Using FLiP with InfluxDB for Edge AI IoT at Scale

Tim Spann | Developer Advocate

Tim Spann, Developer Advocate at StreamNative

Tim Spann
Developer Advocate

StreamNative

- 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, Kafka, Big Data, Cloud, MXNet, IoT 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 · FLIP STACK · EY · HORTONWORKS · airis.DATA · Hewlett Packard Enterprise

Agenda

- Learn how to build an end-to-end streaming edge app
- How to pull messages from Pulsar topics and persists the messages to InfluxDB
- Building a data stream for IoT with NiFi and InfluxDB
- Using Apache Flink + Apache Pulsar
- Using Apache Spark + Apache Pulsar

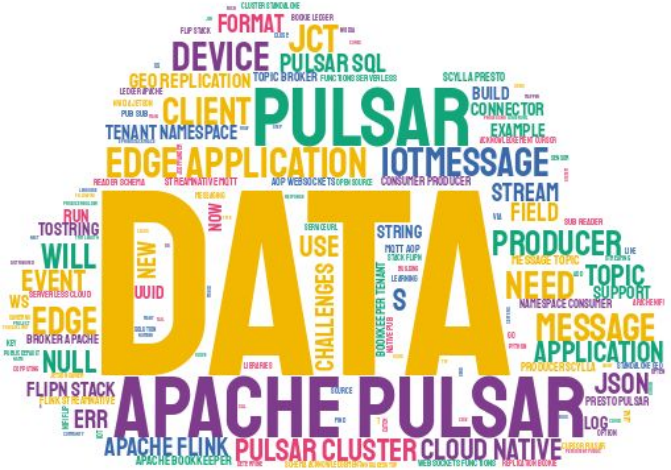
FLiP Stack Weekly

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

https://bit.ly/32dAJft

FLiP(N)(S) Stack

- Apache **Fl**ink
- Apache **P**ulsar
- Pulsar Functions
- StreamNative's Flink Connector for Puls
- Apache **N**iFi
- Apache Spark
- Python, Java, Golang



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

Pulsar Benefits

Unified Messaging Model

Simplify your data infrastructure and enable new use cases with queuing and streaming capabilities in one platform.

Multi-tenancy

Enable multiple user groups to share the same cluster, either via access control, or in entirely different namespaces.

Scalability

Decoupled data computing and storage enable horizontal scaling to handle data scale and management complexity.

Geo-replication

Support for multi-datacenter replication with both asynchronous and synchronous replication for built-in disaster recovery.

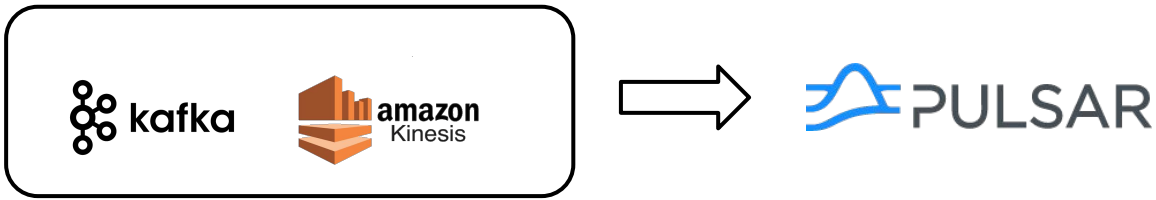
Tiered storage

Enable historical data to be offloaded to cloud-native storage and store event streams for indefinite periods of time.

Top Pulsar Use Cases

#1 Message Queuing

#2 Data Streaming



- ✓ Not built for the cloud
- ✓ Single tenant systems
- ✓ Monolithic architecture couples compute with storage
- ✓ Lack of geo replication support

Pulsar: Unified Messaging + Data Streaming

Messaging

Ideal for work queues that do not require tasks to be performed in a particular order—for example, sending one email message to many recipients.

RabbitMQ and Amazon SQS are examples of popular queue-based message systems.

Pulsar: Unified Messaging + Data Streaming

Messaging

Ideal for work queues that do not require tasks to be performed in a particular order—for example, sending one email message to many recipients.

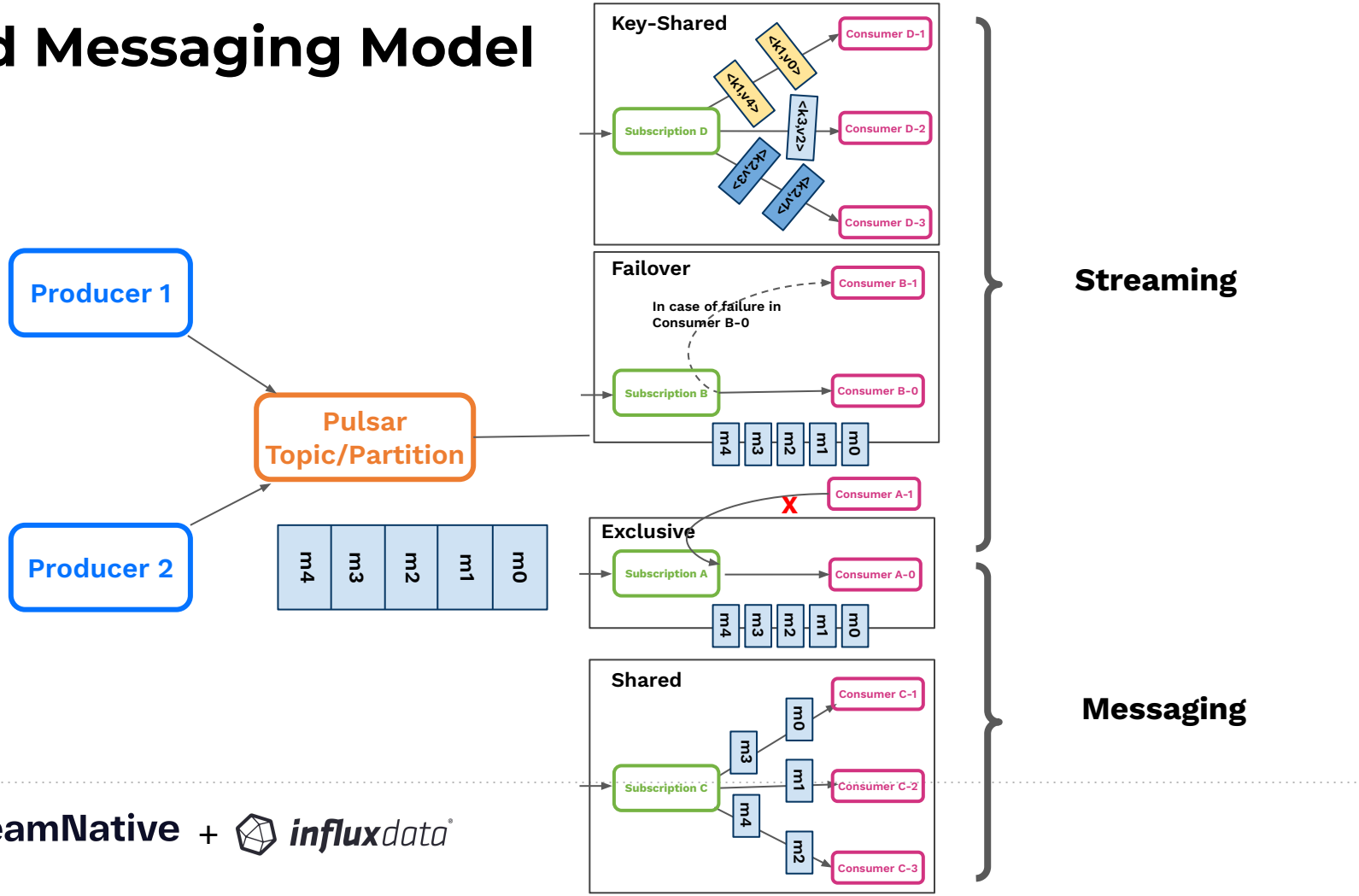
RabbitMQ and Amazon SQS are examples of popular queue-based message systems.

.. and Streaming

Works best in situations where the order of messages is important—for example, data ingestion.

Kafka and Amazon Kinesis are examples of messaging systems that use streaming semantics for consuming messages.

Unified Messaging Model



StreamNative + influxdata

Use Cases

Multi-Tenant Data Infrastructure

AdTech

Fraud Detection

Connected Car

IoT Analytics

Microservices Development

Pulsar Global Adoption

Pulsar Adoption Use Cases

splunk>

Adopted Pulsar to **replace Kafka in their DSP** (Data Streaming Platform).

- 1.5-2x lower in capex cost
- 5-50x improvement in latency
- 2-3x lower in opex due

Tencent 腾讯

Adopted Pulsar to **power their billing platform**, Midas, which processing hundreds of billions of financial transactions daily.

Adoption then expanded to Tencent's Federated Learning Platform and Tencent Gaming.

APPLIED MATERIALS®

Applied Materials is one of the biggest semiconductor hardware and software supplier in the industry.

They adopted Pulsar to enable them to **build a message bus to tie all of their data together**. They previously used Tibco.

Stream Native

Founded by the original developers of Apache Pulsar.

Passionate and dedicated team.

StreamNative helps teams to **capture**, **manage**, and **leverage data** using Pulsar's unified messaging and streaming platform.

streamnative.io

Founded By The Creators Of Apache Pulsar

Sijie Guo

Founder and CEO

ASF Member
Pulsar/BookKeeper PMC

Matteo Merli

CTO

ASF Member
Pulsar/BookKeeper PMC

Jia Zhai

Co-Founder

Pulsar/BookKeeper PMC

- ✓ Original creators of Apache Pulsar & BookKeeper
- ✓ Operated the largest Pulsar/BookKeeper cluster
- ✓ Data veterans with extensive industry experience

StreamNative

StreamNative Cloud

Powered by Apache Pulsar, StreamNative provides a cloud-native, real-time messaging and streaming platform to support multi-cloud and hybrid cloud strategies.

