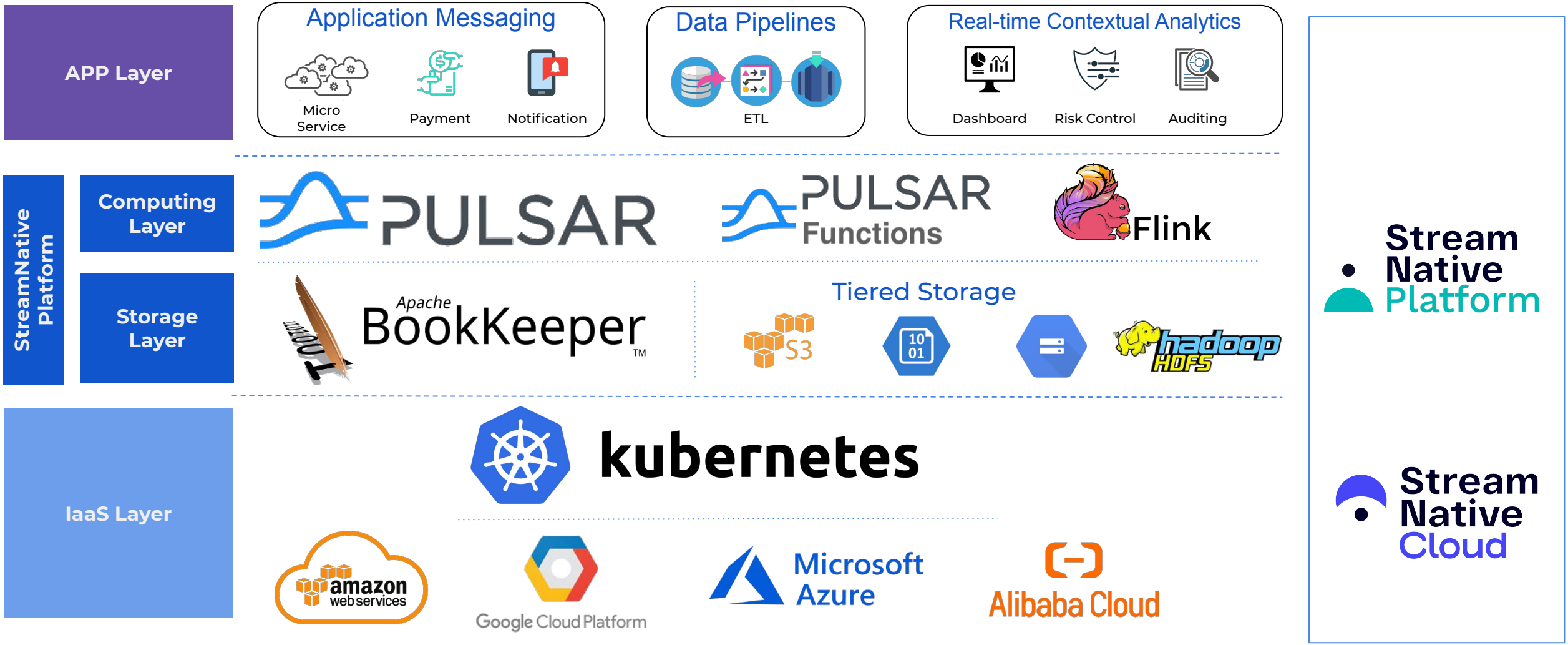
Codeless Pipelines with Pulsar and Flink

