Hortonworks DataFlow

> IoT and Real-Time processing with Open Source technologies

Pedro Algaba
Raúl Marín
Solutions Engineering @ Hortonworks
March 13th, 2018

FUTURE OF DATA

HORTONWORKS®
POWERING THE FUTURE OF DATA™

Contents

- All data needs to be under management: DataPlane Services
- Understanding a Streaming Data Architecture
- IoT Demo: Trucking company with a fleet of international trucks

HORTONWORKS

Contents

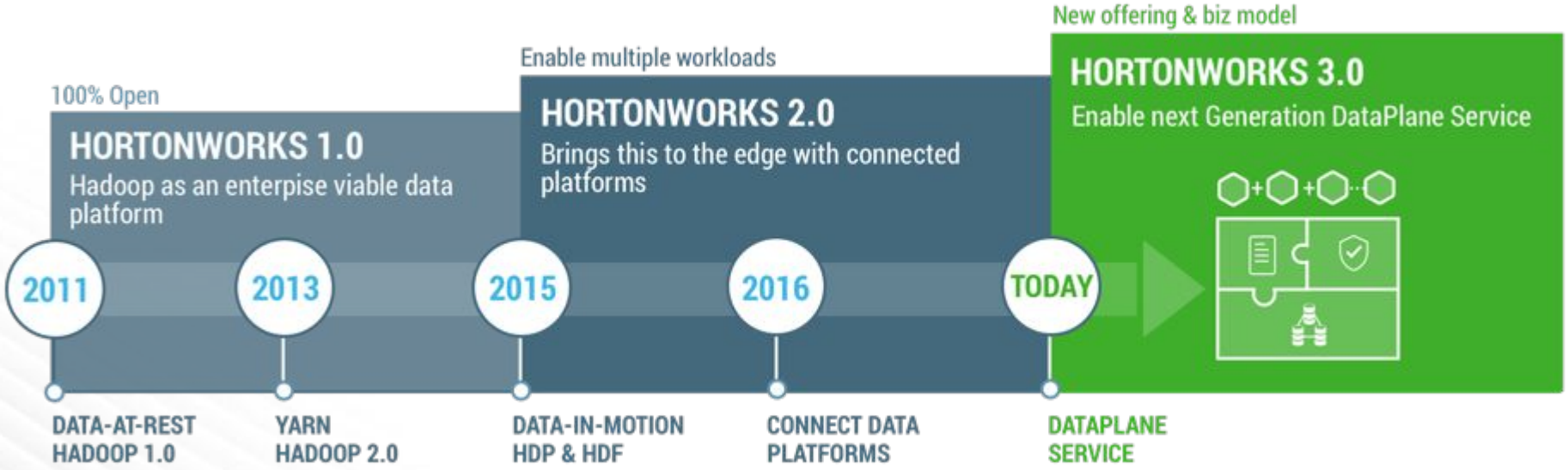
- All data needs to be under management: DataPlane Services
- Understanding a Streaming Data Architecture
- IoT Demo: Trucking company with a fleet of international trucks

Hortonworks: Enabling the Modern Data Architecture

Our mission continues…

- Make Hadoop an enterprise viable data platform
- Bring all data under management – all sources and types
- Extend to Global Data Management

