TECHNOLOGY GUIDE

Intel Corporation

MinIO Object Storage - Performance Benchmark Report on Multi-node Deployment

Authors

Tim Miskell

Seungweon Park

Tushar Gohad

Tom Counihan

Intel Corporation

Daniel Valdivia

MinIO

1 Introduction

MinIO Object Storage is a high performance S3 compatible object storage open-source software that is ready to be deployed in containerized environments and to support native Kubernetes (K8s) and Red Hat OpenShift Container Platform (RHOCP). The MinIO Operator simplifies the management and configuration of deployments in orchestrated, containerized environments, such as K8s. This document discusses the integration of MinIO Object Storage with Intel Network and Edge Cloud Reference Architecture to enable a cloud native, object-storage ready infrastructure in a containers bare metal deployment. MinIO supports a Simple Storage Service (S3) interface and therefore, is competitive to other cloud storage solutions, such as Amazon Web Services (AWS) along with Ceph. For example, compared to file storage, object storage does not organize data in a hierarchical manner in terms of directories and files, but rather as customizable metadata along with segmented chunks of the data itself. Since it is not necessarily to maintain a file system, object storage can readily scale by simply adding more nodes to a storage cluster, with each node in the cluster capable of storing chunks of data representative of the original object. As a result, MinIO as an object store also readily scales by adding more MinIO nodes to the cluster. Furthermore, MinIO object store supports other features such as Erasure Coding (EC), which helps to ensure data availability by distributing individual segments of objects across multiple nodes, a load balancer to help ensure even distribution across MinIO tenant pods, along with Bit rot protection.

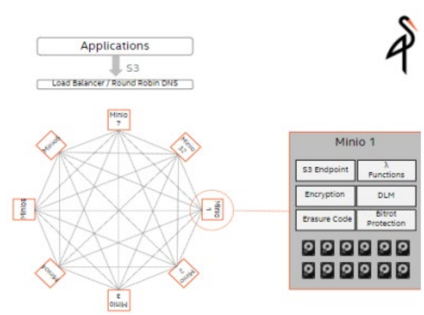

Figure 1. MinIO Object Store

This document describes the benchmarking results for a multi-node deployment of a MinIO Strage Object Cluster. The benchmark environment focuses on two generations of Intel® CPU hardware, specifically the 2nd and 3rd Gen Intel® Xeon® Scalable processors, with each environment leveraging the same storage technology, namely Intel® P5510 NVMe drives. This document provides details regarding, how to set up the benchmark environment, how to run the benchmark tests, and an analysis of the observed results.

The target audience for this document is customers who plan to deploy a storage cluster and are interested in the performance improvements of 3rd Gen Intel Xeon Scalable processors over 2nd Gen Intel Xeon Scalable processors, not only in terms of increased core count, but also by including support for PCI Express (PCIe) Gen 4, which theoretically provides up to 256 MT/s of throughput for a single x16 PCIe slot over Gen 3.

This document is part of the [Network Transformation Experience Kit](#).

Table of Contents

Figures

Tables

Document Revision History

REVISION	DATE	DESCRIPTION
001	October 2022	Initial Release.

1.1 Terminology

Table 1. Terminology

ABBREVIATION	DESCRIPTION
AWS	Amazon Web Services
BMRA	Bare Metal Reference Architecture
CRB	Customer Reference Board
DHCP	Dynamic Host Configuration Protocol
EC	Erasure Coding
HA	High Availability
IPAM	IP Address Manager
K8s	Kubernetes
KPI	Key Performance Indicator
NIC	Network Interface Card
NVMe	Non-Volatile Memory express
Object Storage	Object storage, in contrast to file storage, stores data as metadata and segmented chunks of the data itself.
PCIe	Peripheral Component Interface Express
PF	Physical Function
PV	Persistent Volume
RHOCP	Red Hat OpenShift Container Platform
S3	Simple Storage Service. Amazon's AWS defined interface for object storage.
SDN	Software Defined Networking
SR-IOV	Single Root I/O Virtualization
SUT	System Under Test
VF	Virtual Function

1.2 Reference Documentation

Table 2. Reference Documents

REFERENCE	SOURCE
BMRA Release - GitHub	https://github.com/intel/container-experience-kits
MinIO Object Storage Solution	https://min.io/
MinIO Speedtest	https://blog.min.io/introducing-speedtest-for-minio/
Network and Cloud Edge Container Bare Metal Reference System Architecture User Guide	https://networkbuilders.intel.com/solutionslibrary/network-and-cloud-edge-container-bare-metal-reference-system-architecture-user-guide

2 Benchmark Environment

2.1 Hardware

The multi-node MinIO storage cluster deployment is automated through the aid of the Bare Metal Reference Architecture (BMRA) from RA 22.05 onwards. The BMRA deployment in this case includes an ansible host, multiple HA controller nodes, along with multiple worker or storage nodes. The ansible host is a server from which a set of ansible playbooks and helm charts are run in order to automate the deployment of a fully functional K8s cluster. The controller nodes are arranged in an HA configuration, and each controller node runs the appropriate set of containerized services, such as etcd, the SR-IOV network operator, and the MinIO operator front end, in order to manage the overall K8s cluster. The worker nodes in this case double as storage nodes and run the MinIO tenant pods. Observing the principles of Software Defined Networking (SDN), the cluster includes two separate networks, specifically with one network devoted to control plane traffic and one network devoted to data, that is, storage, plane traffic.

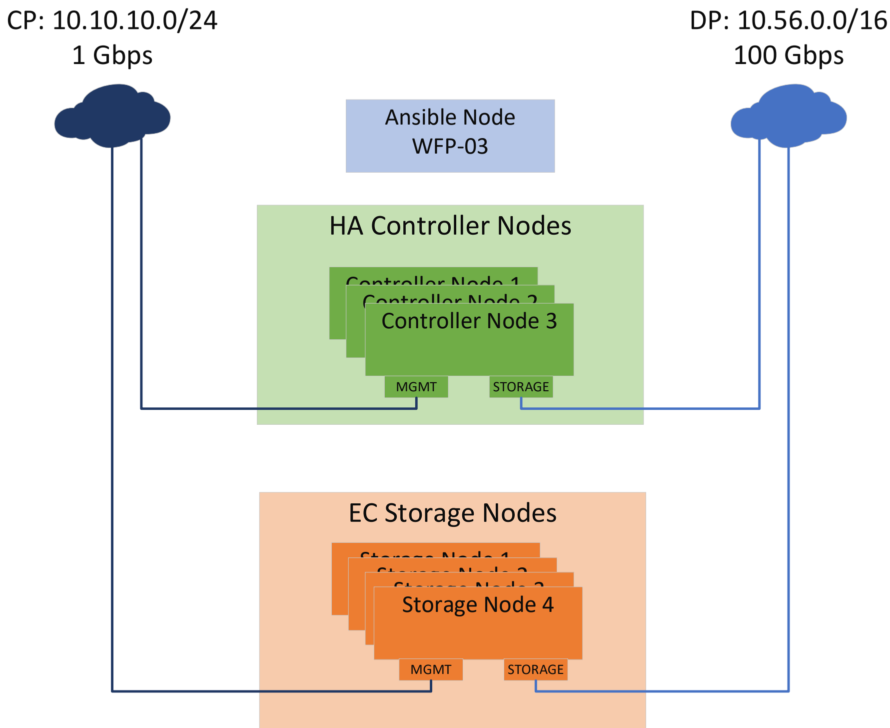

Figure 2. BMRA with Multi-Node MinIO Storage Cluster

For the purpose of comparison, we benchmark the multi-node MinIO storage cluster using two different configurations, specifically one cluster deployed on a set of servers populated with 2nd Gen Intel Xeon Scalable processors, and another cluster deployed on a set of servers populated with 3rd Gen Intel Xeon Scalable processors. Figure 3 presents the server configuration with 2nd Gen Intel Xeon Scalable processors. The table below presents the hardware BOM for the SUT populated with 2nd Gen Intel Xeon Scalable processors.

