

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023



Replacing the Data Platform during hyper-growth

Paweł Kupidura

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Agenda

- Brief history of Data at Bolt
- Hitting the limits
- The new Data Platform
- Challenges
- Lessons learned

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Who are we?

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Bolt

- Estonian startup
- Founded in 2013 (10 years 🐝)
- Uber competitor



WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Leading mobility company

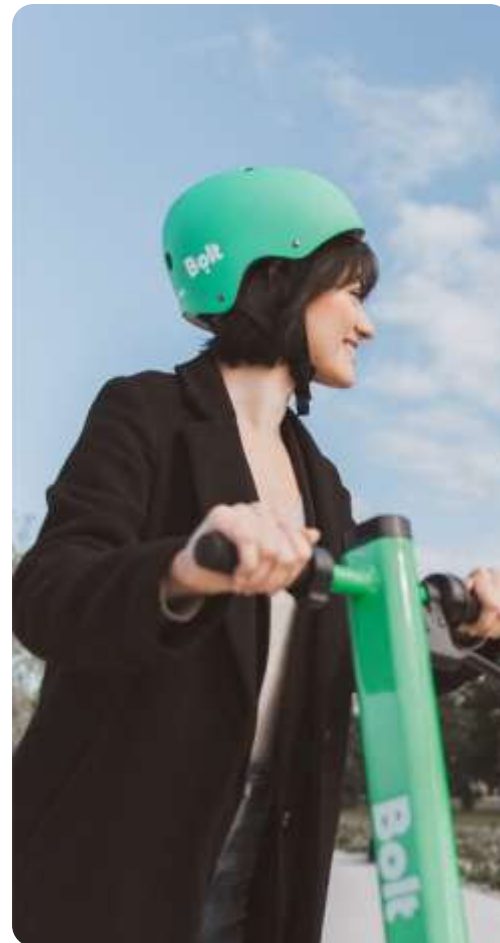
Rides

3+ million partners
in 45 countries



Rentals

260,000 scooters
in 25 countries



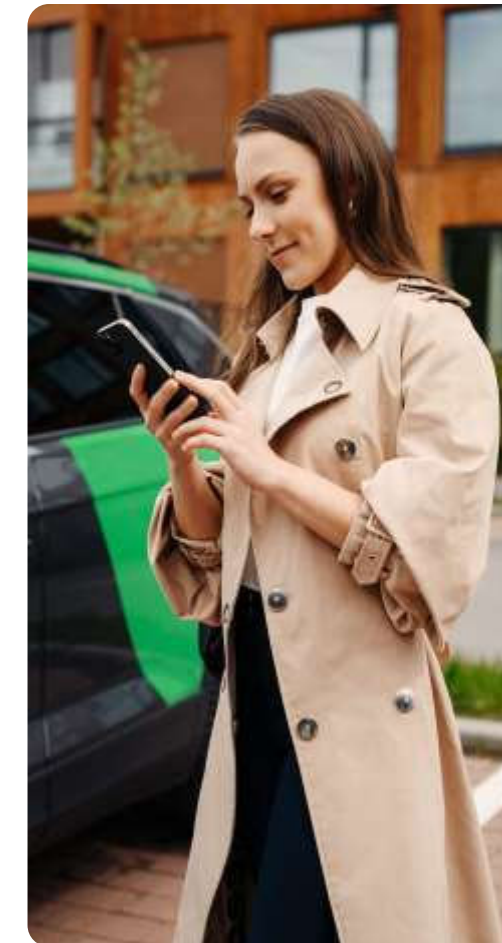
Delivery

30,000+ restaurants in 20
countries



Drive

Available in
5 countries



Market

Fast grocery
delivery



WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

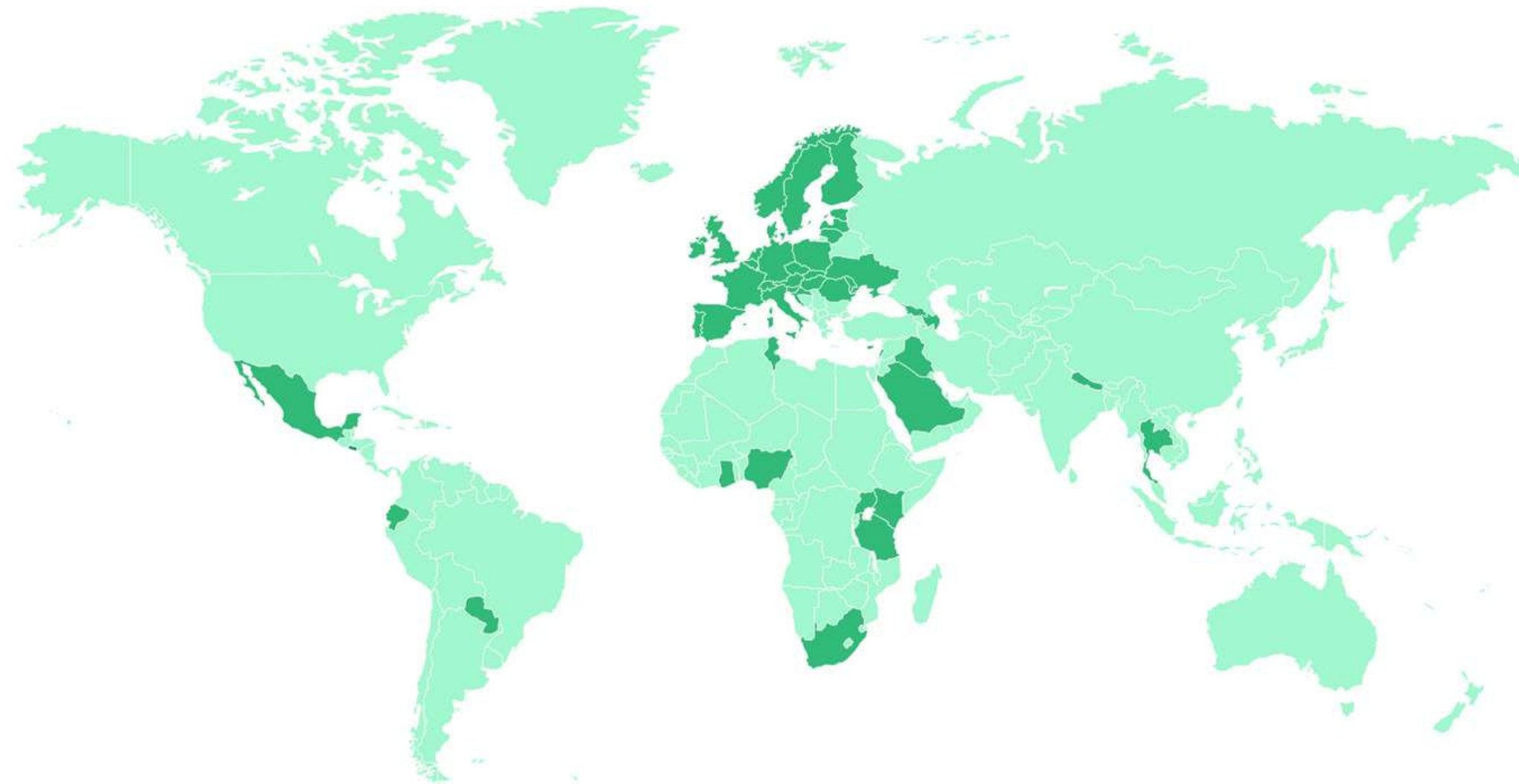
WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Leading mobility company



