parallel Fail Safe

• Special Primary/Secondary Parallel Server configuration optimized for High Availability
  – Application services provided by the primary instance
  – Secondary instance ready to take over in the event of a failure

• Tightly integrated with cluster manager for fast detection and response

• Performs like cold failover from application perspective, but recovers quickly because it’s Parallel Server
Benefits of Oracle Parallel Fail Safe

- Complete fault detection of all types of hard and soft failures including those outside the instance
- Intelligent agents take actions when faults detected
  - automatically restart failed instance when the secondary instance is down for maintenance
  - automatically restart failed listener processes
- Faster resumption after failure
  - send automatic notification of instance up or down to interrupt sessions to reconnect/resume
  - automatically move IP’s to eliminate reconnection delay
Benefits of Oracle Parallel Fail Safe

- Failover agents automatically capture diagnostics information
- Better performance after failover through warming of the library cache on the secondary instance
- Entire hardware and software stack is fully tested under a suite of test scenarios to prove all components are compatible
- Easy to setup and install
  - Includes templates and sample configurations
Oracle Parallel Fail Safe
New 8.1.7 Features

• OPFS 8.0.6/8.1.6 has already shipped on HP. New for 8.1.7…
  – Generic release supporting more UNIX platforms
  – Improved failover performance
  – Fast, event-based detection of instance failure.
  – Secondary IP addresses float, helping to minimize TCP/IP timeouts
  – NLS support
  – 3-node N+1 configuration
Oracle Parallel Fail Safe
New 8.1.7 Features

• Continued…
  – Support for Parallel Server Primary/Secondary configuration
    – Primary/secondary roles enforce access
    – Locks mastered locally to remove runtime overhead
  – Improved user interface
  – Uptime compliance is recorded
  – Better error and message logging
  – Improved oracle_ping timeouts
  – Improved PFS installer
OPFS 8.1.7 Concepts

*Oracle Service*

- Oracle Service is a logical way to represent an application:
  - eg. CRM Oracle Service and ERP Oracle Service
- Parallel Fail Safe uses Oracle Service as the unit of control
  - can boot, move, halt, measure, upgrade, and report
  - can contain planned and unplanned outages
  - can phase planned outages and off-peak processing
OPFS 8.1.7 Concepts

*Oracle Instance Roles*

- OPFS uses the built in “role” functionality in OPS to set the primary and secondary role for the instances.
- OPFS uses Net8 to connect to the appropriate role, even when clients accidentally attempt connection to the wrong instance.
- Advantages:
  - Inadvertent connections to an instance are rejected and forwarded to the correct instance.
  - Low runtime overhead
    - All locks are mastered at the primary instance.
OPFS 8.1.7 Concepts

*Oracle Instance Roles*

- Values for instance_role seen in v$instance when active_instance_count=1:

<table>
<thead>
<tr>
<th>INSTANCE_ROLE</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRIMARY_INSTANCE</td>
<td>First OPS instance to mount. This instance acquires the primary_instance role and masters all the locks locally</td>
</tr>
<tr>
<td>SECONDARY_INSTANCE</td>
<td>Second OPS instance to mount. If the primary instance halts, this instance automatically changes roles from secondary to primary</td>
</tr>
<tr>
<td>UNKNOWN</td>
<td>An instance shows this role when it is in nomount state</td>
</tr>
</tbody>
</table>
OPFS 8.1.7 Concepts
Oracle Instance Roles in NET8

- Values for instance_role in tnsnames.ora file:

<table>
<thead>
<tr>
<th>INSTANCE_ROLE</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRIMARY</td>
<td>Specifies to the listener for sessions to connect only to the instance that has the primary role</td>
</tr>
<tr>
<td>SECONDARY</td>
<td>Specifies to the listener for sessions to connect only to the instance that has the secondary role</td>
</tr>
<tr>
<td>ANY</td>
<td>This role is used to allow sessions to connect to an instance regardless of the current role. (Instance_name specifies which instance.)</td>
</tr>
</tbody>
</table>
OPFS 8.1.7 Concepts

*Preferred Primary/Secondary Node*

- **Preferred Primary Node**
  - The node which the user has designated as primary. At initial startup, this is the node where OPFS will attempt to start the primary instance.

