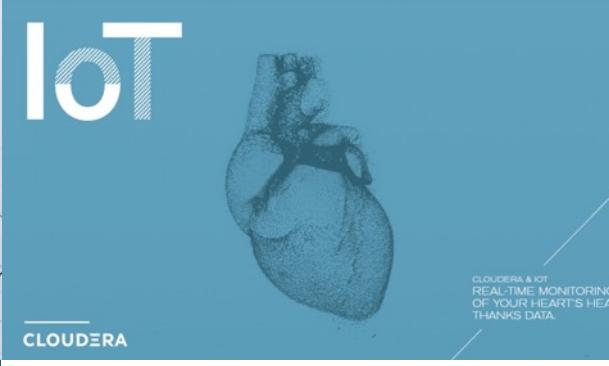




ENTERPRISE DATA CLOUD

CLOUDERA



CLOUDERA Now

CLOUDERA



EDGE 2AI

CLOUDERA



CLOUDERA

LOW CODE REAL TIME ANALYTICS

Open Source Lisbon 2020

Alex Campos | Solutions Engineer

ac@cloudera.com

AGENDA

- Cloudera Data Flow
- Predictive Maintenance
- Time to Play
- Wrap Up

CLOUDERA

DATA

FLOW

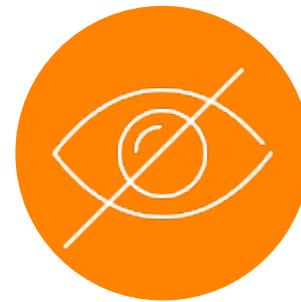
CHALLENGES IN FAST DATA



Data Ingestion: High-volume streaming sources, multiple message formats, diverse protocols and multi-vendor devices creates data ingestion challenges



Real-time Insights: Analyzing continuous and rapid inflow (velocity) of streaming data at high volumes creates major challenges for gaining real-time insights

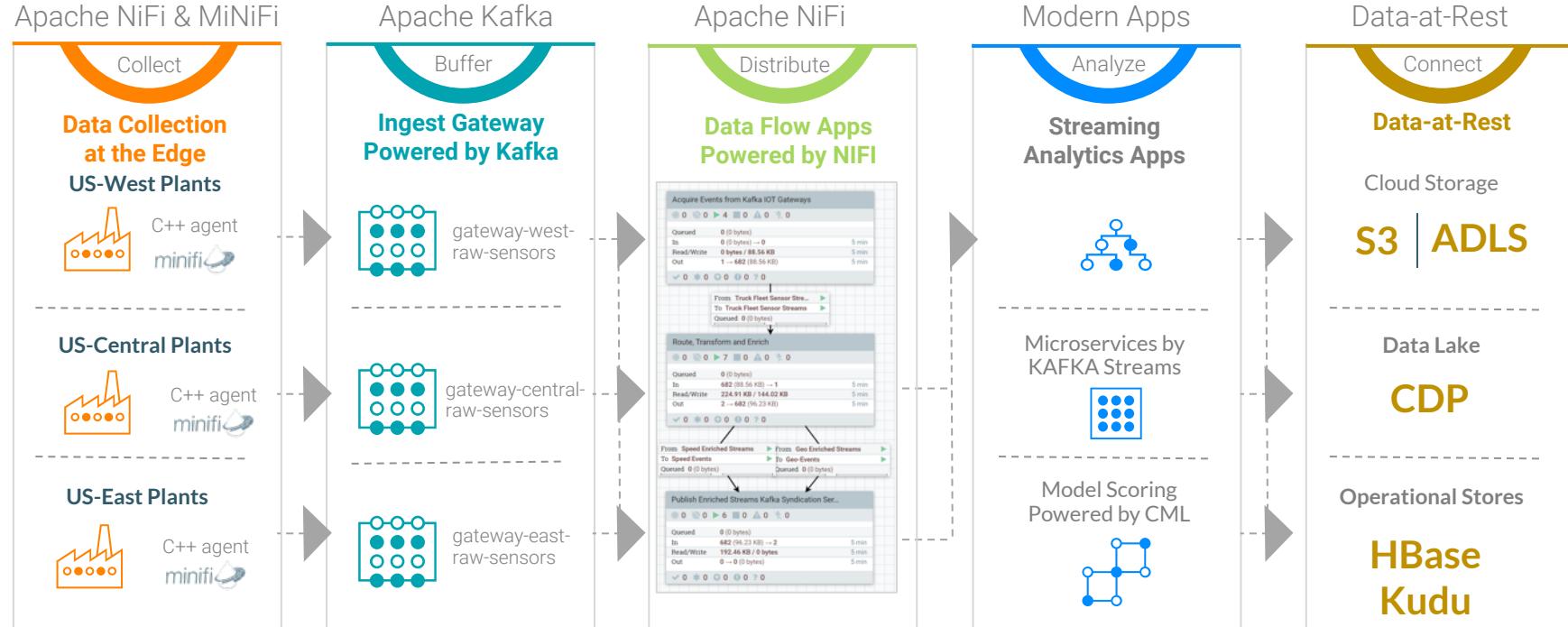


Visibility: Lack visibility of end-to-end streaming data flows, inability to troubleshoot bottlenecks, consumption patterns, secure the data throughout the pipeline etc.

