Abstract

This paper provides a performance testing summary for NetApp and partner systems engineers who are interested in assessing database performance with all-flash FAS8000 systems.

NetApp® all-flash FAS systems uniquely combine the extreme performance capability of flash media with the industry-leading Data ONTAP® platform to provide performance acceleration, operational agility, best-in-class data protection, and business continuance for database deployments.
LIST OF FIGURES

Figure 1) Clustered Data ONTAP..................................................................................................................6
Figure 2) FAS8060 DNFS test configuration. ..................................................................................................10
Figure 3) Oracle on NetApp all-flash FAS8060 100% read test results. .........................................................13
Figure 4) Oracle on NetApp all-flash FAS8060 90% read test results. ...............................................................14
Figure 5) Oracle on NetApp all-flash FAS8060 90% read with Snapshot copies test results..............................14
Figure 6) FAS8080 ASM test configuration. ......................................................................................................16
Figure 7) Oracle ASM disk group layout per database.........................................................................................17
Figure 8) Oracle on NetApp all-flash FAS8080 100% read test results. ............................................................19
Figure 9) Oracle on NetApp all-flash FAS8080 70% read test results. ...............................................................20
Figure 10) Oracle on NetApp all-flash FAS8080 90% read with Snapshot copies test results..........................20
Figure 11) Oracle on NetApp all-flash FAS8060 100% read test results. ............................................................21
Figure 12) Oracle on NetApp all-flash FAS8060 90% read test results. .............................................................22
Figure 13) Oracle on NetApp all-flash FAS8080 100% read test results. ............................................................23
Figure 14) Oracle on NetApp all-flash FAS8080 90% read test results. .............................................................23
1 Introduction

The NetApp flash portfolio is capable of solving database performance and I/O latency problems encountered by many database deployments. The majority of databases have a random I/O workload that creates a problem typically solved by increasing the amount of spinning media.

Many customers are required to deliver improved performance and increase efficiencies while maintaining an enterprise-class infrastructure. NetApp provides an enterprise-class all-flash fabric-attached (AFF) storage solution that can be used to solve complex database performance requirements within a customer's environment.

With this technology, customers can gain significant performance improvements for their Oracle® database workloads when spinning media or hybrid storage does not provide the low latency necessary for those applications. The level of improvement produced by any solution depends upon many variables and the characteristics of the workload.

This report explores how NetApp AFF storage can help overcome the performance challenges of Oracle Database workloads. This document focuses on the NetApp FAS8060 and FAS8080 systems.

2 Audience

This document is intended for the following audiences:

- Users familiar with NetApp clustered Data ONTAP who are also interested in all-flash solutions
- Users familiar with Oracle Database technologies and features

3 NetApp All-Flash Fabric-Attached Storage and Clustered Data ONTAP

The NetApp all-flash FAS solution shares the same unified storage architecture, Data ONTAP OS, management interface, rich data services, and advanced features set as the rest of the fabric-attached storage (FAS) product families. This unique combination of all-flash media with Data ONTAP delivers the consistent low latency and high IOPS of all-flash storage, with the industry-leading clustered Data ONTAP OS. In addition, it offers proven enterprise availability, reliability, and scalability; storage efficiency proven in thousands of database deployments; unified storage with multiprotocol access; advanced data services; and operational agility through tight application integrations.

3.1 FAS8000 Technical Specifications

Table 1 provides the technical specifications for the four FAS series: FAS8080 EX, FAS8060, FAS8040, and FAS8020.

Note: All data in Table 1 applies to active-active dual-controller configurations.
Table 1) FAS8000 storage system technical specifications.

