Storage Architectures for Oracle RAC

Matthew Zito, Chief Scientist
GridApp Systems

Agenda

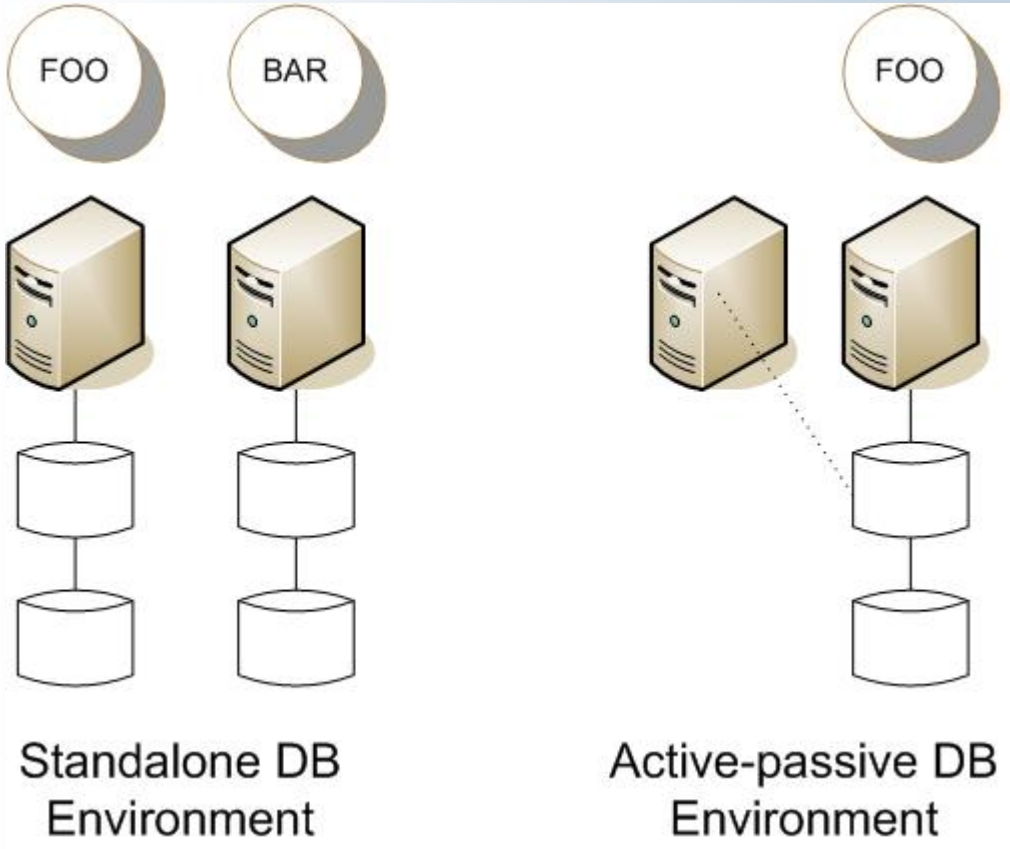
- Oracle RAC Introduction
- Storage Foundations
- Storage and RAC
- Raw Devices
- Clustered Filesystems
- Oracle ASM
- Network File Systems
- Recommended Configuration
- Conclusions/Q&A

Oracle RAC Introduction

- Oracle RAC adoption rates are increasing
- DBAs have come to grips with:
  - Basic changes
  - OS best practices
- Storage in RAC continues to be complex because:
  - Complex support matrix
  - Multitude of different options
  - Storage typically isn't in a DBA's background

