aws SUMMIT

07 June 2018

Data Challenges in an acquisition based world

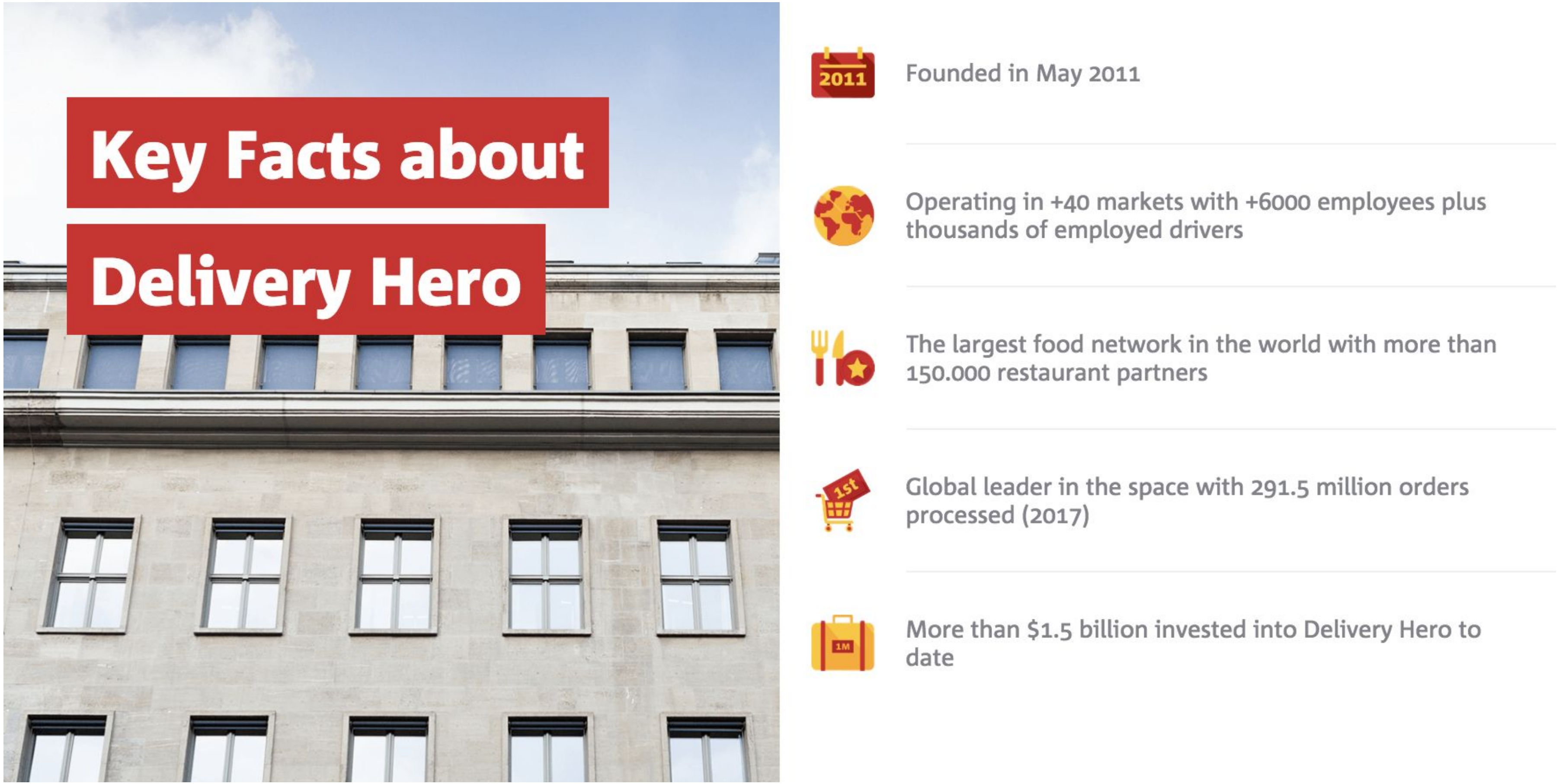
(or how to click a Data Lake together with AWS)

Daniel Manzke

CTO Restaurant Vertical @ Delivery Hero

How do you tackle the challenges of reporting and data quality in a company with a model of local and centralized entities? We will talk about the evolution of Data in Delivery Hero from Data Warehouse to a Data Lake oriented Architecture and will show you how you can click your own Data Lake together with the help of AWS.

We are an Online Food Ordering and Delivery Marketplace



Key Facts about Delivery Hero

- Founded in May 2011
- Operating in +40 markets with +6000 employees plus thousands of employed drivers
- The largest food network in the world with more than 150.000 restaurant partners
- Global leader in the space with 291.5 million orders processed (2017)
- More than $1.5 billion invested into Delivery Hero to date

