

OSA CON 25

Streaming Analytics in Action: Real-World Case Studies from Uber, Razorpay, and Stripe

Jayesh Asrani

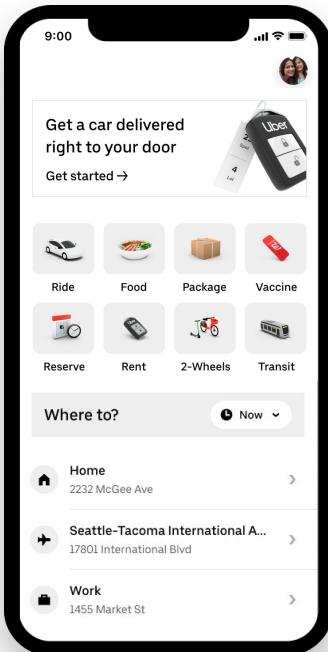
Principal Solutions Architect, StarTree

November 4-5, 2025

Stripe Black Friday Live Microsite

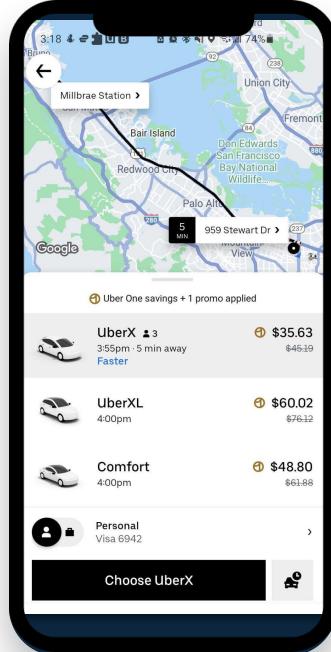


Uber Powering User Experience



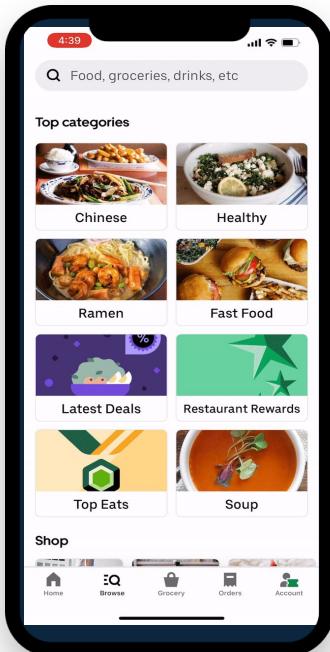
Go anywhere

> 1 Trillion Kafka events/day



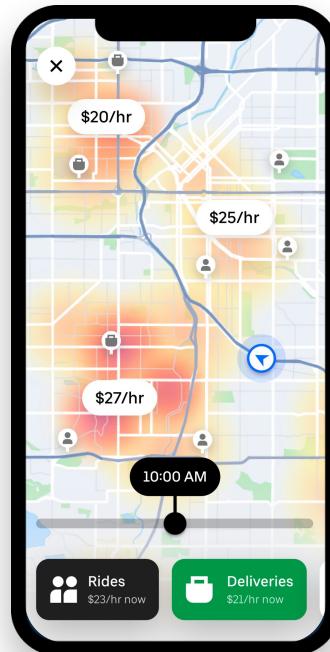
Personalization

~10M QPS ML predictions



Get Anything

3k+ QPS at < 80 ms latency p99



More Work =
More Opportunities

Surge Pricing Heat Map

Razorpay Success Rate Merchant Dashboards

22,376	>
Customer-related ⓘ	
6,220	
Banking-related ⓘ	
0	
Business-related ⓘ	
0	
Other ⓘ	

