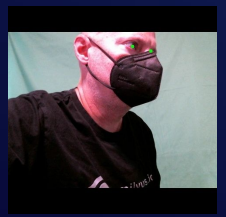
zilliz

AICamp | NEW YORK
IN PERSON MEETUP

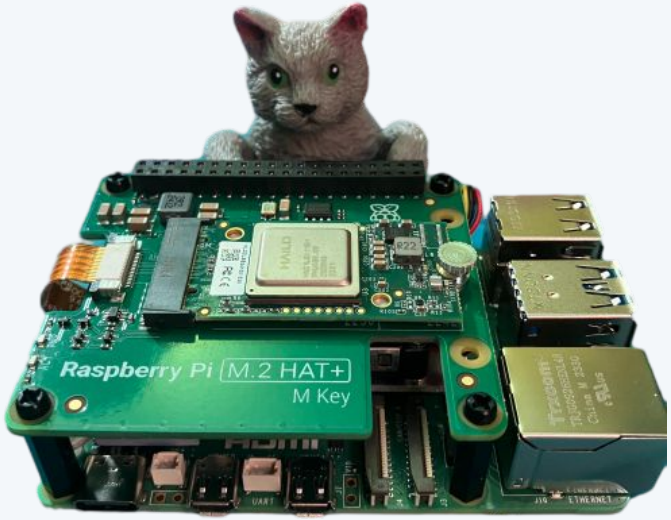
AI, LLMs, ML and Data Meetup

Human Pose Estimation in Real-Time Utilizing Edge AI Accelerated Hardware

August 15, 2024



Slides



https://github.com/tspannhw/AIM-RPIAIKit-PoseEstimation

Speaker

Tim Spann

Principal Developer Advocate,
Zilliz

tim.spann@zilliz.com
https://www.linkedin.com/in/timothyspann/
https://x.com/paasdev
https://github.com/tspannhw
https://github.com/milvus-io/milvus

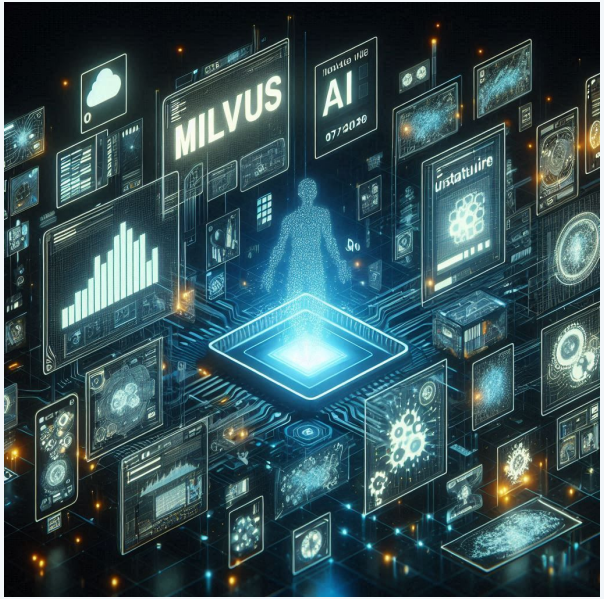
https://proceedings.neurips.cc/paper/2021/hash/299dc35e747eb77177d9cea10a802da2-Abstract.html

SPANN:
Highly-efficient Billion-scale Approximate Nearest Neighborhood Search

Coincidence? I think not.

https://proceedings.neurips.cc/paper/2021/hash/299dc35e747eb77177d9cea10a802da2-Abstract.html

Slides



https://www.slideshare.net/slideshow/08-13-2024-nyc-meetup-unstructured-data-processing-from-cloud-to-edge-milvus/270956288

Agenda

01 **Introduction**

Unstructured data, vector databases, traditional databases, similarity search, Milvus, Vector Database

02 **Human Pose Estimation**

Capture, AI

03 **App and Demo**

Running edge AI application connected to cloud

01

This again?

zilliz

What is Milvus ideal for?

Purpose-built to store, index and query vector embeddings from unstructured data **at scale**.

