SOLUTION BRIEF

Supercharge Apache Kafka with Diamanti Enterprise Kubernetes® Platform

Enterprises are putting immense emphasis on business intelligence, analytics, IoT, web tracking, log aggregation and monitoring. This results in a flood of data and introduces the necessity for large scale data processing from multiple sources. Apache Kafka, often referred to as 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 software applications, it is limited by the underlying infrastructure and storage. This is where the bare-metal hyperconverged infrastructure based on Diamanti Enterprise Kubernetes Platform helps unveil the full potential of Kafka.

Apache Kafka on Diamanti at Unprecedented Performance and Scale

The Diamanti Enterprise Kubernetes Platform provides enterprises with turnkey operational infrastructure using standard virtualization protocols for storage and networking alongside open-standard 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 after racking and stacking 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

DIAMANTI AT A GLANCE

SIMPLICITY
- 15-minute bare-metal deployment
- Easy to manage and scale
- Kubernetes certified
- No vendor lock-in

PERFORMANCE
- Real-world 1,000,000 IOPS per 1U
- Consistent 100-microsecond latency
- Industry-leading application-level transactions per second

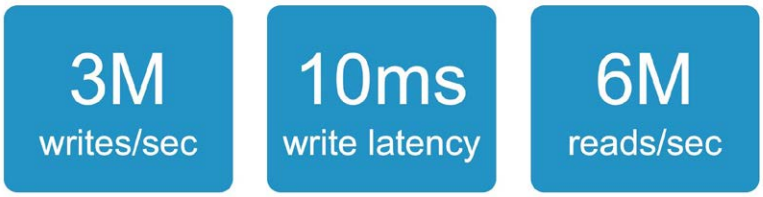
EFFICIENCY
- 70% lower TCO
- 100% host utilization
- 95% usable storage capacity
- No hypervisor needed
- Guaranteed QoS with no overprovisioning

ENTERPRISE READY
- Full-stack support
- Production-grade SLAs
- Secure multi-tenant isolation
- Advanced DR/DP
- On-premises availability zones

Diamanti is Kubernetes certified

CSI and CNI, Diamanti Enterprise Kubernetes Platform is ideal for applications which require low latency network and storage such as Kafka.

An untuned instance of Kafka running on a three node Diamanti cluster can write to 3 million messages per second and read from 6 million messages per second, maintaining a write latency of 10 milliseconds with a minimal amount of system resources. Such unprecedented performance opens up a multitude of possibilities for Kafka in an enterprise environment.



Deploying Apache Kafka on Diamanti Enterprise Kubernetes Platform

Apache Kafka can be deployed on Diamanti Enterprise Kubernetes Platform using Helm charts. A Kafka cluster was deployed with the following configuration:

- 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.

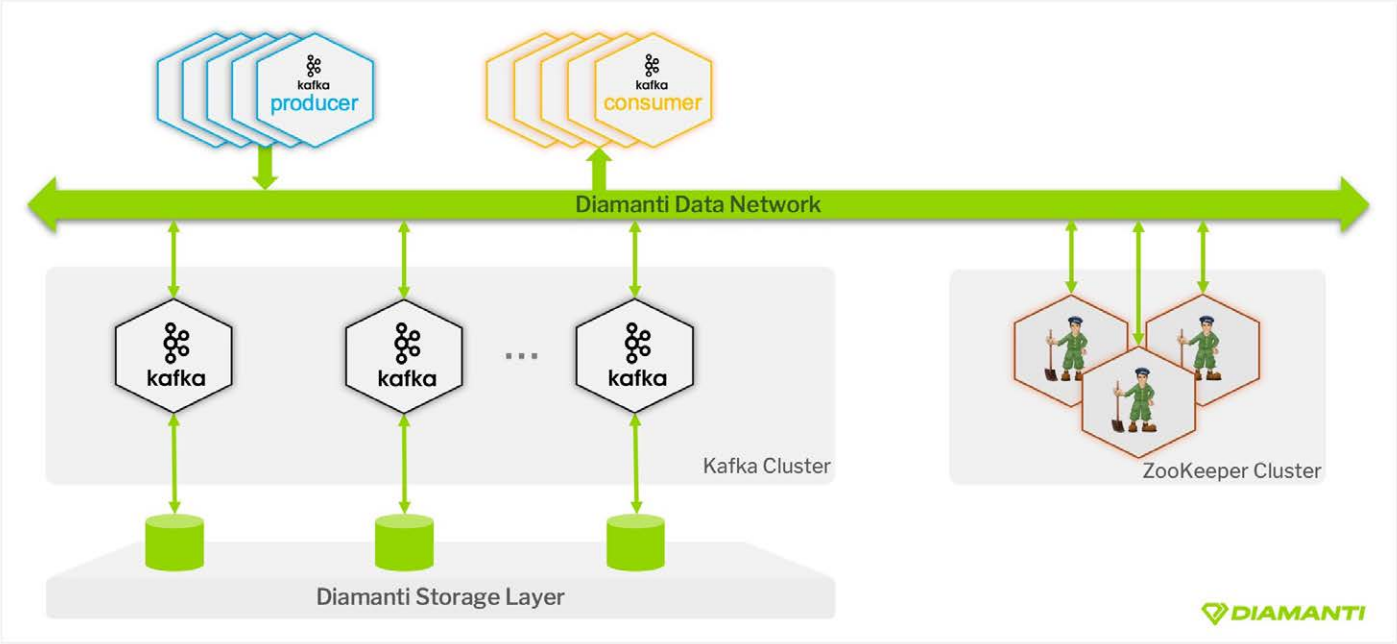


Figure 2: Apache Kafka on Diamanti Enterprise Kubernetes Platform

To benchmark Kafka, containerized Kafka Producer and Consumer clients were deployed on the Diamanti Enterprise Kubernetes Platform. In order to demonstrate the worst case performance, an out-of-the-box Kafka installation was used. Also, to mimic a real world scenario, a mix of replicated and non-replicated 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. First test scenario used two Producers and two Consumers to read and write 60 million records via non-replicated topics with 16 partitions. 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 100 byte message size.

With an untuned three node Kafka cluster running a mix of replicated and non-replicated topics on 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. Table 1 summarizes the results obtained for each Producer and Consumer. These results were achieved with 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.

Table 1: Apache Kafka on Diamanti with a Mix of Replicated and Non-replicated Topics

	PRODUCERS			CONSUMERS	
Topics	Messages per Second	Network Traffic (MB/sec)	Average Latency (ms)	Messages per Second	Network Traffic
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

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 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. Multiple Kafka clusters or other applications can be deployed on the same shared infrastructure without noisy neighbor problems resulting in very high host utilization and significantly reducing total cost of ownership (TCO).

ABOUT DIAMANTI

Diamanti delivers the industry's only purpose-built, fully integrated enterprise Kubernetes platform, spanning on-premises and public cloud environments. It gives infrastructure architects, IT operations, and application owners the performance, simplicity, security, and enterprise features they need to get cloud-native applications to market fast. Diamanti provides the lowest total cost of ownership to enterprise customers for their most demanding applications.

Based in San Jose, California, Diamanti is backed by venture investors ClearSky, CRV, Engineering Capital, Goldman Sachs, GSR Ventures, Northgate Capital, Threshold Ventures (formerly DFJ Venture), and Translink Capital. For more information visit www.diamanti.com or follow @DiamantiCom.

Rev. 110719

Diamanti, Inc. 111 N Market St., Suite 800, San Jose, CA 95113
(408) 645-5111