| Top payment failure reasons: Customer-related

22,214

Payment timed-out

142

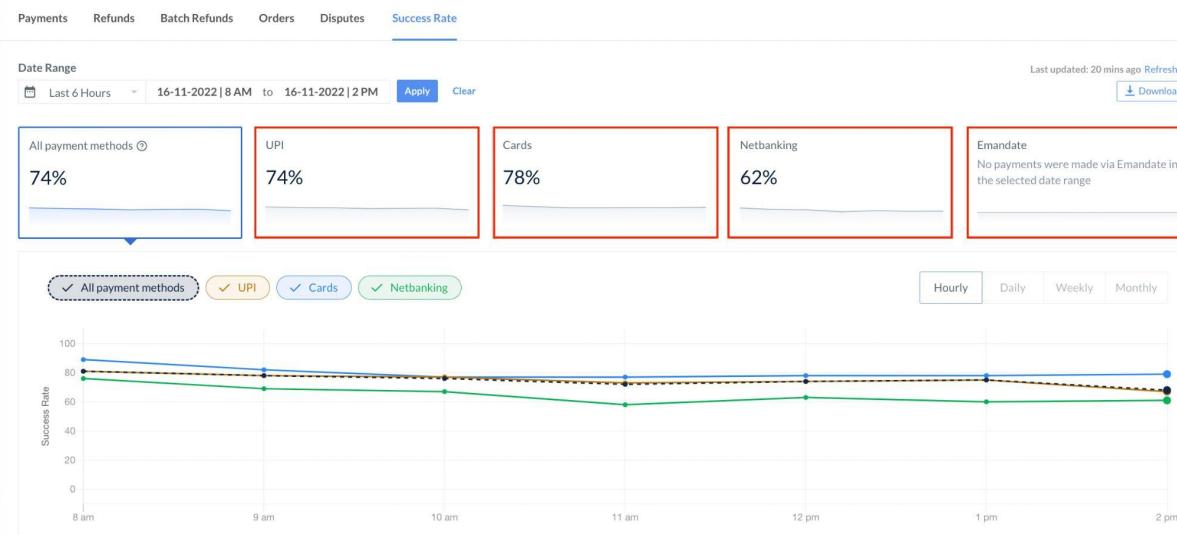
Incorrect UPI ID

16

Bank transaction limit exceeded

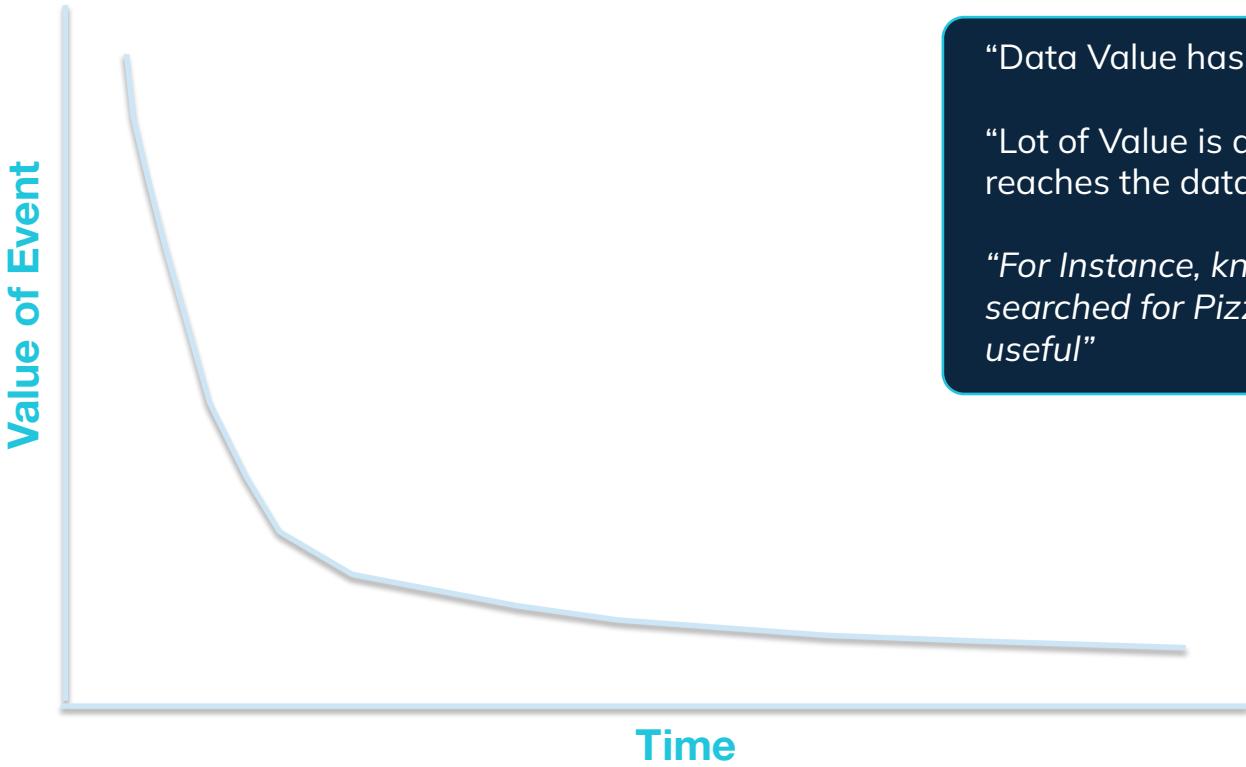
4

Payment was unsuccessful as you may not be registered on the app you're trying to pay with. Try using another method.



