Adding Generative AI to Real-Time Streaming Pipelines

Tim Spann
Principal Developer Advocate

Nov-2024

Introduction

Overview

GenAI Architecture

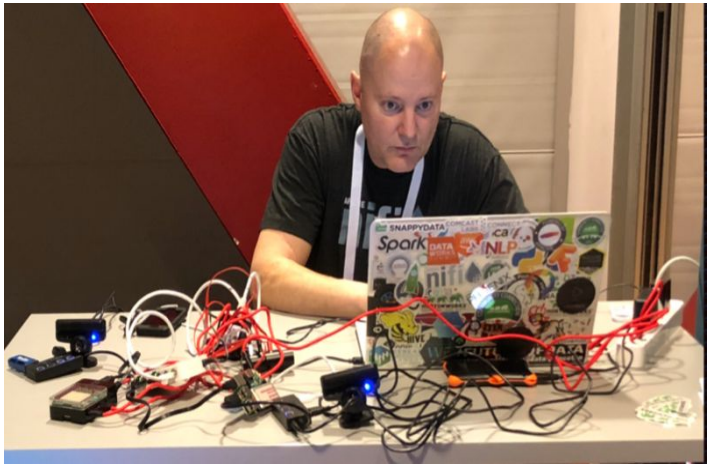
Streaming Projects

Demos

Resources

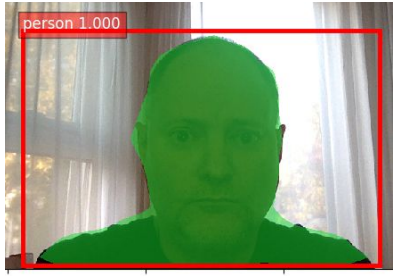
Q&A

Tim Spann

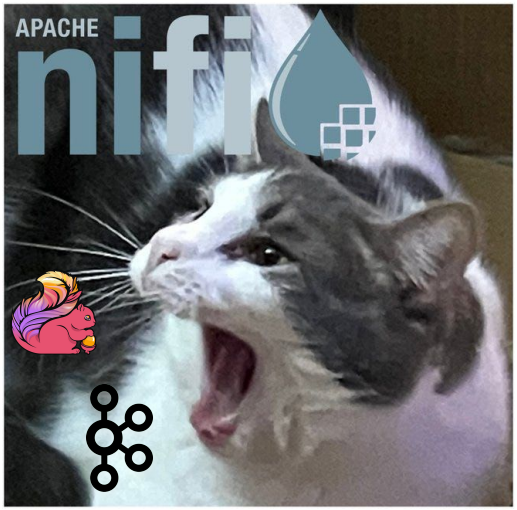


Twitter: @**PaasDev** // Blog: **datainmotion.dev**
Principal Developer Advocate / Field Engineer
NY AI Meetups
ex-Pivotal, ex-Cloudera, ex-StreamNative, ex-PwC, ex-HPE, ex-E&Y.

