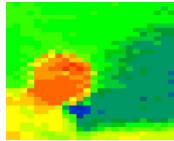




Building Real-Time Requires a Team

Tim Spann
Developer Advocate



Tim Spann

Developer Advocate
at StreamNative

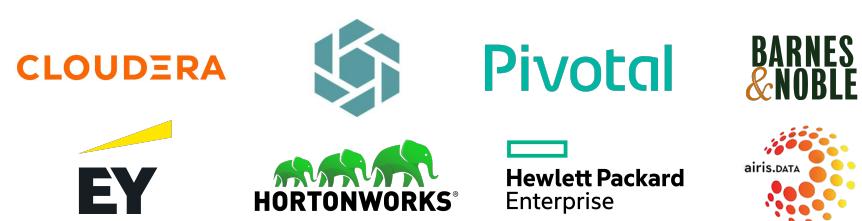


FLIP(N) Stack = Flink, Pulsar and NiFi Stack

Streaming Systems & Data Architecture Expert

Experience:

- 15+ years of experience with streaming technologies including Pulsar, Flink, Spark, NiFi, Big Data, Cloud, MXNet, IoT, Python and more.
- Today, he helps to grow the Pulsar community sharing rich technical knowledge and experience at both global conferences and through individual conversations.





FLIP Stack Weekly

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

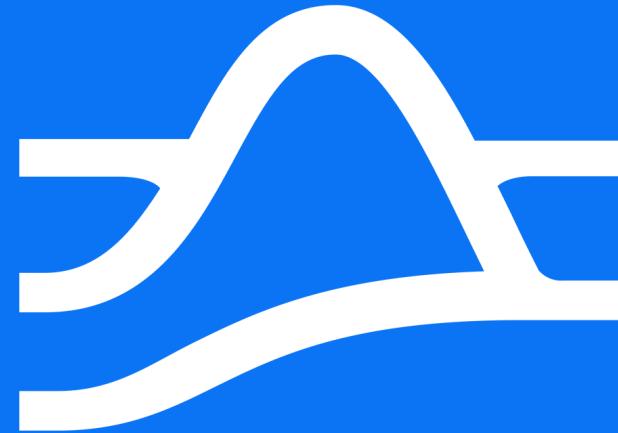
<https://bit.ly/32dAJft>

Building Real-Time Requires a Team



Agenda

- Introduction
- What is Apache Pulsar?
- Pulsar to Pinot
- Demo
- Q&A



Verticals and use cases



FinTech

- Fraud prevention
- Customer 360
- Personalization
- Threat detection



eCommerce

- Dynamic pricing
- Digital payments
- Omnichannel inventory optimization



AdTech

- Real-time bidding
- Ad serving/exchange
- Personalized promos
- Identity graph

Verticals and use cases



IoT

- Predictive maintenance
- Track and trace
- Connected supply chain
- Geo-location based alerts



Telecommunications

- Network optimization
- Churn prevention
- Real-time in-service promos & discounting



Data Lake

- Data pipeline acceleration
- Real-time analytics
- Real-time decisioning



A streaming data platform
for cloud-native,
event-driven applications.

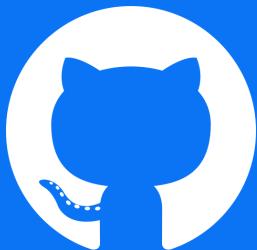


Founded by **the original creators of Apache Pulsar**.

StreamNative employs **more than 50% of the active core committers** to Apache Pulsar.

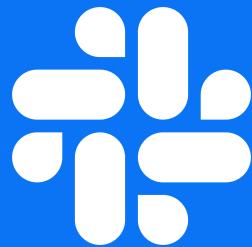
StreamNative has more experience **designing, deploying, and running large-scale Apache Pulsar instances** than any team in the world.

