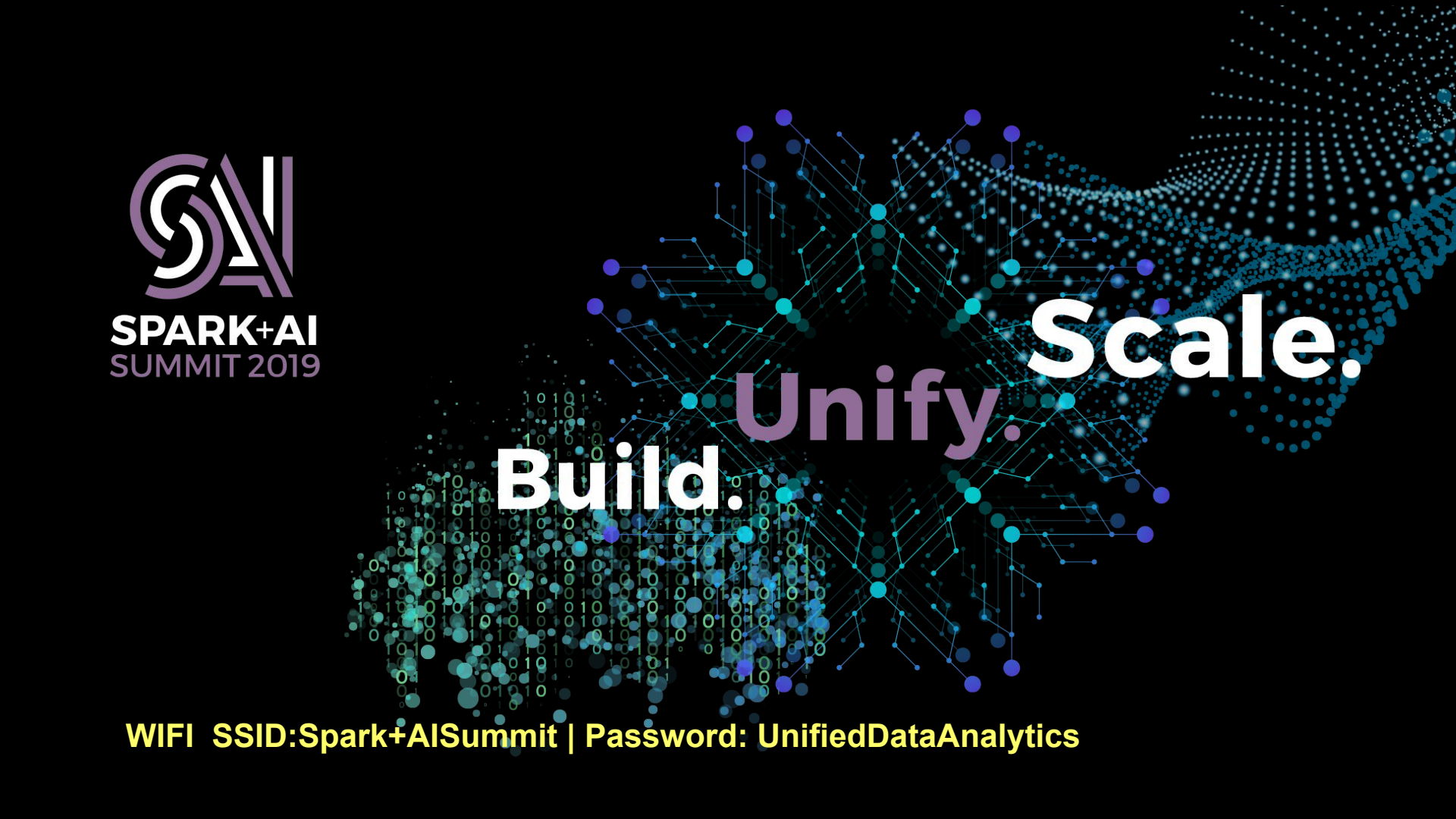

SPARK+AI
SUMMIT 2019
Build.
Unify.
Scale.
WIFI SSID:Spark+AISummit | Password: UnifiedDataAnalytics

AI-Powered Retail Experience with Databricks

Akhil Dhingra, Zalando
Saurav Verma, Zalando

#UnifiedDataAnalytics #SparkAISummit

Zalando SE

- Founded in 2008 in Berlin.
- Europe's leading online fashion platform
- Connects customers, brands and partners.

