Stream, Materialize, Serve

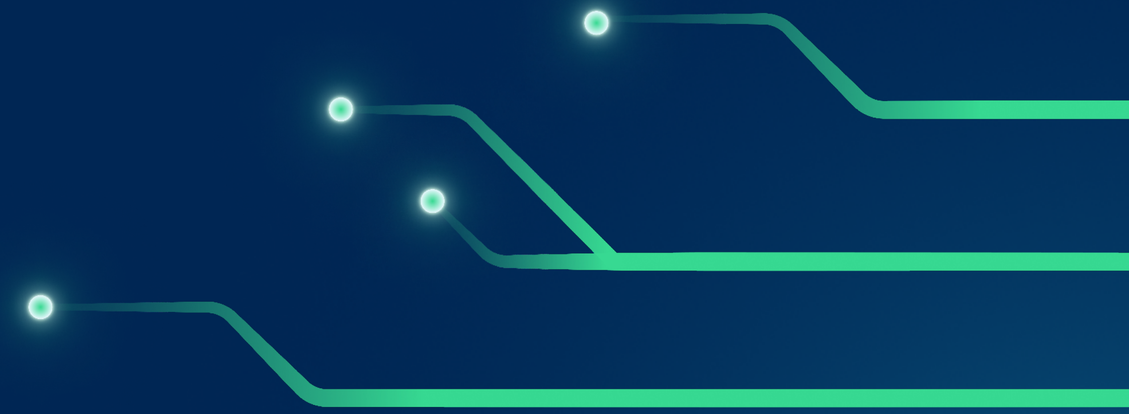
Knitting Flawless Pipelines with Kafka, Flink, and Pinot

Tim Berglund

VP DevRel, Confluent

Viktor Gamov

Principal Developer Advocate



RTA Summit

@gamussa | developer.confluent.io | @tlberglund

What is Apache Pinot™?

“Apache Pinot is a real-time distributed OLAP database, designed to serve OLAP workloads on streaming data with extreme low latency and high concurrency.”

The essence of real-time analytics

LATENCY

The amount of time it takes to execute a query

CONCURRENCY

The ability of a system to handle multiple queries simultaneously

FRESHNESS

The up-to-date nature of data in the system

The essence of real-time analytics

LATENCY

As low as 10ms

CONCURRENCY

As many as 100,000 queries per second

FRESHNESS

Seconds from event time till queryable in Pinot

OLTP

OLTP

- Transaction focused
- Write-heavy workloads
- Often involves a single record per operation

OLAP

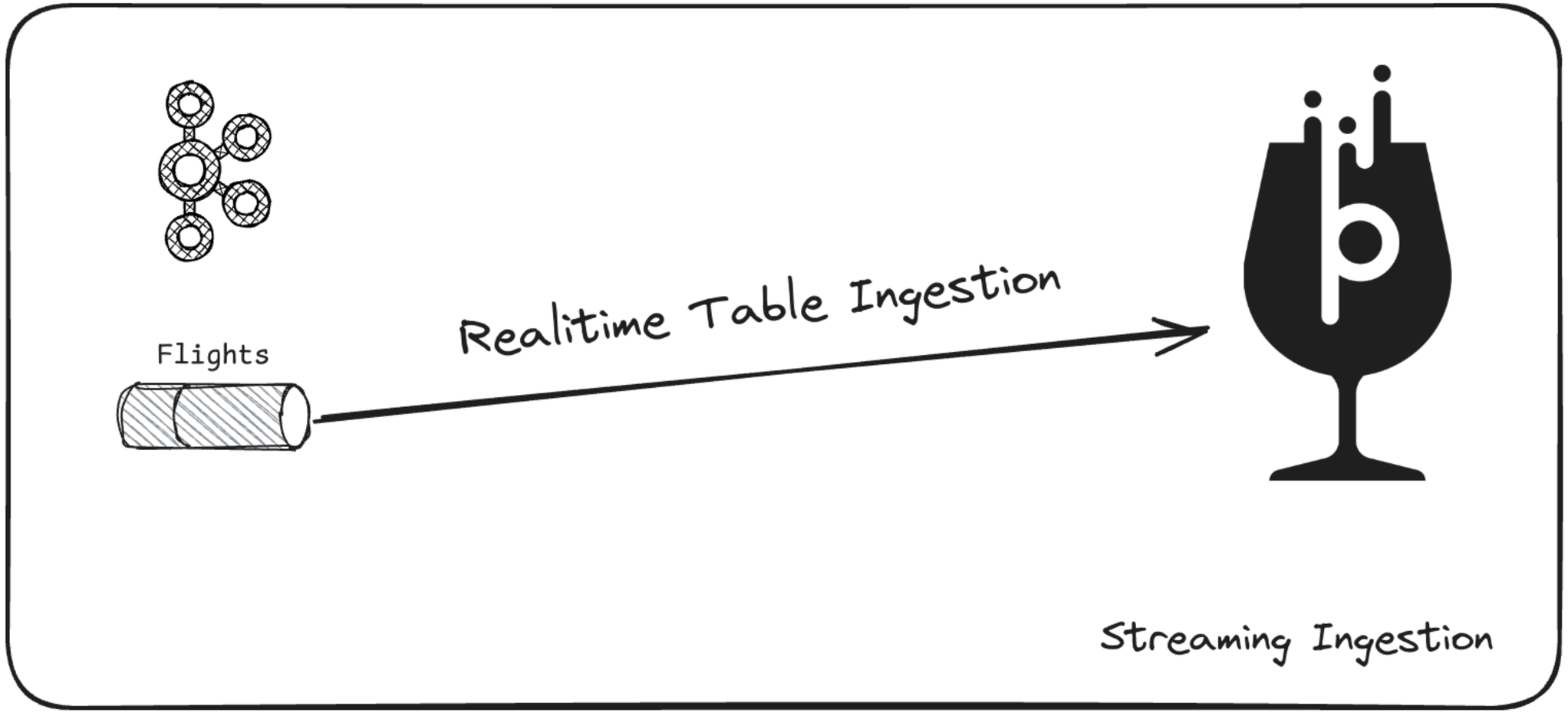
- Aggregation-focused
- Read-heavy workloads
- Often involves many records in one operation

