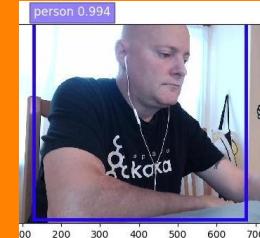




Codeless Generative AI Pipelines

Tim Spann
Principal Developer Advocate

28-March-2024



Tim Spann

Twitter: @PaasDev // Blog: datainmotion.dev

Principal Developer Advocate.

Princeton Future of Data Meetup.

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

ex-PwC, ex-HPE, ex-E&Y.

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

<https://github.com/tspannhw>



 DZone REF CARDS TREND REPORTS EXPERTS

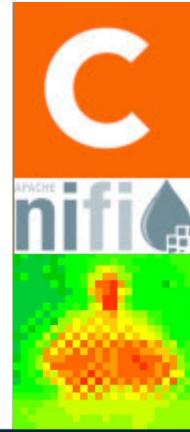
Top IoT Experts

 Tim Spann
Principal Developer Advocate,
Cloudera

<https://github.com/tspannhw/SpeakerProfile/>
Tim Spann is a Principal Developer Advocate in Data in Motion for Cloudera. He works with Apache NiFi, Apache Pulsar, Apache...



Future of Data - NYC + NJ + Philly + Virtual



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

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

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

FLaNK Stack Weekly by Tim Spann



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

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

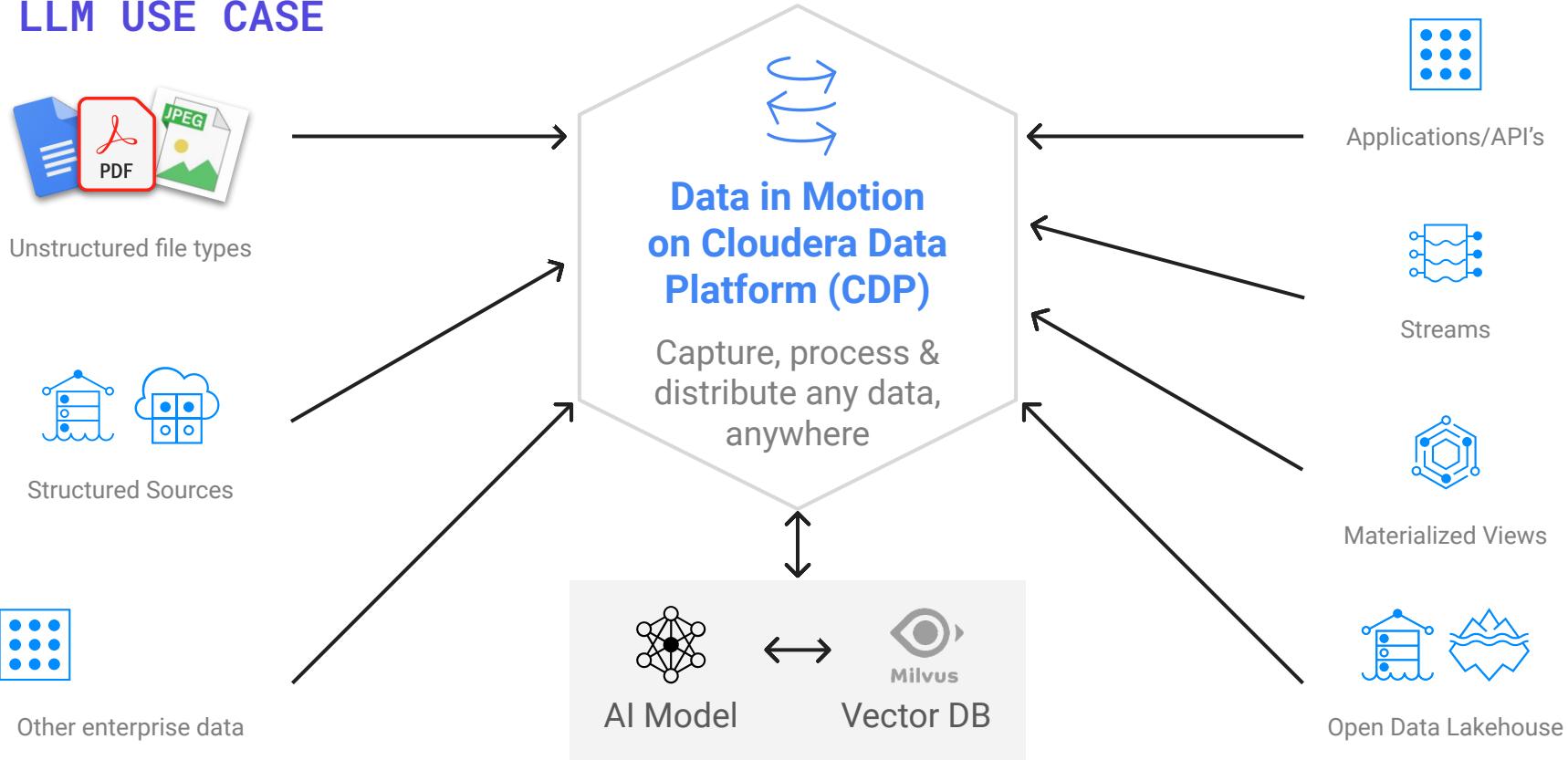
This week in Apache NiFi, Apache Flink, Apache Kafka, ML, AI, Apache Spark, Apache Iceberg, Python, Java, LLM, GenAI, Vector DB and Open Source friends.

Join Our Slack and Interact with LLM

<https://flankworkspace.slack.com/>

https://join.slack.com/t/flankworkspace/shared_invite/zt-2fycjv241-~NRHZDtdfwDjlfvXK_Bz0A

LLM USE CASE



RAPID INNOVATION IN THE LLM SPACE

Too much to cover today.. but you should know the common LLMs, Frameworks, Tools

Notable LLMs

Closed Models	Open Models
 OpenAI GPT3.5 GPT4	 Meta AI Llama2 Code Llama
 Claude2	 MISTRAL AI Mistral7B Mixtral8x7B

++ 100s more... check out the HuggingFace LLM Leaderboard (pretrained, domain fine-tuned, chat models, ...)

Popular LLM Frameworks

LangChain	Llamaindex
 LangChain Langchain is a framework for developing apps powered by LLMs <ul style="list-style-type: none">• Python and JavaScript Libraries• Provides modules for LLM Interface, Retrieval, & Agents	 Llamaindex Llamaindex is a framework designed specifically for RAG apps <ul style="list-style-type: none">• Python and JavaScript Libraries• Provides built in optimizations / techniques for advanced RAG

When to use one over the other? Use Langchain if you need a general-purpose framework with flexibility and extensibility. Consider Llamaindex if you're building a RAG only app (retrieval/search)

Some common Vector DBs

 Chroma	 Pinecone	 Solr
 pgvector	 Milvus	 Weaviate

Open Source vs Self Hosted vs SaaS option

Open Community & Open Models

Hugging Face	HuggingFace
 Hugging Face HuggingFace is an ML community for hosting & collaborating on models, datasets, and ML applications <ul style="list-style-type: none">• Latest open source LLMs are in HuggingFace• + great learning resources / demos	https://huggingface.co/

Cloudera Generative AI Stack

APPLICATIONS

Applied Machine Learning Prototypes (AMPs)

CLOSED-SOURCE FOUNDATION MODELS

APIs: OpenAI (GPT-4 Turbo)
Amazon Bedrock: Anthropic (Claude 2), Cohere...

