TECHNICAL BRIEF

Supercharge Apache Kafka with Confluent Platform and Diamanti

KEY BENEFITS

INCREDIBLE FLEXIBILITY

Scale Kafka messaging queues exponentially with the powerful combination of Diamanti, Kubernetes and the Operator Framework.

EXTREME PERFORMANCE

Get more out of your Kafka cluster. A Kafka cluster running replicated and non-replicated topics on Diamanti yields incredibly high performance of 1M writes/sec (producer) and 2M reads/sec (consumer) per node at 10 ms latency.

REDUCE TCO

Achieve more with less. Reduce infrastructure footprint for running Kafka clusters by at least 25% using the Diamanti Platform.

Introduction

Everything from IoT devices, web transactions and operational intelligence is generating streaming data. Enterprises are putting immense emphasis on business intelligence, analytics, web tracking, log aggregation and monitoring using streaming data as their source. This results in the necessity for large scale data processing from multiple streaming data sources. Thus, today the need for large scale data processing is not just on the rise but has become the de facto reality for an enterprise.

Apache Kafka is the industry-leading, highly scalable, fault-tolerant, distributed streaming platform built to handle trillions of events per day. It provides real-time analysis of large amounts of data with scalability and reliability. Kafka is used in production by more than a third of the Fortune 500. Performance is where Kafka shines ahead of other streaming platforms. However, like all things, it is bound by the underlying infrastructure and storage.

Confluent - The Complete Streaming Platform for Apache Kafka

Founded by the original developers of Apache Kafka, Confluent delivers the most complete distribution of Kafka with Confluent Platform. Confluent Platform is an enterprise-ready platform that complements Kafka with advanced capabilities designed to help accelerate application development and connectivity, enable event transformations through stream processing, simplify enterprise operations at scale and meet stringent architectural requirements.

Why Diamanti helps deliver optimal performance for the Confluent Platform

The Diamanti Enterprise Kubernetes Platform provides enterprises with turnkey operational infrastructure using standard virtualization protocols for storage and networking alongside open-source CNI and CSI plug-ins.

The Diamanti platform includes low-latency and high-performance NVMe flash storage, 40 GbE networking, and open-source Docker and Kubernetes pre-installed. Cloud native applications and other container workloads can be deployed in minutes on the Diamanti cluster, where each pod is assigned a routable IP address due to Diamanti's innovative approach to network virtualization for containers. With its ultra-fast CSI and CNI, the Diamanti Enterprise Kubernetes Platform is ideal for applications such as Kafka that require low latency network and storage.

Implementation of multi-zone clusters on the Diamanti platform allows a Kubernetes cluster to distribute nodes across different zones, ensuring application and infrastructure availability. Diamanti simplifies multi-zone cluster configuration and management with built-in capabilities to protect applications from failures.

Integrating Confluent with the Diamanti Enterprise Kubernetes Platform

Confluent can seamlessly integrate with the Diamanti Enterprise Kubernetes Platform. The Confluent Enterprise Apache Kafka distribution can be installed using Confluent Docker images.

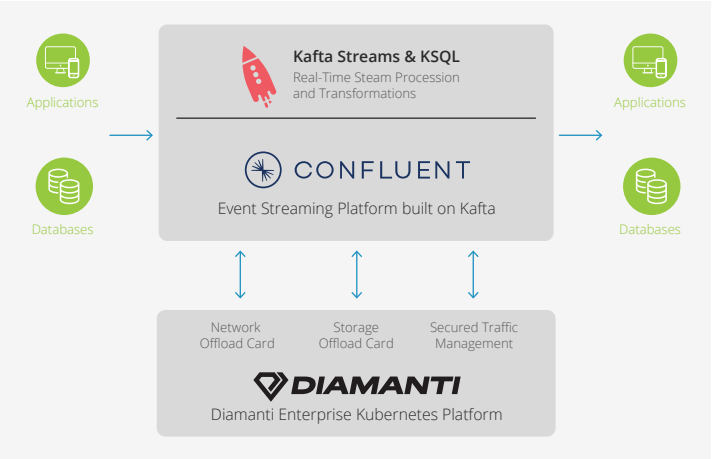

Figure 1: Confluent Platform integration with the Diamanti Enterprise Kubernetes Platform

A Kafka cluster was deployed with the following configuration on Diamanti:

- Three instances of Kafka brokers with 32 GB RAM, 30 GB Heap, 6 CPU cores and 100 GB storage
- Three instances of Apache ZooKeeper with 4 GB RAM, 3 GB Heap, and 4 CPU cores

ZooKeeper is simply used as a coordinating service with minimal requirements.

