



**Stream  
Native**

Real-Time Streaming in Any and  
All Clouds, Hybrid and Beyond

**Timothy Spann | Developer Advocate**



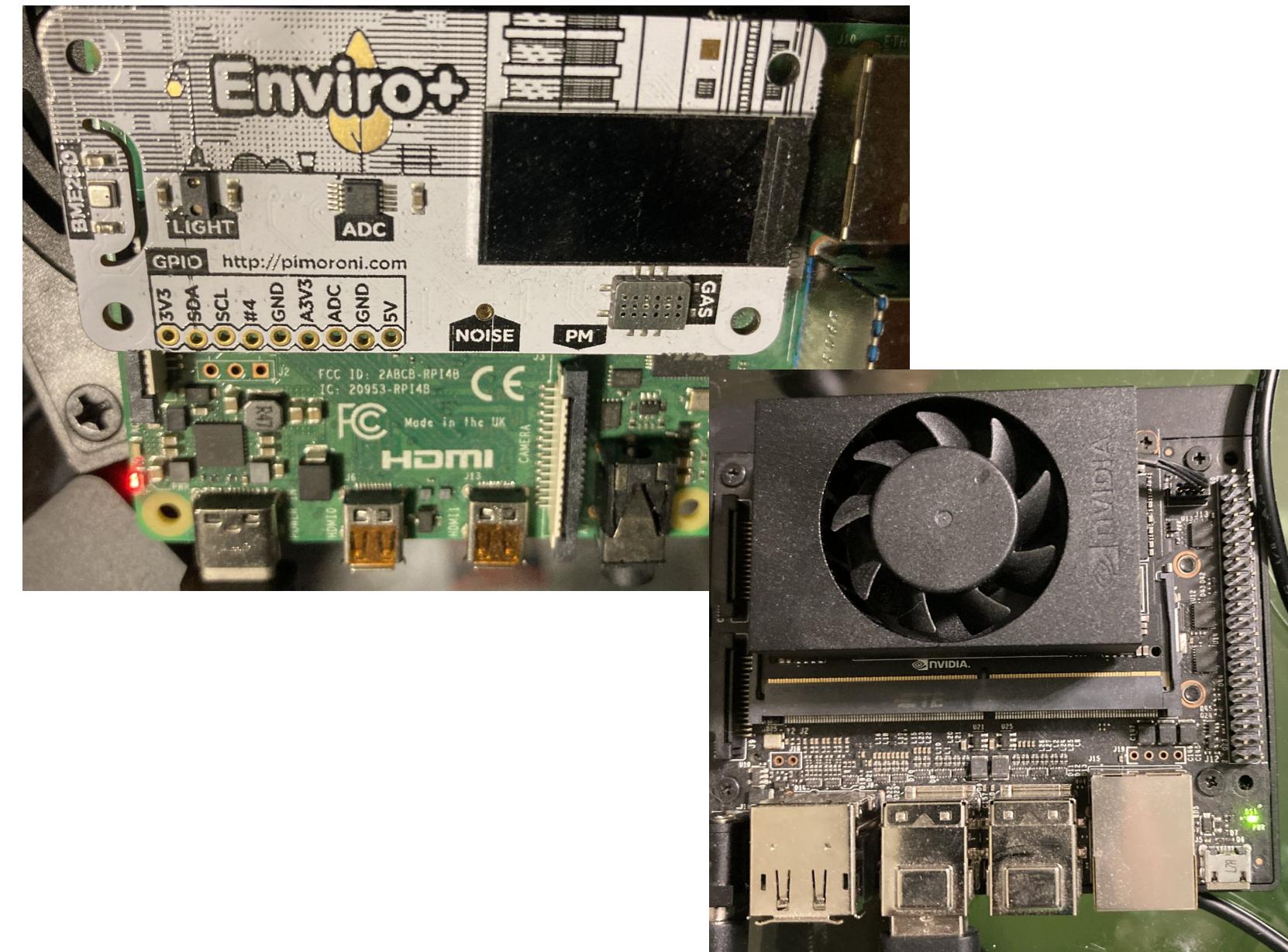
## **Tim Spann, Developer Advocate**

DZone Zone Leader and Big Data MVB Data DJay

# USE CASE

**IoT Ingestion:** High-volume streaming sources, sensors, multiple message formats, diverse protocols and multi-vendor devices creates data ingestion challenges.

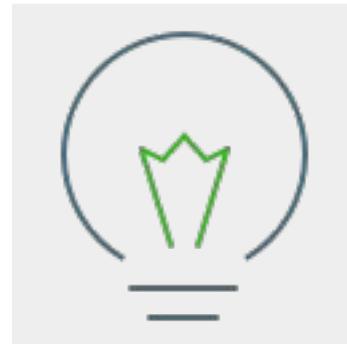
**Other Sources:** Transit data, news, twitter, status feeds, REST data, stock data and more.



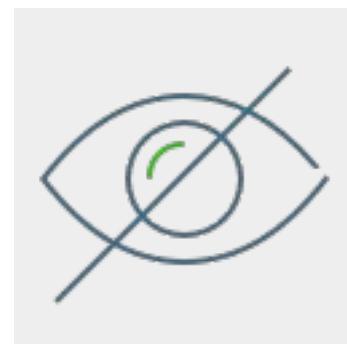
# KEY CHALLENGES



**Data Ingestion:** High-volume streaming sources, multiple message formats, diverse protocols and multi-vendor devices creates data ingestion challenges.



**Real-time Insights:** Analyzing continuous and rapid inflow (velocity) of streaming data at high volumes creates major challenges for gaining real-time insights.



**Visibility:** Lack visibility of end-to-end streaming data flows, inability to troubleshoot bottlenecks, consumption patterns etc.

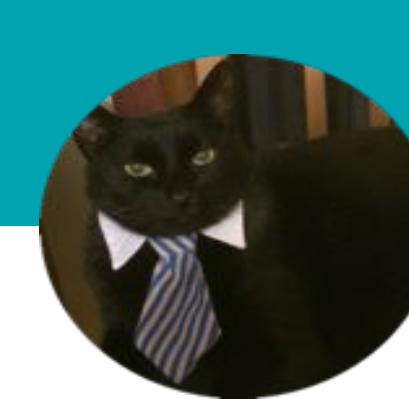
# FLiP(N) Stack for Data Engineers

Multiple users, protocols, frameworks, languages, clouds, data sources & clusters



CLOUD DATA ENGINEER

- Experience in ETL/ELT
- Coding skills in Python or Java
- Knowledge of database query languages such as SQL
- Experience with Streaming
- Knowledge of Cloud Tools



CAT

- Expert in ETL (Eating, Ties and Laziness)
- Edge Camera Interaction
- Typical User
- No Coding Skills
- Can use NiFi
- Questions your cloud spend



AI / Deep Learning / ML / DS

- Can run in Apache NiFi
- Can run in Apache Pulsar Functions
- Can run in Apache Flink
- Can run in Apache Flink SQL
- Can run in Apache Pulsar Clients
- Can run in Apache Pulsar Microservices
- Can run in Function Mesh