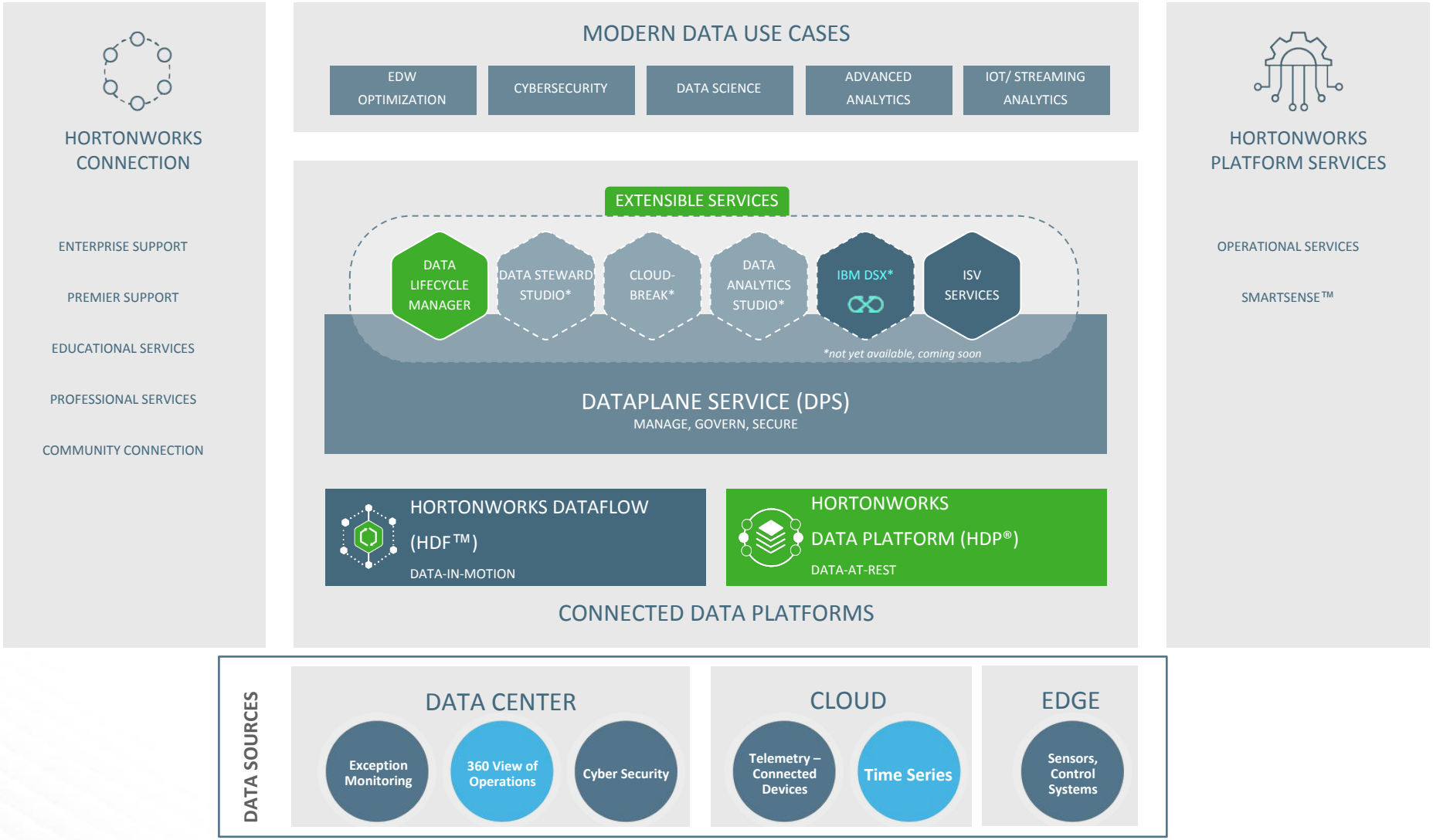
Hortonworks consistent and continuous track record of innovation



