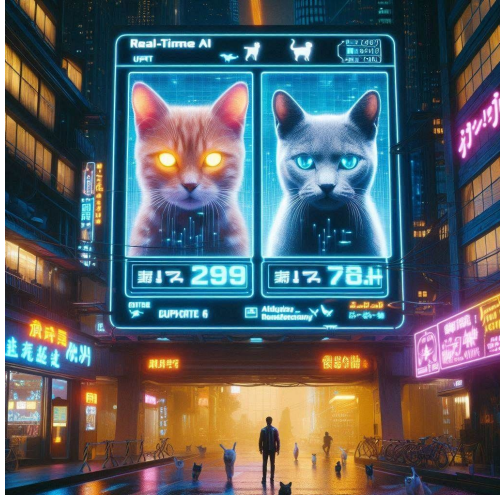
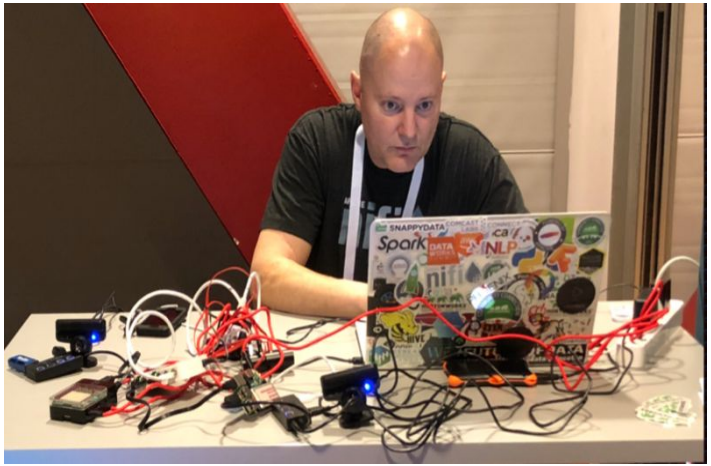


Streaming AI Pipelines with Apache NiFi and Snowflake

Tim Spann, Senior Solutions Engineer

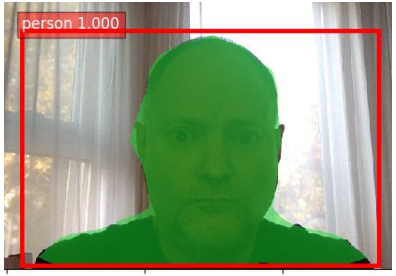
APACHE nifi

Tim Spann



paasdev.bsky.social
@PaasDev // Blog: datainmotion.dev
Senior Solutions Engineer, Snowflake
NY/NJ/Philly - Cloud Data + AI Meetups
ex-Zilliz, ex-Pivotal, ex-Cloudera, ex-HPE, ex-StreamNative, ex-EY, ex-Hortonworks.