<table>
<thead>
<tr>
<th>Features</th>
<th>FAS8080 EX</th>
<th>FAS8060</th>
<th>FAS8040</th>
<th>FAS8020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum raw capacity</td>
<td>5760TB</td>
<td>4800TB</td>
<td>2880TB</td>
<td>1920TB</td>
</tr>
<tr>
<td>Maximum number of drives</td>
<td>1,440</td>
<td>1,200</td>
<td>720</td>
<td>480</td>
</tr>
<tr>
<td>Controller form factor</td>
<td>Two 6U chassis, each with 1 controller and an IOXM</td>
<td>Single-enclosure HA; 2 controllers in single 6U chassis</td>
<td>Single-enclosure HA; 2 controllers in single 6U chassis</td>
<td>Single-enclosure HA; 2 controllers in single 3U chassis</td>
</tr>
<tr>
<td>Memory</td>
<td>256GB</td>
<td>128GB</td>
<td>64GB</td>
<td>48GB</td>
</tr>
<tr>
<td>Maximum Flash Cache™</td>
<td>24TB</td>
<td>8TB</td>
<td>4TB</td>
<td>3TB</td>
</tr>
<tr>
<td>Maximum Flash Pool™</td>
<td>36TB</td>
<td>18TB</td>
<td>12TB</td>
<td>6TB</td>
</tr>
<tr>
<td>Combined flash total</td>
<td>36TB</td>
<td>18TB</td>
<td>12TB</td>
<td>6TB</td>
</tr>
<tr>
<td>NVRAM</td>
<td>32GB</td>
<td>16GB</td>
<td>16GB</td>
<td>8GB</td>
</tr>
<tr>
<td>PCIe expansion slots</td>
<td>24</td>
<td>8</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Onboard I/O: UTA2 (10GbE/FCoE, 16Gb FC)</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Onboard I/O: 10GbE</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Onboard I/O: GbE</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Onboard I/O: 6Gb SAS</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Optical SAS support</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Storage networking supported</td>
<td>FC, FCoE, iSCSI, NFS, pNFS, CIFS/SMB, HTTP, FTP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS version</td>
<td>FAS8080 EX Data ONTAP 8.2.2 RC1 or later, FAS8060, FAS8040, FAS8020 Data ONTAP 8.2.1 RC2 or later</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2 Scale-Out
Data centers require agility. In a data center, each storage controller has CPU, memory, and disk shelf limits. Scale-out means that as the storage environment grows, additional controllers can be added seamlessly to the resource pool residing on a shared storage infrastructure. Host and client connections as well as datastores can be moved seamlessly and nondisruptively anywhere within the resource pool.

The benefits of scale-out include the following:
- Nondisruptive operations
- Ability to increase capacity and performance without downtime
- Operational simplicity and flexibility
Clustered Data ONTAP offers a way to solve the scalability requirements in a storage environment. A clustered Data ONTAP system can scale up to 24 nodes, depending on platform and protocol, and can contain different disk types and controller models in the same storage cluster.

Figure 1) Clustered Data ONTAP.

Note: Storage virtual machines (SVMs), referred to in Figure 1, were formerly known as Vservers.

3.3 Nondisruptive Operations

The move to shared infrastructure has made it nearly impossible to schedule downtime to accomplish routine maintenance. NetApp clustered Data ONTAP is designed to eliminate the planned downtime needed for maintenance operations and lifecycle operations as well as the unplanned downtime caused by hardware and software failures.

Three standard tools make this elimination of downtime possible:

- DataMotion™ for volumes (vol move) allows you to move data volumes from one aggregate to another on the same or a different cluster node.
- Logical interface (LIF) migrate allows you to virtualize the physical Ethernet interfaces in clustered Data ONTAP. LIF migrate lets you move LIFs from one network port to another on the same or a different cluster node.
- Aggregate relocate (ARL) allows you to transfer complete aggregates from one controller in an HA pair to the other without data movement.

Used individually and in combination, these tools offer the ability to nondisruptively perform a full range of operations, from moving a volume from a faster to a slower disk, all the way up to a complete refresh of controller and storage technology.

As storage nodes are added to the system, all physical resources—CPUs, cache memory, network I/O bandwidth, and disk I/O bandwidth—can easily be kept in balance. Clustered Data ONTAP 8.2.1 systems enable users to:

- Add or remove storage shelves (over 23PB in an 8-node cluster and up to 69PB in a 24-node cluster)
- Move data between storage controllers and tiers of storage without disrupting users and applications
- Dynamically assign, promote, and retire storage, while providing continuous access to data as administrators upgrade or replace storage

These capabilities allow administrators to increase capacity while balancing workloads and can reduce or eliminate storage I/O hot spots without the need to remount shares, modify client settings, or stop running applications.
3.4 Availability
The NetApp all-flash FAS solution eliminates sources of downtime and protects critical data against disaster through two key features:

- **High availability (HA).** A NetApp HA pair provides seamless failover to its partner in case of any hardware failure. Each of the two identical storage controllers in the HA pair configuration serves data independently during normal operation. During an individual storage controller failure, the data service process is transferred from the failed storage controller to the surviving partner.

- **RAID DP®.** During any database deployment, data protection is critical because any RAID failure might result in data loss, resulting in lost productivity. RAID DP provides performance comparable to that of RAID 10, yet it requires fewer disks to achieve equivalent protection. RAID DP provides protection against double-disk failure, in contrast to RAID 5, which can protect against only one disk failure per RAID group, in effect providing RAID 10 performance and protection at a RAID 5 price point.