https://medium.com/@tspann
https://github.com/tspannhw



AI + Streaming Weekly by Tim Spann



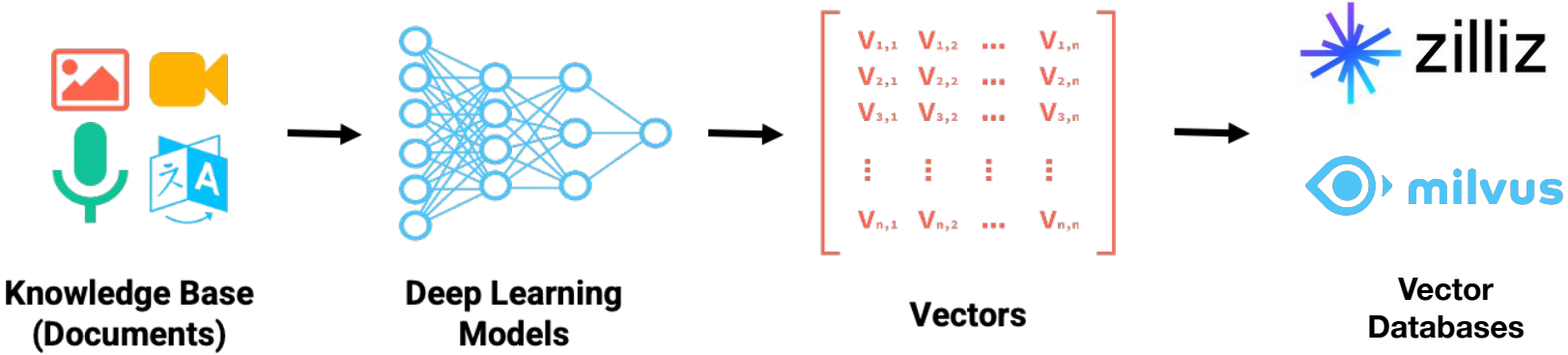
https://bit.ly/32dAJft

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.

The challenge of Unstructured Data

- **Problem:** Unstructured data comes in lots of forms, no easy way to interact with it all
- **Solution:** Vector embeddings
- **How:** Neural networks e.g. embedding models



Unstructured Data is Everywhere

Unstructured data is any data that does not conform to a predefined data model.

rrently, 90% of unstructured data is never analyzed.

Text

Images

Videos

and more!

