



**Stream  
Native**

# Using the FLiP(N) Stack for Edge AI (Flink, NiFi, Pulsar)

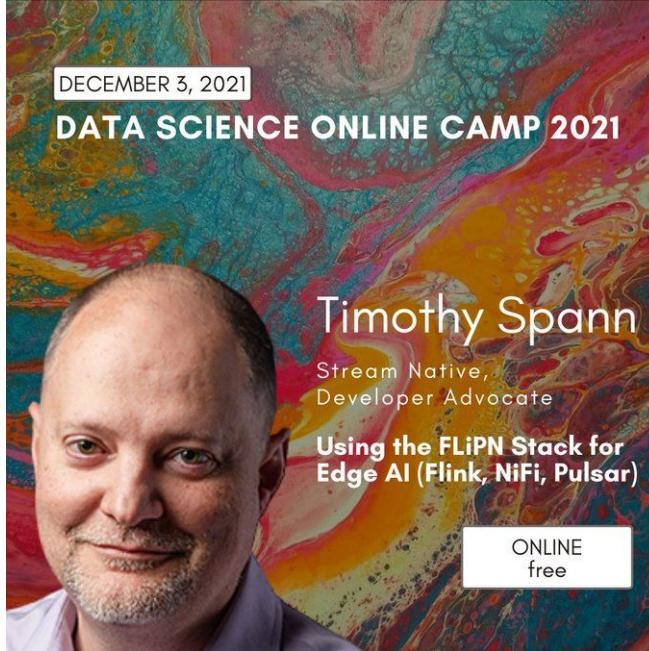
**Tim Spann** | Developer Advocate

FRIDAY, DECEMBER 3

**DATA SCIENCE ONLINE CAMP 2021**

<https://dscamp.org>

Start with vision, continue with passion, be driven by knowledge



<https://twitter.com/paasDev>



<https://github.com/tspannhw>



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



StreamNative

# AGENDA

- Apache NiFi
- Apache Flink
- Apache Pulsar
- Edge AI
- Cloud Native Made Easy

# FLiP(N) Stack

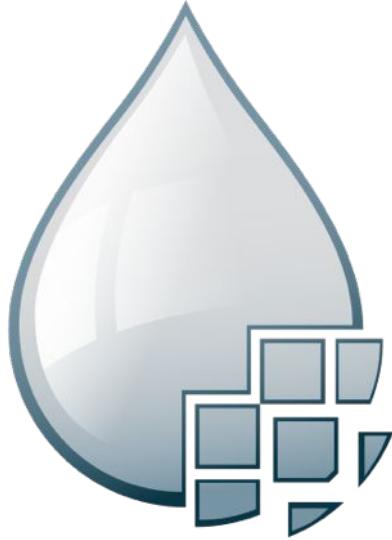
- Apache **Flink**
- Apache **Pulsar**
- Flink Connector for Pulsar
- NiFi Processor for Pulsar
- Apache **NiFi**
- Java, Python, Go



