Data Engineering Project

Automating ELT Data Pipelines with Airflow: Load & Transform data to BigQuery

Extract , Load & Transform 1 Million Records

Aspect	ETL	ELT
Transformation	Performed before loading	Performed after loading
Where It Happens	External tools (e.g., Dataflow, Dataprep)	Inside the data warehouse (e.g., BigQuery SQL)
Scalability	Limited by the ETL tool's capabilities	Highly scalable with BigQuery's processing power
Use Case	Legacy systems, data governance requirements	Cloud-native analytics, large datasets

ETL vs ELT

TechTrapture

Requirements

The Medical Research Team receives a global health statistics file containing disease data for all countries.
Each country's Health Minister should have access only to their respective country's medical data.
Additionally, they need the ability to analyze diseases for which no treatment or vaccination is currently available.

Challenges

- The data is currently provided as a single file containing over 1 million records for all countries.
- Due to the confidential nature of the data, it is not feasible to share the entire file with everyone.
- Analysing such a large CSV file to extract meaningful insights is complex and inefficient.
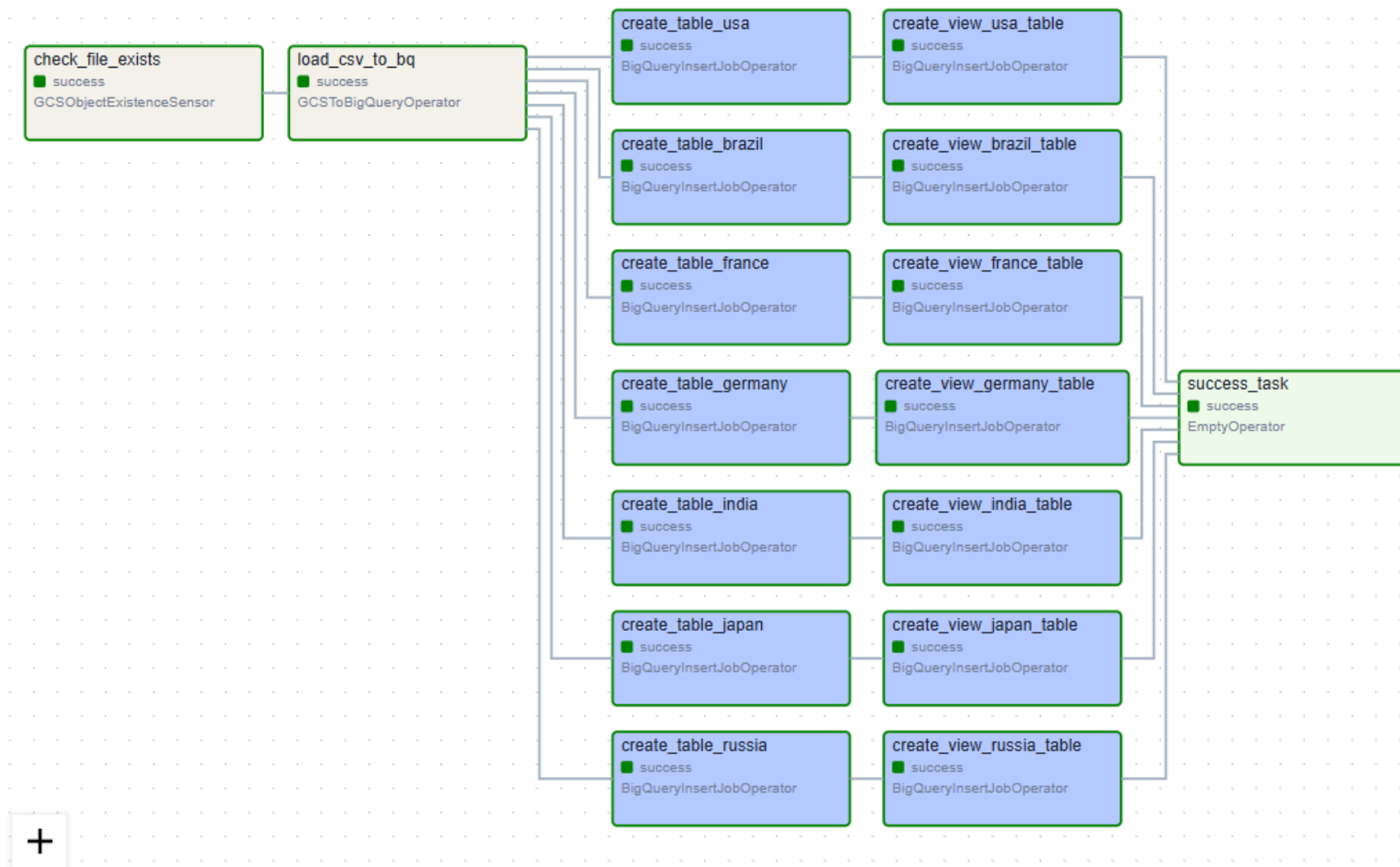-

Objective
Develop a robust data analytics solution to securely manage and filter this data, ensuring restricted access and enabling efficient analysis of diseases without available treatment or vaccination.