Vector Search Overview

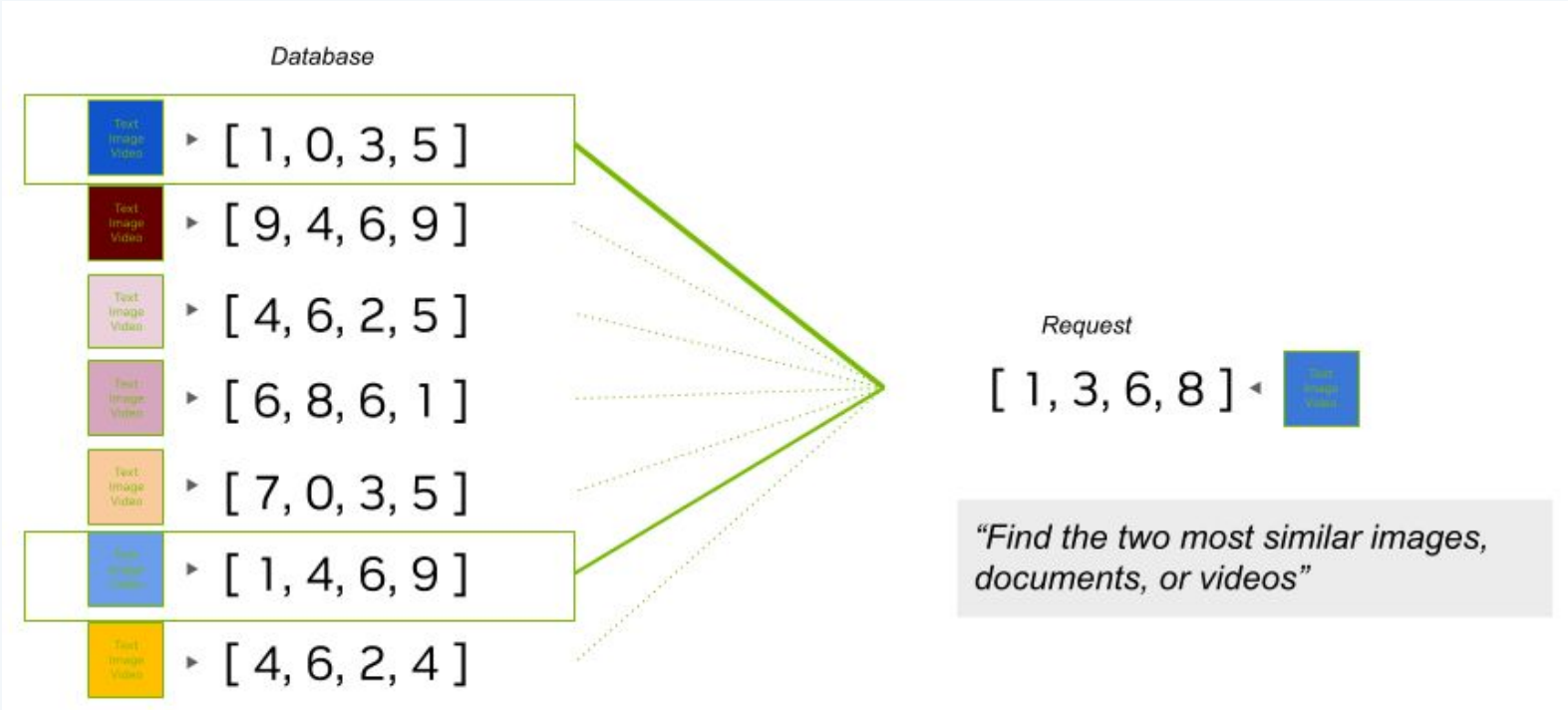