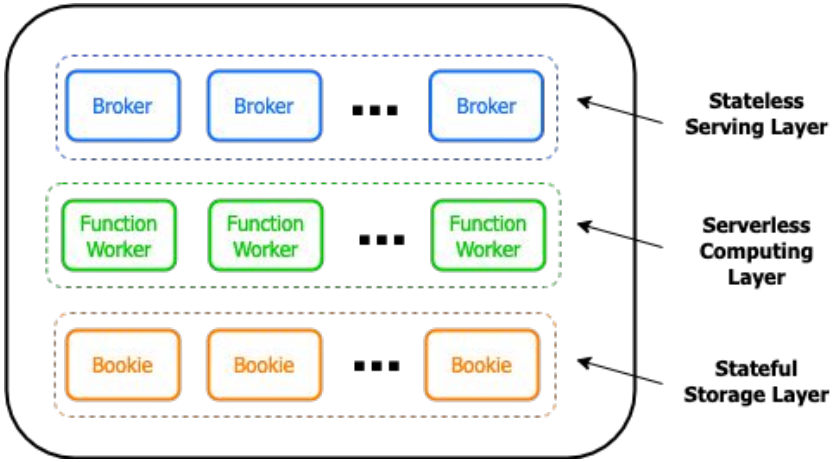
Cloud-Native

Built for Containers

Flink SQL

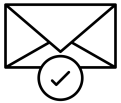
Apache Pulsar

- Serverless computing framework.
- Unbounded storage, multi-tiered architecture, and tiered-storage.
- Streaming & Pub/Sub messaging semantics.
- Multi-protocol support.



Why Apache Pulsar?

Unified Messaging Platform



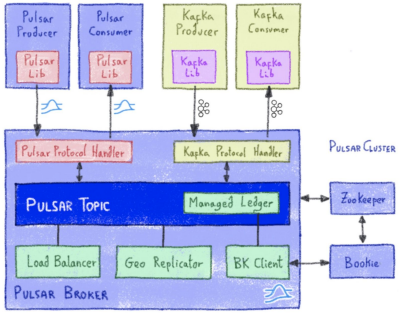
Guaranteed Message Delivery

Resiliency

Infinite Scalability

Connectivity

hub.streamnative.io



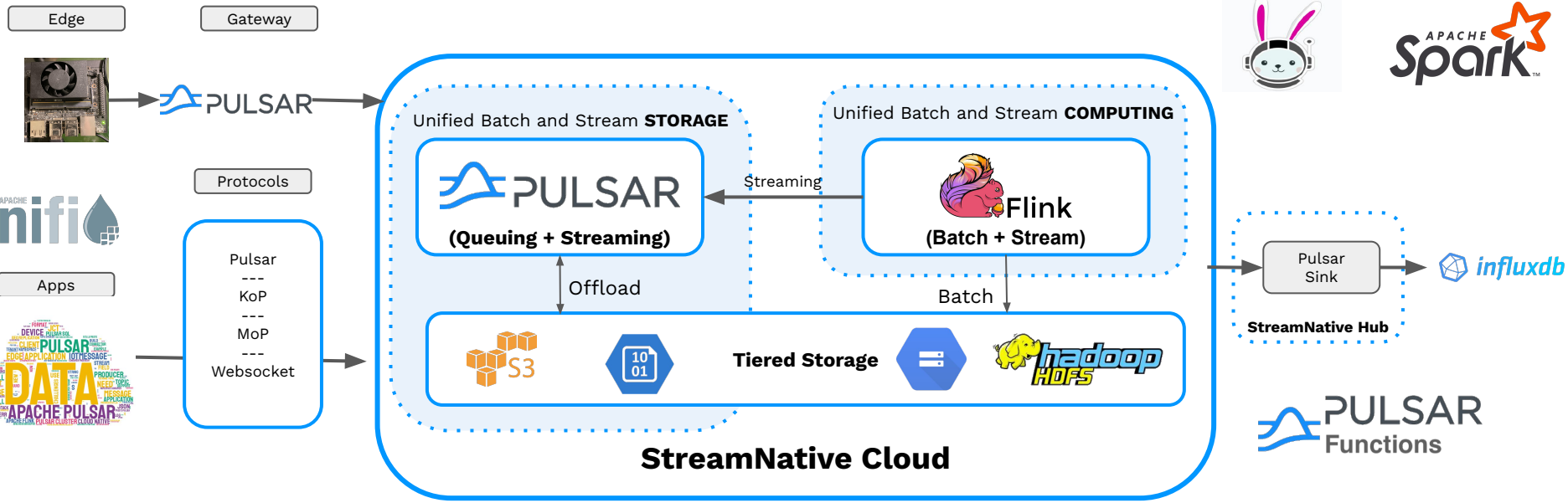
- **Functions** - Lightweight Stream Processing (Java, Python, Go)
- **Connectors** - Sources & Sinks (InfluxDB, Kafka, S3, Kinesis, Lambda, …)
- **Protocol Handlers** - AoP (AMQP), KoP (Kafka), MoP (MQTT)
- **Processing Engines** - Flink, Spark, Presto/Trino via Pulsar SQL
- **Data Offloaders** - Tiered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

Use Cases

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

Streaming Edge Apps

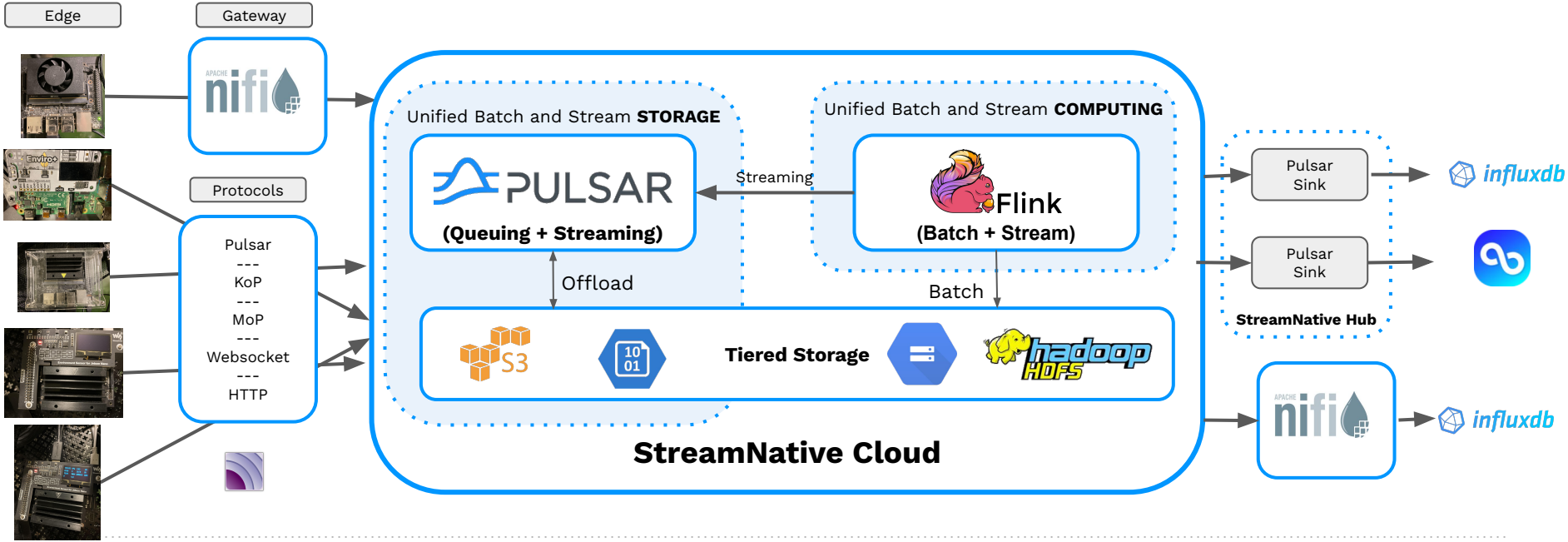
influxdb <-> Sensors <-> PULSAR <-> Sensors <-> Flink