Timothy Spann
Developer Advocate

StreamNative Solution



Speaker Bio

Developer Advocate

DZone Zone Leader and Big Data MVB
@PaasDev

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

FLaNK and FLiP Stacks

- Apache **Fl**ink
- Apache **N**iFi
- Apache **K**afka

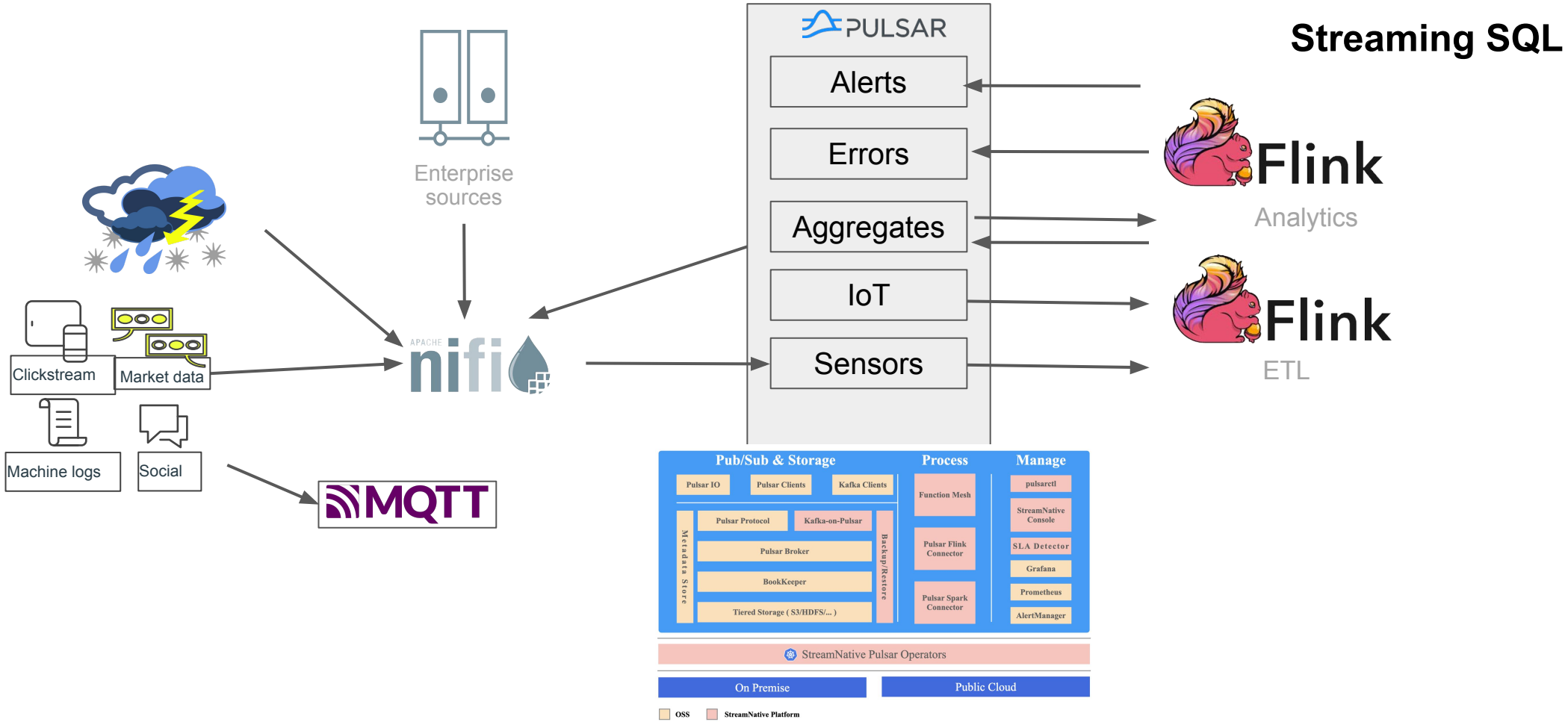
- Apache **Fl**ink
- Apache **P**ulsar
- StreamNative's Flink Connector for Pulsar
- Apache +++

Apache projects are the way for all streaming use cases.

