



# PULSTAR SUMMIT 2020

MESSAGING & STREAMING EVERYWHERE

PRESENTED BY



Welcome to the first-ever Pulsar Summit,  
hosted by:



# A Big Thanks to Our Sponsors

Platinum



---

Gold



---

Community



---

Media



# A Big Thanks to the Program Committee



Sijie Guo



Matteo Merli



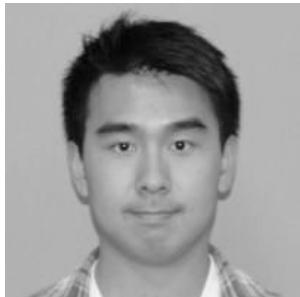
Jia Zhai



Jesse Anderson



Nozomi Kurihara



Jerry Peng



Ben Lorica



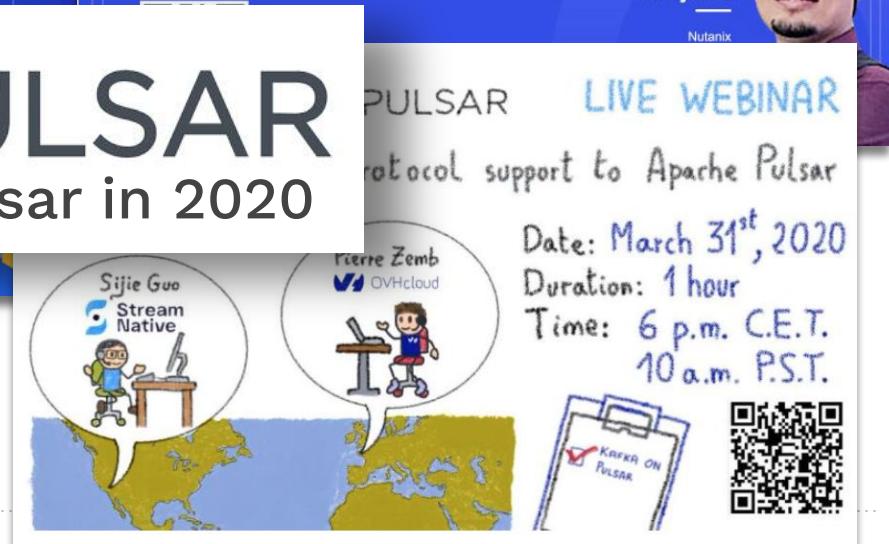
Dave Fisher



Yuvaraj Loganathan

# A Big Thanks to the Speakers





# Pulsar Community & Ecosystem

- Major Product Releases & Updates
- Monthly Webinars
- Weekly Trainings
  - TGIP Every Friday at 1pm PT
- Case Studies, White Papers & Use Cases

Why Zhaopin Chooses Pulsar SQL for Search Log Analysis

How Apache Pulsar Helps Streamline Message System and Reduces O&M Costs at Tuya

Announcing Kafka-on-Pulsar: bring native Kafka protocol support to Apache Pulsar

Announcing AMQP-on-Pulsar: bring native AMQP protocol support to Apache Pulsar

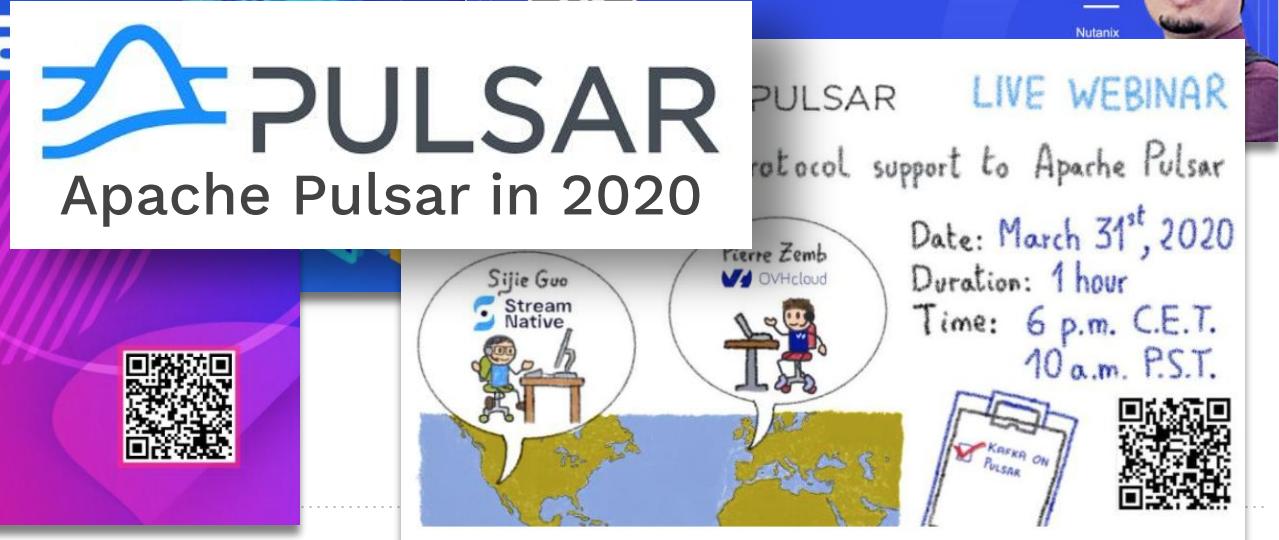
March 24, 2020

June 15, 2020

We are excited to announce that StreamNative and ChinaMobile are open-sourcing AoP. AoP brings the native AMQP protocol support to Apache Pulsar by introducing an AMQP protocol handler on Pulsar brokers.

