Speeding up development with Apache Beam

By Constantin Scacun, Alexander Falca
Adobe Experience Platform

Why Adobe Experience Platform?

An open, intelligent, and real-time platform for customer experience
Adobe Experience Platform Pipeline

Pipeline Service Locations

13 REGIONS

25 PROD
44 CLUSTERS
OVERALL

310 B MSG/DAY
300 TB/DAY IN
900 TB/DAY OUT
How Teams Use the Data

• Filter messages (clients might not be subscribed to certain products)
• Validate messages against a schema
• Map/Transform messages
• Export the data out of Kafka
• Custom business logic

(Most use cases do not even need windowing or aggregations)
Boiler Plate Code

- Runtime/Deployment
- Observability
- Resiliency
- Fault Tolerance
- Optimizing producers, consumers
- Fine-tuning Apache Beam and Apache Flink
- Scaling on traffic bursts
Architecture

UI
- CRUD Definitions
- START/STOP/PAUSE/STATUS
- Live metrics

Runtime APIs
- CRUD Deployments
- Metrics and logs

Centralized Definitions Service

Artifactory
- Packaged User Functions
- Call services

Operator and Runners

Pipeline Kafka
- Publishing data
- Streaming data

Adobe Services
- Provided by our team
- Provided by other team teams inside the Org
- Provided by other Adobe teams
Execution

Download definition → Download custom code → Download secrets → Resolve sources

Resolve sources → Translate to Beam pipeline → Submit for execution
Questions?

https://www.linkedin.com/in/cscacun/
https://www.linkedin.com/in/alexfalca/