
What's New in 11gR2 RAC

Caleb Small, BSc, ISP
Caleb@Caleb.com

Testbed Setup

- 2-node cluster
- Shared raw storage over iSCSI
- 3 gig-ether ports on separate switches
- Oracle Enterprise Linux 5U4, 64-bit
- 11gR2 Grid Infrastructure & RAC
- 4 ASM disk groups
 - DATA
 - FLASH
 - CRS
 - ACFS

Installation

- New (complex) network requirements
- New GRID user + environment
- 2 Oracle Homes (instead of 3)
 - ASM and CRS together in GRID home
 - Must be different ORACLE_BASE
 - Unset env vars before installing
- OCR/Vote ***not supported*** on raw
 - Use ASM or CFS instead
- Cluster Time Synchronization Service
- Software only Grid install possible

Administration

Significant Changes

- User specific actions (grid, oracle, root)
- Startup & shutdown
- Cluster status `crs_stat` deprecated
- Listener configuration
- Client configuration (tnsnames)
- OCR/Vote Disk backup
- ACFS management

User Specific Actions

Three different environments

- User ORACLE

- `srvctl` – manage instances & services

- User GRID

- `lsnrctl` – manage listeners

- `asmcmd` – manage ASM

- User GRID as ROOT

- `crsctl` – manage nodes & cluster

- `ocrconfig` – manage OCR/Vote

Startup & Shutdown

- ASM must start *before* CRS
- More components to manage
 - In dependency order w/ `srvctl`
 - Remain shutdown upon restart
 - Demo at end of presentation
- Easier to start & stop nodes or entire cluster

```
su - grid
```

```
su
```

```
crsctl stop cluster -all
```

(Fails if ACFS file system mounted)

Cluster Status

- **crs_stat** (my favorite) is deprecated
 - No longer shows instance status
- Replace by **crsctl status resource -t**
 - Run as GRID
 - Many more components
 - Disk groups, network, oc4j, scan, etc
 - Local vs. Cluster resources

» Demo

Sat Mar 6 12:03:47 PST 2010

NAME	TARGET	STATE	SERVER	STATE_DETAILS
Local Resources				
ora.ACFS.dg	ONLINE	ONLINE	beta1	
	OFFLINE	OFFLINE	beta2	
ora.CRS.dg	ONLINE	ONLINE	beta1	
	ONLINE	ONLINE	beta2	
ora.DATA.dg	ONLINE	ONLINE	beta1	
	ONLINE	ONLINE	beta2	
ora.FLASH.dg	ONLINE	ONLINE	beta1	
	ONLINE	ONLINE	beta2	
ora.LISTENER.lsnr	ONLINE	ONLINE	beta1	
	ONLINE	ONLINE	beta2	
ora.asm	ONLINE	ONLINE	beta1	Started
	ONLINE	ONLINE	beta2	Started
ora.eons	ONLINE	ONLINE	beta1	
	ONLINE	ONLINE	beta2	
ora.gsd	OFFLINE	OFFLINE	beta1	
	OFFLINE	OFFLINE	beta2	
ora.net1.network	ONLINE	ONLINE	beta1	
	ONLINE	ONLINE	beta2	
ora.net2.network	OFFLINE	OFFLINE	beta1	
	OFFLINE	OFFLINE	beta2	
ora.ons	ONLINE	ONLINE	beta1	
	ONLINE	ONLINE	beta2	

Cluster Resources

ora.LISTENER_SCAN1.lsnr	1	ONLINE	ONLINE	beta1	
ora.LISTENER_SCAN2.lsnr	1	ONLINE	ONLINE	beta2	
ora.beta1.vip	1	ONLINE	ONLINE	beta1	
ora.beta2.vip	1	ONLINE	ONLINE	beta2	
ora.oc4j	1	OFFLINE	OFFLINE		
ora.racdb.db	1	ONLINE	ONLINE	beta1	
	2	ONLINE	ONLINE	beta2	Open
ora.racdb.racdb_taf.svc	1	ONLINE	ONLINE	beta2	
	2	ONLINE	ONLINE	beta1	
ora.scan1.vip	1	ONLINE	ONLINE	beta1	
ora.scan2.vip	1	ONLINE	ONLINE	beta2	

