

Hello Hydrate! From Stream to Lake

Timothy Spann

Developer Advocate @ StreamNative



#AutomationSummit





Tim Spann
Developer Advocate

DZone Zone Leader and Big Data
MVB Data DJay

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



#AutomationSummit

AUTOMATION
DEVOPS

SUMMIT



Stream Native

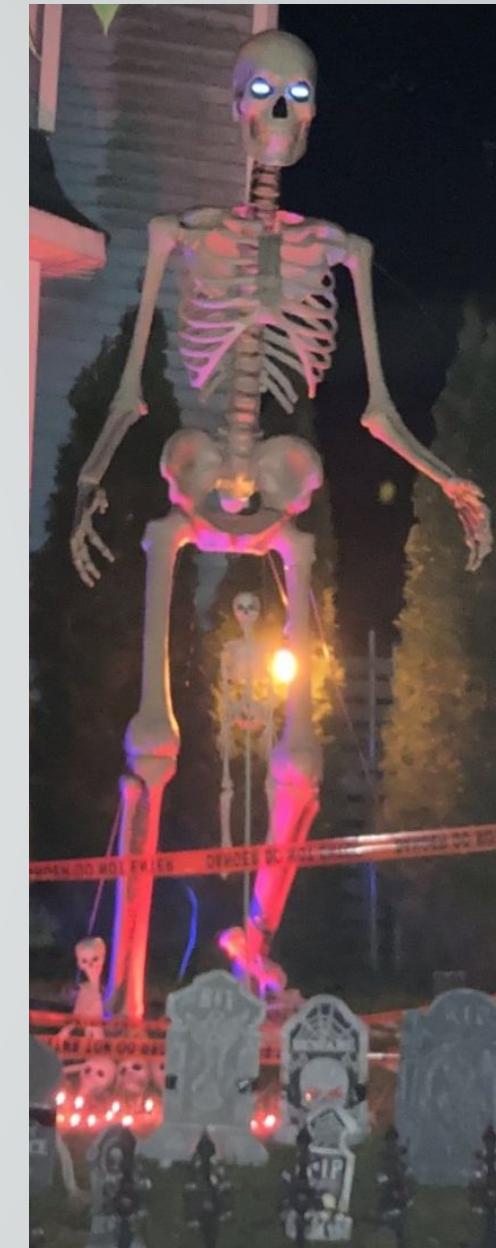
Founded by the original developers of Apache Pulsar and Apache BookKeeper, StreamNative builds a cloud-native event streaming platform that enables enterprises to easily access data as real-time event streams.



#AutomationSummit

AUTOMATION
DEVOPS
 SUMMIT

Don't Be Afraid of Open Source



#AutomationSummit





#AutomationSummit



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



#AutomationSummit

AUTOMATION
DEVOPS

SUMMIT

what are the Benefits of Pulsar?