3.5 Optimized Writes
The NetApp Write Anywhere File Layout (WAFL®) file system enables NetApp to process writes efficiently. When the Data ONTAP OS receives an I/O, it stores the I/O in battery-backed NVRAM and sends back an acknowledgement (or ACK), notifying the sender that the write is committed. Acknowledging the write before writing to disk allows Data ONTAP to perform many functions to optimize the data layout for optimal write/write coalescing. Before being written to disk, I/Os are coalesced into larger blocks because larger sequential blocks require less CPU for each operation.

3.6 Enhancing Flash
Data ONTAP has been leveraging flash technologies since 2009 and has supported solid-state drives (SSDs) since 2010. This relatively long experience in dealing with SSDs has allowed NetApp to tune Data ONTAP features to optimize SSD performance and enhance flash media endurance.

As described in the previous sections, because Data ONTAP acknowledges writes after they are in DRAM and logged to NVRAM, SSDs are not in the critical write path. Therefore, write latencies are very low. Data ONTAP also enables efficient use of SSDs when destaging cache by coalescing writes into a single sequential stripe across all SSDs at once. Data ONTAP writes to free space whenever possible, minimizing overwrites for every dataset, not only for deduped or compressed data.

This wear-leveling feature of Data ONTAP is native to the architecture, and it also leverages the wear-leveling and garbage-collection algorithms built into the SSDs to extend the life of the devices. Therefore, NetApp provides up to a five-year warranty with all SSDs (three-year standard, plus offers an additional two-year extended warranty, with no restrictions on the number of drive writes).

The parallelism built into Data ONTAP, combined with the multicore CPUs and large system memories in the FAS8000 storage controllers, takes full advantage of SSD performance and has powered the test results described in this document.

3.7 Advanced Data Management Capabilities
This section describes the storage efficiencies, multiprotocol support, and replication capabilities of the NetApp all-flash FAS solution.

**Storage Efficiencies**
The NetApp all-flash FAS solution includes built-in thin provisioning and zero-cost cloning with FlexClone® technology that offers storage efficiency and near instantaneous cloning. The comprehensive storage efficiency enables a significantly reduced storage footprint for Oracle databases. Two features make this storage efficiency possible:
• **Thin provisioning** allows multiple applications to share a single pool of on-demand storage, eliminating the need to provision more storage for one application while another application still has plenty of allocated but unused storage.

• **FlexClone** technology offers hardware-assisted rapid creation of space-efficient, writable, point-in-time images of a production database using little to no additional storage capacity. The use of FlexClone technology in database deployments provides high levels of scalability and significant cost, space, and time savings.

**Multiprotocol Support**

By supporting all common NAS and SAN protocols on a single platform, NetApp unified storage enables:

• Direct access to storage by each client
• Network file sharing across different platforms without the need for protocol-emulation products such as SAMBA, NFS Maestro, or PC-NFS
• Simple and fast data storage and data access for all client systems
• Fewer storage systems
• Greater efficiency from each system deployed

Clustered Data ONTAP can support several protocols concurrently in the same storage system. Data ONTAP 7G and 7-Mode versions also include support for multiple protocols.

The following protocols are supported:

- NFS v3, v4, v4.1, including pNFS
- iSCSI
- FC
- Fibre Channel over Ethernet (FCoE)
- CIFS

**Replication**

NetApp SnapMirror® data replication provides fast, efficient data replication and DR for mission-critical data for an enterprise. The network compression that comes with SnapMirror provides lower bandwidth utilization as well as accelerated data transfers and will reduce the RPO for the data center. The Oracle database-consistent Snapshot™ copies that are created on the production environment will be replicated to the destination storage by using SnapMirror. In the event of a failure on the production, the database is activated on the DR environment by using a database-consistent Snapshot copy and recovering the database to the desired recovery point. Optionally, these Snapshot copies can be utilized for provisioning copies of the databases using NetApp FlexClone technology, which allows for space-efficient clone creation of the databases.

