DATATORRENT
Powered by Apache Apex

Streaming OLTP Data to Hadoop using Oracle GoldenGate and DataTorrent RTS Enterprise Platform
Ingestion ready OLTP data

Modern businesses rely on data to make critical decisions in a real-time environment. It is increasingly become evident that businesses need to combine transaction data with unstructured data to gain holistic and deeper insights. As a result, enterprises are now looking to integrate OLTP data into Hadoop. However, enabling Hadoop in a real-time OLTP scenario is challenging because data extraction, transformation, and finally, integration is not an easy task.

Key Customer Challenges:

- Continuously and reliably ingest large volumes of heterogeneous data with very low latency to derive meaningful real-time insights

- Ingest data from and deliver to different types of systems including Kafka, HDFS, Hive, relational and NoSQL databases

- Prevent data loss and system downtime

DataTorrent Solution for OLTP Data Ingestion

DataTorrent RTS Enterprise Platform enables businesses to handle streaming data ingestion scenarios in a Hadoop environment. It works seamlessly with Oracle GoldenGate to extract, transform, and enable ingestion of moving data streams in real-time. Because DataTorrent RTS is designed to offer high throughput, fault-tolerance, and scalability it can handle in-memory processing of billions of events every second.
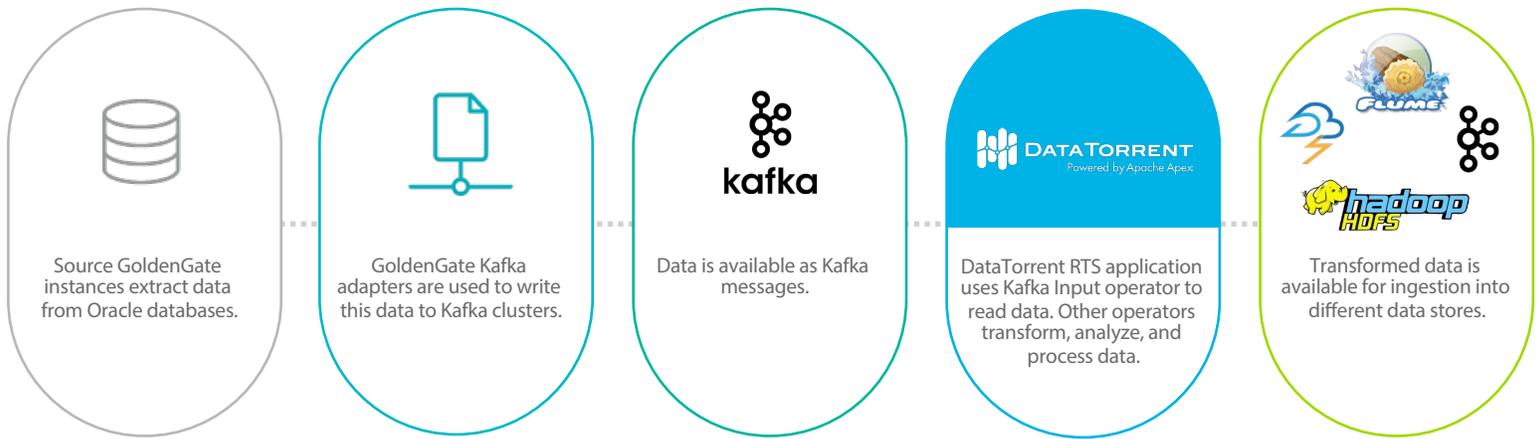
DataTorrent RTS is implemented as a YARN application for ingesting data from Oracle databases. Messaging systems such as Kafka, RabbitMQ on Oracle GoldenGate instances extract real-time data from Oracle databases and push this data to their respective clusters. The data from the messaging system is then parsed, transformed, and written to other data stores by using DataTorrent RTS applications.

Architecture

The DataTorrent-Oracle GoldenGate solution extends the existing GoldenGate configuration by appending the DataTorrent RTS application as a one-stop solution for transforming, processing, and writing data to different data stores. The below listed solution talks about the Datatorrent-Oracle GoldenGate configuration with Apache Kafka. However, DataTorrent is also compatible with a variety of other messaging brokers including RabbitMQ, ActiveMQ, and Amazon SQS.

Oracle GoldenGate Configuration with Apache Kafka

The default GoldenGate data pump functions as the "source pump" for extracting data from the Oracle database. The source data pump sends OLTP data to dedicated Kafka producer pumps, which eventually write the transactions to a single Kafka cluster. The dedicated Kafka producer pumps extend GoldenGate's out-of-the-box JAVA adapter support to create a lightweight Kafka handler, along with a producer. The producer and handler together write OLTP data in JSON format to the Kafka topics of a single Kafka cluster.

Source GoldenGate instances extract data from Oracle databases.	GoldenGate Kafka adapters are used to write this data to Kafka clusters.	Data is available as Kafka messages.	DataTorrent RTS application uses Kafka Input operator to read data. Other operators transform, analyze, and process data.	Transformed data is available for ingestion into different data stores.

The GoldenGate architecture is scalable, which means that there can be multiple instances of source data pumps for data extraction, along with multiple instances of dedicated Kafka producer pumps for representing the extracted data in JSON format. In a production environment, multiple instances of source data pumps will extract data from multiple instances of Oracle database. The dedicated pumps will write this data to Kafka topics of a single Kafka cluster.

DataTorrent RTS Configuration

The DataTorrent RTS application will enable data extracted from Oracle databases to be ingestion-ready for different types of data stores. The application is a directed acyclic graph (DAG) containing the Kafka Input operator, and operators for transforming, processing, and ingesting data. Dedicated GoldenGate pumps continuously write to Kafka topics of a single Kafka cluster. As soon as messages arrive in a topic, the Kafka Input operator consumes these messages in real time, and emits this data in a Java-compatible format. The output of the Kafka Input operator functions as input for the next operator, which typically performs transformations such as enrichment and deduplication. Finally, arrays of transformed data are written to different data stores.

Why DataTorrent
Get better data, faster insights

Seamless GoldenGate Integration

Requires minimal configuration to integrate the DataTorrent RTS application with the existing GoldenGate configuration and deploy in a small or large scale environment.

High Performance

Process billions of events per second with single-digit millisecond latency and without any performance degradation.

One Solution Fits All

Extract transaction data from Oracle databases and make it compatible for ingestion to different types of data stores such as HDFS and Hive with only one implementation of the DataTorrent RTS application and a single GoldenGate adapter.

Failure Recovery and Scalability

DataTorrent RTS application provides processing guarantees for scalability, fault-tolerance, and low latency while ensuring support for transformations such as parsing, deduplication, and enriching. The DataTorrent application also extends support for additional data refining, real-time alerting, and monitoring.

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

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