

ORACLE®

THE **INFORMATION** COMPANY

Clemens Bleile

Senior Principal

Consultant

RAC with ASM on Linux x86-64

Experience with an
extended cluster
(stretched cluster)

The Issue

- Issue
 - Overloaded server
 - Hundreds of applications running against a single database (Oracle 9.2.0.6 on Sun Solaris).
 - Protected by a failover cluster (Veritas)
- Decision: New DWH apps on new DB server
- Requirements
 - Meet cost, stability, availability and performance criteria
 - Do not break internal standards (e.g. multipathing)

The Requirements

- Meet cost criteria
 - RAC on Linux with ASM is an option.
- Meet availability criteria
 - Use cluster technology.
 - Test in a PoC
- Meet stability and performance criteria
 - Customer provided data and queries of their internal systems for a DWH benchmark.
 - Benchmark: Oracle, <DB-vendor 2>, <DB-vendor 3>.
 - Test in a PoC

The Benchmark

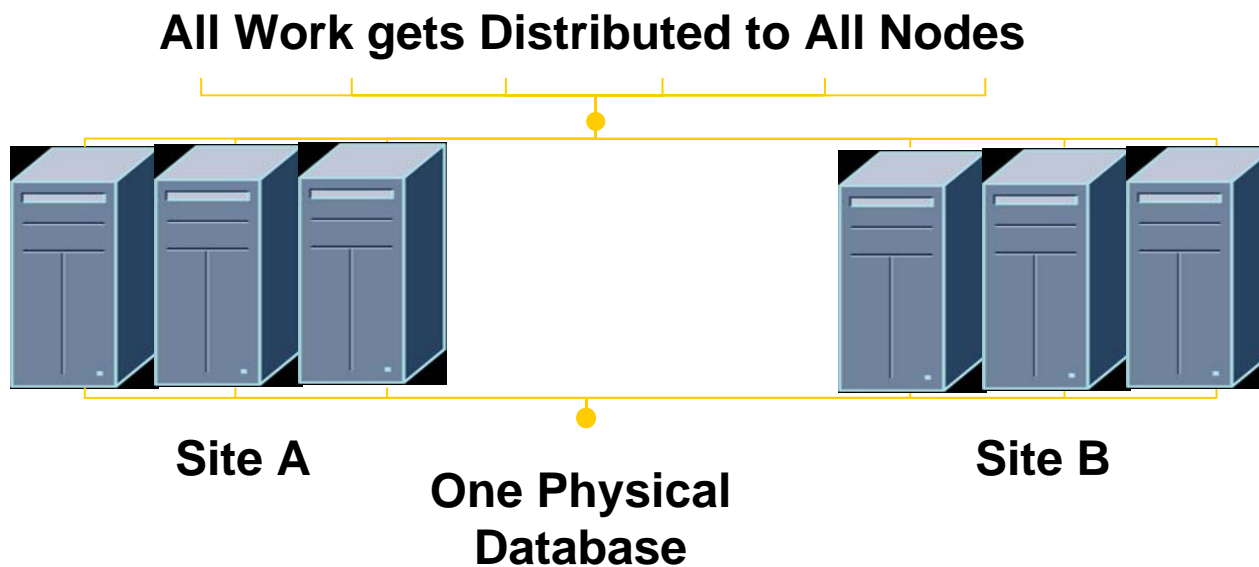
- 1 week at the HP-center in Böblingen
- Hardware:
 - 4 servers with 4 XEON 64Bit CPUs, 16G RAM.
 - EVA 8000 storage with 156 disks.
- Result:
 1. Oracle/<DB-vendor 2>
 2. <DB-vendor 3>
- Why Oracle?