# Apache NiFi

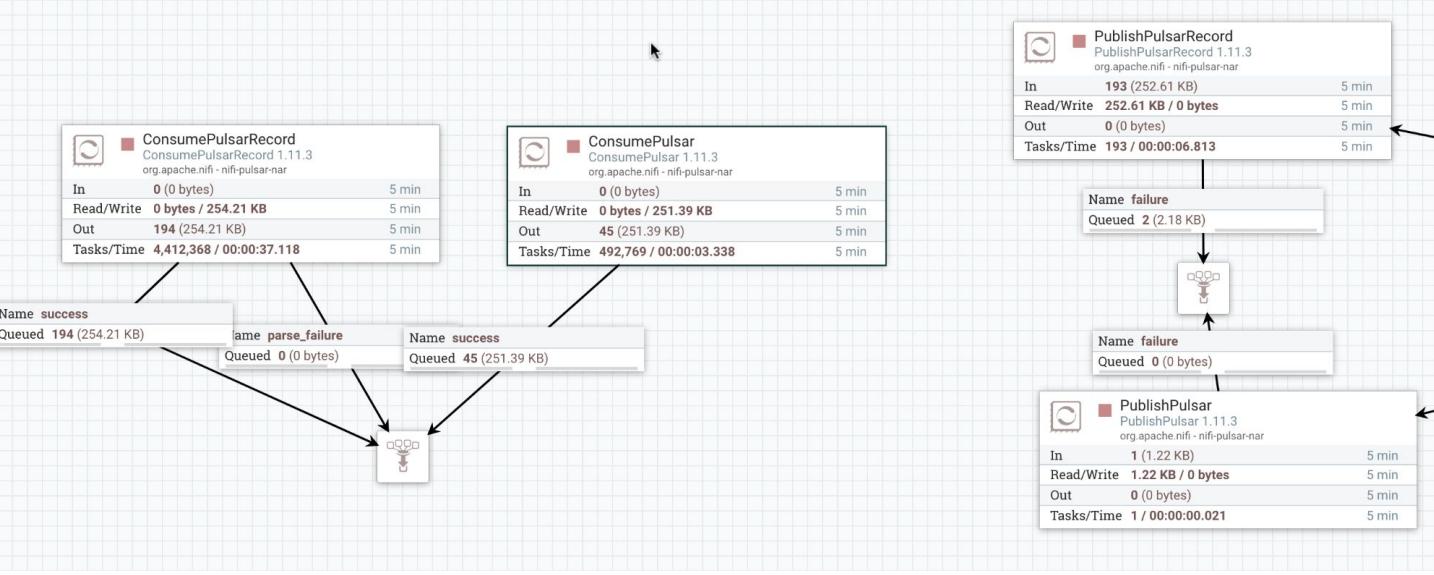


# Why Apache NiFi?



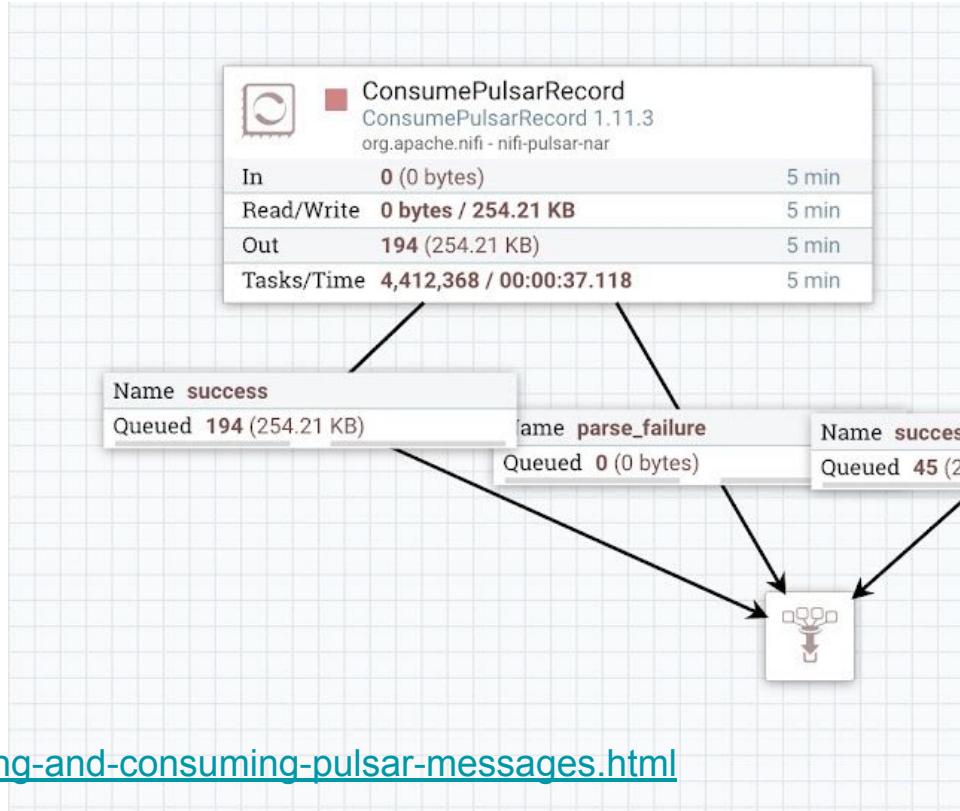
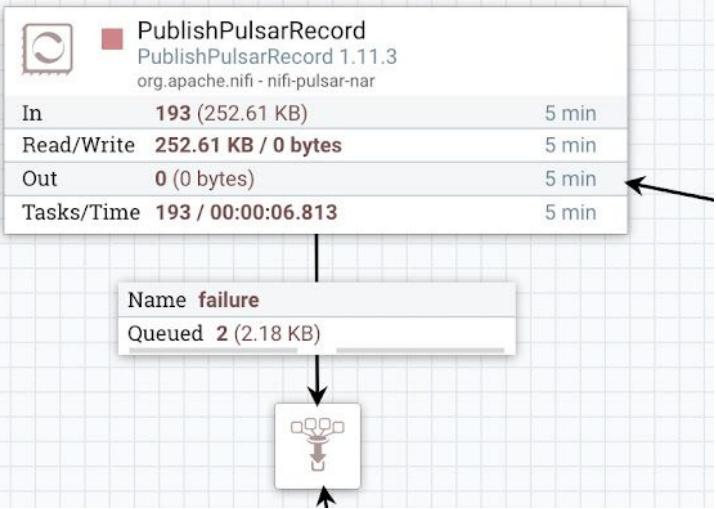
- Guaranteed delivery
- Data buffering
  - Backpressure
  - Pressure release
- Prioritized queuing
- Flow specific QoS
  - Latency vs. throughput
  - Loss tolerance
- Data provenance
- Supports push and pull models
- Hundreds of processors
- Visual command and control
- Over 300 sources
- JSON Flow templates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

# Apache NiFi Pulsar Connector



<https://github.com/david-streamlio/pulsar-nifi-bundle>

# Apache NiFi Pulsar Connector



<https://www.datainmotion.dev/2021/11/producing-and-consuming-pulsar-messages.html>

# Apache NiFi Pulsar Connector

Displaying 12 of 339

pulsar

Type	Version ▲	Tags
ConsumePulsar	1.11.0	PubSub, Consume, ingest, Get, I...
ConsumePulsarRecord	1.11.0	PubSub, Consume, Ingest, Get, ...
PublishPulsar	1.11.0	PubSub, Message, Pulsar, Apac...
PublishPulsarRecord	1.11.0	PubSub, 1.0, Message, csv, json...
ConsumePulsar	1.11.3	PubSub, Consume, Ingest, Get, I...
ConsumePulsarRecord	1.11.3	PubSub, Consume, Ingest, Get, ...
PublishPulsar	1.11.3	PubSub, Message, Pulsar, Apac...
PublishPulsarRecord	1.11.3	PubSub, 1.0, Message, csv, json...
ConsumePulsar	1.14.0	PubSub, Consume, Ingest, Get, I...
ConsumePulsarRecord	1.14.0	PubSub, Consume, Ingest, Get, ...
PublishPulsar	1.14.0	PubSub, Message, Pulsar, Apac...
PublishPulsarRecord	1.14.0	PubSub, 1.0, Message, csv, json...

**ConsumePulsar 1.11.3** org.apache.nifi - nifi-pulsar-nar

Consumes messages from Apache Pulsar. The complementary NiFi processor for sending messages is PublishPulsar.

# Apache NiFi Pulsar Connector

Controller Service Details

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Pulsar Service URL	pulsar+ssl://gke.sndev.snio.cloud:6651
Pulsar Client Authentication Service	PulsarClientOauthAuthenticationService14sn →
Maximum concurrent lookup-requests	5000
Maximum connects per Pulsar broker	1
I/O Threads	1
Keep Alive interval	30 sec
Listener Threads	1
Maximum lookup requests	50000 
Maximum rejected requests per connection	50
Operation Timeout	30 sec
Stats interval	60 sec
Allow TLS Insecure Connection	false
Enable TLS Hostname Verification	false
Use TCP no-delay flag	false

# Apache NiFi Pulsar Connector

## Controller Service Details

SETTINGS

PROPERTIES

COMMENTS

Required field

Property	Value
Audience	urn:sn:pulsar:sndev:gke
Issuer URL	<a href="https://auth.streamnative.cloud">https://auth.streamnative.cloud</a>
Private key file	file:///Users/tspann/Documents/servers/services/apache-pulsar-2.8.0/sndev-tspann.json
Trusted Certificate Filename	?

<https://github.com/david-streamlio/pulsar-nifi-bundle/releases/tag/v1.14.0>

# Apache NiFi Pulsar Connector to StreamNative Cloud

The screenshot shows the StreamNative Cloud interface. On the left, there's a sidebar with navigation links: gke, Tenants, Namespaces, Topics, SQL, Clients, Connector, Manage, Service Accounts, Flink Clusters, Pulsar Clusters, and Instance Setting. The main area has tabs for Message and Properties. A message from user tim.spann@streamnative.io is displayed, showing a JSON payload. The JSON contains weather data for KIAH (Houston Intercontinental Airport) on Oct 19, 2021, at 4:53 pm CDT. The data includes various meteorological parameters like temperature, humidity, wind direction, and pressure.

Message Properties

Updated on Oct 19 2021, 4:53 pm CDT "observation\_time\_rfc822": "Tue, 19 Oct 2021 16:53:00 -0500", "weather": "Mostly Cloudy", "temperature\_string": "80.0 F (26.7 C)", "temp\_f": 80.0, "temp\_c": 26.7, "relative\_humidity": 56, "wind\_string": "Southeast at 1.5 MPH (1 KT)", "wind\_dir": "Southeast", "wind\_degrees": 140, "wind\_mph": 11.5, "wind\_kt": 10, "pressure\_string": "1016.5 mb", "pressure\_mb": 1016.5, "pressure\_in": 30.02, "dewpoint\_string": "63.0 F (17.2 C)", "dewpoint\_f": 63.0, "dewpoint\_c": 17.2, "heat\_index\_string": "81 F (27 C)", "heat\_index\_f": 81, "heat\_index\_c": 27, "visibility\_mi": 10.0, "icon\_url\_base": "http://forecast.weather.gov/images/wtf/small/", "two\_day\_history\_url": "http://www.weather.gov/data/obhistory/KIAH.html", "icon\_url\_name": "bkn.png", "ob\_url": "http://www.weather.gov/data/METAR/KIAH.1.txt", "disclaimer\_url": "http://weather.gov/disclaimer.html", "copyright\_url": "http://weather.gov/disclaimer.html", "privacy\_policy\_url": "http://weather.gov/notice.html"}