- Advanced filtering
- Hybrid search
- Durability and backups
- Replications/High Availability
- Sharding
- Aggregations
- Lifecycle management
- Multi-tenancy

- High query load
- High insertion/deletion
- Full precision/recall
- Accelerator support (GPU, FPGA)
- Billion-scale storage

zilliz

We've built technologies for various types of use cases

Index Types

Offer a wide range of **15 indexes** support, including popular ones like HNSW, PQ, Binary, Sparse, DiskANN and GPU index

Empower developers with tailored search optimizations, catering to performance, accuracy and cost needs

Search Types

Support multiple types such as **top-K ANN, Range ANN, sparse & dense, multi-vector, grouping,** and metadata **filtering**

Enable query flexibility and accuracy, allowing developers to tailor their information retrieval needs

Multi-tenancy

Enable **multi-tenancy** through collection and partition management

Allow for efficient resource utilization and customizable data segregation, ensuring secure and isolated data handling for each tenant

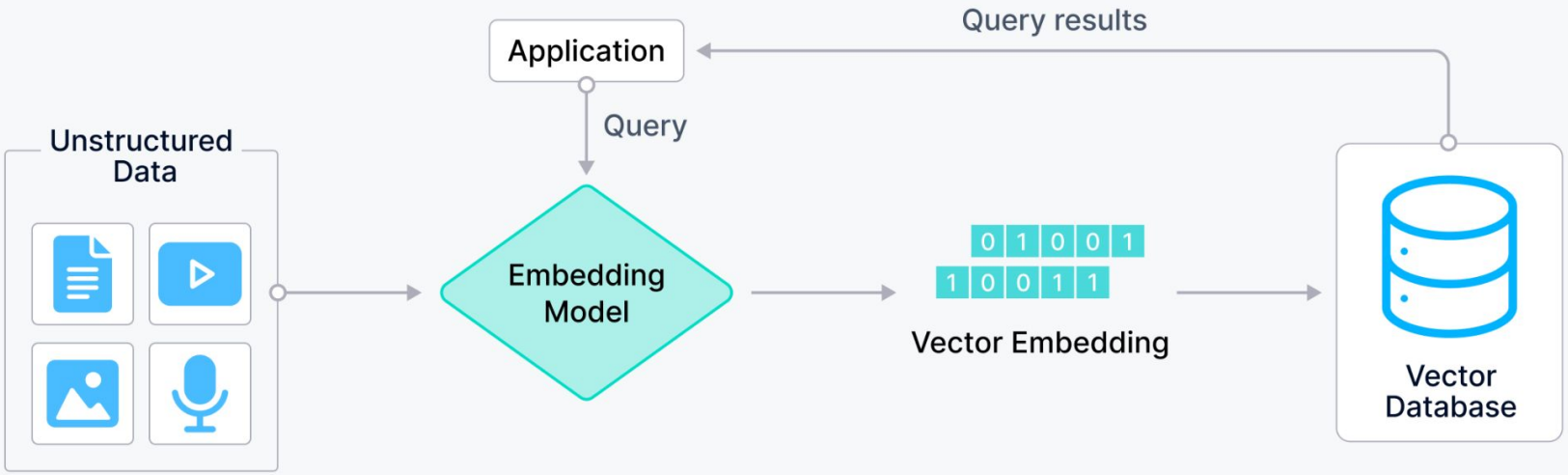
Compute Types

Designed for various compute powers, such as **AVX512, Neon for SIMD, quantization cache-aware optimization** and **GPU**

Leverage strengths of each hardware type, ensuring high-speed processing and cost-effective scalability for different application needs

Vector Database : making sense of unstructured data

A vector database stores embedding vectors and allows for semantic retrieval of various types of unstructured data.



Milvus, OSS vector database since 2019

Originally created by Zilliz, hosted by the Linux Foundation

28K+
GitHub Stars

270+
Contributors

10000+
Enterprise users

70M+
Downloads

02

Human Pose Estimation

zilliz

What is it?

https://paperswithcode.com/task/pose-estimation

1431 papers with code

Human Pose Estimation is a computer vision technique that locates and estimates things like eyes, joints in motion.

It looks pretty cool and has some interesting applications for medical purposes and robotics. For me, it was one of the cool examples that runs on the AI Kit.

zilliz

Pose Estimation by Hailo 8L

Each person is identified and represented by 17 keypoints

Examples

nose, eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles.

