Hybrid cloud data streaming architecture

Kateryna Khotkevych
Software Engineer

Delivery Hero SE This document is strictly confidential and may not be copied, used, made available or be disclosed to third parties without prior written permission.

Prerequisites

Key Facts about Delivery Hero

- Founded in May 2011
- Operating in 40+ markets with c 17.000 employees
- The largest food network in the world with app. 200.000+ restaurant partners
- Delivered a great market debut with our IPO on 30th June 2017
- Global leader in the space with 1 million orders received per day (Q1 2018)
- **2 million orders delivered in one day (July 2019)**

Data Volume in Numbers

10M
Order-Related
Events Per Day

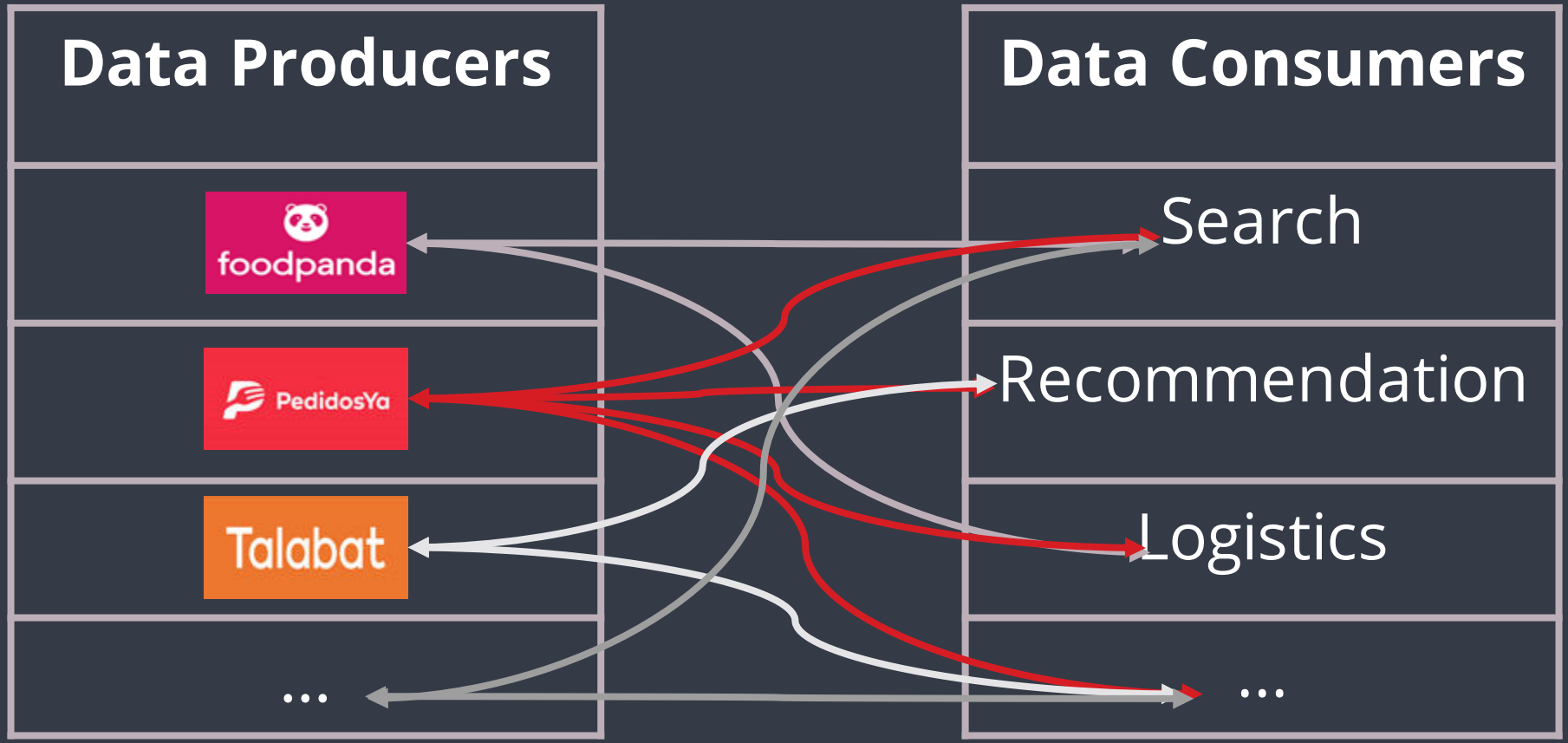
+ User clicks
+ Logistics data
+ Restaurant availability
+ Menu items
+ Customer data

