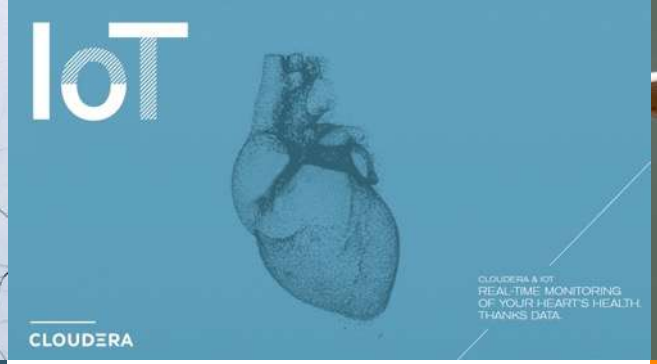


ENTERPRISE DATA CLOUD

CLOUDERA

CLOUDERA Now

EDGE 2AI

CLOUDERA

SDX

CLOUDERA

CLOUDERA

LOW CODE REAL TIME ANALYTICS

#QuarantineMeetup

AGENDA

- Introductions
- Cloudera Data Flow
- Demo Environment
- Predictive Maintenance
- Time to Play
- Wrap Up

INTRO DUCTIONS

WHO WE ARE

WHO WE ARE

Eynar Espinoza

Solutions
Engineering
Manager

Alex Campos

Solutions
Engineer

@campossalex

CLOUDERA DATA FLOW

STREAMING DATA BRINGS GREAT OPPORTUNITIES

Streaming
~$1.65 B

Data Science
~$180 B

Big Data
~$210 B

Cloud
~$410 B

IoT
~$1.2 T

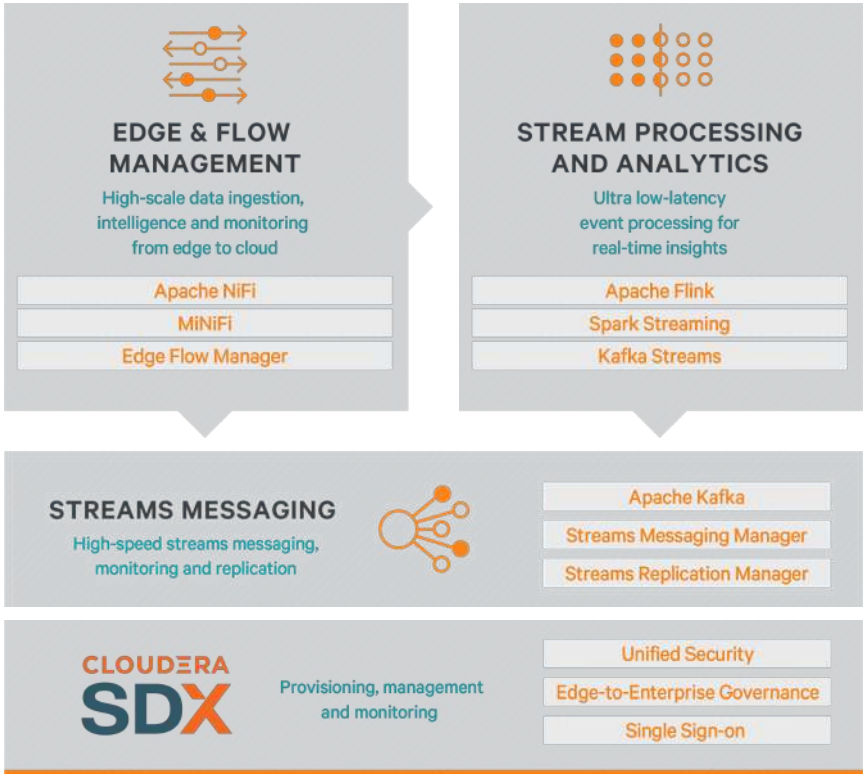
CHALLENGES IN FAST DATA

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

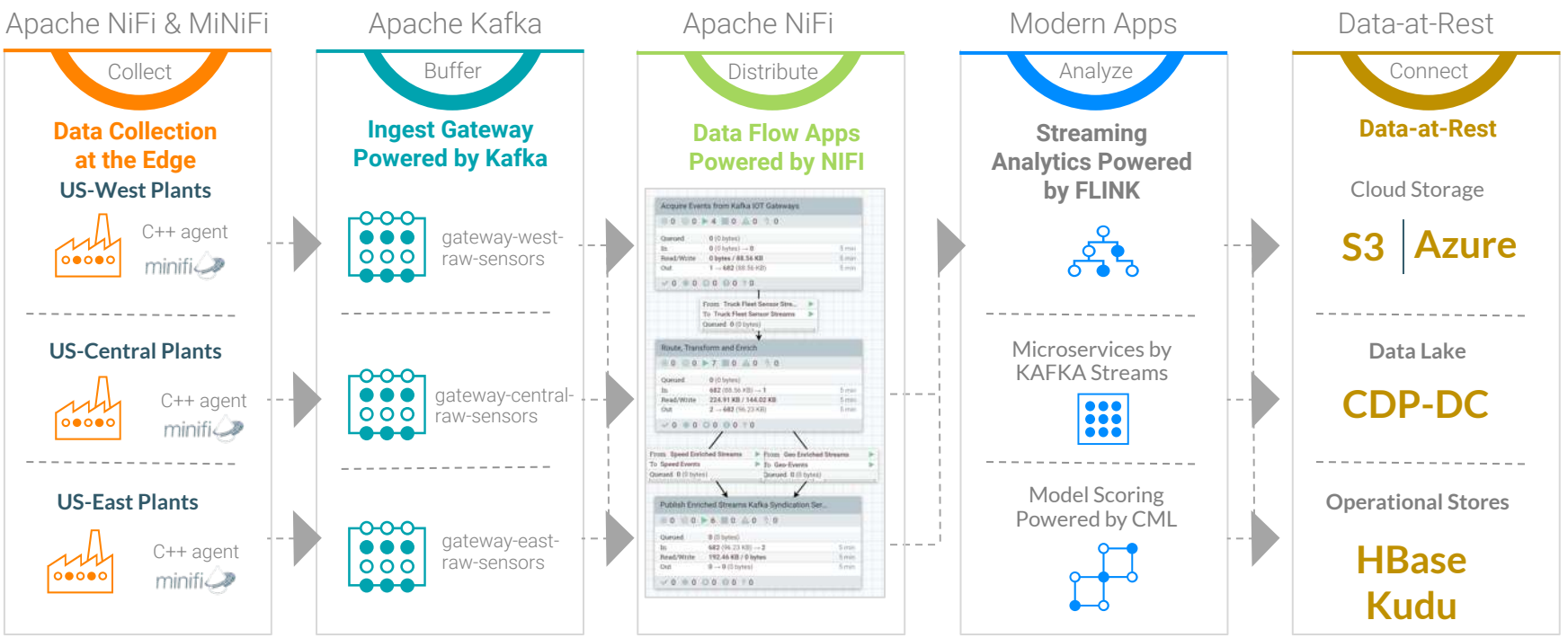
Real-time Insights: Analyzing continuous and rapid inflow (velocity) of streaming data at high volumes creates major challenges for gaining real-time insights

