

INTRODUCING CLUDERA DATAFLOW (CDF)

Dinesh Chandrasekhar

Product Marketing Lead, Data-in-Motion BU

Cloudera

@AppInt4All

George Vetticaden

Product Management Lead, Data-in-Motion BU

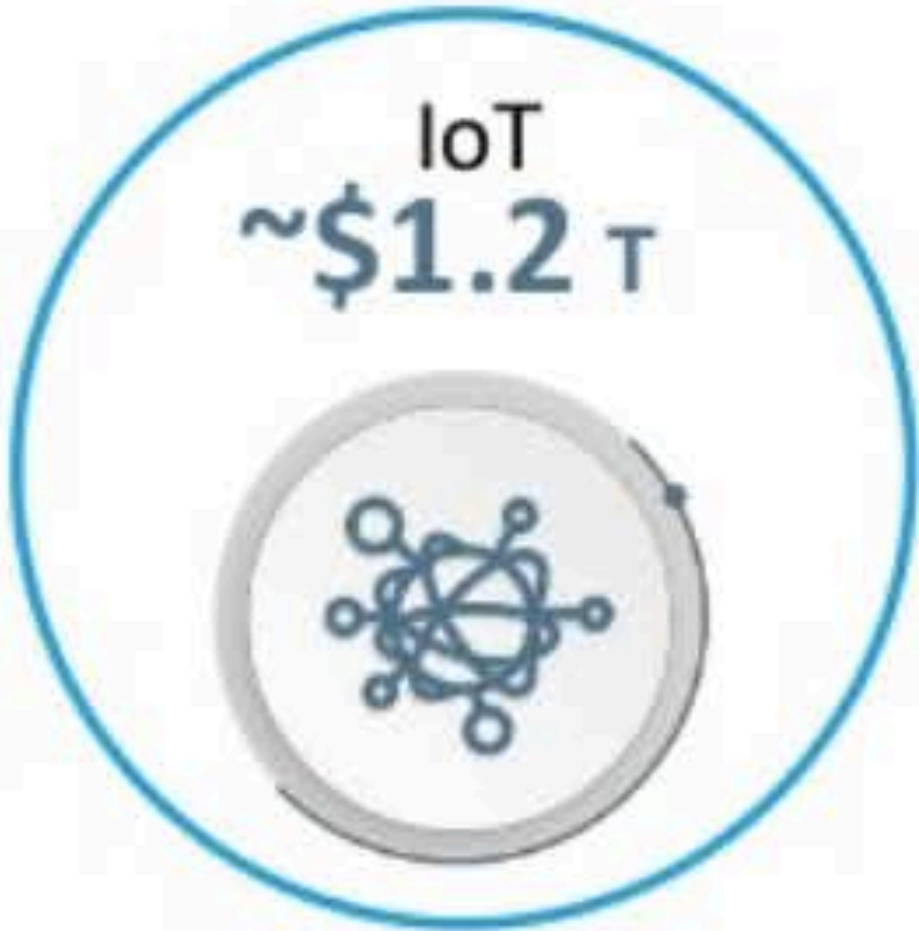
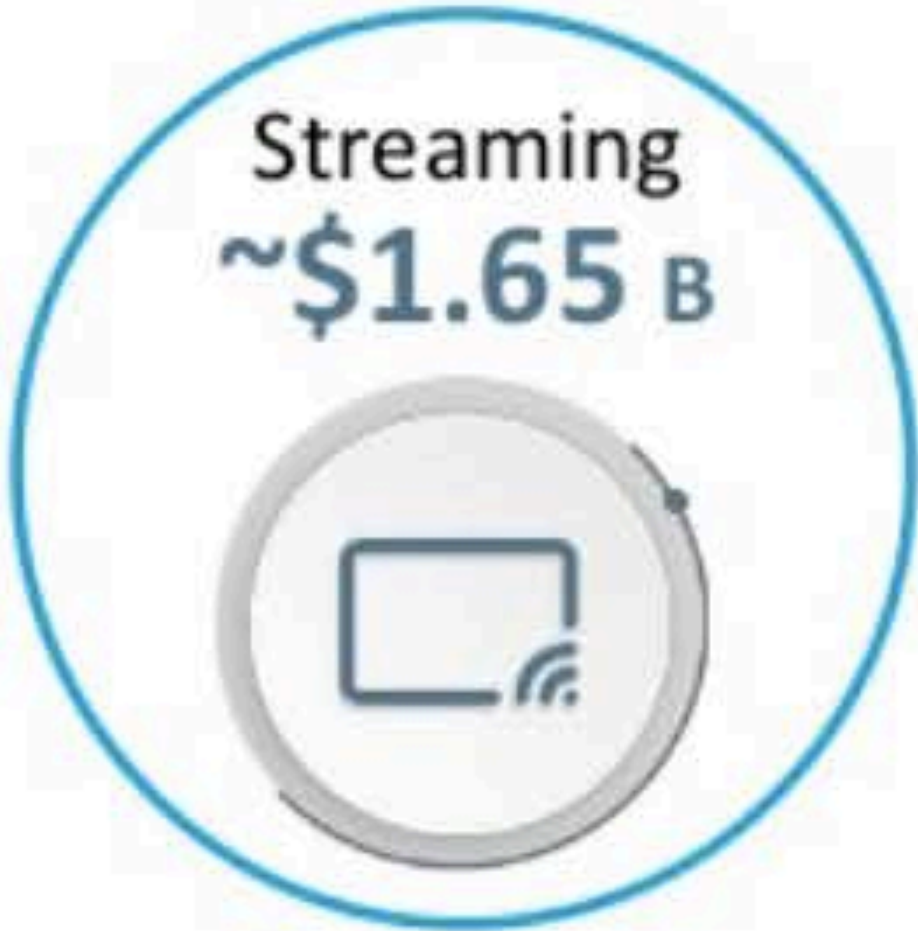
Cloudera

