Assignment Shape

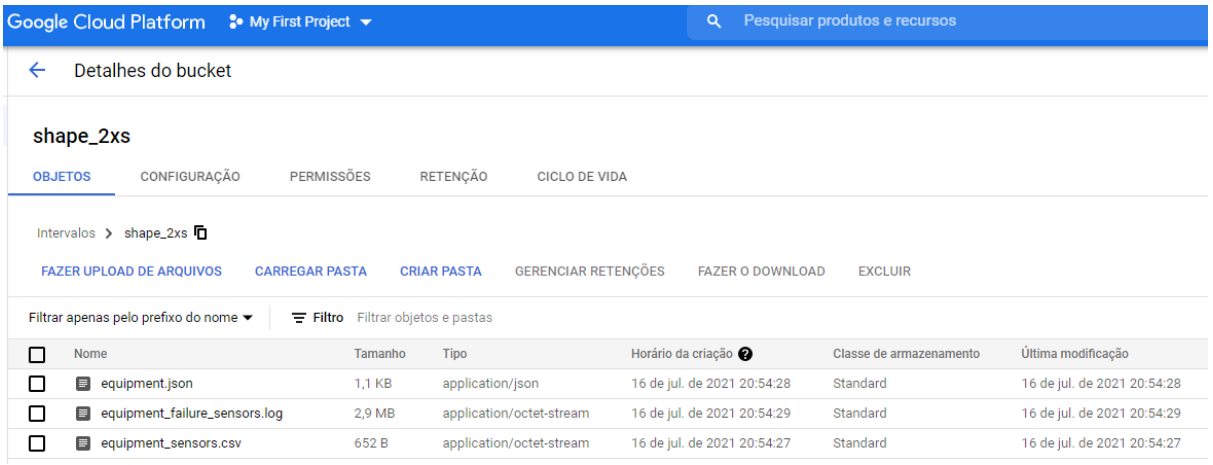
I use Google Cloud Platform (GCP) to store the data on the cloud and I use Python to make the ETL process.

Now I will write the steps of the process for better comprehension:

Steps:

1) Put the archives in a Bucket on Google Cloud Storage (GCP)
2) Construct the code in Python, to collect this data and make small transformation for better usage in BigQuery (GCP) database
3) Use the BigQuery to make the queries and get to the answers of the questions

Step 1: I put the archives in a Bucket on Google Cloud Storage, because it is a good practice to put the data of work in the Cloud.



Step 2: Write the code of extraction and transformation of the data in Python. The code is available in attachment (shape_assignment\main.py) and in my personal GitHub repository (for a good practice of version control). Doing this way, I will have all the information that I need in structured data for analysis. Structured data saved in BigQuery (GCP)

Step 3: For a better way to view the results, I use SQL language in the BigQuery environment to show the results and solve the problems / questions. All the queries done are available in attachment (shape_assignment\queries.txt) and in my personal GitHub repository. In this archive we have the answers for the questions.

Perceptions / Conclusions

1) I think it's a good idea to analyze this data in a temporal form, to see the evolution of errors through the time.
2) It's also a good idea to analyze other relevant information, such as temperature and vibration and its correlation with the amount of failures that happened

Thank you so much