

ORACLE[®] **Empowering Extreme Scalability, Availability & Efficiency for your Applications**

Erik Peterson, Grid Competency Center, RAC Development



Abstract

A large majority of mission critical applications implementations are deployed on Oracle Grid stack (RAC, ASM and Clusterware) and attempt to follow Oracle's Maximum Availability Architecture. While any application implemented on the Grid can benefit from gains in scalability and availability, a fuller integration will offer significant gains from empowering performance managements (including debugging performance problems), better ability to efficiently meet service levels, and potential for even higher levels of availability. This presentation will help you understand potential gains and best practices in how to achieve them.



RAC is Great - Why Optimize?

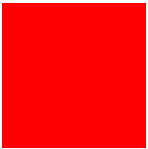
- RAC runs any Application
- 10,000+ RAC Customers
- Hundreds of RAC validated applications
- So why Change your Application?

*Make your Applications Extremely Scalable,
Available & Efficient*



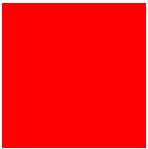
Reasons to Optimize

- **Scalability**
 - Balance Across Unbalanced Servers
 - Immediately Take Advantage of New Capacity
- **Efficiency**
 - Quickly Track Down Source of Bad SQL
 - Manage Performance by Business Area
 - Constrain Lower Priority Work
 - Continuous Application Improvement
- **Availability**
 - Failover Applications Quickly
 - Transaction Failover

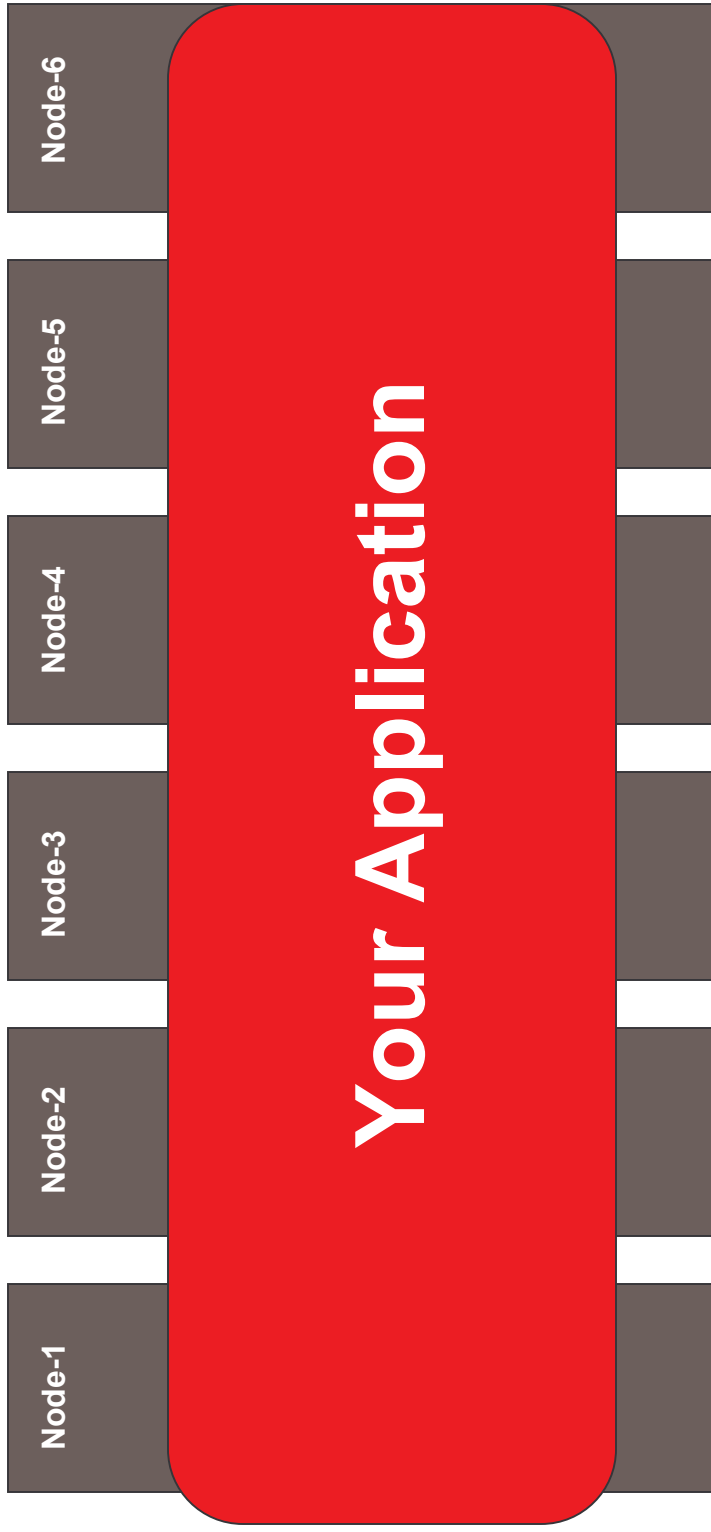


How?

- **Play Tag**
- **Listen**
- **Check Traffic**
- **When All Fails, Try Again**



No Tagging = No Visibility



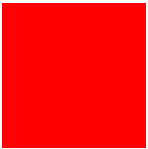
APP_USER



Services

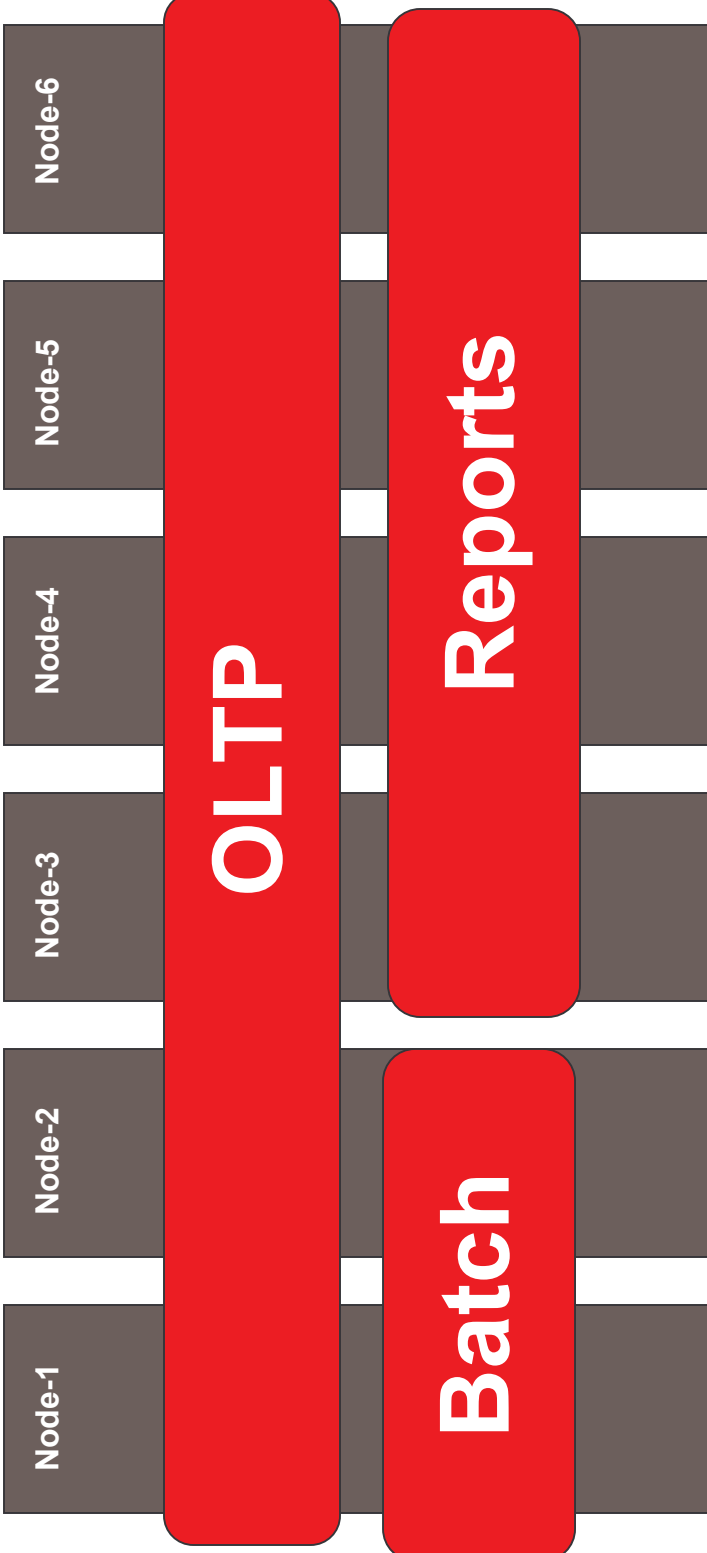
- **Application workloads can be defined as Services**
 - Individually managed and controlled
 - Assigned to instances during normal startup
 - On instance failure, automatic re-assignment
 - Service performance individually tracked
 - Finer grained control with Resource Manager
 - Integrated with other Oracle tools / facilities (E.G. Scheduler, Streams)
 - Managed by Oracle Clusterware