SCAN Single Client Access Name

- A single hostname to access the cluster
- Cluster changes are invisible to clients
- Works best with 11gR2 client (n/a on Windows)

```
racdb_taf =
  (DESCRIPTION =
    (ADDRESS = (PROTOCOL = TCP) (HOST = beta-scan) (PORT = 1521))
    (LOAD_BALANCE = YES)
    (CONNECT_DATA =
      (SERVER = DEDICATED)
      (SERVICE_NAME = racdb_taf)
      (FAILOVER_MODE =
        (TYPE = SELECT) (METHOD = BASIC) (RETRIES = 180) (DELAY = 5)
      )
    )
  )
)
```

Client Configuration

- Old way (10.2 client) still works
- No benefit of server side load balancing

```
racdb_taf_old =
  (DESCRIPTION =
    (ADDRESS = (PROTOCOL = TCP) (HOST = beta1-vip) (PORT = 1521))
    (ADDRESS = (PROTOCOL = TCP) (HOST = beta2-vip) (PORT = 1521))
    (LOAD_BALANCE = yes)
    (CONNECT_DATA =
      (SERVER = DEDICATED)
      (SERVICE_NAME = racdb_taf)
      (FAILOVER_MODE =
        (TYPE = SELECT) (METHOD = BASIC) (RETRIES = 180) (DELAY = 5)
      )
    )
  )
)
```

Listener Configuration

- Two Listeners
 - Both run in GRID home
- Local (database) Listener
 - One on each node
 - Registers local instance (and ASM)
- SCAN Listener
 - Up to 3 per cluster
 - Can migrate around cluster
 - Registers all database instances & services
 - Receives Load Balance Advisory

Demo `lsnrctl stat, lsnrctl stat listener_scan1`

```
Connecting to (DESCRIPTION=(ADDRESS=(PROTOCOL=IPC) (KEY=LISTENER)))
```

```
STATUS of the LISTENER
```

```
-----
```

```
Alias                LISTENER
Version              TNSLSNR for Linux: Version 11.2.0.1.0 - Production
Start Date           05-MAR-2010 12:08:41
Uptime                1 days 23 hr. 21 min. 47 sec
Trace Level           off
Security              ON: Local OS Authentication
SNMP                  OFF
Listener Parameter File /u01/app/11.2.0/grid/network/admin/listener.ora
Listener Log File     /u01/app/grid/diag/tnslsnr/beta1/listener/alert/log.xml
```

```
Listening Endpoints Summary...
```

```
(DESCRIPTION=(ADDRESS=(PROTOCOL=ipc) (KEY=LISTENER)))
```

```
(DESCRIPTION=(ADDRESS=(PROTOCOL=tcp) (HOST=192.168.1.221) (PORT=1521)))
```

```
(DESCRIPTION=(ADDRESS=(PROTOCOL=tcp) (HOST=192.168.1.21) (PORT=1521)))
```

```
Services Summary...
```

```
Service "+ASM" has 1 instance(s).
```

```
Instance "+ASM1", status READY, has 1 handler(s) for this service...
```

```
Service "racdb" has 1 instance(s).
```

```
Instance "racdb1", status READY, has 1 handler(s) for this service...
```

```
Service "racdb_taf" has 1 instance(s).
```

```
Instance "racdb1", status READY, has 1 handler(s) for this service...
```

```
The command completed successfully
```

```
Connecting to (DESCRIPTION=(ADDRESS=(PROTOCOL=IPC) (KEY=LISTENER_SCAN1)))
STATUS of the LISTENER
-----
Alias                               LISTENER_SCAN1
Version                             TNSLSNR for Linux: Version 11.2.0.1.0 - Production
Start Date                          05-MAR-2010 12:08:40
Uptime                              1 days 23 hr. 26 min. 54 sec
Trace Level                         off
Security                            ON: Local OS Authentication
SNMP                                 OFF
Listener Parameter File             /u01/app/11.2.0/grid/network/admin/listener.ora
Listener Log File                   /u01/...diag/tnslnr/beta1/listener_scan1/alert/log.xml
Listening Endpoints Summary...
  (DESCRIPTION=(ADDRESS=(PROTOCOL=ipc) (KEY=LISTENER_SCAN1)))
  (DESCRIPTION=(ADDRESS=(PROTOCOL=tcp) (HOST=192.168.1.26) (PORT=1521)))
Services Summary...
Service "racdb" has 2 instance(s).
  Instance "racdb1", status READY, has 1 handler(s) for this service...
  Instance "racdb2", status READY, has 1 handler(s) for this service...
Service "racdb_taf" has 2 instance(s).
  Instance "racdb1", status READY, has 1 handler(s) for this service...
  Instance "racdb2", status READY, has 1 handler(s) for this service...
The command completed successfully
```

```
SQL> show parameter listener
```