Apache Pulsar has a vibrant community



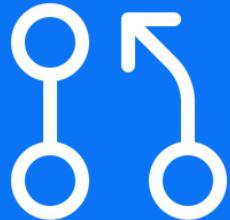
560+

Contributors



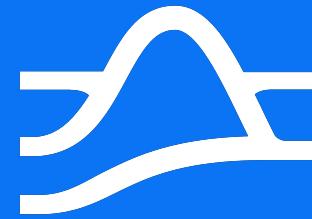
7,000+

Slack Members



10,000+

Commits



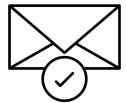
1,000+

Organizations
Using Pulsar

The basics



Unified
Messaging
Platform



Guaranteed
Message
Delivery



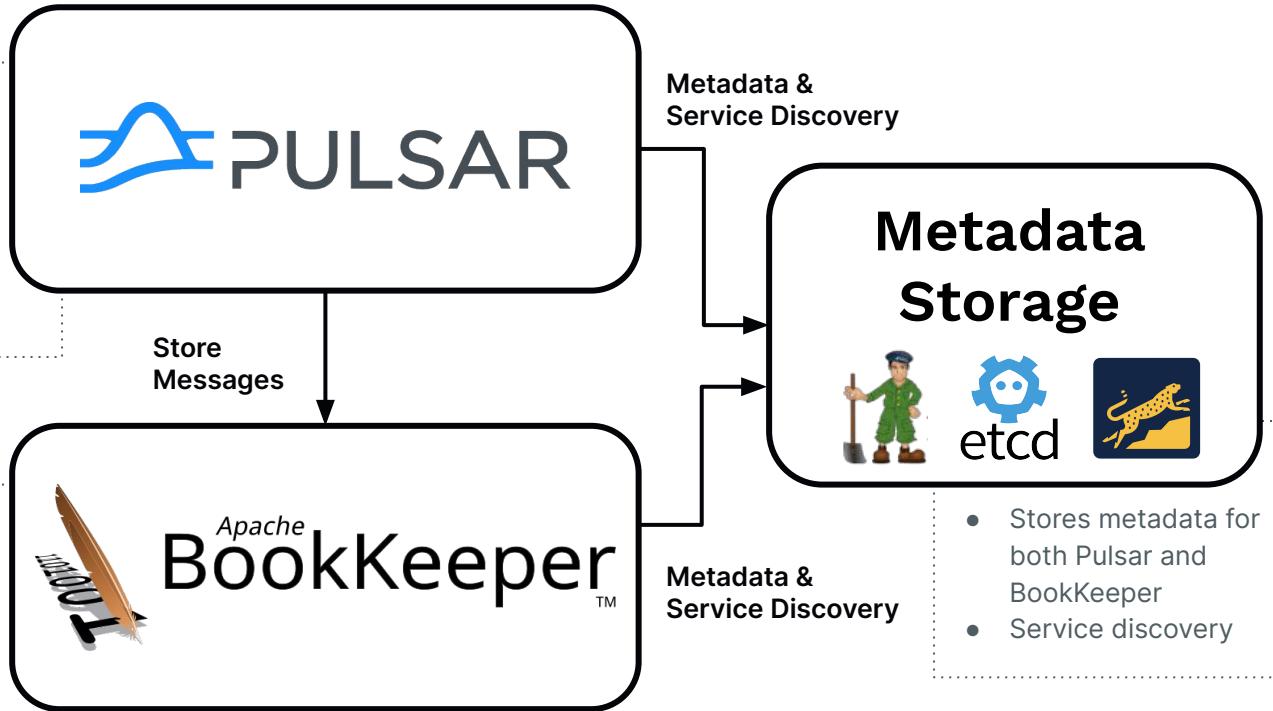
Resiliency



Infinite
Scalability

Pulsar Cluster

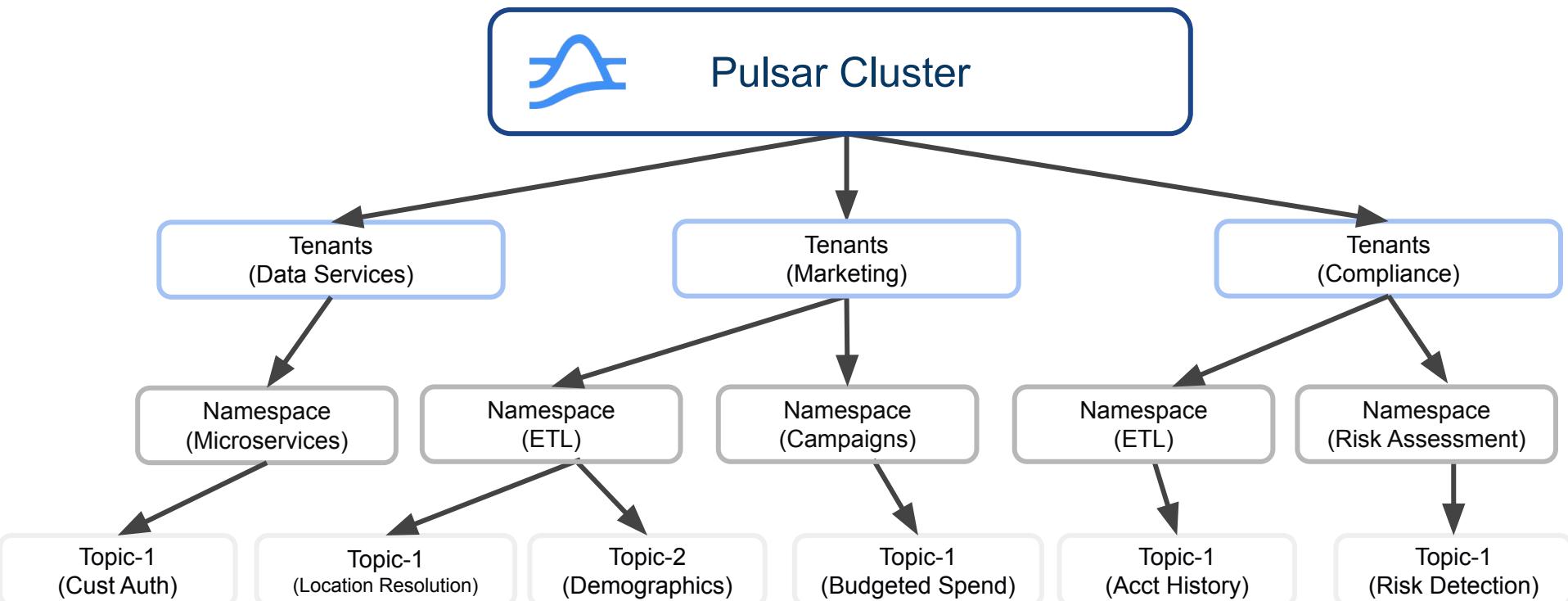
- “Brokers”
- Handles message routing and connections
- Stateless, but with caches
- Automatic load-balancing
- Topics are composed of multiple segments



- “Bookies”
- Stores messages and cursors
- Messages are grouped in segments/ledgers
- A group of bookies form an “ensemble” to store a ledger

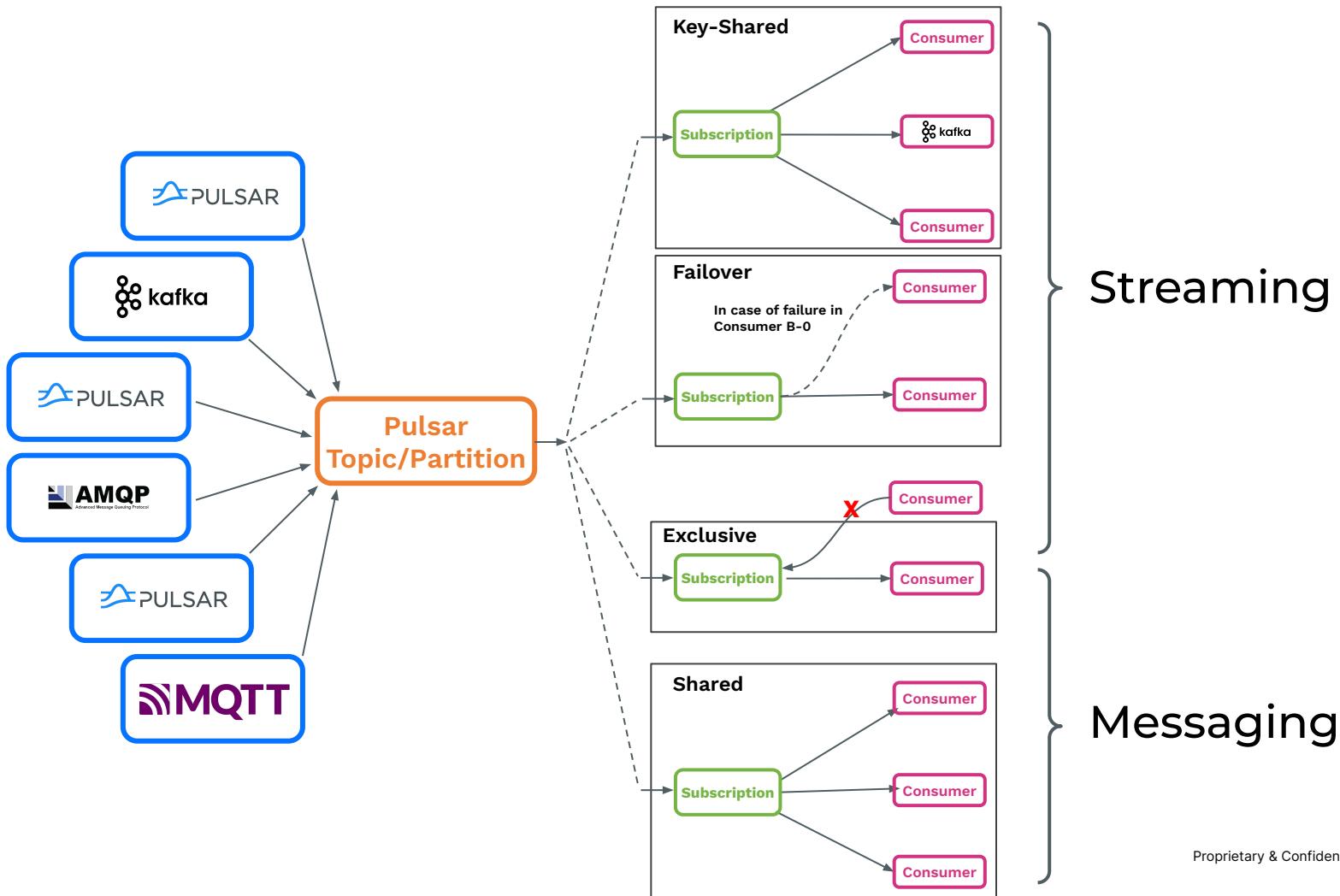
- Stores metadata for both Pulsar and BookKeeper
- Service discovery

