



Delivery Hero

Building Scalable Real-Time Data Pipeline “Data Fridge”

Vicente Valls Rios

Software Engineering Manager

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)



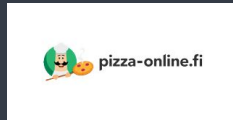
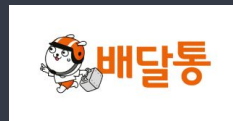
Delivery Hero

Data Volume in Numbers



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

House of Brands & Global Services



Logistics

Search

Recommendation

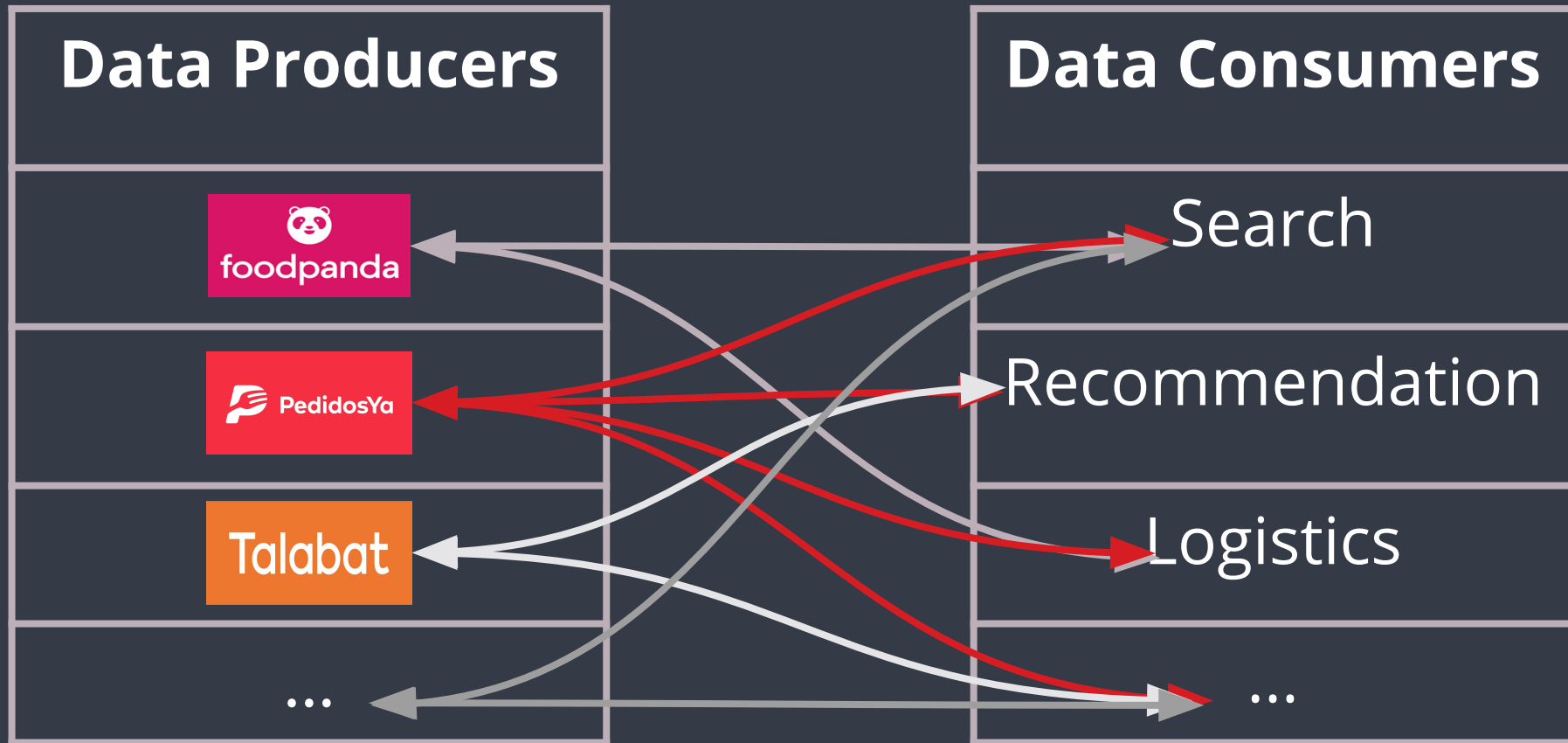
...