StreamNative + influxdata

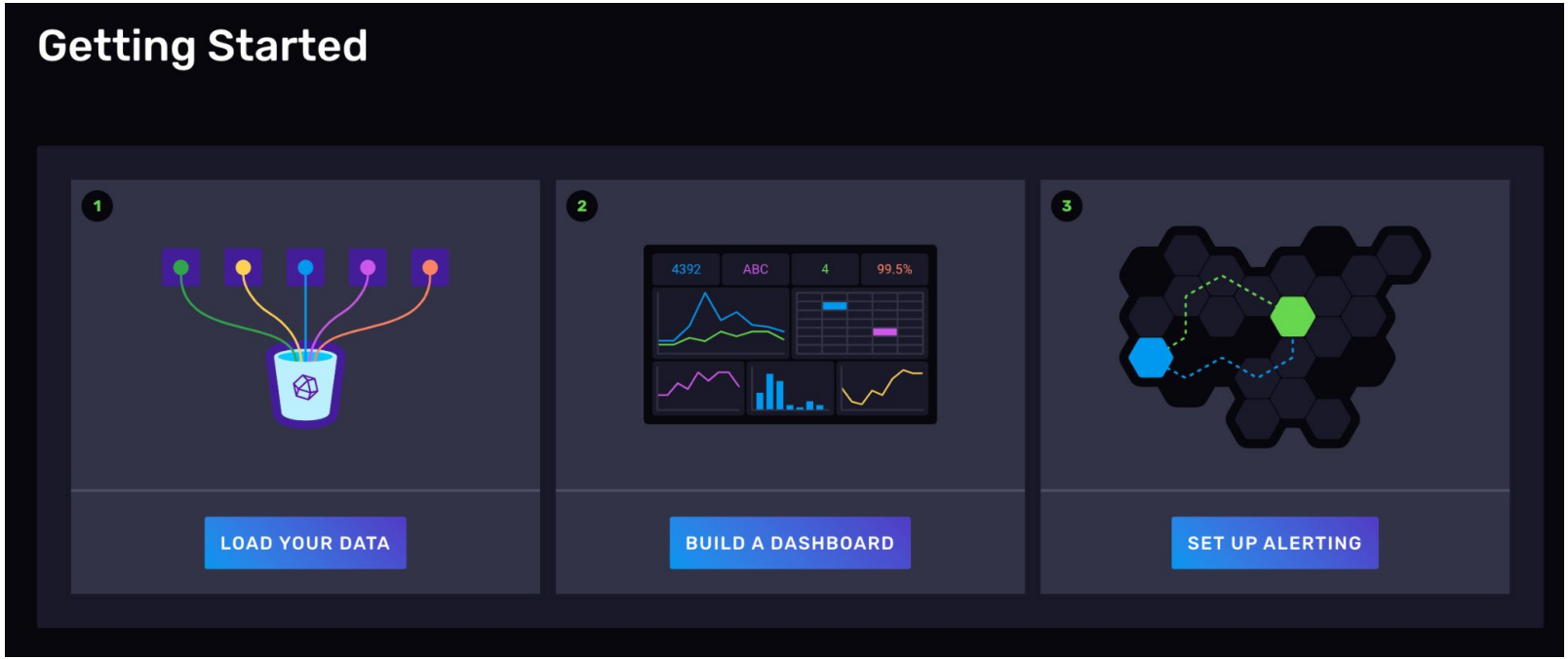
End-to-End Streaming Edge App

Apache Flink - Apache Pulsar - Apache NiFi <-> Devices <-> influxdb



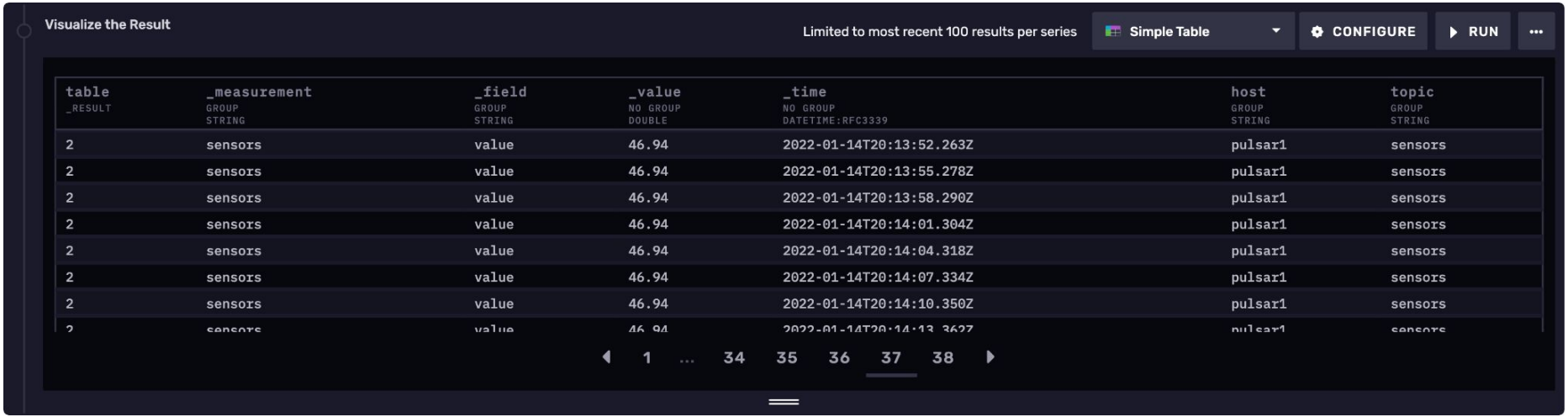
StreamNative + influxdata


Why
influxdb
influxdb cloud



StreamNative + influxdata

Easy Visualization



StreamNative + influxdata

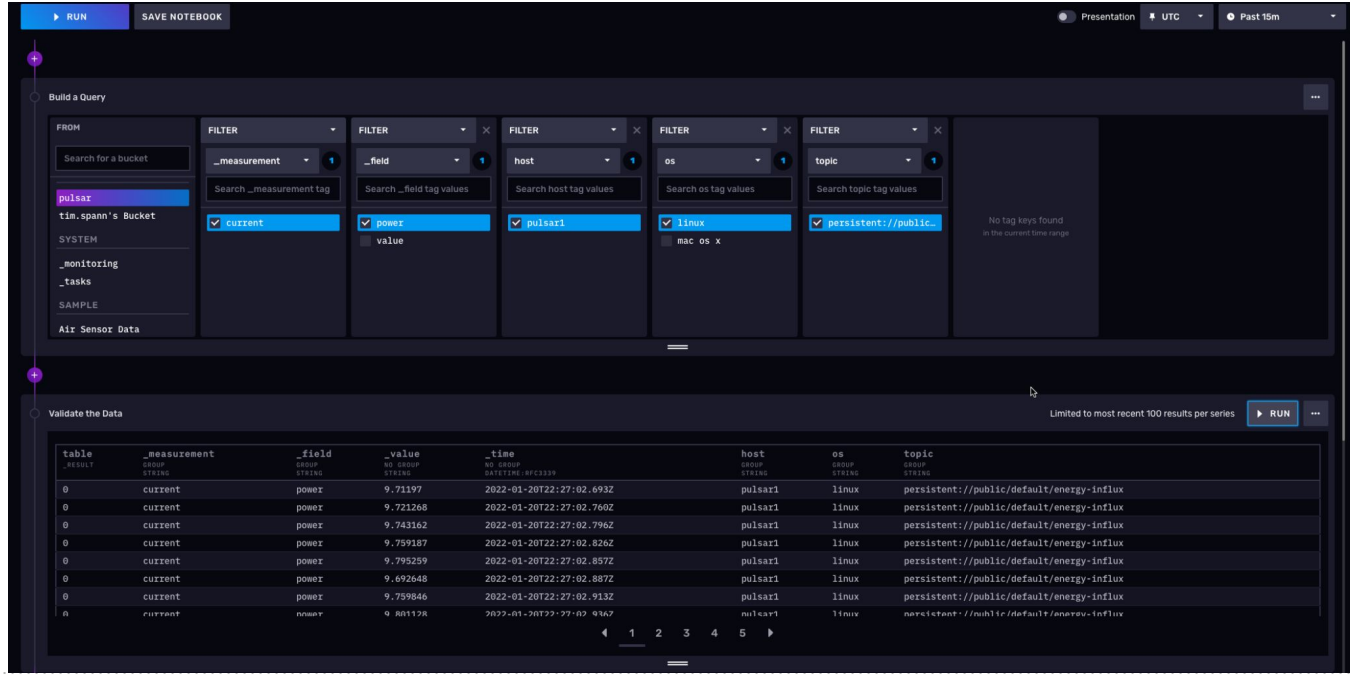

influxdb

Analytics


Pulsar to Influx - Energy
ADD CELL
ADD NOTE
Show Variables
Enable Annotations
REFRESHING EVERY 60S
Local
Past 1h
This dashboard doesn't have any cells with defined variables. Learn How
energy
Electric Current

influxdb

Analytics



StreamNative + influxdata

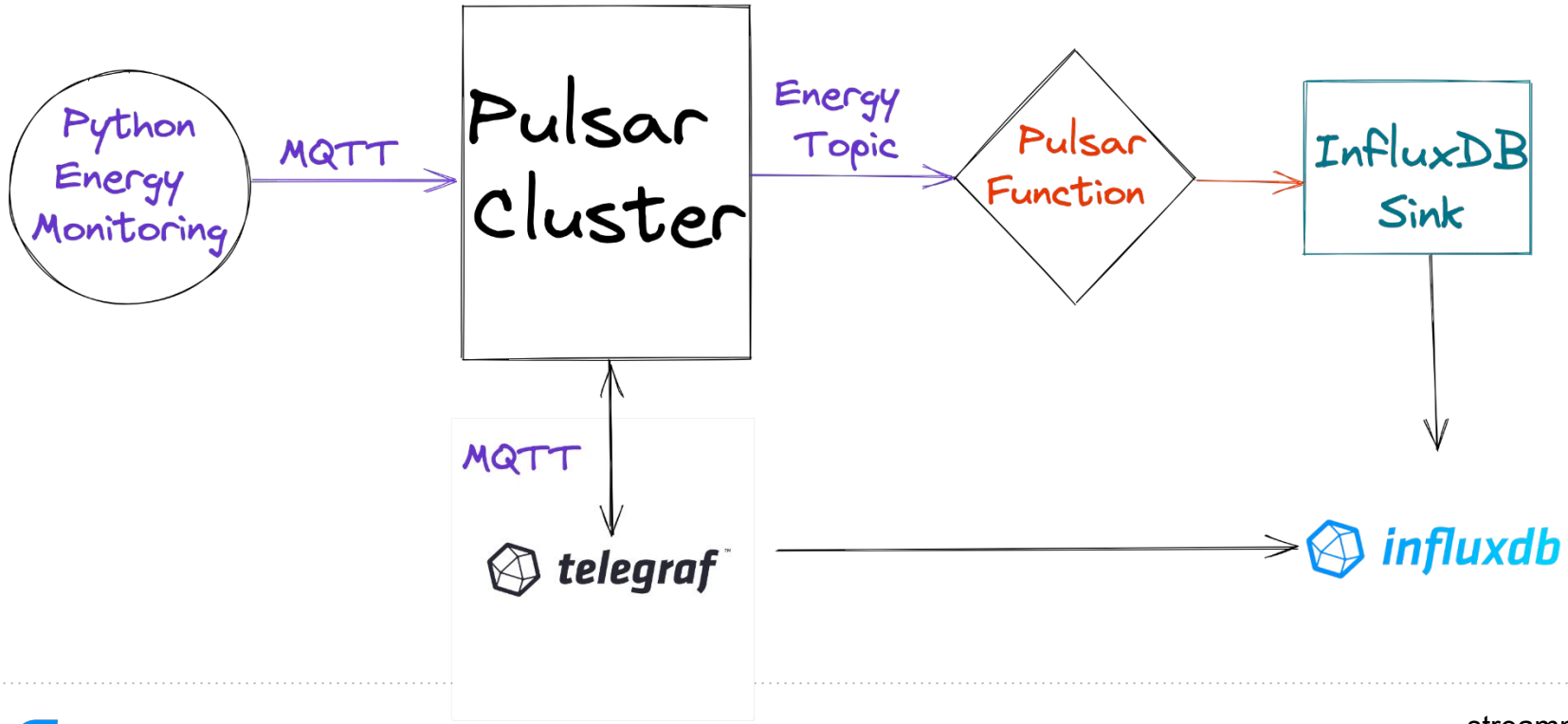
Apache Pulsar InfluxDB Sink

https://pulsar.apache.org/docs/en/io-influxdb-sink/

Apache Pulsar InfluxDB Sink

```
pulsar-admin sinks create --archive
./connectors/pulsar-io-influxdb-2.9.1.nar
    --name inflx-jtsn
    --sink-config-file conf/influxcloud.yml
    --inputs jtsninflx
pulsar-admin sinks get --name inflx-jtsn
pulsar-admin sinks status --name inflx-jtsn
```