Message Properties

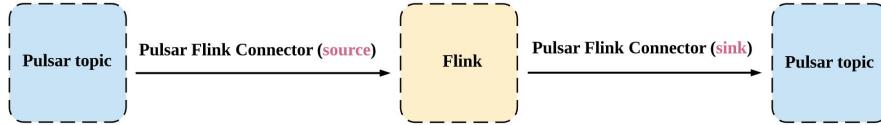
Updated on Oct 19 2021, 4:53 pm CDT "observation\_time\_rfc822": "Tue, 19 Oct 2021 16:53:00 -0500", "weather": "Mostly Cloudy", "temperature\_string": "80.0 F (26.7 C)", "temp\_f": 80.0, "temp\_c": 26.7, "relative\_humidity": 56, "wind\_string": "Southeast at 1.5 MPH (1 KT)", "wind\_dir": "Southeast", "wind\_degrees": 140, "wind\_mph": 11.5, "wind\_kt": 10, "pressure\_string": "1016.5 mb", "pressure\_mb": 1016.5, "pressure\_in": 30.02, "dewpoint\_string": "63.0 F (17.2 C)", "dewpoint\_f": 63.0, "dewpoint\_c": 17.2, "heat\_index\_string": "81 F (27 C)", "heat\_index\_f": 81, "heat\_index\_c": 27, "visibility\_mi": 10.0, "icon\_url\_base": "http://forecast.weather.gov/images/wtf/small/", "two\_day\_history\_url": "http://www.weather.gov/data/obhistory/KIAH.html", "icon\_url\_name": "bkn.png", "ob\_url": "http://www.weather.gov/data/METAR/KIAH.1.txt", "disclaimer\_url": "http://weather.gov/disclaimer.html", "copyright\_url": "http://weather.gov/disclaimer.html", "privacy\_policy\_url": "http://weather.gov/notice.html"}

# Apache Flink

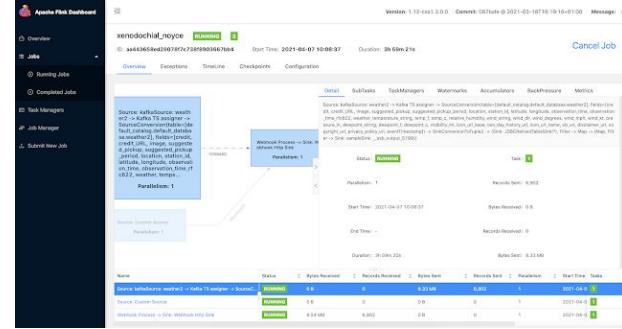


# Flink

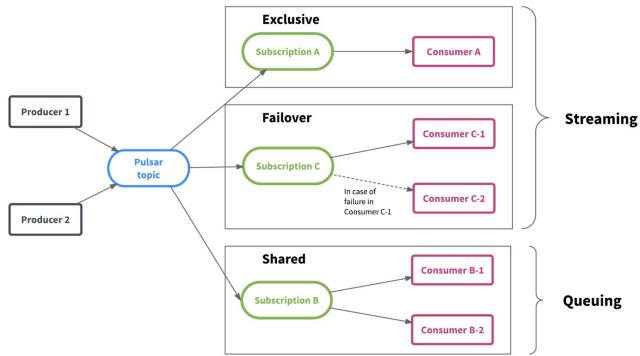
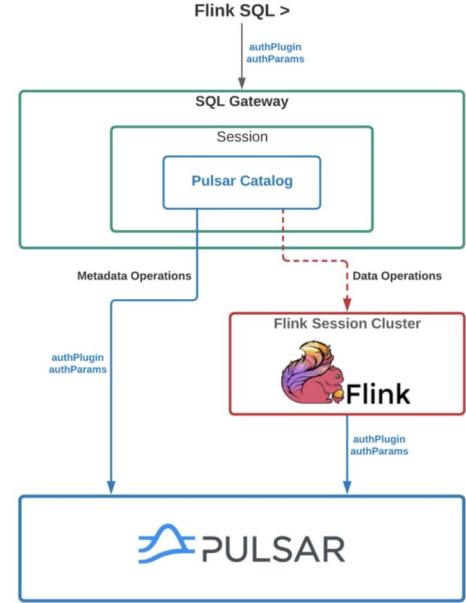
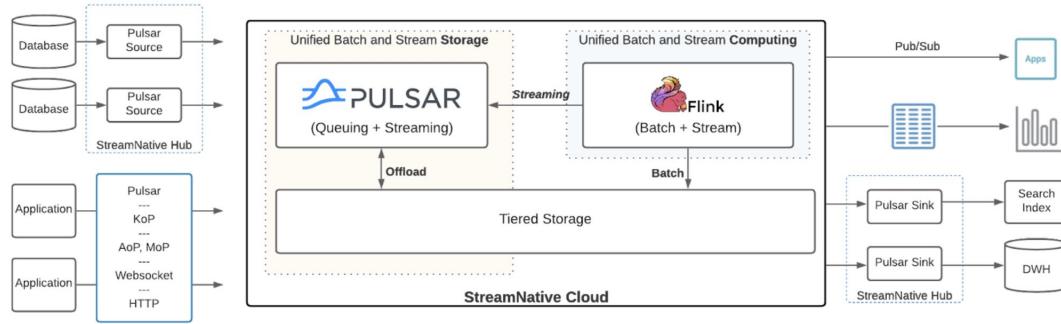
# Why Apache Flink?



- 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 + Pulsar (FLiP)



<https://flink.apache.org/2019/05/03/pulsar-flink.html>  
<https://github.com/streamnative/pulsar-flink>  
<https://streamnative.io/en/blog/release/2021-04-20-flink-sql-on-streamnative-cloud>