Table 3. 2nd Gen Intel® Xeon® Scalable Processor Platform Hardware BOM

Component	Specification
Platform	1x Intel® S2600WFS
Ethernet Adapter	2x Intel® Ethernet Network Adapter E810-2CQDA2
Data Storage	8x Intel® NVMe P5510 4.0 TiB Storage Capacity
PCIe	Gen 3 Support up to 128 MT/s per x16 PCIe Slot
CPU	2x 5218N @2.3 GHz 16C32T Intel® Xeon-SP
DRAM	12x 16 GiB (192 GiB) DDR3 @ 2666 MT/s
Microcode	0x4003302
Test by Intel as of	9/28/2022

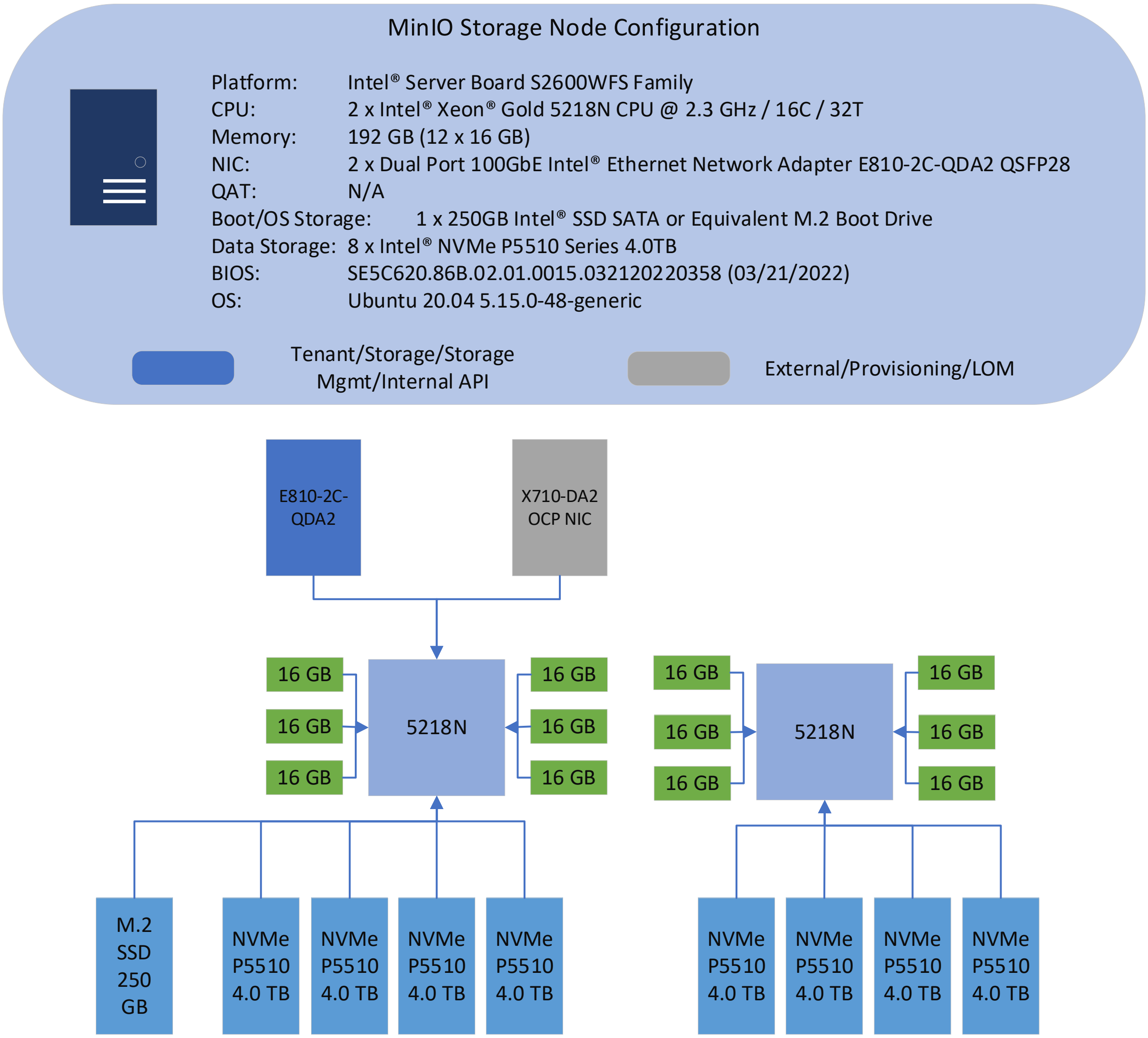

Figure 3. Server configuration with 2nd Gen Intel® Xeon® Scalable Processors

Figure 4 presents the server configuration with 3rd Gen Intel Xeon Scalable processors. Table 4 details the hardware BOM for the platform populated with 3rd Gen Intel® Xeon® Scalable Processors.

Table 4. 3rd Gen Intel® Xeon® Scalable Processor Platform Hardware BOM

Component	Specification
Platform	1x Intel® M50CYP
Ethernet Adapter	1x Intel® Ethernet Network Adapter E810-2CQDA2
Data Storage	8x Intel® NVMe P5510 4.0 TiB Storage Capacity
PCIe	Gen 4 Support up to 256 MT/s per x16 PCIe Slot
CPU	2x 6338N @2.2 GHz 32C64T Intel® Xeon-SP
DRAM	16x 16 GiB (256 GiB) DDR4 @ 2666 MT/s
Microcode	0xd000375
Test by Intel as of	9/28/2022