Penghui Li, Zongtang Hu, Ran Gao

AoP, Pulsar, AMQP, Architecture



# Pulsar Community & Ecosystem

The collage consists of four distinct promotional images:

- Left Image:** A purple and yellow graphic for "TGI Pulsar EP.009". It features a hand-drawn style illustration of two people at desks, one labeled "Sijie Guo Stream Native" and the other "Pierre Zemb OVHcloud". Below them is a world map. The text "kafka on PULSTAR" and "LIVE WEBINAR" is overlaid, along with "Bring native Kafka protocol support to Apache Pulsar". The date "Date: March 31<sup>st</sup>, 2020" and duration "Duration: 1 hour" are also mentioned.
- Middle Left Image:** An orange graphic for "tuya PULSTAR". It features the Tuya logo and the text "How Apache System5000 Chooses Smart". Below it, it says "The success of different Pulsar's add-ons chooses Pulsar". The name "Yonghong Zhang" is at the bottom.
- Middle Right Image:** A pink graphic for "TGI Pulsar EP.011". It features a hand-drawn style illustration of two people at desks, one labeled "Sijie Guo Stream Native" and the other "Pierre Zemb OVHcloud". The text "Apache Pulsar" and "Apache Kafka" is overlaid.
- Right Image:** A blue graphic for "StreamNative Academy | NUTANIX". It features a QR code and the title "Lessons From Managing A Pulsar Cluster". Below the title, it says "Date: May 26, 2020 | Duration: 1 hour | Time: 9 AM PST". A photo of "Shivji Jha" is shown, described as "Nutanix Senior Software Developer". At the bottom, it says "We are excited to announce that StreamNative and OVHcloud are open-sourcing 'Kafka on Pulsar' (KoP). KoP brings the native Apache Kafka protocol support to Apache Pulsar ...". Below this, there are small profile pictures for "Pierre Zemb", "Jia Zhai", and "Sijie Guo".

**PULSTAR SUMMIT** is visible in the bottom left corner.

**PRESENTED BY**

**Stream Native** and **splunk>** are visible in the bottom right corner.

# Get Involved!

#1. Join the Pulsar Slack channel - [Apache-Pulsar.slack.com](https://Apache-Pulsar.slack.com)

- *#PulsarSummit* - connect with fellow attendees in real-time
- *#Job-Board* - post and search Pulsar-related jobs

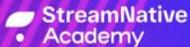
#2. Join Pulsar Summit Newsletter List

- Learn about upcoming webinars, product releases, case studies and more

#3. Follow [@apache\\_pulsar](https://twitter.com/apache_pulsar) on Twitter

#4. Take the Post-Summit Survey!

# Get Involved!



## #1. Join the Pulsar Slack channel - [Apache-Pulsar.slack.com](https://Apache-Pulsar.slack.com)

- **#PulsarSummit** - connect with fellow attendees in real-time
- **#Job-Board** - post and search Pulsar-related jobs

## #2. Join Pulsar Summit Newsletter List

- Learn about upcoming webinars, product releases, case studies and more. At the bottom of the

## #3. Follow [@apache\\_pulsar](https://twitter.com/apache_pulsar) on Twitter



A Twitter profile card for Yonghong Zhang. It features the Pulsar logo at the top left. The name "Yonghong Zhang" is in the center, with "Pulsar SQL, Tuya Smart" below it. There is a small profile picture of a person with a question mark icon.

## #4. Take the Post-Summit Survey!



PRESENTED BY  
 Stream Native  splunk>

# Apache Pulsar in 2020

## First-Ever Apache Pulsar Summit

- 36 speakers
- 35+ sessions
- 550+ attendee sign-ups
- 300+ companies represented



# Conference Overview & Logistics



In-Person Summit  
San Francisco, CA  
March 2020

Virtual Summit  
June 2020

# Conference Overview & Logistics

2 Days / 3 Tracks / 3 Zoom Links

TIME	TRACK 1	TRACK 2	TRACK 3
8:30-9:10	Keynote: Adoption, Use Cases & The Future of Pulsar (TRACK 1)		
9:20-10	Keynote: Why Splunk Chose Pulsar, by Karthik Ramasamy (TRACK 1)		
10:20-10:50	Scaling Customer Engagement...	Kafka on Pulsar...	Pulsar Storage on Bookkeeper...
11-11:40	Getting Pulsar Spinning...	Event Propagation across...	Pulsar Functions Deep Dive...

# Pulsar Summit Track Moderators



Track 1: Carolyn King  
Marketing, StreamNative



Track 2: Jun Wang  
Marketing & Events



Track 3: Rosalie Bartlett  
Sr Community Mgr, Verizon Media

Contact us at [events@streamnative.io](mailto:events@streamnative.io) with any questions!