**4 NetApp All-Flash FAS8000 Performance Validation**

A series of performance tests were conducted to establish the performance capabilities of both a FAS8060 and a FAS8080 with Oracle RDBMS. These were not simulated database tests. A workload generator was used to drive an actual Oracle database with I/O patterns that are likely to be encountered in real-world production environments. The objective of this testing was to determine the maximum IOPS that both systems could deliver while providing exceptional average database read latencies. In addition, the workload tests were performed using multiple read/write percentage I/O mixes, with and without hourly Snapshot copies, to compare the performance impact of using these configurations. Additionally, we tested each configuration at progressively higher workload intensities to determine the maximum
possible IOPS while delivering typically acceptable average read latencies. This was done to address use
cases where customers are seeking very high IOPS density rather than focusing on latency reduction.
Depending on the configuration tested, the I/O patterns used during our tests were as follows, using an
8KB database block size:

- 100% random reads, using an 8KB database block size, without Snapshot copies
- 90% random reads and 10% random updates with and without hourly Snapshot copies enabled
- 70% random reads and 30% random updates, without Snapshot copies

The load generator used for these three sets of tests was capable of driving different levels of a simulated
workload, each generating the specific workloads previously described. (This is how we varied the
workload intensity.) Multiple test iterations using a progressively higher workload were executed to
incrementally increase the load on the database. The load on the database was incrementally increased
to observe the impact on performance at low, medium, and high loads. At each load point, we recorded
the IOPS and average read latency reported by the Oracle database using Automatic Workload
Repository (AWR) reports.

The graphs shown are based on the read latency for several reasons. The most important is that read
latency is demonstrably the primary bottleneck in most databases and is the type of I/O that requires
flash. If a workload requires sequential I/O operations, such as full table scans, there is little value in
using flash as standard SAS drives already offer excellent ability to service such large-block I/O
operations.

Two primary types of write operations occur in these tests: redo logging and datafile updates. Datafile
writes are not generally latency sensitive since they are background operations. Measuring the latency of
such writes is relatively unimportant because higher or lower latency does not affect overall database
performance. Redo writes are usually latency sensitive, but the loads generated by these tests are so
high that the type of logging changes. Rather than a small number of latency-sensitive writes that occur
during a “commit” operation, the redo logs are being written as streams of very large block sequential
writes, which occur as the log buffers fill up. This I/O can be measured as throughput, but it is not
possible to attach a latency value to the number. For further details on write behavior, contact your
NetApp representative.

Although write latency measurements are inapplicable to this document, the impact of the write activity is
important. Any effect is visible in the read latency. Testing write workloads was performed primarily to
evaluate the effect of the background write activity upon the latency-critical foreground read activity.

5 NetApp All-Flash FAS8060 Testing

For testing the AFF FAS8060, we used a 2-node FAS8060 cluster with two Fujitsu RX300s7 servers,
running Oracle 11.2.0.4. For connectivity between hosts and storage, we used 10GbE, two connections
per server. For storage protocols we used NFSv3 on the storage system and Oracle DNFS on the
database host. A discussion of DNFS will be presented in a subsequent section of this paper. The
following is a summary of our storage and host configuration:

- 2x Fujitsu RX300s7 servers, each equipped with:
  - Intel® Xeon® CPU E5-2630 at 2.30GHz, 2 sockets, 6 cores per socket
  - 64GB physical memory
  - 2x Intel 82599EB dual-port 10GbE controllers
  - Red Hat Enterprise Linux® Server release 6.4
  - Oracle Database 11g Enterprise Edition Release 11.2.0.4.0 64-bit with DNFS enabled

- 2-node NetApp FAS8060 cluster equipped with:
  - NetApp clustered Data ONTAP 8.2.1
- 2x FAS8060 storage controllers, each equipped with:
  o 1x DS2246 disk shelf
  o 24x 800GB SSDs
  o 4x 10GbE ports

Figure 2 provides a graphical representation of the NetApp FAS8060 test configuration.

Figure 2) FAS8060 DNFS test configuration.

In this configuration, we created two Oracle Database instances, one on each server. The size of each database, in terms of actual data, was 5.4TB. This was necessary to drive the extreme IOPS to levels required to saturate the FAS8060. The volume layout for each database is provided after a brief discussion of Oracle Direct NFS.

5.1 Oracle Direct NFS

With Oracle Direct NFS (DNFS), client NFS functionality is actually provided by the Oracle RDBMS stack. Unlike native OS NFS clients, DNFS is fully optimized for Oracle data access I/O patterns and provides significantly higher performance than traditional NFS clients.

Key features and benefits of using DNFS include the following:
- Standard NFS client implementation for all hardware and operating system platforms
• Performs concurrent direct I/O:
  - Bypasses OS-level caches, thereby reducing host memory utilization
  - Reduces CPU overhead
• Performs asynchronous I/O, which allows overlap of processing and I/O operations, resulting in improved performance.
• Supports parallel network paths:
  - Automatic load balancing across network paths, which helps avoid network-related performance bottlenecks
  - Automatic failover between network paths, providing network fault tolerance and HA

In summary, Oracle DNFS provides better performance than native OS NFS clients and provides I/O path redundancy.