Function Mesh

<https://functionmesh.io/>

# StreamNative Solution

APP Layer

Application Messaging

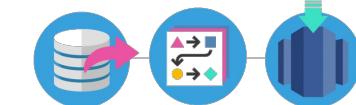


Micro Service

Payment

Notification

Data Pipelines



ETL

Real-time Contextual Analytics



Dashboard

Risk Control

Auditing

StreamNative Platform

Computing Layer



Storage Layer



Tiered Storage



S3



10  
01



IaaS Layer



# kubernetes



Google Cloud Platform



StreamNative

Stream Native Platform

Stream Native Cloud

# FLiP(N) Stack

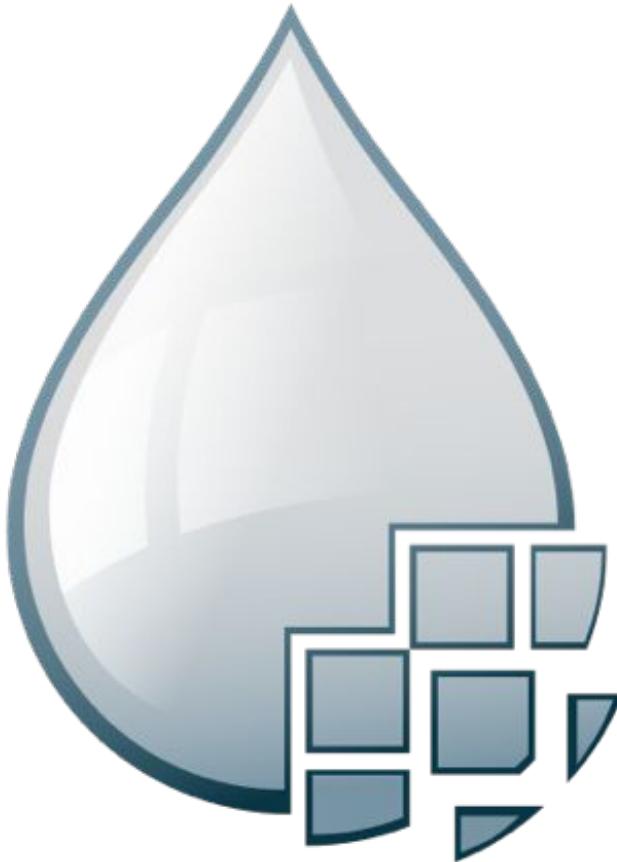
- Apache **Flink**
- Apache **Pulsar**
- StreamNative's Flink Connector for Pulsar
- Apache **NiFi**
- Apache +++



## What is Apache NiFi?

**Apache NiFi** is a scalable, real-time streaming data platform that collects, curates, and analyzes data so customers gain key insights for immediate actionable intelligence.

# 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 sixty sources
- Flow templates
- Pluggable/multi-role security
- Designed for extension
- Clustering
- Version Control

# APACHE NIFI HIGH LEVEL CAPABILITIES

- Scale horizontal and vertically
  - Scale your data flow to millions event/s
  - Ingest TB to PB of data per day
- Adapt to your flow requirements
  - Back pressure & Dynamic prioritization
  - Loss tolerant vs guaranteed delivery
  - Low latency vs high throughput
- Secure
  - SSL, HTTPS, SFTP, etc.
  - Governance and data provenance
- Extensible
  - Build your own processors and Controller services (providers)
  - Integrate with external systems (Security, Monitoring, Governance, etc)

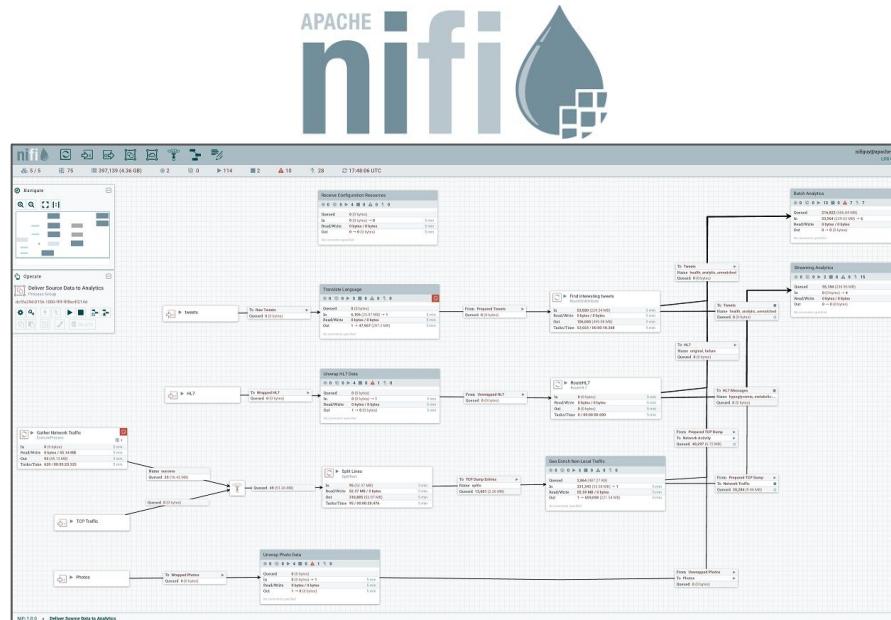
Add Controller Service

Source	Type	Version	Tags
all groups	PulsarClientAthenzAuthenticationService	1.11.0	security, Athenz, client, Pulsar, a...
cache client	PulsarClientJwtAuthenticationService	1.11.0	security, JWT, client, Pulsar, aut...
connection	PulsarClientTlsAuthenticationService	1.11.0	security, client, Pulsar, TLS, aut...
credentials database	StandardPulsarClientService	1.11.0	pool, client, Pulsar
distributed enrich			
invoke javascript js			
jython key lookup			
luaj map parse			
python reader			
record recordset			
restricted row script			
value writer			

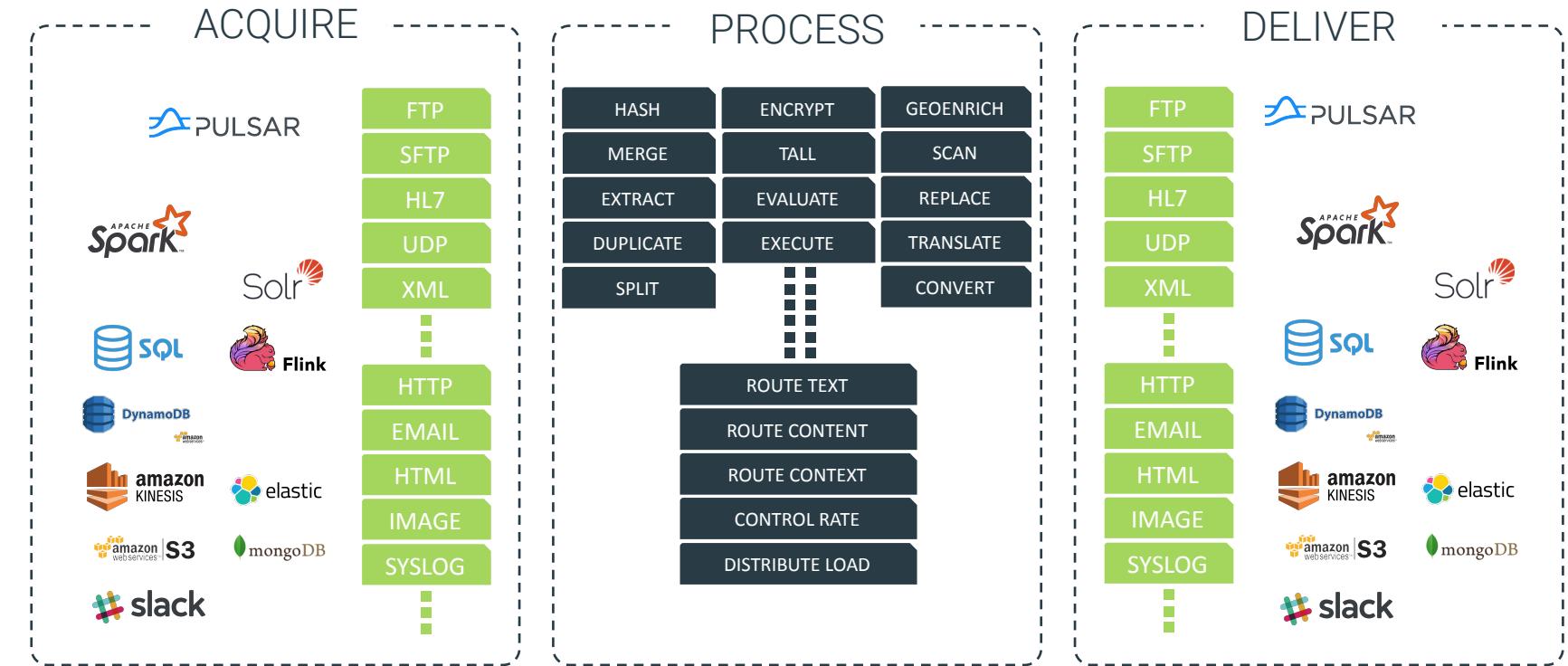
PulsarClientAthenzAuthenticationService 1.11.0 org.apache.nifi - nifi-pulsar-nar  
Implementation with Athenz Authentication of the PulsarClientAuthenticationService. Provides Pulsar clients with the ability to authenticate against a secured Apache Pulsar broker endpoint.

