

Smart Transit

Timothy Spann & David Kjerrumgaard

Developer Advocates / StreamNative

22/Sept/2021 Wednesday



Agenda

Wednesday 15:00 UTC

Smart Transit: Real-Time Transit Information with FLaNK

Timothy Spann

This talk will describe how to use Apache NiFi to read real-time RSS and XML data from transit live event streams and process this data for real-time analytics with Apache Kafka and Apache Flink. We do live ELT with NiFi to add new fields, do lookups and update events as they stream live in the system.

We then store the events in Apache Kudu for querying through Apache Impala via Apache Hue for reporting.

We can then host this data as an API to use with microservices to do real-time route updates and alerts.

My Other Talks & Apache Pulsar Talks

- Tuesday 17:10 UTC - Apache NiFi Deep Dive 300 by Tim Spann
- Tuesday 18:00 UTC - Apache Deep Learning 302 by Tim Spann
- Wednesday 15:00 UTC - Smart Transit: Real-Time Transit Information with FLiP by David Kjerrumgaard & Tim Spann
- Wednesday 15:50 UTC - Replicated Subscriptions: taking Apache Pulsar Geo-Replication to next level by Matteo Merli
- Wednesday 17:10 UTC - Cracking the Nut, Solving Edge AI... by David Kjerrumgaard & Tim Spann
- Wednesday 17:10 UTC - Exclusive Producer: Using Apache Pulsar to build distributed applications by Matteo Merli
- Thursday 14:10 UTC - Apache NiFi 101: Introduction and Best Practices - Tim Spann

Stay Connected With Us!



David Kjerrumgaard
Developer Advocate



<https://twitter.com/DavidKjerrumga1>



<https://github.com/david-streamlio>



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



Tim Spann
Developer Advocate



<https://twitter.com/paasDev>



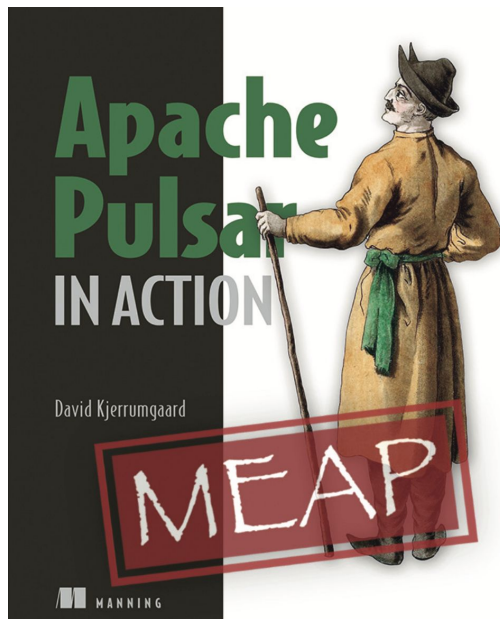
<https://github.com/tspannhw>



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

Speaker Bio

Developer Advocate



<https://github.com/david-streamlio>

<https://pulsar-summit.org/en/event/virtual-conference-2020/speaker/david-kjerrumgaard>

<https://www.slideshare.net/streamnative/using-apache-pulsar-to-provide-realtime-iot-analytics-on-the-edge>