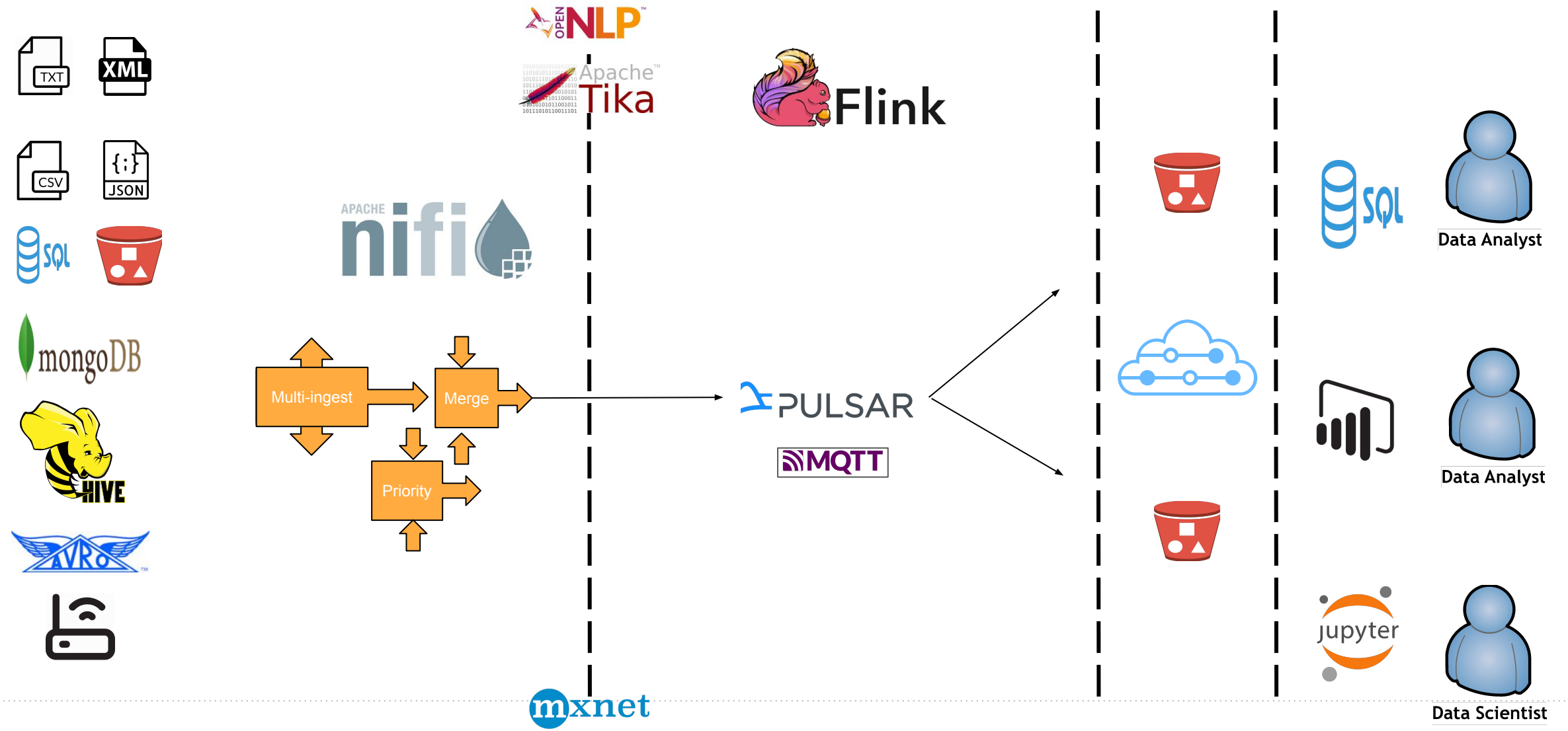
Today's Data. IoT JSON

{"uuid": "rpi4_uuid_vml_20210902151842", "amplitude100": 1.1, "amplitude500": 0.5, "amplitude1000": 0.4, "lownoise": 0.5, "midnoise": 0.2, "highnoise": 0.2, "amps": 0.3, "ipaddress": "192.168.1.244", "host": "rp4", "host_name": "rp4", "macaddress": "3e:f6:33:38:9e:d3", "systemtime": "2021-09-02T11:18:43.765279", "endtime": "1630595923.76", "runtime": "49.71", "starttime": "09/02/2021 11:17:52", "cpu": 0.0, "cpu_temp": "34.0", "diskusage": "37206.6 MB", "memory": 5.5, "id": "20210902151842_b1126bd3-f6e1-4eb1-84d0-bd8557a560cf", "temperature": "22.1", "adjtemp": "16.8", "adjtempf": "42.2", "temperaturef": "51.8", "pressure": 1008.0, "humidity": 37.8, "lux": 156.5, "proximity": 0, "oxidising": 12.4, "reducing": 147.3, "nh3": 25.3, "gasKO": "Oxidising: 12444.44 Ohms\nReducing: 147300.33 Ohms\nNH3: 25266.49 Ohms"}

End to End Streaming Codeless Pipeline



All Data - Anytime - Anywhere - Multi-Cloud - Multi-Protocol



StreamNative

streamnative.io

Python IoT Program Sending Data to MQTT and Pulsar

```
row['gasKO'] = str(readings)
json_string = json.dumps(row)
json_string = json_string.strip()

client.connect("192.168.1.181", 1883, 180)
client.publish("persistent://public/default/mqtt-2", payload=json_string, qos=0, retain=True)

producer = KafkaProducer(bootstrap_servers='192.168.1.181:9092',retries=3)
producer.send('rp4-kafka-1', json.dumps(row).encode('utf-8'))
```