Architecture
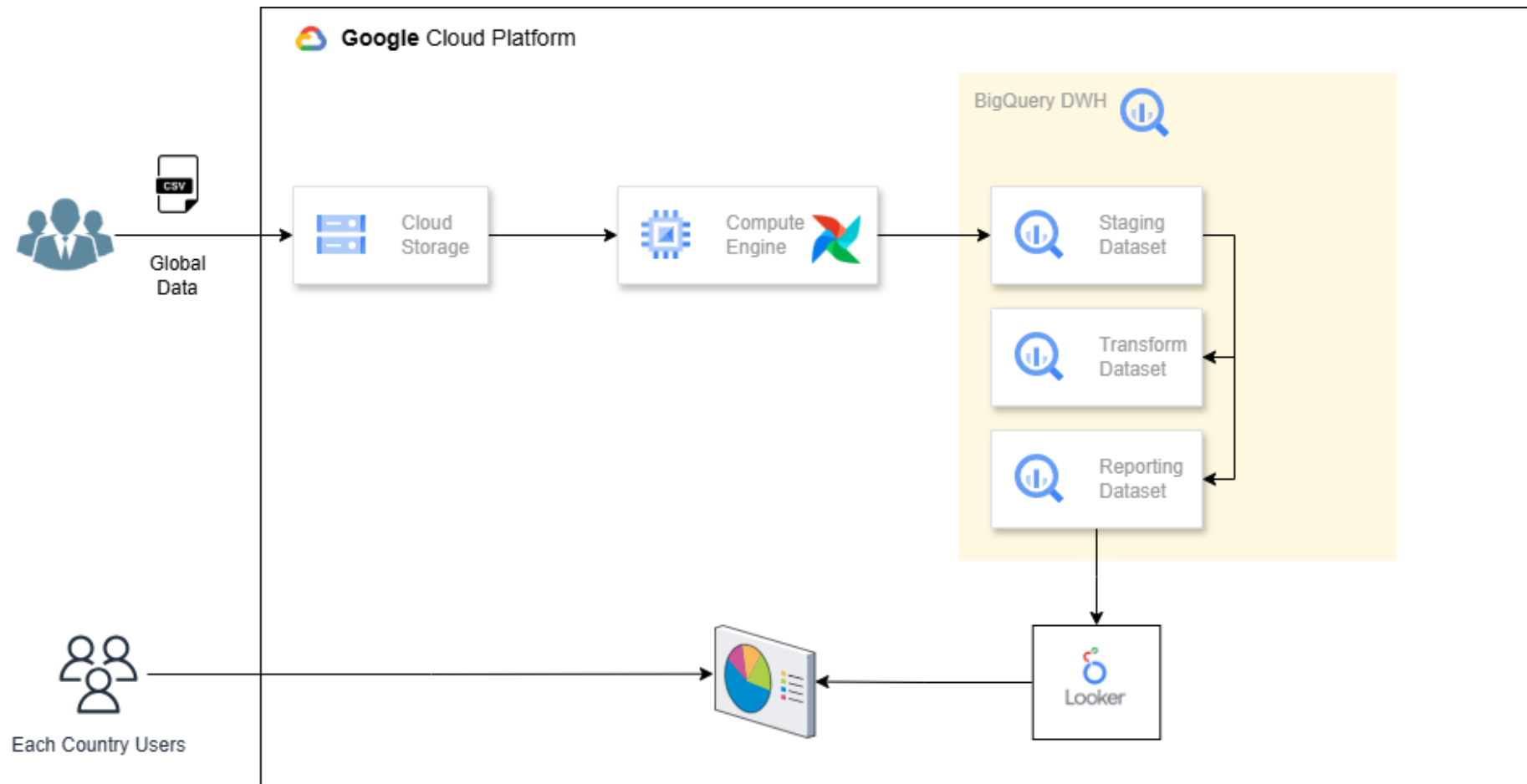
ELT Data Pipeline

TechTrapture

Looker Studio Dashboard

TechTrapture