MODEL HUBS
Hugging Face



FINE-TUNED MODELS

Meta (Llama 2)

OPEN SOURCE FOUNDATION MODELS

MANAGED VECTOR STORE
Pinecone



PRIVATE VECTOR STORE
Milvus, Solr*

CLOUDERA
Open Data Lakehouse



CLOUD INFRASTRUCTURE



SPECIALIZED HARDWARE



NVIDIA



DELL Technologies



DataFlow Pipelines Can Help

External Context Ingest

Ingesting, routing, clean, enrich, transforming, parsing, chunking and vectorizing structured, unstructured, semistructured, binary data and documents

Prompt engineering

Crafting and structuring queries to optimize LLM responses

Context Retrieval

Enhancing LLM with external context such as Retrieval Augmented Generation (RAG)

Roundtrip Interface

Act as a Discord, REST, Kafka, SQL, Slack bot to roundtrip discussions



0 53,639 / 153.08 MB 0 0 ▶ 230 831 546 160 ✓ 0 * 0 0 0 ? 0 22:26:28 EDT



Apache NiFi in a few numbers

A very active project with a dynamic community & comparison with ACEU 2019

2800+ members on the Slack channel (535+ - 5 years ago)

475+ contributors on Github across the repositories (260+ - 5 years ago)

65 committers in the Apache NiFi community (45 - 5 years ago)

Apache NiFi 1.25.0 is the latest release, NiFi 2.0.0-M2 is in beta.

15M+ docker pulls of the Apache NiFi image (1M+ - 5 years ago)

PROVENANCE

Displaying 13 of 104
Oldest event available: 11/15/2016 13:34:50 EST

Showing the most recent events.

ConsumeKafka by component name

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 blue 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

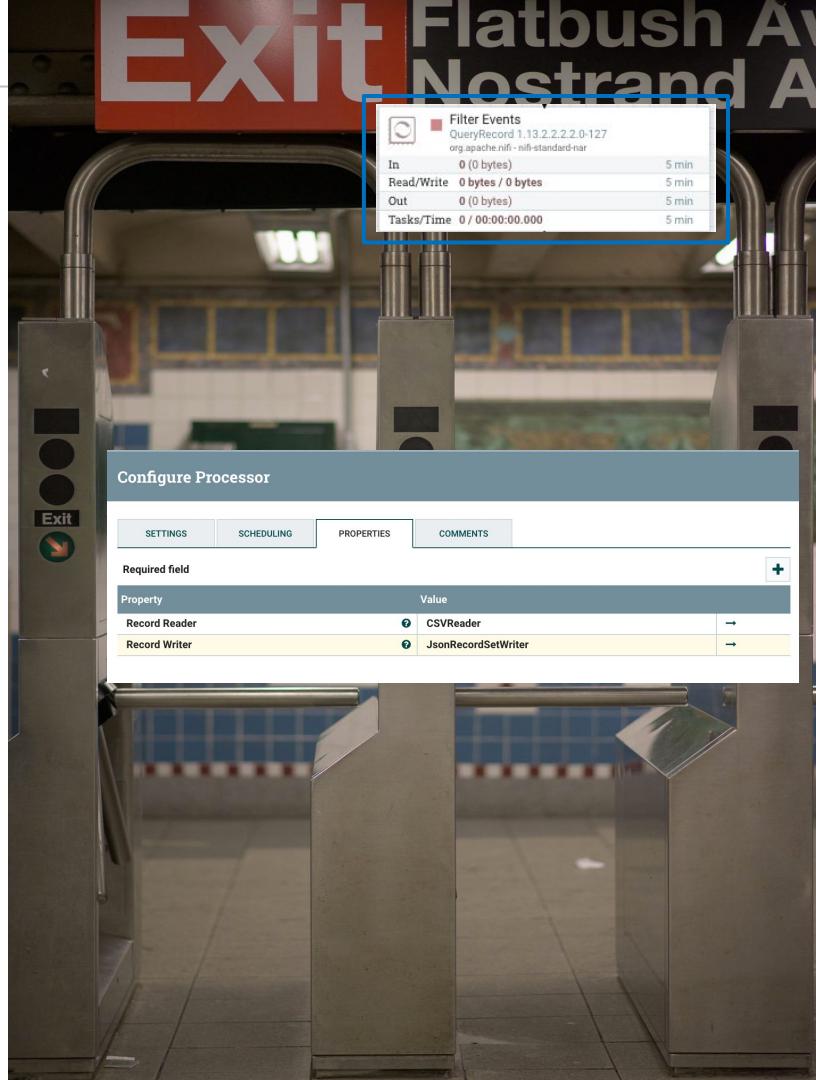
DETAILS ATTRIBUTES CONTENT

Attribute Values

filename	328717796819631
kafka.offset	44815
kafka.partition	6
kafka.topic	nifi-testing
path	/
uuid	32871623852144809510512672385

RECORD-ORIENTED DATA WITH NIFI

- **Record Readers** - Avro, CSV, Grok, IPFIX, JSON1, JSON, Parquet, Scripted, Syslog5424, Syslog, WindowsEvent, XML
- **Record Writers** - Avro, CSV, FreeFromText, Json, Parquet, Scripted, XML
- Record Reader and Writer support referencing a schema registry for retrieving schemas when necessary.
- Enable processors that accept any data format without having to worry about the parsing and serialization logic.
- Allows us to keep FlowFiles larger, each consisting of multiple records, which results in far better performance.



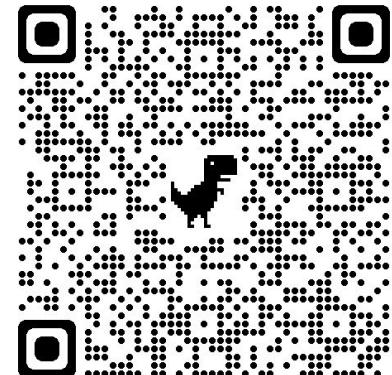
UNSTRUCTURED DATA WITH NIFI

- **Archives** - tar, gzipped, zipped, ...
- **Images** - PNG, JPG, GIF, BMP, ...
- **Documents** - HTML, Markdown, RSS, PDF, Doc, RTF, Plain Text, ...
- **Videos** - MP4, Clips, Mov, Youtube URL...
- **Sound** - MP3, ...
- **Social / Chat** - Slack, Discord, Twitter, REST, Email, ...
- **Identify Mime Types, Chunk Documents, Store to Vector Database**
- **Parse Documents** - HTML, Markdown, PDF, Word, Excel, Powerpoint



CLOUD ML/DL/AI/Vector Database Services

- Cloudera ML
- Amazon Polly, Translate, Textract, Transcribe, Bedrock, ...
- Hugging Face
- IBM Watson X.AI
- Vector Stores Anywhere: Weaviate, Pinecone, Milvus, Chroma DB, SOLR, ...