# Apache NiFi

Enable easy ingestion, routing, management and delivery of any data anywhere (Edge, cloud, data center) to any downstream system with built in end-to-end security and provenance.



Advanced tooling to industrialize flow development  
(Flow Development Life Cycle)



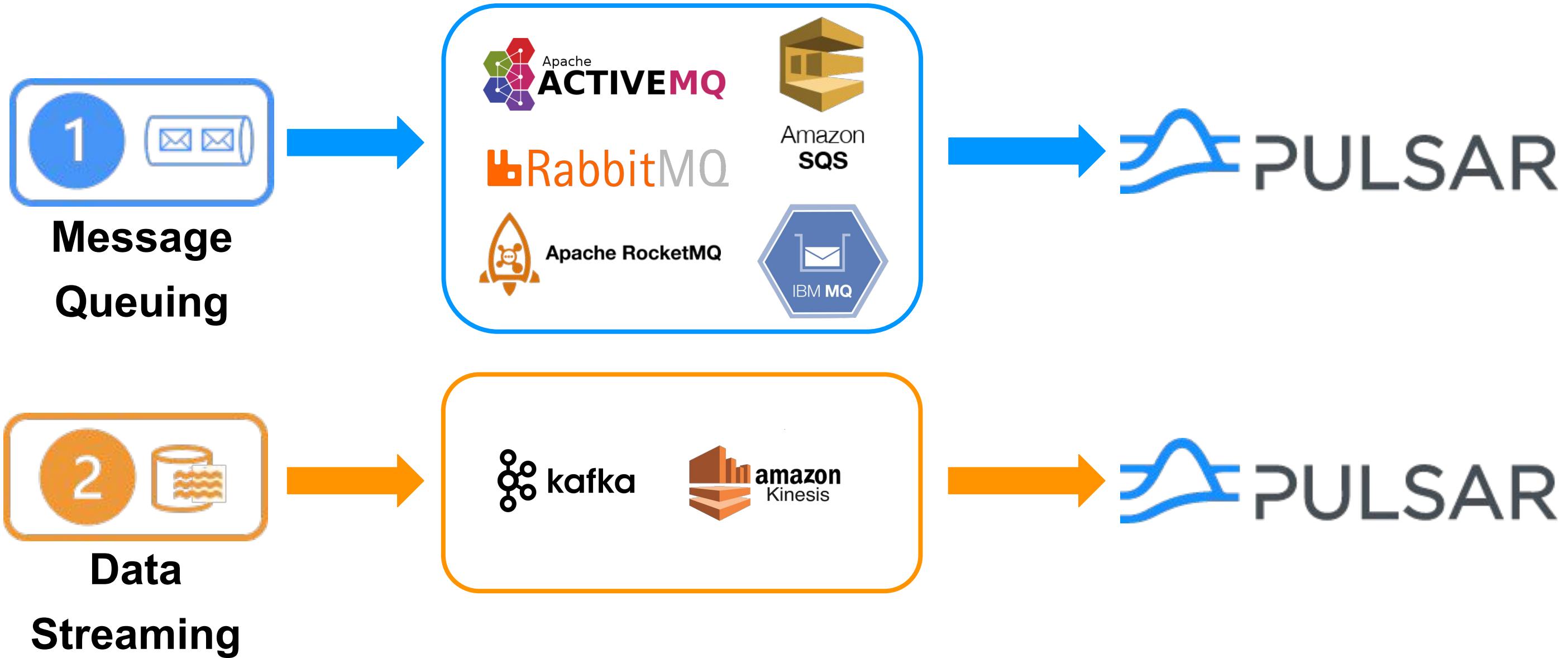
- Over 300 Prebuilt Processors
- Easy to build your own
- Parse, Enrich & Apply Schema
- Filter, Split, Merger & Route
- Throttle & Backpressure
- Guaranteed Delivery
- Full data provenance from acquisition to delivery
- Diverse, Non-Traditional Sources
- Eco-system integration