History

2008
Niklas started OnlinePizza in Sweden

2010
Lieferheld launched

2011
HungryHouse joins

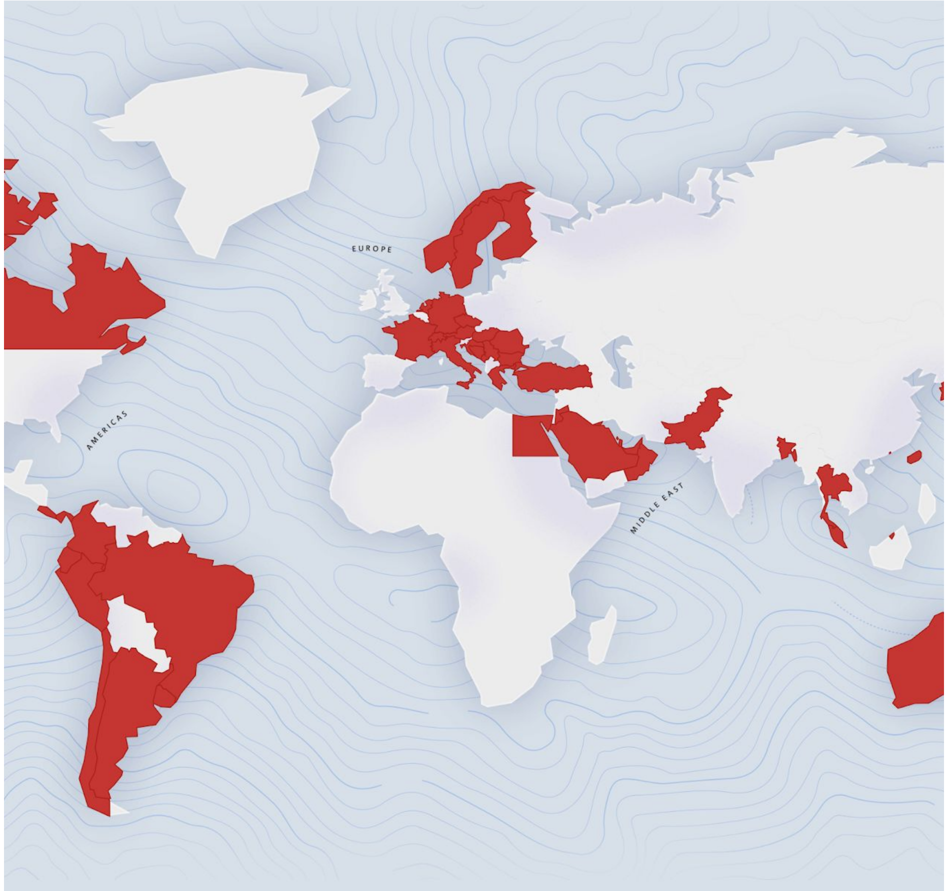
2012
OnlinePizza Norden joins

2014
PedidosYa, pizza.de and Baedaltong joins

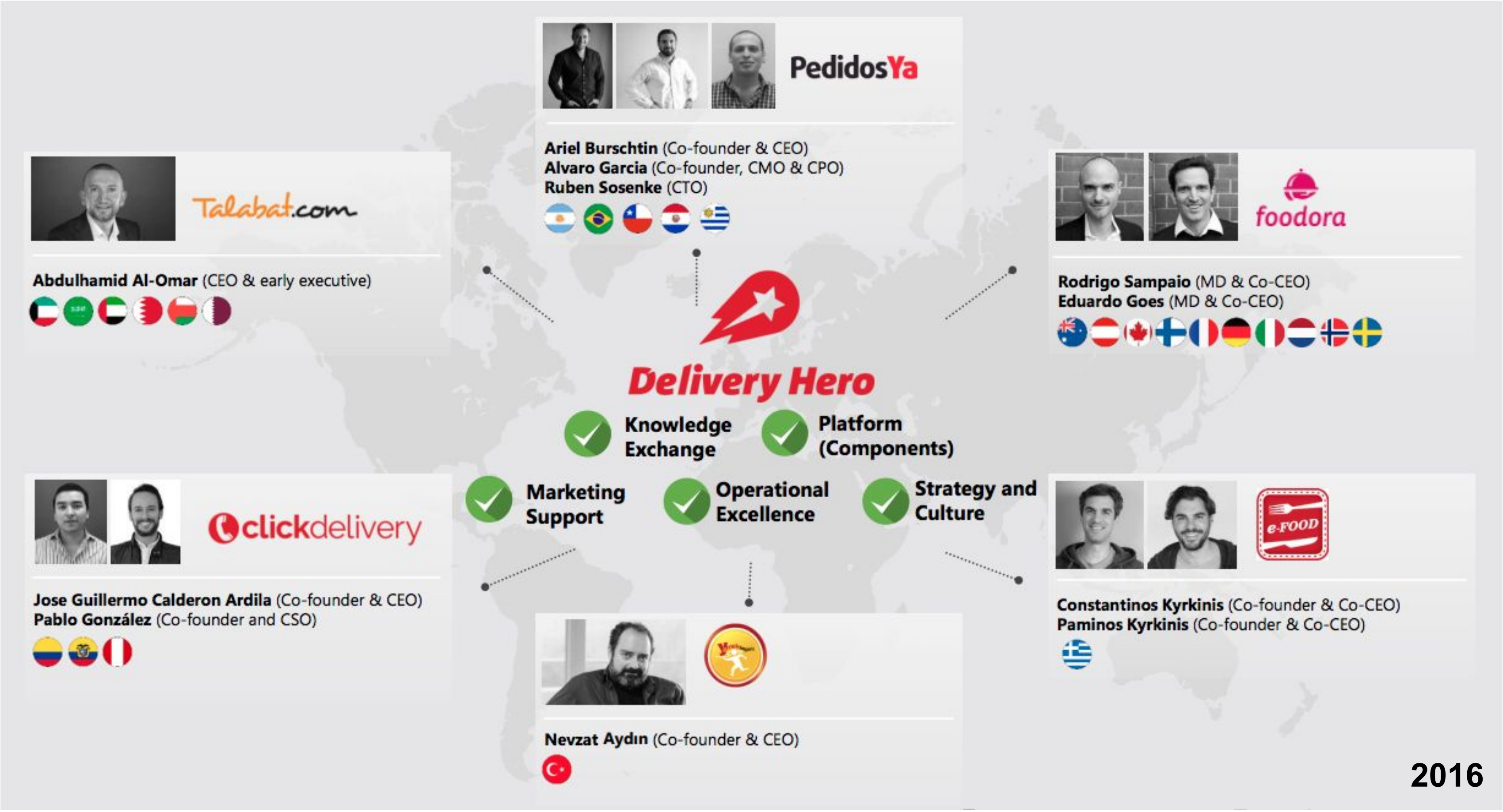
2015
Talabat, Yemeksepti and Foodora joins

2016
foodpanda joins

IPO, M&A, ...



Hyper-Local



PedidosYa

Ariel Burschtin (Co-founder & CEO)
Alvaro Garcia (Co-founder, CMO & CPO)
Ruben Sosenke (CTO)

Talabat.com

Abdulhamid Al-Omar (CEO & early executive)

foodora

Rodrigo Sampaio (MD & Co-CEO)
Eduardo Goes (MD & Co-CEO)

Delivery Hero

- ✓ Knowledge Exchange
- ✓ Platform (Components)
- ✓ Marketing Support
- ✓ Operational Excellence
- ✓ Strategy and Culture

clickdelivery

Jose Guillermo Calderon Ardila (Co-founder & CEO)
Pablo González (Co-founder and CSO)

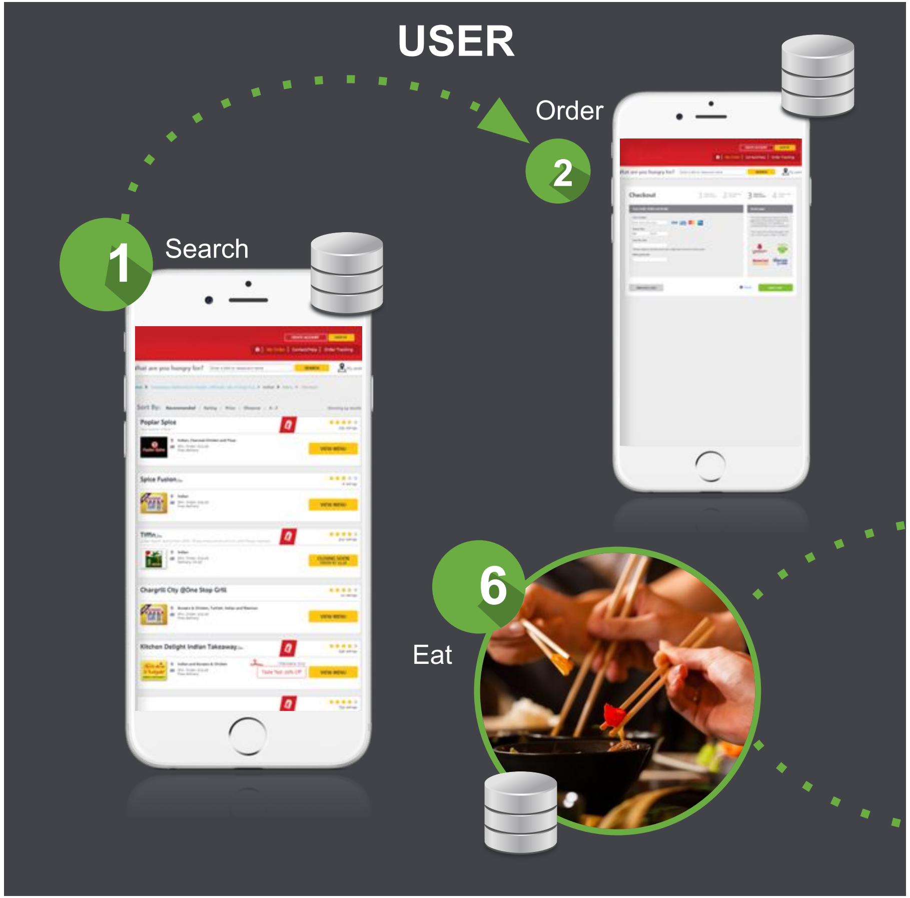
e-FOOD

Constantinos Kyrkinis (Co-founder & Co-CEO)
Paminos Kyrkinis (Co-founder & Co-CEO)

Nevzat Aydın (Co-founder & CEO)