# apache/pulsar

Apache Pulsar - distributed pub-sub messaging system



# Apache Pulsar

465  
Contributors

1k  
Issues

10k  
Stars

2k  
Forks



<https://pulsar.apache.org/docs/en/reference-terminology/>

# What are the Benefits of Pulsar?



Multi-Tenancy

Scalability

Geo-Replication

Unified Messaging  
Model

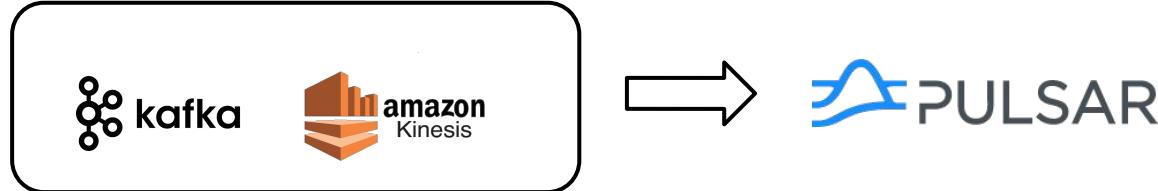
Data Durability

# Message Queuing & Data Streaming

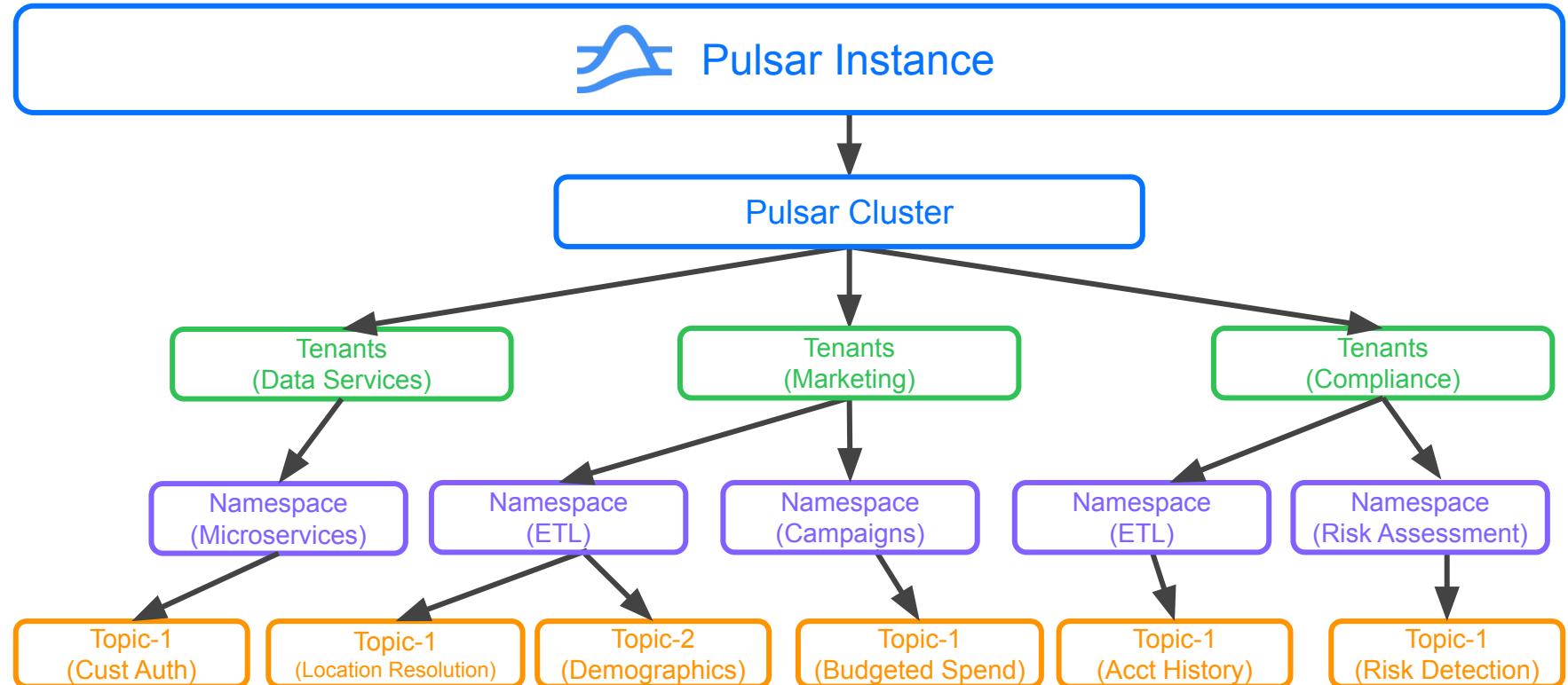
Message  
Queuing



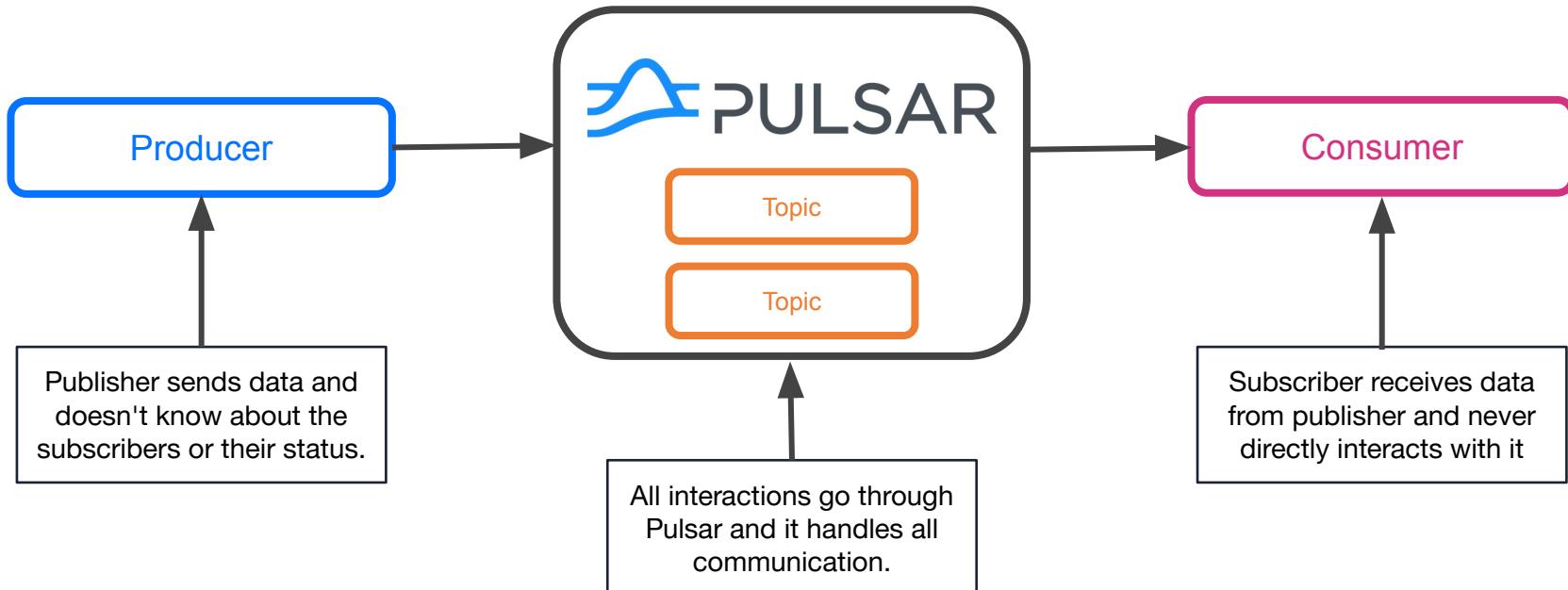
Data  
Streaming



# Topics

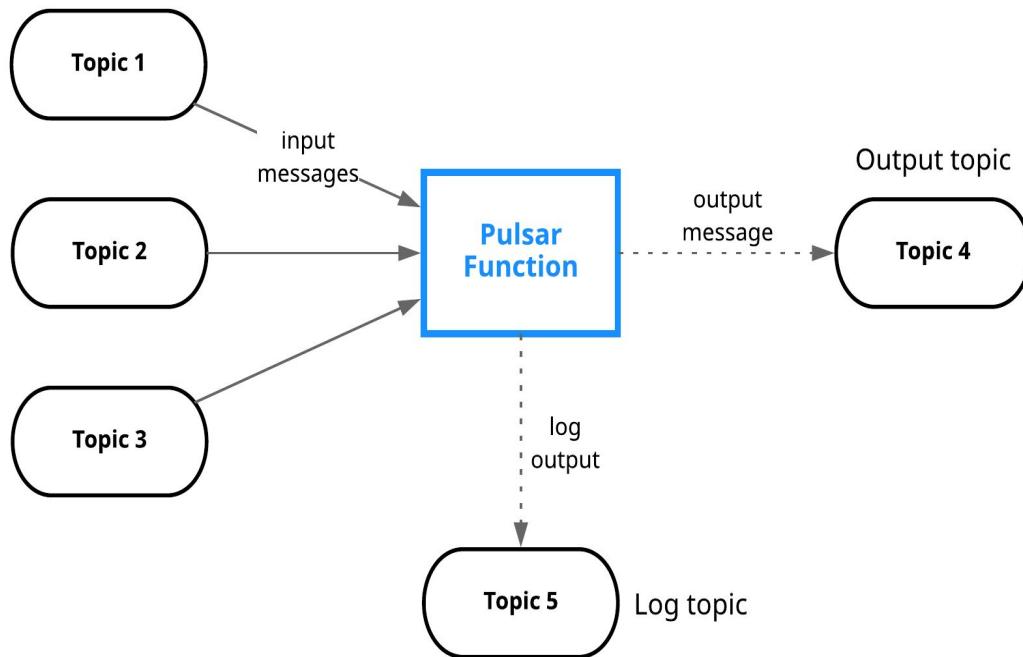


# Producer-Consumer



# Pulsar Functions

Input topics



- Consume messages from one or more Pulsar topics.
- Apply user-supplied processing logic to each message.
- Publish the results of the computation to another topic.
- Support multiple programming languages (Java, Python, Go)
- Can leverage 3rd-party libraries to support the ***execution of ML models on the edge.***

# Moving Data In and Out of Pulsar

IO/Connectors are a simple way to integrate with external systems and move data in and out of Pulsar. <https://pulsar.apache.org/docs/en/io-jdbc-sink/>

