

CLOUDERA

# Building Modern Data Streaming Apps with NiFi, Flink and Kafka

Tim Spann  
Principal Developer Advocate

8-June-2023

June 5-8, 2023 · Data platforms · Data engineering · Data management

## BUDAPEST DATA FORUM

Co-hosted with the Budapest ML Forum



**ENTERPRISE  
DATA CLOUD**

CLOUDERA



**CLOUDERA**

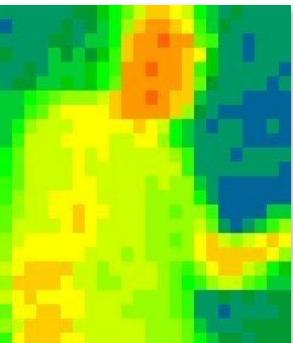


**EDGE  
2AI**

CLOUDERA



# FLiPN-FLaNK Stack



**Tim Spann**

@PaasDev // Blog: [www.datainmotion.dev](http://www.datainmotion.dev)

Principal Developer Advocate.

Princeton Future of Data Meetup.

ex-Pivotal, ex-Hortonworks, ex-StreamNative, ex-PwC

<https://github.com/tspannhw/EverythingApacheNiFi>

<https://medium.com/@tspann>

Apache NiFi x Apache Kafka x Apache Flink x Java



# FLiP Stack Weekly



<https://bit.ly/32dAJft>



This week in Apache NiFi, Apache Flink, Apache Pulsar, Apache Spark, Apache Iceberg, Python, Java and Open Source friends.

# Future of Data - Princeton + Virtual



<https://www.meetup.com/futureofdata-princeton/>

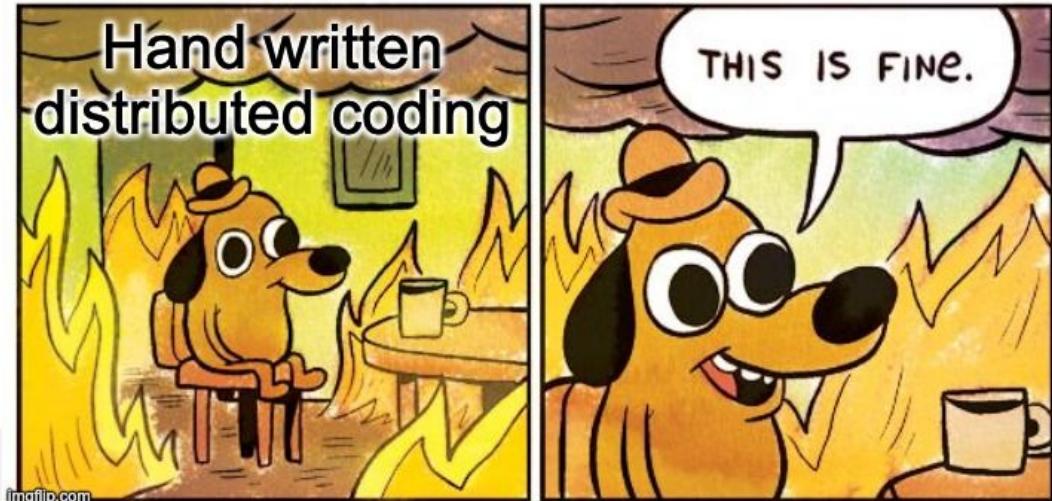
From Big Data to AI to Streaming to Containers to Cloud to Analytics to Cloud Storage to Fast Data to Machine Learning to Microservices to ...



