Technical Report

Performance Characterization of NetApp Cloud Volumes ONTAP for Google Cloud

Oded Berman, NetApp Data Fabric Group
February 2020 | TR-4816

Abstract

This technical report examines the performance of NetApp® Cloud Volumes ONTAP® for Google Cloud for application workloads. NetApp partners, customers, and employees should use the presented information to make informed decisions about which workloads are appropriate for Cloud Volumes ONTAP.

NetApp®

TABLE OF CONTENTS

1 Introduction 3

2 Executive Summary 3

3 Measurement of Storage Performance 3

3.1 Test Methodology 3

3.2 Cloud Volumes ONTAP Configuration 4

3.3 Workloads 5

3.4 Test Environment Diagram 5

4 Performance Assessment Results 6

4.1 Transactional/OLTP 6

4.2 Streaming Reads 7

4.3 Streaming Writes 8

4.4 Mixed Reads/Writes 9

Acknowledgements 10

Where to Find Additional Information 10

Version History 11

LIST OF TABLES

Table 1) Google Compute Engine machine types, performance limits, and selected storage configuration. 4

Table 2) Transactional/OLTP result details. 7

Table 3) Streaming reads result details. 8

Table 4) Streaming writes result details. 9

Table 5) Mixed reads/writes result details. 10

LIST OF FIGURES

Figure 1) Test environment diagram. 6

Figure 2) Transactional/OLTP workload IOPS and latency. 7

Figure 3) Streaming reads throughput. 8

Figure 4) Streaming writes throughput. 9

Figure 5) Mixed reads/writes workload throughput and latency. 10

© 2020 NetApp, Inc. All Rights Reserved.

1 Introduction

NetApp has a long history of providing leading-edge, high-performance, feature-rich data management technologies and solutions that provide the agility and mobility that organizations need in the digital transformation and cloud era. The NetApp® Cloud Volumes ONTAP® software-only storage solution is part of the ONTAP family of products that delivers an organization's data anywhere that it is needed across the hybrid multicloud.

NetApp Cloud Volumes ONTAP is an enterprise-grade data management solution that is deployed on a public cloud infrastructure. It consumes native cloud compute services and manages native cloud block and object storage. Today, Cloud Volumes ONTAP can be deployed in any of the "big three" cloud providers: Google Cloud, Amazon Web Services (AWS), and Azure. Through its unified storage access capabilities, storage efficiency features, and enhanced data protection and security services, Cloud Volumes ONTAP delivers a powerful solution for any enterprise workload. It is optimal for workloads such as databases, DevOps, virtual desktop infrastructure (VDI), Kubernetes, file services, disaster recovery, backup and archiving, and more.

This technical report describes the results of performance assessments that were conducted with Cloud Volumes ONTAP for Google Cloud. When deployed in Google Cloud, Cloud Volumes ONTAP uses Google Compute Engine, a scalable high-performance virtual machine (VM), and Google Persistent Disk, which is durable and high-performance block storage. It also uses Google Cloud Storage, which provides unified, price/performance-optimized object storage.

Cloud Volumes ONTAP is available in two pricing models: pay as you go (PayGo) and bring your own license (BYOL). For the PayGo model, three types of license packages are available: Explore, Standard, and Premium. Each license package provides different compute and storage capacity options.

2 Executive Summary

NetApp performed this study to showcase the storage performance and benefits of NetApp Cloud Volumes ONTAP. Understanding the performance characteristics of Cloud Volumes ONTAP is critical to set expectations and to help stakeholders make educated decisions about which of their workloads Cloud Volumes ONTAP can handle and what configuration is required.

The results are presented in section 4, Performance Assessment Results. The results provide some basic understanding of the Cloud Volumes ONTAP performance capabilities for Google Cloud and its dependency on the underlying infrastructure and the services that are used.

The chosen workloads for this study were transactional/OLTP, streaming reads, streaming writes, and mixed reads/writes. Each of the selected workloads is characterized by different mixture and access patterns of I/O operations.