- Built on top of Pulsar Functions
- Built-in connectors - [hub.streamnative.io](http://hub.streamnative.io)



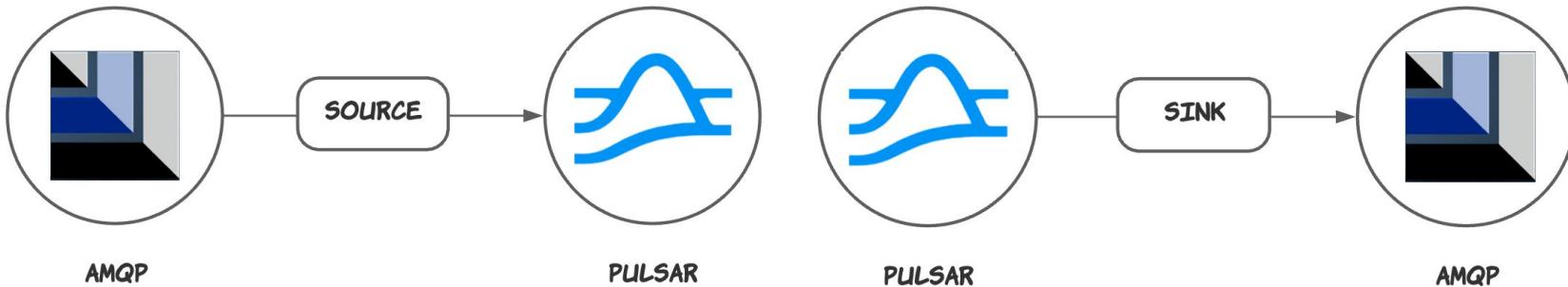
# AMQP / RabbitMQ Protocol

**AMQP on Pulsar (AoP)**

<https://github.com/streamnative/aop>

<https://hub.streamnative.io/connectors/amqp-1-0-sink/>

<https://hub.streamnative.io/connectors/amqp-1-0-source>



# StreamNative Hub

Discover Apache Pulsar Ecosystem

## FILTERS

[Submit a Plugin](#)

### Connectors

Source  
Sink

### Offloaders

### Protocol handlers

Data processing

### Logging

### Monitoring



#### ActiveMQ Sink

The ActiveMQ sink connector pulls messages from Pulsar topics and persist messages to ActiveMQ clusters.



StreamNative



#### ActiveMQ Source

The ActiveMQ source connector receives messages from ActiveMQ clusters and writes messages to Pulsar topics.



StreamNative



#### Aerospike Sink

The Aerospike sink connector pulls messages from Pulsar topics to Aerospike clusters



Apache Software Foundation



#### AMQP1\_0 sink

The AMQP1\_0 sink connector pulls messages from Pulsar topics and persists messages to AMQP brokers.