Proof of Concept (PoC)

- Test extended RAC cluster configuration
 - Installation
 - High Availability
 - Performance
 - Workload Management
 - Backup/Recovery
 - Maintenance
 - Job Scheduling
 - Monitoring

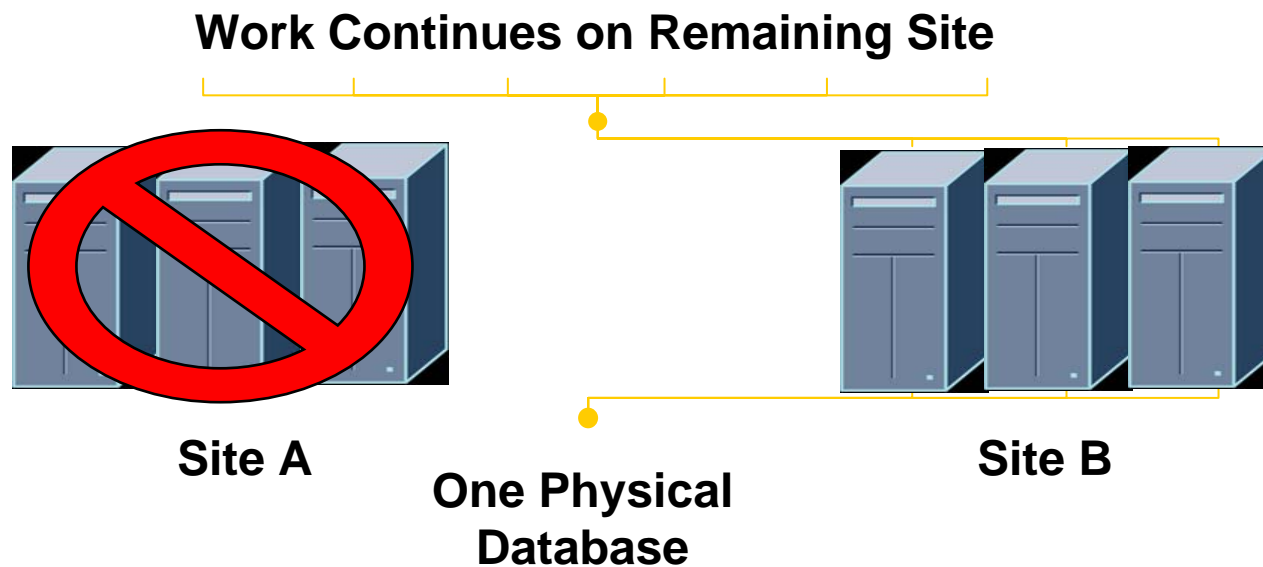
Benefits of an extended cluster

- Full utilization of resources no matter where they are located



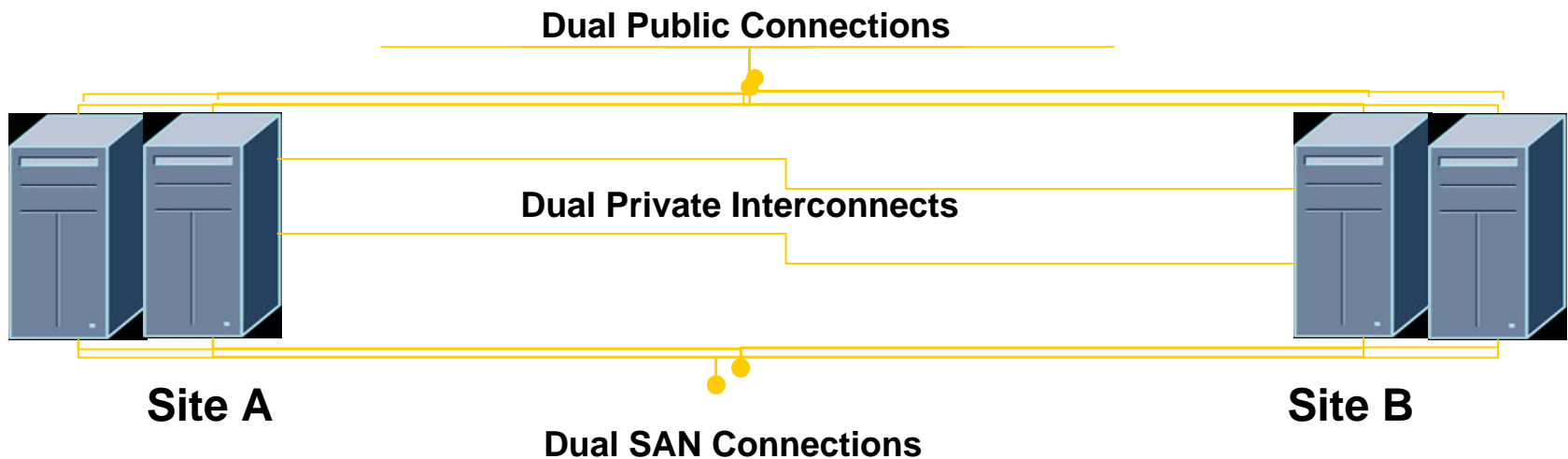
Benefits of an extended cluster

- Faster recovery from site failure than any other technology in the market



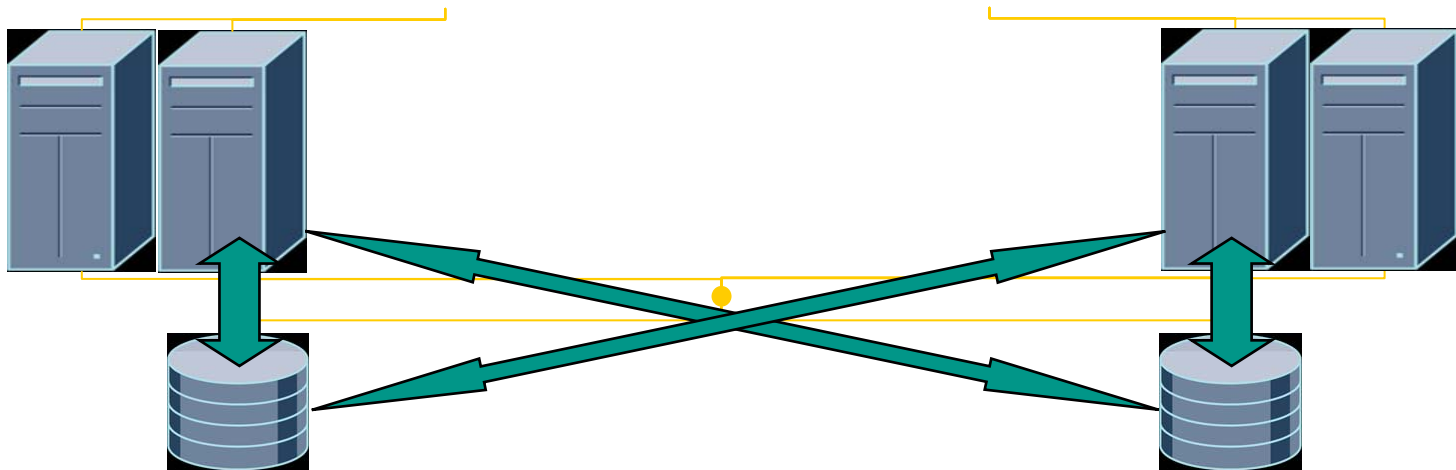