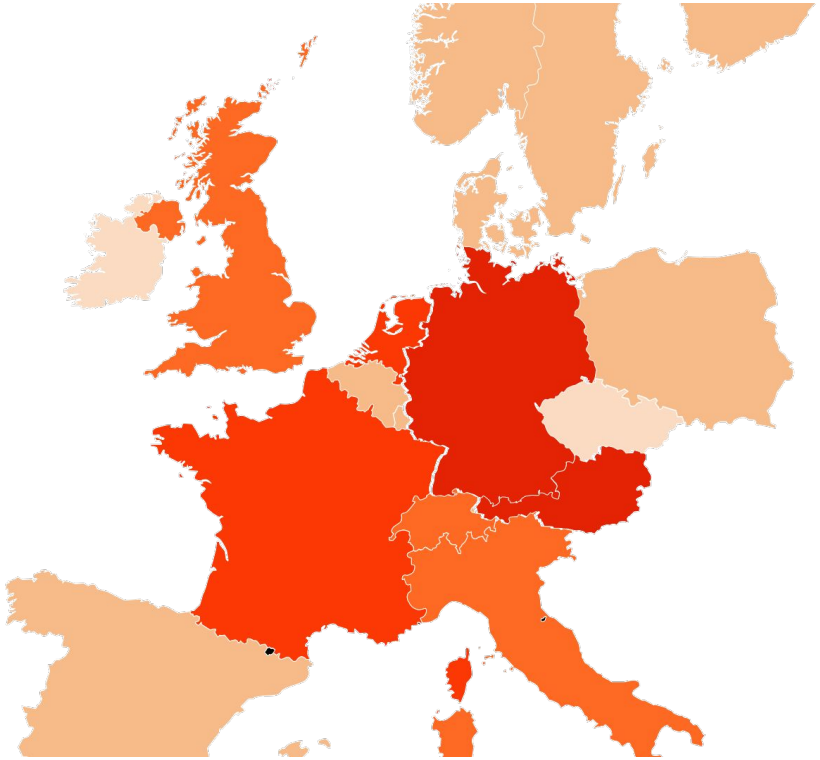
Zalando SE

17
countries

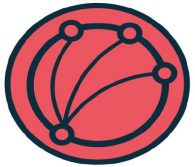
>1,000m
visits last quarter

>80%
mobile traffic

>26m
active customers



Big-Data Stack @ Zalando



About Us

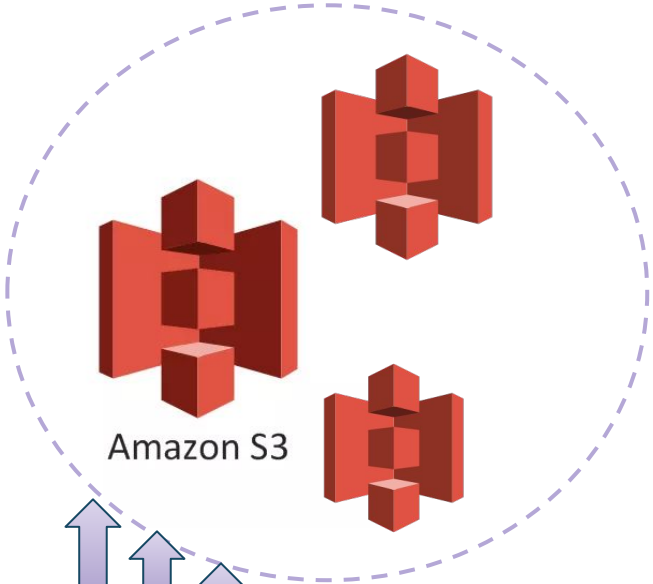
Akhil Dhingra
Product Manager, Data Solutions @Zalando
Exp: 7+ Years, Ex-Groupon, Ex-Wingify | MBA

Saurav Verma
Senior Engineer, Data Lake @Zalando
Exp: 9+ Years , Ex-Visa | Masters NUS

