

Red Hat
Summit

Connect

Cloudera e OpenShift: democratizzare e modernizzare l'analitica di scala dei dati col piu' avanzato Data Lakehouse del mercato

Gabriele Folchi
Senior Partner Solution Engineer
Cloudera EMEA
gabriele.folchi@cloudera.com



CLOUDERA



Red Hat and Cloudera Stronger Together

CLUDERA



- Cloudera is committed to Open Source
- CDP allows you to streamline and simplify data management with a single pane of glass.
- Cloudera serves 9 of the 10 top customers in financial services, retail, manufacturing, and telco.
- Leader in open source enterprise IT solutions
- Open source driven developer community
- Trusted enterprise IT partner to the Fortune 500

CLOUDERA COMMITS TO OPEN DATALAKE INNOVATION

Driving Community OSS Innovation, not just “using” it

CLOUDERA'S COMMITMENT TO OPEN SOURCE

2010-2016

The core big data ecosystem

Hadoop, HBase, Hive

Spark, Kafka

NiFi

2017-2019

SDX

Ranger, Atlas

2020+

Cloud Native Transformation

Iceberg

Ozone

Yunikorn

Airflow

Stable & Predictable

Long Term Support (LTS) Release

- Consistent stream of security and bug fixes
- Backward compatibility
- Low-risk upgrade

500+

Logos on 7.1.7 LTS

New Features & Innovation

Short Term Support (STS) Release

- Major component updates
- New features & behavior
- Core component changes

170

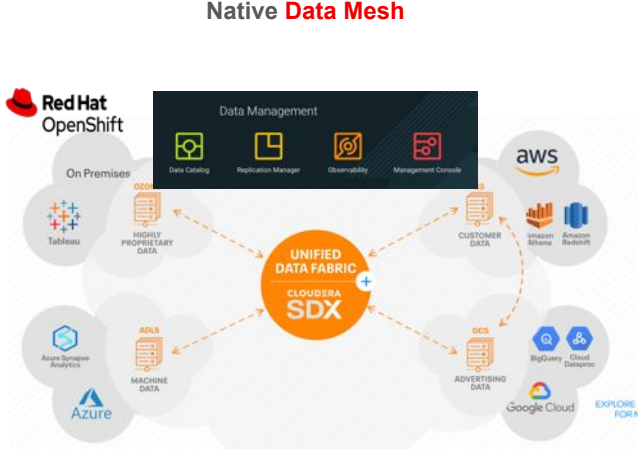
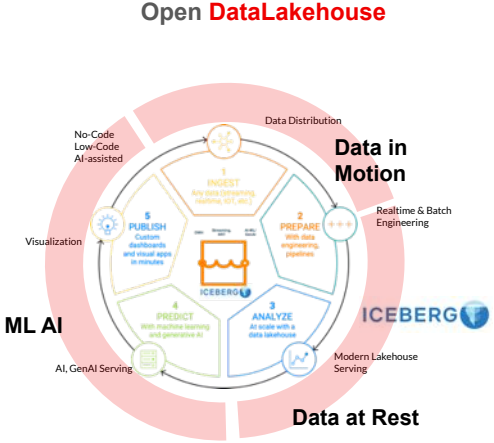
Logos on 7.1.8 STS

