Technical Report

NetApp AFF A400 Performance with Oracle RAC Database

Solution Design

Joe Carter, Scott Lane, Rob McDonald, Dave Krenik, Jeffrey Steiner, Ling Wang, NetApp
March 2020 | TR-4819-DESIGN

Abstract

This report provides a performance summary for NetApp® and partner systems engineers who are interested in assessing Oracle database performance with NetApp AFF A400 storage systems.

NetApp AFF systems combine the extreme performance capability of flash media with NetApp ONTAP® 9.7 software to provide performance acceleration, operational agility, best-in-class data protection, and business continuity for database deployments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

NetApp solutions for Oracle databases are engineered for enterprise workloads with industry-leading performance, superior scalability, continuous data availability, and comprehensive data management for the data center, edge, and cloud. NetApp provides Oracle customers with the next generation of performance and manageability with the industry's first end-to-end, database-to-drive NVMe solution.

Built on ONTAP scale-out architecture, NetApp AFF systems for SAN consistently meet or exceed the high-performance demands of Oracle databases. Designed specifically for flash, AFF A-Series all-flash systems deliver nonstop performance, capacity density, scalability, security, and network connectivity in dense form factors. The NetApp AFF A400 combines low-latency NVMe solid-state drives (SSDs) and NVMe over Fibre Channel (NVMe/FC) connectivity at a mid-range price point and provides the best combination of performance and value.

NVMe/FC is the simplest, most broadly supported NVMe-over-fabrics option. It allows customers to use the same FC fabric for the NVMe protocol and their existing SAN infrastructure. Usually, NetApp customers can nondisruptively upgrade to NVMe/FC with a simple software upgrade, resulting in unprecedented performance without a requirement for network architecture changes. This upgrade allowing customers to run 60% more workloads and reduce application response times.

The benefits of adopting a more efficient storage protocol include dramatically simpler storage management by reducing the number of storage objects required to deliver peak performance. Reducing storage objects translates into reduced CPU demand, further improving database performance and potentially deferring the need for additional Oracle licenses. AFF also provides rich data management capabilities, such as integrated data protection, nondisruptive upgrades, and data migration. These features help eliminate performance silos and seamlessly integrate AFF into a shared infrastructure.

ONTAP software delivers enhanced inline deduplication, inline compression, and inline data compaction capabilities that significantly reduce the amount of flash storage required, with no effect on system performance. It also provides industry-leading ecosystem integration with database applications that makes administration of databases and storage systems far more efficient than with other flash storage solutions on the market.

2 Executive Summary

NetApp performed this study for Oracle applications to showcase the storage performance and the benefits of the AFF A400 with NVMe/FC compared to the Fibre Channel Protocol (FCP) for Oracle applications.

NVMe/FC reached a peak IOPS up to 66% higher than with FCP. Average host CPU utilization was also reduced by up to 29% with a workload over NVMe/FC.

These results show that customers can run more Oracle workloads and at lower latency by upgrading to NVMe/FC host connectivity using existing hardware.

This report will show how, by modernizing the Oracle database infrastructure, NetApp can help enhance customer retention and boost employee productivity.

3 Measuring Storage Performance

NetApp performed the following study to measure the performance of an AFF A400 storage system running ONTAP 9.7. In the following sections, we describe the methodology and design considerations used to test the AFF storage systems running a standard Oracle workload.

3.1 Test Methodology

For this study, we used the SLOB2 v2.5.2.4 load-generation tool to simulate an OLTP workload against the Oracle Database 18c test configuration. We ran three SLOB workloads mixes, one with 100% SQL SELECTs, one with a 75:25 SELECT-to-UPDATE ratio, and the last with 100% UPDATEs.

To allow us to compare FCP and NVMe/FC protocols, we created two SLOB databases. The database created to test FCP was built on volumes which contained LUNs. The database created for testing NVMe/FC was built on volumes which contained NVMe namespaces. Both databases were prepopulated with 2TB of SLOB data prior to running the SLOB workloads.

We ran the three workload mixes individually on each of the two protocol configurations.

We created a 4-node Oracle Real Application Clusters (RAC) environment using Red Hat Enterprise Linux (RHEL) 7.6 hosts with a database connected through FC to the AFF A400. We conducted the NVMe/FC and FCP testing at separate times. However, we used the same four Linux hosts, the same Brocade FC switch, and the same-sized SLOB database (2TB) on each protocol. The SLOB workload driver tool made requests to the Oracle database cluster, which would in turn drive I/O to the AFF A400.

