***MIGRATION IN THE AMERICAS: ECONOMIC INEQUALITY AND FORCED DISPLACEMENT***

**PROJECT DOCUMENTATION**

WORKFLOW IN GOOGLE CLOUD PLATFORM

**CONSULTING TEAM:**

ALEXANDER RIOS - DATA ANALYST

CRISTIAN ANDRES - DATA SCIENTIST

RICARDO SORIA - DATA SCIENTIST

SANTIAGO ZAPATA - DATA ENGINEER - DATA ANALYST

**PROJECT MANAGEMENT:**

MAICO BERNAL - SCRUM MASTER

**MAY 2023**

**Justification for using Google Cloud Platform**

We decided to perform the ETL and Data Warehouse process on Google Cloud Platform because of the flexibility and wide range of technologies it offers, many of them oriented towards Big Data processes, Machine Learning, Data Analytics, among others.

Then it also allows easy connection between different tools including Looker Studio for visualization.

Apart from the mentioned advantages offered by Google Cloud Platform, performing our ETL cloud process also presents many other advantages such as:

**Remote accessibility:** it is accessible from any location both for the collaborative work of the development team and for the use of the proposed solution by the Migrants Without Borders foundation.

**Costs:** it allows a lower cost for the project as it does not require expensive hardware and software infrastructures.

**Security:** GCP provides data security and protection, as well as automatic backups and access control, which improves project security and reduces the risk of data loss.

The Google Cloud Platform tools we are going to use are the following:

**Cloud Storage:** is a cloud storage service integrated with the other tools.

**Dataflow:** facilitates the creation of data pipelines that can be integrated with other Google Cloud and third-party services.

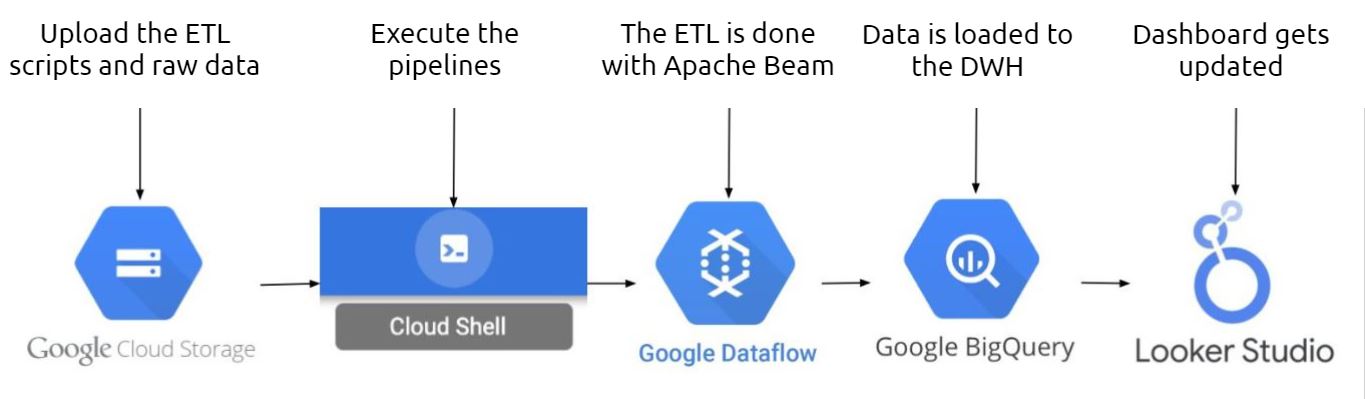
**BigQuery:** is a data analysis service that allows large data sets to be stored and queried with great speed and ease.

**Cloud Functions:** allows code to be executed in response to specific events (triggers), such as file uploads to Cloud Storage.

**Looker Studio:** allows reports and dashboards to be created from data hosted in BigQuery.

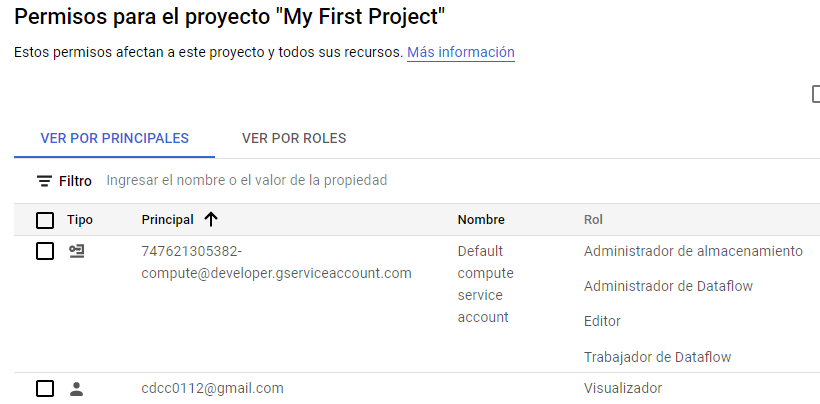
**Vertex AI:** is a machine learning service that enables the creation, training and implementation of Machine Learning models.

**Workflow in Google Cloud Platform**

****

**Step by step of the process to be performed in Google Cloud Platform.**

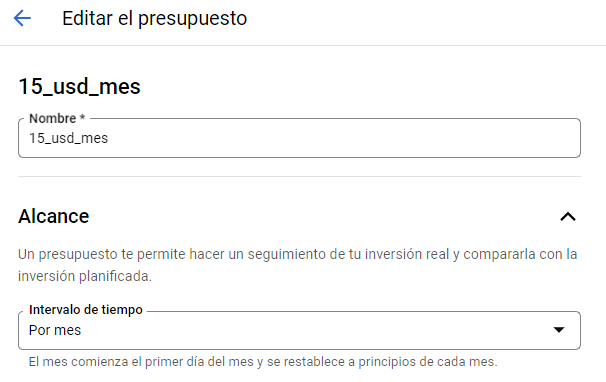
Roles and permissions are created for the applications and for the work team:



The APIs necessary for the correct operation of the various tools are enabled:



The budget for the project is set up:



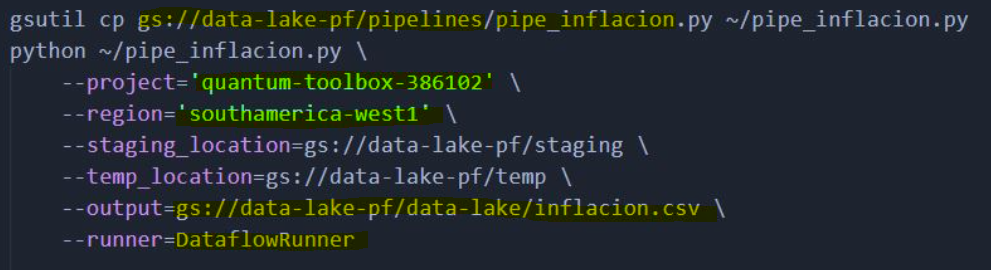
A bucket is created and configured in Cloud Storage:



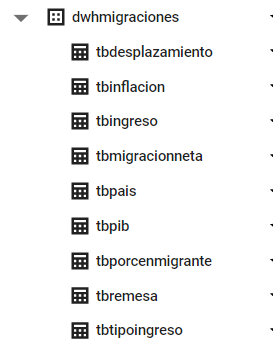
The pipelines to be executed later are stored in the bucket:



The pipelines are executed through the Cloud Shell terminal by defining parameters such as location, project id, temporary folders, output location and "runner" in this case dataflow.



After running the pipelines, the processed files are automatically uploaded to the data warehouse.



Data Warehouse data can be taken from Looker Studio and other tools:

