



Stream Native

Building FLiPN Stack Edge AI
Applications

Tim Spann | Dev Advocate



Thank You Comcast Labs for this event and
your support of Open Source Software and
the Community!

<https://comcast.github.io/>



Tim Spann, Developer Advocate at StreamNative



Tim Spann
Developer Advocate

- **FLiP(N)** Stack = Flink, Pulsar and NiFi Stack
- Streaming Systems & Data Architecture Expert
- Experience:
 - 15+ years of experience with streaming technologies including Pulsar, Flink, Spark, NiFi, Big Data, Cloud, MXNet, IoT, Python and more.
 - Today, he helps to grow the Pulsar community sharing rich technical knowledge and experience at both global conferences and through individual conversations.

CLOUDERA



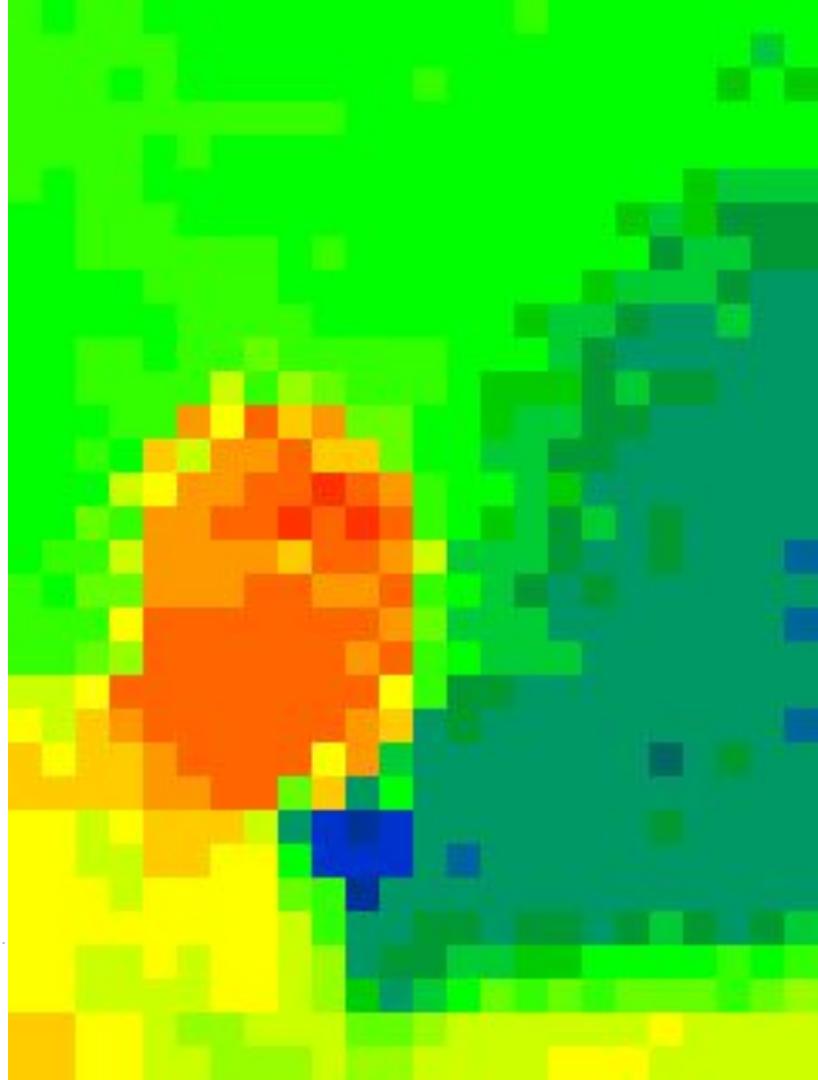
Pivotal

BARNES
&NOBLE



Hewlett Packard
Enterprise

Example Sensor Device





Apache Pulsar is a Cloud-Native
Messaging and Event-Streaming Platform.

Why Apache Pulsar?