3 Measurement of Storage Performance

NetApp performed these assessments to measure the performance of single-node NetApp Cloud Volumes ONTAP 9.6 in different application workloads. This section describes the methodology that NetApp used to assess Cloud Volumes ONTAP performance and the configuration that was used, as well as the workloads that were tested.

3.1 Test Methodology

For this study, NetApp used Flexible I/O Tester (Fio), a tool that can simulate any given workload. Fio generates a number of threads or processes, carrying out the type of I/O combination and patterns that the user specifies to generate the desired workloads. For more information about Fio, see the Fio page on GitHub.

© 2020 NetApp, Inc. All Rights Reserved.

To get the most out of Cloud Volumes ONTAP, NetApp used multiple Fio clients in the various assessments. The I/O workloads were generated by multiple `n1-standard-8` compute engines (each with eight virtual CPUs [vCPUs] and 30GB of memory) running Red Hat Enterprise Linux 7 (for a total of five).

All tests were performed with NFS, which is commonly used in a wide variety of workloads. NFS is a distributed file system, based on a client-server architecture, that enables users to access files and directories on a remote file system, much like the process to access local storage. The NFS protocol provides file-level access to logical storage resources, such as NetApp FlexVol® volumes, on Cloud Volumes ONTAP. A FlexVol volume is an allocation of storage space that serves as a data container. A FlexVol volume is the basic storage unit that can be mounted by one or more NFS clients, providing access to users within the limits of the granted permissions.

In this study, a single FlexVol volume was configured and accessed by seven clients. The goal was to achieve the best performance metrics possible for IOPS, throughput (the amount of data payload), and latency (the response time). All the results that are presented later in this report are the aggregated sum from the clients' side.

All the Compute Engine instances that NetApp used in the study, including the instances that Cloud Volumes ONTAP used, were in the same availability zone, Virtual Private Cloud, and subnet.

3.2 Cloud Volumes ONTAP Configuration

For this study, Cloud Volumes ONTAP was configured as a single-node system. The performance of Cloud Volumes ONTAP depends greatly on the underlying cloud services that are used (the Compute Engine machine type and the type and size of the Persistent Disk). Therefore, NetApp tested Cloud Volumes ONTAP by using different Compute Engine machine types and Persistent Disk configurations. The Compute Engine machine types not only differ in their specifications, but they also incur different costs by Google Cloud and by NetApp and belong to different licensing packages.

NetApp used the following Compute Engine machine types:

- `custom-4-16384` (included in the Explore license; up to 2TB of storage)
- `n1-standard-8` (included in the Standard license; up to 10TB of storage)
- `n1-standard-32` (included in the Premium license; up to 368TB of storage)

All the VM instances were configured with Zonal SSD Persistent Disks. SSD Persistent Disks are backed by SSDs and provide predictable performance, having been designed for single-digit, millisecond (ms) latencies. Because they balance price/performance for a wide variety of transactional workloads, the use of SSD Persistent Disks is preferable. Cloud Volumes ONTAP supports the use of Zonal Standard Persistent Disks as well, which are backed by HDDs and provide an efficient and economical way to handle sequential read/write operations. For more information about Google Cloud Persistent Disks and limits, visit the Google Cloud storage options page.

Table 1 displays the Compute Engine machine types, performance limits, and storage configuration that NetApp used in this study.

Table 1) Google Compute Engine machine types, performance limits, and selected storage configuration.

Instance Type	vCPU	Memory (GiB)	Sustained Random IOPS		Sustained Throughput (MBps)		SSD Persistent Disk
			Read	Write	Read	Write	
`custom-4-16384`	4	16	15,000	15,000	240	240	2 x 1TB

© 2020 NetApp, Inc. All Rights Reserved.

Instance Type	vCPU	Memory (GiB)	Sustained Random IOPS		Sustained Throughput (MBps)		SSD Persistent Disk
			Read	Write	Read	Write	
n1-standard-8	8	30	15,000	15,000	800	400	2 x 4TB
n1-standard-32	32	120	60,000	30,000	1,200	400	4 x 8TB

Note: For more information about the Google Cloud performance limitations, visit the Google Cloud optimizing Persistent Disk performance page.