2016

Customer Experience - Data



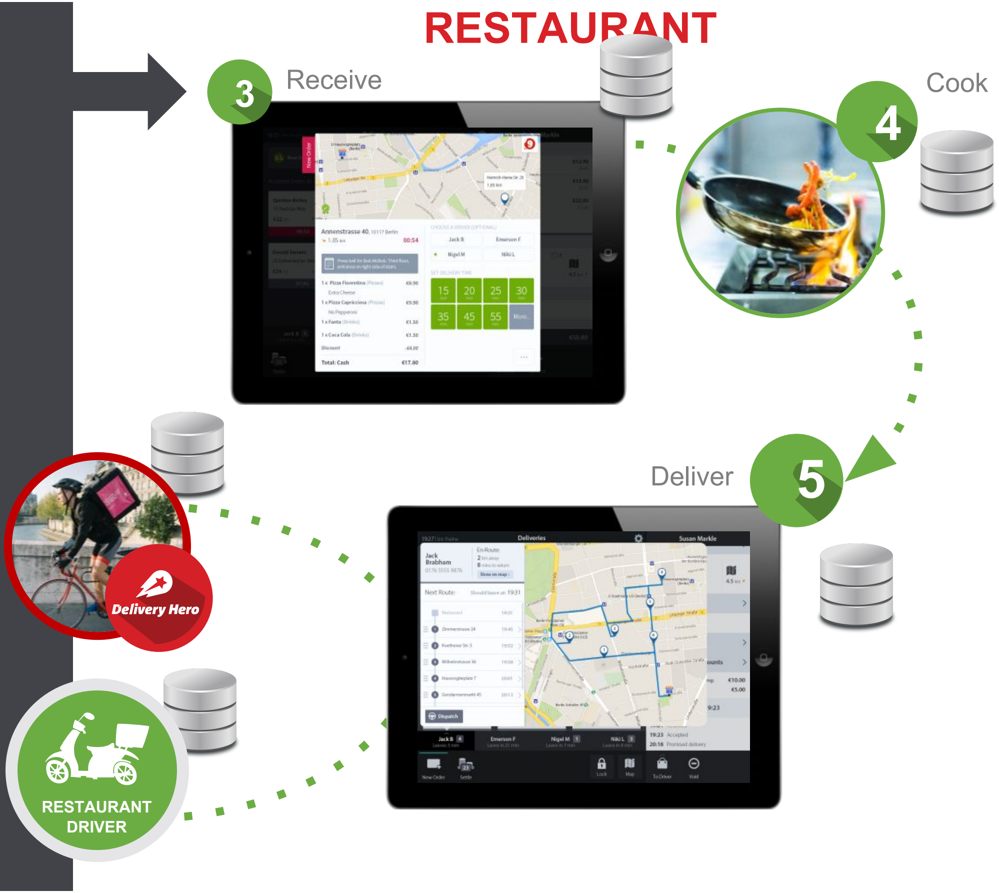
- **The Right Food**
  - Search, Recommendations, Vouchers, Premium Placements
- **Order Placement**
  - Time to Order
  - Estimated Delivery Time
  - Payment Method
- **After Order**
  - Reviews, Ratings, Complaints
  - Where is my food?
  - Reorder Rate

“Understand the whole cycle”

Order delivery - Data

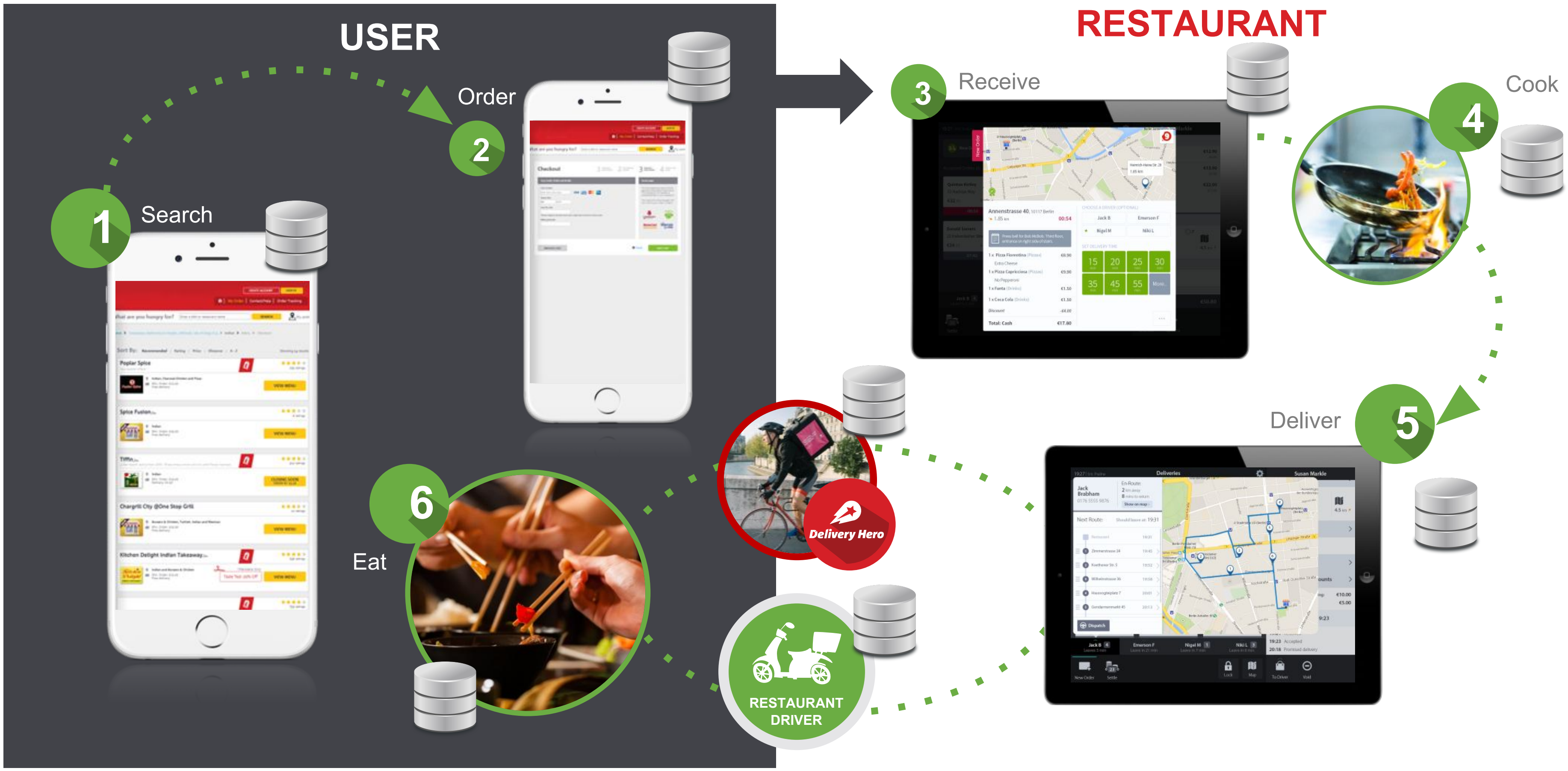
- **Availability**
  - Online, Open, Busy, ...
- **Cooking**
  - Preparation Time
  - Items Unavailable
- **Delivery**
  - Driving Time
  - Driver Shifts
  - Delivered?!