Apache Pulsar InfluxDB Sink



InfluxDB Telegraf MQTT

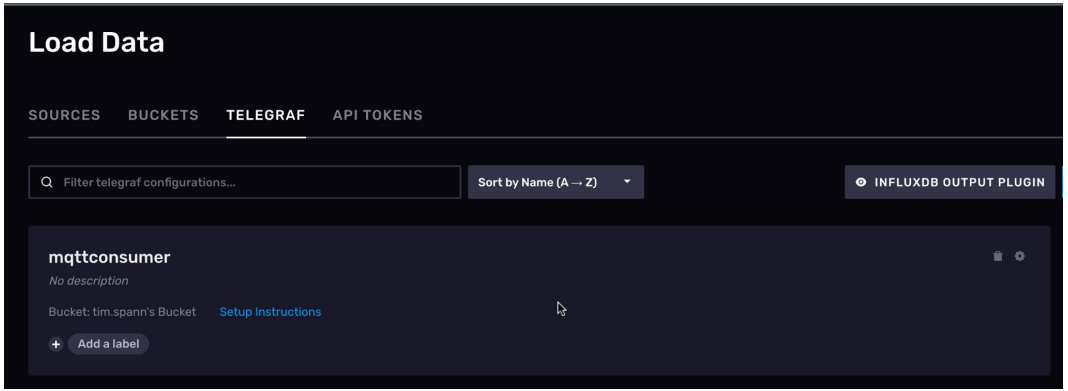
https://www.influxdata.com/integration/mqtt-monitoring/

MQTT on Pulsar (MoP)

https://github.com/streamnative/mop

InfluxDB Telegraf MQTT

telegraf™



InfluxDB Telegraf MQTT

```
# Read metrics from MQTT topic(s)
[[inputs.mqtt_consumer]]
  ## Broker URLs for the MQTT server or cluster.  To connect to multiple
  ## clusters or standalone servers, use a separate plugin instance.
  ##   example: servers = ["tcp://localhost:1883"]
  ##            servers = ["ssl://localhost:1883"]
  ##            servers = ["ws://localhost:1883"]
  servers = ["tcp://192.168.1.230:1883"]

  ## Topics that will be subscribed to.
  topics = [
    "telegrafcpu",
    "telegrafmem",
    "sensors",
  ]
```

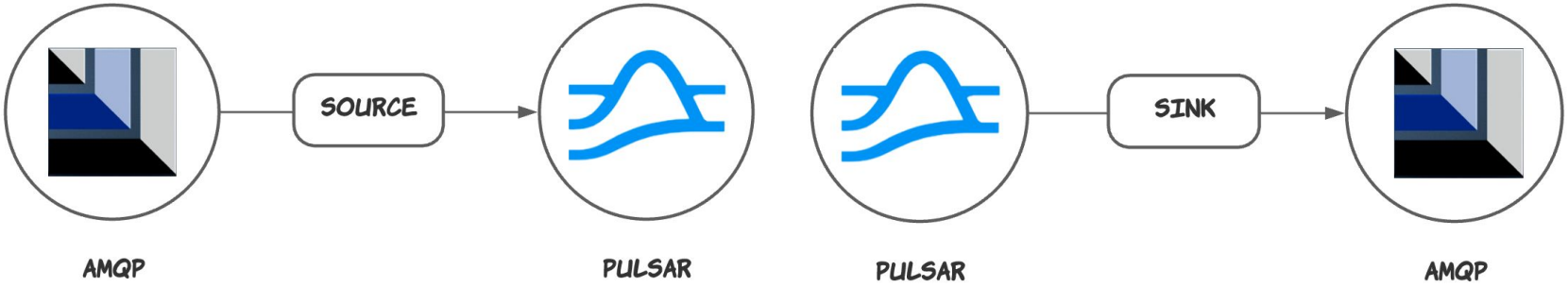
InfluxDB Telegraf MQTT

```
export INFLUX_TOKEN=INFLUXCLOUDTOKENGENERATED
telegraf --config https://us-east-1-1.aws.cloud2.influxdata.com/api/v2/telegrafs/code
```

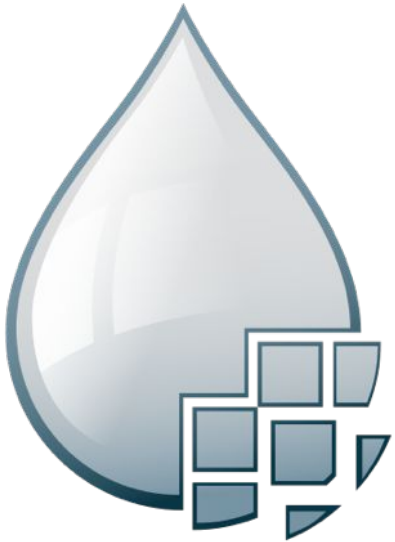
InfluxDB Telegraf RabbitMQ

AMQP on Pulsar (AoP) https://github.com/streamnative/aop

https://hub.streamnative.io/connectors/amqp-1-0-sink/v2.7.1.1
https://hub.streamnative.io/connectors/amqp-1-0-source/v2.7.1.1

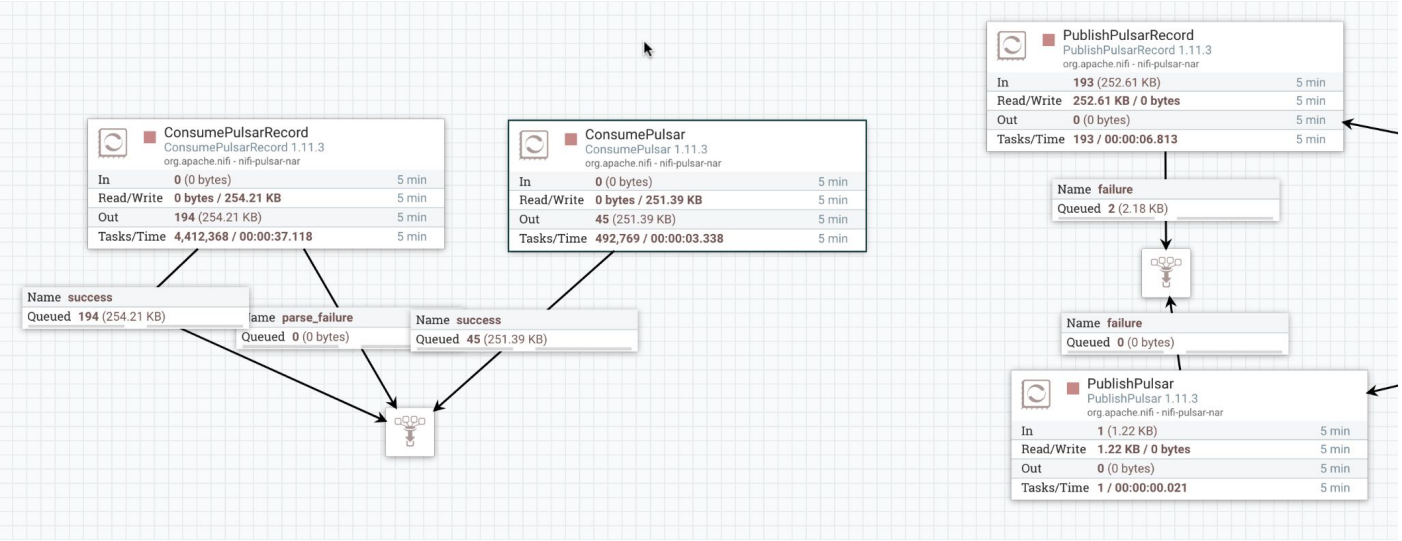


Why Apache NiFi?



