ORACLE OPEN WORLD

experience
OPENWORLD

November 11–15, 2007

ORACLE

ORACLE

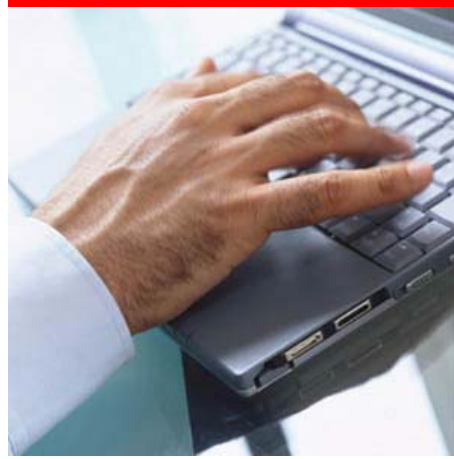
Oracle Active Data Guard
What They Didn't Print in the Manuals

Joe Meeks
Director
HA Product Management

Mike Smith
Principal Member of Technical Staff
MAA Development

Grant McAlister
Principal Database Engineer
Amazon

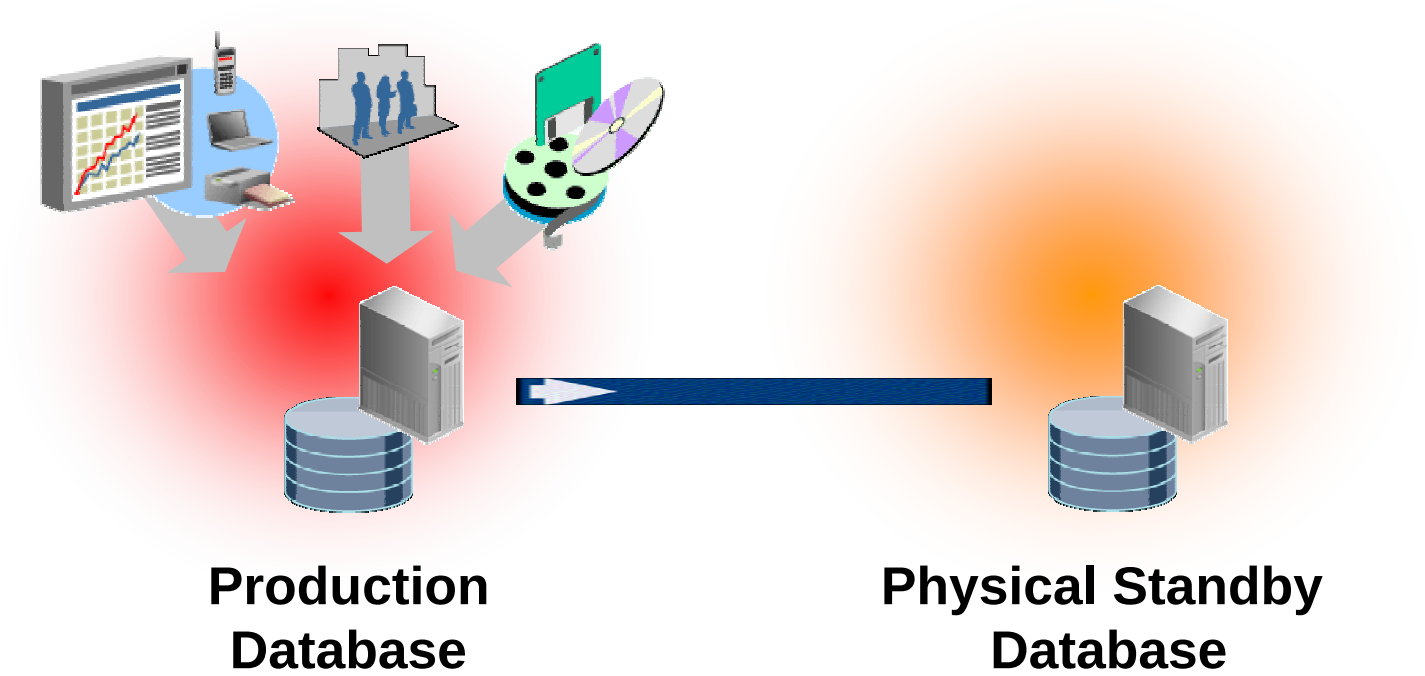
Agenda



- Introduction
- Enabling Active Data Guard
- Best Practices
- Amazon Experience
- Appendix
  - Oracle Database 11g Media Recovery Performance
  - Redo Apply Performance Tuning
  - Client Failover

Traditional Physical Standby Databases

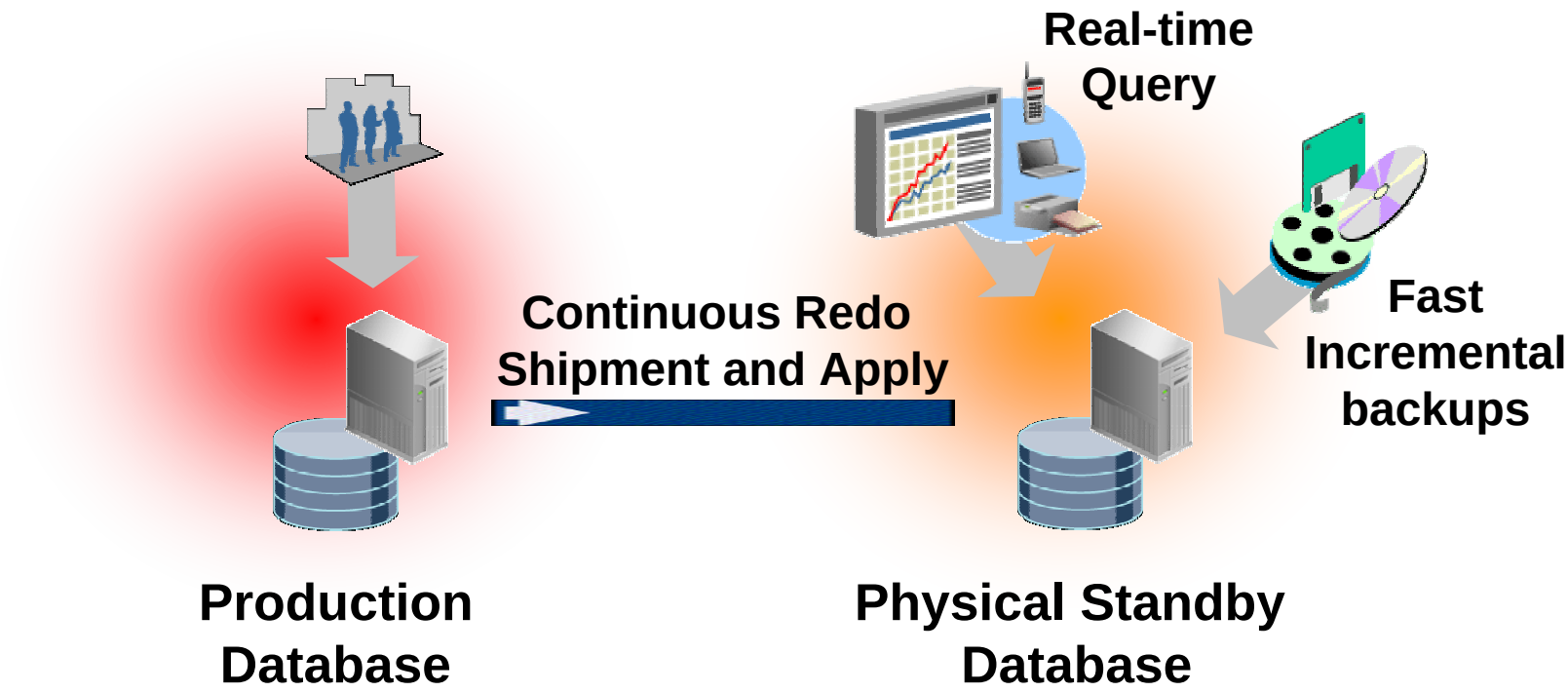
Investment in Disaster Recovery



- Applications, backups, reports run on production only

Active Data Guard 11*g*

Investment in Improved Quality of Service



- Offload read-only queries to an up-to-date physical standby
- Perform fast incremental backups on a physical standby