PoC: Design Considerations

- Connectivity: Redundant connections for public, interconnect and IO traffic

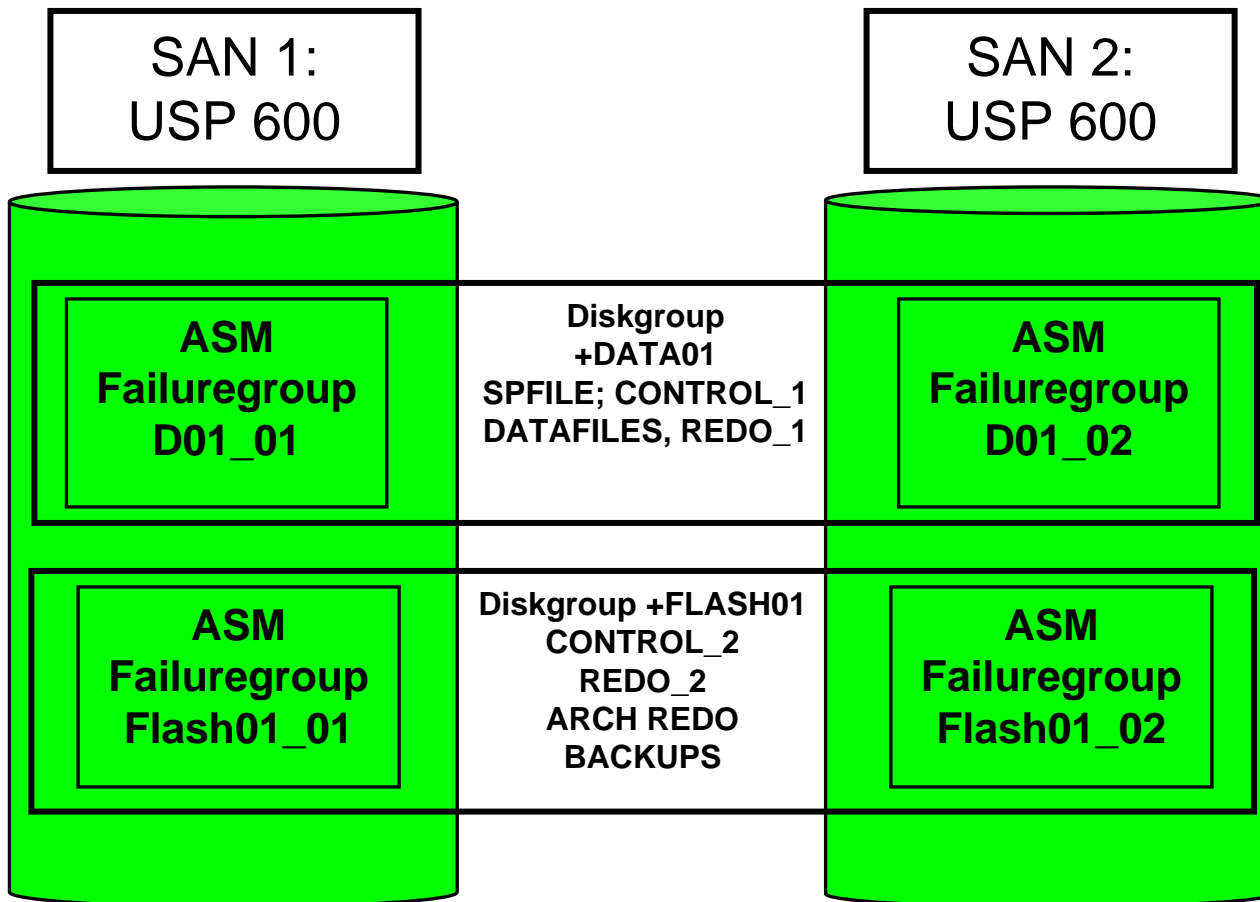


PoC: Design Considerations

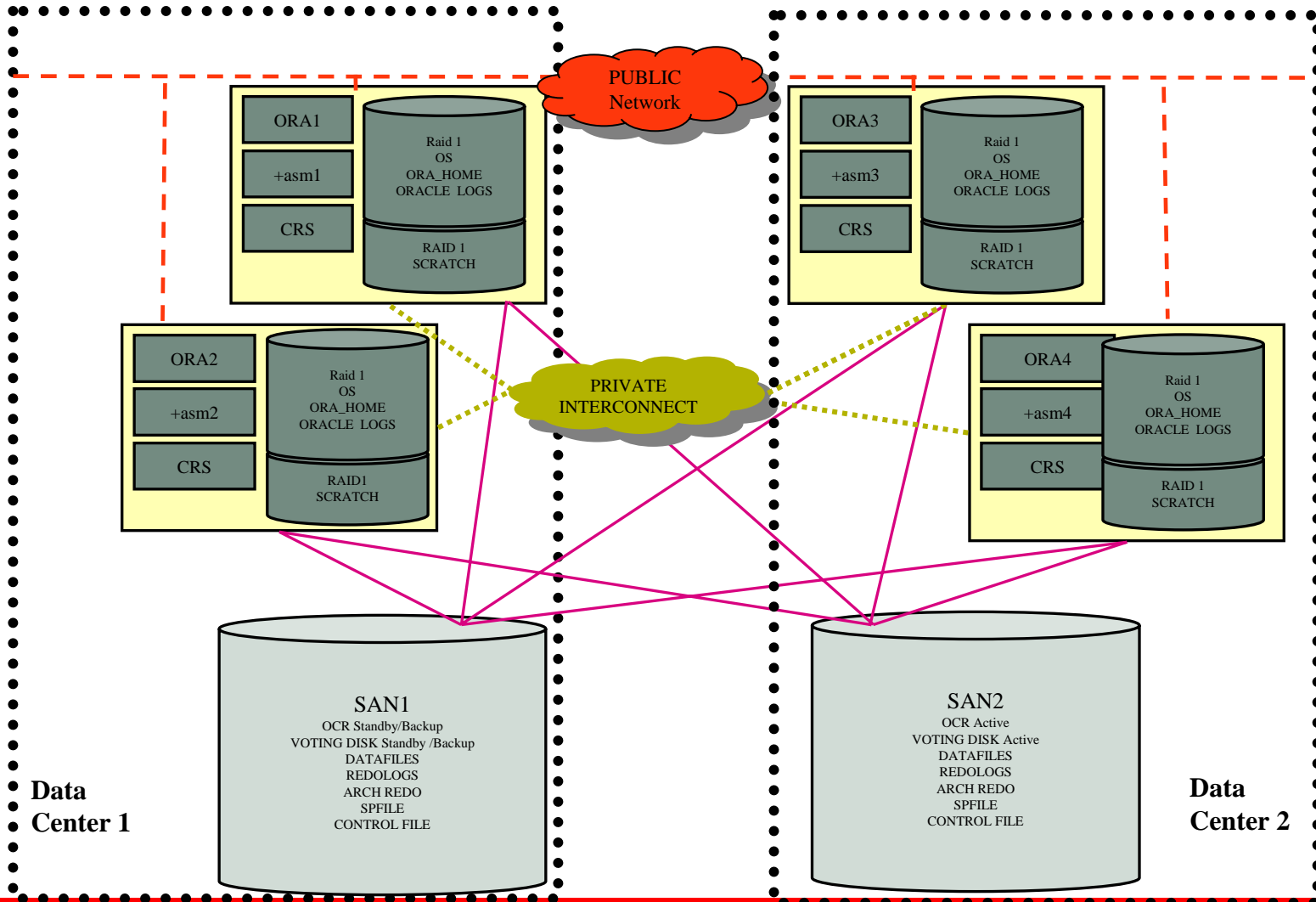
- Disk (or better SAN-) mirroring: Host based mirroring with ASM