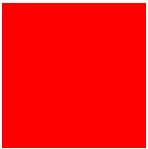
Many features discussed do not apply to default database service



Why Tag? -> Component Visibility

Multiple Services Allows Visibility, Placement & Prioritization





Why Tag? -> Component Visibility

Siebel Component, Component Group & Session ID Mapping
= Full Visibility & Control

HR

Service

Hiring

Module

Payroll

Reject Hire

Approve Hire

Pay Bonus

Print Check

Action

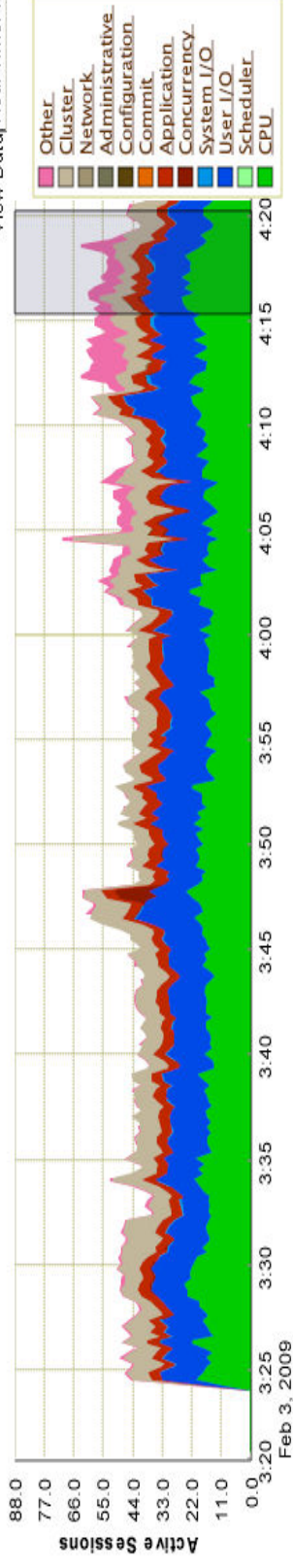
Client Info John Smith Erik Peterson Mike Manager Larry Ellison



ORACLE

Finding the Source of BAD SQL in EBS

Top Activity
 Switch Database Instance [gsi4lap_agsidbs004.us.oracle.com] [Go]
 View Data [Real Time: Manual Refresh]



Detail for Selected 5 Minute Interval

Start Time Feb 3, 2009 4:15:18 PM PST

Top Sessions

Activity (%)	Session ID	User	Name	Program
1.92	6160	APPS		sqlplus@amts681 (TNS V1-V3)
1.92	7075	APPS		sqlplus@amts681 (TNS V1-V3)
1.92	1211	APPS		NODE4_STANDARD@amts681 (TNS V1-V3)
1.92	5699	APPS		sqlplus@amts681 (TNS V1-V3)
1.92	5652	APPS		NODE4_STANDARD@amts681 (TNS V1-V3)
1.92	6381	APPS		NODE4_STANDARD@amts681 (TNS V1-V3)
1.92	3653	APPS		sqlplus@amts681 (TNS V1-V3)
1.92	1234	APPS		sqlplus@amts681 (TNS V1-V3)
1.92	2455	APPS		ALWAYS2@amts681 (TNS V1-V3)
1.92	7009	APPS		NODE4_LONG_RUNNING@amts681 (TNS V1-V3)

Total Sample Count: 14,577

Top SQL

Select Activity (%)	SQL ID	SQL Type
11.57	6t39w1hgmsfa1	SELECT
10.01	7w9twxyz0v2qj	INSERT
5.40	5pjx5pj5xsmnZ	SELECT
4.33	bjgrk8zs4p5fx	SELECT
4.13	3p6c18809b5jn	SELECT
3.94	gx6andxkhggz	PL/SQL EXECUTE
3.55	d9b3jzhkdq6pv	INSERT
2.31	3j5pha0qw7yiv	SELECT
2.31	2sp0ptjh0cdfb	SELECT
2.31	crg3sazcp0f6n	SELECT

Total Sample Count: 12,096

Finding the Source of BAD SQL in EBS

Server

Current Status **ACTIVE**
Serial Number **15371**
DB User Name **APPS**
OS Process ID **21334**
Login Time **Feb 3, 2009 1:21:19 PM**
Login Duration **3:2:44 (hh:mm:ss)**
Connection Type **DEDICATED**
Type **USER**
Resource Consumer Group **Unavailable**

Client

OS User Name **gsiap_a**
OS Process ID **13267**
Host **ams681**
Terminal **Unavailable**
Current Client ID **Unavailable**
Current Client Info **1001 0**

Application

Current SQL **c7w637yuumzsc**
Current SQL Command **UPDATE**
Last Call Duration **3:2:42 (hh:mm:ss)**
SQL Trace **DISABLED**
Open Cursors **202**
Program **NODE4 LONG_RUNNING@ams681 (TNS V1-V3)**
Service **DELTA_BATCH**
Current Module **GSI_GCM_CUST_SYNC**
Current Action **UPDATE_ADD_CUST_DETAILS**

Contention

Blocking Session ID **None**

Wait

Current Wait Event **eng: TX - row lock contention**
Current Wait Class **Application**
Wait Duration **2:53:9 (hh:mm:ss)**
P1 **name(mode 1415053318)**
P2 **usn<<16 | slot 69664775**
P3 **sequence 301074**
Object **GSI_INTF_USER.MIS_HZ_GSI_SYN_CUSTOMER_STG**

Other

Parsing Schema **APPS**
Failover Type **NONE**
Failover Method **NONE**
Failed Over **NO**
PDML Status **DISABLED**
PDDL Status **ENABLED**
PQ Status **ENABLED**
Current Queue Duration **0 (s)**

General

Other EBS Examples

Batch

Application

Current SQL [c7w37tyurnzsc](#)
Current SQL Command **UPDATE**
Last Call Duration **3:2:42 (hh:mm:ss)**
SQL Trace **DISABLED**
Open Cursors [202](#)
Program [NODE4_LONG_RUNNING@amts681](#)
([TNS V1-V3](#))
Service [DELTA_BATCH](#)
Current Module [GSI_GCM_CUST_SYNC](#)
Current Action [UPDATE_ADDL_CUST_DETAILS](#)

OLTP

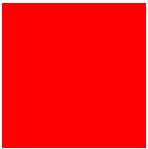
Application

Current SQL [Dwrmwsjy9kd92f](#)
Current SQL Command **SELECT**
Last Call Duration **2:0:50 (hh:mm:ss)**
SQL Trace **DISABLED**
Open Cursors [97](#)
Program [frmweb@amts667 \(TNS V1-V3\)](#)
Service [DELTA_FRM](#)
Current Module [QPXSDDAE](#)
Current Action [FRM:SARISH.MATHEW@ORACLE.COM:US](#)