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

CLOUDERA DATA-IN-MOTION PLATFORM

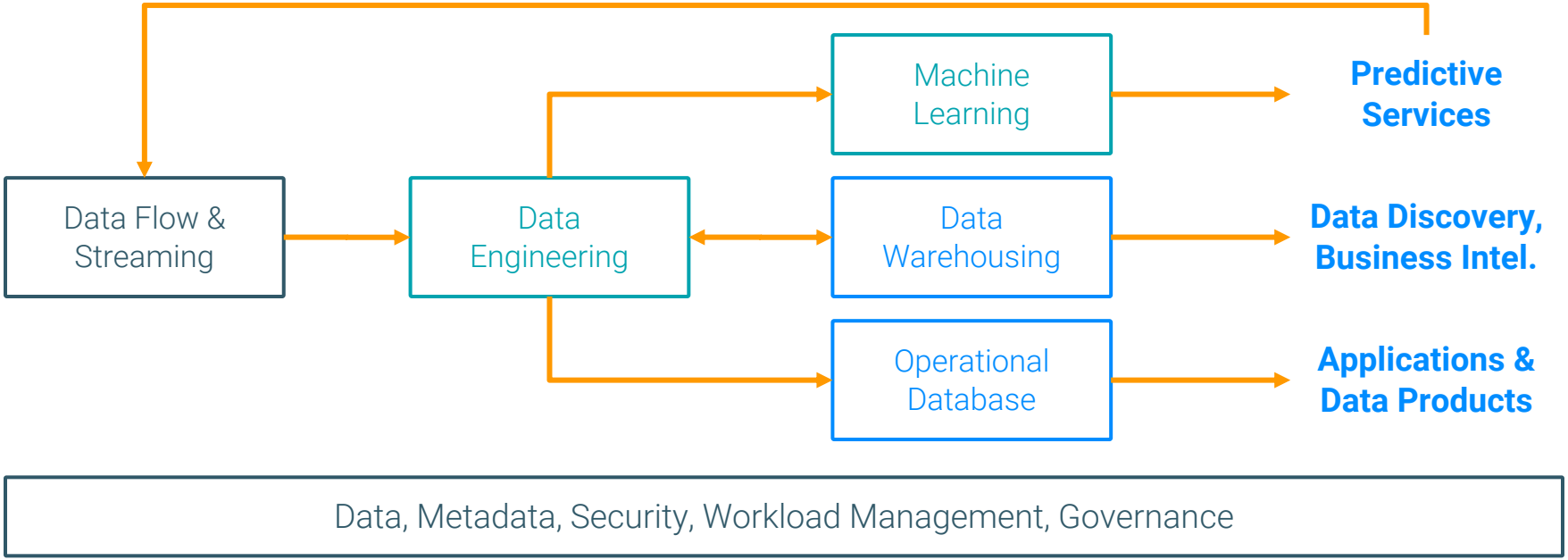


A DATA-IN-MOTION REFERENCE ARCHITECTURE



FOUNDATION OF MODERN DATA MANAGEMENT

CDP A Platform for Multiple Workloads



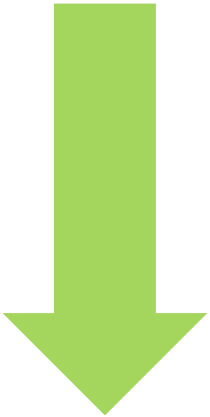

PREDICTIVE
MAINTE
NANCE

PREDICTIVE MAINTENANCE

Business Drivers

Increase operation efficiency

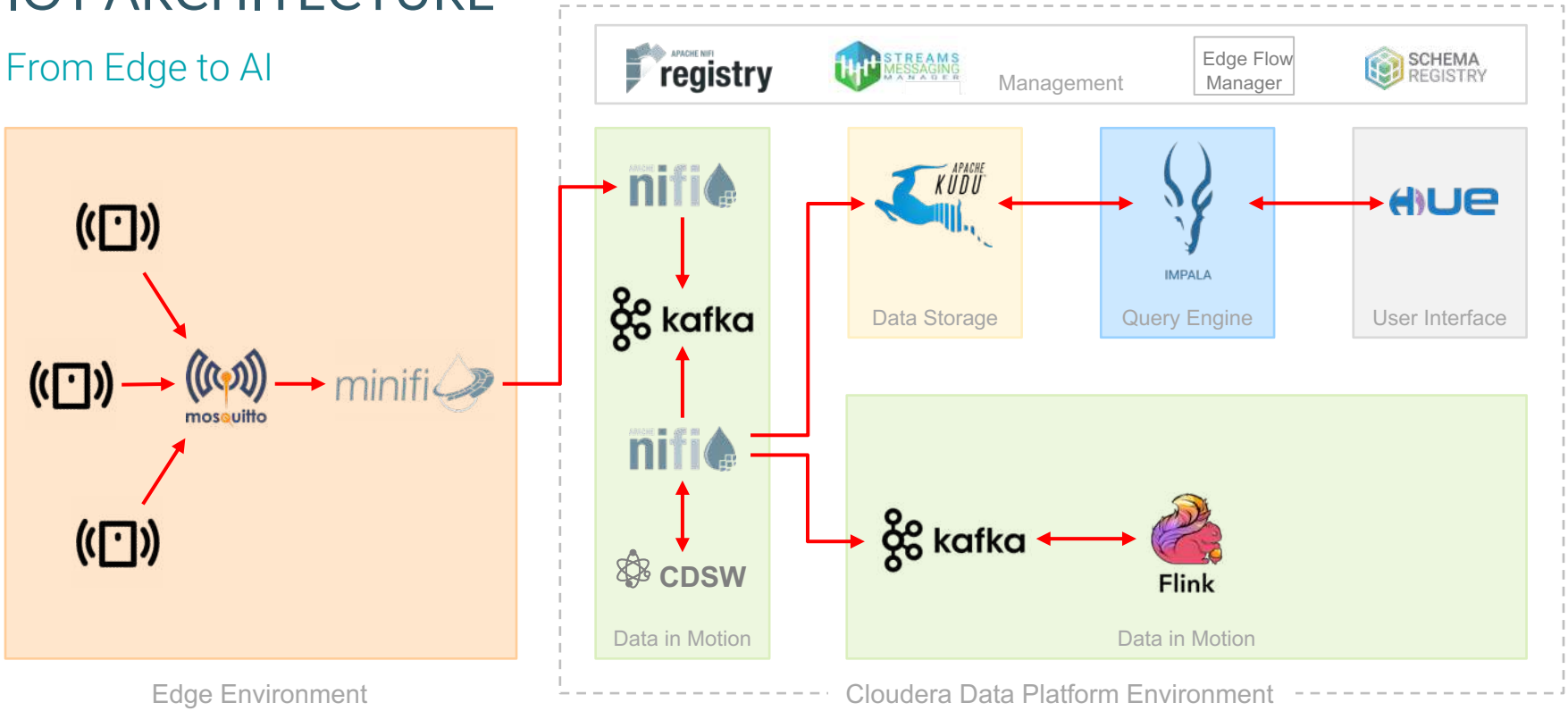
Decrease unplanned outages



From Edge to AI

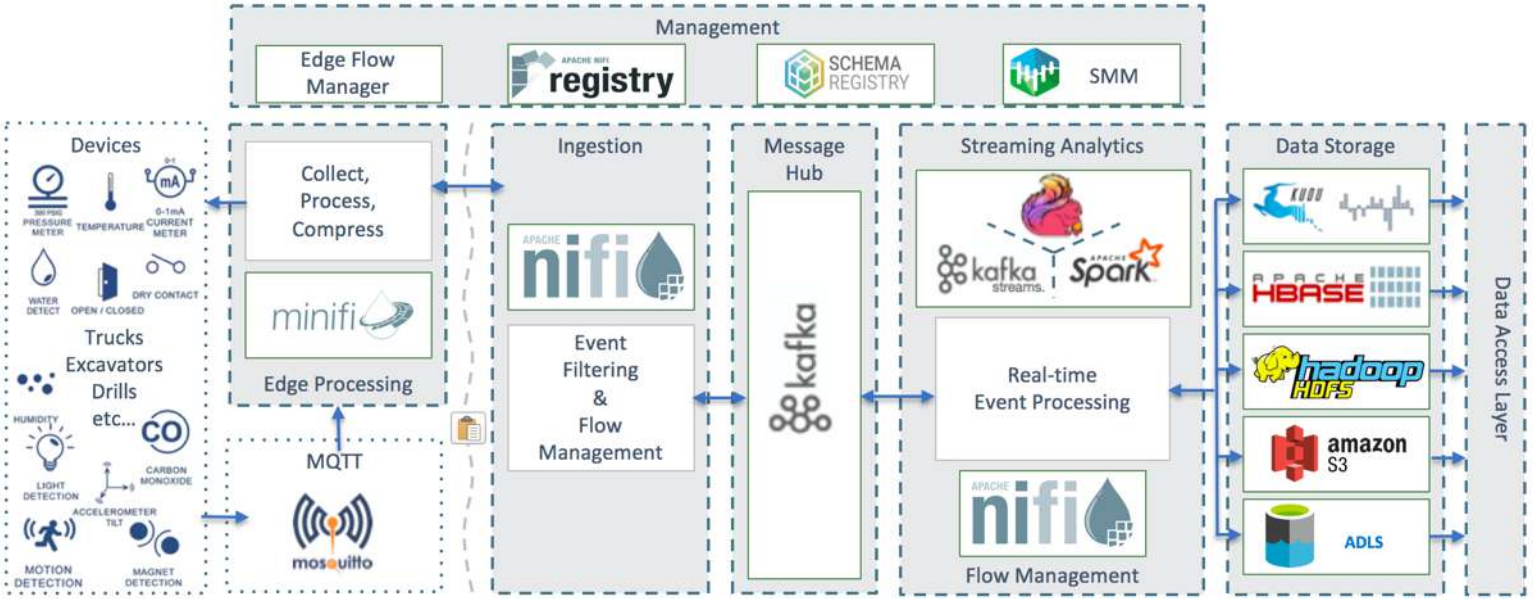
IOT ARCHITECTURE

From Edge to AI

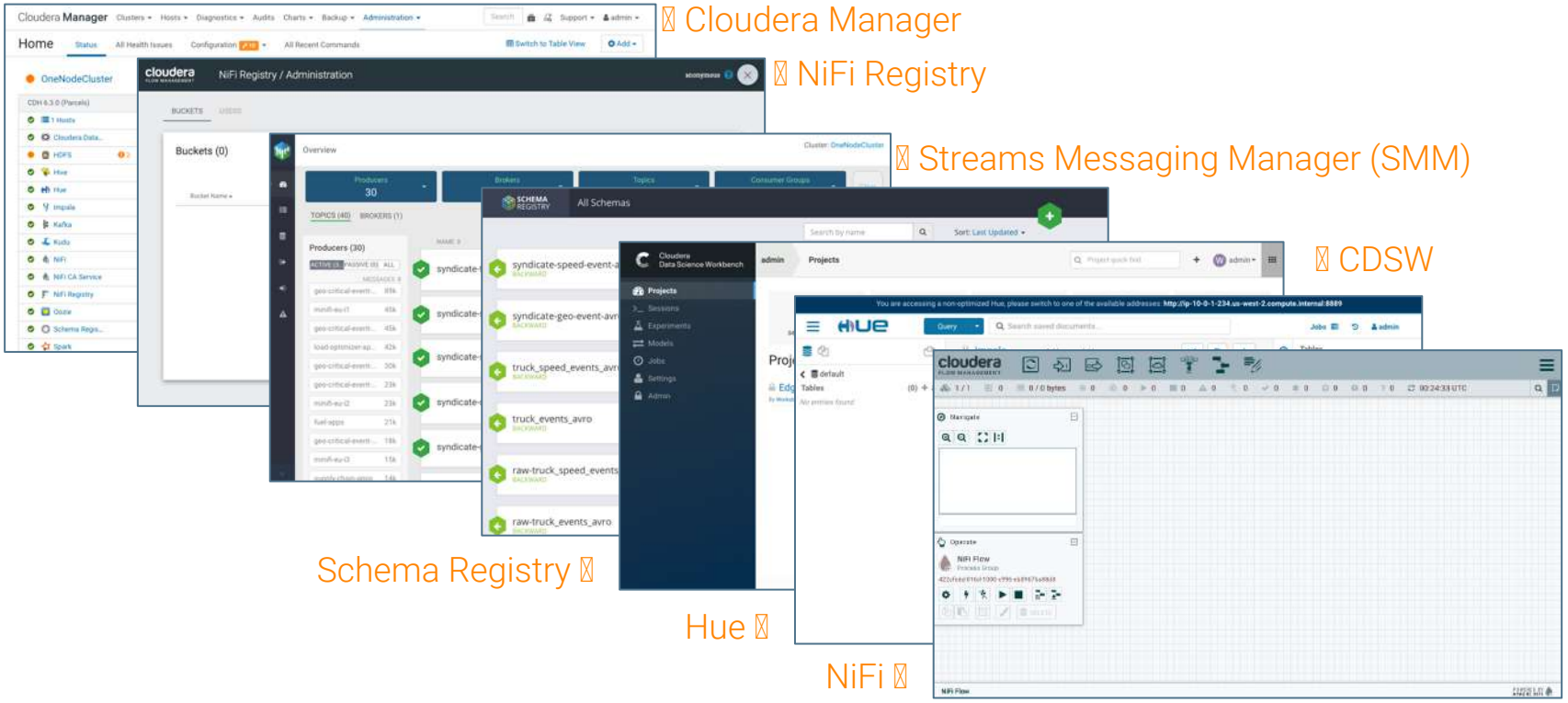


DEMO ENVIRONMENT

DEMO ENVIRONMENT

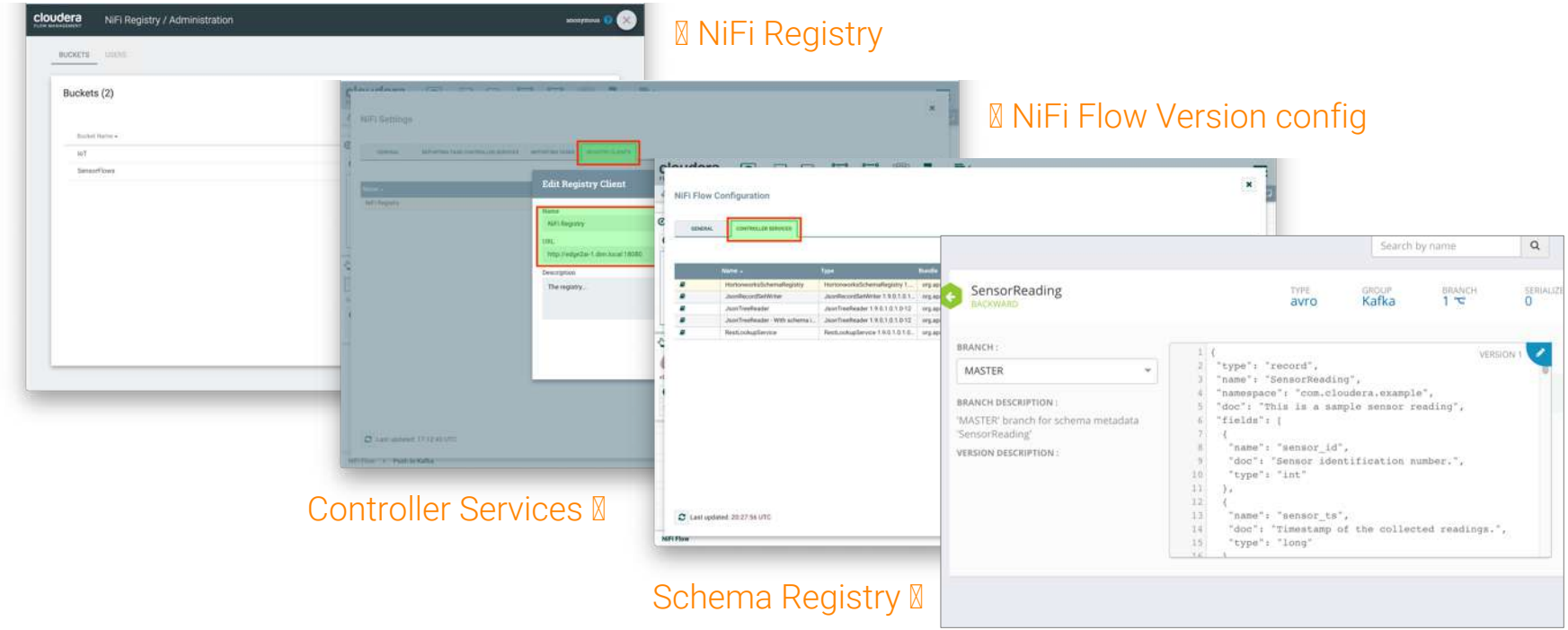


DEMO CLUSTER



CONFIGURATIONS

Pre-configured setup



 NiFi Registry

 NiFi Flow Version config

Controller Services

Schema Registry

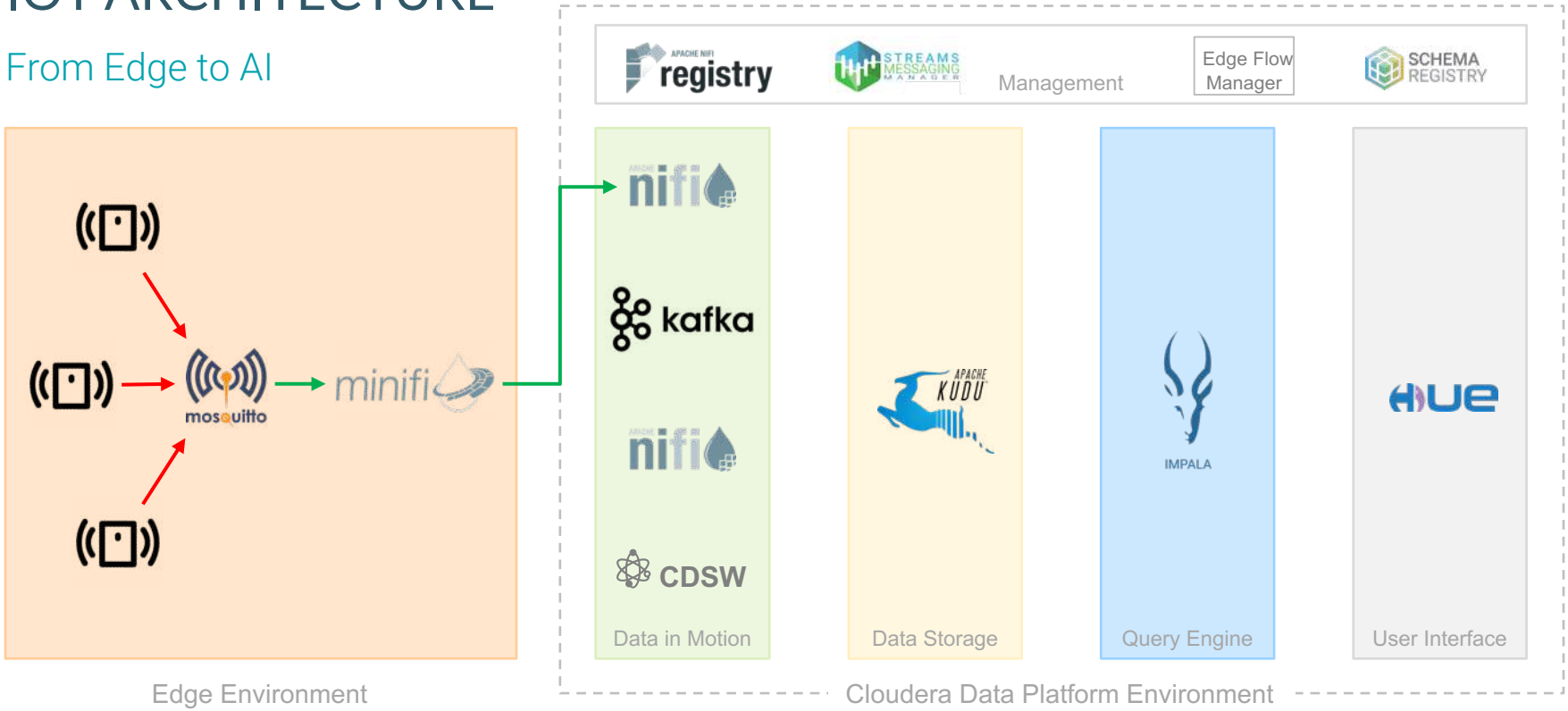
TIME TO PLAY

STEP

1

IOT ARCHITECTURE

From Edge to AI



STEP 1 – OVERVIEW

- Design a flow in Edge Flow Manager to forward data from MQTT Broker to NiFi.
