CLOUDERA

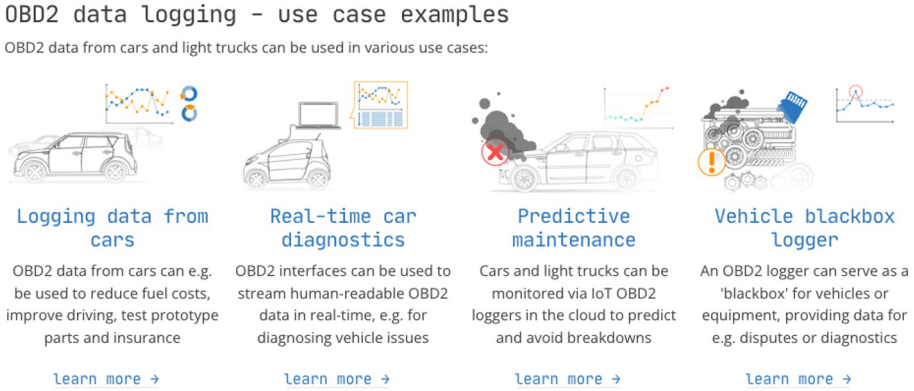
Streaming Demo

Streaming data: The Story

Any kind of data is produced as a stream of events.

Credit card transactions, sensor measurements, machine logs, or user interactions on a website or mobile application, all of these data are generated as a stream.

In our use case …..



© 2019 Cloudera, Inc. All rights reserved.

DRAFT / WIP

Useful data

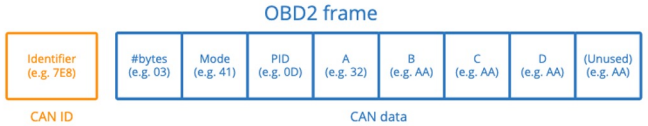
Need useful / proper data …

Let's use OBD2

Link

Raw OBD2 frame details

To get started recording OBD2 data, it is helpful to understand the basics of the raw OBD2 message structure. In simplified terms, an OBD2 message is comprised of an **identifier** and **data**. Further, the data is split in Mode, PID and data bytes (A, B, C, D) as below.



OBD2 message fields explained

Identifier: For OBD2 messages, the identifier is standard 11-bit and used to distinguish between "request messages" (ID 7DF) and "response messages" (ID 7E8 to 7EF). Note that 7E8 will typically be where the main engine or ECU responds at.

Length: This simply reflects the length in number of bytes of the remaining data (03 to 06). For the Vehicle Speed example, it is 02 for the request (since only 01 and 0D follow), while for the response it is 03 as both 41, 0D and 32 follow.

Mode: For requests, this will be between 01-0A. For responses the 0 is replaced by 4 (i.e. 41, 42, … , 4A). There are 10 modes as described in the SAE J1979 OBD2 standard. Mode 1 shows Current Data and is e.g. used for looking at real-time vehicle speed, RPM etc. Other modes are used to e.g. show or clear stored diagnostic trouble codes and show freeze frame data.

PID: For each mode, a list of standard OBD2 PIDs exist - e.g. in Mode 01, PID 0D is Vehicle Speed. For the full list, check out our OBD2 PID overview. Each PID has a description and some have a specified min/max and conversion formula.