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



snowflake

Streamlit

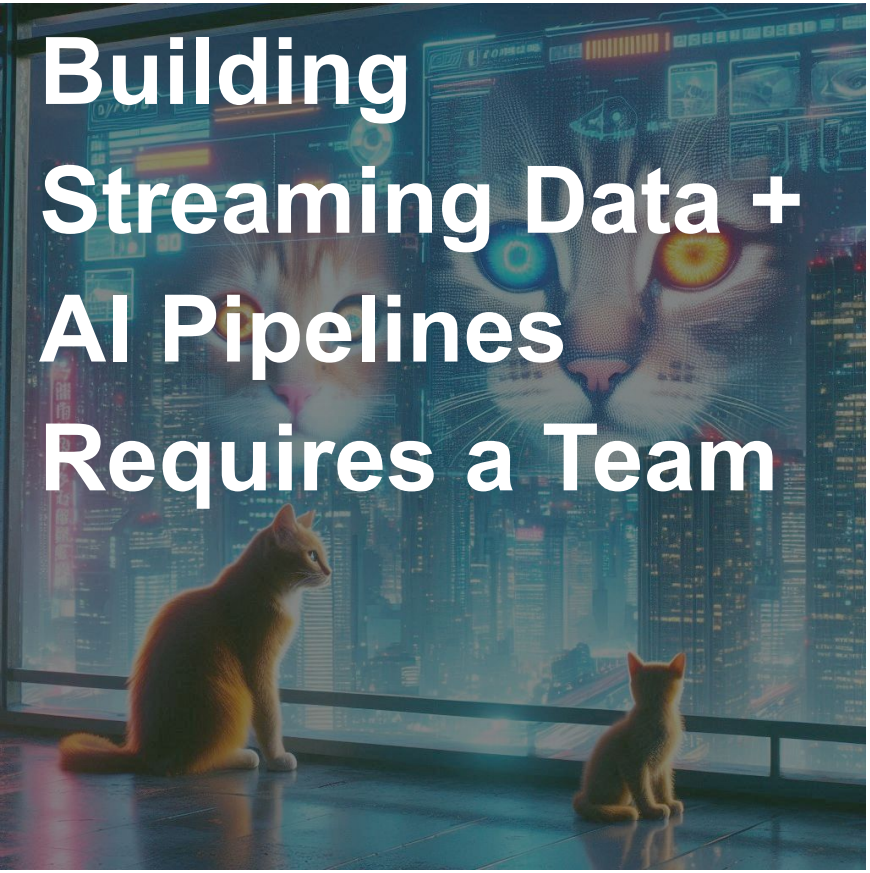
DATA + AI + Streaming Weekly

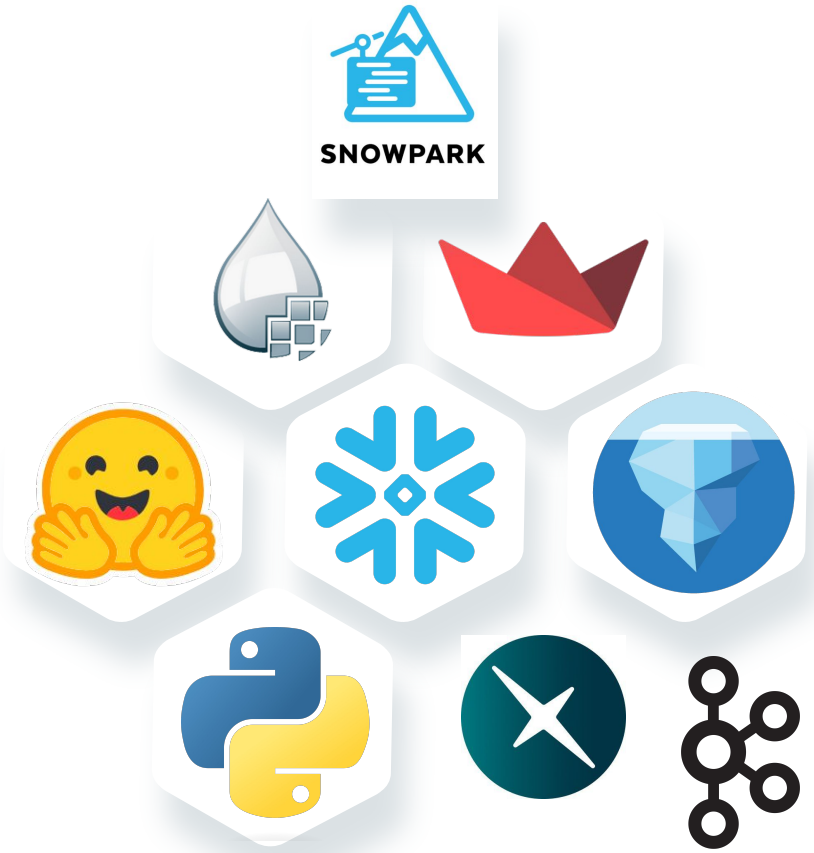
https://bit.ly/32dAJft

This week in Apache NiFi, Apache Polaris, Apache Flink, Apache Kafka, ML, AI, Streamlit, Jupyter, Apache Iceberg, Python, Java, LLM, GenAI, Snowflake, Unstructured Data and Open Source friends.