Challenges



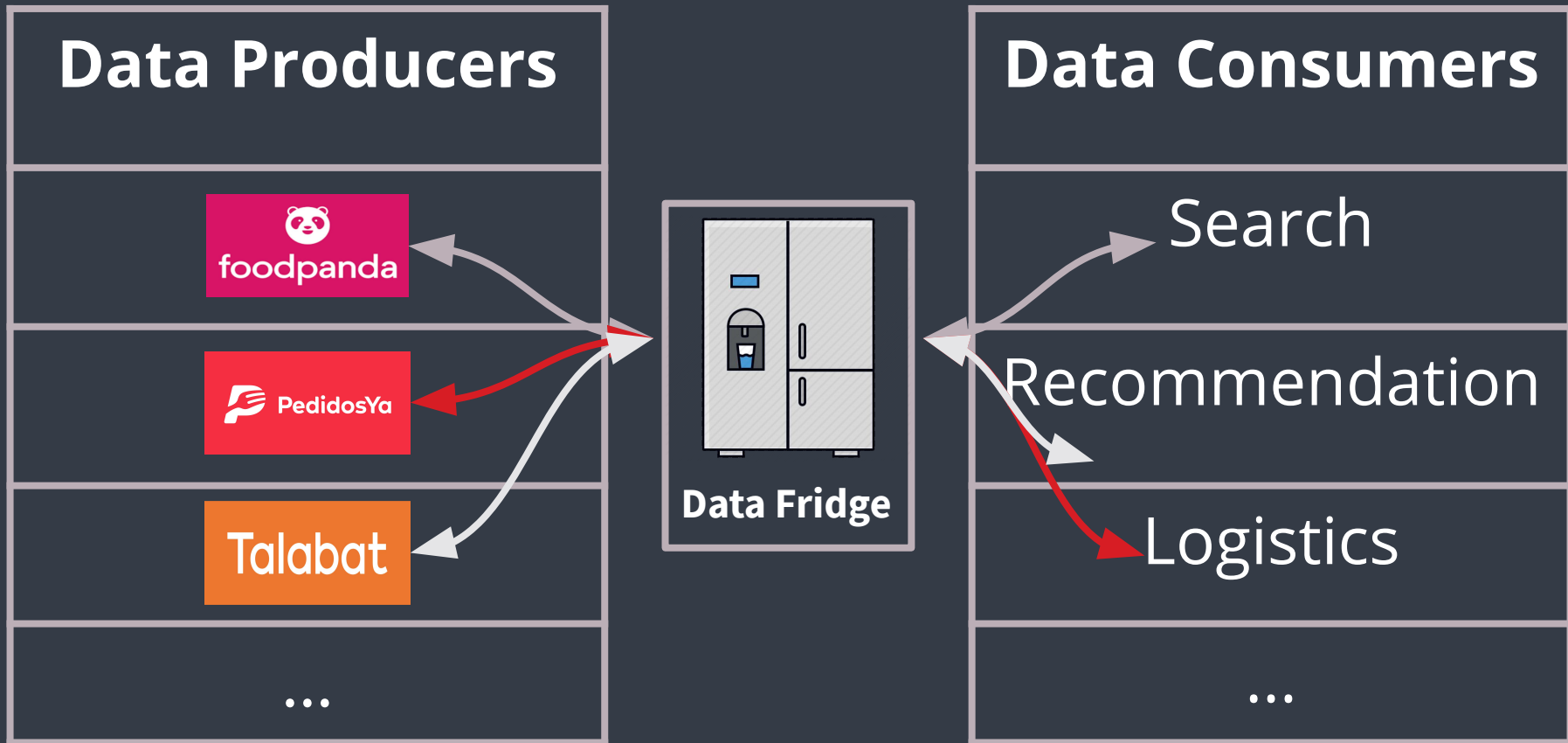


Delivery Hero





Delivery Hero





Delivery Hero

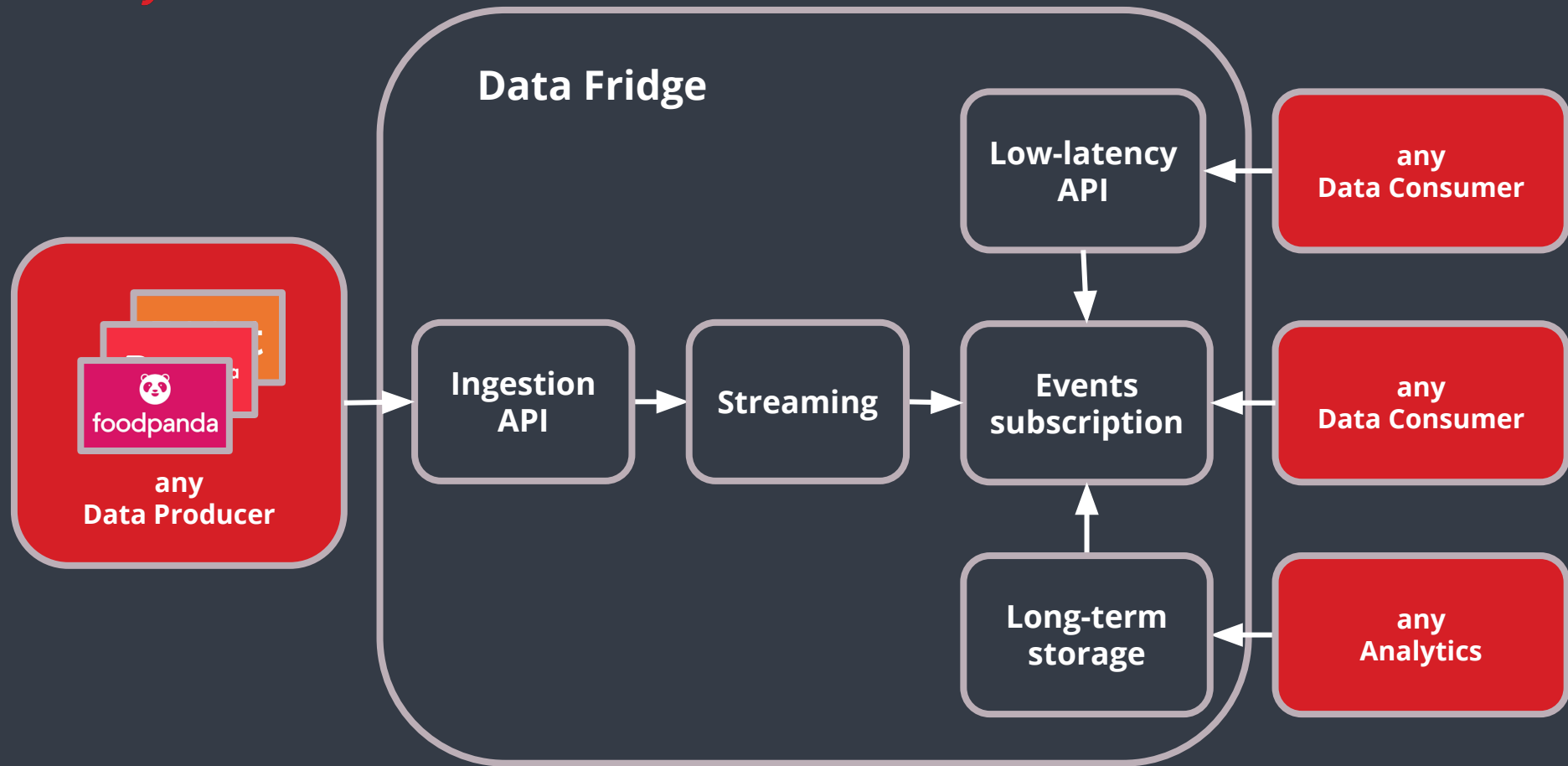
Mission

- **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

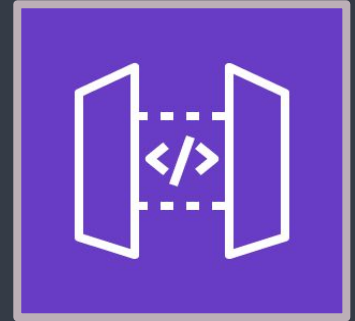