We varied the number of threads to create curves for each workload, while keeping other parameters constant. The number of SLOB virtual users and SLOB SCALE parameter were the same at all load points to ensure the entire 2TB of data was the active working set at all load points. We ran each load point of the workload for 20 minutes with a given number of SLOB threads. We then increased the number of threads and ran for another 20 minutes. We used 2 to 210 SLOB threads for each test suite to create approximately 15 load points for each workload curve.

3.2 Hardware and Software

We configured the Oracle RAC on four Fujitsu PRIMERGY RX300 S7 servers. We connected the four servers to a Brocade G630 switch with 32Gb FC. The AFF A400 nodes were also connected to this switch through 32Gb FC through an added Emulex LPe32004. The AFF A400 HA-pair contained 24 1.9TB internal SSDs.

Table 1 and Table 2 list the hardware and software components that we used for the test configuration.

Table 1) Oracle host hardware and software components.

Hardware and Software Components	Details
Oracle database servers	4 x Fujitsu PRIMERGY RX2540 M4 servers
Server operating system	RHEL 7.6
Oracle database version	18c (RAC)
Processors per server	2 socket, 10-core Intel Xeon Silver 4414@2.20GHz, with 2 threads per core
Physical memory per server	256GB
FC network	32Gb FC with multipathing
FC host bus adapter (HBA)	Emulex LPe32002-M2 32Gb

Hardware and Software Components	Details
32Gb FC switch	Brocade G630 128-port switch
10GbE switch	Cisco Nexus 9446C-FX2

Table 2) NetApp AFF A400 storage system hardware and software.

Component	Details
Storage system	AFF A400 controller, configured as a high-availability (HA) pair
ONTAP version	9.7 RC1
FC card	2 x Emulex LPe32004 (one per node)
FC target ports	8 x 32Gb ports (four per node)
Disk Shelf	NS224
Total number of drives	24
Drive size	1.9TB
Drive type	NVMe-SSD

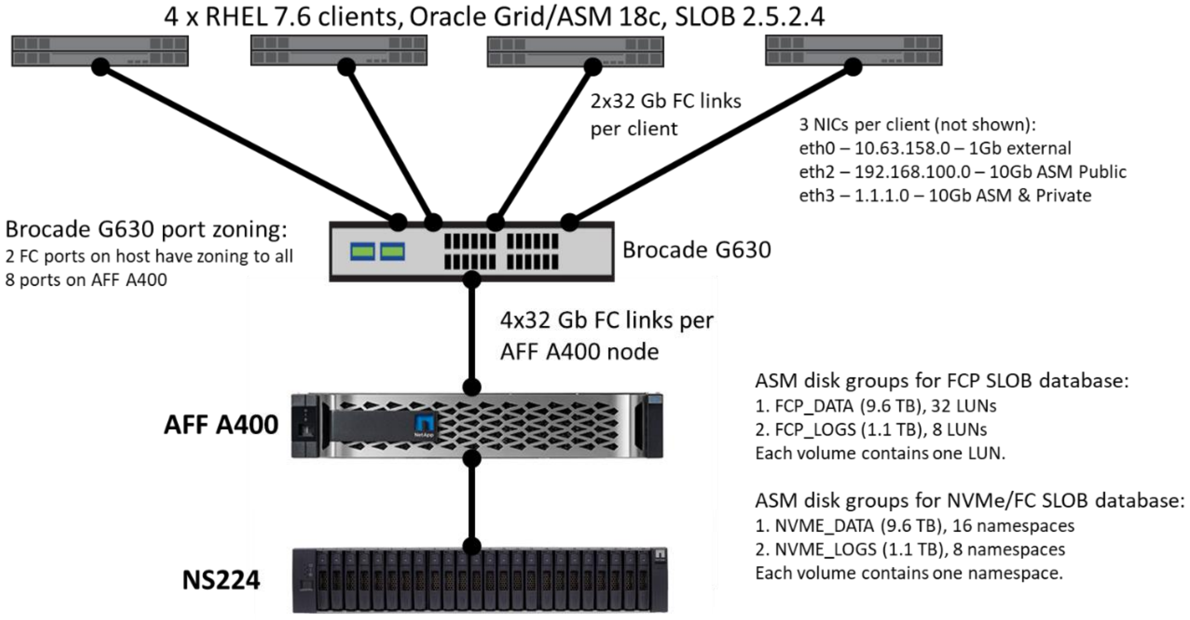
3.3 Network and Database Storage Design

Figure 1 shows the FCP SAN deployed with a Brocade G630 32Gb FCP switch. Each storage AFF A400 node had four target ports connected to the FC switch. Each host had two initiator ports connected to the switch with Broadcom LPe32002 FC HBAs.

For Ethernet connectivity, each of the four hosts had a 1Gb link for external access and a 10Gb link for the Automatic Storage Management (ASM) public network. Each host also had a 10Gb link for ASM and the private network.