Why do we care about fast Analytics ?

The value of data declines over time



“Data Value has half-life”

“Lot of Value is already lost by the time it reaches the data consumer”

“For Instance, knowing that someone searched for Pizza the next day is not useful”

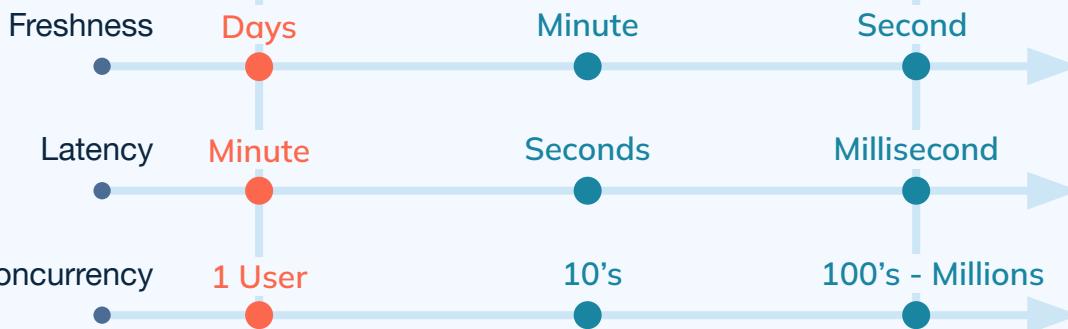
Challenges of Real Time Analytics