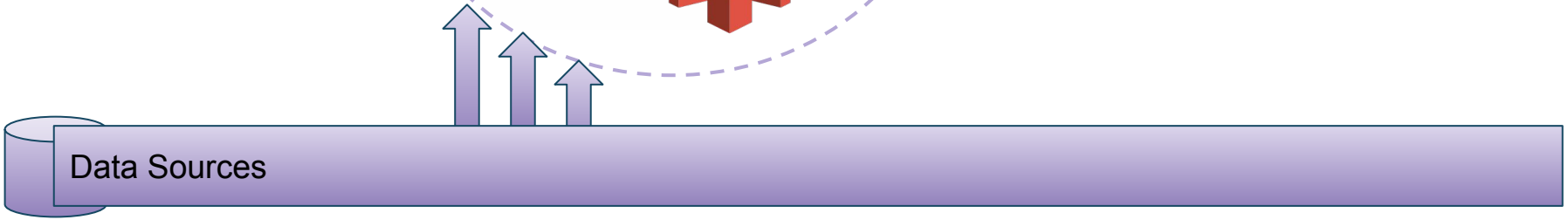
Data Platform

Data Sources

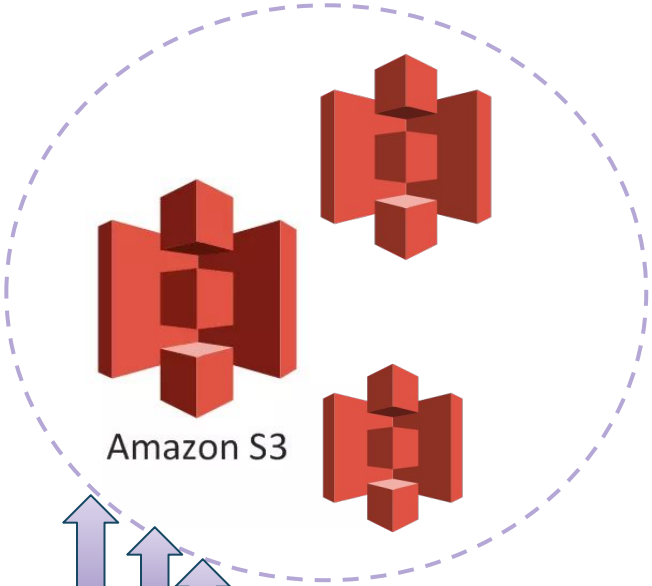
Data Platform



- Data Lake on top of S3



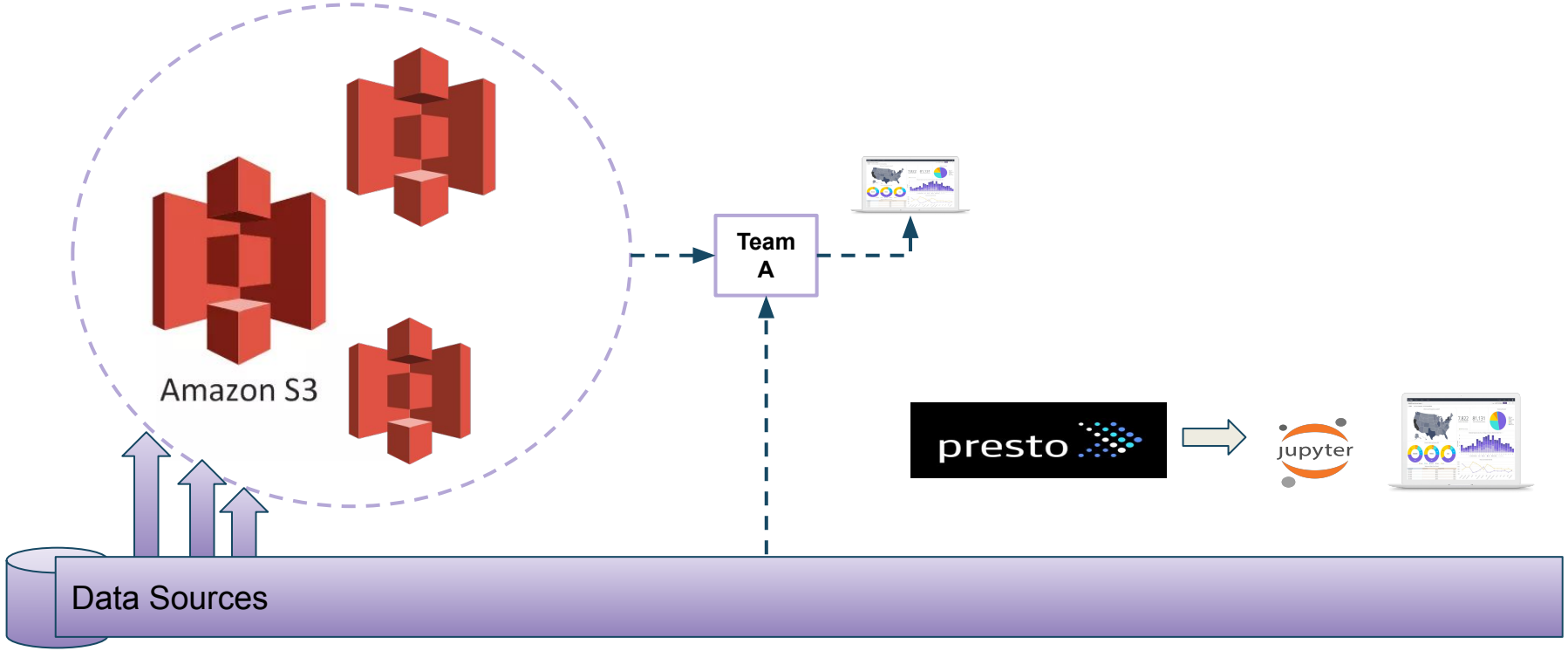
Data Platform



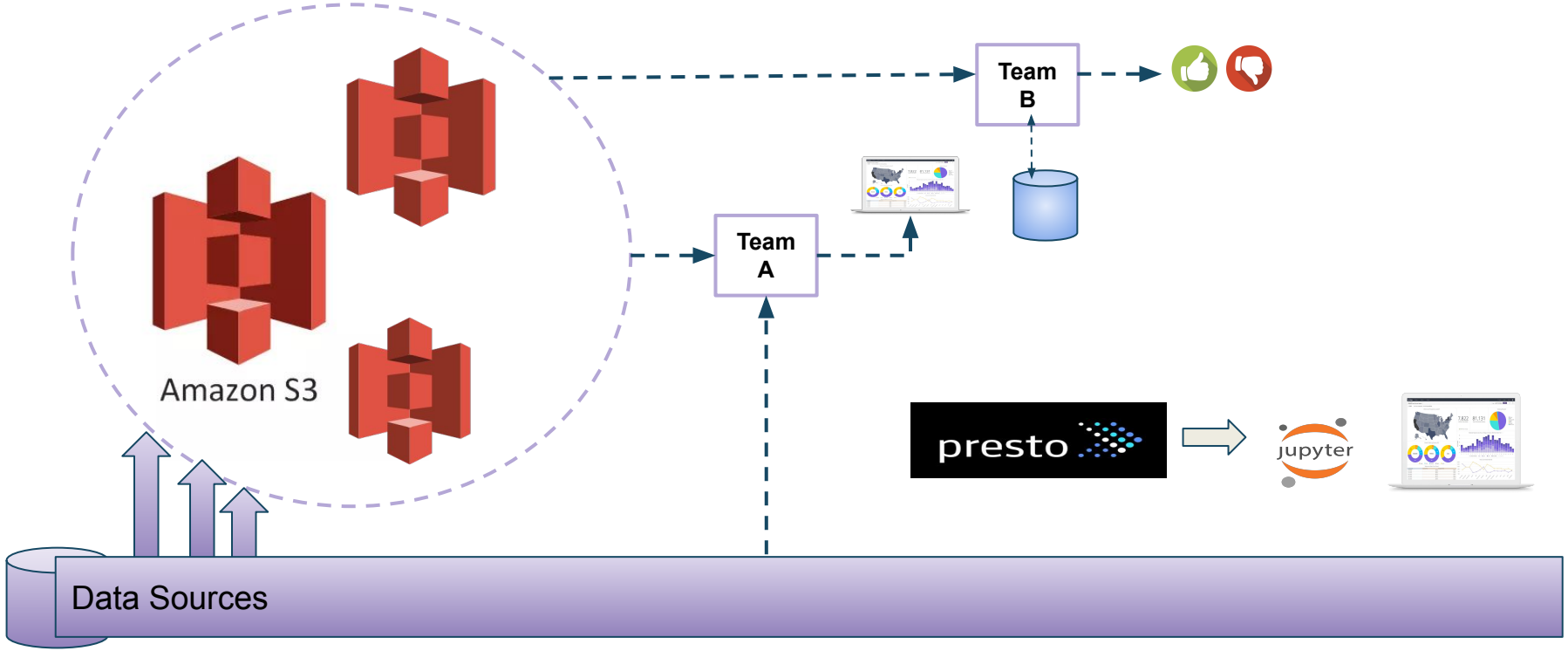
- Multi-tenant / single compute: more ingestion pipelines