HORTONWORKS

Global Data Management With Hortonworks

Globally Manage, Secure, Govern, Consume



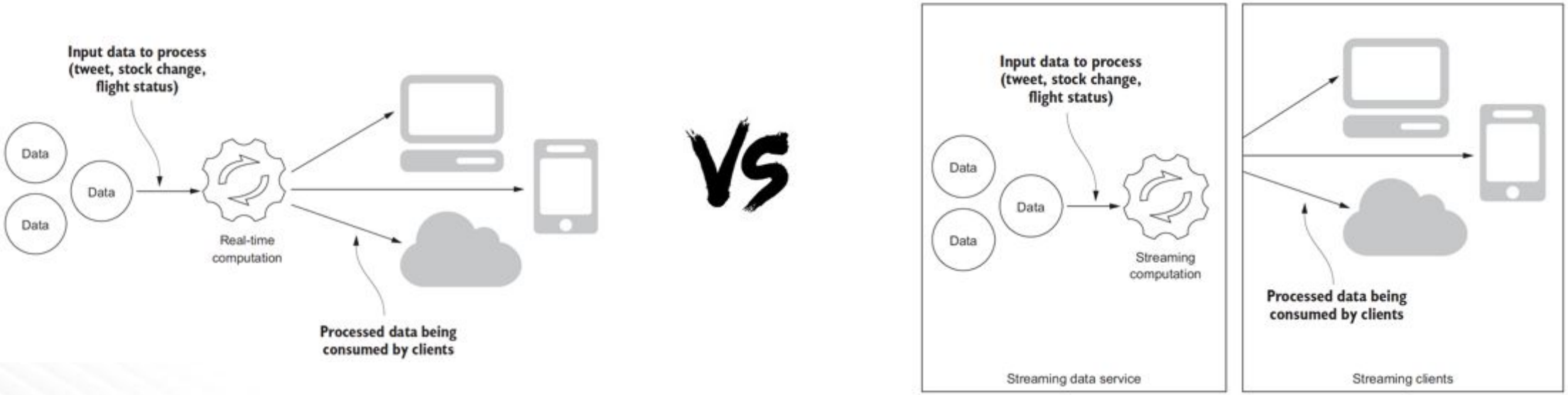
Contents

- All data needs to be under management: DataPlane Services
- Understanding a Streaming Data Architecture
- IoT Demo: Trucking company with a fleet of international trucks

Understanding a Streaming Data Architecture

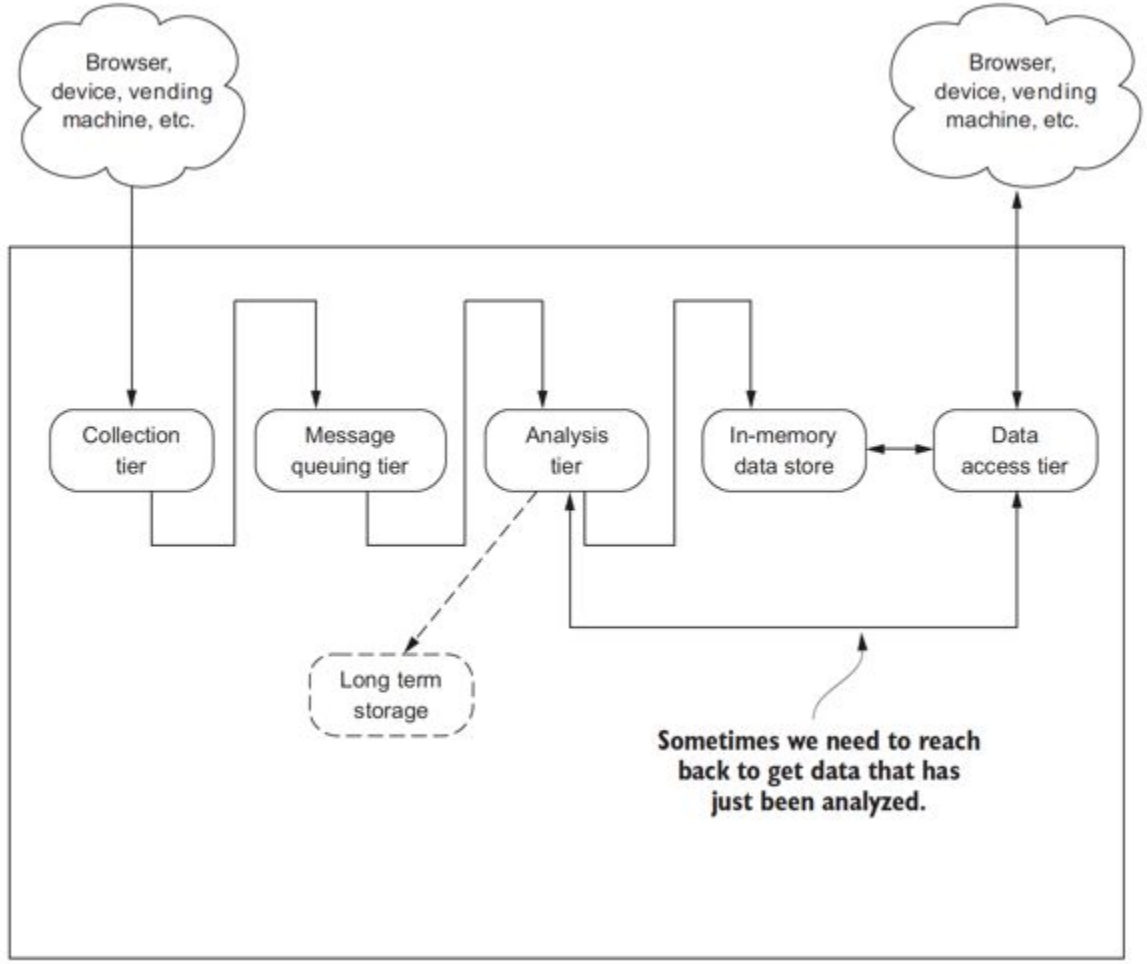
Real-time computing can be actually classified as*:

- **Hard real-time –** latency measured in ns/ms ; no tolerance for delay → potential loss of life
- **Soft real-time –** latency measured in ms/s ; low tolerance for delay → no life at risk
- **Near real-time –** latency measured in s/mins ; high tolerance for delay → no life at risk



* From ***"Stream Data – Understanding the real-time pipeline"*** by Andrew G. Psaltis

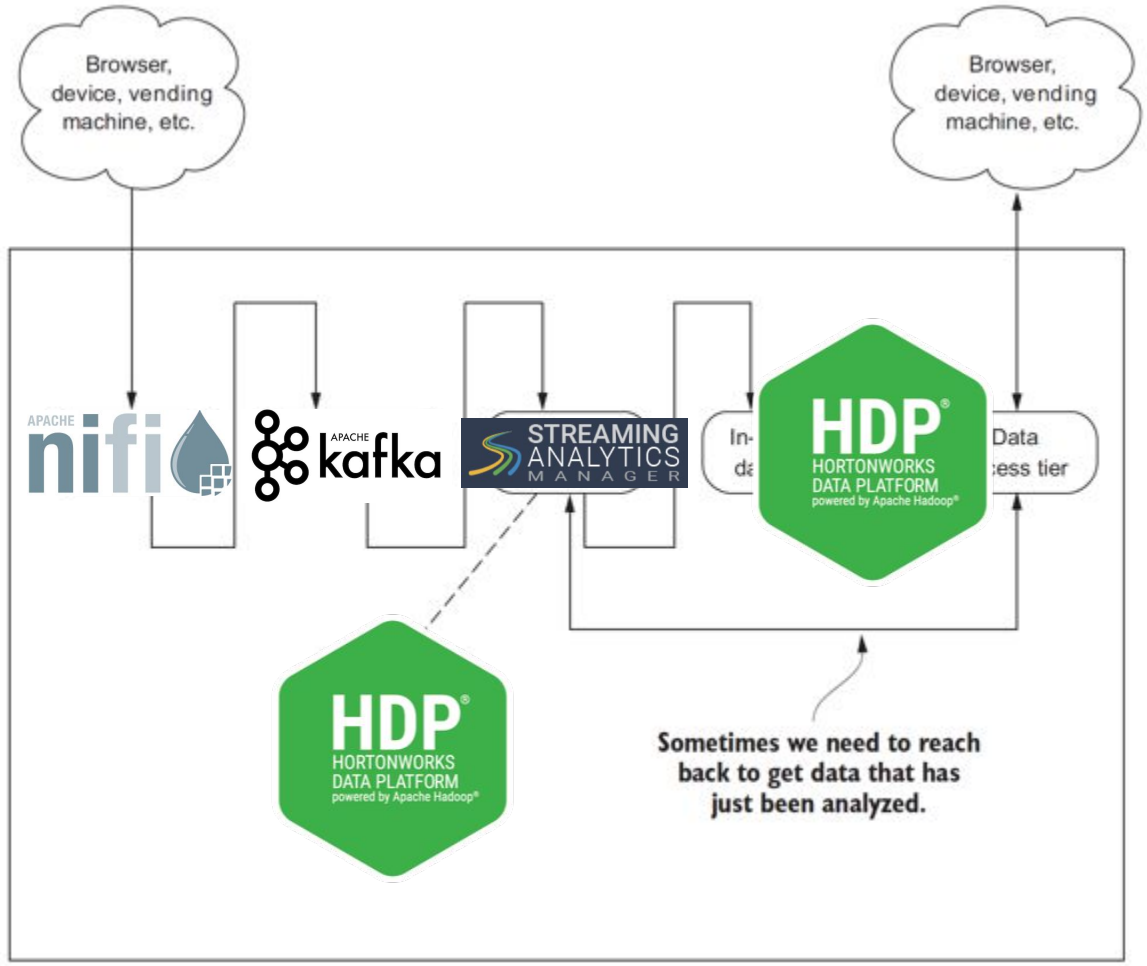
The Streaming Data Architectural blueprint*