@gvetticaden





MARKET OPPORTUNITIES





IOT MARKET

24.9B

By 2024 more than 24.9 Billion IoT connections will be established

\$70B

An estimated \$70 billion will be spent by global manufacturers on IoT solutions in 2020

646M

An estimated 646 million healthcare devices (excluding fitness trackers and wearable devices) will be connected by 2020

78%

An estimated 78% of cars shipped globally will be built with hardware that connects to the internet by 2020

50%

50% of decision-makers in IT, services, utilities, and manufacturing have either deployed IoT, or will deploy it in the next 12-24 months



KEY CUSTOMER CHALLENGES



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 DATAFLOW

Cloudera DataFlow Data-in-Motion Platform



EDGE DATA MANAGEMENT

Edge data collection, routing and monitoring

Apache MINIFI
Edge Flow Manager



FLOW MANAGEMENT

Enterprise data ingestion, transformation and enrichment

Apache NIFI
NIFI Registry



STREAM PROCESSING

Real-time streams processing at IoT scale

Apache Kafka
Streams Messaging Manager



STREAMING ANALYTICS

Predictive analytics and real-time insights

Apache Storm
Streaming Analytics Manager
Kafka Streams



ENTERPRISE SERVICES

Provisioning, Management and Monitoring

Unified Security
Edge-to-Enterprise Governance
Single Sign-on

Schema Registry

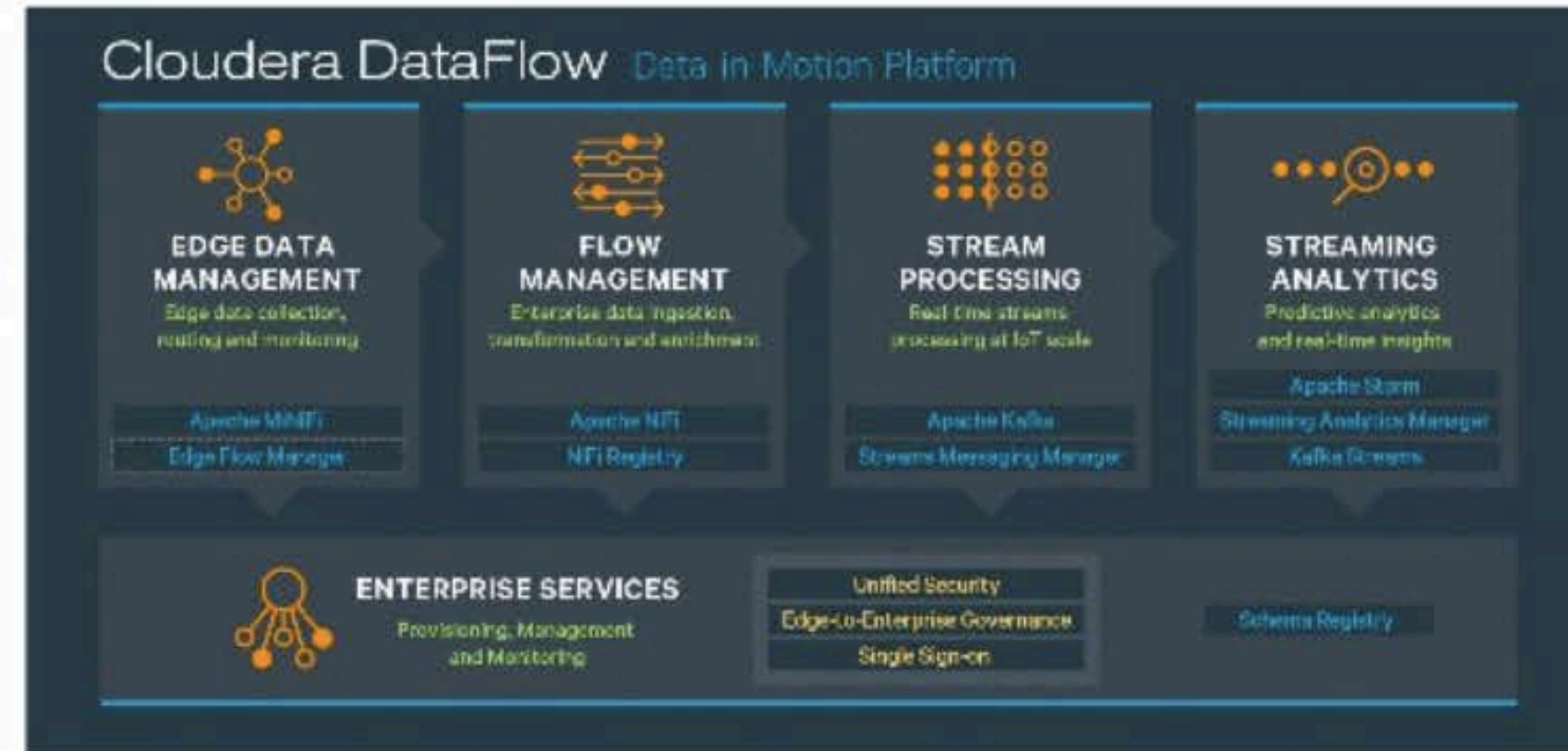
WHAT IS CLUDERA DATAFLOW (CDF)?