100+ million
customers

500+ cities
across 45 countries

3000+
employees

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Data at Bolt

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

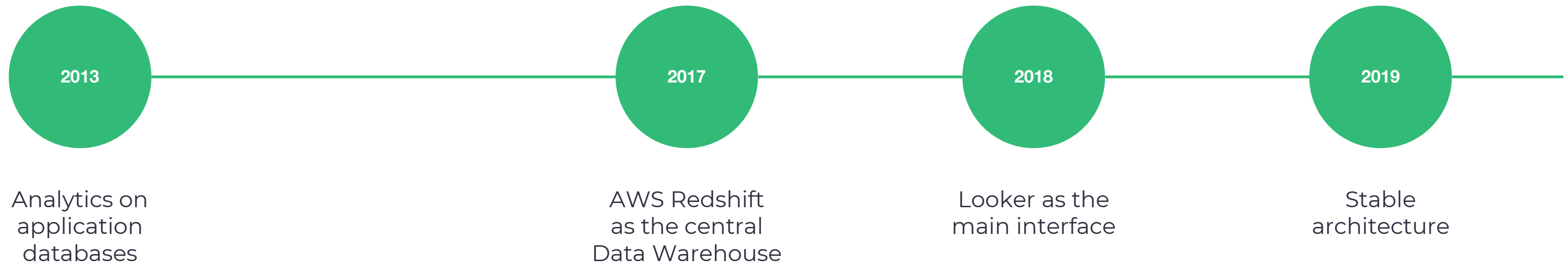
WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA

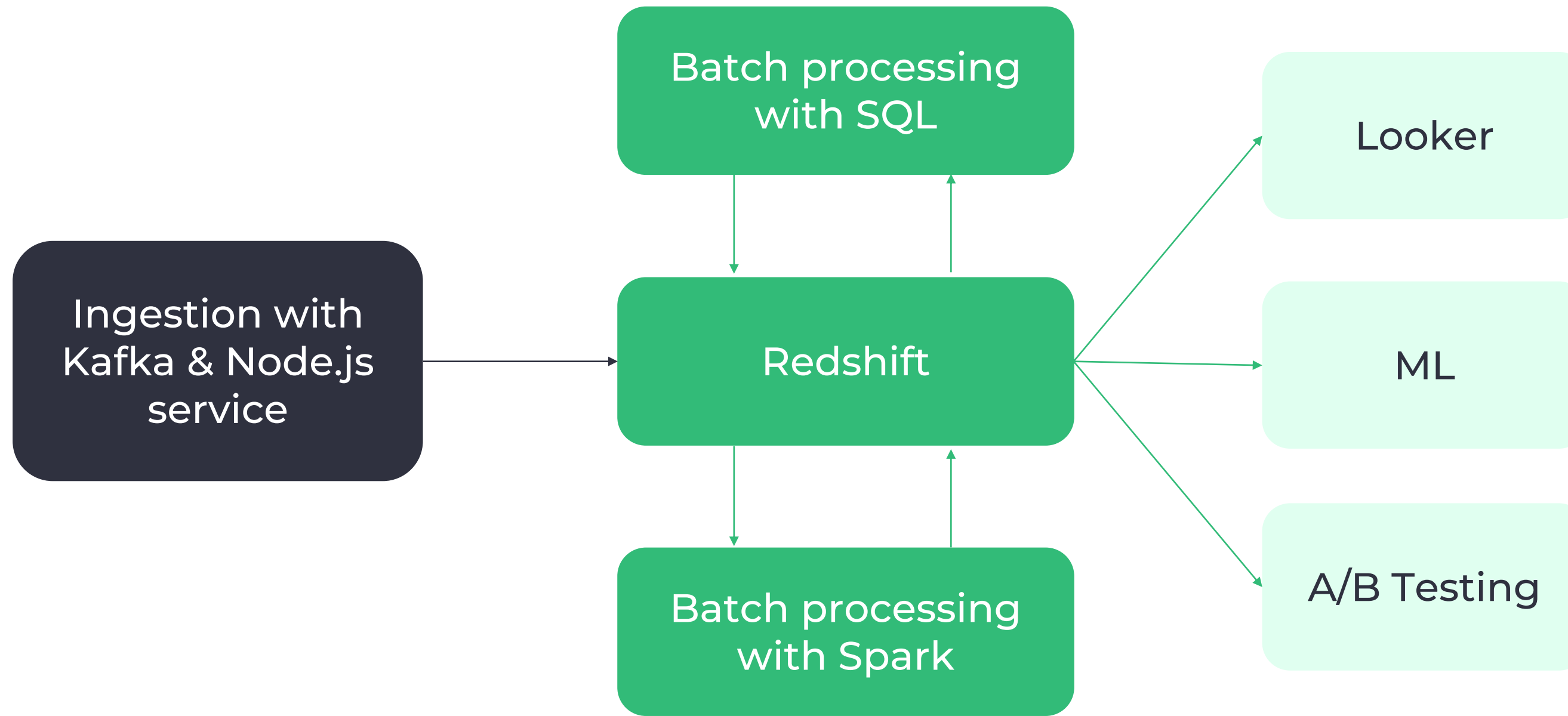


Data Platform 1.0





Data Platform 1.0



WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

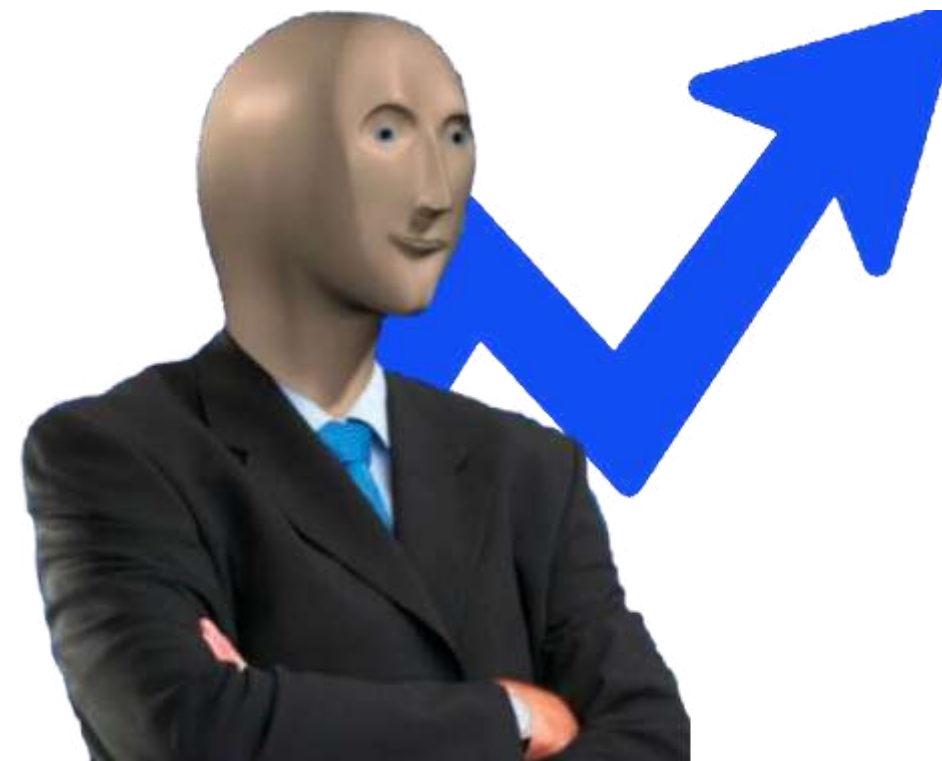
WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Hyper-Growth