# KEYNOTE SESSION:

# Messaging and Event Streaming

## Adoption, Use Cases and the Future of Pulsar

Matteo Merli / Splunk  
Sijie Guo / StreamNative

# Who are we?



- Matteo Merli ([@merlimat](#))
- Sr. Principal Engineer, Splunk
- Co-creator and PMC chair of Pulsar
- Ex Co-Founder, Streamlio



- Sijie Guo ([@sijieg](#))
- Co-Founder, StreamNative
- PMC Member of Pulsar/BookKeeper
- Ex Co-Founder, Streamlio

# Splunk

Splunk provides operational intelligence software that monitors, reports, and analyzes real-time machine data.

Splunk acquired Streamlio in Nov 2019 as part of an expanded investment in the data streaming space.

Splunk is using Pulsar in multiple product lines and it's deeply committed in further developing Pulsar and fostering its community.

# StreamNative

Founded by the developers of Apache Pulsar and Apache BookKeeper, StreamNative enables companies to access enterprise data as real-time event streams.

- Contributing and helping the Pulsar/BK community grow
- Helping people resolve business problems using Pulsar
- Providing managed Pulsar services and enterprise support

# Agenda

- How Organizations are Using Pulsar Today / Sijie Guo
- What is Driving Pulsar Adoption / Sijie Guo
- The Future of Apache Pulsar / Matteo Merli

# What is Apache Pulsar?

# Pulsar is a cloud-native messaging and event streaming platform

# Pulsar's Global Adoption





# Splunk

The Data-to-Everything™  
Platform

- **Industry:** IT
- **Adoption:** +6 months
- **Market Cap:** 29B

## Case Study Highlights

#1 Splunk Data Stream Processor  
based on Apache Pulsar

#2 Streaming and Batch connectors

#3 Pulsar as a Service for Splunk cloud  
products



# Narvar

Intelligent Customer  
Experience Platform

- **Industry:** Retail
- **Adoption:** 1.5 years
- **Scale:** 50k txns/second
- Mission Critical Applications

## Case Study Highlights

- #1 Real time transactional messaging
- #2 Data integration with Data Lake
- #3 Complex event processing
- #4 Heavy Pulsar Functions user

# Instructure

Educational Technology  
Company

- **Industry:** Education
- **Adoption:** 1+ years
- **Scale:** 8 AWS regions, 50k  
msgs/sec in the busiest region

## Case Study Highlights

#1 Low-cost

#2 Easy to manage long term retention

#3 Unified messaging model



clever cloud

# Clever Cloud

PaaS Company

- **Industry:** Cloud Computing
- **Adoption:** 1+ years
- **Use Cases:** log ingestion pipeline  
and Function as a Service

## Case Study Highlights

#1 Multi-tenant queue system

#2 Proxy Architecture

#3 Presto and S3 integration

# Tencent

The Wechat Company

- **Industry:** Internet
- **Use case:** Financial & Log pipeline
- **Adoption:** 2+ years
- **Scale:** 10s billions of financial txns  
every day

## Case Study Highlights

#1 Powering Tencent Billing Platform

#2 Data transfer layer for federated machine learning platform

#3 Replace Kafka for its logging pipeline in Tencent Games



# Huya Live

Live Streaming Service

- **Market Cap:** 4B
- **Use case:** Log collection
- **Adoption:** 1+ year
- **Scale:** 15 millions msgs/sec

## Case Study Highlights

- #1 Replace Kafka in its log pipeline
- #2 Instant scalability
- #3 Multi Tenancy



YumChina

# Yum China

American Fortune 500  
fast-food company

- **Revenue:** 8B
- **Use case:** Notification & Order Processing
- **Adoption:** 6+ months

