# Competitive Snapshot

## Choices in Storage Architecture for Oracle Environments
### EMC vs. Network Appliance

By Clay Ryder

The Sageza Group, Inc.
July 25, 2007

**The Sageza Group, Inc.**
32108 Alvarado Blvd #354
Union City, CA 94587
510·675·0700 fax 650·649·2302

sageza.com
info@sageza.com

# Choices in Storage Architecture for Oracle Environments

## ABSTRACT

*Database systems have always been at the core of the IT landscape. Companies typically put their most valuable information in databases and use it to drive key business processes. Not only is storage an increasingly large cost component of database investments, but storage architecture can significantly and directly impact performance, availability, and recoverability of any database installation.*

*Making matters more difficult, today's data centers have multiple, distinct application workloads, each with different I/O profiles and distinct levels of value to the business. Because of this, IT organizations can expect that requirements for performance, high availability, and capacity planning will vary between production, development/test, and data warehouse environments. Over time, as IT organizations grow and database storage infrastructure scales the complexity of managing storage requirements across the environment will increase. This leads to a preference for a holistic view of storage architecture requirements for databases, and begs the question: how can IT architects to design a storage environment that is cost-effective yet meets changing needs for performance, availability, and recoverability at different points in a database's lifecycle? Additionally, and perhaps even more importantly, enterprises may demand that IT architects exploit synergy between the database software and the storage architecture.*

*Oracle RAC 10*g *is a particularly interesting case study in this regard. Whereas Oracle's "grid" architecture offers customers many useful features and the ability to use server resources more effectively, it is incumbent on storage vendors and IT architects to extend these concepts to the storage architecture in a useful way.*

*In this paper, we explore the interaction between Oracle databases and EMC and Network Appliance storage architectures and how this impacts storage architecture requirements, given the specific needs of Oracle databases; how EMC's and NetApp's approaches differ from one another with respect to performance, scalability, and high availability; and which storage platforms Oracle itself has chosen to run its Global IT consolidated applications.*

# Choices in Storage Architecture for Oracle Environments

## TABLE OF CONTENTS

# Understanding Common Oracle Database Use Cases

Creating a framework for understanding storage requirements around individual Oracle databases is at once both easy and difficult. If one looks at one user group, with a specific usage profile, at a given point at time, the answers may at first appear relatively straightforward: this much capacity, that much performance, this sort of availability and recoverability requirements, and so on. If it were only this simple. In the real world, several factors conspire to make such a "bottom up" approach extremely difficult. Even in a simplistic model with one database supporting one user group, requirements tend to evolve over time. Usage levels might increase, or users might find themselves using the database in a more sophisticated way.

The classic example has been data warehousing. Despite IT's best efforts, as users have become more proficient in using query tools, they have exercised the underlying database in new and more demanding ways. More subtly, databases that are deemed less critical at the outset have a way of becoming more critical over time as users (and the business) come to depend on their output.

Databases tend to find new audiences within an organization, with different needs and expectations than those envisioned by the original use case. A customer database might find audiences in marketing, sales, or customer services, and so forth, or a data warehouse might end up being the place where everyone goes for information once it is understood what it can do. Even modest databases can create substantial ecosystems around them for development, test, QA, and the like, especially as they become more popular (and important) to the business. It always seems there is a new version, a new application, or a new feature. More resources become invested in the supporting function, and disruptions in production become less tolerated by the business.

An external change in the business environment can create entirely new architectural concerns. Recent examples include retaining compliance records, or securing information within databases. If one views the evolution of business processes supported within IT, they tend to start with internal users and, over time, end up exposed to customers and partners outside the organization, creating entirely new concerns around performance, availability, and security.

Finally, IT has a vested interest in using the smallest number of technology providers in an effort to tame complexity. Over-optimizing "best of breed" approaches for each and every IT requirement can lead to a complex landscape that is difficult to manage, and ultimately ineffective.

Given these factors, IT professionals increasingly are leaning towards a holistic view of designing storage architectures for database environments. Simply put, this is starting with the known requirements at hand, but allowing for a rapid change in visible parameters (e.g., capacity, performance, usage models, availability, recoverability, development, and test support, security, and so forth) without forcing a rip-and-replace re-architecting of the environment.

We identify three distinct components of the database landscape:

- **Production:** the classic database use case, characterized by a preponderance of transactional updates and significant fact-based reporting or queries.
- **Decision support:** also known as data warehousing or business intelligence. This is characterized by sophisticated queries, analysis, and reporting against historical data, but also with the need to process significant amounts of updates, often in a batch fashion.
- **Development**: including testing, quality assurance, integration, and related concepts. This is characterized by the need to incorporate new functionality while ensuring that business operations are not impacted in the process.

Depending on the size of the environment, these components might translate into different numbers of database instances. A small requirement might use a single database for all three purposes, while larger environments might have multiple database instances for each role. Regardless of the number of database instances, the same requirements exist.

# Choices in Storage Architecture: EMC vs. Network Appliance

Oracle databases run on servers that communicate with storage. The choice of how the server connects to the storage (e.g., the storage network) has significant implications in cost, performance, availability, and manageability. As there is no true consensus on the best way to do this, each organization's IT architects must decide for themselves which approach works best for their installation.

