

ORACLE DATABASE HIGH AVAILABILITY STRATEGY, ARCHITECTURE AND SOLUTIONS

DOAG Nuremberg - 17/09/2013
Kirill Loifman
Oracle Certified Professional DBA
email: loifmkir@gmail.com
www: dadbm.com
Twitter: [@loifmkir](https://twitter.com/loifmkir)

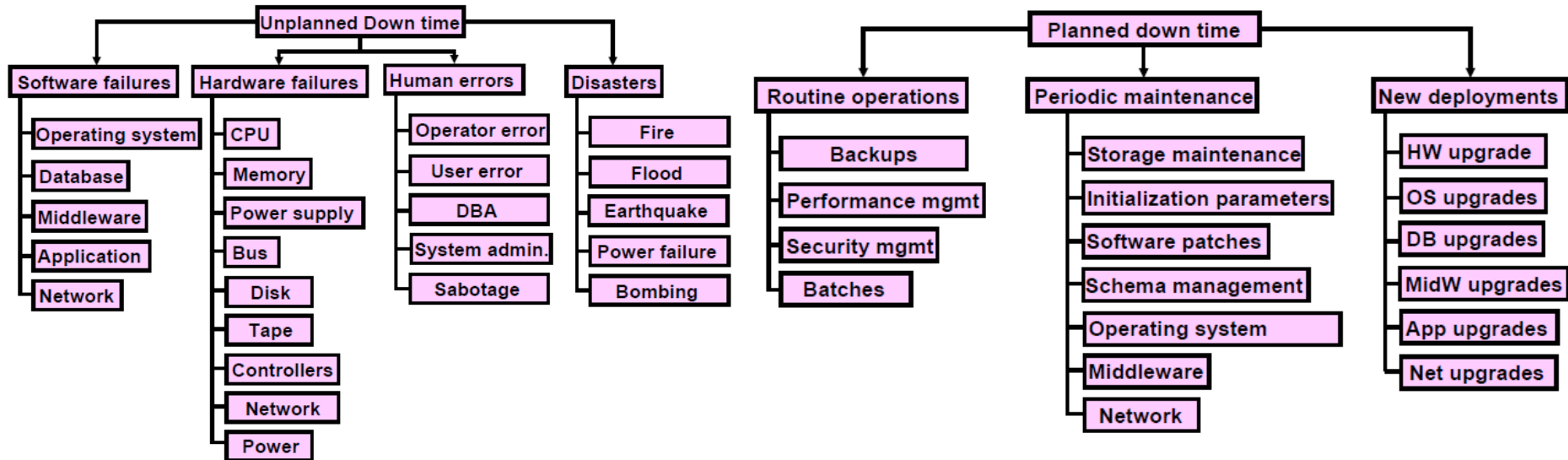
ELEMENTS OF HIGH AVAILABILITY

- **High availability**
 - Eliminating single points of failure through redundancy
 - Examples: redundant HW, SAN/ASM, RAC databases
- **Disaster recovery**
 - Recovering from failure
 - Examples: Extended clusters, standby databases
- **Oracle Maximum Availability Architecture (MAA)**
 - Combining HA and DR elements
- **Downtime**
 - How much time you can tolerate your system being unavailable
 - Includes unexpected as well as planned down time
 - 99.9% availability ~ 8 hours and 45 minutes of down time per year



CAUSES OF DOWNTIME

- 2 ways of dealing with the potential for disaster:
 - a) expect and plan for it*
 - b) ignore it, head to the pub and hope for the best*



UNDERSTANDING SERVICE LEVEL AGREEMENTS (SLA)

- 1. Understanding *Service Level Agreements* - customer's service requirements
- 2. Understanding elements of SLAs:
 - *System Availability* e.g. The database servers must be available 7 days a week, from 6am to midnight
 - *Acceptable Data Loss* e.g. No more than 15 minutes of data entry can be lost
 - *Recovery Time* e.g. In the event of a disaster, the systems should be back up and running within one hour.
 - *Performance* e.g. Transaction response time should not exceed 2 seconds.
- 3. Map your SLAs to appropriate High Availability Levels / Tiers
- 4. Design systems and processes to meet SLA expectations
- 5. Do not forget about cost!

AVAILABILITY TARGET VS COST

- **How many 9s do you need?**

Availability Target	Downtime Per Year (Approx.)
90 %	36 days
99 %	3.5 days
99.9 %	8 hours
99.99 %	52 minutes
99.999 %	5 minutes

- - Sharp decrease in downtime for each additional 9 in the availability target.

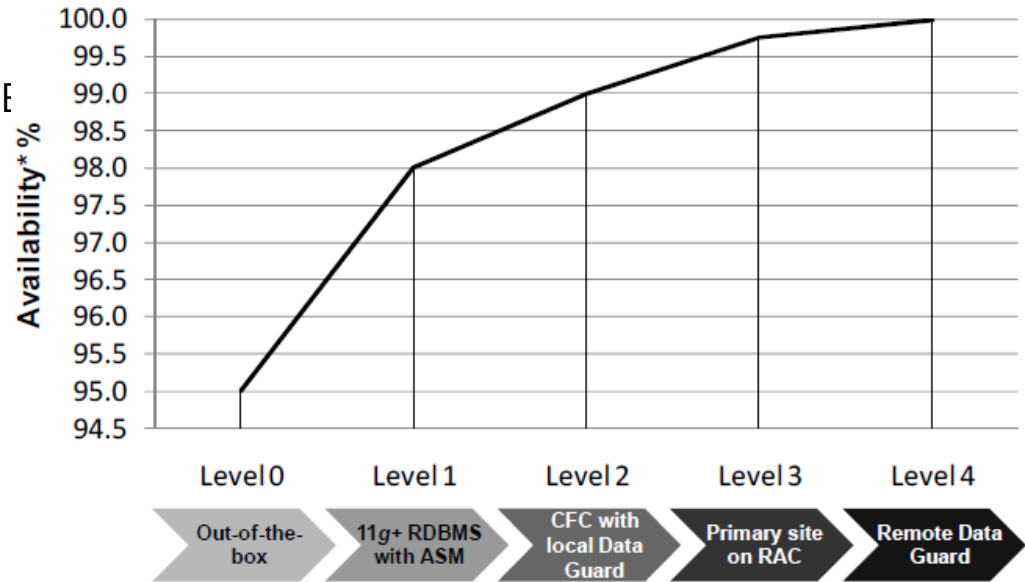
- **What will it cost?**

- - Cost comparison of options helps in achieving realistic expectations when developing SLAs

Item	SAN Snapshot backup	Native DB backup
Licensing	€28,000	
Training	€14,000	
Storage	€45,000	€10,000
Total Cost	€87,000	€10,000
Recovery Time	about 5 minutes	about 45 minutes
Benefits	Near-instant recovery	Reduced complexity