## Case Study Highlights

#1 Replace RabbitMQ

#2 Instant scalability

#3 Multi Tenancy



# Global Adoption



</ndustrial.io>

**NUTANIX**™

**THE HUT GROUP**®



**proxyclick**

**OVHcloud**

**tuya.com**

**IoTium**

**OKCOIN**

**COMCAST**

**Tencent** 腾讯

**splunk**®

**Huya**

**mercado libre**

**narvar**

**toast**

**verizon media**

**YAHOO! JAPAN**



**ITERABLE**

**overstock**

**中国电信**  
CHINA TELECOM

**360**  
www.360.cn

**BestPay**

**BrandsEye**

**VIP KID**

**5G+** | **中国移动**  
China Mobile

**Kingsoft Cloud**  
KS CLOUD

**Max Kelsen**

**环球易购**  
Globalegrow E-Commerce

**EMQ**

**PULSTAR  
SUMMIT**

PRESENTED BY

**Stream Native** **splunk**®

# What is driving Pulsar Adoption?

---

-

## Insights from the Pulsar User Survey 2020

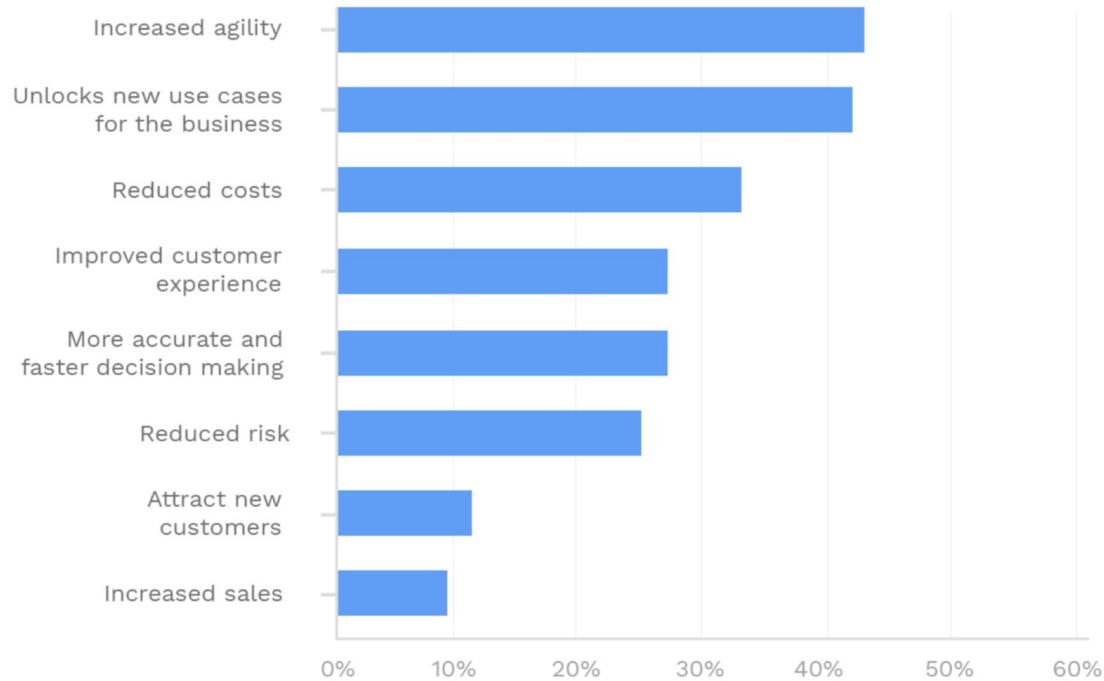
# What value do Pulsar bring to your organization?

#1 Increased Agility

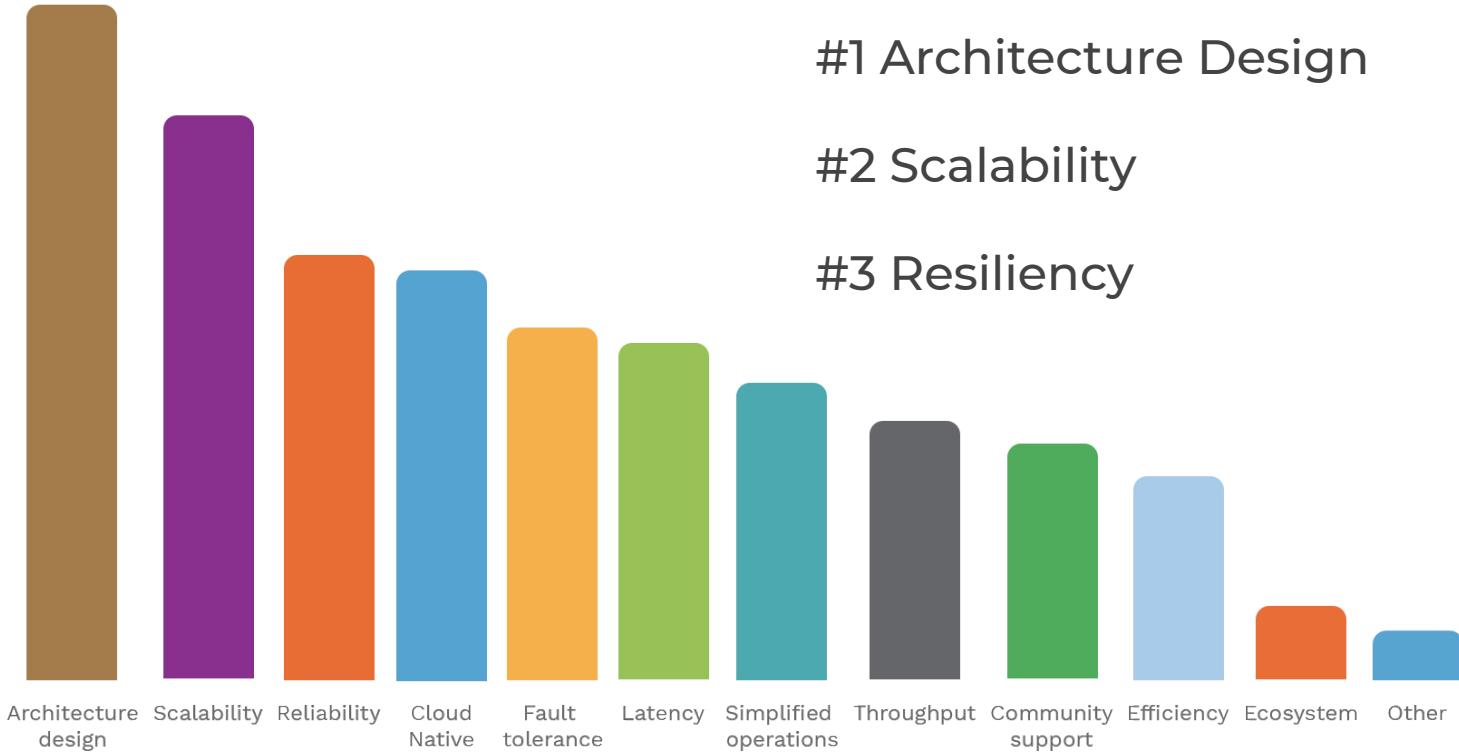
#2 Unlocks New Use Cases for the Business