Unified
Messaging
Platform



Guaranteed
Message
Delivery



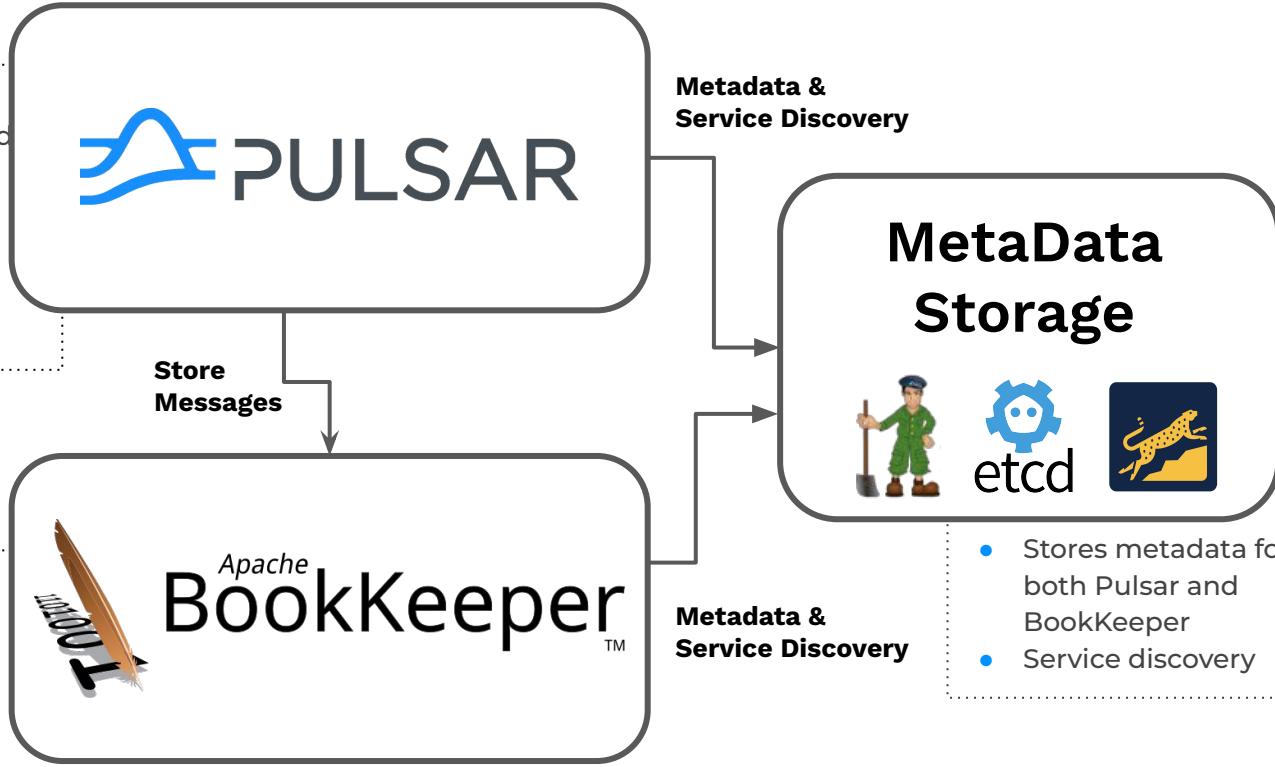
Resiliency



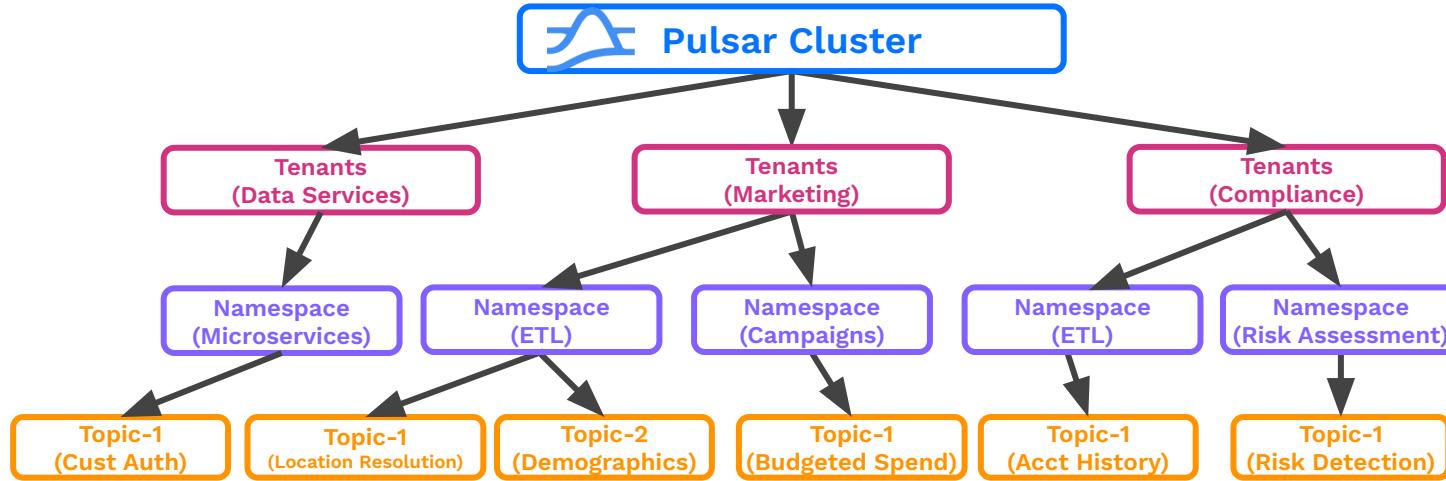
Infinite
Scalability

Key Pulsar Concepts: Architecture

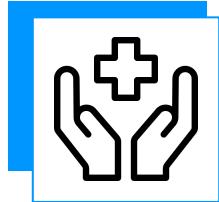
- “Brokers”
- Handles message routing and connections
- Stateless, but with caches
- Automatic load-balancing
- Topics are composed of multiple segments



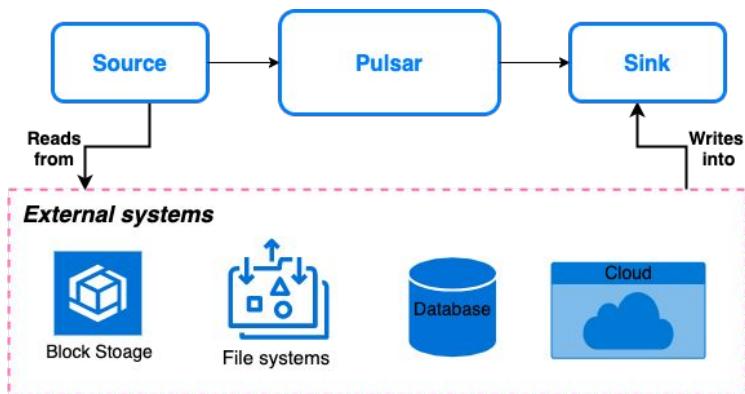
Multi-Tenancy Model



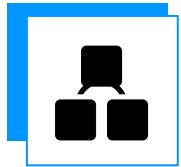
Topic URI structure: `persistent:// [tenant] / [namespace] / [topic]`



Connectivity



- **Functions** - Lightweight Stream Processing (Java, Python, Go)
- **Connectors** - Sources & Sinks (Cassandra, Kafka, ...)
- **Protocol Handlers** - AoP (AMQP), KoP (Kafka), MoP (MQTT)
- **Processing Engines** - Flink, Spark, Presto/Trino via Pulsar SQL
- **Data Offloaders** - Tiered Storage - (S3)



Use Cases

- Unified Messaging Platform
- AdTech
- Fraud Detection
- Connected Car
- IoT Analytics
- Microservices Development

FLiP(N)(S) Stack

- Apache Flink
 - Apache Pulsar
 - Pulsar Functions
 - Apache NiFi
 - Apache Spark
 - Python, Java, Golang