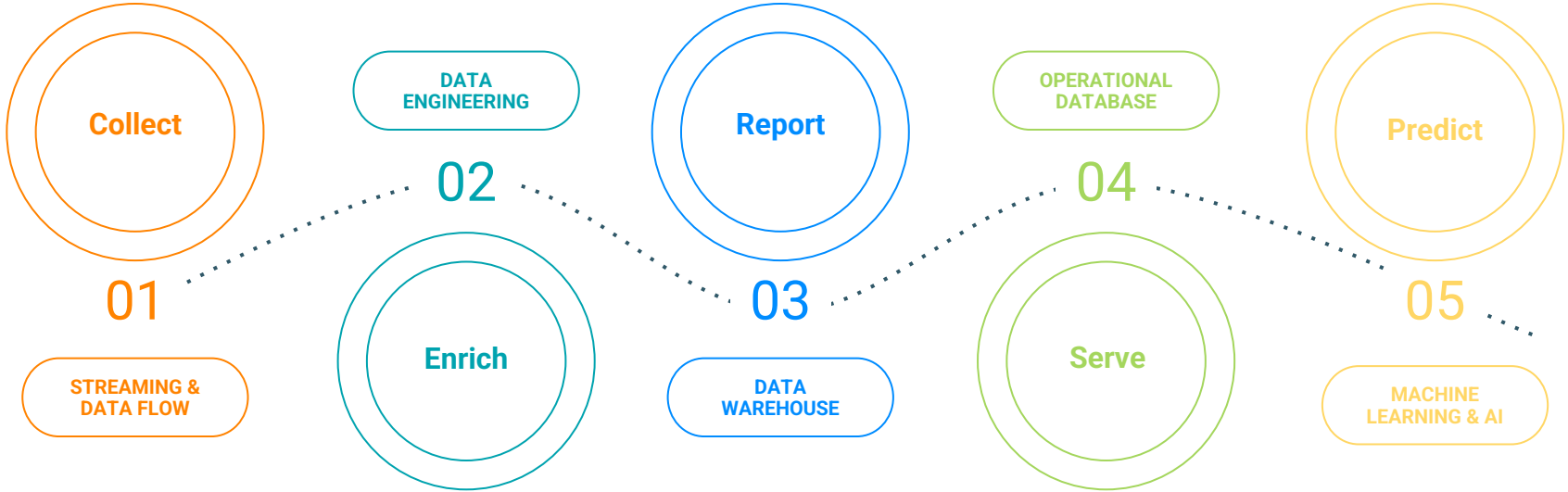
The formula for speed is e.g. simply A, meaning that the A data byte (which is in HEX) is converted to decimal to get the km/h converted value (i.e. 32 becomes 50 km/h above). For e.g. RPM (PID 0C), the formula is (256*A + B) / 4.

A, B, C, D: These are the data bytes in HEX, which need to be converted to decimal form before they are used in the PID formula calculations. Note that the last data byte (after Dh) is not used.

© 2019 Cloudera, Inc. All rights reserved.

Streaming Data Scenario: Zero Downtime

IoT for predictive maintenance



© 2019 Cloudera, Inc. All rights reserved.

DEMO FLOW

topic:
iot

topic:
kafka_LookupWeatherCondition

JSON message
> **append only**

CSV message
> **update / Key: StationId**

Kudu table:
RefData_GeoLocation

Virtual Table:
iot_enriched_source

Virtual Table:
weather_condition_upsert

Virtual Table:
refdata_geolocation

JOIN

Data Visualization

Apache Flink

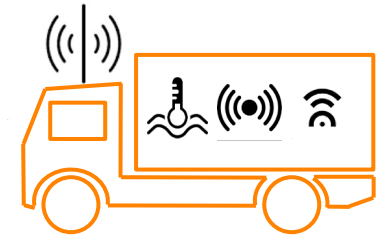
Virtual Table:
sensors_joined

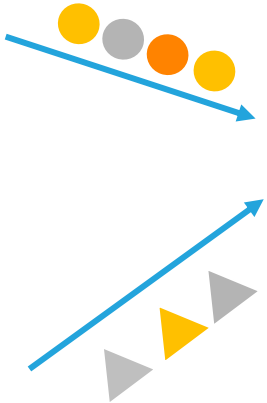
Kudu table:
sensors_joined

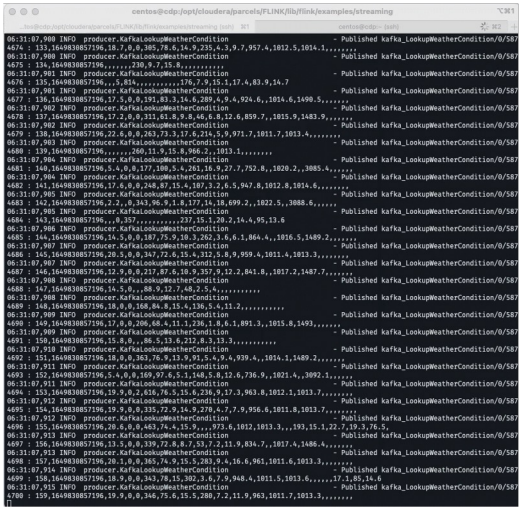
© 2019 Cloudera, Inc. All rights reserved.