Each of the four RHEL hosts had one FC port connected to the Brocade switch. Each AFF A400 storage node had four FCP target ports that were also connected to the same switch, for eight total connected target ports. We configured the Brocade switch with port zoning to map the initiator port of each RHEL host to all four target ports of each AFF A400 storage node.

Figure 1) Network design.



3.4 Database Layout and Storage Provisioning Design

In an NVMe/FC configuration, a namespace is nonvolatile memory storage that is formatted for block access. A namespace is analogous to an FCP LUN, which resides within a volume in the storage array. Both LUNs and namespaces show up as block storage devices at the host. An ASM disk group can be created across a set of LUNs or, similarly, a set of namespaces.

Figure 1 shows the layout of LUNs, ASM disk groups, and SLOB databases for the NVMe/FC and FCP test configurations. One storage virtual machine (SVM) was created for the FCP configuration. This SVM contained all logical interfaces (LIFs), LUNs, and volumes for FCP. In the NVMe/FC configuration, eight SVMs were used. Each of the NVMe/FC dedicated SVMs contained one LIF and one eighth of the namespaces.

Of the 24 SSDs in the AFF A400, 23 of the drives were used to create a single NetApp RAID DP® aggregate and one was left as a spare drive. An aggregate was created for each of the two AFF A400 nodes.

In the FCP configuration, we created 32 176GB data volumes for a total size of 5.5TB, and 8 138GB redo log volume for a total size of 1.1TB. One LUN was created in each volume.

In the NVMe/FC configuration, we created 16 352GB data volumes for a total size of 5.5TB, and 8 138GB redo log volumes for a total size of 1.1TB. One namespace was created in each volume.

The Oracle ASM Configuration Assistant was used to create ASM disk groups. In both configurations, an ASM disk group was created for data and redo logs, spanning the associated data and redo volumes.

The SLOB databases were generated and populated using the SLOB toolkit before running the performance workloads. A 3TB tablespace was created in each data disk group. The SLOB database was populated with data for 4 users and a SLOB SCALE factor of 524,288M. This arrangement resulted in 2TB of SLOB data. A 2TB temporary tablespace was also created in each data disk group to allow for initial loading of the SLOB data.

Another disk group was created for the Oracle Grid repository (CRS and Voting), on a single 50GB namespace. The Grid repository was shared on both configurations.

The settings used to create ASM disk groups are listed in Table 3.

Table 3) ASM disk group settings.

Setting	Value
ASM compatibility	18.0.0.0
Database compatibility	18.0.0.0
Sector size	512B for FCP; 4KB for NVMe/FC
Logical sector size	512B for FCP; 4KB for NVMe/FC
Allocation unit size	64MB

3.5 Workload Design

In this study, SLOB 2.5.2.4 was used as an Oracle I/O workload generation tool. SLOB can drive massive scale SQL execution against an Oracle database to simulate an OLTP workload.

We designed a set of SLOB workloads to ramp from 2 to 210 threads with approximately 15 intermediate points. Each data point ran a fixed number of threads for 20 minutes. This setup allowed us to gather performance metrics at a range of different load points and determine peak performance. We collected metrics by SLOB in Oracle Automatic Workload Repository (AWR) reports. We ran each set of data points three or more times to generate repeatable results. We ran all sets of workloads on two configurations: NVMe/FC and FCP.

We ran three different workload mixes:

- 100% SELECTs (100% reads)
- 75% SELECTs with 25% UPDATEs (an approximately 80:20 read-to-write ratio)
- 100% UPDATEs (an approximately 50:50 read-to-write ratio)

Keep in mind that this test was not designed to have high levels of caching on the six Linux hosts in the Oracle cluster. We wanted to demonstrate the capabilities of the AFF storage controller serving I/O in this workload. If we wanted to increase the SLOB throughput even further, additional caching could be configured on the Oracle servers. This setup would service more requests (especially reads) from memory on the Oracle servers, reduce the percentage of requests going to the AFF storage, and increase overall SLOB throughput.

SLOB 2.5.2.4 has a new parameter called OBFUSCATE_COLUMNS. When enabled, the SLOB setup tool will use randomized character strings instead of a repeated single character. We chose to enable this parameter in order to have a data set compressed to a realistic level. Previous versions of SLOB only allowed for a repeated single character. A repeated single character resulted in a highly compressible data set which resulted in mostly cached reads on the storage controller. With a randomized data set, storage side compression is greatly reduced, and most of the read requests are served from SSDs, not cache.