- Version and publish flow to MiNiFi agent.
- Review data queue in NiFi to make sure data is incoming.

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

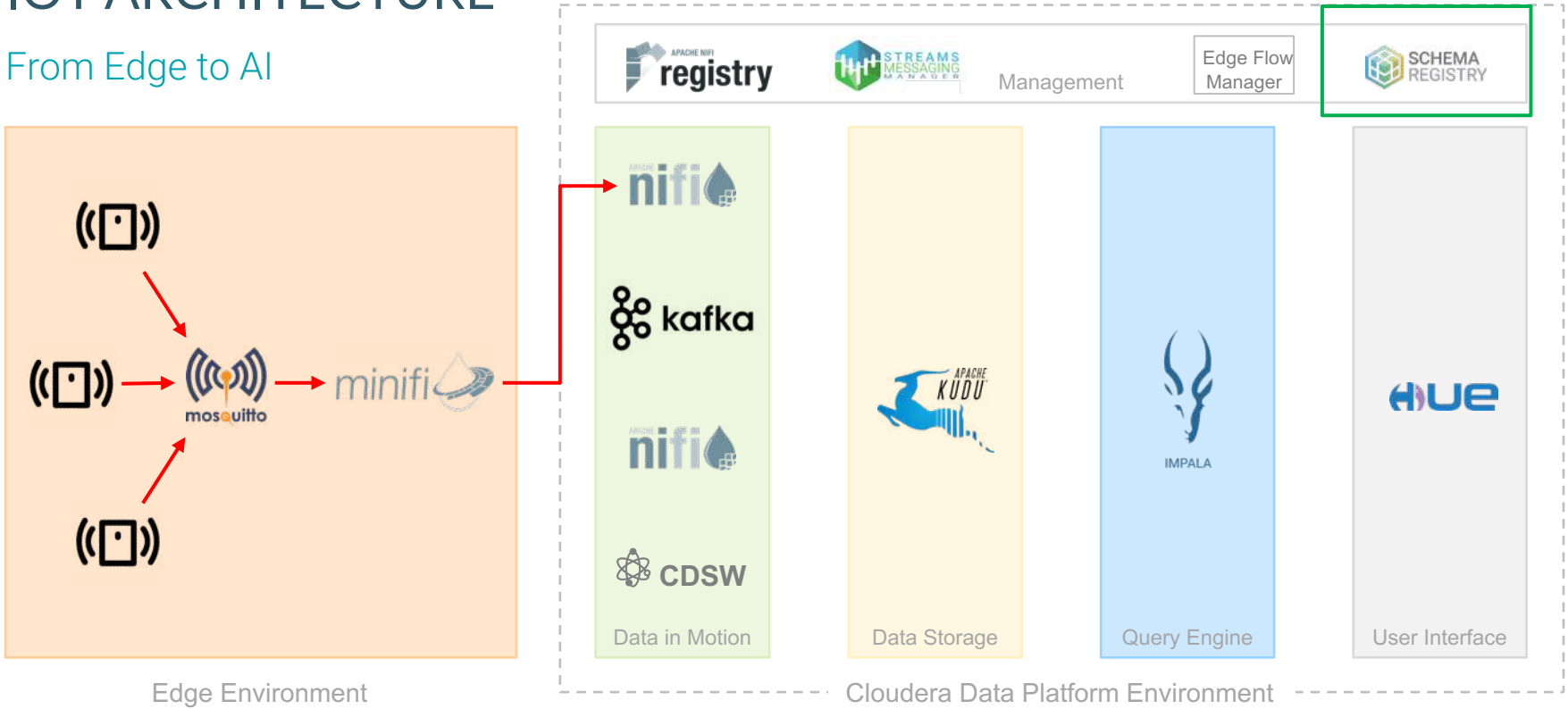
Agent management hub that supports a graphical flow-based programming model to develop, deploy, and monitor edge flows on thousands of MiNiFi agents

Edge Flow Manager

STEP 2

IOT ARCHITECTURE

From Edge to AI



STEP 2 – OVERVIEW

- Define a schema in Schema Registry.