Hyper-Growth

Data growth

- Tens of millions of customers
- New products

Usage growth

- Data Platform users doubled every year
- New ML models
- More experiments

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

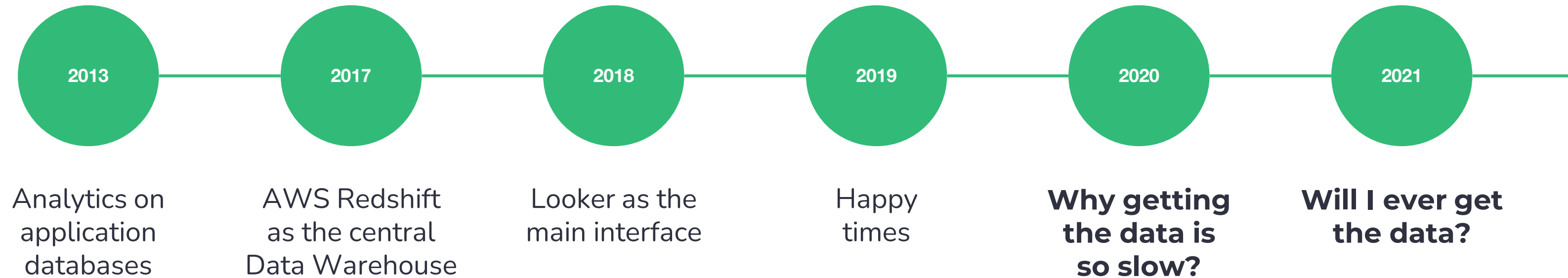
CLOUD AGAINST DATA



Did it scale?



Did it scale?



WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Problems

- Shared resource

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Problems

- Shared resource
- **Limited scaling options**
- Coupled storage and compute
- Limited scaling options



Problems

- Shared resource
- Limited scaling options
- **Ingestion delays**
- Hundreds of tables
- Users got used to fast data delivery

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Problems

- Shared resource
- Limited scaling options
- Ingestion delays
- “Looker Mondays”



DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Problems

- Shared resource
- Limited scaling options
- Ingestion delays
- “Looker Mondays”
- **Costs**



WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Data Platform 2.0



Options considered

Cloud Data Warehouses

- Snowflake
- BigQuery

Decoupled approach

Query engines

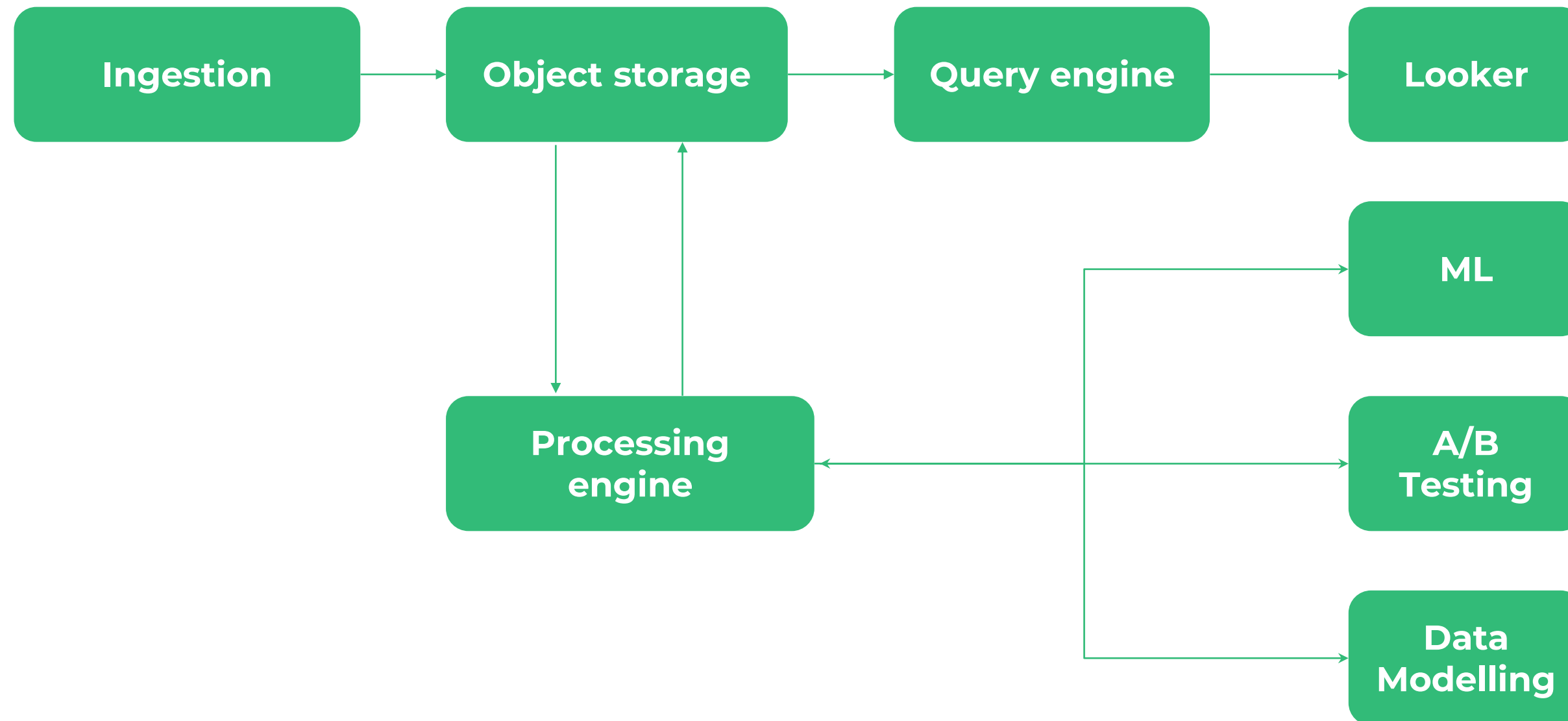
- Presto
- Trino
- Spark SQL

Storage layers

- Delta Lake
- Apache Hudi

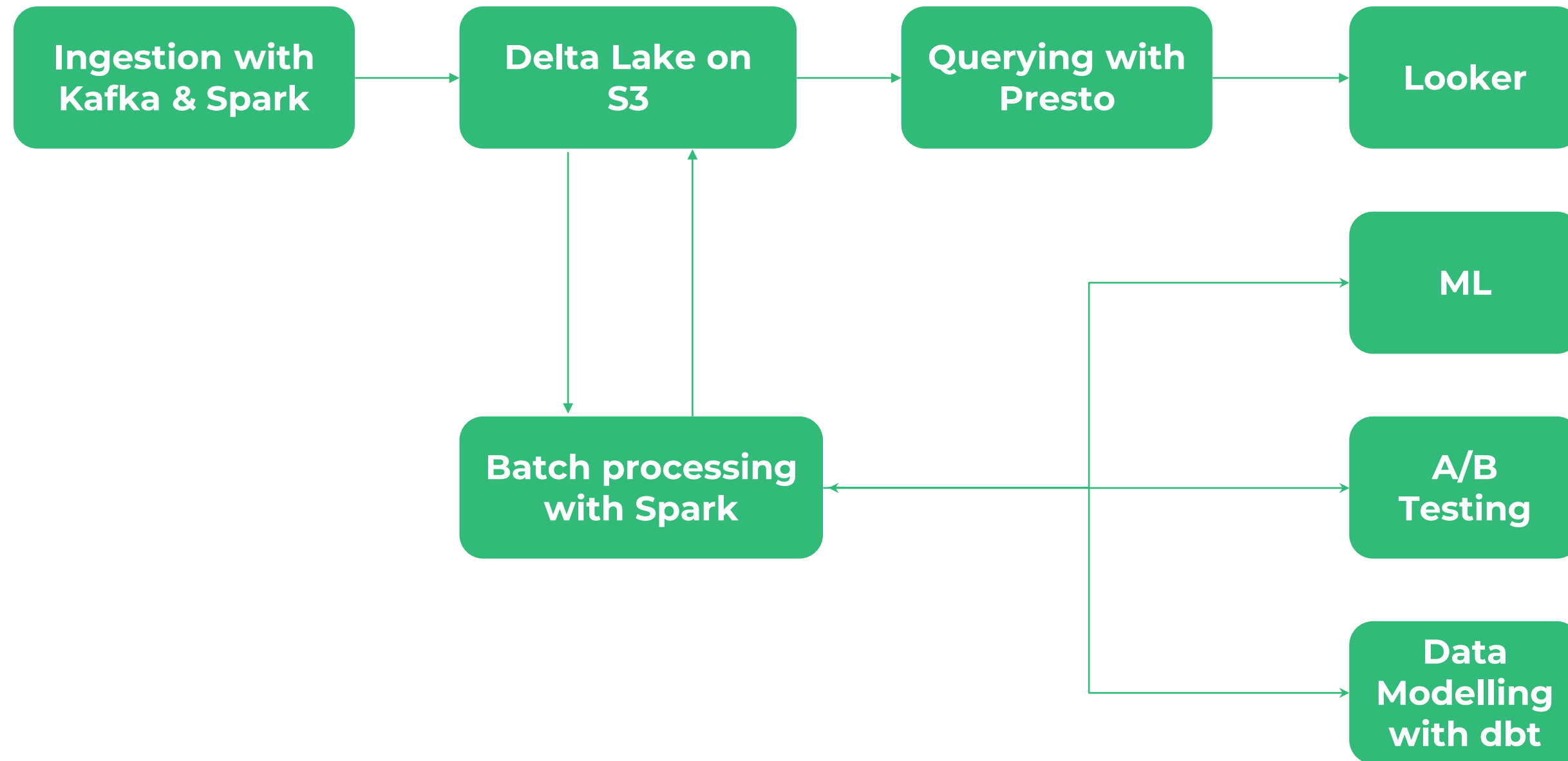


New architecture





New architecture





Main components



Open-source distributed processing system.

Industry standard, compatible with many other tools.

In our stack since 2019.



Open-source storage layer, built with Apache Spark.

ACID transactions on top of Parquet data in S3.

Upserts and deletes on columnar formats.



Open-source distributed query engine.

Built by Meta to handle analytical workloads at their scale.

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Challenges

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Challenges Migration

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Migration challenges

- **No interruptions**
- Business critical service
- Thousands of direct users

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Migration challenges

- No interruptions
- **Adapting queries**
- Thousands of queries
- Batch jobs & ML models
- Not just a syntax change



Migration challenges

- No interruptions
- Adapting queries
- **Engaging owners**
- Diverse user base
- Influencing priorities of owners



Migration approach

- **No interruptions**
- Adapting queries
- Engaging owners
- Maintain two parallel platforms
- Gradual migration



Migration approach

- No interruptions
- **Adapting queries**
- Engaging owners
- Automating whenever possible
- Better tools for data modelling
- Ruthless cleanup



Migration approach

- No interruptions