We need to install MQTT, Kafka and Pulsar libraries. You can choose your protocol to communicate with the messaging cluster.

```
pip3 install paho-mqtt
pip3 install kafka-python
pip3 install pulsar-client
```

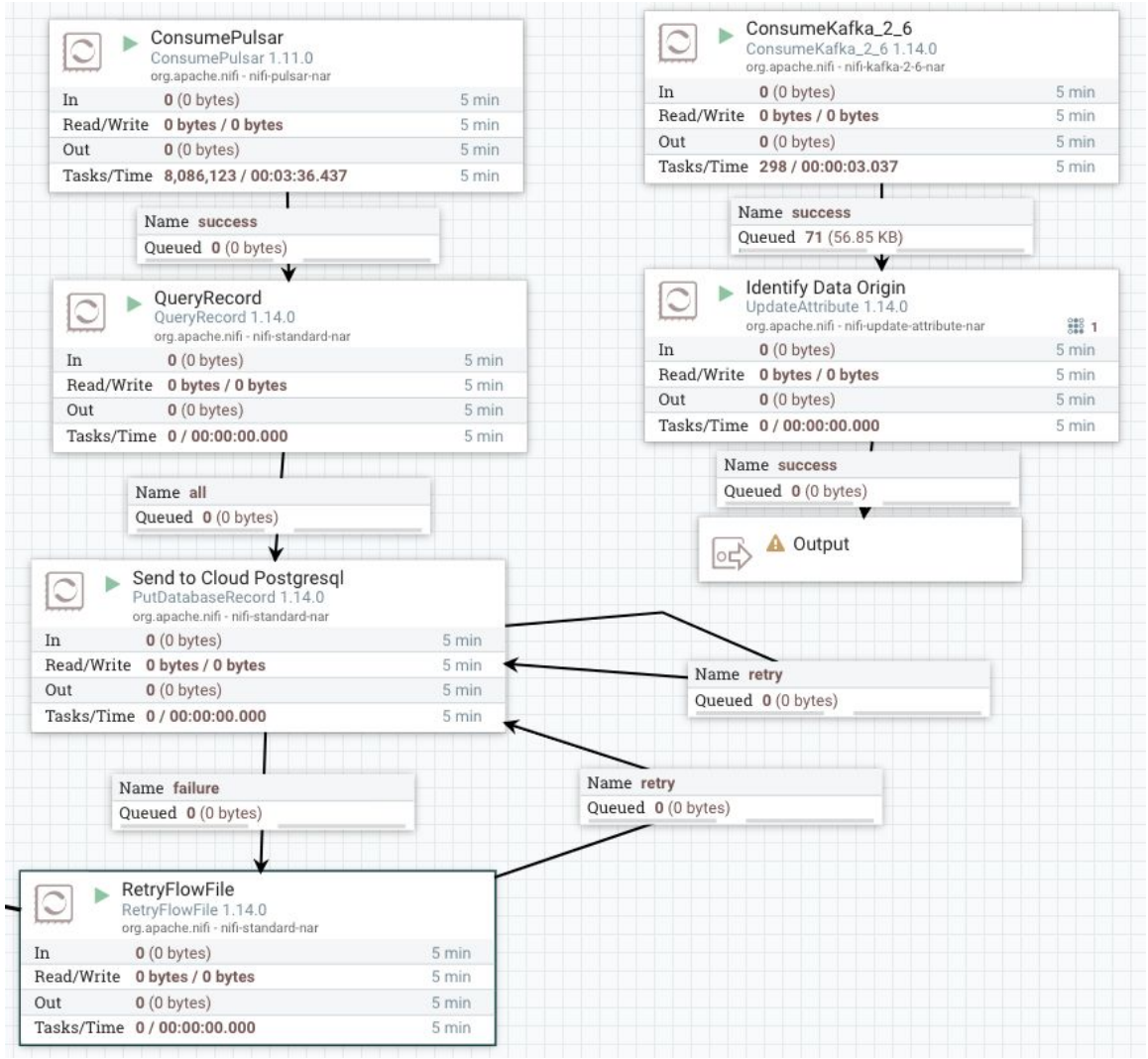
Running This All Yourself

See: https://github.com/tspannhw/FLiP-SQL https://github.com/tspannhw/FLiP-IoT

Run Apache Pulsar Standalone - locally, docker or native cloud (https://console.streamnative.cloud/)
Run Apache Flink Standalone - locally, docker or native cloud (https://console.streamnative.cloud/)
Run Apache NiFi Single - locally, docker or cloud