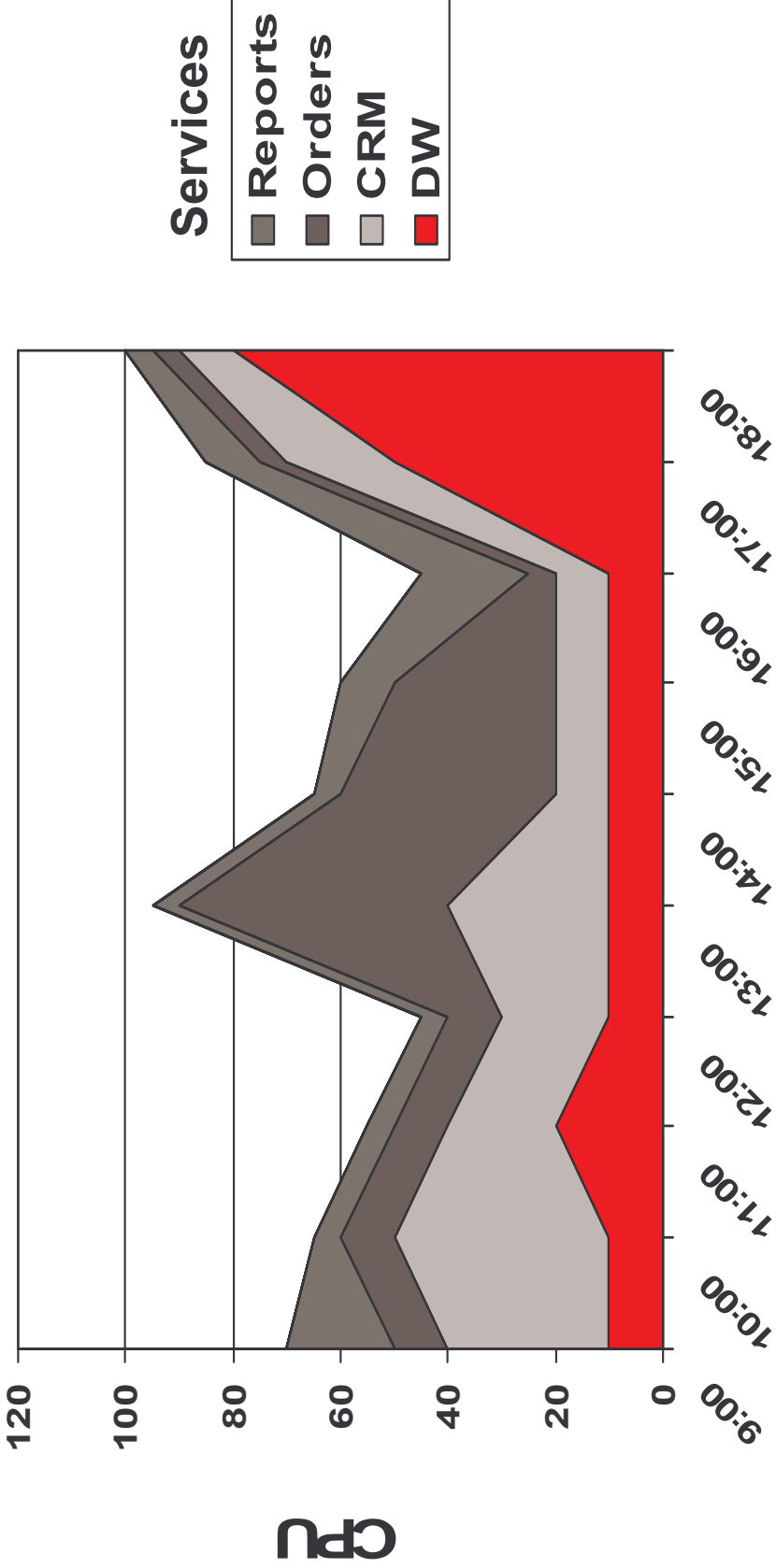


Workload tagging in the database

- Service
 - High level application info
 - *Used for load balancing and node selection*
 - *Default statistics aggregation*
 - Can be basis for tracing
 - Performance Analysis
 - **Module/Action**
 - Lower level application info
 - *Used to identify source of database call*
 - *Optional statistics aggregation*
 - Can be basis for tracing
 - Performance Analysis

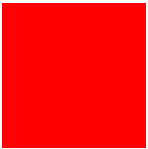


Why Tag? Track & Manage Performance



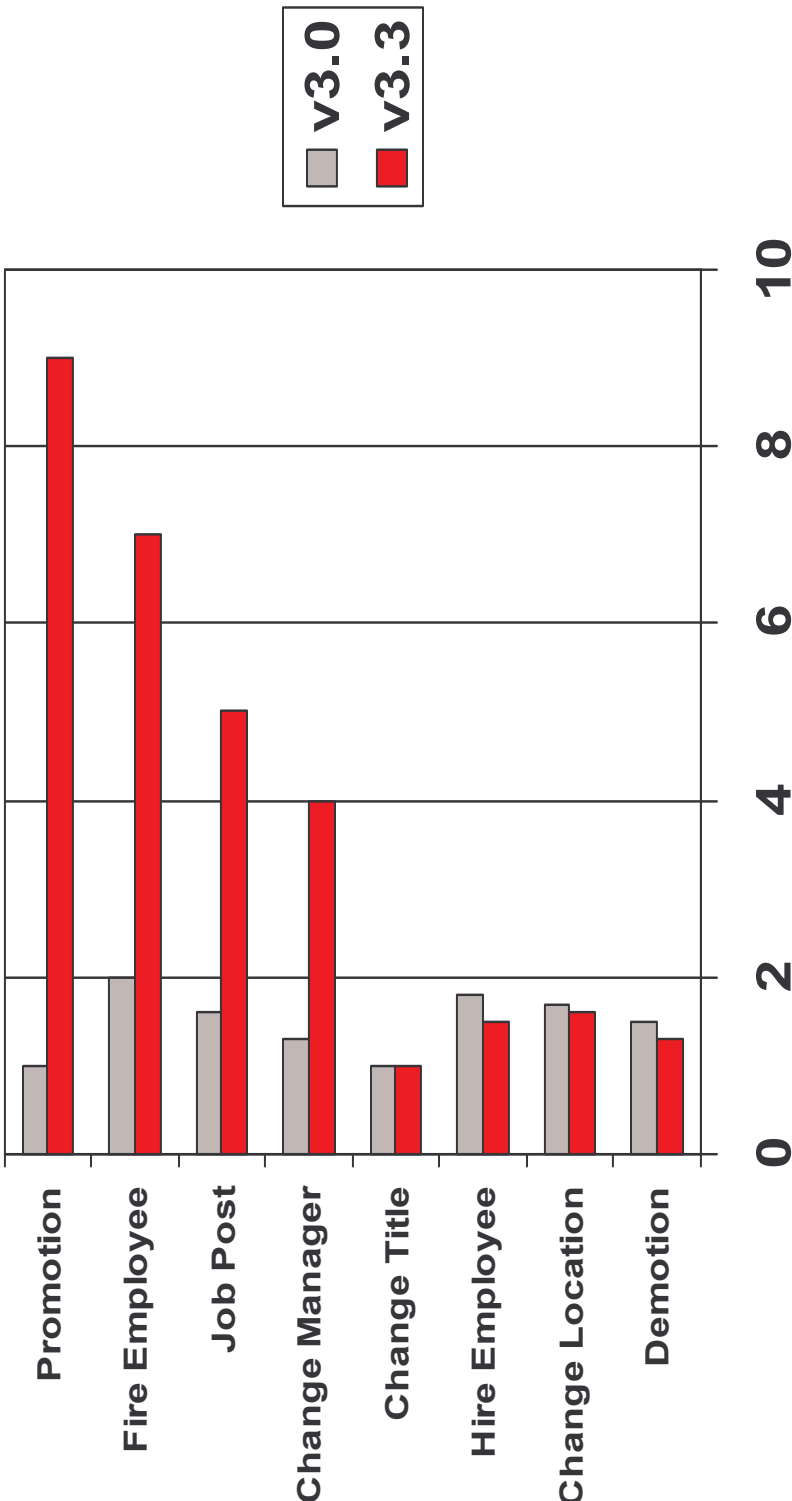
Service, Module & Action Statistics

- user calls
- DB time – response time
- DB CPU – CPU/service
- parse count (total)
- parse time elapsed
- parse time cpu
- execute count
- sql execute elapsed time
- sql execute cpu time
- opened cursors cumulative
- session logical reads
- physical reads
- physical writes
- redo size
- user commits
- workarea executions - optimal
- workarea executions - onepass
- workarea executions - multipass
- session cursor cache hits
- user rollbacks
- db block changes
- gc cr blocks received
- gc cr block receive time
- gc current blocks received
- gc current block receive time
- cluster wait time
- concurrency wait time
- application wait time
- user I/O wait time



