Hortonworks DataFlow Overview

Date of Publish: 2020-04-28

Contents

Overview

To build real-world data in motion apps such as those based on the Internet of Things (IoT), you need both flow management and stream processing capabilities. What is the difference between the two?

Data in motion apps typically have the following key requirements:

- Acquisition of Data from data sources within the data center, and across cloud environments and edge devices.
- Moving and Filtering of Data from edge devices (such as telematic panels on trucks), and across cloud environments and core data centers.
- Intelligent and Dynamic Routing of Data across regional data centers to core processing data centers.
- Delivering Data to different downstream systems.
- Joining and Splitting Streams of Data as they move.
- Detecting complex patterns in the streams of data.
- Scoring/Executing Analytics Models within the stream.
- Creating Custom Dashboards to visualize and analyze the streams and insights.

To explain how flow management and stream processing relate to these requirements, we employ a fictitious use case for trucking company X, which installed sensors on its fleet of trucks. These sensors emit streams of event data such as speed, braking frequency, and geo-code location. In this use case, the trucking company is building an IoT trucking app that monitors trucks in real time.

The following diagram illustrates how each of these requirements would be implemented in the context of stream processing and flow management:
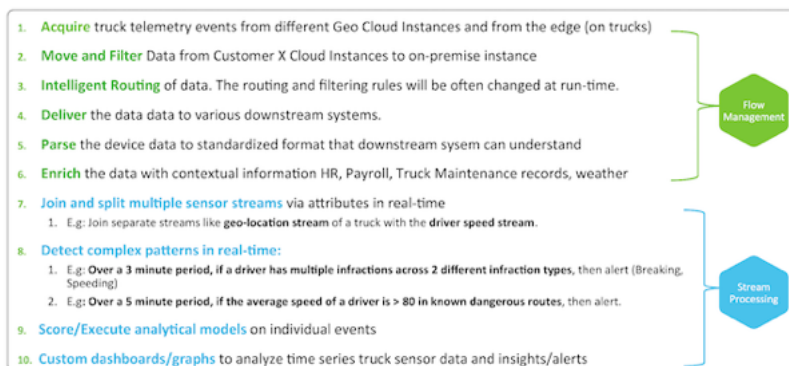
As part of the stream processing suite available in HDF, Streaming Analytics Manager provides capabilities for implementing the requirements outlined in blue in the previous diagram.

To summarize, Streaming Analytics Manager provides the following core capabilities:

- Building stream apps, using the following primitives:
    - Connecting to streams
    - Joining streams
    - Forking streams
    - Aggregating over windows
    - Extensibility: adding custom processors and user-defined-functions (UDFs)
    - Stream analytics: descriptive, predictive, and prescriptive
    - Rules engine
    - Transformations
    - Filtering and routing
    - Notifications and alerts
- Deploying stream apps:

- Deploying the stream app on a supported streaming engine:
- Monitoring the stream app with application-specific metrics.
- Exploring and analyzing streaming data; discovering insights:

  - Creating dashboards of streaming data
  - Exploring streaming data
  - Creating streaming cubes