Collection Phase

Collect data from the edge





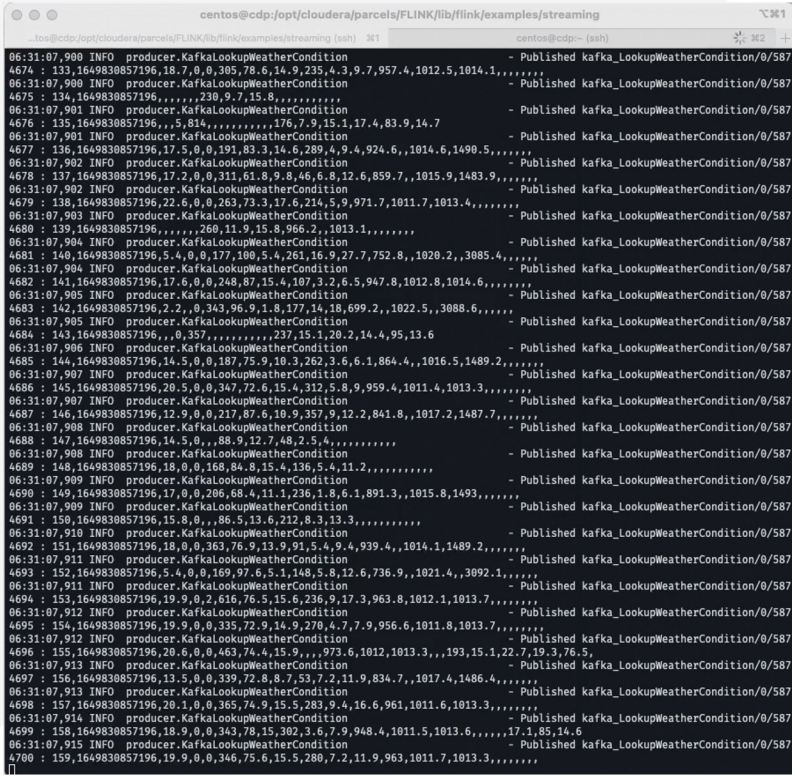


© 2019 Cloudera, Inc. All rights reserved.

Kafka Inputs

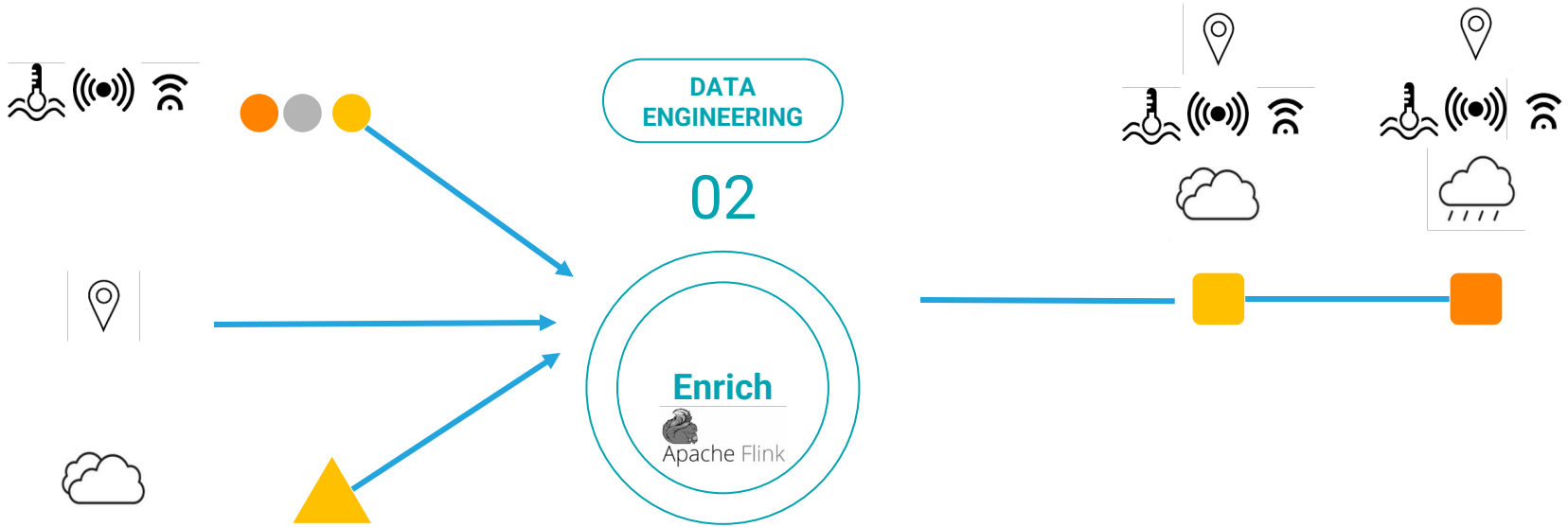
```
- Published iot/8/8469 :
{"sensor_ts":1649831184077,"truck_id":12,"sensor_0
":351,"sensor_1":5,"sensor_2":14,"sensor_3":12,"se
nsor_4":22,"sensor_5":17,"sensor_6":61,"sensor_7":
66,"sensor_8":78,"sensor_9":17,"sensor_10":308,"se
nsor_11":708}
```

```
- Published kafka_LookupWeatherCondition/12/5843963
:
109,1649830857196,15.7,0,0,318,94.5,14.8,176,3.2,6.1
,919.4,,1015.3,1490.8,,,,,,,
```