Tenant - Namespaces - Topics

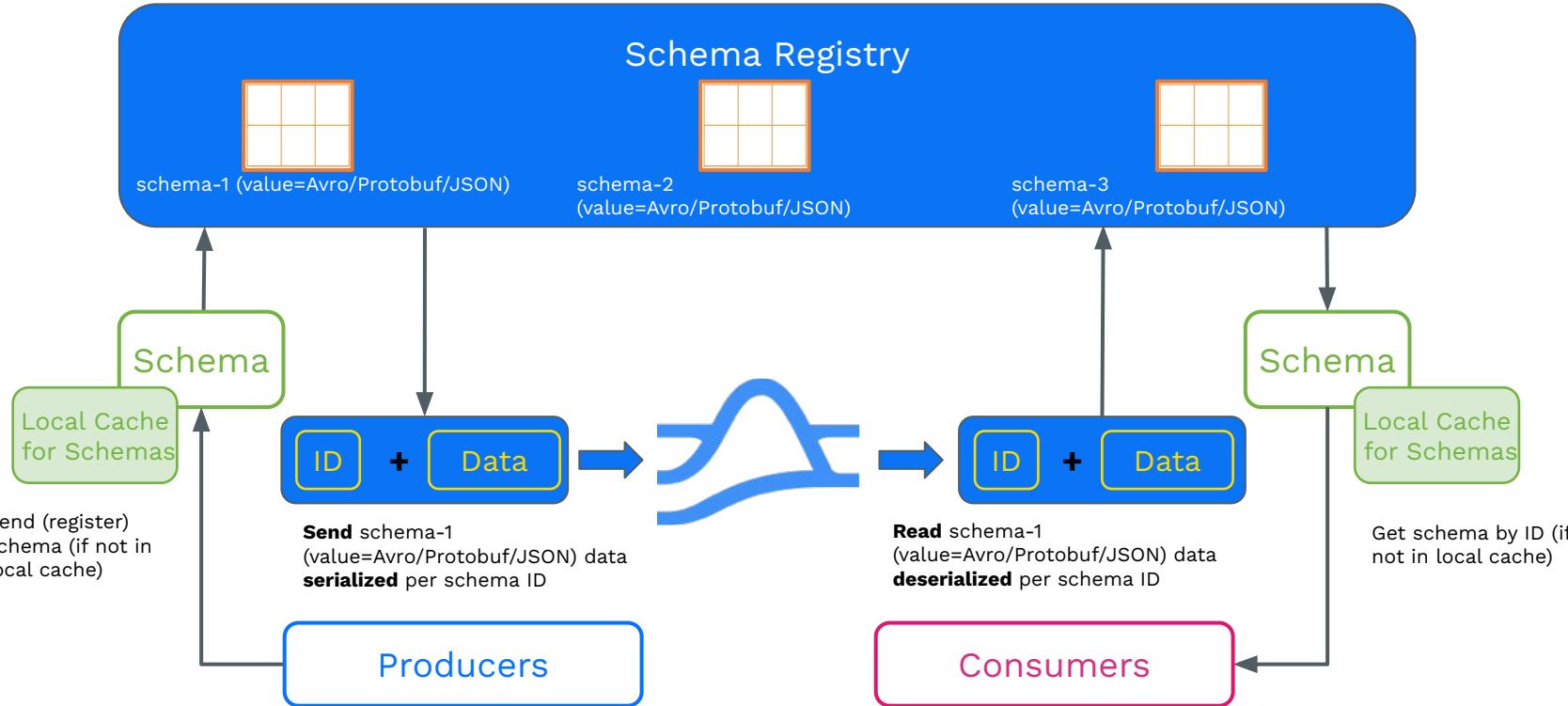


Messages - the basic unit of Pulsar

Component	Description
Value / data payload	The data carried by the message. All Pulsar messages contain raw bytes, although message data can also conform to data schemas.
Key	Messages are optionally tagged with keys, used in partitioning and also is useful for things like topic compaction.
Properties	An optional key/value map of user-defined properties.
Producer name	The name of the producer who produces the message. If you do not specify a producer name, the default name is used.
Sequence ID	Each Pulsar message belongs to an ordered sequence on its topic. The sequence ID of the message is its order in that sequence.