Cloudera DataFlow (CDF) is a scalable, real-time streaming data platform that collects, curates, and analyzes data so customers gain key insights for immediate actionable intelligence.

HISTORY OF CDF

Data-in-Motion:

- Comprehensive real-time streaming data platform
- Manage data-in-motion from edge-to-enterprise
- Power IoT-scale streaming architectures



Bring this to the edge with
connected platforms

Enable next generation
Modern Data Architecture

Tomorrow:
Edge-to-AI

Mid-2000's
NiFi was developed
and used at NSA

2015
Onyara is acquired
HDF is born

2018
Strong Streaming Platform

- Support for Kafka 2.0
- SMM is introduced

2019
Cloudera merger
Enable Edge Intelligence



COMMON USE CASES

Data Movement

Optimize resource utilization by moving data between data centers or between on-premises infrastructure and cloud infrastructure

Optimize Log Collection & Analysis

Optimize log analytics solutions by using CDF as a single platform to collect and deliver multiple data sources

Gain key insights with Streaming Analytics

Accelerate big data ROI by analyzing streaming data for patterns, comparing with ML models and delivering actionable intelligence

Single view / 360° view of customer

Ingest, transform and combine customer data from multiple sources into a single data view / lake

Stream Processing

Combine multiple streams of data in real-time, enrich the data and route it to different end points based on rules

Capture IoT Data

Ingest sensor data from IoT devices and stream it for further processing and comprehensive analysis



COMMON IOT USE CASES BY INDUSTRY

Top 5
Use cases



Public Sector

Transportation

Utilities

Healthcare

Manufacturing

Retail

- IoT is a **\$1.13T** market opportunity in 2021.
- Americas - **\$329B** IoT spending. Manufacturing and Transportation are top industries, accounting for 26% of total spending.
- APAC - **\$500B** IoT spending. Manufacturing, Utilities and Transportation are top industries.
- EMEA - **\$264B** IoT spending. Manufacturing is top industry, powered by Industry 4.0 initiatives.
- Worldwide IoT Analytics and Information Management Market = **\$573M**

CUSTOMERS



Improving Healthcare with SMART data

CHALLENGE

Combine multi-format data streams, with hundreds of sources, into one platform

- Needed a platform that could combine multi-format data streaming
- Data scarcity & latency problems
- Machine learning & data science

IMPACT

Lack of medical expertise around patient care, post surgery

- Patient Code Blue status
- Possible cardiac arrest 4–6 hours post surgery

SOLUTION

Cloud-based systems architected to deliver SMART data, using HDP and CDF

- First to deliver SMART real-time streaming data
- Clearsense's Inception™ product enables fast decisions for clinicians
- Customers have access to all data sources with HDP & CDF

RESULT

Mission-critical data and relevant insight for 2,000 rural providers

- Mission critical data is now available for doctors to make critical decisions
- Cost efficiencies led to access for 2,000 rural providers
- Real-time data helps prevent "Code Blue"





Positioning technology products & services empower companies worldwide

CHALLENGE

Provide accurate data for small carriers to improve business results

- 95% of small carriers (less than 50 trucks) have a deficit of data available
- Estimated data, price points and revenue base opportunity for controlling fuel cost
- Understanding of freight and lane movement

