

## Achieve Real-Time Data Processing with Confluent® Platform and Apache Kafka®

In collaboration with:



### Tech Note by

Todd Mottershead

*Todd.mottershead@dell.com*

Seamus Jones

*Seamus.jones@dell.com*

Brian Walters

*Brian.walters@intel.com*

Krzysztof Cieplucha

*krzysztof.cieplucha@intel.com*

Murali Madhanagopal

*Murali.Madhanagopal@intel.com*

### Summary

Enabling mission critical application, system and connecting data to the entire organization with real-time data flow and process means that the system and software stack must be optimized. In this document Intel and Dell discuss key considerations and sample configurations for PowerEdge server deployments to ensure your Confluent Kafka architecture is robust and takes advantage of the most recent advancements in server technology.

Mission-critical applications need to analyze large amounts of data in real time, but this requires refined tools built on scalable platforms. Originally developed at LinkedIn by the founders of Confluent, Apache Kafka® is an open-source, high-throughput message broker that fills this need. It quickly decouples, queues, processes, stores and consumes high-volume streams of event data. With Apache Kafka, enterprises can acquire data once and consume it multiple times.

Confluent continues to enhance the Kafka platform with tools like cluster management, additional security, and more connectors. Companies like Square, Bosch and The Home Depot use Confluent's distribution of Apache Kafka to identify actionable patterns within business data<sup>1</sup>. Intel created an Apache Kafka data pipeline based on Confluent® Platform for faster security threat detection and response for its Cyber Intelligence Platform (CIP). Data flows to a Kafka message bus and then into the Splunk® platform.

Organizations that are looking for a solution to enable real-time processing of massive data streams should consider Confluent Platform and Apache Kafka running on Dell EMC™ PowerEdge™ servers with high-performing Intel compute, storage and networking technologies.

### Key Considerations

- **Compute.** 3rd Generation Intel® Xeon® Scalable processors ingest and analyze massive quantities of data fast in the decoupling work common to Apache Kafka broker nodes.
- **Storage.** The Intel SSD P5500 is recommended for storage for all node types. Architected with 96-layer TLC and Intel 3D NAND Technology, it optimizes performance and capacity. The Dell™ PowerEdge RAID Controller (PERC) H755N is recommended for Brokers + Apache ZooKeeper™ nodes. It offers expandable storage capacity to improve performance.
- **Networking.** Network speed is one of the most important factors in Kafka performance. Intel Ethernet 800 Series network adapters enable scaling from 10 gigabit Ethernet (GbE) to 100 GbE for accelerated packet processing.

## Available Configurations

Configurations for the control center node, ksqlDB + Kafka Connect + Schema Registry, and Brokers + Apache ZooKeeper are shown below.

	<b>Control Center Node (One Node Required)</b>	<b>ksqlDB + Apache Kafka® Connect + Schema Registry (Minimum of Two Nodes Required)</b>	<b>Brokers + Apache ZooKeeper™ (Minimum of Three Nodes Required)</b>
Platform	Dell EMC™ PowerEdge™ R650 or R750 chassis supporting NVM Express® (NVMe®) drives		
CPU <sup>ii</sup>	2 x Intel® Xeon® Silver 4316 processor (20 cores at 2.3 GHz)	2 x Intel® Xeon® Gold 6330 processor (28 cores at 2.0 GHz)	2 x Intel® Xeon® Silver 4316 (20 cores at 2.3 GHz)—small throughput clusters 2 x Intel® Xeon® Gold 6338 (32 cores at 2.0 GHz)—medium throughput clusters 2 x Intel® Xeon® Platinum 8368 (38 cores at 2.4 GHz)—high throughput clusters with full encryption enabled
DRAM <sup>iii</sup>	64 GB (4 x 16 GB)	128 GB (8 x 16 GB)	128 GB (8 x 16 GB) or more
Boot device	Dell EMC™ Boot Optimized Server Storage (BOSS)-S2 with 2 x 480 GB Intel® SSD D3-S4510 M.2 Serial ATA (SATA)		
Storage controller <sup>iv</sup>	None		Dell™ PERC H755N Front NVMe
Storage <sup>v</sup>	2 x 3.84 TB Intel® SSD P5500		4 x 3.84 TB Intel® SSD P5500
Network interface controller (NIC)	Intel® Ethernet Network Adapter E810-XXVDA2 for OCP3 (dual-port 25 Gb)		Intel® E810-XXVDA2 for OCP3 (dual-port 25 Gb) or Intel® E810-CQDA2 PCIe® (dual-port 100 Gb) for high-throughput clusters

### Learn More

Contact your dedicated Dell or Intel account team. [1-877-289-3355](tel:1-877-289-3355)

Download the solution briefs and white papers below:

- [Enabling Real-Time Processing of Massive Data Streams](#)
- [IT@Intel: Building a Modern, Scalable Cyber Intelligence Platform with Apache Kafka](#)
- [IT@Intel: Transforming Intel's Security Posture with Innovations in Data Intelligence](#)

The information in this publication is provided as is. Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2021 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, PowerEdge and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners.

Dell Inc. believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided as is. Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2021 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, PowerEdge and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners.

Dell Inc. believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

---

<sup>i</sup> Confluent. "Set Your Data in Motion." 2021. [www.confluent.io/](http://www.confluent.io/).

<sup>ii</sup> Small throughput: less than 10 gigabits per second (Gbps), medium throughput: less than 25 Gbps, high throughput: more than 25 Gbps

<sup>iii</sup> Brokers and Apache ZooKeeper™: More memory might be required to accommodate traffic bursts.

<sup>iv</sup> Brokers and Apache ZooKeeper™: An NVMe® RAID controller is optional for small- and medium-throughput clusters.

<sup>v</sup> Brokers and Apache ZooKeeper™: Add more drives or add higher capacity drives as needed for higher throughput, extended data-retention periods or desired (optional) RAID configurations.