## What is Apache Pulsar?

**Apache Pulsar** is an open source, cloud-native distributed messaging and streaming platform.

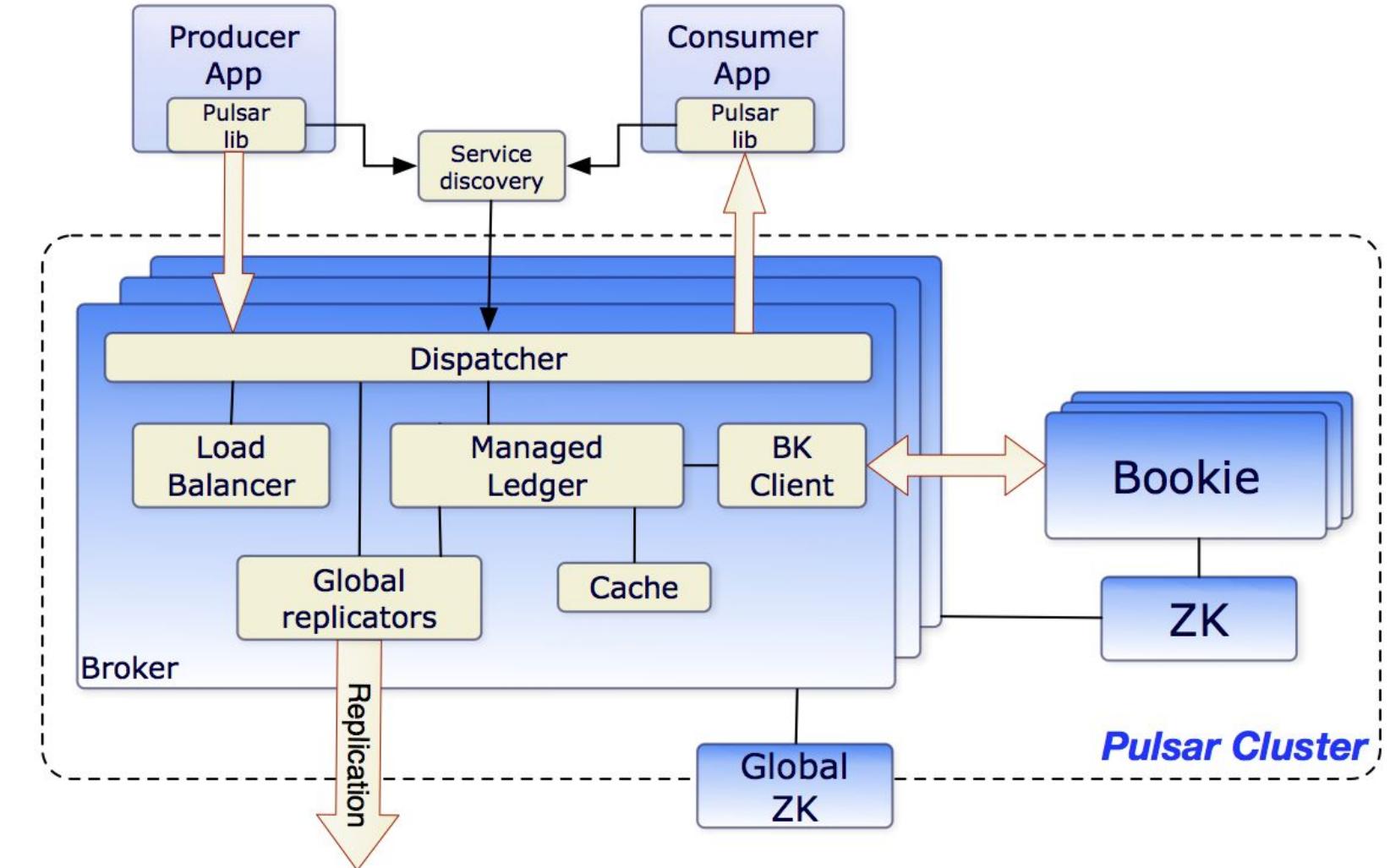


# A Unified Messaging Platform

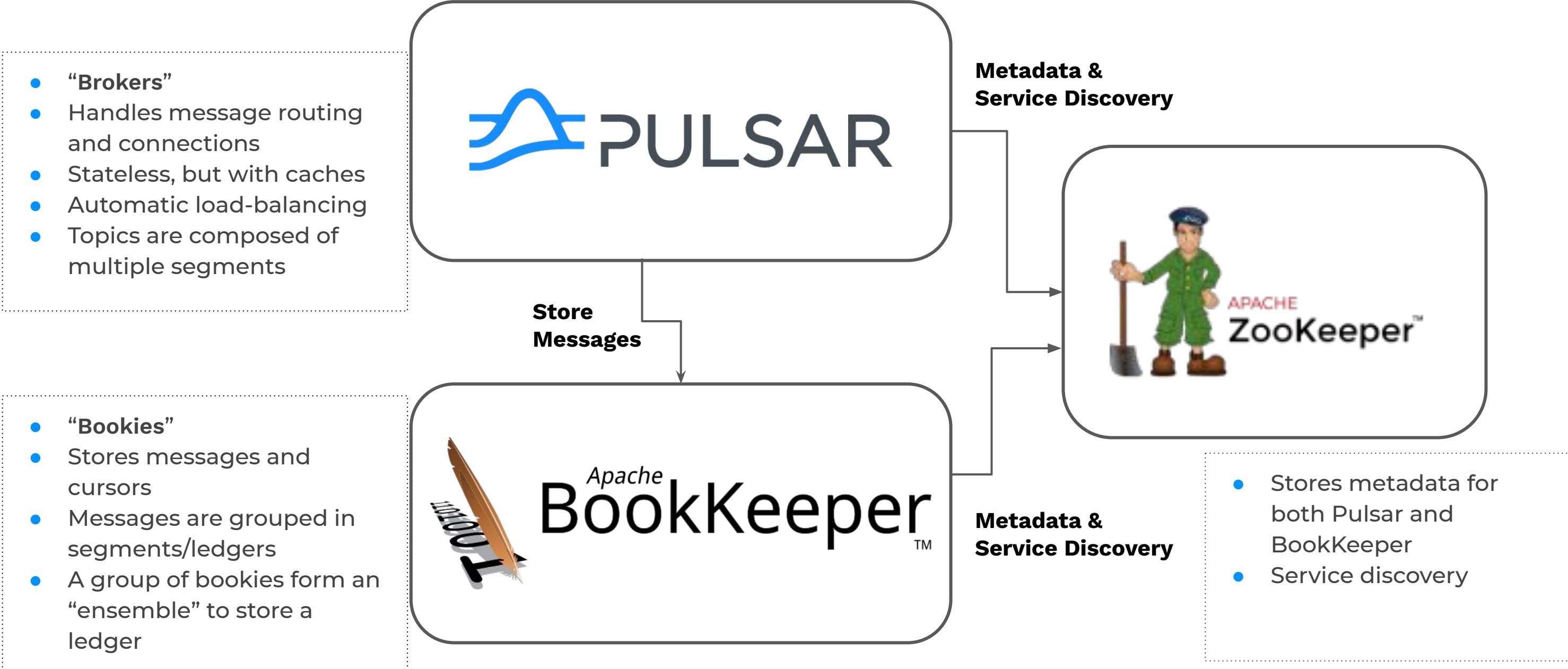


# Apache Pulsar

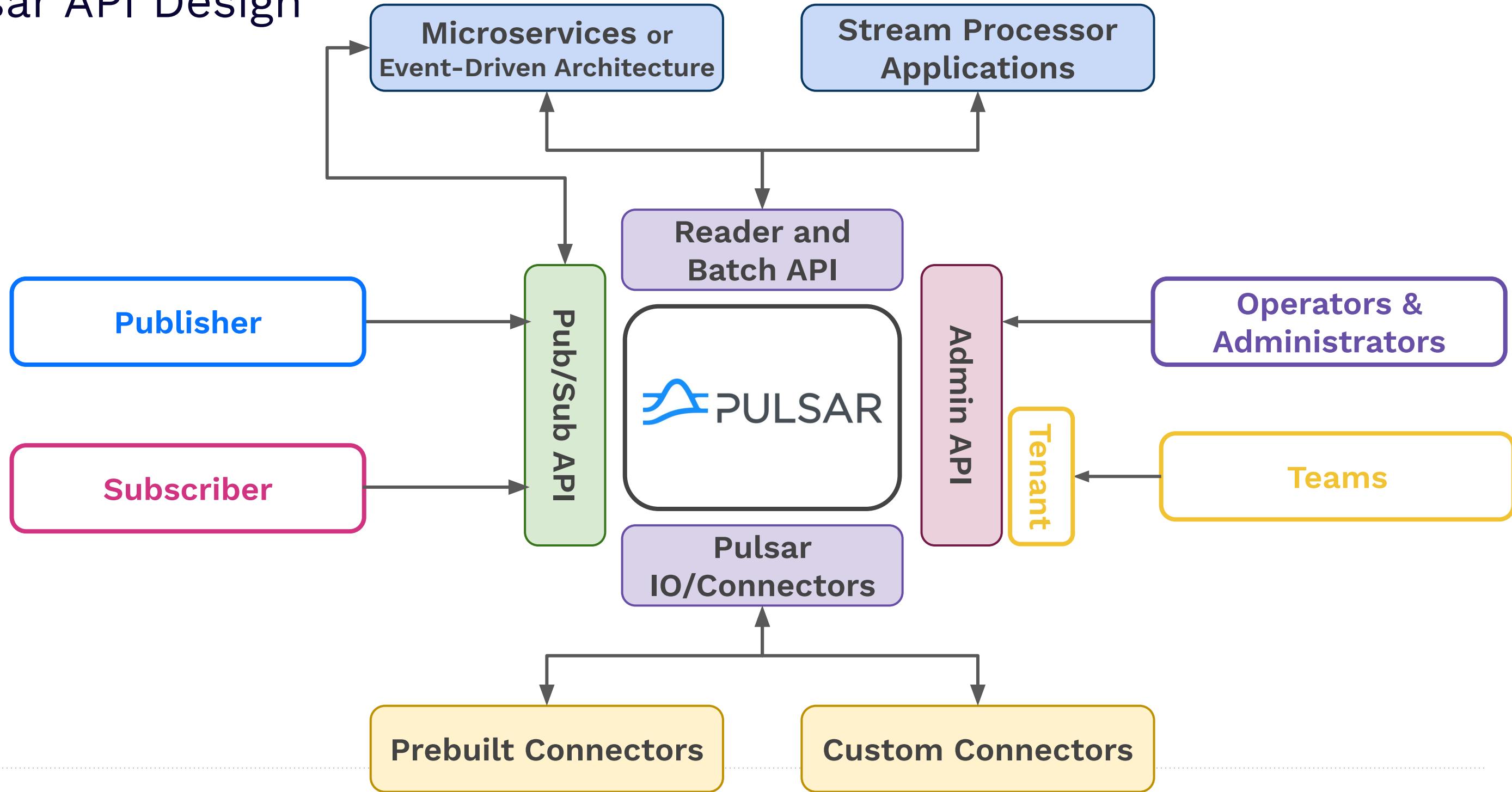
- Pub-Sub
- Geo-Replication
- Pulsar Functions
- Horizontal Scalability
- Multi-tenancy
- Tiered Persistent Storage
- Pulsar Connectors
- REST API
- CLI
- Many clients available
- Four Different Subscription Types
- Multi-Protocol Support
  - **MQTT**
  - AMQP
  - JMS
  - **Kafka**
  - ...