CLOUDERA CDP is a Modern Data Platform, ready for Data-Driven Innovation

Right Technology

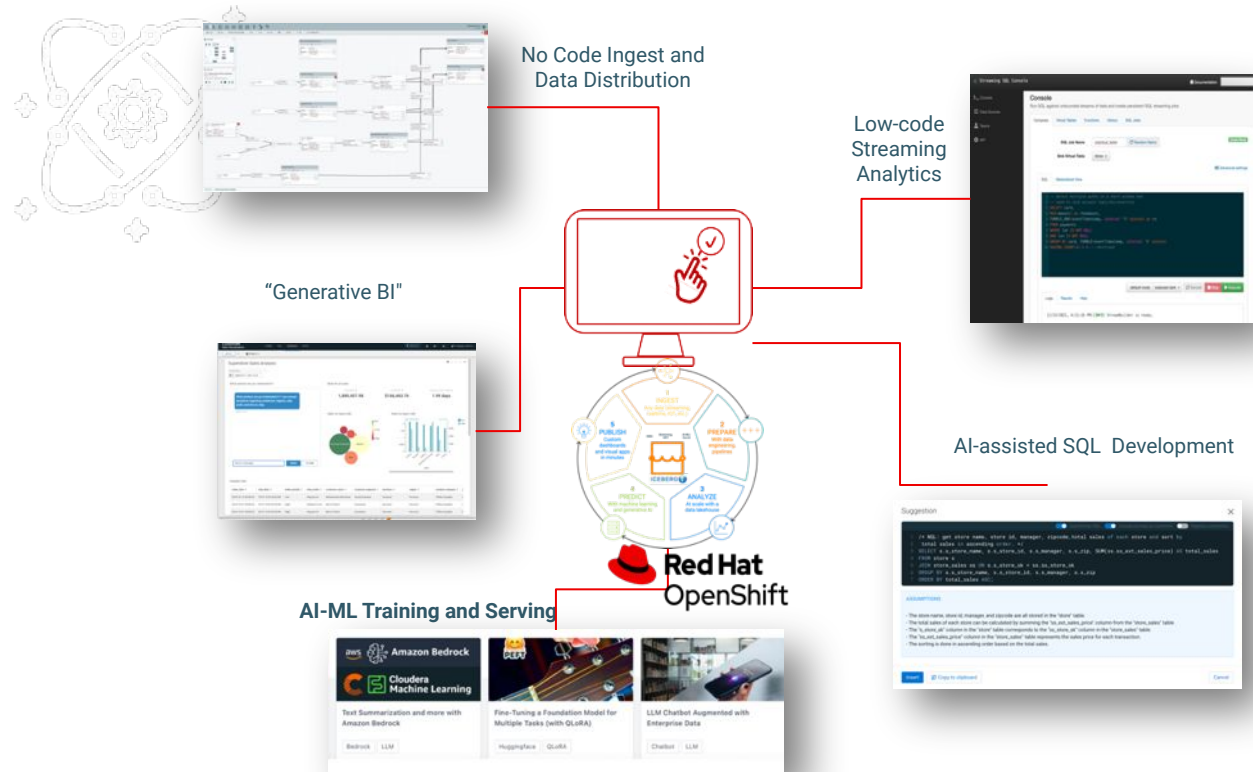
Right Enablers

Right Architecture

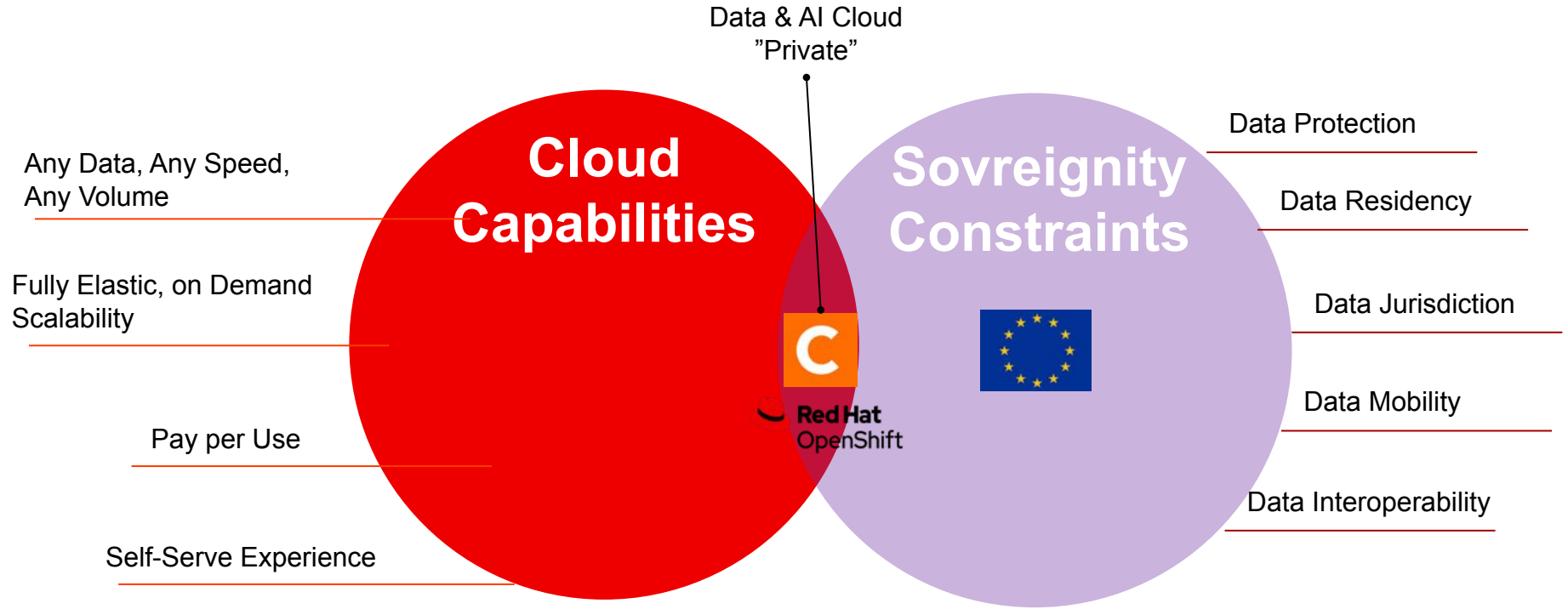


Right Posture

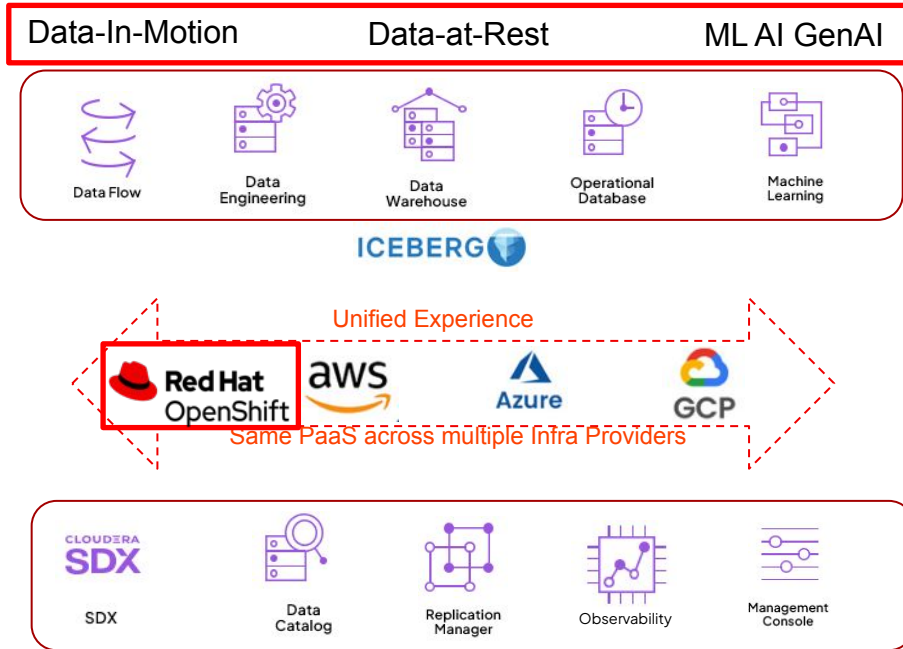
Cloudera on-prem Data Services powered by Openshift: the most advanced data platform on customer premises



Sovereignty and Innovation: a play of capabilities & constraints



Cloudera and OpenShift – Uniquely positioned for Strictest Sovereign Needs



Develop **most advanced** near real-time AI services, GenAI plays

... Under your strict **jurisdiction and residency control**

... without missing a **self-service, PaaS elastic development experience**

... with **open, interoperable standard language & data format** which is reusable anywhere

Cloudera and OpenShift – Uniquely positioned for **most advanced innovation** under Strictest Sovereign Needs