- **Preferred Secondary Node**
  - The node which the user has designated as secondary. At initial startup, this is the node where OPFS will attempt to start the secondary instance.
OPFS 8.1.7 Concepts

*Oracle Services and Roles*

- How it works:
  1. OPFS mounts the preferred primary instance first acquiring the service and primary role
  2. OPFS mounts the preferred secondary instance second acquiring the service and secondary role
  3. At failover, the service and primary role register at the secondary instance
  4. After restore, the roles are reversed

- Applications connect to a primary instance, to a secondary instance, or to a named instance
OPFS 8.1.7 Concepts

Oracle PFS Packs

• Two Oracle PFS Pack per Oracle Service
  – one supporting primary instance
  – one supporting secondary instance

• Oracle PFS Pack has the following components:
  – Resources
  – Pack profile
    – (run script, halt script, failover policy, package switching)
  – Monitors

• The behavior of the Oracle PFS Pack is determined by the node it is running on - home or foreign
Oracle PFS Packs

Resources

- Resources
  - Floating IP Addresses
  - Database Volume Groups (optional)
  - Oracle Instances
  - Public and Private Listeners
    - Each pack contains one public listener and one private listener
    - Public listener provides session access to the Oracle instance
Oracle PFS Packs

Home / Foreign Node

• On the home node, the pack exists with all resources and package switching enabled
• On the foreign node, the pack exists with the IP addresses only and package switching disabled
  – Keeping the IP addresses available eliminates TCP/IP timeouts
Oracle PFS Packs

Run Script on the Home Node

- Executed by the cluster service under timer
  - Activates Volume Groups
  - Enables IP addresses
  - Starts public and private listeners
  - Starts the Oracle instance
  - Starts event monitors
  - Sends up notification and records service availability
  - Enables package switching
Oracle PFS Packs

Halt Script on the Home Node

• Executed by the cluster service under timer
  – Halts event monitors
  – Issues first fault data capture, archiving, and checkpoint
  – Halts Oracle instance, public and private listeners
  – Sends down notification and records service unavailability
  – Starts the watchdog process
  – Deactivates volume groups
  – Disables IP addresses
Oracle PFS Packs

**Monitors**

- Monitor the instance, listener, the ability of the application to connect and do work
  - Oracle Instance Monitor - Oracle_Up
  - Oracle Listener Monitor - Oracle_Lsnr
  - Oracle Heartbeat Monitor - Oracle_Ping
    - application function for application service levels
    - heartbeat for local and remote instances

- If the Oracle monitors exit, they will initiate recovery and signal the cluster service to halt the pack
Oracle Monitors

*Oracle Instance Monitor - Oracle_Up*

- Connection-based monitor that instantly detects instance termination
- Each instance has its own table created by catpfs.sql
- Oracle_Up has 2 routines: routine 0 and routine 1
  - Routine 0 updates a row in a table and sleeps without issuing commit
  - Routine 1 updates the same row, and is blocked, waiting for routine 0 to commit, rollback, or exit if the instance crashes
  - Routine 1 contains the logic to exit and initiate the halt method
Oracle Monitors

Oracle Heartbeat Monitor - Oracle_Ping

- Monitors the accessibility of the Oracle Service
- Connects as a regular user
  - updates a heartbeat table and verifies the other Oracle instance heartbeat table is changing
  - executes a customized PL/SQL function representing the application function
- Takes recovery action if...
  - The local heartbeat is not updated on time
  - The remote heartbeat is not changing
  - The custom application function does not complete on time
Oracle Monitors