Delivery Hero

House of Brands Concept

Challenges

Delivery Hero



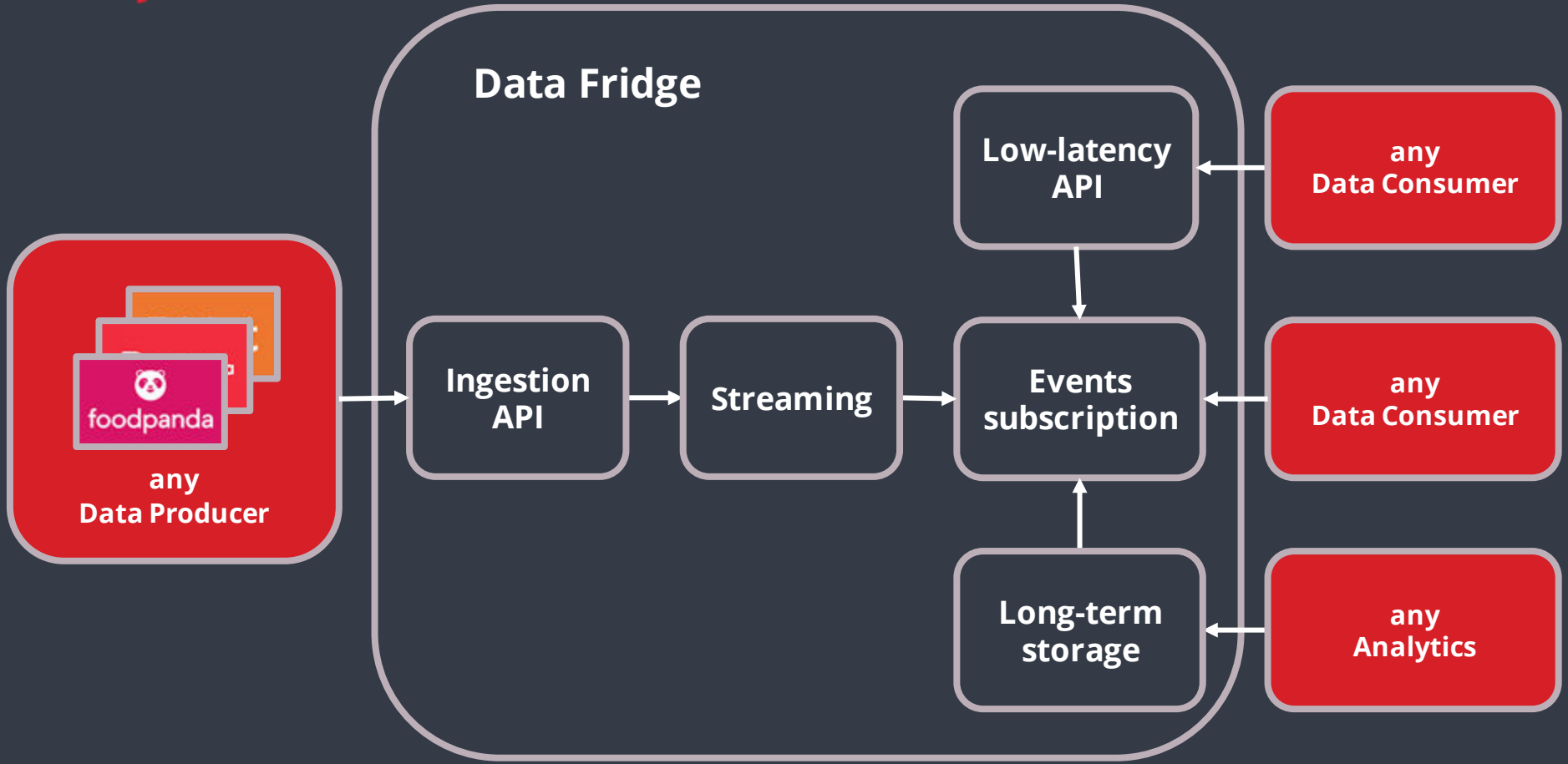
Delivery Hero

Mission

- Unify the data structure across all entities
- Provide different data consumptions types:
  - Near real-time
  - Low-latency
  - High-latency (analytics)
- Become a data producer for ML applications

Architecture

Architecture



Ingestion API

How to validate input from various data producers?