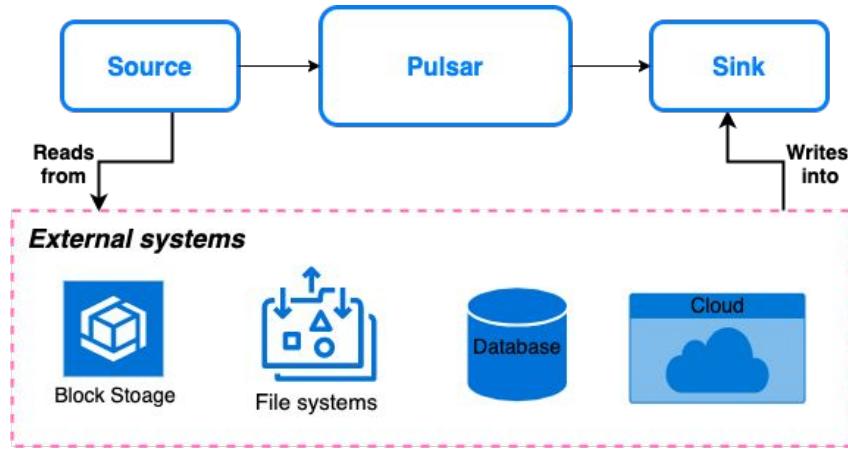


Integrated Schema Registry



Connections

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



The FliPN kitten crosses the stream 4 ways with Apache Pulsar

MoP



AoP



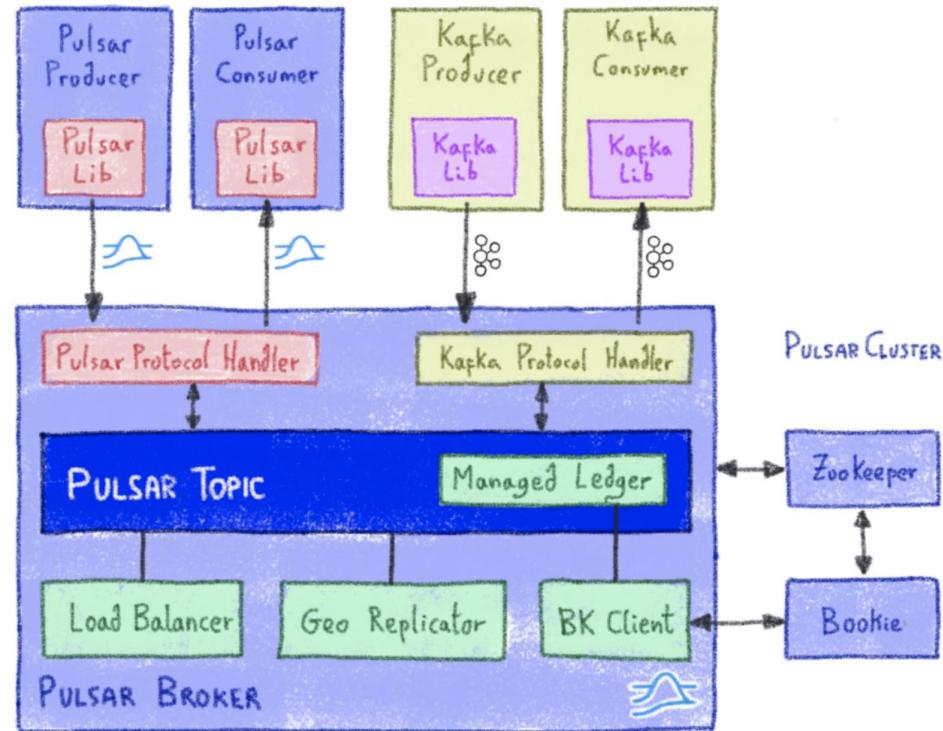
KoP



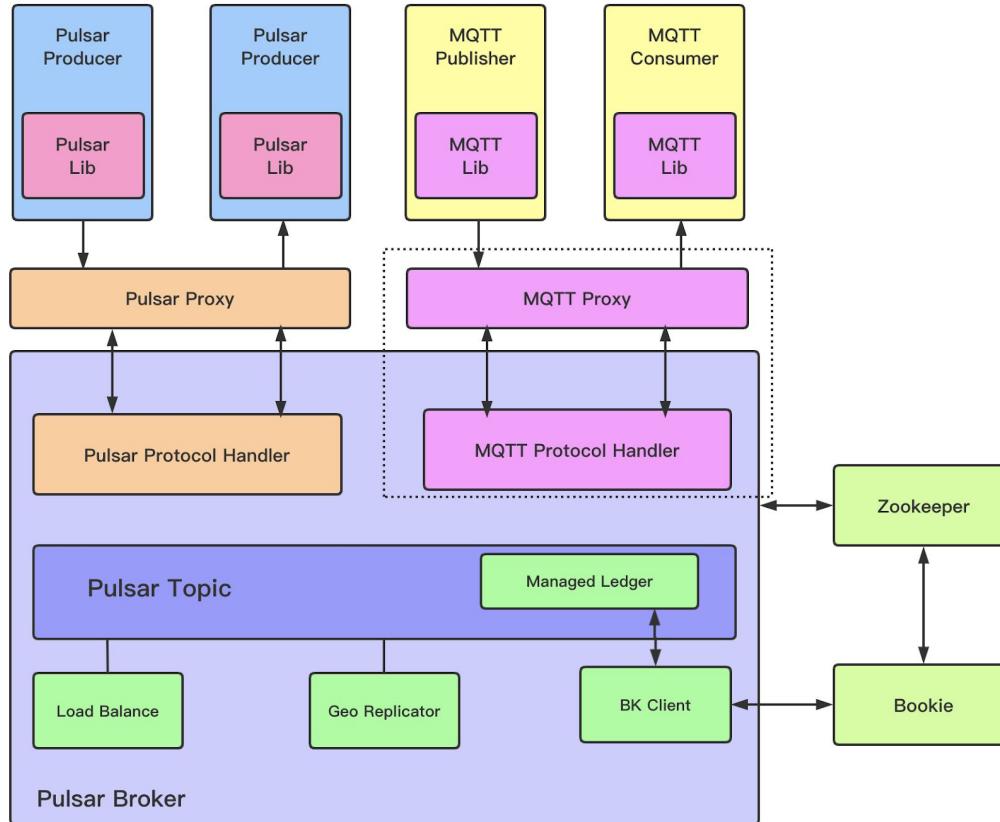
WebSockets



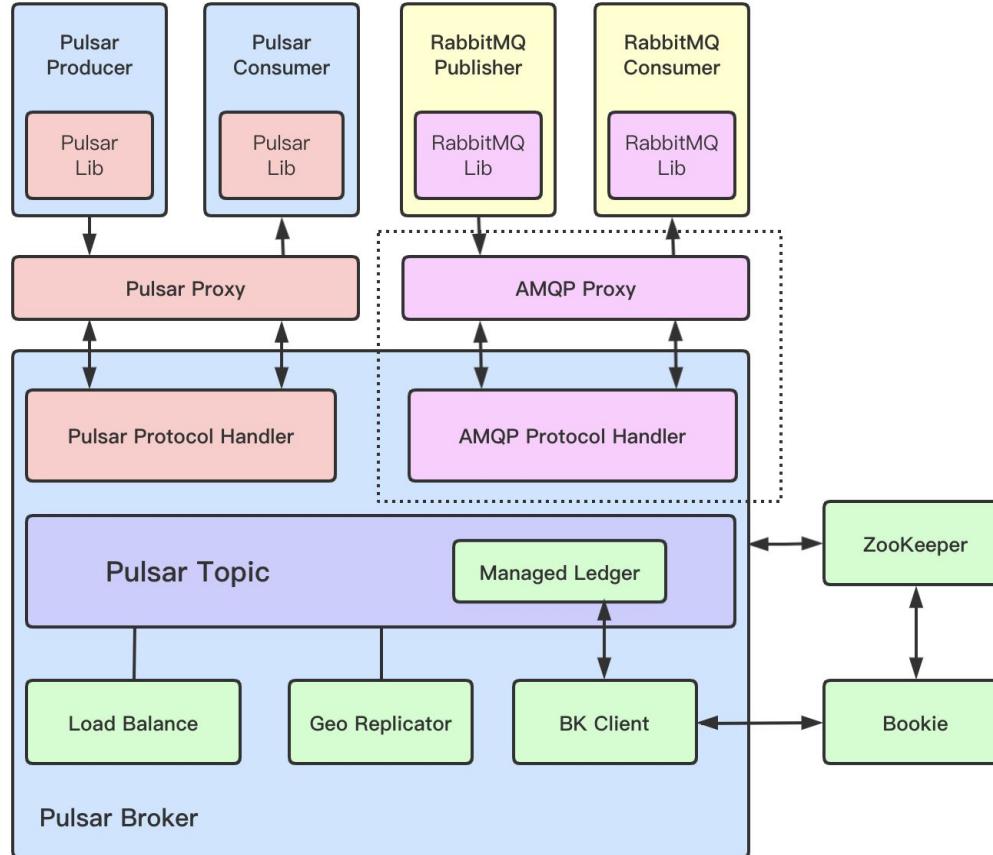
Kafka on Pulsar (KoP)