# Pulsar Cluster



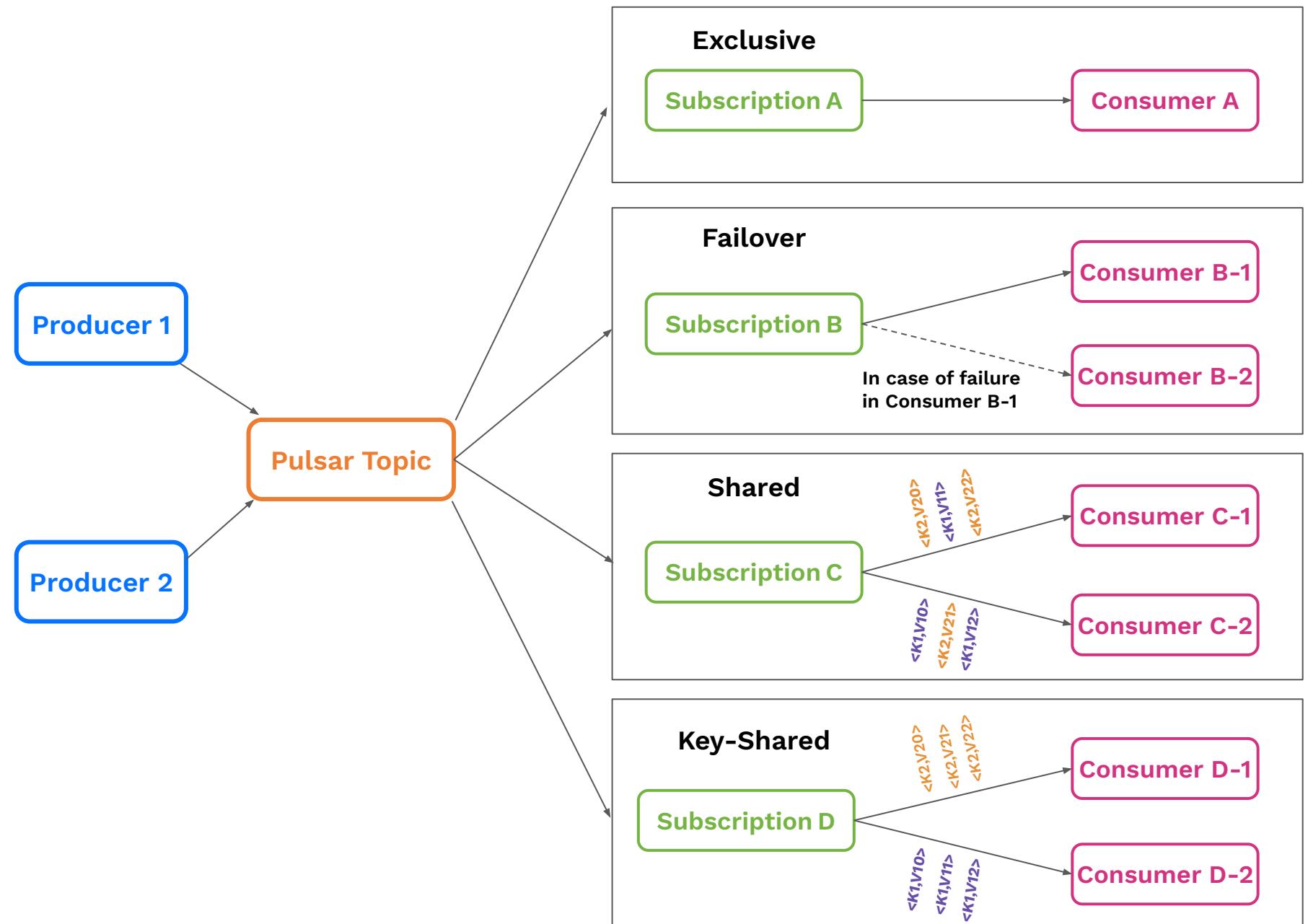
# Pulsar API Design



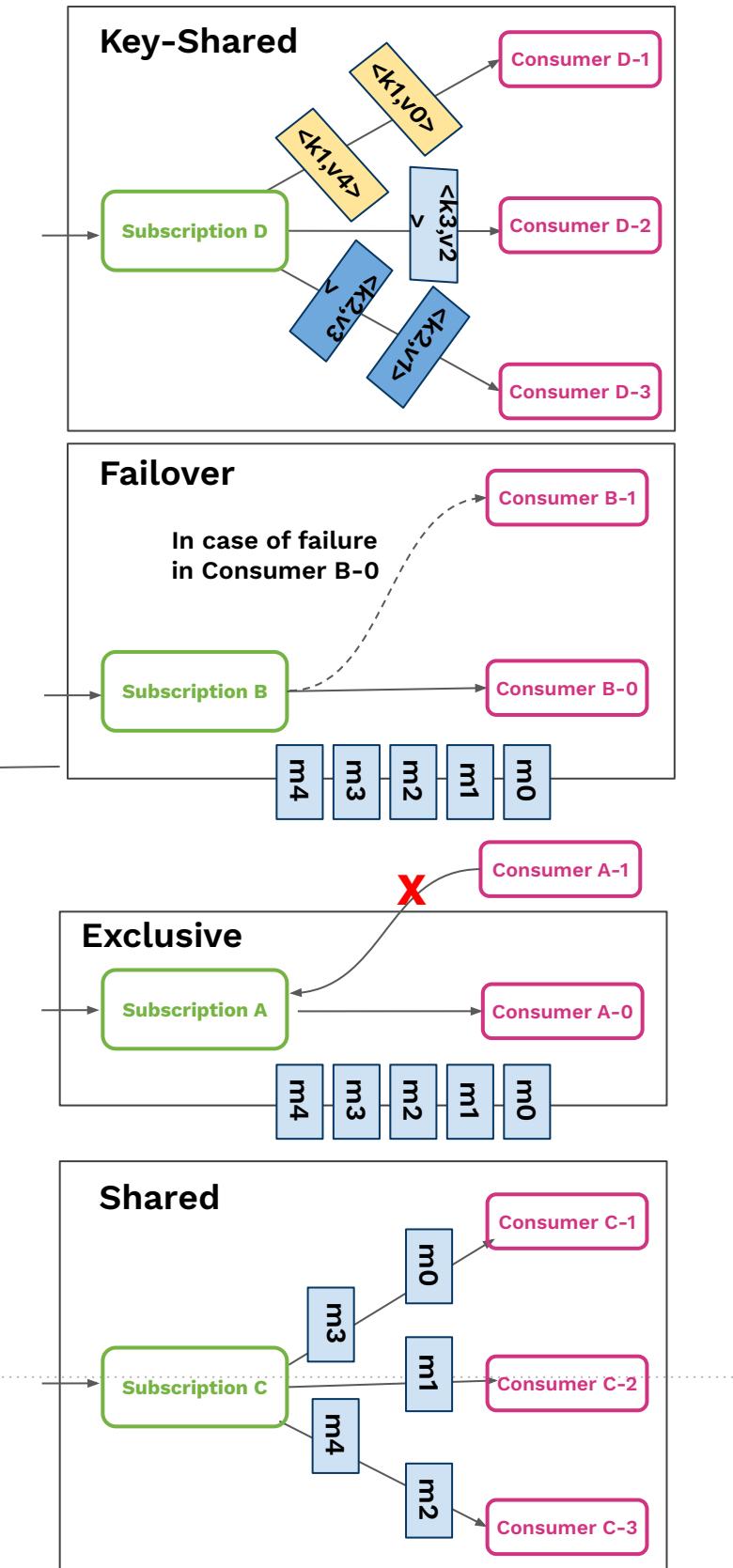
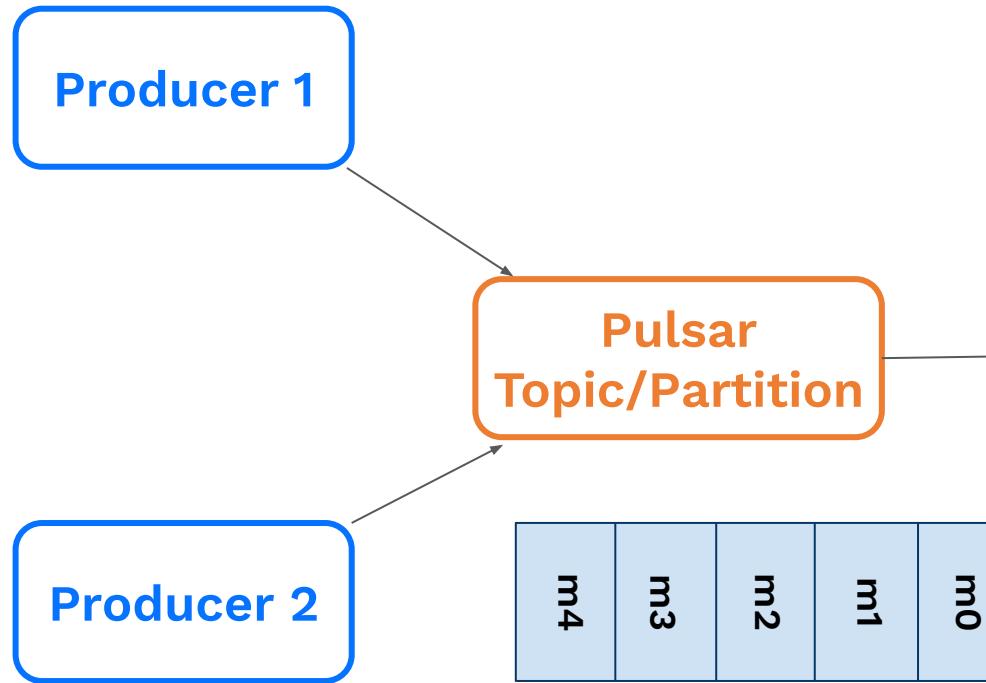
# Subscription Modes

Different subscription modes have different semantics:

- Exclusive/Failover** - guaranteed order, single active consumer
- Shared** - multiple active consumers, no order
- Key\_Shared** - multiple active consumers, order for given key

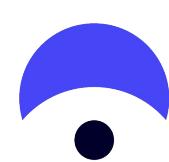


# Unified Messaging Model



**Streaming**

**Messaging**



# Stream Native Cloud

A cloud-native, real-time messaging and streaming platform to support multi-cloud and hybrid cloud strategies.