Delivery Hero

Ingestion API

- Single entrypoint
- Batch import
- AWS Lambda for event processing
- IP Whitelisting
- JWT authentication
- Data quality dashboards

AWS API Gateway

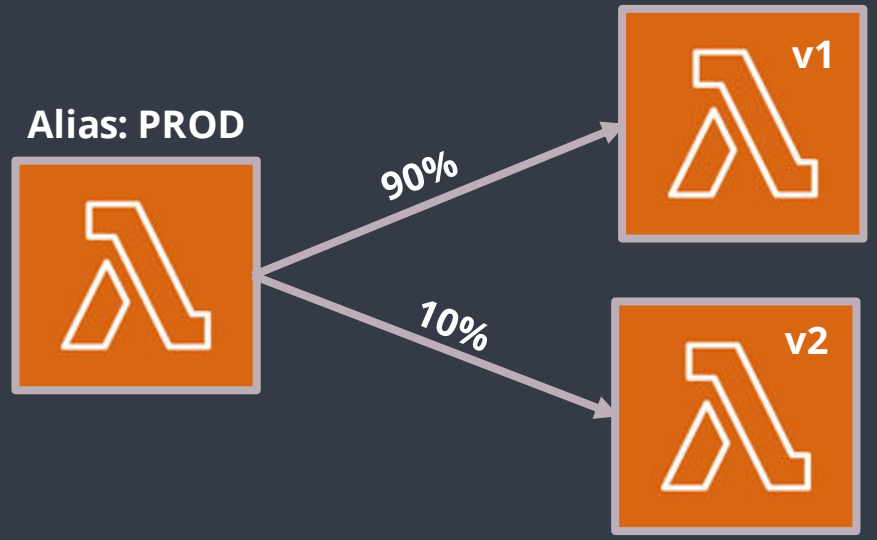
AWS Lambda

Canary Deployment

Delivery Hero

Canary Deployment

- Custom solution
- 2 versions under the same alias
- Metrics monitoring

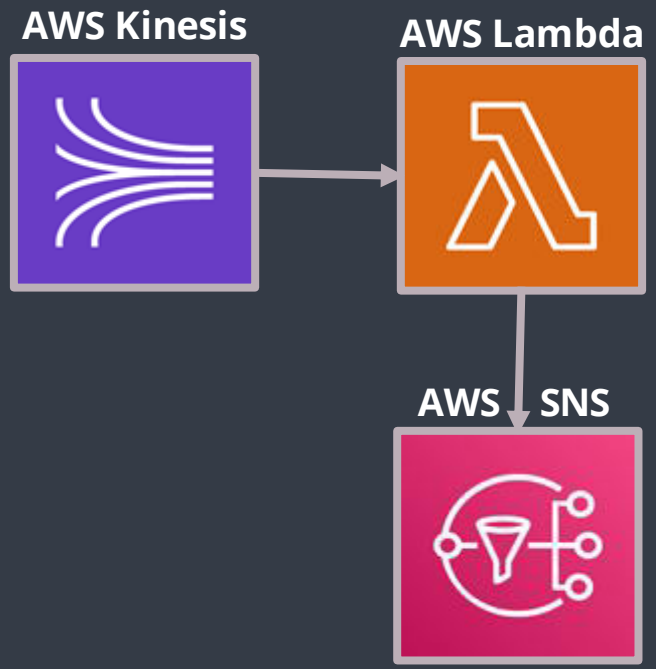


Streaming

Delivery Hero

Streaming

- Kinesis stream preserves messages up to 7 days
- Ability to replay data

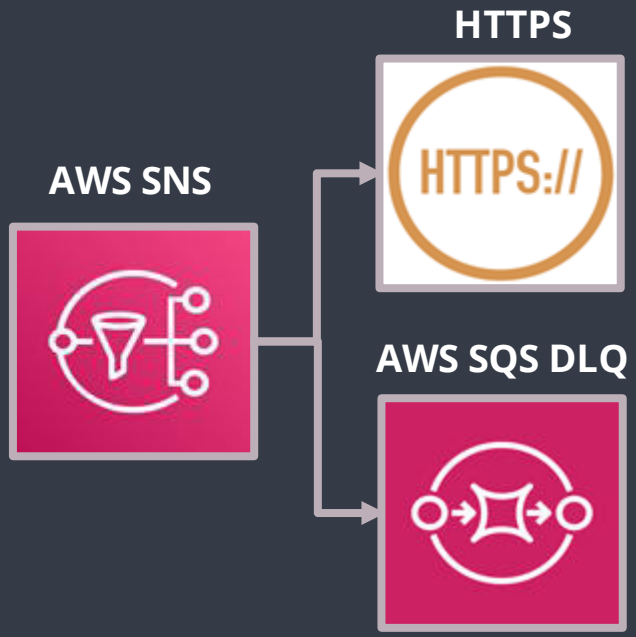


Near real-time streaming (1,2 ms)