Multi-Tenancy

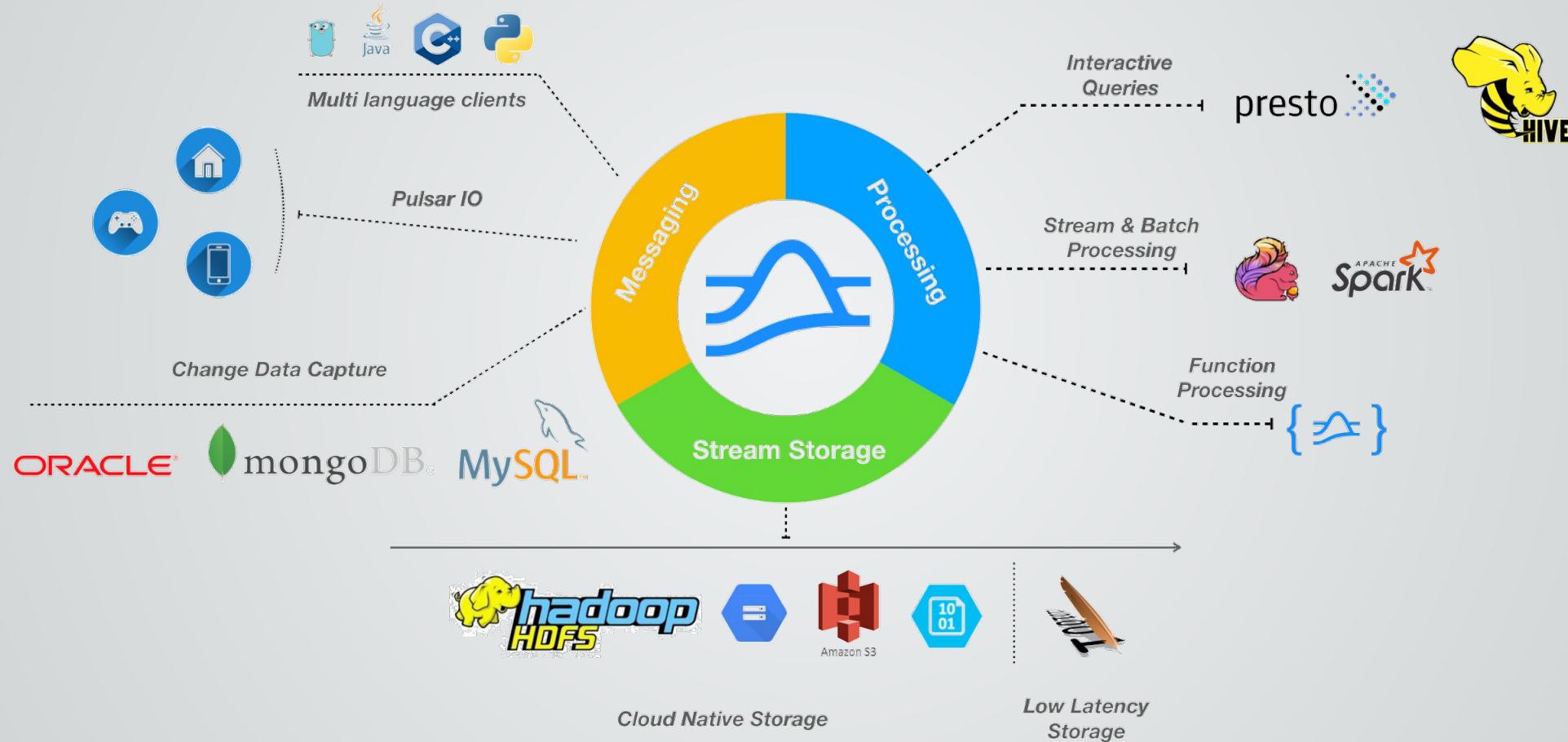
Scalability

Geo-Replication

Unified Messaging
Model

Data Durability

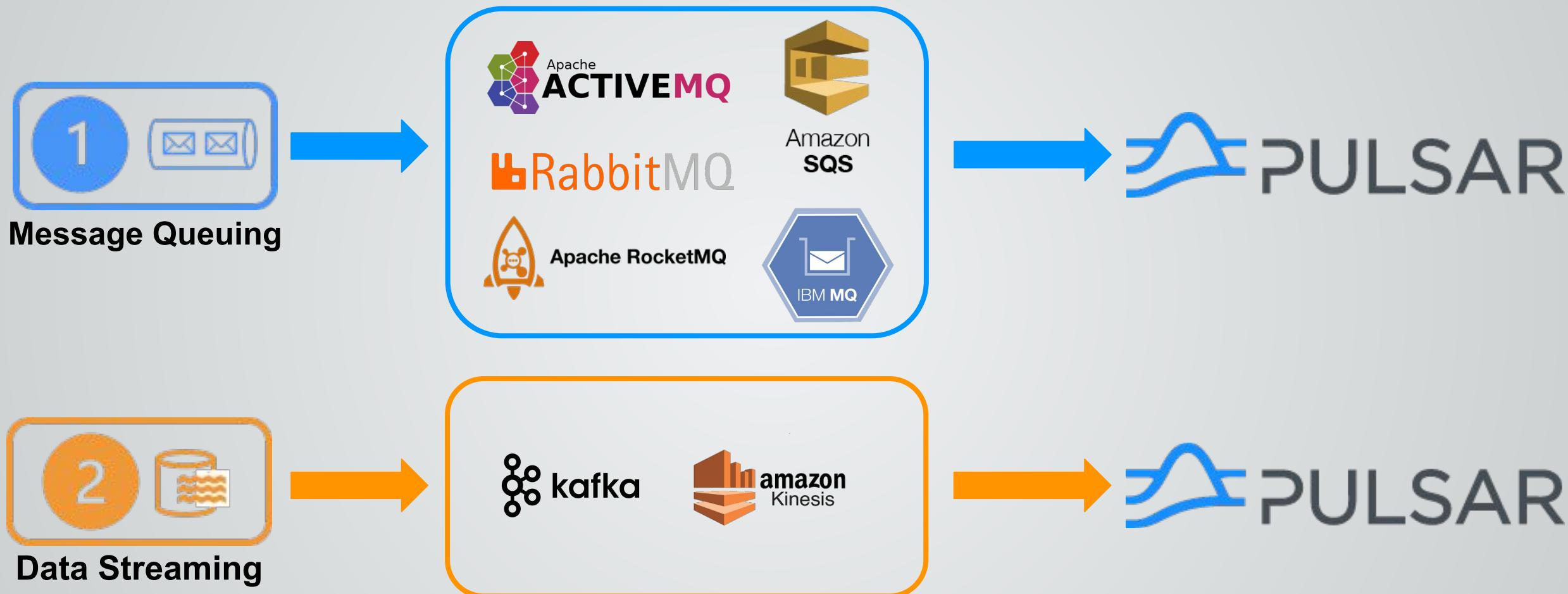
Apache Pulsar



#AutomationSummit

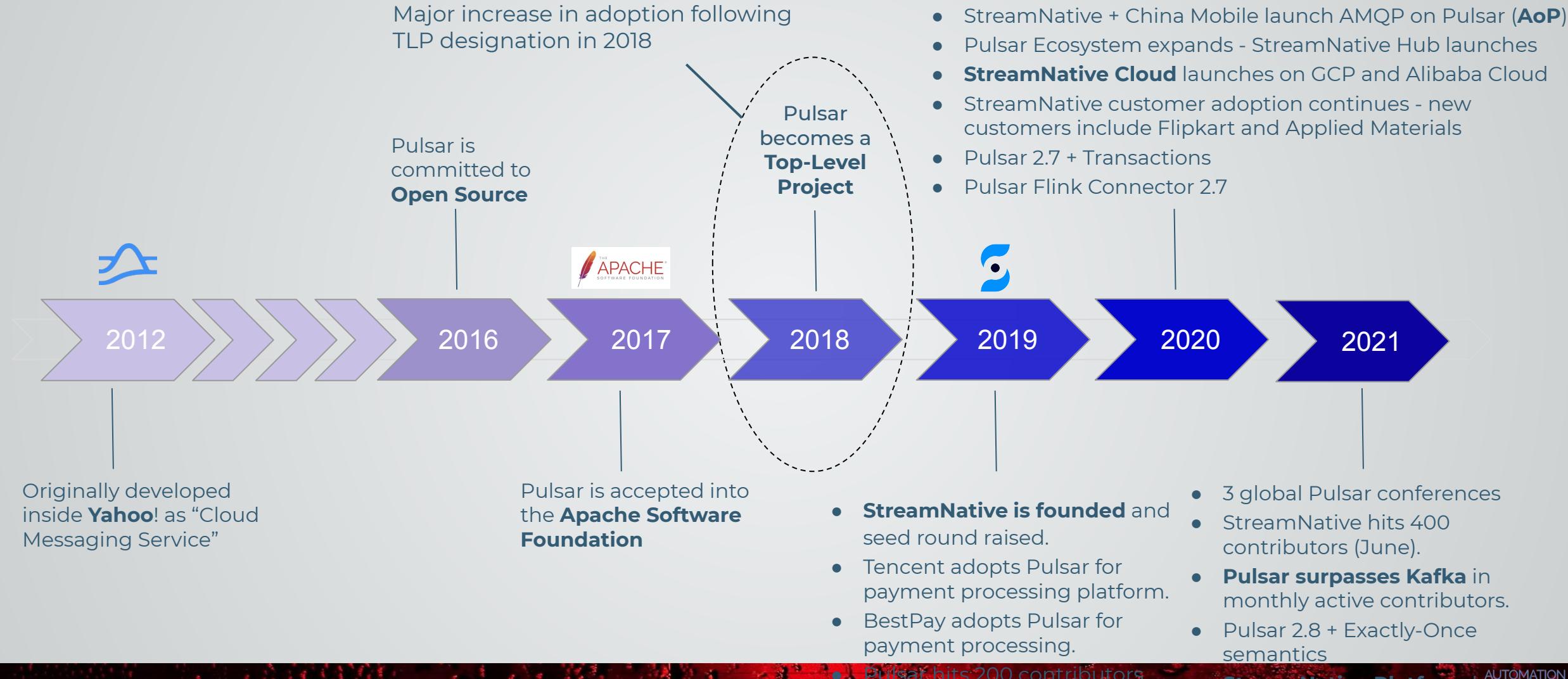


A Unified Messaging Platform



#AutomationSummit

Key Milestones



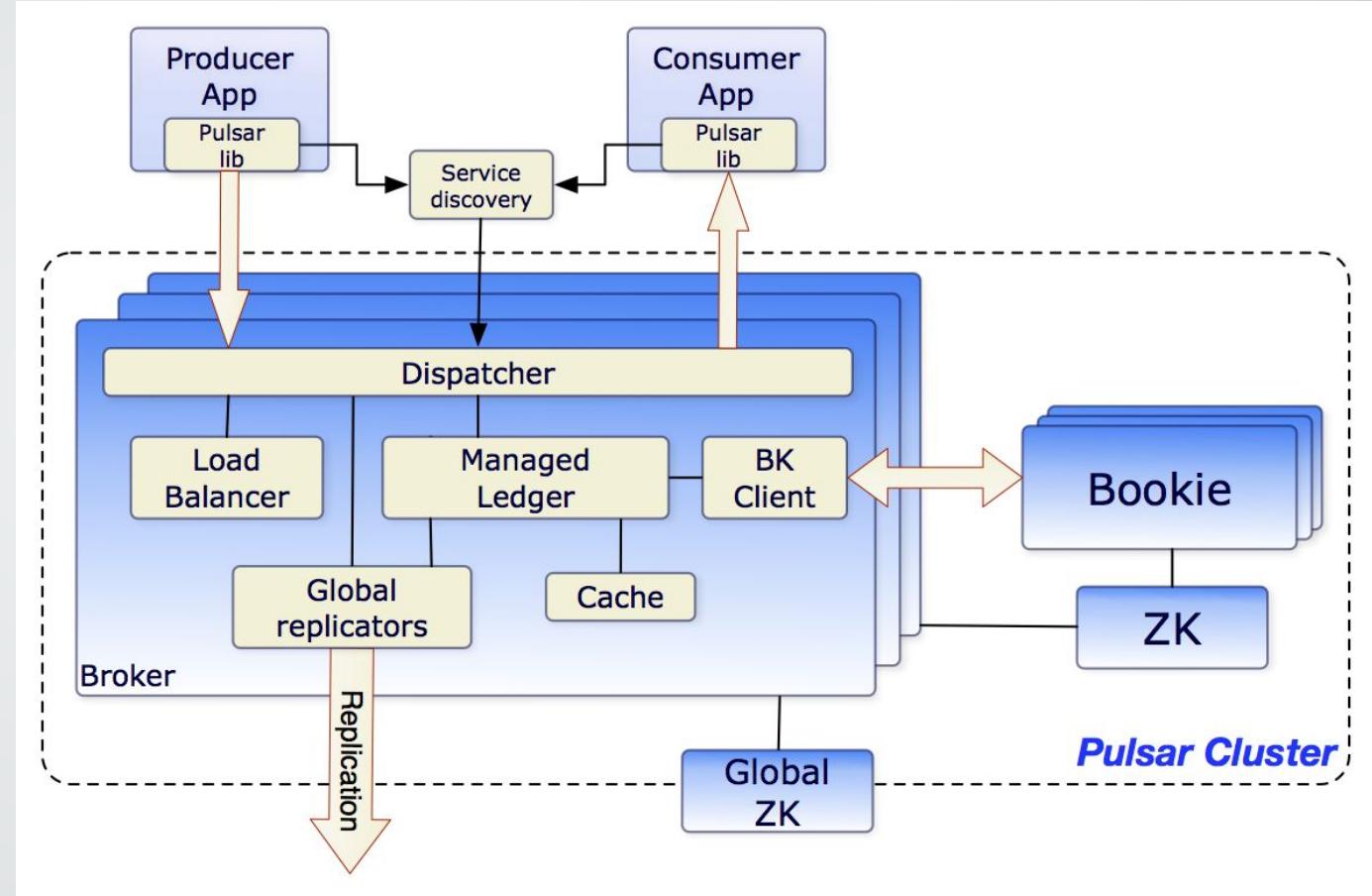
#AutomationSummit



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
 - ...

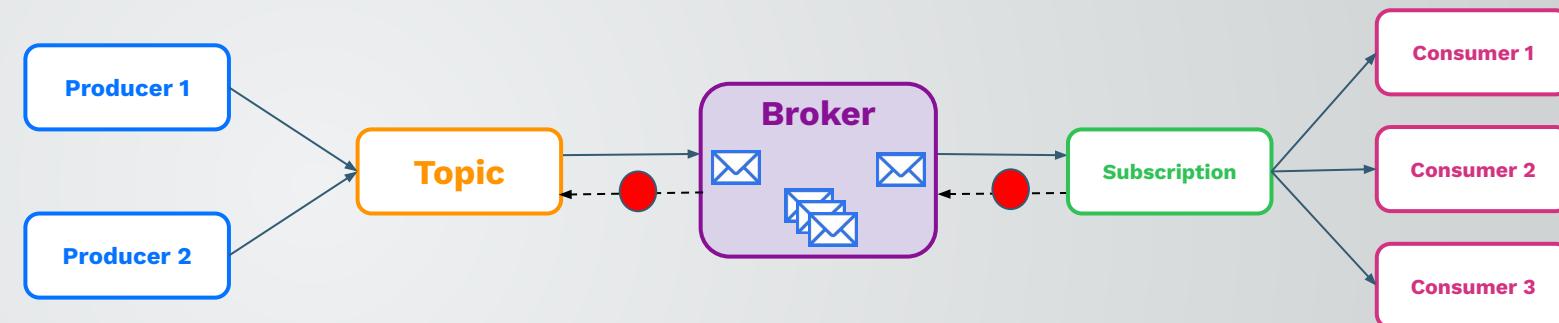


#AutomationSummit



Pulsar's Publish-Subscribe model

- **Producers** send messages.
- **Topics** are an ordered, named channel that producers use to transmit messages to subscribed consumers.
- **Messages** belong to a topic and contain an arbitrary payload.
- **Brokers** handle connections and routes messages between producers / consumers.
- **Subscriptions** are named configuration rules that determine how messages are delivered to consumers.
- **Consumers** receive messages.