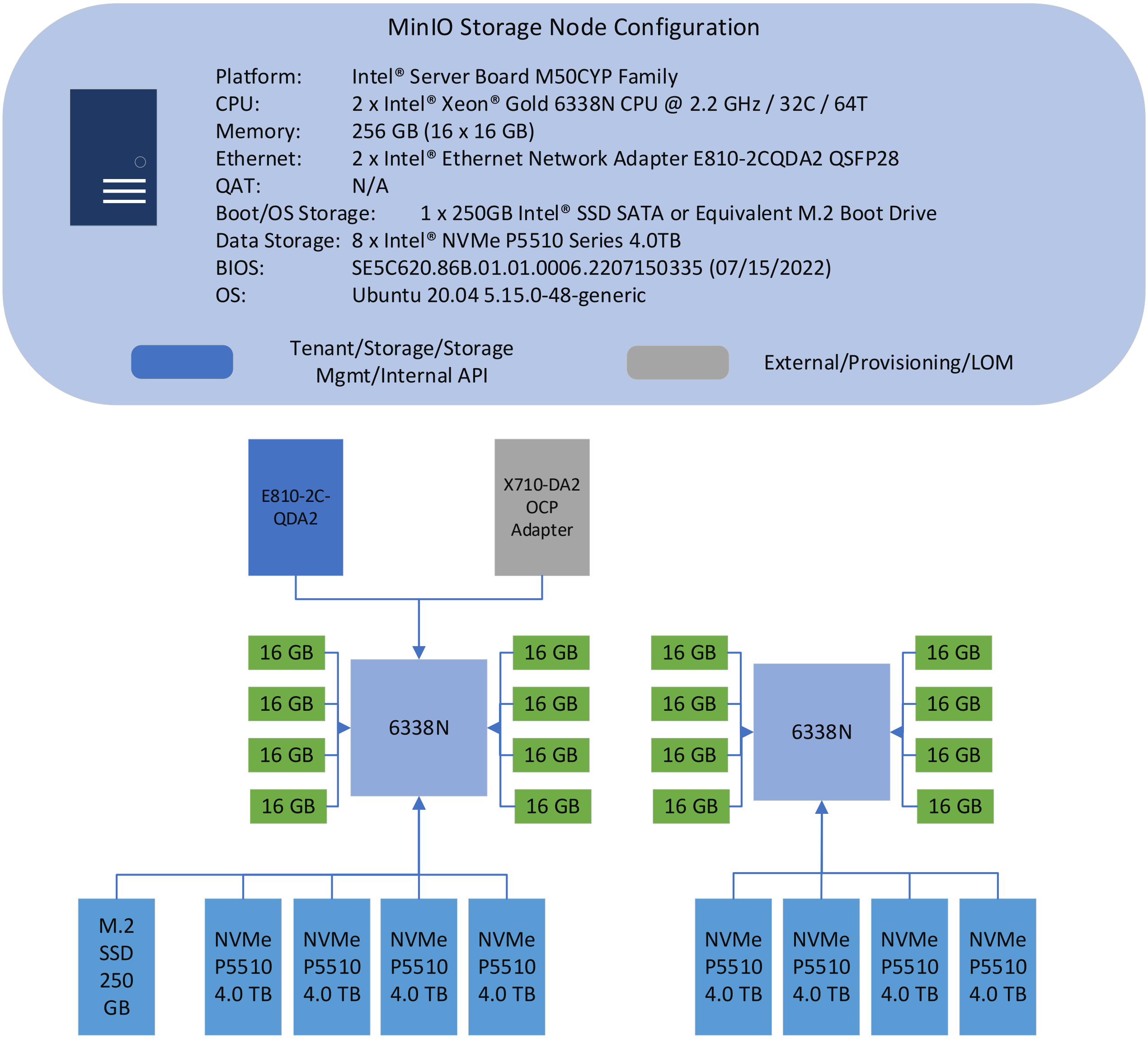

Figure 4. Server configuration with 3rd Gen Intel® Xeon® Scalable Processors

As shown in Figure 5, in terms of collecting Key Performance Indicators (KPIs) from the System Under Test (SUT), we leverage the MinIO Speedtest benchmarking tools. In this case, a separate client server connects to the MinIO tenant pods directly over the storage network. The Speedtest Client benchmarks the cluster using a variety of object sizes from 1KiB up to 64 MiB. For each benchmark, we vary the number of servers assigned to the MinIO cluster from one server up to four servers.

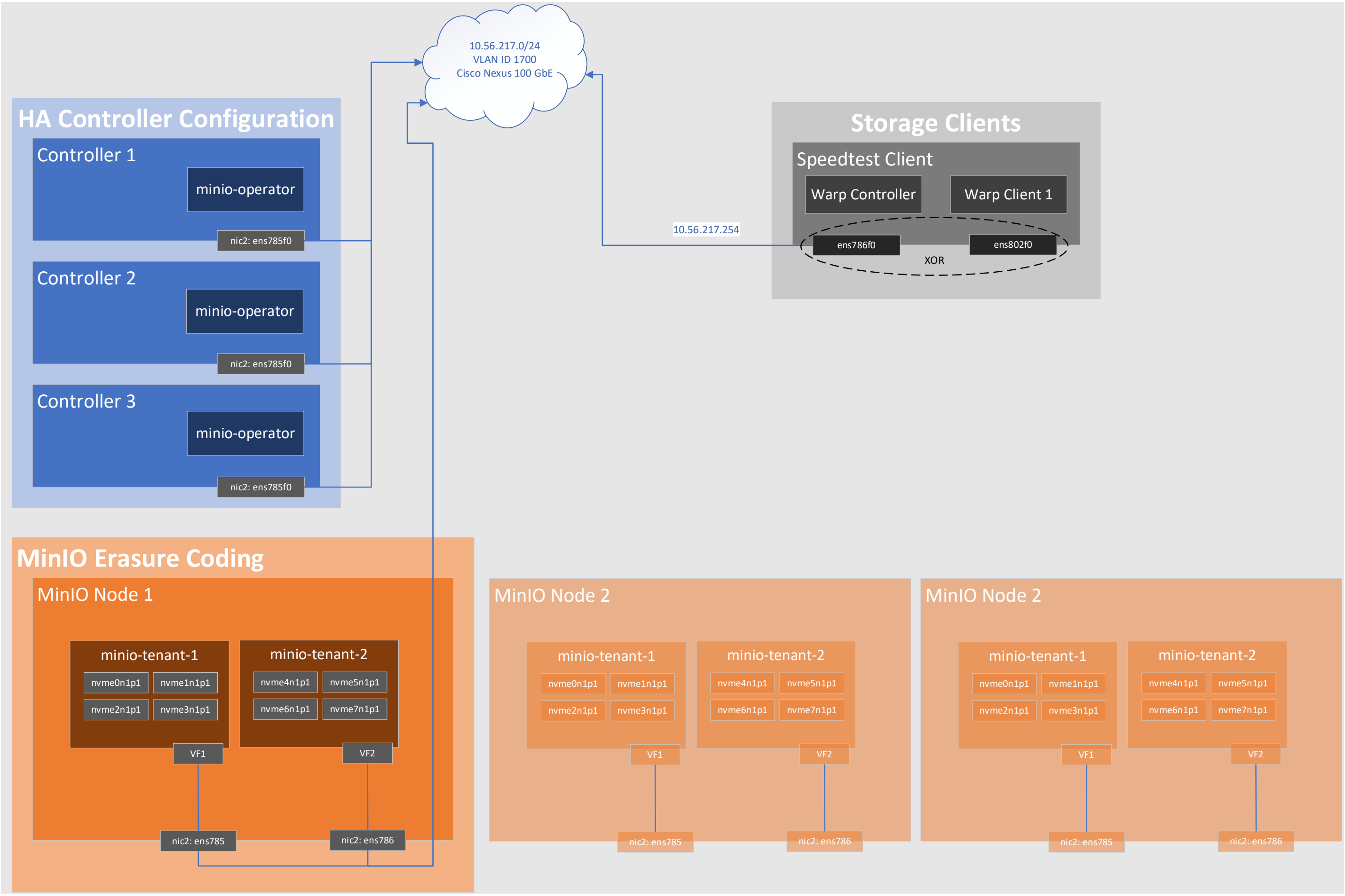

Figure 5. BMRA Deployment with MinIO Storage and MinIO Speedtest Client Server

2.2 Software

The table below details the specific software configuration for the BMRA cluster.