CLOUDERA DATA-IN-MOTION PLATFORM

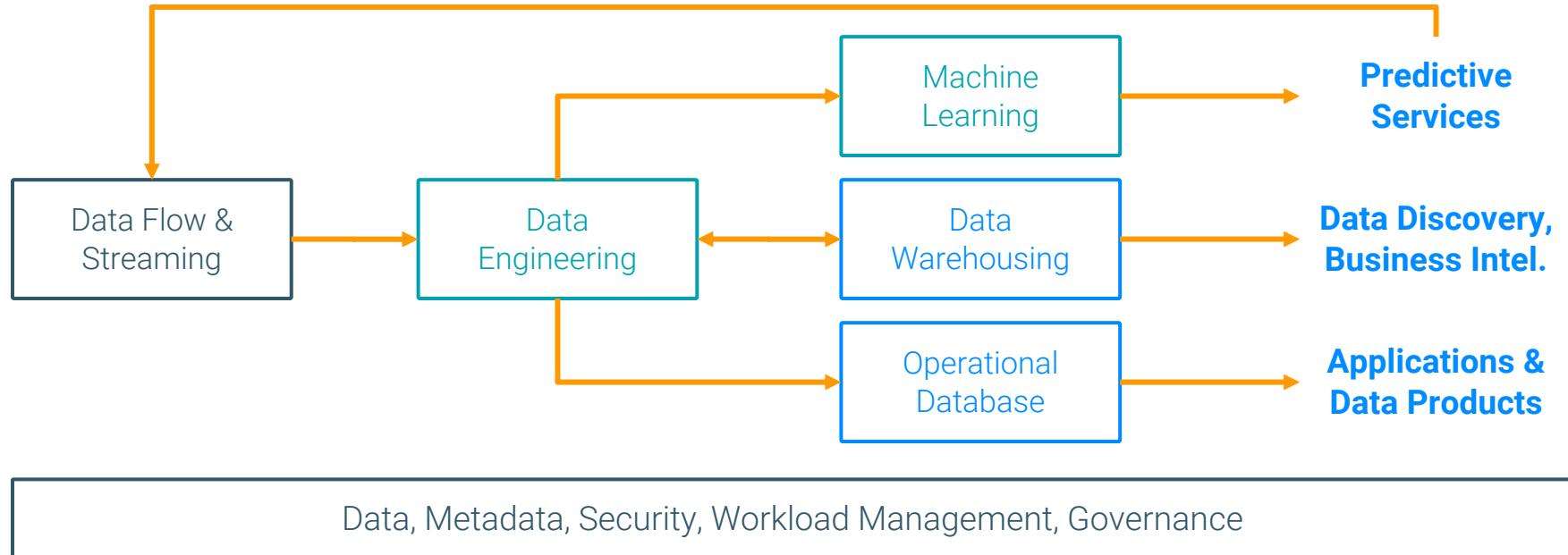


A DATA-IN-MOTION REFERENCE ARCHITECTURE



FOUNDATION OF MODERN DATA MANAGEMENT

CDP A Platform for Multiple Workloads



ONE PLATFORM – TWO FORM FACTORS

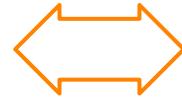
CDP Public Cloud
(platform-as-a-service)

CDP On-Prem
(installable software)