How Snowflake and Apache NiFi work with Streaming Data and AI

Building Streaming Data + AI Pipelines Requires a Team

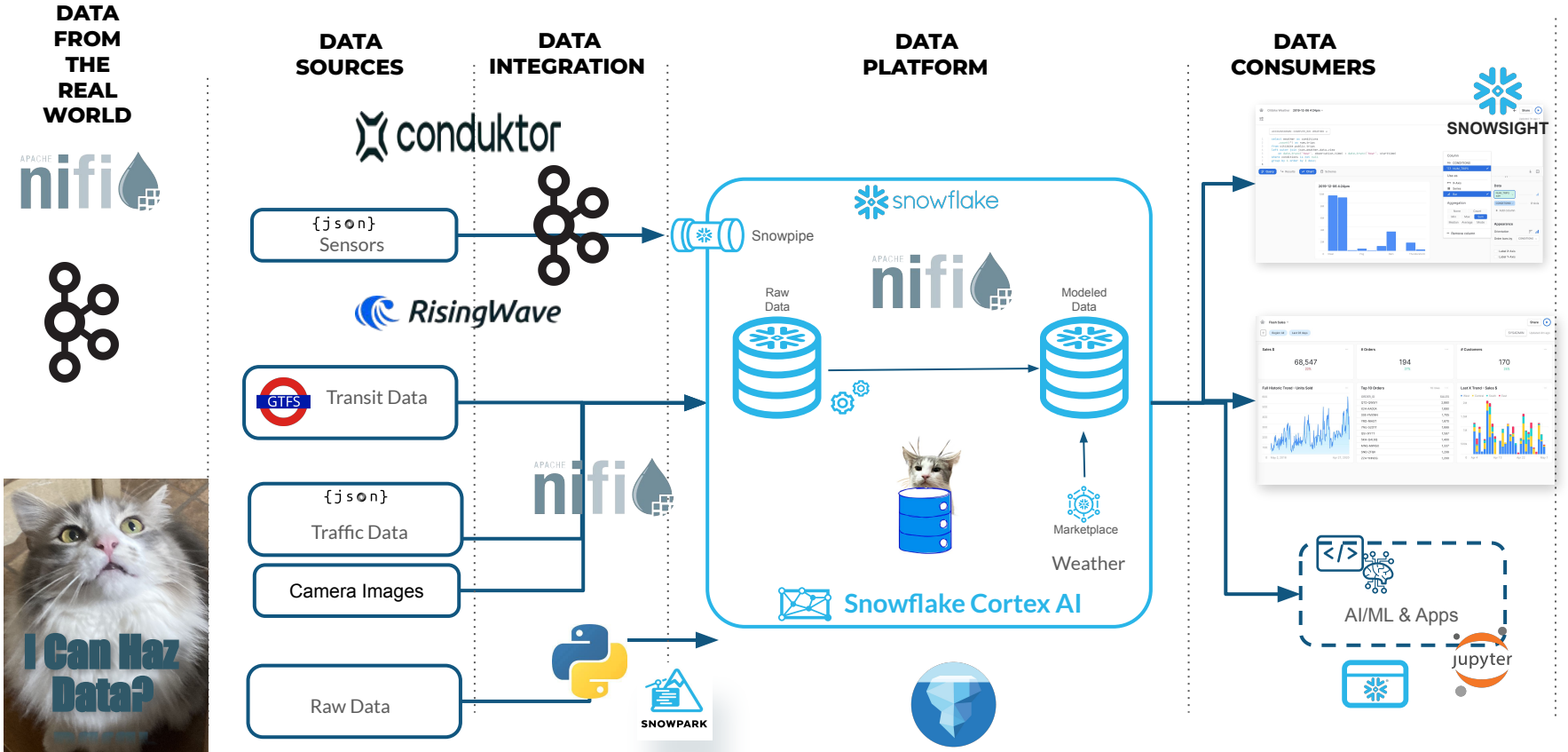




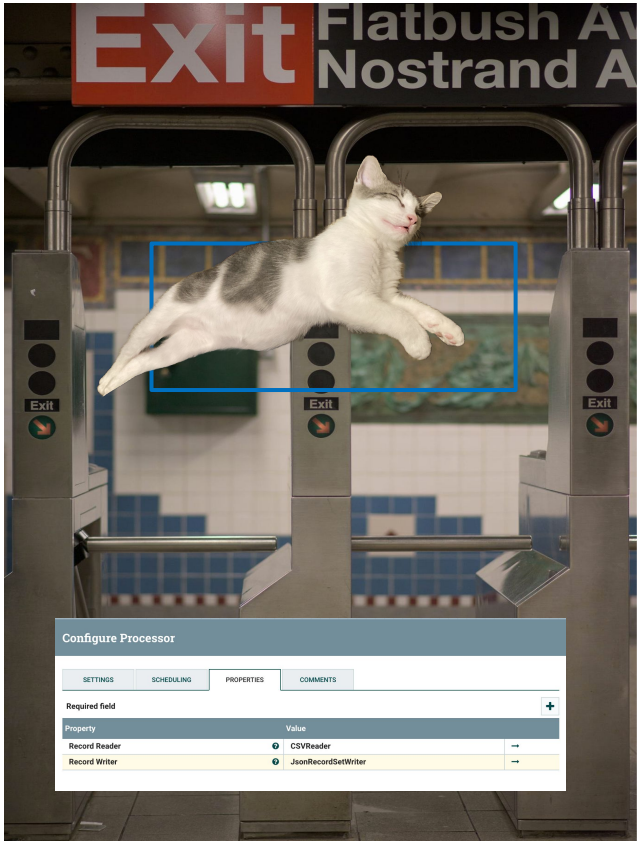
RisingWave

conduktor

Example Smart City Architecture



Apache NiFi



- From laptop to 1,000 nodes
- Ingest, Extract, Split
- Enrich, Transform
- Mature, 10 years+
- Any Data, Any Source
- LLM Calls
- Data Provenance
- Back Pressure
- Guaranteed Delivery

Unstructured Data

Unstructured

- Lots of formats
- Text, Documents, PDF
- Images, Videos, Audio
- Email, Slack, Teams
- Logs
- Binary Data Formats
- Zip
- Variants

Semi-Structured Data

- Open Data like Open AQ - Air Quality Data
- Location, Time,Sensors
- Apache Avro, Parquet, Orc
- JSON and XML
- Hierarchical Data
- Logs
- Key-Value

https://docs.snowflake.com/en/sql-reference/data-types-semistructured

Semi-structured

Structured Data

- Snowflake Tables
- Snowflake Hybrid Tables
- Apache Iceberg Tables
- Relational Tables
- Postgresql Tables
- CSV, TSV

Structured

Open LLM Options

- Arctic Instruct
- Arctic-embed-m-v2.0
- Llama-3.3-70b
- Mixtral-8x7b
- Llama3.1-405b
- Mistral-7b
- Deepseek-r1


SPIRIT
Business
Cat
ADULT
Size Costume
ONE SIZE FITS MOST


AUTOMATE EVERYTHING
imgflip.com