NetApp FAS systems fully support the features of DNFS. In addition to that, they provide I/O redundancy and high-performance I/O as follows:

• NFS shared volumes are created in storage aggregates, which are made up of disk RAID groups. Those RAID groups provide disk redundancy, protecting data against disk failures. With RAID DP, data is actually protected against double-disk failures.
• The NFS shared volumes, along with the data files they contain, are spread across all the disk drives in the aggregate:
  - I/O performance benefits from the total bandwidth of disks in the aggregate.
  - I/O is load balanced across all disks in the aggregate.

For each of the two databases used in our configuration, we set up 2 x 10GbE paths to be shared by all data volumes and log volumes in each of our databases, providing each database with 2 x 10GbE paths, shared by its data and log volumes.

Implementation of Oracle DNFS with NetApp FAS systems provides optimized I/O performance, network and disk redundancy, and load balancing across network paths and disk drives. Best practices for use of NetApp storage with Oracle DNFS are the same as with native OS NFS clients and can be found in TR-3633: Best Practices for Oracle Databases on NetApp Storage.

5.2 Storage and Database Layout for Oracle DNFS

In terms of procedure, provisioning storage for an Oracle database using SSDs is no different from provisioning with spinning disks. We started by creating a single aggregate on each FAS8060 controller (in addition to the existing root volumes). Each aggregate was created from 20 SSDs with a RAID group size of 20, using NetApp RAID-DP. The required volumes for each database instance were created in those two aggregates. The following is a summary of our storage configuration:

• Database 1: 1 x aggregate made up of 20 x 800GB SSDs
  - 4x data volumes, 2.75TB each, thin provisioned
  - 1x temp tablespace volume, 750GB, thin provisioned
  - 2x log volumes, 1TB each, thin provisioned
• Database 2: 1x aggregate made up of 20x 800GB SSDs
  - 4x data volumes, 2.75TB each, thin provisioned
  - 1x temp tablespace volume, 750GB, thin provisioned
  - 2x log volumes, 1TB each, thin provisioned

In our tested configuration, each data volume contained a single tablespace for tables and data. Each of those tablespaces used a single 1625GB data file for physical storage. The temp tablespace volume...
provided storage for the database’s temporary tablespace, and each log volume contained 3x redo log group members, each member having a size of 40GB.

Table 2 summarizes the storage configuration.

**Table 2)** Storage layout for Oracle DNFS.

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Aggregate</th>
<th>Volume</th>
<th>Volume Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>aggr1_1</td>
<td>datavol1</td>
<td>2.75TB</td>
</tr>
<tr>
<td>db1</td>
<td></td>
<td>datavol2</td>
<td>2.75TB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>datavol3</td>
<td>2.75TB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>datavol4</td>
<td>2.75TB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>logvol1</td>
<td>1TB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>logvol2</td>
<td>1TB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tempvol</td>
<td>750GB</td>
</tr>
<tr>
<td>db2</td>
<td>aggr1_2</td>
<td>datavol1</td>
<td>2.75TB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>datavol2</td>
<td>2.75TB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>datavol3</td>
<td>2.75TB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>datavol4</td>
<td>2.75TB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logvol1</td>
<td>1TB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logvol2</td>
<td>1TB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tempvol</td>
<td>750GB</td>
</tr>
</tbody>
</table>

5.3 Test Methodology and Results for FAS8060

This section provides the results of the AFF FAS8060 performance tests. When evaluating performance of storage arrays, it is critical to establish the definition of terms such as IOPS. For example, the IOPS values for databases are often taken from the physical reads statistic, located in the Load Profile section of an Oracle AWR report. This can be highly misleading, because at the storage layer, 128 physical reads could be performed as 128 individual random I/O operations, or they could be read as a 1MB unit as part of a longer sequential read.

In our tests, the read IOPS were calculated based on the physical reads statistic, while making sure that essentially all I/O was performed as random “db file sequential read” or “db file parallel read” I/O. This makes sure that the IOPS level in terms of Oracle block operations is an accurate reflection of the IOPS levels that occurred at the storage layer.

Write performance was not a focus of these tests. Although low-latency redo logging is critical to a production database, virtually all arrays on the market from any vendor already commit inbound writes to some type of solid-state cache. The use of SSD drives for persistent storage has little impact on this activity because the write is complete from the database viewpoint as soon as it is stored in the array cache itself. Likewise, database block write operations are generally background operations and are not latency sensitive.

