DATATORRENT
Powered by Apache Apex

Speed your Transition to a Data-Driven Business
Faster, Reliable Data Ingestion and Analytics with Apache Apex and Apache Kafka

Data Drives Business Success

To succeed in today's data-driven economy, businesses must keep pace with rapidly, continuously evolving customer needs and market trends. Capturing opportunity in those trends—and transforming insights into business decisions that better meet customer needs—is key to your success.

To capture opportunity, businesses need to analyze large amounts of streaming data in real-time. And, timing is everything: ideally, all business-critical data need to be ingested as soon as it is generated and made available for analytics and decision-making. With Apache Apex's powerful data ingestion and analytics capabilities, you'll be positioned to harness data's power to drive decisions.

## Complementary Technologies, Highly Performant Solution

Together, Apache Apex and Apache Kafka give you powerful data ingestion and analytics capabilities. Apache Apex is a fast, scalable, and fault-tolerant computational platform that enables businesses to ingest and analyze moving data streams at scale without any data loss. Apache Kafka is a real-time, distributed, publish-subscribe messaging system used for processing messages at scale. Kafka stores metadata, such as topics and brokers, in Apache Zookeeper.

Apache Kafka and Apache Apex, used together, offer tremendous business potential: these complementary open-source technologies provide you a highly performant solution to transform streaming data into real-time insights. And, they offer such a solution with minimal developer effort and seamless integration using the Kafka input operator—along with a guarantee of zero data loss.

Apex-Kafka Integration
Seamless Integration for Powerful Ingestion and Insights

Kafka and Apex's seamless integration is made possible through an out-of-the-box Kafka input operator. Apex supports this input operator, which allows ingestion of data into Apex applications. The Kafka input operator consumes data from the Kafka messaging system, which is later processed in the Apex environment. The Kafka input operator consumes data from the Kafka broker and also interacts with the Zookeeper to monitor metadata change and adjust partitions accordingly at runtime. The result is a powerful solution for ingestion of streaming data and its translation into real-time insights.
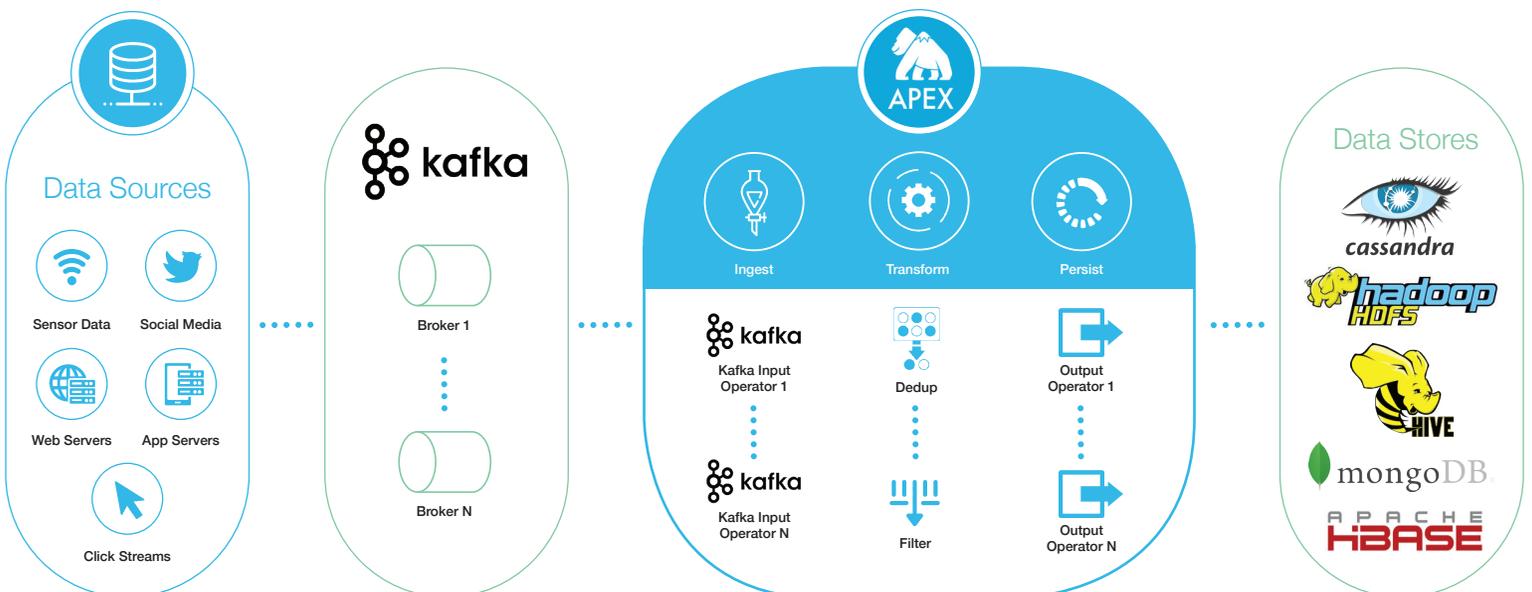
Unlock Powerful Data Capabilities with Apex-Kafka Integration

Kafka is a robust messaging system that can handle a large volume of data and enables you to process messages at scale. Pairing it with a real-time, production-ready, stream-processing platform helps you realize Kafka's full potential. Apache Apex perfectly fits the bill. Apex ensures dynamic scalability, high fault-tolerance, and low latency while guaranteeing zero data loss, enabling you to achieve real-time ingestion and analytics on moving data streams.

Fault Tolerance

Failure recovery is essential for the Kafka input operator. Apex makes failure recovery on the Kafka input operator easy, as the operator is natively supported by Apex. In the event of one instance going down, another is brought to life by the Apex engine. Other instances are not affected, and the operator mechanisms proceed without interruption. Additionally, the recovered operator can resume from the safe check-pointed offsets.

Zero Data Loss, Guaranteed

Kafka uses at-least-once delivery semantics to protect against data loss, meaning that it can sometimes deliver the same message multiple times to ensure successful delivery. Designed for exactly-once delivery semantics, the Kafka input operator in Apex guarantees that all data is delivered without any duplicates, thereby addressing a characteristic of Kafka that creates problems for some applications.

Dynamic Partitioning

Partitioning helps make the Kafka input operator scalable. A single-node consumer will likely encounter bottlenecks with multiple partitions for a Kafka topic. To prevent these bottlenecks, Apex supports dynamic partitioning, a key feature that allows the Kafka input operator to scale with Kafka partitions Two partition strategies are supported by default: one-to-one and one-to-many.

Better data, better business decisions with Apache Apex

Apache Apex gives you powerful data ingestion and analytics capabilities to harness data and insights for decision-making—helping you capture opportunity and all the upside that comes with it.

Imagine what you could achieve with real-time data-in-motion ingestion and analytics.

Download the Sandbox, Community, Enterprise, or Cloud version of the DataTorrent platform, and explore our difference today. Learn more about our products and solutions at www.datatorrent.com.