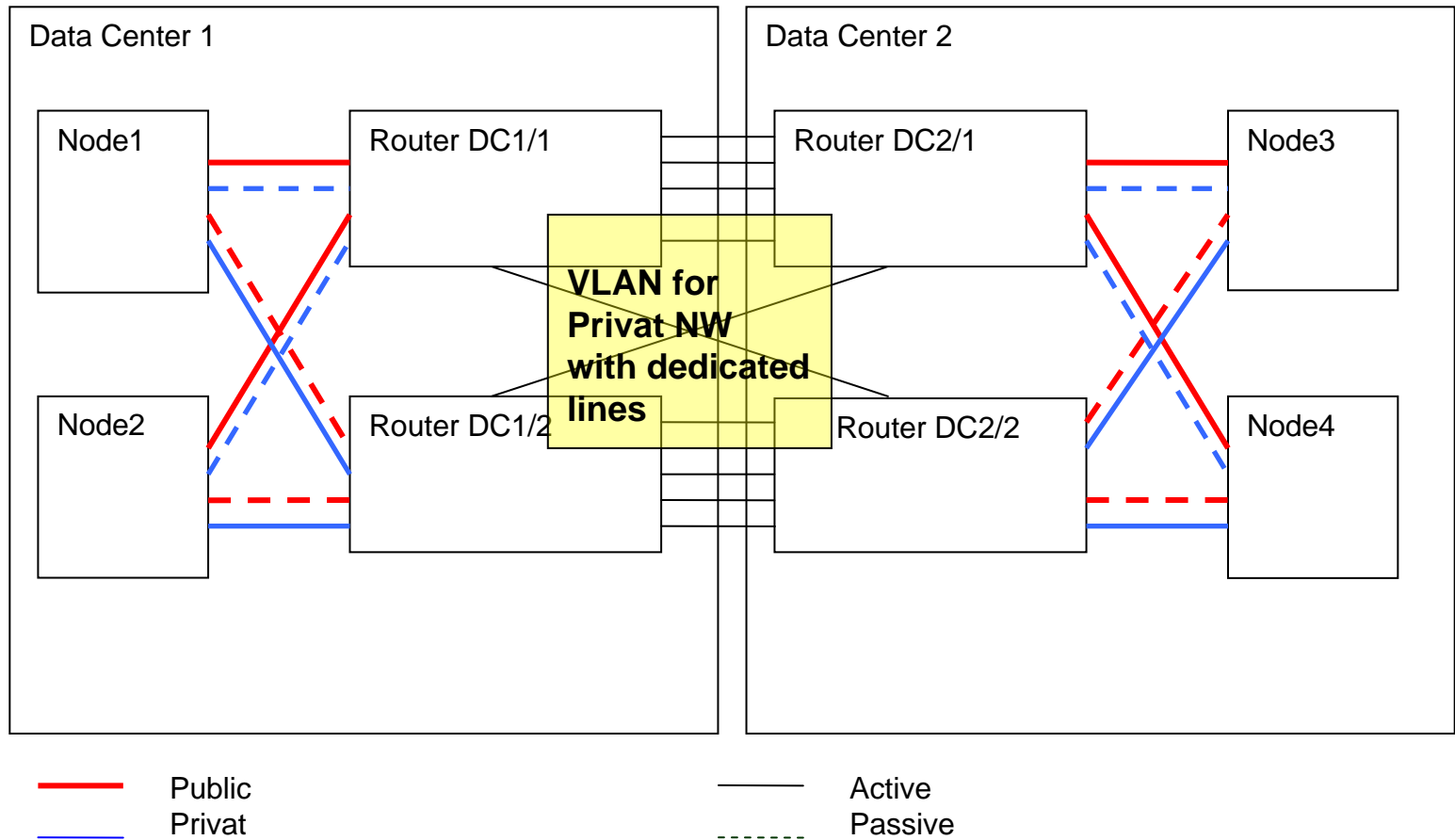
ASM setup



PoC: Architecture



PoC: Network Layout



Used Software

- Linux SLES-9 SP2
 - Native Multipathing: Device Mapper
 - 32Bit or 64Bit?
 - 32Bit: Availability of versions and patches
 - 64Bit: Addressable memory, no midterm-migration
