Technology Guide

Extended Detection & Response (XDR) - Accelerate XDR Data Streaming using Apache Flink* on Intel® Platforms

Authors

Yiwen Li

Dhaval Patel

Jingdu Hou

Aparna Balachandran

Heqing Zhu

1 Introduction

Extended Detection & Response (XDR) is a Software as a Service (SaaS)-based, vendor-specific, security threat detection and incident response platform that natively integrates multiple security products into a cohesive security operations system. XDR unifies security-relevant endpoint detections with telemetry from security and business tools such as network analysis and visibility (NAV), email security, identity and access management, cloud security, security information and event management (SIEM), and more. It is a cloud-native platform built on big data infrastructure to provide security teams with flexibility, scalability, and opportunities for automation.

Processing and conducting advanced analytics on big data is one of the core building blocks of XDR. For enterprise customers to effectively address growing cybersecurity threats, such as zero-day attacks, and to protect the security of data and assets, it is critical for their XDR platform to be able to process large amounts of real-time data quickly and efficiently.

Apache Flink* is an open-source, distributed engine for stateful processing over unbounded (streams) and bounded (batches) data sets. Stream processing applications are designed to run continuously, with minimal downtime, and process data as it is ingested. Apache Flink is designed for low latency processing, performing computations in-memory, for high availability, removing single point of failures, and to scale horizontally. Flink is one popular design choice for building the data streaming infrastructure of XDR systems.

In this paper, we tested and validated the XDR data streaming workload by running the Flink Serialization benchmarks on Intel® Xeon® platforms. Our testing results demonstrated how Flink Serialization benchmarks can be tuned and optimized on Intel® Xeon® Scalable processors, essentially maximizing all the available compute power (90%+) CPU utilization. For various Flink Serializers, when compared against a 3rd Gen Intel® Xeon® Scalable processor, the 4th Gen Intel® Xeon® Scalable processor was able to achieve up to a 38% performance improvement in terms of throughput. This demonstrates the superior capability of 4th Gen Intel® Xeon® Scalable processors and shows that it can significantly help accelerate XDR data streaming using Apache Flink. This helps improve the data processing capabilities of the customer's XDR pipeline, which could potentially lead to higher data ingestion rates and faster threat detection and responses.

This document is part of the Network Transformation Experience Kits.

Table of Contents

Figures

Tables

Document Revision History

Revision	Date	Description
001	January 2024	Initial release.

1.1 Terminology

Table 1. Terminology

Abbreviation	Description
CPU	Central Processing Unit
CASB	Cloud Access Security Broker
CWPP	Cloud Workload Protection Platform
DLP	Data Loss Prevention
EDR	Endpoint Detection and Response
EPP	Endpoint Protection Platform
IAM	Identity and Access Management
NDR	Network Detection and Response
SEG	Secure Email Gateway
SWG	Secure Web Gateway
UEM	Unified Endpoint Management
XDR	Extended Detection and Response

1.2 Reference Documentation

Table 2. Reference Documents

Reference	Source
Palo Alto Networks: What is XDR?	https://www.paloaltonetworks.com/cyberpedia/what-is-xdr
Apache Flink	https://flink.apache.org/
Flink Benchmarks (v1.17)	https://github.com/apache/flink-benchmarks/tree/dev-1.17

2 Overview

2.1 XDR Architecture

XDR is an evolving technology that can offer unified threat prevention, detection and response capabilities. An XDR platform integrates, correlates and contextualizes data and alerts from multiple security prevention, detection and response components. XDR can be delivered on-premises or as a SaaS offering.

An XDR platform has the front end and the back end. The front end includes multiple security points, such as firewall, SWG, EPP/EDR, SEG, etc. These are the data sources for the XDR system. Typically, the front end will be monitored and the relevant data, such as server logs and security scanning results, will be sent to the XDR back end. The back end will ingest data from multiple front end devices and is responsible for processing the data and performing advanced analytics to detect any anomalies that occur and further generate security responses or recommendations. The high-level overview of an XDR system is illustrated in Figure 1.