Data Sources

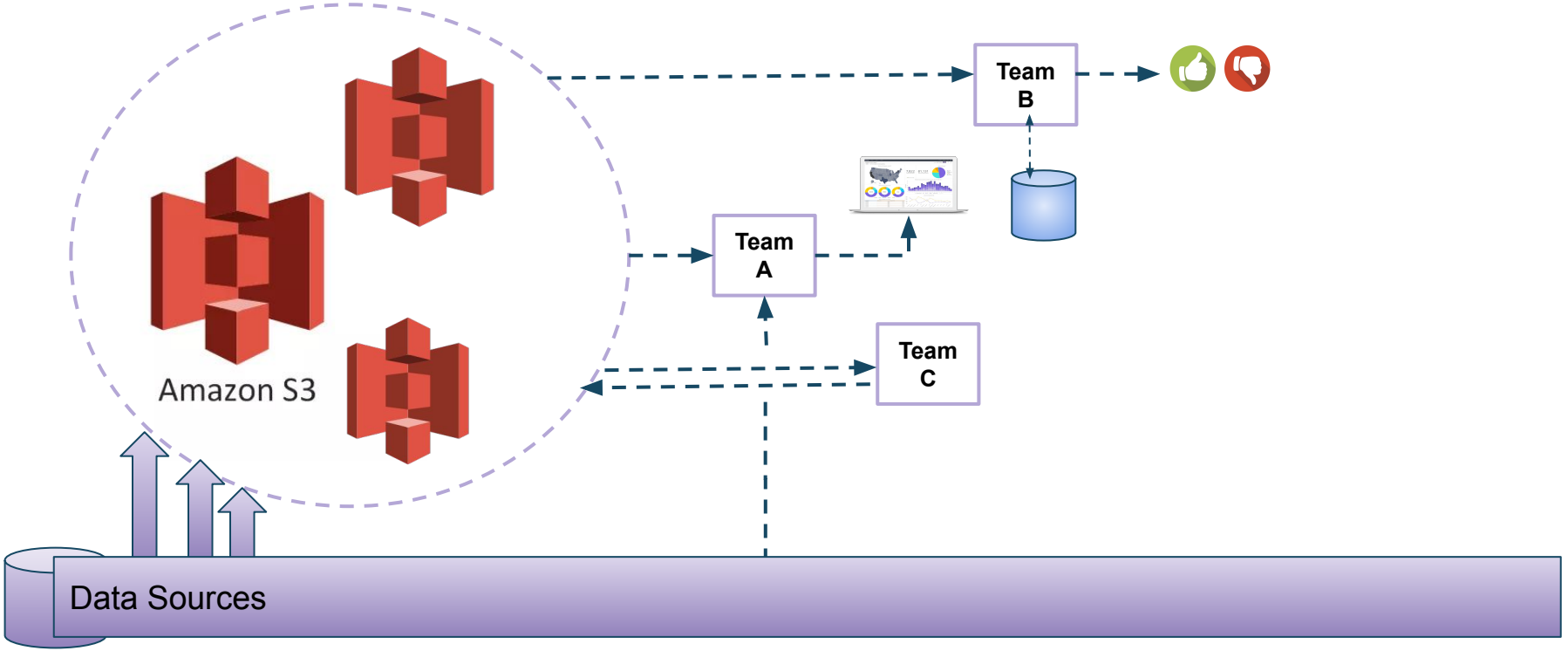
Many Use Cases



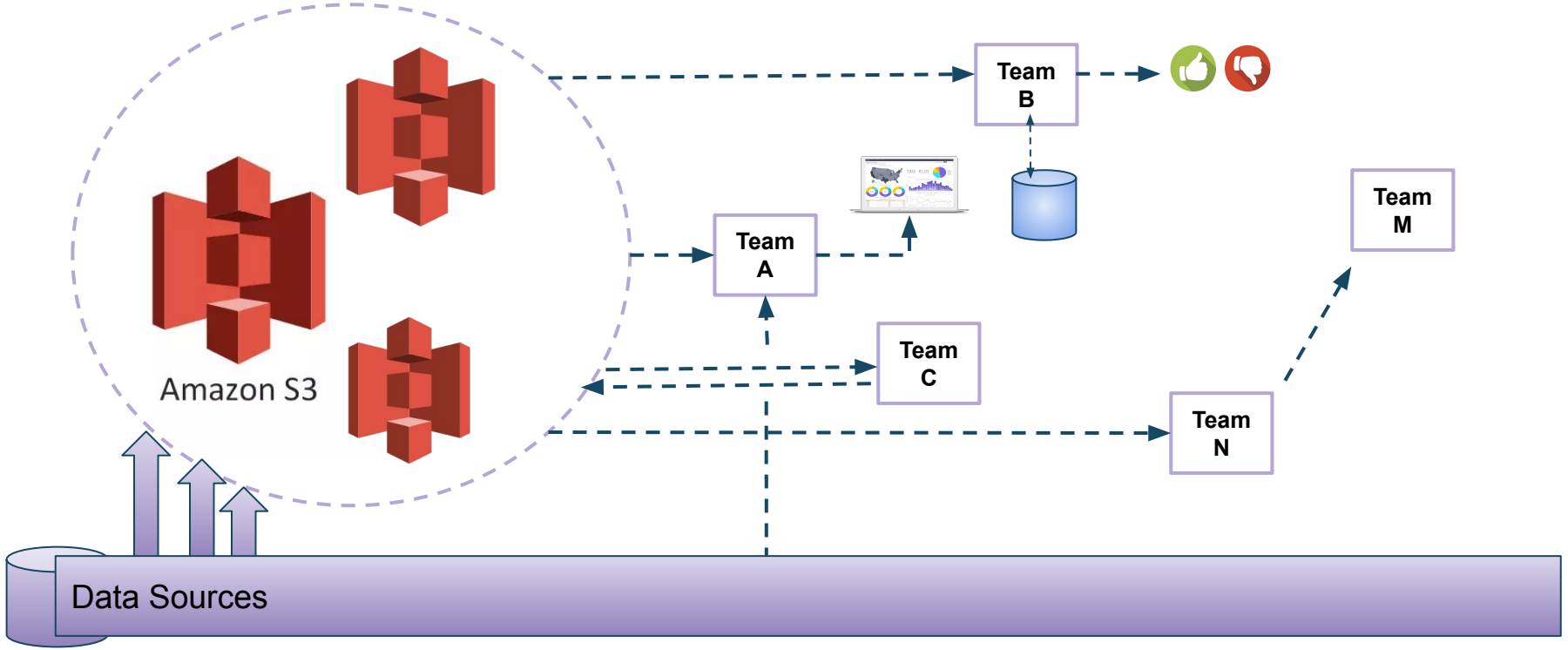
Many Use Cases



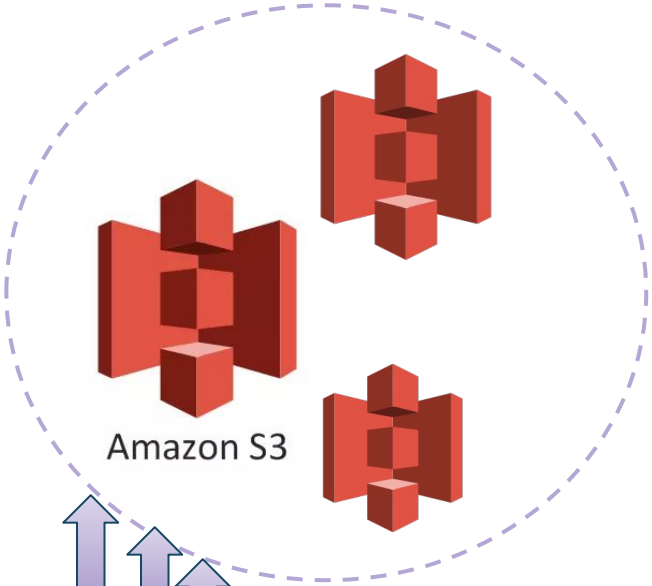
Many Use Cases



Too Many Use Cases



Too Many … Compute



Stream

Team B

Training

Team M

Auto-Scale

Team A

Compute

Team C

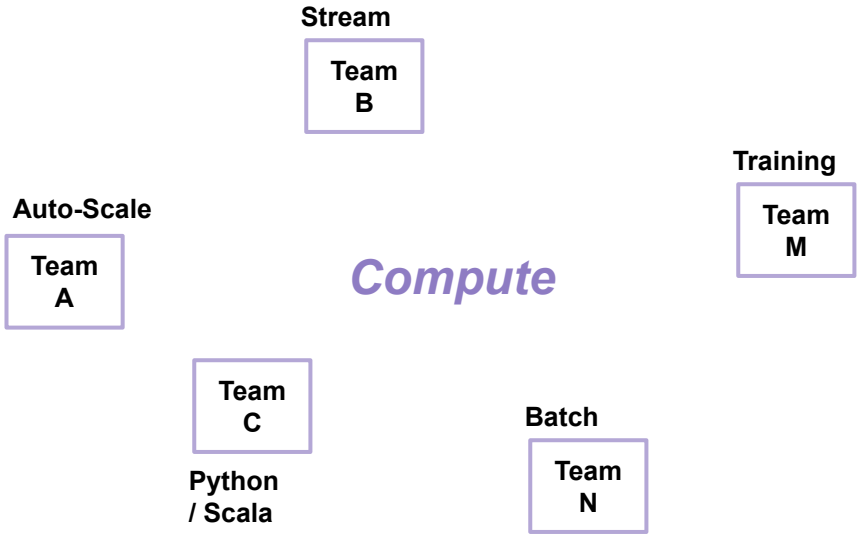
Python / Scala

Batch

Team N

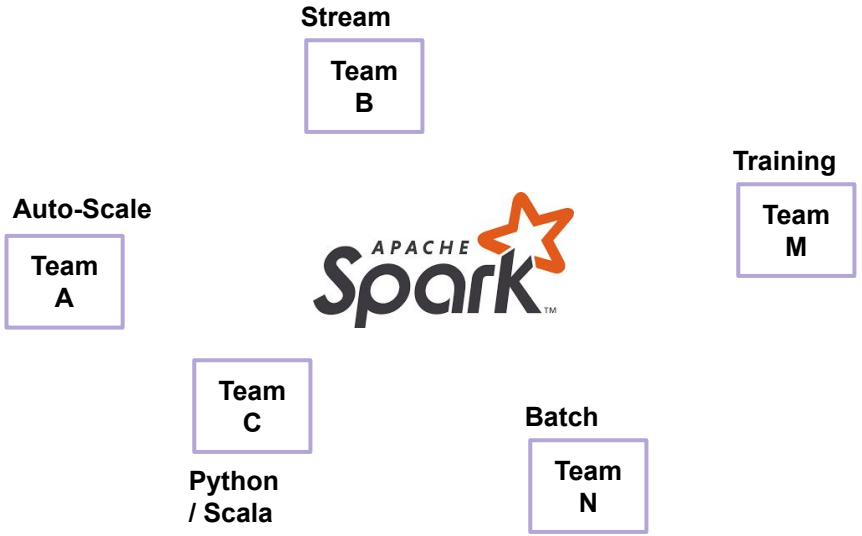
