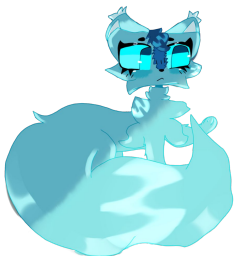


# Citizen Streaming Engineer - A How To

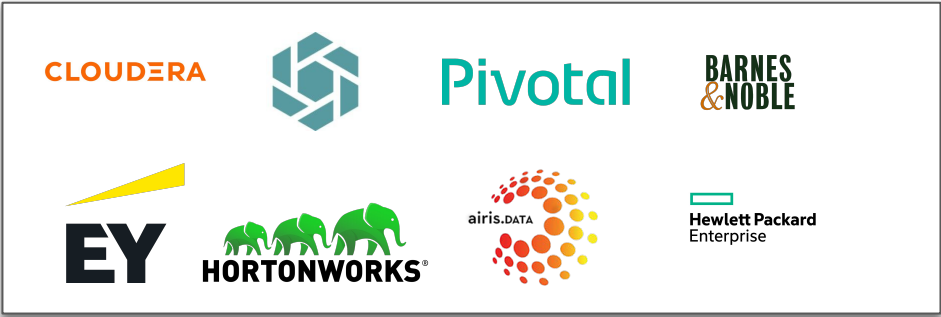




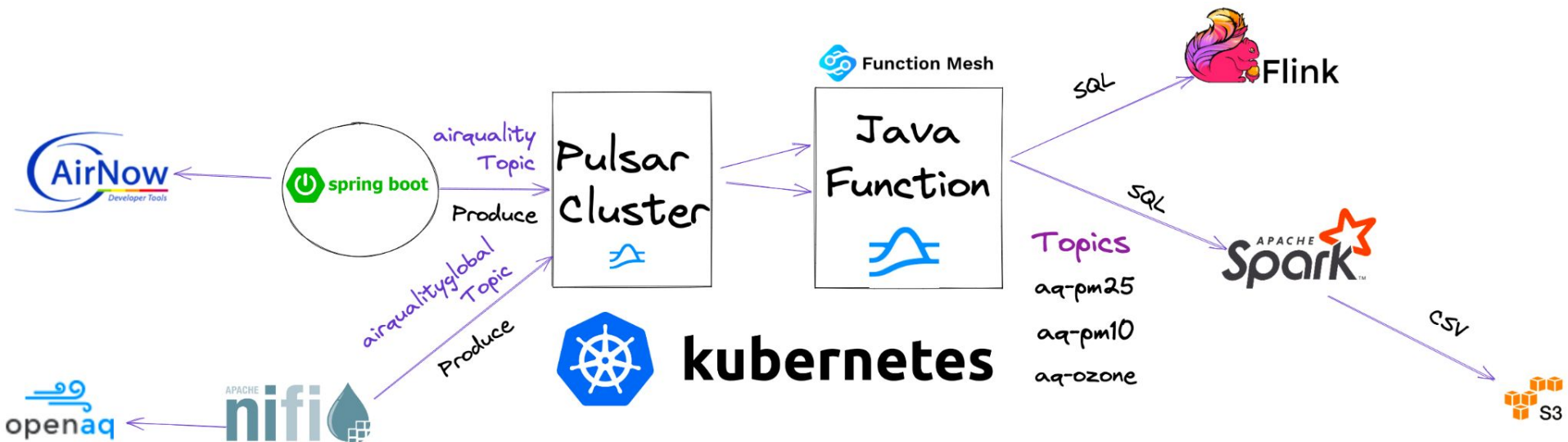
**Tim Spann**  
Developer Advocate



- FLiP(N) Stack = Flink, Pulsar and NiFi Stack
- Streaming Systems/ Data Architect
- Experience:
  - 15+ years of experience with batch and streaming technologies including Pulsar, Flink, Spark, NiFi, Spring, Java, Big Data, Cloud, MXNet, Hadoop, Datalakes, IoT and more.



# Demo







# Demo



PM 2.5

Show 10 entries

publishTime	Date	Hour	TZ	Area	State	Lat	Long	PM 2.5 AQI
2022-04-26T12:02:37.299-04:00	2022-04-26	10	EST	Atlanta	GA	33.65	-84.43	63
2022-04-26T12:02:47.304-04:00	2022-04-26	10	EST	Atlanta	GA	33.65	-84.43	63
2022-04-26T12:02:57.294-04:00	2022-04-26	10	EST	Atlanta	GA	33.65	-84.43	63
2022-04-26T12:03:07.295-04:00	2022-04-26	10	EST	Atlanta	GA	33.65	-84.43	63
2022-04-26T12:05:24.794-04:00	2022-04-26	10	EST	Atlanta	GA	33.65	-84.43	63
2022-04-26T12:05:25.853-04:00	Sat, 09 Apr 2022 12:45:34 GMT	-1	EST	MiraculousMess:US		42.038597	-87.69545	0
2022-04-26T12:05:25.886-04:00	Sat, 09 Apr 2022 12:45:34 GMT	-1	EST	Fruitdale:US		39.778137	-105.11771	190
2022-04-26T12:05:25.949-04:00	Sat, 09 Apr 2022 12:45:34 GMT	-1	EST	Willow Glen West:US		37.304996	-121.91792	730
2022-04-26T12:05:26.024-04:00	Sat, 09 Apr 2022 12:45:34 GMT	-1	EST	Greengate school exterior:US		38.68429	-121.79665	340
2022-04-26T12:05:26.087-04:00	Sat, 09 Apr 2022 12:45:34 GMT	-1	EST	Golden Bear:US		37.87195	-122.27165	819