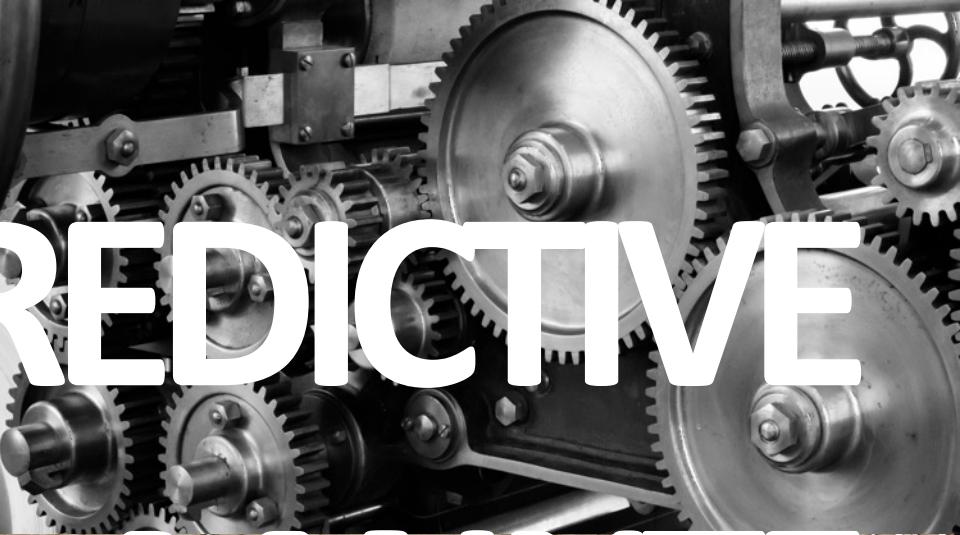
Control Plane



CLOUDERA
SDX



Cloudera Runtime



PREDICTIVE MAINTENANCE

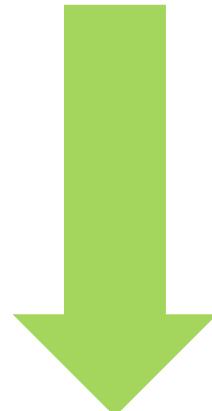
PREDICTIVE MAINTENANCE

Business Drivers



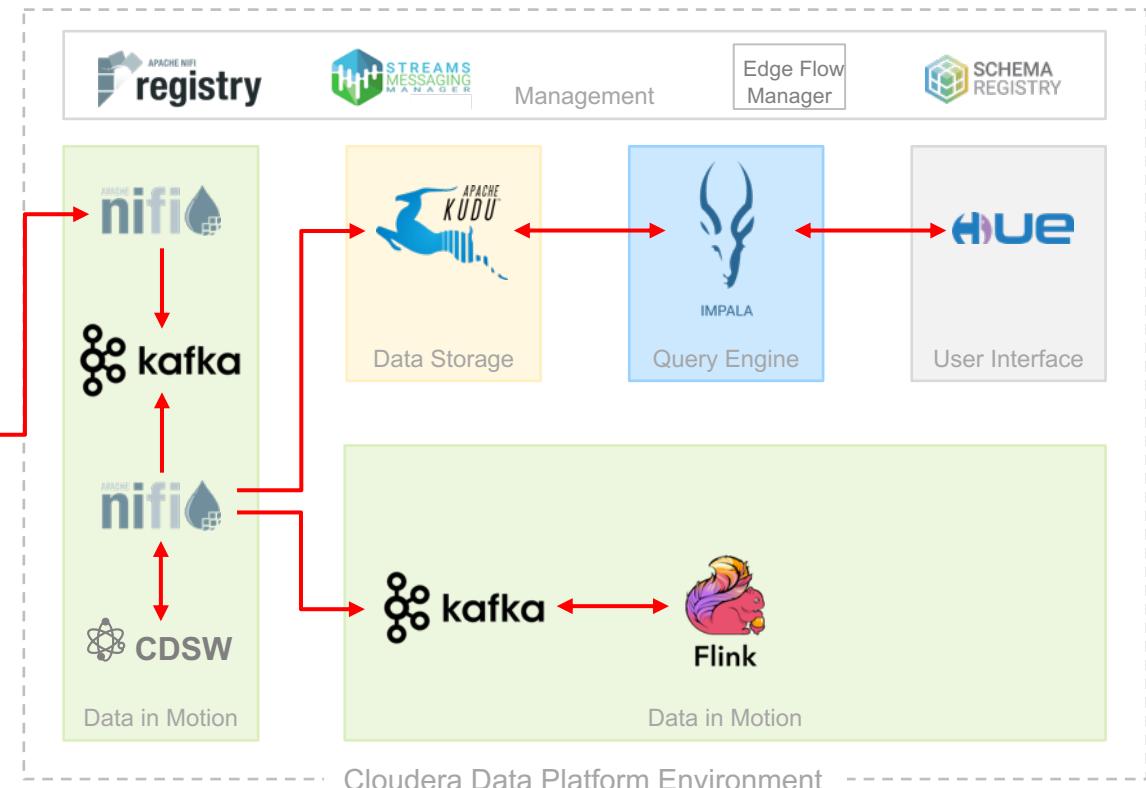
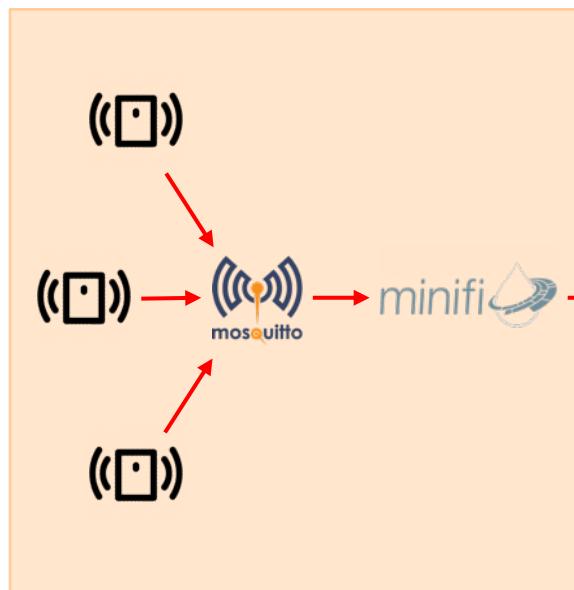
Increase
operation
efficiency

Decrease
unplanned
outages



IOT ARCHITECTURE

From Edge to AI



Edge Environment

Cloudera Data Platform Environment

TIME

TO

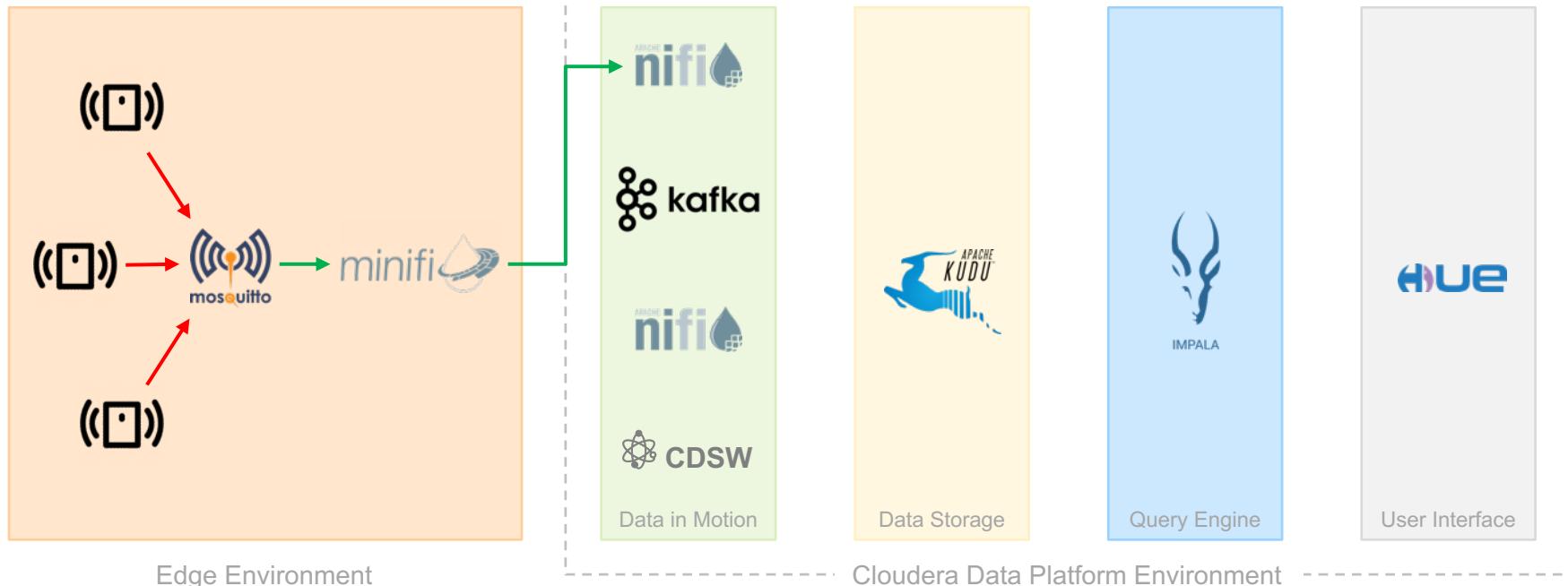
PLAY

STEP

1

IOT ARCHITECTURE

From Edge to AI



STEP 1 – COMPONENTS



Apache NiFi is an open source software for automating and managing the flow of data between systems. It is a powerful and reliable system to process and distribute data. It provides a web-based User Interface for creating, monitoring, & controlling data flows.