* From ***"Stream Data – Understanding the real-time pipeline"*** by Andrew G. Psaltis

HORTONWORKS

The Streaming Data Architectural blueprint* with Hortonworks

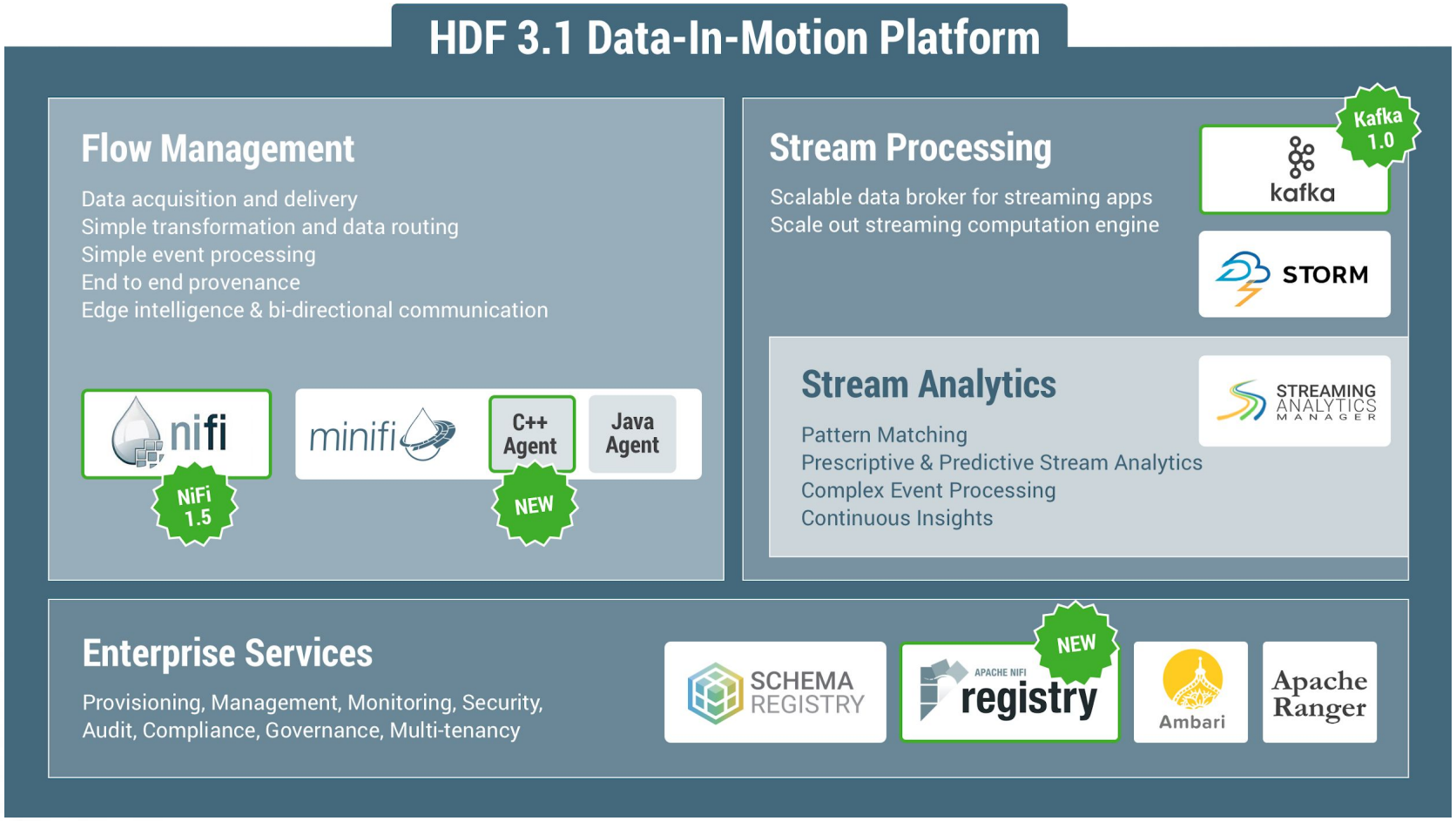


* From ***"Stream Data – Understanding the real-time pipeline"*** by Andrew G. Psaltis

HORTONWORKS

Hortonworks Data Flow (HDF)

Platform to build dataflow management and streaming analytics solutions that allow you to *collect*, *curate*, *analyze* and *act on* data in motion across the data center and cloud.