LEVELS OF HIGH AVAILABILITY - EXAMPLE

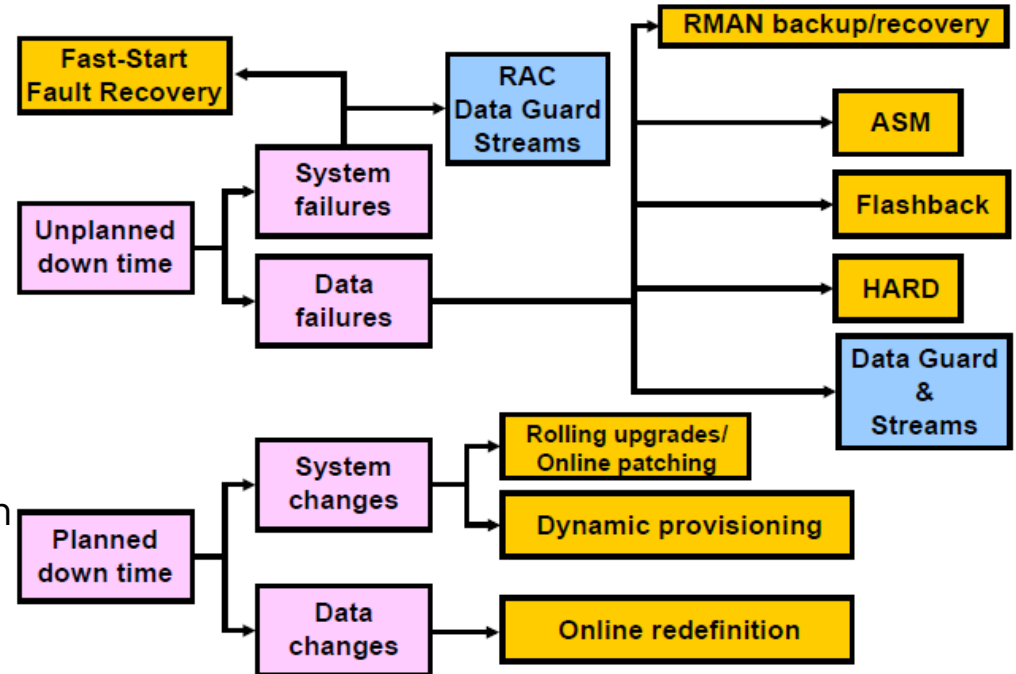
- **Level 0: Out-of-the-box** (no specific high-availability elements)
- **Level 1: Storage-level protection** (database instance with protected storage)
- **Level 2: CFC or/and Standby**
(Cold Failover Cluster or/and Standby Database at the same physical location)
- **Level 3: Recovery via redundant components**
(Hot Failover with multi-node RAC at the same physical location)
- **Level 4: Active and passive recovery**
(2 site concept, MAA, RAC, Standby Database, Storage mirroring)



Note: HA Level combinations possible; * percentages shown are for illustrative purposes only.

ORACLE'S SOLUTION TO DOWN TIME

- Oracle Cluster-ware & ASM
- RAC, RAC 1 Node
- Oracle Restart
- Flashback Technology
- Recovery Manager (RMAN)
- Data Guard
- Streams and GoldenGate
- Fast-Start Fault Recovery
- Data Recovery Advisor
- Online reorganization & redefinition
- Patching in rolling fashion
- Online patching
- Dynamic database reconfiguration
- Automatic Diagnostic Repository (ADR)
- HARD - Oracle Hardware Assisted Resilient Data
- ...

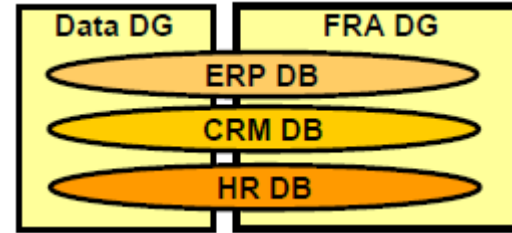


ORACLE DATABASE HIGH AVAILABILITY: BEST PRACTICES

Use SPFILE.	Create two or more control files.	Set CONTROL_FILE_RECORD_KEEP_TIME long enough.	Multiplex production and standby redo logs
Log checkpoints to the alert log.	Use auto-tune checkpointing.	Enable ARCHIVELOG mode and use a flash recovery area.	Enable Flashback Database.
Enable block checking.	Use Automatic Undo Management.	Use locally managed tablespaces.	Use Automatic Segment Space Management.
Use resumable space allocation.	Use Database Resource Manager.	Register all instances with remote listeners.	Use temporary tablespaces with tempfiles.

STORAGE-LEVEL PROTECTION CONSIDERATION

- **2 file systems (DATA & FRA) are enough** to share storage between multiple DBs
- **Which RAID Configuration for HA to use?**
 - based on business requirements, costs, etc.
 - a) ASM mirroring (best choice for low cost storage; enables extended clustering solutions)
 - b) Hardware RAID1 (mirroring; best performance; best choice for modern SAN)
 - c) RAID5 (parity protection, more economical solution, not for write intensive workloads)
 - **d) Both ASM mirroring and Hardware mirroring**



- **What Type of Striping Works Best?**
 - a) ASM only striping
 - many LUNs to manage at the storage level
 - LUNs should be equal and limited in size
 - b) RAID0 and ASM striping (stripe-on-stripe)

Disk I/O desing best practices

- Use external RAID protection when possible
- Use LUNs with the same performance characteristics
- Use LUNs with the same capacity
- Maximize the number of spindles in your ASM disk group - pain for SAN admin ☹

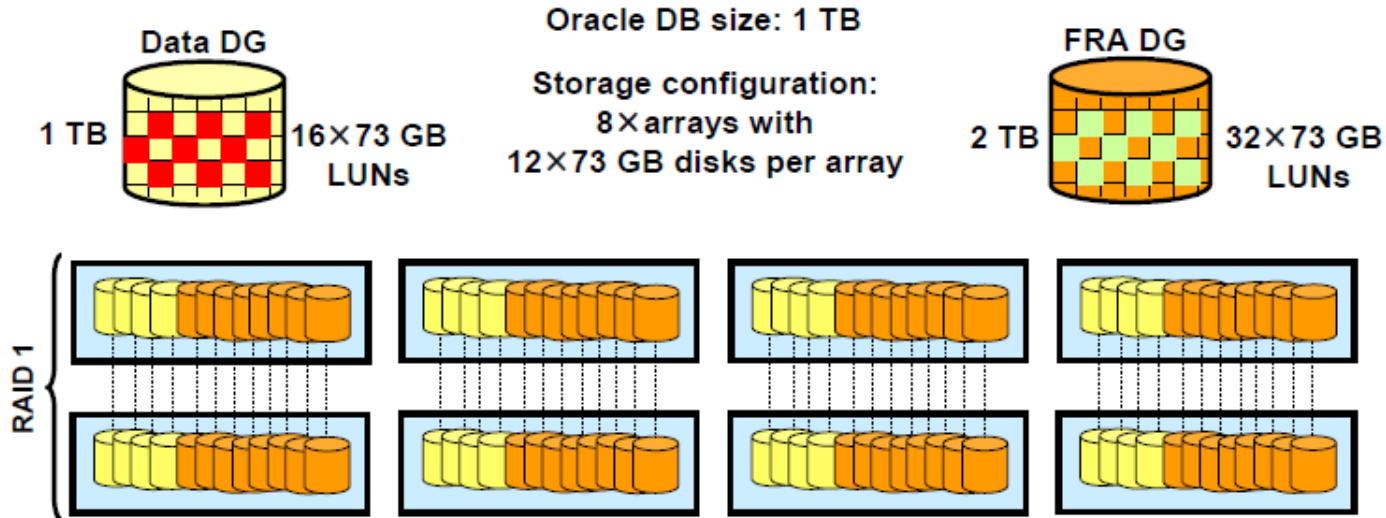
ASM STRIPING ONLY

Pros:

- Drives evenly distributed for Data & FRA
- Higher bandwidth
- Allows small incremental growth (73 GB)
- No drive contention

Cons:

- Not well balanced across ALL disks
- LUN size limited to disk size



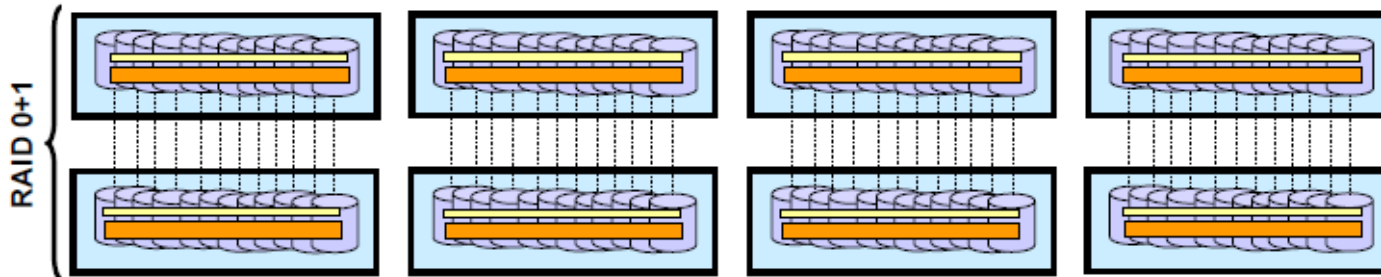
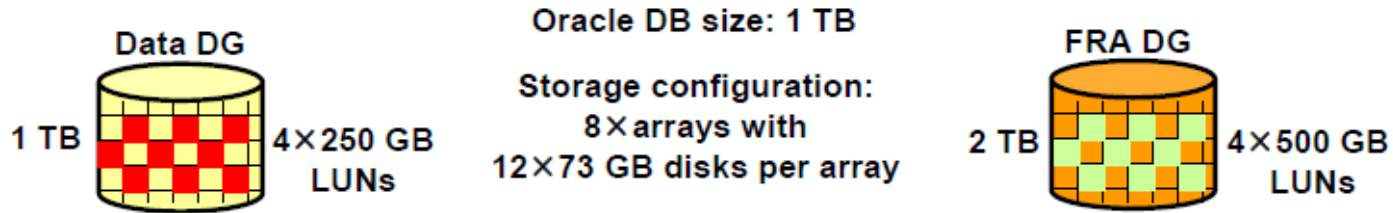
HARDWARE RAID-STRIPED LUNS

Pros:

- Fastest region for Data DG
- Balanced data distribution
- Fewer LUNs to manage while max spindles

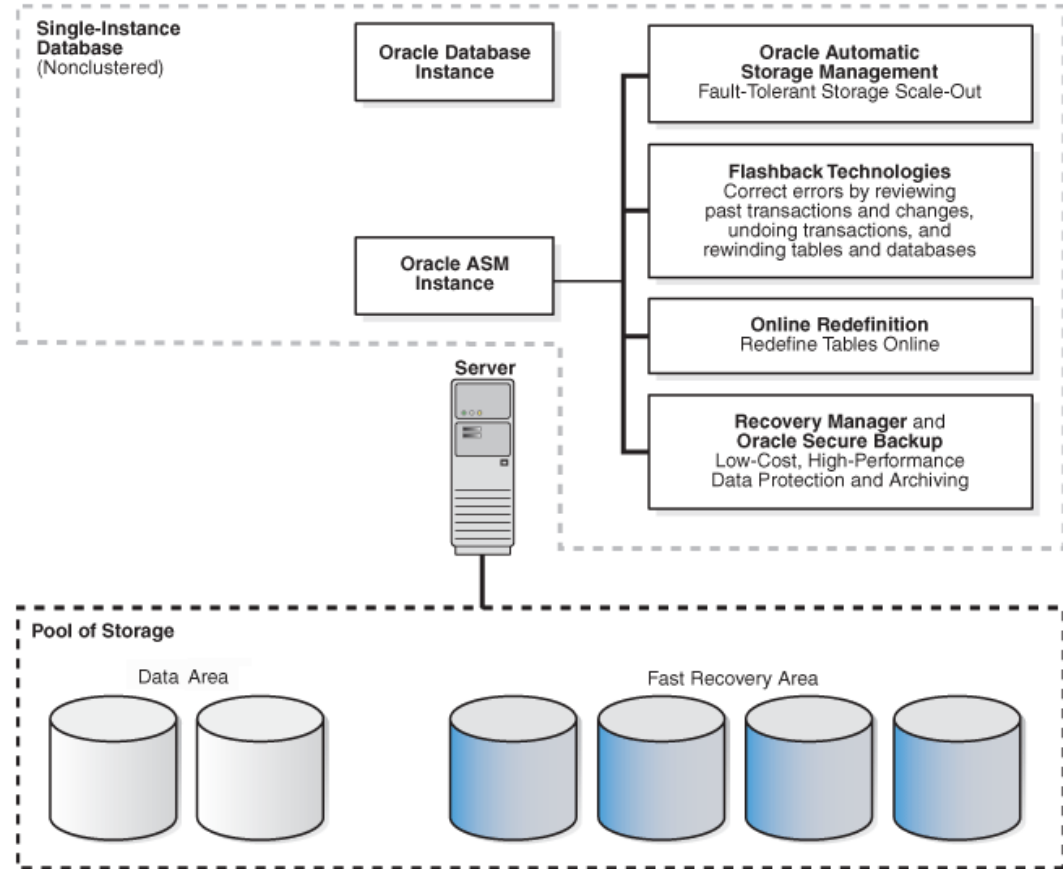
Cons:

- Large incremental growth
- Data & FRA "contention"