#3 Reduced Costs

#4 Improved Customer Experience



# What are the top 3 highlights for Pulsar?

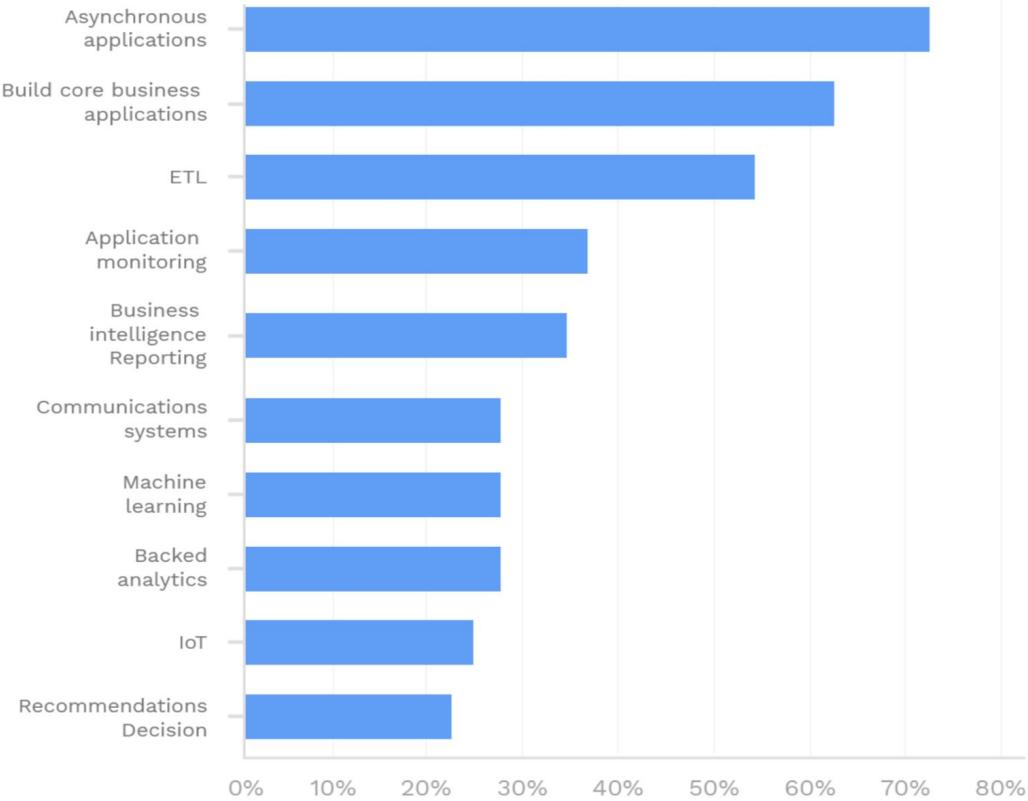


# Top Use-Cases

#1 Asynchronous Applications

#2 Building Core Business Applications

#3 ETL / Data Pipelines



# Most-Used Features

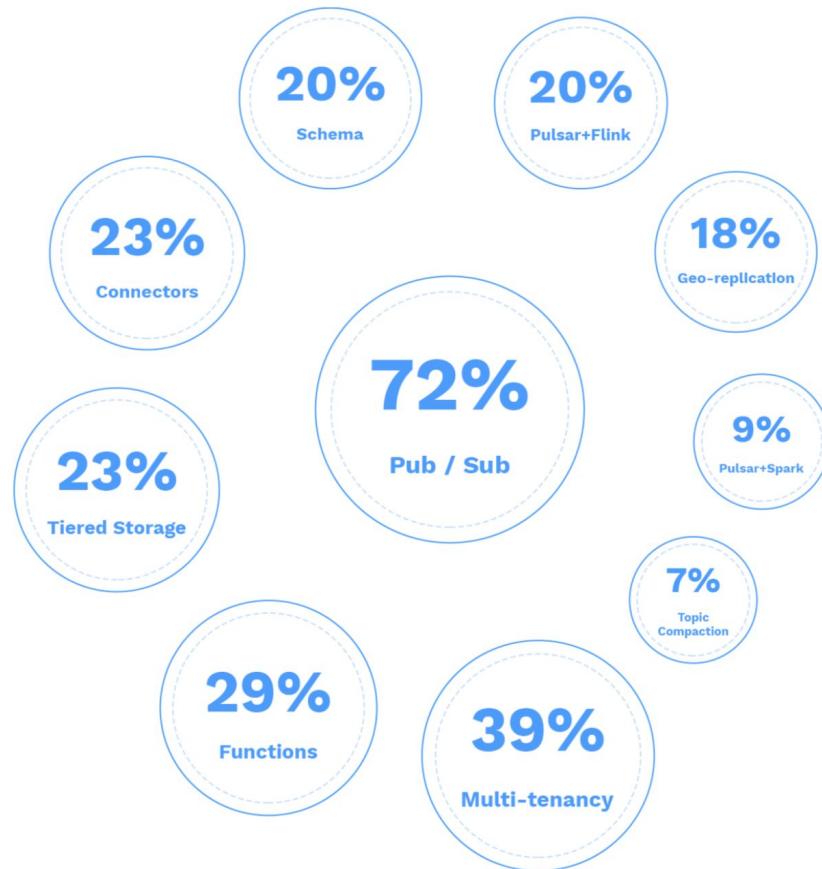
#1 Pub/Sub

#2 Multi-Tenancy

#3 Functions

#4 Tiered Storage

#5 Connectors

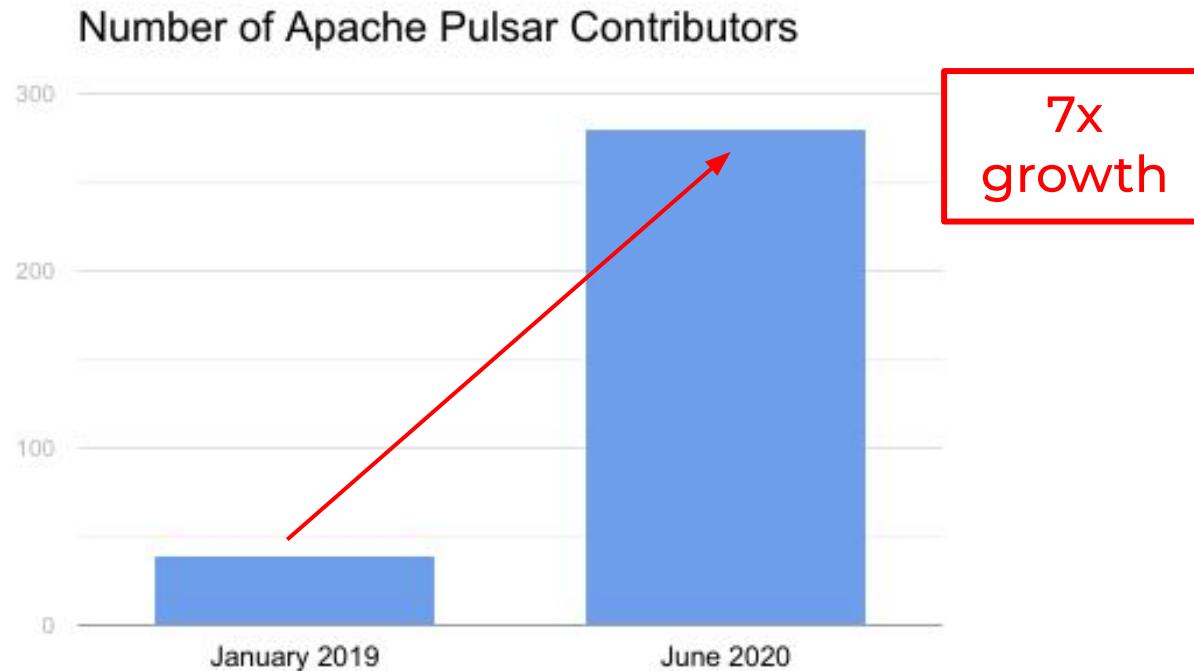


# 42%

Consider Pulsar to replace two or more messaging systems,

Because Pulsar is a **unified messaging and event streaming** platform.

# Apache Pulsar: 18-mo Growth Rate



# Apache Pulsar today

The Apache Pulsar community has shaped the current Pulsar as -

**A cloud-native messaging and event streaming platform**

-

**Pub/Sub**

**Store**

**Process**

# The future of Apache Pulsar

# History

- ✓ 2012 — Pulsar inception at Yahoo
  - ✓ Mandate: scalable, multi-tenant messaging service
- ✓ 2016 — Open-Sourced by Yahoo
- ✓ 2017 — Project migrated to Apache Software Foundation
- ✓ 2018
  - ✓ Promoted to Top-Level Apache project
  - ✓ Apache Pulsar 2.0 released
- ✓ 2019 — Adoption increase

# Pulsar today

- Huge growth for the project
  - Scope and features
  - Community
- We have been able to build a lot without changing the core of the system
- The path has been linear:
  - Focus on providing the best infra for messaging and event streaming
  - Build on the architectural strength of Pulsar
  - Listen to the users, make their life easier

# Pulsar evolution

1. Pub-Sub implemented over distributed log-storage
2. Schema
3. Pulsar Functions
4. Pulsar IO
5. Pulsar SQL
6. Tiered Storage

# Messaging

---

Publish and consume events at scale **from anywhere** using **any protocols and languages**

# Protocol Handler (\*oP)

- KoP: Kafka-on-Pulsar
- AoP: AMQP-on-Pulsar
- MoP: MQTT-on-Pulsar



# New Features

- Transaction Support: lot of progress, coming in 2.7
- REST API to produce / consume
- Readonly brokers:
  - High fanout
  - Scale brokers on demand without affecting ownership
- Exclusive Producer
  - Single writer to provide fencing and leader election for applications

# Partitions auto-scaling

- Partitions are an artifact of scaling
- System complexity should be hidden
- Pulsar should be able to automatically manage partitions:
  - Increase / Decrease based on load
  - Retain ordering
- Remove duality of partitioned/non-partitioned topics

# Pluggable metadata store

- Pulsar uses ZooKeeper as a metadata store and coordination service
- Work in ongoing to abstract the metadata access layer
- Soon it will be possible to choose from different backend implementations
- In future, we would provide out of the box metadata support in Pulsar