Ingestion API

Ingestion API

- Verifies quality of data using complex validations
- Single endpoint
- Batch import
- AWS Lambda for event processing
- IP Whitelisting
- JWT authentication



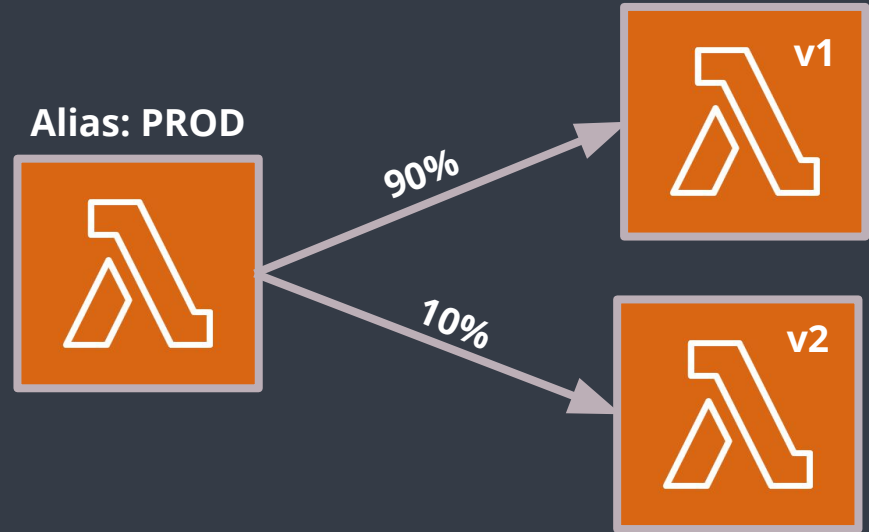
AWS API Gateway



AWS Lambda