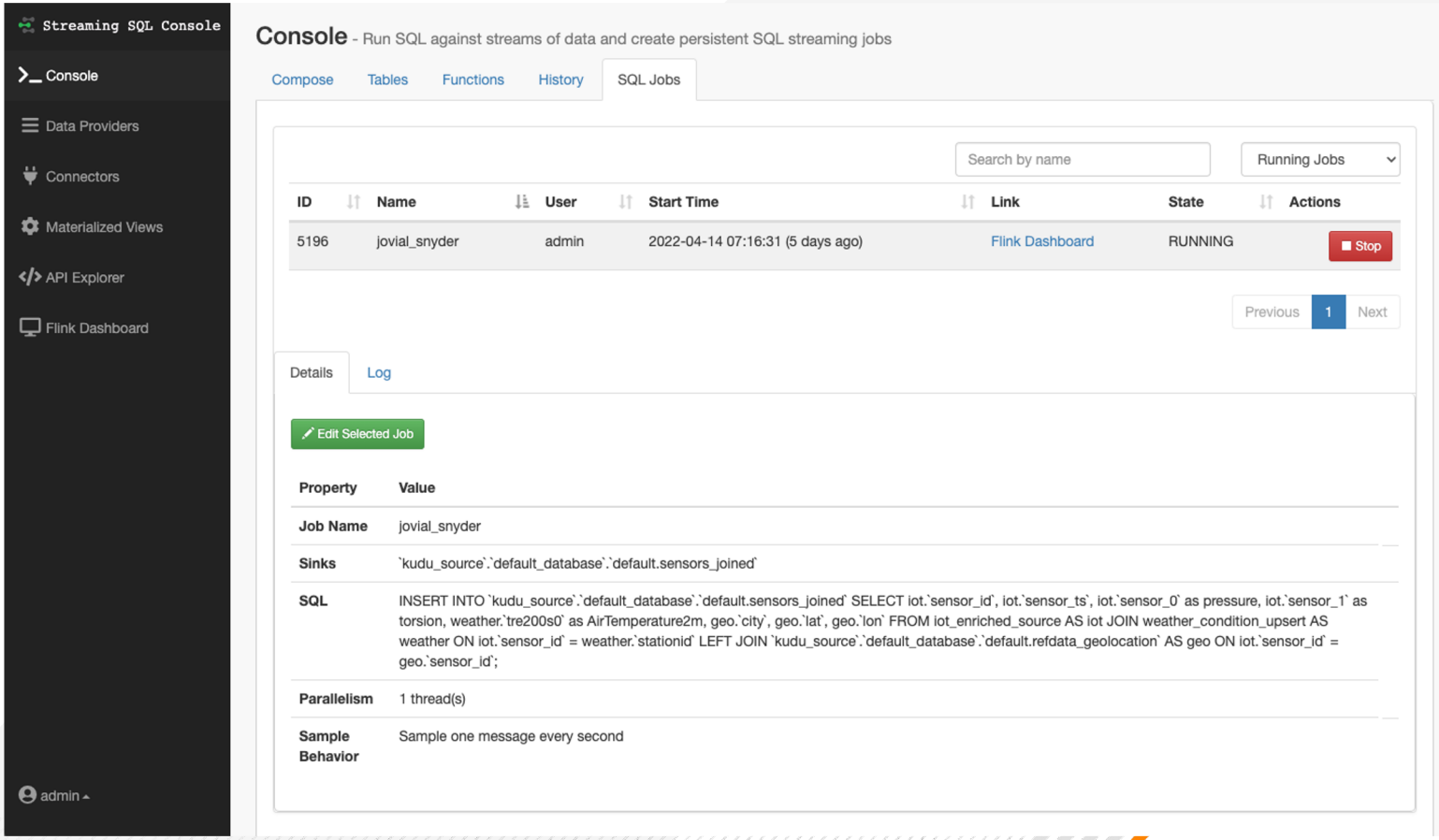
© 2019 Cloudera, Inc. All rights reserved.