Why Tag? – Find Release Digressers

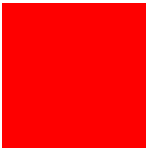
HR Activity Completion Time



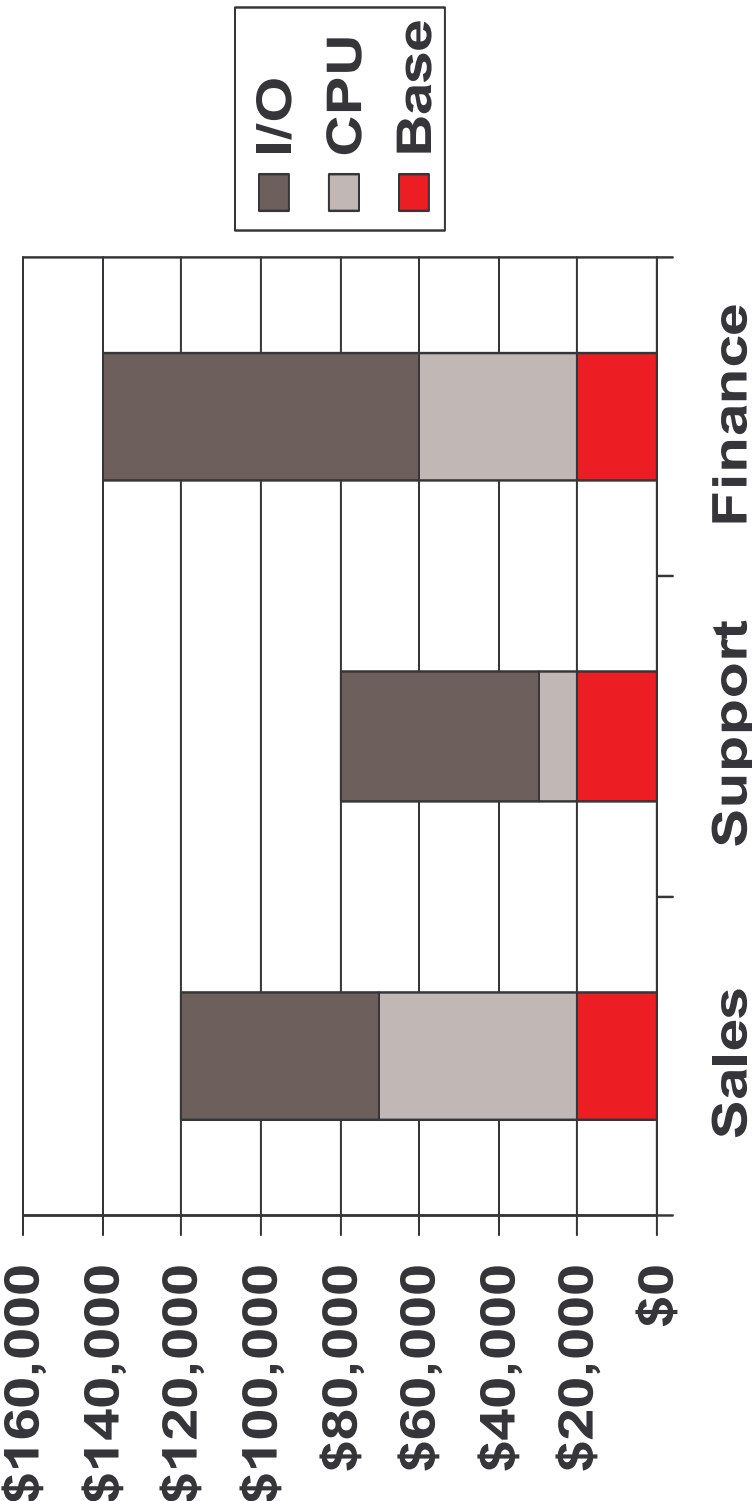


Why Tag? - Limit Resources

- Classify operations into Consumer Groups based on
 - Service, Module
 - User name
- Per Consumer Group, you can
 - Allocate **CPU** resources
 - Limit number of concurrently **active sessions**
 - Limit **degree of parallelism**
 - Terminate or reprioritize **long-running operations**
 - Limit the amount of **undo space** used
 - Limit the **idle time** of a session

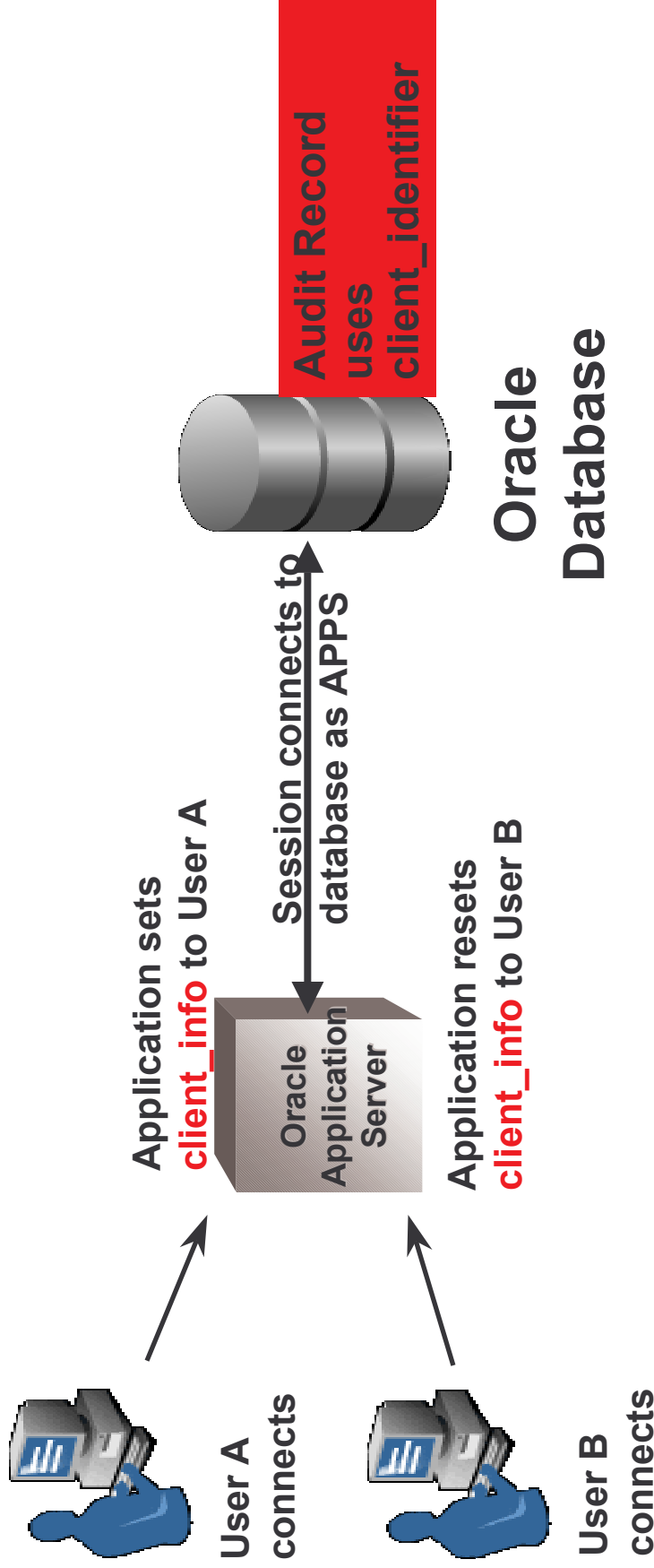


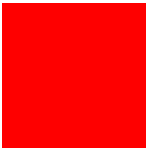
Why Tag? - Chargeback



Why Tag: Auditing

- Any application that sits on top of an Oracle database is integrated with Audit Vault by auditing the base tables and using native database auditing.
- Audit Vault and applications are **VALIDATED** by default





Create Services with Enterprise Manager

The screenshot displays the Oracle Enterprise Manager 11g interface. The 'Availability' tab is highlighted with a green circle. The 'Instances' table shows the following data:

Name	Status	Alerts	Policy Violations	Compliance Score (%)	ASM Instance	ADDM Findings
orcl_orcl1		0	2	95	+ASM1_bpo47.oracle.com	0
orcl_orcl2		0	2	95	+ASM2_bpo48.oracle.com	0

The 'Related Links' section includes: Access, All Metrics, Advisor Central, Backouts, Alert History, and Deployments.

Create Services with Enterprise Manager

Oracle Enterprise Manager (SYS) - Create Service - Windows Internet Explorer

https://pmrac1.us.oracle.com:1159/em/console/acr/acServices?target=sales.us.oracle.com&type=rac_databat Certificate Error

File Edit View Favorites Tools Help

Google Yahoo! Ask.com LookSmart Highlight WeatherBug Companion 21*

Cluster Database: sales.us.oracle.com > Cluster Managed Database Services >

Create Service

Define a highly available service by specifying preferred and available instances. You can also specify service properties to customize failover mechanisms, monitoring thresholds and resource management.

* Service Name

Start service after creation

High Availability Configuration

Instance Name	Service Policy
sales1	Preferred
sales2	Available

TIP Must select at least one preferred instance.

Service Properties

Transparent Application Follower (TAF) Policy

Enable Distributed Transaction Processing
Choose this option for all Distributed Transactions including XA, JTA. Services with exactly one preferred instance can enable this.

Connection Load Balancing Goal Short Long
Load balance connections based on elapsed time (Short) or number of sessions (Long).

Notification Properties

Enable Load Balancing Advisory

Service Time Throughput
Enable advisory for load balancing based on service quality.