What's New

Data Guard 11*g*

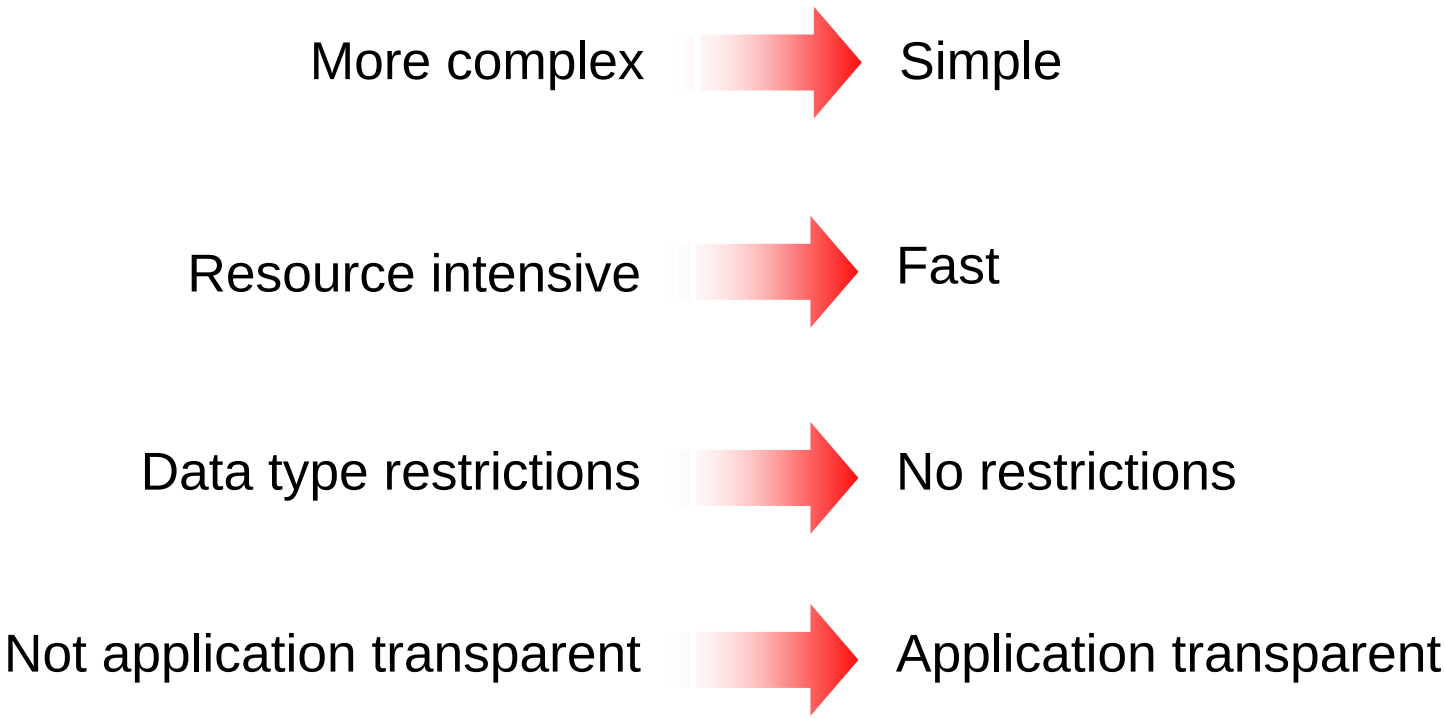
- Recovery (redo apply) must be stopped to open a standby read-only
  - Same functionality as previous Data Guard releases
- Redo Apply has exclusive access to data files – reads not allowed
- Not possible to guarantee read consistency while redo apply is active

Data Guard 11*g* with the Active Data Guard Option

- Physical Standby is open read-only while redo apply is active
- Read consistency is guaranteed
- Redo apply is not adversely affected by read-only workload