- 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



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

Apache NiFi Pulsar Connector

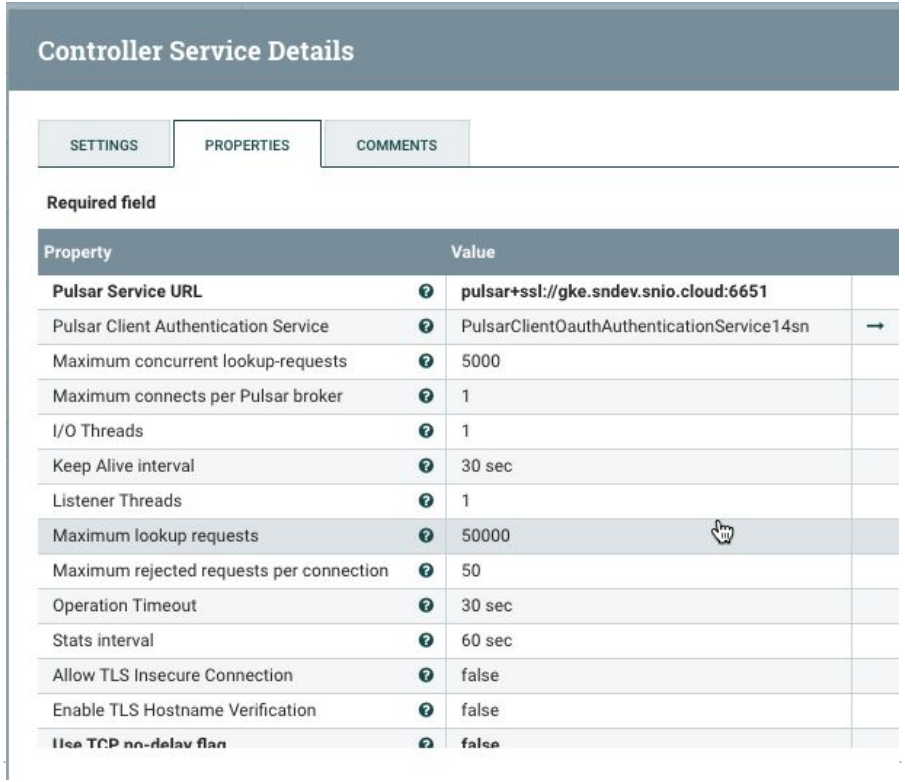
Displaying 12 of 339 | pulsar

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

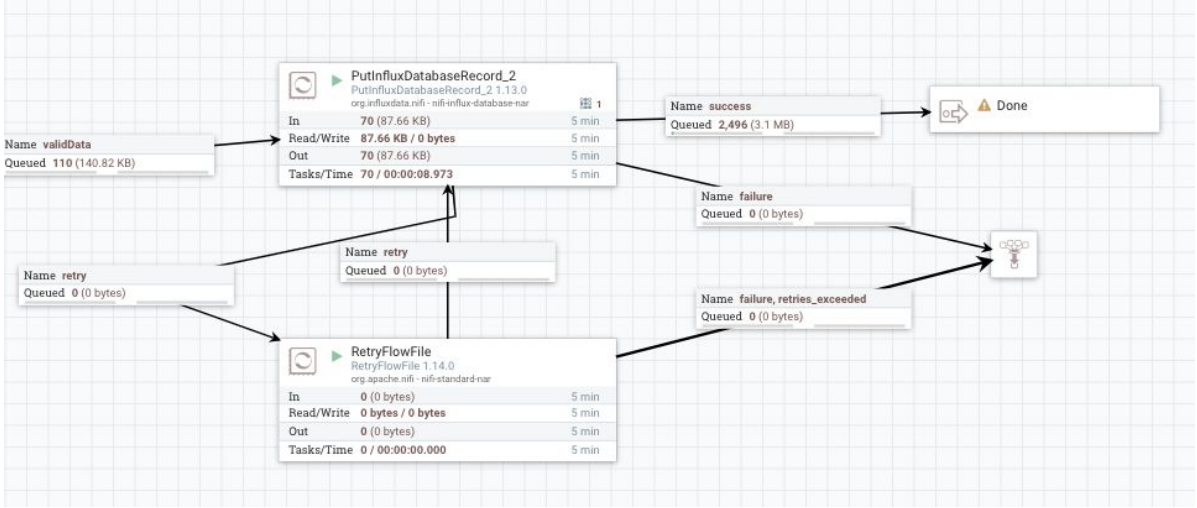
Consumes messages from Apache Pulsar. The complementary NiFi processor for sending messages is PublishPulsar.

Apache NiFi Pulsar Connector



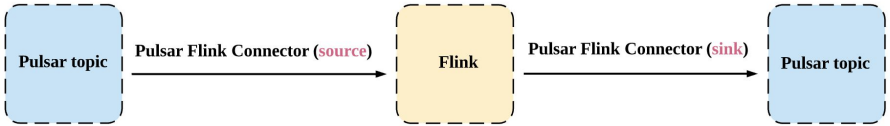
Apache NiFi Pulsar Connector

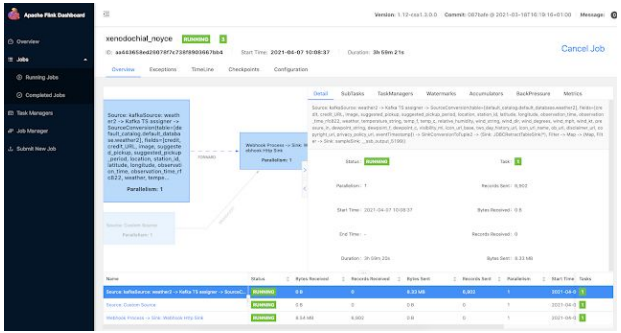
InfluxDB via Apache NiFi



https://github.com/influxdata/nifi-influxdb-bundle
https://www.influxdata.com/blog/building-a-data-stream-for-iot-with-nifi-and-influxdb/

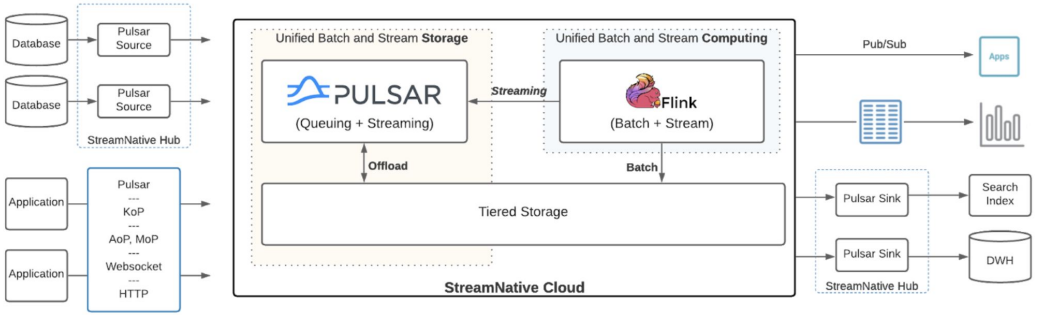
Why Apache Flink?

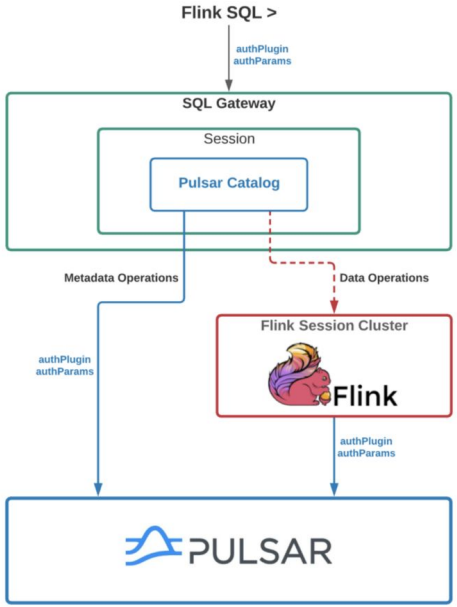


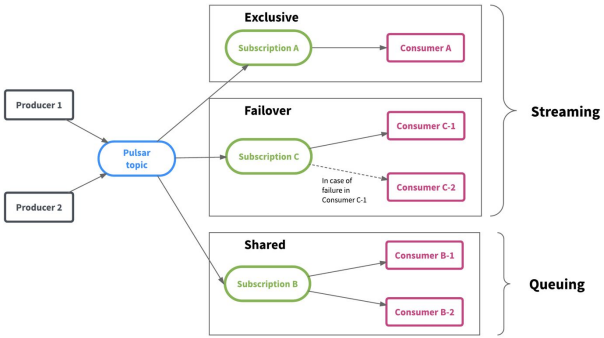


- Unified computing engine
- Batch processing is a special case of stream processing
- Stateful processing
- Massive Scalability
- Flink SQL for queries, inserts against Pulsar Topics
- Streaming Analytics
- Continuous SQL
- Continuous ETL
- Complex Event Processing
- Standard SQL Powered by Apache Calcite

Flink + Pulsar

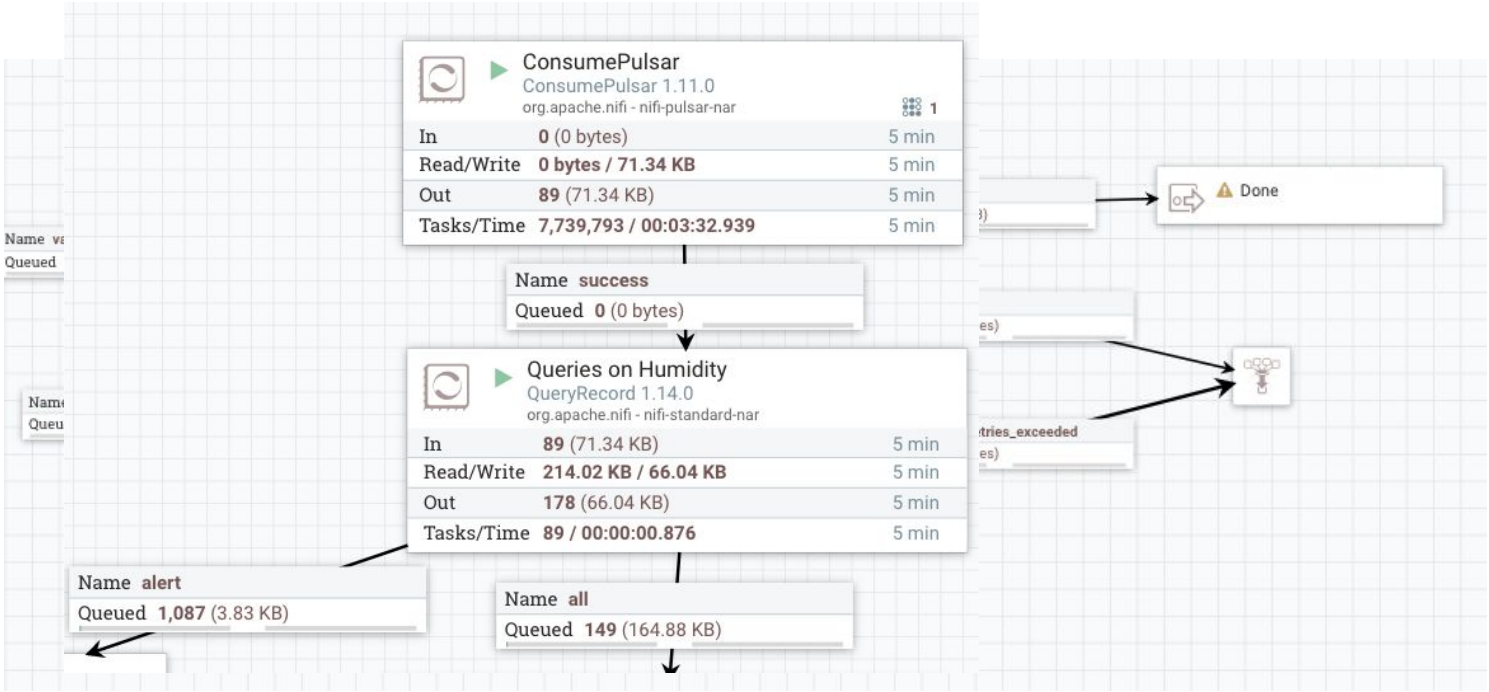






https://flink.apache.org/2019/05/03/pulsar-flink.html
https://github.com/streamnative/pulsar-flink
https://streamnative.io/en/blog/release/2021-04-20-flink-sql-on-streamnative-cloud