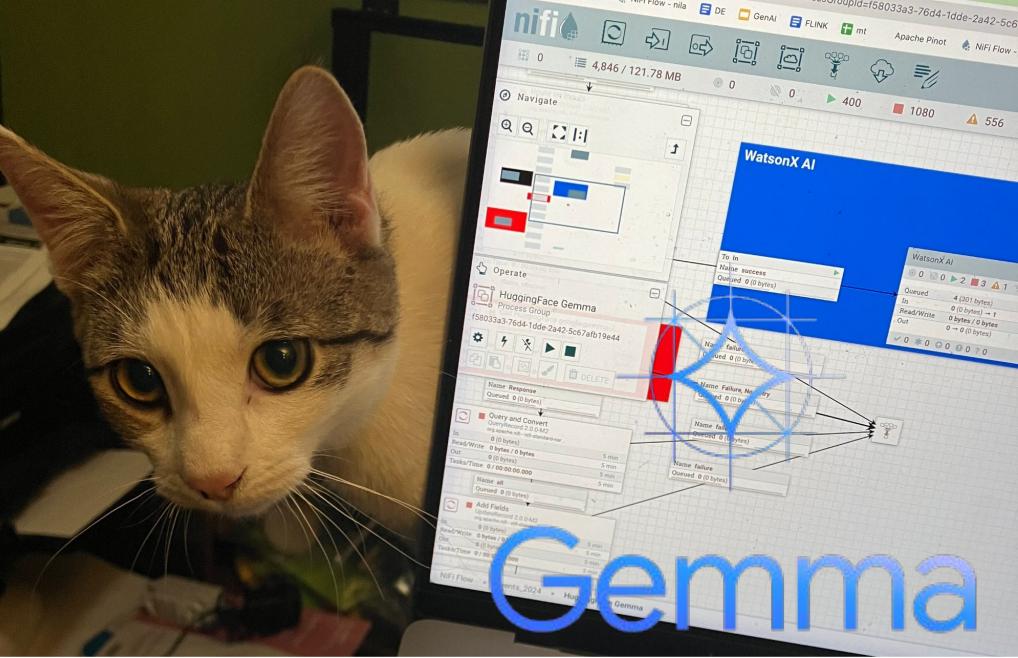
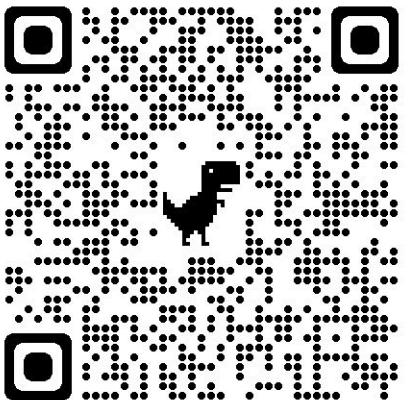
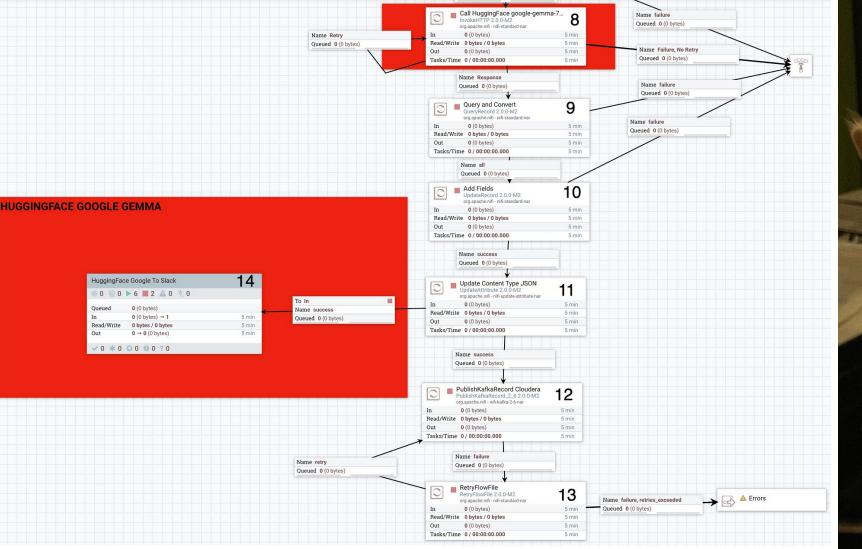


NiFi 2.0.0 Features

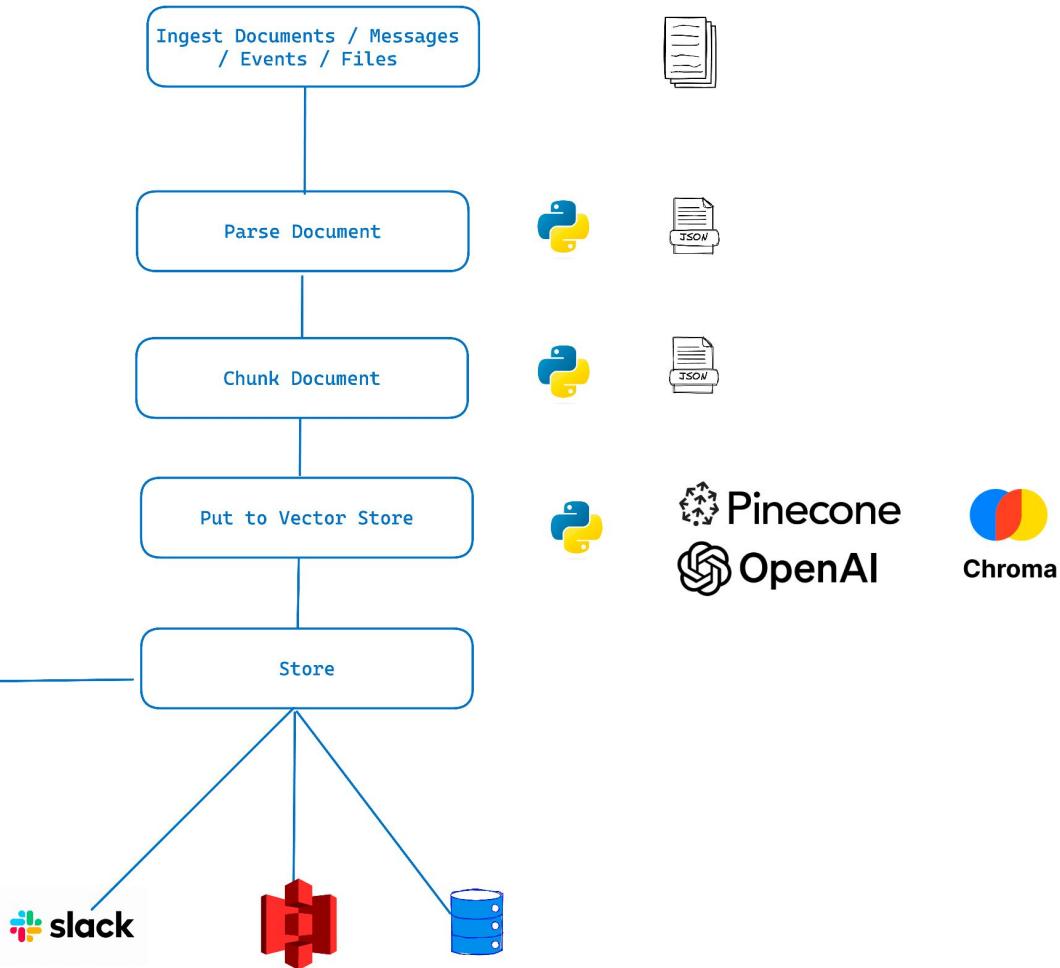
- Python Integration
- Parameters
- JDK 21+
- JSON Flow Serialization
- Rules Engine for Development Assistance
- Run Process Group as Stateless
- flow.json.gz