Canary Deployment

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



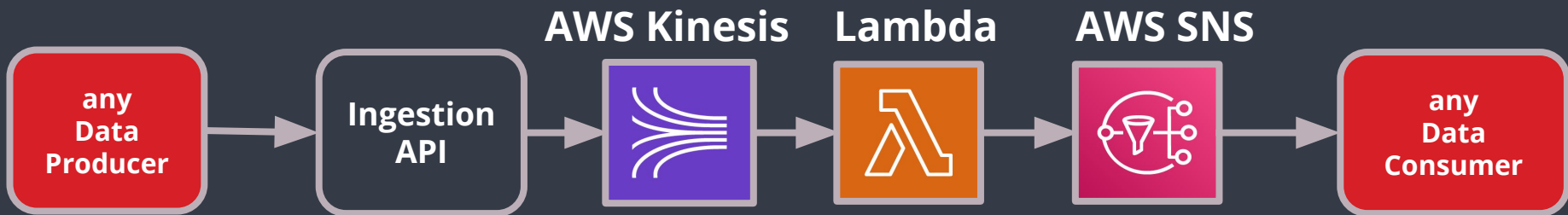
Streaming





Delivery Hero

Streaming



- **Kinesis stream preserves messages up to 7 days**
- **Ability to replay data**
- **Ability to scale up/down**