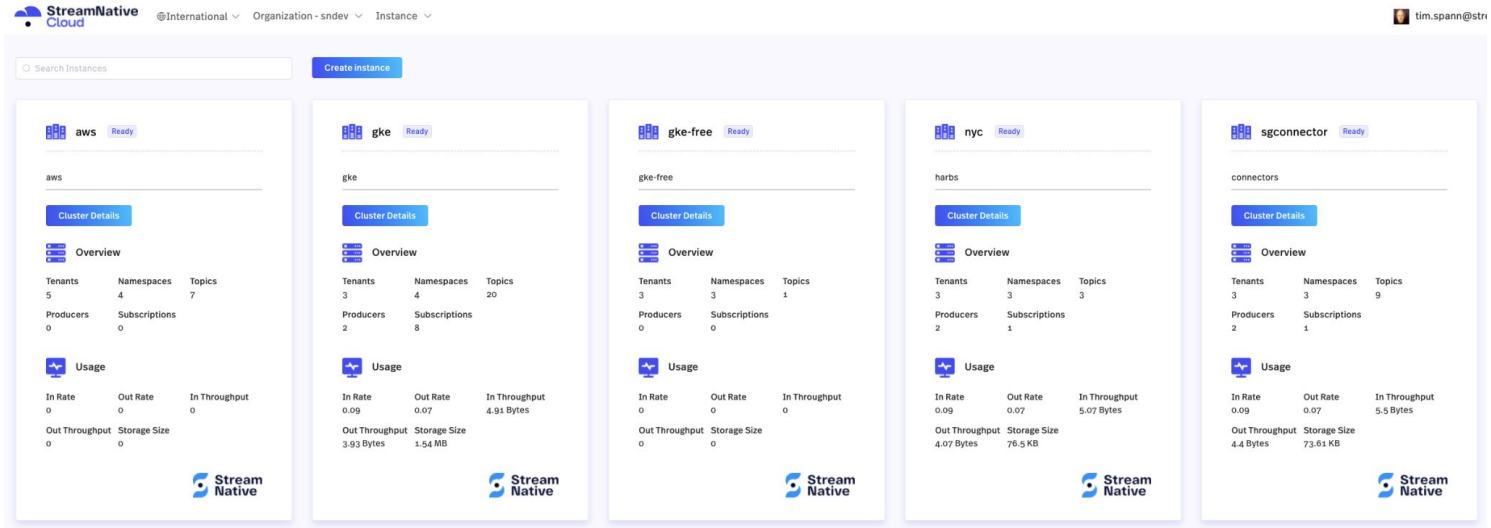
Apache NiFi


ConsumePulsar
ConsumePulsar 1.11.0
org.apache.nifi - nifi-pulsar-nar
1
In 0 (0 bytes) 5 min
Read/Write 0 bytes / 71.34 KB 5 min
Out 89 (71.34 KB) 5 min
Tasks/Time 7,739,793 / 00:03:32.939 5 min
Done
Name success
Queued 0 (0 bytes)
Queries on Humidity
QueryRecord 1.14.0
org.apache.nifi - nifi-standard-nar
In 89 (71.34 KB) 5 min
Read/Write 214.02 KB / 66.04 KB 5 min
Out 178 (66.04 KB) 5 min
Tasks/Time 89 / 00:00:00.876 5 min
Name alert
Queued 1,087 (3.83 KB)
Name all
Queued 149 (164.88 KB)

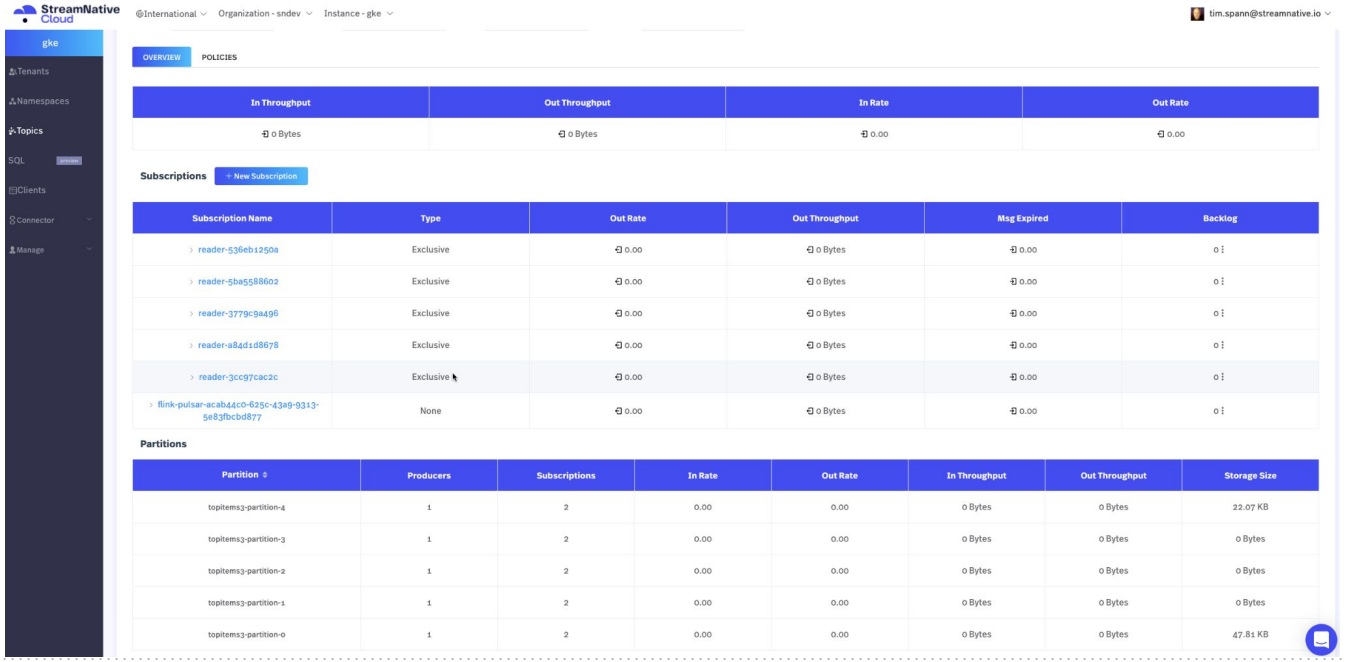
Apache Pulsar

```
tspann@Timothys-mbp apache-pulsar-2.8.0 % ./influxstatus.sh
{
  "numInstances" : 1,
  "numRunning" : 1,
  "instances" : [ {
    "instanceId" : 0,
    "status" : {
      "running" : true,
      "error" : "",
      "numRestarts" : 0,
      "numReadFromPulsar" : 12,
      "numSystemExceptions" : 0,
      "latestSystemExceptions" : [ ],
      "numSinkExceptions" : 0,
      "latestSinkExceptions" : [ ],
      "numWrittenToSink" : 12,
      "lastReceivedTime" : 1633550560478,
      "workerId" : "c-standalone-fw-127.0.0.1-8080"
    }
  } ]
}
```