Past — Present



Dashboards

- Multiple Tools
- Information Overload
- Limited Audience



Present — Future

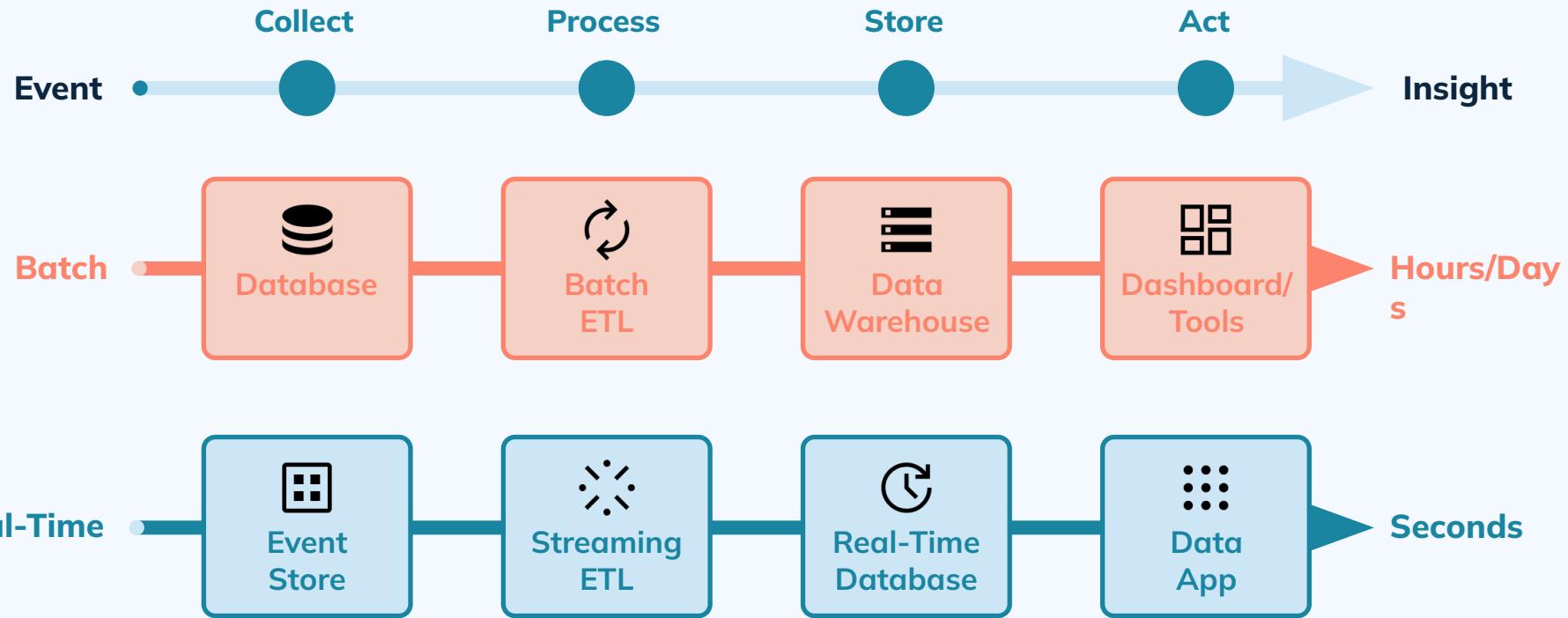


Data Apps

- Live
- Contextual
- Everyone

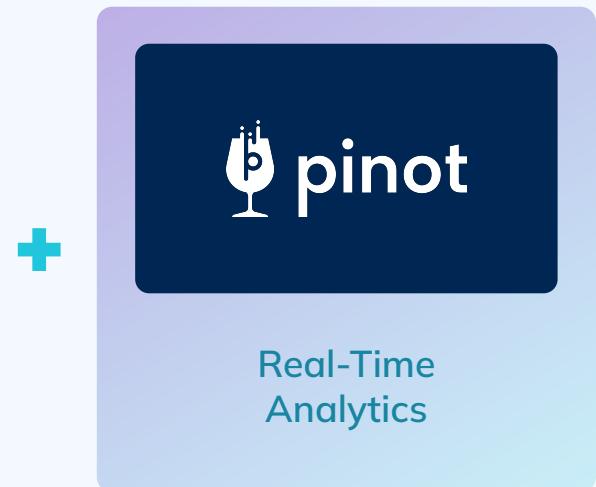
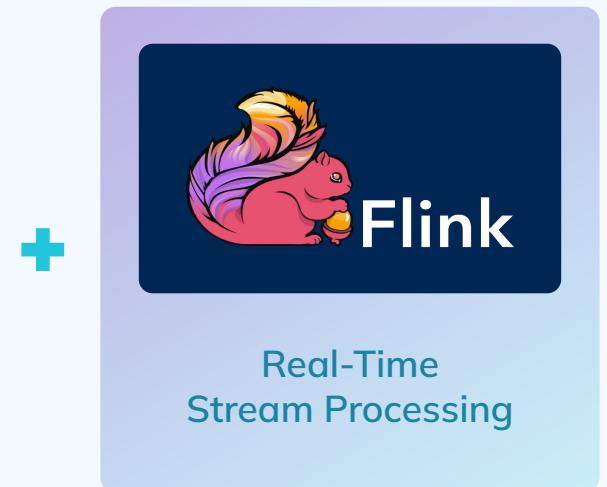
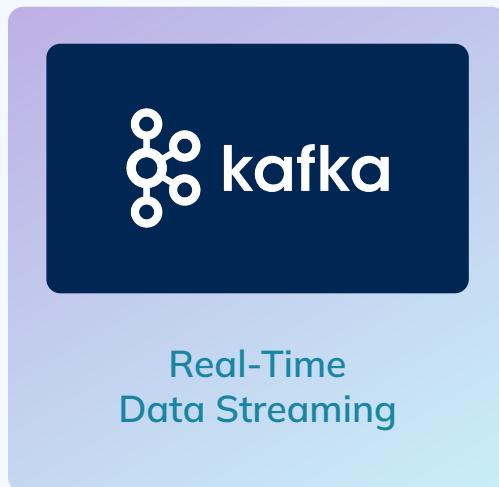