Data-In-Motion

Data-at-Rest

ML AI GenAI

Develop **most advanced** near real-time AI services, GenAI plays



ICEBERG

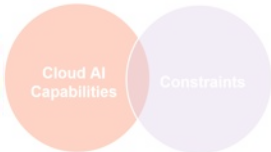
Unified Experience



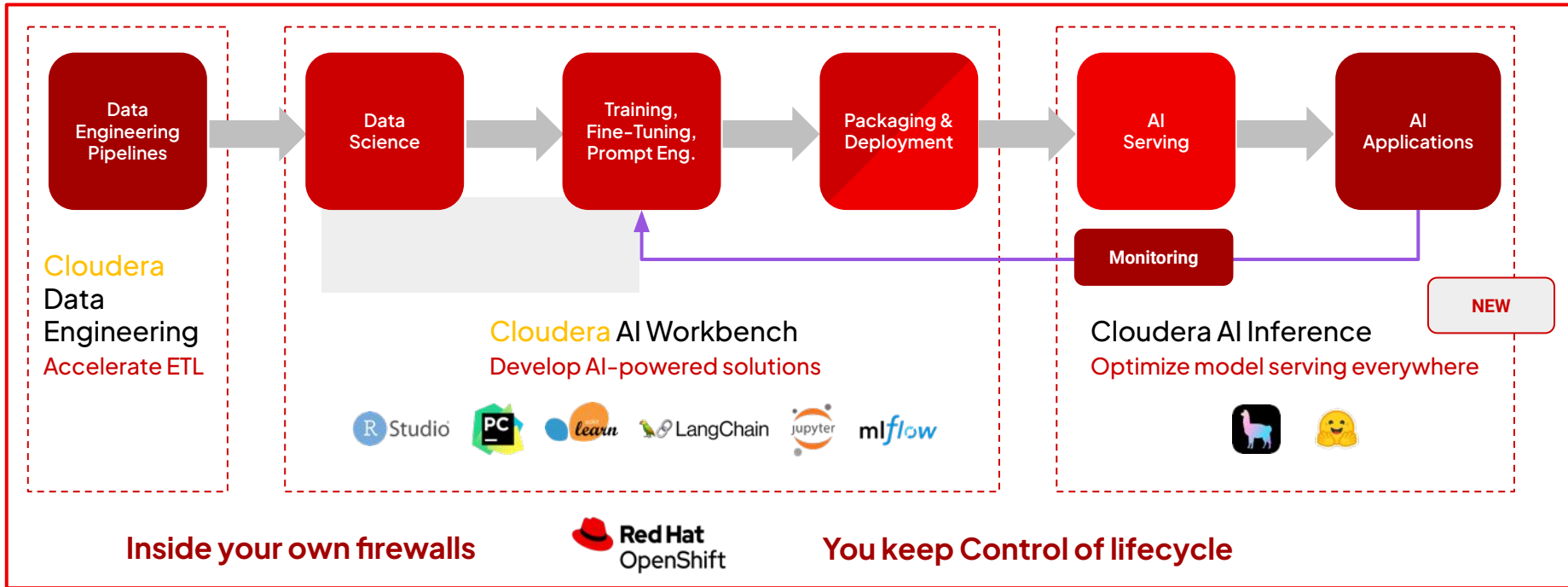
Same PaaS across multiple Infra Provider



<https://cloudera.github.io/Applied-ML-Prototypes/#/cloudera>



Cloudera and OpenShift – Uniquely positioned for **most advanced innovation** under Strictest Sovereign Needs





2nd

Largest financial services group in Southeast Asia

Top 50

Ranked among the World's Safest Banks by Global Finance

One

Of the most highly rated banks in the world

Cloudera Data Platform (CDP) to build a data lake in a private cloud environment

CDP Machine Learning running on Red Hat® OpenShift® to power its enterprise data science platform

Utilized **Data Lakehouse Architecture**

The Challenge

- Pressure from rivals to be more innovative
- Market in Southeast Asia becoming increasingly digitized
- Consumer looking for better experiences

The Goal

- Leverage private cloud for data platform
- Deliver secure, controlled, personal data experience.
- Use (AI/ML) to make data-driven decisions to improve customer experience

The Results

Using Red Hat OpenShift with Cloudera CDP

100

Million SGD

OCBC Bank's Next Best Conversation platform helped earn annually by using data to curate personalized experiences for customers.