## EMC Networked Storage Systems

EMC storage systems are designed for flexibility in terms of storage architecture, supporting Fibre Channel SAN, IP SAN (iSCSI), and NAS.

EMC platforms consolidate disk drives into RAID groups. Multiple types of RAID are supported including RAID 0, 1, 1/0, 3, or 5. For parity RAID types, the data-to-parity ratio depends on the number of drives in the group. Within these RAID groups, users create logical units of storage called LUNs. Multiple LUNs can be striped or concatenated together and presented to a host as a single large LUN, called a metaLUN. This allows a single LUN to be striped across more than one RAID group, increasing the number of disk spindles and providing a wider range of performance configuration options.

EMC platforms are designed with an in-place write approach to storage layout, which aligns strongly with the "locality of reference" model championed by Oracle databases.

## Network Appliance FAS Series

Network Appliance (NetApp) introduced its file server appliance in the early 1990s. These storage computers were specifically designed to access file shares over IP networks. Several years ago NetApp modified and repositioned its filers to serve both block and file storage protocols. The current NetApp family of products is the Fabric Attached Storage (FAS) series. The base offering is a 2U single-node model that can be upgraded to a dual node system. Redundant fans and power supplies are standard in the FAS series.

The FAS system organizes disks in groups called aggregates. Unused drives not in an aggregate are marked as hot spares. An aggregate may consist of one or more RAID groups. Two types of RAID are supported: Raid 4 and RAID-DP (double parity). RAID 4 uses one disk for parity and the rest for data. RAID-DP uses two disks for parity. An aggregate can contain one or more disk volumes called FlexVols. All offerings in the FAS series leverage a built-in file system known as the Write Anywhere File Layout (WAFL). WAFL was designed for file server applications and is optimized for sequential writes and reads in such environments. WAFL was not designed with an in-place write approach to storage layout. When NetApp added Fibre Channel and iSCSI support to the FAS series, it was required to force-fit a file server into the role of a block storage device.

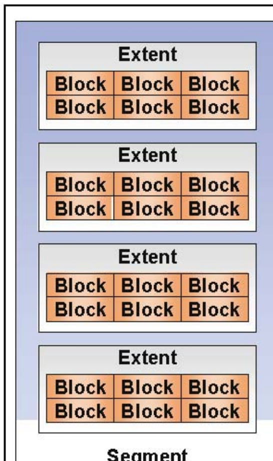


Figure 1: Oracle's Database Layout Approach

## Understanding Oracle Database Architecture Performance

Oracle 10*g* has specific strategies for optimizing write performance. These strategies are predicated around Oracle's logical model of physical storage: understanding which blocks are adjacent, what portions of the disk are easier to get to, and so forth.

Storage vendors can either support or hinder Oracle's write optimization strategies, depending on their approach. Use of non-volatile storage cache can aid performance, as can certain striping algorithms. However, simply abstracting a virtual presentation of physical devices can interfere with Oracle's optimizations, unless care is taken to preserve Oracle's logical model of physical storage.

Oracle's Database layout approach benefits from locality of reference. In this, the DBMS makes the assumption that all blocks will be contiguous to each other within the storage subsystem, i.e., there is no fragmentation. This is illustrated in Figure 1.

Some vendors offer storage products that do not align well with Oracle's database model. This can result in a performance penalty that can become more severe as the database environment scales. To illustrate this we now review the approaches taken by EMC and NetApp.

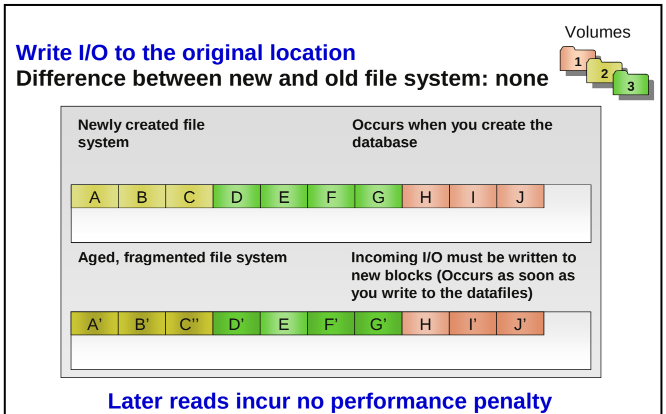


Figure 2: EMC's In-Place Write Architecture

## EMC's Write-In-Place Architecture

All EMC platforms were designed with in-place writes as a guiding principle. (See figure 2.) Data blocks are written to disk in a contiguous fashion. As Oracle expects to find real disks (not emulated ones), EMC's in-place write approach supports Oracle's performance optimization for the read-intensive environments that are typical of Oracle databases.

## Network Appliance's WAFL and Fragmented Writes

Architecturally, the FAS operating system (ONTAP) creates a file system (WAFL) that can then be presented to Oracle either as a file system (NAS), or—optionally—through block emulation (either FC or iSCSI). FAS block data is stored in WAFL files. WAFL spreads data and metadata everywhere on the LUN (hence "Write Anywhere").