Enable Fast Application Notification (FAN) for OCI and ODP.NET Applications

Service Threshold Levels

If thresholds are specified, alerts will be published when the service elapsed response time and/or CPU time exceed the threshold.

Elapsed Time Threshold (milliseconds)

CPU Time Threshold (milliseconds)

Warning

Critical

Resource Management Properties

Associate this service with a predefined consumer group or job class.

Consumer Group Mapping

Job Scheduler Mapping

Done

Setting Module, Action, Client Info

- Set `MODULE / ACTION` using OCI
 - No extra message exchanged – “bundled”

```
OCIAttrSet(session, OCI_HTYPE_SESSION, (dvoid *)  
    "set salary", (ub4)strlen("set salary"),  
    OCI_ATTR_ACTION, error_handle);
```

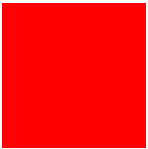
- Set `MODULE / ACTION` using PL/SQL
 - Does require extra message exchanges

```
DBMS_APPLICATION_INFO.SET_MODULE 9  
    module_name => 'add_employee'  
    action_name => 'record contact info');
```



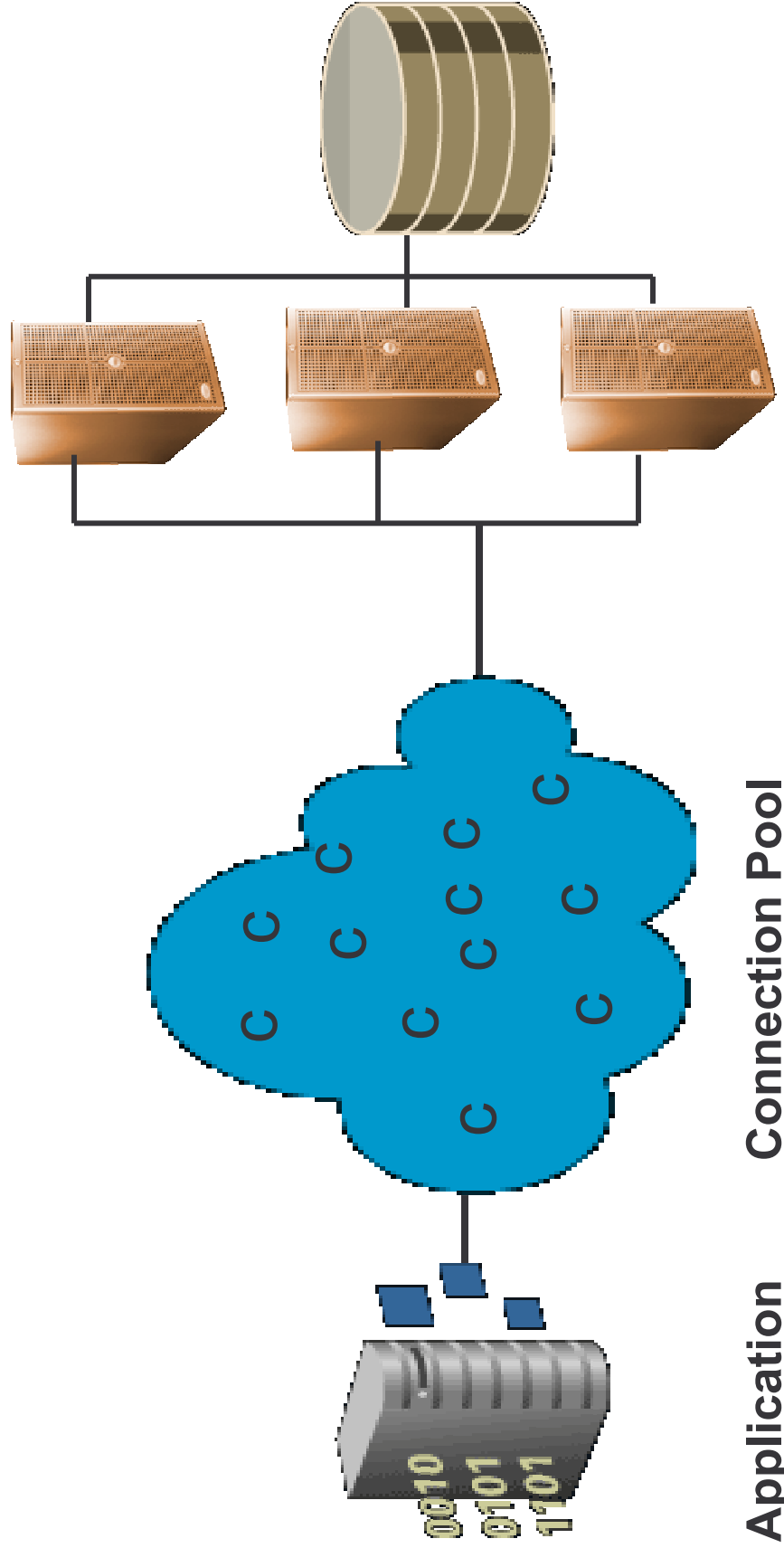
Listen

Fast Application Notification/ Fast Connection Failover



Connection Pools

Fast Connection Failover

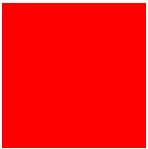


Application

Connection Pool

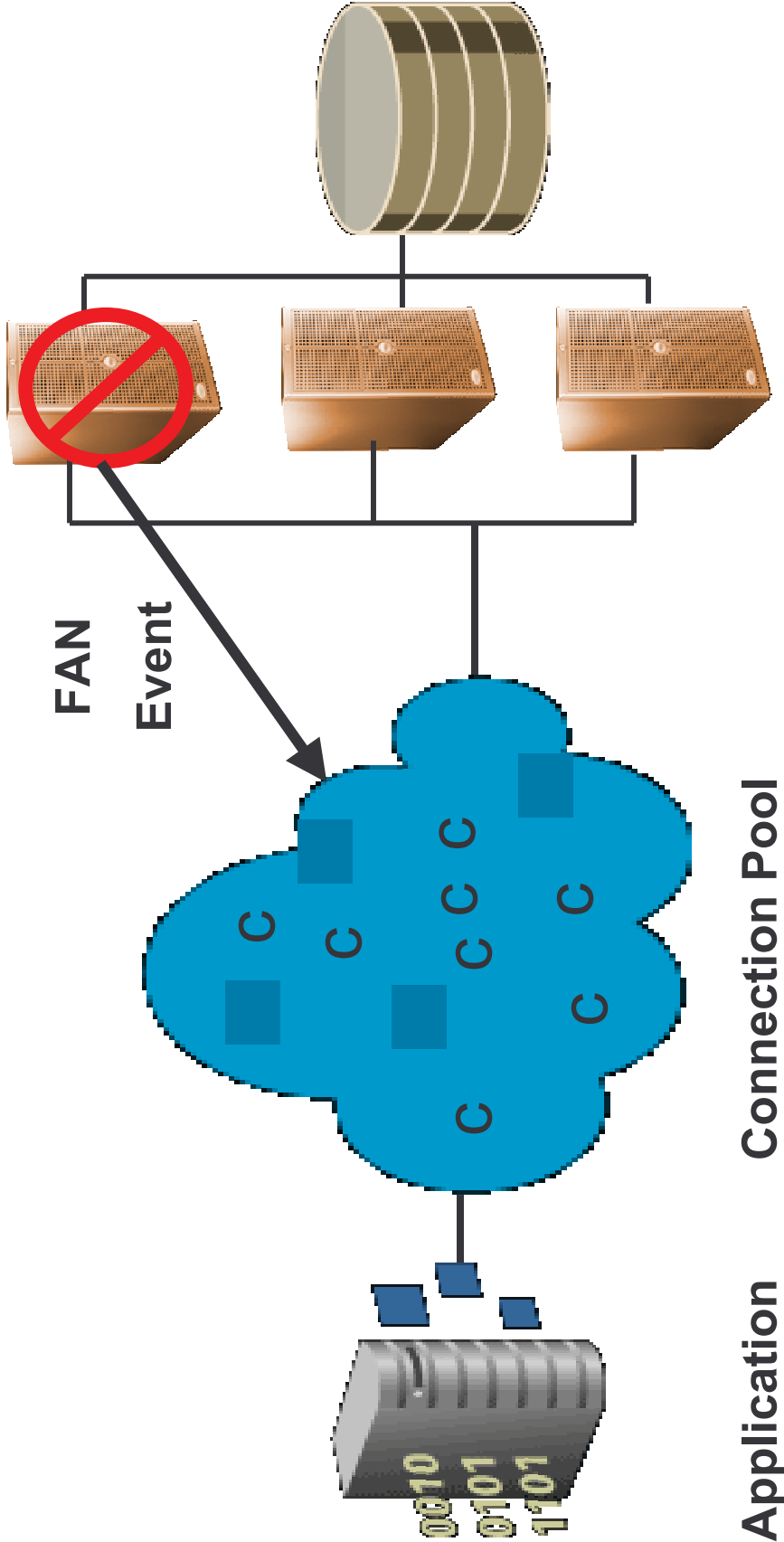
Real Application Clusters





Connection Pools

What happens with a failure?