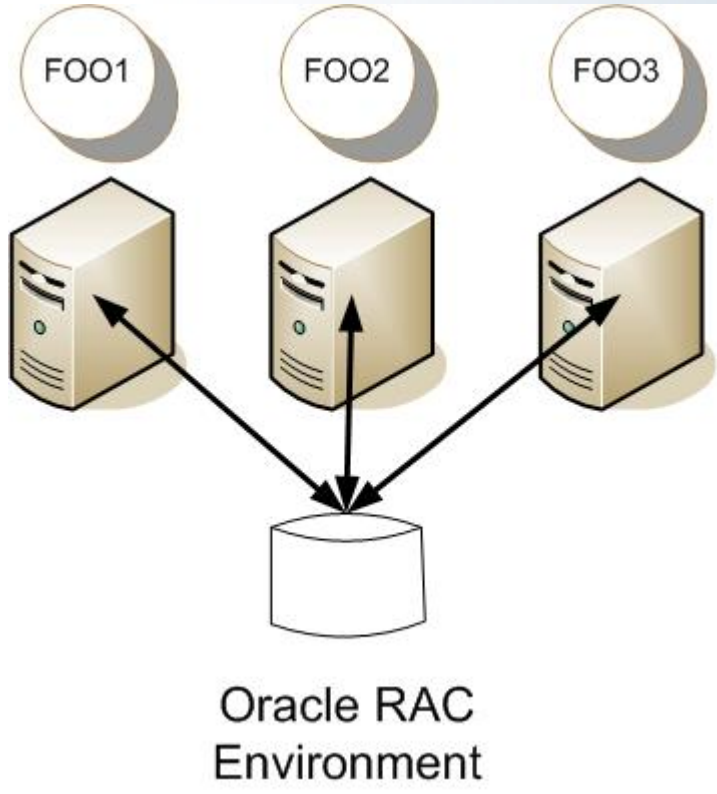
Storage Foundations

- Traditional database model
  - One server, one set of disks
- Active/passive model
  - N servers,
  - one set of disks



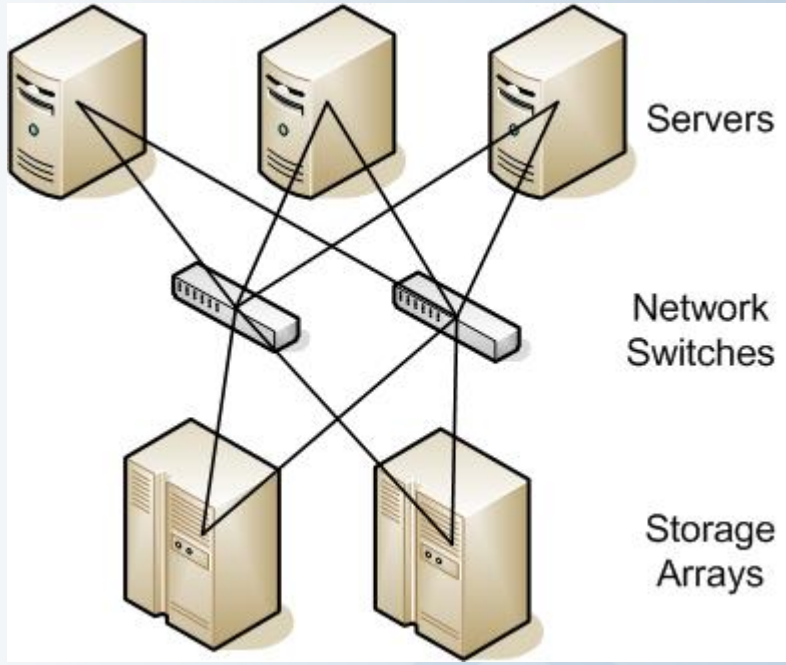
Storage Foundations – Oracle RAC

- Oracle RAC requires *shared* disk access
- N servers, all with concurrent access to the storage



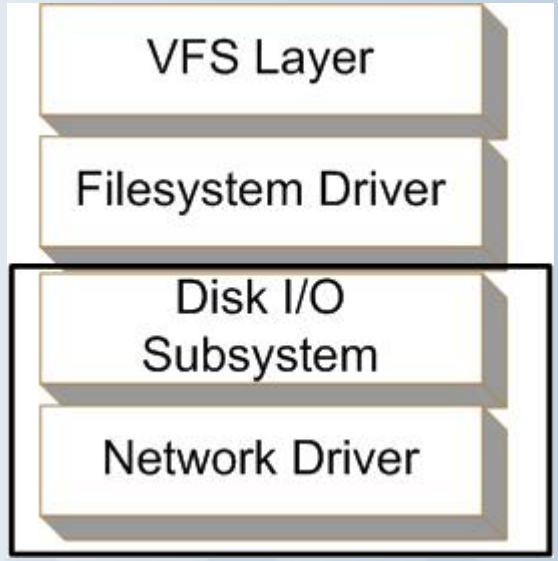
Shared Disk Access

- Requires some sort of networked storage
  - iSCSI
  - Fibre Channel/SCSI
  - NFS
- Typical Network Technologies
  - Ethernet
  - Fibre Channel
- Networked Storage
  - Centralized pool
  - Storage admins allocate it out
  - Designed for scale efficiencies
  - Block- or file-based