STEP 2 – COMPONENTS

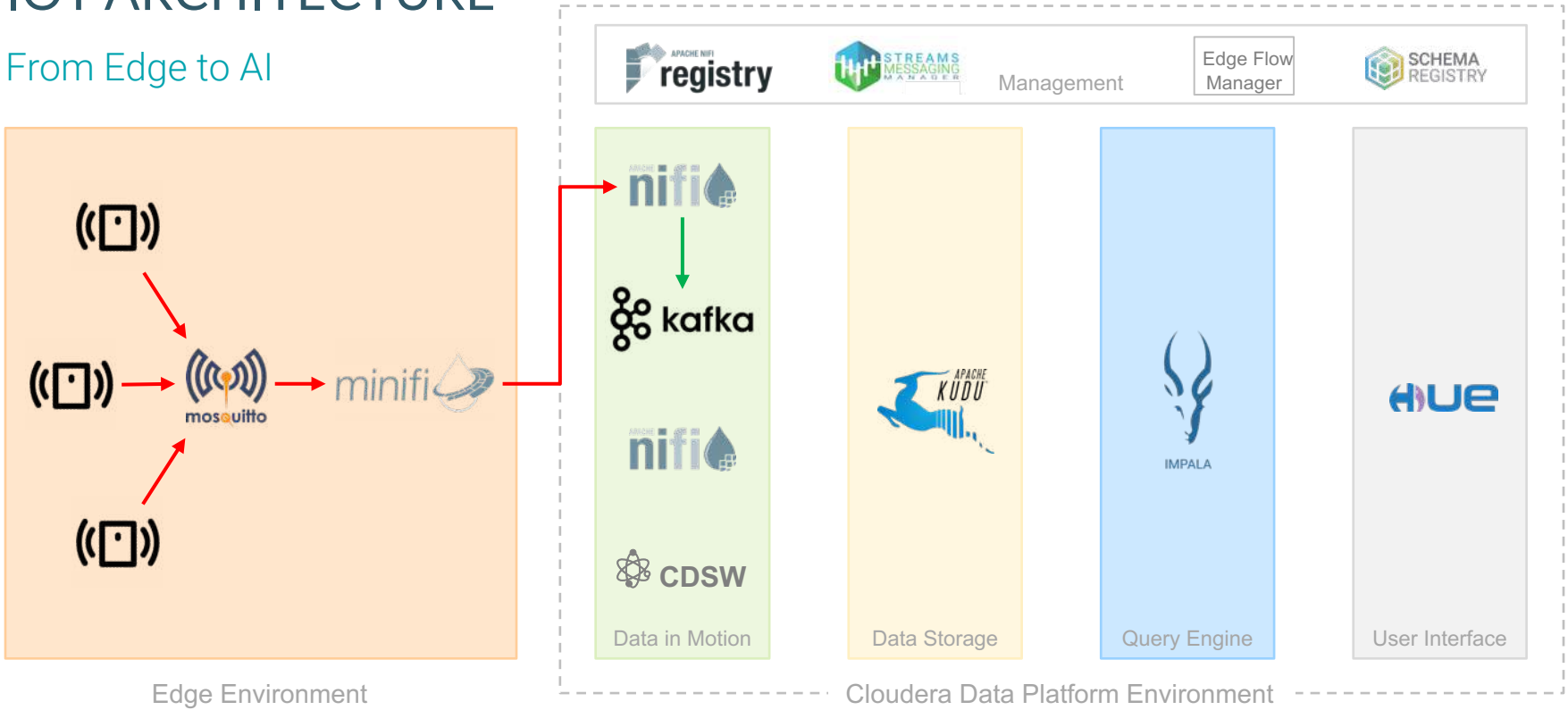
Provides a central repository to store, manage and version data schemas between components, allowing evolution and reusability

STEP

3

IOT ARCHITECTURE

From Edge to AI



STEP 3 – OVERVIEW

- Version a NiFi Process Group, using NiFi Registry
- Design a flow in NiFi to publish the data events to Kafka, including related schema

STEP 3 – COMPONENTS

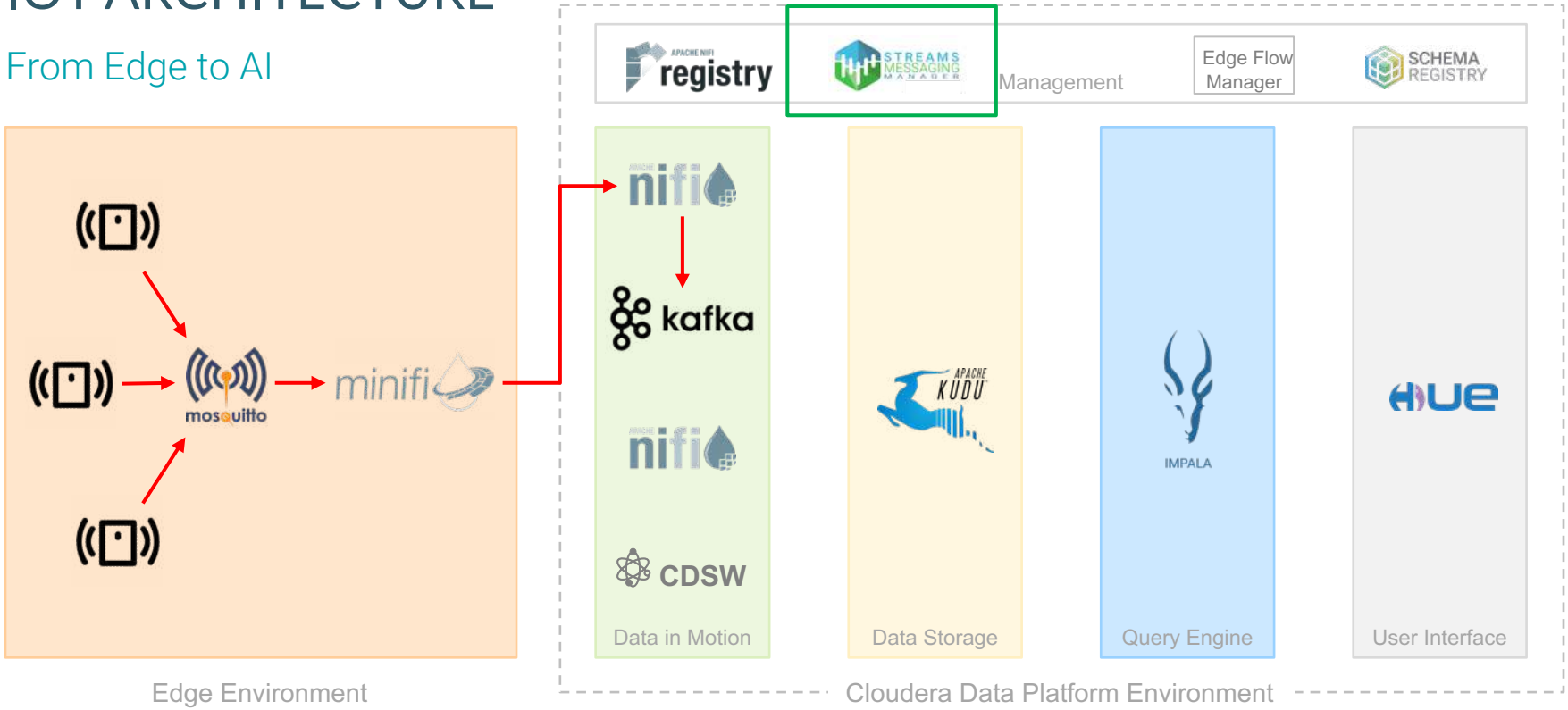
Provides a central location for storage and management of shared resources across one or more instances of NiFi and/or MiNiFi

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 – OVERVIEW

- Check in SMM that data is flowing to Kafka.
- Review ***iot*** topic configuration

