

The logo consists of a large red square. Inside this square is a smaller white square. The word "Astrafy" is written in a white, bold, sans-serif font, centered within the white square. Below it, the words "Analytics Solution" are written in a smaller, red, bold, sans-serif font, also centered within the white square.

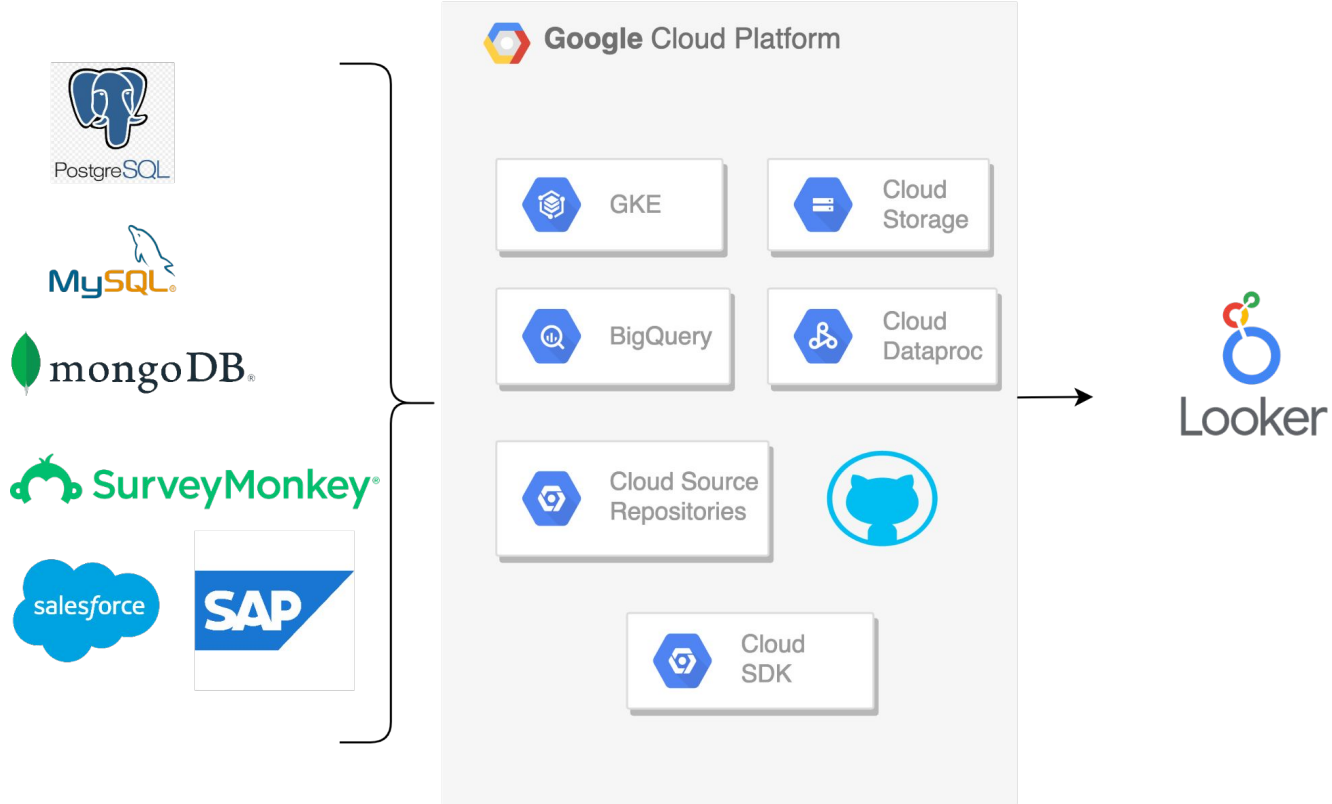
# Astrafy

**Analytics Solution**

# Motivation & Assumption

- To revamp the analytics backbone of “client A” on Google Cloud with a cloud agnostics architecture.
- Focus on data processing scalability for future with structured & non-structured data.
- Cost optimized pipelines with tolerable data latency in Intelligence Dashboard.
- Designed with continuous integration of data pipeline.