# Storage

---

Continue to push the boundaries for truly scalable stream storage

Performance - Operability - Cost

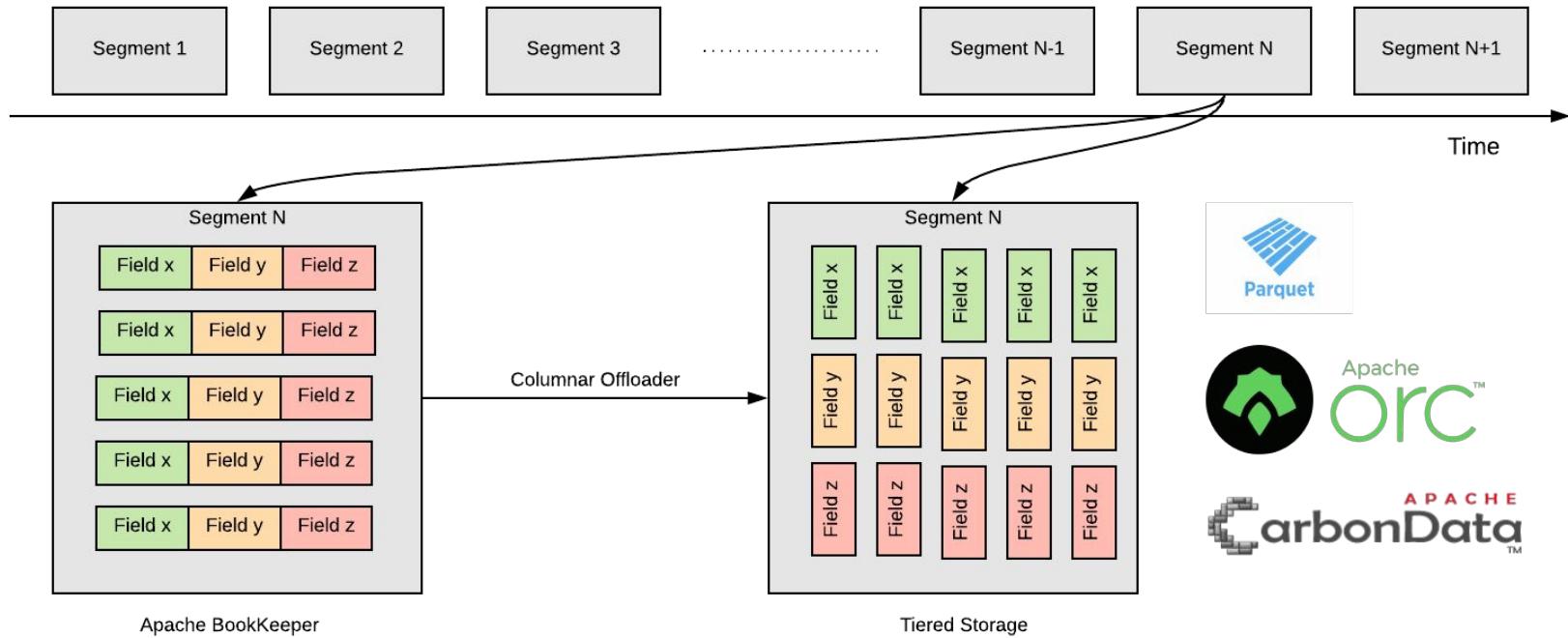
# Storage

- A lot of behind the scenes work has been happening on performance
- Continuously ensure that Pulsar + BookKeeper are the most effective platform to store data in every environment:
  - Cloud / Multi-Cloud
  - On-Prem
- Efficiently support wildly different requirements:
  - Strong consistency and durability
  - Low cost and huge throughput

# Storage evolution

1. Distributed log-storage
2. Schema — Structured storage
3. Tiered Storage — Infinite stream capacity
4. Topic Compaction — Table & Stream duality
5. Key-Value — Functions state access

# Storage - Columnar Offloader



# State Store

Key-Value store used by Pulsar Functions

- Maturing Global State
  - Hot-replicas for super-fast failovers
  - Monitoring
  - Extensive testing
- Change data capture of state updates
  - Local read access to cached values
  - Efficiently support read-intensive data accesses

# Processing

---

Built-In processing + Integrate with existing platforms

# Processing

Process event streams in real-time at scale

- Pulsar Functions → lightweight / serverless compute
- Pulsar-Flink / Pulsar-Spark → batch and stream processing
- Pulsar SQL → interactive queries with Presto

# Pulsar-Flink

- [FLIP-72: Introduce Pulsar Connector](#)
- Batch reader: for batch processing
  - Segment reader
  - Bypassing brokers
  - Read segments from Apache BookKeeper and Tiered Storage
- Sub-stream reader: for scale-out stream processing
  - Key\_Shared subscription & readers
  - Read from brokers
  - Scale the processing parallelism beyond the number of partitions