Showing 1 to 10 of 751 entries

Previous 1 2 3 4 5 ... 76 Next

PM 10

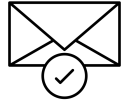
Show 10 entries

publishTime	Date	Hour	TZ	Area	State	Lat	Long	PM 10 AQI
2022-04-26T12:14:33.898-04:00	Sun, 10 Apr 2022 00:31:24 GMT	-1	EST	Edina MN:US		44.87834	-93.35441	120
2022-04-26T12:14:34.01-04:00	Sun, 10 Apr 2022 00:31:24 GMT	-1	EST	Cherry Hill:US		48.11406	-123.44409	0
2022-04-26T12:14:34.02-04:00	Sun, 10 Apr 2022 00:31:24 GMT	-1	EST	60 Oakland Hts:US		29.838072	-98.603134	110
2022-04-26T12:14:34.041-04:00	Sun, 10 Apr 2022 00:31:24 GMT	-1	EST	Healdsburg:US		38.62931	-122.86675	500
2022-04-26T12:14:34.061-04:00	Sun, 10 Apr 2022 00:31:24 GMT	-1	EST	rex manor outside:US		37.402657	-122.087944	690
2022-04-26T12:14:34.123-04:00	Sun, 10 Apr 2022 00:31:24 GMT	-1	EST	Jeanne Circle:US		37.976818	-122.0929	660
2022-04-26T12:14:34.148-04:00	Sun, 10 Apr 2022 00:31:24 GMT	-1	EST	Uplands:US		37.854927	-122.24137	1500
2022-04-26T12:14:34.199-04:00	Sun, 10 Apr 2022 00:31:23 GMT	-1	EST	Murray 500W Cherry St. powered by UTOPIA Fiber:US		40.672577	-111.90616	370
2022-04-26T12:14:34.296-04:00	Sun, 10 Apr 2022 00:31:23 GMT	-1	EST	BR09-KR		36.408836	126.53104	3510
2022-04-26T12:14:34.446-04:00	Sun, 10 Apr 2022 00:31:23 GMT	-1	EST	Vista:US		34.168556	-118.03021	280