I run MQTT on Pulsar (MoP) to allow Pulsar to use MQTT protocol (https://github.com/streamnative/mop/releases/tag/v2.8.0.10)
I run Kafka on Pulsar (KoP) to allow Pulsar to use Kafka protocol (https://github.com/streamnative/kop)

Apache NiFi Consuming From Pulsar Cluster via Multiple Protocols

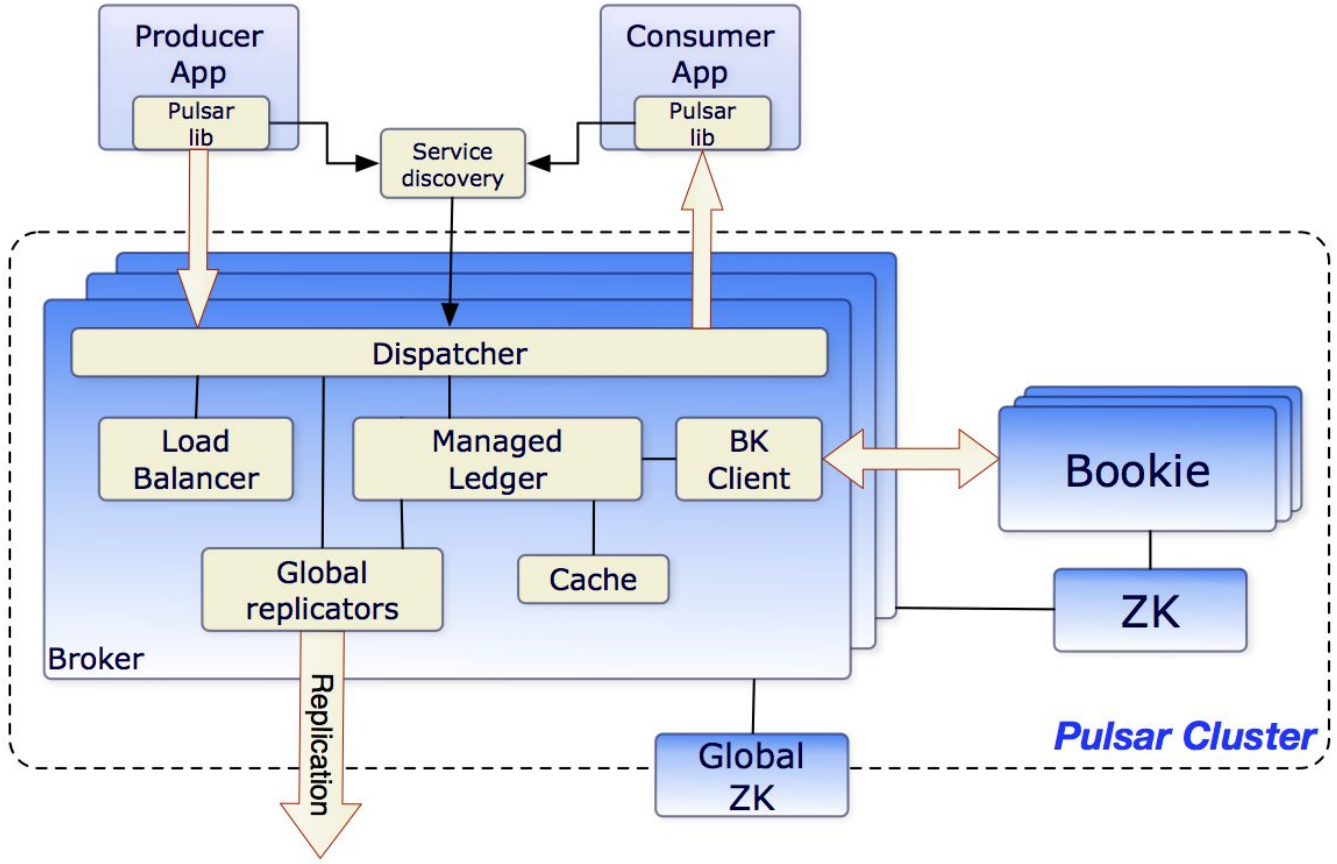


Apache Pulsar is Cloud-Native Messaging and Event-Streaming Platform

Apache Pulsar Overview

Enable Geo-Replicated Messaging