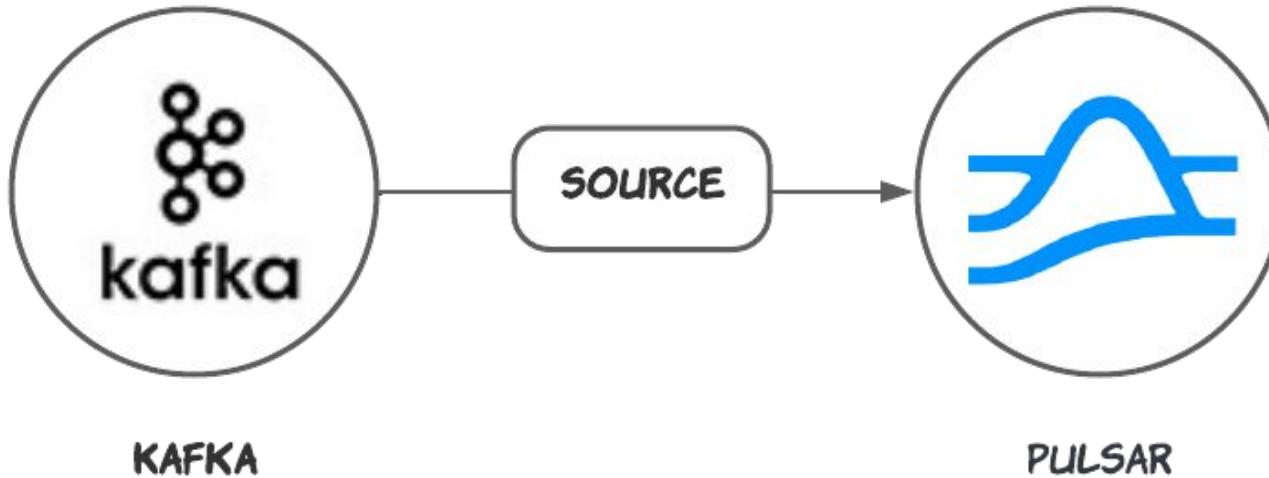
MQTT on Pulsar (MoP)



AMQP on Pulsar (AoP)



Kafka to Pulsar

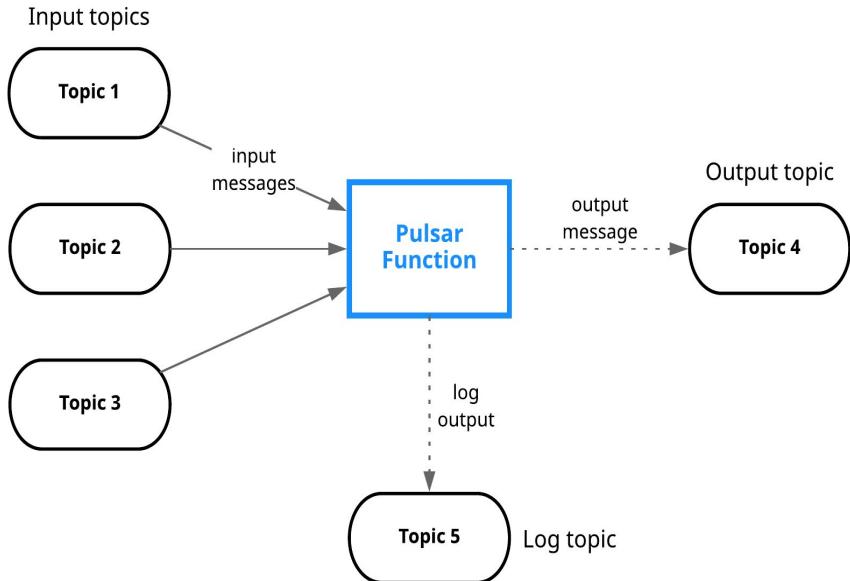


Pulsar Functions

A serverless event streaming framework

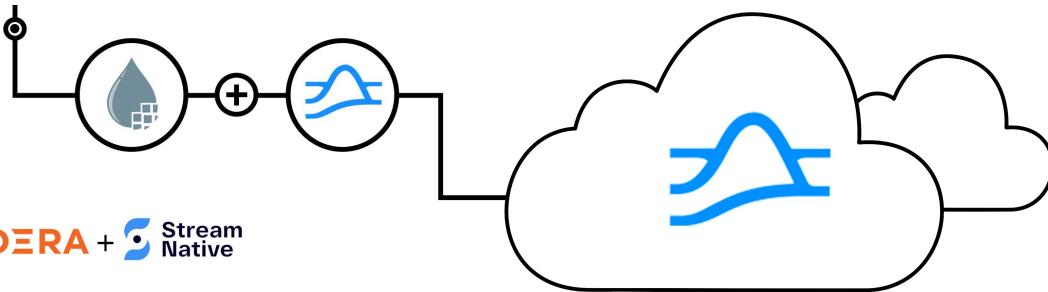
- Lightweight computation similar to AWS Lambda.
- Specifically designed to use Apache Pulsar as a message bus.
- Function runtime can be located within Pulsar Broker.
- Java Functions

