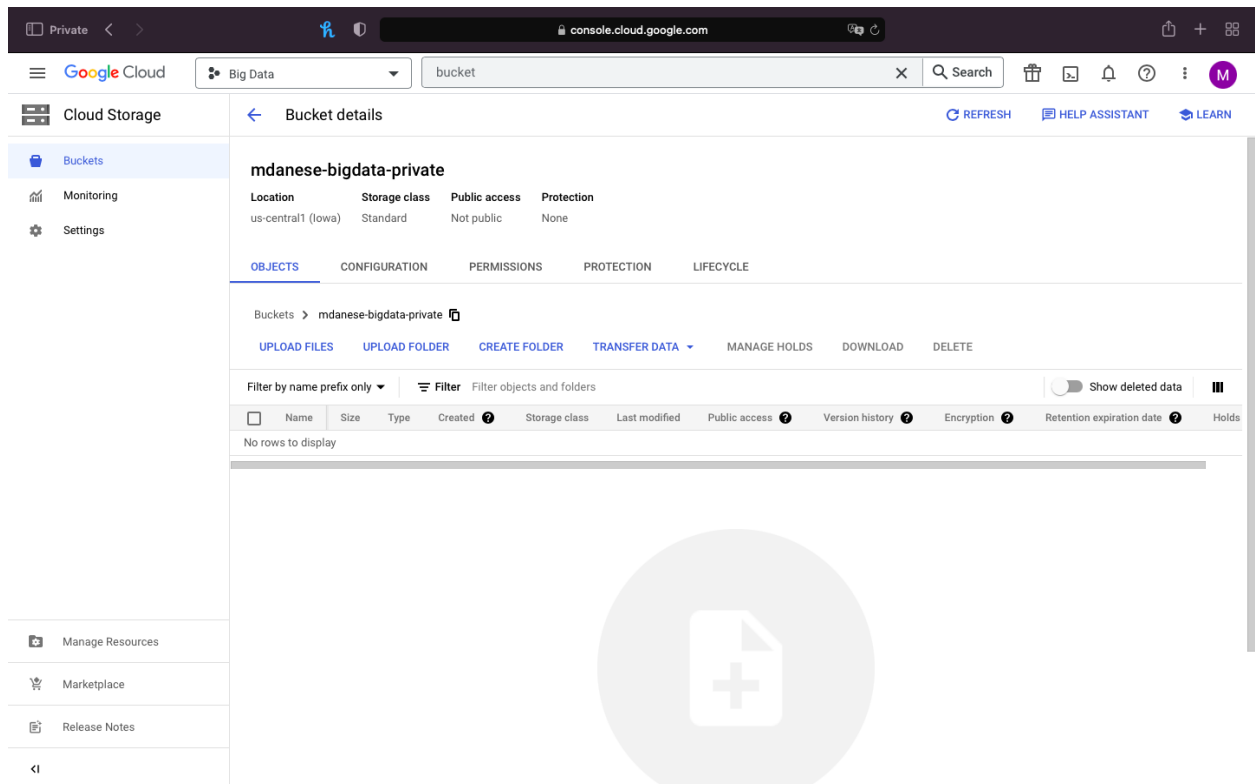


Mattia Danese  
CS 119 - Big Data  
Professor Singh  
Quiz 2: The Opioid Files with BigQuery

### Question 1

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



## Question 2

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.

The screenshot shows the BigQuery Console interface. On the left, the Explorer pane shows a project named 'studied-brace-363320' with two datasets: 'alice\_2022' and 'arcos\_2007'. The 'arcos\_2007' dataset is selected. The top navigation bar shows 'arcos\_2007' with tabs for SCHEMA, DETAILS, and PREVIEW. The SCHEMA tab is active, displaying a table with 20 columns. Each column has a checkbox, a field name, a type (all are STRING), and a mode (all are NULLABLE). The columns are: REPORTER\_DEA\_NO, REPORTER\_BUS\_ACT, REPORTER\_NAME, REPORTER\_ADDL\_CO\_INFO, REPORTER\_ADDRESS1, REPORTER\_ADDRESS2, REPORTER\_CITY, REPORTER\_STATE, REPORTER\_ZIP, REPORTER\_COUNTY, BUYER\_DEA\_NO, BUYER\_BUS\_ACT, BUYER\_NAME, BUYER\_ADDL\_CO\_INFO, BUYER\_ADDRESS1, BUYER\_ADDRESS2, BUYER\_CITY, BUYER\_STATE, BUYER\_ZIP, and BUYER\_COUNTY. At the bottom, there are buttons for 'EDIT SCHEMA' and 'VIEW ROW ACCESS POLICIES'.

	Field name	Type	Mode	Collation	Default Value	Policy Tags	Description
<input type="checkbox"/>	<a href="#">REPORTER_DEA_NO</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">REPORTER_BUS_ACT</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">REPORTER_NAME</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">REPORTER_ADDL_CO_INFO</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">REPORTER_ADDRESS1</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">REPORTER_ADDRESS2</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">REPORTER_CITY</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">REPORTER_STATE</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">REPORTER_ZIP</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">REPORTER_COUNTY</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">BUYER_DEA_NO</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">BUYER_BUS_ACT</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">BUYER_NAME</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">BUYER_ADDL_CO_INFO</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">BUYER_ADDRESS1</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">BUYER_ADDRESS2</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">BUYER_CITY</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">BUYER_STATE</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">BUYER_ZIP</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">BUYER_COUNTY</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">TRANSACTION_CODE</a>	STRING	NULLABLE				
<input type="checkbox"/>	<a href="#">DRUG_CODE</a>	STRING	NULLABLE				

### 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-363320: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"
```