We are tracking eyes.

https://github.com/tensorboy/centerpose
https://softwaremill.com/human-pose-estimation-2023-guide/

https://github.com/hailo-ai/hailo_model_zoo/blob/master/docs/public_models/HAILO8/HAILO8_pose_estimation.rst

zilliz

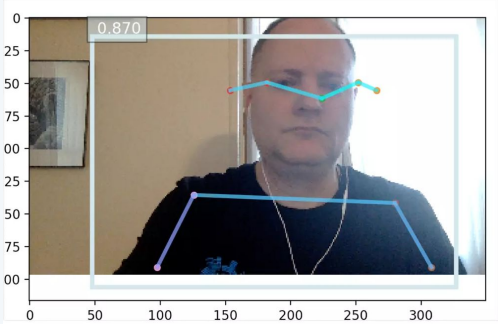
Pose Estimation on Hailo 8L

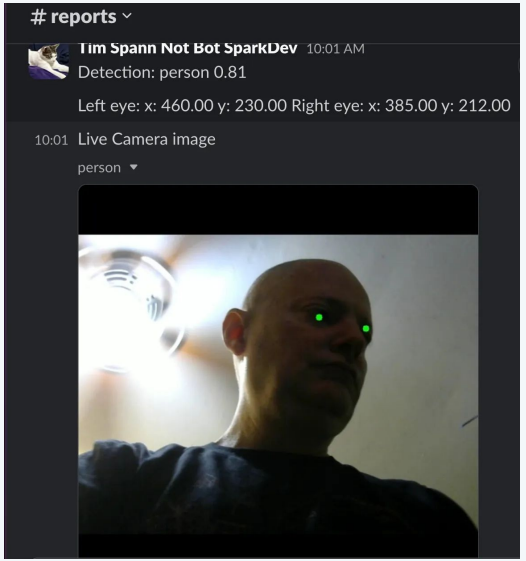
Pose Estimation COCO

Yolov8s_pose

Hailo-8L

https://github.com/ultralytics/ultralytics





 |

zilliz

03

Just Show Me

zilliz

Show Me The Source Code

```
lefteye = (f"x: {left_eye_x:.2f} y: {left_eye_y:.2f}")
righteye = (f"x: {right_eye_x:.2f} y: {right_eye_y:.2f}")

try:
        imageembedding = extractor(strfilename)
         milvus_client.insert( COLLECTION_NAME, {"vector": imageembedding,
"lefteye": lefteye,
            "righteye": righteye, "label": label, "confidence": confidence})
except Exception as e:
        print("An error:", e)
```

https://github.com/tspannhw/AIM-RPIAIKit-PoseEstimation
https://github.com/hailo-ai/hailo-rpi5-examples

zilliz

Edge Unstructured Data

- Vision to Images and Videos
- Audio from Cameras and Microphones
- Raw Text
- Edge Neural Networks and Gen AI
- Unstructured Data Processing and Vector DB