“Tons of Data to collect”



We are an Online Food Ordering and Delivery Marketplace

*Delivery*Tech



USER

1 Search

2 Order

6 Eat

RESTAURANT

3 Receive

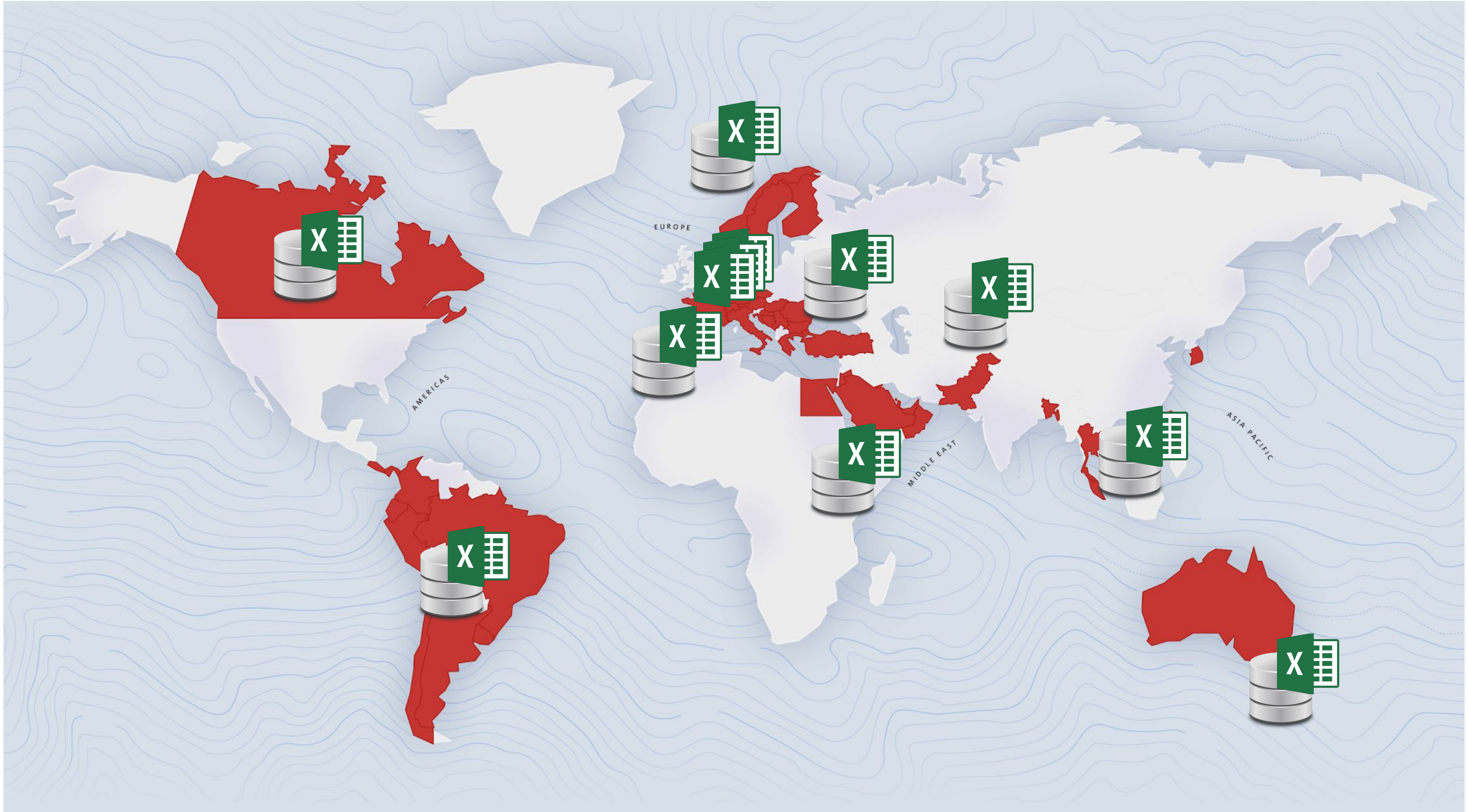
4 Cook

5 Deliver

Delivery Hero

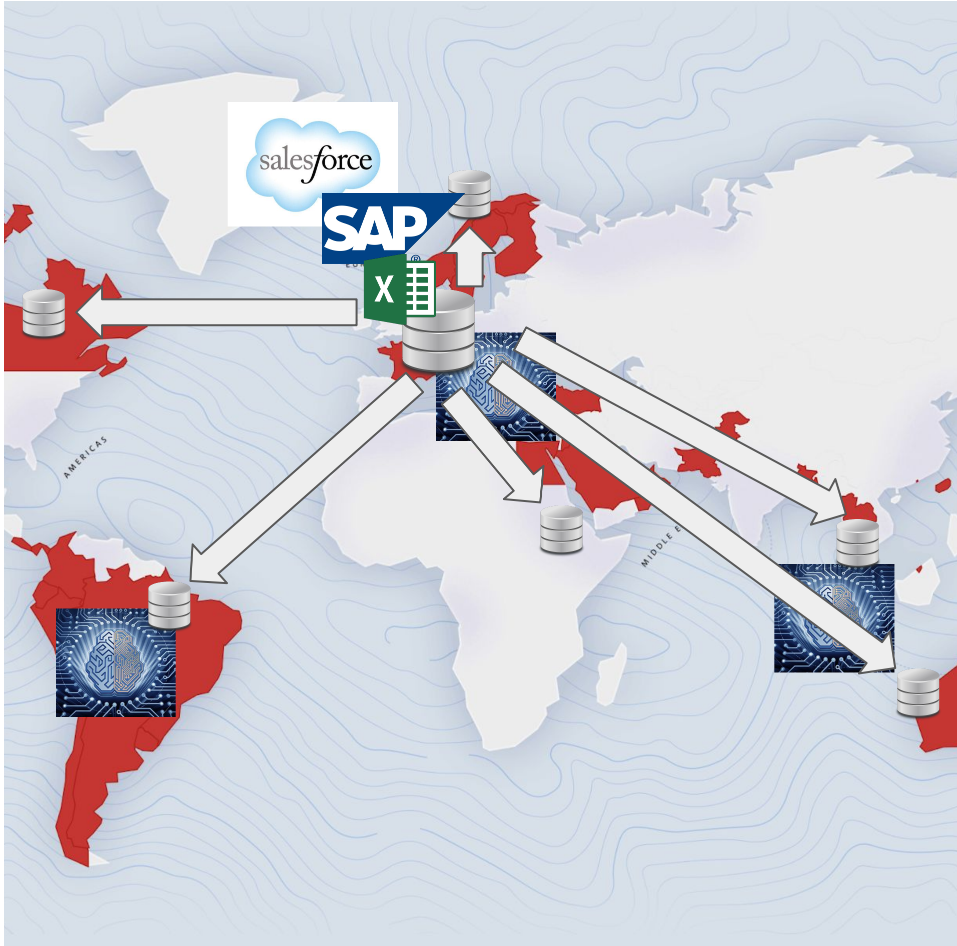
RESTAURANT DRIVER

KPIs, Reporting, ...



SAP, Salesforce, DH Services, ...

- **Global Data Warehouse**
  - Airflow, Redshift, S3
  - ETLs everywhere
  - Standard Data Structures
- **Schemas & Definitions**
  - Evolving, Interpretation
- **Scalability & Realtime**
  - 24 - 48 hours
  - Reporting vs Access
- **Data Quality**



DeliveryTech

Time to Rethink

Requirements

- Near **Realtime** (ideally <5sec)
- Scalable, reliable
- **Long-term** persistence
- Cloud based
- Event based subscription and batch download
- **Consistent** with source of truth
- Column level ACLs
- **GDPR compliant**
- No over-engineering
- Fast access by index or time or entity
- Cost efficient
- Data representation adapted to use-cases

Limitations

- API only access
- No guaranteed order of events
- Should not be used for critical applications (i.e. applications that are required to place & transmit an order)

Potential First Scope

- Billing in SAP
- Increased response rate for Surveys
- Customer Segmentation for Marketing
- Reports for Vendor Portal
- Forecasting for Logistics
- Standardized Monitoring & Alerting
- Fraud Detection for Reviews
- ….

