

## Assignment Shape

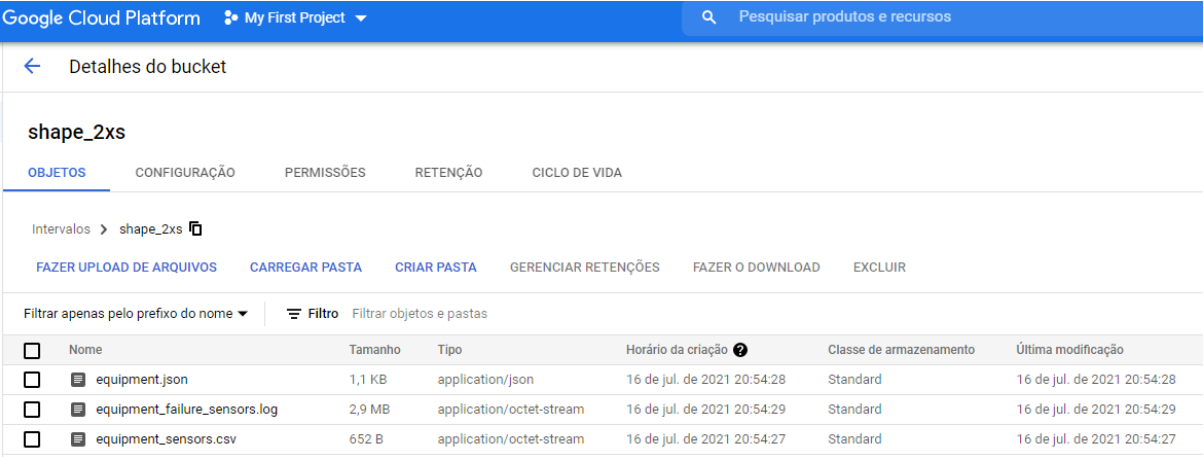
I use Google Cloud Platform (GCP) to store the data on the cloud and I use Python to make the ETL process.

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.



The screenshot shows the Google Cloud Platform interface for a bucket named 'shape\_2xs'. The bucket is located in 'My First Project'. The interface includes tabs for 'OBJETOS', 'CONFIGURAÇÃO', 'PERMISSÕES', 'RETENÇÃO', and 'CICLO DE VIDA'. The 'OBJETOS' tab is selected, showing a list of objects. The objects are filtered by name prefix. The table lists three objects: 'equipment.json' (1,1 KB, application/json), 'equipment\_failure\_sensors.log' (2,9 MB, application/octet-stream), and 'equipment\_sensors.csv' (652 B, application/octet-stream). All objects were created on 16 de jul. de 2021 at 20:54:28, 20:54:29, and 20:54:27 respectively, and are stored in the 'Standard' storage class.

Nome	Tamanho	Tipo	Horário da criação	Classe de armazenamento	Última modificação
equipment.json	1,1 KB	application/json	16 de jul. de 2021 20:54:28	Standard	16 de jul. de 2021 20:54:28
equipment_failure_sensors.log	2,9 MB	application/octet-stream	16 de jul. de 2021 20:54:29	Standard	16 de jul. de 2021 20:54:29
equipment_sensors.csv	652 B	application/octet-stream	16 de jul. de 2021 20:54:27	Standard	16 de jul. de 2021 20:54:27

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