Delivery Hero

Near real-time streaming

- 2 types for subscriptions
- Aggregations
- Filtering based on event attributes

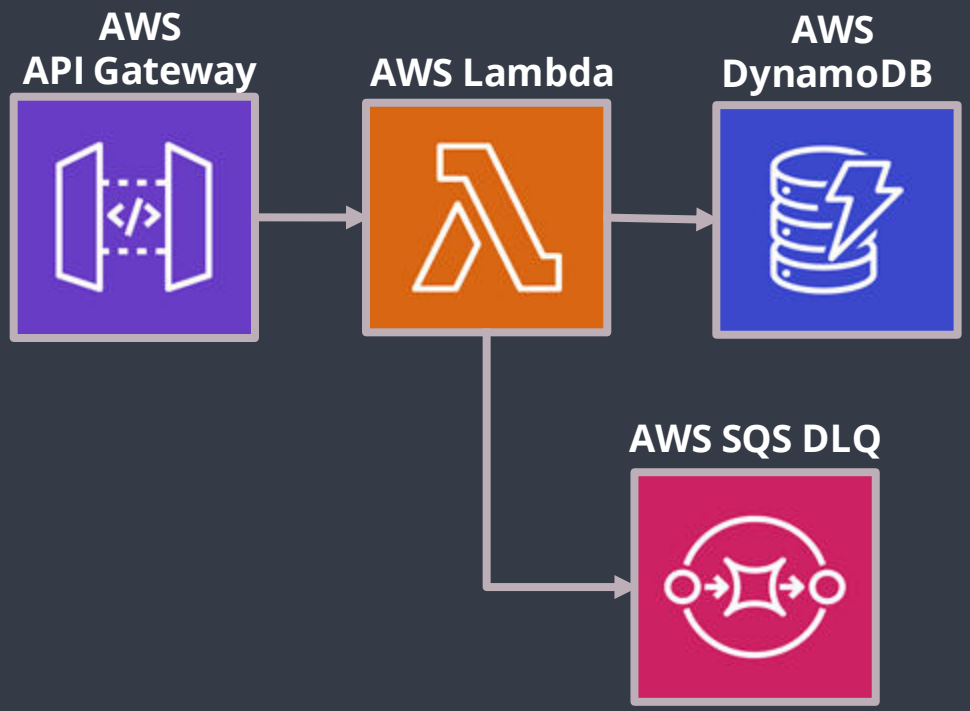


Low-latency API

Delivery Hero

Low-latency API

- Dead Letter Queue
- On-demand scaling
- Retains PII data for 30 days



Analytics

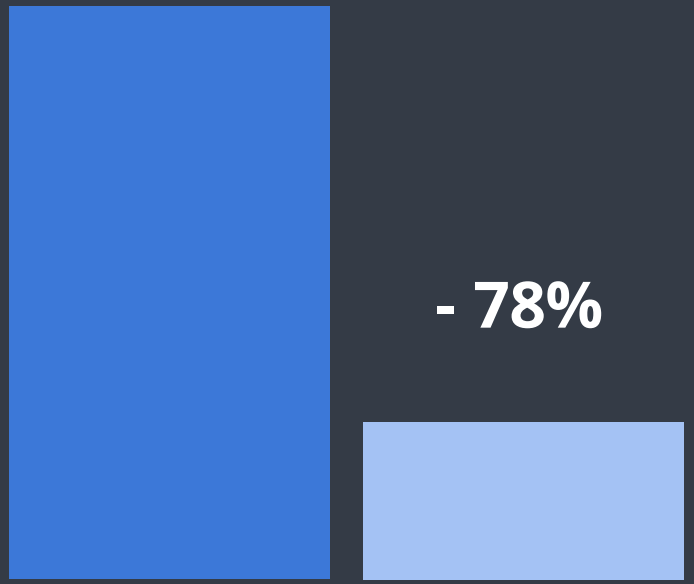
AWS S3?
AWS Redshift?
GCP BigQuery?

Delivery Hero

AWS to GCP transfer

Google Pub/Sub

Google Dataflow

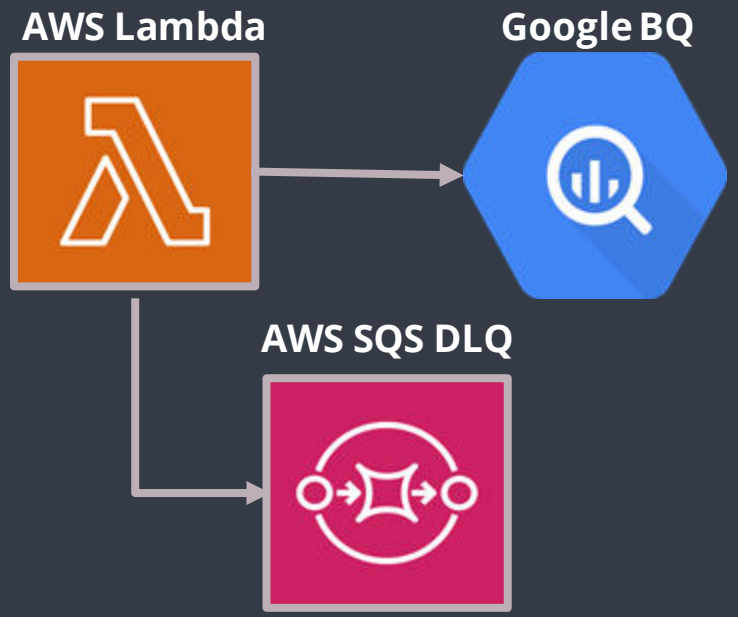


Monthly Expenses

Delivery Hero

Analytics

- GDPR compliance
- Data quality visualization
- Restricted data access