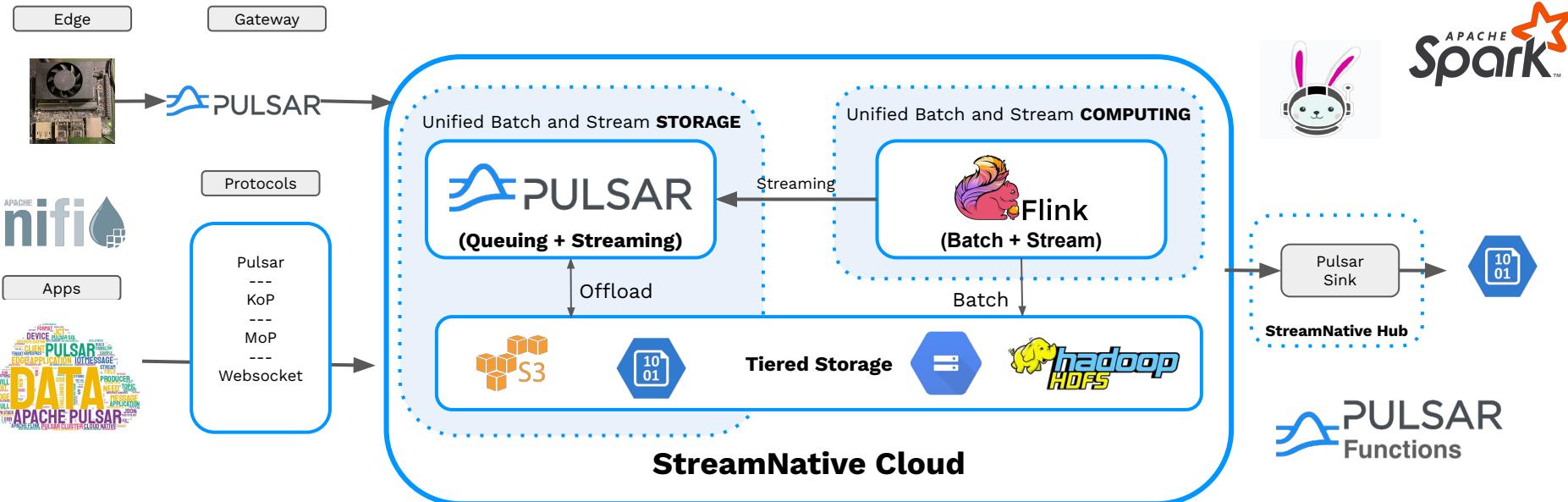


Streaming Edge Apps

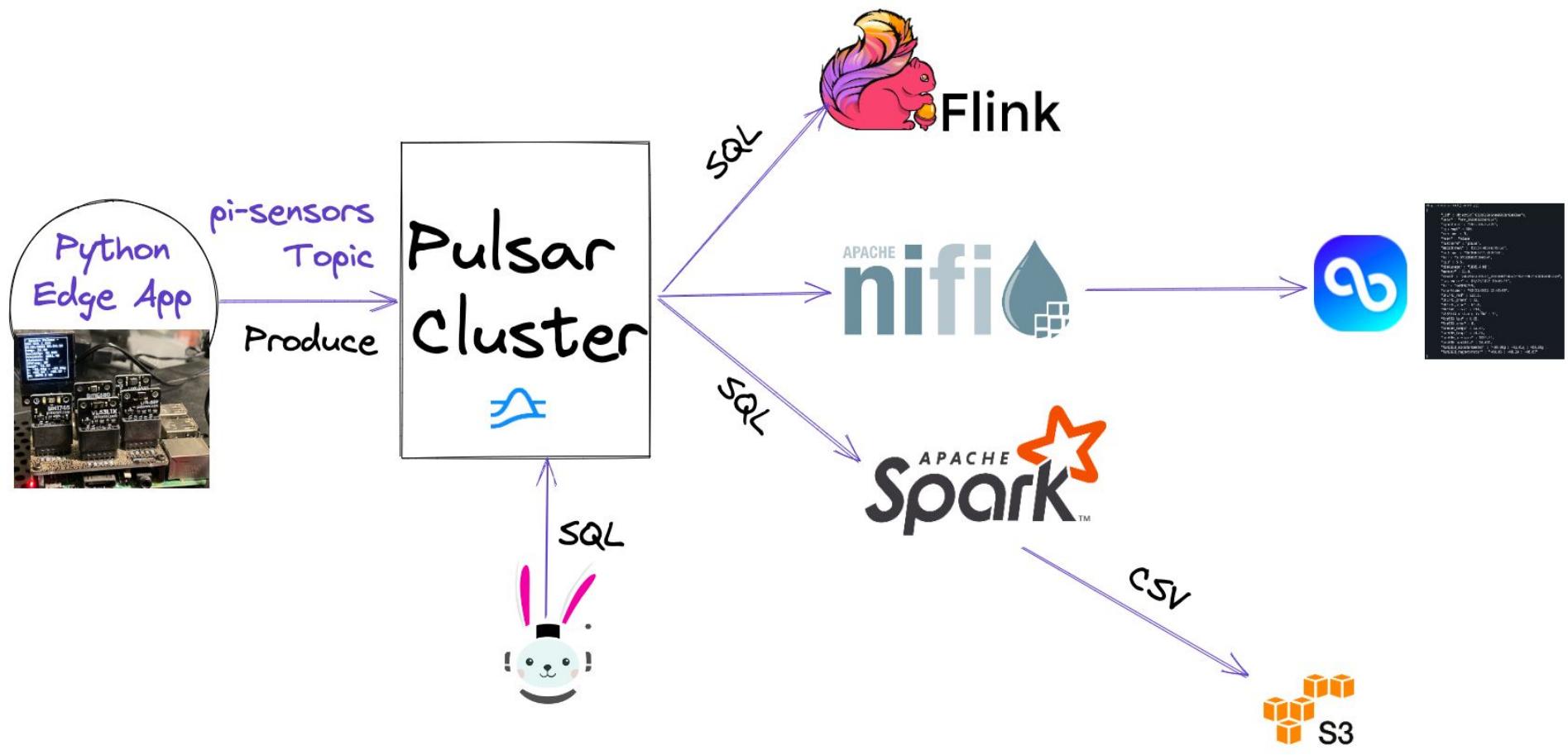
Sensors <->



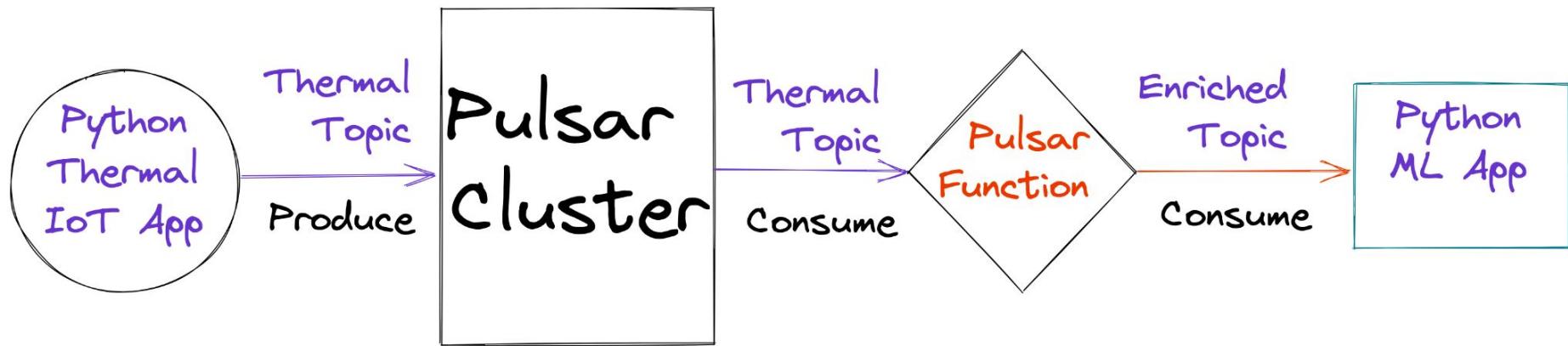
<-> Sensors <->



StreamNative

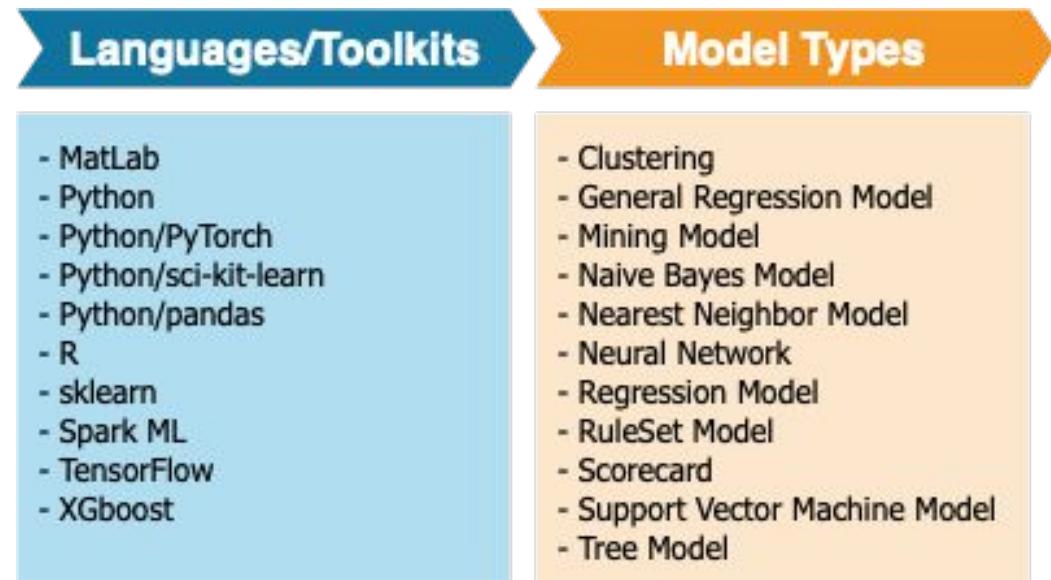


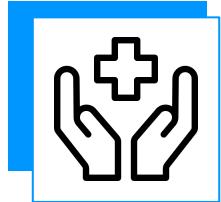
Pulsar ML Functions In Python



Pulsar Function – 3rd Party Libraries