**Powered  
by Pulsar**



**Cloud  
Native**



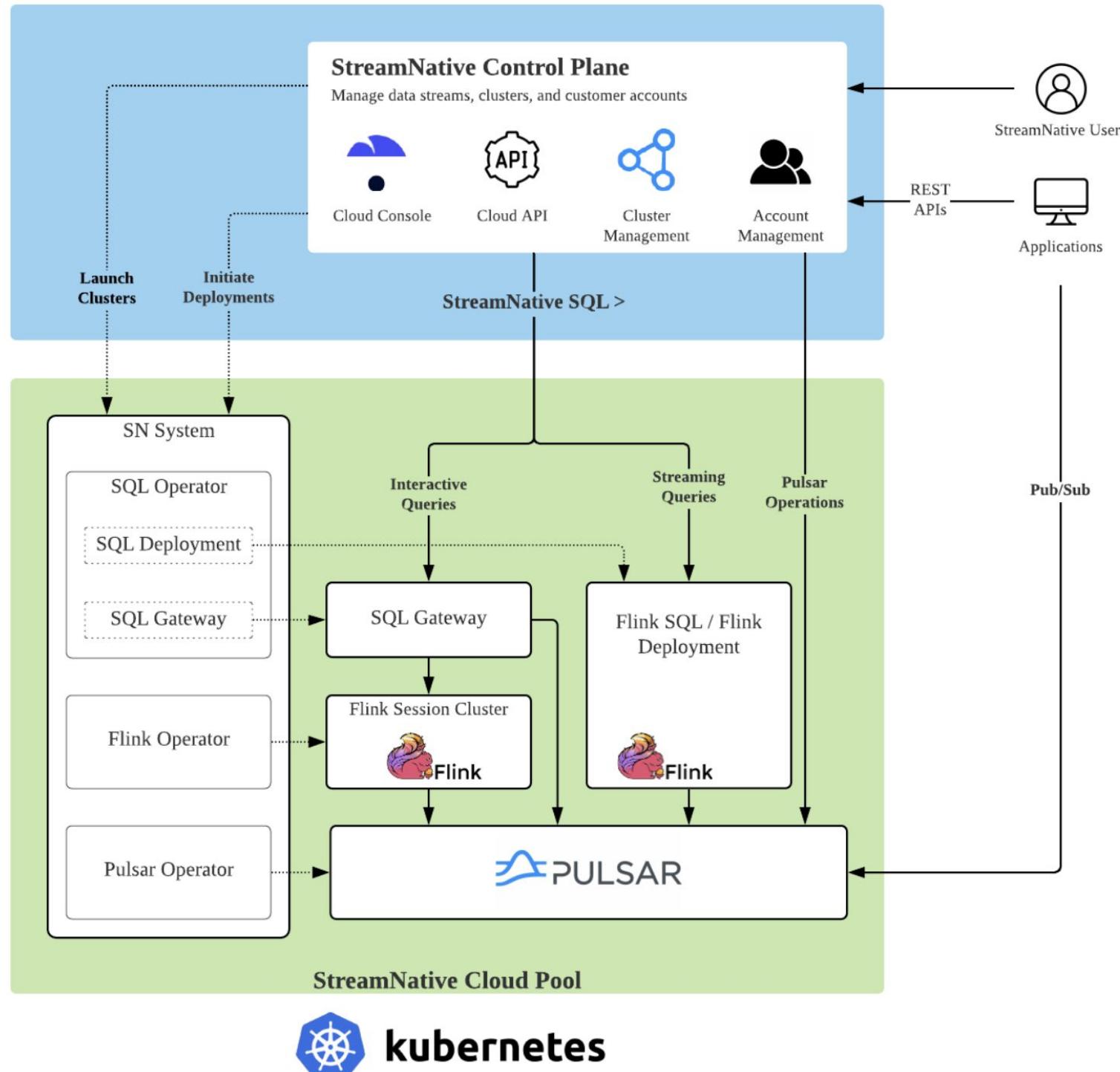
**Flink SQL**



**Built for  
Containers**



**StreamNative**



Google Cloud Platform



```

| mqtt-python | 
| mqtt-rp4 | 
| my-topic | 
| nvidia-kafka-1 | 
| rp4-kafka-1 | 
| rwar | 
| scada | 
| stocks | 
| stonks | 
| stonkss | 
| topic82547611 | 
+-----+
28 rows in set

Flink SQL> use catalog default_catalog;
[INFO] Execute statement succeed.

Flink SQL> show tables;
+-----+
| table name |
+-----+
| scada2 |
| stocks |
+-----+
2 rows in set

Flink SQL> exit;
[INFO] Exiting Flink SQL CLI Client...