Delivery Hero

Any Producer

Ingestion API		
Streaming		
First use case: Near real-time	Second use case: Low-latency API	Third use case: Analytics

Any Consumer

Tech Stack

Delivery Hero

		Collaboration	Monitoring & Logging
Backend	Java, TypeScript	GitHub, Lucidchart, JIRA	CloudWatch, AWS X-Ray, looker
Data Management	Amazon DynamoDB, Amazon S3, amazon SQS, Google Cloud Pub/Sub, Google BigQuery, amazon KINESIS, amazon SNS, Amazon Lambda, Apache Atlas		
Infra	Amazon API Gateway, aws, Google Cloud Platform, HashiCorp Terraform		

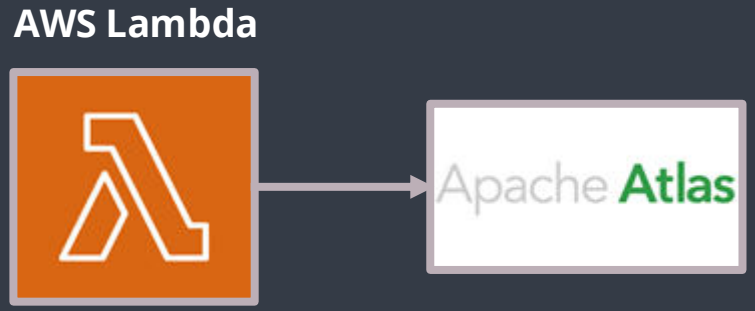
Data Catalog

How to pick Data Catalog?