For each set of tests, the workloads were incrementally increased in intensity and executed across both Oracle databases to demonstrate throughput scalability of our storage cluster. We then measured the
total IOPS delivered along with the associated average read latency at each workload level. The specifics of the workloads used are as follows:

- 100% random reads by using 8KB request size without Snapshot copies enabled.
- 90% random reads by using 8KB request size without Snapshot copies enabled.
- 90% random reads by using 8KB request size with hourly Snapshot copies enabled and a Snapshot copy retention policy of 6. With a Snapshot retention policy of 6, only 6 Snapshot copies are retained. If there are 6 Snapshot copies, creation of a new one will result in deletion of the oldest existing Snapshot copy.

Figures 3-5 shows the total IOPS and average read latency observed by the database for all workload configurations.

For tested workloads, we observed the following items of interest:

- All workloads were tested to the saturation point of one or more aspects of the configuration.
- At a low to moderate workload intensity, all three workload mixes delivered comparable overall average read latencies, due to relatively low resource utilization.
- For all workloads tested, the total IOPS for the 100% random read workload remained much higher than for 90% reads (with and without Snapshot copies).
- The 100% read workload reached about 250,000 IOPS. At this point, both CPU utilization and SSD utilization were at 100%.
- Read latencies increased for all three workload mixes as workload intensity increased. This was mainly the result of corresponding increases in CPU and SSD utilization.
- Read latency began to increase at a higher rate after ~1.5ms for all three workload mixes as FAS CPU utilization reached a range of 90% to 97%, and SSD utilization increased.
- Regardless of workload intensity, both IOPS and read latency remained very close for the 90% read workloads with and without Snapshot copies, clearly demonstrating that Snapshot copies are, for all practical purposes, free with regard to performance.

Figure 3) Oracle on NetApp all-flash FAS8060 100% read test results.
Figure 4) Oracle on NetApp all-flash FAS8060 90% read test results.

Figure 5) Oracle on NetApp all-flash FAS8060 90% read with Snapshot copies test results.
6 NetApp All-Flash FAS8080 Testing

For testing the all-flash FAS8080, NetApp used a 2-node FAS8080 cluster with two Fujitsu RX300s7 servers running RHEL 6.4 and Oracle 11.2.0.4. For connectivity between hosts and storage, NetApp used 16Gb/s FC, two connections per server. On the host side, NetApp used Oracle Automatic Storage Management (ASM). A discussion of ASM is presented in a subsequent section of this document. The following is a summary of the NetApp storage and host configuration:

- 4x Fujitsu RX300s7 servers, each equipped with:
  - Intel Xeon CPU E5-2630 at 2.30GHz, two sockets, six cores per socket
  - 64GB physical memory
  - 2x QLogic QLE2672 dual-port 16Gb/s FC adapters, configured for HA and load balancing using Linux DM-Multipathing
  - Red Hat Enterprise Linux Server release 6.4
  - Oracle Database 11g Enterprise Edition Release 11.2.0.4.0 64-bit with Oracle ASM

- 2-node NetApp FAS8080 cluster equipped with:
  - NetApp clustered Data ONTAP 8.2.2
  - 2x FAS8080 storage controllers, each equipped with:
    - 2x DS2246 disk shelves
    - 48x 800GB SSDs
    - 4x 16Gb/s FC ports

Figure 6 provides a graphical representation of the NetApp FAS8080 test configuration.
In this configuration, NetApp created two Oracle databases, one on each server. The size of each database was 5.4TB, in terms of actual data. Even though the database size remained the same as that for our FAS8060, NetApp used twice as many SSDs to avoid encountering an SSD bottleneck and limiting using the expanded I/O bandwidth of the FAS8080 controllers. This was necessary to drive the extreme IOPS to levels required to saturate the FAS8080. The volume/LUN layout for each database is provided after a brief discussion of Oracle ASM (following).

### 6.1 Oracle ASM

Oracle ASM provides integrated cluster file system and volume management features, removing the need for third-party volume management tools, while reducing the complexity of the overall architecture.

Some of the key Oracle ASM features include:

- Automatic file and volume management
- Database file system with the performance of raw I/O
- Automatic distribution and striping of data across ASM disks
• A choice of external (array-based) data protection, two-way, and three-way mirror protection
• Control over which copy of mirrored data should be used preferentially

With these capabilities, Oracle ASM provides an alternative to third-party file systems and volume management solutions for database storage management tasks, such as creation or layout of databases, and disk space management. Volume management tasks on the Oracle server host can be performed by using familiar create, alter, and drop SQL statements, simplifying the job of database administration with regard to database storage provisioning. Load balancing avoids performance bottlenecks by enabling the I/O workload to use all of the available disk drive resources.