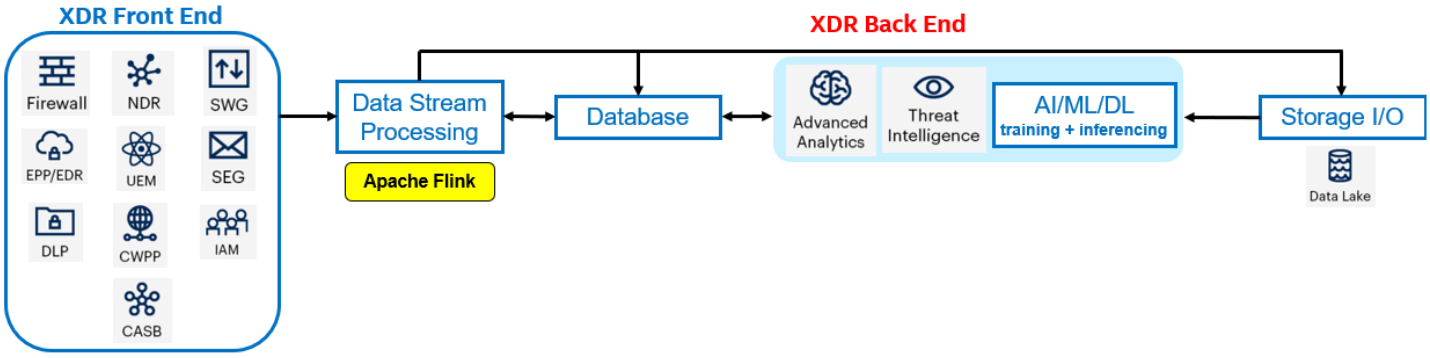

Figure 1. XDR Architecture Overview

2.2 XDR Data Streaming using Apache Flink*

The data stream processing is a key component in the XDR pipeline. It is responsible for processing all the real-time data ingested from multiple front-end points, thus it needs to be efficient and scalable.

Apache Flink is a popular choice for building the XDR data streaming component. It is an open-source framework and distributed processing engine for stateful computations over unbounded and bounded data streams. Flink has been designed to run in all common cluster environments, perform computations at in-memory speed and at any scale. Flink can help achieve low latency and high throughput through distributed processing.

As illustrated in Figure 2, in the XDR pipeline, real-time data from the front-end points will be ingested through Apache Flink into data streams. The processed data can then be stored into an in-memory database or into storage and be passed to applications to perform Advanced Analytics.

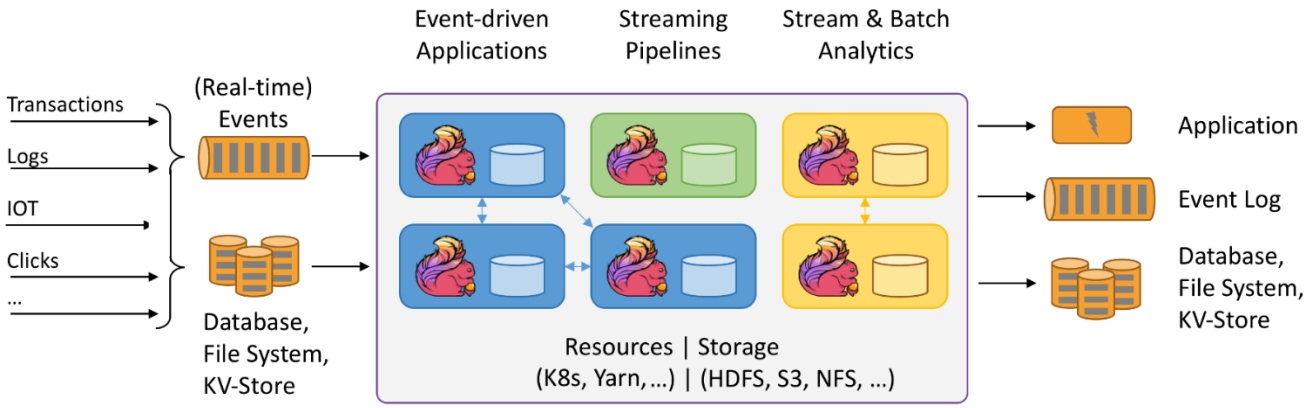

Figure 2. Data Streaming using Apache Flink

In this paper, we tested Apache Flink benchmark on different generations of Intel Xeon Scalable processors to demonstrate that running the Flink workload on Intel Xeon processors can achieve efficient utilization of CPU resources. When compared against the 3rd Gen Intel Xeon Scalable processor, the 4th Gen Intel Xeon Scalable processor has significant performance improvement in Flink benchmark, which shows that 4th Gen Intel Xeon Scalable processor can help build more efficient, performant, and scalable XDR systems.