IMPACT

Continuing on current path would slow organizational growth and impact customers

- Being unable to predict weather patterns would lead to delays and decreased product quality
- Operational inefficiencies prevent reaching business revenue goals, lack of insights
- Loss of product during transportation

SOLUTION

Big Data in the Cloud with HDP, CDF, and Microsoft Azure

- Leveraging big data powering Blockchain, with machine learning, to revolutionize Transportation and Logistics industries
- Analyzed fuel data; can consolidate data set for small carriers to generate community data lake

RESULT

Double digit revenue increase, year over year

- Managing for 4 million trucks daily
- \$31 billion dollars in freight movement guides customers to profitability
- Blockchain driven architecture

PRODUCT OVERVIEW

CLOUDERA DATAFLOW

Cloudera DataFlow Data-in-Motion Platform



EDGE DATA MANAGEMENT

Edge data collection, routing and monitoring

Apache MINIFI
Edge Flow Manager



FLOW MANAGEMENT

Enterprise data ingestion, transformation and enrichment

Apache NIFI
NIFI Registry



STREAM PROCESSING

Real-time streams processing at IoT scale

Apache Kafka
Streams Messaging Manager



STREAMING ANALYTICS

Predictive analytics and real-time insights

Apache Storm
Streaming Analytics Manager
Kafka Streams



ENTERPRISE SERVICES

Provisioning, Management and Monitoring

Unified Security
Edge-to-Enterprise Governance
Single Sign-on

Schema Registry

CLOUDERA DATAFLOW

Data-in-motion platform



EDGE DATA MANAGEMENT

Edge data collection,
routing and monitoring

Apache MiNiFi

Edge Flow Manager

EDGE DATA MANAGEMENT

- Edge data collection powered by Apache MiNiFi
- MiNiFi – smaller footprint than NiFi
 - Guaranteed delivery
 - Data buffering
 - Prioritized queuing
 - Flow-specific QoS
 - Data provenance
 - Designed for extension
 - C++ / Java agents
- Designed for IoT