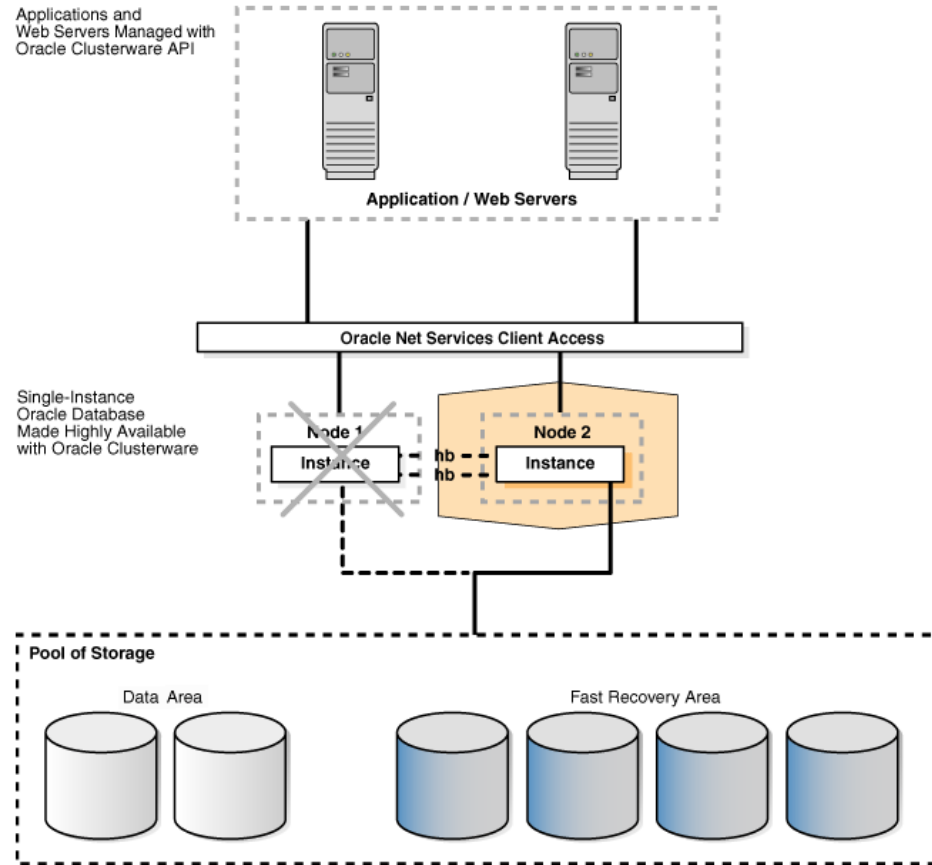
SINGLE-NODE, NON-CLUSTERED ORACLE DATABASE

- HA Level 1 ~ 95-98% Availability
- Single-instance, standalone (noncluster) database
- Protected storage via SAN / ASM
- Oracle Restart to enhance HA
- Other HA features like: Flashback Database, Online Redefinition, Recovery Manager, and Oracle Secure Backup, etc.
- 2 file systems / ASM data groups



ORACLE DATABASE WITH CLUSTERWARE (COLD FAILOVER)

- HA Level 2 ~ 98-99% Availability
- 2 or multiple-nodes cluster
- Single DB instance in cold failover mode
- Oracle GRID Infrastructure
- Oracle RAC One Node as Oracle recommended cold failover solution
- Shared storage on ASM, OCFS2, NFS, ...
- Oracle RAC one Node pros and cons:
 - + better DB availability than the others
 - + Online migration & patching
 - + Single vendor solution
 - + Ready to scale to full RAC
 - + Supports DataGuard
 - + Host-based mirroring with ASM
 - - Works only from 11gR2
 - - Licensed separately (+25% on CPU)

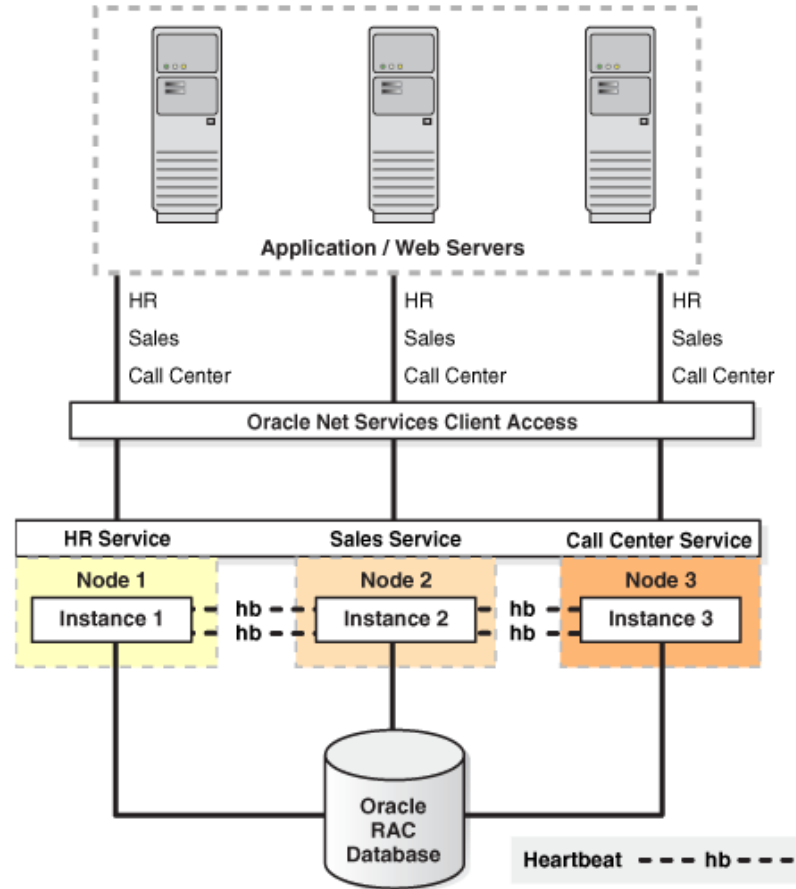


ORACLE DATABASE COLD FAILOVER - OTHER OPTIONS

- Oracle Grid Infrastructure with Oracle clusterware scripts
 - + Not licensed separately in case you cluster an Oracle licensed product (database)
 - + Works for Oracle Clusterware and Database before 11gR2
 - + Host-based mirroring with ASM possible
 - No online relocation possible as with RAC One Node
 - Oracle does not directly support action scripts as custom code fragments.
- RedHat Failover Cluster
 - + Inexpensive
 - + Make sense using Linux RedHat
 - + Simple to manage but a major upgrade can be a pain
 - Storage not shared but mounted to active node
 - Host-based mirroring in a stretched cluster requires different Linux LVM with more overhead
- HP Serviceguard
 - + Complex but stable
 - + Available on HP-UX and Linux
 - + Host-based mirroring with HP-UX LVM possible
- - Expansive