# Event storage API

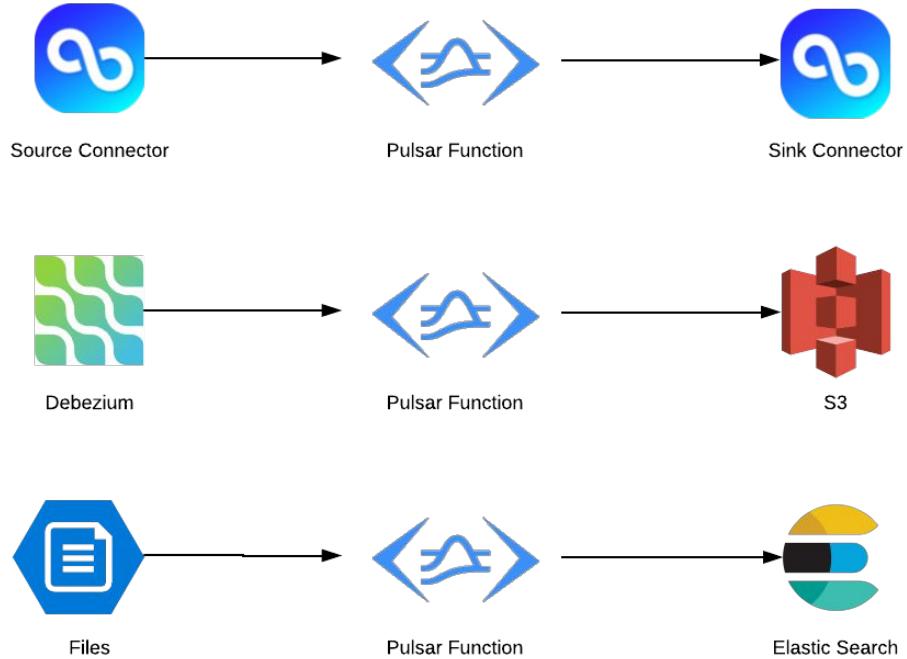
- Provide multiple access layers to the data
  - Stream Reader: read events in a partition-based order
  - Sub-stream Reader: read events in key-based order
  - Segment Readers: read segments from Apache BookKeeper and Tiered Storage
- Integrations
  - Pulsar-Flink
  - Pulsar-Spark
  - Pulsar-Presto

# Pulsar Functions

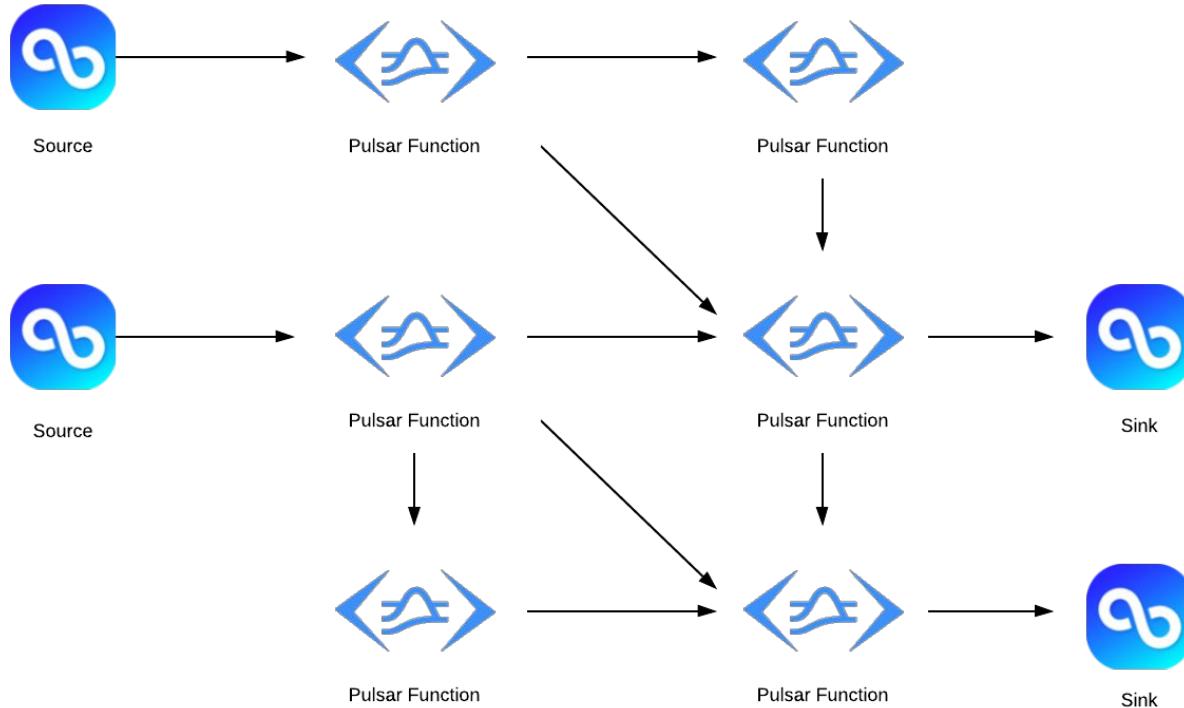
- Pluggable Language Runtime
- Function registry
  - Share and reuse functions
- Operability
  - Functions Versioning
  - Upgrade / Rollback
  - A/B testing
- Connectors (Batch / Stream connectors)

# Function Mesh

- PIP-66: [Function Mesh](#)
- Compose multiple sources, sinks and functions together
- YAML Config or DSL



# Function Mesh



# Managements Tools

- Pulsar Manager
  - Support Schema, Functions & Connectors
  - Integrate with BookKeeper Visual Manager
- PulsarCtl - Go based CLI admin tool
- Pulsar Helm Chart
- Kubernetes Operator
- Tenant & Topic level configuration policies
- Broker interceptors
  - Pluggable provider for tracing and extending broker capabilities
- System Events/Topics - Change Event Streams



# PULSTAR SUMMIT 2020

MESSAGING & STREAMING EVERYWHERE

PRESENTED BY