STEP 4 – COMPONENTS

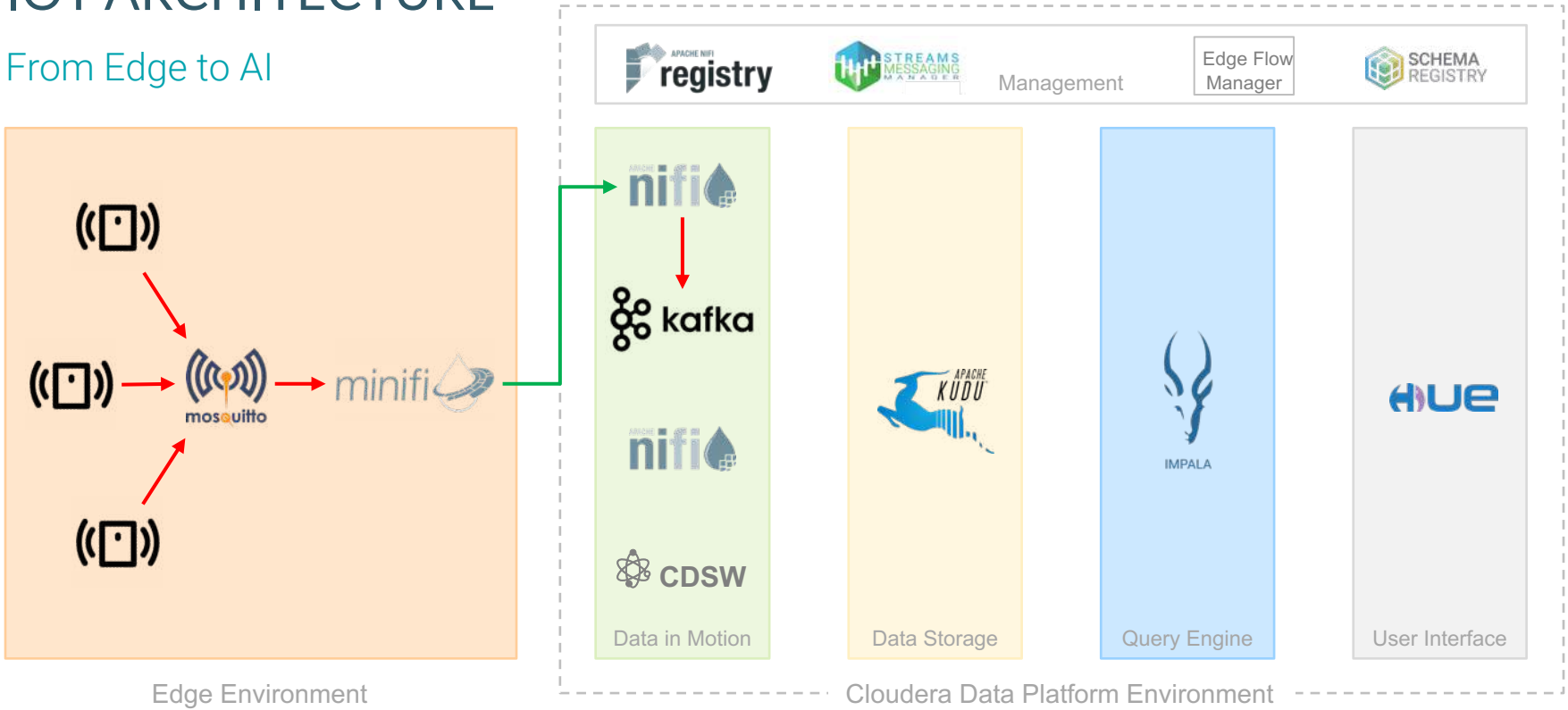
As a monitoring and management tool that provides end-to-end visibility in an enterprise Apache Kafka environments. With SMM, it is possible to gain clear insights about Kafka clusters, like brokers, topics, producers and consumers.

STEP

5

IOT ARCHITECTURE

From Edge to AI



STEP 5 – OVERVIEW

- Update MiNiFi flow to filter noisy/erroneous data, pushing additional logic to the edge

STEP 5 – 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

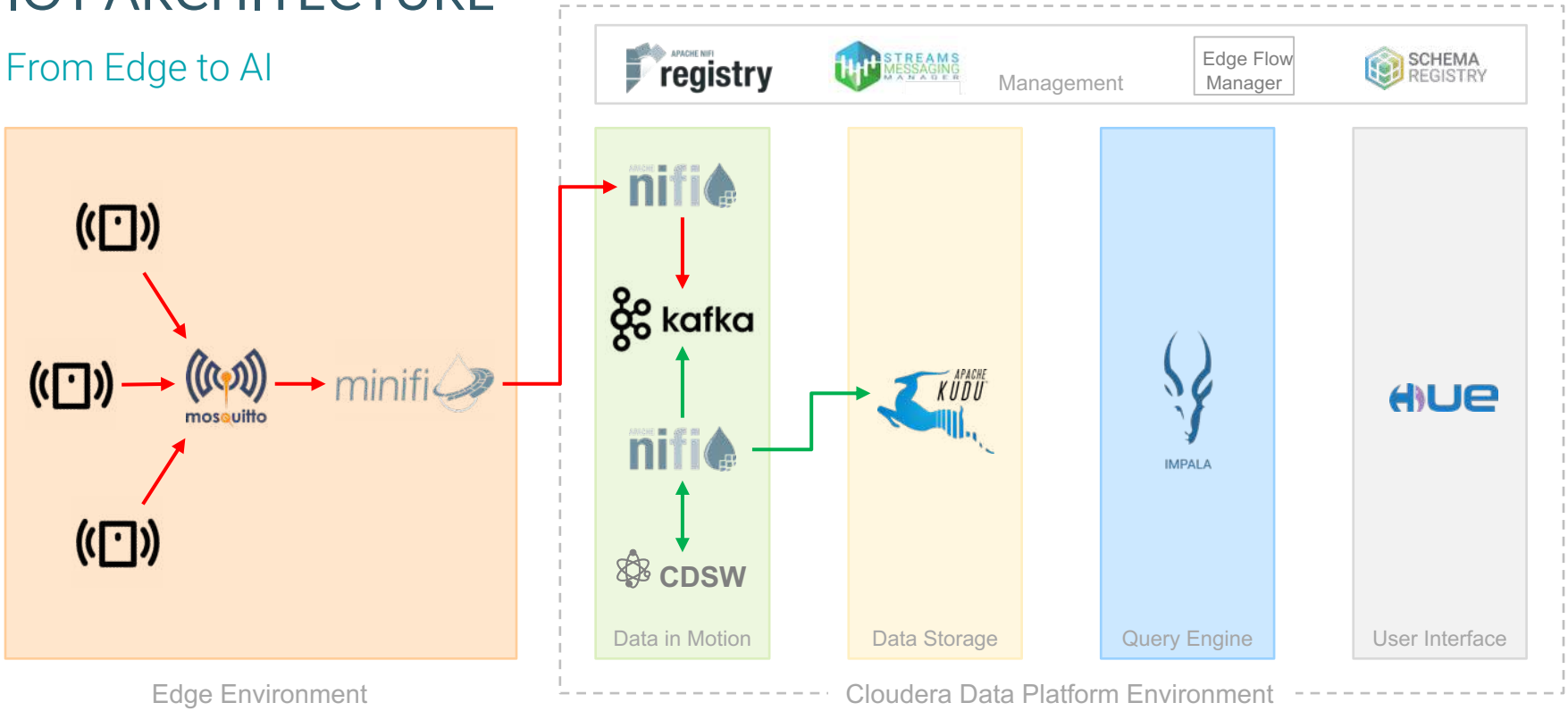
Edge Flow Manager

STEP

6

IOT ARCHITECTURE

From Edge to AI



STEP 6 – OVERVIEW

- From NiFi:
  - Consume Kafka events.
  - Call a Machine Learning model exposed in CDSW to predict device health (if the device is likely to break or not).
  - Store each event in a Kudu table