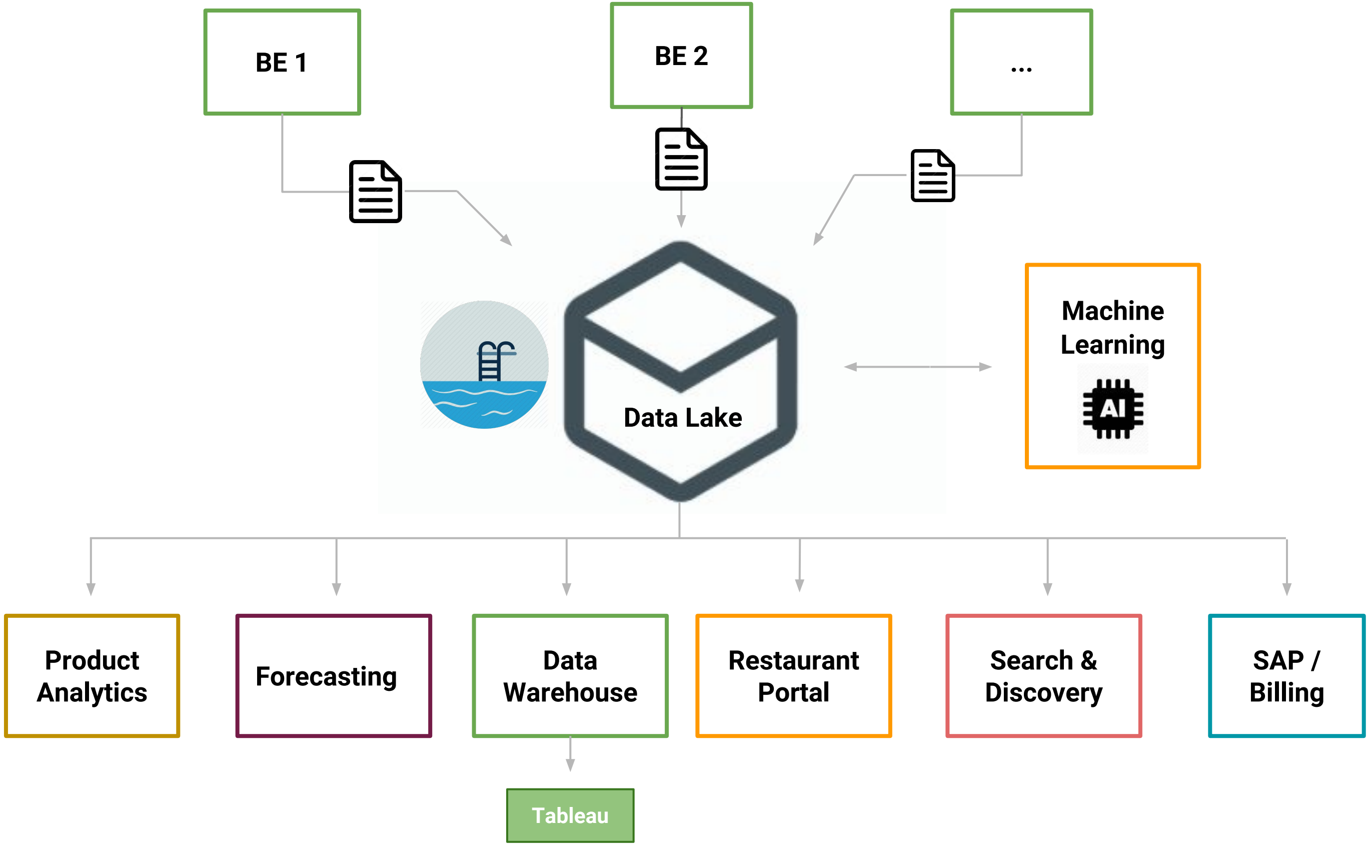
Data Lake - A normalized, consistent, near real-time data platform for global services...

Goals

- Simplify integration
- Make data available for machine learning
- Allow data-exchange between entities and services
- Simplify migrations

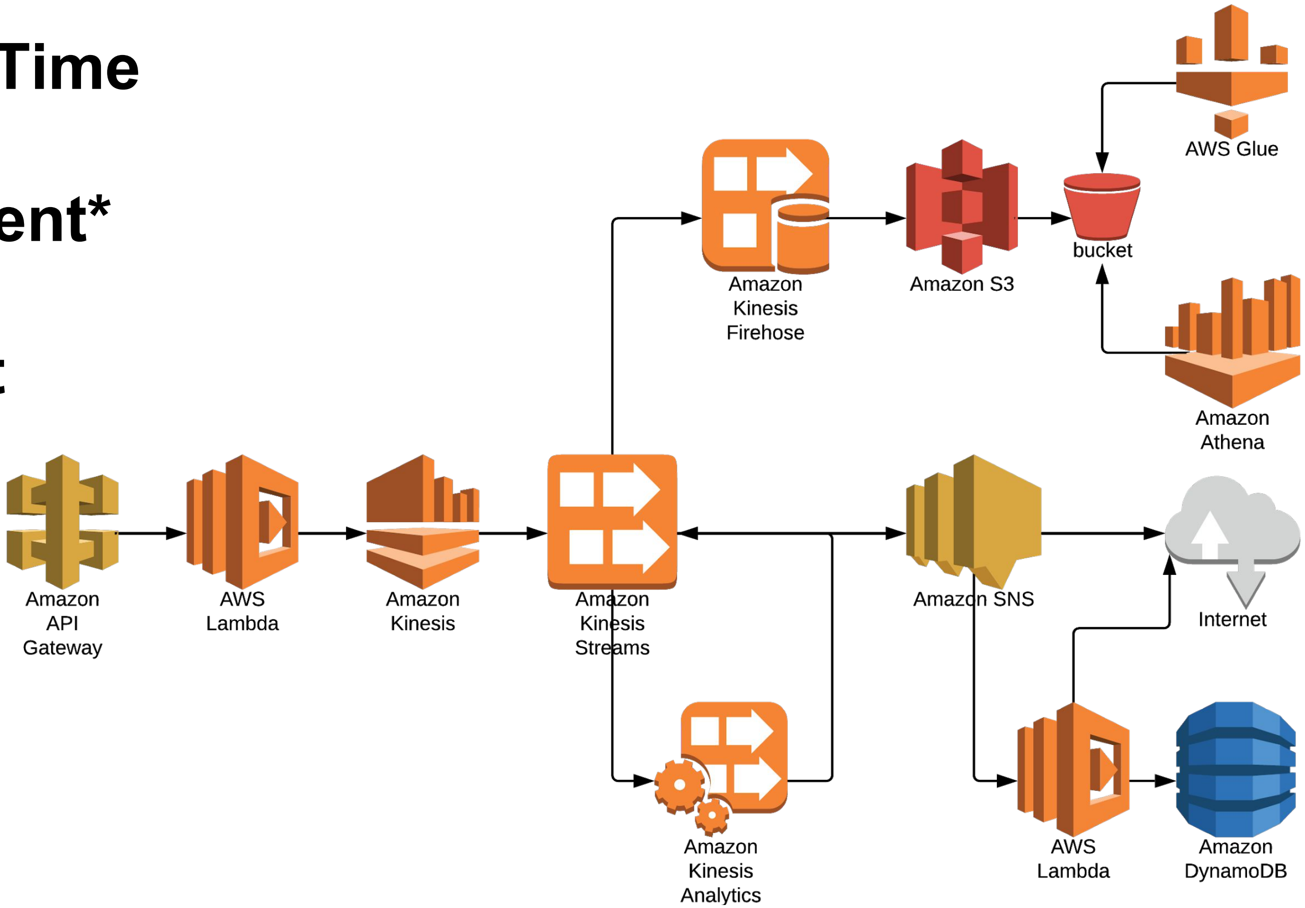
Requirements

- Normalized backend data
- Near real-time data
- Consistent with source of truth
- Data access both via event subscription and batch processing