**Oracle Heartbeat Monitor - Oracle_Ping**

- **Oracle_Ping** flow executed under timer:
  - Connect as sys
  - Check for special conditions (example: instance recovery, logon storms)
  - Connect as a regular user (OPS$) via the public listener in the Oracle Pack
  - Update the local Oracle heartbeat
  - Validate the remote Oracle heartbeat, when resilient
  - Execute the function to validate the application on the primary
  - Disconnect and sleep to the end of the cycle
Oracle Monitors

**Oracle Listener Monitor - Oracle_Lsnr**

- Monitors the public and private listeners
- Restarts the listener should it fail
  - The listener is re-startable up to MAX_LSNR_RETRIES (default is 3)
Recovery Policy

• Resilient mode
  – The secondary instance exists
  – If instance fails, move pack to surviving node, enabling that pack’s IP on the surviving node
  – If the primary instance fails, Parallel Server will internally migrate the primary role to the secondary instance

• Non-resilient mode
  – No secondary instance exists
  – recycle (restart) the pack
OPFS Services - Normal operation

Preferred Primary Service

Oracle Pack
Oracle instance
public and private listeners
service and primary role
floating ip address A
Oracle monitors

service publication

ipA

Fast failover

ipB

Oracle Pack
Oracle instance
public and private listeners
service and secondary role
floating ip address B
Oracle monitors

service publication

Preferred Secondary Service

Middle-tier (OES/IAS/Tuxedo..)

TNS Names
OPFS 8.1.7 Concepts
First Fault Data Capture

• After a failure, OPFS captures the following state information for later analysis
  – system states
  – DLM lock dump
  – netstat

• Diagnostics capture is done under a timer and is aborted if user-specified bound is exceeded
OPFS 8.1.7 Concepts

Warming the Library Cache

Cluster Manager

Primary Instance

Library Cache

SELECT
PL/SQL
INSERT
UPDATE
DELETE

Secondary Instance

Library Cache

SELECT
PL/SQL
INSERT
UPDATE
DELETE

DBMS_Libcache
Failover of the Application

- The application must restart transactions on receipt of an error or signal that recovery has occurred
  - restart or resume if preconnected
  - determining the outcome of the last transactions submitted
  - clear in-doubt transactions, if any
- For instance failure - application failover is straightforward using Oracle Parallel Server
- Application receives errors
  - new calls -- ora-1089, ora-1034
  - current calls -- ora-1041, ora-3113, ora-3114
Failover of the Application

For node failure, application failover is NOT straightforward

• TCP/IP time-outs occur
  – whenever a node fails without closing sockets
  – whenever an IP address is unavailable

• Sessions making new calls wait to time-out
• Sessions blocked in IO wait to time-out
Failover of the Application

Oracle Parallel Fail Safe addresses TCP time-outs

• With mobile IP addresses, active calls never time-out
  – eliminates tcp_ip_interval, tcp_ip_abort_interval delays

• With service notification, calls blocked in IO can be interrupted
  – eliminates tcp_keepalive_interval delays
Summary

- Oracle Parallel Fail Safe takes the complexity and cost out of building highly available data servers
  - Pre-packaged
  - Pre-configured
  - Pre-tested
- Parallel Fail Safe provides the highest possible availability for off-the-shelf applications
- Holistic solution addresses all aspects from prevention through analysis
Parallel Fail Safe Operation

Single Database on a cluster:

Server 1
- Primary IP
- Primary Instance A

Server 2
- Secondary IP
- Secondary Instance A

DB-A
Parallel Fail Safe Operation

Single Database on a cluster: Failure of Server 1
Parallel Fail Safe Operation

Multiple Databases on a cluster:

Server 1

- Primary IP
- Primary Instance A
- Secondary IP
- Secondary Instance A
- DB-A

Server 2

- Secondary IP
- Secondary Instance A
- Primary IP
- Primary Instance B
- DB-B
Parallel Fail Safe Operation
N+1 Cluster

Server 1
- Primary IP
- Primary Instance A
- DB-A

Server 2
- Primary IP
- Primary Instance B
- DB-B

Server 3
- Secondary IP
- Secondary Instance A
- Secondary Instance B