Pulsar Functions



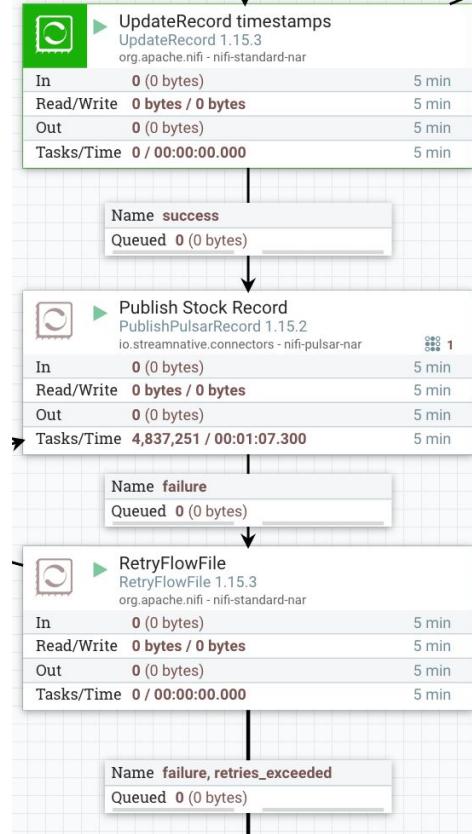
- Consume messages from one or more Pulsar topics.
- Apply user-supplied processing logic to each message.
- Publish the results of the computation to another topic.
- Support multiple programming languages (**Java, Python, Go**)
- Can leverage 3rd-party libraries to support the **execution of ML models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

Apache NiFi - Apache Pulsar Connector



Announcing the Integration of
Apache NiFi and Apache Pulsar

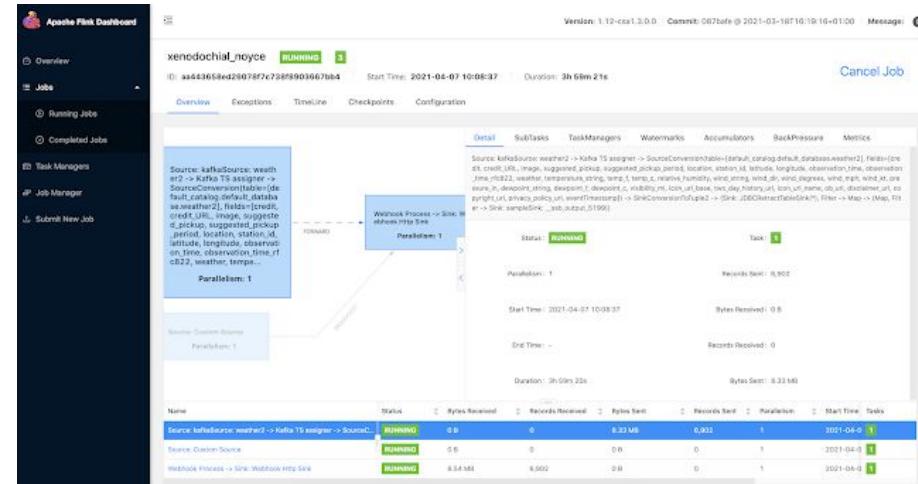
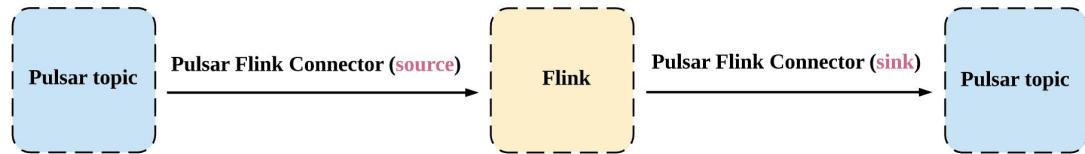
<https://github.com/streamnative/pulsar-nifi-bundle>



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





pinot-pulsar
consumer



pinot

<https://dev.startree.ai/docs/pinot/recipes/pulsar>

<https://docs.pinot.apache.org/basics/data-import/pinot-stream-ingestion/apache-pulsar>

 Cluster Manager Query Console Zookeeper Browser Swagger REST API

OPERATIONS

 Add Schema Add Offline Table Add Realtime Table

TABLES

 Search...

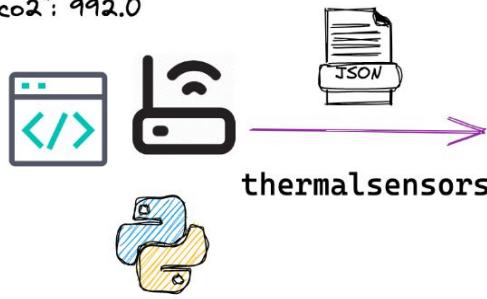
Table Name	Reported Size	Estimated Size	Number of Segments	Status
thermal_REALTIME	0 Bytes	0 Bytes	1 / 1	Good

SCHEMAS

 Search...

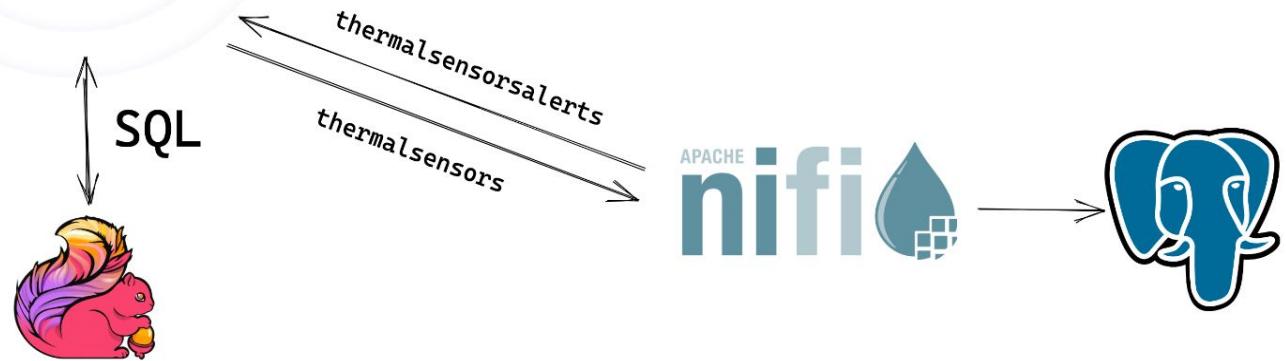
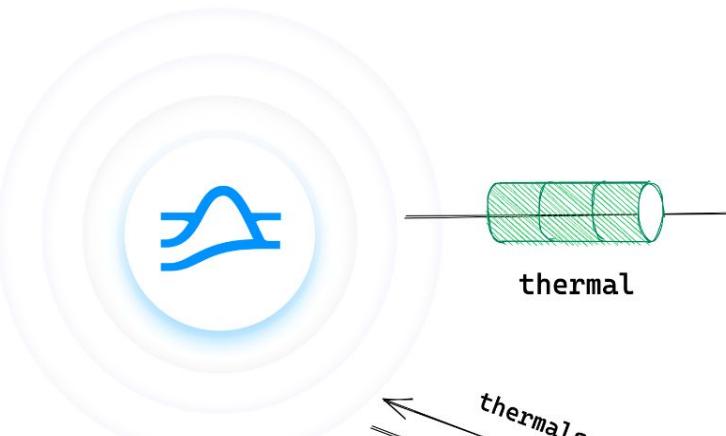
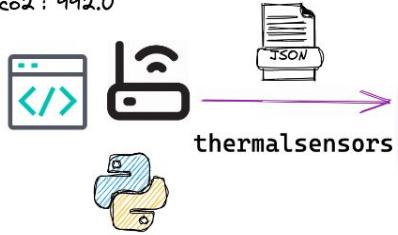
Schema Name	Dimension Columns	Date-Time Columns	Metrics Columns	Total Columns
thermal	16	1	7	24

```
"temperature": 28.238,  
"humidity": 29.61,  
"co2": 992.0
```