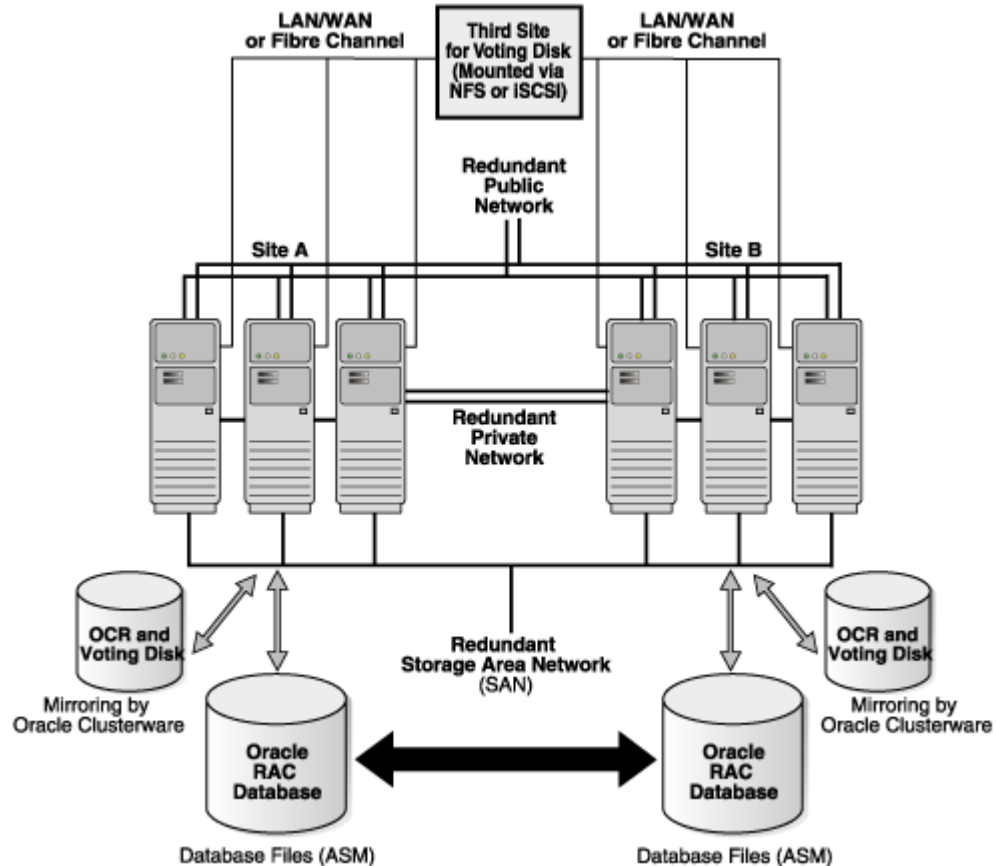
ORACLE RAC DATABASE

- HA Level 3 ~ 99,5% Availability
- 2 or multiple-nodes cluster
- Oracle GRID Infrastructure + Oracle RAC DB
- Oracle RAC pros and cons:
 - + Hot failover functionality
 - + Scalability across database instances
 - + Ability of increasing capacity without downtime
 - + Optimal computer resource usage
 - + Rolling upgrades and patching
 - + DB connection load-balancing
 - Might affect DML intensive performance
 - Licensed separately (~50% on top of CPU license)
 - Strict requirements for the cluster interconnect



ORACLE RAC DATABASE ON EXTENDED CLUSTERS

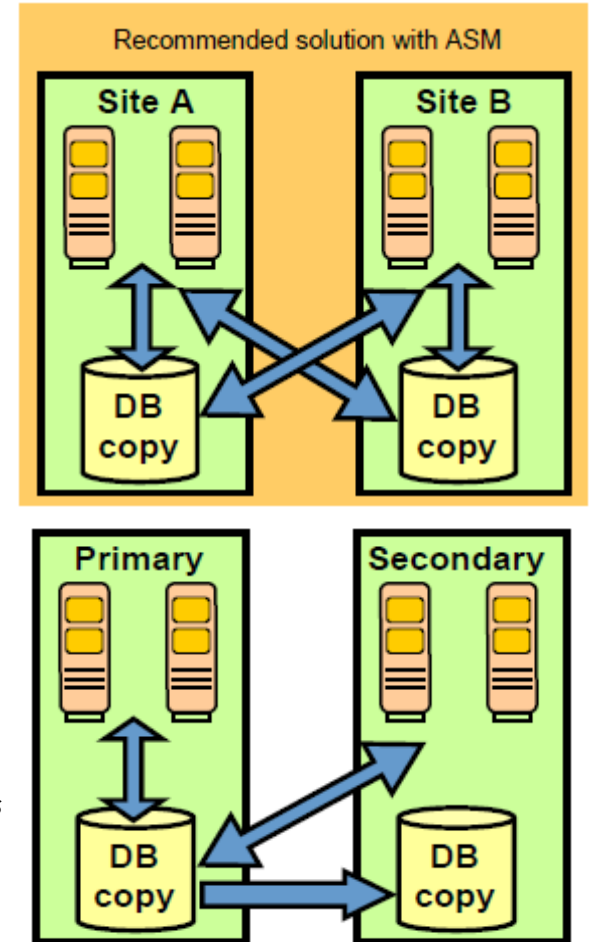
- HA Level 4 ~ 99,7% Availability
- Same configuration as local Oracle RAC
- 2 Data Center locations
- Distances over ten kilometers require dark fiber
- Redundant storage
- Third voting disk on an inexpensive NFS
- Pros and Cons:
 - + Greater availability than local Oracle RAC cluster
 - + Disaster recovery possibilities for some disasters
 - + Extremely rapid recovery if one site fails
 - Cannot protect against all data corruptions or comprehensive disasters



EXTENDED RAC DISK MIRRORING

- Need copy of data at each location
- Two options:
 - 1) Host-based mirroring
 - + Works with Oracle ASM, LRH HA-LVM, HP-UX LVM, ...
 - + No downtime in case of storage failure
 - Extra overhead on server resources
 - More complexity for cluster maintenance
 - 2) Remote array-based mirroring
 - + Works with HP CLX, EMC SRDF, HITACHI HAM, ...
 - + Done on hardware level
 - + No overhead on server
 - Outage in case of primary storage failure for some (HP CLX)
 - Storage HA software license required
- Disk mirroring is vendor specific