When Oracle decides to write a specific block to a specific location, ONTAP intercepts the write, and puts it in a location of its choosing (illustrated in Figure 3). Additionally, in the case of a NetApp NAS implementation, when taking snapshots ONTAP and WAFL do not overwrite existing data blocks, but write new blocks in unused space. Older blocks are retained to provide point-in-time snapshots, or "snaps."

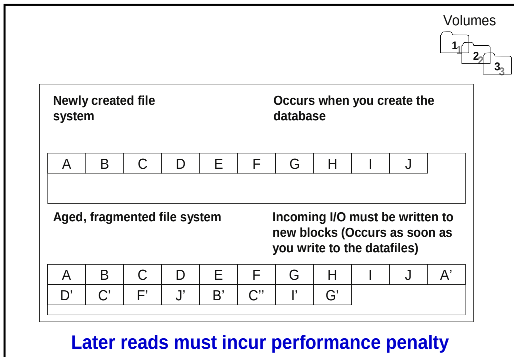


Figure 3: NetApp's WAFL and Fragmented Writes

As disk space is consumed, these older blocks (and their corresponding snaps) are deleted. The result is disk fragmentation. Oracle thinks it is writing data in sequential locations but ONTAP interferes with Oracle's intent, and simply places updated disk blocks in unused locations. Although initial performance on empty disks may appear reasonable, the results degrade over time. As databases change information with writes and the database size grows, fragmentation increases. Read performance may suffer, as Oracle is expecting sequentially organized data on disk. As database size increases to allocated capacity, write performance suffers, and WAFL must work harder to find new locations (and free old locations) with every new write.

Initially this performance impact may be negligible, but over time as more data is added and databases become more read-intensive vs. write-intensive, applications will experience longer seek times and performance will suffer. In very high read-intensive environments, such as decision support/data warehouse applications, this performance impact will be more significant.

This is in contrast with EMC's in-place write approach, where data that Oracle intends to be stored sequentially is actually stored sequentially, thus improving read performance over time, and disk writes have predictable performance characteristics at any age in the database's lifespan. This can be of particular value in Oracle decision support or data warehouse environments where large sequential I/O requests are common and where performance is critical.[1]

### Oracle Performance Implications of EMC vs. NetApp Architectures

Performance is very dependent on the exact behavior of a customer's environment. While precise numbers are difficult to define without detailed measurements and requirements, there are specific design and architectural factors that significantly impact performance.

The basic design of the WAFL file system leads to performance degradation over time due to WAFL fragmentation. It is for this reason that upon examining many NetApp public benchmark tests[2] we see the tested configurations typically leverage only a small fraction (~10%) of the configured storage capacity. EMC storage systems are designed to allow for predictable and scalable performance over time, with a high level of utilization, while providing tightly aligned locality of reference for Oracle databases.

Beyond baseline performance, users may opt to employ performance management solutions to monitor and fine-tune their storage platforms. While NetApp FAS series provides performance monitoring software, there are few functions available to tune the FAS array.

As an example, for EMC CLARiiON systems, EMC introduced the Navisphere Quality of Service Manager,[3] providing performance monitoring capabilities by extending users the ability to also proactively set policies that dynamically allocate system resources to meet service levels. In an Oracle database environment serving multiple applications with varying workloads, this functionality enables a storage administrator to set performance targets for high-priority applications or performance limits for low-priority applications. These performance targets and limits can be set for one of three key performance characteristics: response time, bandwidth, and throughput. Users can combine multiple workloads on a single array, yet still deliver required performance levels, avoiding contention.

### Oracle Automated Storage Management Implications

Another consideration is that the Oracle 10*g* provides functionality that centralizes the management of database information. Oracle Automated Storage Management (ASM) is a Logical Volume Manager and file system that was developed for Oracle 10*g* databases, especially those deployed in Oracle Grid Computing environments. Since ASM provides file system management functionality, users are often advised to not configure ASM with NFS-based files. Although it will work, for optimal performance this is not considered best practice. It is most often recommended that Oracle 10*g* ASM environments be deployed on Fibre Channel SAN or other block-based I/O storage architectures.[4]

### Summary

Oracle databases have a block-oriented and contiguous storage strategy. NetApp's WAFL architecture inherently undermines this Oracle model through random writes and fragmentation. EMC platforms are designed for in-place write schemes, which match the architectural assumptions of Oracle's Locality of Reference. These differences should be carefully considered when deploying Oracle database applications in the enterprise to ensure performance and cost efficiency at scale as the use and importance of the Oracle system grows to the business.

## High Availability: Challenges and Concerns

Most production databases are mission-critical, and as individual databases tend to become more critical over time, IT architects must give thought to availability concerns. Deploying redundancy to achieve high availability can be expensive. For this reason, many IT architects are interested in using these additional resources and multiple copies of information for workload compression (doing additional work) during normal operation. Ideally, the benefit from additional processing capability would mitigate the additional expense of redundancy, leading to a best-of-both-worlds scenario for Oracle database environments. These additional resources needed for both high availability and, ideally, workload compression comprise servers, Host Bus Adapters (HBAs), storage network paths, and the array controller as well as disk drives.