What is the Pulsar Ecosystem?

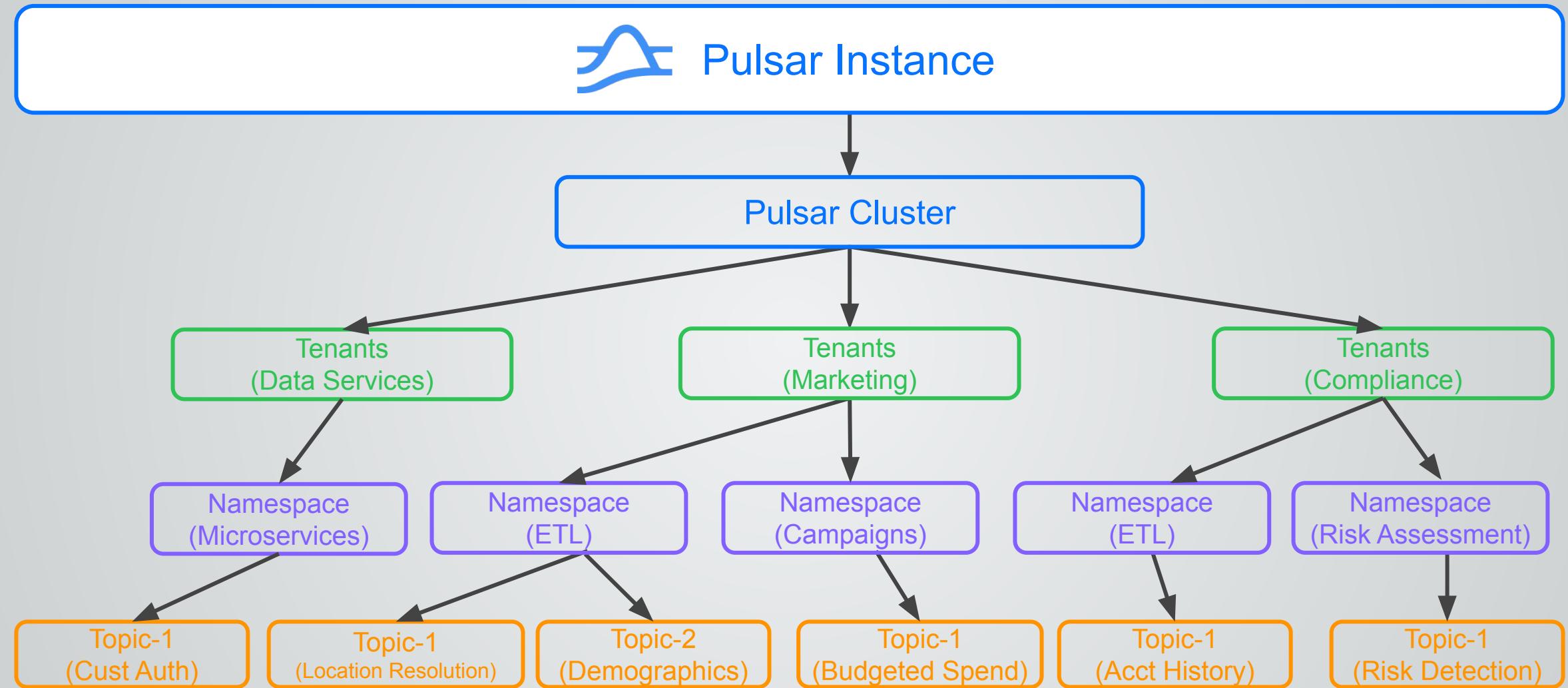
- **Functions and Connectors**
 - Functions: Lightweight stream processing
 - Connectors: Part of “Pulsar IO”, includes “Source” and “Sink” APIs
 - Files, Databases, Data tools, Cloud Services, etc
- **Protocol Handlers**
 - Allows Pulsar to handle additional protocols by an extendable API running in the broker
 - AoP (AMQP), KoP (Kafka), MoP (MQTT)