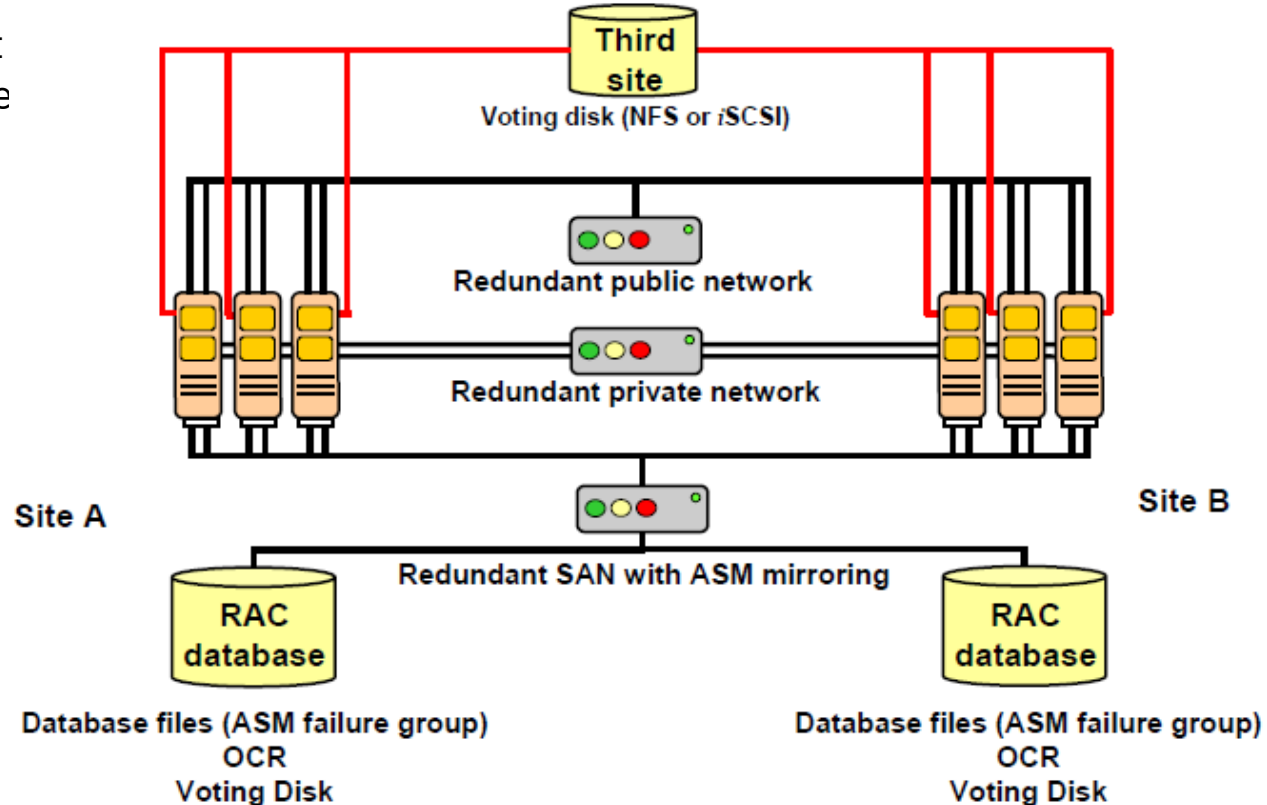
RedHat: When using a stretched cluster solution, not all storage configurations are supported by Red Hat, i.e. clvm and cmirror are not supported, but HA-LVM (without clvmd) may be used



ACHIEVING QUORUM WITH EXTENDED RAC

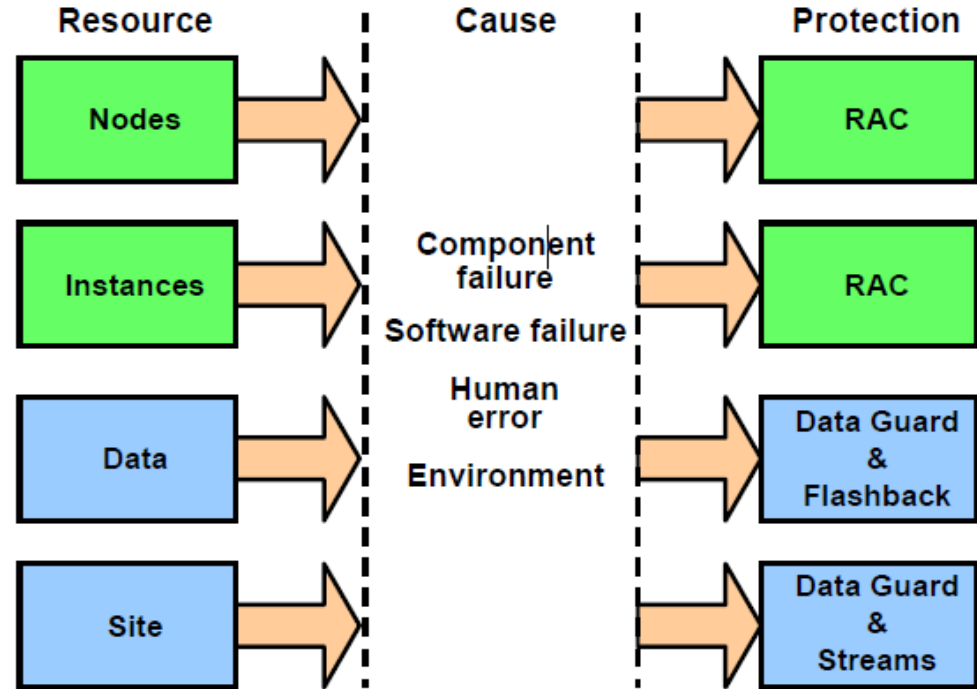
- Every Cluster node must access more than half of the voting disks at any time or that node will be evicted from the cluster

* The site that houses the majority of the voting disks is a potential single point of failure for the entire cluster.



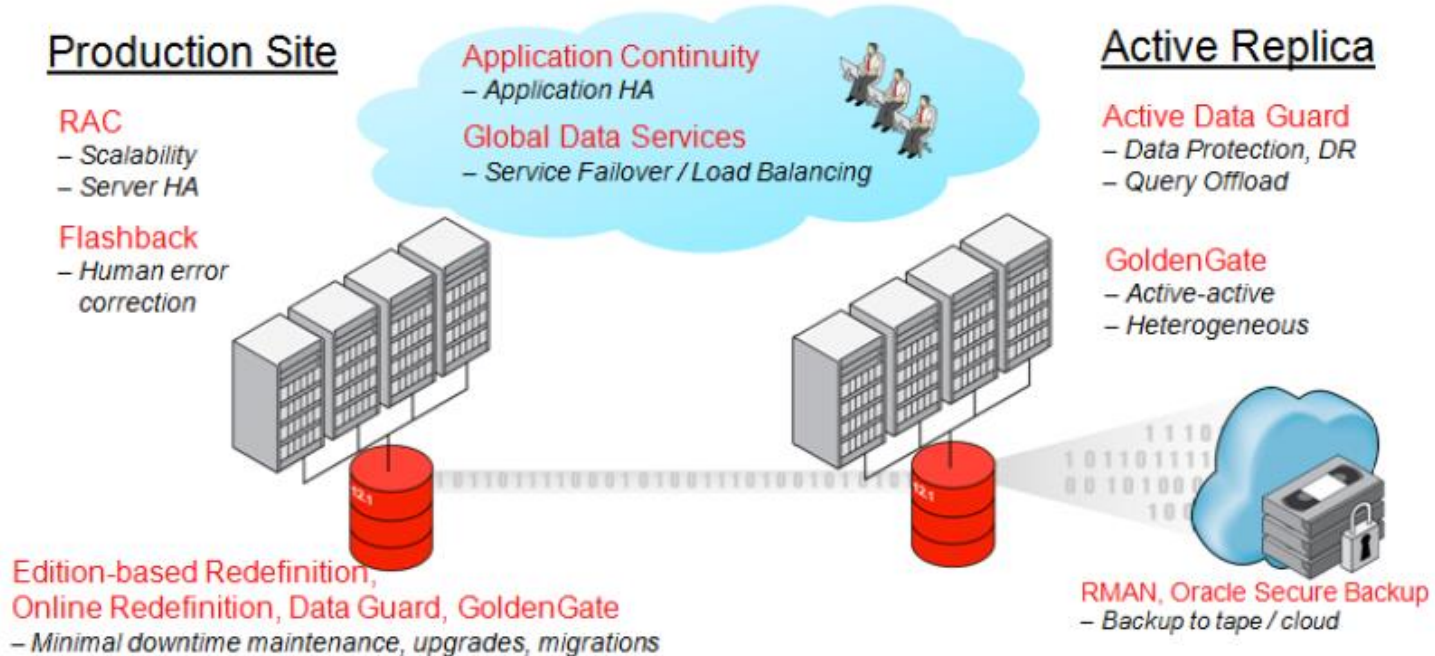
ORACLE RAC AND DATA GUARD COMPLEMENTARITY

- RAC addresses system failures by providing rapid and automatic recovery from failures, such as node failures and instance crashes.
- Data Guard addresses site failures and data protection through transactional consistent primary and standby databases that do not share disks, enabling recovery from site disasters and data corruption.