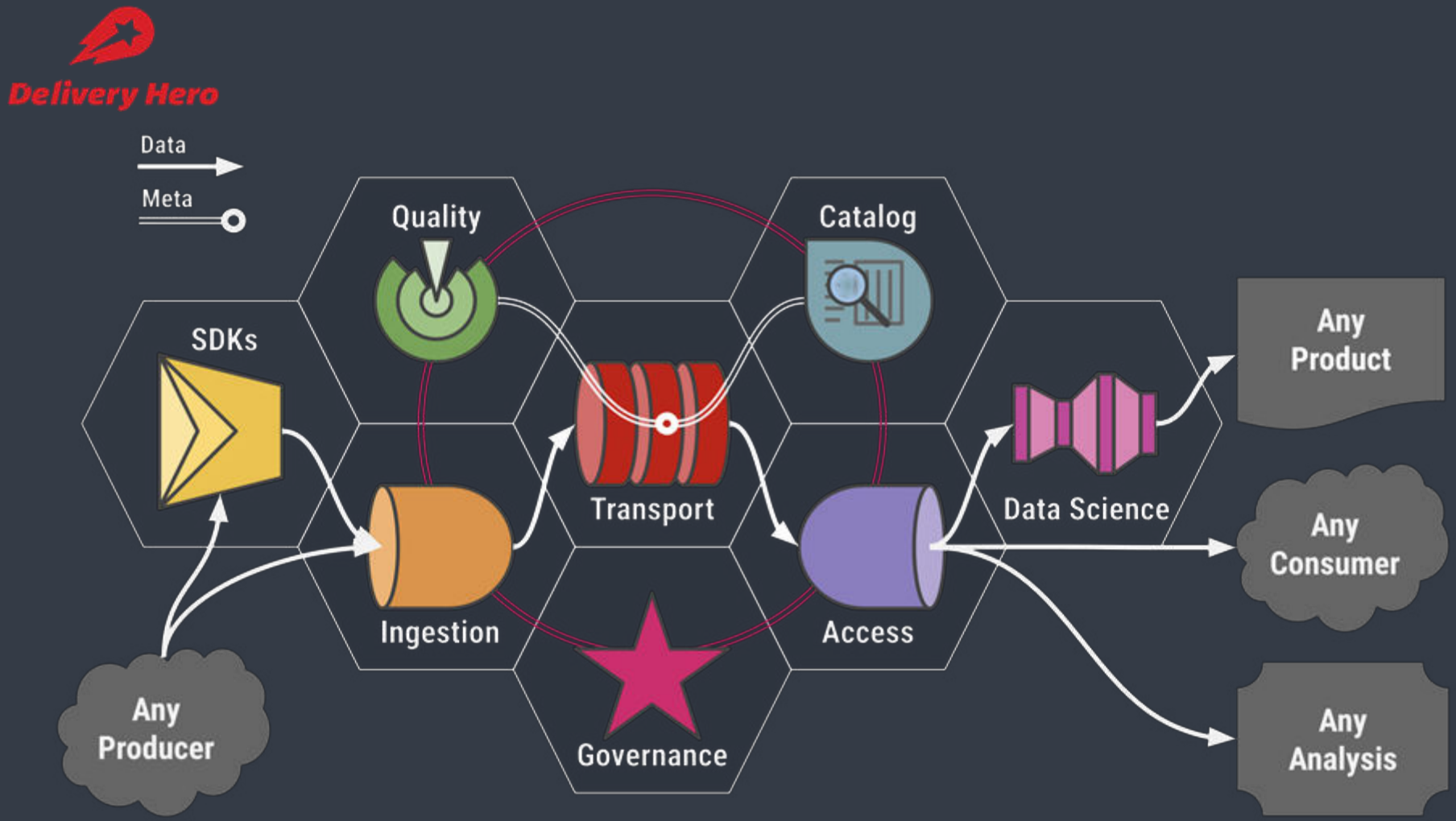
Delivery Hero

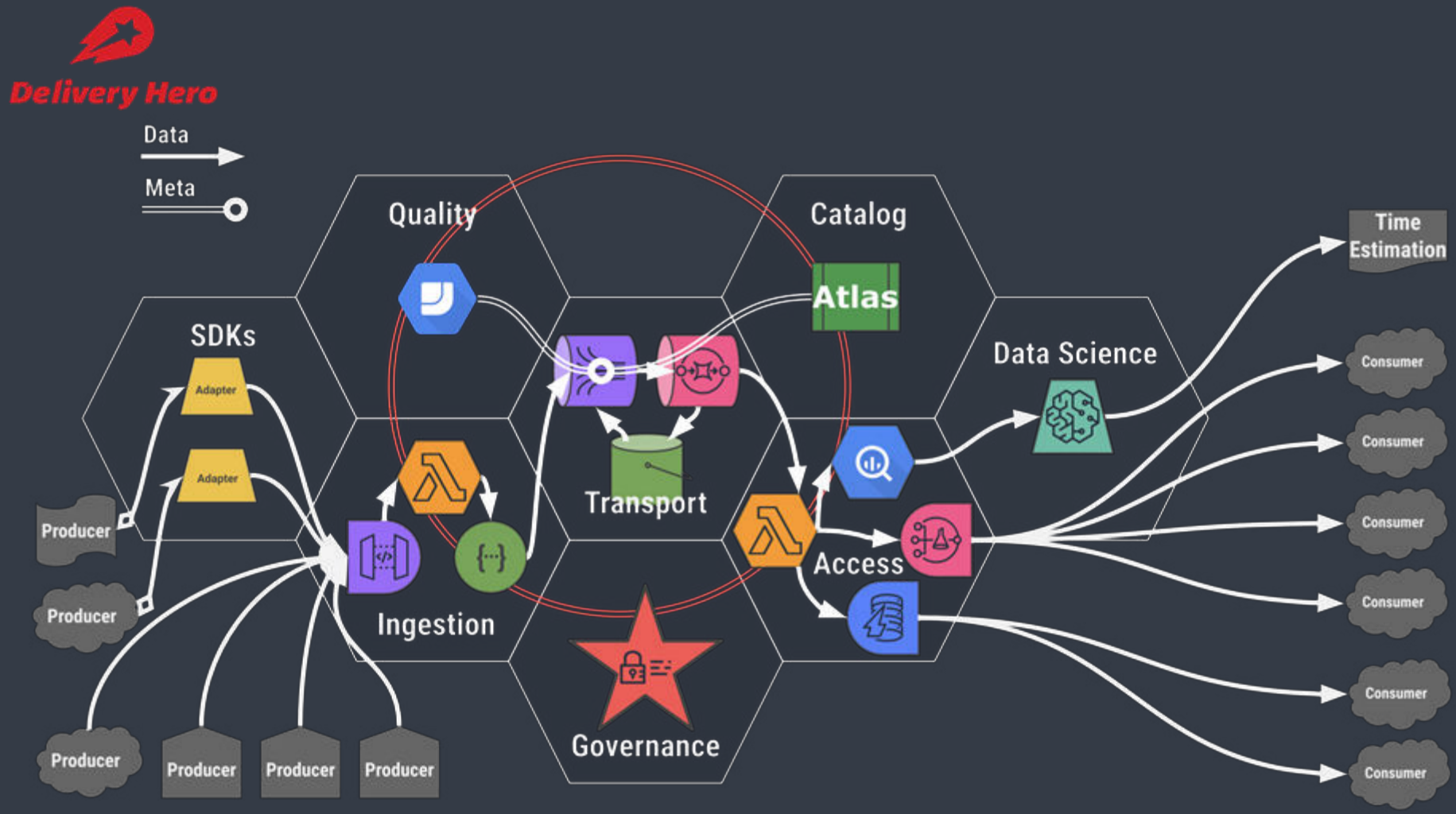
Data Catalog

- Lineage
- Relations between data models
- Can work with various data sources



Further development


Delivery Hero
Data
Meta
Quality
Catalog
SDKs
Transport
Data Science
Any Product
Any Consumer
Any Analysis
Ingestion
Access
Governance
Any Producer


Delivery Hero
Data
Meta
Quality
Catalog
Atlas
Time Estimation
SDKs
Adapter
Adapter
Data Science
Consumer
Consumer
Consumer
Consumer
Consumer
Consumer
Transport
Producer
Producer
Ingestion
Access
Governance
Producer
Producer
Producer
Producer

We Are Hiring!

www.deliveryhero.com