 |

zilliz

edgenote.ipynb

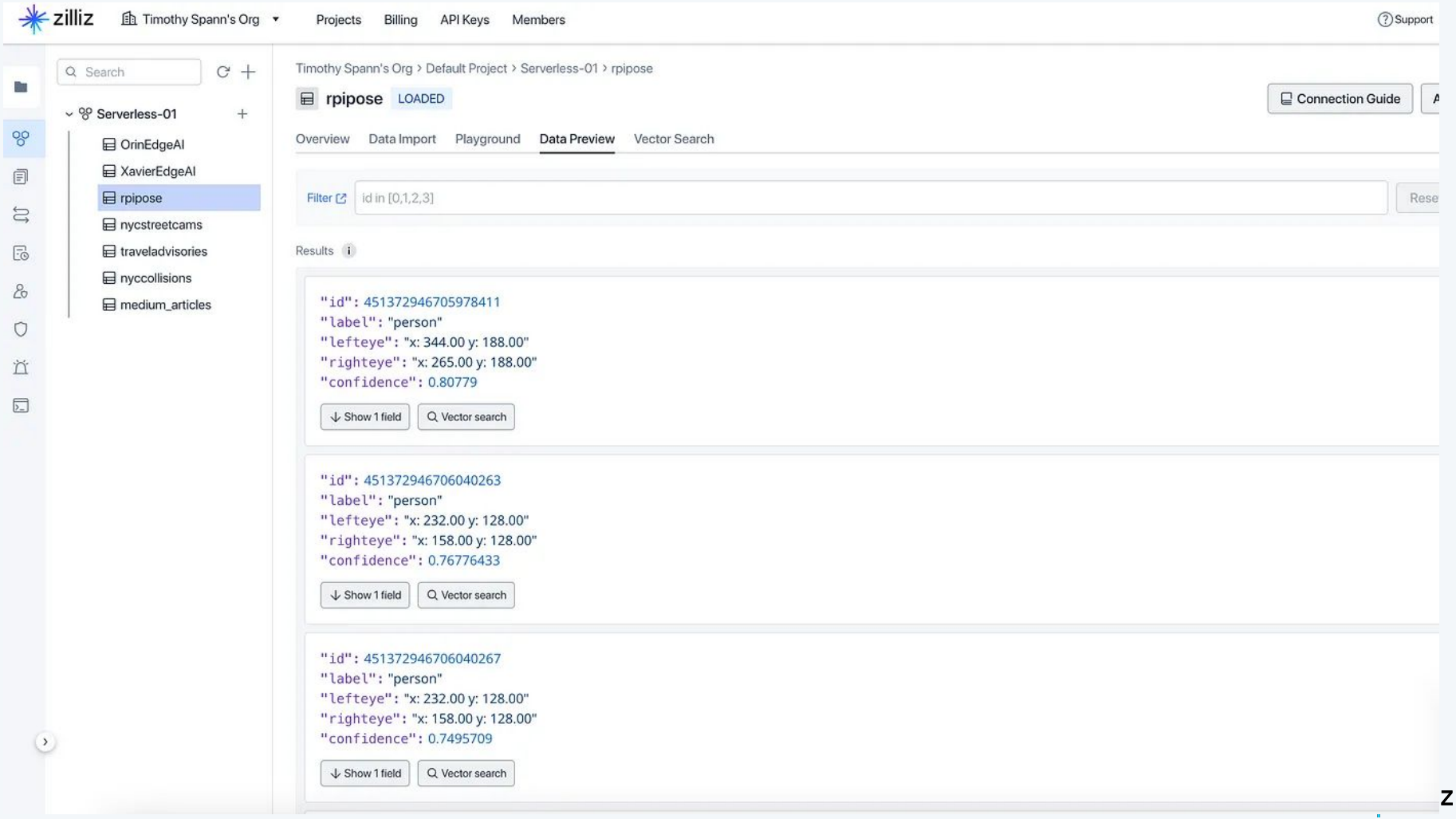
Code

[9]:
```
# ------------------------------------------------------------------------------
# Search Milvus for that vector and filter by a label
# ------------------------------------------------------------------------------
results = milvus_client.search(
    COLLECTION_NAME,
    data=[extractor(search_image_name)],
    filter='label in ["person"]',
    output_fields=["label", "confidence", "id", "lefteye", "righteye"],
    search_params={"metric_type": "COSINE"},
    limit=5
)

# ------------------------------------------------------------------------------
# Iterate through last five results and display metadata and image
# ------------------------------------------------------------------------------
for result in results:
    for hit in result[:5]:
        label = hit["entity"]["label"]
        confidence = hit["entity"]["confidence"]
        righteye = hit["entity"]["righteye"]
        lefteye = hit["entity"]["lefteye"]

        print(f"Detection: {label} {confidence:.2f} Left Eye: {lefteye}, Right Eye: {righteye}" )

        try:
            response = client.chat_postMessage(
                channel="C06NE1FU6SE",
                text=(f"Search Result Detection: {label} {confidence:.2f}")
            )
        except SlackApiError as e:
            # You will get a SlackApiError if "ok" is False
            assert e.response["error"]

        try:
            response = client.chat_postMessage(
                channel="C06NE1FU6SE",
                text=(f" Search Result Left eye: {lefteye} Right eye: {righteye} ")
            )
        except SlackApiError as e:
            # You will get a SlackApiError if "ok" is False
            assert e.response["error"]
```

```
Detection: person 0.79 Left Eye: x: 437.00 y: 138.00, Right Eye: x: 381.00 y: 124.00
Detection: person 0.61 Left Eye: x: 284.00 y: 142.00, Right Eye: x: 219.00 y: 114.00
Detection: person 0.77 Left Eye: x: 232.00 y: 128.00, Right Eye: x: 158.00 y: 128.00
Detection: person 0.75 Left Eye: x: 232.00 y: 128.00, Right Eye: x: 158.00 y: 128.00
Detection: person 0.81 Left Eye: x: 344.00 y: 188.00, Right Eye: x: 265.00 y: 188.00
```