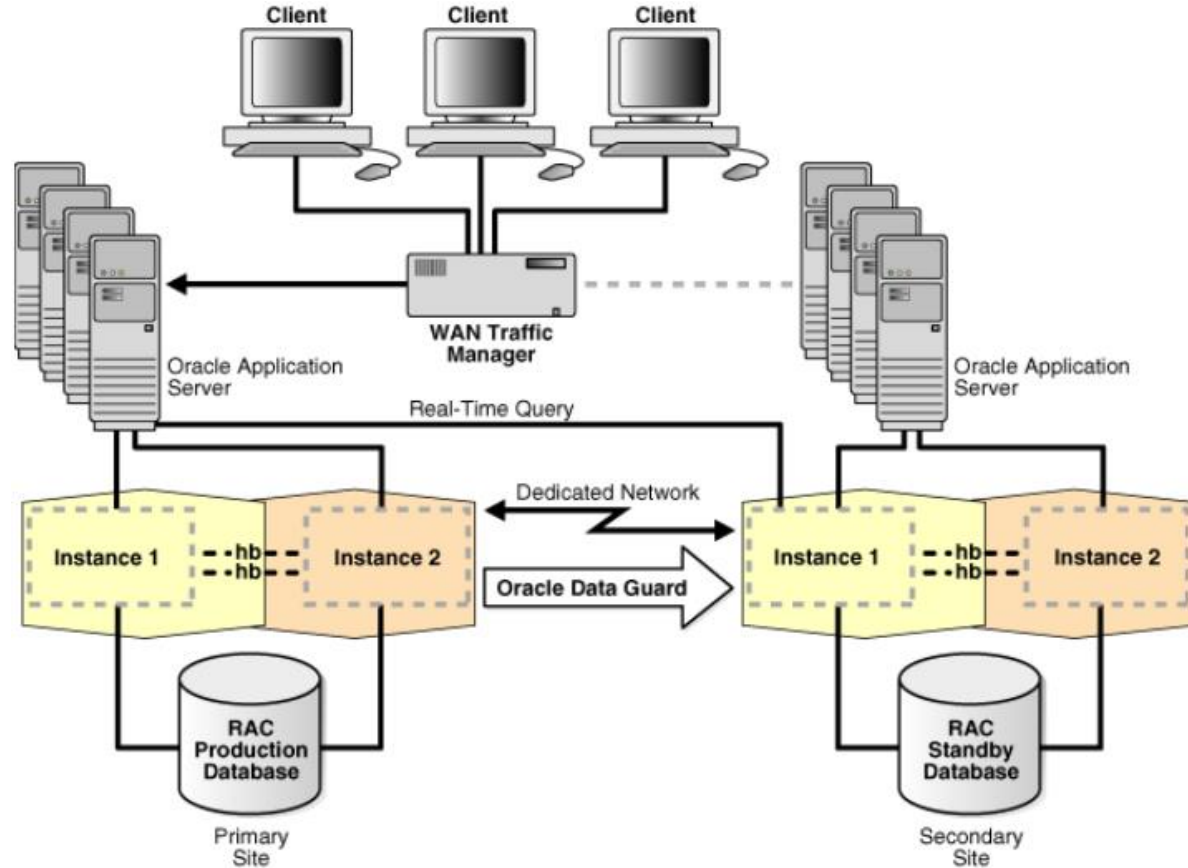
ORACLE MAXIMUM AVAILABILITY ARCHITECTURE (MAA)

- MAA is a set of best practice blueprints for the integrated use of Oracle HA technologies
- MAA provides guidance on optimal HA architectures designed to achieve the different service level objectives
- MAA resources: oracle.com/goto/maa

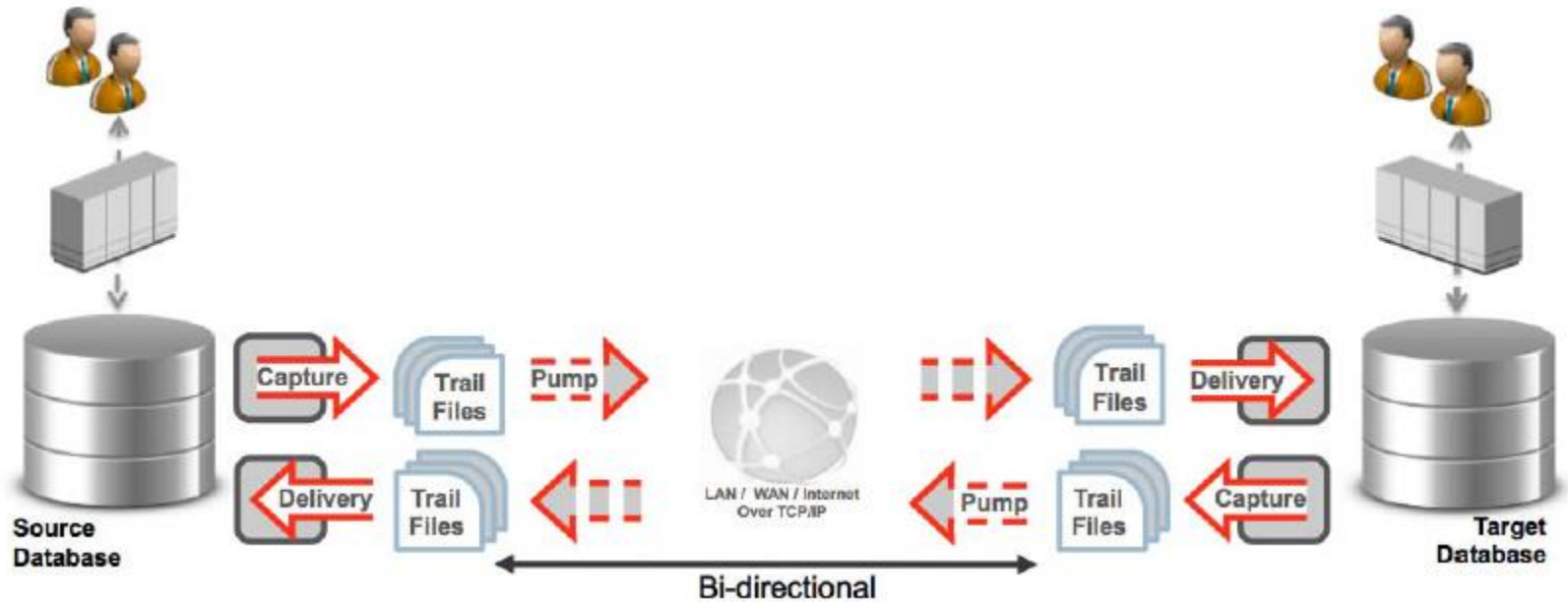


ORACLE MAA EXAMPLE – RAC & DATA GUARD

- HA Level 4 ~ 99.9% availability
- MAA provides the most comprehensive architecture for reducing down time for scheduled and unscheduled outages
- Two identical sites
- Primary site contains the RAC database
- Secondary site contains Physical Standby database or both Physical and Logical one on RAC
- Active DataGuard can be used for online reporting (extra licensed)
- Data Guard switchover and failover functions allow the roles to be traded between sites