STEP 6 – COMPONENTS

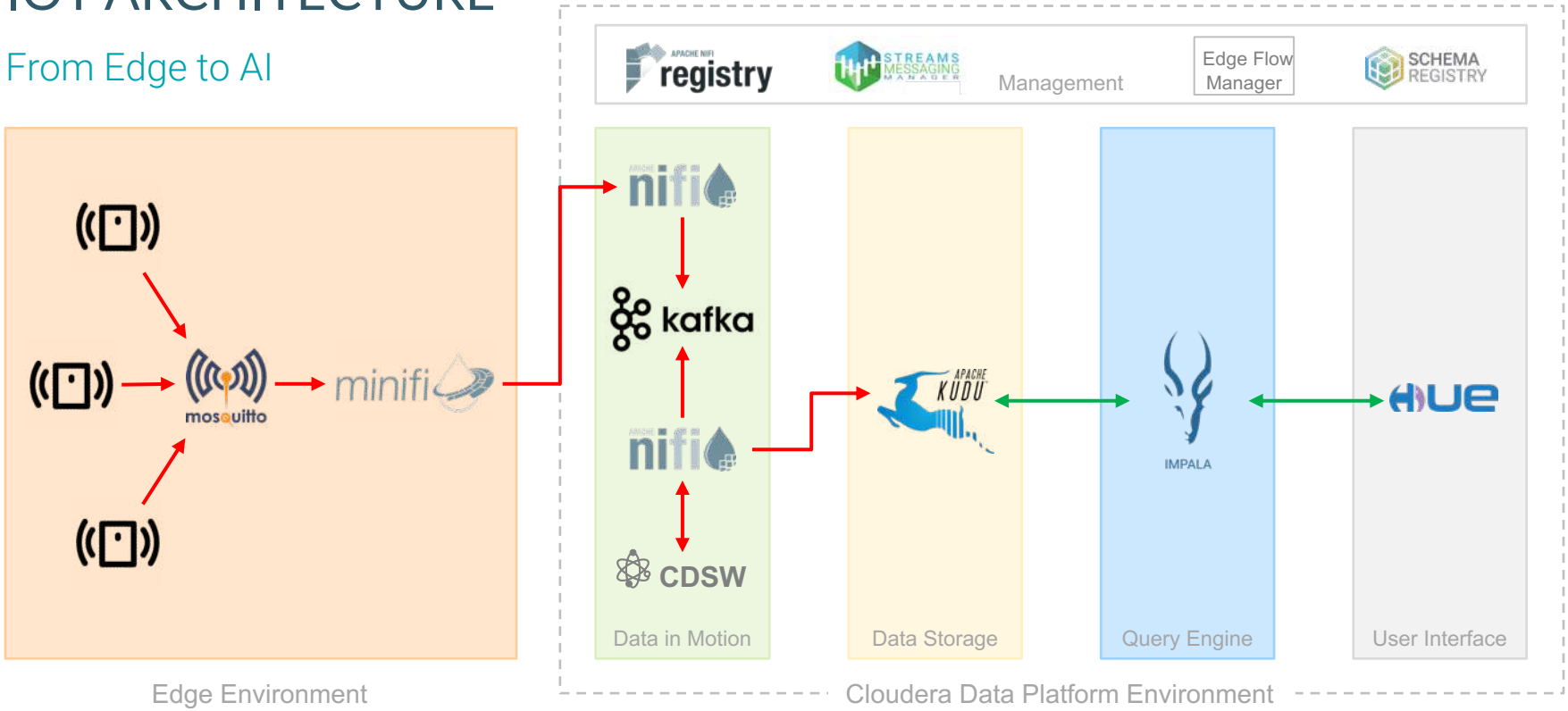
Self-service enterprise data science platform that lets data scientists manage their own analytics pipelines, thus accelerating machine learning projects from exploration to production

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 7

IOT ARCHITECTURE

From Edge to AI



STEP 7 – OVERVIEW

- From Hue, query the data stored in Kudu running an Impala query (SQL)

STEP 7 – COMPONENTS

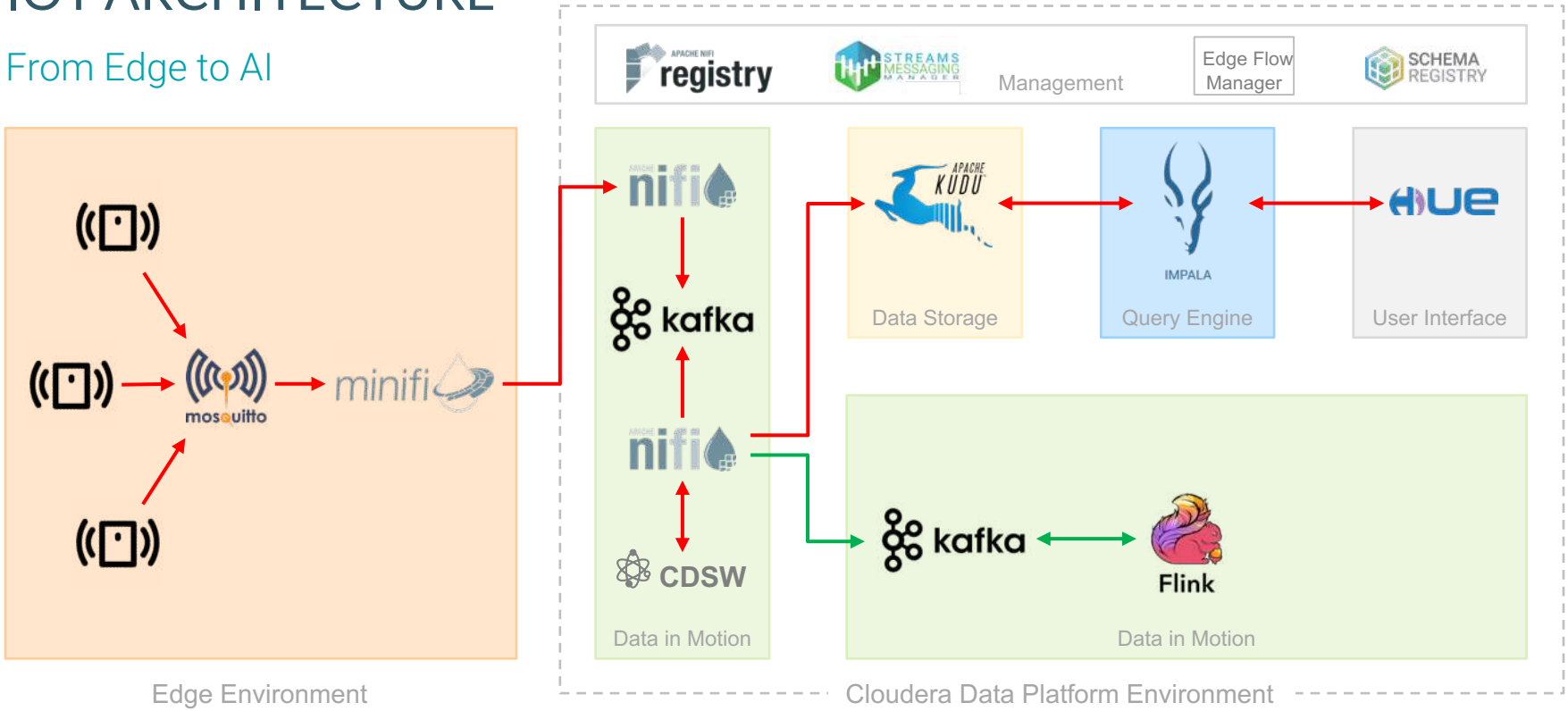
Hue is the open source analytics workbench designed for fast data discovery, intelligent query assistance, and seamless collaboration. Bridge the gap between IT and the business for trusted self-service analytics.