Some of the features includes **data provenance**, **extensible**, **secure** and others.

STEP 1 – COMPONENTS



MiNiFi is a complementary data collection approach that supplements the core tenets of NiFi in dataflow management, focusing on the collection of data at the source of its creation, mainly in the edge

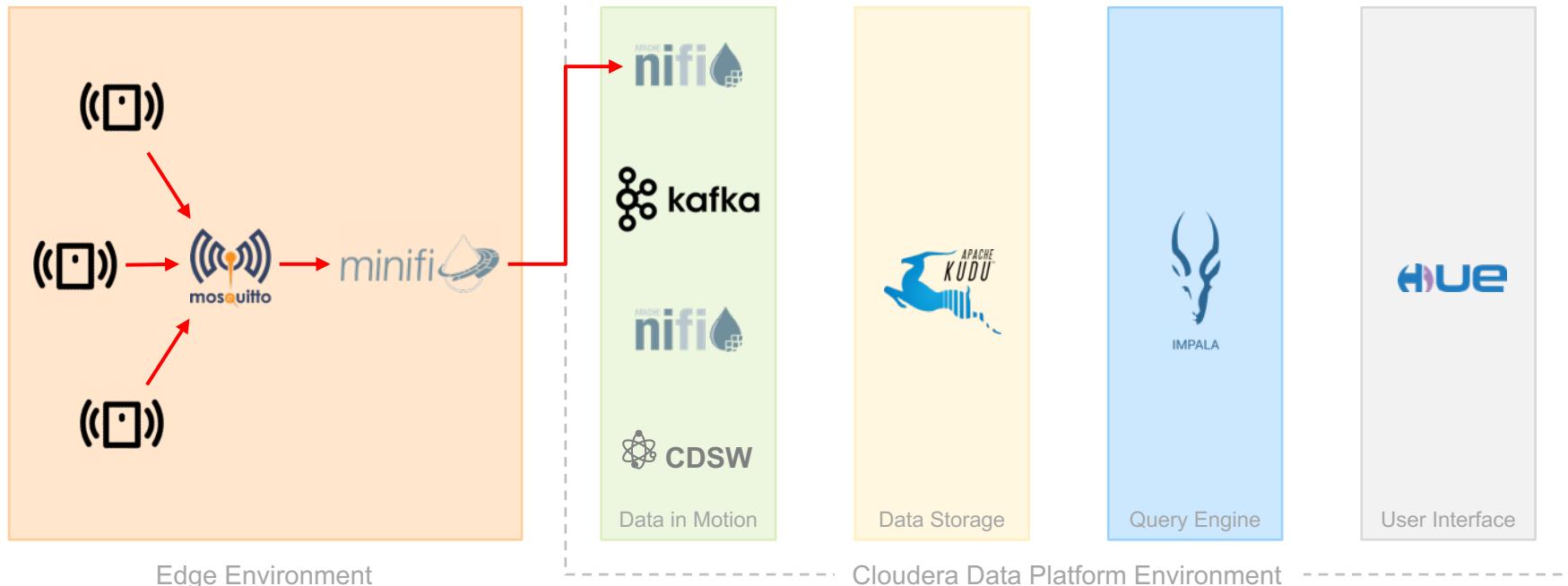
Agent management hub that supports a graphical flow-based programming model to develop, deploy, and monitor edge flows on thousands of MiNiFi agents

Edge Flow Manager

STEP 2

IOT ARCHITECTURE

From Edge to AI



STEP 2 – COMPONENTS

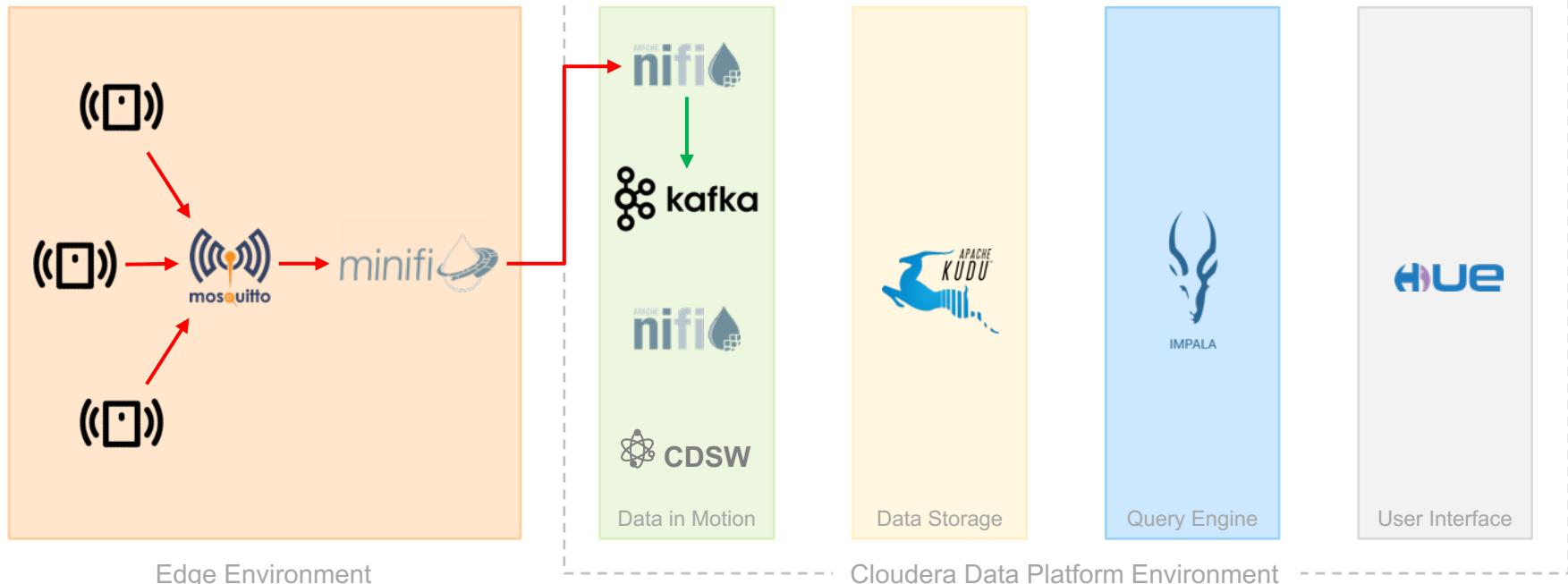


Provides a central repository to store, manage and version data schemas between components, allowing evolution and reusability