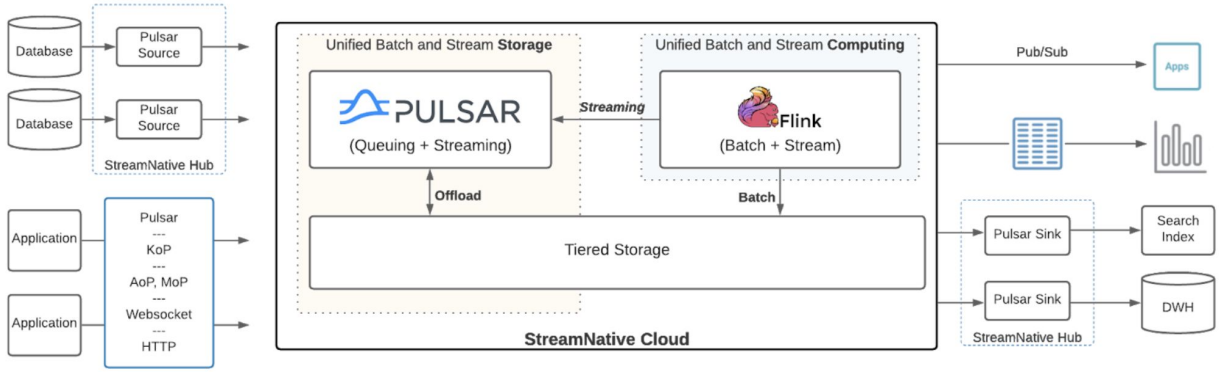
- Pub-Sub
- Geo-Replication
- Pulsar Functions
- Horizontal Scalability
- Multi-tenancy
- Tiered Persistent Storage
- Pulsar Connectors
- REST API
- CLI
- Many clients available
- Four Different Subscription Types
- Multi-Protocol Support
  - MQTT
  - AMQP
  - JMS
  - Kafka
  - ...

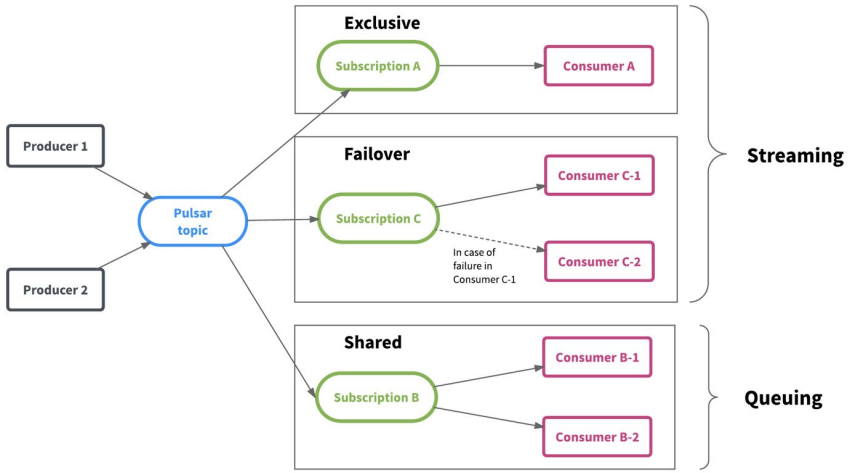


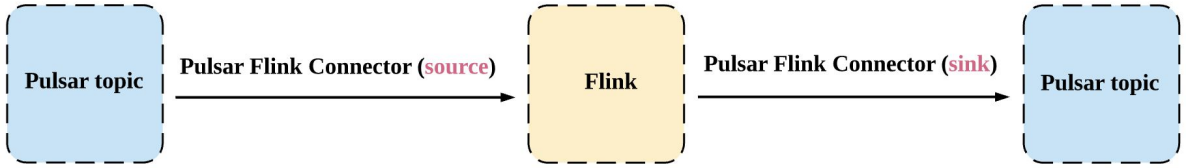
What are the Benefits of Pulsar?