# Why Apache Pulsar?



**Unified  
Messaging  
Platform**



**Guaranteed  
Message  
Delivery**

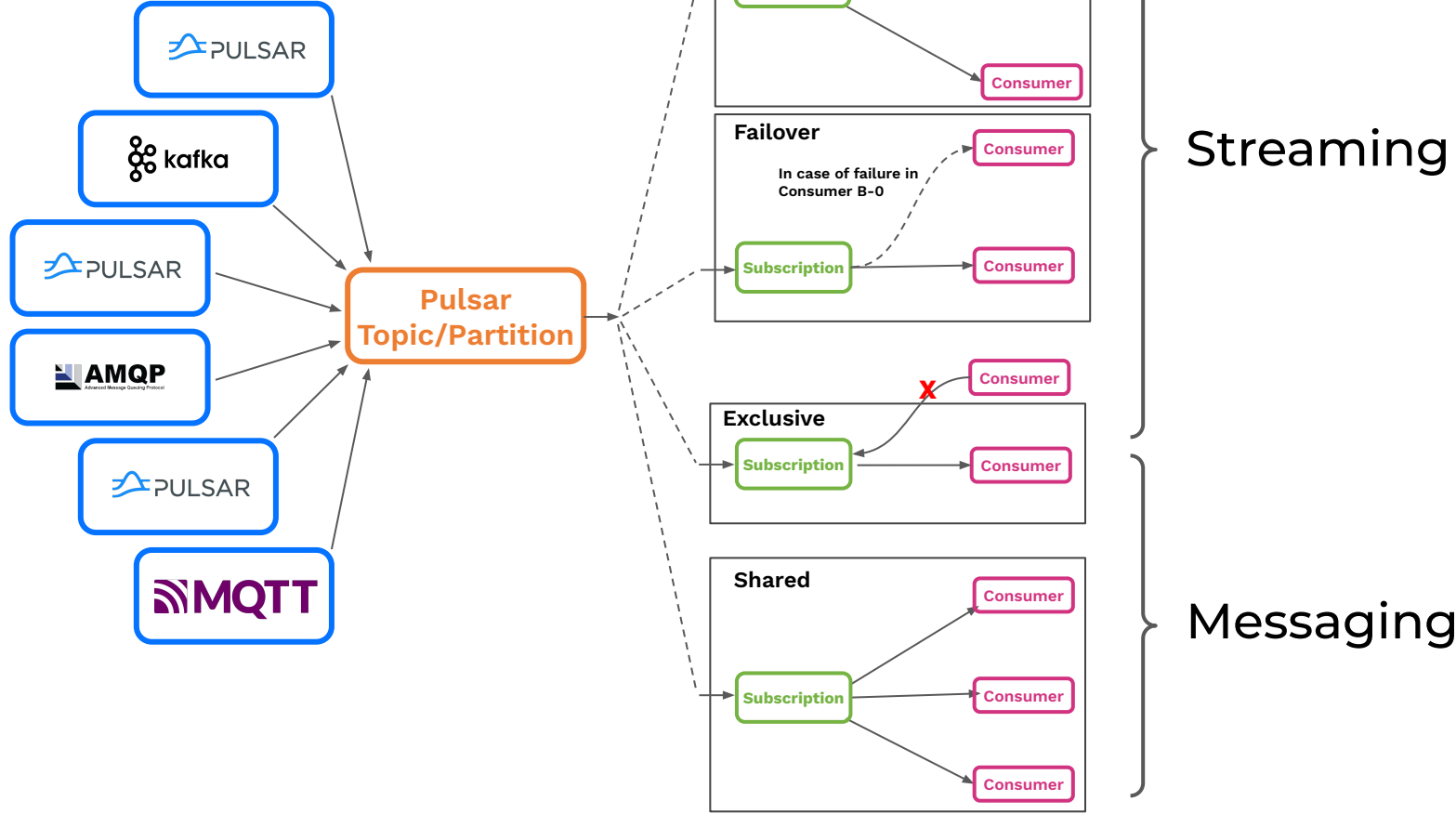


**Resiliency**

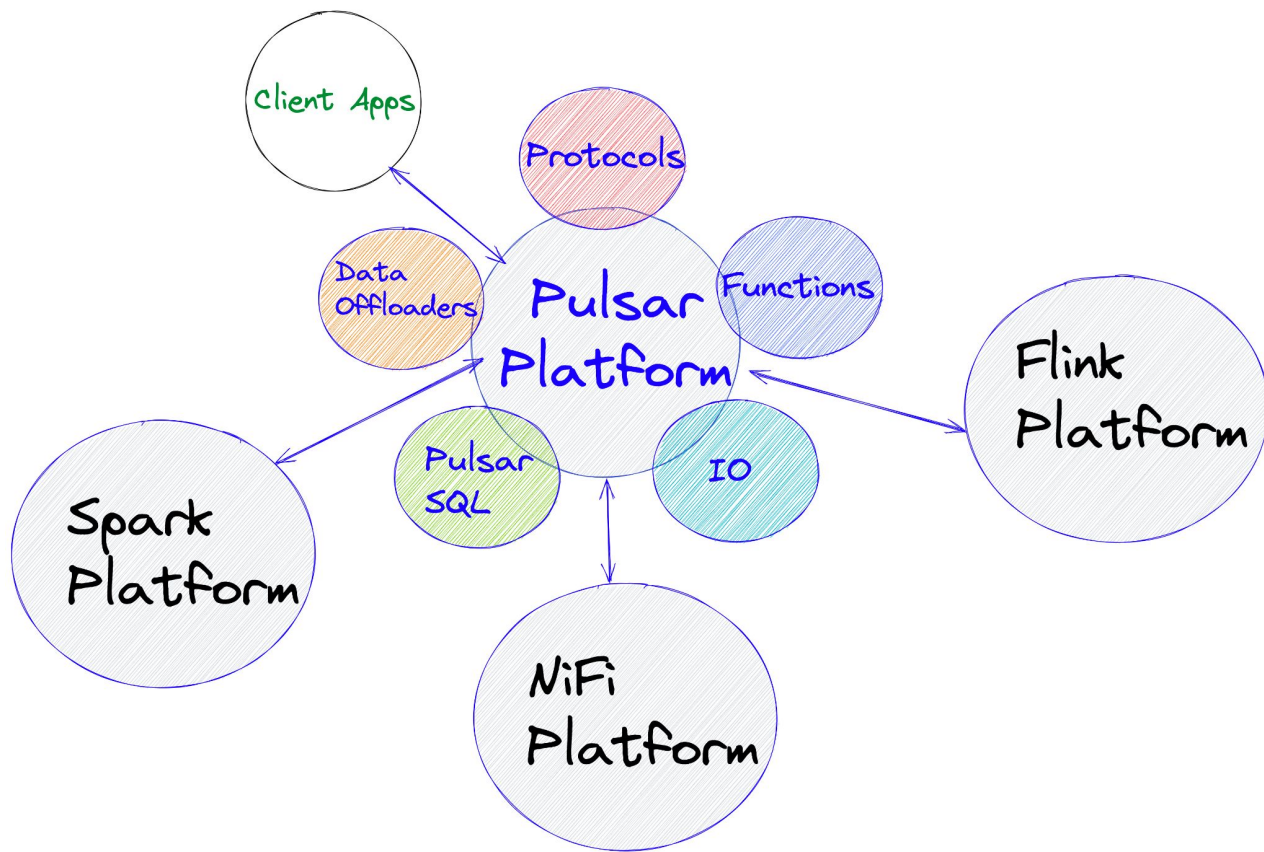


**Infinite  
Scalability**

# Unified Messaging Model



# Ecosystem





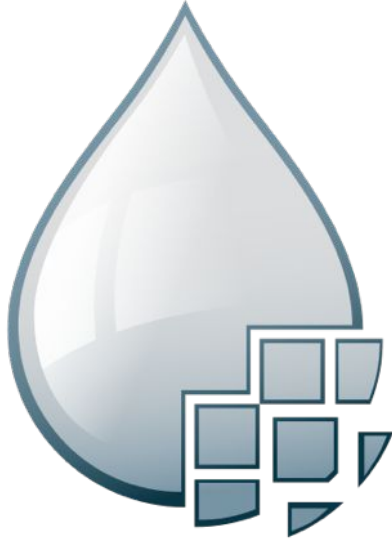
# Easy to Build Streaming Data Pipelines



- Ingest Data
- Route, Transform, Enrich
- Join Data
- ML Model Access
- Store



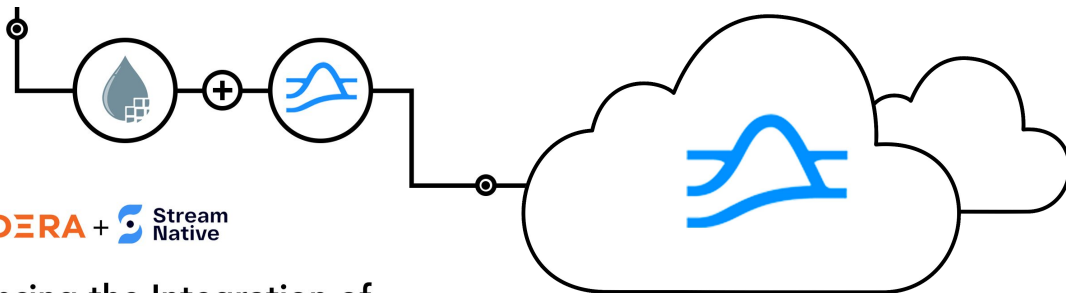
# Why Apache NiFi?