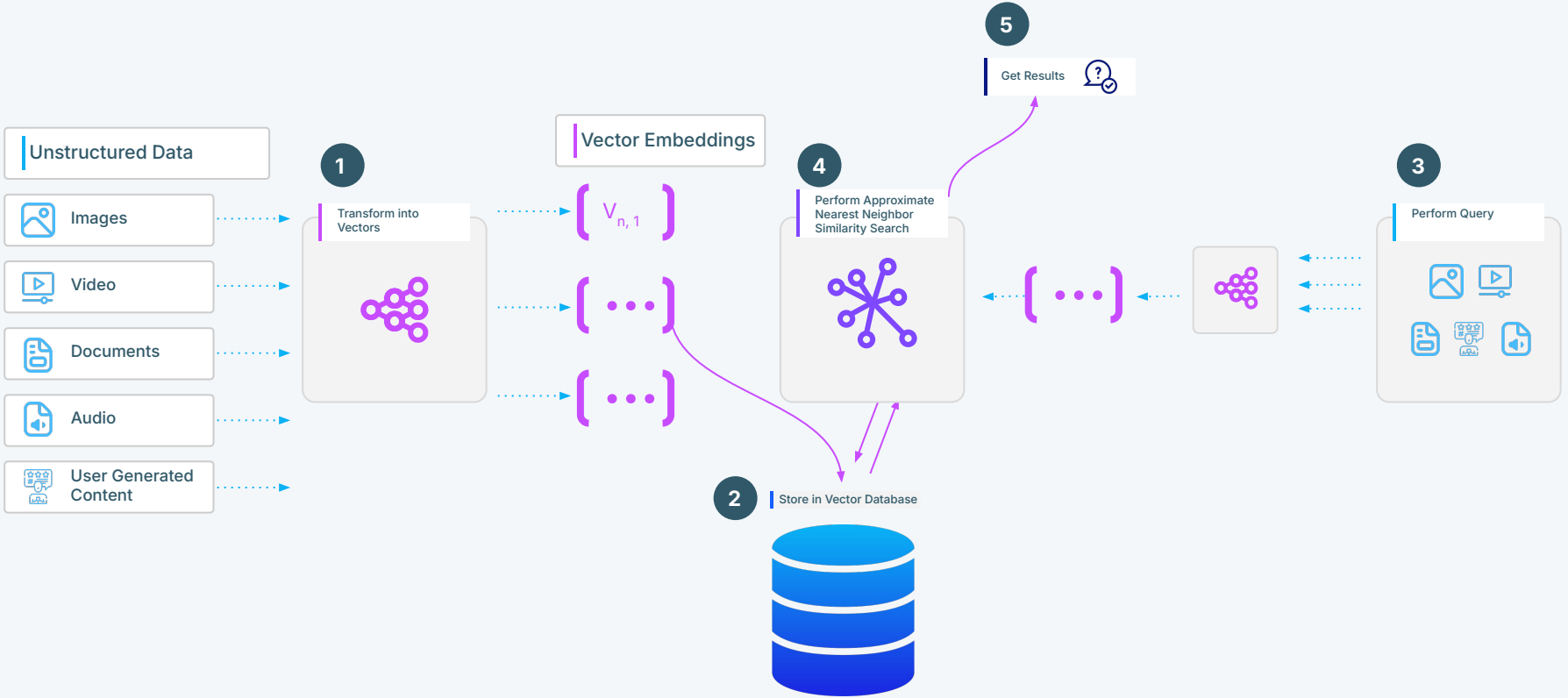
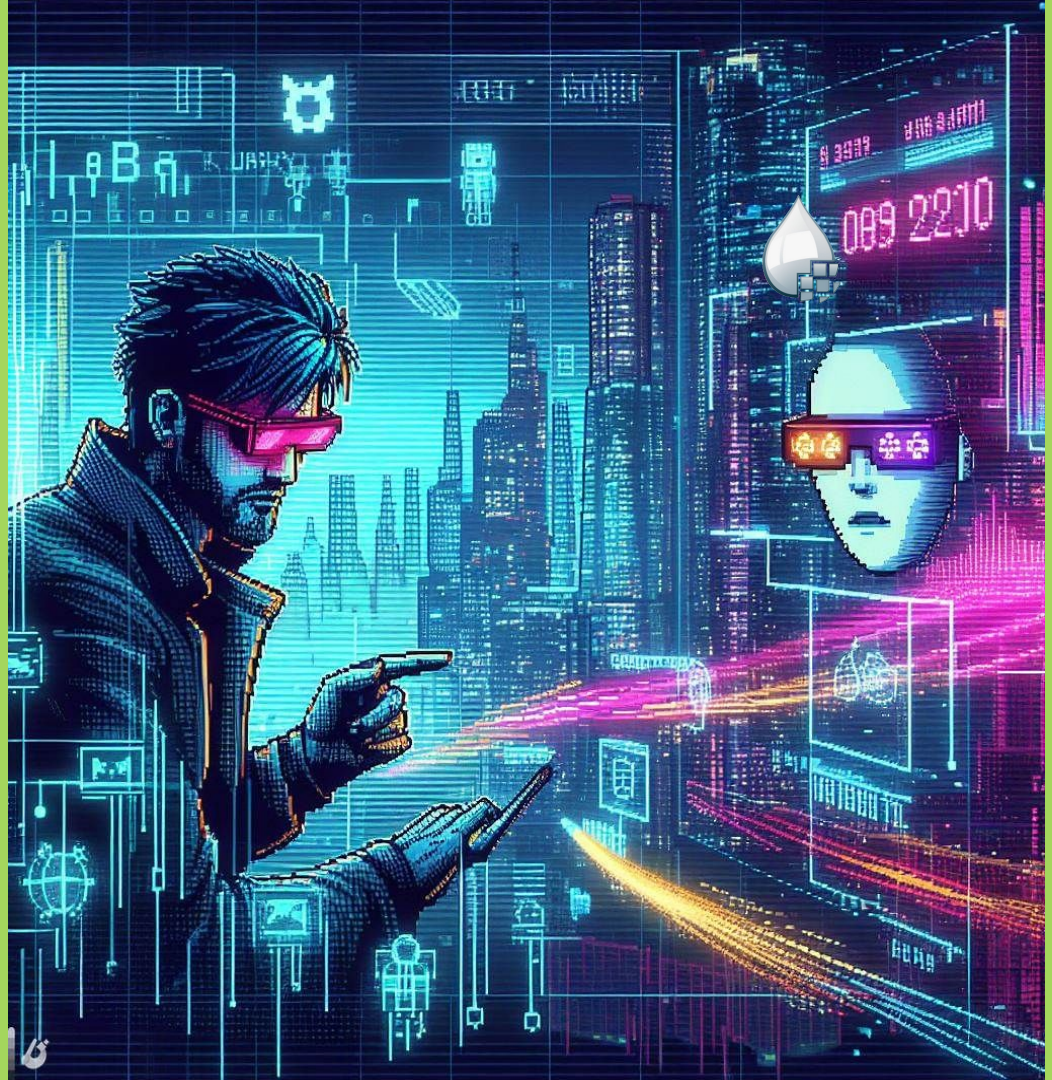