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



STEP 8 – OVERVIEW

- In NiFi:
  - Push scored event data to a new Kafka topic.
- In SMM:
  - Review new Kafka topic
- In Flink:
  - Create a stream data table, using Kafka topic as source
  - Run analytics queries to filter and aggregate data

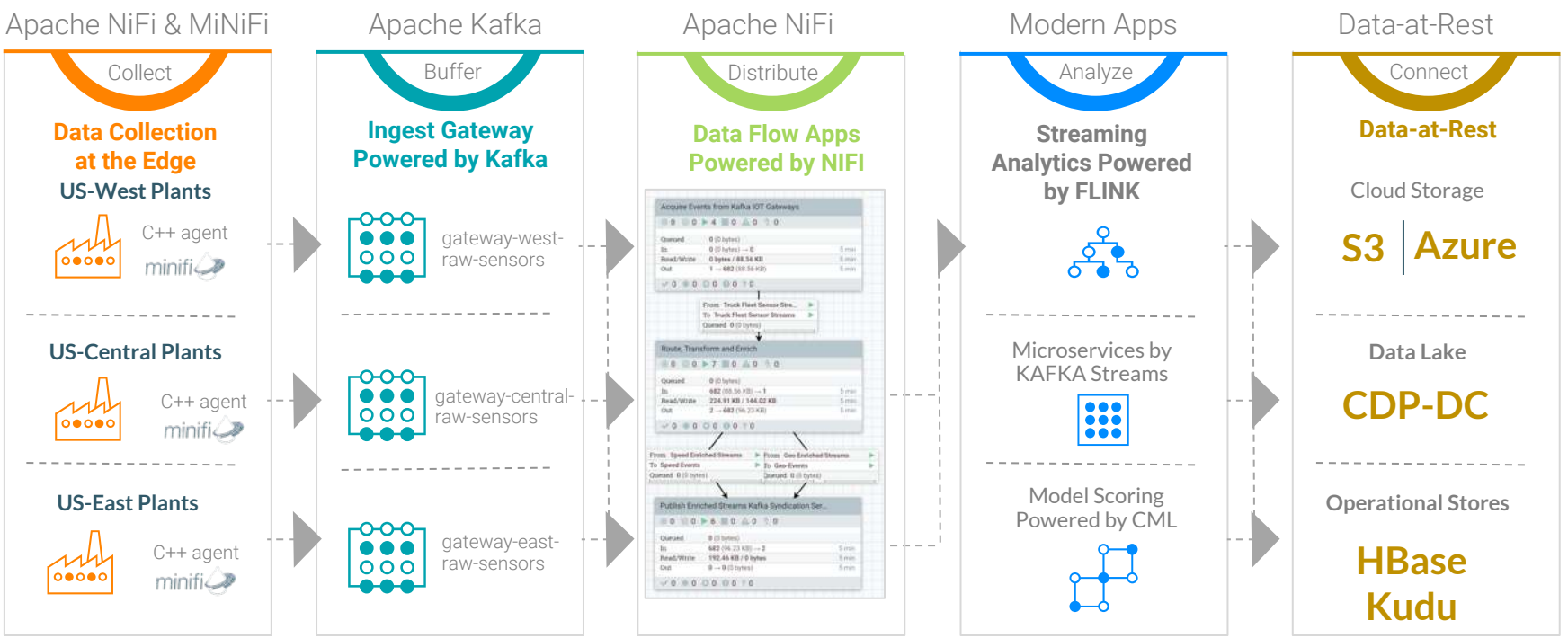
STEP 8 – COMPONENTS

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.

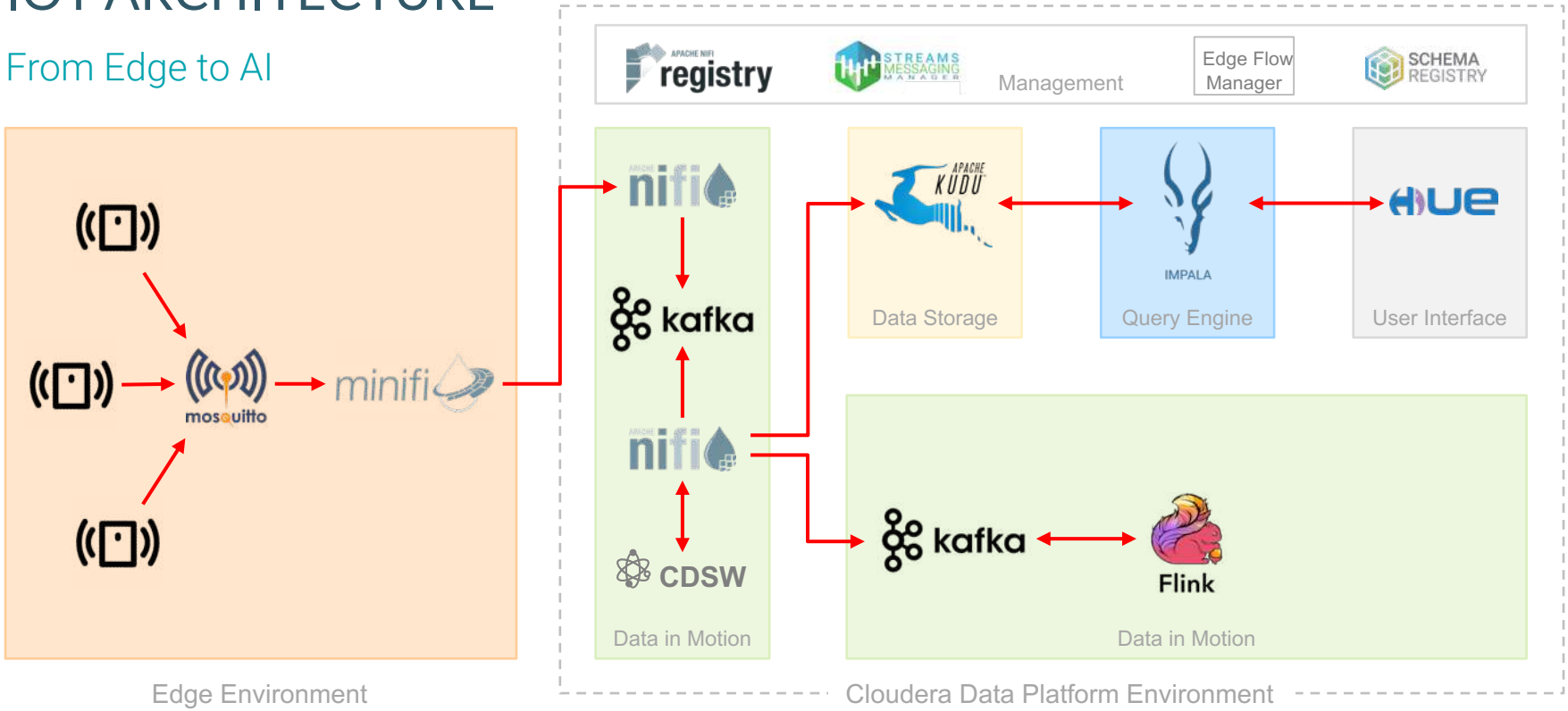
WRAP UP

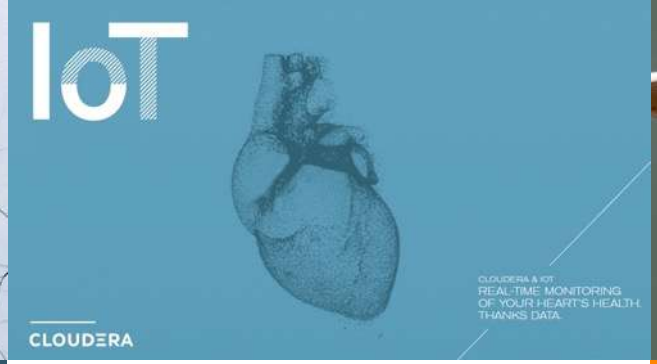
A DATA-IN-MOTION REFERENCE ARCHITECTURE



IOT ARCHITECTURE

From Edge to AI





ENTERPRISE DATA CLOUD

CLOUDERA

CLOUDERA Now

EDGE 2AI

CLOUDERA

SDX

CLOUDERA

THANK YOU