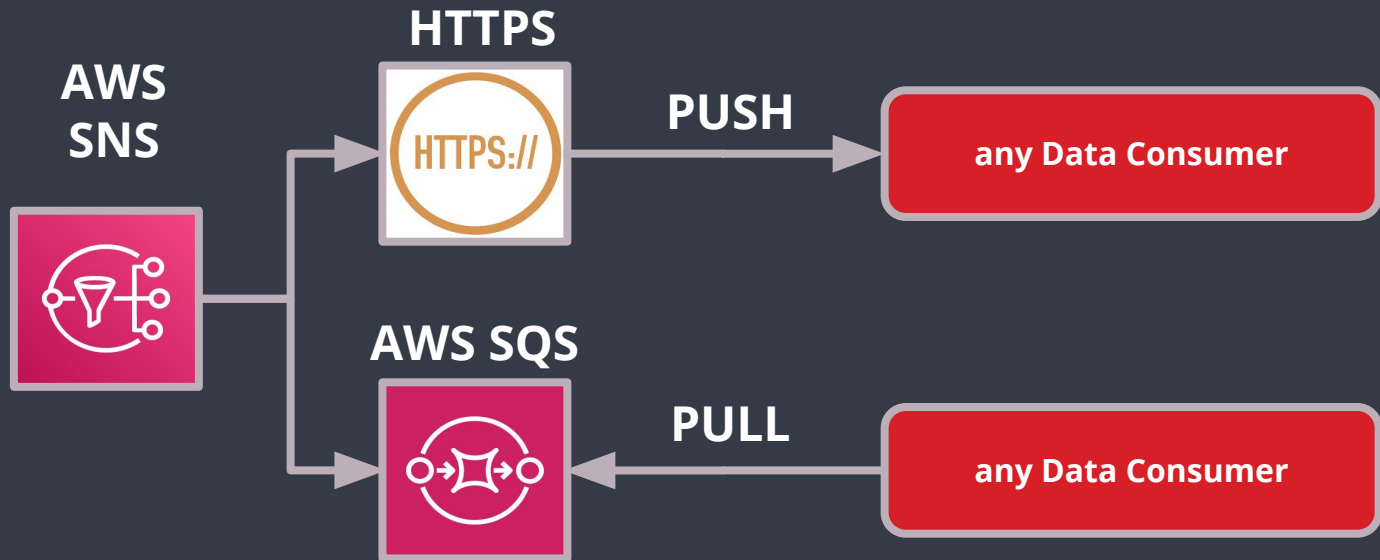


SNS instead of Kinesis ?