NetApp FAS storage systems automatically load balance I/O across all the disk drives making up an underlying aggregate. All LUNs placed within a single aggregate can use all of the aggregate’s disk drives in a balanced manner. Oracle ASM provides further load balancing of I/O across all LUNs or files in an Oracle ASM disk group by distributing the contents of each data file evenly across the entire pool of storage in the disk group based on a 64MB stripe size. This provides even performance through the available SCSI devices at the host and network layers.

When used together, NetApp load balancing allows multiple LUNs and file system data to share common disk drives. This reduces the number of LUNs required for each Oracle ASM disk group, which improves manageability without compromising performance. As a result, read and write I/O in high-transaction Oracle database environments is optimized.


6.2 Storage and Database Layout for Oracle ASM

After storage was provisioned on the FAS8080, two separate Oracle database instances were configured, each on a separate server. This was necessary to drive the extreme IOPS to levels required to saturate the FAS8080. The layout of the databases was designed such that Oracle distributed the overall load evenly across both FAS8080 storage controllers.

Figure 7 shows the Oracle ASM disk group layout for each of the two database instances deployed for testing. One Oracle ASM disk group was created for the data, one for the redo logs, and one for the temp tablespace on each server. Table 3 shows the specifics of the disk groups in terms of the number of LUNs used for each, the total usable capacity, and total used capacity. The same configuration was used for each set of tests to allow an accurate comparison of results.

Figure 7) Oracle ASM disk group layout per database.
### 6.3 Test Methodology and Results (FAS8080)

This section provides the results of our FAS8080 performance tests. As previously explained, when evaluating performance of storage arrays, it is critical to establish the definition of terms such as “IOPS.” For example, the IOPS values are often taken from the database physical reads statistic, located in the Load Profile section of an Oracle AWR report. This can be highly misleading, because at the storage layer, 128 physical reads could be performed as 128 individual random I/O operations, or they could be read as a 1MB unit as part of a longer sequential read.

In our tests, the read IOPS were calculated based on the physical reads statistic, while making sure that essentially all I/O was performed as random “db file sequential read” or “db file parallel read” I/O. This makes sure that the IOPS level in terms of Oracle block operations is an accurate reflection of the IOPS levels that occurred at the storage layer.

Write performance was not a focus of these tests. Although low-latency redo logging is critical to a production database, virtually all arrays on the market from any vendor already commit inbound writes to some type of solid-state cache. The use of SSD drives for persistent storage has little impact on this activity because the write is complete from the database viewpoint as soon as it is stored in the array cache itself. Likewise, database block write operations are generally background operations and are not latency sensitive.

For each set of tests, the workloads were incrementally increased in intensity and executed across both Oracle databases to demonstrate throughput scalability of our storage cluster. (Loads were scaled higher than with the FAS8060 to demonstrate the higher workload capacity of the FAS8080.) We then measured the total IOPS delivered along with the associated average read latency at each level of workload intensity. The specifics of the workloads used are as follows:

- 100% random reads by using 8KB request size without Snapshot copies enabled
- 70% random reads by using 8KB request size without Snapshot copies enabled
- 90% random reads by using 8KB request size with hourly Snapshot copies enabled and a Snapshot copy retention policy of 6

Figure 8-10 shows the IOPS and average read latency observed by the database for all workload configurations.

For tested workloads, NetApp noted the following items of interest:

- At a low to moderate workload intensity, all three scenarios delivered average read latencies below 1ms. At this load level, both CPU utilization and SSD utilization were relatively light.
- For all workloads tested, the total IOPS for the 100% random read workload remained noticeably higher than for 90% reads with Snapshot copies and even higher than throughput for the 70% read workload.
- All workloads were tested to the saturation point of one or more aspects of the configuration.
- A 100% read workload was tested up to the level of about 300,000 IOPS. At this point, both CPU utilization and SSD utilization were at 100%.

- Read latencies increased for all three workload mixes as workload intensity increased. This was mainly the result of corresponding increases in storage CPU and SSD utilization.

- The read latency for the 100% read workload began to increase at a higher rate after ~1.25ms, as FAS CPU utilization reached a range of 90% to 97% and SSD utilization increased.

- As previously noted, throughput for the 100% read workload reached 300,000 IOPS. Compared to the FAS8060, which reached 250,000 IOPS, the FAS8080 outperformed the FAS8060 by about 20%.

- Having already demonstrated with the FAS8060 that Snapshot copies are free, NetApp chose not to include the 90% read without Snapshot copies test in the FAS8080 test plan.