CLOUDERA DATAFLOW

Data-in-motion platform



EDGE DATA MANAGEMENT

Edge data collection,
routing and monitoring

Apache MiNiFi

Edge Flow Manager



FLOW MANAGEMENT

Enterprise data ingestion,
transformation and enrichment

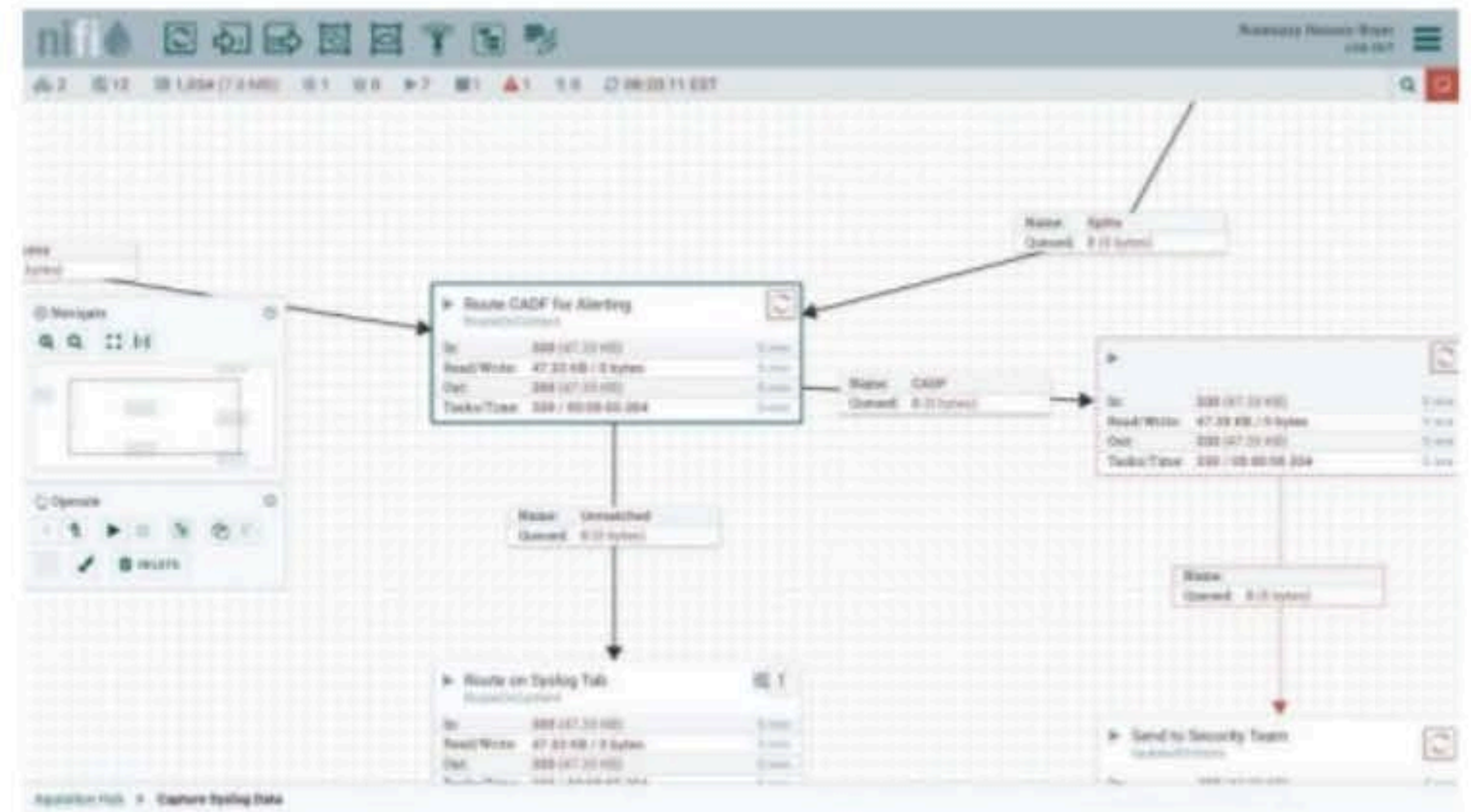
Apache NiFi

NiFi Registry

FLOW MANAGEMENT



- Web-based user interface
- Highly configurable
- Out-of-the-box data provenance
- Designed for extensibility
- Secure
- NiFi Registry
 - DevOps support
 - FDLC
 - Versioning
 - Deployment



280+ PROCESSORS FOR DEEPER ECOSYSTEM INTEGRATION



All Apache project logos are trademarks of the ASF and the respective projects.