Note: We took care in these test steps to simulate more realistic database and customer workloads. However, we acknowledge that workloads can vary across databases. In addition, these test results were obtained in a closed lab environment with no competing workloads on the same infrastructure. In a typical shared-storage infrastructure, other workloads share resources. Therefore, your results might vary from the results that are found in this report.

3.6 Performance Test Results

We measured the performance of our Oracle database implementation with the AFF A400, using both the FCP and NVMe/FC protocols, with ONTAP 9.7. All other hardware and software were configured identically.

Figure 2 shows the results of these tests with a 100% SELECT workload, Figure 3 with a 75% SELECT/ 25% UPDATE workload, and Figure 4 with a 100% UPDATE workload.

The x-axis represents the total physical reads and writes per second (IOPS). Note that the IOPS metric is from the perspective of the Oracle database servers. It includes only the IOPS that were directed to the AFF storage, not the IOPS that were serviced directly by the cache on the four Linux hosts running the Oracle database cluster.

The y-axis represents the server read latency. Again, this is from the perspective of the Oracle RAC nodes and includes the FC transport time. Triangle markers represent FCP protocol tests. Circle markers are NVMe/FC.

Figure 2) AFF A400 Oracle database performance with 100% SELECT workload.

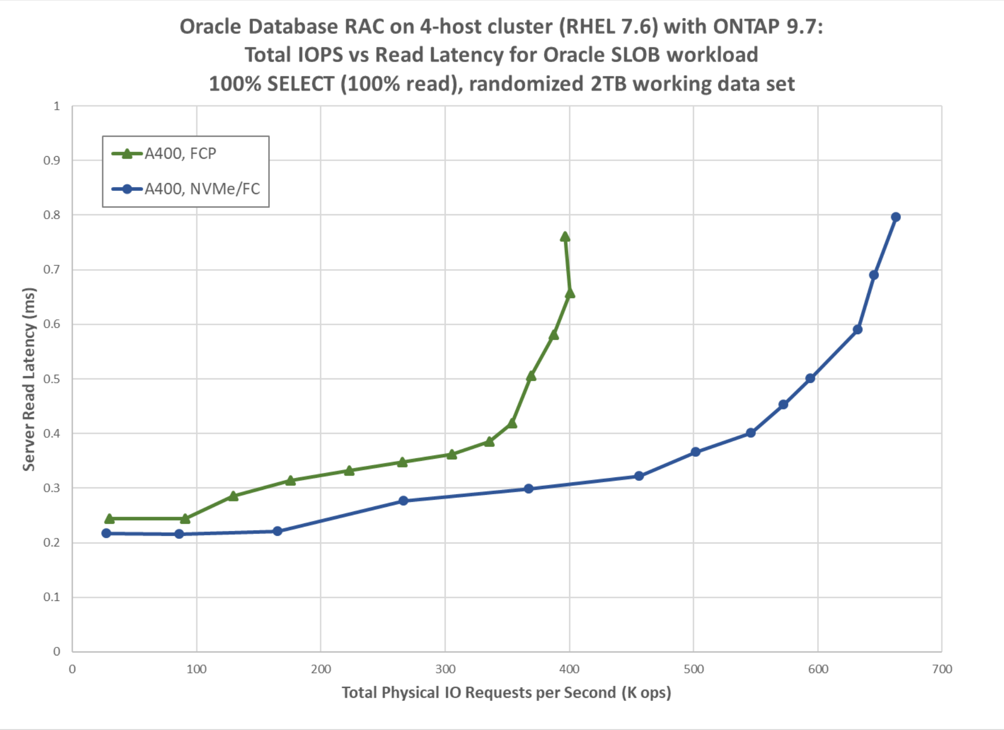


Figure 3) AFF A400 Oracle database performance with 75% SELECT / 25% UPDATE workload.