HORTONWORKS

Contents

- All data needs to be under management: DataPlane Services
- Understanding a Streaming Data Architecture
- IoT Demo: Trucking company with a fleet of international trucks

IOT Demo: Trucking company w/ fleet of international trucks

A truck generates millions of events for a given route; an event could be:

- 'Normal' events: starting / stopping of the vehicle
- 'Violation' events: speeding, excessive acceleration and breaking, unsafe tail distance
- 'Speed' Events: The speed of a driver that comes in every minute.

Company uses an application that monitors truck locations and violations from the truck/driver in real-time

Route?

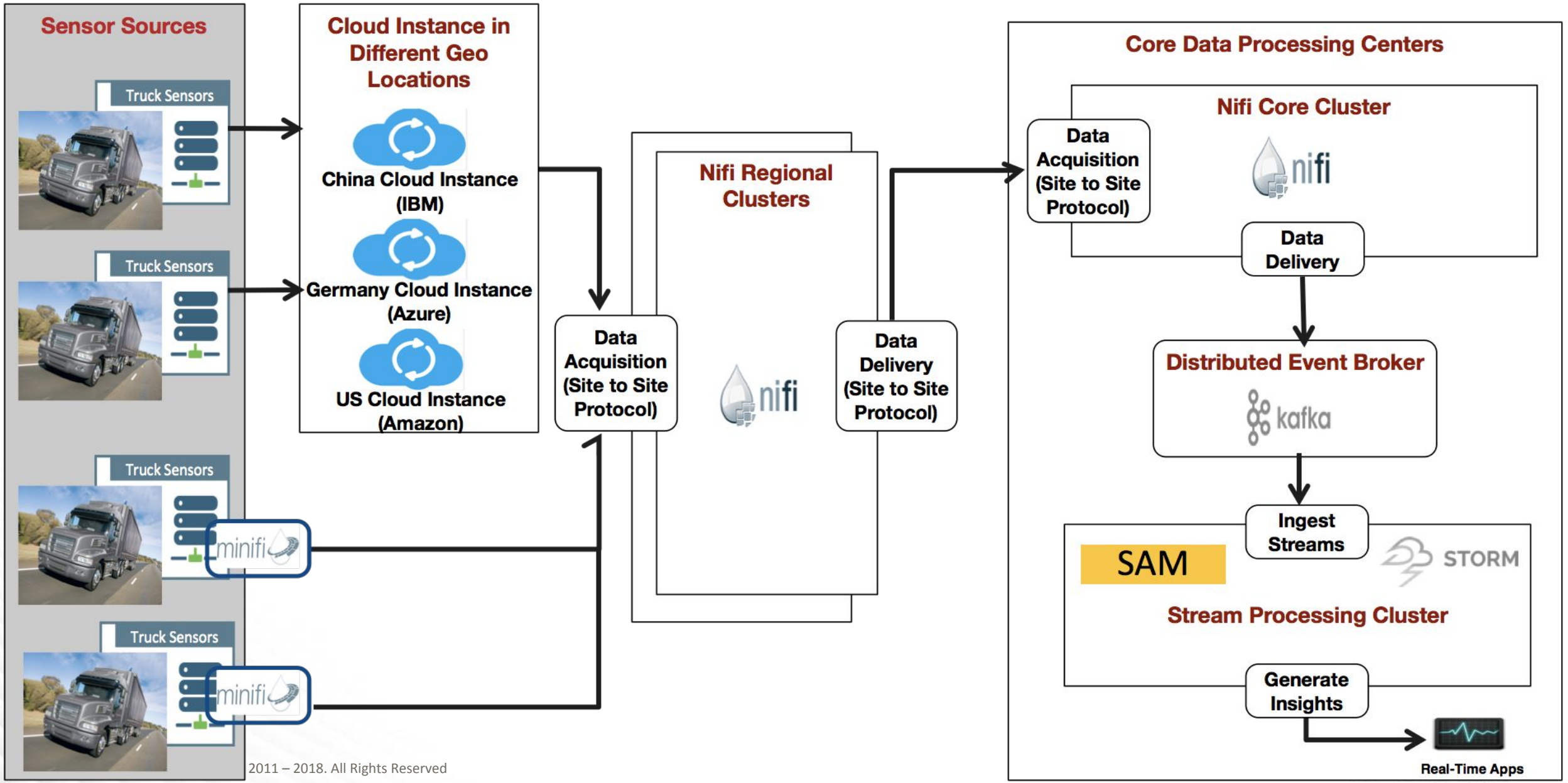
Truck?