NAME	TYPE	VALUE
Local_listener	string	(DESCRIPTION= (ADDRESS_LIST= (ADDRESS= (PROTOCOL=TCP) (HOST=beta1-vip) (PORT=1521))))
remote_listener	string	beta-scan.erpbackup.com:1521

- Do not use tns alias for remote_listener

SCAN Configuration

- Requires network configuration in place prior to install
- Requires either DNS *or* GNS (Grid Naming Service)
- GNS requires 3 IPs acquired from DHCP
- DNS recommended for “manual” configuration
- 3 additional IPs on public network for SCAN-VIPs
- Single SCAN hostname resolves to 3 IPs
- During installation, DNS resolution provides 3 IPs which are used to create 3 SCAN-VIP / Listener pairs scattered across the cluster

*See Metalink Doc ID Doc ID 887522.1 -
11gR2 Grid Infrastructure SCAN Explained*

DNS Configuration

- Use round robin for up to 3 IPs
- Set Time To Live (TTL) very short
 - Especially for pre 11gR2 clients
- Beware of
 - Routers w/caching DNS
 - Windows DNS client
- Test with repeated `nslookup` / `dig` commands

DNS Round Robin Configuration

```
[oracle@beta1 ~]$ nslookup beta-scan.erpbackup.com
```

```
Server:          64.59.160.13  
Address:         64.59.160.13#53
```

```
Non-authoritative answer:
```

```
Name:   beta-scan.erpbackup.com
```

```
Address: 192.168.1.26
```

```
Name:   beta-scan.erpbackup.com
```

```
Address: 192.168.1.25
```

```
[oracle@beta1 ~]$ nslookup beta-scan.erpbackup.com
```

```
Server:          64.59.160.13  
Address:         64.59.160.13#53
```

