influxdata®
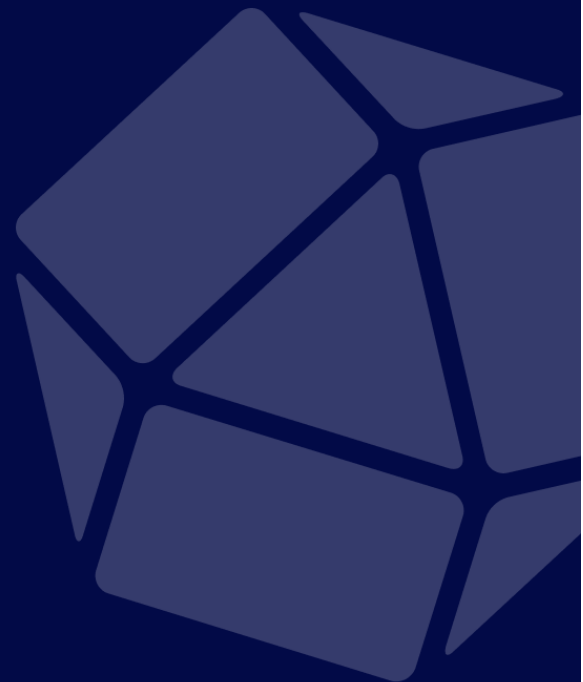
How to Build a Streaming Pipeline for Massive IoT Flows with Lenses.io and InfluxDB

Stefan Bocutiu

Head of Engineering, Lenses.io

LENSES

Company in brief

Landoop, creator of Lenses, brings the fastest and simplest ways for companies to get complete control of their data and transform the way teams collaborate to access, process, and analyze their data streams.

Lenses is a streaming data management platform for Apache Kafka. Lenses supports the core elements of Kafka with a web user interface and vital enterprise capabilities that enable engineering and data teams to query real-time data and to create and monitor Kafka topologies with rich integrations to other systems. Lenses Box, Landoop's free offering for the Lenses platform, provides the entire Apache Kafka ecosystem in one Docker command.

Landoop provides enterprise- grade software products, has also contributed more than 35 open source components related to Apache Kafka technologies, and is continuously open sourcing parts of its work. Landoop is a strong supporter of Apache Kafka — as well as Cloud Native Computing Foundation projects Kubernetes & Prometheus.

Case overview

Landoop wanted to design a platform, using Apache Kafka®, to enable teams to build an end-to-end data streaming pipeline for massive IoT data flows. In the IoT space, companies face challenges collecting and analyzing the ever-increasing amount of IoT data, and extracting value for increasing data volumes, to improve service quality and speed.

Landoop created Lenses: a streaming data management platform that uses the Kafka ecosystem to orchestrate the cornerstones of IoT data platforms (MQTT, a protocol optimized for connected devices and the M2M protocol, a protocol to remotely perform service enablement and application), analyze device data on the fly, and persist the results into an InfluxDB time series database.

Using InfluxDB, Landoop was able to create the easiest and fastest way for teams to gain complete control over their streaming data with Apache Kafka as well as create and monitor Kafka topology visualizations showing the entire data flows landscape in one interactive graph.
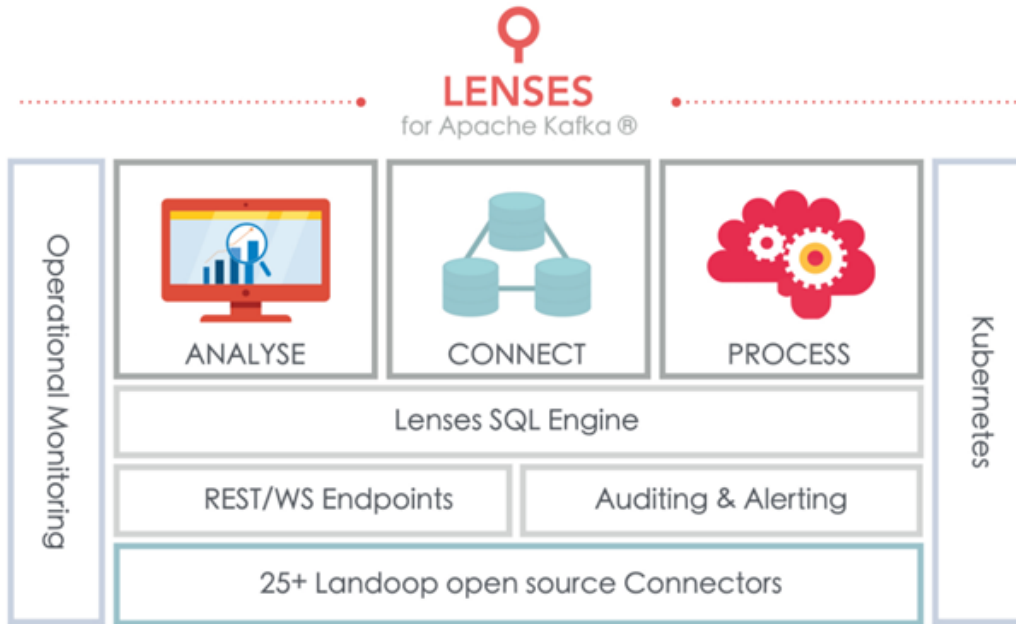
*Landoop's Streaming Data Management Platform for
building and deploying IoT data streaming pipelines in minutes*