@PaasDev

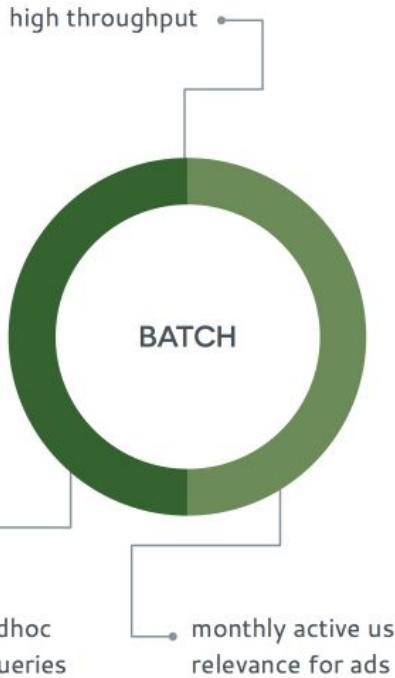
---

# STREAMING

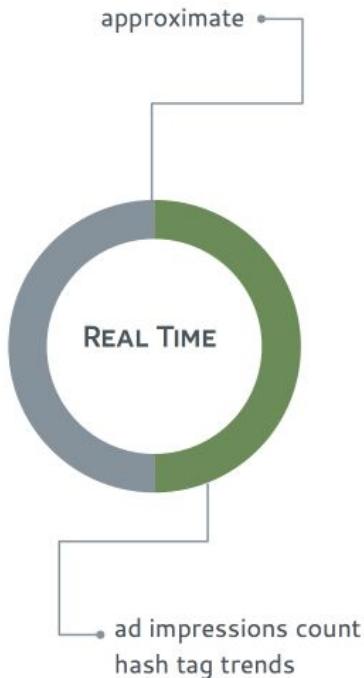


# What is Real-Time?

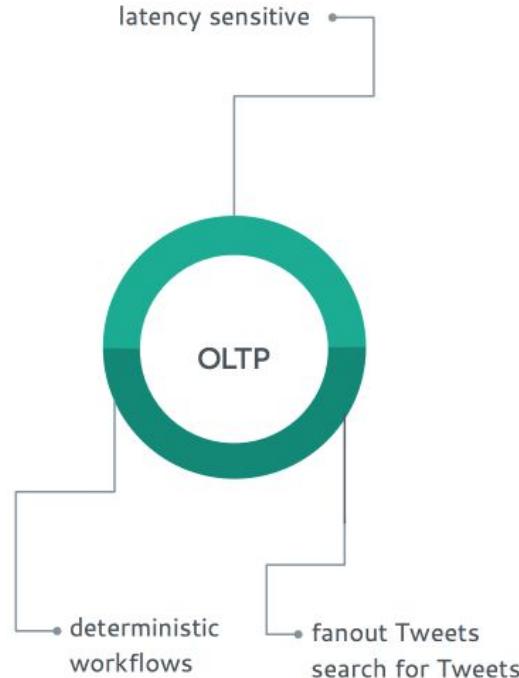
> 1 HOUR



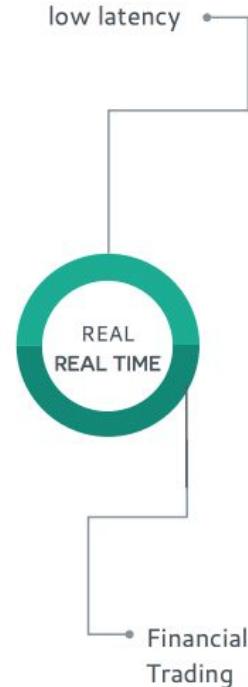
10 MS - 1 SEC



< 500 MS

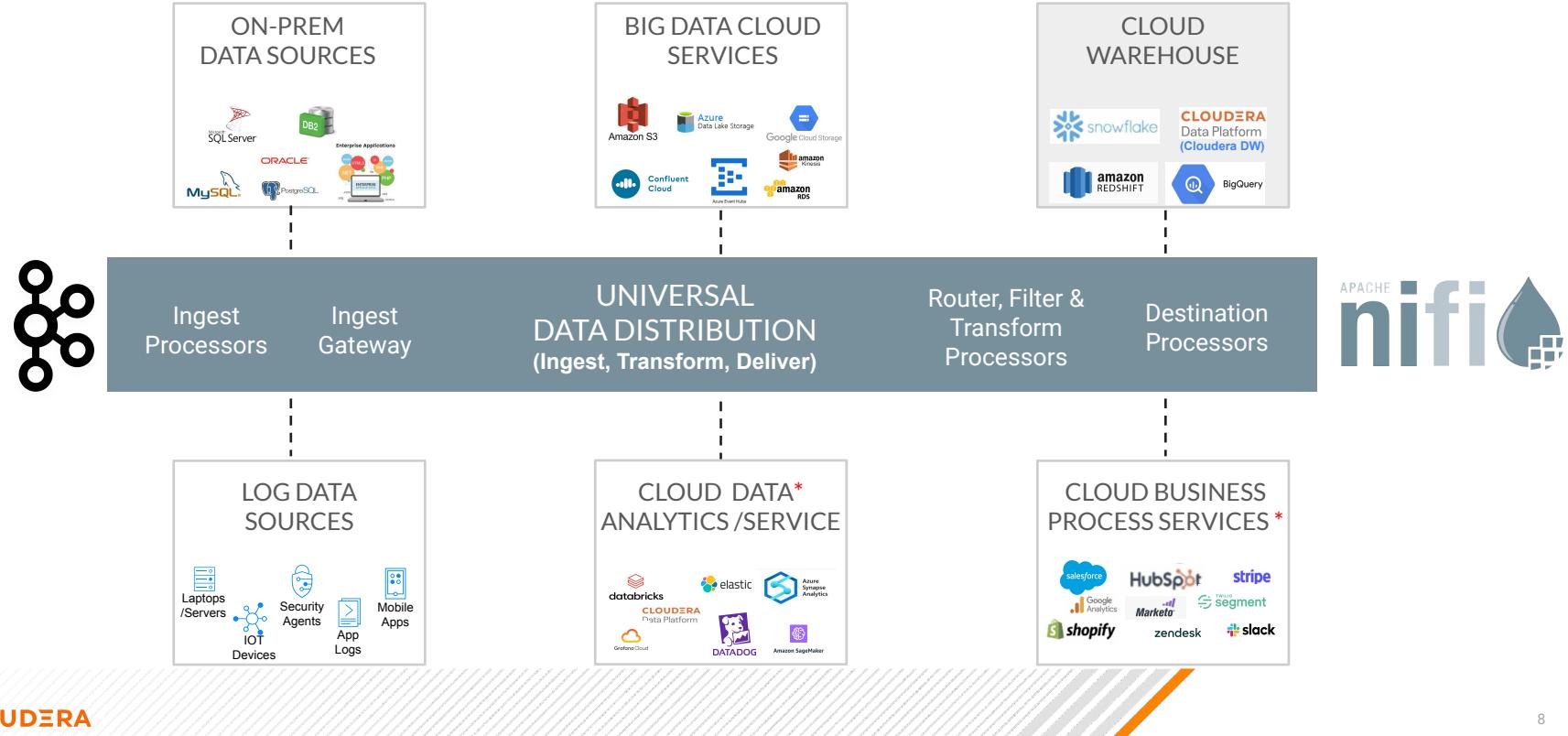


< 1 MS



# Streaming From ... To ...

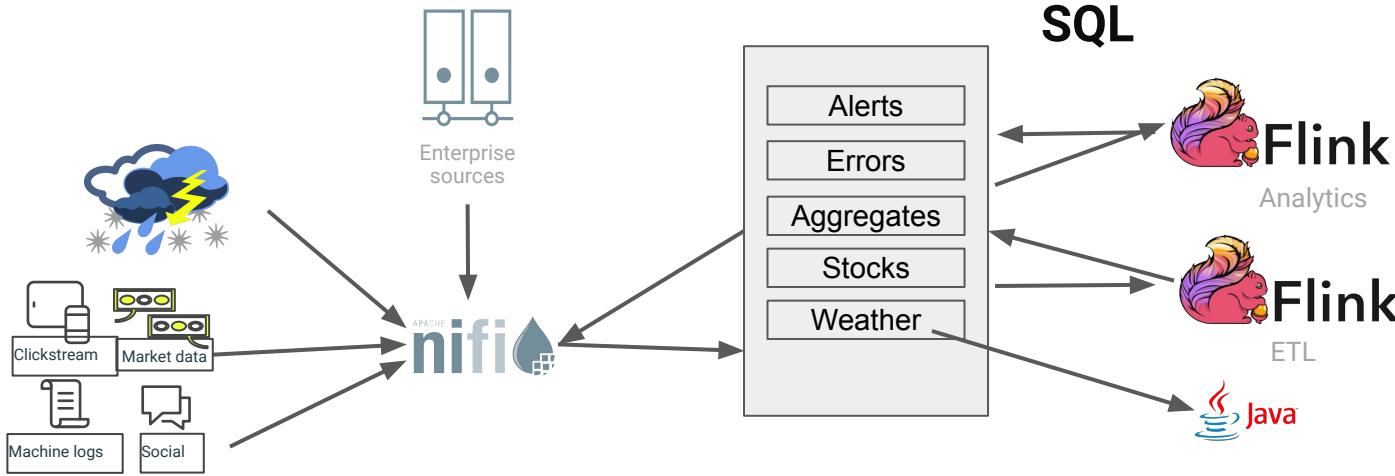
Data distribution as a first class citizen