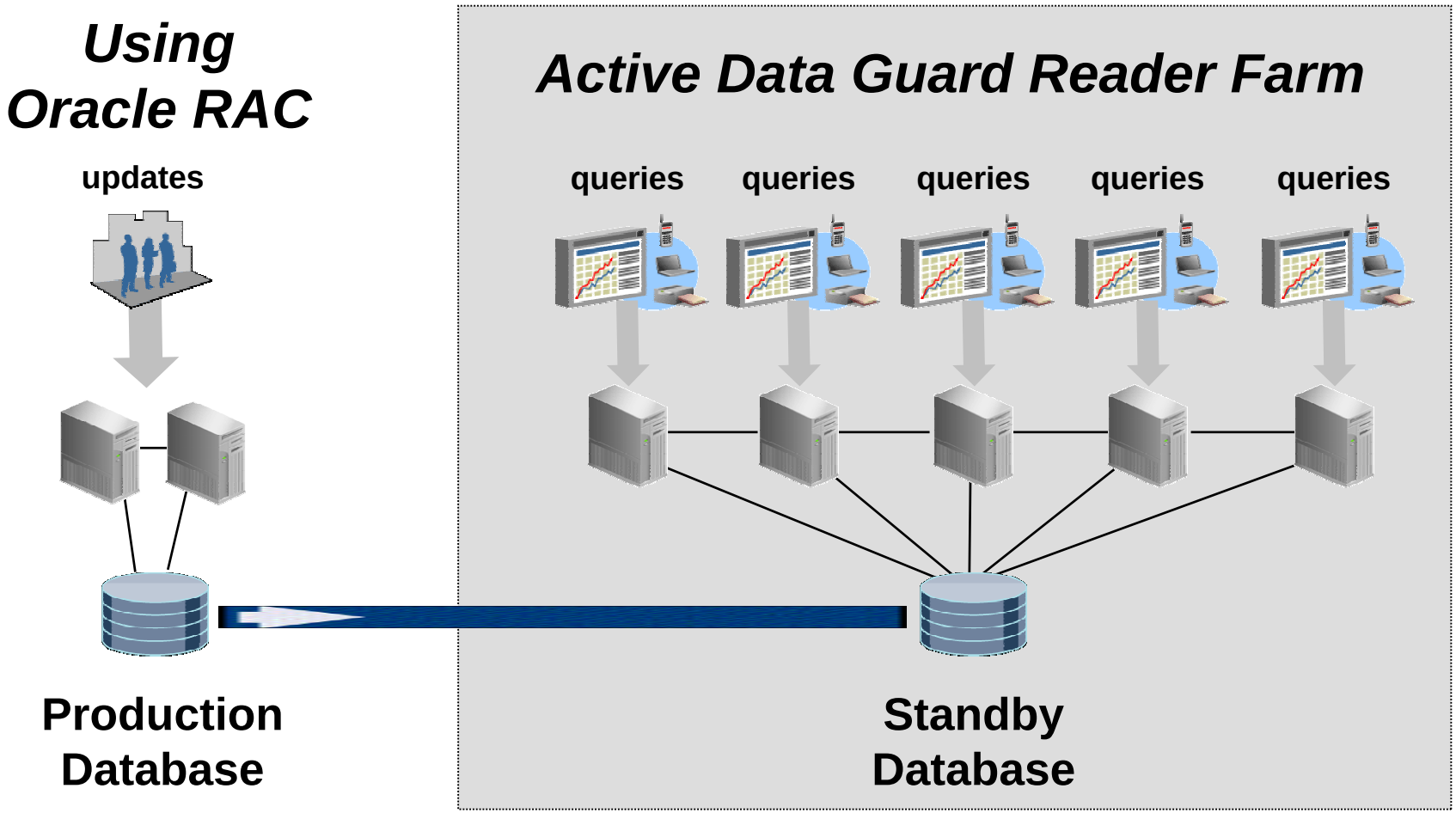
Active Data Guard Difference

Compared to Traditional Replication Methods



Active Data Guard 11*g*

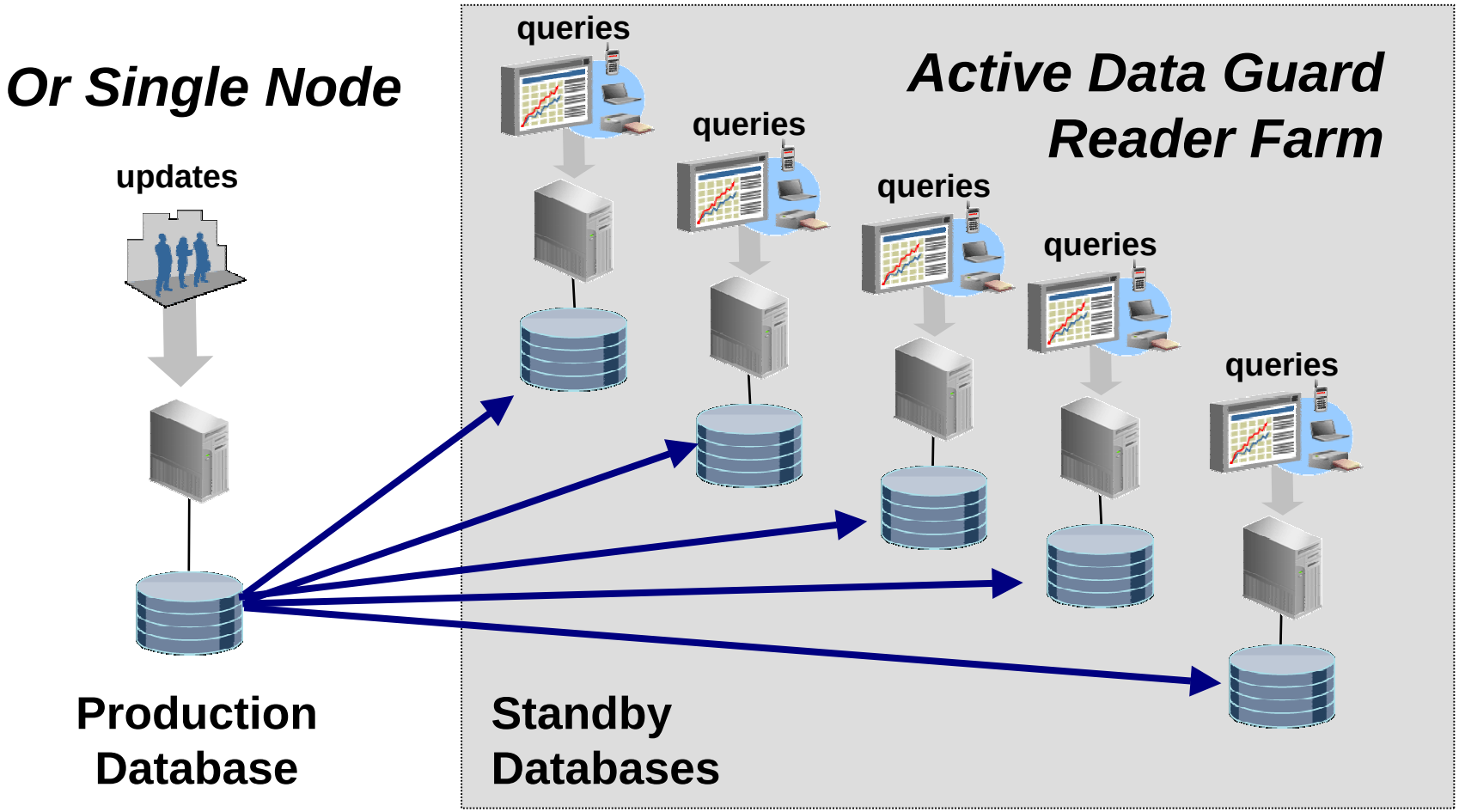
Scale-out Query Performance to Web-Scale*



*DR included **

Active Data Guard 11*g*

Scale-out Query Performance to Web-Scale*

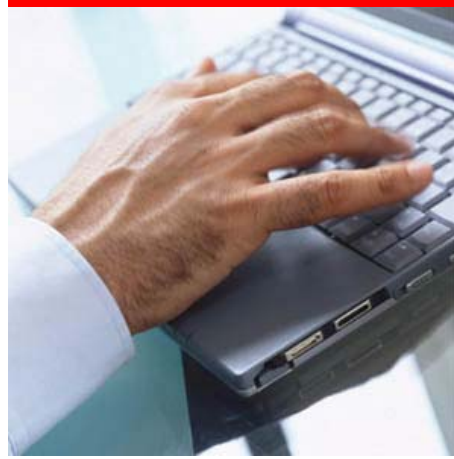


*DR included **

Active Data Guard - Licensing

- A Database Option for Oracle Enterprise Edition
- Active Data Guard license is required when using either:
  - Real-time Query
  - RMAN block-change tracking on a standby database
- Active Data Guard is 100% compatible with new Data Guard functionality included with Oracle Database 11*g* Enterprise Edition
  - S291915 - What's New in Oracle Data Guard 11g: Revolutionizing Data Protection and Availability

Agenda



- Introduction
- Enabling Active Data Guard
- Best Practices
- Amazon Experience
- Appendix
  - Oracle Database 11g Media Recovery Performance
  - Redo Apply Performance Tuning
  - Client Failover

Enabling Active Data Guard

- Begin with a Data Guard 11*g* physical standby database
  - If redo apply is running, stop redo apply
  - Open the standby database read-only
  - Start redo apply
- If open read-only fails because standby instance was aborted or datafiles were restored then…
  - Bring to mount state and start redo apply
  - Stop redo apply and open read-only
  - Restart redo apply
- If standby database is Oracle RAC
  - Make sure redo apply is running on the apply instance in read-only mode BEFORE opening subsequent instances read-only

Data Guard Broker & Enterprise Manager

- Data Guard Broker CLI
  - Stop redo apply with the following command

    EDIT DATABASE 'RTQ' SET STATE='APPLY-OFF'
  - Open standby read-only via SQL*Plus

    SQL> alter database open read only;
  - Restart redo apply via broker CLI

    EDIT DATABASE 'RTQ' SET STATE='APPLY-ON'
- Oracle Enterprise Manager 10*g*
  - Stop redo apply within Data Guard GUI
  - Open standby in read-only mode in Advanced Startup Options
  - Restart redo apply within Data Guard GUI

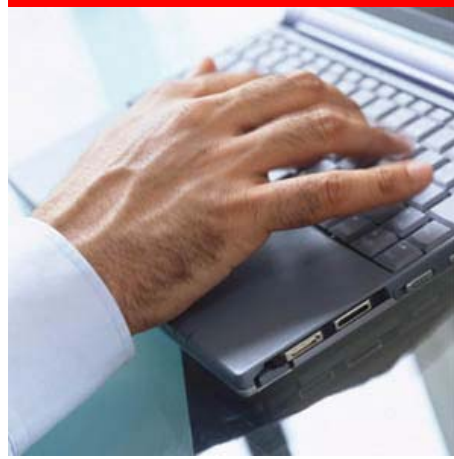
Guarantee of Consistent Reads

- Maintained through Query SCN
  - Identifies most recent read point
  - Used by queries to insure consistent reads
  - Current value given by **`current_scn`** from v$database on standby
- Redo Apply advances the Query SCN
  - After all dependent changes have been fully applied
  - Propagated to all other instances in standby RAC

Supported Operations for Read Only

- When connected to an Active Data Guard standby database, read-only applications can perform/use:
  - Selects
  - Alter session / system
  - Set role
  - Lock table
  - Call stored procedures
  - DBlinks to write to remote databases
  - Stored procedures to call remote procedures via DBlinks
  - SET TRANSACTION READ ONLY for transaction level read consistency
  - Complex queries e.g. grouping set queries and with clause queries

Agenda



- Introduction
- Enabling Active Data Guard
- Best Practices
- Amazon Experience
- Appendix
  - Oracle Database 11g Media Recovery Performance
  - Redo Apply Performance Tuning
  - Client Failover

MAA Best Practices

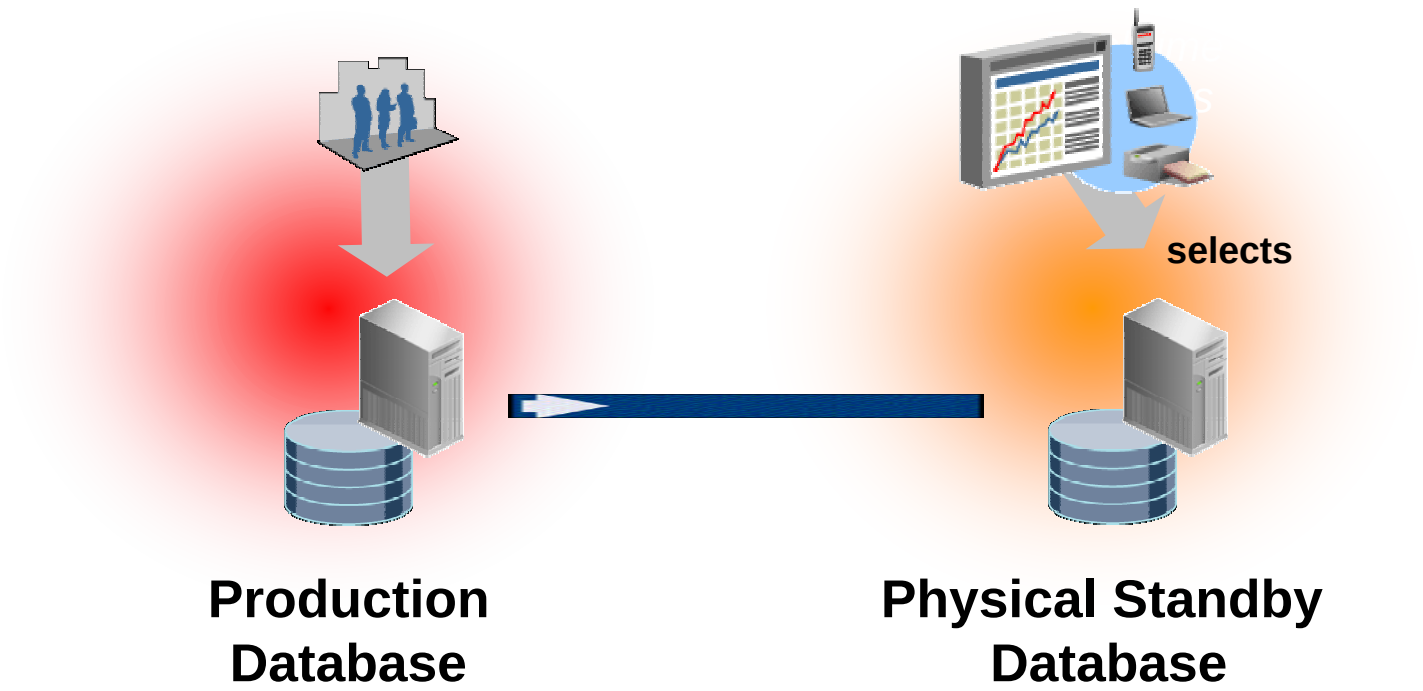
- Reporting application work load
  - Selects only
  - DML
- Routing user connections
  - New connections
  - Role transitions
- Optimizing performance
  - SQL Tuning
  - Redo Apply Tuning