2X

campaign conversion rates

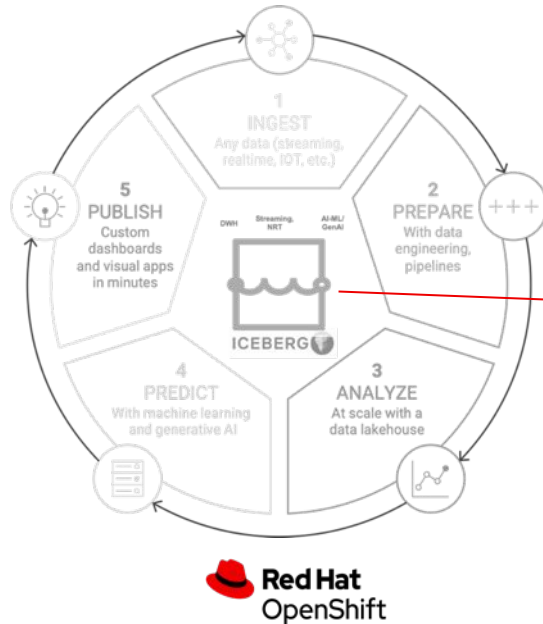
Is how much Next Best Conversation increased campaign conversion rates with more relevant personalized offers.

10%

Chatbot handles online transactions

Banking customers enjoy faster transactions with the aid of the bank's chatbots, which handles 10% of its customers interactions online

Cloudera and OpenShift – Uniquely positioned for **most advanced innovation** under Strictest Sovereign Needs



Consolidate proprietary DSS RDBMS into DLH with Iceberg

Modernize existing DL DWH with Iceberg

PRACTITIONER



Snapshot Isolation



Time Travel / Table Rollback



In-place Table Evolution



Table Maintenance



ACID Transaction

ARCHITECT



Engine Agnostic



SDX Integration



Ease of Adoption



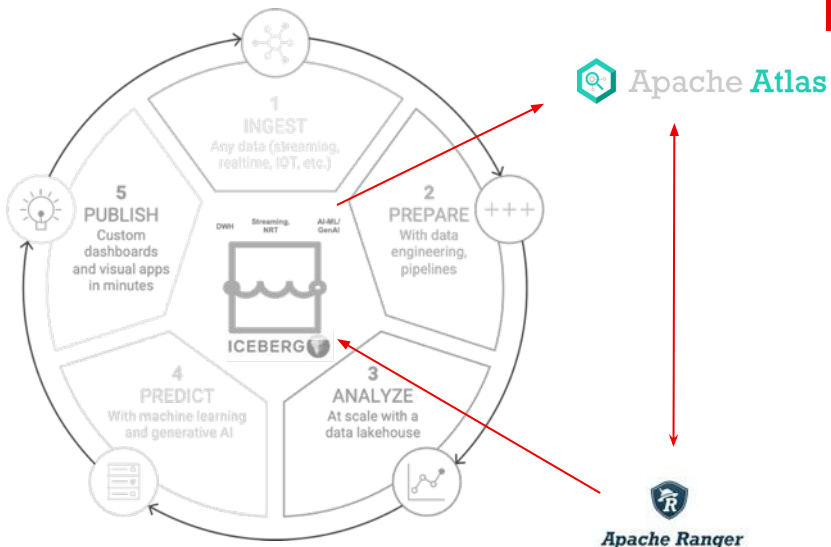
Multi-Cloud



Performance / Scalability

Cloudera and OpenShift – Uniquely positioned for **most advanced innovation** under Strictest Sovereign Needs

Data Stewardship and Governance out of the box



All Assets are **automatically harvested** (with their **lineage**) and **searchable**

out-of-the-box **Tag-based policy management** (ABAC)