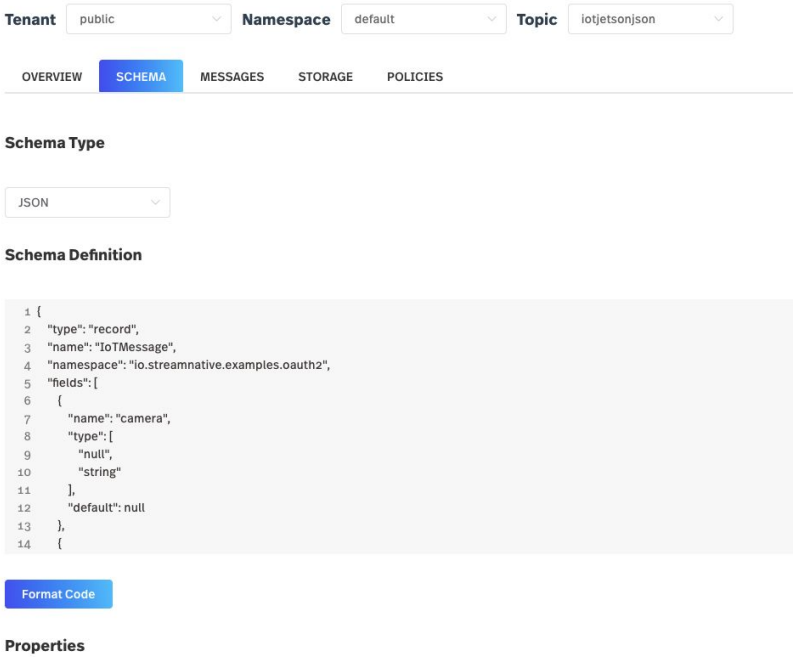
StreamNative Cloud



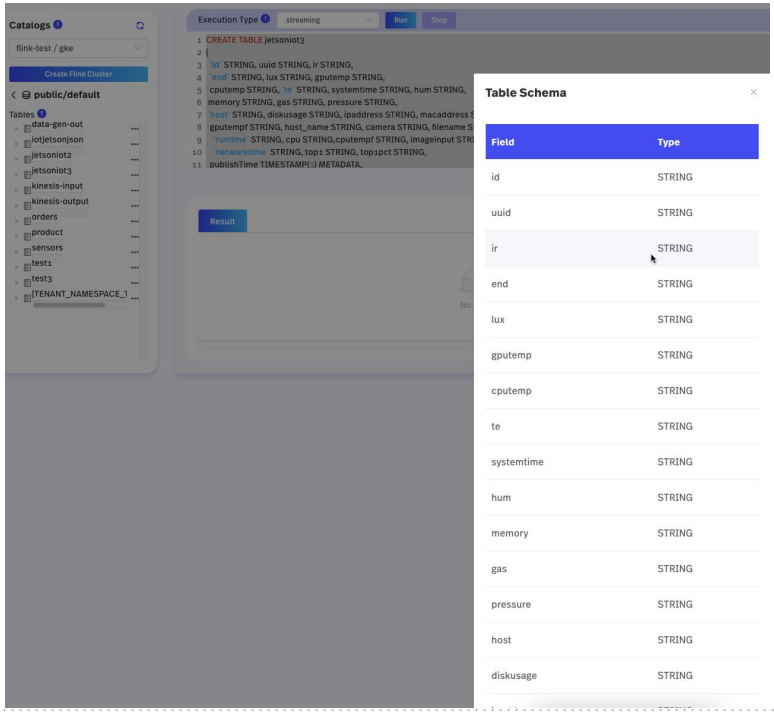
StreamNative Cloud



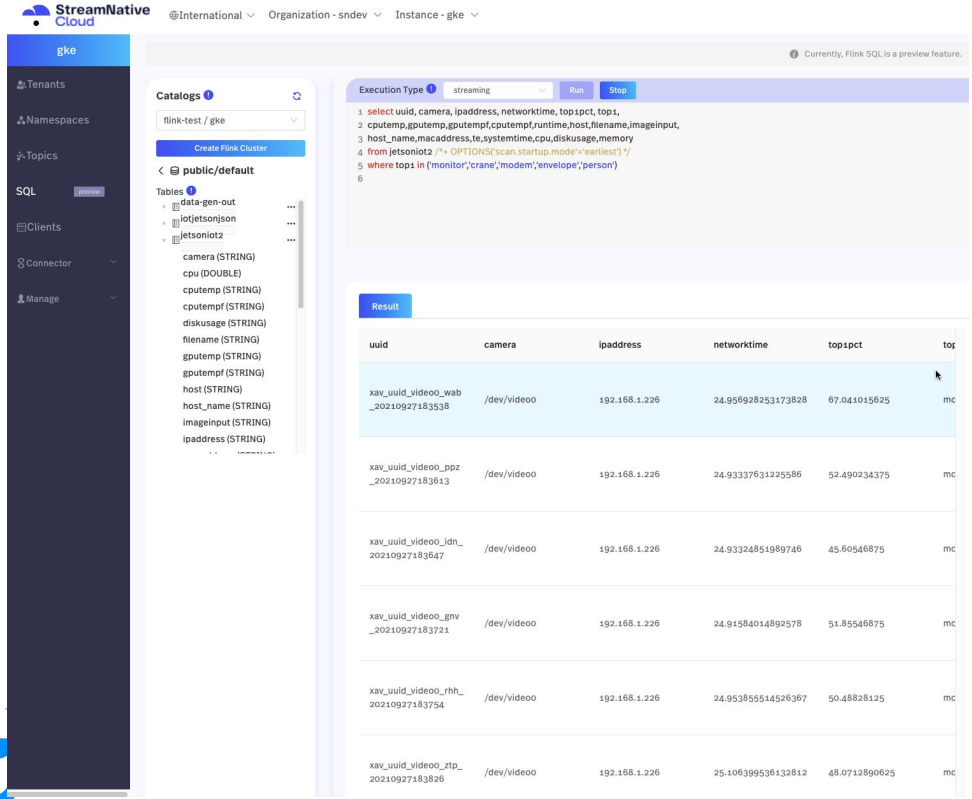
StreamNative Cloud



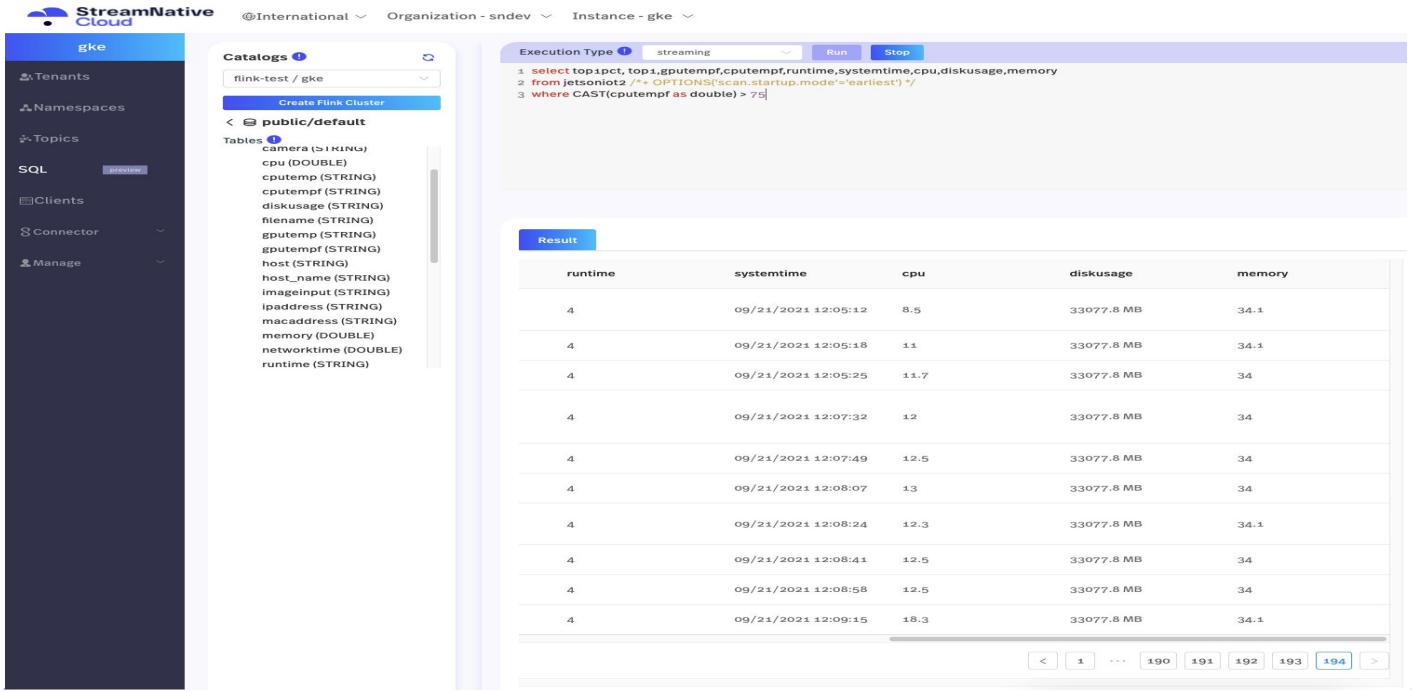
StreamNative Cloud



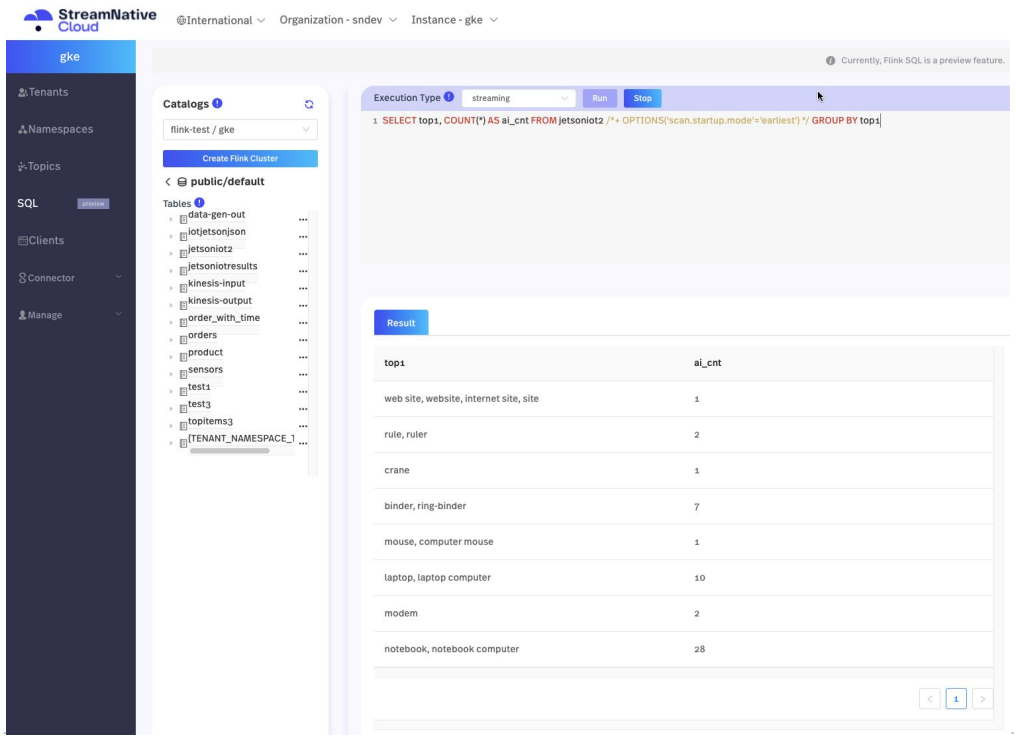
StreamNative Flink SQL



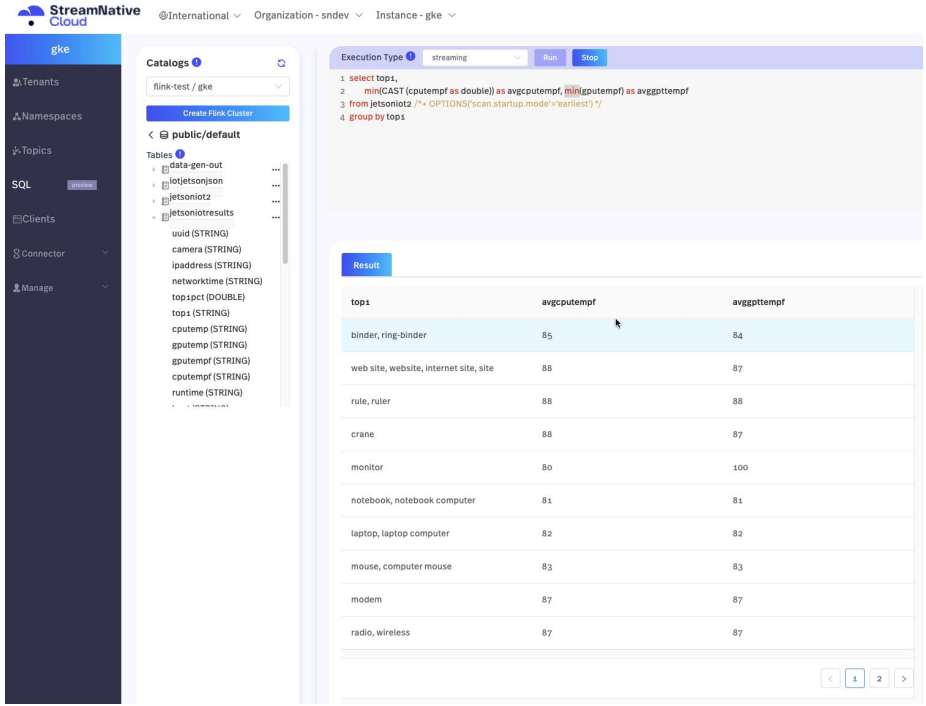
StreamNative Flink SQL



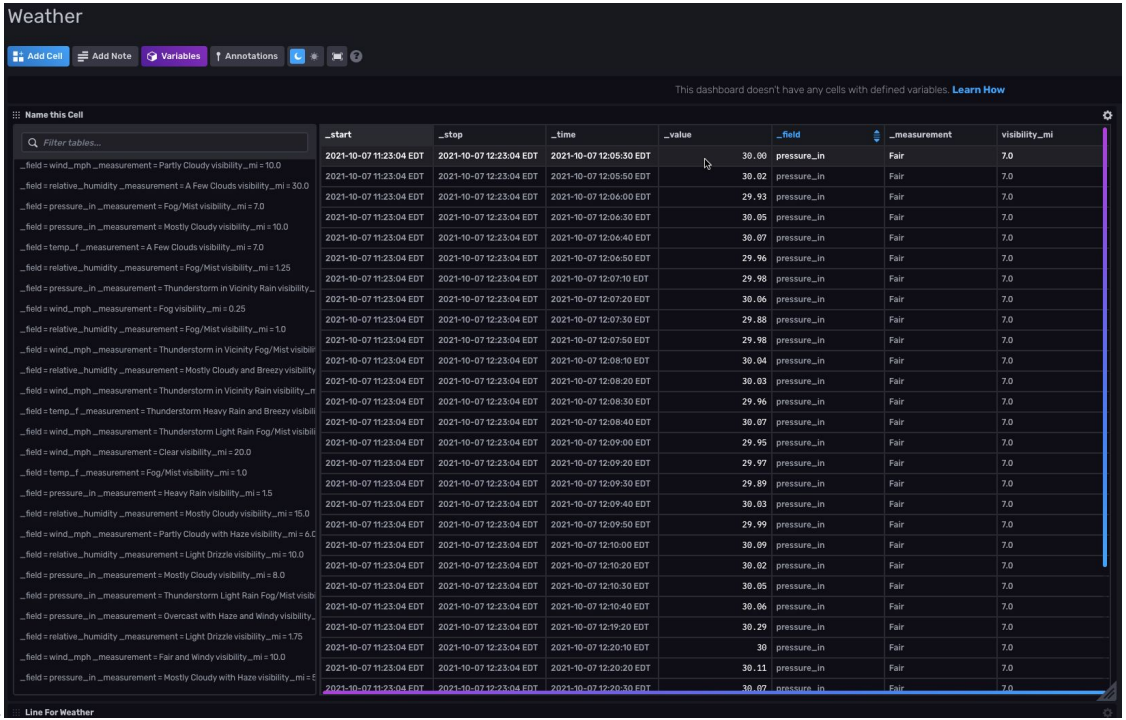
StreamNative Flink SQL



StreamNative Flink SQL

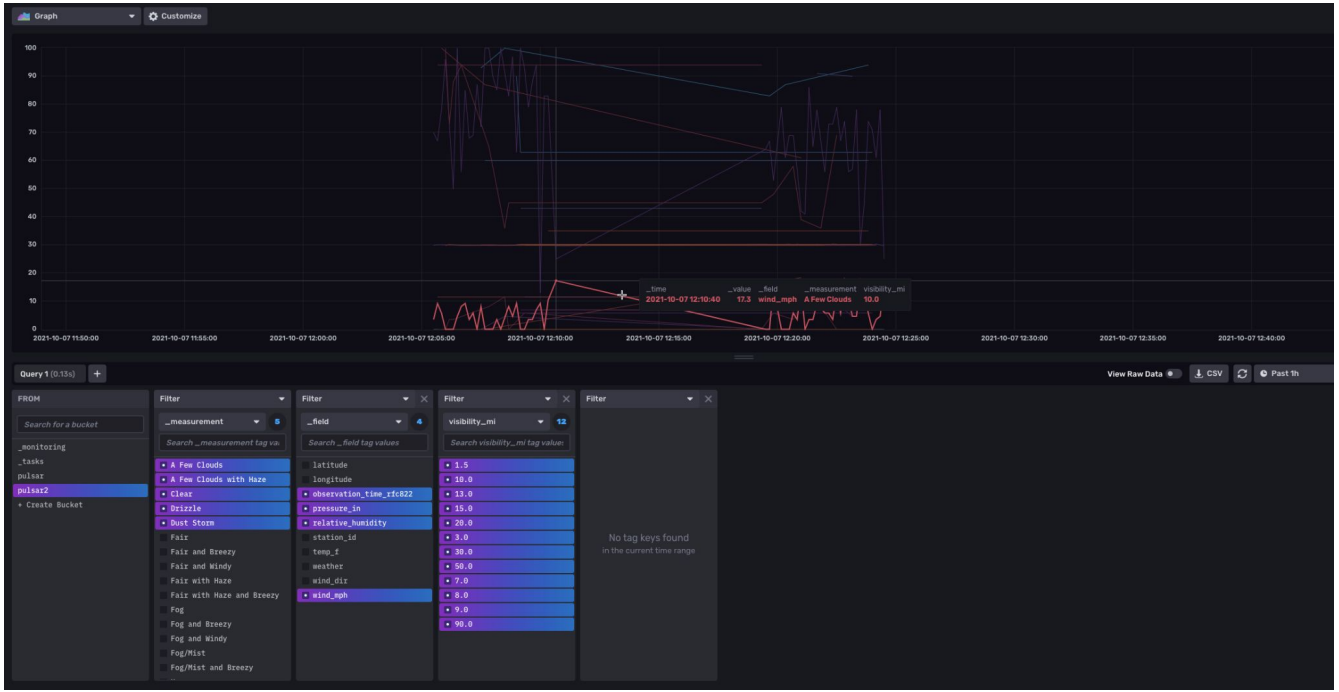


Influx Data Explorer



_start	_stop	_time	_value	_field	_measurement	visibility_mi
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:05:30 EDT	30.00	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:05:50 EDT	30.02	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:06:00 EDT	29.93	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:06:30 EDT	30.05	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:06:40 EDT	30.07	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:06:50 EDT	29.96	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:07:10 EDT	29.98	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:07:20 EDT	30.06	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:07:30 EDT	29.88	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:07:50 EDT	29.98	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:08:10 EDT	30.04	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:08:20 EDT	30.03	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:08:30 EDT	29.96	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:08:40 EDT	30.07	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:09:00 EDT	29.95	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:09:20 EDT	29.97	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:09:30 EDT	29.89	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:09:40 EDT	30.03	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:09:50 EDT	29.99	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:10:00 EDT	30.09	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:10:20 EDT	30.02	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:10:30 EDT	30.05	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:10:40 EDT	30.06	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:19:20 EDT	30.29	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:20:10 EDT	30	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:20:20 EDT	30.11	pressure_in	Fair	7.0
2021-10-07 11:23:04 EDT	2021-10-07 12:23:04 EDT	2021-10-07 12:20:30 EDT	30.07	pressure_in	Fair	7.0

Influx Data Explorer



Web Socket Visualization

StreamNative

User: Data Engineer 7

Question: Why is Pulsar and Scylla together such a great idea?