- Oracle for x86-64:
 - 10.1.0.3
 - 10.1.0.4
 - 10.2.0.1

Used Hardware

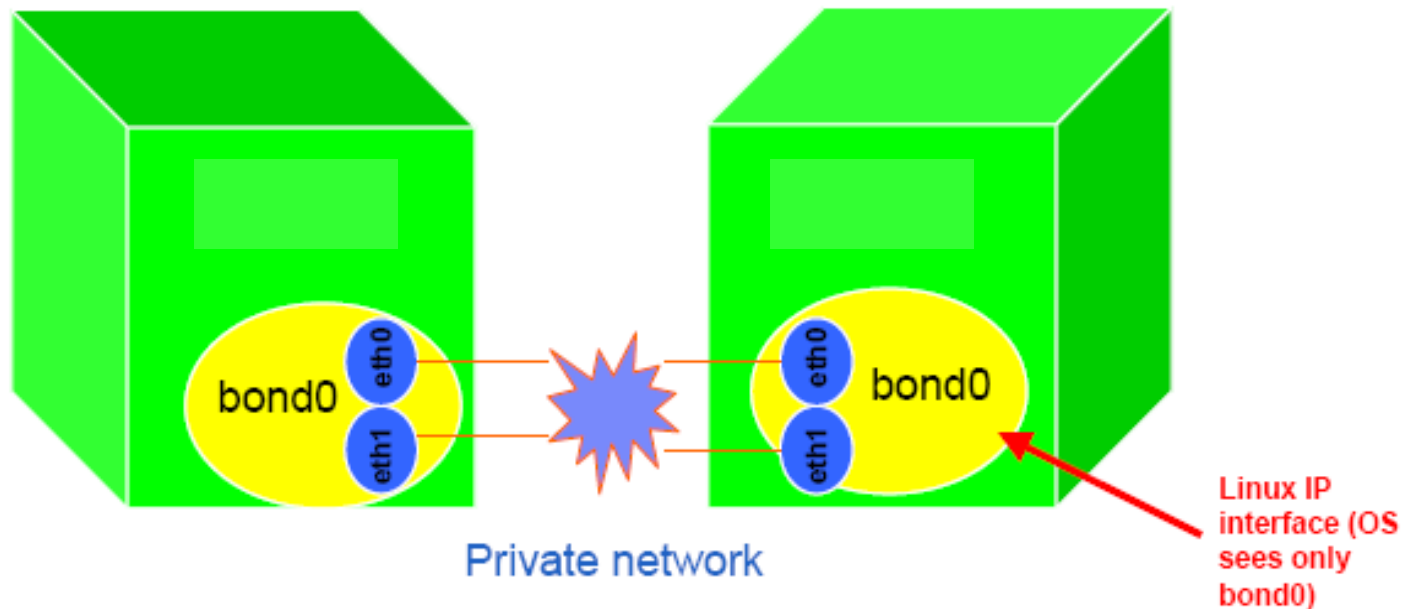
- HP Proliant DL585 with 4 AMD Opteron 800 series processor running at 2.2 GHz, 32GB RAM.
- 2 Fibre-channels per server
- Storage: 2 Hitachi USP 600