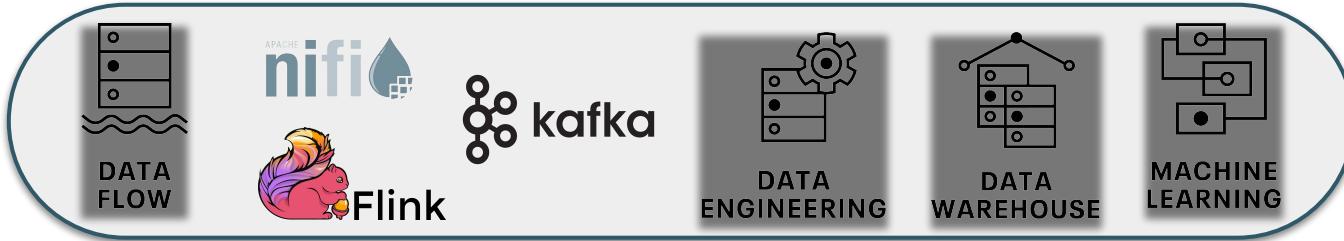
# BUILDING REAL-TIME REQUIRES A TEAM



# End to End Streaming Pipeline Example



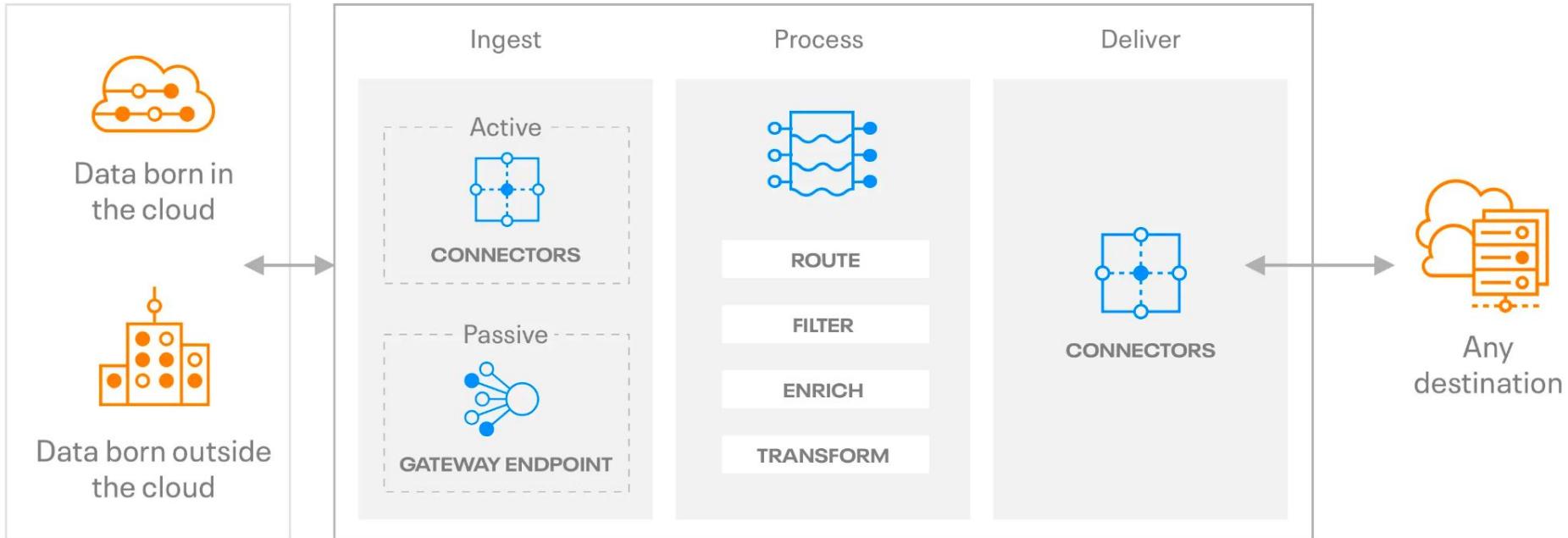
# CDP: AN OPEN DATA LAKEHOUSE



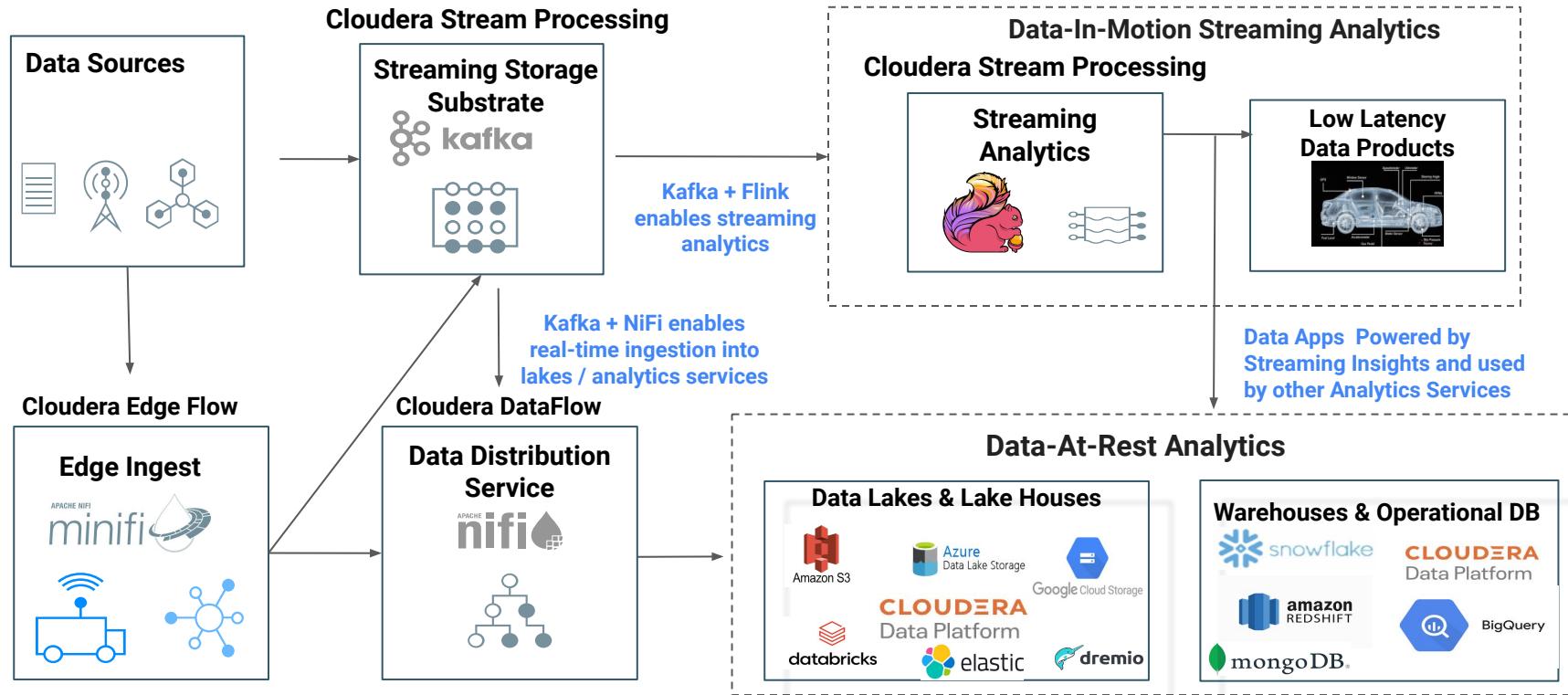
ICEBERG 



# DATAFLOW FOR THE PUBLIC CLOUD

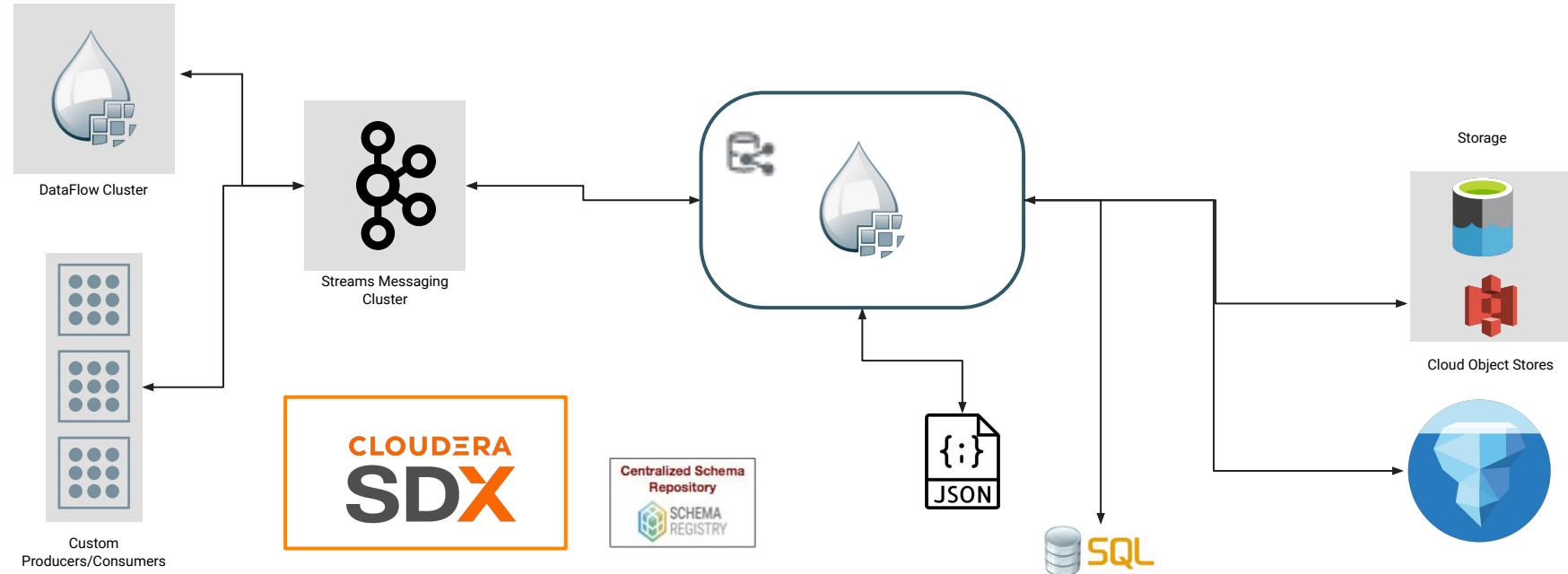


# Analytics-in-Stream

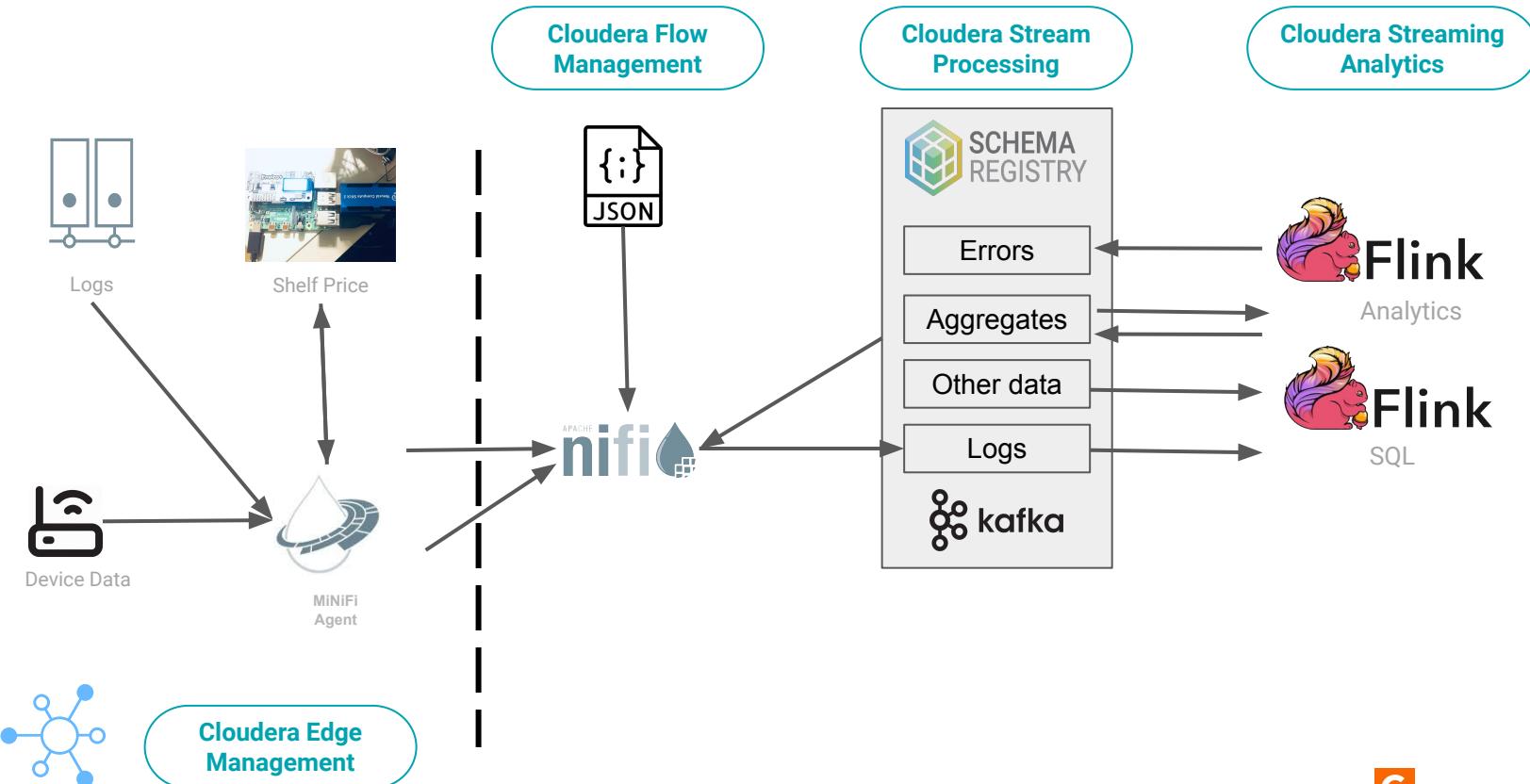


# Retail Use Case: Ingest Retail Goods Prices

## Codeless Data Movement



# Pricing Pipeline



C

---

# APACHE KAFKA

---

# What is Apache Kafka?

**Distributed:** horizontally scalable

**Partitioned:** the data is split-up and distributed across the brokers

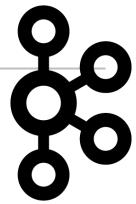
**Replicated:** allows for automatic failover