<https://cwiki.apache.org/confluence/display/NIFI/NiFi+2.0+Release+Goals>

<https://medium.com/cloudera-inc/getting-ready-for-apache-nifi-2-0-5a5e6a67f450>



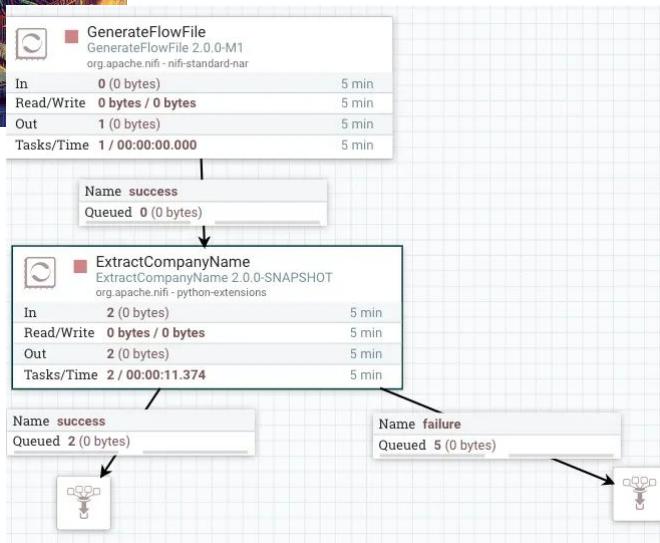
<https://medium.com/cloudera-inc/google-gemma-for-real-time-lightweight-open-llm-inference-88efe98e580f>





Extract Company Names

- Python 3.10+
- Hugging Face, NLP, SpaCY, PyTorch



Attribute Values

companylist

["Amazon", "Microsoft", "Cloudera", "DataSQLR", "Google", "IBM"]

filename

36fb4ae6-701a-4e1d-b890-c93b44f2200b

parsedcompany

Amazon

path

./

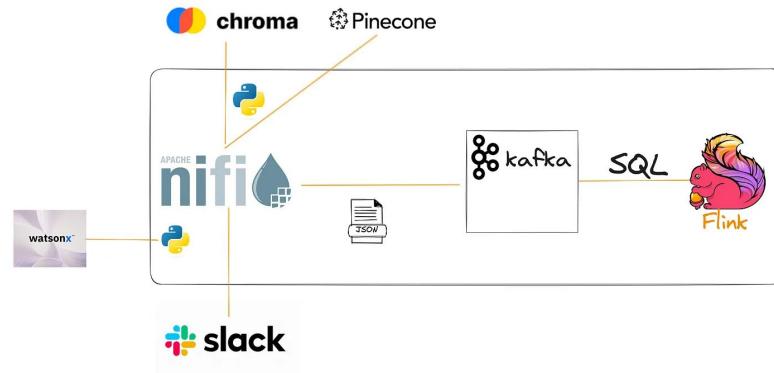
uuid

6366a2c9-3dd4-4e8f-8825-83189d403b92



WatsonX SDK To Foundation

- Python 3.10+
- LLM
- WatsonX.AI Foundation Models
- Inference
- Secure
- Official SDK from IBM

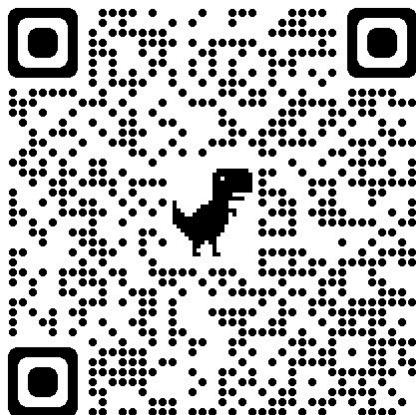


<https://github.com/tspannhw/FLaNK-python-watsonx-processor>



CaptionImage

- Python 3.10+
- Hugging Face
- Salesforce/blip-image-captioning-large
- Generate Captions for Images
- Adds captions to FlowFile Attributes
- Does not require download or copies of your images

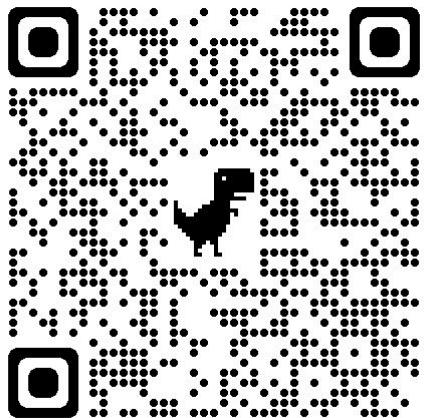


<https://github.com/tspannhw/FLaNK-python-processors>



RESNetImageClassification

- Python 3.10+
- Hugging Face
- Transformers
- Pytorch
- Datasets
- microsoft/resnet-50
- Adds classification label to FlowFile Attributes
- Does not require download or copies of your images

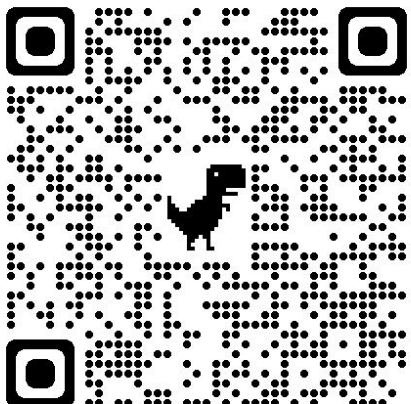


<https://github.com/tspannhw/FLaNK-python-processors>



NSFWImageDetection

- Python 3.10+
- Hugging Face
- Transformers
- Falconsai/nsfw_image_detection
- Adds normal and nsfw to FlowFile Attributes
- Gives score on safety of image
- Does not require download or copies of your images

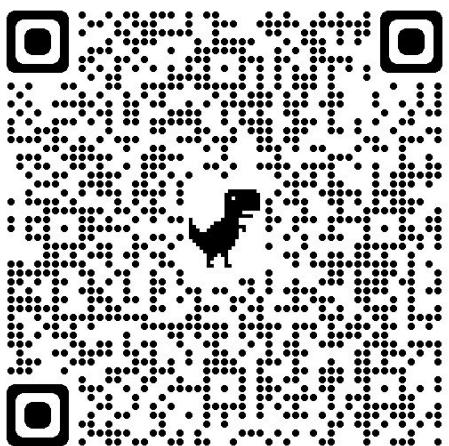


<https://github.com/tspannhw/FLaNK-python-processors>



FacialEmotionsImageDetection

- Python 3.10+
- Hugging Face
- Transformers
- facial_emotions_image_detection
- Image Classification
- Adds labels/scores to FlowFile Attributes
- Does not require download or copies of your images



<https://github.com/tspannhw/FLaNK-python-processors>

FLINK SQL -> CLOUDERA MACHINE LEARNING MODELS

The screenshot displays two main windows from the Cloudera Machine Learning platform.