StreamNative



#### AMQP1\_0 source

The AMQP1\_0 source connector



#### Canal Source

The Canal source connector pulls



#### Cassandra Sink

The Cassandra sink connector

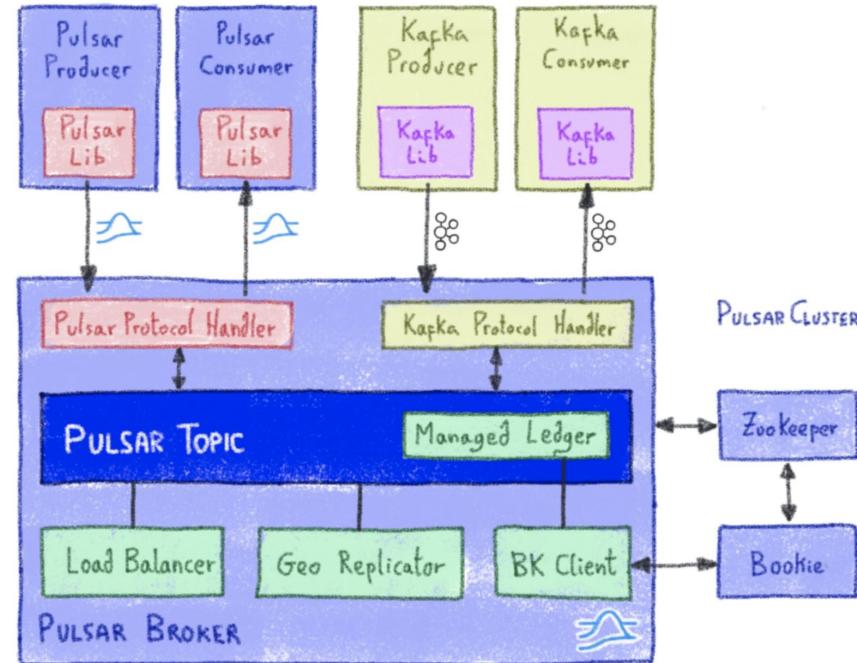


#### Cloud Storage Sink

The Cloud Storage sink

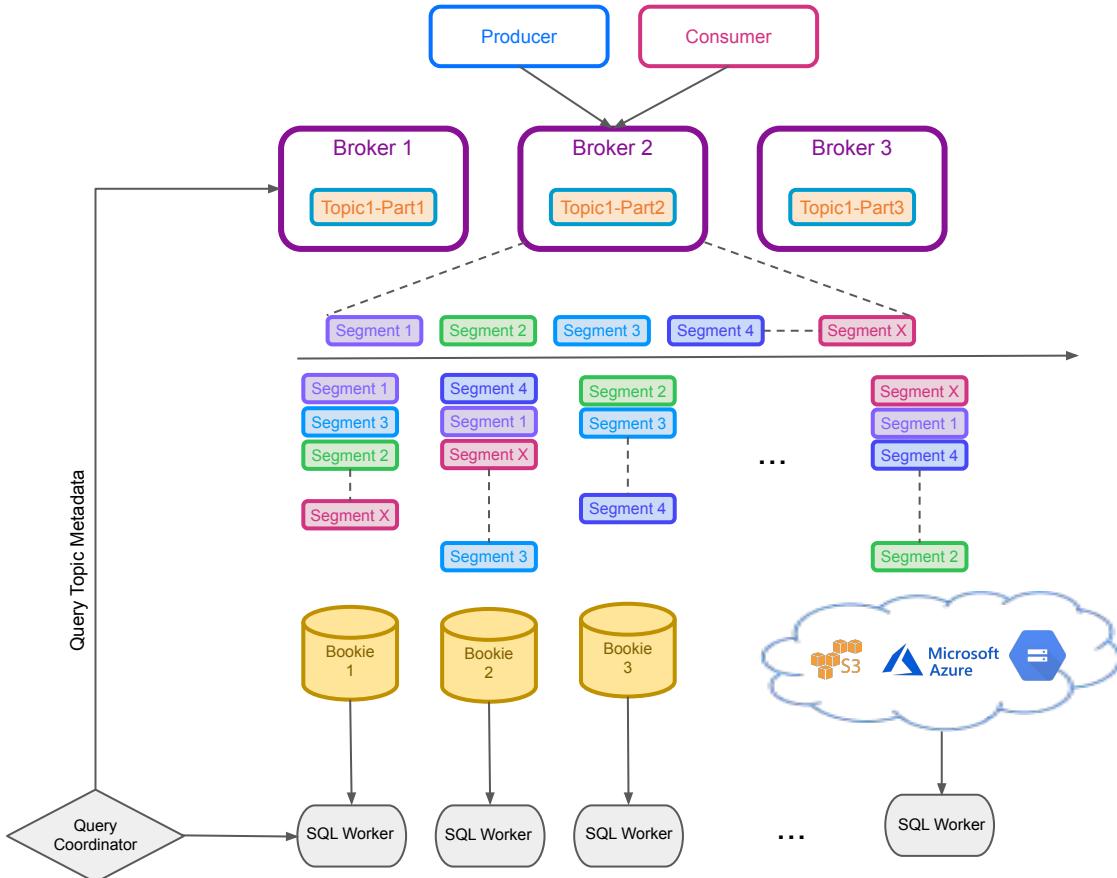
<https://hub.streamnative.io/>

# Kafka-on-Pulsar (Kop)



# Pulsar SQL

Presto/Trino workers can read segments directly from bookies (or offloaded storage) in parallel.



# Edge AI



StreamNative



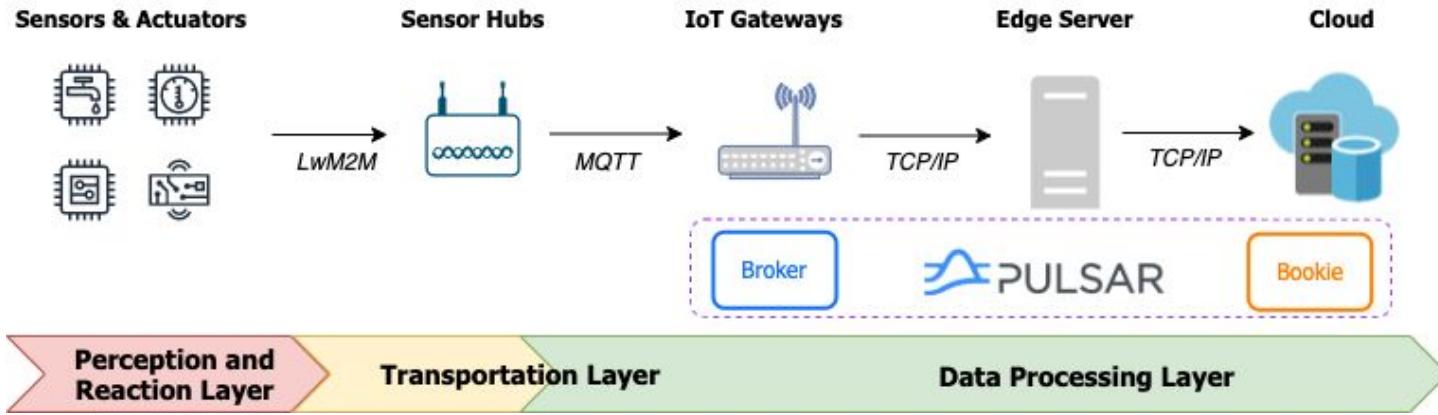
# Pulsar Edge SME



**David Kjerrumgaard**  
Developer Advocate

- Author of ***Pulsar in Action***
- Committer on the Apache Pulsar project.
- Over a decade of experience working with customers to solve Big Data / Streaming use cases

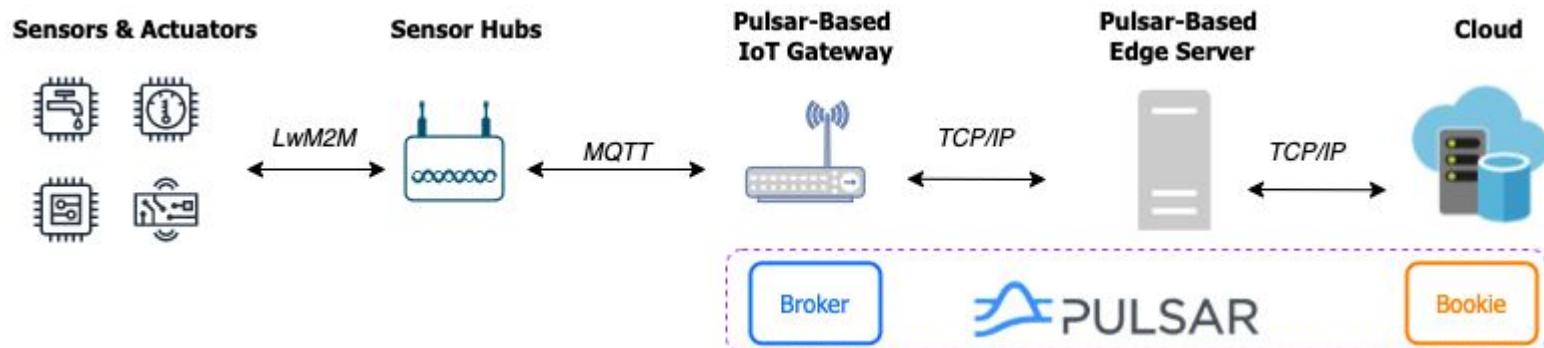
# Edge Computing with Pulsar



- Apache Pulsar's two-tier architecture separates the compute and storage layers, and interact with one another over a TCP/IP connection. This allows us to run the computing layer (Broker) on either Edge servers or IoT Gateway devices.
- Pulsar's serverless computing framework, known as Pulsar Functions, can run inside the Broker as threads. Effectively "stretching" the data processing layer.