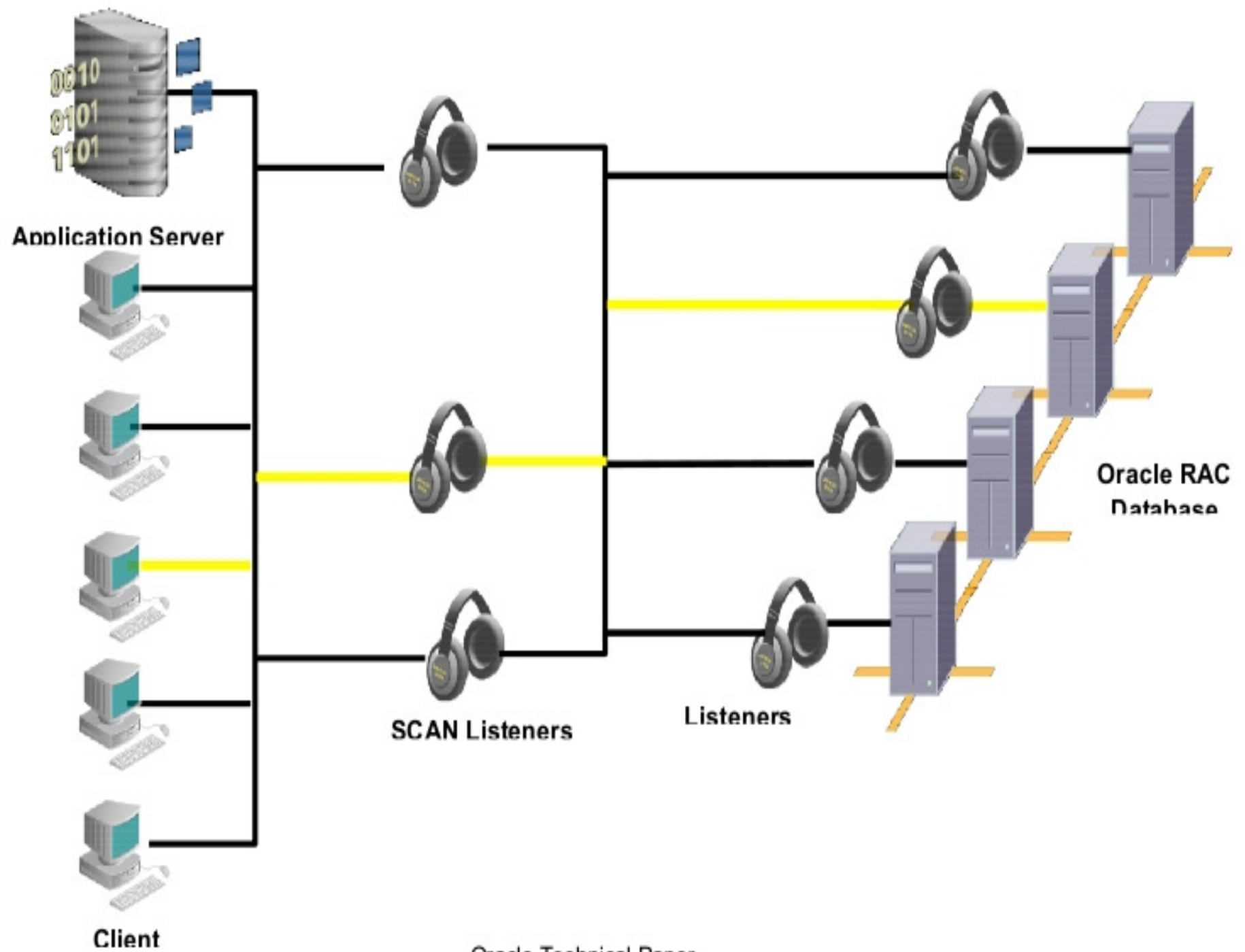
```
Non-authoritative answer:
```

```
Name:   beta-scan.erpbackup.com
```

```
Address: 192.168.1.25
```

```
Name:   beta-scan.erpbackup.com
```

```
Address: 192.168.1.26
```



SCAN – How It Works

1. Client requests DNS resolution of SCAN hostname
2. DNS responds with circulating list of 3 IPs
3. Client chooses first, or random, IP in list
4. Client connects to SCAN Listener at chosen IP and requests database service
5. SCAN Listener is receiving service registrations and load balance advisories from all instances in cluster
6. SCAN Listener chooses least loaded instance offering requested service
7. SCAN Listener re-directs connection request to Local Listener on that node
8. Local Listener accepts request and establishes session.

Testing SCAN

- Determine which nodes are running SCAN Listeners
- Tail the SCAN Listener log files
- Make repeated connections to cluster database from remote client
- Query to determine which instance gets the connection

>>Demo

Testing SCAN - example

```
[grid@beta1 ~]$ crsctl status resource -t
```

```
ora.LISTENER_SCAN1.lsnr
```

```
1 ONLINE ONLINE beta1
```

```
ora.LISTENER_SCAN2.lsnr
```

```
1 ONLINE ONLINE beta2
```

```
[grid@beta1 trace]$ tail -f
```

```
/u01/app/11.2.0/grid/log/diag/tnslsnr/beta1/listener_scan1/trace/  
listener_scan1.log
```

```
[grid@beta2 trace]$ tail -f
```

```
/u01/app/11.2.0/grid/log/diag/tnslsnr/beta1/listener_scan2/trace/  
listener_scan2.log
```

```
[oracle@beta1 ~]$ cat c.sql
```

```
connect system/oracle@racdb_taf
```

```
select instance_name from v$instance;
```

```
SQL> @c
```

```
Connected.
```

```
INSTANCE_NAME
```

```
-----
```

```
racdb1
```