CLOUDERA DATAFLOW

Data-in-motion platform



EDGE DATA MANAGEMENT

Edge data collection,
routing and monitoring

Apache MiNiFi

Edge Flow Manager



FLOW MANAGEMENT

Enterprise data ingestion,
transformation and enrichment

Apache NiFi

NiFi Registry



STREAM PROCESSING

Real-time streams
processing at IoT scale

Apache Kafka

Streams Messaging Manager

Streaming Analytics Reference Architecture

Kafka is Everywhere. Critical Component of Streaming Architectures

Clip slide

Kafka Producers

Data Collection
at the Edge

US West Fleet



US Central Fleet

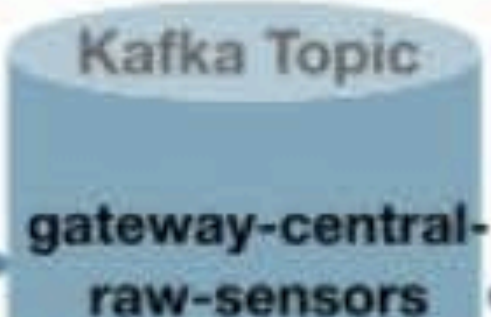


US East Fleet



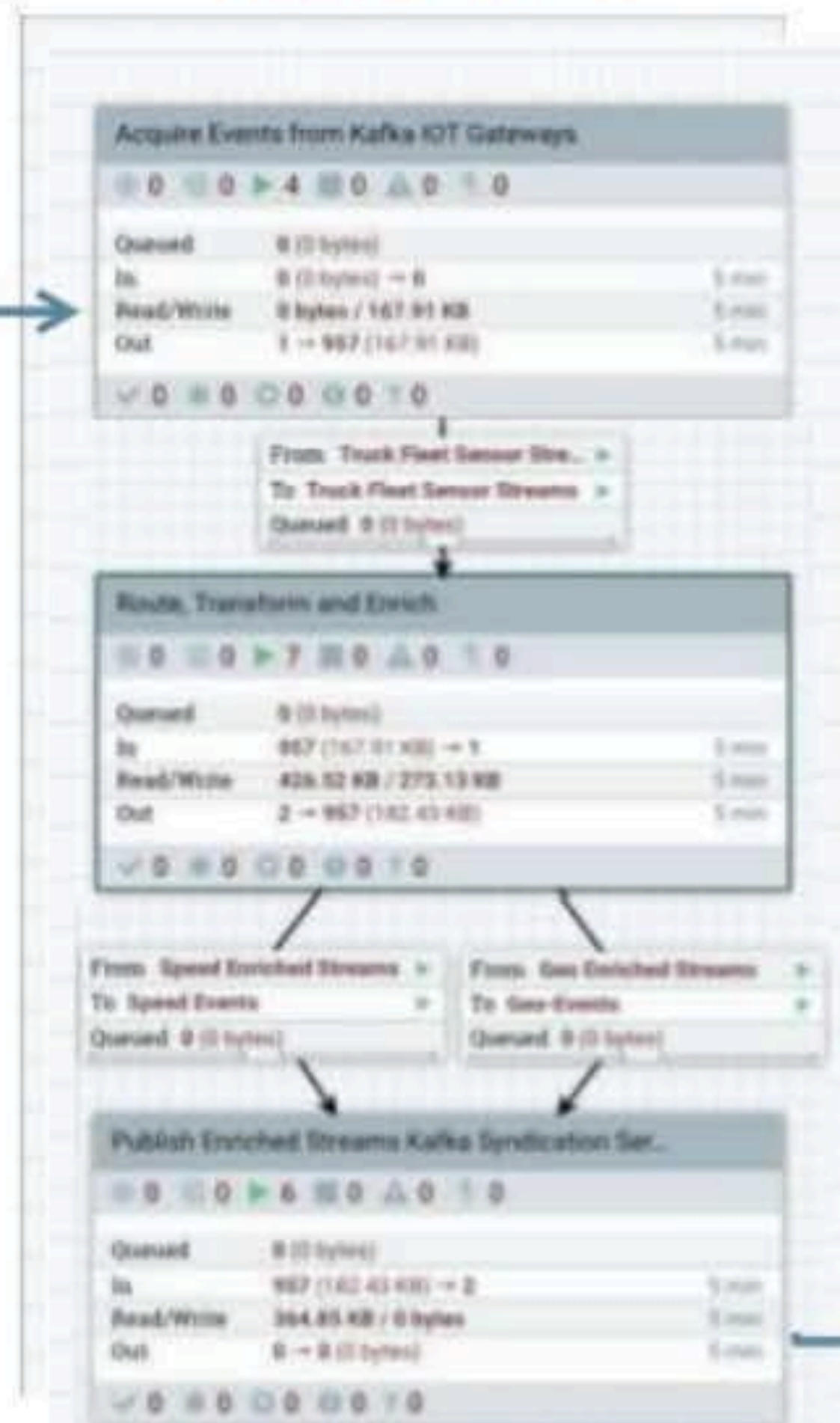
Kafka Topics

IOT Ingest Gateway
Powered by Kafka