- You can leverage 3rd party libraries within Pulsar Functions
 - DeepLearning4J
 - JPMML
 - DJL.AI
 - Keras
- Pulsar Functions are able to support:
 - A variety of ML model types.
 - Models developed with different languages and toolkits





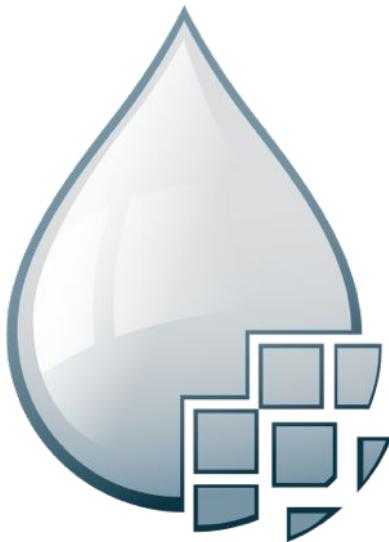
ML

- **Visual Question and Answer**
- **NLP (Natural Language Processing)**
- **Sentiment Analysis**
- **Text Classification**
- **Named Entity Recognition**
- **Content-based Recommendations**
- **Predictive Maintenance**
- **Fault Detection**
- **Fraud Detection**
- **Time-Series Predictions**
- **Naive Bayes**



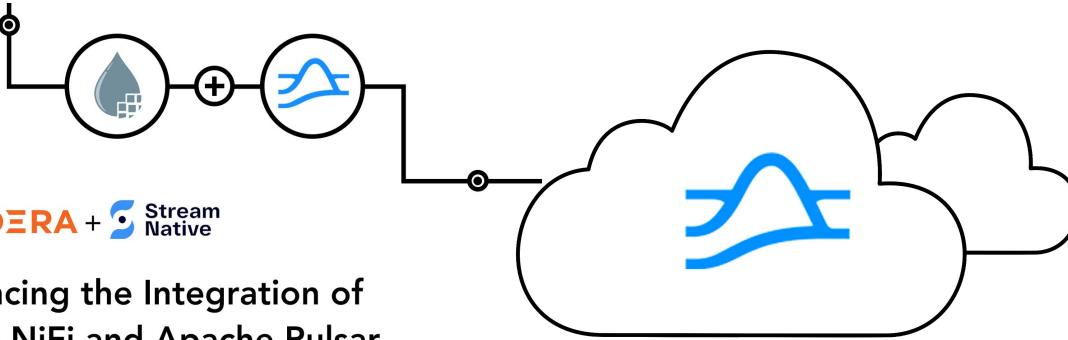
StreamNative

Why Apache NiFi?



