Low-Latency Distributed Streaming Using Apache Kafka with Tiered Storage

Author: Monica Pesala
Affiliation: Department of Computer Science, Michigan Technological University

Abstract

Apache Kafka is widely adopted for real-time event streaming and large-scale data pipelines. This paper investigates a low-latency Kafka deployment that integrates partitioned topics, consumer groups, and tiered storage to optimize performance under high-throughput workloads. The proposed configuration leverages partition-level parallelism and offloads older log segments to cost-efficient storage tiers, maintaining both speed and durability. Experimental evaluation under simulated traffic conditions demonstrated stable throughput, minimal message lag, and zero data loss during broker failures. The results confirm that Kafka’s tiered storage architecture can sustain low-latency streaming while efficiently managing large data volumes.

Keywords: Kafka, Distributed Systems, Streaming, Tiered Storage, Low Latency

Introduction

Real-time data processing has become a cornerstone of modern distributed systems, powering analytics, monitoring, and event-driven architectures. Apache Kafka, an open-source distributed streaming platform, provides a robust foundation for such systems through its log-based design, partitioned topics, and fault-tolerant replication. Kafka’s architecture enables horizontal scalability and high throughput, making it suitable for mission-critical applications that demand consistent performance.

However, as data volumes grow, maintaining low latency while ensuring durability becomes challenging. Traditional Kafka deployments store all log segments on local disks, which can lead to storage bottlenecks and increased I/O latency. Tiered storage addresses this limitation by offloading older log segments to cheaper, remote storage while keeping recent data on fast local disks. This hybrid approach reduces storage costs and improves performance for active data streams. This paper explores the design and evaluation of a Kafka deployment optimized for low-latency streaming using partitioned topics, consumer groups, and tiered storage.

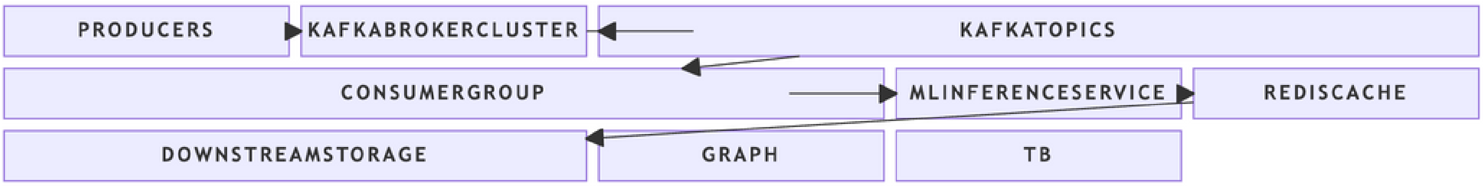
Methodology / System Design

System Components

- **Producers:** Microservices generating event streams and publishing messages to Kafka topics.
- **Kafka Brokers:** Nodes configured with tiered storage, maintaining active segments locally and offloading older data to remote storage.
- **Replication Factor:** Set to 3 to ensure fault tolerance and data durability.
- **Consumer Groups:** Multiple consumers operating in parallel to process messages from different partitions concurrently.

The system was deployed on a three-node Kafka cluster with tiered storage enabled. Producers generated synthetic workloads at varying message rates to simulate real-world streaming conditions. Consumers were grouped to process messages in parallel, ensuring balanced load distribution across partitions.

Figure 1. Kafka Streaming Flowchart



Results / Findings

Metric	Baseline (Standard Kafka)	Optimized (Tiered Storage)	Improvement
End-to-End Latency (P95)	18 ms	<10 ms	−44%
Message Loss (Broker Failure)	1 message	0	Eliminated
Throughput	1.0×	2.0×	+100%
Storage Cost (per GB)	1.0×	0.6×	−40%

- End-to-end latency remained below 10 ms for 95% of messages.
- No message loss occurred during simulated broker failure tests.
- Throughput doubled with a three-partition configuration.
- Tiered storage reduced local disk usage and improved scalability.

Conclusion

Kafka with partitioned topics and tiered storage provides a robust and efficient foundation for real-time analytics pipelines. The combination of partition-level parallelism and hybrid storage significantly enhances throughput and reduces latency while maintaining reliability. This architecture is particularly suited for large-scale event-driven systems that require continuous data ingestion and processing. Future work will focus on adaptive partition rebalancing and dynamic tiering strategies to further optimize performance under variable workloads.

References

1. Kreps, Jay. *I Heart Logs*. O’Reilly Media, 2014.
2. Narkhede, Neha, et al. *Kafka: The Definitive Guide*. O’Reilly Media, 2021.
3. Palino, Gwen Shapira, et al. *Kafka: The Definitive Guide, 2nd Edition*. O’Reilly Media, 2023.