Data Sources

Too Many … Compute



- Cost control problem at Scale
- More Time To Production
- No Best Practices
- Duplication of work / Data
- Dependencies
- Inconsistent Environment
- No Community Knowledge
- Accidental Complexity

Spark as a Service



- Foundational piece of Zalando's Big Data Infrastructure
- GitOps Management, Decentralized Clusters
- Security / Compliance / CI-CD
- XX clusters/Jobs
- ~20 teams in production
- Thriving **#Databricks** community in Zalando

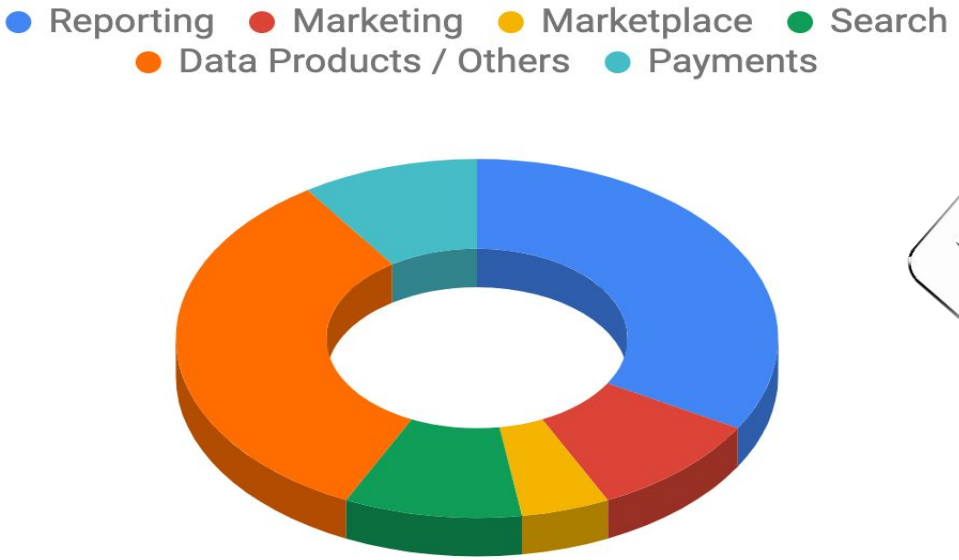
Spark as a Service

Migration Projects
ETLs | Data Preparation in Spark-S3

Spark as a Service

Others:
Structured Streams |
Traceability

Spectrum of use cases



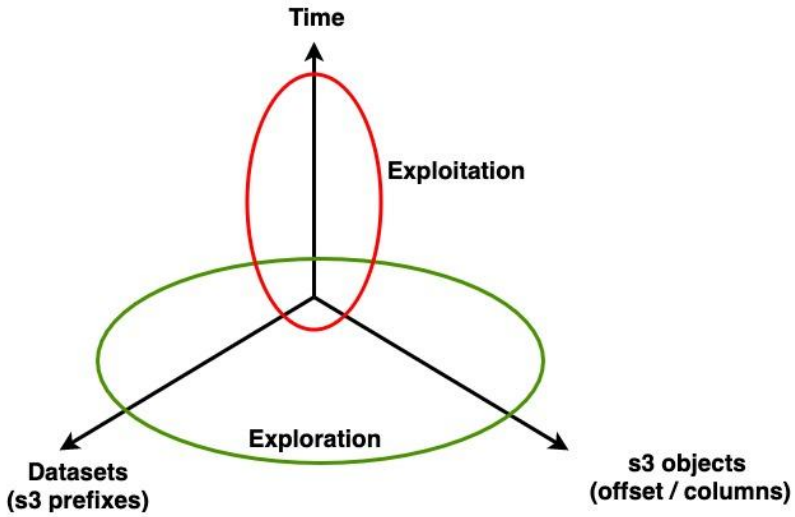
GDPR and Antitrust

Compliance with GDPR and antitrust laws

GDPR and Antitrust

Probe (pilot)