Figure 8) Oracle on NetApp all-flash FAS8080 100% read test results.
Figure 9) Oracle on NetApp all-flash FAS8080 70% read test results.

Figure 10) Oracle on NetApp all-flash FAS8080 90% read with Snapshot copies test results.

7 NetApp All-Flash FAS Testing Using Clustered Data ONTAP 8.3

Data ONTAP 8.3 is the latest release of the Data ONTAP software, which delivers nondisruptive operations, improved performance, proven efficiency, and seamless scalability. Data ONTAP 8.3 delivers up to 60% better performance at consistently low latency in comparison to previous Data ONTAP
versions. Additionally, Data ONTAP 8.3 all-flash FAS systems also provide up to 17% additional usable capacity with the new Advanced Drive Partitioning technology.

This section provides the results of additional performance testing that utilized Data ONTAP 8.3 with the same configurations described previously for both DNFS on the FAS8060 and SAN using Oracle ASM on the FAS8080. For these tests, only the version of Data ONTAP was changed so the results are directly comparable between Data ONTAP versions.

For each set of tests, the workloads were incrementally increased in intensity and executed across both Oracle databases to demonstrate throughput scalability of the storage cluster. We then measured the total IOPS delivered along with the associated average read latency at each workload level. The specifics of the workloads used are as follows:

- 100% random reads by using 8KB request size without Snapshot copies enabled.
- 90% random reads by using 8KB request size with hourly Snapshot copies enabled and a Snapshot copy retention policy of 6. With a Snapshot retention policy of 6, only 6 Snapshot copies are retained. If there are 6 Snapshot copies, creation of a new one will result in deletion of the oldest Snapshot copy.

Figure 11 and Figure 12 show the total IOPS and average read latency observed by the database for all workload configurations using the FAS8060 with DNFS using Data ONTAP 8.2.1 and Data ONTAP 8.3.

For the tested workloads, the following items of interest were observed:

- All workloads were tested to the saturation point of one or more aspects of the configuration.
- At a low to moderate workload intensity, both workload mixes delivered comparable overall average read latencies, due to relatively low resource utilization.
- For all workloads tested, Data ONTAP 8.3 showed significantly better performance compared to Data ONTAP 8.2.1, delivering higher levels of IOPS and lower average latencies at comparable workload intensities.

Figure 11) Oracle on NetApp all-flash FAS8060 100% read test results.
Figure 12) Oracle on NetApp all-flash FAS8060 90% read test results.

Figure 13 and Figure 14 show the total IOPS and average read latency observed by the database for all workload configurations using the FAS8080 with Fibre Channel Protocol (FCP) or ASM using Data ONTAP 8.2.2 and Data ONTAP 8.3.

For the tested workloads, we observed the following items of interest:

- All workloads were tested to the saturation point of one or more aspects of the configuration.
- At a low to moderate workload intensity, both workload mixes delivered comparable overall average read latencies, due to relatively low resource utilization.
- For all workloads tested, Data ONTAP 8.3 showed significantly better performance compared to Data ONTAP 8.2.2, delivering higher levels of IOPS and lower average latencies at comparable workload intensities.

![Graph showing comparison between Data ONTAP 8.2.1 and Data ONTAP 8.3 for Oracle on FAS8060 with DNFS: 90% Reads w/ Snapshot]
8 Summary

The performance test results highlighted in this technical report prove that the NetApp all-flash FAS8000 is a highly compelling storage solution for any Oracle database deployment demanding high IOPS and
ultralow latency. Combine this with the field-proven storage efficiency, data protection, and other data management capabilities inherent in the FAS systems, and customers can deploy a very cost-efficient, high-performing, and simple NetApp storage solution for Oracle.

**References**

The following links provide additional information and reference material for the subjects contained in this document:

- TR-3982: NetApp Clustered Data ONTAP 8.2  
- TR-3633: Best Practices for Oracle Databases on NetApp Storage  

**Acknowledgements**

The authors of this paper thank Saad Jafri, Keith Griffin, Larry Touchette, Bobby Oommen, and Jeff Steiner for their contributions to this report.
Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer’s installation in accordance with published specifications.

Copyright Information

Copyright © 1994–2014 NetApp, Inc. All rights reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.227-7103 (October 1988) and FAR 52-227-19 (June 1987).

Trademark Information


Cisco and the Cisco logo are trademarks of Cisco in the U.S. and other countries. All other brands or products are trademarks or registered trademarks of their respective holders and should be treated as such.

TR-4322-1214