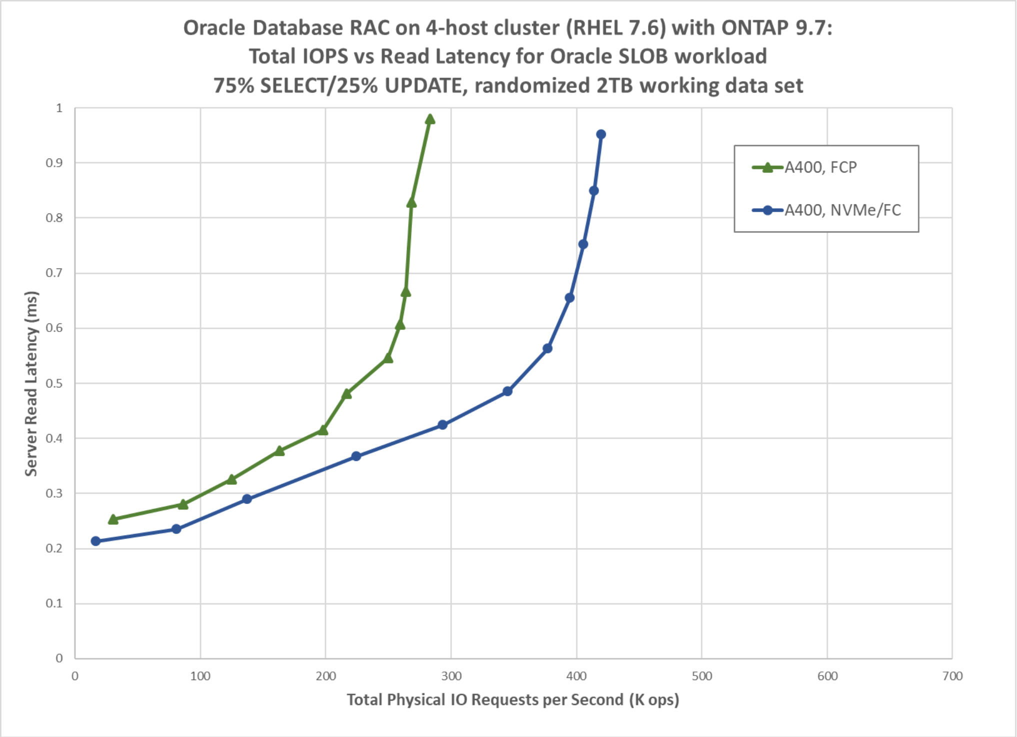
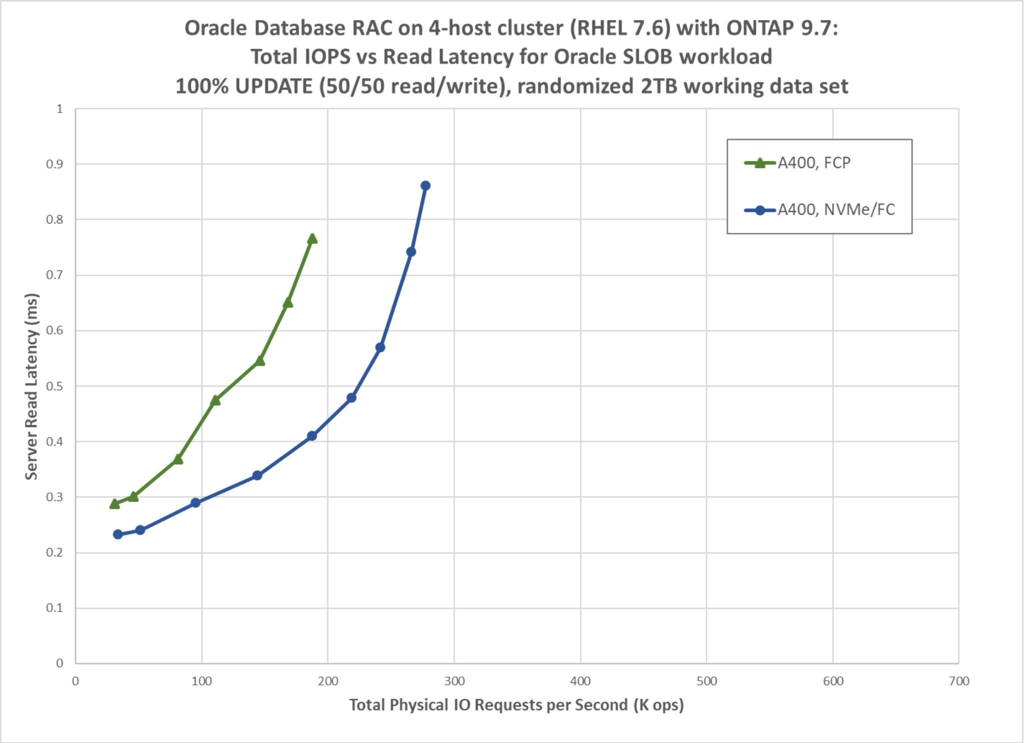