Data Model

- Pinot uses the completely familiar **tabular data model**
- **Table** creation and **schema** definition **expressed in JSON**
- **Queries** expressed in **SQL**

Kafka + Pinot

Streaming Ingestion


Flights
Realitime Table Ingestion
Streaming Ingestion

Kafka + Flink + Pinot

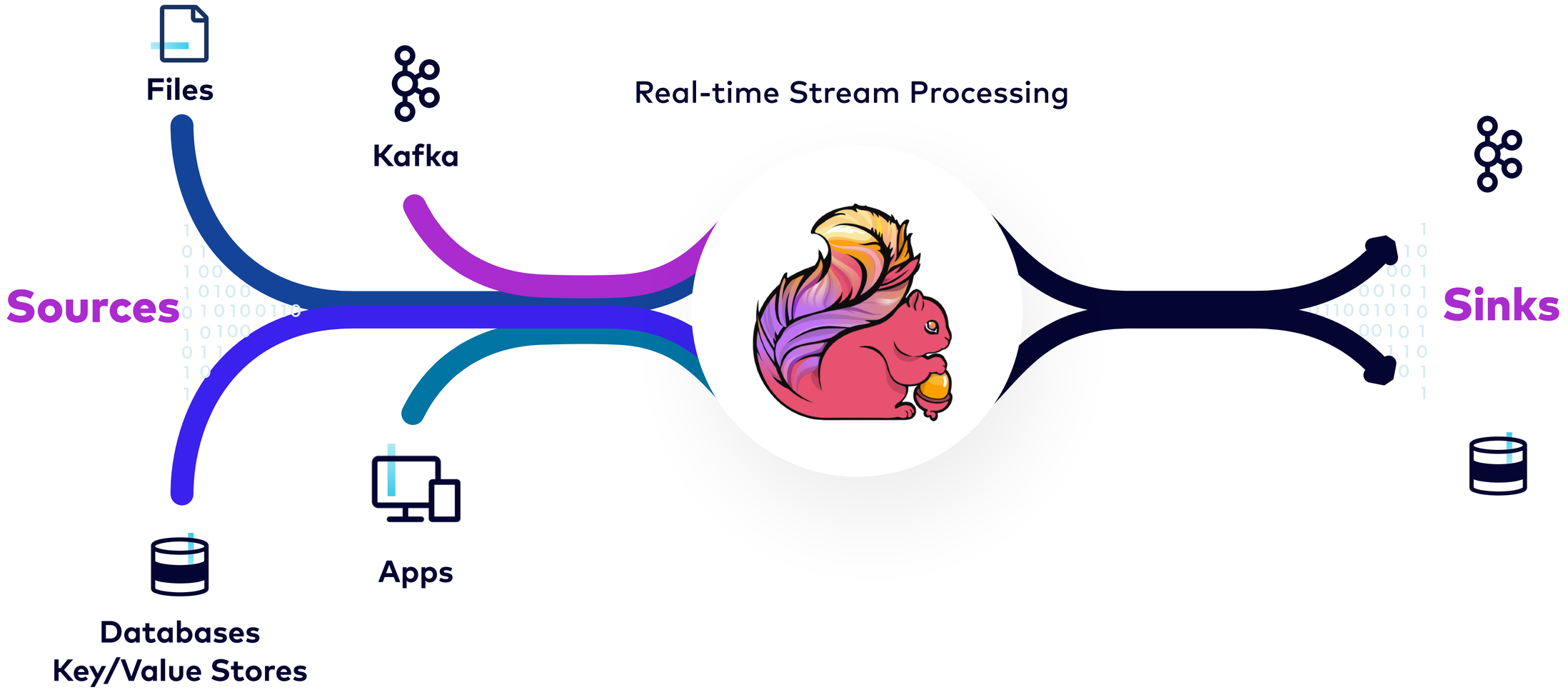
Knitting Flawless Pipelines

@gamussa | developer.confluent.io | @tlberglund

Flink 101

«Apache Flink is a *framework* and *distributed processing engine* for *stateful* computations over *unbounded* and *bounded* data streams.»

Real-time services rely on stream processing



What is Flink SQL

A standards-compliant SQL engine for processing both batch and streaming data with the scalability, performance, and consistency of Apache Flink

How does Flink work with Kafka?