Enrich the data



© 2019 Cloudera, Inc. All rights reserved.

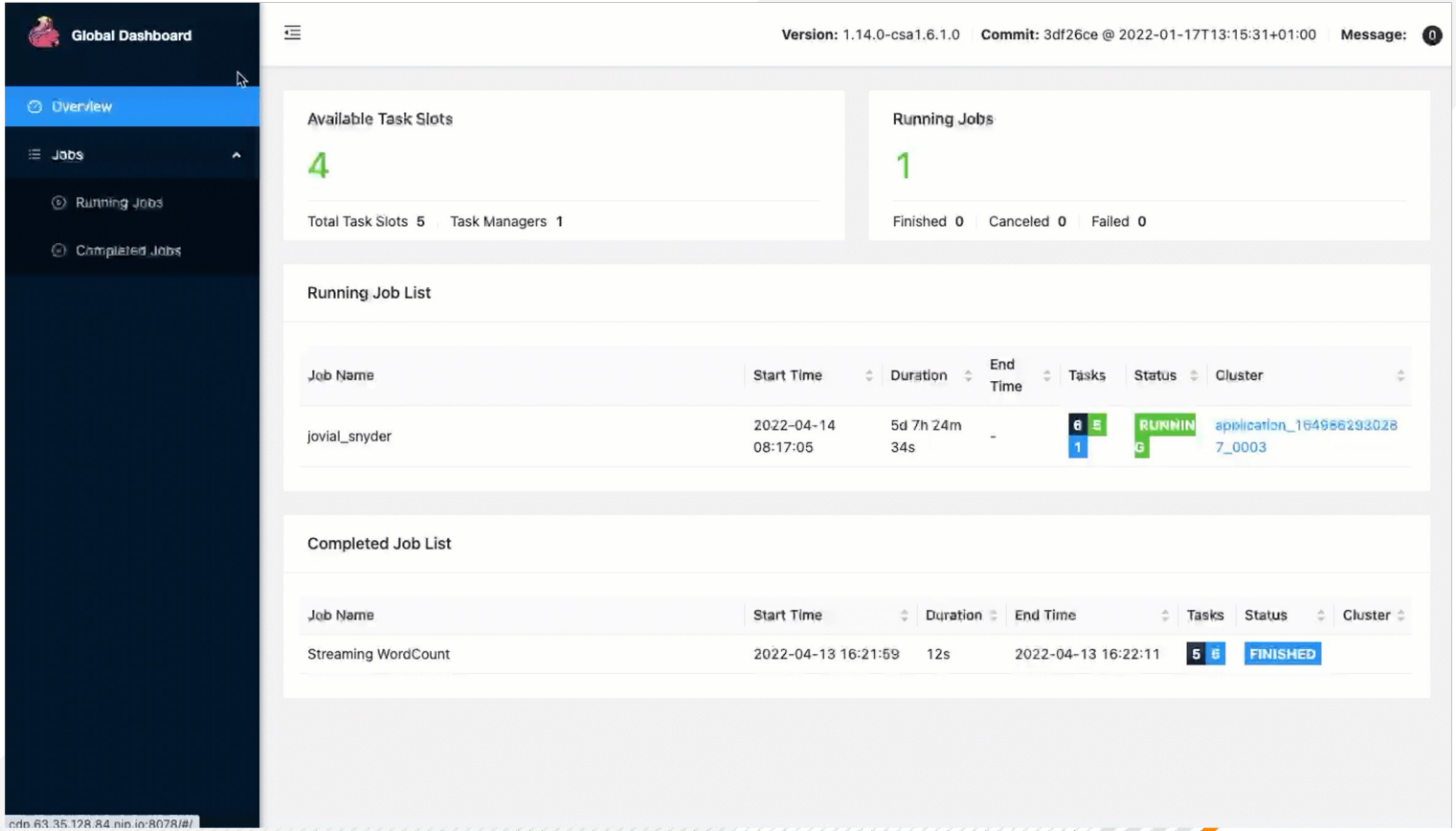
SSB



Link

© 2019 Cloudera, Inc. All rights reserved.