Rise of Real-Time Architecture



Kafka + Flink + Pinot = The “KFP” Stack

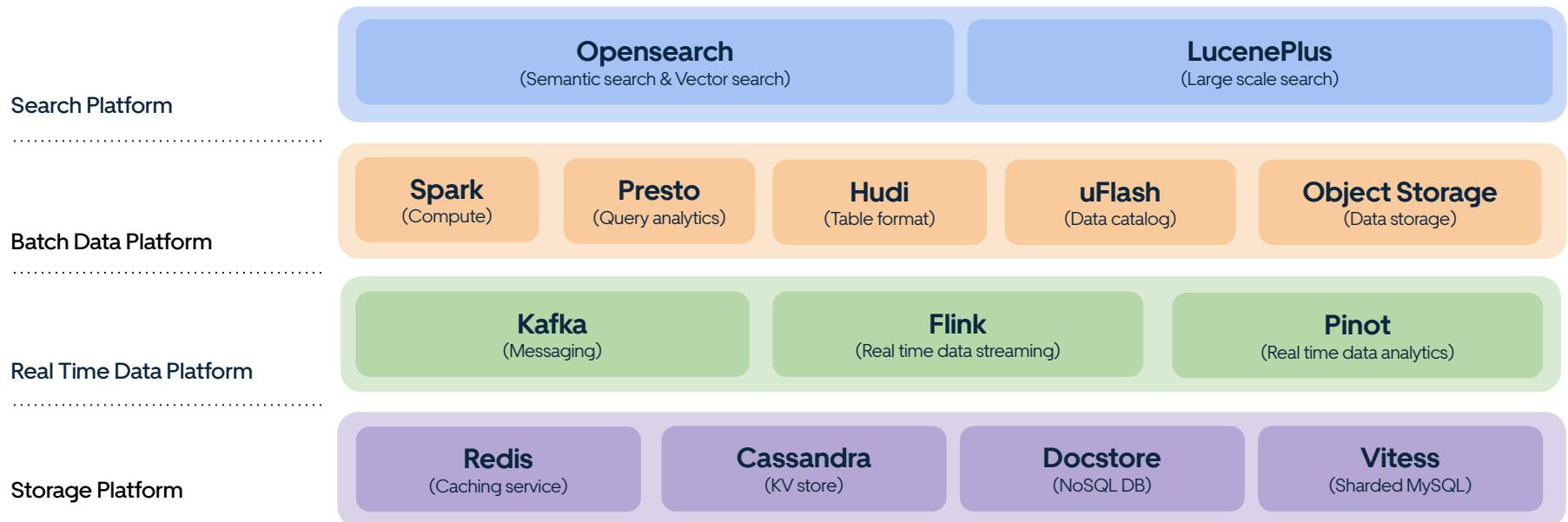
For End-to-End Real-Time Data Architectures



Uber!

Platform Engineering

Technology Overview



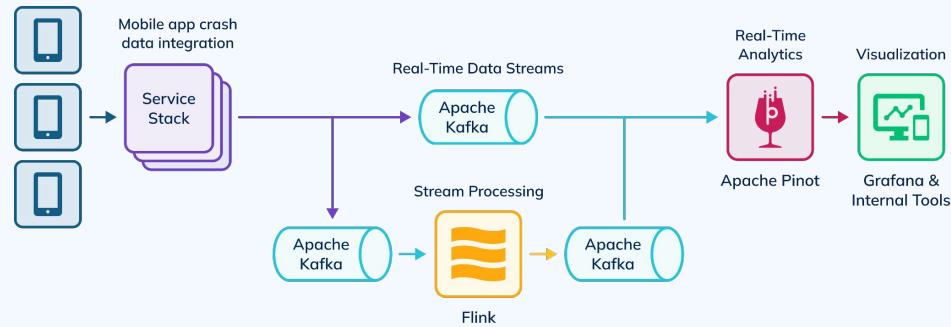
Uber Healthline



Serving Real-Time App Crash Analytics while Saving \$2M+ With Apache Pinot

After moving to Pinot:

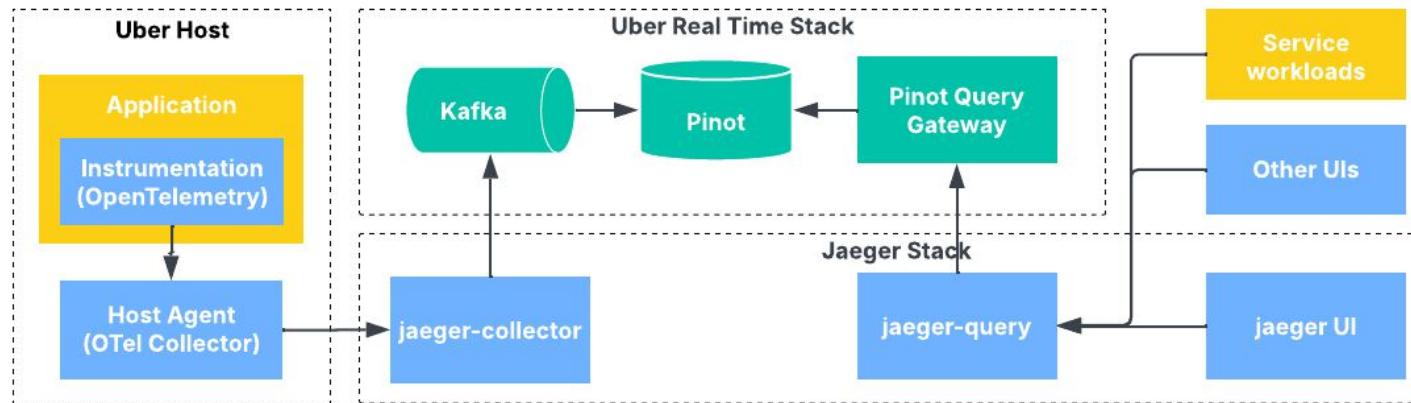
- Removed 19K+ CPU cores
- 50% reduction in DB cores
- Page load time reduced from 14 secs to 5 secs
- Reduced ingestion lag to <10 milliseconds
- Decline in query timeouts and elimination of data loss issue