**Unique:** Kafka does not track the consumption of messages (the consumers do)

**Fast:** designed from the ground up with a focus on performance and throughput

Kafka was built at LinkedIn in 2011

Open sourced as an Apache project



# Yes, Franz, It's Kafka

Let's do a metamorphosis on your data. Don't fear changing data.

**You don't need to be a brilliant writer to stream data.**



Franz Kafka was a German-speaking Bohemian novelist and short-story writer, widely regarded as one of the major figures of 20th-century literature. His work fuses elements of realism and the fantastic.

[Wikipedia](#)



# What is Can You Do With Apache Kafka?

Web site activity: track page views, searches, etc. in real time

Events & log aggregation: particularly in distributed systems where messages come from multiple sources

Monitoring and metrics: aggregate statistics from distributed applications and build a dashboard application

Stream processing: process raw data, clean it up, and forward it on to another topic or messaging system

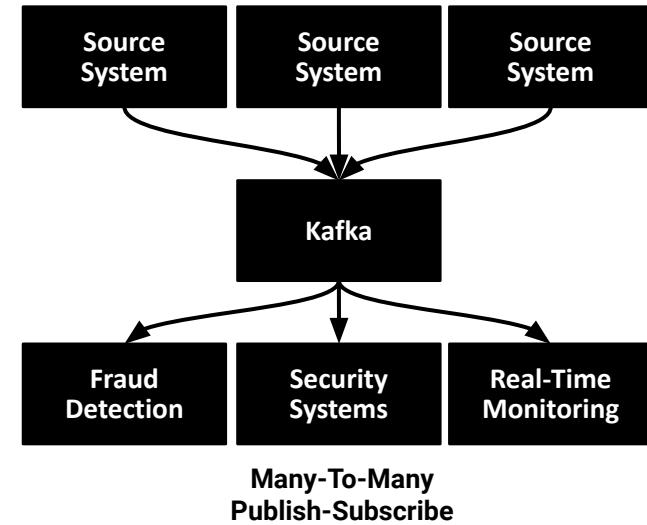
Real-time data ingestion: fast processing of a very large volume of messages

# Kafka Terms

- Kafka is a publish/subscribe messaging system comprised of the following components:
  - **Topic:** a message feed
  - **Producer:** a process that publishes messages to a topic
  - **Consumer:** a process that subscribes to a topic and processes its messages
  - **Broker:** a server in a Kafka cluster

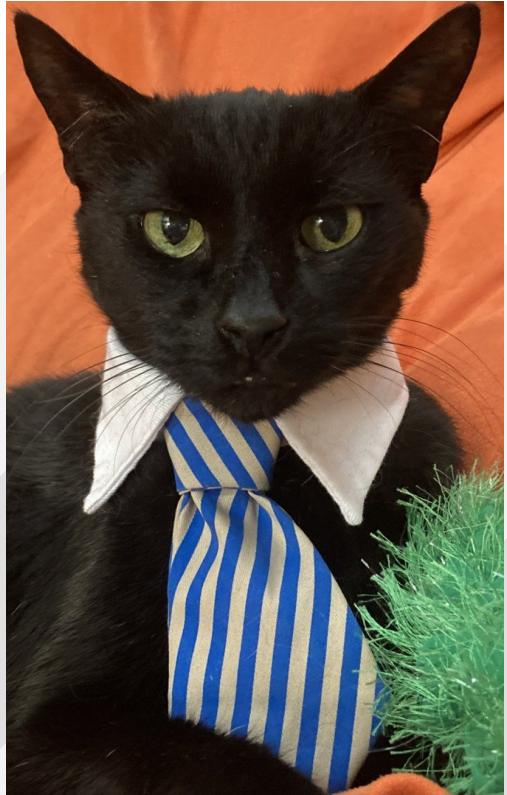
- Highly reliable distributed messaging system
- Decouple applications, enables many-to-many patterns
- Publish-Subscribe semantics
- Horizontal scalability
- Efficient implementation to operate at speed with big data volumes
- Organized by topic to support several use cases

## EVENTS



---

# APACHE FLINK



# Flink SQL



- Streaming Analytics
- Continuous SQL
- Continuous ETL
- Complex Event Processing
- Standard SQL Powered by Apache Calcite

The screenshot shows the Apache Flink Dashboard interface. A job named "xenodochial\_noyce" is running. The job configuration includes a Kafka source, a Sink, and various processing steps like SourceConversionTablet, Map, Filter, and Reduce. Task details show the start time (2021-04-07 10:08:37), duration (3h 56m 21s), and performance metrics such as bytes received, records received, and bytes sent.

<https://www.datainmotion.dev/2021/04/cloudera-sql-stream-builder-ssb-updated.html>

# Flink SQL

**Key Takeaway: Rich SQL grammar with advanced time and aggregation tools**

```
-- specify Kafka partition key on output
SELECT foo AS _eventKey FROM sensors

-- use event time timestamp from kafka
-- exactly once compatible
SELECT eventTimestamp FROM sensors

-- nested structures access
SELECT foo.'bar' FROM table; -- must quote nested
column

-- timestamps
SELECT * FROM payments
WHERE eventTimestamp > CURRENT_TIMESTAMP-interval
'10' second;

-- unnest
SELECT b.* , u.*
FROM bgp_avro b,
UNNEST(b.path) AS u(pathitem)

-- aggregations and windows
SELECT card,
MAX(amount) as theamount,
TUMBLE_END(eventTimestamp, interval '5' minute) as
ts
FROM payments
WHERE lat IS NOT NULL
AND lon IS NOT NULL
GROUP BY card,
TUMBLE(eventTimestamp, interval '5' minute)
HAVING COUNT(*) > 4 -- >4==fraud

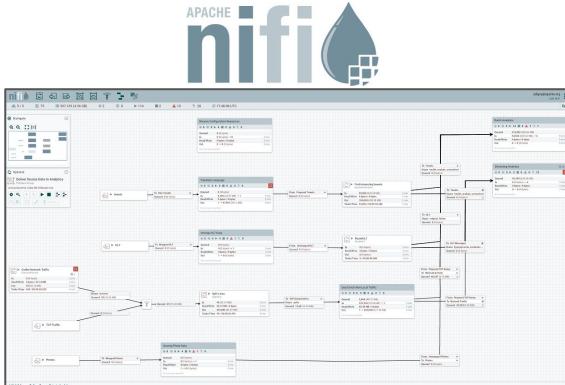
-- try to do this ksql!
SELECT us_west.user_score+ap_south.user_score
FROM kafka_in_zone_us_west us_west
FULL OUTER JOIN kafka_in_zone_ap_south ap_south
ON us_west.user_id = ap_south.user_id;
```

---

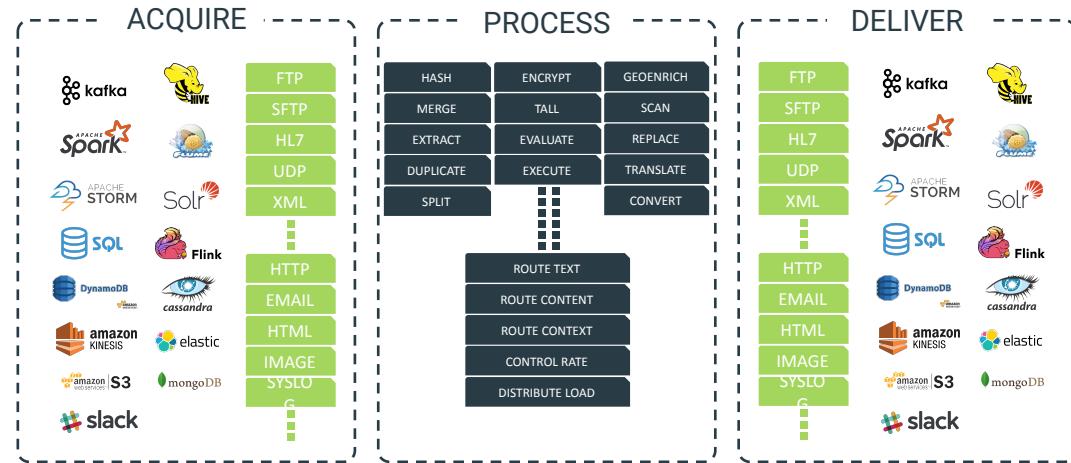
# DATAFLOW APACHE NIFI

# Apache NiFi

Enable easy ingestion, routing, management and delivery of any data anywhere (Edge, cloud, data center) to any downstream system with built in end-to-end security and provenance



Advanced tooling to industrialize flow development  
(Flow Development Life Cycle)



- Over 300 Prebuilt Processors
- Easy to build your own
- Parse, Enrich & Apply Schema
- Filter, Split, Merger & Route
- Throttle & Backpressure

- Guaranteed Delivery
- Full data provenance from acquisition to delivery
- Diverse, Non-Traditional Sources
- Eco-system integration

# Provenance

Displaying 13 of 104  
Oldest event available: 11/15/2016 13:34:50 EST  
Showing the most recent events.