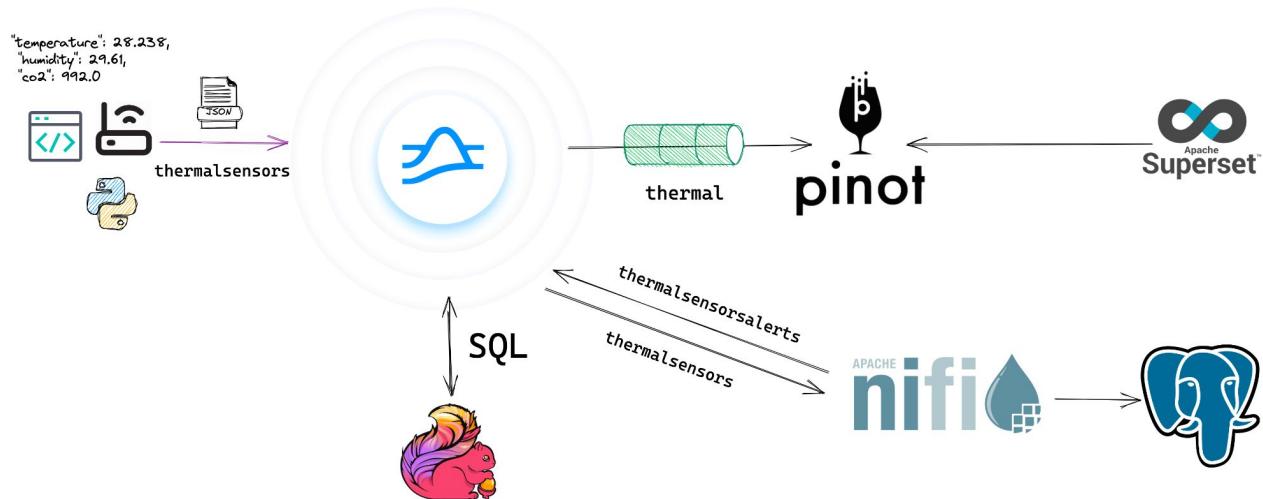
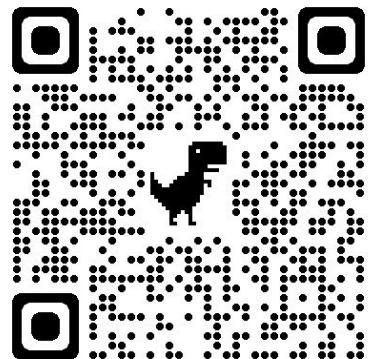


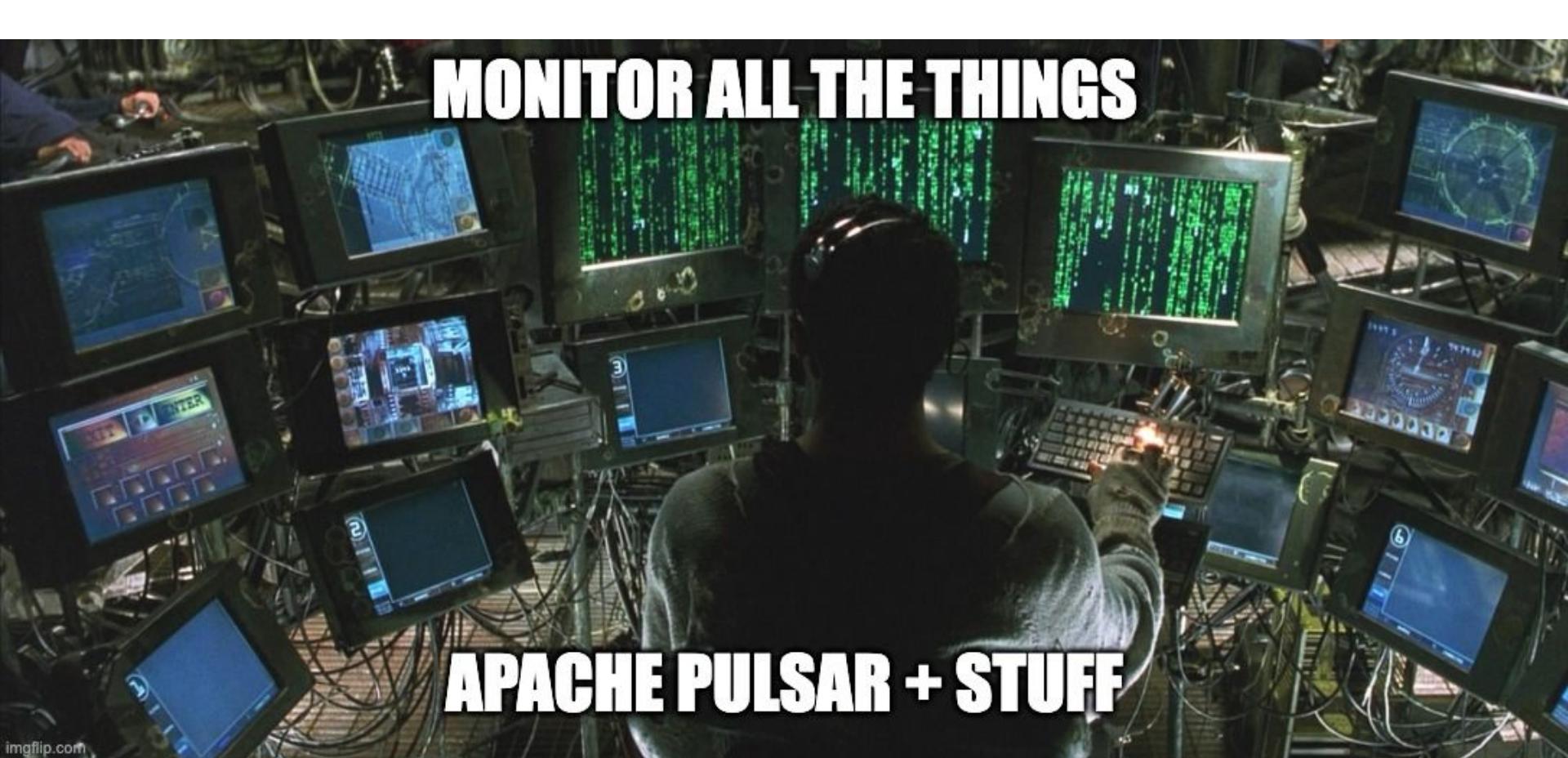
QUERY RESULT						
Search...						
systemtime	totalvocppb	temperature	cputempf	humidity	co2	equivalentco2ppm
11/21/2022 18:35:52	8	24.8547	118	28.63	941	65535
11/21/2022 18:35:48	11	24.9055	118	28.52	941	65535
11/21/2022 18:35:43	2	24.9535	118	28.47	942	65535
11/21/2022 18:35:38	7	24.9983	117	28.5	942	65535
11/21/2022 18:35:33	7	25.0443	119	28.34	942	65535
11/21/2022 18:35:28	3	25.0951	119	28.27	943	65535
11/21/2022 18:35:24	10	25.1511	118	28.2	943	65535
11/21/2022 18:35:19	6	25.2099	118	28.07	943	65535

```
"temperature": 28.238,  
"humidity": 29.61,  
"co2": 992.0
```