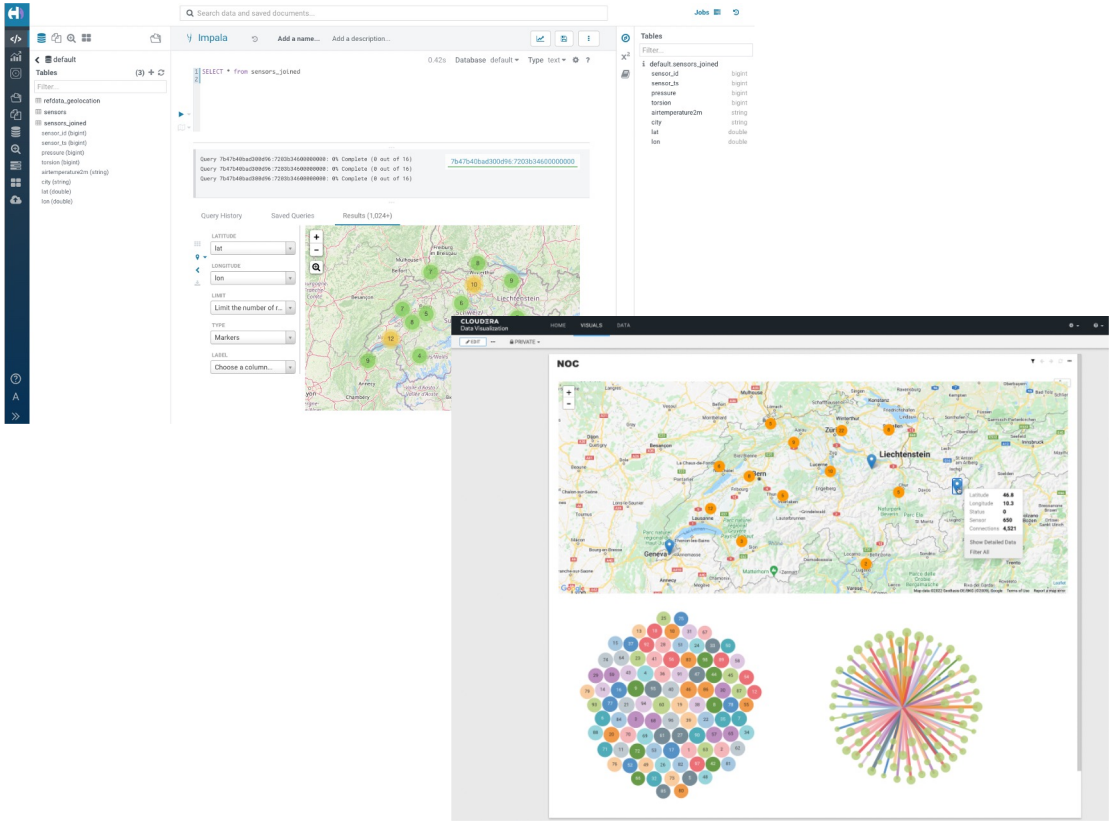
Flink



Link

© 2019 Cloudera, Inc. All rights reserved.