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>



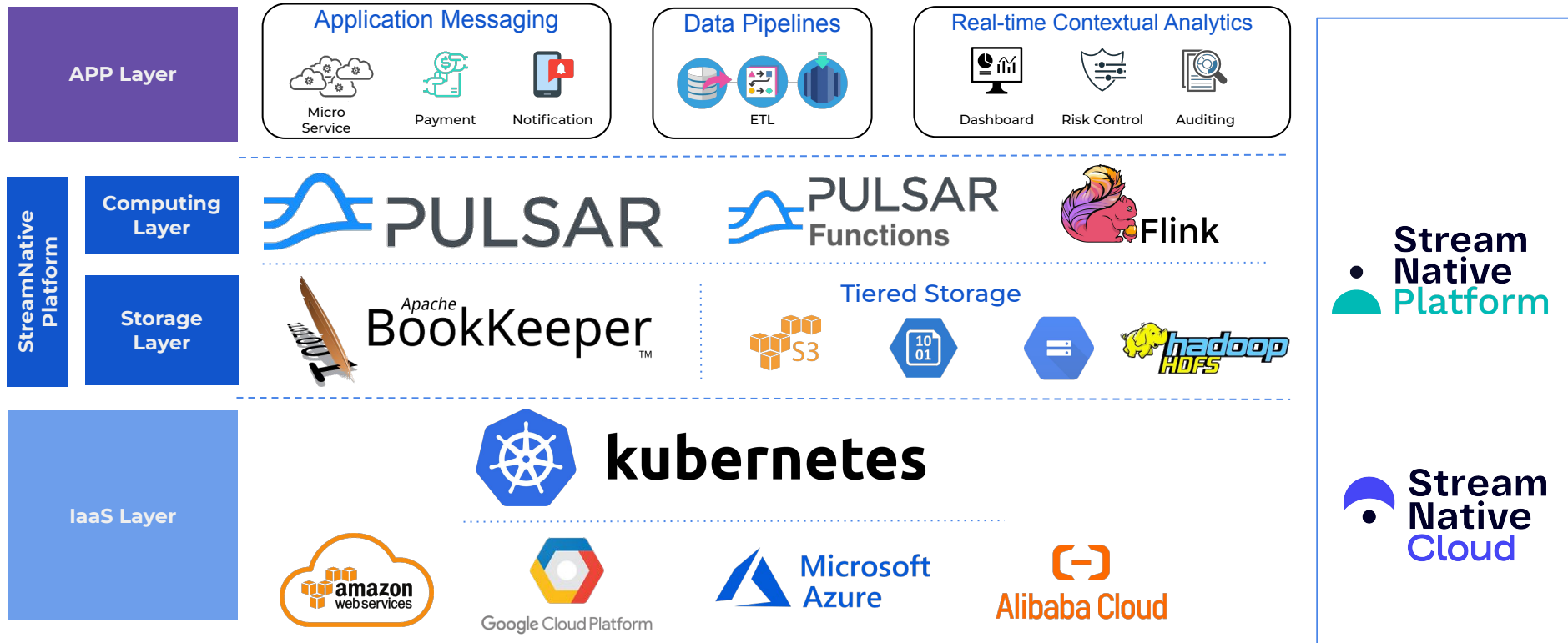
FLiP Stacks

- Apache **F**link
- Apache **P**ulsar
- StreamNative's Flink Connector for Pulsar
- Apache **+**+

Apache projects are the way for all streaming



StreamNative Solution



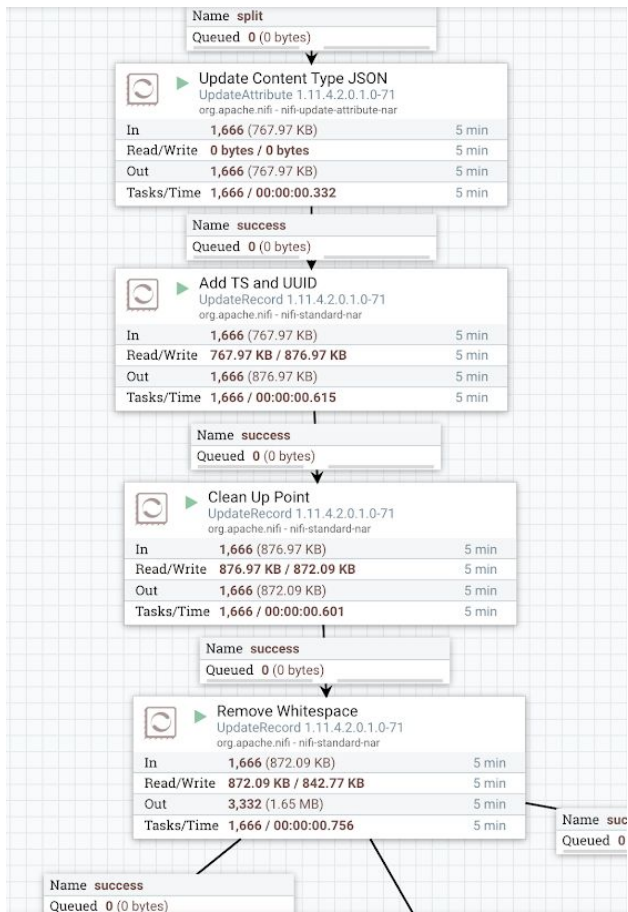
Open Source Cloud Native Frameworks



MQTT on Pulsar (MoP)