[ ]:

zilliz

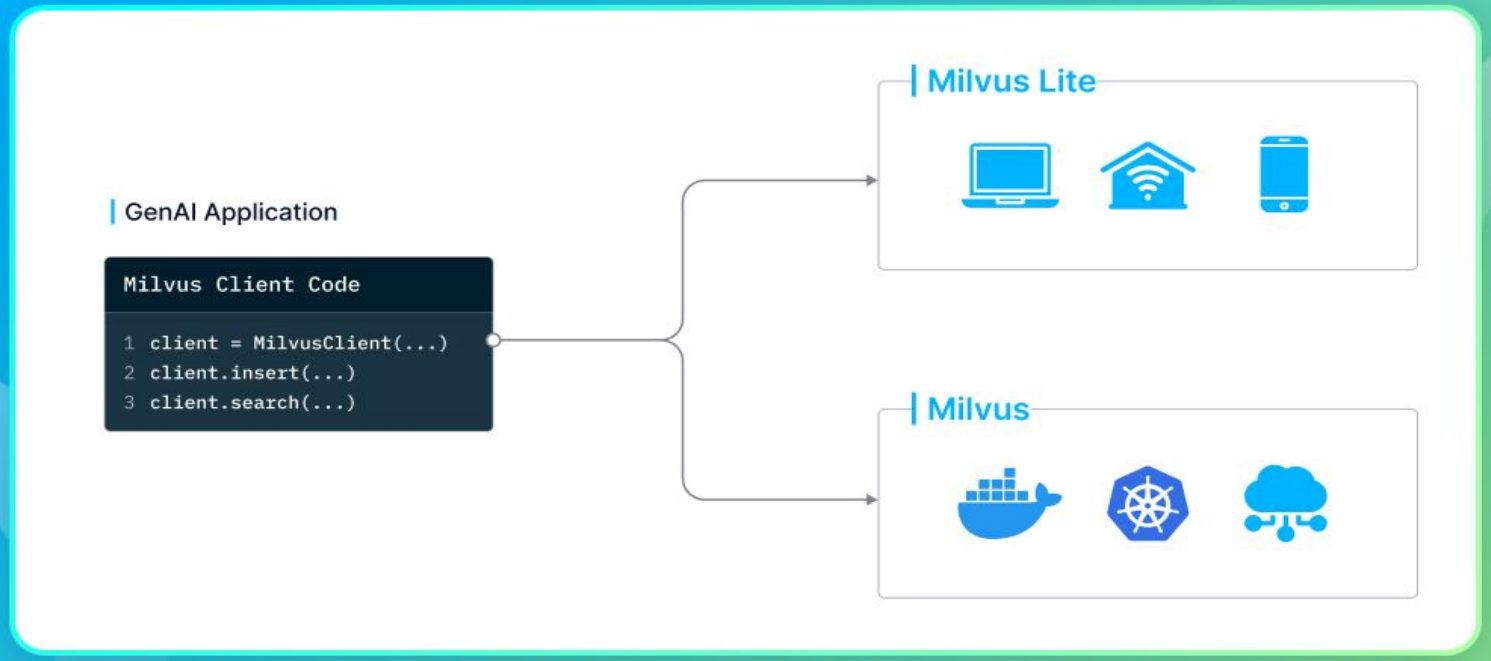

zilliz
Timothy Spann's Org
Projects
Billing
API Keys
Members
Support
Search
Serverless-01
OrinEdgeAI
XavierEdgeAI
rpipose
nycstreetcams
traveladvisories
nyccollisions
medium_articles
Timothy Spann's Org > Default Project > Serverless-01 > rpipose
rpipose
LOADED
Connection Guide
Overview
Data Import
Playground
Data Preview
Vector Search
Filter
id in [0,1,2,3]
Results
"id": 451372946705978411
"label": "person"
"lefteye": "x: 344.00 y: 188.00"
"righteye": "x: 265.00 y: 188.00"
"confidence": 0.80779
Show 1 field
Vector search
"id": 451372946706040263
"label": "person"
"lefteye": "x: 232.00 y: 128.00"
"righteye": "x: 158.00 y: 128.00"
"confidence": 0.76776433
Show 1 field
Vector search
"id": 451372946706040267
"label": "person"
"lefteye": "x: 232.00 y: 128.00"
"righteye": "x: 158.00 y: 128.00"
"confidence": 0.7495709
Show 1 field
Vector search

04

Next Steps

zilliz

Build Once
Deploy Anywhere



milvus

zilliz

Vector Database Resources

Give Milvus a Star!

Chat with me on Discord!

https://github.com/milvus-io/milvus

 |

zilliz

Takeaway:

- Closer is better
- Empowering AI Robots
- Vector Search Everywhere
- Keep your data and computation close

