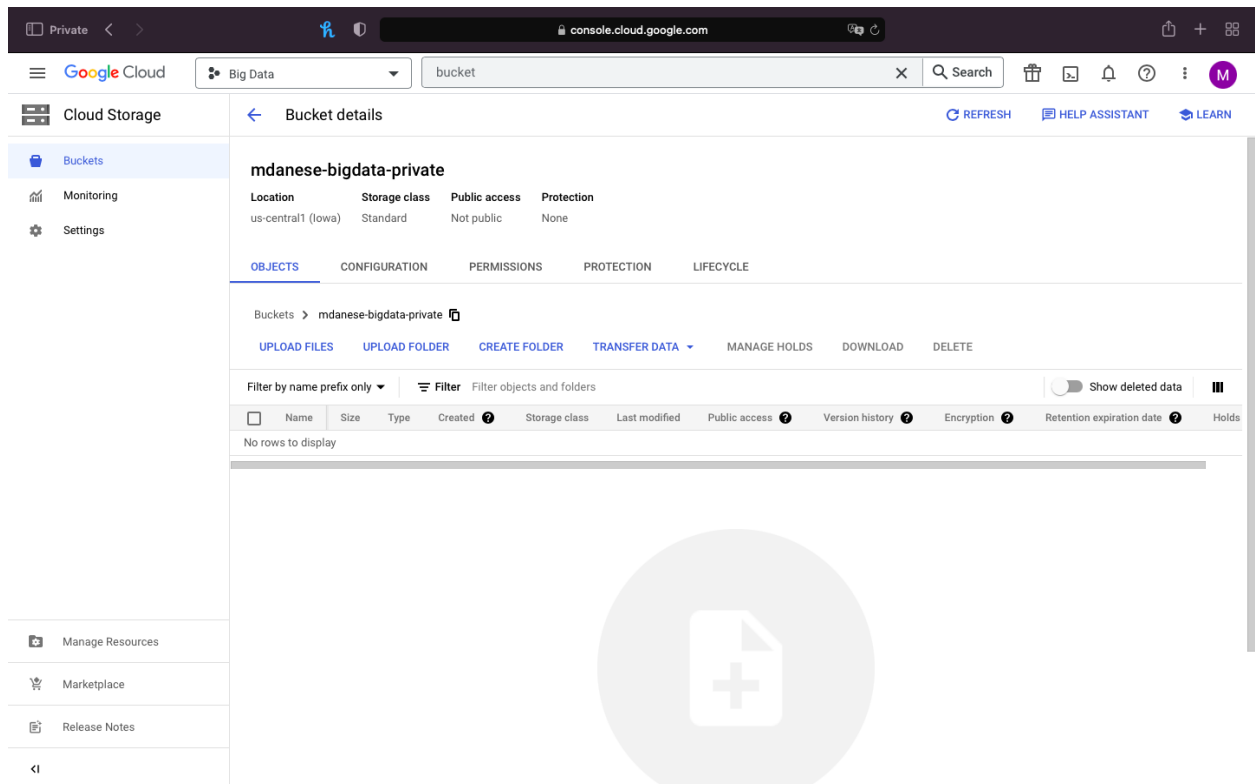Mattia Danese
CS 119 - Big Data
Professor Singh
Quiz 2: The Opioid Files with BigQuery

Question 1

I activated my Google Account, made a project called `Big Data` and made a storage bucket called `mdanese-bigdata-private`.

<u>Question 2</u>

I populated the file `arcos_2007.tsv` with 5000 random lines that all have transaction dates in 2007.

The commands I ran:
- `zcat /comp/119/arcos_all_washpost.tsv.gz | awk -F '\t' '($31 % 10000) == 2007' | shuf -n 5000 > arcos_2007.tsv`

Question 3

I uploaded `arcos_2007.tsv` to my `mdanese-bigdata-private` storage bucket and checked that it was uploaded successfully.

```
vm-hw02{mdanes01}67: ~/google-cloud-sdk/bin/gcloud storage ls gs://mdanese-bigdata-private/
gs://mdanese-bigdata-private/arcos_2007.tsv
vm-hw02{mdanes01}68:
```

The commands I ran:
- `~/google-cloud-sdk/bin/gcloud storage cp ./arcos_2007.tsv gs://mdanese-bigdata-private/`
- `~/google-cloud-sdk/bin/gcloud storage ls gs://mdanese-bigdata-private/`

Question 4

I made a dataset called `alice_2022` and table called `arcos_2007` with the correct schema on the BigQuery Console.

Question 5

I uploaded the `arcos_2007.tsv` file into the `arcos_2007` table using the BigQuery
PowerShell

```
mattia_danese@cloudshell:~ (studied-brace-363320)$ gcloud config set project studied-brace-363320


Updated property [core/project].
mattia_danese@cloudshell:~ (studied-brace-363320)$ bq load --field_delimiter=tab --source_format=CSV studied-brace-363
320:alice_2022.arcos_2007 gs://mdanese-bigdata-private/arcos_2007.tsv
Waiting on bqjob_r238b3e04e9872f46_0000018386752b56_1 ... (1s) Current status: DONE
mattia_danese@cloudshell:~ (studied-brace-363320)$ ▊
```

The commands I ran:
- `gcloud config set project studied-brace-363320`
- `bq load --field_delimiter=tab --source_format=CSV`
  `studied-brace-363320:alice_2022.arcos_2007`
  `gs://mdanese-bigdata-private/arcos_2007.tsv`

Question 6

The names of all the drugs that were delivered during 2007 are:

```
[{
  "DRUG_NAME": "HYDROCODONE"
}, {
  "DRUG_NAME": "OXYCODONE"
}]
```

The queries I ran on the `arcos_2007` table using BigQuery are:

- ```
  SELECT DISTINCT DRUG_NAME
  FROM "alice_2022.arcos_2007"
  ```