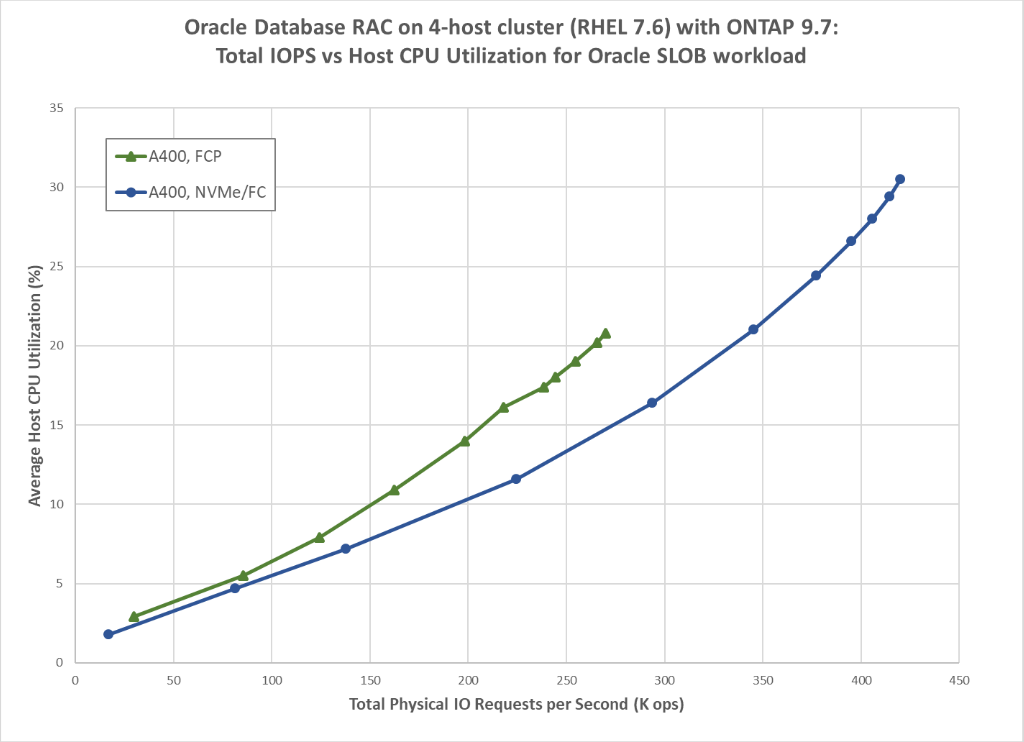
Figure 4) AFF A400 Oracle database performance with 100% UPDATE workload.



Oracle AWR reports were collected during all workloads. The Database Summary section of the AWR shows the elapsed time and the database time for that specific performance point. The Top Timed Events section shows the top 10 events and their respective latencies. The System Statistics - Per Second section shows the number of physical reads and writes per second and the number of redo log operations in kilobytes per second. For details from an example AWR report, see Appendix A: AWR Report.

Figure 5 shows average host system CPU utilization from a single host in the 4-host RAC. This metric is recorded in the AWR reports. The host CPU utilization reflects the Oracle database server processes, RHEL system processes, and the host side of the IO requests being made to the attached AFF A400 storage.

Figure 5) Host CPU utilization.



4 Conclusion

We found that the NetApp AFF A400 running ONTAP 9.7 with NVMe/FC generated very high IOPS at consistently low read latencies when serving an Oracle Database 18c OLTP workload. ONTAP 9.7 with NVMe/FC achieved 59% more IOPS than FCP while at the same latency of 400µs. The 266K IOPS load point of the 100% SELECT workload showed 70µs lower latency, or 25% faster with NVMe/FC than FCP. A peak of 663K IOPS was reached with NVMe/FC and 100% SELECT, with latency under 800µs. With the 75/25% SELECT/UPDATE workload, 420K IOPS was achieved.

We found the FCP 100% SELECT workload was able to serve up to 401K IOPS and up to 283K IOPS with a 75/25% SELECT/UPDATE workload. At 100% SELECT load points below 340K IOPS, we found latency with FCP was between 245µs and 500µs.

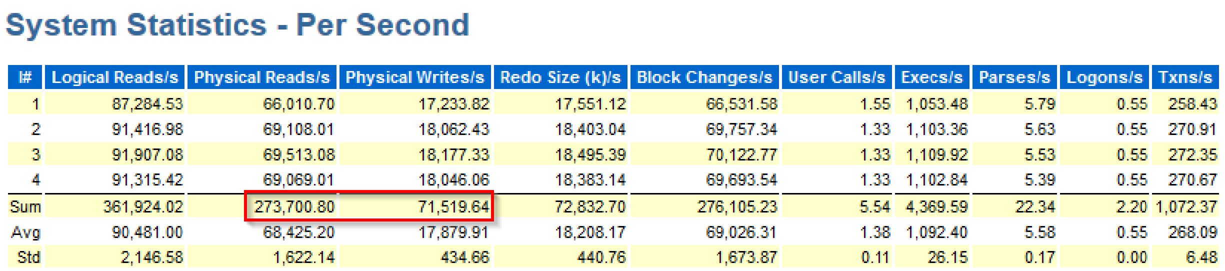
Keep in mind, that with the large highly randomized data set used in this test, 89% to 98% of these reads were coming from SSD, not from cache.