### Hardware Redundancy

Oracle databases run on servers, making the protection and leverage of those servers critical. One of the most interesting features of Oracle RAC 10*g* is its support for dynamic server clustering. This features supports N+1 clustering that can be used to either gain additional performance, or failover a malfunctioning server from the same pool of resources. Thus, Oracle RAC 10*g* is an attractive option for many Oracle users wanting both high performance and high availability from the same investment.

For those customers deploying Oracle 10*g* Grid Computing environments with Real Application Clusters, storage hardware high availability is a key consideration to ensure no weak link exists in the IT stack. Multiple storage controllers within an array provide both higher availability and additional performance. Storage traffic is typically directed over both controllers during normal production; in the event of a storage controller malfunction, work is automatically transferred to the surviving storage controller environment.

This general discussion is also valid with respect to NAS filers as well as block storage devices. The choice of storage controller architecture has implications as to whether the additional controller can be used to increase performance, or simply to guard against a storage controller failure.

### HBA and Network Path Redundancy

Another area that is a candidate for an investment in both high availability and workload compression is host bus adaptors and their associated network paths. MPIO software (multi-path I/O) creates a volume abstraction on servers that provides the basis for both. MPIO software supports failover of either HBA or network path in the event of a malfunction. Additionally, advanced MPIO implementations support load balancing over available network paths, which can provide additional performance benefits (e.g., workload compression) during peak periods.

The choice of MPIO software has implications as to whether the additional ports and paths can be used for workload compression to achieve increased performance, or simply to guard against a failed component.

### Snapshots / Clones / Consistency Groups

Many production sites guard against data corruption using snaps and/or clones. These are logical copies of production volumes kept in reserve should the primary database become corrupted and a quick restore be needed. Snaps and clones can also be used for workload compression. A logical copy of the database can be presented to a tape backup application, minimizing database downtime, or a logical copy of the database can be used for running reports, conducting tests, or other uses.

Snapshots are not full copies of actual production data, but rather point-in-time images (lists of files) of the production data from which they were taken. As such, they require less disk space than a full clone, which is a full disk-to-disk copy of the entire production database and associated files. Given the low resource cost of creating snaps, they are an attractive means of point-in-time recovery from minor errors, while clones are deployed for more intensive workload compression activities or for full database backup/recovery from scenarios that have corrupted the entire database.

Snap and clone technologies differ vastly between storage vendors in their ability to provide full recoverability as well as workload compression.

### Disk Drive Redundancy

Storage arrays guard against disk failure by storing redundant copies of information, using either RAID 1 (mirroring), or RAID 5/6 (parity). Although both provide adequate protection of data for most cases, there are performance implications during recovery from a failed disk drive.

## A Comparative Look: How EMC and NetApp Address High Availability

As discussed previously, there are many capabilities requisite in providing a high-availability environment for Oracle databases and associated applications. While the needs may be relatively simple to identify, implementing a robust storage solution that meets these needs in a cost-effective and scalable fashion requires significant architectural design and implementation planning. While each storage vendor has a predisposition towards its own solutions, the underlying architectures and approaches to deliver such a storage solution vary, even within a given vendor. Ultimately, the end-user organization needs to make a decision about the architecture that best fits its philosophical approach to storage implementations; however, the degree of architectural alignment between the storage vendor and Oracle's strategy for database storage is an important consideration in the selection process.

What follows is a comparative look at how the EMC CLARiiON, EMC Celerra, and NetApp FAS solutions implement high availability in their product offerings.

## EMC CLARiiON Block-Oriented Storage

CLARiiON's architecture is designed for full end-to-end redundancy. Within a single storage array, storage controllers are replicated, as are other key components such as power supplies. This ensures that failover from any one component is transparent and fast. EMC has tested CLARiiON with Oracle 10*g* RAC in a variety of operating system environments and server hardware combinations, and has produced an extensive set of best practices regarding configuring such environments to enhance performance, availability, and manageability.

At the HBA and network path level, CLARiiON supports most popular MPIO software packages, including EMC's PowerPath, which provides a single multipathing product that runs on a wide range of operating systems.

CLARiiON provides clones that reside on disk drives physically separate from the production database LUN, enabling a variety of workload compression applications including accelerated backup to tape, independent testing of the database, reporting, and so on. A CLARiiON clone is always in sync with the source LUN until it is split off or fractured. Creating clones with CLARiiON provides a variety of benefits:

- Spindle independence approach follows Oracle database layout best practices[5] which recommend separating data files and recovery files onto different volumes for performance and recoverability. In the event that a source LUN experiences a multiple-drive failure, or the database becomes corrupted, the clone would remain accessible.
- After initial synchronization, subsequent resyncs copy only data that has changed on the source LUN. It is also possible to write to the clone and then propagate only the changes back to the source LUN (reverse-synchronization).
- Protected restore mode option ensures that the clone will not change even as Oracle Redo is actively changing data blocks on the source LUN. In a situation where database recovery needs to be reapplied, the process can be restarted using the clone gold copy.