STEP 3

IOT ARCHITECTURE

From Edge to AI



STEP 3 – COMPONENTS



Provides a central location for storage and management of shared resources across one or more instances of NiFi and/or MiNiFi

A streaming message platform. It is designed to be high performance, highly available, and redundant, ideal for real-time and streaming applications

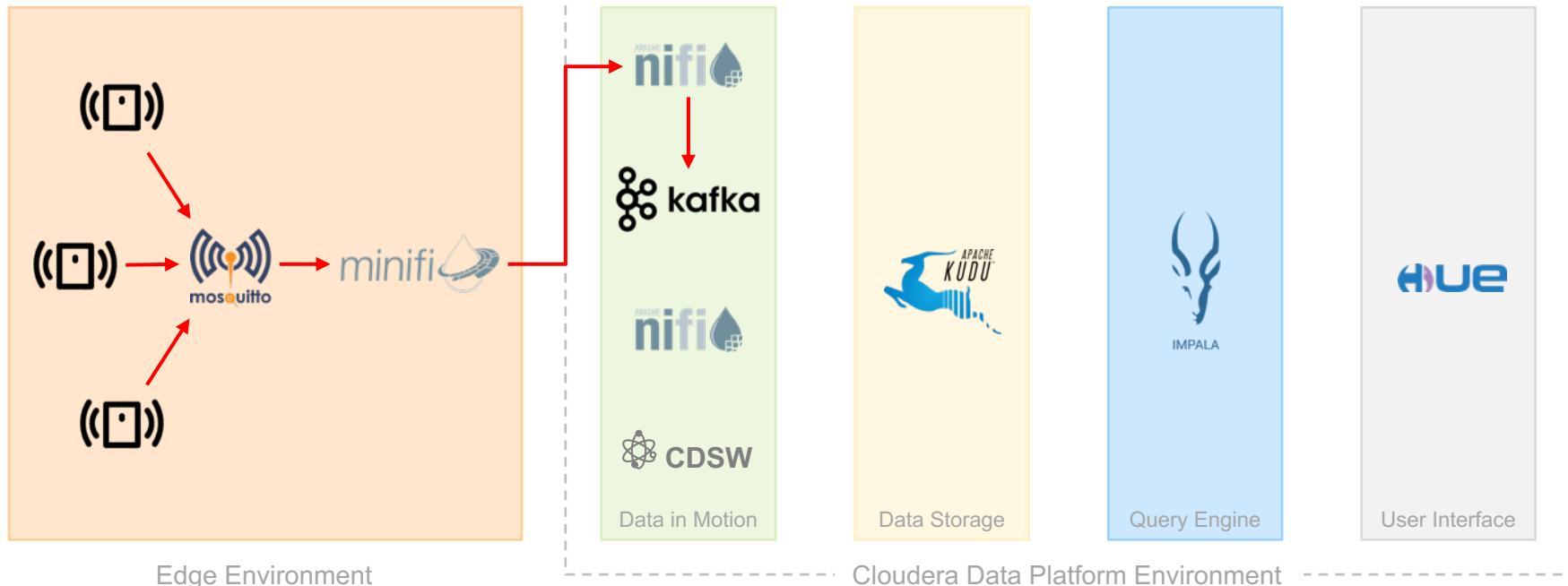


STEP

4

IOT ARCHITECTURE

From Edge to AI



STEP 4 – COMPONENTS

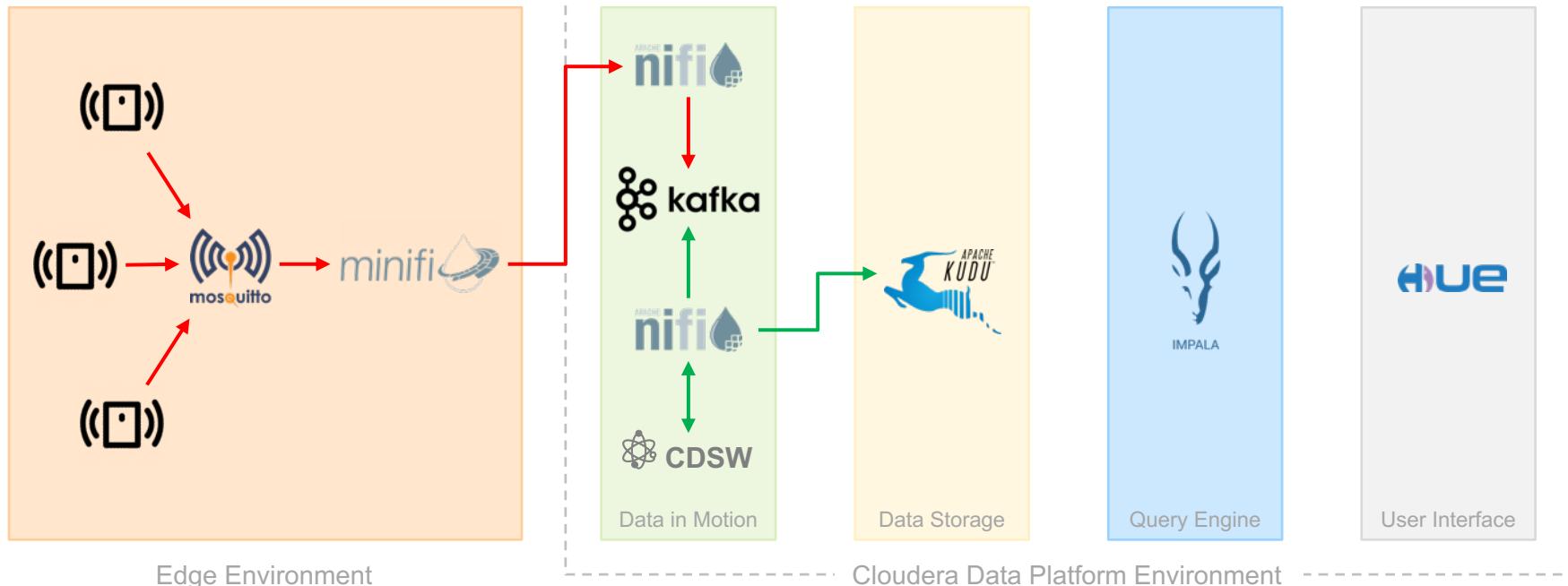


As a monitoring and management tool that provides end-to-end visibility in an enterprise Apache Kafka environments. With SMM, it is possible to gain clear insights about Kafka clusters, like brokers, topics, producers and consumers.

STEP 5

IOT ARCHITECTURE

From Edge to AI



STEP 5 – COMPONENTS