[Blog](#) on Uber Healthline Crash Analytics



Tracing System Overview



Current State

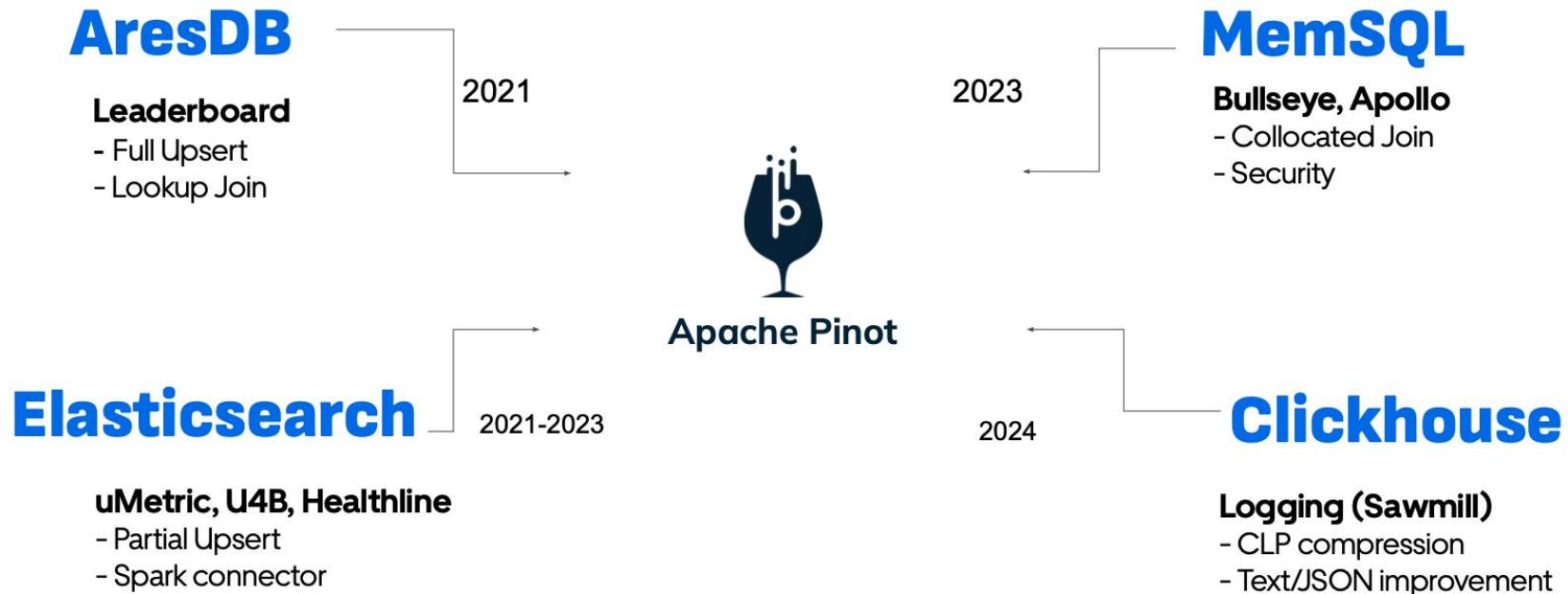
3M
Spans ingested / s

3.2 GBps
Daily peak ingestion rate

100
Traces retrieved/ s

0.6 PB
Storage Footprint
(7 days retention)

Database Consolidation : Uber



Scale By The Numbers

14T

Kafka

Messages/day

5000+

Flink

Jobs across Uber

600M

Pinot

Queries/day
20PB of data size

600K

Presto

Queries/day
80PB a day

1.4 EB

HDFS

Physical capacity
in-use

94M

Docstore

Queries/second
180+ PB data size

34M

Cassandra

Queries/second

Stripe!