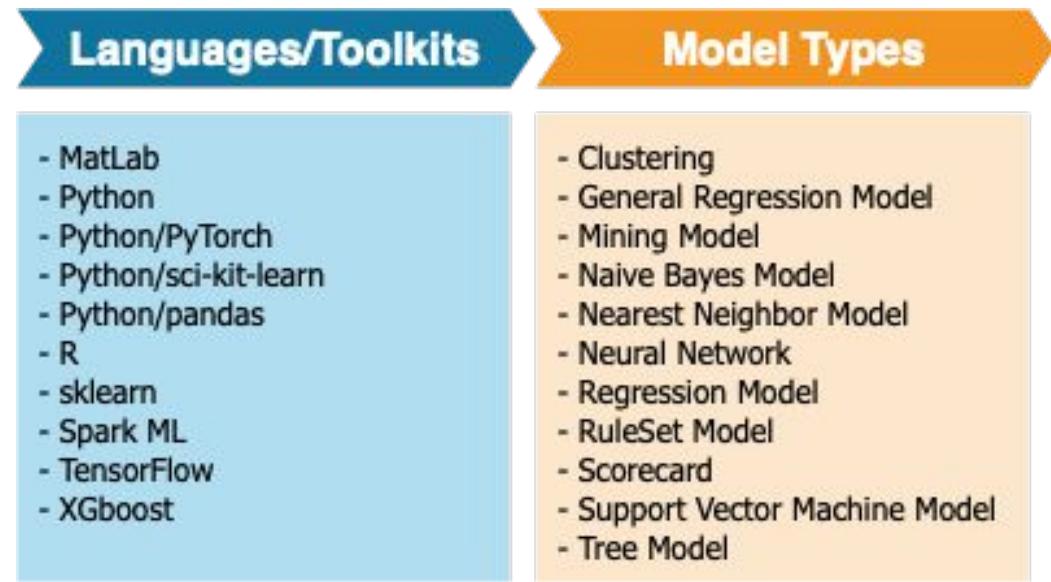
# Benefits of Running Pulsar Broker on the Edge

- Pulsar's Serverless computing framework can run inside the Pulsar Broker as a thread pool. This framework can be used as the execution environment for ML models.
- The Apache Pulsar Broker supports the MQTT protocol and therefore can directly receive incoming data from the sensor hubs and store it in a topic.



# Pulsar Function – Third Party Library Support

- You can leverage 3<sup>rd</sup> 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



Apache MXNet Native Processor through DJL.AI for Apache NiFi



#workshop

☆ | ▲ 2 | ❁ Add a topic

11:30 AM =====

Deep Learning Class Label: person

File: ccoa469f-c108-42c7-95c6-10e5fda95006.person.png

Probability: 0.96

UUID: 32ef65a3-0650-42cd-965c-ba25597eb1ad

Rank: 1

Bounding Box (Height/Width, X,Y)

0.74 / 0.69

0.27, 0.25

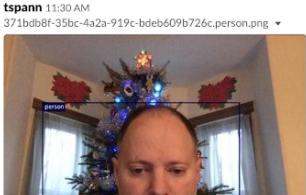
Image (Height/Width, X,Y)

480 / 640

0, 0

=====

tspann 11:30 AM  
371bdb8f-35bc-4a2a-919c-bdeb609b726c.person.png ▾



## Attribute Values

## boundingbox\_height\_1

0.99

No value set

### boundingbox\_width\_1

0.90

No value set

boundingbox\_x\_1

0.09

No value set

boundingbox\_y\_1

0.01

### class\_1

No value set

2020-08-

2020-08-26\_1330.jpg (previous)

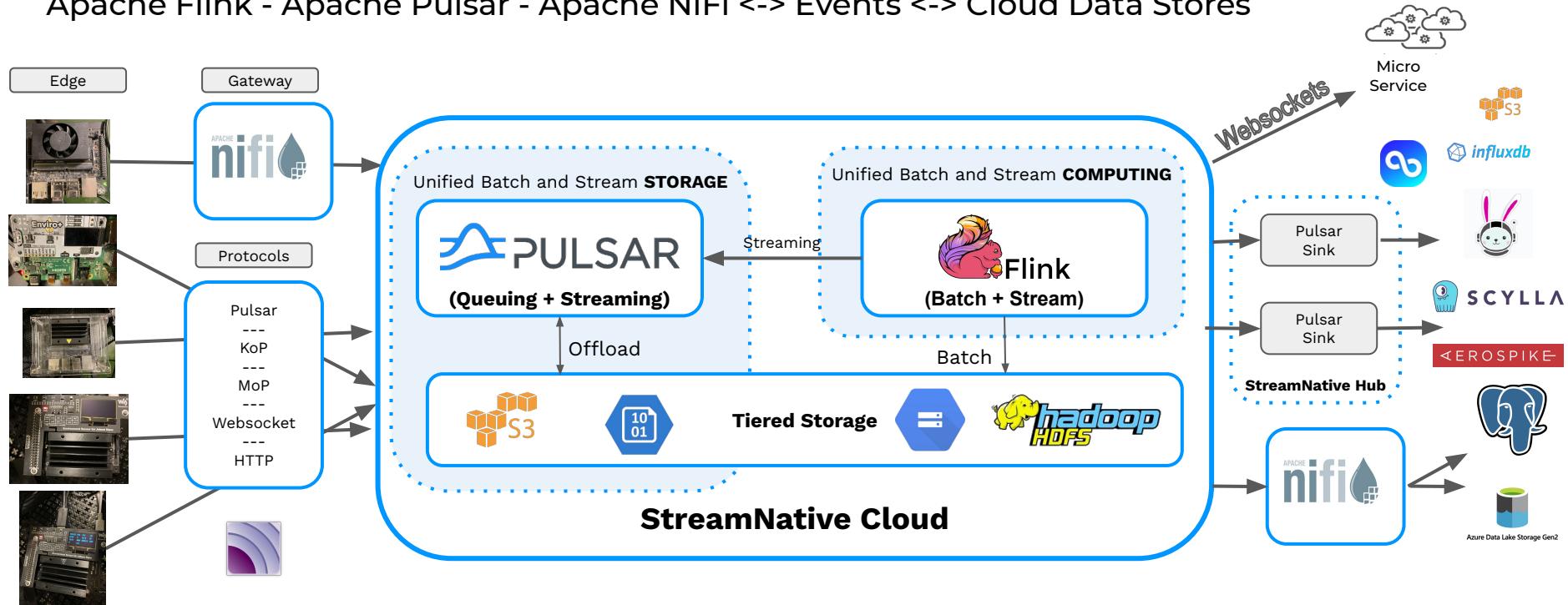
This processor uses the DJL.AI Java Interface