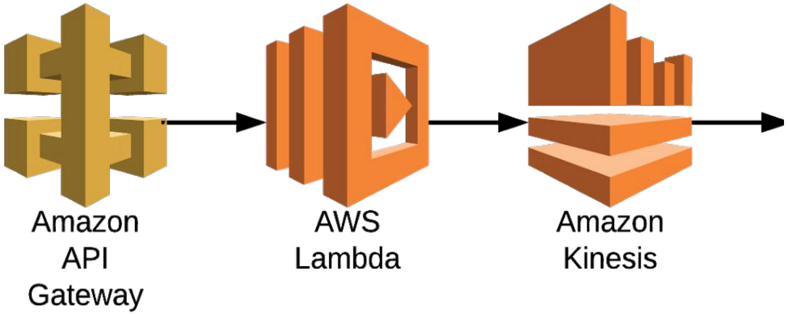
Drawing Board

- **Near Real-Time**
- **Cost Efficient***
- **Permanent**
- **Scalable**



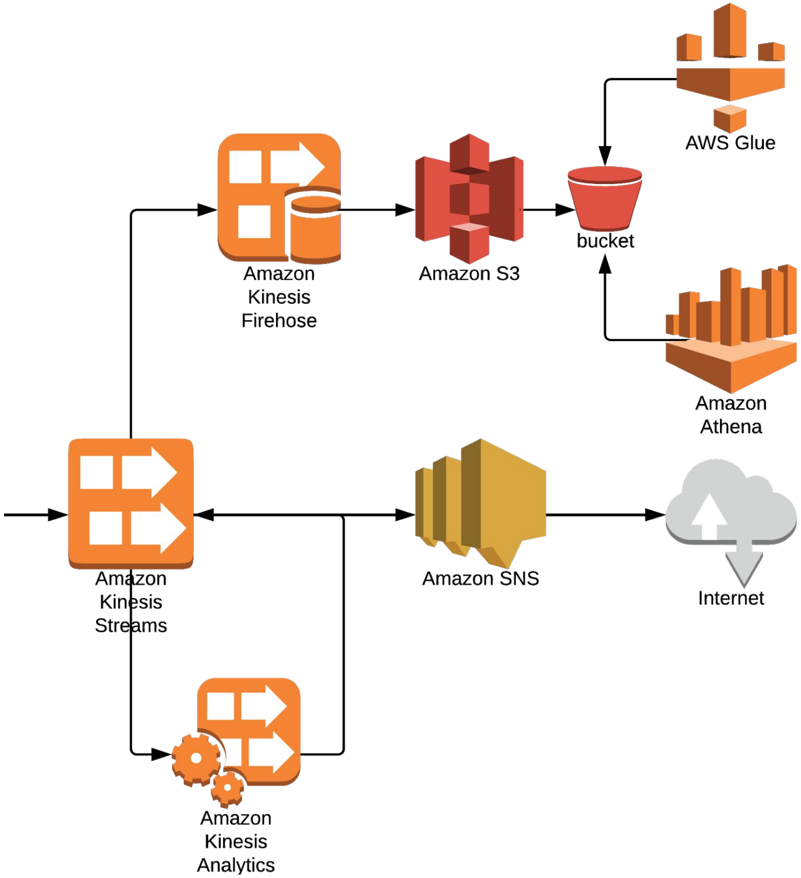
Ingestion

- **API Gateway**
  - Swagger
  - Versioning
  - Authentication
- **Lambda**
  - Scales
  - Validation
  - Serverless
- **SDKs**
  - Client-Side validation
  - Direct ingestion into kinesis



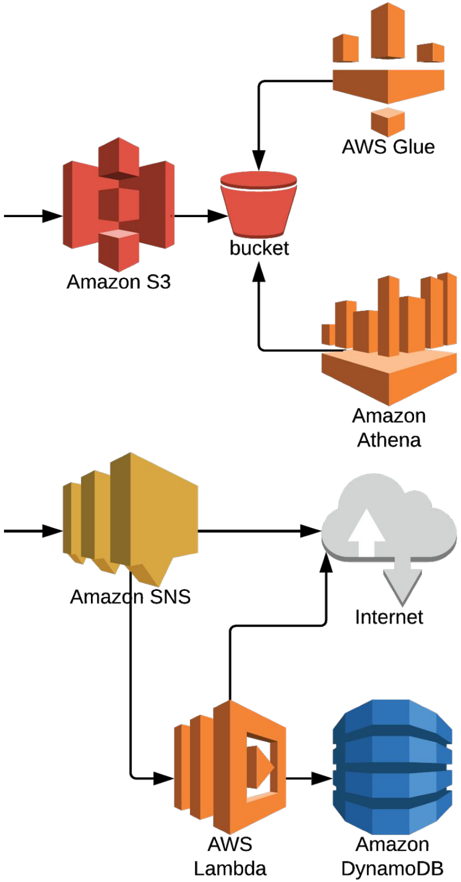
Processing

- **Kinesis**
  - Stream-based, Simple, Scales, …
- **Kinesis Firehose**
  - S3, Redshift, Elastic Search
- **Kinesis Analytics**
  - Window-based SQL Queries
  - Merge Streams
- **S3***
  - S3 storage incl. lifecycle, replication and publish/access



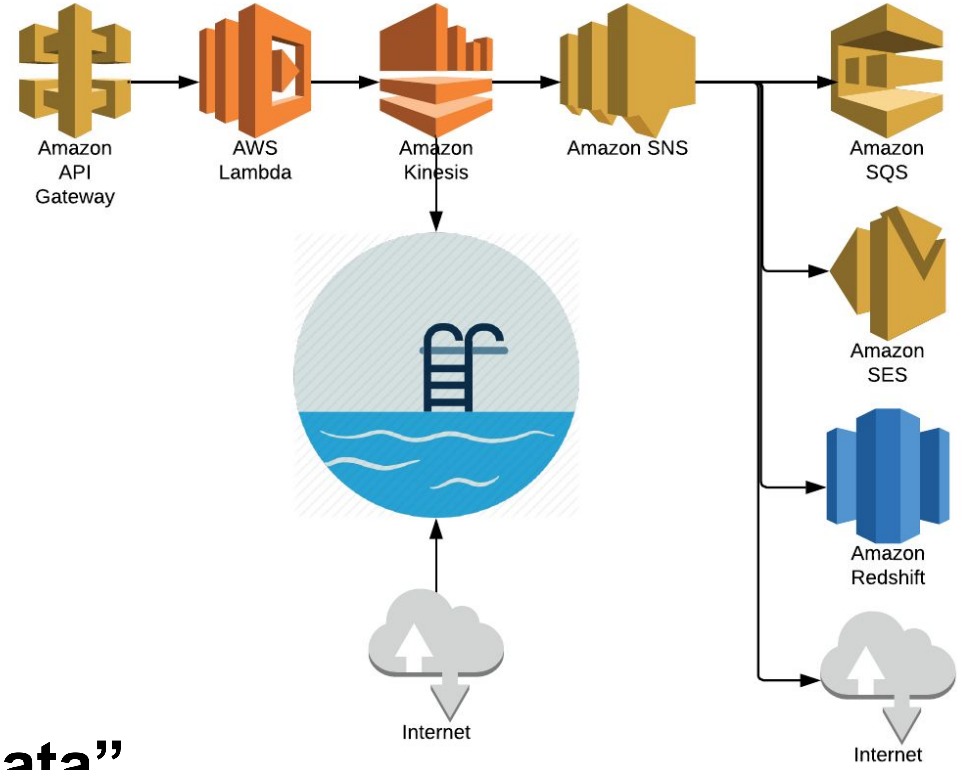
Consume

- **SNS**
  - Filtering!
  - Fanout Events
- **DynamoDB**
  - Enrich Data
- **Athena & Glue**
  - On Demand Queries
  - Detect Schemas On-the-fly
- **Further Features**
  - APIs for Single & Batch
  - Subscriptions
  - Replay
  - ...