Compute

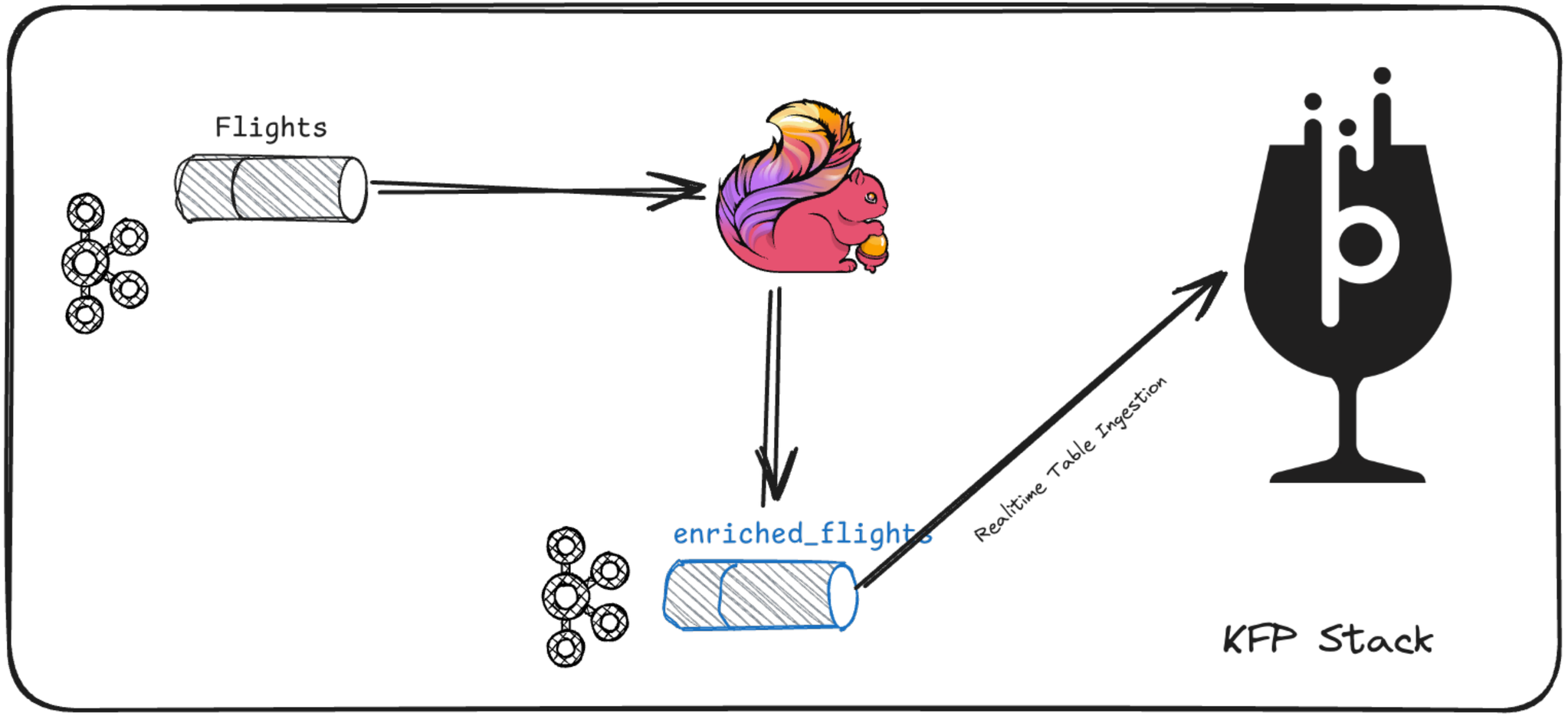
Apache Flink

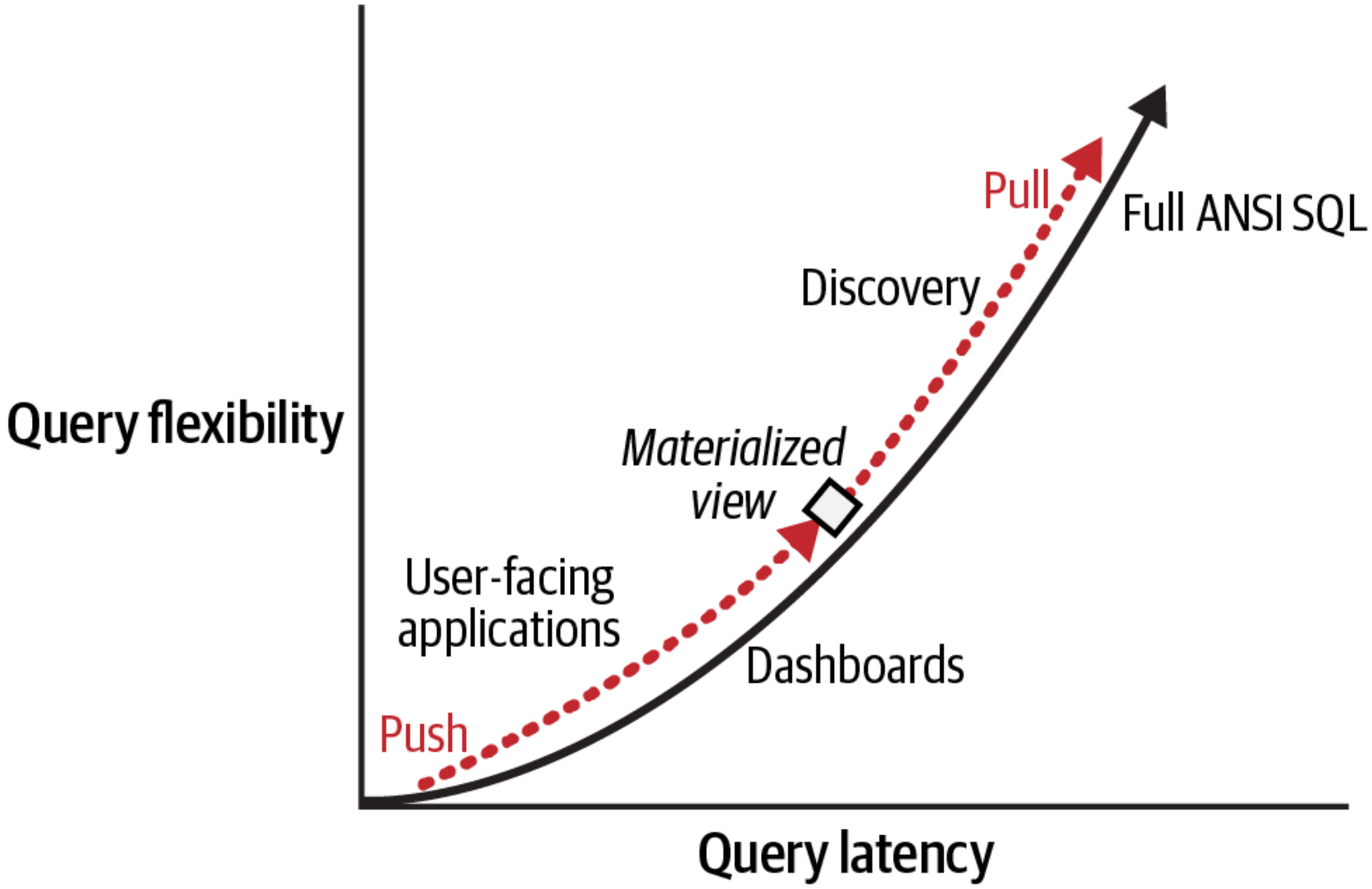
Explore | Enrich | Transform | Aggregate

Storage

Apache Kafka

Produce | Store | Consume


Flights
enriched_flights
Realtime Table Ingestion
KFP Stack



Source: Streaming Databases, Hubert Dulay, Ralph Matthias Debusmann

Check out developer.confluent.io

@tlberglund | @gamussa