3 Apache Flink Benchmark and Testing Methodology

3.1 Flink JMH Benchmark

As illustrated in Figure 3, data serialization is the core workload of Apache Flink, it is the process of converting an object into a stream of bytes to more easily store or transmit data. We ran serializer functions in Flink as our test cases for data streaming.

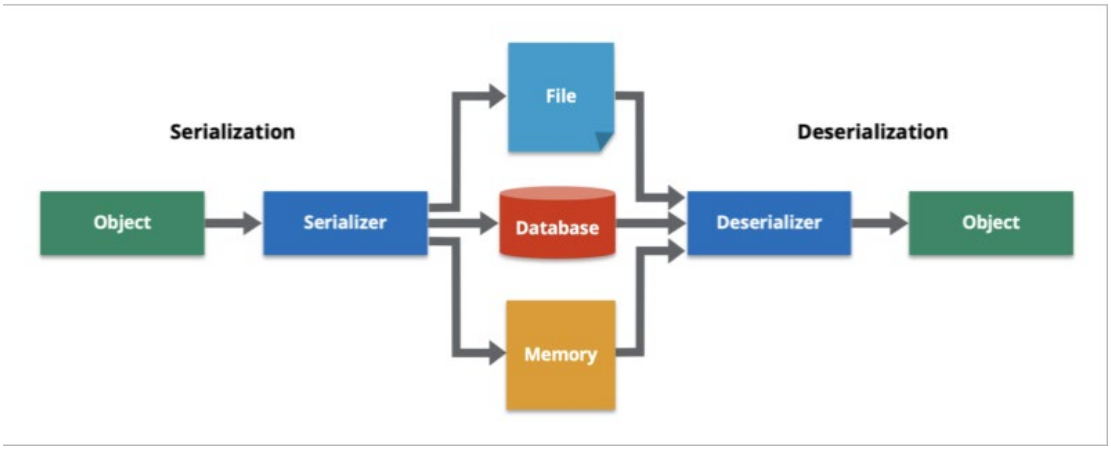

Figure 3. Data Serialization

For our testing, we leverage the Flink-JMH benchmark suite. JMH is a Java harness for building, running, and analyzing smaller benchmarks written in Java and targeting the JVM (Java virtual machine). JMH is maintained by Java and is the industry standard benchmarking tool for Java projects. Flink-JMH is the official benchmark from Apache. It uses the JMH micro

benchmark suite to define and execute test cases. Our testing software stack is shown in Table 3. The tests were setup in a containerized environment, as illustrated in Figure 4.

Table 3. Flink Testing SW Stack

Component	Version
OS Distribution	Ubuntu 22.04 Jammy Jellyfish
Kernel	5.15.0-75-generic
Kubernetes/Containerd	v1.26.6/v1.6.24
JRE/JDK	2:1.11-72build2
Flink	v1.17.1 (latest stable)
Flink-Benchmark	1.17