Managing space reserves associated with snaps and clones is critical. CLARiiON supports the identification of specific capacity pools with snaps and clones, typically constituting a space reserve of 20% of the production volume. CLARiiON snaps reside on disk drives physically separate from the production database LUN, stored in a global area called the reserved LUN pool. Even with this space reserve, sometimes there is insufficient space to create a copy. In this scenario, the spindle independence afforded by CLARiiON ensures that only the snap and/or clone procedure fails if there is insufficient disk space. There is no impact on production volumes.

CLARiiON allows snaps to be taken at the LUN level. In an Oracle database environment, certain LUNs, such as backups, may only need to be snapped once per day, while others, such as reports, might be needed more frequently. This fine level of granularity in what data is snapped and how often ensures optimized management and the lowest amount of storate needed for snaps.

There is overhead with any replication approach. CLARiiON snapshots require additional I/O processing to create snapshots using a copy-on-first-write method. A changed block is copied once; however, subsequent writes to that block do not need to be copied, preserving performance over time as more and more snapshots are created.

Many Oracle database implementations will span multiple LUNs or multiple storage platforms for performance or scalability reasons. Creating transactionally consistent copies of data can be complex to achieve as transactions are in process during replication. There are several considerations to ensure transaction consistency in Oracle database environments:

- Oracle database applications that have interrelated data spread across multiple LUNs (such as an Oracle 10*g* database with logs) require a dependent-write consistent image to be created for restart.
- Hot backup may not be possible due to a heavy server workload and cold backup may require more downtime than is permissible. Consistency technology can minimize host-side operational impact in these cases and enable more frequent backups.
- Oracle ASM can trigger periodic rebalances of the underlying storage volumes. If this takes place during a replication procedure, inconsistency in data might result if no consistency software were deployed.

CLARiiON provides consistent split technology for snaps and clones, enabling customers to start a snapshot session or fracture a set of clones across a set of LUNs at the same time. Using clones, for example, EMC SnapView will defer write requests to the source LUNs for the Oracle database until fracturing has completed on all of the clones. This can deliver a simplified solution to ensure transactionally consistent Oracle database replication with minimized impact to production applications. EMC and Oracle have developed joint best practices around the use of this technology.[6]

CLARiiON supports both mirrored (RAID 1) and parity RAID approaches enabling solution architects to precisely specify the performance levels required during drive rebuilds. In accordance with Oracle's Stripe And Mirror Everything (SAME) methodology for achieving High Availability,[7] RAID 1 mirroring is recommended for data protection.

## EMC Celerra File-Oriented Storage

EMC's block-oriented storage products can also be presented as NAS through the use of Celerra in addition to CLARiiON and/or Symmetrix. Celerra inherits many of the attributes of the underlying storage platform, and adds a few of its own.

Larger Celerra NAS configurations support N+1 clustering of NAS processors, enabling more cost-effective failover as well as better aggregate performance. Celerra has an N+1 shared cluster approach as part of an enclosed data mover architecture that is coupled with a back-end storage infrastructure that shares resources. In the Celerra architecture, all data movers have access to all underlying storage as illustrated in Figure 4. In this architecture, failover transparently occurs to the passive standby resources, thus avoiding a performance hit on the components. Volumes can be moved between data movers as deemed necessary.

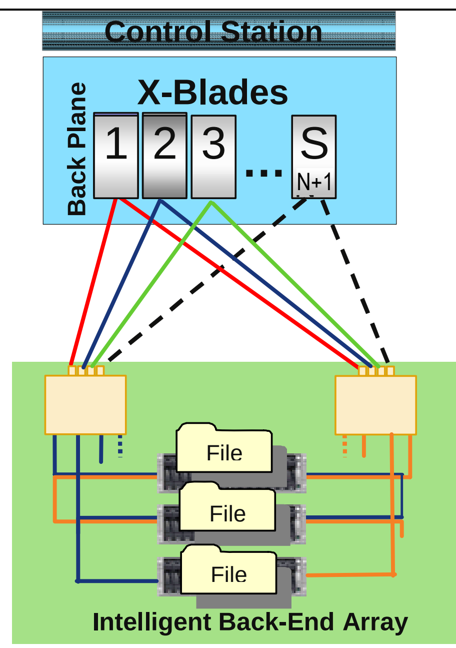


Figure 4:
EMC Celerra Architecture N+1 Shared Cluster

The N+1 shared cluster approach EMC Celerra takes with regards to storage architecture directly complements the N+1 cluster approach taken by Oracle RAC 10*g*.[8] Used together, the Oracle I/O stack benefits from shared, clustered hardware resources across servers and Data Movers which are quickly and easily scalable, documented in several Oracle and EMC joint best practices. The choice of whether to deploy Celerra or CLARiiON is driven by the organization's preference for either a Fibre Channel or IP NFS connectivity.