The Cloud Volumes ONTAP write speed, which affects write performance, was set to `Normal`. When the write speed is set to `Normal`, data is first written to Persistent Disk storage and then an acknowledgment is sent back to the client. This process eliminates the chance of data loss if an unplanned outage occurs.

When the write speed is set to `High`, data is buffered in memory and an acknowledgment is sent back to the client before the data is committed to Persistent Disk storage. A high write speed enhances write performance but also increases the potential for data loss if an unplanned outage occurs. Users should enable a high write speed only if enhanced write performance is required and if data loss can either be tolerated or be handled by the application. For example, data loss from an unplanned outage can be tolerated, or the application (such as Microsoft SQL Server for Always On availability groups or MySQL replication) can handle data loss.

3.3 Workloads

In this study, four different workloads were configured and tested. Each tested workload consisted of a unique I/O mixture and access pattern that simulated workloads that are used in widely deployed applications:

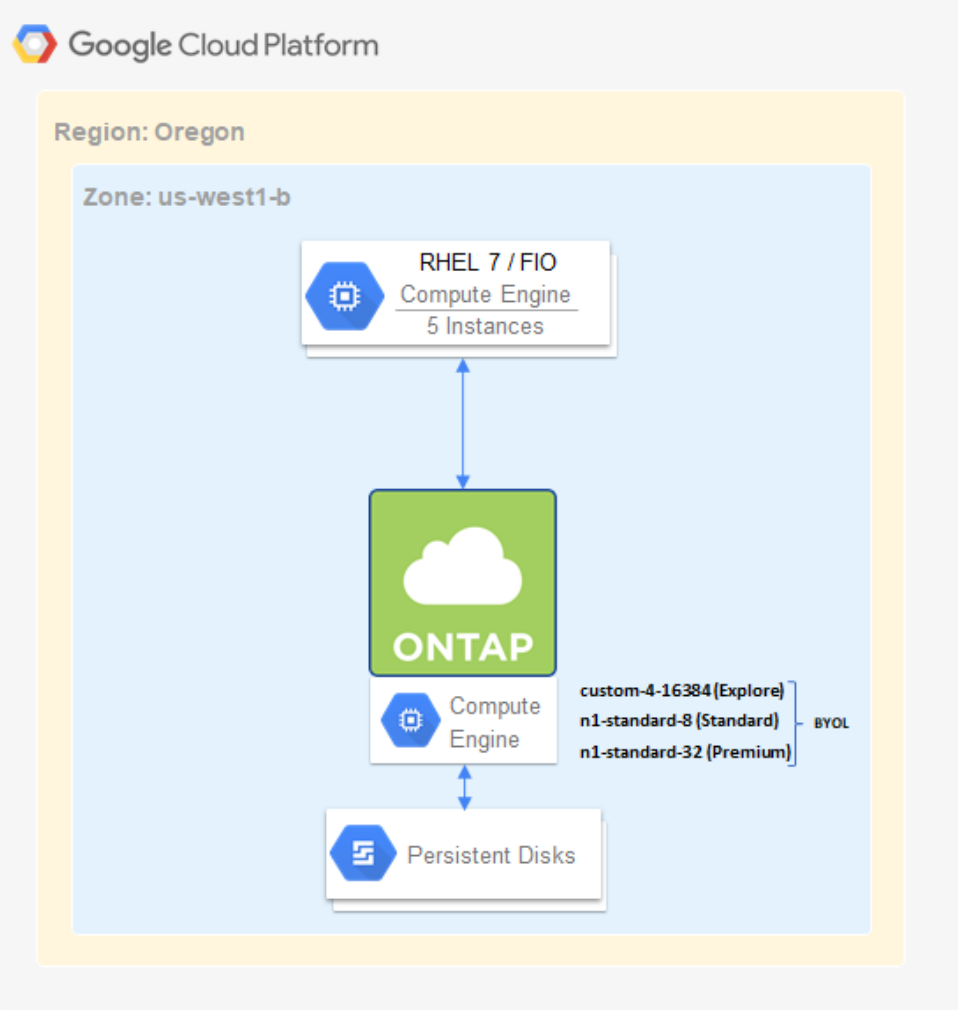
- **Transactional/OLTP** (8KB block size, 80% reads, and 100% random access I/O). A transactional workload is typically identified by a database that receives numerous requests for data and multiple changes to this data from several users over time. These modifications are called *transactions*. The transactions are small and primarily random in nature with high concurrency. This type of workload is generated by transactional database applications such as SAP, Oracle, and SQL Server and MySQL databases.
- **Streaming reads** (64KB block size, 100% reads, and 100% sequential access I/O). In a streaming reads workload, only read operations are performed (and they are concurrent, large, contiguous requests). This workload is typically generated by applications such as media servers (for example, video on demand) and virtual tape libraries.
- **Streaming writes** (64KB block size, 100% writes, and 100% sequential access I/O). In a streaming writes workload, only write operations are performed (and they are concurrent, large, contiguous requests). This workload is typically generated by applications such as media capture, virtual tape libraries, medical imaging, backup and archiving, and video surveillance.
- **Mixed reads/writes** (16KB block size, 50% reads, and 100% random access I/O). Many applications generate a mix of read/write operations that differ in the reads versus writes percentage, block size, and random versus sequential percentage and that typically require high throughput and low latency. A mix of 50% reads and 50% writes is a common starting point for measuring. That kind of workload can be found, for example, in a VDI.

