

Scalable Computing Project

Product recommendation system for an ecommerce site using large data transactions data set

Carlos Huerta | s3743071

Michael Yuja | s3766365

Ankit Mittal | s3683974

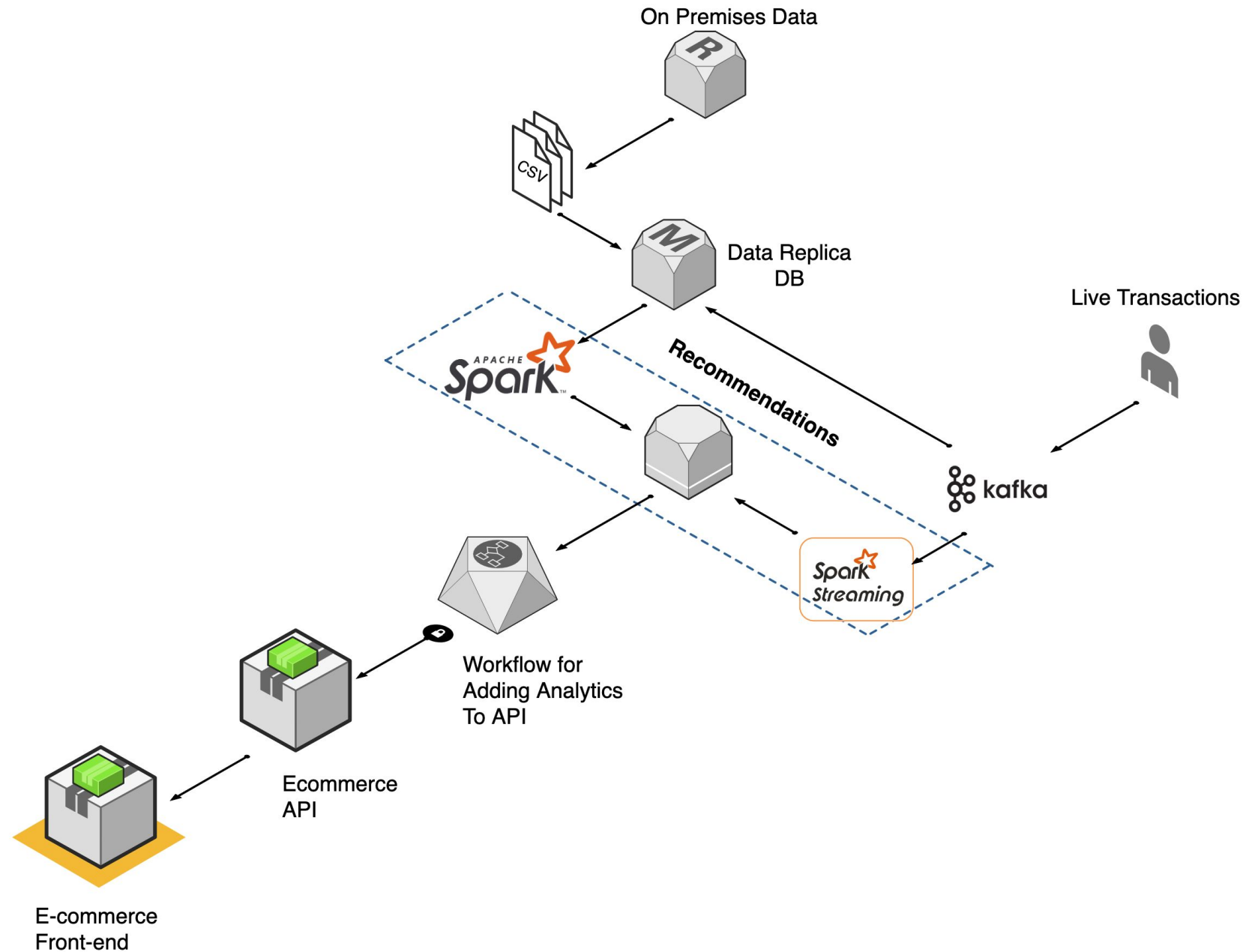
Feature/Workflow Proposal

- › For our project, we aim to implement a recommendation feature in an ecommerce website, www.agglobal.com, similar to Amazon's "People who bought this also bought that" using transactional purchase data accumulated since 2013.
- › There are three types of approaches for recommender systems - Collaborative filtering, Content-based filtering, and Hybrid filtering. For this feature, we will use collaborative filtering, which means we will use past purchases to relate products.
- › Our contact at the company will provide us with whatever data we need. They use SAP B1 for their business processes and SQL as a database, but we will get the data as CSV as we do not plan to integrate directly with them (to avoid security issues.)
- › Output will be a lookup table with recommendations for every product.

Technology stack

- Python
- Spark/PySpark
- Spark Streaming/Kafka
- Spark Standalone for Clusters
- Docker
- Cassandra

Proposed Architecture



Thank you for your attention