Report



Apache IMPALA

APACHE KUDU

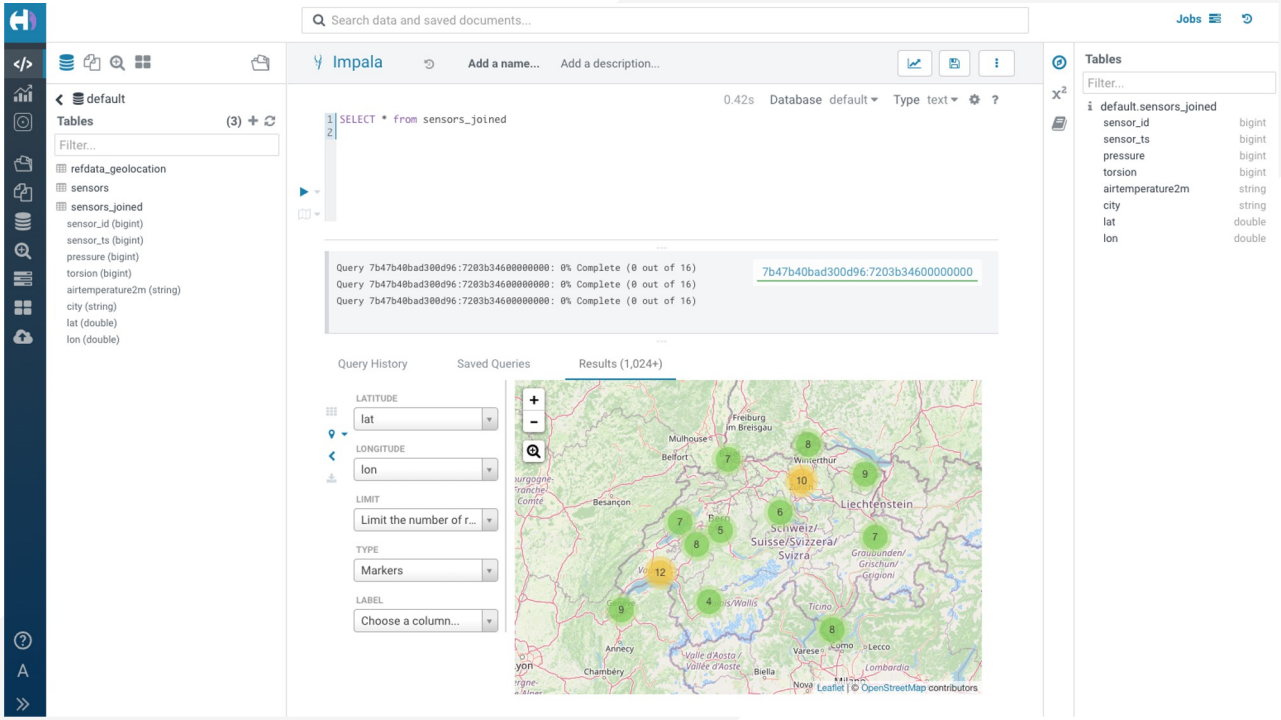
Report

03

DATA WAREHOUSE

© 2019 Cloudera, Inc. All rights reserved.