#AutomationSummit



Topics



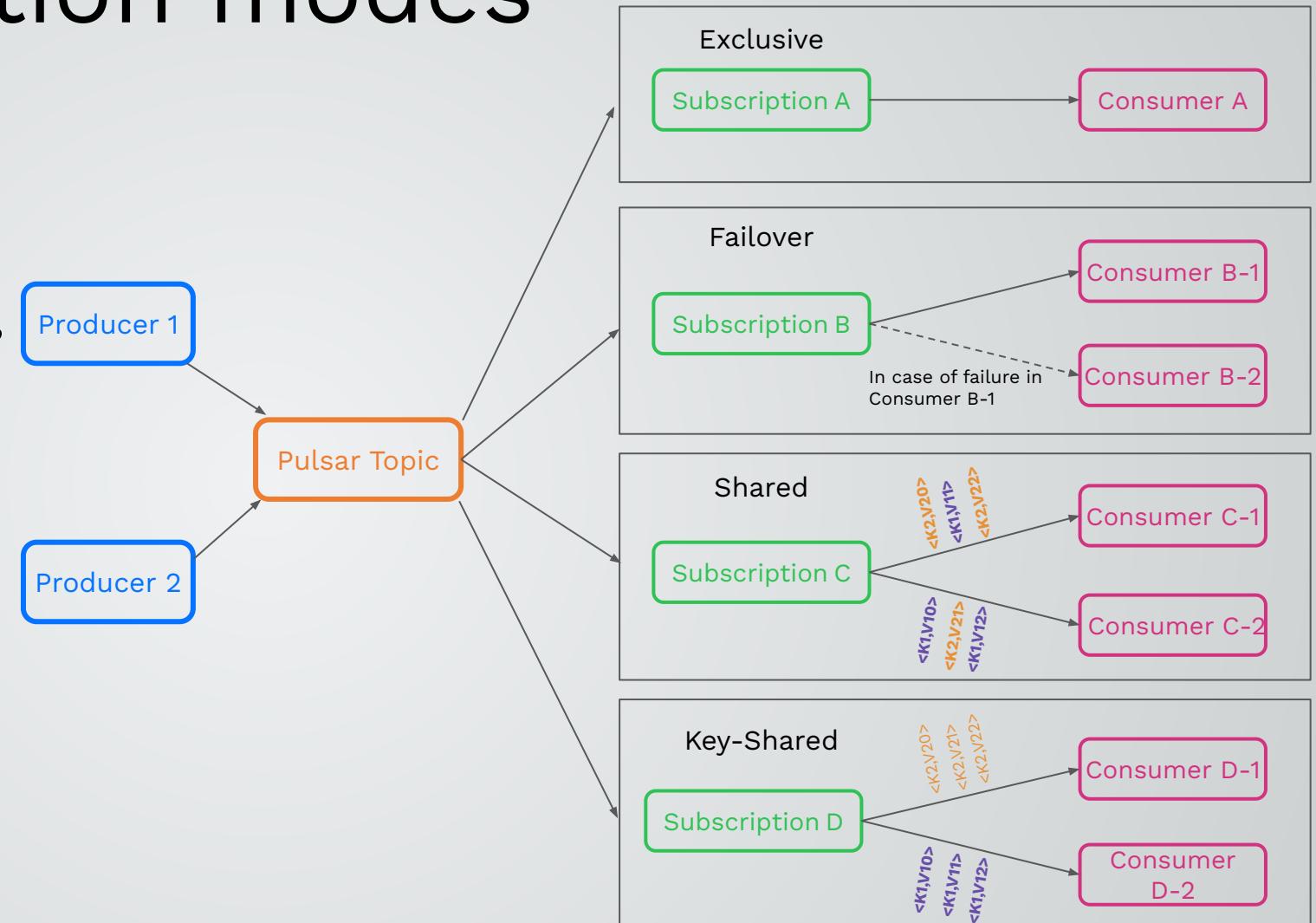
Pulsar subscription modes

Different subscription modes have different semantics:

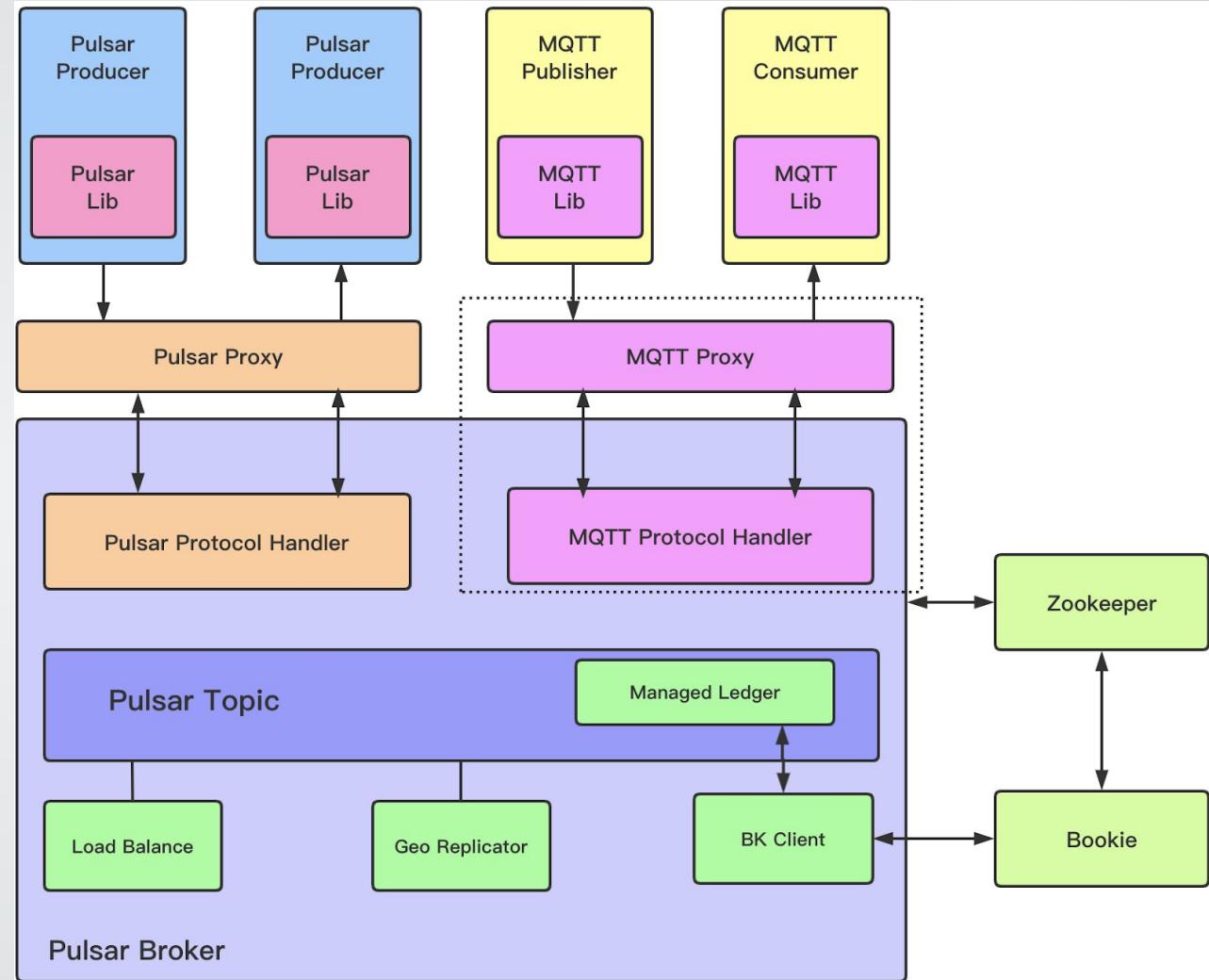
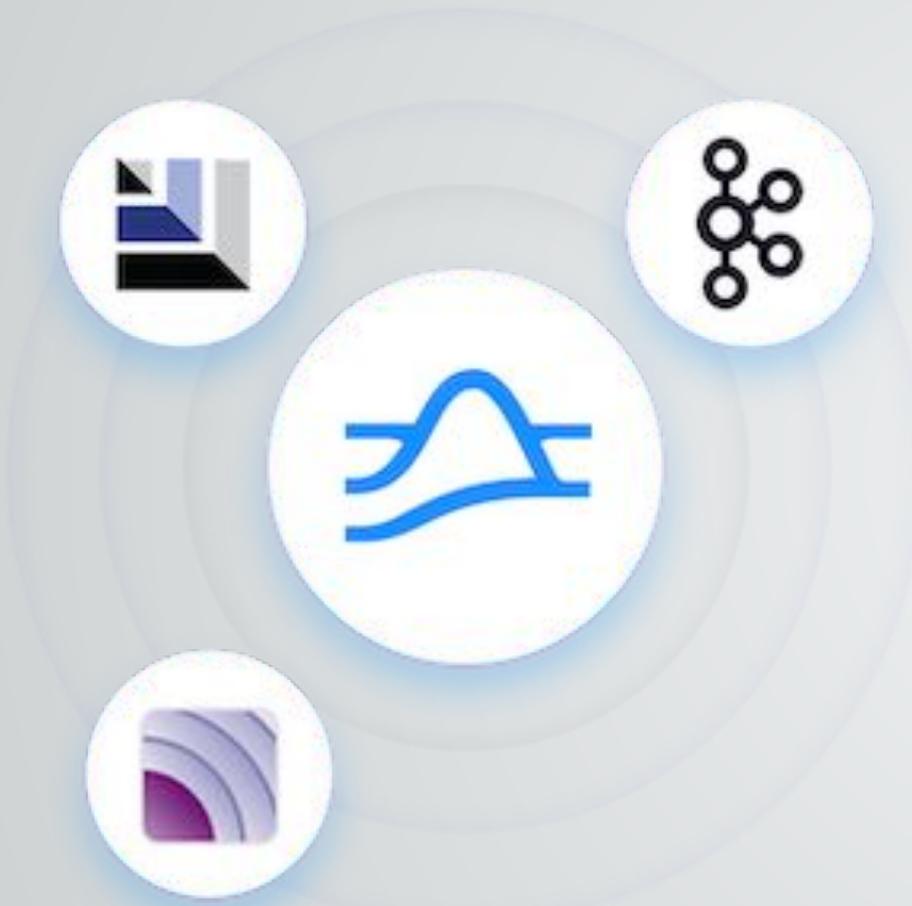
Exclusive/Failover - guaranteed order, single active consumer