```



# Apache Pulsar - Cloud Storage Sink

- Ensure exactly-once delivery. Records, which are exported using a deterministic partitioner, are delivered with exactly-once semantics regardless of the eventual consistency of cloud storage.
- Support data formats with or without a Schema. The Cloud Storage sink connector supports writing data to objects in cloud storage in either Avro, JSON, or Parquet format. Generally, the Cloud Storage sink connector may accept any data format that provides an implementation of the Format interface.
- Support time-based partitioner. The Cloud Storage sink connector supports the TimeBasedPartitioner class based on the publishTime timestamp of Pulsar messages. Time-based partitioning options are daily or hourly.
- Support more kinds of object storage. The Cloud Storage sink connector uses jclouds as an implementation of cloud storage. You can use the JAR package of the jclouds object storage to connect to more types of object storage. If you need to customize credentials, you can register `org.apache.pulsar.io.jcloud.credential.JcloudsCredential` via the Service Provider Interface (SPI).

# Apache Pulsar - Other Sinks



- mongoDB



- AWS Lambda



- redis



- AWS S3



- GCS



<https://hub.streamnative.io/connectors/cloud-storage-sink/2.5.1/>



## Apache Pulsar - Other Sinks



mongoDB



AWS Lambda



redis



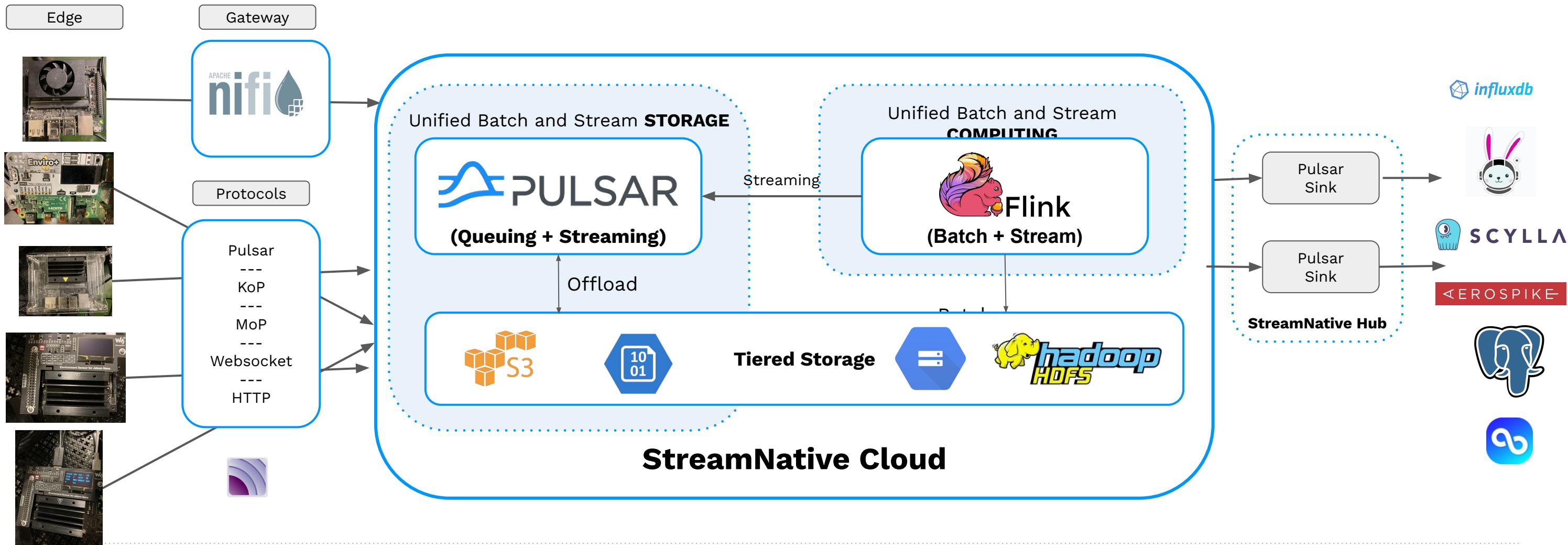
AWS S3



GCS

# End-to-End Streaming FLiPN Edge AI Application

Apache Flink - Apache Pulsar - Apache NiFi <-> Devices - GPU/TPU - Python/Go/Java



# Show Me The Data!

```
{"ir": "252.0", "id": "20210914001822_5e4882ee-22d9-432c-9074-19f12be62006", "end": "1631578962.03", "uuid": "nano_uuid_koo_20210914001822", "lux": "0", "gputemp": "26.0", "cputemp": "25.5", "te": "259.676094055", "systemtime": "09/13/2021 20:22:42", "hum": "52.31", "memory": 20.1, "gas": "29671.0", "pressure": "1013.62", "host": "nano2gb-desktop", "diskusage": "33312.5 MB", "ipaddress": "192.168.1.170", "macaddress": "1c:bf:ce:1a:7f:a0", "temp": "22.92", "uv": "0.01", "gputempf": "79.0", "host_name": "nano2gb-desktop", "runtime": "260.0", "cpu": 3.8, "cputempf": "78.0"}
```

# Show Me More Data

The screenshot shows the StreamNative Cloud interface with the 'sgconnector' selected in the sidebar. The main area displays log entries with timestamps and detailed JSON log messages. The 'Message' tab is selected at the bottom.