## Network Appliance FAS

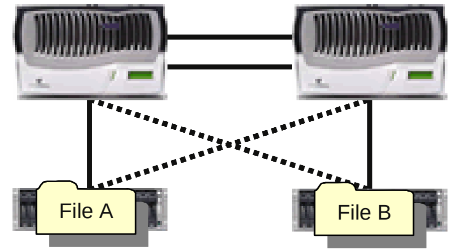


Figure 5
NetApp FAS Device Architecture
Active/Active Share Nothing Cluster

The Network Appliance FAS device is a single-node that does not inherently support high availability. It can be upgraded to a dual-node system. (See Figure 5.) The dual-node system is a bounded two-node cluster that provides failover and failback. Cluster support is built into NetApp's Data ONTAP operating system.

When running in an active-active controller configuration, NetApp best practices recommend to keep the CPU utilization at under 50%,[9] so that if failover occurs, the workload will not overwhelm the surviving node. In addition, it is recommended that the active-active controllers do not exceed the single storage controller limit. Since NetApp cluster failover is an active-active failover system it does not allow an individual node to run higher than 50% in the active-active mode.

The result of this architectural design is that organizations running a NetApp solution in an HA configuration often are encouraged to purchase twice the anticipated storage. This dramatically affects the TCO of the storage solution.

| |
|---|
| Windows: MPIO |
| Solaris: VERITAS DMP |
| AIX: Dot Hill |
| Linux: Qlogic SANsurfer |
| HP-US: HP PV links |

While NetApp FAS devices provide limited redundancy for back-end failovers, multipath software is needed should protection from front-end failures (such as HBA, cable, switch, network, and front-end port failures) be desired. With no native multipathing software, NetApp is forced to utilize a variety of different multipath software products, depending on the operating system employed in the Oracle database environment. This adds an additional layer of vendor complexity to the environment. As customers deploy multiple Oracle databases across a variety of operating systems, this scenario will become more complex.

In the NetApp architecture, all snaps are integrated with WAFL and bundled with Data ONTAP. They are shared-spindle; there is no provision for ensuring that the copies land on disk spindles separate from production volumes as there is no global storage pool for WAFL snap space. Therefore, any use of FAS snaps for workload compression will compete for the same computing resources as production data, which can significantly impact production database performance. It is often recommended best practice to first replicate the production data to different spindles and then create snaps and clones, which would defeat the cost savings of leveraging clones in the first place.

As mentioned previously, even with space reserve approaches, sometimes there is insufficient disk space to create a copy. In the case of a WAFL snapshot, not only will the snapshot operation fail, but the Oracle production instance itself will fail since there is no spindle independence.

Rather than managing shapshots at the LUN level, WAFL snapshots must replicate all the LUNs within a volume. In an Oracle database environment, certain LUNs may only need to be snapped once per day (e.g., backup), while others are required multiple times per day (e.g., reports). NetApp FAS devices require the entire file system be snapped as a unit, which can double the storage requirement and increase cost.

Given the lack of snapshot spindle independence, if a disaster occurs that damages the production data to the point of no return, the NetApp snap can not be used to restore since the snap was simply pointing to the production data that is now gone.

NetApp introduced FlexClone as a means to provide cloning capabilities. However, FlexClone does not provide a true clone capability, but is merely a writeable snapshot consisting of pointers back to a FlexVol. The FlexVol is not a LUN, but rather a volume that can contain one or more LUNs. In addition, the FlexClone exists in the same aggregate as its parent volume, meaning it shares the same disk spindles as the parent FlexVol. Because of this approach, catastrophic disk failures can wipe out Oracle data on FlexVols, snapshots, and LUNs in one stroke. The inability to create a truly independent relationship between production and snapshots/clones is also in conflict with Oracle best practices of separating data files and recovery files onto different spindles for the purposes of ensuring recoverability.[10]

In order to separate FlexClone from its parent volume, NetApp does provide a "FlexClone Split" operation, which is an separate copy operation requiring additional time to complete. Confusion can arise because FAS implements FlexClones and LUN clones, requiring different commands. Users would need to first create a snapshot, then create a LUN clone. In order to place data onto separate spindles, NetApp must provide users with additional software to copy the data. This is an additional copy procedure requiring time to complete.

The split LUN clone has no relationship to the source LUN, making it very complex to perform an incremental resync to the FAS LUN clone. In a scenario of frequent Oracle backups or refreshes to a reporting system, this inability to perform incremental refreshes adds time and management complexity in creating clones and can lead to excessive use of storage capacity through excessive replication.

The lack of uniform consistent split functionality poses significant challenges as users are left to achieve this through a manual approach, and the recommended best practices of NetApp (and Oracle) raise questions as to the viability of using NetApp FAS devices to achieve transactional consistency during replication.

NetApp best practices recommend the use of different FlexVols within the same aggregate for all Oracle data files versus recovery files.[11] This means that the various Oracle files, with their possibly very different I/O characteristics, will all share the same spindles as part of the same larger aggregate. However, NetApp snaps occur at a volume level, not an aggregate level. So, if users adhere to NetApp best practices, and separate Oracle database files into different FlexVols, database consistency during replication cannot be effectively managed or maintained in the array.