Application

Connection Pool

Real Application Clusters



Fast Connection Failover

- Fast and reliable high availability for connections in an Oracle Real Application Clusters 10g environment
- Enable it and forget it
- Application can make it transparent to user by trapping SQL Exception and retrying
- Supported by Oracle JDBC, OCI, and ODP.NET



FAN/FCF Client Integration

JDBC

- When DOWN signal received from RAC 10g
 - First pass: Connections are marked as down
 - Second pass: Aborts and removes connections that are marked as down
 - Routes new requests to surviving instances
 - Throws exception if application was in midst of transaction
- When UP signal received from RAC 10g
 - Creates new connections to new instances
 - Distributes new work requests evenly to all available instances

Step 1. Enable Cache and Fast Connection Failover, Remote ONS

- Set data source properties

```
OracleDataSource ods = new OracleDataSource()  
...  
ods.setUser("scott");  
ods.setPassword("tiger");
```



```
ods.setConnectionCachingEnabled(True);  
ods.setFastConnectionFailoverEnabled(True);  
setONSConfiguration("nodes=host1:6200,host2:6200");
```

```
ods.setConnectionCacheName(MyCache);  
ods.setConnectionCacheProperties(cp);  
ods.setURL("jdbc:oracle:thin:@(DESCRIPTION=  
  (LOAD_BALANCE=on)  
  (ADDRESS=(PROTOCOL=TCP) (HOST=VIP1) (PORT=1521))  
  (ADDRESS=(PROTOCOL=TCP) (HOST=VIP2) (PORT=1521))  
  (CONNECT_DATA=(SERVICE_NAME=MYSERVICE)))");
```

- or set system properties

```
-D oracle.jdbc.FastConnectionFailover=true
```





Step 2. Verify ONS on RAC nodes

- `$ORACLE_HOME/opmn/conf/ons.config`

```
localport=6100    # port ONS is writing to
remoteport=6200  # port ONS is listening on
loglevel=3
useocr=on
```



Step 3. When starting the application..

- Specify system property
 - Doracle.ons.oraclehome=<ORACLE_HOME-on-client>
- Ensure ons.jar file is on the CLASSPATH.

See [Workload Management with Oracle Real Application Clusters \(FAN, FCF, Load Balancing\)](#) white paper for OCI and ODP.NET details



ODP.NET Integration

- DOWN event: cleans up sessions in the connection pool that go to the stopped instance proactively disposes connections that are no longer valid.
- Establishes connections to existing RAC instances total number of connections below the value that is set for the MIN POOL SIZE parameter.
- Must be using connection pool

```
user id=scott;password=tiger;data
source=erp;HA events=true;pooling=true"
```
- Service must have `aq_ha_notifications=>true`



Oracle Call Interface (OCI)

- Down event for an instance or node:
 - Terminate affected connections at the client
 - Remove connections from the OCI connection pool and OCI session pool
 - If TAF is configured, the connection will failover, if not, the client receives an error such as ORA-12543
 - If a TAF callback has been registered, then the failover retries and failover delay are ignored. If an error occurs, TAF will continue to attempt to connect and authenticate as long as the callback returns a value of `OCI_FO_RETRY`. Any delay should be coded into the callback logic.
- Must initialize the OCI Environment in `OCI_EVENTS` mode and link with a thread library.

Using FAN with OCI

OCI connection pool, OCI session pool, TAF, OCI client/server

1. Set HA_AQ_Notification using DBMS_SERVICE

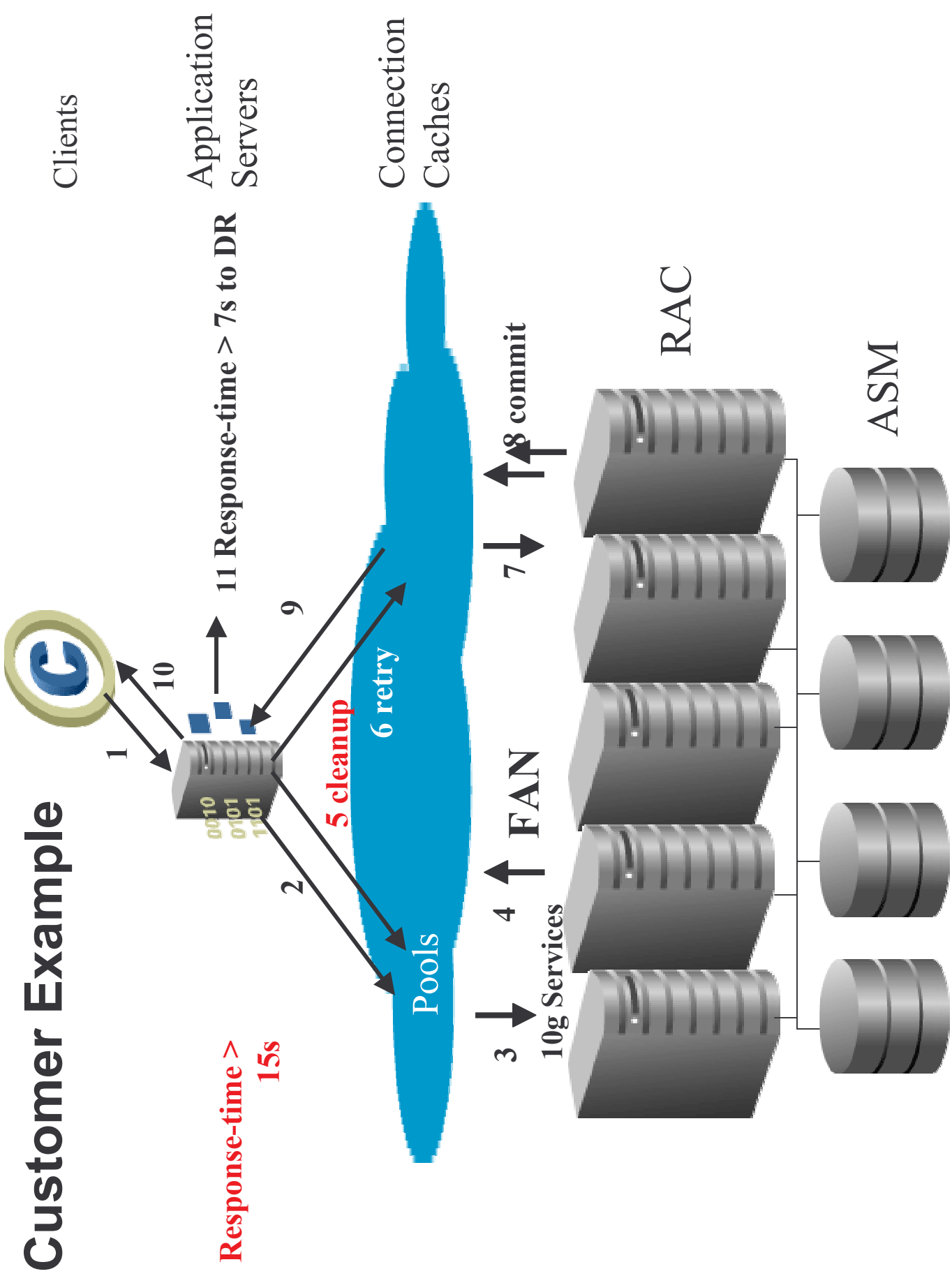
```
execute dbms_service.modify_service
      (service_name => 'test' -
      , aq_ha_notifications => true -
      , clb_goal => dbms_service.clb_goal_short);
```

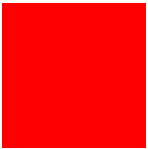
2. Enable OCI_EVENTS on the client

```
( OCIEnvCreate(...) )
```