Shared - multiple active consumers, no order

Key_Shared - multiple active consumers, order for given key

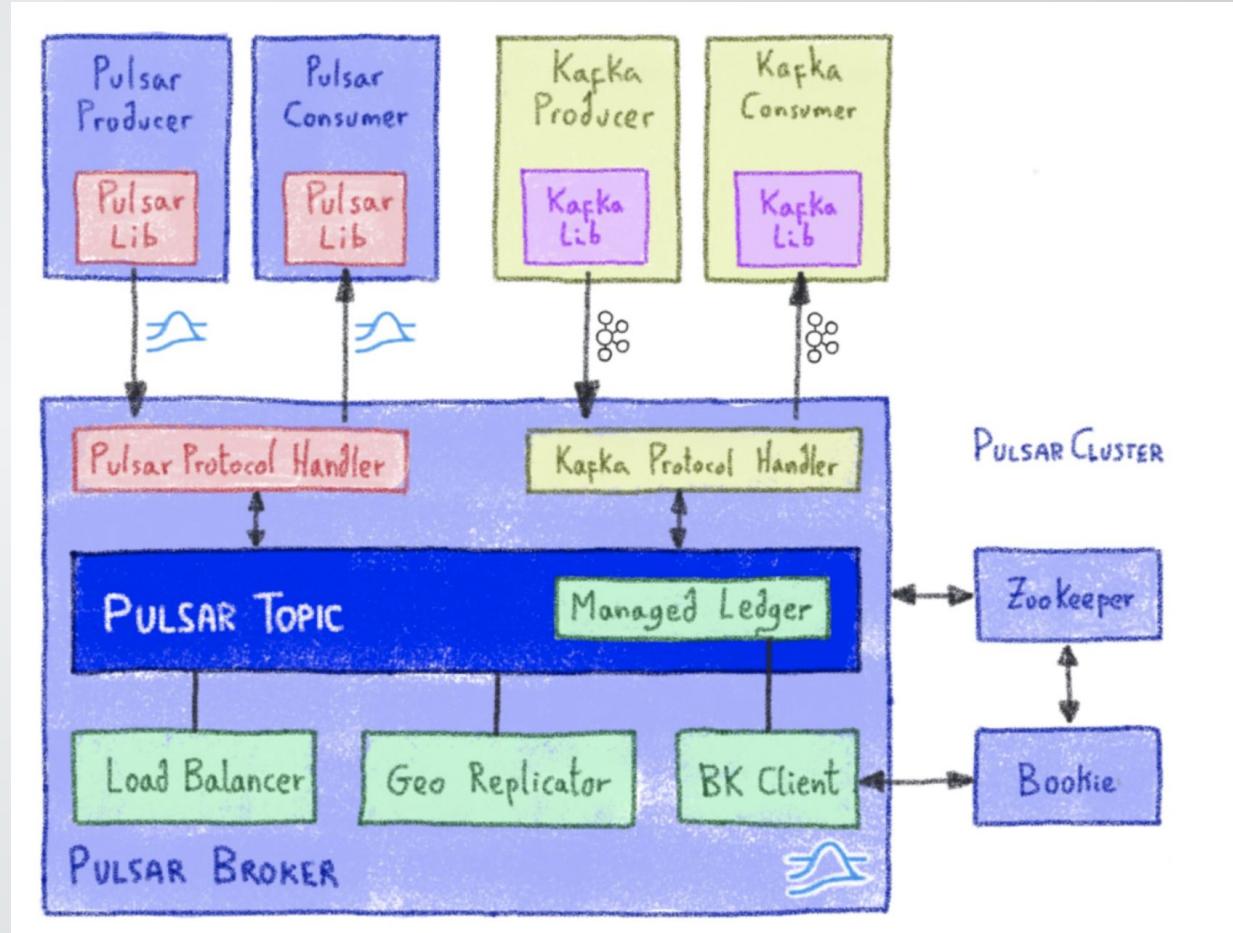
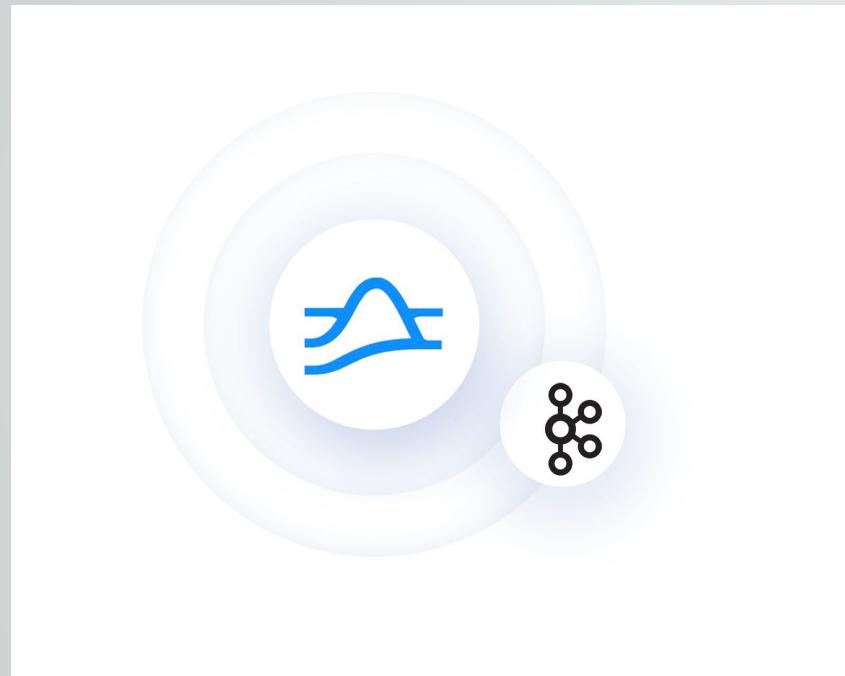