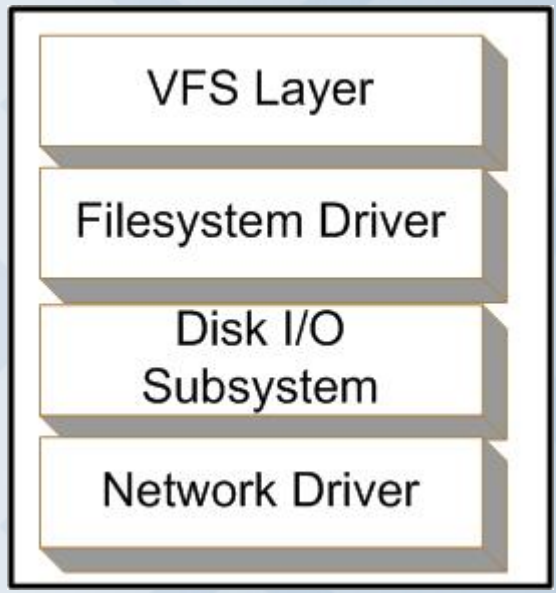
Block Storage

- Fundamentals
  - Traditional disk devices
  - Operates at a low-level
- Internals
  - Below the filesystem
  - Basic IO primitives – read, write, “how big is the device”
  - Provides a generic way to access block storage, abstracts underlying tech
  - Fibre Channel, SCSI, iSCSI



File-based Storage

- Fundamentals
  - Higher up the stack at an OS level
  - More intelligence resides in the OS
- Internals
  - NFS and CIFS (CIFS not Oracle-supported!)
  - Metadata lives within the protocol
    - Creation, access time
    - File sizes, owners, permissions
  - Much richer set of semantics:
    - opendir, read, write, stat
    - File locking



Storage and RAC

- RAC clusters have access to a shared set of storage – problematic because:
  - Not as common a configuration
  - Traditional technologies are not compatible with shared storage access
  - Specialty storage technologies are required
- Four general classes of RAC-suitable solutions for storage on RAC:
  - Raw devices
  - Clustered filesystems
  - ASM
  - NFS

Raw Devices

- Fundamentals
  - A disk or partition where all I/O operations to that device bypass any OS-level caches or buffers and are delivered immediately to the driver subsystem
  - Examples: /dev/raw/raw1, /dev/rdsk/c0t0d0s0, /dev/sda1 when opened with O_DIRECT

```
[root@rh44-ma-012 tmp]# raw /dev/raw/raw1 /dev/sda1
/dev/raw/raw1:  bound to major 8, minor 1
[root@rh44-ma-012 tmp]# ls -l /dev/sda1
brw-rw----  1 root disk 8, 1 Jul  7 08:38 /dev/sda1
[root@rh44-ma-012 tmp]# ls -l /dev/raw/raw1
crw-rw----  1 root disk 162, 1 Sep  8 14:17 /dev/raw/raw1
```

- Advantages:
  - Removes double-buffering problem
  - Guaranteed writes
  - Minimal OS overhead from performance perspective

Raw Devices - continued

- Disadvantages
  - Oracle treats each raw device or raw partition as one file – can result in many many raw devices required
  - There's no way to get an accurate picture at an OS level of how much disk space is in use – no df, find, ls -l
  - Backup and recovery solutions that do backups at an OS level are unaware of raw devices
  - Can only support database files – ocr, voting, dbf files, redo logs, etc.
  - As of 12g, raw devices are no longer supported by Oracle

Clustered Filesystem Basics

- Fundamentals
  - Most familiar environment; resembles the traditional filesystems used in non-RAC environments
  - Emulates a traditional filesystem with extra intelligence to handle shared negotiation of metadata, etc.
- Advantages:
  - Simplified day-to-day administration, all existing tools, scripts work as before
  - Simplified storage configuration
  - Can be used to store non-database files
- Disadvantages:
  - Additional initial configuration complexity
  - Adds another product/solution to the database stack
  - Can add performance overhead