Date/Time	Type	FlowFile Uuid	Size	Component Name	Component Type
11/15/2016 13:35:03.8...	RECEIVE	379fc4f6-60e0-4151-9743-28...	44 bytes	ConsumeKafka	ConsumeKafka
11/15/2016 13:35:02.7...	RECEIVE	78f8c38b-89fc-4d00-a8d8-51...	44 bytes	ConsumeKafka	ConsumeKafka
11/15/2016 13:35:01.6...	RECEIVE	2bcd5124-bb78-489f-ad8a-7...	44 bytes	ConsumeKafka	ConsumeKafka

- Tracks data at each point as it flows through the system
- Records, indexes, and makes events available for display
- Handles fan-in/fan-out, i.e. merging and splitting data
- View attributes and content at given points in time

The diagram illustrates a data flow process. It starts with a red circle labeled "RECEIVE", which has an arrow pointing down to a grey circle labeled "JOIN". From the "JOIN" circle, an arrow points down to a grey circle labeled "DROP". Two green arrows originate from the "RECEIVE" and "JOIN" circles and point to a separate "Provenance Event" panel on the right.

**Provenance Event**

DETAILS	ATTRIBUTES	CONTENT
Attribute Values		
filename	328717796819631	No value previously set
kafka.offset	44815	No value previously set
kafka.partition	6	No value previously set
kafka.topic	nifi-testing	No value previously set
path	/	No value previously set
uuid	328717796819631-44800-10519073-0E	

# Extensibility

- Built from the ground up with extensions in mind
- Service-loader pattern for...
  - Processors
  - Controller Services
  - Reporting Tasks
  - Prioritizers
- Extensions packaged as NiFi Archives (NARs)
  - Deploy NiFi lib directory and restart
  - Same model as standard components

The screenshot shows the IntelliJ IDEA interface with the project 'nifi-mxnetinference-processors' open. The code editor displays `InferenceProcessorTest.java` with annotations for `LinkProcessor`, `UpdateAttribute`, and `PutHDFS`. Below the code editor is a flowchart illustrating the data processing pipeline:

```
graph TD; In[MockFlowFile] --> LinkProcessor[LinkProcessor]; LinkProcessor --> UpdateAttribute[UpdateAttribute]; UpdateAttribute --> PutHDFS[PutHDFS];
```

**LinkProcessor** (InferenceProcessorTest.java):

Operation	In	Read/Write	Out	Tasks/Time
LinkProcessor	0 bytes	5 min	0 bytes	5 min
In	0 bytes	5 min	0 bytes	5 min
Read/Write	0 bytes / 31.45 KB	5 min	0 bytes	5 min
Out	2 (31.45 KB)	5 min	0 bytes	5 min
Tasks/Time	2 / 00:00:04.808			

**UpdateAttribute** (InferenceProcessorTest.java):

Operation	In	Read/Write	Out	Tasks/Time
UpdateAttribute	2 (31.45 KB)	5 min	0 bytes	5 min
In	2 (31.45 KB)	5 min	0 bytes	5 min
Read/Write	0 bytes / 0 bytes	5 min	0 bytes	5 min
Out	2 (31.45 KB)	5 min	0 bytes	5 min
Tasks/Time	2 / 00:00:00.005			

**PutHDFS** (InferenceProcessorTest.java):

Operation	In	Read/Write	Out	Tasks/Time
PutHDFS	2 (31.45 KB)	5 min	0 (0 bytes)	5 min
In	2 (31.45 KB)	5 min	0 (0 bytes)	5 min
Read/Write	31.45 KB / 0 bytes	5 min	0 (0 bytes)	5 min
Out	0 (0 bytes)	5 min	0 (0 bytes)	5 min
Tasks/Time	2 / 00:00:00.603			

# Parquet Reader/ Writers

- Native Record Processors for Apache Parquet Files!
- CSV <-> Parquet
- XML <-> Parquet
- AVRO <-> Parquet
- JSON <-> Parquet
- More...

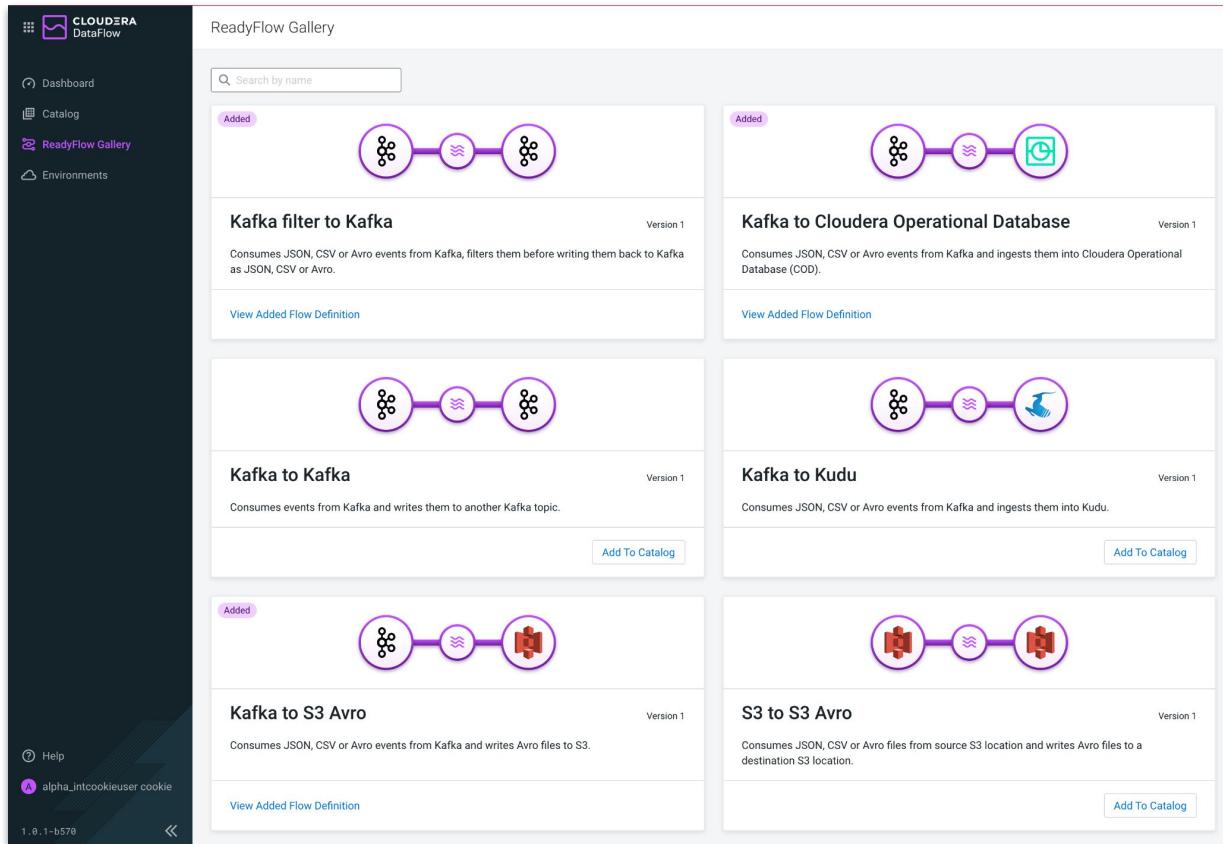
Property	Value
Record Reader	JsonTreeReader
Record Writer	ParquetRecordSetWriter
Merge Strategy	Bin-Packing Algorithm
Correlation Attribute Name	No value set
Attribute Strategy	Keep Only Common Attributes
Minimum Number of Records	10
Maximum Number of Records	
Minimum Bin Size	Requires Controller Service RecordReaderFactory 1.10.0.2.0.0.0-35 from org.apache.nifi - nifi-standard-services-api-nar
Maximum Bin Size	
Max Bin Age	Compatible Controller Services
Maximum Number of Bins	ParquetReader 1.10.0.2.0.0.0-35
Controller Service Name	ParquetReader
Bundle	org.apache.nifi - nifi-parquet-nar
Tags	reader, record, parse, row, parquet
Description	

Add Controller Service

CANCEL CREATE

# ReadyFlow Gallery

- Cloudera provided flow definitions
- Cover most common data flow use cases
- Optimized to work with CDP sources/destinations
- Can be deployed and adjusted as needed



# Flow Catalog

- Central repository for flow definitions
- Import existing NiFi flows
- Manage flow definitions
- Initiate flow deployments

The screenshot shows the Cloudera DataFlow interface with the 'Catalog' tab selected. The main area is titled 'Flow Catalog' and displays a list of available flow definitions. A search bar at the top allows users to search by name. A blue button labeled 'Import Flow Definition' is located in the top right corner. The catalog table has columns for Name, Type, Versions, and Last Updated. Each row contains a link icon. At the bottom of the table, there are pagination controls for items per page (set to 10) and a total count of 10 items.