**Log Entries:**

- 46:20: {"uuid": "xav\_uuid\_video0\_uul\_20200921211433", "camera": "/dev/video0", "ipaddress": "192.168.1.78", "networktime": 25.096160888671875, "top1pct": 18.29833984375, "top1": "desk", "cputemp": "30.5", "gputemp": "31.0", "gputempf": "88", "cputempf": "87", "runtime": "6", "host": "nvidia-desktop", "filename": "/home/nvidia/nvme/images/out\_video0\_xsx\_20200921211433.jpg", "imageinput": "/home/nvidia/nvme/images/img\_video0\_nzj\_20200921211433.jpg", "host\_name": "nvidia-desktop", "macaddress": "70:66:55:15:b4:a5", "end": "1600722879.891937", "te": "6.490077018737793", "systemtime": "09/21/2020 17:14:39", "cpu": 48.0, "diskusage": "42145.1 MB", "memory": 45.9, "id": "video0\_20200921211433\_ff3757d5-52eb-4a8a-8b40-5da2fc3edca"}
- 46:21: {"uuid": "xav\_uuid\_video2\_cg\_z\_20200921211450", "camera": "/dev/video2", "ipaddress": "192.168.1.78", "networktime": 29.682527542114258, "top1pct": 33.203125, "top1": "ski", "cputemp": "31.0", "gputemp": "31.0", "gputempf": "88", "cputempf": "88", "runtime": "8", "host": "nvidia-desktop", "filename": "/home/nvidia/nvme/images/out\_video2\_zsz\_20200921211450.jpg", "imageinput": "/home/nvidia/nvme/images/img\_video2\_mgi\_20200921211450.jpg", "host\_name": "nvidia-desktop", "macaddress": "70:66:55:15:b4:a5", "end": "1600722898.9325902", "te": "8.286669492721558", "systemtime": "09/21/2020 17:14:58", "cpu": 45.3, "diskusage": "42145.1 MB", "memory": 46.0, "id": "video2\_20200921211450\_b5c396ad-82ff-48aa-ac9b-9caeef4f42855"}
- 46:22: {"uuid": "xav\_uuid\_video0\_prb\_20200921211601", "camera": "/dev/video0", "ipaddress": "192.168.1.78", "networktime": 6.291071891784668, "top1pct": 16.02783203125, "top1": "desk", "cputemp": "32.0", "gputemp": "32.5", "gputempf": "90", "cputempf": "90", "runtime": "5", "host": "nvidia-desktop", "filename": "/home/nvidia/nvme/images/out\_video0\_gwe\_20200921211601.jpg", "imageinput": "/home/nvidia/nvme/images/img\_video0\_oon\_20200921211601.jpg", "host\_name": "nvidia-desktop", "macaddress": "70:66:55:15:b4:a5", "end": "1600722966.4952142", "te": "5.116875648498535", "systemtime": "09/21/2020 17:16:06", "cpu": 44.5, "diskusage": "42145.1 MB", "memory": 46.3, "id": "video0\_20200921211601\_935b5e90-4bcd-47dc-84eb-f85846fae190"}
- 46:23: {"uuid": "xav\_uuid\_video0\_zwh\_20210804181811", "camera": "/dev/video0", "ipaddress": "192.168.1.214", "networktime": 48.33692932128906, "top1pct": 40.91796875, "top1": "Model T", "cputemp": "41.5", "gputemp": "42.0", "gputempf": "108", "cputempf": "107", "runtime": "11", "host": "nvidia-desktop", "filename": "/home/nvidia/nvme/images/out\_video0\_tes\_20210804181811.jpg", "imageinput": "/home/nvidia/nvme/images/img\_video0\_prs\_20210804181811.jpg", "host\_name": "nvidia-desktop", "macaddress": "70:66:55:15:b4:a5", "end": "1628101102.2976809", "te": "10.955556154251099", "systemtime": "08/04/2021 14:18:22", "cpu": 12.0, "diskusage": "33277.7 MB", "memory": 33.5, "id": "video0\_20210804181811\_67e85abc-9085-41c3-b27f-eb7f7aa36daa"}
- 46:24: {"uuid": "xav\_uuid\_video0\_zwh\_20210804181811", "camera": "/dev/video0", "ipaddress": "192.168.1.214", "networktime": 48.33692932128906, "top1pct": 40.91796875, "top1": "Model T", "cputemp": "41.5", "gputemp": "42.0", "gputempf": "108", "cputempf": "107", "runtime": "11", "host": "nvidia-desktop", "filename": "/home/nvidia/nvme/images/out\_video0\_tes\_20210804181811.jpg", "imageinput": "/home/nvidia/nvme/images/img\_video0\_prs\_20210804181811.jpg", "host\_name": "nvidia-desktop", "macaddress": "70:66:55:15:b4:a5", "end": "1628101102.2976809", "te": "10.955556154251099", "systemtime": "08/04/2021 14:18:22", "cpu": 12.0, "diskusage": "33277.7 MB", "memory": 33.5, "id": "video0\_20210804181811\_67e85abc-9085-41c3-b27f-eb7f7aa36daa"}

**Selected Log Message:**

```
{"uuid": "xav_uuid_video0_zwh_20210804181811", "camera": "/dev/video0", "ipaddress": "192.168.1.214", "networktime": 48.33692932128906, "top1pct": 40.91796875, "top1": "Model T", "cputemp": "41.5", "gputemp": "42.0", "gputempf": "108", "cputempf": "107", "runtime": "11", "host": "nvidia-desktop", "filename": "/home/nvidia/nvme/images/out_video0_tes_20210804181811.jpg", "imageinput": "/home/nvidia/nvme/images/img_video0_prs_20210804181811.jpg", "host_name": "nvidia-desktop", "macaddress": "70:66:55:15:b4:a5", "end": "1628101102.2976809", "te": "10.955556154251099", "systemtime": "08/04/2021 14:18:22", "cpu": 12.0, "diskusage": "33277.7 MB", "memory": 33.5, "id": "video0_20210804181811_67e85abc-9085-41c3-b27f-eb7f7aa36daa"}
```

# 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-iot>

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

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

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

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

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



## Deeper Content

- <https://www.datainmotion.dev/2020/10/running-flink-sql-against-kafka-using.html>
- <https://www.datainmotion.dev/2020/10/top-25-use-cases-of-cloudera-flow.html>
- <https://github.com/tspannhw/EverythingApacheNiFi>
- <https://github.com/tspannhw/CloudDemo2021>
- <https://github.com/tspannhw/StreamingSQLExamples>
- <https://www.linkedin.com/pulse/2021-schedule-tim-spann/>
- <https://github.com/tspannhw/StreamingSQLExamples/blob/8d02e62260e82b027b43abb911b5c366a3081927/README.md>
- <https://www.pulsardeveloper.com/>



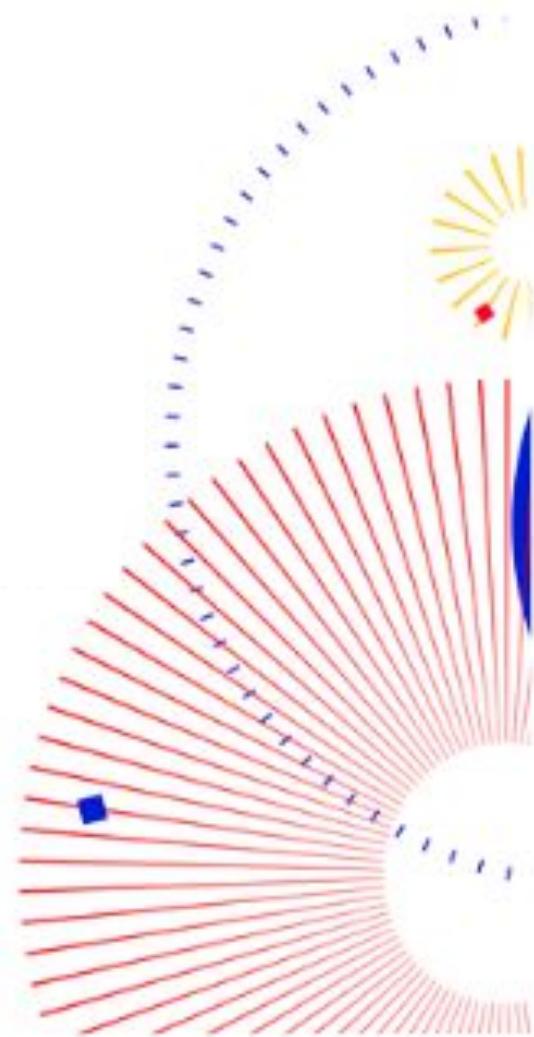
## Connect with the Community & Stay Up-To-Date

- 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



# Pulsar Summit Europe

October 6, 2021



# Pulsar Summit Asia

November 20-21, 2021

Contact us at [partners@pulsar-summit.org](mailto:partners@pulsar-summit.org) to become a sponsor or partner

# Let's Keep in Touch!



**Tim Spann**  
Developer Advocate

-  <https://twitter.com/paasDev>
-  <https://github.com/tspannhw>
-  <https://www.linkedin.com/in/timothyspann>

## Other Resources:

- <https://www.datainmotion.dev/>
- <https://github.com/tspannhw/SpeakerProfile>
- <https://dev.to/tspannhw>
- <https://sessionize.com/tspann/>
- <https://www.slideshare.net/bunkertor>

# Q & A