<https://github.com/tspannhw/nifi-djl-processor>

<https://dev.to/tspannhw/easy-deep-learning-in-apache-nifi-with-dl-2d79>

# End-to-End Streaming FLiP(N) Apps

Apache Flink - Apache Pulsar - Apache NiFi <-> Events <-> Cloud Data Stores



# Using NVIDIA Jetson Devices With Pulsar

## DEMO TIME

<https://github.com/tspannhw/minifi-xaviernx/>

<https://github.com/tspannhw/minifi-jetson-nano>

<https://github.com/tspannhw/FLiP-Jetson>

<https://github.com/tspannhw/Flip-iot>

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

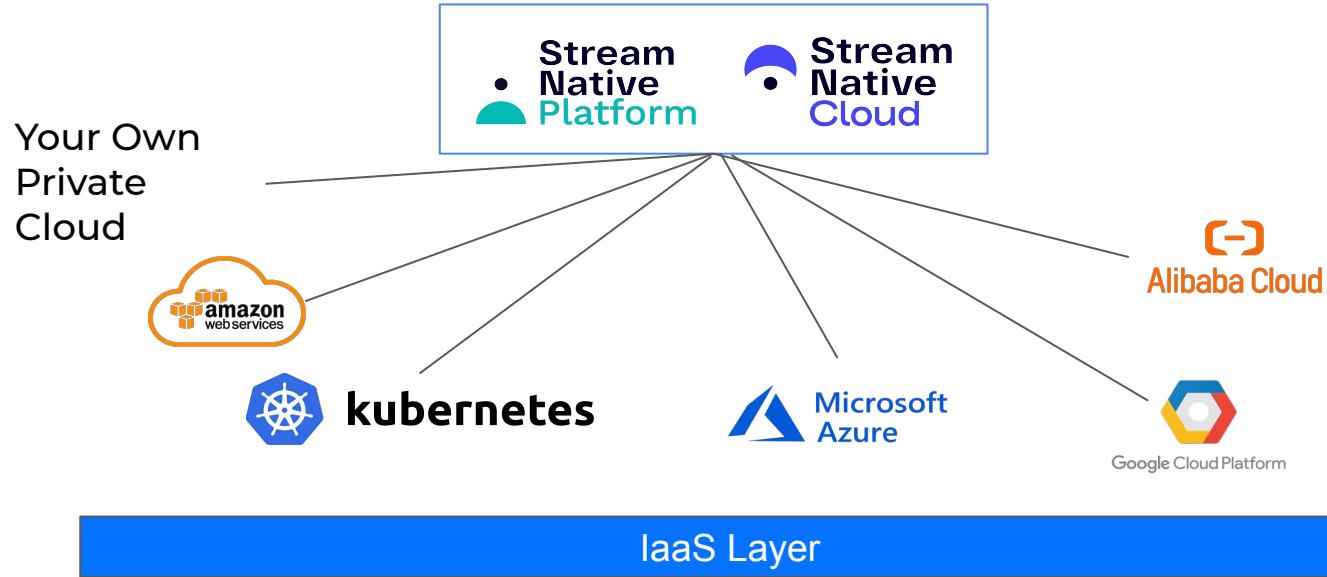
<https://www.datainmotion.dev/2020/10/flank-streaming-edgeai-on-new-nvidia.html>

StreamNative  
Cloud



**Stream  
Native  
Cloud**

# The StreamNative Offering



gke

Tenants

Namespaces

Topics

SQL preview

Clients

Connector ▾

Manage ▾

Service Accounts

Flink Clusters

Pulsar Clusters

Instance Setting

## Storage Size



## Entries



## Segments



## Segments

Ledger ID	Entries	Size	Status	Offload
3275	395	556.94K	opening	false

## Cursors

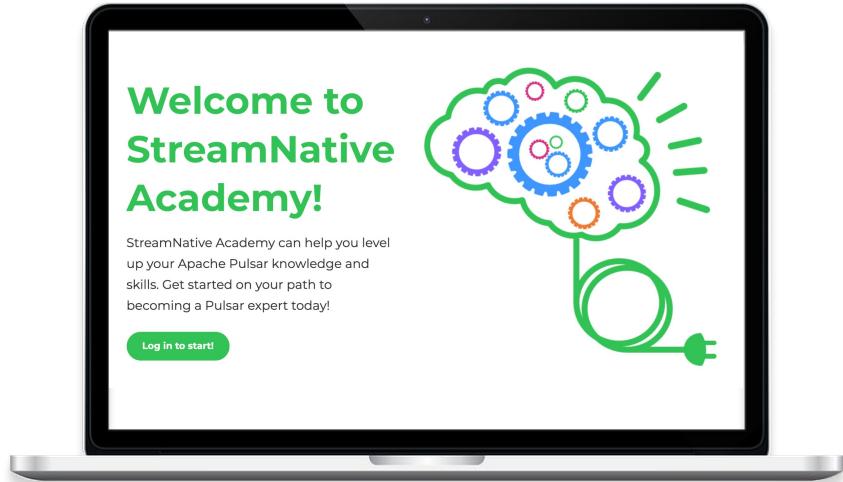
Cursor	Mark Delete Position	Read Position	Waiting Read Op	Pending Read Op	Entries Since First Not AckedMessage
weather-nifi	3275:394	3275:395	false	0	1
weather-nifi-file	3275:394	3275:395	false	0 	1
weatherman	3275:-1	3275:0	false	0	1



# Learn More

# Now Available On-Demand Pulsar Training

[Academy.StreamNative.io](https://Academy.StreamNative.io)



Platform Engineer [Remote]

San Francisco

Platform Engineer (Flink/Spark) [Remote]

San Francisco

Product Engineer - Cloud [Remote]

San Francisco

Platform Engineer (Flink/Spark) [Remote]

San Francisco

Product Engineer - Cloud [Remote]

San Francisco

Sr. Product Manager [Remote]

San Francisco

# We're Hiring

[streamnative.io/careers/](https://streamnative.io/careers/)



StreamNative

# Connect with the Community & Stay Up-To-Date

- Join the Pulsar Slack channel - [Apache-Pulsar.slack.com](https://Apache-Pulsar.slack.com)
- Follow [@streamnativeio](https://twitter.com/streamnativeio) and [@apache pulsar](https://twitter.com/apache_pulsar) on Twitter
- [Subscribe](#) to Monthly Pulsar Newsletter for major news, events, project updates, and resources in the Pulsar community

# Interested In Learning More?



## Resources

[Flink SQL Cookbook](#)

[The Github Source for Flink SQL Demo](#)

[The GitHub Source for Demo](#)



## Free eBooks

[Manning's Apache Pulsar in Action](#)

[O'Reilly Book](#)



## Upcoming Events

[Jan/2022] Data Minutes

[Jan/2022] devfest 2029

[Feb/2022] ScyllaDB Summit

# Community Resources



Webinars



e-Books



Tutorials

# Upcoming Events



# Pulsar Summit Asia 2021

**January 15-16, 2022 | Hosted By**  Stream Native

Join us for this two-day hybrid event. On the first day, we will be hosting talks live in Beijing and also broadcasting online. On the second day, we will have one whole day virtual event.

[Register Now](#)

<https://pulsar-summit.org/>

# Upcoming Events



<https://datagrillen.com/dataminutes/>  
21-January-2022 Virtual

## SCYLLA SUMMIT 2022

<https://www.scylladb.com/scylla-summit-2022/>  
09-February-2022 Virtual



## SCYLLA

The slide features a blue header with 'devfest 2021' and 'devfest-uki.com'. Below the header is a white box containing the title 'Using Apache NiFi with Apache Pulsar for Fast Data On-Ramp' and the speaker's name 'Timothy Spann, Developer Advocate @ StreamNative'. To the right is a portrait photo of Timothy Spann. At the bottom, there is binary code '011010 010001 001100' and a small logo for 'Google Developer Groups UK & Ireland'.

<http://devfest-uki.com>  
29-January-2022 Hybrid



# FLIP Stack Weekly

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

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

# Building Edge Applications with Apache Pulsar

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

# Let's Keep in Touch!



**Timothy Spann**  
Developer Advocate



@PassDev



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



<https://github.com/tspannhw>



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