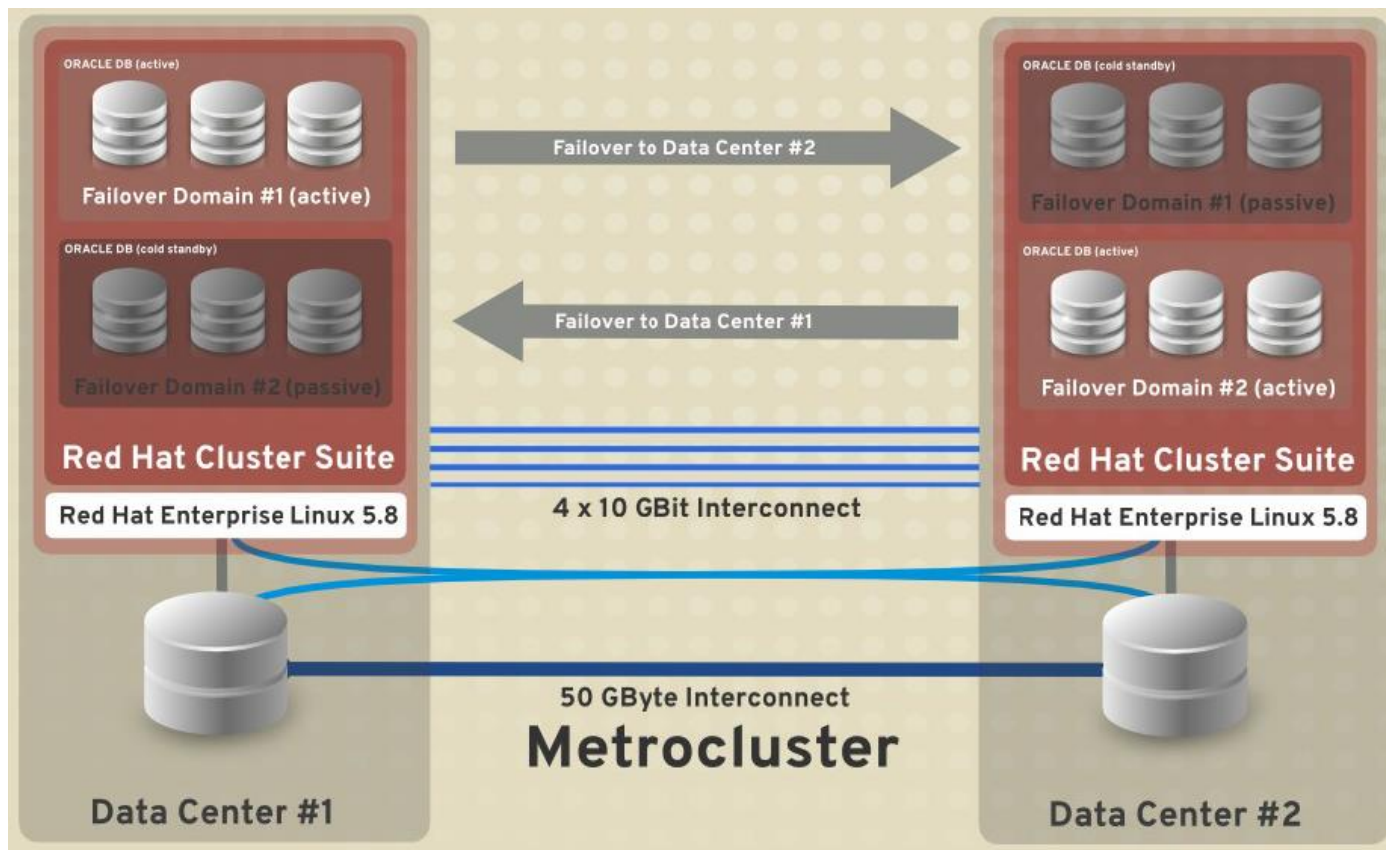


ACTIVE-ACTIVE HA AND ZERO-DOWNTIME UPGRADES: GOLDEN GATE



DATABASE HA SOLUTIONS – LIFE EXAMPLE

- Stretched RHCS with 2 Nodes
- Single DB instance(s) in cold failover mode
- 1 Oracle binary set per DB instance
- Storage mirroring with DRBD
- HA Level 3 - 99.7%



ORACLE EXADATA VERSION 3

Exadata X3 Database for High Availability

- Same architecture and price as X2
- 4x larger flash memory -> 22TB PCI flash memory per rack
- 20x more write performance -> Exadata Smart Flash Write Caching
- 33% more data throughput -> 100GB/sec running SQL
- 33% faster DB CPUs -> 8-Core Xeon SandyBridge E5-2690
- 75% more memory -> 2-4 TB DRAM
- Full 10Gb Ethernet to Data Centre
- 10-30% lower power -> Up to 3 Kilowatt Reduction per Rack
- New lower cost 1/8 rack solution



Exadata limitations

- Can not install other software
- Can not spread cluster nodes
- Can not mirror storage
- ...

New Oracle Platinum Services on X3

- Applied to Oracle Engineered Systems
- Delivered at no additional costs for Premier Support customers
- 24/7 oracle monitoring and remote diagnostics
- 5 Minute Fault Notification
- 15 Minute Restoration or Escalation to Development
- 30 Minute Joint Debugging with Development
- Quarterly system updates

ORACLE 12C HIGH AVAILABILITY FEATURES

- Oracle Multitenant architecture
- Application Continuity - improve application recovery (former TAF)
- Global Data Services - support global database services
- Active Data Guard Far Sync - extend zero-data-loss protection
- Flex ASM - enable inter-node storage failover and reduce ASM-related resource consumption
- SYSDG, SYSBACKUP – new privileges to avoid granting SYSDBA
- Active Duplicate Cloning – rman feature
- More online operations: datafile move, partition move
- Database Rolling Upgrades using Active Data Guard
- OEM12c Cloud Control:
 - Oracle Site Guard – extends automation of disaster recovery to the rest of the Oracle stack
 - High Availability Console - integrates monitoring of various HA areas
 - MAA Configuration Advisor
 - Automation of migration a database to ASM and RAC

....

???

THANK YOU!

Kirill Loifman

www.dadbm.com