- 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 components
- Flow templates
- Pluggable/multi-role security
- Designed for extension
- Clustering
- Version Control

# Use Apache NiFi For Ingest

- Ingest Data
- Cleanse



**CLUSTERA** + **Stream Native**

Announcing the Integration of Apache NiFi and Apache Pulsar

UpdateRecord timestamps  
UpdateRecord 1.15.3  
org.apache.nifi-nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name success  
Queued 0 (0 bytes)

Publish Stock Record  
PublishPulsarRecord 1.15.2  
to streamnative-connectors - nifi-pulsar-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	4,837,251 / 00:01:07.300	5 min

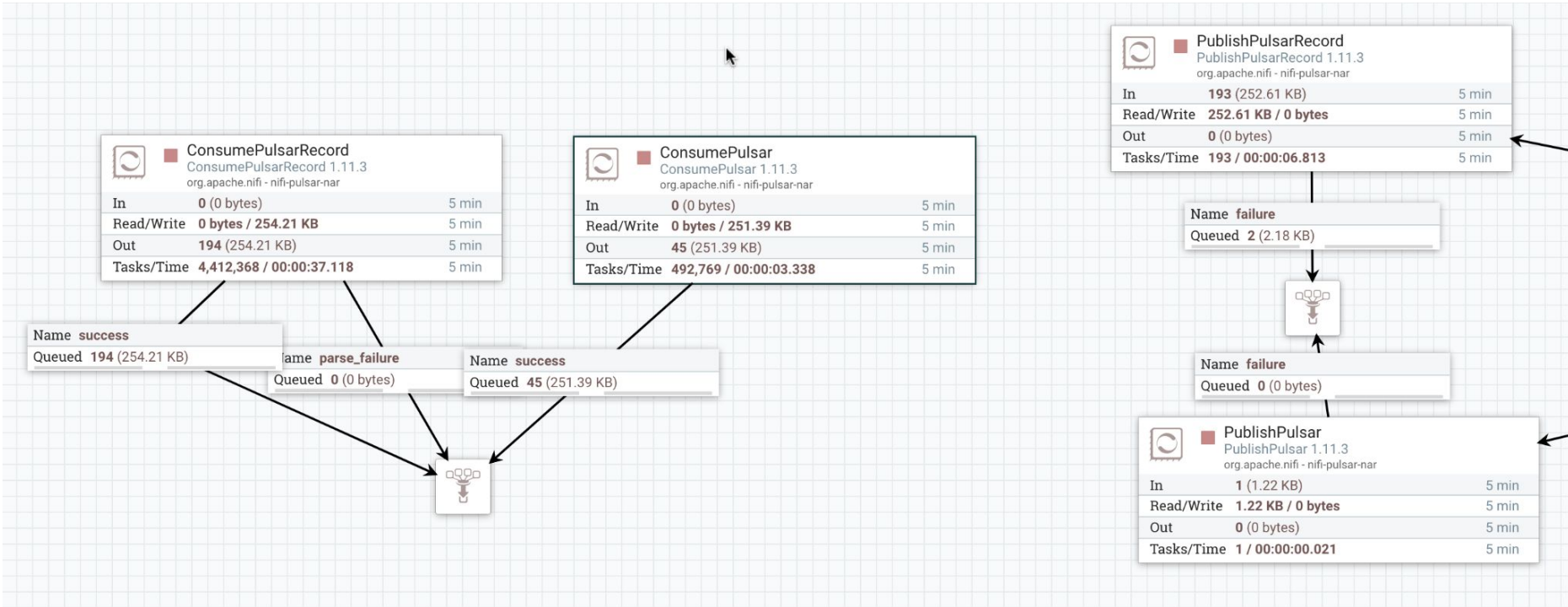
Name failure  
Queued 0 (0 bytes)

RetryFlowFile  
RetryFlowFile 1.15.3  
org.apache.nifi-nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

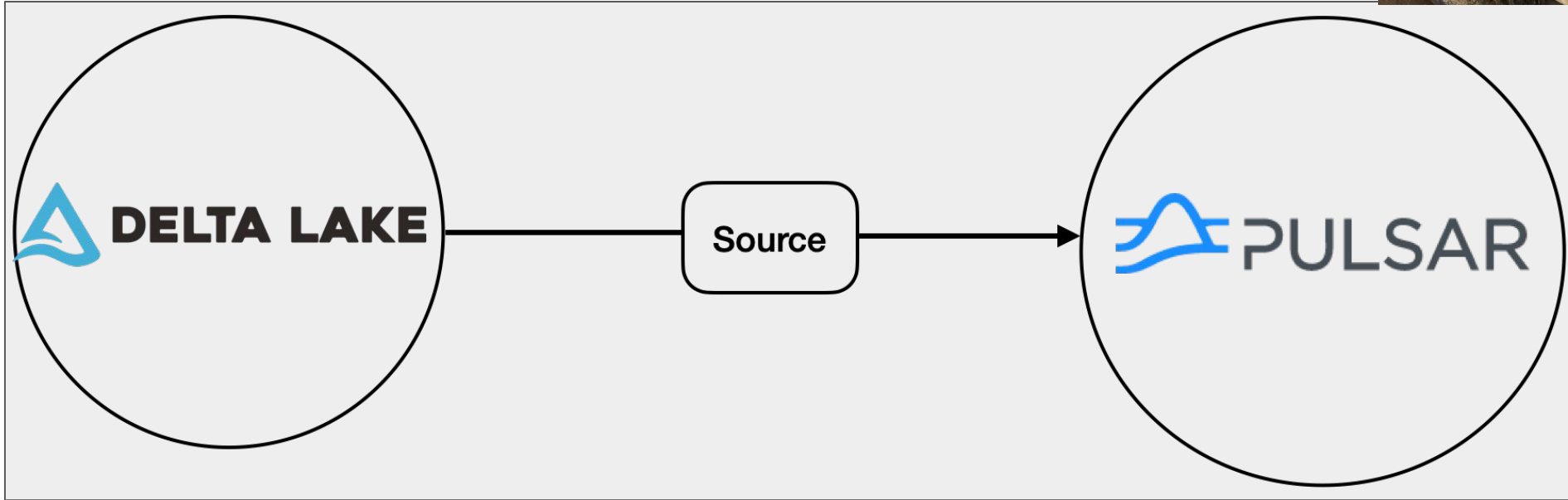
Name failure, retries\_exceeded  
Queued 0 (0 bytes)