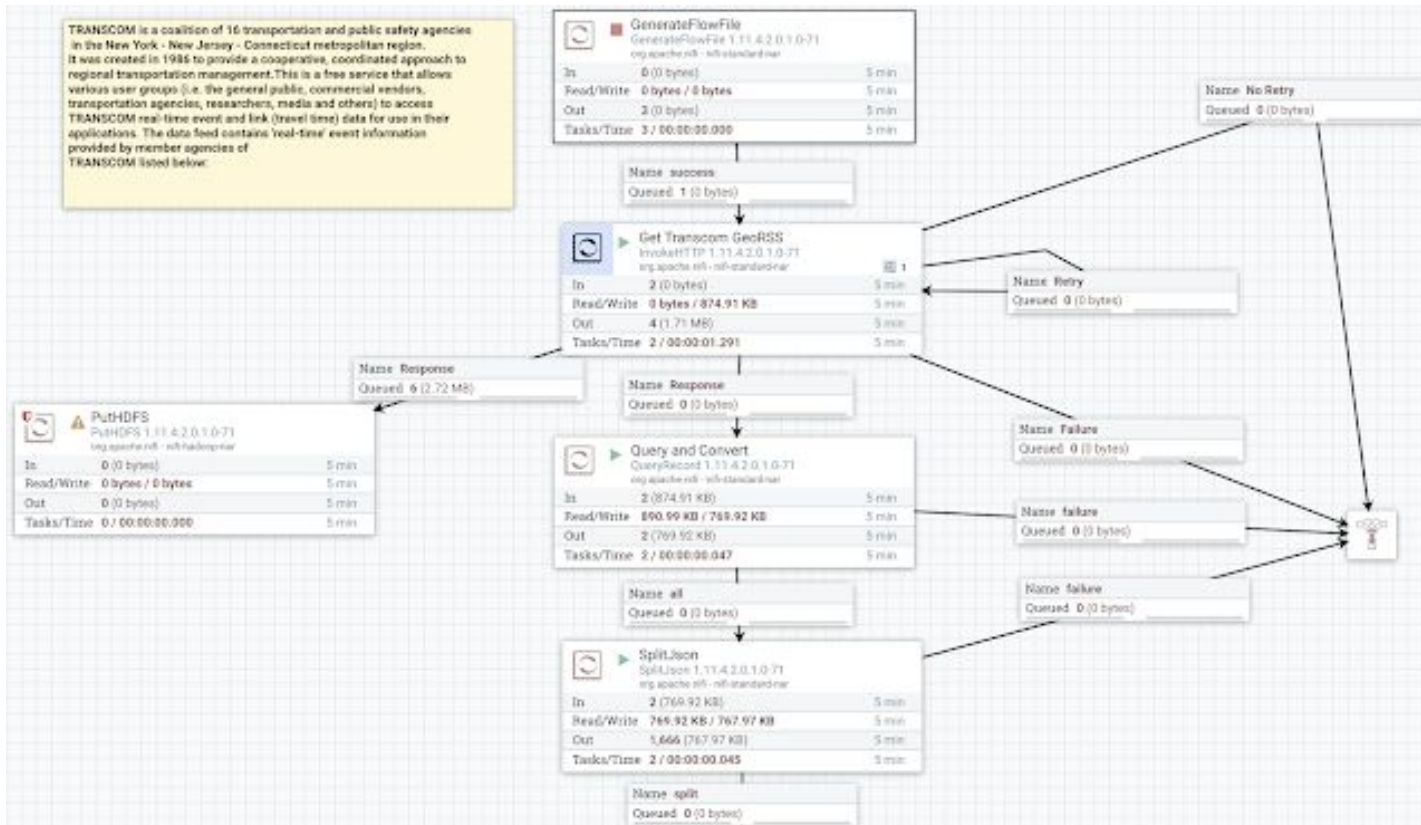
Kafka on Pulsar (KoP)

Clean Up Data In-Stream with Apache NiFi



Ingest Real-Time Transcom GeoRSS Data

TRANSCOM is a coalition of 16 transportation and public safety agencies in the New York - New Jersey - Connecticut metropolitan region. It was created in 1986 to provide a cooperative, coordinated approach to regional transportation management. This is a free service that allows various user groups (i.e. the general public, commercial vendors, transportation agencies, researchers, media and others) to access TRANSCOM real-time event and link (travel time) data for use in their applications. The data feed contains real-time event information provided by member agencies of TRANSCOM listed below:



Using GTFS Data In Apache NiFi

Add Processor

Source

dev.datainmotion



Displaying 1 of 307

Filter

Type ▴

Version

Tags

GTFSRealTimeProcessor

1.0

gtfs,feeds,real-time gtfs,protoc...