Candidates for Active Data Guard

- Read-only applications and ad-hoc queries
  - Do not write or modify database state
  - Do not generate redo
- Read-mostly applications
  - Applications that perform many more reads than writes
    - e.g. reporting applications with ancillary writes
  - All writes must be redirected to a database that is open read-write

Active Data Guard 11*g*

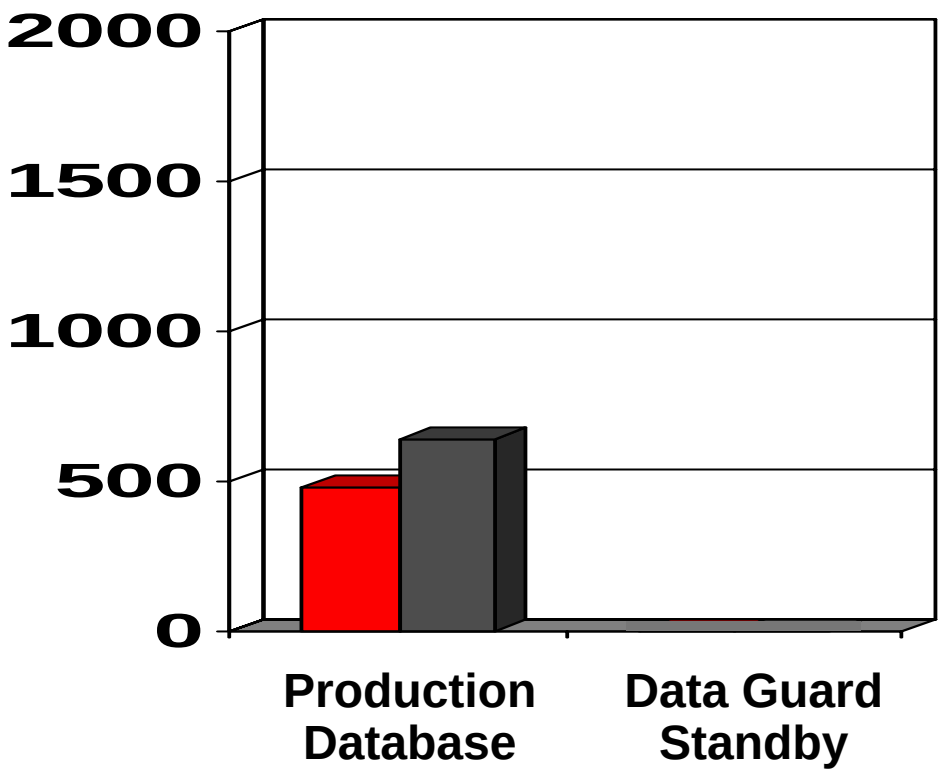
Read-Only Application Model



- Application directs read-only selects to the standby

Yesterday's Paradigm

All Workloads Run on Production



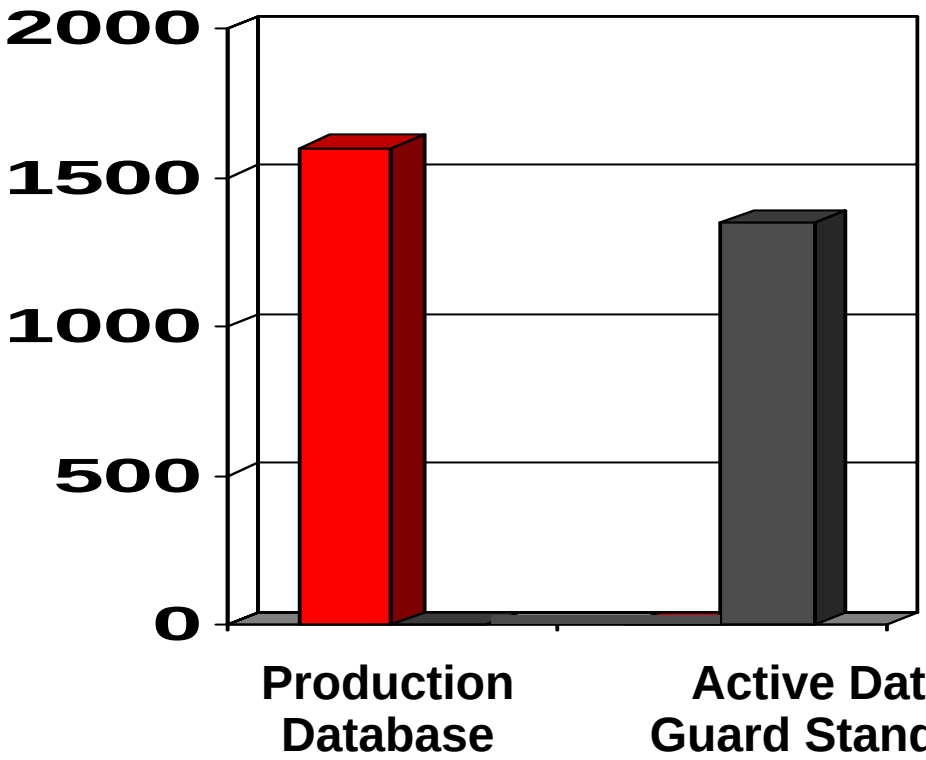
Read-write service
480 tps

Read-only service
640 tps

Primary server at
100% of capacity

Active Data Guard 11*g*

Easily Optimize Performance for All Workloads



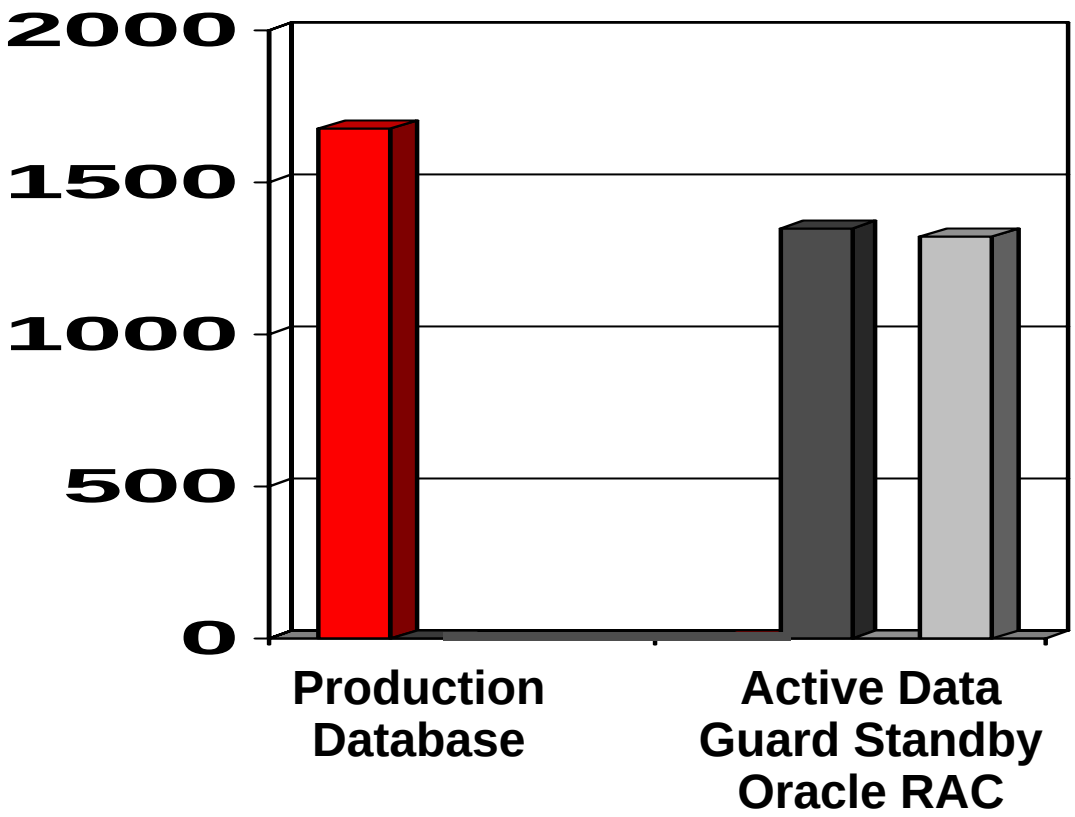
read-write service
1,680 tps
+250%

read-only service
1,350 tps
+110%

Primary server at 60% of capacity

Active Data Guard 11*g*

Scale Read Performance by Adding Standby Nodes



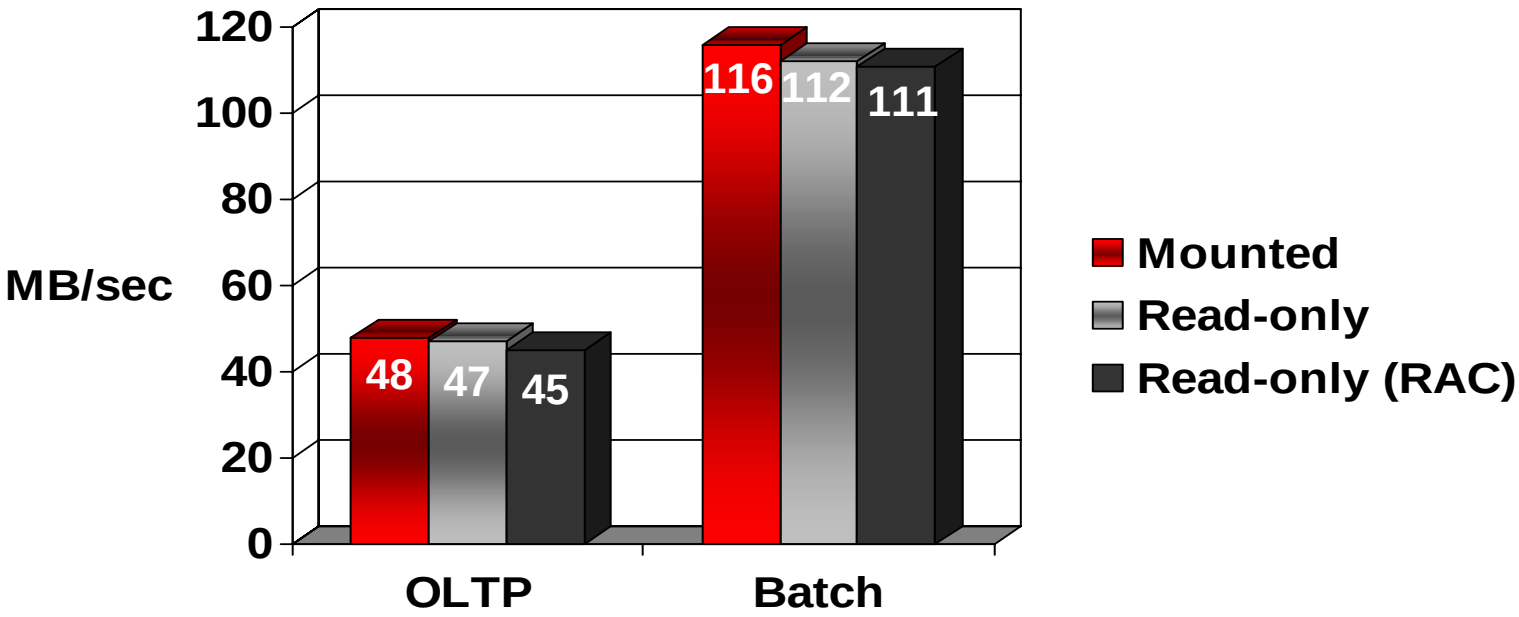
read-write service
1,680 tps
+250%

read-only service
node1: 1,350 tps
node2: 1,325 tps

Adding the 2nd node scales read-only performance *another* 100%

Redo Apply Performance

With Active Data Guard Enabled

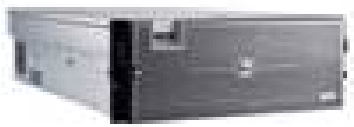


- No significant performance impact when open read-only

