Transforming IIoT Data into Opportunity with Data Torrent using Apache Apex

DataTorrent delivers better business outcomes for customers using industrial of things (IIoT) data
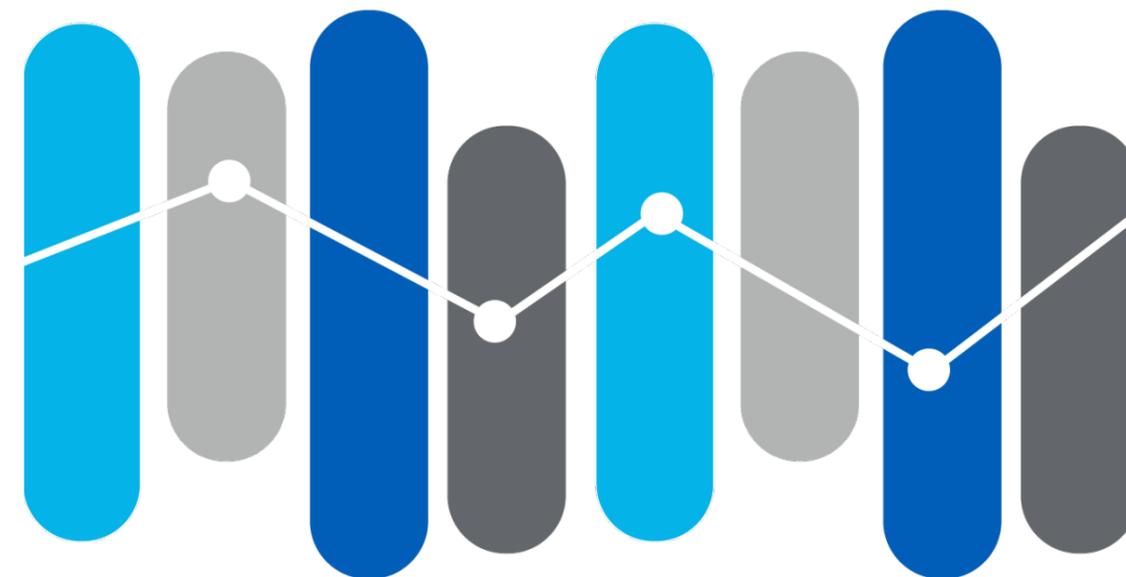
Challenge

The industrial internet revolution is transforming every industry, from energy and transportation to healthcare and aviation. Connected devices and sensor-enabled machines are proliferating rapidly. These devices include medical equipment, drill rigs, wind turbines, assembly line components, and aircraft engines, among more—and all generate data in real-time. **The challenge is how to ingest data from millions of connected machines and sensors in real-time with zero data loss.**

To solve this challenge, GE built a cloud-based analytics platform, Predix, to help organizations leverage data from industrial machines and use it to build industrial internet applications. These applications transform machine data into actionable insights— ultimately to build smart, agile businesses.

This deluge of data needs to be processed to manage and service these devices. Traditional technology stacks were not equipped to handle massive amounts of data in an automated way. In addition to automation, enterprises also needed to reduce the time to extract value from these systems and reduce the total cost of ownership in servicing the devices. Today the bulk of this data is simply dropped as legacy systems are not able to keep pace. The biggest problem occurs when the data is ingested, prepared, transformed, alerted on, and converted to structured data. This happens before traditional analytics are