 |

zilliz

Q&A

 |

zilliz

Questions

- Edge AI
- Edge Hardware
- Milvus
- Vector Databases

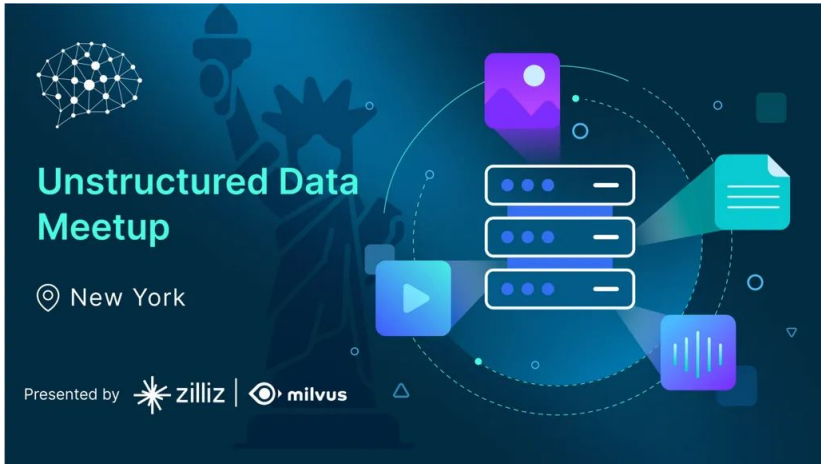
 |

zilliz

RESOURCES

zilliz

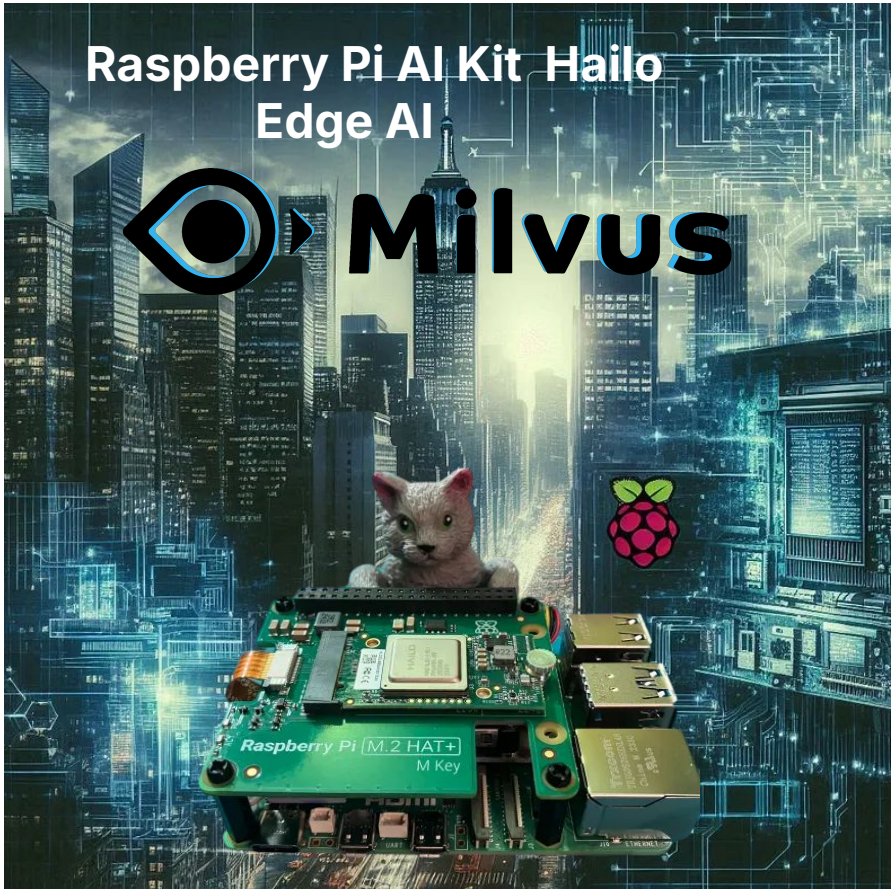
Unstructured Data Meetup



https://www.meetup.com/unstructured-data-meetup-new-york/

This meetup is for people working in unstructured data. Speakers will come present about related topics such as vector databases, LLMs, and managing data at scale. The intended audience of this group includes roles like machine learning engineers, data scientists, data engineers, software engineers, and PMs.

This meetup was formerly Milvus Meetup, and is sponsored by Zilliz maintainers of Milvus.



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

https://medium.com/@tspann/unstructured-street-data-in-new-york-8d3cde0a1e5b

Extracting Value from Unstructured Data