Image from Nvidia

How Similarity Search Works





Real-Time 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

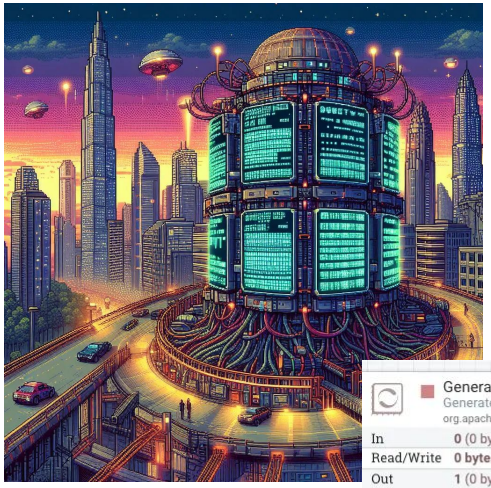
Act as a Discord, REST, Kafka, SQL, Slack bot to roundtrip discussions

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

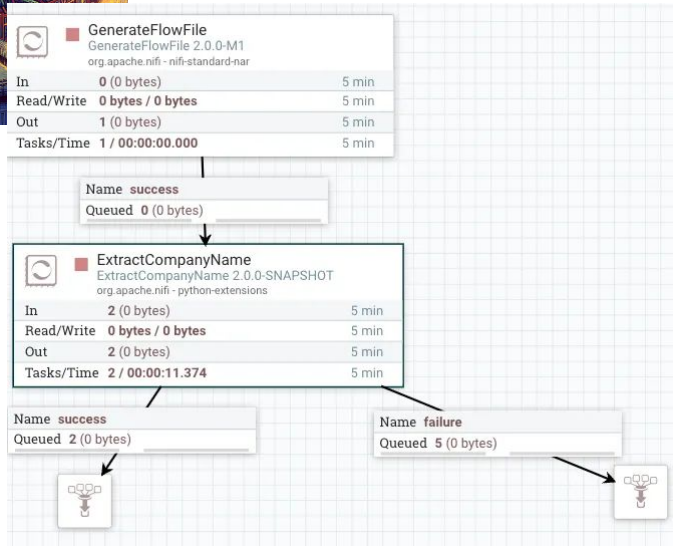
https://cwiki.apache.org/confluence/display/NIFI/NiFi+2.0+Release+Goals

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



Extract Company Names

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

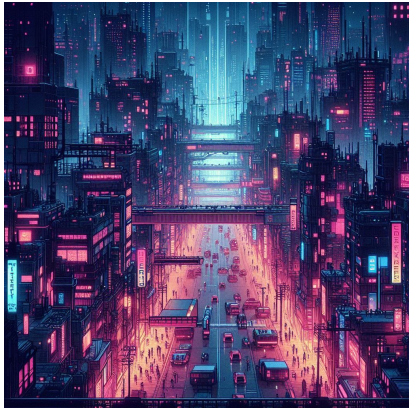


https://github.com/tspannhw/FLaNK-python-ExtractCompanyName-processor

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

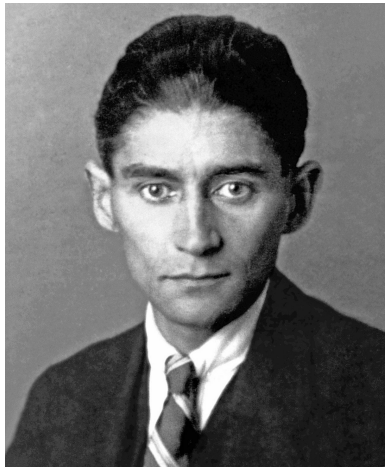
- Python 3.10+
- Hugging Face
- Transformers
- facial_emotions_image_detection
- Image Classification
- Adds labels/scores to FlowFile Attributes
- Does not require download or copies of your images

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

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.