# Apache NiFi <-> Apache Pulsar





# Use Apache Pulsar For Ingest



# Use Pulsar to Route/Transform/Enrich

- Libraries
- Functions
- Connectors
- AMQP, Kafka, MQTT
- Tiered Storage

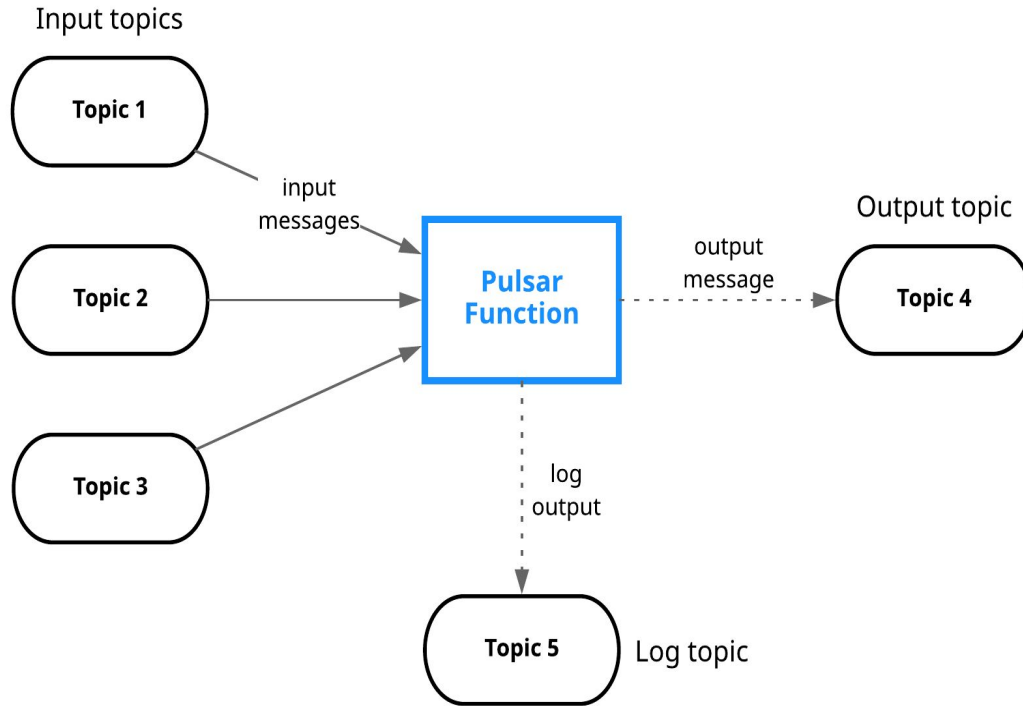


# Use Schemas

- Utilizing JSON Data with a JSON Schema
- Consistency, Contracts, Clean Data
- This enables easy SQL:
  - Pulsar SQL (Presto SQL)
  - Flink SQL
  - Spark Structured Streaming



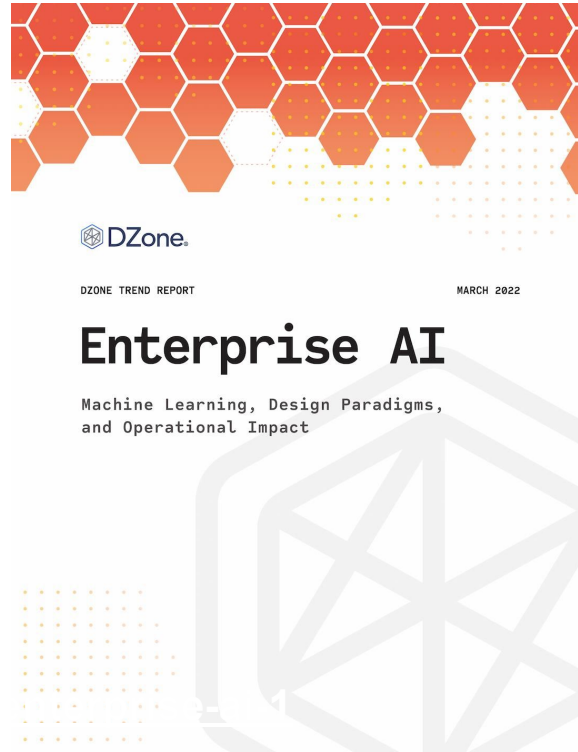
# Use Pulsar Functions



- Use Java, Python or Go
- Simple way to add functionality
- Route / Filter / Transform
- Call Machine Learning Models



# Deploying AI With an Event-Driven Platform

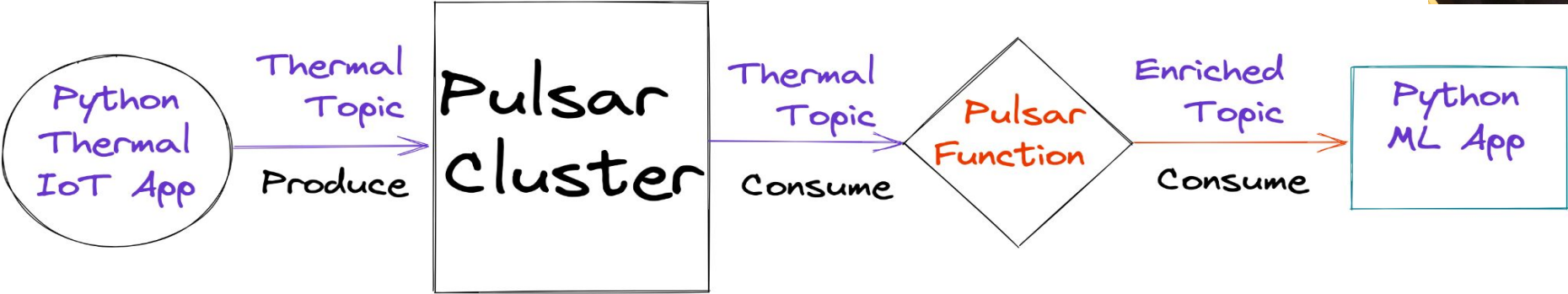
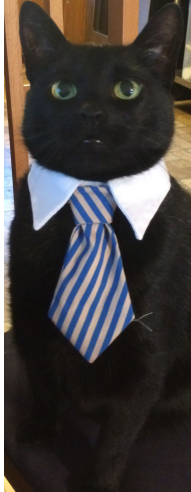


# ML Models via Python / Java FN



- Visual Question and Answer
- Natural Language Processing
- Sentiment Analysis
- Text Classification
- Named Entity Recognition
- Content-based Recommendations
- Predictive Maintenance
- Fault Detection
- Fraud Detection
- Time-Series Predictions
- Naive Bayes