Self-service enterprise data science platform that lets data scientists manage their own analytics pipelines, thus accelerating machine learning projects from exploration to production

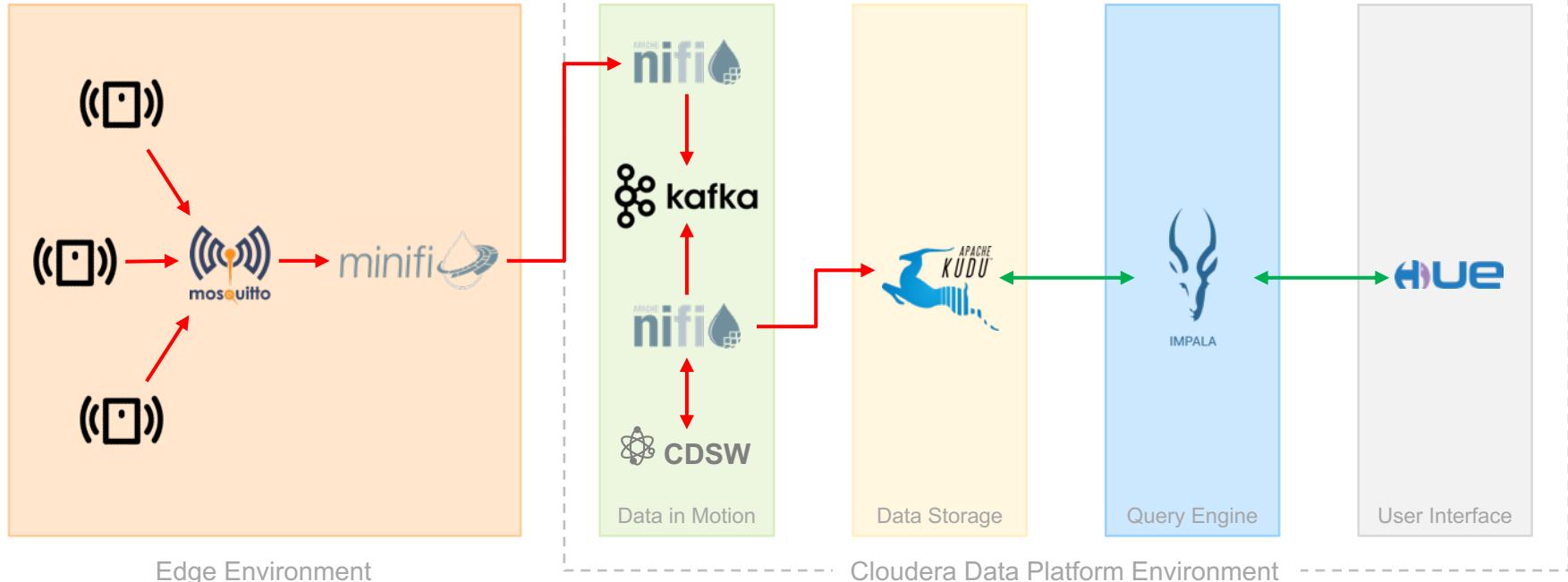
Kudu is storage for fast analytics on fast data—providing a combination of fast inserts and updates alongside efficient columnar scans to enable multiple real-time analytic workloads across a single storage layer.



STEP 6

IOT ARCHITECTURE

From Edge to AI



STEP 7 – COMPONENTS



Hue is the open source analytics workbench designed for fast data discovery, intelligent query assistance, and seamless collaboration. Bridge the gap between IT and the business for trusted self-service analytics.

Impala provides high-performance, low-latency SQL queries on data storage layer. The fast response for queries enables interactive exploration and fine-tuning of analytic queries.