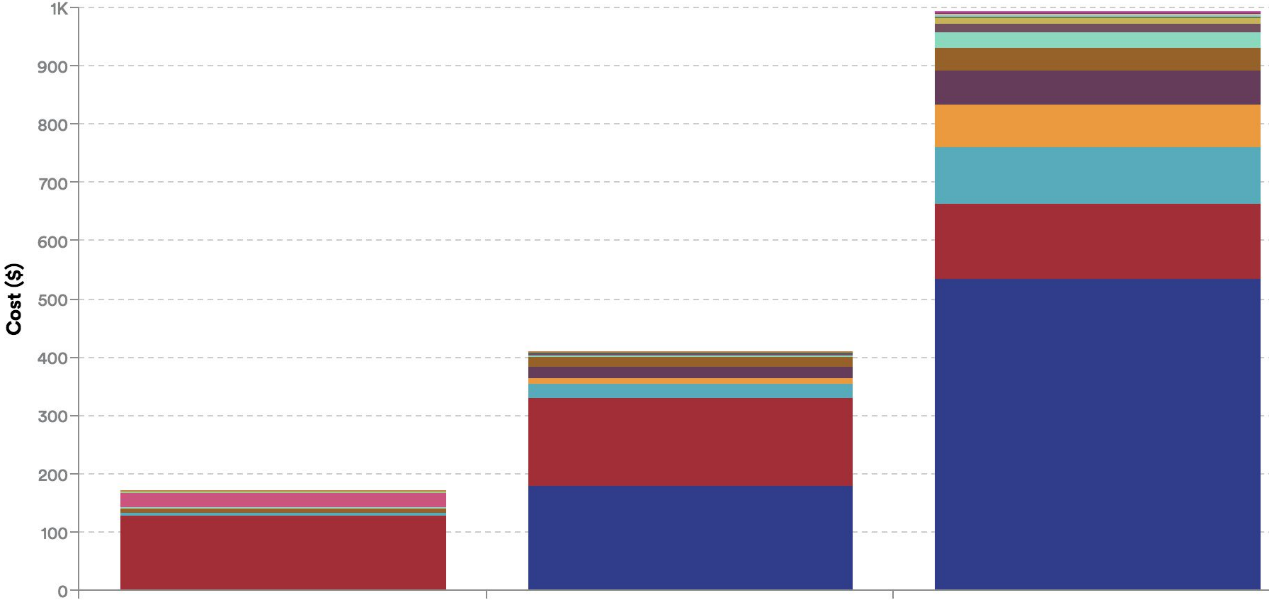
Consumer Blueprints

- **Access**
  - 80% API
- **Store**
  - SNS Consumer
  - REST Service
  - DB
- **Push**
  - SDKs
  - Subscription



"Unleash the Power of Data"

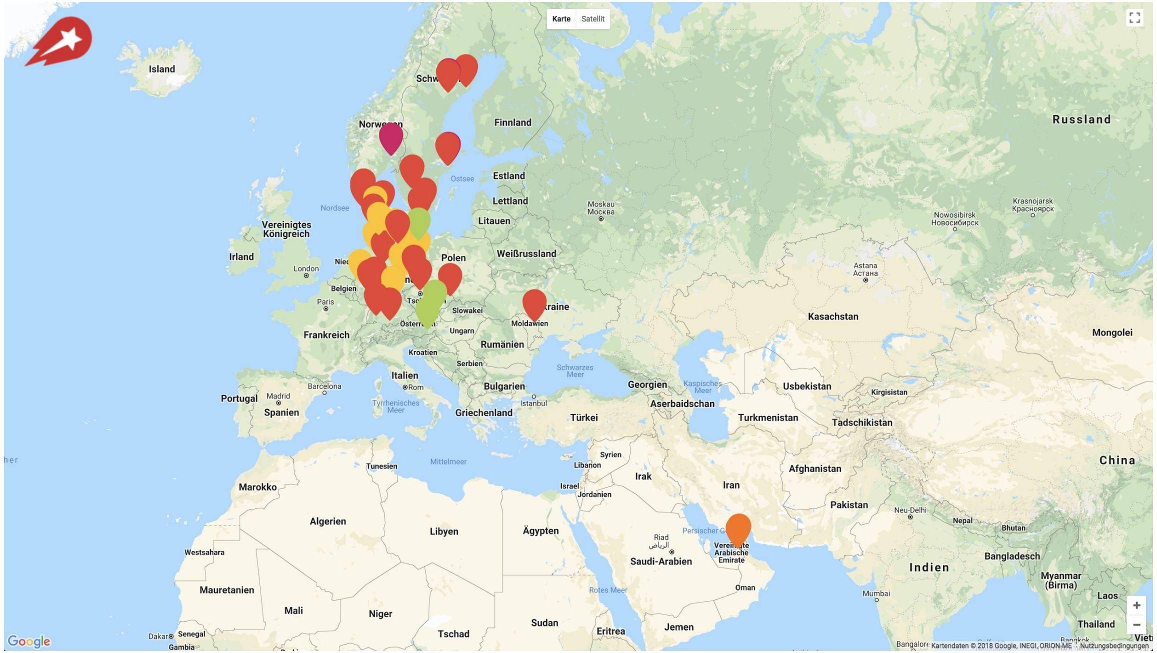
- **>250.000 Orders per Day**
- **5 - 10 Events per Order**
- **New Team of 4**
- **< 3 Month**
- **< 500 $***



Order Map

- **Visualize new Orders**
- **All Platforms***
- **Stack**
  - SNS to SQS
  - NodeJS, Socket.io

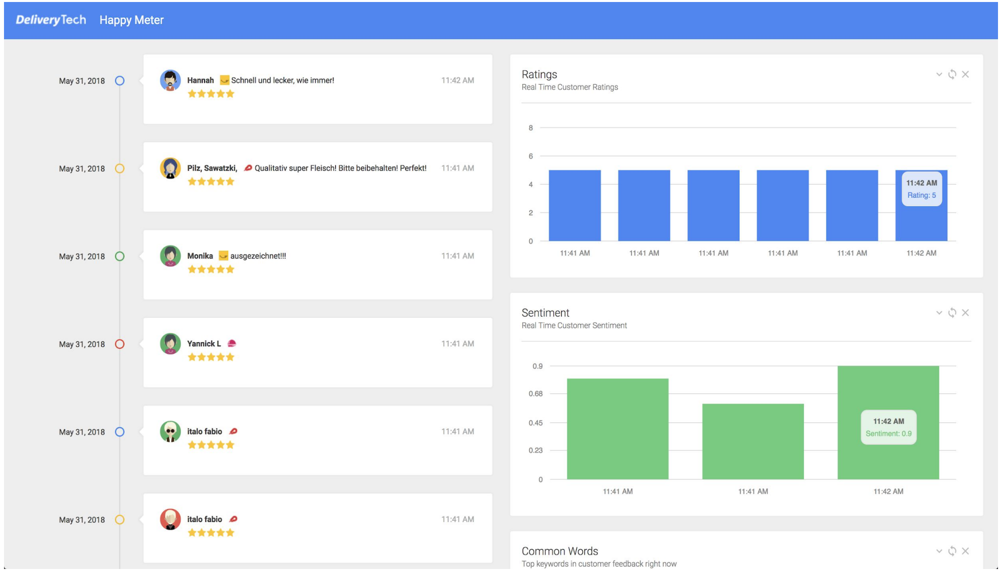
https://ordermap.rps-ops.com/ (internal only)



Happy Meter

- **Customer feedback**
- **Sentiment Analysis**
- **Consumes Data from within another cloud***

https://happymeter.deliveryhero.com/ (internal only)



DeliveryTech Happy Meter

May 31, 2018 — Hannah Schnell und lecker, wie immer! ★★★★★ 11:42 AM

May 31, 2018 — Pilz, Sawatzki, Qualitativ super Fleisch! Bitte beibehalten! Perfekt! ★★★★★ 11:41 AM