> *"Our solution leverages Apache Kafka, Kafka Connect and
> Kafka Streams, Lenses SQL, and the popular choice for
> storage, the time series database in the space of IoT,
> InfluxDB."*
>
> **Stefan Bocutiu,** *head of data engineering*

The business problem

In an ever-more connected world, more data means more complexity. The rapidly growing number of interconnected devices confirms the Internet of Things (IoT) is a fast-maturing technology. Yet the digital economy's currency—data—is valuable only if you can use it. While the IoT is a driver for being data-rich, you need to be able to analyze the data and take appropriate timely action.

Proceeding from that, Landoop sought to build a data integration platform on top of Apache Kafka (a distributed streaming platform) and Kubernetes (a powerful container orchestrator) that connects, reshapes, and delivers actionable data. Landoop wanted their platform to enable engineering and data teams to seamlessly collect and perform real-time, at-scale analytics on the IoT generated data.

"Using the right tools enables you to build IoT streaming pipelines in a couple of minutes and allows you to focus on extracting value out of your data rather than building infrastructure."

The technical problem

Landoop realized that for IoT data flows, real-time data ingestion plus immediate queries equals actionable/timely insights. They wanted to use Apache Kafka to build a platform that helps developers boost productivity by providing the tools to build end-to-end IoT data pipelines with just a few lines of SQL code. Apache Kafka is an open source streaming framework with messaging semantics and constitutes a good fit for building IoT systems, particularly when leveraging Kafka Streams API (for processing and analyzing data in motion) and Kafka Connect (to ingest data into Kafka reliably and with fault tolerance).

Landoop wanted their platform to:
- Make it easy to write a stream processing application without having to be an Apache Kafka expert or a JVM developer
- Provide running, monitoring and scaling Kafka Streams apps out of the box
- Enable visibility for customers into their entire Apache Kafka-native data pipelines in one interactive graph
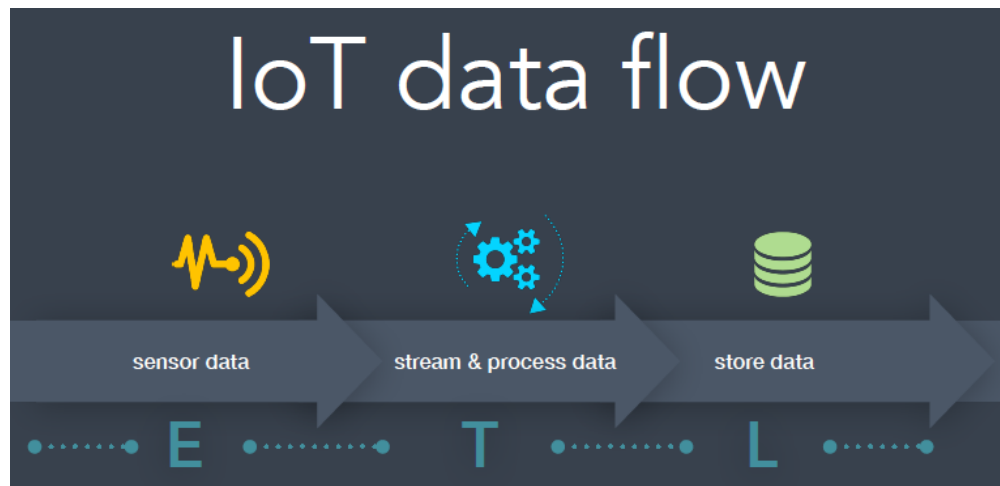
A typical IoT data flow involves a vast network of sensors publishing data periodically. That flux of information is then processed on the fly to extract patterns, trends, or execute predictions. Those end

results and the initial data might be stored for further analysis. As IoT data is times series data, and IoT use cases involve large amounts of time-stamped device and sensor data typically arriving at infrequent intervals and massive volumes, Landoop's platform required a scalable time series database to store streaming data in real time.