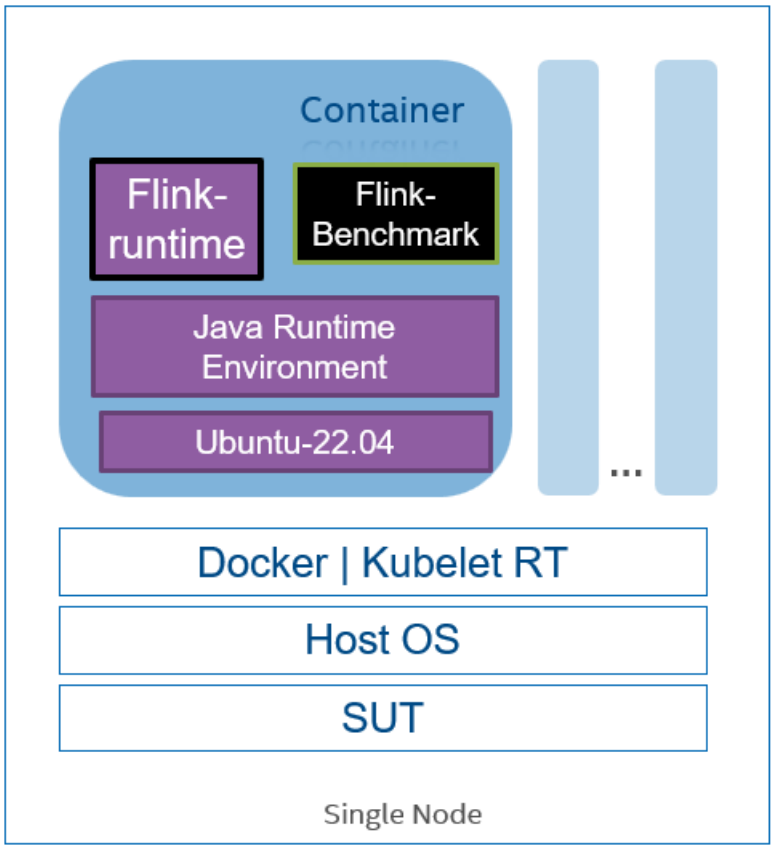

Figure 4. Testing Environment Setup

3.2 Apache Flink Core Scaling Experiment

We conducted CPU core scaling experiment to try to identify the best core number assignment that would achieve high CPU utilization and yield the best throughput per core.

In our test, we used serializerKryo-1.17.1 testcase from the Flink-JMH benchmark.. The software configuration is listed in Table 4. As illustrated in Table 5 and Figure 5, we found out that when processing the dataset of 300,000 records on a 3rd Gen Intel Xeon Scalable processor, eight CPU cores is the best assignment option, as it can achieve the highest "Throughput per Core" with a high CPU utilization.

Thus, when handling large amounts of data, we recommend splitting the workload into multiple containers, each assigned with the number of CPU cores that can achieve the highest "Throughput per Core." In our testcase, we assigned eight CPU cores to a single container, handling 300,000 records, and spin up multiple containers to scale up the workload.

Table 4. Flink-JMH Benchmark Software Configuration

Tunable parameters	ITERATIONS=30 | WARMUP=20 | THREADS=16 | FORKS=4
Data size	300,000 copies of records

Table 5. CPU Core Scaling for Flink Serialization Workload on a 3rd Gen Intel® Xeon® Scalable processor

Cores Available	Throughput (ops/ms)	Throughput per Core	CPU Utilization
4	222.971	55.74	99.05%
8	532.350	66.54	95.92%
16	913.915	57.12	95.23%
32	1795.567	56.11	90.82%
64	2968.021	46.38	77.13%
72	3179.313	44.16	73.84%

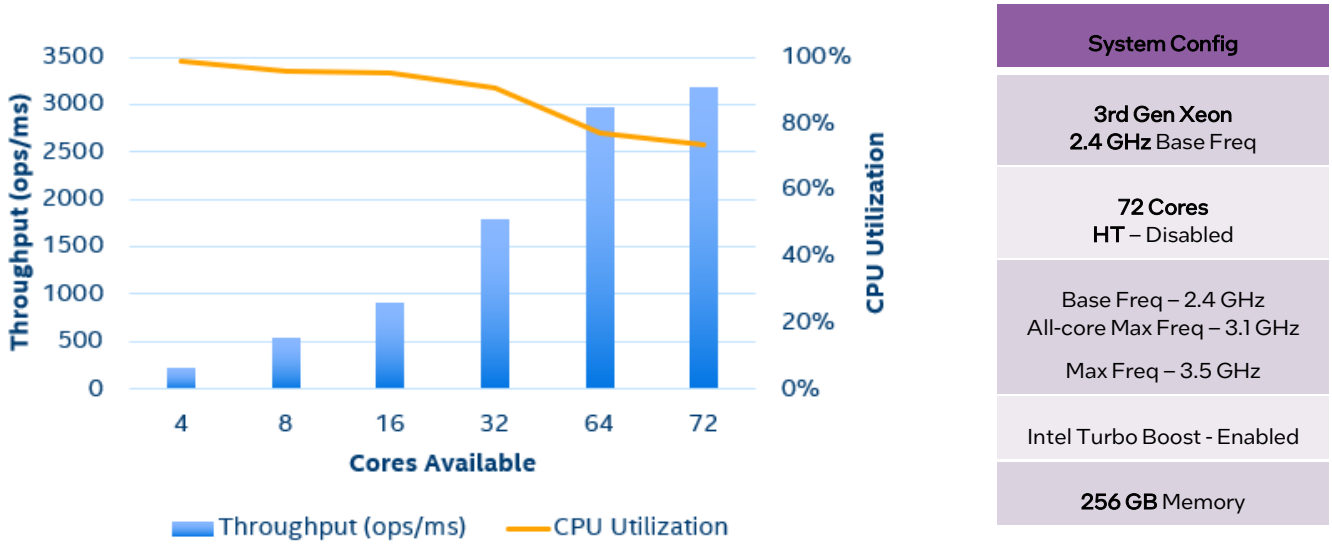

Figure 5. CPU Core Scaling for Flink Serialization Workload on a 3rd Gen Intel® Xeon® Scalable processor

4 Performance Data and Results

Using the benchmarking tools and testing methodology described above (in Section 3), we ran the Apache Flink Serialization tests on Intel® Xeon® Platinum 8360Y @2.4GHz (3rd Gen) and a 4th Gen Intel® Xeon® Scalable processor (56-cores) @2.4GHz. (CPU frequency was fixed at the same level to give a fair comparison.)