amazon attributes

avro aws

consume csv

delete fetch get

hadoop ingest

ingress insert json

kafka listen logs

message pubsub

put record

restricted source

text update

Attribute Values

filename

95fb53d2-1442-4b84-8693-5914d4b1619c

gtfs.alert.279

informed_entity { trip { trip_id: "121500_2..S01R" route_id: "2" 1001: { 1: "02 2015 241/FLA" 2: 1 } } } header_text { translation { text: "Train delayed" } }

gtfs.id.1

000001

gtfs.id.10

000010

gtfs.id.100

000100

gtfs.id.101

000101

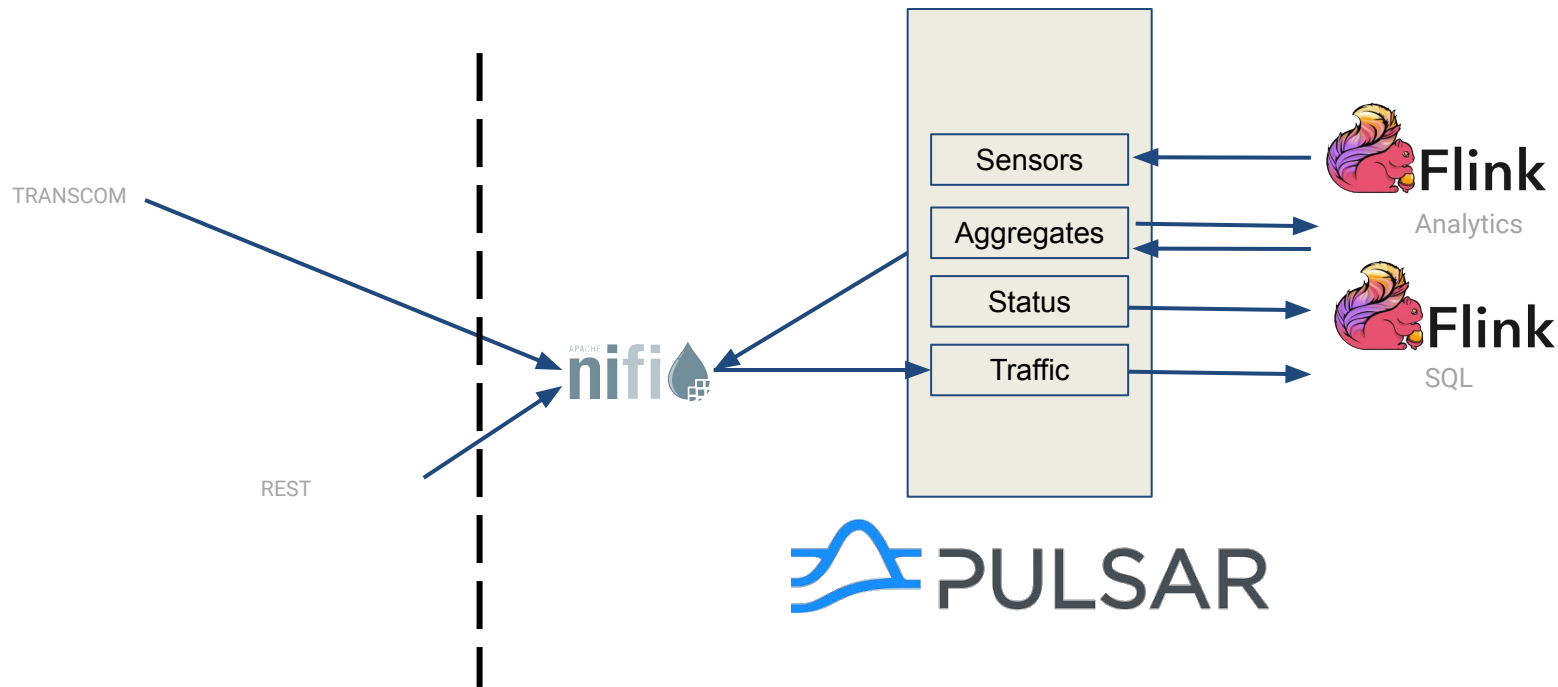
GTFSRealTimeProcessor 1.0 dev.datainmotion - nifi-gtfs-nar