Left Window (Project Overview):

- Title:** afrank / afrank-SKO-demo
- Status:** Demo by afrank@for SKO
- Project Creation Message:** Project creation succeeded! Step 3 of 3 model deployment step View details
- Models:** Twitter sentiment analysis model (Deployed), genai (Deployed)
- Jobs:** This project has no jobs yet. Create a new job to document your analytics pipelines.
- Files:** Name, templates, ai.py, cds-build.sh, inference.py, README.md, requirements.txt
- Workspace:** go01-aws-workspace
- Cloud Provider:** AWS (AWS)

Right Window (UDF Tester):

- Title:** Projects / Meetups (dev)
- Search Bar:** Search in SSB
- Explorer:** Shows a tree view of Meetups (active), Jobs (47), Virtual Tables (42), Functions (8), and Data Sources.
- Function Details:** f() CALLAI User Defined Function
- Properties:** Name: CALLAI
- Function (JavaScript):** Usage: (STRING)
```function CALLAI(input) { try {```
- UDF Tester (CALLAI):**
  - Parameters:** STRING, Value: "test"
  - Result:** Success, Response: {"Success":true,"Response":{"RecommendationText":"I'm sorry to hear about the issues you've been experiencing with your internet speed, Alex. It sounds frustrating. To improve your internet experience, have you considered upgrading to a faster plan like AirSpeed Advanced? It offers speeds up to 70 Mbps, which should provide a more reliable connection for your work and gaming needs."}, "ReplicaID": "genai-951-1818-7cb7cf6f66-cklgv", "Size": 387, "StatusCode": 200}}
  - Run Button:** Run
  - Close Button:** Close
- Code Editor:** Shows the function code: 1 function CALLAI(input) { 2 try { 3 ("https://modelservice.ml-f... 4 intent("http://modelservice.ml-f... 5 outputStream(con.outputStream); 6 bufferedReader(new java.io.InputStreamReader(... 7 StringBuilder(builder("... 8 builder.readLine()) != null) 9 } 10 } 11 CALLAI(\$p0); // this line must exist

# FLINK SQL -> NIFI -> HUGGING FACE GOOGLE GEMINI

Properties

Name \*

CALLLLM

Description

CALLLLM: call LLM

Output Type \*

STRING

Add Input Type

STRING

Test Save



Properties

Name \*

CALLLLM

Description

CALLLLM: call LLM

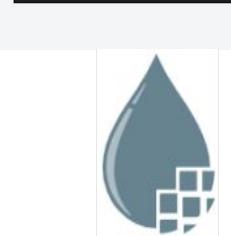
Output Type \*

STRING

Add Input Type

STRING

```
1 function CALLLLM(input) {
2 try {
3 var c = new java.net.URL("http://192.168.1.158:9676/query?calltype=llm&key=" + java.net.URLEncoder.encode(input));
4 c.requestMethod="GET";
5 var reader = new java.io.BufferedReader(new java.io.InputStreamReader(c.getInputStream()));
6 var inputLine = new java.lang.String();
7 var out = new java.lang.StringBuilder();
8 if (reader != null) {
9 while ((inputLine = reader.readLine()) != null)
10 out.append(inputLine);
11 }
12 reader.close();
13 return out.toString();
14 } catch(err) {
15 return "Unknown: " + err;
16 }
17 }
18 CALLLLM($p0);
```



## Hugging Face

UDF Tester (CALLLLM)

Parameters

STRING

What is apache flink?

Result

CALLLLM(What is apache flink?)

Run

Close

What is apache flink? <strong>Flink</strong> is an open-source project built for complex distributed data processing on the Hadoop platform. It is mainly developed by a cluster of 43 contributors in Cloudera. In recent years, we have witnessed a technology boom due to Cloud, Big data, Internet of Things, and machine learning. Notably, it emphasizes a few proprietary and open-source components. As a result, technological advancements helped streamline processes within the industry and increased the importance of analytics within this field. F

# SSB UDF JS/JAVA + GenAI = Real-Time GenAI SQL

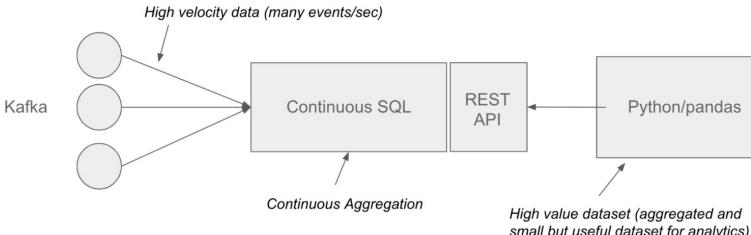


```
SELECT CALLLLM(CAST (messagetext as
STRING)) as generatedtext,
messagerealname, messageusername,
messagetext,messageusertz,
messageid, threadts, ts
FROM flanksslackmessages
WHERE messagetype = 'message'
```

<https://medium.com/cloudera-inc/adding-generative-ai-results-to-sql-streams-513e1fd2a6af>

# SSB MATERIALIZED VIEWS

Key Takeaway; MV's allow data scientist, analyst and developers consume data from the firehose



```
SELECT userid,
 max(amount) as max_amount,
 sum(amount) as sum_amount,
 count(*) as thecount,
 tumble_end(eventTimestamp, interval '5' second) as ts
 FROM authorizations
 GROUP BY userid, tumble(eventTimestamp, interval '5' second)
 HAVING count(*) > 1
```

```
[90]: import pandas as pd
[91]: mv = "https://xxxxxxxxxx"
[92]: df = pd.read_json(mv)
[93]: len(df.keys())
[93]: 5
[95]: df['ts'] = pd.to_datetime(df['ts'])
[97]: df.dtypes
[97]: max_amount int64
sum_amount int64
thecount int64
ts datetime64[ns]
userid int64
dtype: object
[98]: df.set_index('userid').sort_values(by=['thecount'], ascending=False).head()
[98]:
```

userid	max_amount	sum_amount	thecount	ts
787	34911	57304	10	2020-06-16 19:52:15
744	77407	95407	9	2020-06-16 19:52:15
78	88761	330397	9	2020-06-16 19:52:15
541	78762	282682	8	2020-06-16 19:52:15
926	85636	129728	8	2020-06-16 19:52:15

# Apache Flink SQL

Democratize access to real-time data with just SQL



