Data Pipeline Architecture

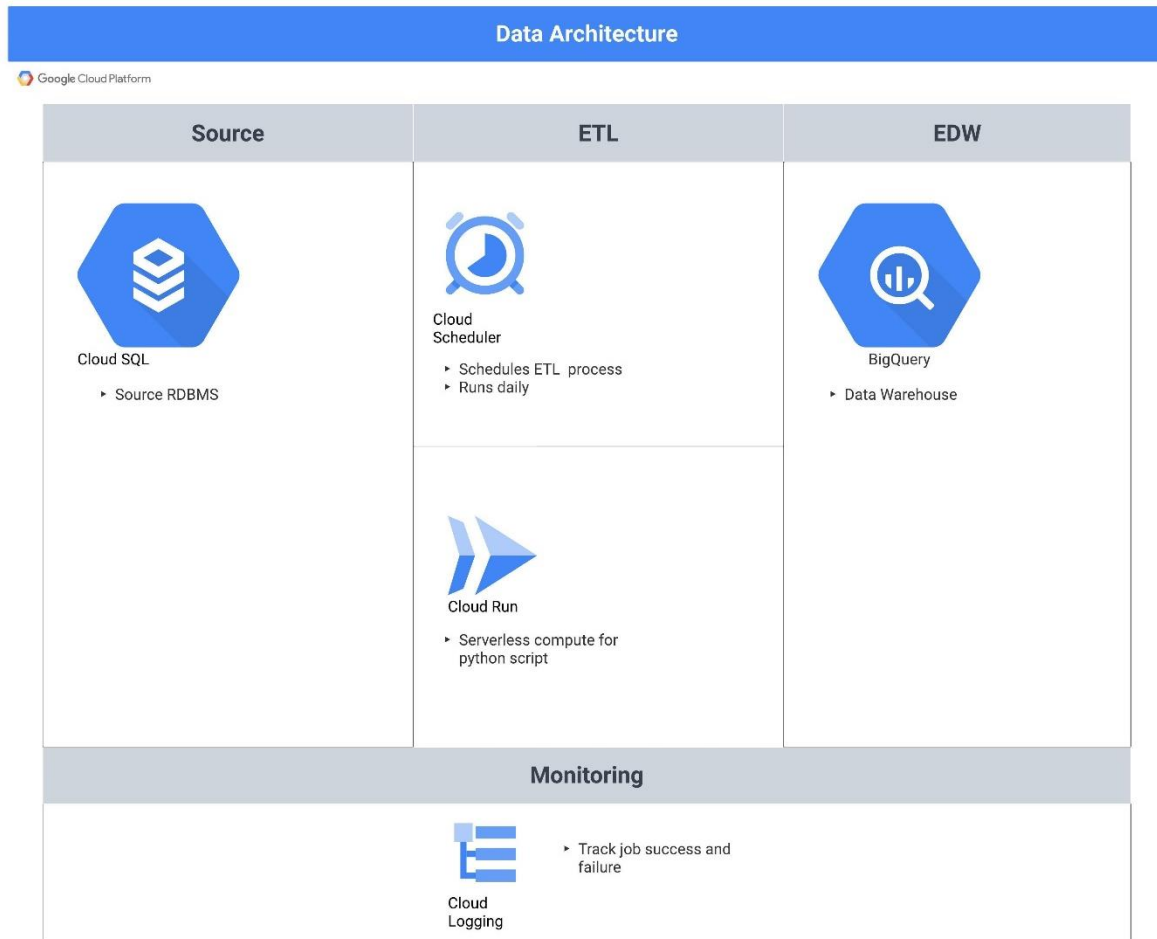**Author:** Obiechina Iyi

Figure 1: Data Pipeline Architecture Using Google Cloud Platform Services

Overview

The workflow is triggered daily using cloud scheduler, which invokes the python script deployed on Cloud Run. The script extracts the data from the Cloud SQL RDBMS, performs the transformation in memory and writes the resulting dataset to BigQuery directly using the BigQuery API. The entire process is monitored using Cloud Logging to ensure reliability.

The system is simple and minimizes cost. It is appropriate for the stated task because of the **moderate data volume** and **simple data transformation**.