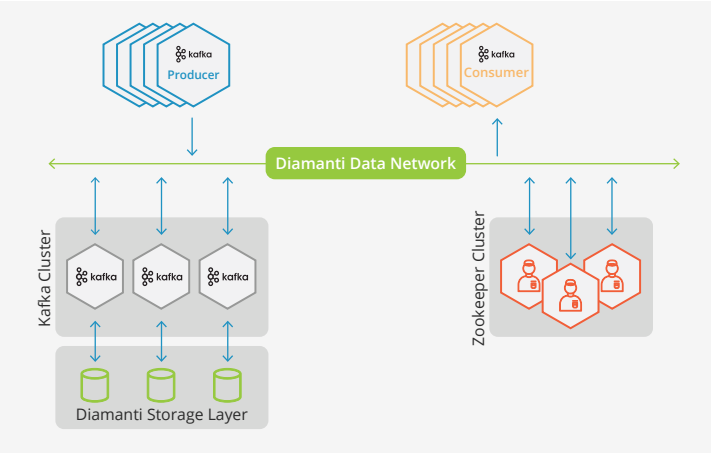

Figure 2: Apache Kafka deployment on Diamanti Enterprise Kubernetes Platform

To benchmark Kafka, containerized Kafka Producer and Consumer clients were deployed on the Diamanti Enterprise Kubernetes Platform. To mimic a real-world scenario, a mix of replicated and unreplicated topics were used. For this benchmarking exercise, five Producer and five Consumer Kafka client instances each consuming 1 GB RAM and 4 CPU cores were used. Two test scenarios were run in parallel. The first test scenario used two Producers and two Consumers to read and write 60 million records via non-replicated topics with 16 partitions. The second test scenario used three Producers and three Consumers to read and write 60 million records asynchronously via 3x replicated topics with 16 partitions. Both test scenarios were executed in parallel with a 100-byte message size.

Figure 3: Realized benefits of Confluent Platform with the Diamanti Enterprise Kubernetes Platform

With an untuned three-node Kafka cluster running a mix of replicated and non-replicated topics on the Diamanti Enterprise Kubernetes Platform, Producers were able to achieve 3 million writes per second with an average latency of 10 milliseconds. Meanwhile, Consumers were able to read topics at approximately 6 million messages per second. Such unprecedented performance opens up a multitude of possibilities for Kafka in an enterprise environment.

Table 1 summarizes the results obtained for each Producer and Consumer. These results were achieved with a CPU usage of just 4.5 cores per Kafka instance with very low and deterministic latency. The results can be further enhanced with better tuning and scaling of Kafka.

	Producers			Consumers	
Topics	Messages per Second	Network Traffic (MB/sec)	Average Latency (ms)	Messages per Second	Network Traffic (MB/sec)
non-replication-1	478,491	45.63	10.59	1,549,586	147.78
non-replication-2	670,653	63.96	8.42	1,361,533	129.84
replication-1	764,448	72.90	16.66	716,452	68.32
replication-2	673,597	64.24	7.83	1,502,484	143.28
replication-3	477,828	45.57	13.10	717,875	68.46
Total	3,065,017	292.3	11.32	5,847,930	557.68

Table 1: Apache Kafka on Diamanti Platform with a Mix of Replicated and Unreplicated Topics

Summary

Apache Kafka has become the de-facto standard for building real-time streaming data pipelines for big data, analytics and other large data systems. With its zero-copy architecture, Kafka provides a fast, scalable and fault-tolerant distributed messaging system. The Confluent Platform simplifies connecting data sources to Kafka, building applications with Kafka, as well as securing, monitoring, and managing Kafka. The Diamanti Enterprise Kubernetes Platform provides a turnkey solution for deploying containerized Kafka clusters and associated applications. With its bare-metal architecture, guaranteed QoS, and PCIe-level isolation for storage and networking, the Diamanti platform provides unprecedented throughput for Kafka with low and deterministic latency. With seamless deployment of Confluent on Diamanti Enterprise Kubernetes Platform multiple Kafka clusters can be deployed without noisy neighbor problems resulting in very high host utilization and significantly reducing total cost of ownership (TCO).

Ready to get started? Contact us today.