- Multi-Tenancy
- Scalability
- Geo-Replication
- Unified Messaging Model
- Data Durability

Upcoming - 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

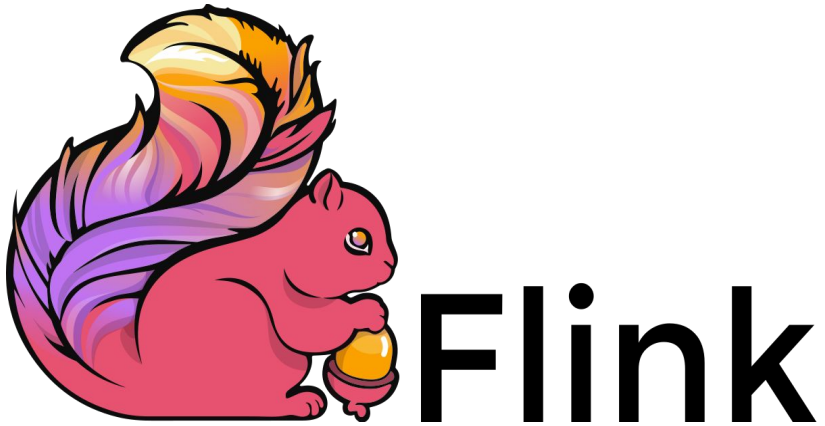
Apache Flink

Apache Flink is a distributed stream processing system.

It is capable of providing high throughput, near real-time processing of streams from Pulsar.

It is ideal for *ambitious* Stream Processing compared to Pulsar's model of lightweight Stream Processing.

End-to-end exactly-once stream processing



https://streamnative.io/en/blog/release/2021-06-14-exactly-once-semantics-with-transactions-in-pulsar

Connect with the Community & Stay Up-To-Date

- Join the Pulsar Slack channel - *Apache-Pulsar.slack.com*
- Follow @streamnativeio and @apache_pulsar on Twitter
- Subscribe to Monthly Pulsar Newsletter for major news, events, project updates, and resources in the Pulsar community

Deeper Content

- https://github.com/tspannhw/FLiP-SQL
- https://github.com/tspannhw/StreamingSQLExamples
- https://github.com/streamnative/pulsar-flink
- https://www.linkedin.com/pulse/2021-schedule-tim-spann/
- https://github.com/tspannhw/SpeakerProfile/blob/main/2021/talks/20210729_HailHydrate!FromStreamtoLake_TimSpann.pdf
- https://streamnative.io/en/blog/release/2021-04-20-flink-sql-on-streamnative-cloud
- https://docs.streamnative.io/cloud/stable/compute/flink-sql

@PaasDev

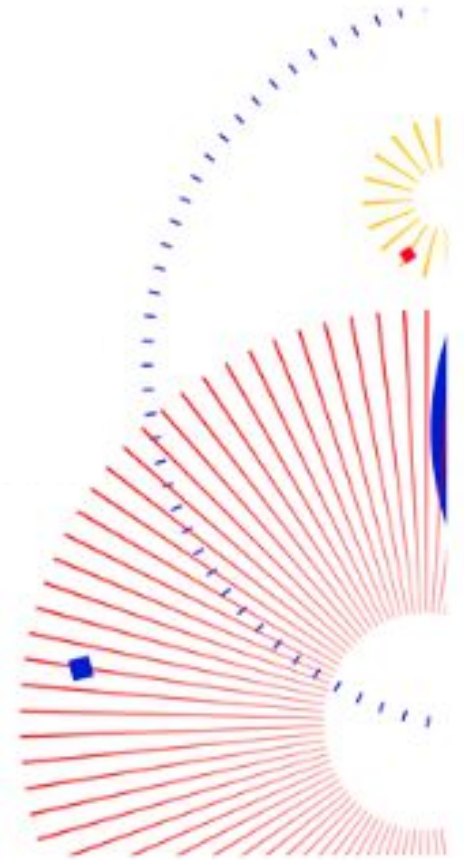
timothyspann

https://www.pulsardeveloper.com/

PULSAR SUMMIT

Pulsar Summit Europe

October 6, 2021



Pulsar Summit Asia

November 20-21, 2021

Contact us at partners@pulsar-summit.org to become a sponsor or partner

Announcing

Flink SQL on StreamNative Cloud