MQTT on Pulsar (MoP)



#AutomationSummit

Kafka-on-Pulsar (Kop)



#AutomationSummit

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



#AutomationSummit

Tencent

- Customer 360 analytics
- Payment reconciliation
- Security analytics
- Adoptions are growing across different business units, such as Gaming, Wechat, Music, etc.



BIGO LIVE

- ETL
- Recommendation
- Data Analysis
- Real-time data warehouse

 StreamNative

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

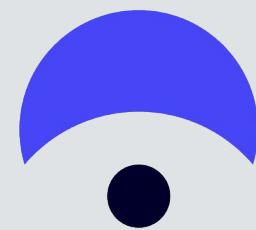


#AutomationSummit

AUTOMATION
DEVOPS

SUMMIT

StreamNative Cloud



Stream Native Cloud



#AutomationSummit



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



kubernetes

Built for Containers



Flink

Flink SQL

The unified messaging and streaming platform made by the creators of Apache Pulsar.

Built for Kubernetes. Made for the cloud. Enables multi-cloud and hybrid

Take A Tour

Contact Sales

Product Update

StreamNative Cloud on AWS Marketplace. Leverage Pulsar on the largest cloud provider with StreamNative

StreamNative, Powered by Apache Pulsar

StreamNative
Cloud

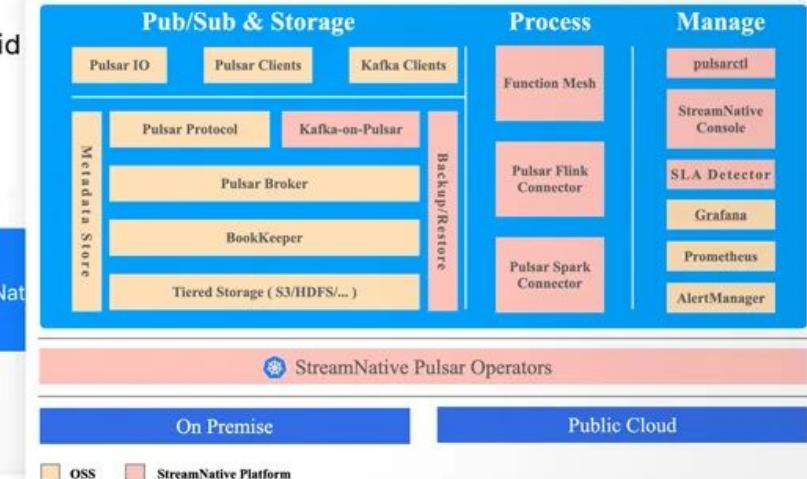
Apache Pulsar as a service, StreamNative Cloud delivers a resilient and scalable messaging and event streaming service deployable in minutes.

StreamNative
Platform

StreamNative Platform is a cloud-native messaging and event streaming platform built by the original creators for Apache Pulsar.

StreamNative
Pro Services

Accelerate your messaging and streaming platform development and drive business results with help from StreamNative's Pulsar experts.