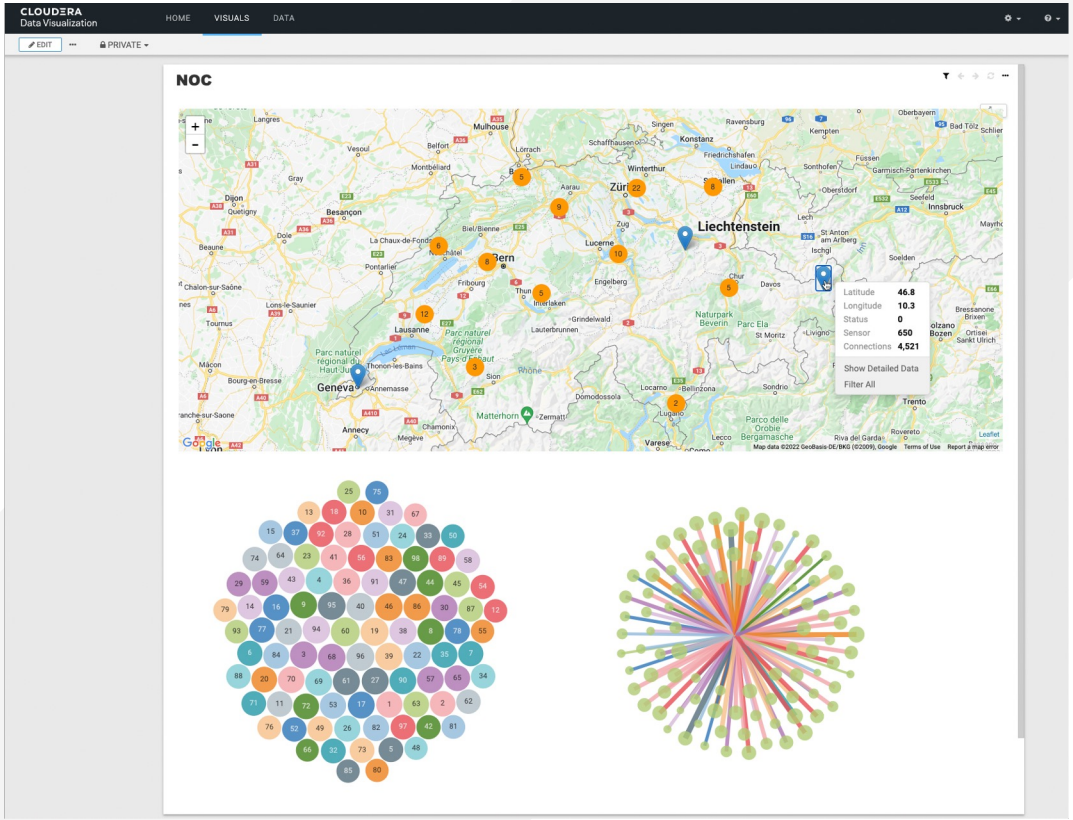
Impala + Kudu



Link

© 2019 Cloudera, Inc. All rights reserved.

Data Visualization and Beyond

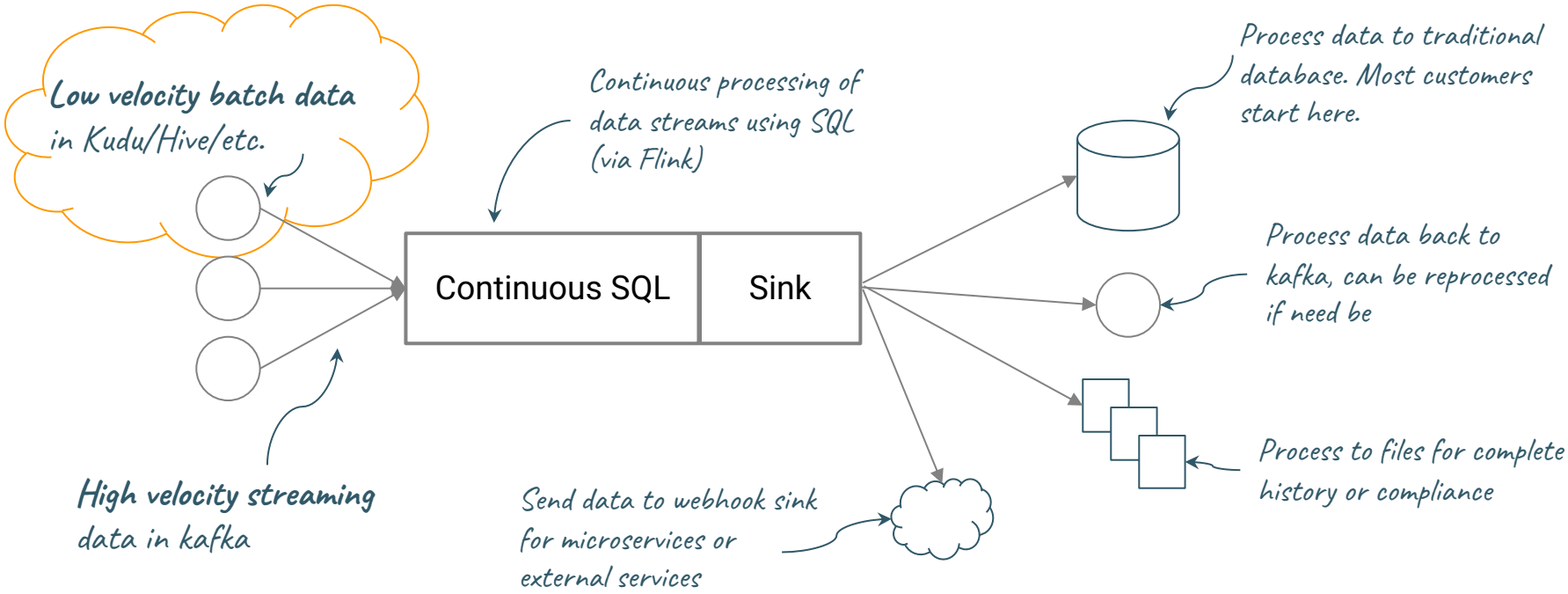


Link

© 2019 Cloudera, Inc. All rights reserved.

SQL Stream Builder

Continuous SQL allows for processing data using simple grammar.



© 2019 Cloudera, Inc. All rights reserved.