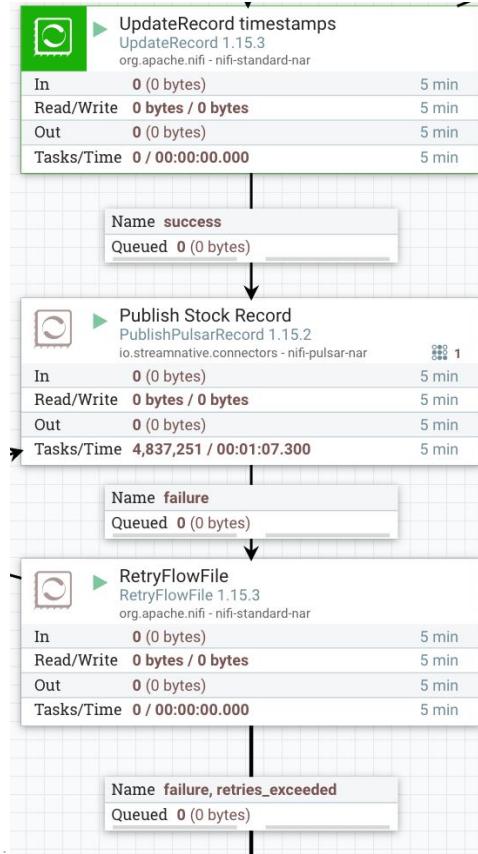
- Guaranteed delivery
- Data buffering
 - Backpressure
 - Pressure release
- Prioritized queuing
- Flow specific QoS
 - Latency vs. throughput
 - Loss tolerance
- Data provenance
- Supports push and pull models
- Hundreds of processors
- Visual command and control
- Over a 300 sources
- Flow templates
- Pluggable/multi-role security
- Designed for extension
- Clustering
- Version Control

Apache NiFi Pulsar Connector



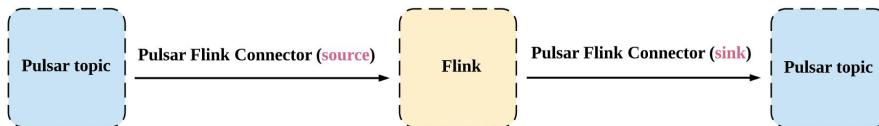
Announcing the Integration of
Apache NiFi and Apache Pulsar

<https://streamnative.io/apache-nifi-connector/>

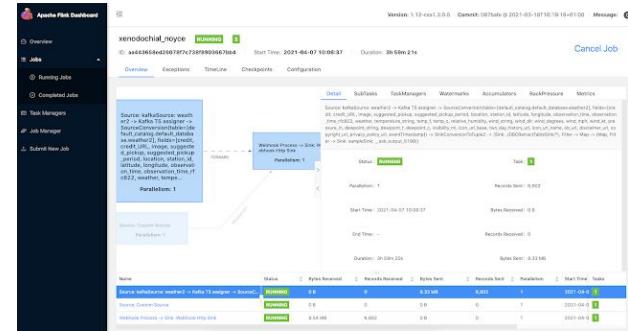




Why Apache Flink?



- Unified computing engine
- Batch processing is a special case of stream processing
- Stateful processing
- Massive Scalability
- Flink SQL for queries, inserts against Pulsar Topics
- Streaming Analytics
- Continuous SQL
- Continuous ETL
- Complex Event Processing
- Standard SQL Powered by Apache Calcite





Flink SQL

Welcome! Enter 'HELP;' to list all available commands. 'QUIT;' to exit.

```
Flink SQL> CREATE CATALOG pulsar WITH (
>   'type' = 'pulsar',
>   'service-url' = 'pulsar://pulsar1:6650',
>   'admin-url' = 'http://pulsar1:8080',
>   'format' = 'json'
> );
[INFO] Execute statement succeed.

Flink SQL>
```