Environment Details - EMC

2 x CX3-40F UltraScale Storage System

- Flare Release 26
- 4 GB RAM per SP
  - Write Cache = 2GB
  - SPA & SPB = 1GB
- 60 146GB FC drives @ 15K RPM
- All LUNs bound as 1+1 Raid 10
  - Non Vault DATA LUNs 133 GB
  - Vault DATA LUNS 99 GB
  - LUN Prefetch set to Variable with default settings

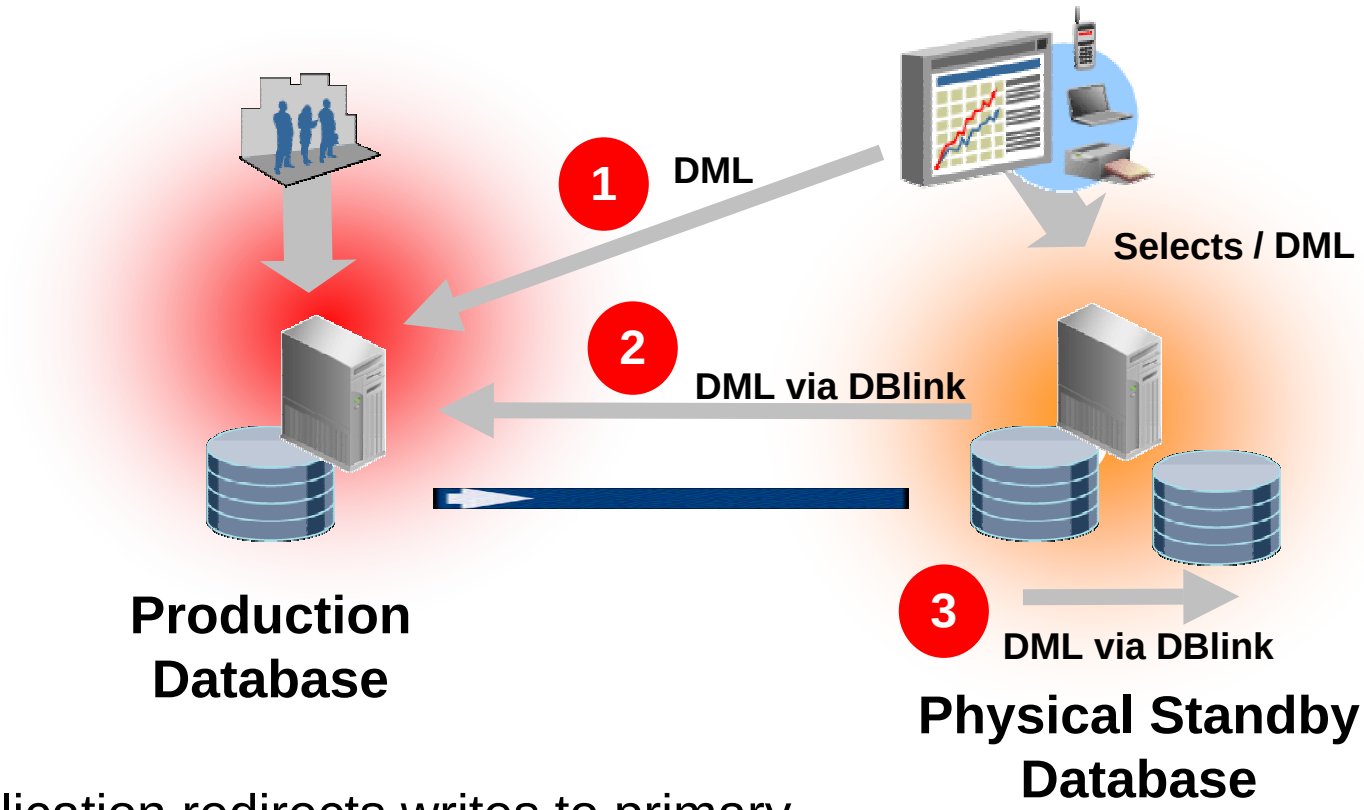


Dell 6950s

- 4 way Dual-Core AMD Opteron Processor 8212
- 8 GB RAM
- OEL 4.5 x86_64 (2.6.9-55.0.0.0.2.ELsmp)

Active Data Guard 11*g*

Three Read-Mostly Application Models



1. Application redirects writes to primary
2. Writes redirected to primary via database link
3. Writes redirected to a separate database via a database link

Redirecting Writes at the Database Level

- If writes must be persistent and available to all client applications accessing the database
  - Redirect writes to the primary database
  - Writes will be protected by the standby
  - Writes available to all production application users
- Otherwise . . .
  - Use a second, local database for best performance
    - Zero impact on primary performance
  - Reduce overhead on standby with DBlink using IPC protocol
    - See Oracle Database Net Services Reference

      http://download.oracle.com/docs/cd/B28359_01/network.111/b28317/protocoladd.htm#sthref574