During these tests, host CPU utilization was up to 30% lower with NVMe/FC than FCP. During the 75/25% SELECT/UPDATE workload, up to 28% higher IOPS were achieved using the NVMe/FC protocol while at the same CPU utilization as the FCP protocol.

Customers should investigate how they can move their workloads to NVMe/FC to increase IOPS and decrease latency. These improvements will enhance user experience, increase customer retention, and improve employee productivity.

Appendix A: AWR Report

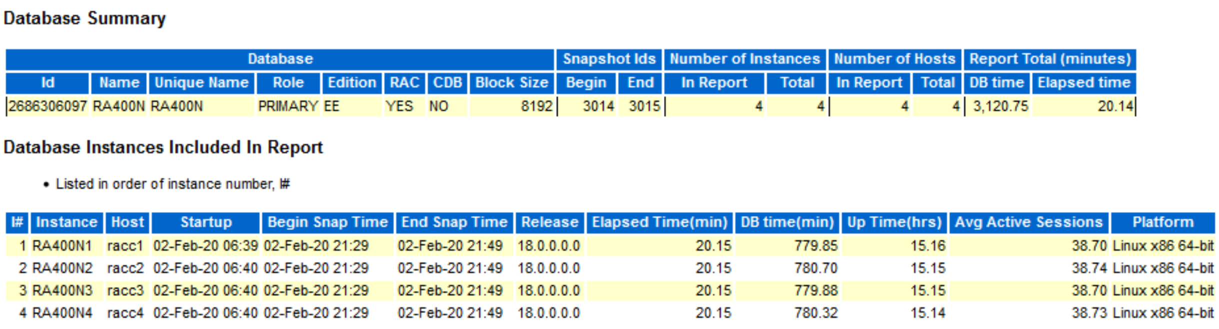
The following three screenshots show the AWR report that we collected at the 390K IOPS point of the NetApp AFF A400 performance test.



System Statistics - Per Second

I#	Logical Reads/s	Physical Reads/s	Physical Writes/s	Redo Size (k)/s	Block Changes/s	User Calls/s	Execs/s	Parses/s	Logons/s	Txns/s
1	87,284.53	66,010.70	17,233.82	17,551.12	66,531.58	1.55	1,053.48	5.79	0.55	258.43
2	91,416.98	69,108.01	18,062.43	18,403.04	69,757.34	1.33	1,103.36	5.63	0.55	270.91
3	91,907.08	69,513.08	18,177.33	18,495.39	70,122.77	1.33	1,109.92	5.53	0.55	272.35
4	91,315.42	69,069.01	18,046.06	18,383.14	69,693.54	1.33	1,102.84	5.39	0.55	270.67
Sum	361,924.02	273,700.80	71,519.64	72,832.70	276,105.23	5.54	4,369.59	22.34	2.20	1,072.37
Avg	90,481.00	68,425.20	17,879.91	18,208.17	69,026.31	1.38	1,092.40	5.58	0.55	268.09
Std	2,146.58	1,622.14	434.66	440.76	1,673.87	0.11	26.15	0.17	0.00	6.48

Total IOPS going to the AFF system are reported by physical reads per second and by physical writes per second in the Oracle AWR report.



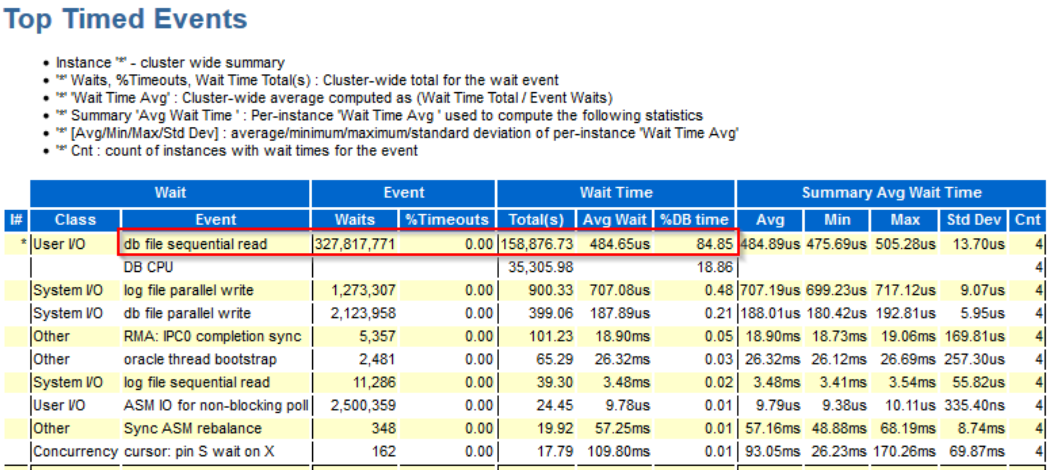
Database Summary

Database								Snapshot Ids		Number of Instances		Number of Hosts		Report Total (minutes)	
Id	Name	Unique Name	Role	Edition	RAC	CDB	Block Size	Begin	End	In Report	Total	In Report	Total	DB time	Elapsed time
2686306097	RA400N	RA400N	PRIMARY	EE	YES	NO	8192	3014	3015	4	4	4	4	3,120.75	20.14