3.4 Test Environment Diagram

Figure 1 displays the environment and Google Cloud resources that NetApp used for the performance assessments.

© 2020 NetApp, Inc. All Rights Reserved.

Figure 1) Test environment diagram.



4 Performance Assessment Results

This section describes the values of the performance metrics that were obtained in the process of assessing NetApp Cloud Volumes ONTAP.

4.1 Transactional/OLTP

The significant metrics for this type of workload are IOPS and latency, as shown in Figure 2 and as described in Table 2.

© 2020 NetApp, Inc. All Rights Reserved.

Figure 2) Transactional/OLTP workload IOPS and latency.

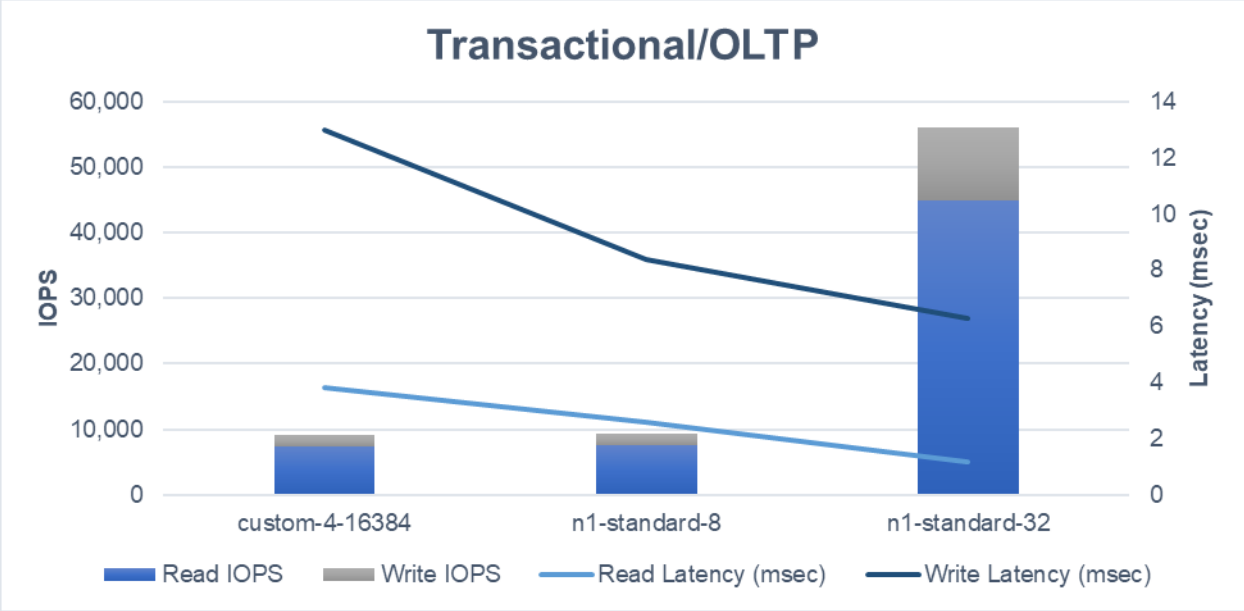


Table 2) Transactional/OLTP result details.

Machine Types	IOPS		Throughput (MBps)		Latency (ms)	
	Read	Write	Read	Write	Read	Write
`custom-4-16384`	7,300	1,800	61.6	15.7	3.8	13
`n1-standard-8`	7,500	1,800	61.8	15.6	2.6	8.4
`n1-standard-32`	45,000	11,000	369.3	92.4	1.15	6.3

4.2 Streaming Reads

The significant metric for this type of workload is throughput, as shown in Figure 3 and as described in Table 3.

© 2020 NetApp, Inc. All Rights Reserved.

Figure 3) Streaming reads throughput.

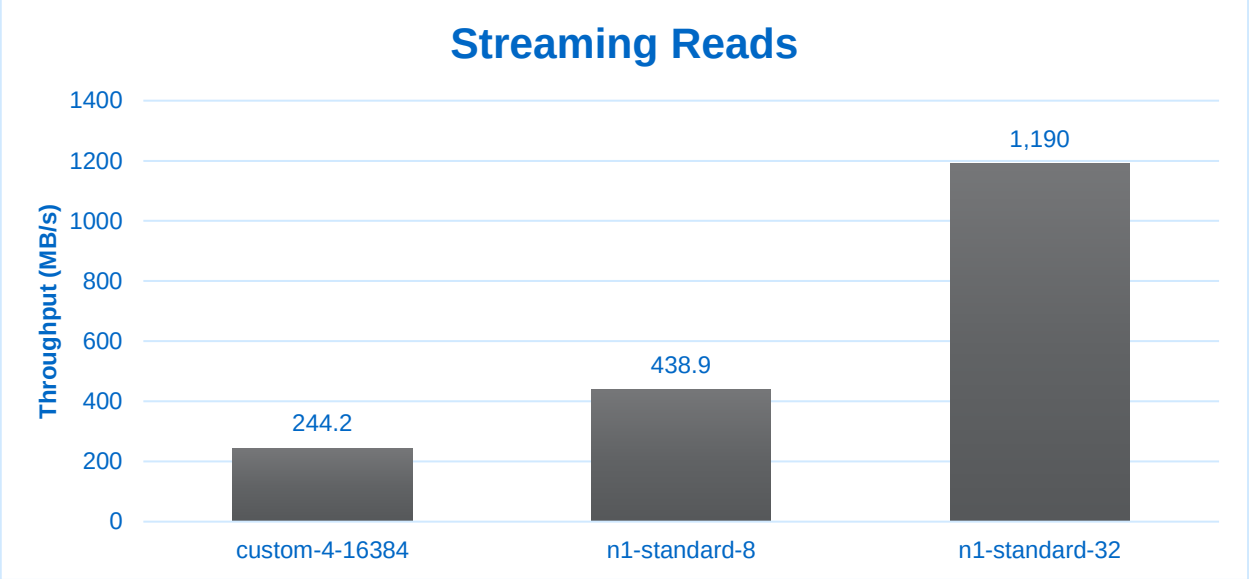


Table 3) Streaming reads result details.