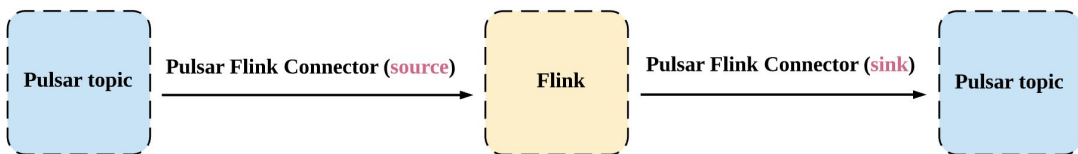
# Functions for Enrichment



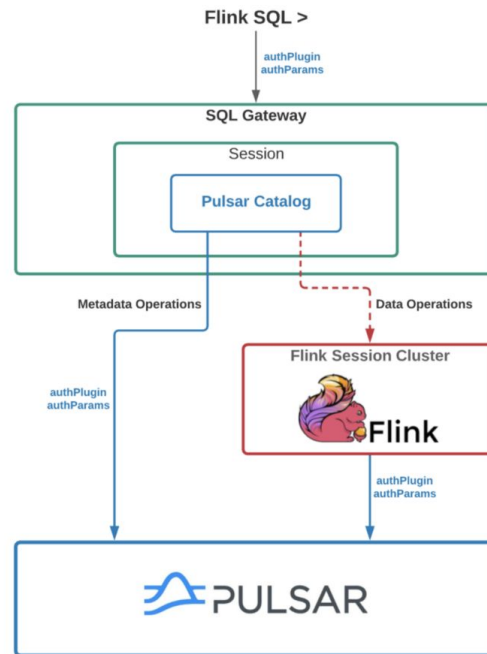
# Use Apache Flink to Join / Aggregate



## Continuous SQL



```
BETA
Flink SQL Client
Welcome! Enter 'HELP;' to list all available commands. 'QUIT;' to exit.
Flink SQL> CREATE CATALOG pulsar WITH (
  'type' = 'pulsar',
  'service-url' = 'pulsar://pulsar1:6650',
  'admin-url' = 'http://pulsar1:8080',
  'format' = 'json'
);
[INFO] Execute statement succeed.
Flink SQL> █
```



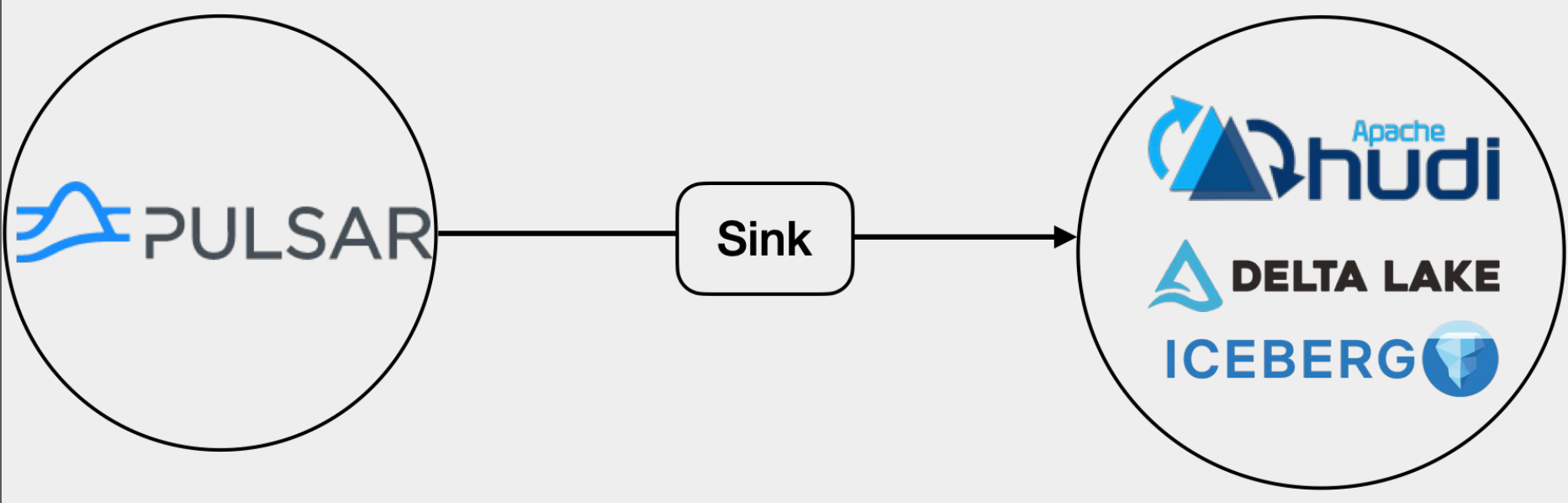


# Use Apache Spark To Store



```
val dfPulsar = spark.readStream.format("pulsar")  
    .option("service.url", "pulsar://pulsar1:6650")  
    .option("admin.url", "http://pulsar1:8080")  
    .option("topic", "persistent://public/default/airquality").load()  
  
val pQuery = dfPulsar.selectExpr("*")  
    .writeStream.format("parquet")  
    .option("truncate", false).start()
```