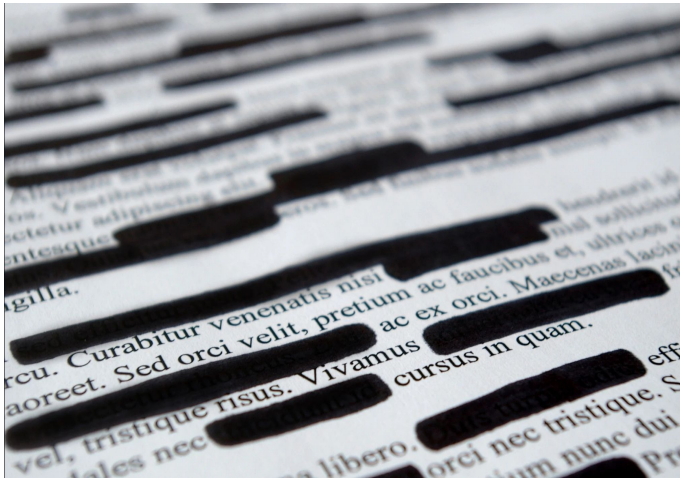
- Use marker event to create heat map of the data path.
- List of all datasets within the heat map.



GDPR and Antitrust

Pseudonymize/Remove

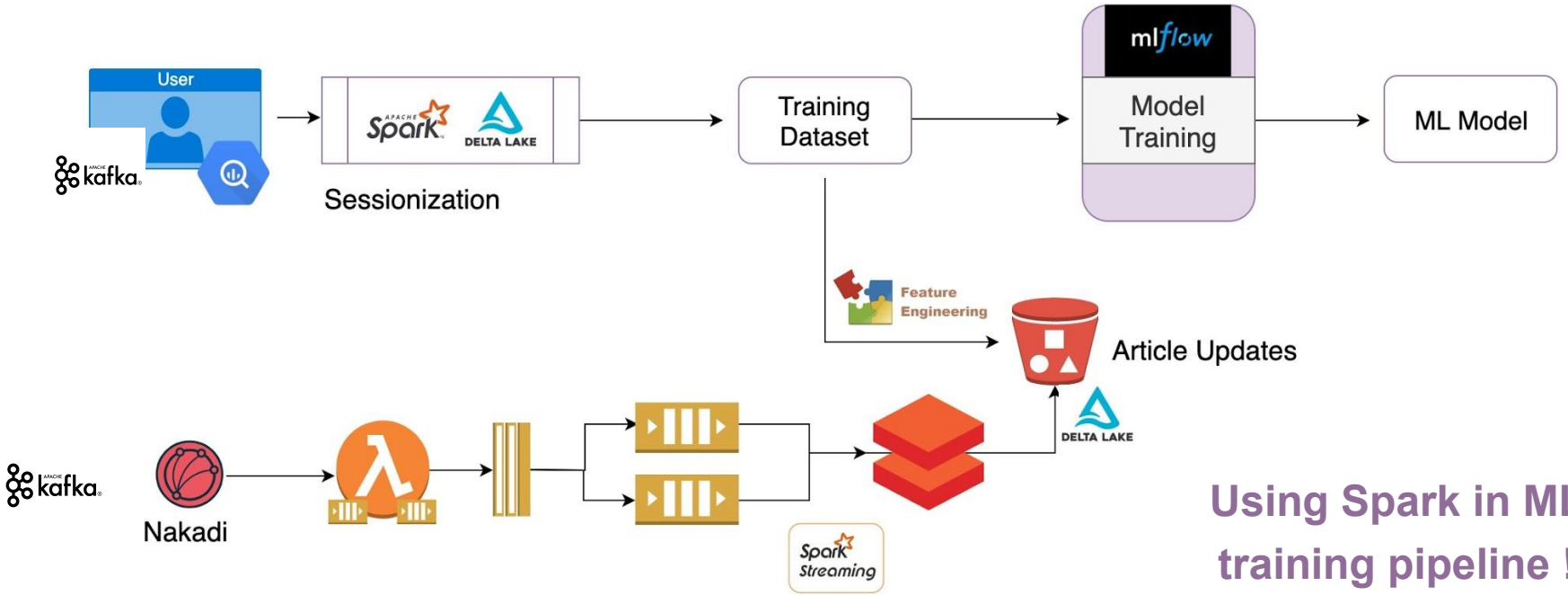
- Identifier based, on-demand, in-place record updater with field precision
- Great for semi-structured formats like JSON
- Use S3 Inventory + Streaming