Example

- A company has 100,000s+ pages of proprietary documentation to enable their staff to service customers.

Problem

- Searching can be slow, inefficient, or lack context.

Solution

- Create internal chatbot with ChatGPT and a vector database enriched with company documentation to provide direction and support to employees and customers.

https://osschat.io/chat

We provide deployment flexibility for different operational, security and compliance requirements

SELF MANAGED SOFTWARE

Milvus

Most widely-adopted open source vector database

Self hosted on any machine with community support

Local

Docker

K8s

FULLY MANAGED SERVICE

Zilliz Cloud

Milvus Re-engineered for the Cloud

Available on the leading public clouds

BRING YOUR OWN CLOUD

Zilliz BYOC

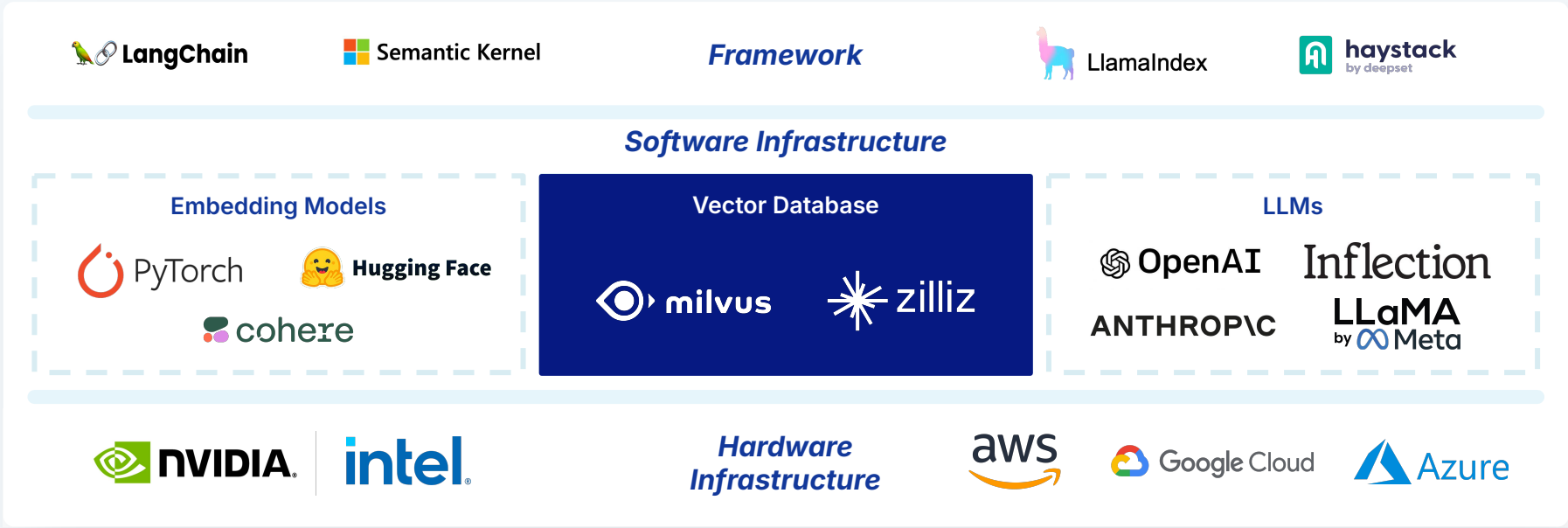
Enterprise-ready Milvus for Private VPCs