Table 5. Software Configuration for BMRA Cluster

Component	Version
Operating System	Ubuntu 20.04.4 LTS (Focal Fossa)
Kernel	5.4.0-125-generic
BMRA Release	22.05
MinIO Release	DEVELOPMENT RELEASE.2022-09-09T21-05-00Z
Ice Kernel Driver	1.7.16
Iavf Kernel Driver	4.1.1
E810-2C-QDA2 Firmware	3.00
nvme	1.0

With reference to Figure 6, at a high level, the deployment process for the MinIO storage cluster can essentially be grouped into the following four steps:

1. The MinIO operator pods are deployed onto each of the three controller nodes.
2. For each of the Intel® P5510 NVMe drives in the cluster:
    a. The drive is low-level formatted with a sector size of 4096 B.
    b. A single partition is created.
    c. The drive is XFS formatted with a sector size of 4096 B to match the NVMe drive low-level format.
    d. The drive is added to the pool of PVs managed by K8s.

3. For each Intel® Ethernet 800 Series Network Adapter, namely Intel® Ethernet Network Adapter E810-2CQDA2, in the cluster:
   a. A Virtual Function (VF) is created from each Physical Function (PF), with each VF supporting up to 100 GbE of network bandwidth.
   b. The VF is added to the pool of VFs managed by the SR-IOV Network Operator and in turn by K8s.
   c. The MTU of each VF is increased from 1500 B to 4096 B in order to match the low level formatting of the NVMe drives.
   d. An IP address is assigned to each VF on the DHCP storage plane network through the use of IPAM.
4. The MinIO tenant pods are deployed onto the storage nodes. Each MinIO tenant pod is assigned a pair of VFs from the storage plane network, and thus each pod is capable of sustaining up to 200 GbE of network throughput. Furthermore, each MinIO tenant pod is assigned eight PVs.

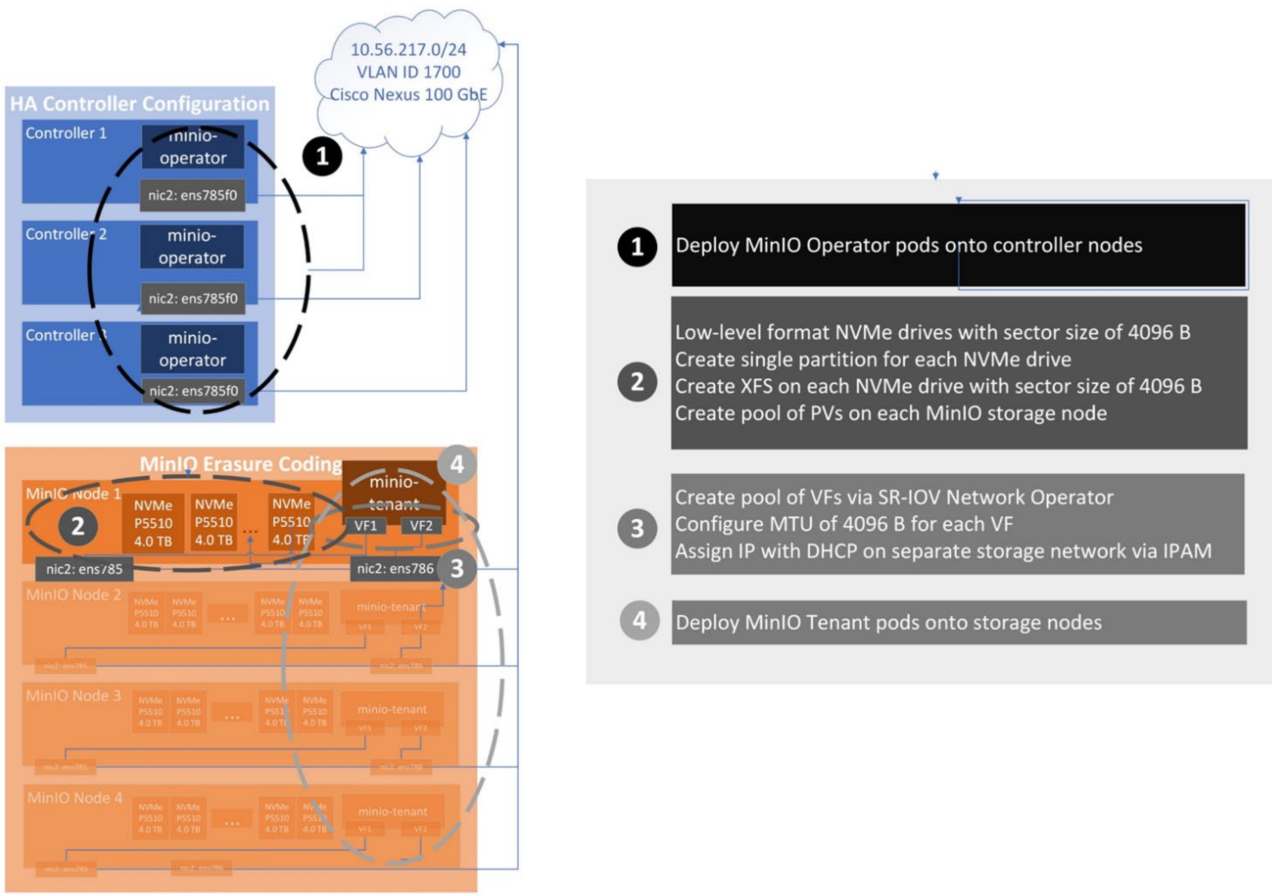

Figure 6. MinIO Deployment Process

2.3 Test Harness

The test harness in this case is the Speedtest tool provided by MinIO. The Speedtest tool is included with the MinIO Client tool, specifically mc. The MinIO Client provides a set of administrative commands that can be used to monitor and query the status of the MinIO cluster. The Speedtest tool runs a set of GET and PUT operations and presents the throughput results both on a per server and per cluster basis. The KPIs collected include the following:

1. GET throughput both on a per server basis as well as the aggregate throughput for the cluster.
2. PUT throughput both on a per server basis as well as the aggregate throughput for the cluster.
3. CPU utilization on a per server basis.
4. Memory utilization on a per server basis.