Kafka Consumers & Producers

Data Flow Apps
Powered by NiFi



Kafka Topics

Data Syndication
Services Powered by
Kafka



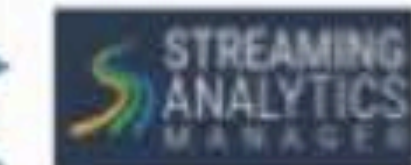
Kafka Consumers

Subscribing
Streaming Analytics
Apps

Analytics App 1



Analytics App 2



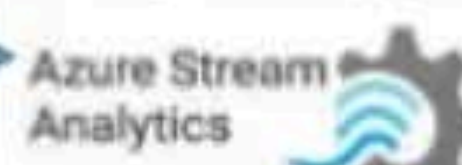
Analytics App 3



Analytics App 4



Analytics App 5

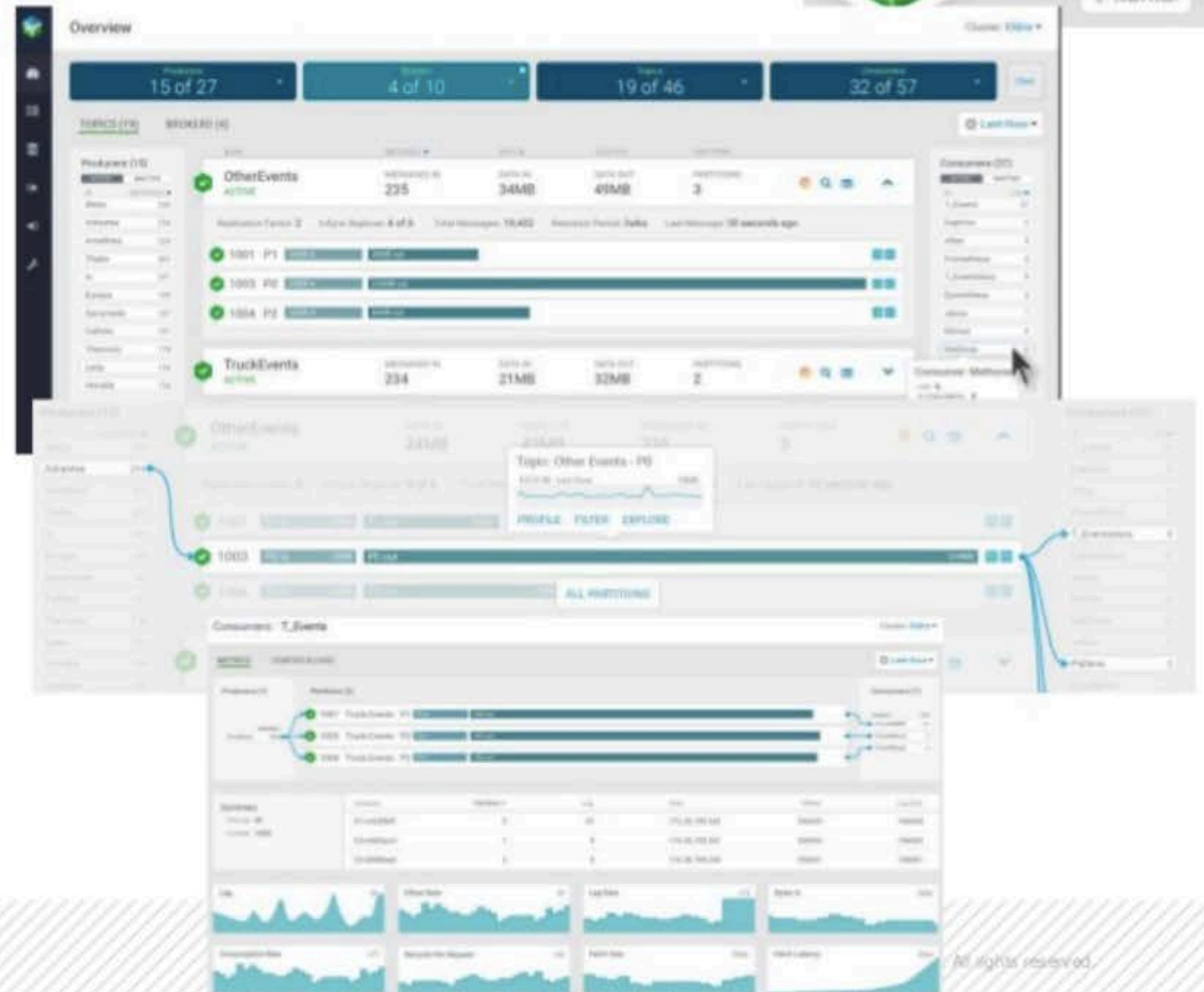


Cloudera Streams Messaging Manager (SMM)



What is SMM?

- Kafka Management and Monitoring tool
- Cure the “Kafka Blindness”
- Single Monitoring Dashboard for all your Kafka Clusters across 4 entities
 - Broker
 - Producer
 - Topic
 - Consumer
- REST as a First Class Citizen
- Alerting
- Schema Management
- Integration with Schema Registry