Creating DBlinks for the Standby

- DBlinks used by the standby to redirect writes are created on the primary and propagated to the standby via redo
  - On the Primary

```
SQL> CREATE DATABASE LINK rtq_prmy USING 'rtq';
```

  - On the standby

```
SQL> insert into emp@rtq_prmy values (999,'SMITH','GEEK',999,sysdate,1,0,30);
1 row created.
SQL> select * from emp where empno=999;

     EMPNO ENAME      JOB              MGR HIREDATE         SAL       COMM
---------- ---------- --------- ---------- --------- ---------- ----------
       999 SMITH      GEEK             999 23-OCT-07          1          0
```

Configuring DBlinks

- Include all hosts in the Oracle Net `ADDRESS_LIST` to enable connect time failover
- Oracle Net alias used by DBlink should reference role specific service names
- Set `OUTBOUND_CONNECT_TIMEOUT` in sqlnet.ora for fast `ADDRESS_LIST` transversal
- Evaluate setting `RECV_TIMEOUT` and `SEND_TIMEOUT`
  - See Oracle Database Net Services Reference

    http://download.oracle.com/docs/cd/B28359_01/network.111/b28317/protocoladd.htm#sthref574

Session State with DBlinks

- Many applications perform connection auditing using session state and USERENV function
- When using a DBlink USERENV values are reset to the DBlink session with the remote database
- Use a stored procedure to collect local USERENV values and insert into the primary database
- The NLS_% parameters of the local session are automatically propagated to remote sessions

Note: An example will be provided in a future Best Practices Paper for Active Data Guard at http://www.oracle.com/technology/deploy/availability/htdocs/maa.htm

Redirecting DML

- Use synonyms to hide DBlinks from the application
  - On the primary:

```
SQL> rename emp to emp_hidden;
Table renamed.
SQL> create synonym emp for emp_hidden@rtq_prmy;
Synonym created.
```

  - On the standby:

```
SQL> insert into emp values
(999,'SMITH','GEEK',999,sysdate,1,0,30);
1 row created.
```

- Be sure to test primary performance impact
  - Alternatively - modify the application to redirect writes to the primary

Redirecting DDL

- Remote procedure uses dynamic SQL to do DDL on the primary
  - On the primary:

```
CREATE OR REPLACE PROCEDURE do_ddl(STRING IN varchar2) AS
 BEGIN
 execute immediate string;
END;
```

  - On the standby:

```
SQL> exec do_ddl@rtq_prmy('create table mts.foo (col1 number)');
BEGIN do_ddl@rtq_prmy('create table mts.foo (col1 number)'); END;

ERROR at line 1:
ORA-04053: error occurred when validating remote object MTS.DO_DDL@RTQ_PRMY
ORA-00604: error occurred at recursive SQL level 1
ORA-16000: database open for read-only access
```

  - Why did it fail?

Redirecting DDL

- A remote procedure can not be invoked from a read-only database
- Workaround: place the remote procedure call in a stored procedure
  - On the primary:

```
CREATE OR REPLACE PROCEDURE call_do_ddl(STRING IN varchar2) AS
 begin
  do_ddl@rtq_prmy(string);
end;
```

  - On the standby

```
SQL> exec call_do_ddl('create table foo2 (col1 number)');
PL/SQL procedure successfully completed.
SQL> select * from foo2@rtq_prmy;
no rows selected
```

Redirecting Writes – Additional Info

- Applications that require unique values can use SYS_GUID SQL on the read-only standby
  - Alternatively you may use the primary to get sequences

```
SQL> select customers_seq.nextval from dual@rtq_prmy;
   NEXTVAL
----------
      1003
```

- To minimize PL/SQL invalidation because of remote procedure calls set the following database parameter in the parameter file:

```
REMOTE_DEPENDENCIES_MODE=SIGNATURE
```

MAA Best Practices

- ✓ Reporting application work load
  - ✓ Selects only
  - ✓ DML
- Routing user connections
  - New connections
  - Role transitions
- Optimizing performance
  - SQL Tuning
  - Redo Apply Tuning

Routing New User Connections

- Primary and reporting applications should connect using role specific service name
- The Oracle Net alias should list all hosts in the ADDRESS_LIST to accommodate role changes

Primary Application

```
Sales_RW =
  (ADDRESS_LIST=
    (ADDRESS=(PROTOCOL=TCP)
    (HOST=hasun01)
    (PORT=1521))
    (ADDRESS=(PROTOCOL=TCP)
    (HOST=hasun02)
    (PORT=1521))
        (CONNECT_DATA =
      (SERVER = DEDICATED)
      (SERVICE_NAME = sales_rw)))
```

Reporting Application

```
Sales_RO =
  (ADDRESS_LIST=
    (ADDRESS=(PROTOCOL=TCP)
    (HOST=hasun01)
    (PORT=1521))
    (ADDRESS=(PROTOCOL=TCP)
    (HOST=hasun02)
    (PORT=1521))
        (CONNECT_DATA =
      (SERVER = DEDICATED)
      (SERVICE_NAME = sales_ro)))
```

Routing User Connections

Role Transitions

- User connections to the read only standby will be disconnected as part of the failover/switchover
- Once new primary is up the primary service is enabled automatically via service management trigger
- Clients connected to old primary are notified via FAN to reconnect
- Reconnection logic routes connection quickly to the new primary
- Read only service is started manually after the role transition completes

note: see appendix for more details

MAA Best Practices

- ✓ Reporting application work load
  - ✓ Selects only
  - ✓ DML
- ✓ Routing user connections
  - ✓ New connections
  - ✓ Role transitions
- Optimizing performance
  - SQL Tuning
  - Redo Apply Tuning

Determining Query Latency

- From Primary (requires database link)
  - Create database link to Active Data Guard Standby and use the query below

```
select scn_to_timestamp((select current_scn from
v$database))-scn_to_timestamp((select current_scn
from v$database@adg)) from dual;
```

- If you do not wish to connect to the Primary - determine the value for APPLY LAG for a “best estimate”
  - Use Enterprise Manager monitoring
  - Query **V$DATAGUARD_STATS**

```
select value,unit,time_computed from
v$dataguard_stats where name='apply lag';
```

SQL Tuning and Active Data Guard

- Tuning standby queries can be performed from the primary database
  - Primary and standby have identical structure
  - Execution plan will be the same between the two databases
- Use of SQL Trace is supported on a read only database
- ASH and Tkprof are not supported on a read-only database
- Use V$SQLAREA and V$SQL_PLAN on the standby to identify expensive queries

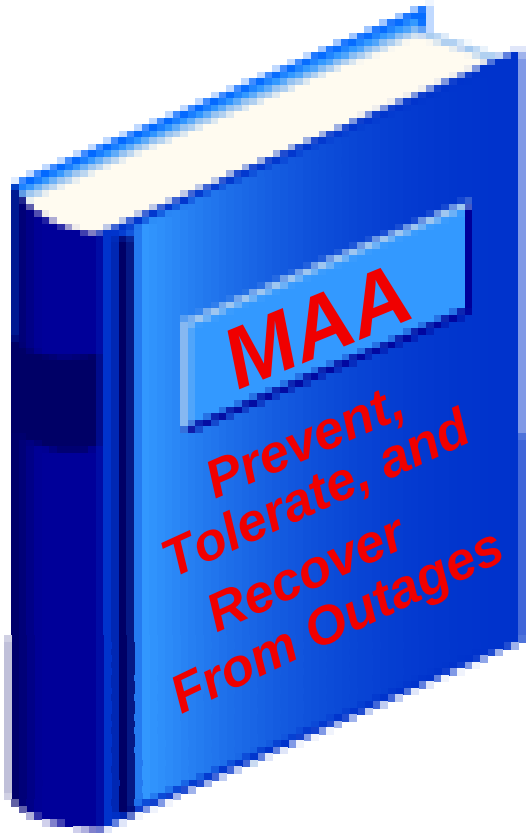
Redo Apply - Tuning Media Recovery

- Big performance boost in Oracle Database 11*g*
  - Up to 100% increase in redo apply performance
  - Tuning should be unnecessary unless standby is undersized
- New standby statspack in Oracle Database 11*g*
  - Create stdbyperf user on primary
  - Add standby databases and instances
  - Execute snaps
  - Generate reports
  - See MetaLink Note 454848.1
- Requires perfstat user and statspack installation

note: see appendix for detailed Redo Apply 11g tuning procedures

Maximum Availability Architecture (MAA)