To investigate the relation between throughput and object size, we benchmark a range of object sizes including:

- 1 KiB
- 64 KiB
- 128 KiB
- 256 KiB
- 1 MiB
- 4 MiB

- 8 MiB
- 16 MiB
- 64 MiB

Furthermore, we benchmark using a variety of server allocation combinations to the MinIO cluster, including:

- Single node deployment
- 2 node deployments
- 3 node deployments
- 4 node deployments

3 Benchmark Results

The following diagrams present the comparison results of running MinIO Speedtest against an Intel® CRB platform, that is, S2600WFS, populated with 2nd Gen Intel Xeon Scalable processors against an M50CYP platform populated with 3rd Gen Intel Xeon Scalable processors for a single node MinIO deployment for a range of object sizes[1].

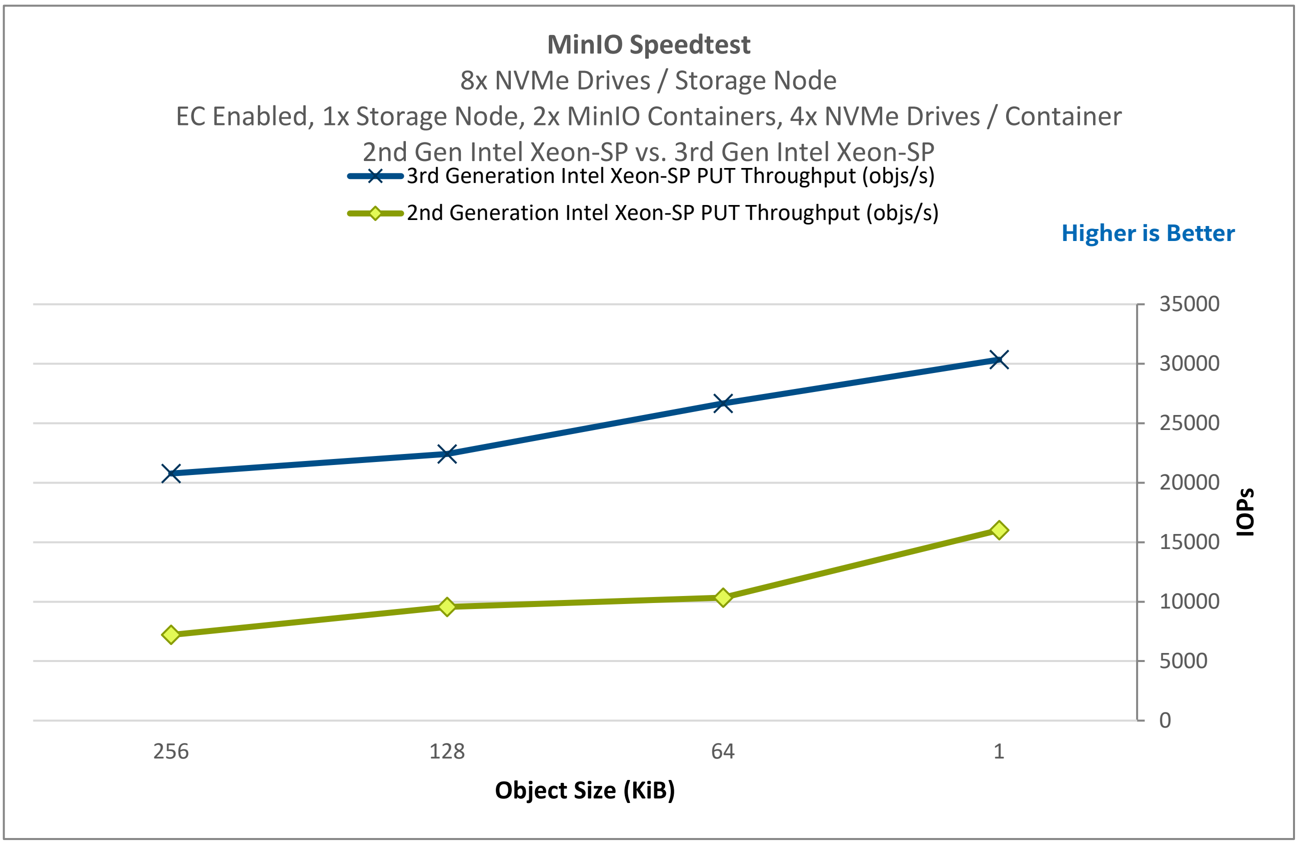

Figure 7. MinIO Speedtest, Aggregate PUT Throughput Comparison for up to 256 KiB object sizes, 2nd Gen Intel® Xeon® Scalable Processors versus 3rd Gen Intel® Xeon® Scalable Processors, Single Storage Node

[1] See backup for workloads and configurations. Results may vary.

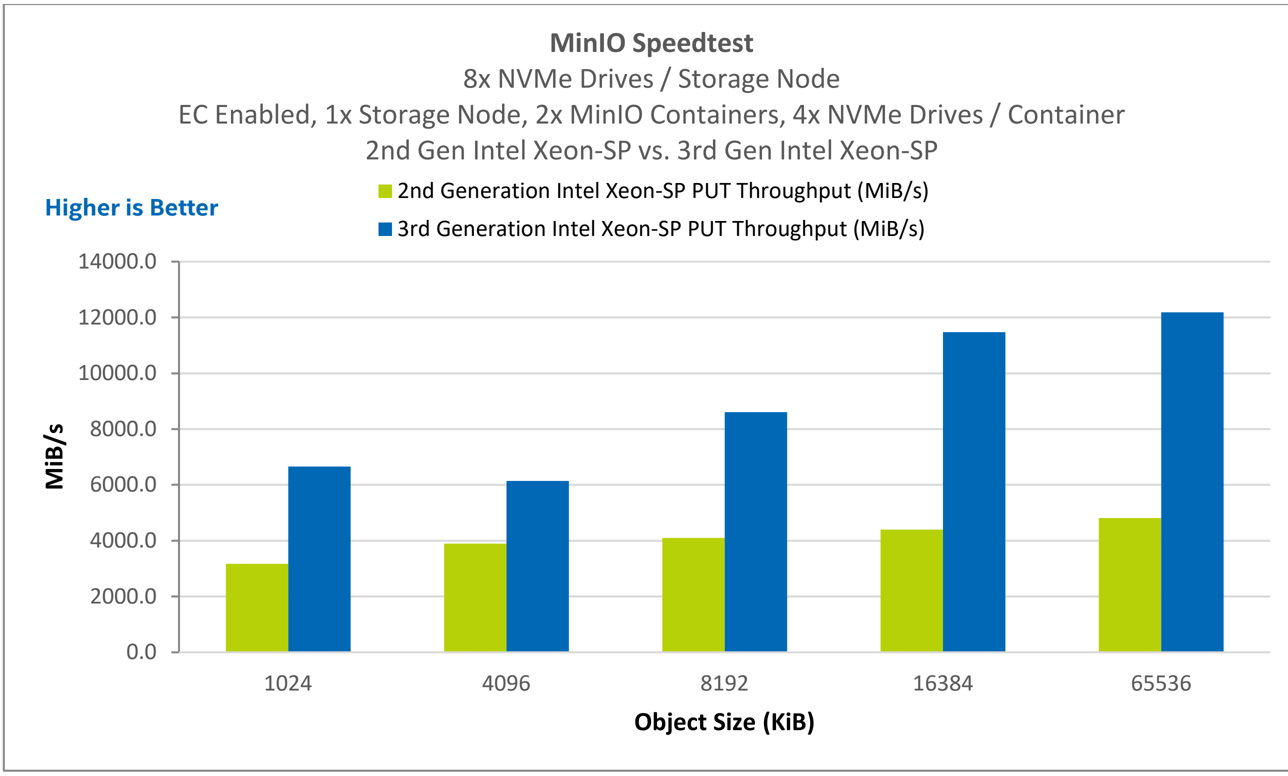

Figure 8. MinIO Speedtest, Aggregate PUT Throughput Comparison for up to 64 MiB object sizes, 2nd Gen Intel® Xeon® Scalable Processors versus 3rd Gen Intel® Xeon® Scalable Processors, Single Storage Node[2]