Driver?

Analysts query a broad history to understand if today's violations are part of a larger problem with specific routes, trucks, or drivers

HORTONWORKS

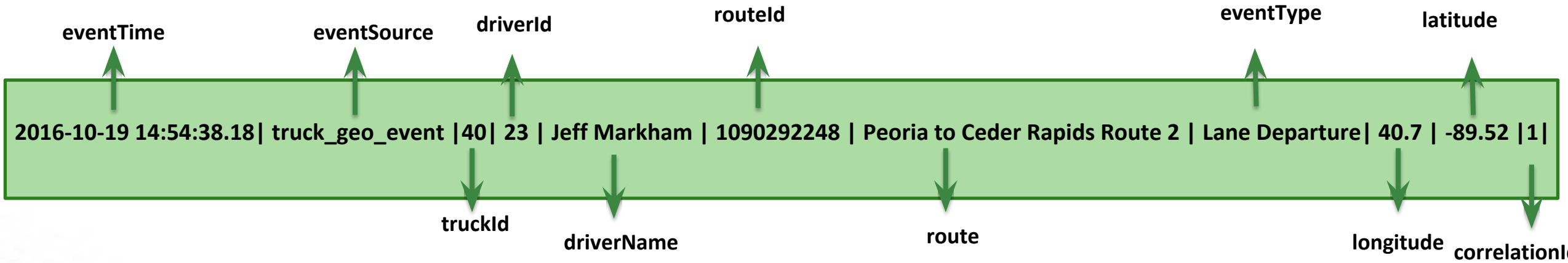
IOT Demo: Real-time Analytics Architecture with HDF



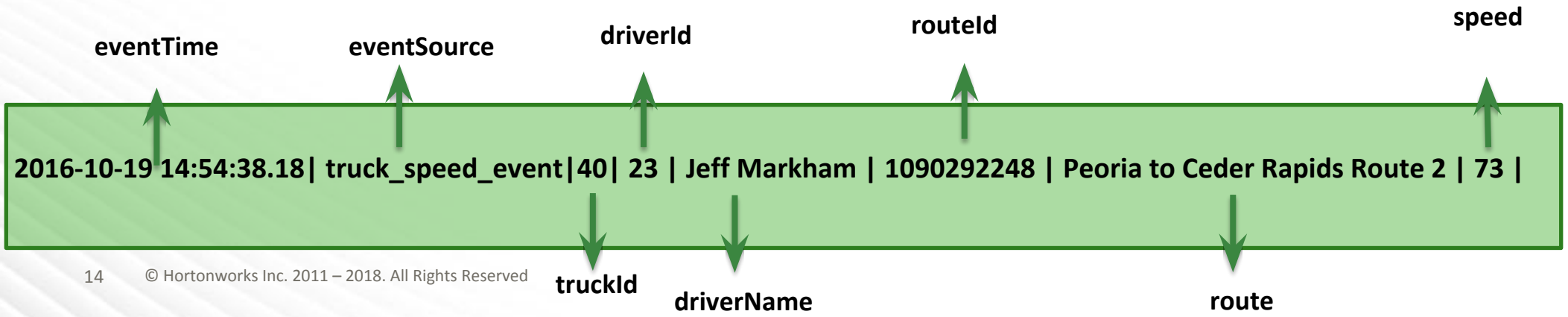
IOT Demo Data Sources: TruckGeoEvent and TruckSpeedEvent Streams

- Each Truck emits different event stream
  - Truck Geo Event
  - Truck Speed Event

Truck Geo Event:



Truck Speed Event:



DEMO

DataWorks Summit Berlin

April 16–19, 2018

Call for Abstracts is now open!

Call for Abstracts closes December 15,2017

https://dataworkssummit.com/abstracts/
https://dataworkssummit.com/berlin-2018/#tracks