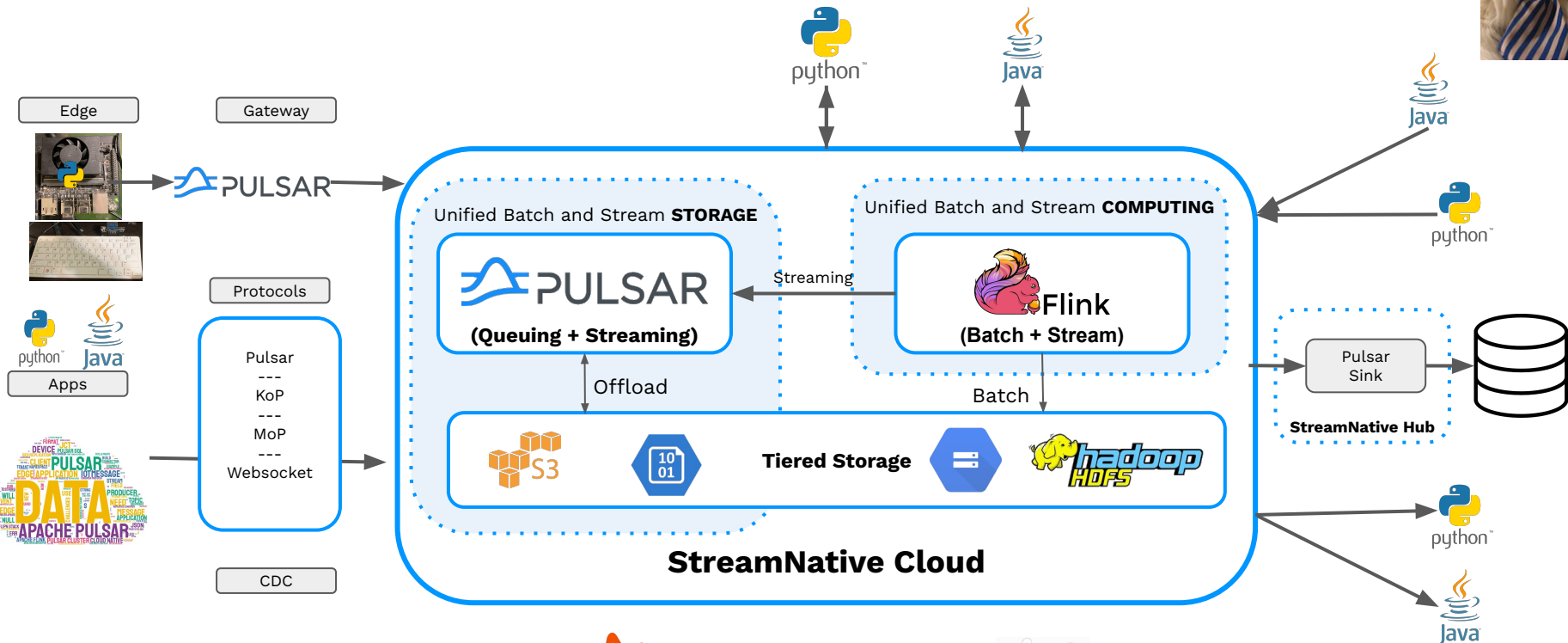
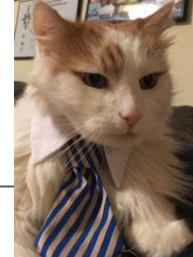
# Use Pulsar to Stream to Lakehouses



# Simple Data Pipeline



# Streaming FLiP-ML Apps



# Continuous Air Quality Aggregate Monitoring

Refresh: 1 s

SQL Query Result (Table)  
Page: Last of 1

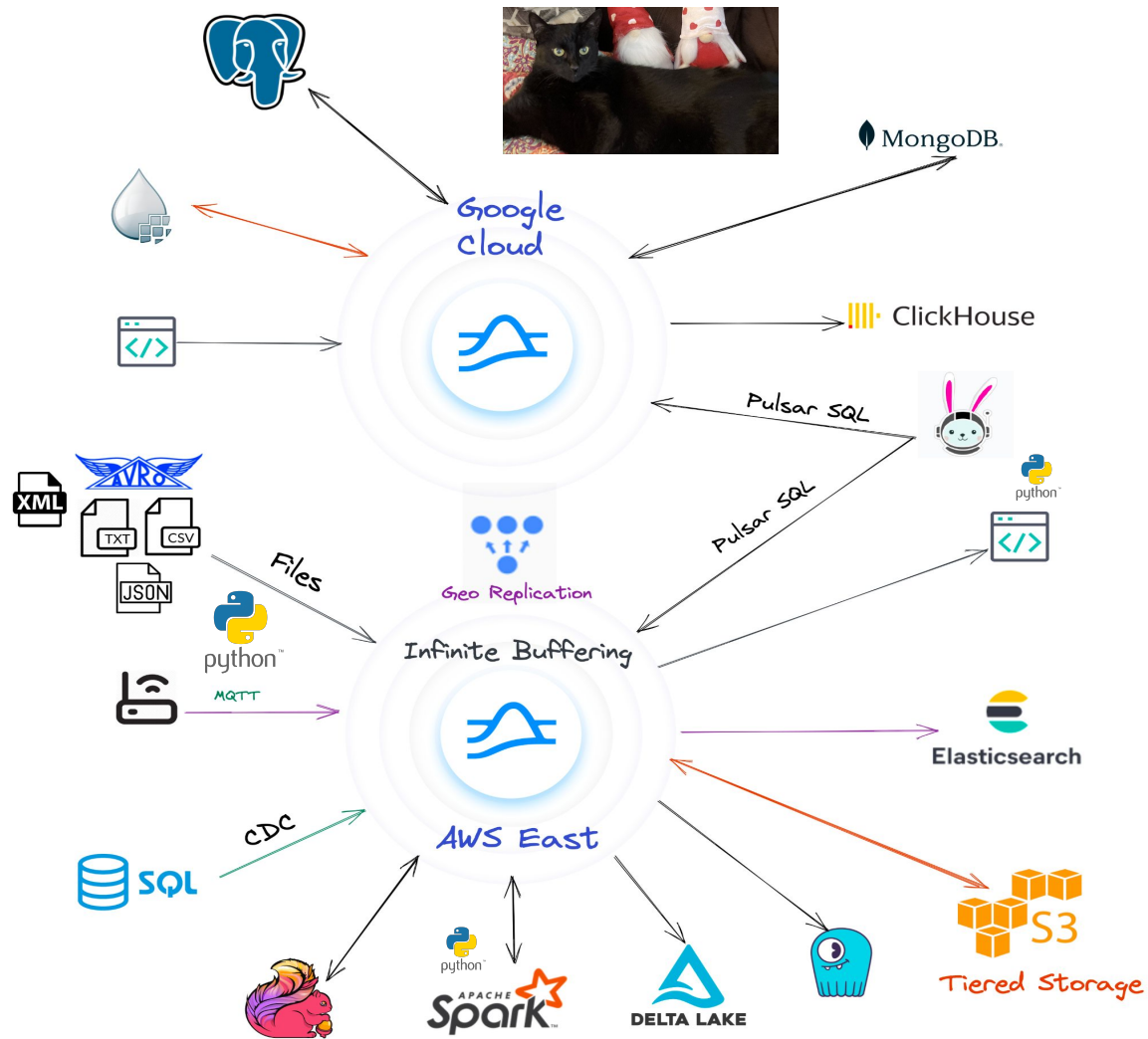
MaxAQI	parameterName	reportingArea
33	O3	Atlanta
30	PM2.5	Atlanta
14	PM10	Atlanta