Stripe Merchant Dashboards



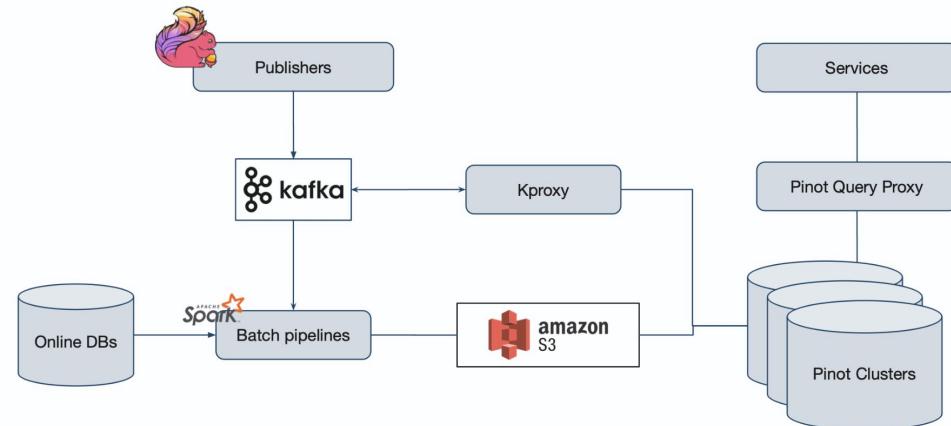
Powers dozens of Merchant facing dashboard chats

SLA:

- 20k QPS with 200 distinct queries
- p99 query latency < 70ms
- p99 ingestion of < 30s
- 99.99% availability

Challenges with in-house system “Decibel” :

- Multiple hops to Kafka to limit the write load on MongoDB
- Duplicates would occur under failure conditions
- leading to inaccurate metrics on the dashboards



[Video](#) of Johan Adami (Software Engineer, Stripe)

Use Cases Overview : Stripe

Customer Facing Analytics

- Stripe Dashboards
- Billing Analytics
- Sigma Reports
- Developer Analytics

Internal Analytics

- Security Tools
- Financial Data Reporting
- Risk Dashboards

80B
Kafka

Messages/day

1 PB+
Data

Data In Pinot

1.5B
Pinot

Queries/day

Razorpay!

Razorpay



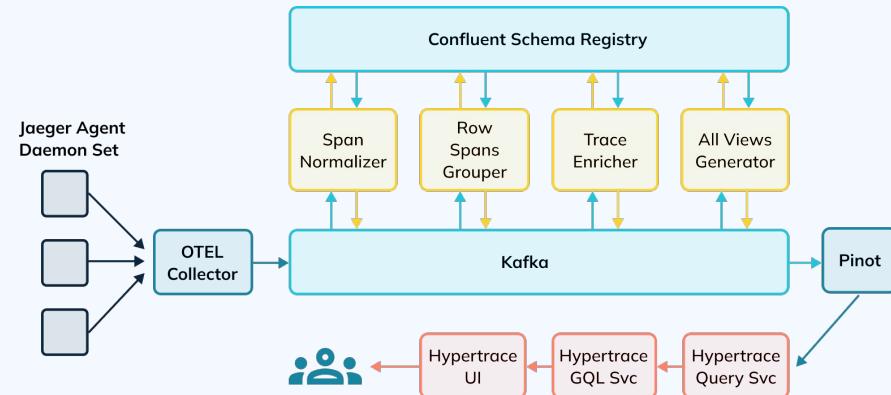
Powers Merchant Dashboards, Health of Banks and Merchants

Outcome:

- Save 50% on AWS infrastructure Cost compared to OSS Pinot Cluster
- Consolidated analytical tech stack onto Pinot, **300K - 1M events/sec, TB+/day**
- Sub-second query response