SCAN Maintenance

To re-configure later (not documented), see Metalink Doc ID 972500.1 *How to Modify SCAN Setting after Installation*

Example: Modify DNS to add additional SCAN VIP and change SCAN hostname

```
[root@beta1 bin]# srvctl config scan
SCAN name: beta-scan.psoug.org, Network: 1/192.168.1.0/255.255.255.0/eth0
SCAN VIP name: scan1, IP: /192.168.1.25/192.168.1.25

[root@beta1 bin]# srvctl stop scan_listener
[root@beta1 bin]# srvctl stop scan

[root@beta1 bin]# srvctl modify scan -n beta-scan.erpbackup.com
[root@beta1 bin]# srvctl modify scan_listener -u

[root@beta1 bin]# srvctl start scan
[root@beta1 bin]# srvctl start scan_listener

[root@beta1 bin]# srvctl config scan
SCAN name: beta-scan.erpbackup.com, Network: 1/192.168.1.0/255.255.255.0/eth0
SCAN VIP name: scan1, IP: /192.168.1.26/192.168.1.26
SCAN VIP name: scan2, IP: /192.168.1.25/192.168.1.25
```

OCR/Vote Disk

- Use ASM or CFS, not raw
- Vote Disk backup with `dd` *not supported*
- Vote Disk automatically backed up when:
 - Config parameters changed
 - Add or delete disk
- New OLR Oracle Local Registry
 - Accessible even if CRS is not fully functional
- Automatic OCR backup as before
- Manual backup with

```
ocrconfig -manualbackup
```

OCR/Vote Disk on ASM

- ASM diskgroups now start *before* ASM
- Use a diskgroup with at least 3 disks
 - ASM “disk” = raw LUN
- Use “normal” redundancy for DG and OCR/Vote files
- Vote Disk is written in block headers and not visible as a file
- OCR is visible with ASMCMD
- Explored “Cluster Vulnerability”
 - Could not reliably duplicate

CRS Diskgroup Contents

ASM spfile

```
SQL> show parameter spfile
```

NAME	TYPE	VALUE
spfile	string	+CRS/beta-cluster/asmparameterfile/registry.253.709422419

Vote Disk

```
[grid@beta1 ~]$ crsctl query css votedisk
```

##	STATE	File Universal Id	File Name	Disk group
1.	ONLINE	6e9b66c60dc04fb5bf3fc0fd2de5785f	(ORCL:ASMCRS)	[CRS]

Located 1 voting disk(s).

OCR

```
ASMCMD> ls -al +CRS/beta-cluster/ocrfile
```

Type	Redund	Striped	Time	Sys	Name
OCRFILE	UNPROT	COARSE	MAR 05 12:00:00	Y	none => REGISTRY.255.709422421

Cluster Time Synchronization Service

- Runs automatically all the time
- Addresses common problem of time synchronization within cluster
- If ntpd is running:
 - Runs in observer mode
 - Steps in if time drifts
- If ntpd is not running:
 - Synchronizes all nodes to node1

```
crsctl check ctss
```

ACFS Management

- Optional general purpose cluster file system
- Not for database, nor grid home files
- Looks, acts and feels like regular file system
- Supports snapshots, ASM redundancy, etc
- *Demo:*
 - Create volume
 - Create mount point
 - Register volume (for startup)
 - Mount volume
 - Create files!

References

- Grid Infrastructure Installation Guide for Linux
- RAC Administration and Deployment Guide
- RAC Installation Guide for Linux and UNIX
- Storage Administrator's Guide

Sridhar Avantsa,
Senior Management Consultant, Rolta-TUSC

Hamish Robertson,
Network Specialist, ERP Services Group

Build Your Own Oracle RAC 11g Cluster on Oracle
Enterprise Linux and iSCSI
- by Jeffrey Hunter

<http://www.oracle.com/technology/pub/articles/hunter-rac11gr2-iscsi.html>