The solution

> *"InfluxDB Sink Connector is by far the most used connector of all of our Kafka Sink Connectors."*

Data from Landoop's community and Google Analytics showed that InfluxDB is quite popular; add to this that Landoop's InfluxDB Sink is the most used connector out of all of their 25+ Kafka sink and source connectors, and so they identified that InfluxDB, an open source time series database able to handle high write and query loads, is an optimal solution for the IoT world. Written in Go, it can handle large amounts of data such as application metrics, IoT sensor data, and real-time analytics. Landoop also liked the ease of API client libraries in InfluxDB.

Using InfluxDB to meet its storage needs, Landoop built Lenses to be an enterprise and scalable streaming platform that supports the core elements of Kafka with a rich user interface, endpoints and vital enterprise capabilities that enable engineering and data teams to query real time data and to create and monitor Kafka topologies with rich integrations with other systems. Lenses was built and operated in production systems in various industries (from finance to IoT to retail).

Lenses enables users to:
- **Analyze** - Explore data in motion with point-in-time and ad-hoc real time queries with full SQL support
- **Connect** - Connect to any data store with over 25+ connectors available with SQL and monitoring support
- **Process** - Register SQL Stream Processors, visualize topologies, scale via cluster mode and Kubernetes

Technical architecture

> *"Having the data in InfluxDB is great because with the TICK Stack that InfluxData provides, you can set up your rich dashboards."*

Apache Kafka's Connect framework allows moving data in and out of Kafka reliably and at scale. Kafka connectors are included in the Lenses platform. The Lenses Docker image available with Lenses Box comes with the Stream Reactor component: Landoop's opensource collection of Kafka Connect sink connectors (to bring data from Kafka into other systems) and source connectors (to publish data into Kafka). Kafka connectors cover the major IoT Protocols.

Stream Reactor Includes Kafka Sink and Source Connectors
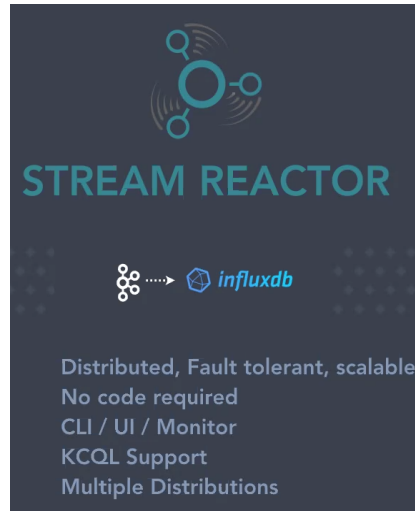
InfluxDB Deployment at Landoop
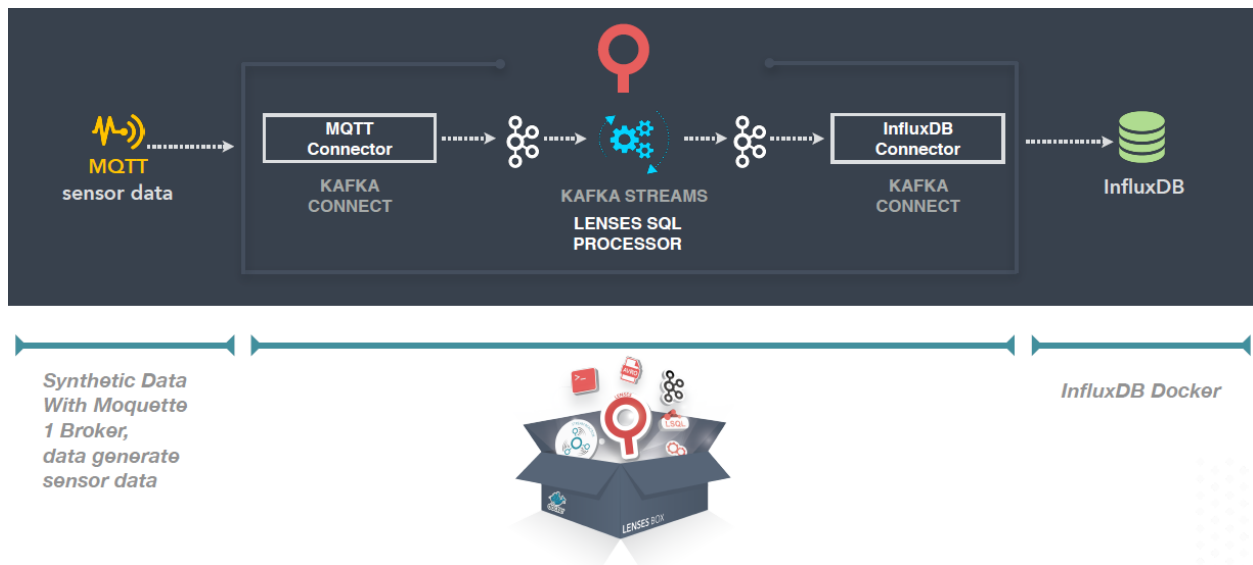
IoT data in Lenses takes the following trajectory:

- Device data probes Apache Kafka via Connect.
- The MQTT source connector ingests the sensor data into Kafka. MQTT, a very simple and lightweight publish-subscribe messaging protocol designed for low bandwidth, high-latency, and unreliable networks, is ideal for IoT data.
- The connector subscribes to the MQTT topic and writes the data it receives to a Kafka topic.
- Kafka, run as a cluster, stores streams of records in categories called topics. Each record consists of a key, a value, and a timestamp.
- To run stream analytics, Lenses SQL engine for Apache Kafka is used to declare, deploy, scale and monitor SQL processors. For each processor, you get the visual topology that shows how the data is being manipulated and transformed.
- Landoop's InfluxDB Sink Connector allows writing data from Kafka into InfluxDB—the connector takes the value from Kafka Connect SinkRecords and inserts a new entry to InfluxDB.
- The InfluxDB Sink supports Kafka Connect Query Language (KCQL), an open source component of Lenses SQL Engine that provides an elegant and simple SQL-like syntax for selecting fields and routing from sources or topics to Kafka or the target system.
- To complete the data pipeline, the Kafka Connect InfluxDB sink will persist the stream analysis result.

Sample Task Configuration

```
connector.class=com.datamountaineer.streamreactor.connect.influx.InfluxSinkConnector
connect.influx.db=iot
tasks.max=1
topics=sensor_data_avg
connect.influx.kcql=INSERT INTO sensorMeasure SELECT * FROM sensor_data_avg WITHTAG (id, poweredBy=Lenses)
connect.influx.username=""
connect.influx.url=http://localhost:8086
key.converter.schemas.enable=false
task.class=com.datamountaineer.streamreactor.connect.influx.InfluxSinkTask
value.converter.schemas.enable=false
name=iot-influxdb-sink
value.converter=org.apache.kafka.connect.json.JsonConverter
key.converter=org.apache.kafka.connect.json.JsonConverter
```

- The InfluxDB Sink Connector comes with full support for AVRO and JSON payloads in Kafka and also allows setting up various error policies (such as Noop, Retry, and Throw options).
- Through the Lenses UI, users can see and manage the entire end-to-end topology and visualize their Kafka-based data flows in real time. This is not a static graph but one in which they can interact with any nodes in the topology, for a 360 view of their data.

Visualizing the Journey IoT Data Takes to InfluxDB

Lenses also comes with the Javascript library to tap into database data as it is computed and present it in your Web application. It allows you to connect to Kafka and get live data into your browser leveraging Redux and LSQL.

Results

> *"MQTT. Kafka. InfluxDB. SQL. IoT Harmony."*

With Lenses, Landoop is enabling teams to build, deploy and govern real-time Kafka-based IoT pipelines in minutes—users can stream, analyze, and react in real time to extract actionable value from their IoT data and inform business processes and outcomes. Utility and power companies are already using Lenses to track various parameters of their machines at nanosecond precision to gain valuable real-time business insights. Lenses is also being deployed across a variety of market sectors and use cases, from tracking smart home devices to tracking farming vehicles.

Landoop's Lenses platform marks a series of "firsts" as it is the first to provide:
- Enterprise-grade SQL Engine over Kafka
- Native Kubernetes integration for streaming apps
- Topology visualizations over Kafka

A best-in-class all-in-one platform for Apache Kafka, Lenses enriches the Kafka ecosystem by providing the tools to easily and quickly set up a data pipeline. Using the Lenses platform, companies are delivering IoT streaming solutions a lot faster and a lot easier. With its real-time processing and monitoring, enterprise-grade security, scalability and ease of integration through Kafka connectors, Lenses accelerates time to value.

Through Lenses, teams streaming IoT data can:

- Demystify complexity and focus on what's important
- Optimize engineering resources
- Gain robust, multi-level production tooling for extreme resilience, at scale

Using InfluxData, Landoop is fulfilling its mission of "Apache Kafka Made Easy".

About InfluxData

InfluxData is the creator of InfluxDB, the leading time series platform. We empower developers and organizations, such as Cisco, IBM, Lego, Siemens, and Tesla, to build transformative IoT, analytics and monitoring applications. Our technology is purpose-built to handle the massive volumes of time-stamped data produced by sensors, applications and computer infrastructure. Easy to start and scale, InfluxDB gives developers time to focus on the features and functionalities that give their apps a competitive edge. InfluxData is headquartered in San Francisco, with a workforce distributed throughout the U.S. and across Europe. For more information, visit influxdata.com and follow us @InfluxDB.