```
CREATE TABLE `ssb`.`Meetups`.`hfbloom` (
 `generated_text` VARCHAR(2147483647),
 `ts` VARCHAR(2147483647),
 `x_compute_type` VARCHAR(2147483647),
 `inputs` VARCHAR(2147483647),
 `x_compute_time` VARCHAR(2147483647),
 `x_inference_time` VARCHAR(2147483647),
 `uuid` VARCHAR(2147483647),
 `x_time_per_token` VARCHAR(2147483647),
 `x_compute_characters` VARCHAR(2147483647),
 `eventTimeStamp` TIMESTAMP(3) WITH LOCAL TIME ZONE METADATA FROM 'timestamp',
 WATERMARK FOR `eventTimeStamp` AS `eventTimeStamp` - INTERVAL '3' SECOND
) WITH (
 'scan.startup.mode' = 'group-offsets',
 'properties.request.timeout.ms' = '120000',
 'properties.auto.offset.reset' = 'earliest',
 'format' = 'json',
 'properties.bootstrap.servers' = 'kafka:9092',
 'connector' = 'kafka',
 'properties.transaction.timeout.ms' = '900000',
 'topic' = 'hfbloom',
 'properties.group.id' = 'llmBloomProps'
)
```

# Infer Tables from Kafka Topics with JSON or Avro

## Kafka Table

Table Name \*

Kafka Cluster \*

Data Format \*

Topic Name \*

Schema Definition

Event Time

Data Transformation

Properties

Deserialization

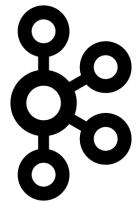
```
1 {
2 "type": "record",
3 "name": "inferredSchema",
4 "fields": [
5 {
6 "name": "bssid",
7 "type": "string",
8 "doc": "Type inferred from '\"\\\"'"
9 },
10 {
11 "name": "channel",
12 "type": "string",
13 "doc": "Type inferred from '\"52\\\"'"
14 },
15 {
16 "name": "channel_band",
17 "type": "string",
18 "doc": "Type inferred from '\"S\\\"'"
19 },
20 {
21 "name": "channel_width",
22 "type": "string",
23 "doc": "Type inferred from '\"80\\\"'"
24 },
25 {
26 "name": "country_code",
27 "type": "string",
28 "doc": "Type inferred from '\"\\\"'"
29 },
30 {
31 "name": "interface",
32 "type": "string",
33 "doc": "Type inferred from '\"en0\\\"'"
34 },
35 }
```

 Schema is valid

Detect Schema

Cancel

Create and Review



# 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.



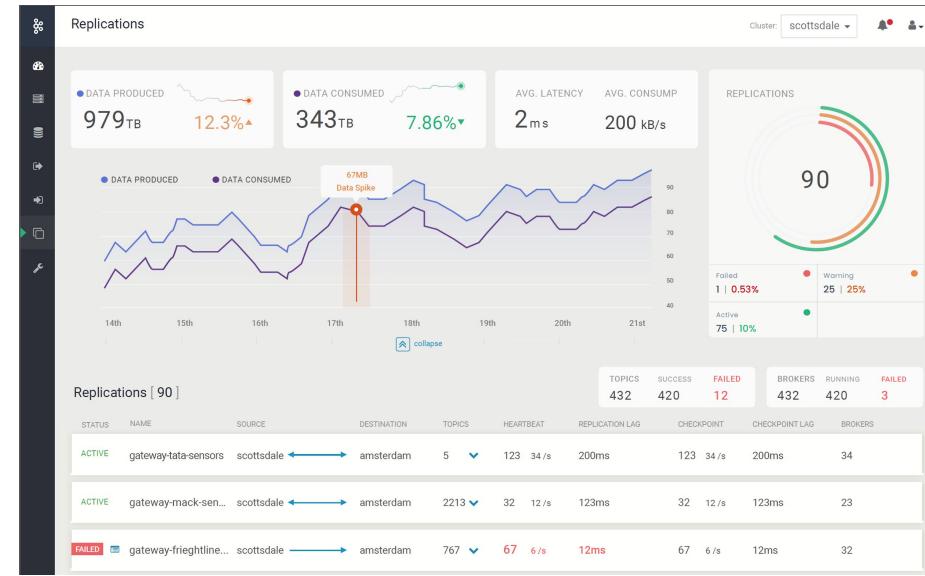
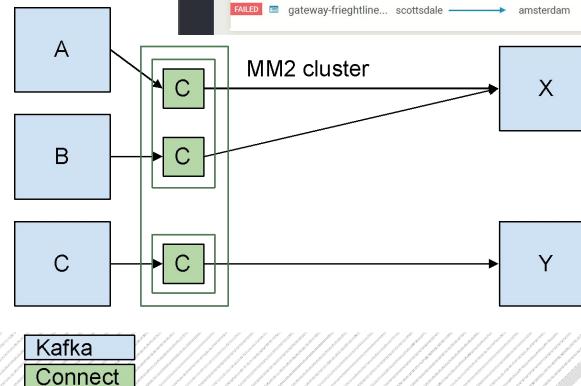
Franz Kafka was a German-speaking Bohemian novelist and short-story writer, widely regarded as one of the major figures of 20th-century literature. His work fuses elements of realism and the **fantastic**.

[Wikipedia](#)

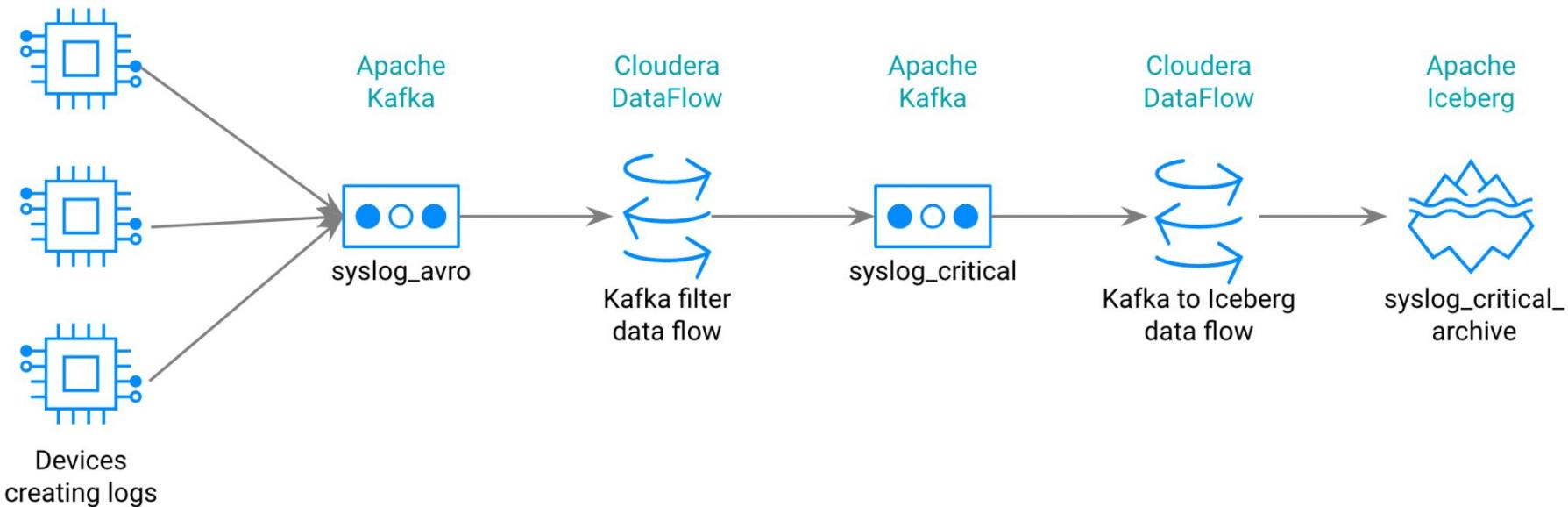


# Streams Replication Manager (SRM)

- Event Replication engine for Kafka
- Supports active-active, multi-cluster, cross DC replication scenarios
- Leverage Kafka Connect for scalability and HA
- Replicate data and configurations (ACL, partitioning, new topics, etc)
- Offset translation for simplified failover