Tests

- Installation tests
- High Availability
- Performance
- Workload Management
- Backup/Recovery
- Maintenance
- Job Scheduling
- Monitoring

High Availability Tests

Linux bonding driver: Network redundancy



The Linux bonding driver provides the ability to detect link failure and reroute network traffic around a manner transparent to the application.

The linux bonding is used to protect the Oracle Interconnect network.

High Availability Tests

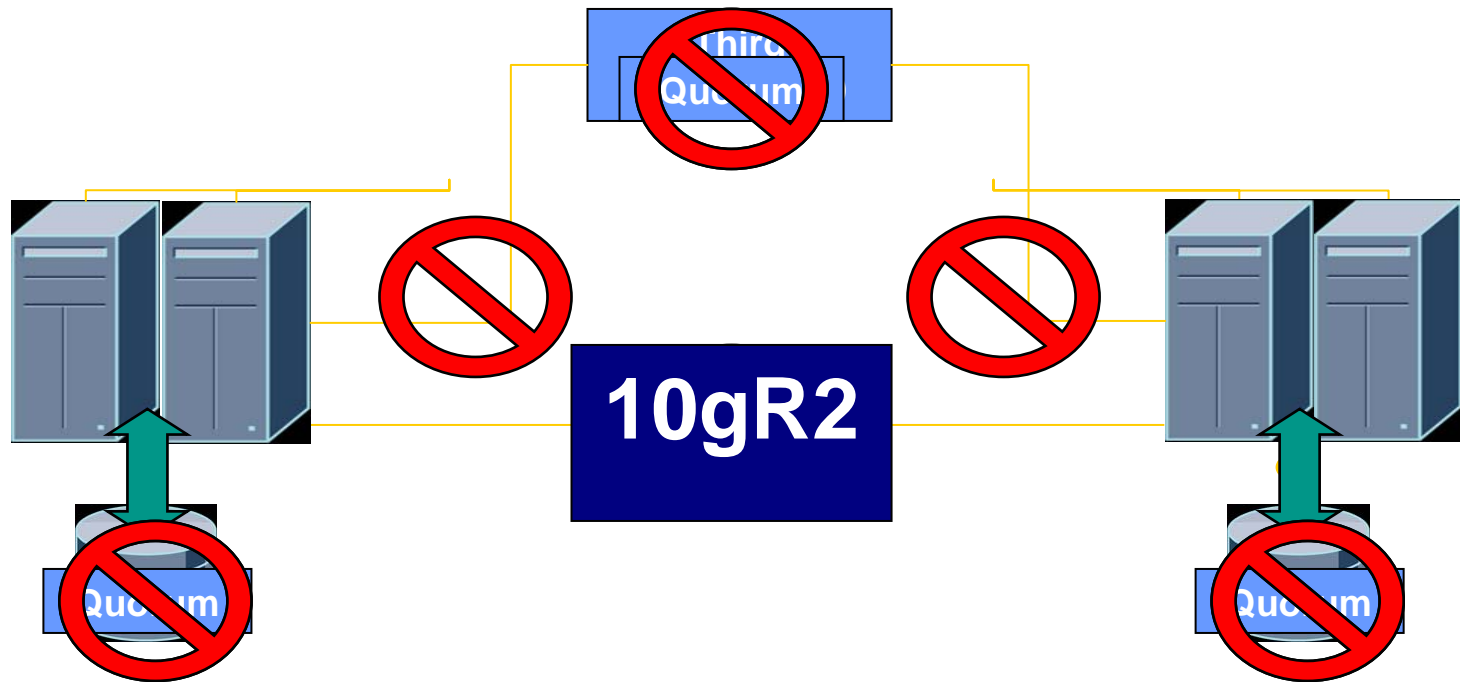
- Bonding
 - NIC failover tests successful.
- Multipathing
 - Device mapper available in Linux 2.6 (SLES9 SP2).
 - Multiple paths to the same drive are auto-detected, based on drive's WWID.
 - We successfully tested failover-policy (fault tolerance) and multibus-policy (fault tolerance and throughput).
 - Check with your storage vendor on which multipath software (in our case Device Mapper or HDLM) and policy to use.