Name ↑	Type	Versions	Last Updated	
cc_fraud_template_int101run	Custom Flow Definition	2	a day ago	>
cc_fraud_template_int101run2	Custom Flow Definition	1	9 days ago	>
JSON_Kafka_To_Avro_S3	Custom Flow Definition	2	a day ago	>
Kafka filter to Kafka	ReadyFlow	1	2 days ago	>
Kafka to Cloudera Operational Database	ReadyFlow	1	2 days ago	>
Kafka to S3 Avro	ReadyFlow	1	14 hours ago	>
nifi_flows	Custom Flow Definition	1	2 months ago	>
Weather Data Flow	Custom Flow Definition	1	a day ago	>
Weather_Data	Custom Flow Definition	1	15 days ago	>
Weather_JSON_Kafka_To_Avro_S3	Custom Flow Definition	1	21 days ago	>

Items per page: 10 | < < > > | 1 – 10 of 10

# Apache NiFi with Python Custom Processors

## Python as a 1st class citizen

```
import cv2
import numpy as np
import json
from nifiapi.properties import PropertyDescriptor
from nifiapi.properties import ResourceDefinition
from nifiapi.flowfiletransform import FlowFileTransformResult

SCALE_FACTOR = 0.00392
NMS_THRESHOLD = 0.4 # non-maximum suppression threshold
CONFIDENCE_THRESHOLD = 0.5

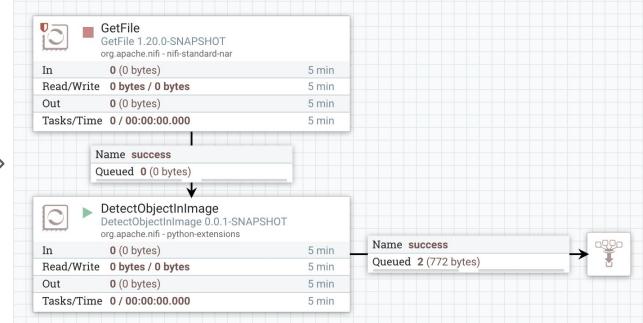
class DetectObjectInImage:
    class Java:
        implements = ['org.apache.nifi.python.processor.FlowFileTransform']
        class ProcessorDetails:
            version = '0.0.1-SNAPSHOT'
            dependencies = ['numpy >= 1.23.5', 'opencv-python >= 4.6']

    def __init__(self, jvm=None, **kwargs):
        self.jvm = jvm

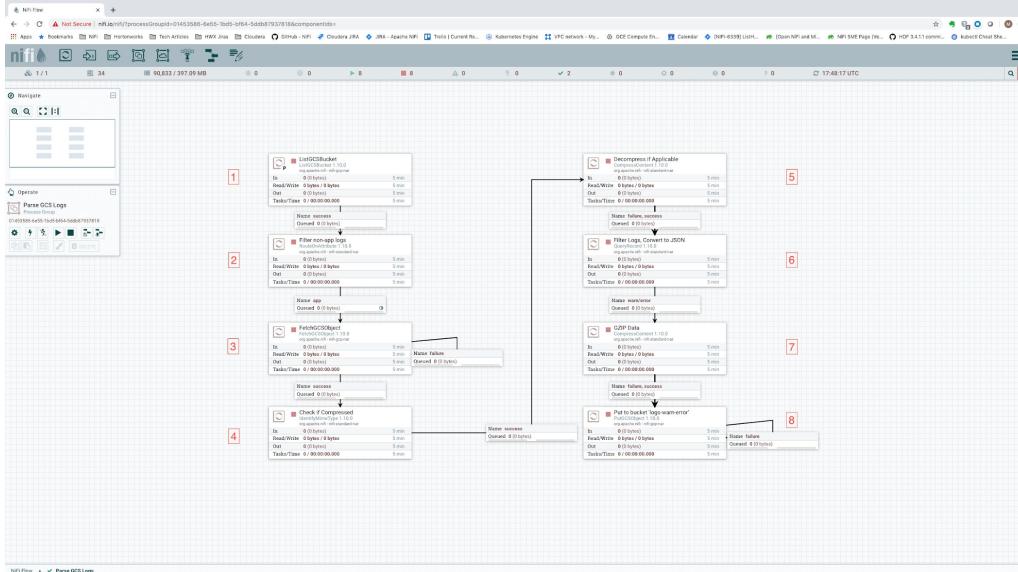
    # Build Property Descriptors
    self.model_file = PropertyDescriptor(
        name = 'Model File',
        description = 'The binary file containing the trained Deep Neural Network weights. Supports Caffe (*.caffemodel), TensorFlow (*.pb), Torch (*.t7, *.net), Darknet (*.weights), ' +
                    'DLDT (*.bin), and ONNX (*.onnx)',
        required = True,
        resource_definition = ResourceDefinition(allow_file = True)
    )
    self.config_file = PropertyDescriptor(
        name = 'Network Config File',
        description = 'The text file containing the Network configuration. Supports Caffe (*.prototxt), TensorFlow (*.pbtxt), Darknet (*.cfg), and DLDT (*.xml)',
        required = False,
        resource_definition = ResourceDefinition(allow_file = True)
    )
    self.class_name_file = PropertyDescriptor(
        name = 'Class Names File',
        description = 'A text file containing the names of the classes that may be detected by the model. Expected format is one class name per line, new-line terminated.',
        required = True,
        resource_definition = ResourceDefinition(allow_file = True)
    )
    self.descriptors = [self.model_file, self.config_file, self.class_name_file]

    def getPropertyDescriptors(self):
        return self.descriptors

    def onScheduled(self, context):
        # read class names from text file
        class_name_file = context.getProperty(self.class_name_file.name).getValue()
        if class_name_file is None:
```



# Processing one million events per second with NiFi



Nodes	Data rate/sec	Events/sec	Data rate/day	Events/day
1	192.5 MB	946,000	16.6 TB	81.7 Billion
5	881 MB	4.97 Million	76 TB	429.4 Billion
25	5.8 GB	26 Million	501 TB	2.25 Trillion
100	22 GB	90 Million	1.9 PB	7.8 Trillion
150	32.6 GB	141.3 Million	2.75 PB	12.2 Trillion

---

# SOURCES AND SINKS



# APACHE ICEBERG

A Flexible, Performant & Scalable Table Format

- Donated by **Netflix** to the Apache Foundation in 2018
- Flexibility
  - Hidden partitioning
  - Full schema evolution
- Data Warehouse Operations
  - Atomic Consistent Isolated Durable (ACID) Transactions
  - Time travel and rollback
- Supports best in class SQL performance
  - High performance at Petabyte scale





AMQP



AWS Lambda



Airtable



Amazon  
API Gateway



Amazon  
CloudWatch



Amazon DynamoDB



amazon  
KINESIS FIREHOSE



Amazon Kinesis  
Data Streams



amazon  
SQS

Amazon SQS



amazon  
SNS

Amazon Simple Notification Services  
(SNS)



Amazon S3



Apache Accumulo



Apache Cassandra



Apache HBase



Apache Hive



Apache Iceberg

Apache Iceberg



Apache Ignite



Apache Kafka



Apache Kudu



Apache Solr

---

**FREE LEARNING ENVIRONMENT**

# CSP Community Edition



- Kafka, KConnect, SMM, SR, Flink, and SSB in Docker
- Runs in Docker
- Try new features quickly
- Develop applications locally

- Docker compose file of CSP to run from command line w/o any dependencies, including Flink, SQL Stream Builder, Kafka, Kafka Connect, Streams Messaging Manager and Schema Registry
  - \$>docker compose up
- Licensed under the Cloudera Community License
- **Unsupported**
- Community Group Hub for CSP
- Find it on [docs.cloudera.com](https://docs.cloudera.com) under Applications



<https://www.cloudera.com/downloads/cdf/csp-community-edition.html>



CSP Community Edition

A readily available, dockerized deployment of Apache Kafka and Apache Flink that allows you to test the features and capabilities of Cloudera Stream Processing.