All Assets are **securable** from a single Policy Engine

CSM Operator - Key Capabilities



Deploys & Manages Kafka

Create Kafka clusters in minutes in your Kubernetes infrastructure

```
NAME ↑
demo-kafka-cruise-control-54cd6b6bfd-9ktmz
demo-kafka-demo-kafka-first-pool-0
demo-kafka-demo-kafka-first-pool-1
demo-kafka-demo-kafka-first-pool-2
demo-kafka-entity-operator-cbc9795f-n8c95
demo-kafka-zookeeper-0
demo-kafka-zookeeper-1
demo-kafka-zookeeper-2
kafka-admin
strimzi-cluster-operator-7676fdd8cc-kfq5f
```



Includes Cruise Control

Cruise Control takes care of Kafka health and cluster rebalances

```
Name: my-rebalance
Namespace: myproject
Labels: strimzi.io/cluster=my-cluster
Annotations: API Version: kafka.strimzi.io/v1alpha1
Kind: KafkaRebalance
Metadata:
# ...
Status:
Conditions:
  Last Transition Time: 2020-06-04T14:36:11.900Z
  Status: ProposalReady
  Type: State
Observed Generation: 1
Optimization Result:
  Data To Move MB: 0
  Excluded Brokers For Leadership:
  Excluded Brokers For Replica Move:
  Excluded Topics:
  Intra Broker Data To Move MB: 12
  Monitored Partitions Percentage: 100
  Num Intra Broker Replica Movements: 0
  Num Leader Movements: 24
  Num Replica Movements: 55
  On Demand Balancedness Score After: 82.91290759174306
  On Demand Balancedness Score Before: 78.01176356230222
  Recent Windows: 5
  Session Id: a4f833bd-2055-4213-bfdd-ad21f95bf184
```



Easily Scale Up/Down

Manually scale up/down by adjusting # of brokers in Kafka node pools and applying the change:

```
kind: KafkaNodePool
spec:
  replicas: 3 4
```

> kubectl apply -f prod.yaml

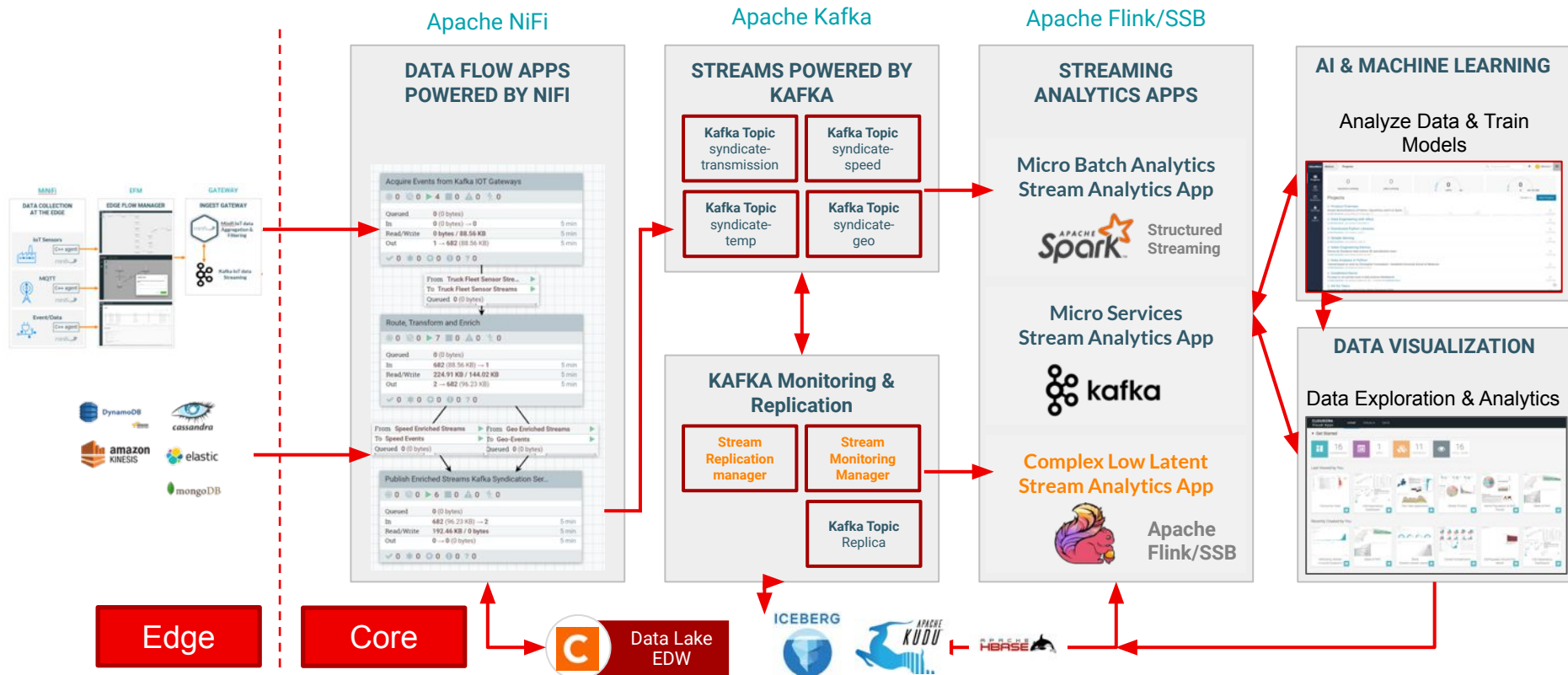
Or use the kubectl scale command:

```
> kubectl scale \
  kafkanodepool prod-node-pool \
  --replicas=4
```