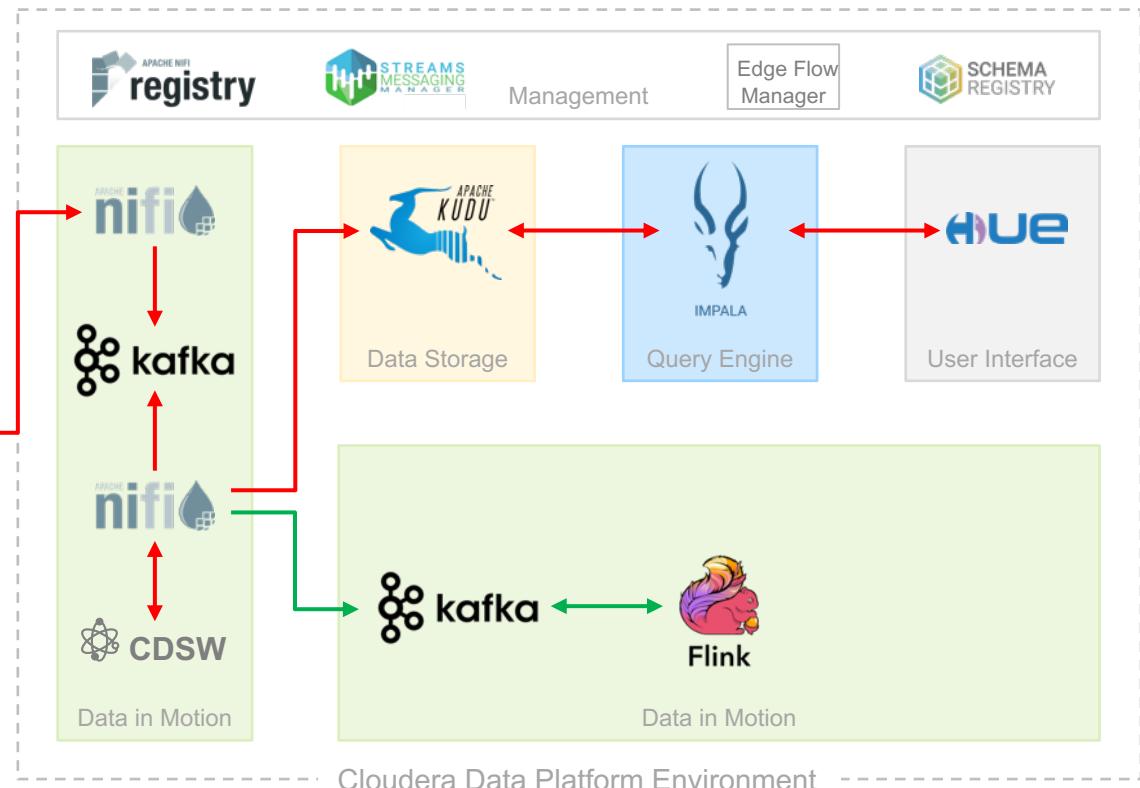
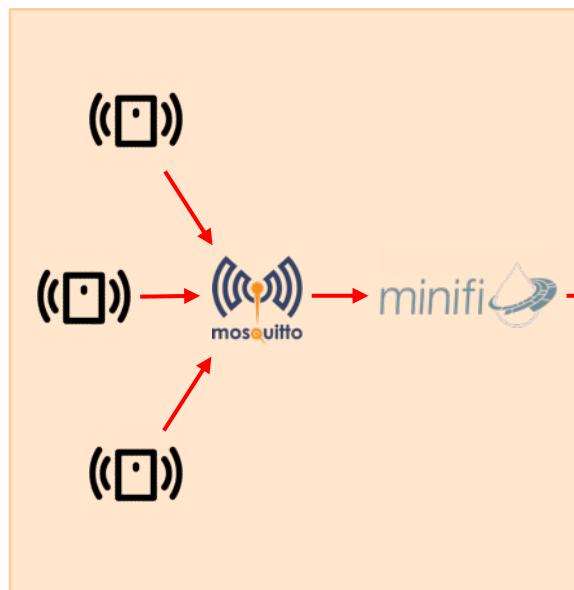


STEP

8

IOT ARCHITECTURE

From Edge to AI



Edge Environment

Cloudera Data Platform Environment

STEP 8 – COMPONENTS

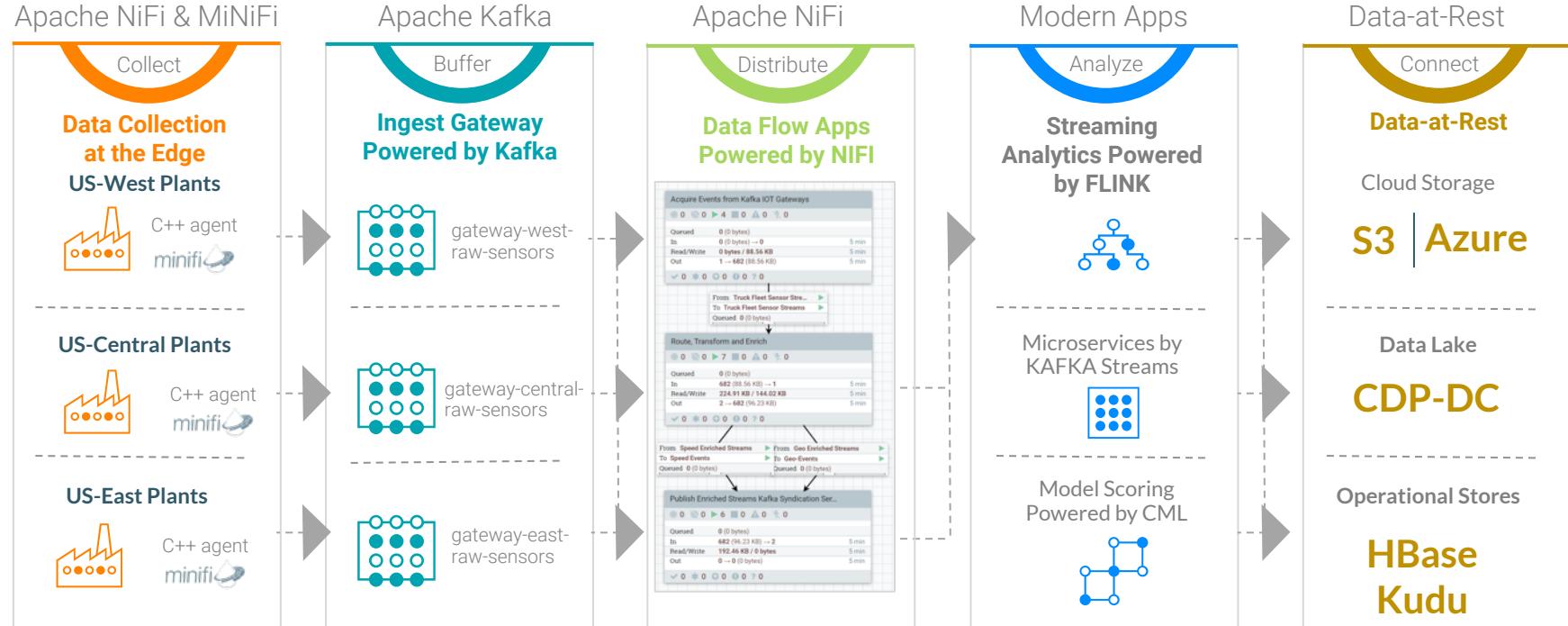


Flink is a distributed processing engine and a scalable data analytics framework that can deliver data analytics in real-time.