[Learn More](#)

# Open Source Edition



- Apache NiFi in Docker
- Runs in Docker
- Try new features quickly
- Develop applications locally
- Docker NiFi
  - `docker run --name nifi -p 8443:8443 -d -e SINGLE_USER_CREDENTIALS_USERNAME=admin -e SINGLE_USER_CREDENTIALS_PASSWORD=ctsBtRBKHRAx69EqUghvvgEvjnaLjFEB apache/nifi:latest`
  - Licensed under the ASF License
  - **Unsupported**

<https://hub.docker.com/r/apache/nifi>

---

# DEMO AND CODE



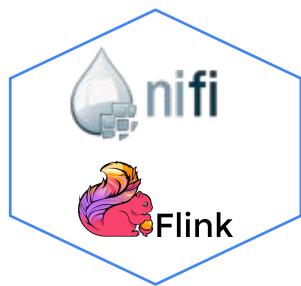
## Collect: Bring Together

Aggregate all data from sensors, drones, logs, geo-location devices, images from cameras, results from running predictions on pre-trained models.



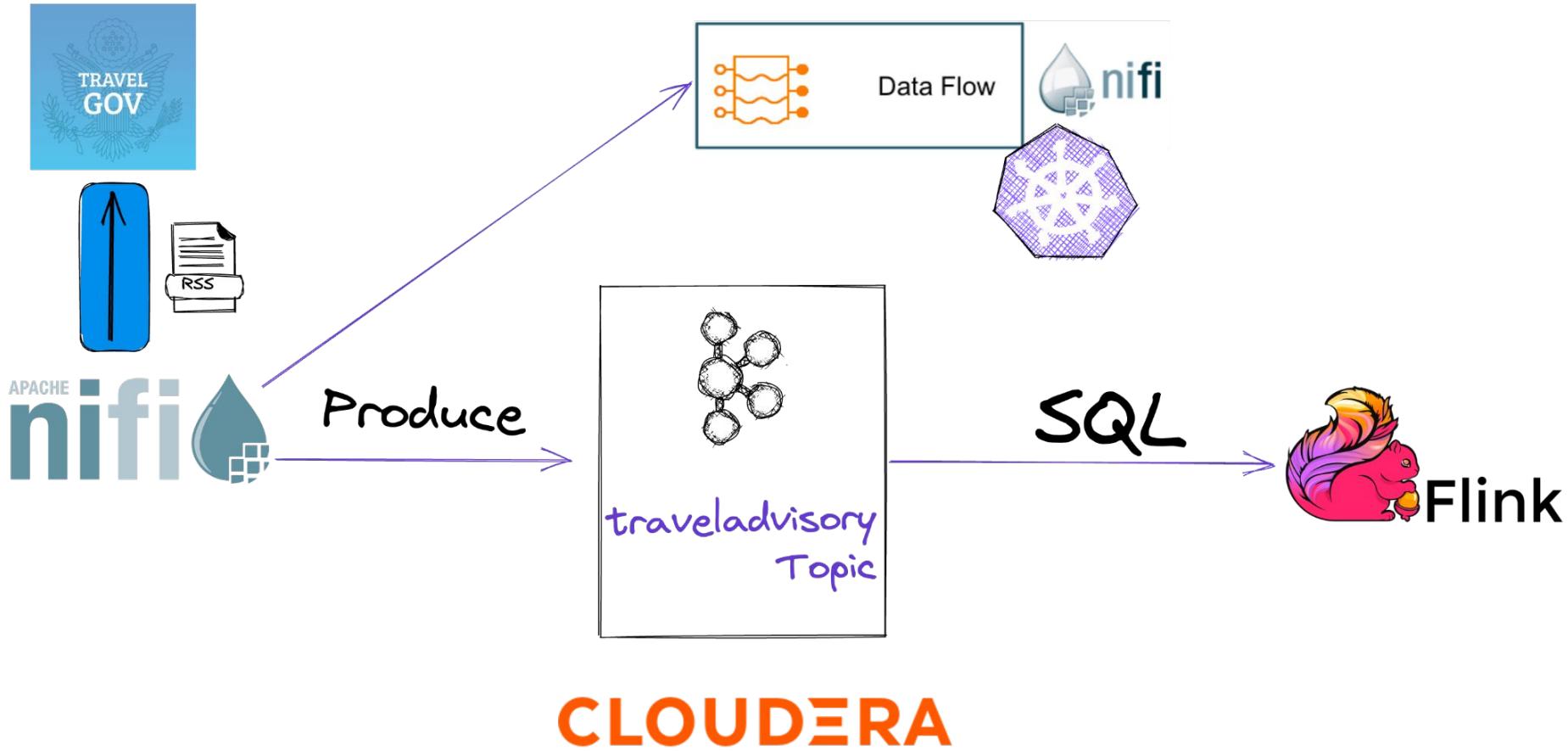
## Conduct: Mediate the Data Flow

Mediate point-to-point and bi-directional data flows, distribute, delivering data reliably to Apache Iceberg, S3, SnowFlake, Slack and Email.



## Curate: Gain Insights

Orchestrate, parse, merge, aggregate, filter, join, transform, fork, query, sort, dissect, store, enrich with weather, location, sentiment analysis, image analysis, object detection, image recognition and more with Apache Tika, Apache OpenNLP and Machine Learning.





METRICS ASSIGNMENT DATA EXPLORER CONFIGS LATENCY

ISOLATION LEVEL:  ▾DESERIALIZER:  ▾  ▾

VALUE SCHEMA NAME:

traveladvisory

VALUE SCHEMA VERSIONS:

1

Show schema text

RECORD LIMIT

15



Partition 0

FROM OFFSET

162

0

59

118

177

Offset	Timestamp	Key	Value
162	Fri, Mar 31 2023, 10:09:34	null	{"title": "Afghanistan - Level 4: Do Not Travel", "pubdate": "Thu, 20 Oct 2022", "link": "http://travel.state.gov/content/travel/en/traveladvisories/1 show more"}
163	Fri, Mar 31 2023, 10:09:34	null	{"title": "Cura\u00e7ao - Level 1: Exercise Normal Precautions", "pubdate": "Tue, 04 Oct 2022", "link": "http://travel.state.gov/content/travel/en/travelad show more"}
164	Fri, Mar 31 2023, 10:09:34	null	{"title": "Cura\u00e7ao - Level 1: Exercise Normal Precautions", "pubdate": "Tue, 04 Oct 2022", "link": "http://travel.state.gov/content/travel/en/travelad show more"}
165	Fri, Mar 31 2023, 10:09:34	null	{"title": "Cura\u00e7ao - Level 1: Exercise Normal Precautions", "pubdate": "Tue, 04 Oct 2022", "link": "http://travel.state.gov/content/travel/en/travelad show more"}
166	Fri, Mar 31 2023, 10:09:34	null	{"title": "Azerbaijan - Level 2: Exercise Increased Caution", "pubdate": "Tue, 15 Nov 2022", "link": "http://travel.state.gov/content/travel/en/travel show more"}
167	Fri, Mar 31 2023, 10:09:34	null	{"title": "Azerbaijan - Level 2: Exercise Increased Caution", "pubdate": "Tue, 15 Nov 2022", "link": "http://travel.state.gov/content/travel/en/travel show more"}
168	Fri, Mar 31 2023, 10:09:34	null	{"title": "Cameroon - Level 2: Exercise Increased Caution", "pubdate": "Wed, 12 Oct 2022", "link": "http://travel.state.gov/content/travel/en/travelad show more"}
169	Fri, Mar 31 2023, 10:09:34	null	{"title": "Colombia - Level 3: Reconsider Travel", "pubdate": "Wed, 04 Jan 2023", "link": "http://travel.state.gov/content/travel/en/traveladvisories/ show more"}
170	Fri, Mar 31 2023, 10:09:34	null	{"title": "Colombia - Level 3: Reconsider Travel", "pubdate": "Wed, 04 Jan 2023", "link": "http://travel.state.gov/content/travel/en/traveladvisories/ show more"}
171	Fri, Mar 31 2023, 10:09:34	null	{"title": "Georgia - Level 1: Exercise Normal Precautions", "pubdate": "Tue, 04 Oct 2022", "link": "http://travel.state.gov/content/travel/en/travelad show more"}





traveladvisories X

searchplanes

RUNNING



Flink Dashboard

Templates

Editor

Materialized View

Job Settings

Job Actions

```
1 select title, domain, category, link, pubdate, ts, uuid, advisoryId
2 FROM
3 `sr1`.`default_database`.traveladvisory
4
```

 Restart  Stop  Stop Polling Polling samples...