Database Instances Included In Report

- Listed in order of instance number, I#

I#	Instance	Host	Startup	Begin Snap Time	End Snap Time	Release	Elapsed Time(min)	DB time(min)	Up Time(hrs)	Avg Active Sessions	Platform
1	RA400N1	racc1	02-Feb-20 06:39	02-Feb-20 21:29	02-Feb-20 21:49	18.0.0.0.0	20.15	779.85	15.16	38.70	Linux x86 64-bit
2	RA400N2	racc2	02-Feb-20 06:40	02-Feb-20 21:29	02-Feb-20 21:49	18.0.0.0.0	20.15	780.70	15.15	38.74	Linux x86 64-bit
3	RA400N3	racc3	02-Feb-20 06:40	02-Feb-20 21:29	02-Feb-20 21:49	18.0.0.0.0	20.15	779.88	15.15	38.70	Linux x86 64-bit
4	RA400N4	racc4	02-Feb-20 06:40	02-Feb-20 21:29	02-Feb-20 21:49	18.0.0.0.0	20.15	780.32	15.14	38.73	Linux x86 64-bit



Top Timed Events

- Instance '*' - cluster wide summary
- '*' Waits, %Timeouts, Wait Time Total(s) : Cluster-wide total for the wait event
- '*' 'Wait Time Avg' : Cluster-wide average computed as (Wait Time Total / Event Waits)
- '*' Summary 'Avg Wait Time ' : Per-instance 'Wait Time Avg ' used to compute the following statistics
- '*' [Avg/Min/Max/Std Dev] : average/minimum/maximum/standard deviation of per-instance 'Wait Time Avg'
- '*' Cnt : count of instances with wait times for the event

	Wait		Event		Wait Time			Summary Avg Wait Time				
I#	Class	Event	Waits	%Timeouts	Total(s)	Avg Wait	%DB time	Avg	Min	Max	Std Dev	Cnt
*	User I/O	db file sequential read	327,817,771	0.00	158,876.73	484.65us	84.85	484.89us	475.69us	505.28us	13.70us	4
		DB CPU			35,305.98		18.86					4
	System I/O	log file parallel write	1,273,307	0.00	900.33	707.08us	0.48	707.19us	699.23us	717.12us	9.07us	4
	System I/O	db file parallel write	2,123,958	0.00	399.06	187.89us	0.21	188.01us	180.42us	192.81us	5.95us	4
	Other	RMA: IPC0 completion sync	5,357	0.00	101.23	18.90ms	0.05	18.90ms	18.73ms	19.06ms	169.81us	4
	Other	oracle thread bootstrap	2,481	0.00	65.29	26.32ms	0.03	26.32ms	26.12ms	26.69ms	257.30us	4
	System I/O	log file sequential read	11,286	0.00	39.30	3.48ms	0.02	3.48ms	3.41ms	3.54ms	55.82us	4
	User I/O	ASM IO for non-blocking poll	2,500,359	0.00	24.45	9.78us	0.01	9.79us	9.38us	10.11us	335.40ns	4
	Other	Sync ASM rebalance	348	0.00	19.92	57.25ms	0.01	57.16ms	48.88ms	68.19ms	8.74ms	4
	Concurrency	cursor: pin S wait on X	162	0.00	17.79	109.80ms	0.01	93.05ms	26.23ms	170.26ms	69.87ms	4

Where to Find Additional Information

To learn more about the information that is described in this document, review the following documents and/or websites:

- The Silly Little Oracle Benchmark v2.5.2.4 (SLOB2)
  http://kevinclosson.net/2012/02/06/introducing-slob-the-silly-little-oracle-benchmark/
  https://github.com/therealkevinc/SLOB_distribution/tree/2.5.2.4
- NetApp AFF A-Series All Flash Array product webpage
  http://www.netapp.com/us/products/storage-systems/all-flash-array/aff-a-series.aspx

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 © 2020 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 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

Trademark Information

NETAPP, the NETAPP logo, and the marks listed at http://www.netapp.com/TM are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.

TR-4819-DESIGN