- Integrate replication monitoring with SMM



# ICEBERG





## ReadyFlow Gallery



Iceberg X

Added



### Kafka to Iceberg

Version 1

Consumes JSON, CSV or Avro events from Kafka and writes them as Parquet files to a destination Iceberg table.

[View Added Flow Definition](#)

[Create New Draft](#)

# Cloudera's Open Data Lakehouse



Metadata | Security | Encryption | Control | Governance



**Iceberg Tables**



**Multi-Hybrid Cloud**

- ❑ Multi-function analytics for **Streaming**, **Data Engineering**, **Data Warehouse** and **AI/ML** with integrated data services
- ❑ Common security and **governance** policies and data lineage with SDX integration
- ❑ Common dataset with all **CDP** analytics engines without data duplication and movement
- ❑ Deployment freedom with **Multi-Hybrid Cloud**

# Compute Engine Interoperability & SDX Integration



CLOUDERA  
**SDX**



Iceberg Tables



Ranger



Atlas

- **Snapshot isolation** ensures **consistent** data access and processing with various compute engines including **Hive**, **Spark**, **Impala** and **Nifi**
- **Security & Governance** support (e.g. FGAC) through **Ranger** integration
- Data **lineage** support through **Atlas** integration

# FLINK & ICEBERG INTEGRATION

Robust Next Generation Architecture for Data Driven Business



Unified Processing Engine

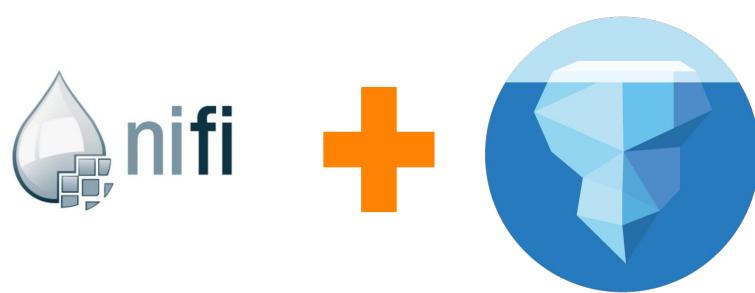
- Maximally open
- Maximally flexible
- Ultra high performance for MASSIVE data
- Can be used as Source and Sink
- Supports batch and streaming modes
- Supports time travel

Massive Open table format

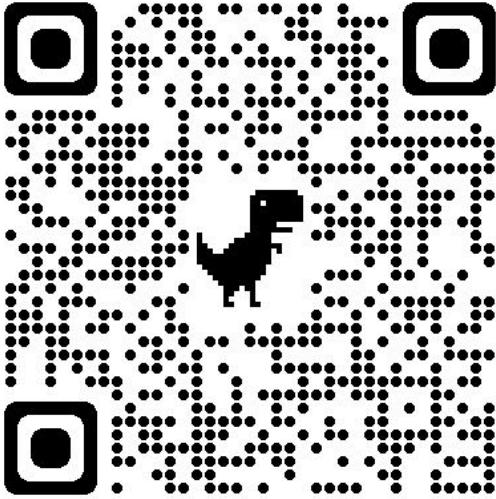
## Iceberg Support for Flink APIs through SSB

Feature support	Flink	Notes
SQL create catalog	✓	
SQL create database	✓	
SQL create table	✓	
SQL create table like	✓	
SQL alter table	✓	Only support altering table properties, column and partition changes are not supported
SQL drop_table	✓	
SQL select	✓	Support both streaming and batch mode
SQL insert into	✓	Support both streaming and batch mode
SQL insert overwrite	✓	
DataStream read	✓	
DataStream append	✓	
DataStream overwrite	✓	
Metadata tables		Support Java API but does not support Flink SQL
Rewrite files action	✓	

# NIFI & ICEBERG INTEGRATION



- **PutIceberg** processor in CFM 2.1.6
- **PutIcebergCDC**



#### 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](#)

# CSP Community Edition

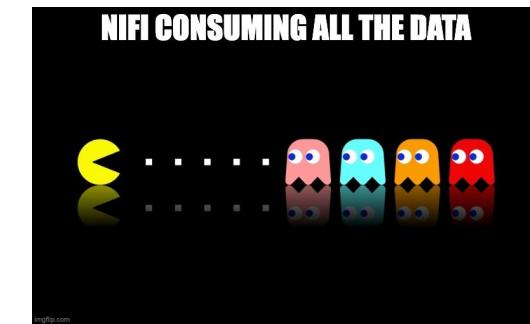


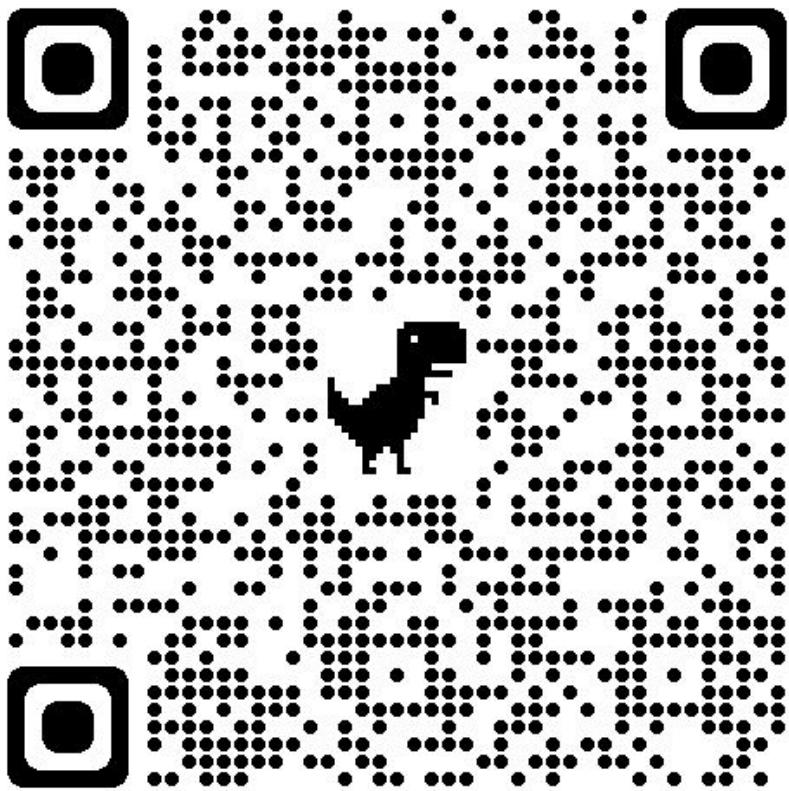