BOSSIE

2021 AWARDS

InfoWorld

#AutomationSummit



Best Practice Architectures

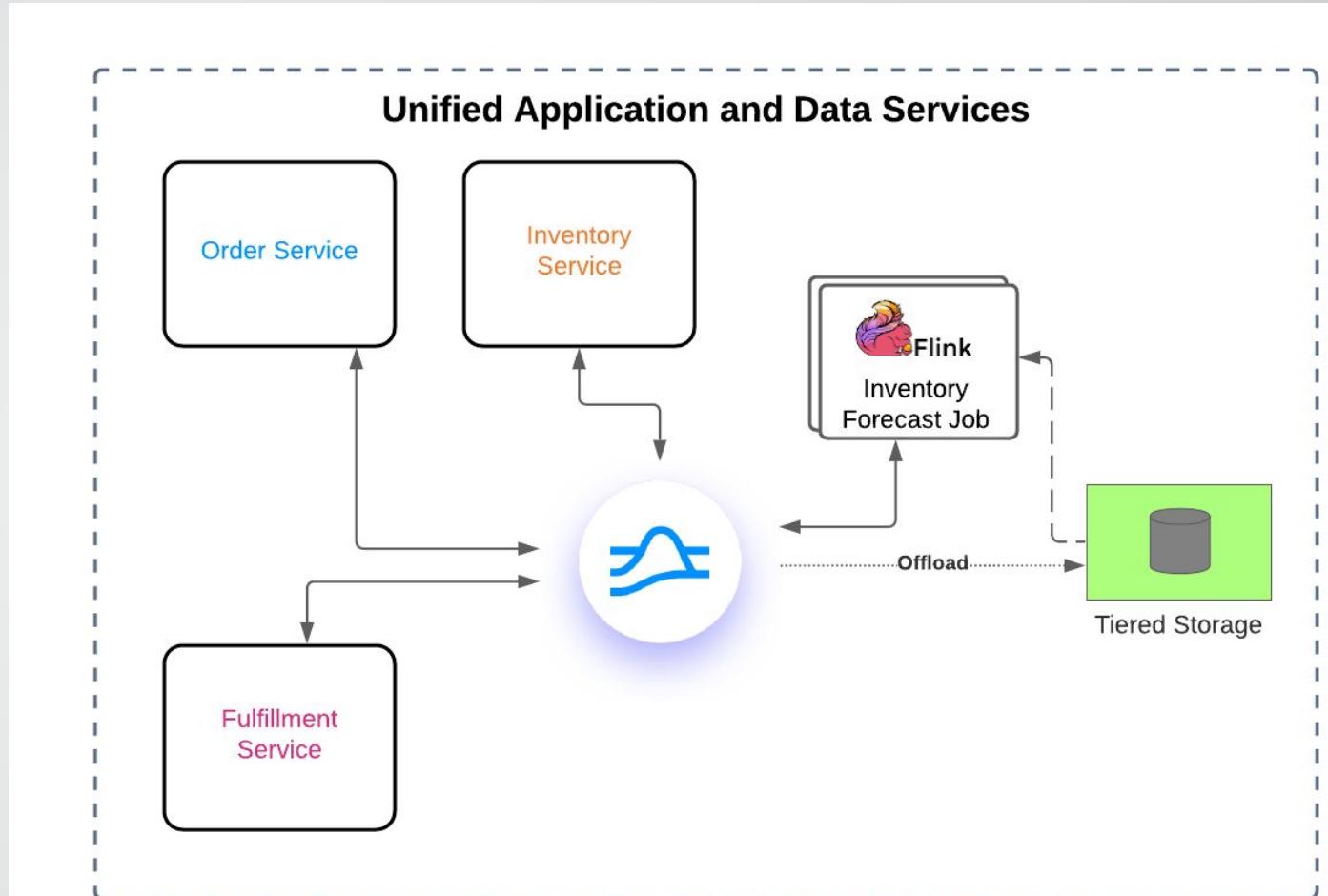


#AutomationSummit



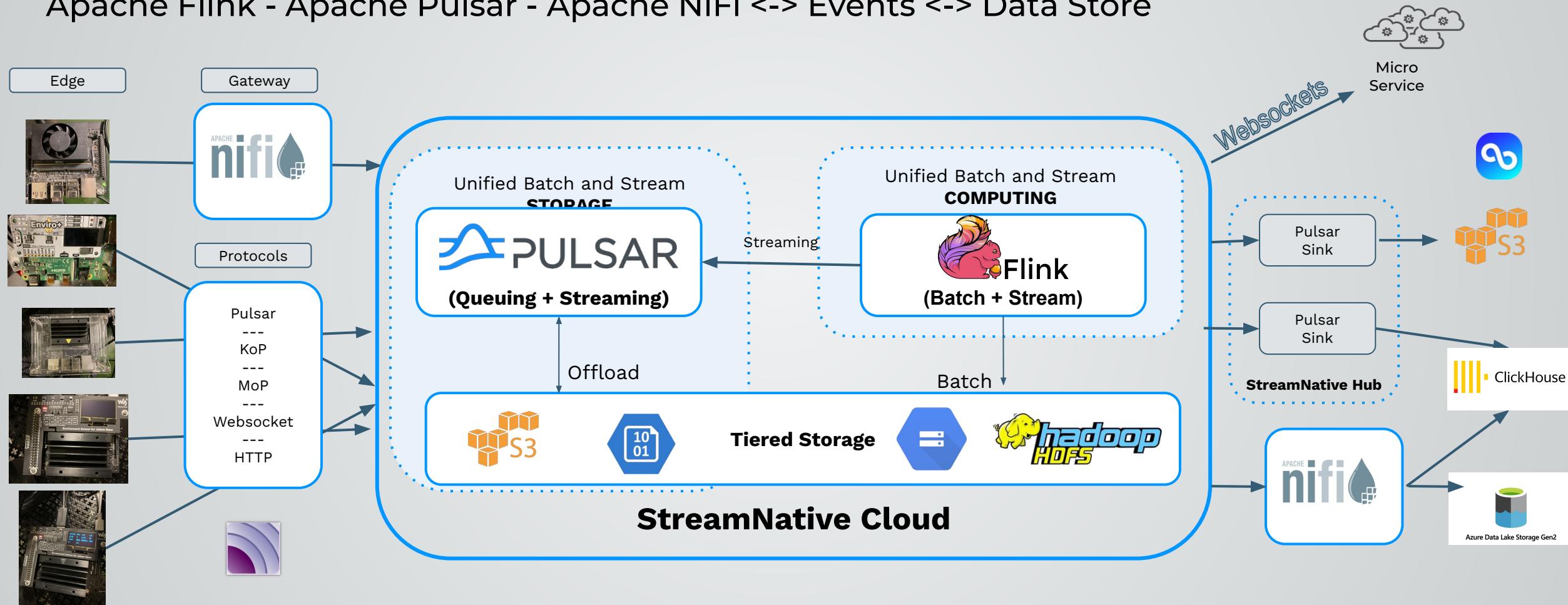
Example: E-Commerce with Pulsar

- Unified storage with access to underlying data
- Native tiered storage
- Single system to exchange data
- Teams share toolset



End-to-End Streaming FLiP(N) Apps

Apache Flink - Apache Pulsar - Apache NiFi <-> Events <-> Data Store



#AutomationSummit



Demo



#AutomationSummit



Wrap-Up



#AutomationSummit



Connect with the Community & Stay Up-To-Date

- Join the Pulsar Slack channel - 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



#AutomationSummit



Deeper Content

- <https://www.datainmotion.dev/2020/04/building-search-indexes-with-apache.html>
- <https://github.com/tspannhw/nifi-solr-example>
- <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/>



#AutomationSummit



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

[11/18] [Developer Week Austin](#)

[11/19] [Porto Tech Hub Con](#)

[12/3] [Data Science Camp](#)

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/



#AutomationSummit

AUTOMATION
DEVOPS
CLOUD
SUMMIT

Let's Keep in Touch!

Tim Spann

Developer Advocate



@[PaasDev](https://twitter.com/PaaSDev)



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



<https://github.com/tspannhw>



#AutomationSummit

AUTOMATION
DEVOPS
 A stylized orange and blue 'CD' icon.
SUMMIT

Questions



#AutomationSummit