Instance Type	IOPS	Throughput (MBps)	Latency (ms)
custom-4-16384	3,700	244.2	3.2
n1-standard-8	6,700	438.9	1.8
n1-standard-32	18,000	1,190	1.3

4.3 Streaming Writes

The significant metric for this type of workload is throughput, as shown in Figure 4 and as described in Table 4.

© 2020 NetApp, Inc. All Rights Reserved.

Figure 4) Streaming writes throughput.

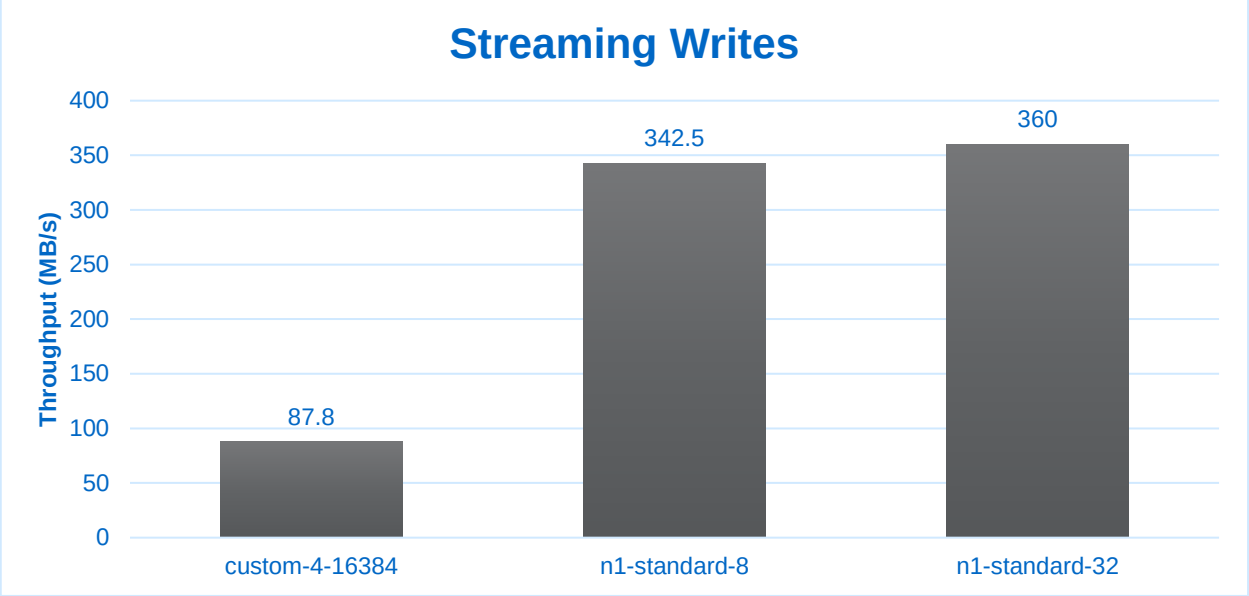


Table 4) Streaming writes result details.

Instance Type	IOPS	Throughput (MBps)	Latency (ms)
custom-4-16384	1,300	87.8	8.8
n1-standard-8	5,200	342.5	4.5
n1-standard-32	5,500	360	4.4

4.4 Mixed Reads/Writes

The significant metrics for this type of workload are throughput and latency, as shown in Figure 5 and as described in Table 5.

© 2020 NetApp, Inc. All Rights Reserved.

Figure 5) Mixed reads/writes workload throughput and latency.