Integrated set of HA best practices



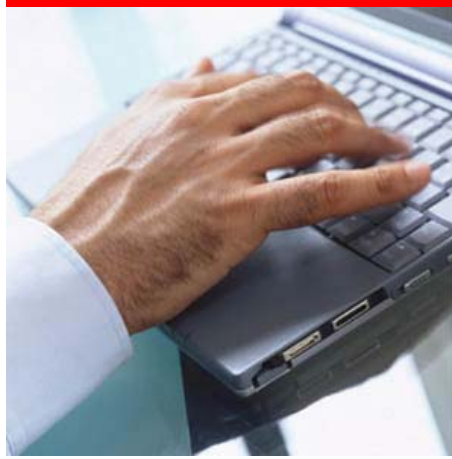
MAA provides a blueprint for achieving HA

- Operational best practices
- Prevent, tolerate, and recover

- Tested, validated, and documented
  - Database, storage, cluster, network
  - Applications and mid-tier
- Active Data Guard Best Practices will be published shortly

otn.oracle.com/deploy/availability

Agenda



- Introduction
- Enabling Active Data Guard
- Best Practices
- Amazon Experience
- Appendix
  - Oracle Database 11g Media Recovery Performance
  - Media Recovery Performance Tuning
  - Client Failover

Active Data Guard

@Amazon.com

Grant McAlister
Principal Database Engineer

amazon.com

Uses of Active Data Guard

- Reporting copy of a primary system
  - Typically different queries than primary system
  - Gap may be in hours
    - (i.e. data from last night/last business day)
  - One copy

- Scalable read store
  - Subset of queries that run on primary
  - Gap in seconds/minutes - not hours
  - As many copies as needed to handle read load

amazon.com

Physical standby for reporting

- 8i
  - Destructive copy of database rolled forward to point in time.
  - Could use 3rd mirror to make this simpler
- 9i – 10g
  - Read-Only Mode – non destructive
- 11g
  - Real Time Query – Active Data Guard

amazon.com

Active Data Guard for Scalable Read

- Allow read only queries to scale beyond single db
- Higher availability for read only queries
- Can be configured to shed extra reads
- More efficient use of hardware

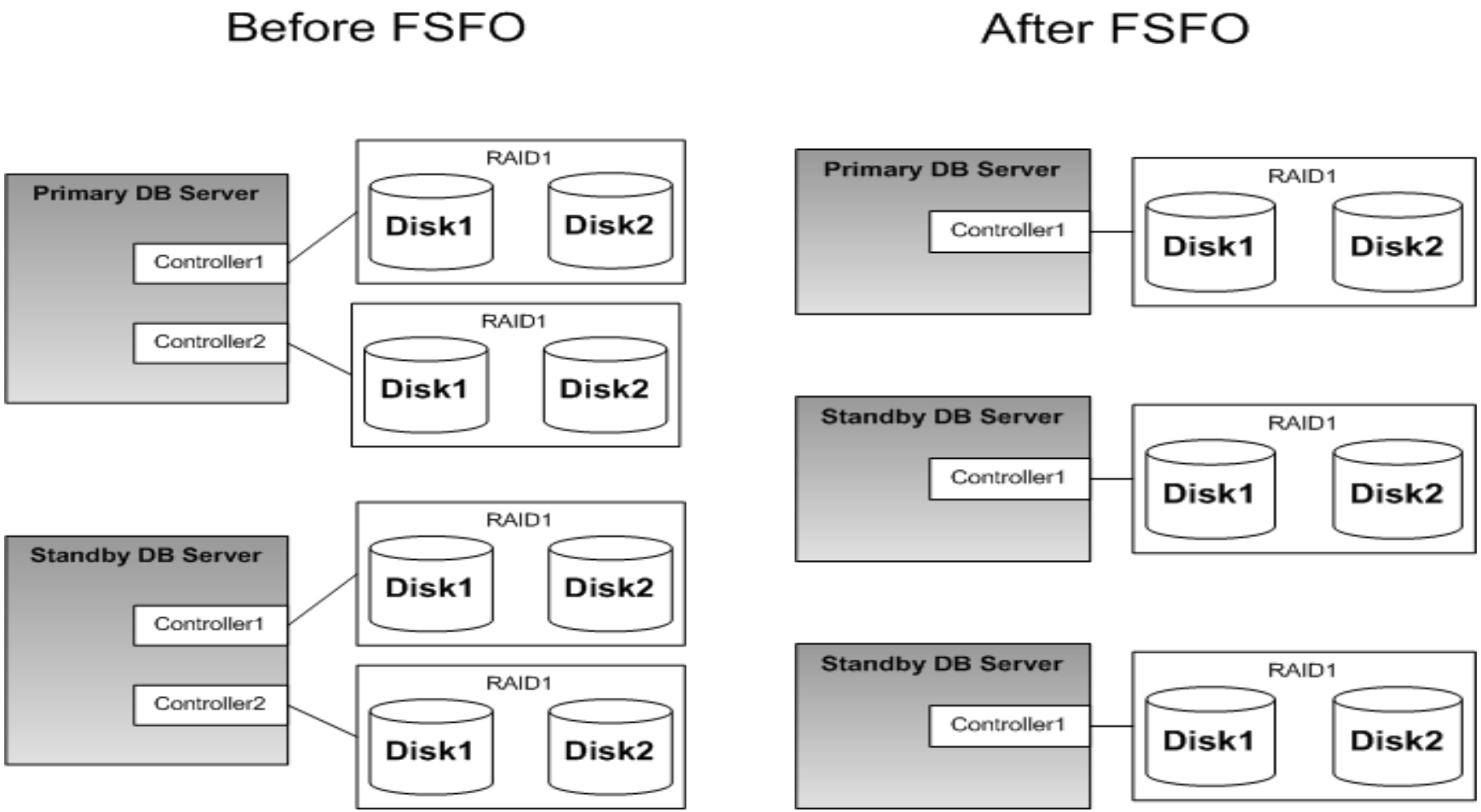
amazon.com

Data Guard Fast-Start Failover

- Ability to use less expensive equipment
  - Reduced need for multiple controllers for redo mirroring
  - Possible to depend on standby servers for protection
- Need 1 standby but using 2 is much better
  - Efficiency is <50% with 1 standby and <33% with 2
- How to use this extra equipment efficiently?
  - Active Data Guard makes 60+% utilization possible

amazon.com

Possible Hardware Changes



How to track data state of the standby

- Comparing CURRENT_SCN from v$database between primary and standby along with SYSTIMESTAMP
- Only accurate as your systems clock drift (ntpd)
- Example

Primary SCN	Primary Time	Standby SCN	Standby Time
4248440	1,193,156,891,753	-	-
4248446	1,193,156,892,764	-	-
4248**452**	1,193,156,89**3,774**	-	-
4248**459**	1,193,156,89**4,786**	4248**453**	1,193,156,89**4,788**

amazon.com

Clients view of Active Data Guard

- How do we inform the client about how far behind the standby is?

- Many possible methods
  - Interceptor layer that rejects queries that exceed window
  - Broadcast current state to the clients
  - Take database out of service when gap exceeds threshold

amazon.com

Our Experience

- It works!!
  - Throughput
    - Can push hundreds of Megabytes of redo per minute
  - Data gap
    - Average is less than 1 second (test measurement granularity)
    - Spikes of less than 10 seconds

amazon.com

Database HA Sessions from Oracle Development

Monday, Nov 12

- S291483 - The Fastest and the Most Cost-Effective Backup for Oracle Database: What's New in Oracle Secure Backup 10.2, 11:00 am - 12:00 pm, Moscone South 304
- S291492 - Oracle Database 11*g*: Next-Generation High Availability, 12:30 - 1:30 pm, Moscone South 103
- S291923 - Implementing Oracle Maximum Availability Architecture (MAA) at Allstate Insurance Using Oracle 10*g* RAC, ASM, Oracle Data Guard and Oracle Grid Control, 3:15 - 4:15 pm, Moscone South 304
- S291484 - Oracle Database 11*g* Data Repair Technologies: Comprehensive, Intelligent Recovery, 4:45 - 5:45 pm, Moscone South 304

Tuesday, Nov 13

- S290710 - Maximum Availability Architecture Best Practices: Oracle E-Business Suite 12, 12:15 - 1:15 pm, Marriott Salon 10 & 11

Wednesday, Nov 14

- S291915 - What's New in Oracle Data Guard 11*g*: Revolutionizing Data Protection and Availability, 9:45 - 10:45 am, Moscone South 304

Database HA Sessions from Oracle Development

Wednesday, Nov 14

- S291487 - Backup and Recovery Best Practices for Very Large Databases (VLDB), 11:15 am - 12:15 pm, Moscone South 304
- S291920 - Oracle Active Data Guard: How to Utilize Your Standby Databases for Production Workload - What They Didn't Print in the Manuals, 3:00 - 4:00 pm, Moscone South 304
- S291917 - Oracle Data Guard Tips and Tricks: Direct From Oracle Development, 4:30 - 5:30 pm, Moscone South 102