Our results (shown in Figure 6.) show that for various serializers in Flink, 4th Gen Intel Xeon Scalable processor can improve the total throughput performance by around 23% up to 38%, when compared against 3rd Gen Intel Xeon Scalable processor.

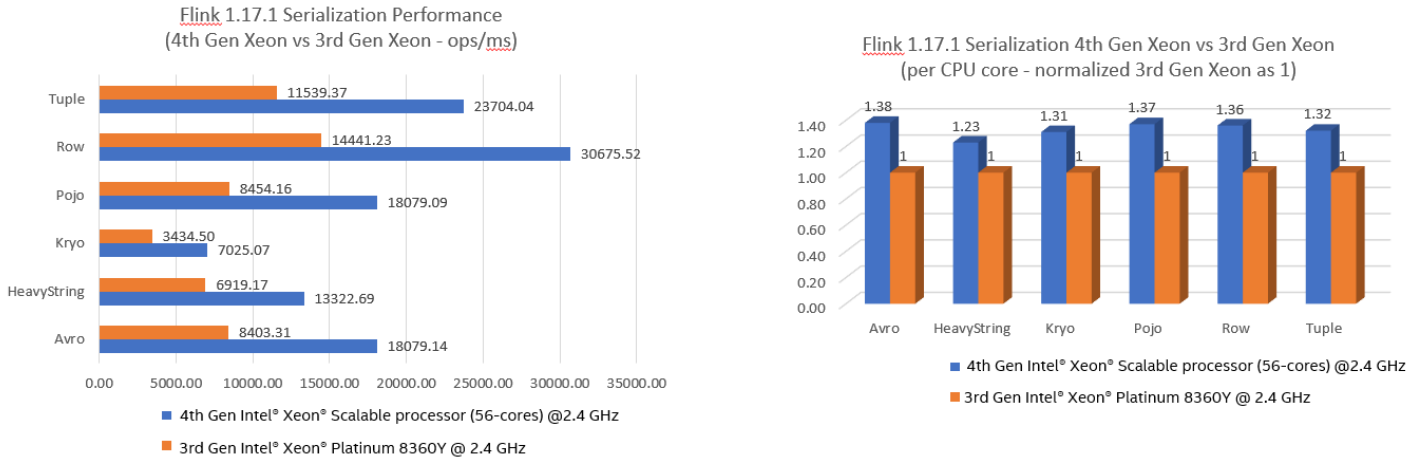

Figure 6. Flink Serialization Results (4th Gen Intel Xeon Scalable processor vs. 3rd Gen Intel Xeon Scalable processor)

5 Summary

XDR is an emerging technology and a key area for network security. Apache Flink open-source software is being used as a popular design choice for building the data streaming component for XDR systems. We tested Flink's core serialization function on Intel Xeon processors. Our results show that when compared against a 3rd Gen Intel Xeon Scalable processor, the 4th Gen Intel Xeon Scalable processor can achieve up to around 38% throughput performance improvement with high CPU utilization. This demonstrates the superior capability of 4th Gen Intel Xeon Scalable processor to help improve the data ingestion rates and data analytics speed for XDR pipeline to more effectively detect and respond to security threats and help protect data and digital assets of end users.