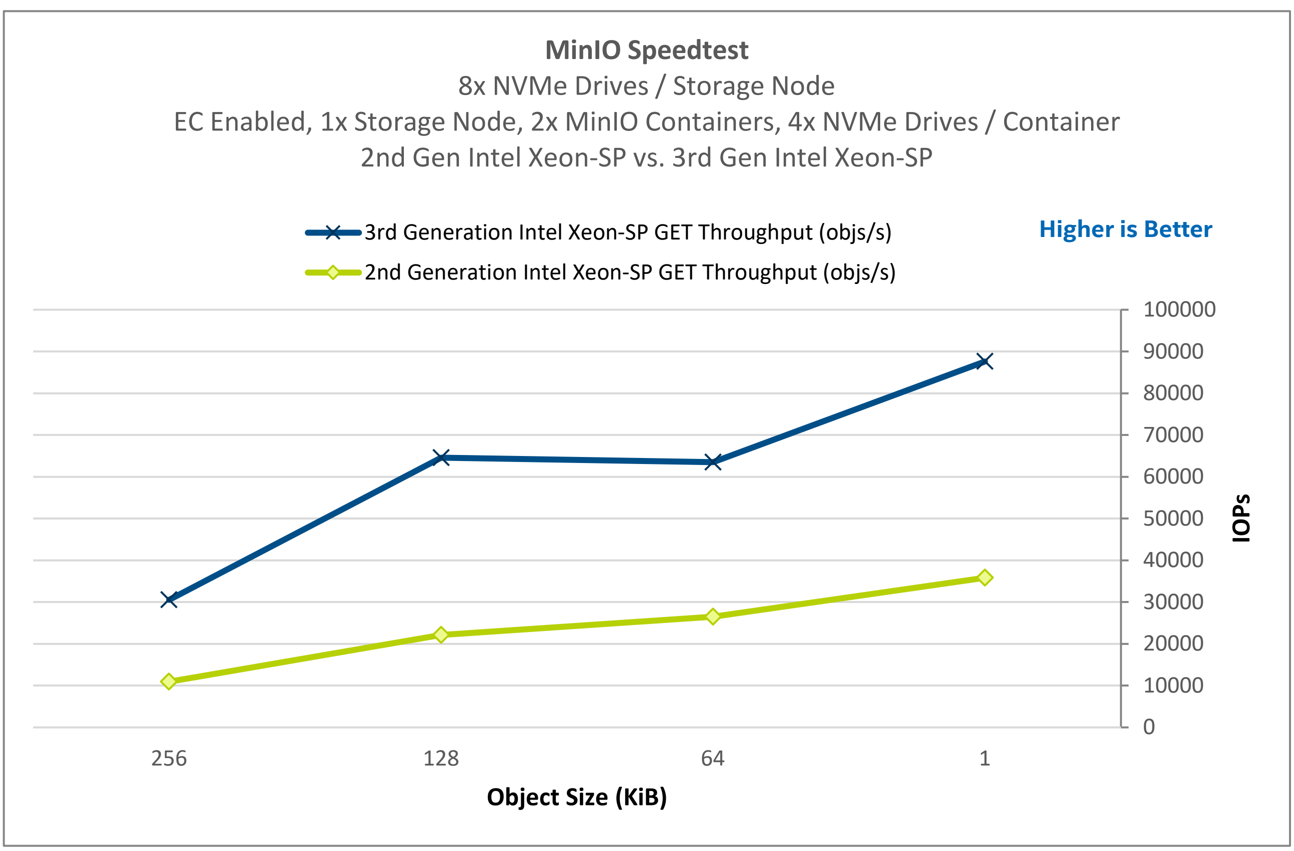

Figure 9. MinIO Speedtest, Aggregate GET Throughput Comparison for up to 256 KiB object sizes, 2nd Gen Intel® Xeon® Scalable Processors versus 3rd Gen Intel® Xeon® Scalable Processors, Single Storage Node

[2] See backup for workloads and configurations. Results may vary.