To use the array to maintain consistency, all of the Oracle files would have to be stored in the same single FlexVol. This type of configuration would conflict with NetApp's best practices noted previously and conflicts with Oracle best practices of separating data files and recovery files onto different spindles.

All disk protection on FAS is parity-based; RAID 1 is not an option. Many Oracle documents, including their Maximum Availability Architecture best practices, recommend RAID 1 for transaction log files for optimized performance as well as ensuring complete recovery. If transaction logs are lost, database recovery (if still possible) can become extremely time consuming and costly; many higher-end Oracle databases use RAID 1 for production data stores as well.

## Summary

Through the use of built-in clustering, flexibility of RAID options, and native I/O multipathing, EMC platform architectures are designed to meet the requirements of Oracle databases. EMC architectures provide spindle independence of both snaps and clones, ensuring IT organizations the flexibility to leverage snaps or clones for recovery, without impacting production volumes, and with full recoverability even if the production database becomes completely unavailable.

NetApp architectures were designed originally as file-serving appliances, which translates well into certain environments, but for Oracle database environments, poses serious challenges for high availability. Lack of inherent clustering and sharing of resources between FAS devices creates the need for excess storage infrastructure to achieved desired levels of HA with sustained performance. Inflexibility around spindle independence for snaps/clones as well as RAID options expose the business to risk in unrecoverability of Oracle production database data.

Overall, the comparative differences between EMC's and NetApp's fundamental architecture design illustrate their alignment with Oracle's best practices in database storage. EMC's archicture aligns well with Oracle's, whereas NetApp solutions force complex decisions and tradeoffs in implementing high-availability production in Oracle database environments.

## Understanding True Acquisition and Utilization Cost

Sophisticated users of IT technology typically go beyond vendor claims, and try to determine whether the proposed solution can meet their organization's requirements both now and in the future, and what the true costs will be to the organization.

It is our opinion that database storage architectural requirements evolve over time. What may appear to meet requirements today in terms of performance, availability, growth, replication, security, and so forth may not do so in the future.

Costs also evolve over time. What might appear financially attractive at the outset might become less attractive over time as hidden costs pile up. IT solution architects should ask about costs associated with specific use cases and outcomes over longer time frames, rather than just initial product costs.

Starting over with a new solution is usually a very expensive proposition. Therefore, it is important to choose a storage solution provider who can provide complete solutions against all Oracle database requirements, offering a single source of support and consistent architectural approach.

Some key issues that organizations need to consider include:

- The storage architecture must be able to scale and perform tomorrow as the database grows and the workload increases. Thought should be given to how the existing storage architecture will be able to expand for new projects, such as data warehousing, which may pose vastly different I/O usage requirements on the storage itself.
- If high availability is needed for the environment, additional hardware and utilization costs must be considered. Such costs include availability of applications through a failover, and flexibility in RAID and other hardware protection to properly serve the Oracle Applications.
- If the application requires the performance accorded from native block access without a storage-based file system, the organization must view the available options in light of how the approach works in the context of an Oracle 10*g* ASM file system.
- If snapshots and clones are leveraged for high availability and workload compression, the cost of the solution, including hardware and efficiency, must be weighed.
- As the Oracle database applications grow in significance to the business, it is vital to ensure transactional integrity and full recoverability in a timely manner.

Table 1: EMC vs. NetApp for Oracle Database Environment Requirements

| Oracle Database Environment Requirements | EMC | NetApp |
|---|---|---|
| In-place-write approach to storage architecture for maximized performance | Y | N |
| Built-in hardware/controller clustering without complexity of a "virtual" clustered environment | Y | N |
| Independence of snaps/clones from production spindles for performance and recoverability | Y | N |
| Ability to snap database copies at LUN level for higher granularity and storage cost reduction | Y | N |
| Incremental resync capability for updating Oracle database blocks | Y | N |
| Consistent split software to simplify creation of transaction consistency during replication | Y | N |
| Native I/O multipathing for flexibility of server/OS deployment and maximized HA | Y | N |
| Parity RAID flexibility for data protection / recoverability of Oracle database logs | Y | N |
| Adherence to Oracle's database best practices without forcing deign tradeoffs | Y | N |

# Implementation of Mission-Critical Databases and Applications with Oracle

It is easy for a vendor to make claims about its product's abilities; however, vendor claims alone are not enough to assure that an organization's substantial investment will deliver on its promises. Third-party testing and validation goes a long way to assuage the concerns of purchasers, but when a vendor makes substantial investments in another vendor's technologies it is an endorsement that carries considerable weight in the buying community.

Oracle has substantial numbers of databases, application development, and test environments, as well as other file and data stores thorough its enterprise. Not surprisingly, the company has purchased storage technology from many vendors and has deployed this technology in ways that are commensurate with its best practices. Network Appliance has a large, multi-petabyte storage footprint within Oracle. For many potential customers this might be viewed as a ringing endorsement of the platform. However, this storage is typically used for file sharing development environments or in support of hosted services for independent third-party customer environments, not in support of Oracle's Global IT production applications.