Appendix A Platform Configuration

Table 6. Platform Configuration

	Config 1	Config 2
Name	ad07-15-cyp	aj09-01-fcp
Time	Fri Sep 8 08:30:21 UTC 2023	Mon Sep 25 11:43:21 PM UTC 2023
System	Intel Corporation WHITLEY	Intel Corporation M50FCP2SBSTD
Baseboard	Intel Corporation WHITLEY	Intel Corporation M50FCP2SBSTD
Chassis	Intel Corporation Rack Mount Chassis	..... Rack Mount Chassis
CPU Model	Intel(R) Xeon(R) Platinum 8360Y processor @ 2.40GHz	4th Gen Intel(R) Xeon(R) Scalable processor (56-cores)
Sockets	2	2
Cores per Socket	36	56
Hyperthreading	Disabled	Disabled
CPUs	72	112
Intel Turbo Boost	Enabled	Enabled
Base Frequency	2.4GHz	2.0GHz
All-core Maximum Frequency	3.1GHz	3.0GHz
Maximum Frequency	3.5GHz	3.8GHz
NUMA Nodes	2	2
Prefetchers	L2 HW, L2 Adj., DCU HW, DCU IP	L2 HW, L2 Adj., DCU HW, DCU IP
PPINs		
Accelerators (4th Gen Only)	Intel® DLB 0 [0], Intel® DSA 0 [0], Intel IAX 0 [0], Intel® QAT 0 [0]	Intel® DLB 2 [0], Intel® DSA 2 [0], Intel® IAX 2 [0], Intel® QAT 2 [0]
Installed Memory	256GB (16x16GB DDR4 3200 MT/s [3200 MT/s])	256GB (16x16GB DDR5 4800 MT/s [4800 MT/s])

Hugepagesize	2048 kB	2048 kB
Transparent Huge Pages	madvise	madvise
Automatic NUMA Balancing	Enabled	Enabled
Ethernet Adapter	2x Intel® Ethernet Network Adapter E810 (QSFP), 2x Intel® Ethernet Network Adapter X710 (10GBASE-T), 2x Intel® Ethernet Network Server Adapter I350	2x Intel Ethernet Network Adapter E810 (QSFP), 2x BCM57416 NetXtreme-E Dual-Media 10G RDMA Ethernet Controller
Disk	1x 931.5G Intel® SSD DC P4510 SSDPE2KX010T8	3x 1.7T SAMSUNG MZQL21T9HCJR-00A07, 1x 447.1G Micron_7450_MTFDKBA480TFR
BIOS	SE5C620.86B.01.01.0007.2210270543	SE5C741.86B.01.01.0004.2303280404
Microcode	0xd000389	0x2b0001b0
OS	Ubuntu 22.04.2 LTS	Ubuntu 22.04.2 LTS
Kernel	5.15.0-71-generic	5.19.0-32-generic
TDP	250 watts	350 watts
Power & Perf Policy	Normal (6)	Normal (6)
Frequency Governor	performance	performance
Frequency Driver	intel_pstate	intel_pstate
Max C-State	9	9
System Summary	1-node, 2x Intel(R) Xeon(R) Platinum 8360Y processor @ 2.40GHz, 36 cores, HT Off, Turbo On, NUMA 2, Integrated Accelerators Available [used]: Intel DLB 0 [0], Intel DSA 0 [0], Intel IAX 0 [0], Intel QAT 0 [0], Total Memory 256GB (16x16GB DDR4 3200 MT/s [3200 MT/s]), BIOS SE5C620.86B.01.01.0007.2210270543, microcode 0xd000389, 2x Intel Ethernet Network Adapter E810 (QSFP), 2x Intel Ethernet Network Adapter X710 (10GBASE-T), 2x Intel Ethernet Network Server Adapter I350, 1x 931.5G Intel SSD DC P4510 SSDPE2KX010T8, Ubuntu 22.04.2 LTS, 5.15.0-71-generic, WORKLOAD+VERSION, COMPILER, LIBRARIES, OTHER_SW, score=?UNITS. Test by COMPANY as of 09/08/23.	1-node, 2x Intel(R) Xeon(R) Platinum 8480L processor, 56 cores, HT Off, Turbo On, NUMA 2, Integrated Accelerators Available [used]: Intel DLB 2 [0], Intel DSA 2 [0], Intel IAX 2 [0], Intel QAT 2 [0], Total Memory 256GB (16x16GB DDR5 4800 MT/s [4800 MT/s]), BIOS SE5C741.86B.01.01.0004.2303280404, microcode 0x2b0001b0, 2x Intel Ethernet Network Adapter E810 (QSFP), 2x BCM57416 NetXtreme-E Dual-Media 10G RDMA Ethernet Controller, 3x 1.7T SAMSUNG MZQL21T9HCJR-00A07, 1x 447.1G Micron_7450_MTFDKBA480TFR, Ubuntu 22.04.2 LTS, 5.19.0-32-generic, WORKLOAD+VERSION, COMPILER, LIBRARIES, OTHER_SW, score=?UNITS. Test by COMPANY as of 09/25/23.

0124/DN/WIT/PDF 795812-001US