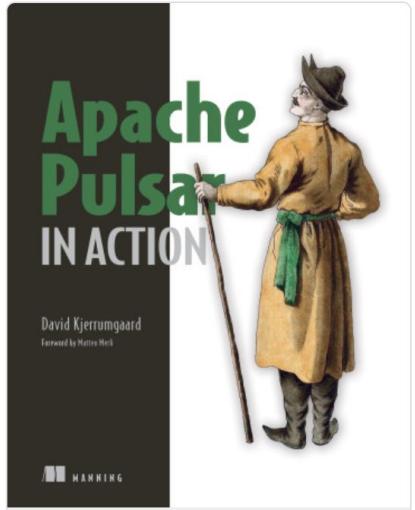
<https://github.com/tspannhw/pulsar-thermal-pinot>





MONITOR ALL THE THINGS

APACHE PULSAR + STUFF

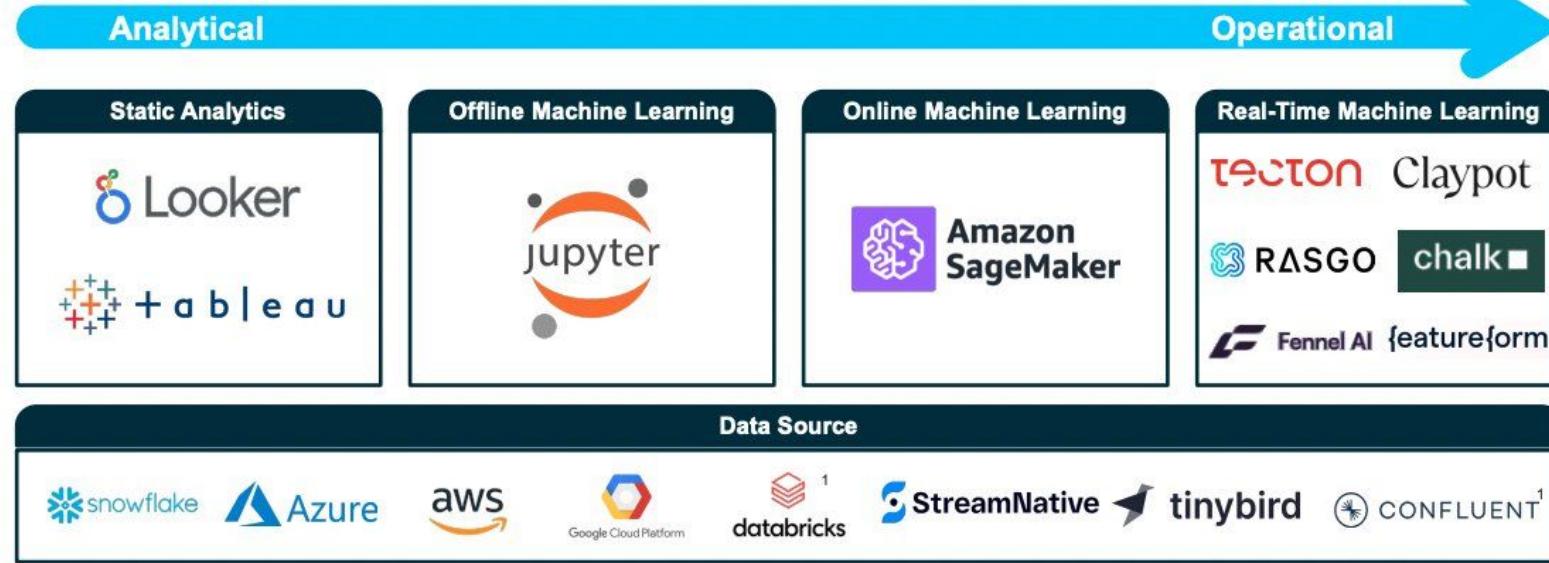


Apache Pulsar in Action

Please enjoy David's complete book which is the ultimate guide to Pulsar.



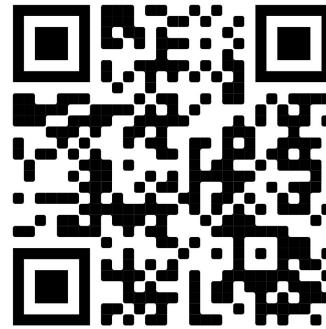
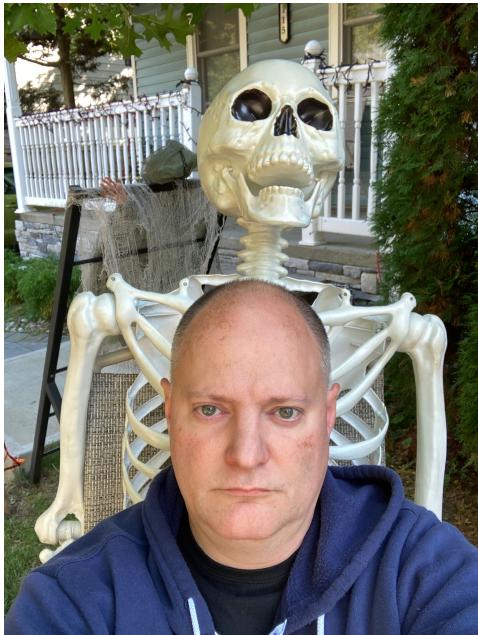
Machine learning is going real time



As ML adoption continues to mature, more companies are deploying models to power mission-critical operational use cases, driving the need for real-time capabilities for greater accuracy / reliability and a shorter feedback loop.

Tim Spann

Developer Advocate
at StreamNative



<https://streamnative.io/pulsar-python/>



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



@PassDev



<https://github.com/tspannhw>