And let Cruise Control rebalance it:

```
kind: KafkaRebalance
spec:
  mode: add-brokers
  brokers: [3]
```

Data in Motion: e2e enablers for most advanced streaming analytics



Cloudera and OpenShift – Uniquely positioned for **most advanced innovation** under Strictest Sovereign Needs

<https://www.cloudera.com/products/cloudera-data-platform.html?tab=1>

Data-In-Motion

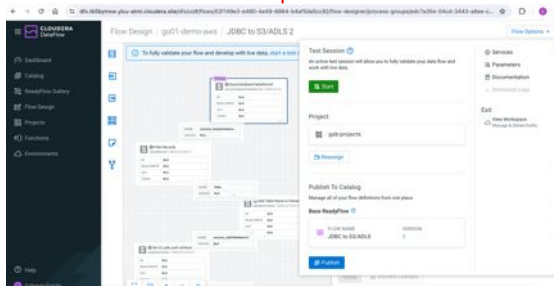
Data-at-Rest

ML AI GenAI

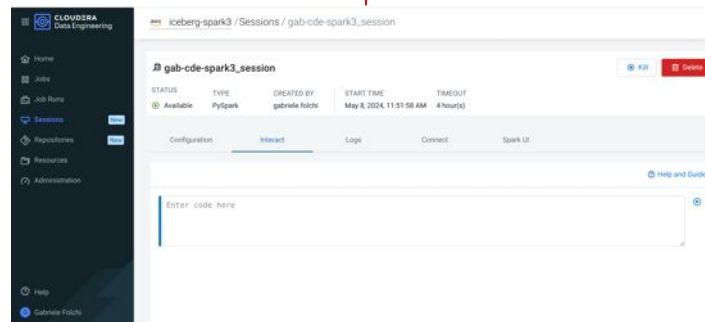
Data Scientist



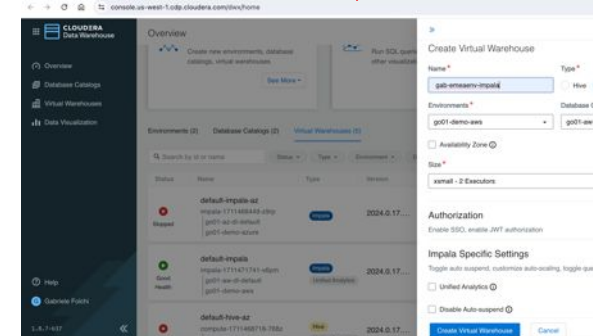
**Self-Serve Dev & Deploy
MPP SQL procs**



Self-Serve Dev & Deploy DataFlows



**Self-Serve Dev & Deploy Spark and
Airflow apps**



Red Hat
Summit

Connect

Thank you

Partner logo
here