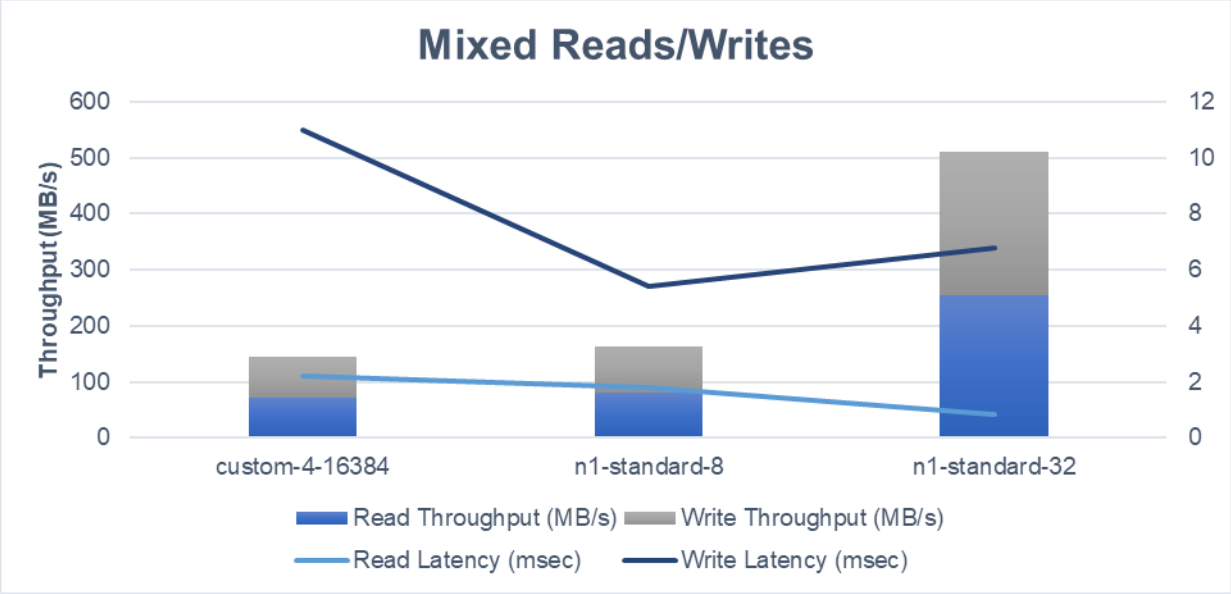


Table 5) Mixed reads/writes result details.

Instance Type	IOPS		Throughput (MBps)		Latency (ms)	
	Read	Write	Read	Write	Read	Write
`custom-4-16384`	4,400	4,400	72	72	2.2	11
`n1-standard-8`	5,000	5,000	82	82	1.8	5.4
`n1-standard-32`	15,000	15,000	255	255	0.85	6.8

Acknowledgements

The author gratefully acknowledges the contributions that were made to this technical report by key NetApp team members: Ishai Avramovich, Liron Bar-Mashiah, Aviv Degani, Yaron Haimsohn, Michele Pardini and Oren Inbar. My sincere appreciation and thanks go to all these individuals, who designed and performed the assessments and provided the insight and expertise that greatly assisted in the creation of this paper.

Where to Find Additional Information

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

- Cloud Volumes ONTAP (features, architecture, TCO calculator, pricing, and more)
  https://cloud.netapp.com/ontap-cloud
- NetApp Cloud Manager (formerly OnCommand® Cloud Manager) and Cloud Volumes ONTAP documentation
  https://docs.netapp.com/us-en/occm/

© 2020 NetApp, Inc. All Rights Reserved.

- Getting Started with Cloud Volumes ONTAP for Google Cloud
  https://docs.netapp.com/us-en/occm/task_getting_started_gcp.html
- Fio page on GitHub
  https://github.com/axboe/fio
- Google Cloud storage options
  https://cloud.google.com/compute/docs/disks/performance
- Block storage performance page in Google Cloud
  https://cloud.google.com/compute/docs/disks/performance#ssd-pd-performance
- Cloud Manager for Cloud Volumes ONTAP in Google Cloud Platform Marketplace
  https://console.cloud.google.com/marketplace/details/netapp-cloudmanager/cloud-manager

Version History

Version	Date	Document Version History
Version 1.0	February 2020	Initial release.

© 2020 NetApp, Inc. All Rights Reserved.

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.

Data contained herein pertains to a commercial item (as defined in FAR 2.101) and is proprietary to NetApp, Inc. The U.S. Government has a non-exclusive, non-transferrable, non-sublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b).

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-4816-0220