CLOUDERA DATAFLOW

Data-in-motion platform



EDGE DATA MANAGEMENT

Edge data collection,
routing and monitoring

Apache MiNiFi

Edge Flow Manager



FLOW MANAGEMENT

Enterprise data ingestion,
transformation and enrichment

Apache NiFi

NiFi Registry



STREAM PROCESSING

Real-time streams
processing at IoT scale

Apache Kafka

Streams Messaging Manager



STREAMING ANALYTICS

Predictive analytics
and real-time insights

Apache Storm

Streaming Analytics Manager

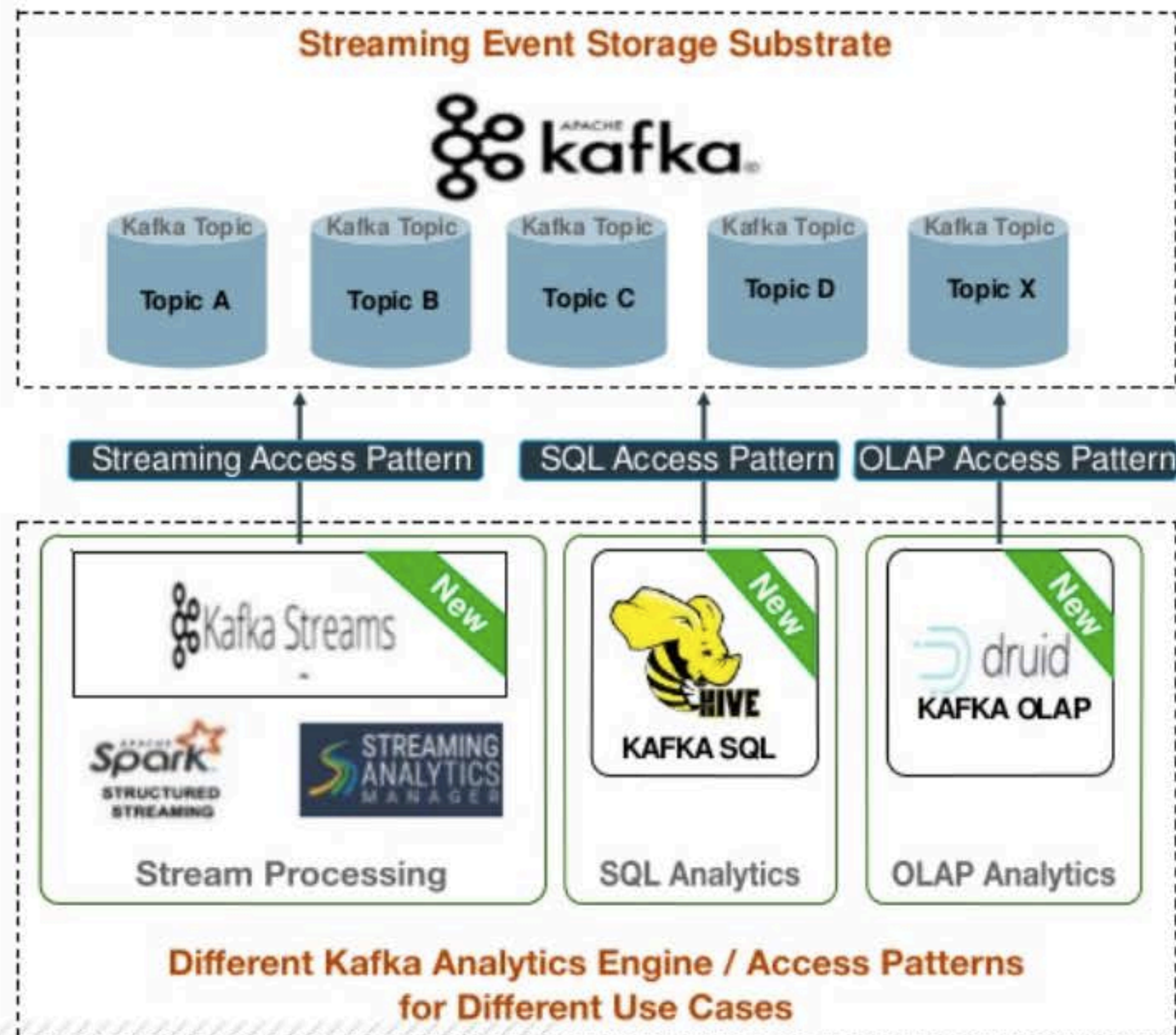
Kafka Streams

STREAMING ANALYTICS

- Pattern matching
- Predictive and Prescriptive Analytics
- Complex Event Processing
- Continuous & Real-time Insights



3 Kafka Analytics Access Patterns



CLOUDERA DATAFLOW

Data-in-motion platform



EDGE DATA MANAGEMENT

Edge data collection,
routing and monitoring

Apache MiNiFi

Edge Flow Manager



FLOW MANAGEMENT

Enterprise data ingestion,
transformation and enrichment

Apache NiFi

NiFi Registry



STREAM PROCESSING

Real-time streams
processing at IoT scale

Apache Kafka

Streams Messaging Manager



STREAMING ANALYTICS

Predictive analytics
and real-time insights

Apache Storm

Streaming Analytics Manager

Kafka Streams



ENTERPRISE SERVICES

Provisioning, Management
and Monitoring

Unified Security

Edge-to-Enterprise Governance

Single Sign-on

Schema Registry

ENTERPRISE SERVICES

- Provisioning
- Management
- Monitoring
- Unified Security
- Single Sign-on
- Audit
- Compliance
- Edge-to-Enterprise Governance

CLOUDERA DATAFLOW

Data-in-motion platform



EDGE DATA MANAGEMENT

Edge data collection,
routing and monitoring

Apache MiNiFi

Edge Flow Manager



FLOW MANAGEMENT

Enterprise data ingestion,
transformation and enrichment

Apache NiFi

NiFi Registry



STREAM PROCESSING

Real-time streams
processing at IoT scale

Apache Kafka

Streams Messaging Manager



STREAMING ANALYTICS

Predictive analytics
and real-time insights

Apache Storm

Streaming Analytics Manager

Kafka Streams



ENTERPRISE SERVICES

Provisioning, Management
and Monitoring

Unified Security

Edge-to-Enterprise Governance

Single Sign-on

Schema Registry



KEY DIFFERENTIATORS

100% open source technology – Only vendor with this strategy; prevents vendor lock-in



280+ pre-built processors – Only product to offer such comprehensive connectivity from edge to enterprise



3 Streaming analytics engines – Only vendor to offer a choice of three streaming analytics engines to customers for all their streaming architecture needs



Built-in data provenance – Only product in the market to offer out-of-the-box data provenance on data-in-motion



Comprehensive streaming platform – Only big data vendor to offer a comprehensive streaming platform from real-time data ingestion, transformation, routing to descriptive, prescriptive and predictive analytics.