3. Link application with O/S client thread library

Customer Example





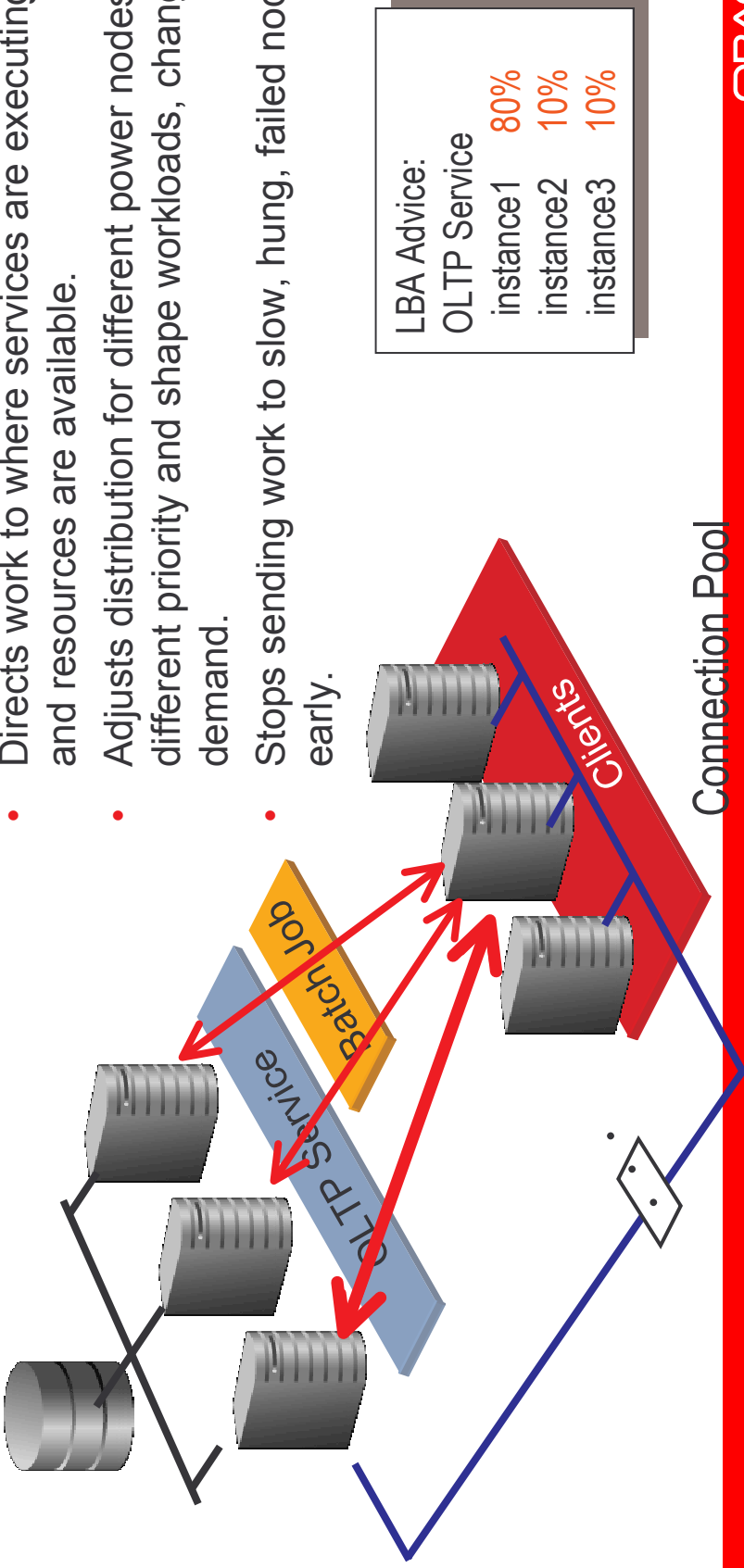
Check Traffic

Load Balancing Advisory



Load Balancing Advisory (LBA)

- Load Balancing Advisory is an advisory for balancing work across RAC instances.
- Load balancing advice
 - Is available to ALL applications that send work.
 - Directs work to where services are executing well and resources are available.
 - Adjusts distribution for different power nodes, different priority and shape workloads, changing demand.
 - Stops sending work to slow, hung, failed nodes early.





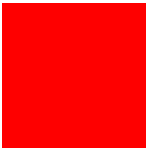
Runtime Connection Load Balancing

- Solves the Connection Pool problem!
- Easiest way to take advantage of Load Balancing Advisory
- No application changes required
- No extra charge software to buy
- Enabled by parameter on datasource definition
- Supported by JDBC, OCI and ODP.NET



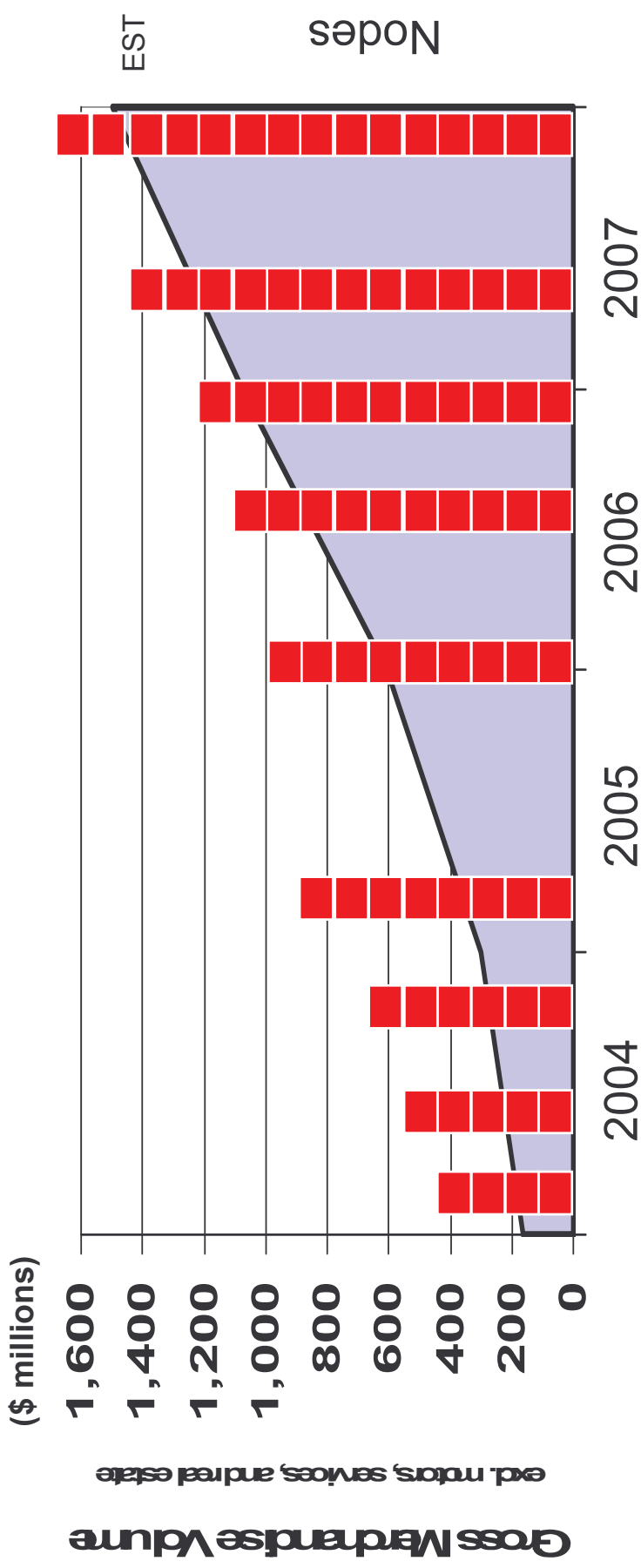
Runtime Connection Load Balancing

- Client connection pool is integrated with RAC load balancing advisory
- When application does “getConnection”, the connection given is the one that will provide the best service.
- Policy defined by setting GOAL on Service
- Need to have Connection Load Balancing



MercadoLibre

- E-commerce leading company in Latin America
- Runs marketplace and payments platform on RAC
- Scaled incrementally as marketplace grew



Create Services with Enterprise Manager

Oracle Enterprise Manager (SYS) - Create Service - Windows Internet Explorer

Cluster Database: sales.us.oracle.com > Cluster Managed Database Services >

Create Service

Define a highly available service by specifying preferred and available instances. You can also specify service properties to customize failover mechanisms, monitoring thresholds and resource management.

* Service Name: test

Start service after creation

High Availability Configuration

Instance Name	Service Policy
sales1	Preferred
sales2	Available

TIP Must select at least one preferred instance.

Service Properties

Transparent Application Follower (TAF) Policy: None

Enable Distributed Transaction Processing

Choose this option for all distributed transactions including XA, JTA. Services with exactly one preferred instance can enable this.

Connection Load Balancing Goal: Short Long
Load balance connections based on elapsed time (Short) or number of sessions (Long).

Notification Properties

Enable Load Balancing Advisory

Service Time Throughput
Enable advisory for load balancing based on service quality.