applied. Ideally, you would want to analyze the data as soon as it is transformed and before it is loaded in databases. This will speed time to insights.
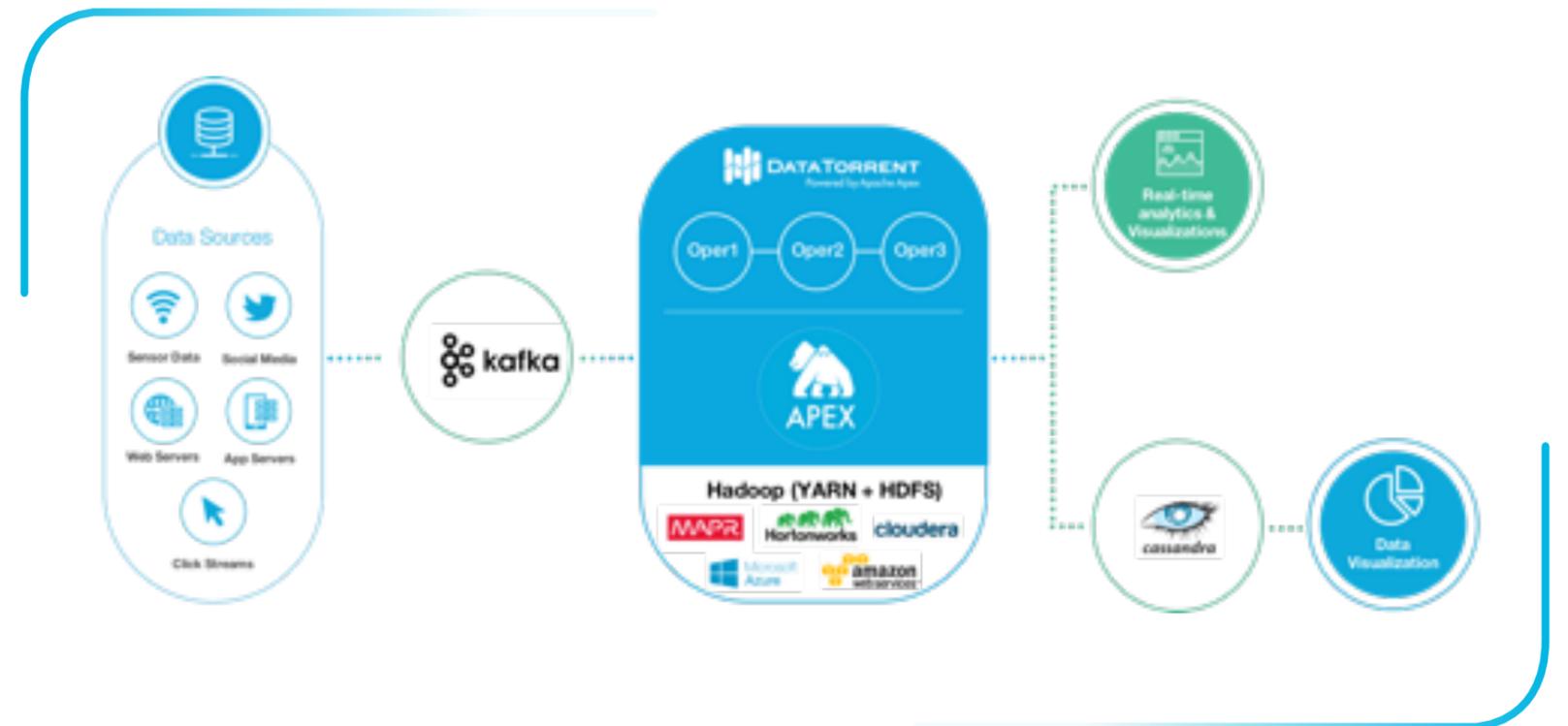
The applications need to scale linearly to handle the predicted growth in data results from the explosion of devices. The cost structure of managing this huge data also had to be viable. Therefore, the solution had to work both in the cloud and on-premise. The solution also needed to be operationalized in a way that would allow it to run in various countries, subject to different security laws. And finally, all of this needed to be achieved with available expertise, and reduced time to market.

Solution

The Time Series application consists of a pipeline that has Apache Kafka as the input source, Apache Apex as the real-time engine powered by DataTorrent, and Apache Cassandra as its final data store. HDFS is used to archive raw data. The input sources of GE Predix customers send their data to the message bus. Apex application ingests this data, cleans it and prepares it, does all the checks, and encryption, does transformations as needed and sends alerts if come errors are found.

After transformation, this data is loaded into Cassandra for analysis. The pipeline runs continuously on a 24x7 basis. DataTorrent Real Time Streaming (RTS) is used to monitor and manage the entire application. DataTorrent RTS consists of operators, tools, and metrics. Dashboards for all the metrics are part of DataTorrent RTS. Commodity hardware is utilized, thereby allowing for the cost structure to go down with Moore's law. The solution has no need for specialized hardware. The system metrics and logs from aggregation from DataTorrent RTS is integrated with DevOps tools, to ensure that they do not need to change their current operational technology.

Results

Leveraging DataTorrent with GE Predix delivers the following capabilities:

- Ingest and analyze high-volume, high-speed data from thousands of real-time data sources, including connected machines and sensors, with zero data loss.

- Ability to build industrial internet applications quickly and easily to lower barriers to entry, accelerate time-to-market, and drive innovation faster.

- Provide mission critical sensor alerts.

- Ability to scale automatically to accommodate any data processing needs with single-digit millisecond latency and high-throughput.

- Easily manage and monitor application performance and debug applications.

- Power predictive analytics with the right data at the right time to reduce costly maintenance and unplanned downtime, improve asset performance, and lower costs.

- Unify monitoring of all connected sensors and devices to minimize disruptions.

- Reduce total cost of ownership on development and operations.

About Data Torrent

DataTorrent is a fast data analytics applications company, focused on accelerating customer business outcomes through enterprise-grade, data-in-motion analytics that provide true real-time views of information for actionable insights.

DataTorrent improves time to value and lowers total cost of ownership by delivering pre-built, application pipelines that include everything needed for production from ingestion to visualization built on open-source innovations.

datatorrent.com

Real-time ingestion and analytics, zero data loss

Unified monitoring of all connected devices

Sub-millisecond latency—preserving data's fleeting value

About GE

GE (NYSE: GE) is the world's Digital Industrial Company, transforming industry with software-defined machines and solutions that are connected, responsive and predictive. GE is organized around a global exchange of knowledge, the "GE Store," through which each business shares and accesses the same technology, markets, structure and intellect. Each invention further fuels innovation and application across our industrial sectors. With people, services, technology and scale, GE delivers better outcomes for customers by speaking the language of industry.

Contact Information

Americas: 1-855-YOUR1GE (1-855-968-7143)

gedigital@ge.com

www.ge.com/digital