Quit Refresh

Inc Refresh  
Dec Refresh

Goto Page  
Last Page

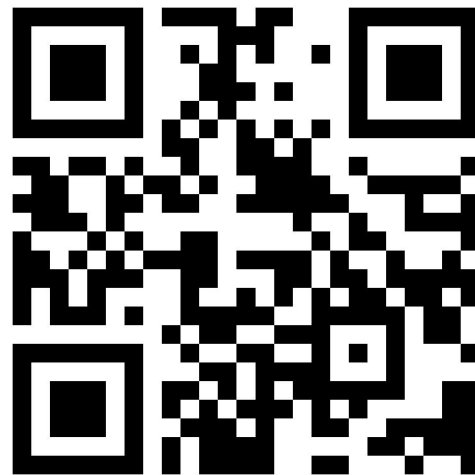
- Buffer
- Batch
- Route
- Filter
- Aggregate
- Enrich
- Replicate
- Dedupe
- Decouple
- Distribute



# FLiP Stack Weekly



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



This week in Apache Flink, Apache Pulsar, Apache NiFi, Apache Spark, Java and Open Source friends.



Let's Keep  
in Touch!



**Tim Spann**

Developer Advocate



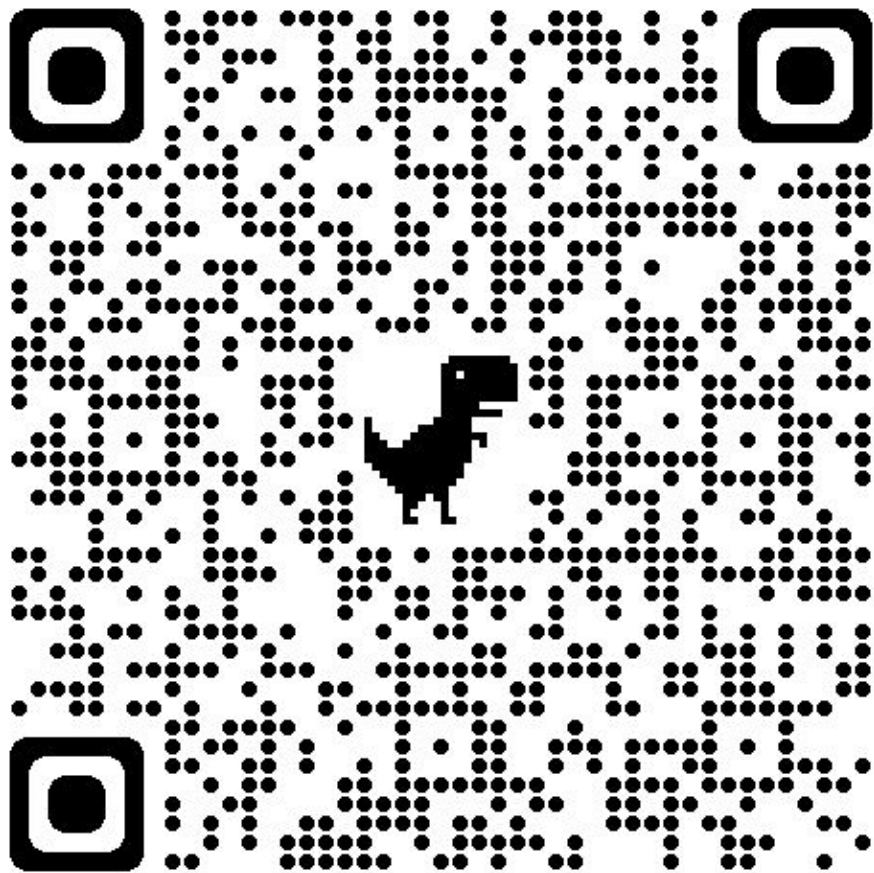
[PaaSDev](#)



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



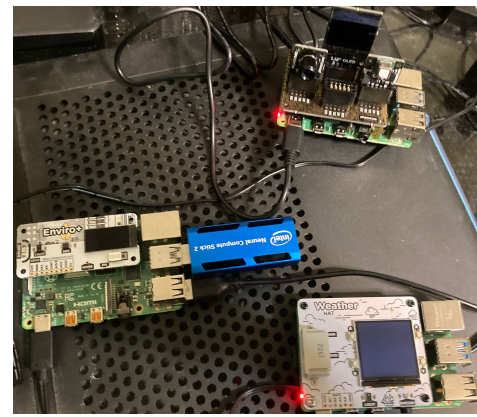
<https://github.com/tspannhw>



**SLIDES**

# Python For Pulsar on Pi

- <https://github.com/tspannhw/FLiP-Pi-BreakoutGarden>
- <https://github.com/tspannhw/FLiP-Pi-Thermal>
- <https://github.com/tspannhw/FLiP-Pi-Weather>
- <https://github.com/tspannhw/FLiP-RP400>
- <https://github.com/tspannhw/FLiP-Py-Pi-GasThermal>
- <https://github.com/tspannhw/FLiP-PY-FakeDataPulsar>
- <https://github.com/tspannhw/FLiP-Py-Pi-EnviroPlus>
- <https://github.com/tspannhw/PythonPulsarExamples>
- <https://github.com/tspannhw/pulsar-pychat-function>
- <https://github.com/tspannhw/FLiP-PulsarDevPython101>





Thanks