- Adapting queries
- **Engaging owners**
- Self-service whenever possible
- Detailed documentation
- Coaching sessions
- Annoying owners

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Was it worth it?



Results

- **Shared resource**
- Limited scaling options
- Ingestion delays
- “Looker Mondays”
- Costs
- **Isolated environments**
- **Decoupled components**



Results

- Shared resource
- **Limited scaling options**
- Ingestion delays
- “Looker Mondays”
- Costs

Flexible scalability



Results

- Shared resource
- Limited scaling options
- **Ingestion delays**
- “Looker Mondays”
- Costs

Stable, predictable data delivery

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Results

- Shared resource
- Limited scaling options
- Ingestion delays
- **“Looker Mondays”**
- Costs
- **Dedicated Presto clusters**
- **Working autoscaling**

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Results

- Shared resource
 - Limited scaling options
 - Ingestion delays
 - “Looker Mondays”
 - **Costs**
- **Better understanding of costs**
 - **Flattened growth curve**

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Key takeaways

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Key takeaways

- **Overcommunicate**
- Benchmark dilligently
- Clean up ruthlessly

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Key takeaways

- Overcommunicate
- **Benchmark dilligently**
- Clean up ruthlessly

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Key takeaways

- Overcommunicate
- Benchmark dilligently
- **Clean up ruthlessly**

WORKSHOP DAY ONSITE
OCTOBER 4, 2023

CONFERENCE DAY ONSITE
OCTOBER 5, 2023

WORKSHOP DAY ONLINE
OCTOBER 6, 2023

DATAMASS GDAŃSK SUMMIT

CLOUD AGAINST DATA



Thanks for joining!