- Docker compose file of CSP to run from command line w/o any dependencies, including Flink, SQL Stream Builder, Kafka, Kafka Connect, Streams Messaging Manager and Schema Registry.
  - \$>docker compose up
  - Licensed under the Cloudera Community License
  - **Unsupported Commercially (Community Help - Ask Tim)**
  - Community Group Hub for CSP
  - Find it on [docs.cloudera.com](http://docs.cloudera.com) (see QR Code)
  - Kafka, Kafka Connect, SMM, SR, Flink, Flink SQL, MV, Postgresql, SSB
  - Develop apps locally

# Open Source Edition



- Apache NiFi 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
- NiFi 1.25 and NiFi 2.0.0-M2





Street Cameras



<https://medium.com/cloudera-inc/streaming-street-cams-to-yolo-v8-with-python-and-nifi-to-minio-s3-3277e73723ce>

## **CEM, CDF, CSP**

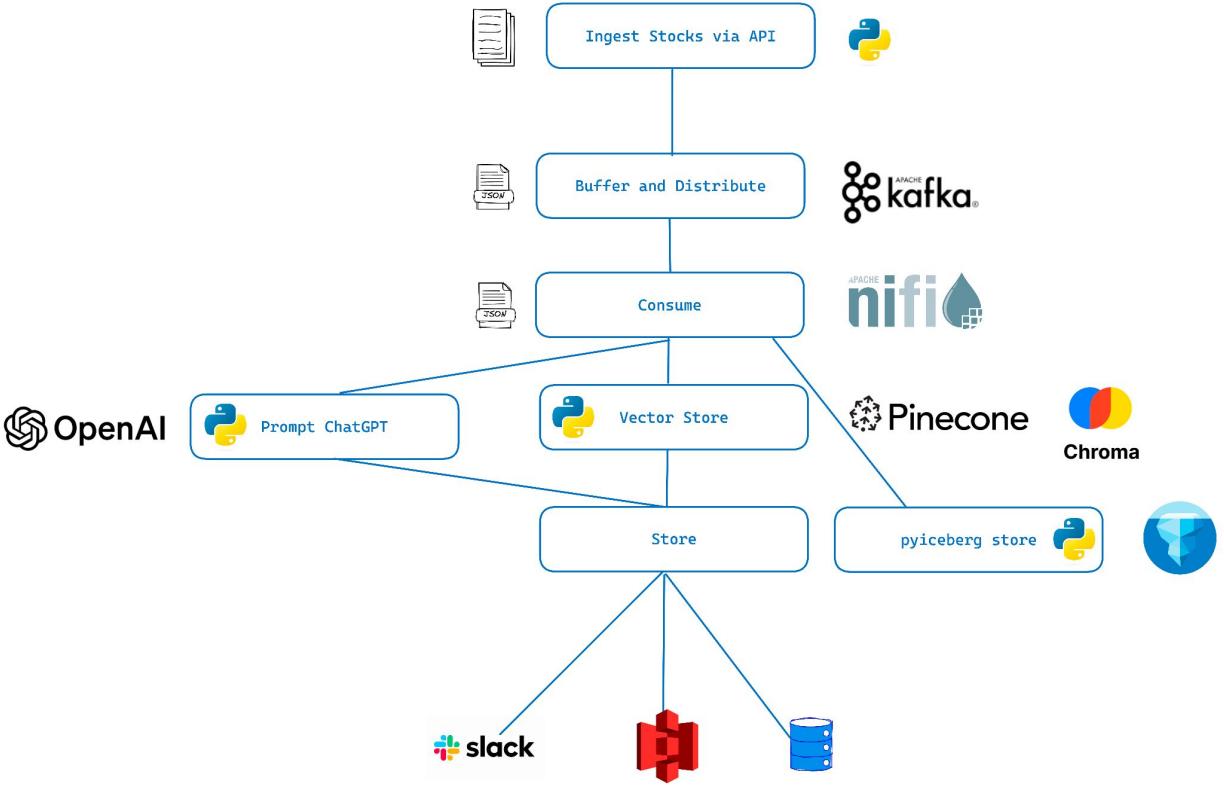
- <https://medium.com/cloudera-inc/watching-airport-traffic-in-real-time-32c522a6e386>
- <https://medium.com/cloudera-inc/building-a-real-time-data-pipeline-a-comprehensive-tutorial-on-mi-nifi-kafka-and-flink-ee03ee6722cb>
- <https://medium.com/cloudera-inc/finding-the-best-way-around-7491c76ca4cb>
- <https://medium.com/cloudera-inc/nyc-traffic-are-you-kidding-me-6d3fa853903b>

## **CDF**

- <https://medium.com/@tspann/building-a-travel-advisory-app-with-apache-nifi-in-k8-969b44c84958>

## **LLM, GenAI, HuggingFace, WatsonX, OLLAMA, Mistral, NiFi, Python, Slack, Pytorch**

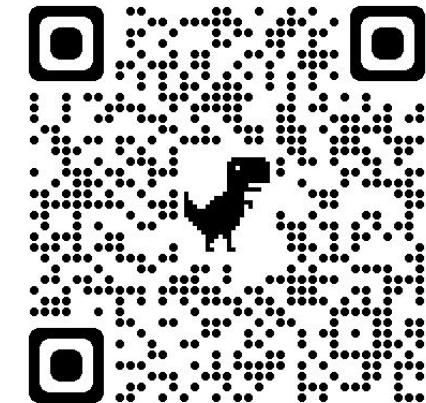
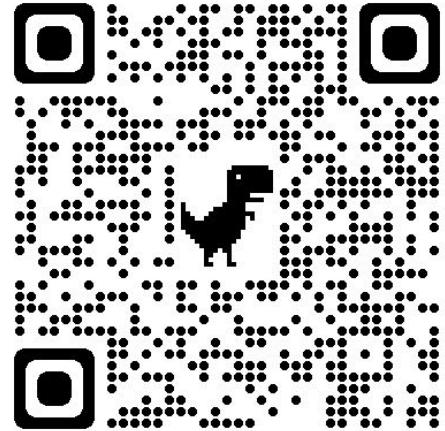
- <https://medium.com/@tspann/using-ollama-with-mistral-and-apache-nifi-720c17f5ff12>
- <https://medium.com/cloudera-inc/google-gemma-for-real-time-lightweight-open-llm-inference-88efe98e580f>
- <https://medium.com/@tspann/image-processing-with-custom-python-and-nifi-2-0-06eadc62c03c>
- <https://medium.com/@tspann/ai-augmented-devrel-part-1-4058af905a89>
- <https://medium.com/cloudera-inc/mixtral-generative-sparse-mixture-of-experts-in-dataflows-59744f7d28a9>
- <https://medium.com/@tspann/building-an-llm-bot-for-meetups-and-conference-interactivity-c211ea6e3b61>
- <https://medium.com/@tspann/yet-another-python-processor-45aaaae6fe406>



<https://github.com/tspannhw/PaK-Stocks>

<https://github.com/tspannhw/FLaNK-Py-Stocks>

<https://medium.com/cloudera-inc/let-nifi-worry-about-those-stocks-for-you-57d5f16b5e6b>



**CONF42**



**startup  
grind**



**DATA**  
SUMMIT  
DATA DRIVEN DECISIONS. DATA-LED INNOVATION.

MAY 8–9  
BOSTON, MA

**NLITR'24**  
SUMMIT  
COLLABORATE, EXPLORE, LEARN

**Design, Code,  
Learn Generative AI  
& Compete  
creating chatbots  
for a  
nonprofit cause**

help others. give first. make friends.