Thursday, Nov 15

- S291495 - Oracle Streams Replication and Advanced Queuing (AQ): What's New in Oracle Database 11*g*, 8:30 - 9:30 am, Moscone South 304
- S291499 - Best Practices for Implementing Replication with Oracle Streams in Oracle Database 10*g* and 11*g*, 10:00 - 11:00 am, Moscone South 304
- S291525 - Maximum Availability Architecture (MAA) Best Practices: Online Patching, Rolling Upgrades and Planned Maintenance with Minimal Downtime with Oracle Database, 11:30 am - 12:30 pm, Moscone South 104
- S290542 - Maximum Availability Architecture (MAA) Best Practices for Siebel 8.0, 2:30 pm - 3:30 pm, Marriott Salon 10 & 11

Database HA Demos From Oracle Development

Monday, Nov 12 – Thursday, Nov 15
Oracle DEMOgrounds, Moscone West

Oracle Active Data Guard

Oracle Streams: Replication and Advanced Queuing

Oracle Secure Backup

Recovery Manager (RMAN) and Flashback Technologies

Maximum Availability Architecture

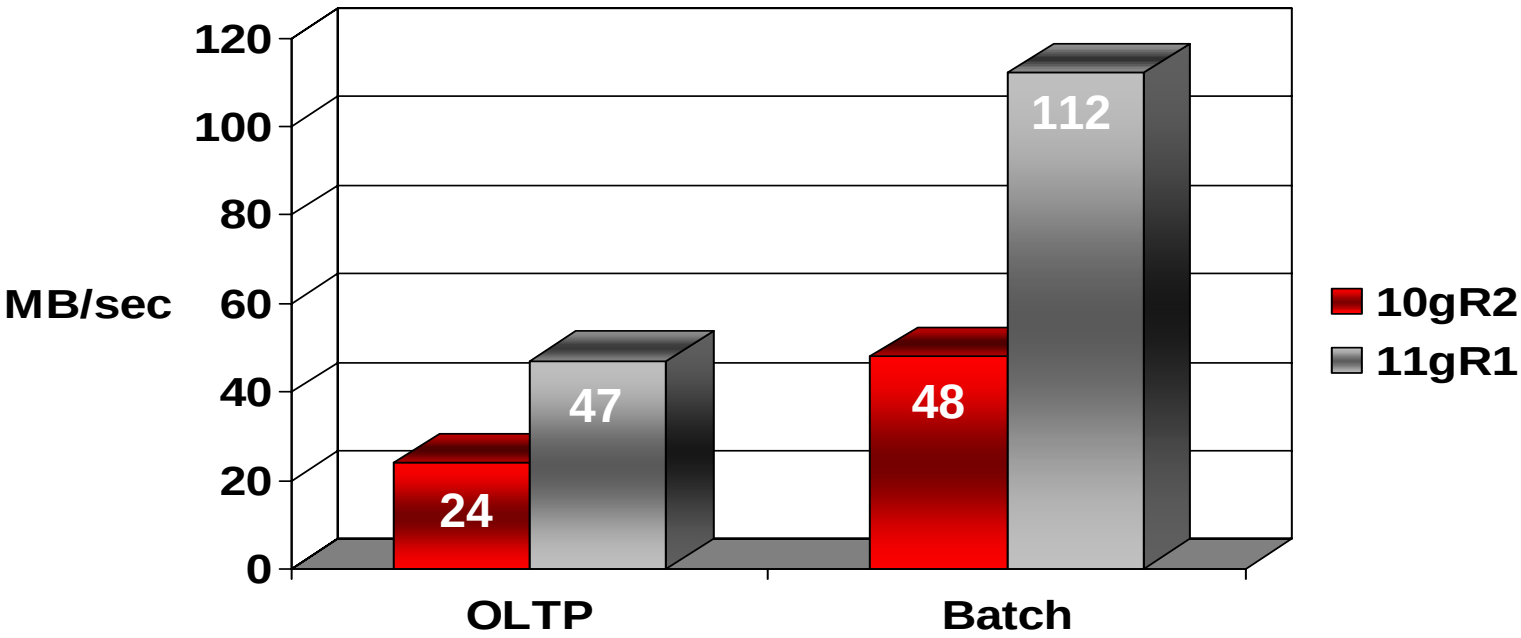

Q&A
QUESTIONS
ANSWERS

Appendix

- Oracle Database 11g Media Recovery Performance
- Media Recovery Performance Tuning
- Client Failover

Oracle Database 11*g*

Media Recovery Performance Improvements



- Up to 100% Performance Improvement

Tuning Media Recovery

- In the rare cases where media recovery doesn't maintain pace
  - Analyze top wait events
  - Identify I/O bottlenecks in recovery area and data area
  - Large gain when adding more spindles
  - Monitor CPU usage

Tuning Media Recovery

- Standby statspack includes the same information as regular statspack
  - Top 5 wait events
  - Memory and database statistics
  - Latch information
  - Global Cache statistics
  - I/O and OS stats
- Also includes information specific to a standby
  - Output from V$RECOVERY_PROGRESS
  - Output from V$MANAGED_STANDBY

Tuning Media Recovery

```
Top 5 Timed Events                                                    Avg %Total
~~~~~~~~~~~~~~~~~~                                                   wait   Call
Event                                            Waits    Time (s)   (ms)   Time
---------------------------------------- ------------ ----------- ------ ------
latch free                                  1,047,968         498      0   72.8
log file sequential read                        6,989          63      9    9.2
checkpoint completed                              121          30    246    4.3
parallel recovery read buffer free              1,168          14     12    2.0
parallel recovery change buffer free           46,717          10      0    1.4
          ------------------------------------------------------------
------snip--------
Recovery Start Time Item                        Sofar Units   Redo Timestamp
------------------- ----------------- -------------- ------- ------------------
09-Oct-07 06:50:52  Active Apply Rate          50,159 KB/sec
09-Oct-07 06:50:52  Active Time                   242 Seconds
09-Oct-07 06:50:52  Apply Time per Lo              11 Seconds
09-Oct-07 06:50:52  Average Apply Rat          51,142 KB/sec
09-Oct-07 06:50:52  Checkpoint Time p               0 Seconds
09-Oct-07 06:50:52  Elapsed Time                  266 Seconds
09-Oct-07 06:50:52  Last Applied Redo    404,648,753 SCN+Tim 09-Oct-07 05:24:32
09-Oct-07 06:50:52  Log Files                     20 Files
09-Oct-07 06:50:52  Redo Applied               12,286 Megabyt
          ------------------------------------------------------------
```

Tuning Media Recovery

- Monitor recovery rate using “Recovery Progress Stats” section of standby statspack
  - **Average Apply Rate**: Redo Applied / Elapsed Time: includes time spent actively applying redo and time spent waiting for redo to arrive.
  - **Active Apply Rate**: (Redo Applied / Active Time) moving average over the last 3 minutes. Does not include time spent waiting for redo to arrive.
  - **Apply Time per Log**: Average time spent actively applying redo in a logfile.
  - **Checkpoint Time per Log**: Average time spent for a log boundary checkpoint.

Routing User Connections

Failover of the read-write service on the primary

- User connections to the “sales_ro” service on the read-only standby will be disconnected as part of the failover
- Once the standby transitions to be the new primary and is started, the “sales_rw” service is enabled automatically via a manage_service trigger, described in the paper below
- Clients connected to old primary are notified via FAN to reconnect
- Reconnection logic routes connections quickly to the “sales_rw” service on the new primary.
- Best Practices for Automating Client Failover details in:

  http://www.oracle.com/technology/deploy/availability/pdf/MAA_WP_10gR2_ClientFailoverBestPractices.pdf

Routing User Connections

Failover of the read-only service on the standby

- Failover of the reporting application services running on the Active Data Guard Standby is done manually
- Following a failover determine where the reporting application service should be started
  - If new primary can support all sevices, start “sales_ro” on primary
  - If new primary cannot support both
    - Start “sales_ro” on another Active Data Guard Standby
    - Start “sales_ro” after old primary is reinstated as a new standby
- In all cases, the “sales_ro” service is started manually

  alter system set service_names=‘sales_ro”;
- Once service is running restart reporting application

Client Failover Best Practices

- JDBC clients configured for Fast Connection Failover and FAN ONS
  - ONS daemons on primary and standby clusters
  - JDBC client uses remote subscription to all daemons
- OCI client configured for FAN OCI
  - AQ_HA_NOTIFICATIONS
- Implement fast ADDRESS_LIST transversal
  - OCI – OUTBOUND_CONNECT_TIMEOUT
  - JDBC - **SQLnetDef.TCP_CONNTIMEOUT_STR**

ORACLE IS THE INFORMATION COMPANY