<input type="checkbox"/> title	domain	category	link	pubdate	ts	uuid
<input type="checkbox"/> Bhutan - Level 1: Exercise Normal Precautions	BT,advisory	Level 1: Exercise Normal ...	http://travel.state.gov/co...	Wed, 05 Oct 2022	1680277517680	0412509-8e00-4000-95...
<input type="checkbox"/> China - Level 3: Reconsider Travel	CH,advisory,MC,HK	CH	http://travel.state.gov/co...	Fri, 10 Mar 2023	1680277517682	79e7912a-5d40-4afb-96...
<input type="checkbox"/> China - Level 3: Reconsider Travel	CH,advisory,MC,HK	HK	http://travel.state.gov/co...	Fri, 10 Mar 2023	1680277517682	528c584a-e2cc-4119-ac...
<input type="checkbox"/> Tajikistan - Level 2: Exercise Increased Caution	TI,advisory	Level 2: Exercise Increas...	http://travel.state.gov/co...	Wed, 05 Oct 2022	1680277517683	24fef95e-42a9-4011-9f3...
<input type="checkbox"/> Zambia - Level 1: Exercise Normal Precautions	ZA,advisory	advisory	http://travel.state.gov/co...	Tue, 28 Mar 2023	1680277517684	a4e8106e-5f55-4ef9-a5e...
<input type="checkbox"/> Taiwan - Level 1: Exercise Normal Precautions	TW,advisory	advisory	http://travel.state.gov/co...	Mon, 24 Oct 2022	1680277517688	ed3bad9e-96a0-42ca-a6...
<input type="checkbox"/> Chad - Level 3: Reconsider Travel	CD,advisory	Level 3: Reconsider Travel	http://travel.state.gov/co...	Tue, 04 Oct 2022	1680277517690	1ac6673c-dd29-4186-b8...

Logs

Results

Events

1 to 7 of 7

&lt;

&gt;

Page 1 of 1

»

## 🔍 Materialized View

### Configuration

#### Primary Key ⓘ

uuid

 Enable MV ⓘ

#### Retention (Seconds) ⓘ

 Recreate on Job Start ⓘ

#### Min Row Retention Count ⓘ

10000

 Ignore NULLs ⓘ

#### API Key ⓘ

traveladvisory1



### Queries

[⊕ Add New Query](#)

```
/api/v1/query/5201/travel?key=66ba91a9-507f-422c-bbb4-86250a9f7bb1&limit=100
```



## Weather Data For USA

Location: USA

City: Los Angeles

Country: United States

Timezone: Pacific Time (US & Canada)

Latitude: 34.0522

Longitude: -118.2437

Altitude: 334 meters

Wind Speed: 10.0 km/h

Humidity: 45%

Cloudiness: 10%

UV Index: 4

Pressure: 1012 hPa

Cloud Type: Cumulus

Cloud Cover: 10%

Cloud Height: 1000 meters

Cloud Distance: 10 km

## Live Transit Feeds

Location: USA

City: New York City

Country: United States

Timezone: Eastern Time (US & Canada)

Latitude: 40.7128

Longitude: -74.0060

Altitude: 10 meters

Wind Speed: 15.0 km/h

Humidity: 55%

Cloudiness: 20%

UV Index: 5

Pressure: 1010 hPa

Cloud Type: Stratocumulus

Cloud Cover: 20%

Cloud Height: 1500 meters

Cloud Distance: 15 km

## Carried Data Fields

Location: USA

City: San Francisco

Country: United States

Timezone: Pacific Time (US & Canada)

Latitude: 37.7749

Longitude: -122.4194

Altitude: 10 meters

Wind Speed: 12.0 km/h

Humidity: 40%

Cloudiness: 15%

UV Index: 3

Pressure: 1011 hPa

Cloud Type: Cumulus

Cloud Cover: 15%

Cloud Height: 1000 meters

Cloud Distance: 10 km

---

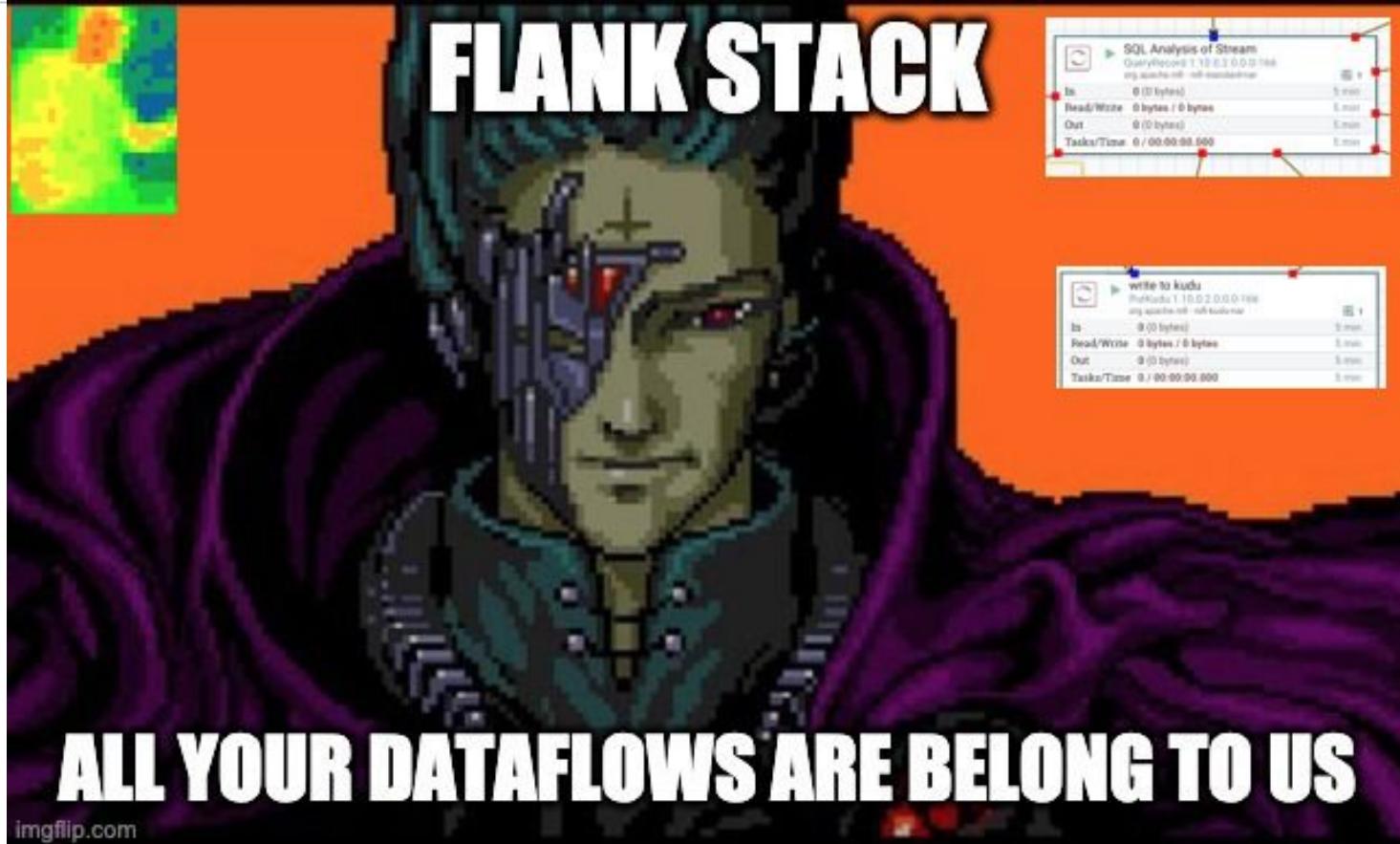
# RESOURCES AND WRAP-UP

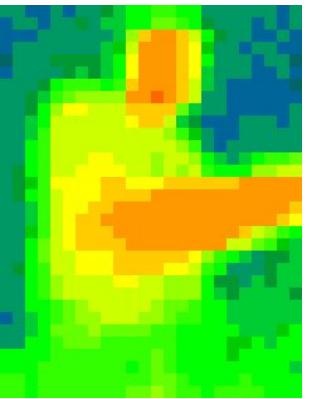
# Streaming Resources

- <https://dzone.com/articles/real-time-stream-processing-with-hazelcast-and-streamnative>
- <https://flipstackweekly.com/>
- <https://www.datainmotion.dev/>
- <https://www.flankstack.dev/>
- <https://github.com/tspannhw>
- <https://medium.com/@tspann>
- <https://medium.com/@tspann/predictions-for-streaming-in-2023-ad4d7395d714>
- [https://www.apachecon.com/acna2022/slides/04\\_Spann\\_Tim\\_Citizen\\_Streaming\\_Engineer.pdf](https://www.apachecon.com/acna2022/slides/04_Spann_Tim_Citizen_Streaming_Engineer.pdf)

## Resources







TH<sub>N</sub>O Y<sub>U</sub>