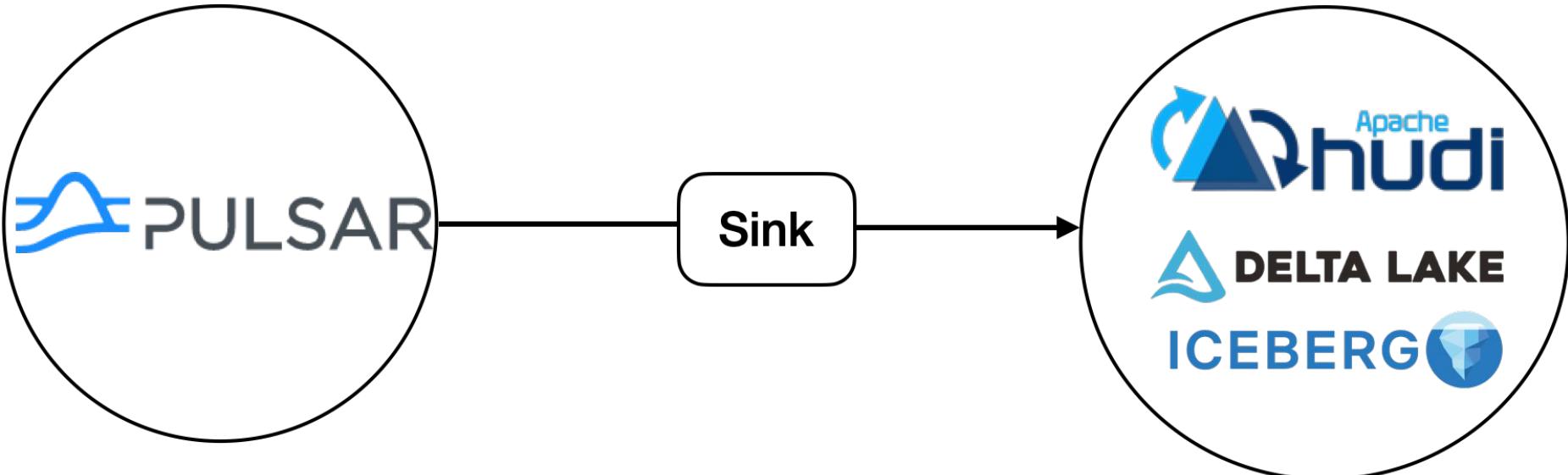
```
select aqi, parameterName, dateObserved, hourObserved, latitude,
longitude, localTimeZone, stateCode, reportingArea from airquality
```

```
select max(aqi) as MaxAQI, parameterName, reportingArea from airquality
group by parameterName, reportingArea
```

```
select max(aqi) as MaxAQI, min(aqi) as MinAQI, avg(aqi) as AvgAQI,
count(aqi) as RowCount, parameterName, reportingArea from airquality
group by parameterName, reportingArea
```



StreamNative

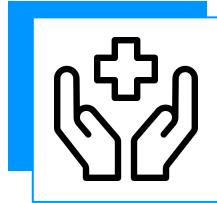


humidity	co2	datetimestamp	cputempf	ts	uuid
36.56	1127.0	2022-07-15 13:56: ...	106	1657893362	thrml_xlh_2022071 ...
36.69	1127.0	2022-07-15 13:56: ...	107	1657893367	thrml_cuv_2022071 ...



CODE +
COMMUNITY
DEMO

<https://github.com/tspannhw/FLiP-EdgeAI-PHLAI>



DEPLOYING AI WITH AN EVENT-DRIVEN PLATFORM

<https://dzone.com/articles/deploying-ai-with-an-event-driven-platform>



 DZone®

DZONE TREND REPORT

MARCH 2022

Enterprise AI

Machine Learning, Design Paradigms,
and Operational Impact

Building Edge Applications with Apache Pulsar

<https://streamnative.io/blog/engineering/2021-11-17-building-edge-applications-with-apache-pulsar/>



FLIP Stack Weekly

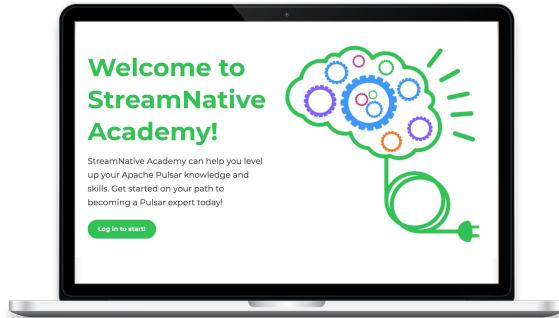
This week in Apache Flink, Apache Pulsar, Apache NiFi, Apache Spark and open source friends.

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

StreamNative Academy

Apache Pulsar Training

- ✓ Instructor-led courses
 - ✓ Pulsar Fundamentals
 - ✓ Pulsar Developers
 - ✓ Pulsar Operations
- ✓ On-demand learning with labs
- ✓ 300+ engineers, admins and architects trained!



**Now Available
On-Demand
Pulsar Training**

Academy.StreamNative.io



Let's Keep in Touch!



Tim Spann
Developer Advocate



[PaaSDev](#)



<https://www.linkedin.com/in/timothyspann>



<https://github.com/tspannhw>