Search & Ranking

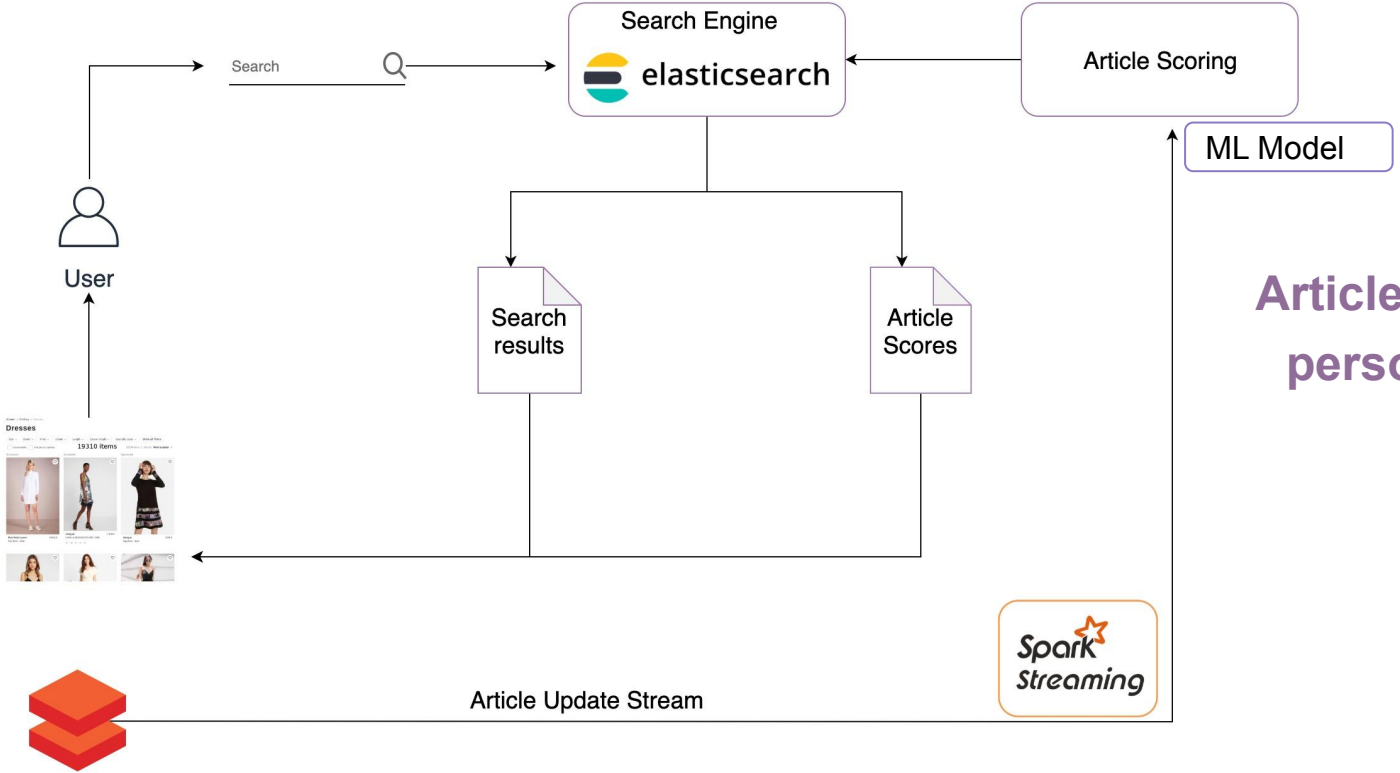
Personalized article ranking for relevance and user engagement.

Search & Ranking



Using Spark in ML training pipeline !

Search & Ranking



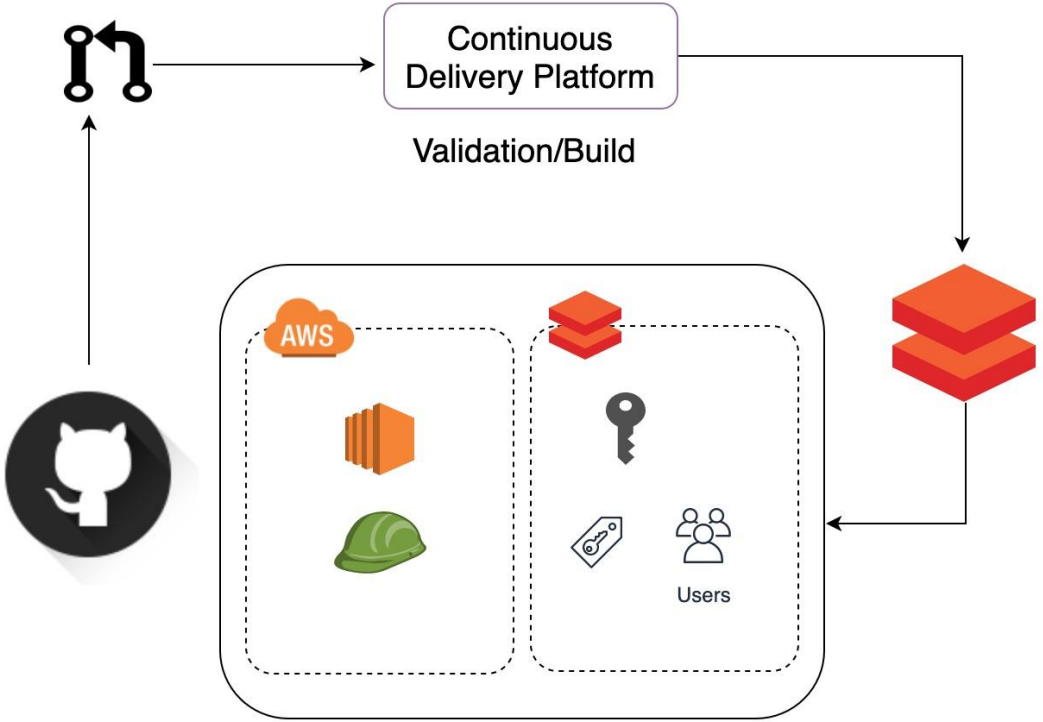
Article Scoring and personalization !

Others

- Sizing: Reducing return rates due to size and fit issues.
- Experimentation @Scale
- Merchant Analytics
- Marketing Services

First Impressions

- GitOps | Self Service



First Impressions

- Multi-Tiered support system
- Delta Adoption | But few readers outside Databricks ecosystem
- Communicating pricing downstream
- Exploding Usage is Good
- Fits all Size?

Thank you.

AI- Powered Retail Experience with Databricks

Akhil Dhingra
Saurav Verma

www.zalando.com
www.jobs.zalando.com/tech

SPARK+AI SUMMIT 2019

DON'T FORGET TO RATE AND REVIEW THE SESSIONS

SEARCH SPARK + AI SUMMIT

Download on the App Store

GET IT ON Google Play