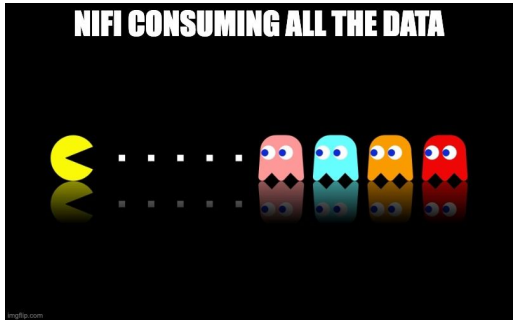
Franz Kafka was a German-speaking Bohemian novelist and short-story writer, widely regarded as one of the major figures of 20th-century literature. His work fuses elements of realism and the **fantastic**.
Wikipedia

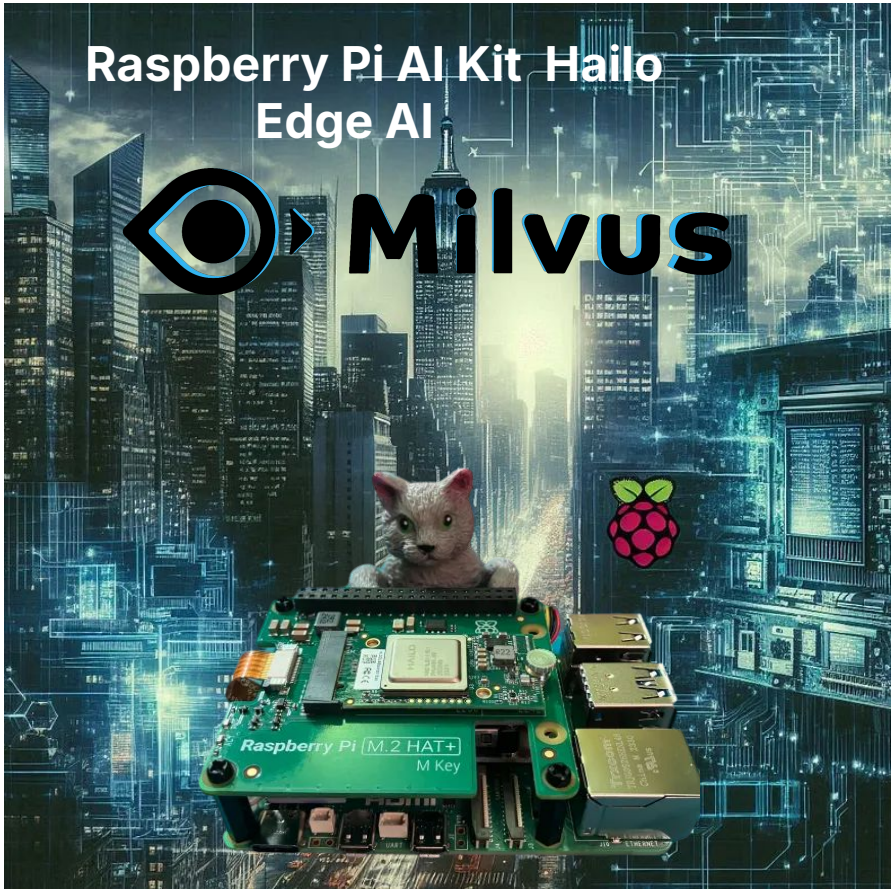
Open Source Edition

- 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





https://medium.com/@tspann/unstructured-data-processing-with-a-raspberry-pi-ai-kit-c959dd7fff47

Raspberry Pi AI Kit Hailo
Edge AI Pose Estimation

Milvus

https://medium.com/@tspann/from-the-edge-to-the-cloud-and-back-again-01095e95a783

Spann

Street Cameras

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