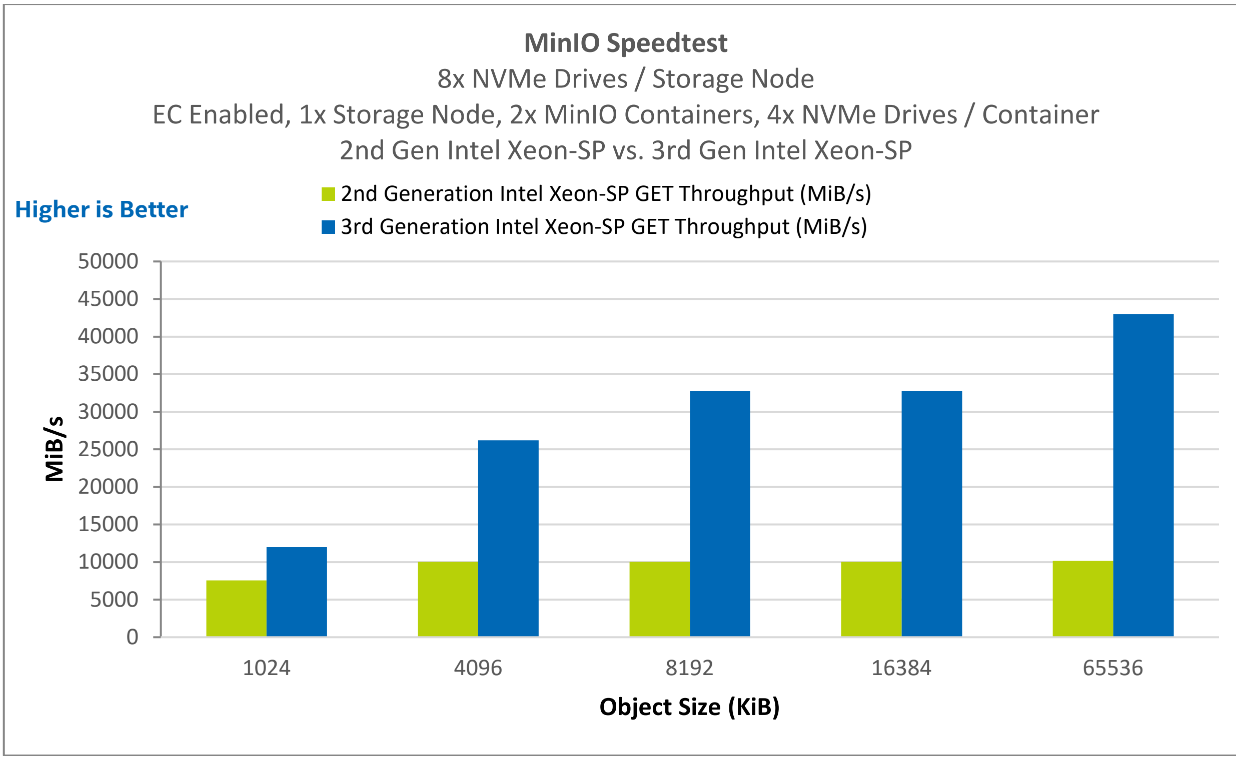

Figure 10. MinIO Speedtest, Aggregate GET Throughput Comparison for up to 64 MiB object sizes, 2nd Gen Intel® Xeon® Scalable Processors versus 3rd Gen Intel® Xeon® Scalable Processors, Single Storage Node[3]

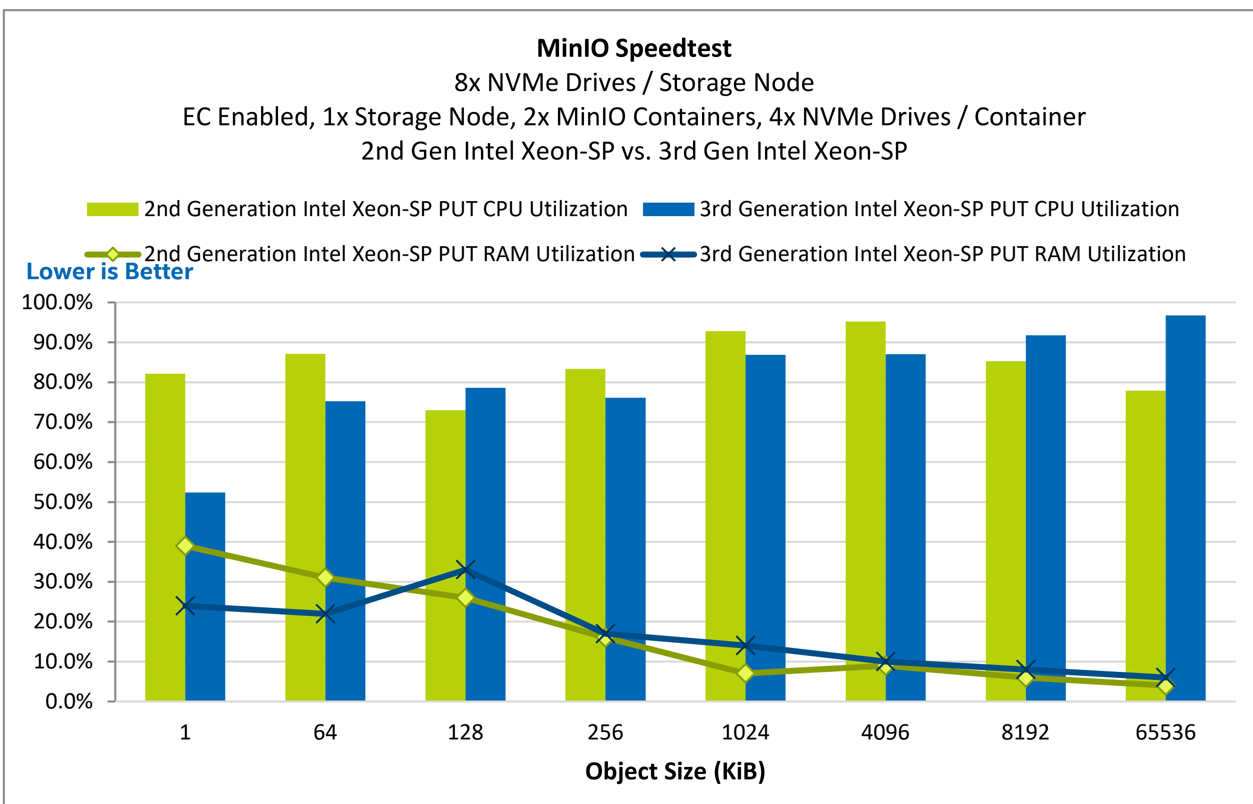

Figure 11. MinIO Speedtest, PUT CPU and Memory Utilization, 2nd Gen Intel® Xeon® Scalable Processors versus 3rd Gen Intel® Xeon® Scalable Processors, Single Storage Node[4]

[3] See backup for workloads and configurations. Results may vary.

[4] See backup for workloads and configurations. Results may vary.