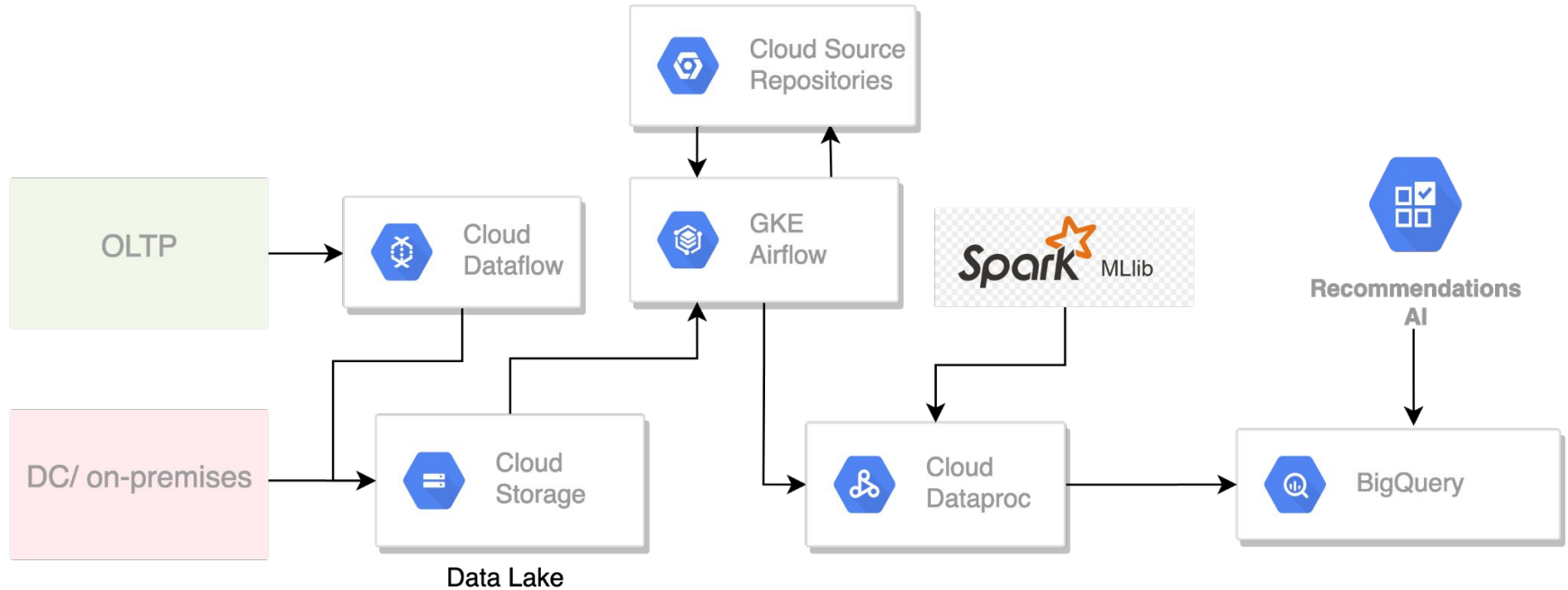
# GCP Involved



# Data Journey

- All the data is collected in cloud storage and separated based on domain-Customers, Products & Maisons.
- All data jobs are scheduled by Apache Airflow running on GKE cluster.
- Spark jobs are scheduled on Dataproc by Airflow.
- Curated data is pushed to DWH for to be consumed by looker dashboard.

# Data Journey-2



# Data Mesh for BI

- Separate Domain Data from the start.
- Transformation & cleaning are carried out by Apache Spark on Dataproc.
- ML predictions and recommendation are made on curated data.
- BigQuery manages data governance.