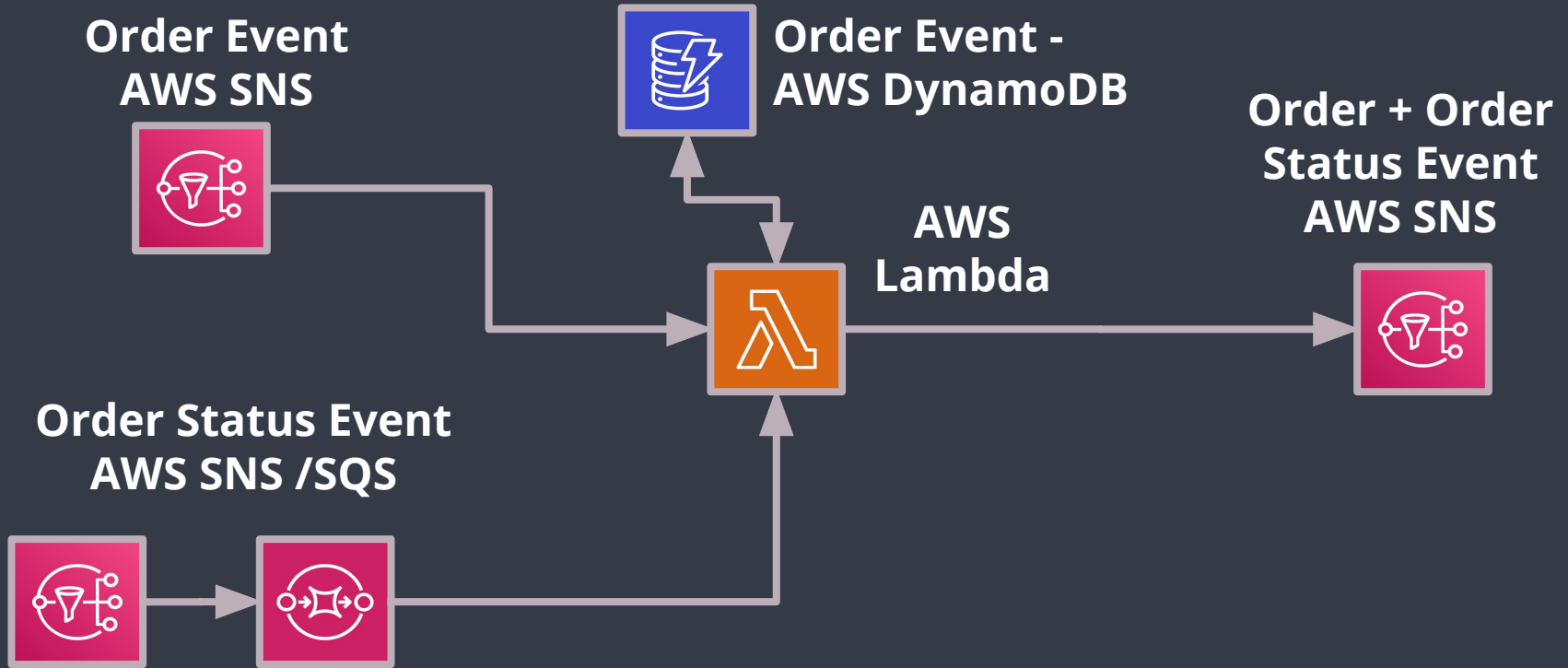
- We do not care about order of events.
- Having Message Filtering.
- Kinesis requires to scale up shards as our Consumers grow.
- Scale up Kinesis is harder and more expensive than SNS.
- SNS->HTTP/SQS service provides PUSH-PULL data consumption.