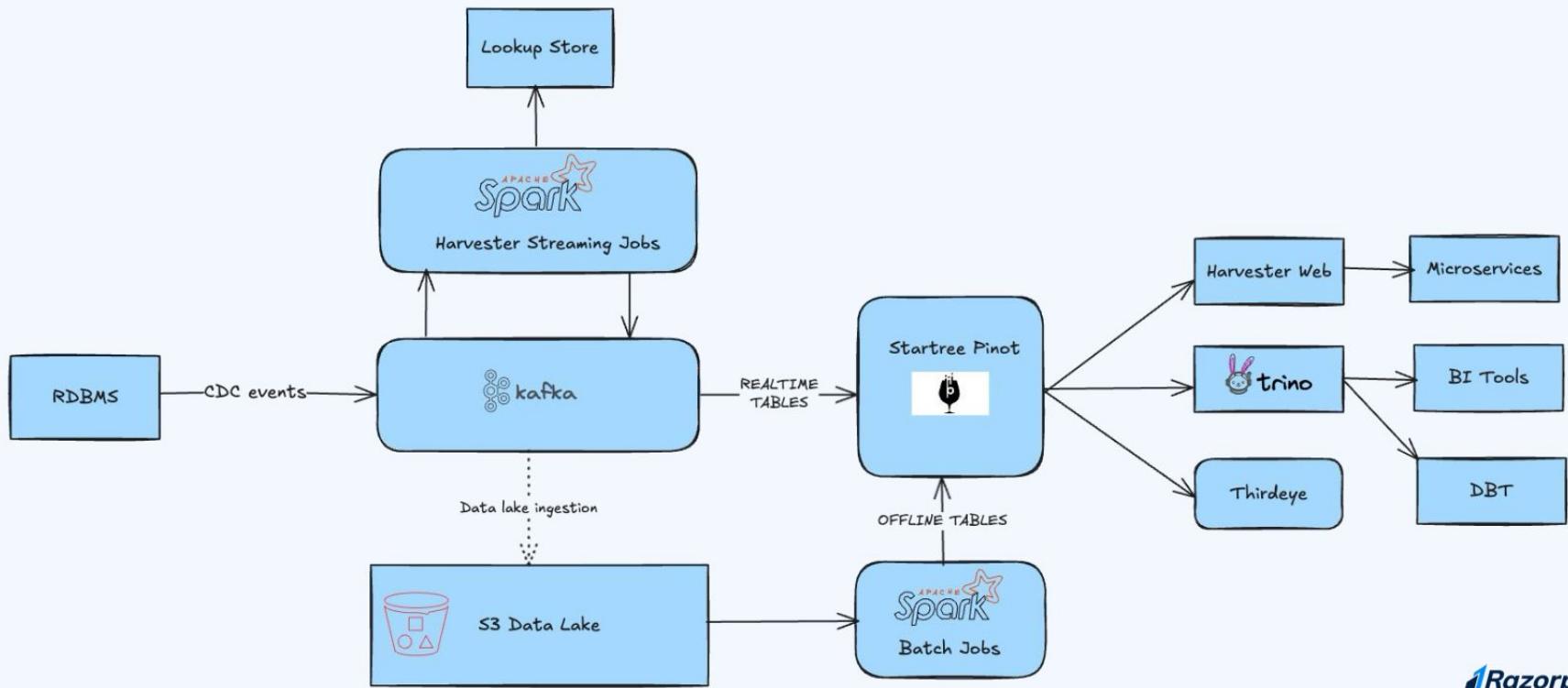
Selected Pinot over Elastic, Druid, and Presto

[Video](#) - Monitoring Payment Success Rates with Apache Pinot



Razorpay easily powers their real-time analytics stack with Apache Pinot

Streaming Architecture - Razorpay



Use Cases Overview : Razorpay

Customer Facing Analytics

- Payment Analytics on Merchant Dashboards
- Microservices Supporting Merchant-Facing Flows

Internal Analytics

- Real Time Analytics & Monitoring
- Anomaly Detection
- Operations Automation

30+

Pinot Servers

50+

Production Tables

~28TB

Data Size

~30,000

records/sec Ingestion Throughput

>5M

Queries/Day

500ms-2s

P99 Query Latency

Keep up with the Community!

Connect with Me!



<https://www.linkedin.com/in/jayesh-asrani/>

Join Pinot OSS Slack



<stree.ai/pinot-slack>



Copyright © 2024 StarTree, Inc. All Rights Reserved.

Your Center for All
Things Apache Pinot™

Explore resources, events,
and latest updates.



<dev.startree.ai>

Thank You