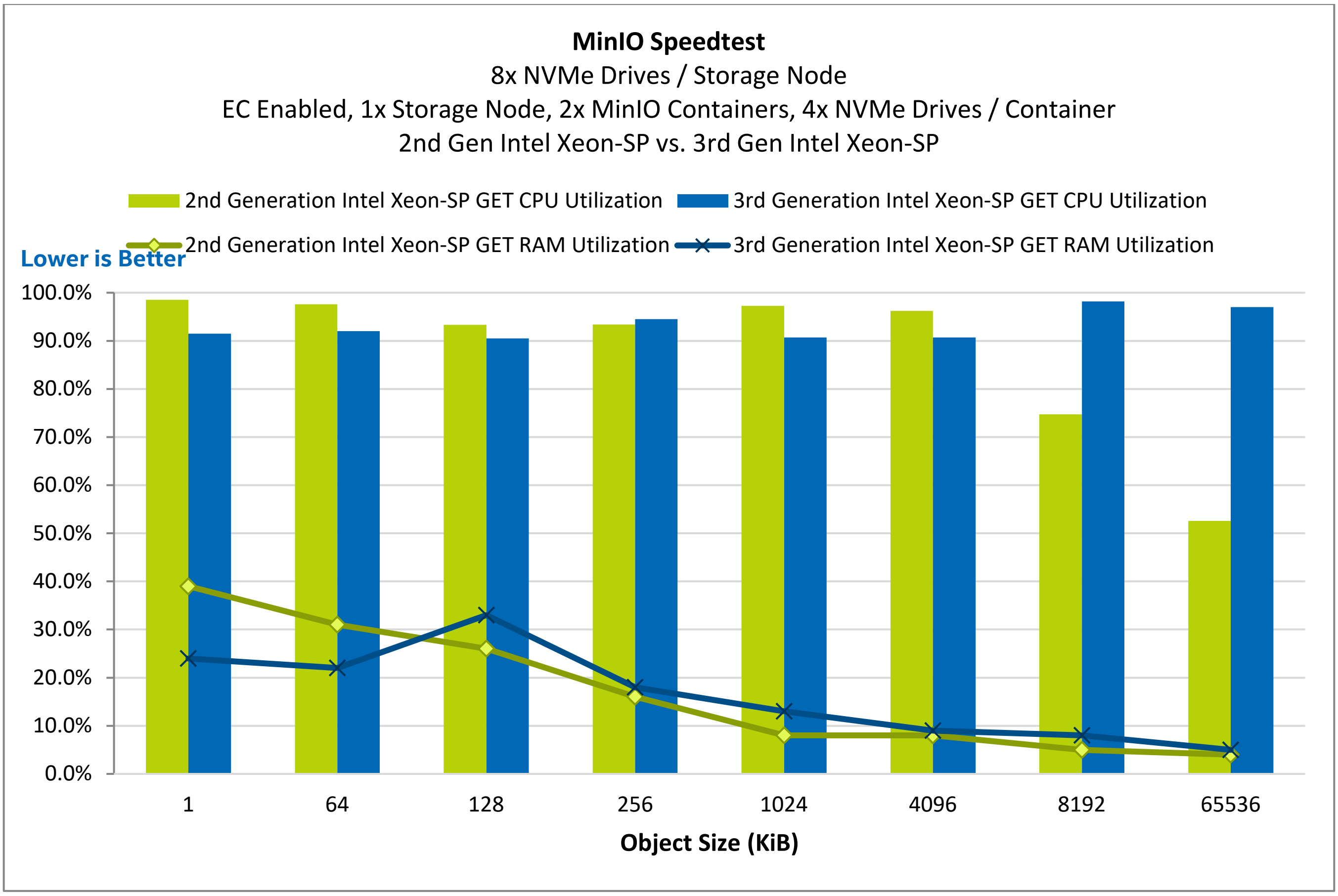

Figure 12. MinIO Speedtest, GET CPU and Memory Utilization, 2nd Gen Intel® Xeon® Scalable Processors versus 3rd Gen Intel® Xeon® Scalable Processors, Single Storage Node[5]

3.1 Analysis of Results

The following diagram presents the percentage performance improvement for the 3rd Gen Intel Xeon Scalable processor platform over the 2nd Intel Xeon Scalable processor platform.

[5] See backup for workloads and configurations. Results may vary.

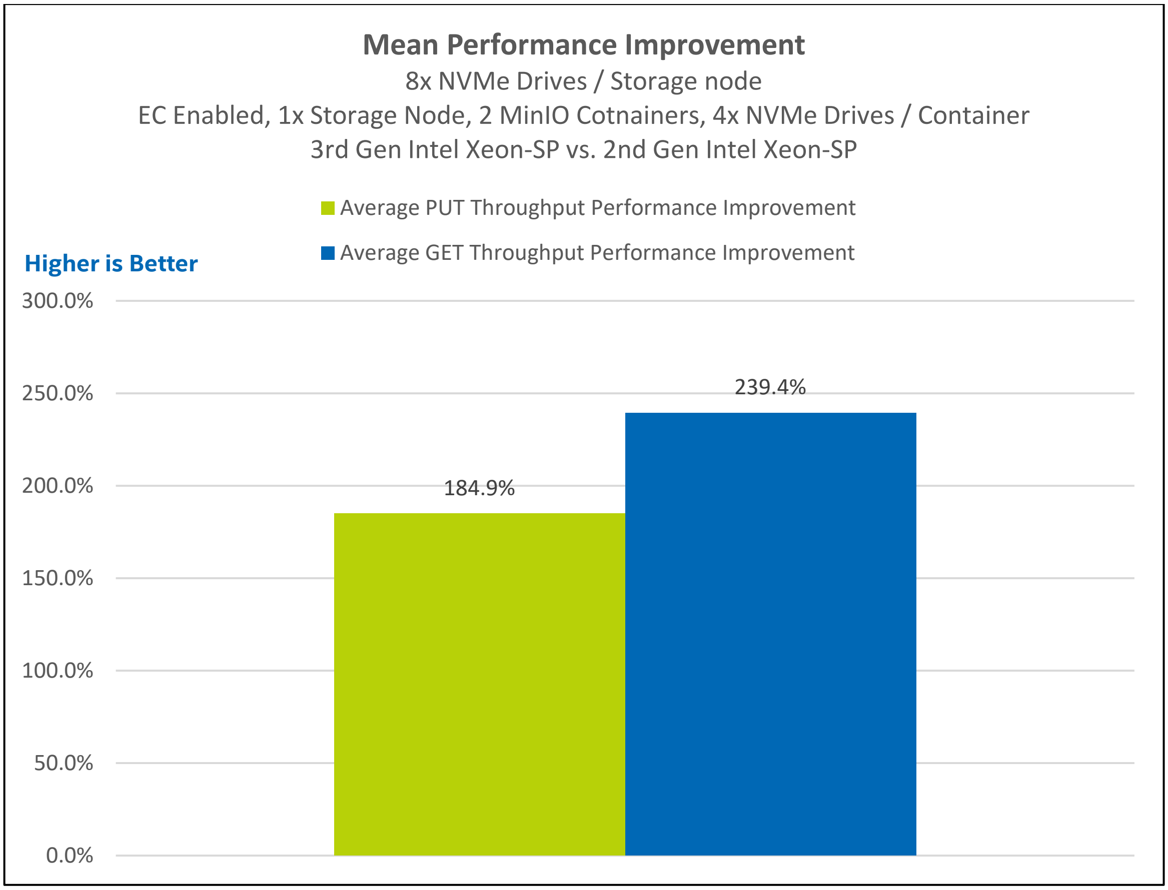

Figure 13. Mean PUT and Mean GET throughput performance improvement across all object sizes for 3rd Gen Intel® Xeon® Scalable Processors over 2nd Gen Intel® Xeon® Scalable Processors.[6]

In this case, we observe, on average, approximately up to 184.9% (1.849x) PUT throughput improvement across all object sizes along with approximately up to 239.4% (2.394x) GET throughput improvement across all object sizes.

For the majority of the PUT throughput rates, we observe that the CPU utilization is lower for 3rd Gen Intel Xeon Scalable processors. For example, for a 64KiB object size, the CPU utilization is approximately 13.7% lower for 3rd Gen Intel Xeon Scalable processors. In addition, for the majority of the GET throughput rates, the CPU utilization is lower for 3rd Gen Intel Xeon Scalable processors. For example, for a 1 KiB object size, the CPU utilization is approximately 7.4% lower for 3rd Gen Intel Xeon Scalable processors.

4 Summary

In this guide, we presented MinIO performance on both 2nd and 3rd Gen Intel Xeon Scalable processor platforms. In this case, observe on average up to 2.3x performance improvement in GET throughput with 3rd Gen Intel Xeon Scalable processors for MinIO storage workloads.

In terms of next steps, we plan to explore potential performance gains when enabling Intel® Advanced Vector Extensions 512 (Intel® AVX-512) library support for Transport Layer Security (TLS) for the MinIO storage application. In addition, we plan to perform a gen2gen performance comparison between 3rd Gen Intel Xeon Scalable processors and 4th Gen Intel® Xeon® Scalable processors.

[6] See backup for workloads and configurations. Results may vary.

1022/DN/WIPRO/PDF 742734-001US