High Availability Tests 10gR1

Crash Test	Impact Service	Main component	Behavior
Power off Server		TAF, CRS	second instance is up and running
Disk failure - LUN_data		RAID/ASM	instances up and running
Disk failure - LUN_voting		RAID	instances up and running
Disk failure - LUN_ocr		RAID	instances up and running
Unplug fiber attachment		Multipathing	instances up and running
Kill Oracle Instance		TAF, CRS	second instance is up and running
Kill ASM Instance		TAF, CRS	second instance is up and running
Halt -q on Linux		TAF, CRS	second instance is up and running
Unplug Interconnect cable		Bonding	instances up and running
Unplug all Interconnect cables		TAF,CRS/Voting	second instance is up and running
Power off - SAN (NoVoting)		ASM	instances up and running
Power off - SAN (voting)			cluster down

HA Design Considerations

- What about quorum?
 - What happens if the site with the quorum fails or all communications between sites are lost?



High Availability Tests 10gR2

Crash Test	Impact Service	Main component	Behavior
Power off Server		TAF, CRS	second instance is up and running
Disk failure - LUN_data		RAID/ASM	instances up and running
Disk failure - LUN_voting		RAID	instances up and running
Disk failure - LUN_ocr		RAID	instances up and running
Unplug fiber attachment		Multipathing	instances up and running
Kill Oracle Instance		TAF, CRS	second instance is up and running
Kill ASM Instance		TAF, CRS	second instance is up and running
Halt -q on Linux		TAF, CRS	second instance is up and running
Unplug Interconnect cable		Bonding	instances up and running
Unplug all Interconnect cables		TAF, CRS/Voting	second instance is up and running
Power off - SAN (voting)		ASM, 3rd voting	instances up and running
Power off - 3rd voting disk		2 remaining votes	instances up and running

High Availability Tests

- Switch off SAN (i.e. switch off one HDS USP 600)
 - Disk mount status changes from “CACHED” to “MISSING” in ASM (v\$asm_disk)
 - Processing continues without interruption
- Switch on SAN again
 - Disks are visible twice: Mount status MISSING and CLOSED.
 - Add disks again
 - alter diskgroup <dg> failgroup <fg> disk ‘<disk>’ name <new name> force rebalance power 10;
 - Disks with status “MISSING” disappear. Disks with previous status “CLOSED” have status “CACHED” now.
- Result: No downtime even with a SAN unavailability.

Performance considerations

- Need to minimize latency
 - Direct effect on ASM mirroring and cache fusion operations
 - It's better to use direct connections. Additional routers, hubs or extra switches add latency.

Extended Cluster versus MAA

	Extended RAC	RAC + DG
Minimum Nodes Needed	2	3
Active Nodes	All	One Side Only
Recovery from Site Failure	Seconds, No Intervention Required	Manual Restart Usually Required
Network Requirements	High cost direct dedicated network w/ lowest latency. Much greater network bandwidth	Shared commercially available network. Does not have low latency requirements
Effective Distance	Campus & Metro	Continent
Disaster Protection	Minor Site or Localized Failures User Errors	Database Corruptions Wider Disasters User Errors

Summary

- RAC on Extended Cluster with ASM
 - It works!
 - Good design is key!
 - Data Guard offers additional benefits

ORACLE®