Flink is designed to run in all common cluster environments, perform computations at in-memory speed and at any scale. Furthermore, Flink provides communication, fault tolerance, and data distribution for distributed computations over data streams.

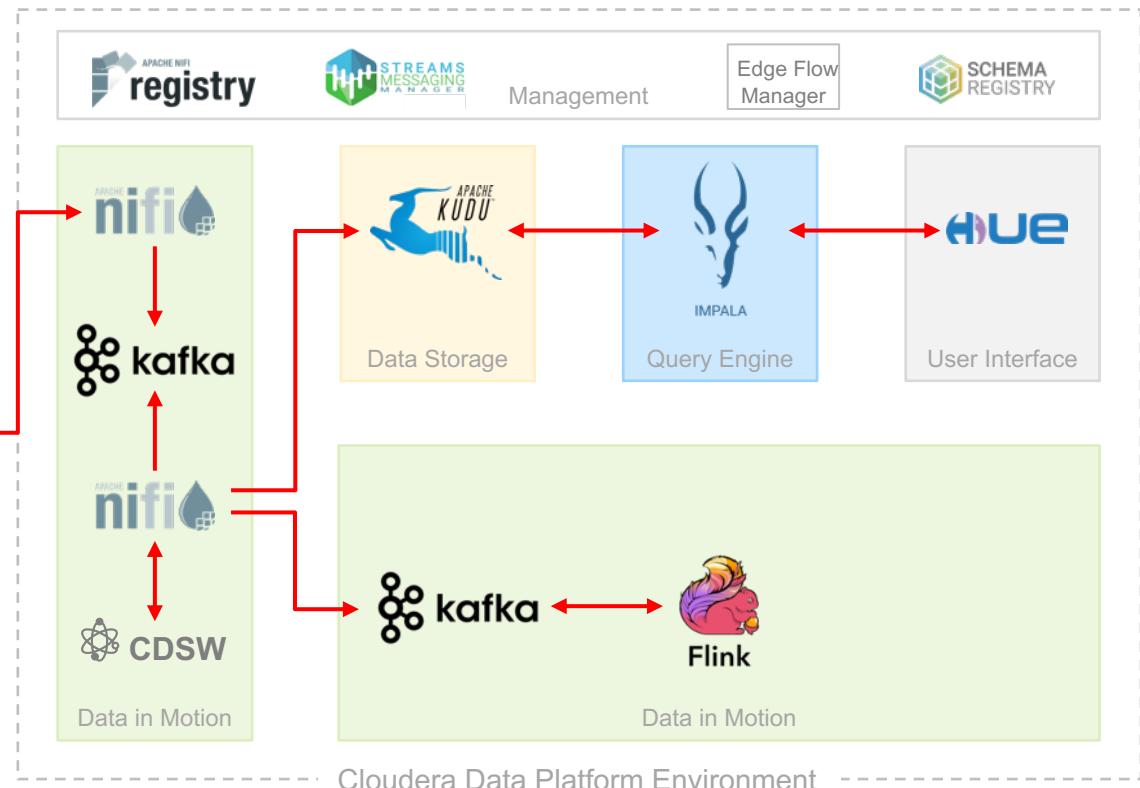
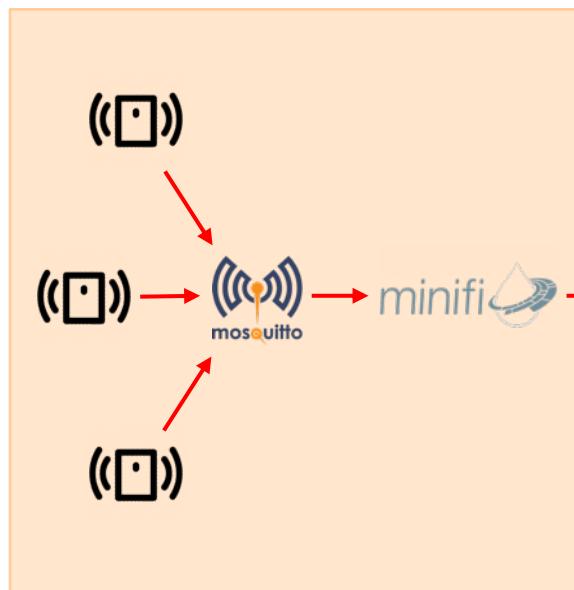
WRAP UP

A DATA-IN-MOTION REFERENCE ARCHITECTURE



IOT ARCHITECTURE

From Edge to AI



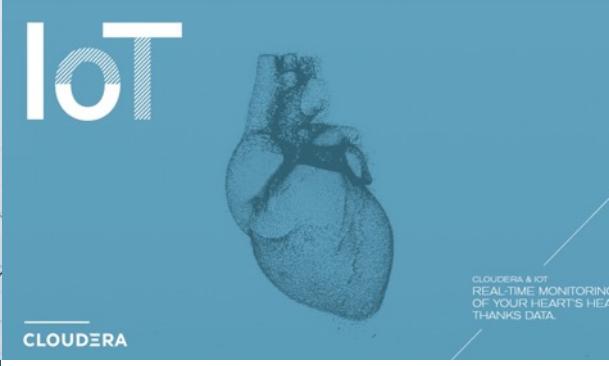
Edge Environment

Cloudera Data Platform Environment



ENTERPRISE DATA CLOUD

CLOUDERA



IoT

CLOUDERA



EDGE 2AI

CLOUDERA



WANT MORE?

- CDP (Cloudera Data Platform) weekly demos: Connect the Data Lifecycle
- Industry events
- Deep dive:
 - Flink
 - MLOps
 - SDX
 - Machine Learning

THANK YOU