Contact Info: Data Engineer at Fortune 500 Company

Send to Pulsar

Show 10 entries Search:

Sentiment	Publish Time	Message	User	Contact Info
Positive	2021-12-14T14:36:35.868-05:00	Why is Pulsar and Scylla together such a great idea?	Data Engineer 7	Data Engineer at Fortune 500 Company
Positive	2021-12-14T14:36:34.041-05:00	Why is Pulsar and Scylla together such a great idea?	Data Engineer 7	Data Engineer at Fortune 500 Company
Positive	2021-12-14T14:16:55.776-05:00	We all know that Apache Pulsar along with Apache Flink, Apache Spark, Apache NiFi and Scylla DB are the best options for modern streaming data applications.	Tim Spann	Tim Spann - Developer Advocate @ StreamNative
Positive	2021-12-14T14:16:34.858-05:00	We all know that Apache Pulsar along with Apache Flink, Apache Spark, Apache NiFi and Scylla DB are the best options for modern streaming data applications.	Tim Spann	Tim Spann, Developer Advocate @ StreamNative
Positive	2021-12-14T14:16:32.717-05:00	We all know that Apache Pulsar along with Apache Flink, Apache Spark, Apache NiFi and Scylla DB are the best options for modern streaming data applications.	Tim Spann	Tim Spann, Developer Advocate @ StreamNative
Neutral	2021-12-14T09:53:12.787-05:00	this is a test	Porto	Porto
Positive	2021-12-14T09:51:38.726-05:00	pulsar and flink with spark are amazing	Tim Spann	Tim @ Data Science Camp
Positive	2021-12-14T09:50:50.835-05:00	pulsar and flink with spark are amazing	Tim Spann	Tim @ Data Science Camp
Neutral	2021-12-14T09:43:27.878-05:00	test	Tim Spann	test
Positive	2021-12-14T09:16:29.737-05:00	Why is Apache Pulsar so amazing?	Tim Spann	Tim @ StreamNative

Showing 1 to 10 of 23 entries Previous 1 2 3 Next

StreamNative

- https://github.com/tspannhw/FLiP-InfluxDB
- https://pulsar.apache.org/docs/en/io-influxdb-sink/
- https://pulsar.apache.org/docs/en/io-overview/
- https://github.com/tspannhw/Flip-solr
- https://github.com/tspannhw/Flip-transit
- https://github.com/tspannhw/Flip-iot
- https://www.datainmotion.dev/2020/10/flank-streaming-edgeai-on-new-nvidia.html
- https://github.com/tspannhw/minifi-jetson-nano
- https://github.com/tspannhw/SpeakerProfile/
- https://hub.streamnative.io/connectors/influxdb-sink/2.5.1/
- https://github.com/tspannhw/minifi-xaviernx/

Resources

Now Available

On-Demand Pulsar Training

Academy.StreamNative

Live 3-day

Developers Training

Save Your Spot!

Free ebook

Pulsar in Action

Read it now

StreamNative

influxdata

Let’s Keep in Touch!

Timothy Spann
Developer Advocate

@PaasDev

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

https://github.com/tspannhw

StreamNative + influxdata

Questions?