May 31, 2018 — Monika ausgezeichnet!!! ★★★★★ 11:41 AM

May 31, 2018 — Yannick L ★★★★★ 11:41 AM

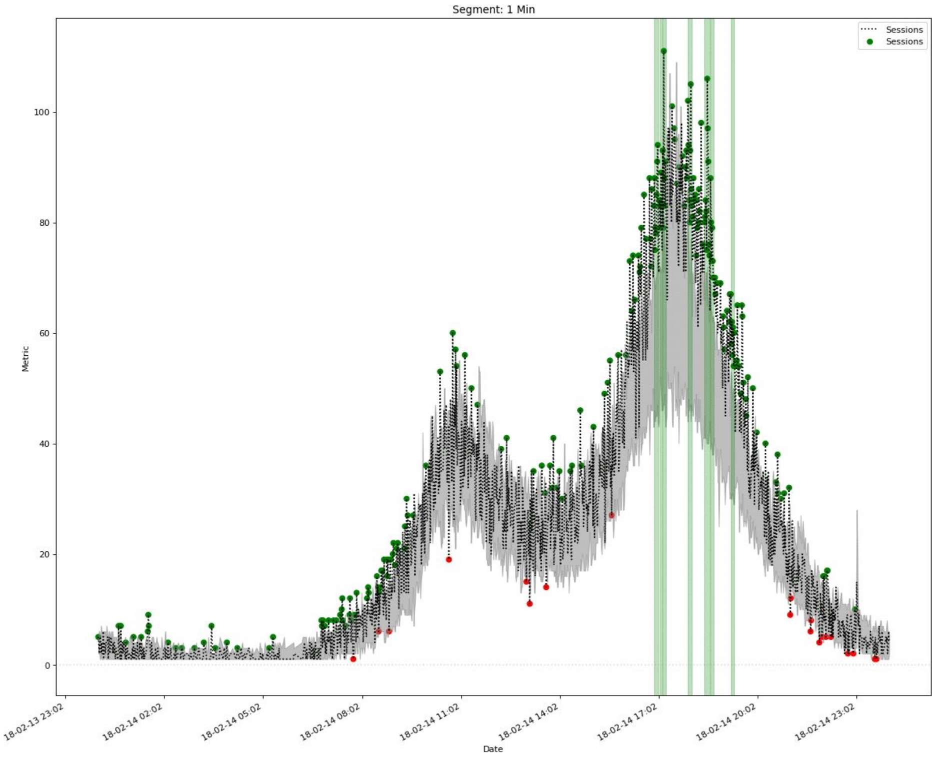
May 31, 2018 — italo fabio ★★★★★ 11:41 AM

May 31, 2018 — italo fabio ★★★★★ 11:41 AM

Predictions - Prophet Alghorythm | UpTime (Ongoing)

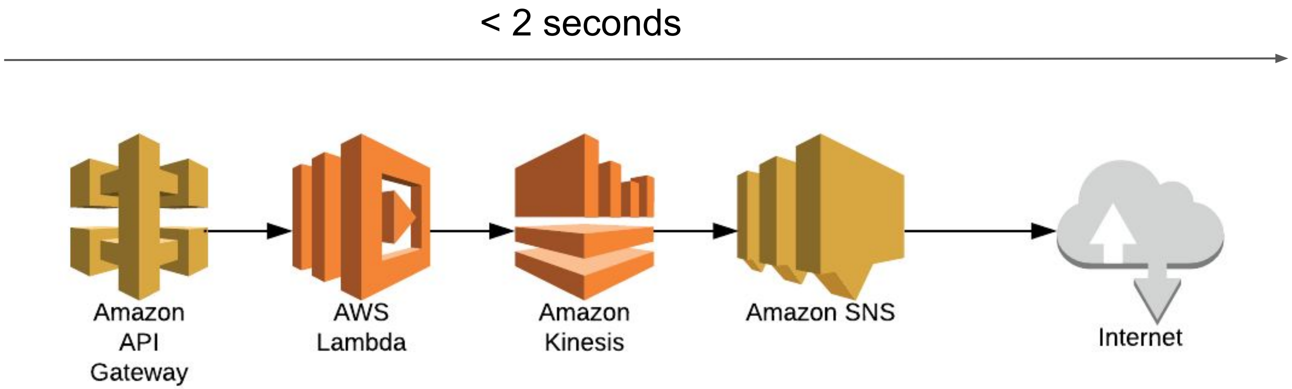
DeliveryTech

- **Predict upcoming Downti**
- **Based on Historical and (near) Real-Time data**



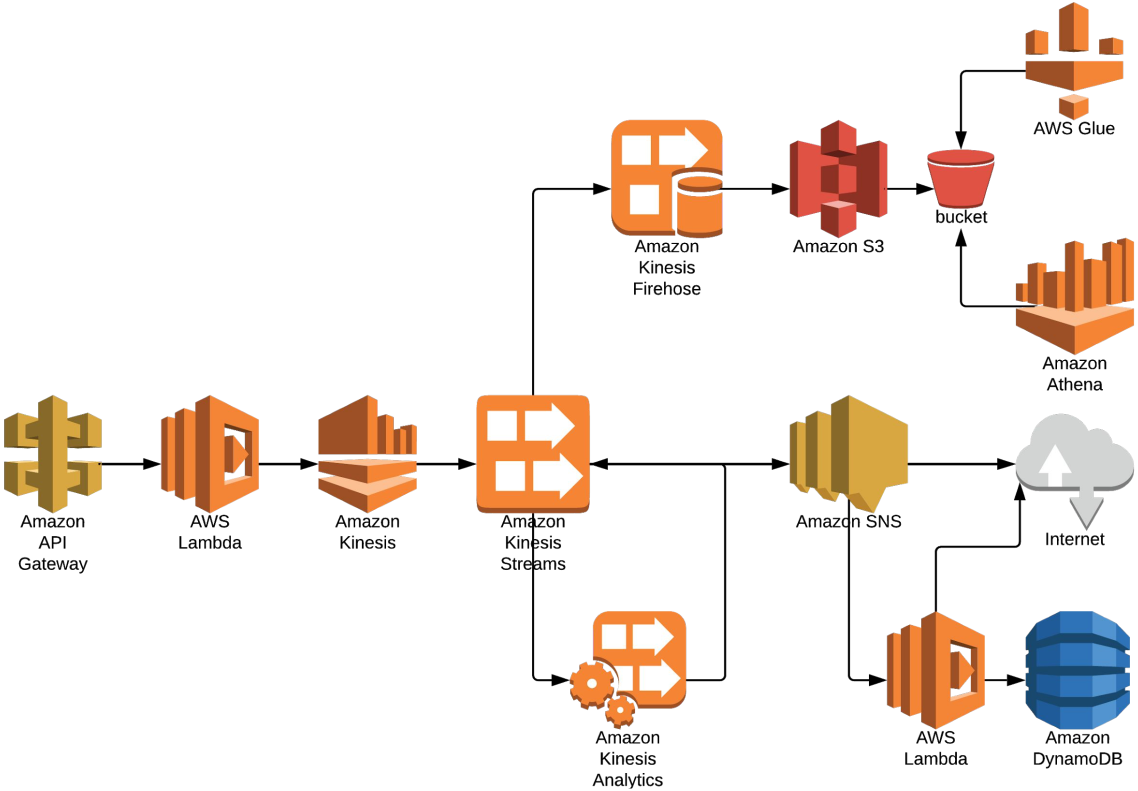
Near Real-Time

- **How real-time do you need?**
- **How much control do you need?**
- **What’s your critical path?**



Click it together!

- < 1 Day
- No Code*
- No Machines
- No Automation
- No Consistency



Conclusion

- **AWS your One-Stop-Shop**
- **Serverless for faster Go-to-Market**
- **Use Case driven because “No Value, No Support”**
- **Duplication, Consistency and Ordering**
- **Permanent Storage vs GDPR**
- **Event Bus first, Data Lake later?**
- **Know your Customers (Consumers)**
- **Be Aware of the Limits**
    - No control over Partitioning with Kinesis Firehose*
    - Kinesis Consumer Count

Daniel Manzke in