Clustered Filesystems & Oracle

- Multitude of Oracle-supported cluster filesystems
  - Specific support matrix
  - ALL CFS solutions require an additional clustering technology to run on the system
- Except for Linux, all of the clustered filesystem options are provided by a third-party vendor
- On Linux, Oracle has written its own CFS, OCFS2 (Oracle Clustered Filesystem version 2)
  - Supports datafiles, ORACLE_HOME, and general purpose file storage
  - Integrated into the mainline Linux kernel
  - OCFS2 is lacking in online scalability compared to some third-party vendors

Clustered Filesystem RAC Configurations

- Shared ORACLE_HOME:
  - Some CFS architectures support sharing ORACLE_HOME installs across nodes
  - Reduces total disk space requires, and number of homes to manage across nodes
  - Creates SPOFs and increases patch complexity (impossible to patch one home without patching all)
- Oracle files on CFS:
  - Datafiles, ocr, voting, all on CFS
  - Reduces number of disk devices
  - May run into limitations on the CFS concerning sizing, scaling, etc.
- Logs, admin directories, scripts:
  - Useful to put on a CFS for centralization purposes
  - CDSL an option, but more complex – better to name directories based on node name

Oracle ASM

- ASM is a stripped down Oracle instance or RAC database
- ASM's concept of volume management is very simplistic compared to "traditional" volume managers
  - Disks are grouped together as named "disk groups"
  - Disks can be added to disk groups online
  - No concept of plexes, snapshots, subdisks
- Primary advantages of ASM over raw devices are
  - It removes the "one disk, one datafile" requirement
  - Adds limited support for RAID
- ASM is cross-platform – works with all OS vendors on 10g+

ASM and 11g

- Oracle 11g Enhancements:
  - Now with ASM mirroring, Oracle does not need to completely rebuild all of the data on that disk if it fails
  - Addition of the “sysasm” group – separates out administrative overhead
- Future releases of Oracle are expected to extend the ability of ASM to hold non-database files
- 10g and 11g Standard Edition *require* ASM

NFS

- Network File System (NFS)
  - Started by Sun Microsystems as a generic fileserver solution
  - Originally UNIX-only, some windows support available today
  - In the database world, generally dismissed as slow and unreliable
- In NFS environments, the NFS server or array acts as a CFS, arbitrating access, locks, and metadata updates
  - Think of it like a CFS with the cluster and intelligence running on the storage array
  - Frees the server to focus on driving IO to the storage
  - NFS servers sometimes have additional functional capabilities over traditional block storage arrays

NFS Continued

- NFS & RAC
  - Looks like a clustered filesystems
  - All database components can live on NFS, but only certain OS and NAS array configurations are supported – check Metalink
  - Specific mount options are required
- Disadvantages:
  - Per MB, NFS storage is often more expensive than Fibre Channel or iSCSI
  - Certain workloads may not scale well on NFS platforms, though most will.

Recommended Configuration

- Certainly, no one size fits all
- However, GridApp has seen one configuration consistently offer a blend of manageability, scalability, and performance
- Three core components in use:
  - Local disks of the servers
  - Clustered Filesystem (OCFS2)
  - ASM

Recommended Configuration

- Local disks:
  - ORACLE_HOME – separate per-node, and per database
- OCFS2:
  - OCR
  - Voting
  - (optionally) archive_log_dest
- ASM:
  - Local disks of the servers
  - Clustered Filesystem
  - ASM

Recommended Configuration

- Advantages:
  - Minimum of disk devices required
  - Allows scripts, etc. to be centrally shared
  - ASM provides storage and capacity growth
- Disadvantages:
  - Multiple moving parts
  - Additional complexity

Conclusions

- Oracle RAC dramatically increases the infrastructure complexity surrounding its configuration
- With storage, there is a particular concern due to the breadth of options available
- Raw devices, NFS, CFS, and ASM all have particular advantages and disadvantages
- A recommended storage infrastructure uses all of these technologies

Q & A