Enable Fast Application Notification (FAN) for OCI and ODP.NET Applications

Service Threshold Levels

If thresholds are specified, alerts will be published when the service elapsed response time and/or CPU time exceed the threshold.

Elapsed Time Threshold (milliseconds):

CPU Time Threshold (milliseconds):

Warning:

Critical:

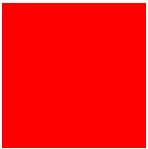
Resource Management Properties

Associate this service with a predefined consumer group or job class.

Consumer Group Mapping: None

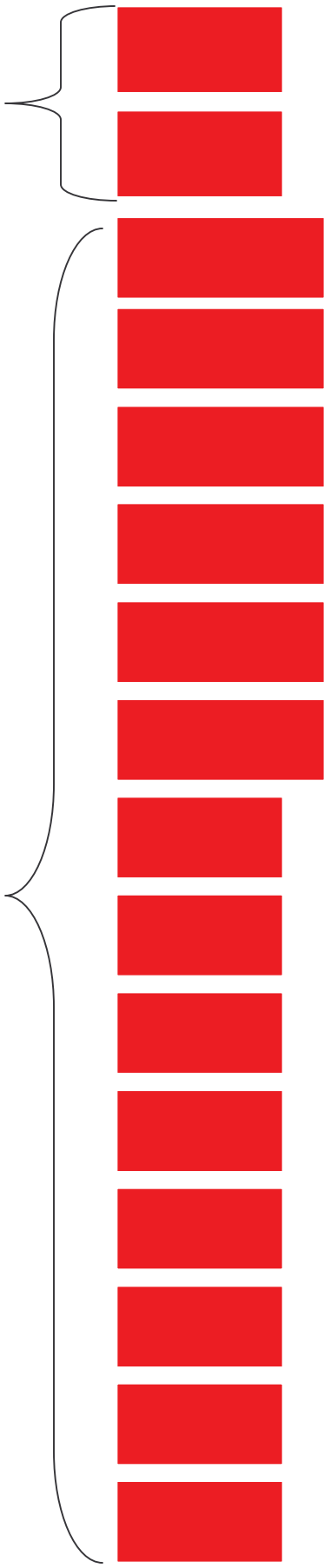
Job Scheduler Mapping: None

Done



MercadoLibre Load Distribution

Marketplace / MercadoPago (payments platform) Customer Service
Searches, Listings, Bidding, Reporting

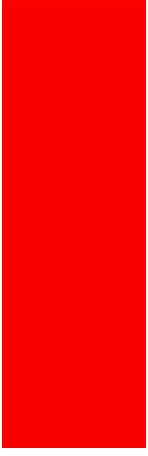


- Newer nodes are faster and have more memory
- Marketplace workload is not partitioned



When All Fails, Try Again

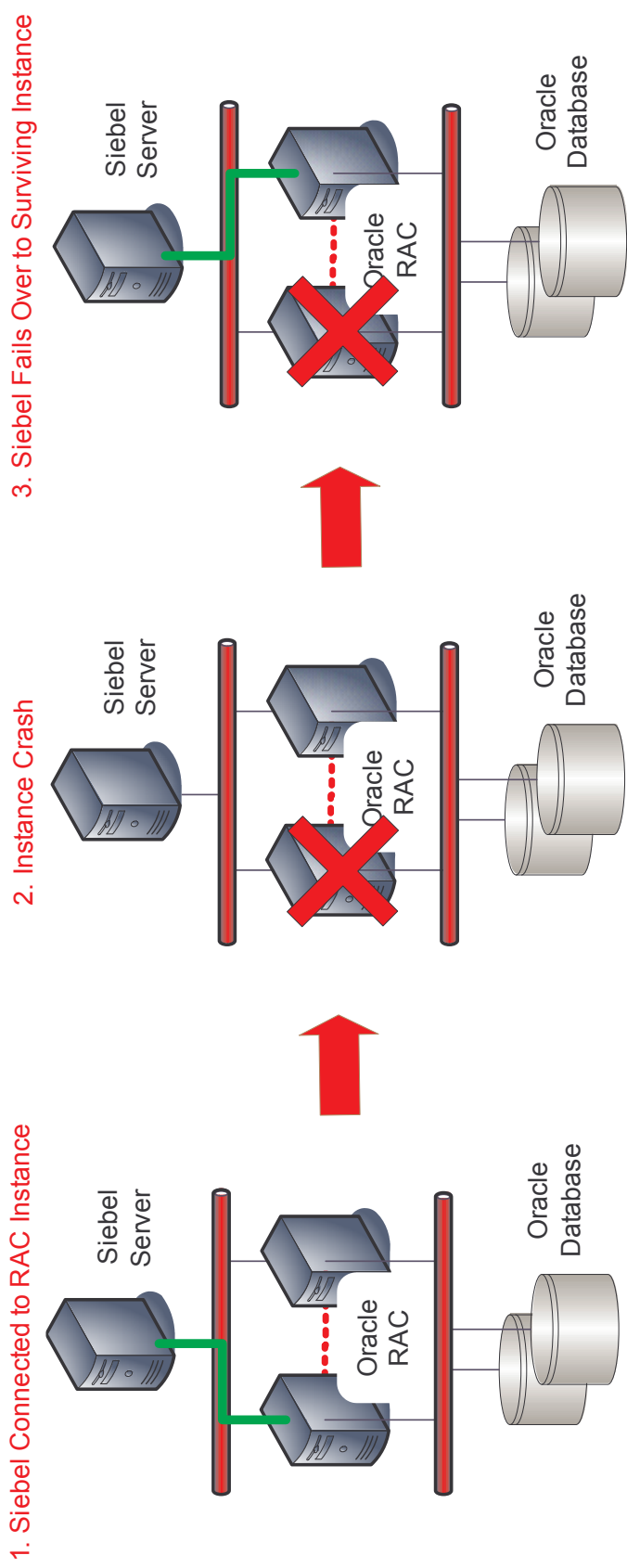
Application Retry



**IF AT FIRST
YOU DON'T
SUCCEED,
TRY, TRY
AGAIN!**

Siebel MAA

Transparent Application Failover & Retry



Works for:

- RAC Instance or Node Failure
- Local Data Guard Standby Failover and Switchover
- Database Shutdown/Startup

TAF & Retry Transactions

Siebel Client Behavior on Failover or Switchover

Client Operation	Behavior
Updating data and steps-off (saves) the updates during or just after the DB failure	<ul style="list-style-type: none">• Oracle reconnects and reconstructs the database session on a surviving node and Siebel resubmits the update• Failure is transparent to the end-user
Paging through queried data when the DB failure occurs	<ul style="list-style-type: none">• Oracle reconnects and reconstructs the database session on a surviving node, re-executes the query, repositions the SQL cursor, and returns the next set of rows• Failure is transparent to the end-user
New query or switch screens after the DB failure	<ul style="list-style-type: none">• Oracle reconnects and reconstructs the database session on a surviving node• Failure is transparent to the end-user



Demo Siebel RAC Failover





Application Code

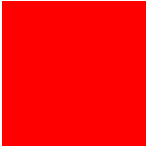
Catch SQL Exception and Retry

```
try {
    conn= getConnection();
    // do some work
} catch (SQLException e) {
    handleSQLException(e);
}
...
handleSQLException (SQLException e)
{
    if
    (OracleConnectionCacheManager.isFatalConnectionError(e))
        // check and retry transaction
        ConnRetry = true; // Fatal Connection error detected,
}
}
```



Reasons to Optimize

- **Scalability**
 - Balance Across Unbalanced Servers (LBA/RCLB)
 - Immediately Take Advantage of New Capacity (FAN)
- **Efficiency**
 - Quickly Track Down Source of Bad SQL (Tagging)
 - Manage Performance by Business Area (Tagging)
 - Constrain Lower Priority Work (Tagging/Resource Mgr)
 - Continuous Application Improvement (Tagging)
- **Availability**
 - Failover Applications Quickly (FAN)
 - Transaction Failover (Retry)



Where Change

- **Infrastructure Changes**
 - **Fast Application Notification (FAN)**
 - **Load Balancing Advisory (LBA)**
 - **Services**
- **Application Changes**
 - **Module, Action, Client Info**
 - **Retry**



References

- **Oracle Real Application Clusters Administration and Deployment Guide – Chapter 4: Introduction to Automatic Workload Management**
- **Workload Management with Oracle Real Application Clusters (FAN, FCF, Load Balancing)**
<http://otn.oracle.com/rac>
- **Siebel Maximum Availability Architecture**
- **MAA Best Practices for Client Failover**
 - <http://otn.oracle.com/goto/maa>
- **RAC Sample Code Page**

http://www.oracle.com/technology/sample_code/products/rac/index.html