Streaming

- 2 types for subscriptions
- Filtering based on event attributes



Stream Aggregation

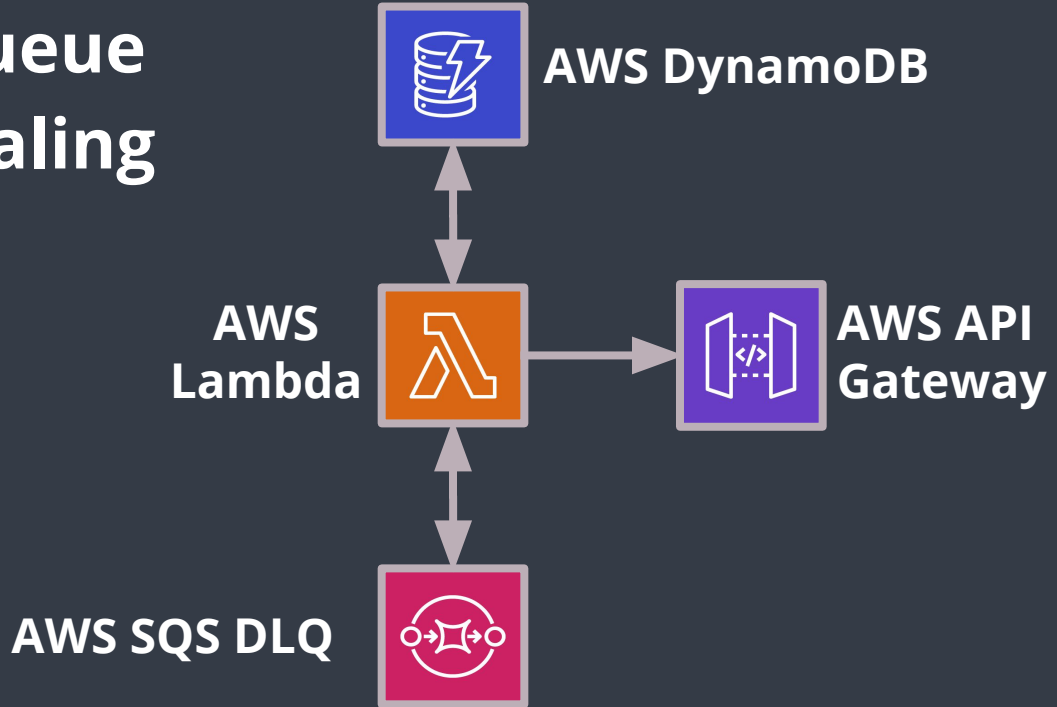


Low-latency API



Low-latency API

- Dead Letter Queue
- On-demand scaling

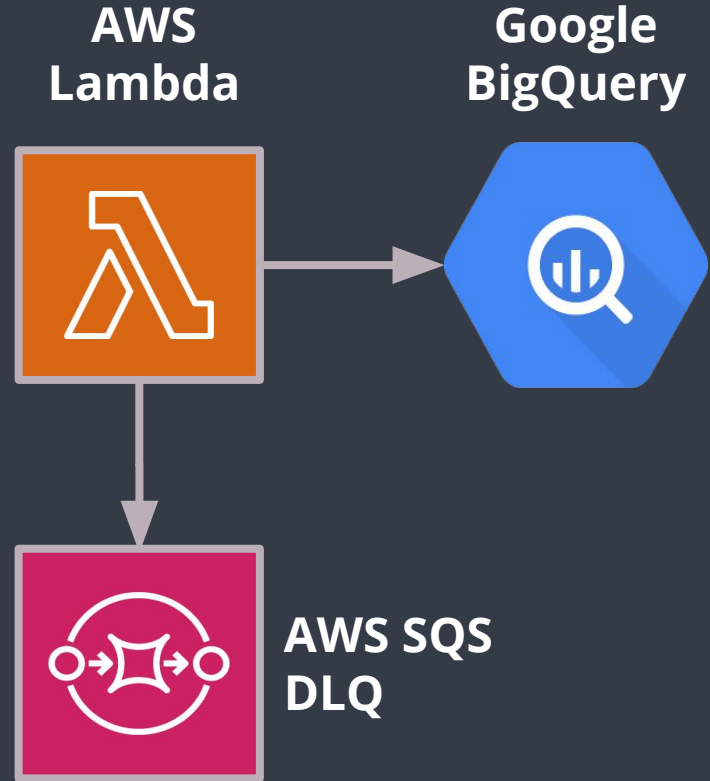








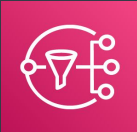
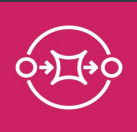

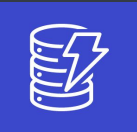
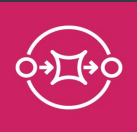





Analytics

Analytics

- **Bigquery as OLAP db**
- **Data quality visualization**
- **Bigquery is scalable DWH solution**



Any Producer

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

Any Consumer









Delivery Hero

Tech Stack

Collaboration

Monitoring & Logging

Backend	 Java	 TypeScript	 GitHub		 Lucidchart	 JIRA	 CloudWatch	 AWS X-Ray	 Stackdriver
Data Management	 Amazon DynamoDB	 amazon SQS		 Amazon API Gateway	 Google BigQuery	 Google Data Studio			
Infra		 HashiCorp Terraform	 aws	 Google Cloud Platform					



Challenges

Challenges

- **SLAs (latency, durability, etc.) for some cloud services.**
- **Data Quality & Data documentation**
- **GDPR**
- **Automating new pipeline creation**
- **Automating SNS subscriptions / BigQuery Access**



We Are Hiring!



www.deliveryhero.com