What is more telling is how Oracle supports its three primary revenue-generating applications, including Global Single Instance Global Mail and Worldwide Database Development. To meet the needs of these production environments, Oracle has deployed over 3 petabytes of EMC storage solutions. In fact, Oracle Global IT runs the overwhelming majority of its revenue-generating applications on EMC hardware, the most significant of which is Oracle Global Single Instance, promoted by Oracle as the driving force behind a multi-billion dollar IT cost savings. Oracle has deployed all varieties of EMC platforms, including Symmetrix, CLARiiON, and Celerra to serve all tiers of storage functionality and access for production, data warehousing, and development activities.

Similarly, EMC is one of the five largest Oracle customers in the world, based on the number of Oracle Application modules, users, and database instances deployed. Given this unique bond, EMC and Oracle IT executives talk with one another regularly, exchanging information and experiences not only as vendors, but most importantly as customers. EMC's first-hand experience with Oracle directly feeds Oracle product requirements and assessments; conversely, Oracle's experience with EMC technology provides valuable input to the future of EMC storage solutions. This level of collaboration and strategic investment into another's technology yields considerable benefit for any organization seeking the highest ROI on their database and storage deployments.

## What Does It All Mean?

Not all databases are created equal. Oracle database solutions have specific performance and data storage requirements. Oracle has articulated its own vision for storage, which involves clustering, disk management, and replication, as well as promoting the Linux OS. Organizations that deploy storage solutions that are not in alignment with Oracle's strategy may find themselves at risk. The choice of storage architecture impacts database performance, storage utilization, suitability as a high-availability solution, and the ability to tune and optimize the solution to meet organizational goals. Quite simply, some storage solutions are designed to meet the needs of Oracle environments; others are not.

There are out-of-the-box offerings from NetApp that can be deployed to support Oracle environments. However, for a growing organization, the architectural approach taken by Network Appliance can show its design limitations quickly. The initial performance and utilization experienced will degrade over time as fragmentation occurs due to WAFL. To implement high availability, organizations will need to purchase additional hardware that will only be partially utilized in order to achieve operational continuity at consistent performance levels in the case of an operational failure. While some may reap immediate benefit from the relative simplicity of such a solution, the long-term strategic impact of this platform results in higher capital expenditure and operational costs with decreased utilization and flexibility.

EMC solutions feature an architecture that is optimized for common Oracle database workloads including OLTP production, data warehousing, and related applications. These solutions offer inherent high availability and customer choice of storage protocols in order to meet current customer needs, and also offer cost-effective growth scenarios to support future business needs. EMC offers a range of platforms and functionality to enable a customer solution that is the best match and best ROI for each specific database/application environment. This is especially true for organizations that have made a strategic commitment to Oracle's storage model as instantiated by Oracle 10*g*, ASM, and RAC, and who will find more synergy with EMC's architecture and a greater divergence with Network Appliance's architecture.

Oracle has selected EMC as its storage vendor of choice for the overwhelming majority of its revenue-generation applications and databases. EMC and Oracle meet regularly, as each other's customer, to discuss technical and operational issues associated with their respective products and provide valuable input into each product's roadmap. EMC's first-hand experience with Oracle uniquely differentiates its offerings from the competition. It is for these reasons EMC platforms are deployed in over 55,000 Oracle installations worldwide.

Organizations that are considering upgrading or installing new Oracle environments are strongly advised to review the suitability of their storage infrastructure for database-oriented solutions. If after review, it is found that the data architecture is not the best match for a database scenario, organizations are well advised to consider the options afforded by the EMC family of data storage solutions.

---

[1] Further reading: *Oracle Information Appliance Initiative.* http://www.oracle.com/solutions/business_intelligence/oiai.html

[2] *Linux 64 bit performance with NFS, iSCSI and FCP using an Oracle Database on NetApp Storage.* http://www.netapp/com/library/tr/3495.pdf (12,096 GB of raw storage configured: 1,000 GB utilized for actual test)

[3] *Using Navisphere QoS Manager in Oracle Databases.* www.emc.com/techlib/pdf/H2488_using_navisphere_mgr_oracle_wp_ldv.pdf

[4] *Using Oracle 10*g *Automatic Storage Management with Network Appliance Storage, A Joint NetApp and Oracle Whitepaper,* June 2004.

[5] *Using Navisphere QoS Manager in Oracle Databases.* www.emc.com/techlib/pdf/H2488_using_navisphere_mgr_oracle_wp_ldv.pdf; *Oracle Database Installation Guide 10g Release 2 (10.2)*

[6] Further reading: *EMC CLARiiON SnapView and MirrorView for Oracle Database 10g Automated Storage Management.* http://www.emc.com/techlib/pdf/H2259_clariion_ snapview_ mirrorview_Oracle_10g_wo_ldv.pdf

[7] *Oracle High Availability Reference Architecture and Best Practice*

[8] *EMC Solutions for Oracle 10*g *RAC, EMC Celerra NS Series NFS, Best Practices Planning.* March 2007

[9] NetApp *Active-Active Controller Configuration Overview and Best Practices Guidelines*

[10] *Oracle Database Installation Guide 10g Release 2 (10.2)*

[11] *Network Appliance Best Practices Guidelines for Oracle*