Deploy in your virtual private cloud

Coming Soon!

Coming Soon!

Well-connected in LLM infrastructure to enable RAG use cases

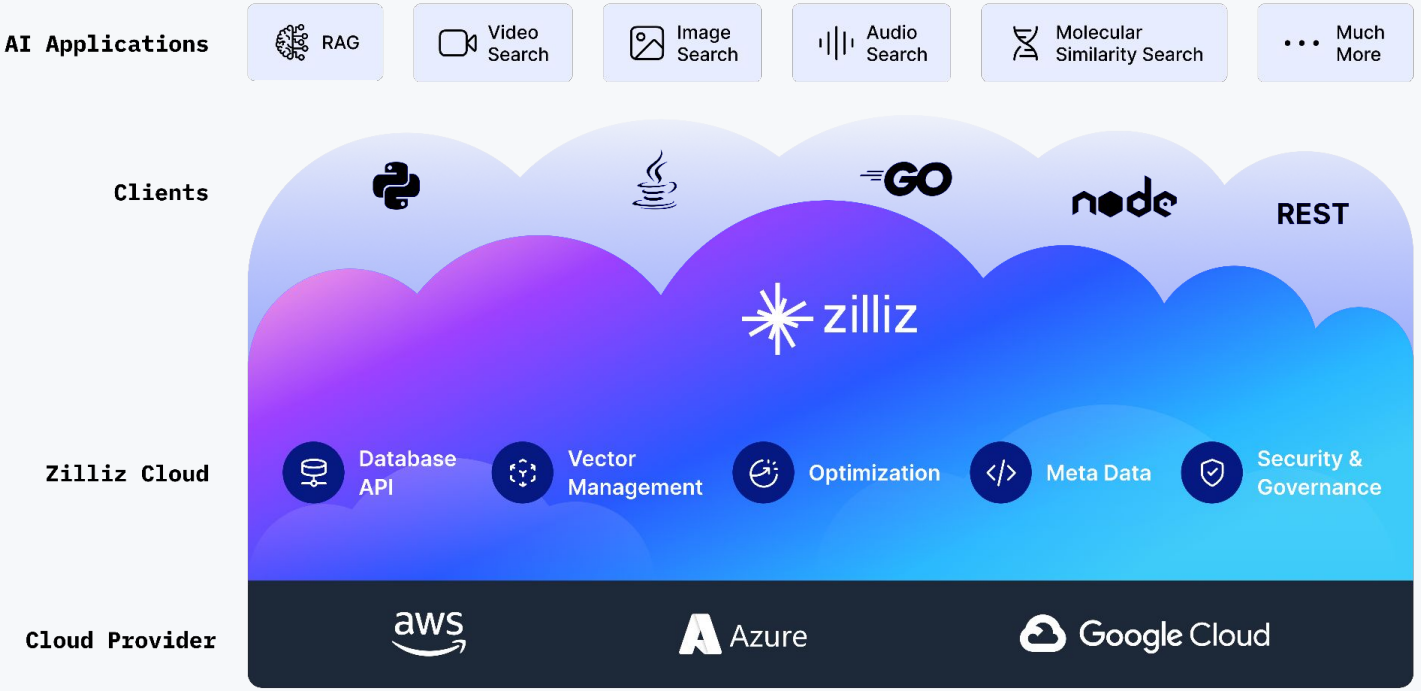


 |

zilliz

Multi-cloud: Zilliz Cloud is built atop of OSS Milvus

AWS, GCP, Azure



THANK YOU