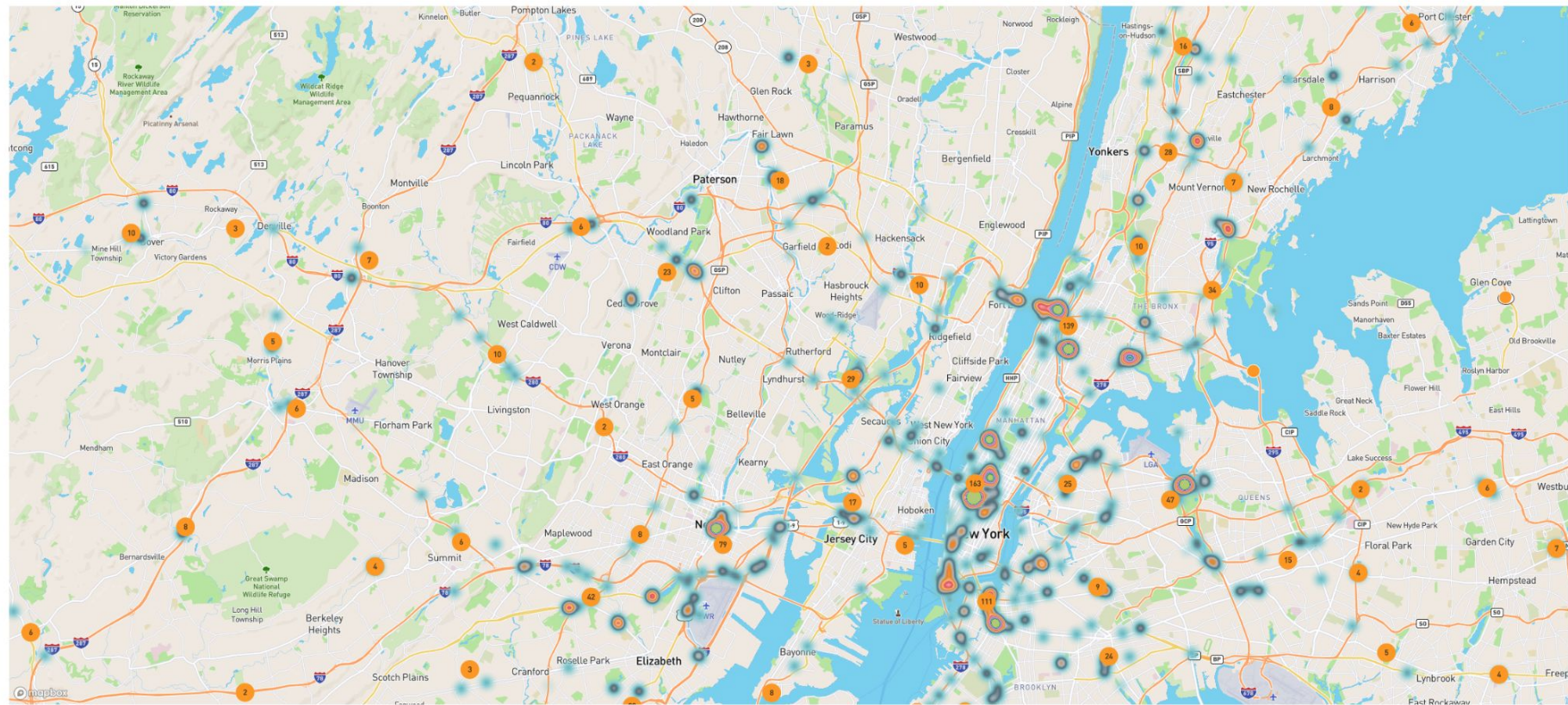
Uses Google GTFS Java library to read and translate GTFS feeds

Demo

Real-Time Cloud Streaming Pipeline

APIs





<https://www.clouddataops.dev/transit-monitoring>

DEMO SOURCE CODE

- <https://github.com/tspannhw/FLiP-Transit>
- <https://github.com/tspannhw/gtfs/>
- <https://github.com/tspannhw/minifi-xaviernx>
- <https://github.com/tspannhw/minifi-jetson-nano>
- <https://github.com/tspannhw/minifi-envioplus>
- <https://github.com/tspannhw/EverythingApacheNiFi>
- <https://github.com/tspannhw/CloudDemo2021>
- <https://www.datainmotion.dev/2019/07/powering-edge-ai-with-powerful-jetson.html>
- <https://www.datainmotion.dev/2021/07/upcoming-apache-pulsar-and-apache-flink.html>
- <https://www.datainmotion.dev/2019/08/gtfs-real-time-streaming-with-apache.html>

Deeper Content

- <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>
- <https://dzone.com/articles/real-time-transit-feed-data-processing>



@PaasDev



timothyspann

<https://www.pulsardeveloper.com/>

Interested In Learning More?



Resources

[Function Mesh - Simplify Complex Streaming Jobs in Cloud](#)

[The GitHub Source Code for Demo](#)



Free eBooks

[Manning's Apache Pulsar in Action](#)

[O'Reilly Book](#)



Upcoming Events

[\[10/6\] Pulsar Summit Europe](#)

Announcing

Flink SQL on StreamNative Cloud



**Stream
Native
Cloud**

