**What is GCP:**

Google Cloud Platform (GCP) is a suite of cloud computing services provided by Google. It allows you to use Google's infrastructure to build, deploy, and scale applications, websites, and services. GCP offers a range of services, including computing power, storage options, data analytics, machine learning tools, and more, all accessible over the internet. This means you can run your projects on Google's powerful hardware without needing to maintain your own servers.

**Create GCP Project:**

**Setting GCP CLI and Set Project:**

Commands:

* *gcloud auth login*
* *gcloud projects list*
* *gcloud config set project graphic-adapter-425011-f4*
* *gcloud projects create PROJECT\_ID --name="PROJECT\_NAME"*
* *gcloud projects delete PROJECT\_ID*

**GCP Bucket:**

Commands:

* **List Buckets:** *gsutil ls*
* **Create Bucket:** *gsutil mb gs://your-bucket-name*
* **Delete Bucket:** *gsutil rb gs://your-bucket-name*
* **Upload file to Bucket:** *gsutil cp local-file.txt gs://your-bucket-name*
* **Download file from Bucket:** *gsutil cp gs://your-bucket-name/remote-file.txt local-file.txt*
* **Copy files within Cloud Storage:** *gsutil cp gs://your-bucket-name/source-file.txt gs://your-bucket-name/destination-file.txt*
* **List Files: List files in a bucket or directory:** *gsutil ls gs://your-bucket-name*
* **Delete Files: Delete a file from a bucket:** *gsutil rm gs://your-bucket-name/remote-file.txt*
* **Set Bucket Permissions: Apply a predefined ACL to a bucket:** *gsutil acl set private gs://your-bucket-name*
* **Set File Permissions: Apply a predefined ACL to a file:** *gsutil acl set public-read gs://your-bucket-name/remote-file.txt*
* **View Bucket or File Permissions: Display ACL for a bucket or file:** 
  + *gsutil acl get gs://your-bucket-name*
  + *gsutil acl get gs://your-bucket-name/remote-file.txt*
* **Sync Local Directory to Bucket: Synchronize a local directory with a bucket:**
  + *gsutil rsync -r local-directory gs://your-bucket-name*
* **Set Lifecycle Management Policy: Apply a lifecycle management policy to a bucket:**
  + *gsutil lifecycle set policy.json gs://your-bucket-name*
* **Enable Object Versioning: Enable versioning for a bucket:**
  + *gsutil versioning set on gs://your-bucket-name*
* **Disable Object Versioning: Disable versioning for a bucket:**
  + *gsutil versioning set off gs://your-bucket-name*
* **Check Bucket Usage: Get information about bucket usage:**
  + *gsutil du -s gs://your-bucket-name*
* **View Object Metadata: Display metadata for an object:**
  + *gsutil stat gs://your-bucket-name/remote-file.txt*

**Big Query:**

Commands

* **Listing Datasets:**
  + *bq ls*
* **List all the dbs in specific project:**
  + *bq ls --project\_id=PROJECT\_ID*
* **Creating a Dataset:**
  + *bq mk my\_dataset*
* **Deleting a Dataset:**
  + *bq rm -r -d my\_dataset (The -r flag ensures that all tables in the dataset are removed.*
* **Listing Tables in a Dataset:**
  + *bq ls my\_dataset*
* **Creating a Table:**
  + *bq mk --table my\_dataset.my\_table column1:STRING,column2:INTEGER*
* **Deleting a Table:**
  + *bq rm -t my\_dataset.my\_table*
* **Inserting Data into a Table:**
  + *bq load --source\_format=CSV my\_dataset.my\_table path\_to\_csv\_file*
* **Querying Data:**
  + *bq query "SELECT column1, column2 FROM my\_dataset.my\_table WHERE column2 > 10"*
* **Exporting Query Results:**
  + *bq query --format=csv --use\_legacy\_sql=false "SELECT column1, column2 FROM my\_dataset.my\_table WHERE column2 > 10" > results.csv*
* **Creating a Partitioned Table:**
  + *bq mk --table --time\_partitioning\_field column1 my\_dataset.my\_partitioned\_table column1:TIMESTAMP,column2:STRING*
* **Loading Data from Google Cloud Storage:**
  + *bq load --source\_format=CSV my\_dataset.my\_table gs://my\_bucket/path\_to\_csv\_file*
* **Queries:**
  + **Show DBS:** *SELECT schema\_name FROM INFORMATION\_SCHEMA.SCHEMATA;*
  + **Show tables:** *SELECT table\_name FROM `my\_dataset.INFORMATION\_SCHEMA.TABLES`;*
  + **Show table schema:** *SELECT column\_name, data\_type FROM `my\_dataset.INFORMATION\_SCHEMA.COLUMNS` WHERE table\_name = 'my\_table';*
  + **DB statistics:** *SELECT table\_name, row\_count, size\_bytes FROM `my\_dataset.INFORMATION\_SCHEMA.TABLES`;*
  + **Create non-partitioned table from CSV file:**

*CREATE OR REPLACE EXTERNAL TABLE my\_dataset.my\_external\_table*

*PARTITION BY DATE(transaction\_date)*

*OPTIONS (*

*format = 'CSV',*

*uris = ['gs://my\_bucket/path/to/file.csv'],*

*skip\_leading\_rows = 1*

*);*

* + **Creating table from parquet file:**

*CREATE OR REPLACE EXTERNAL TABLE my\_dataset.my\_external\_parquet\_table*

*OPTIONS (*

*format = 'PARQUET',*

*uris = ['gs://my\_bucket/path/to/file.parquet']*

*);*

* + **Update does not support but can use create or replace**
* **Query Optimization?**
* **Scheduling Queries**
* **Big Query Machine Learning**
  + **Classification**
    - **Create Model**

CREATE OR REPLACE MODEL `my\_dataset.purchase\_classification\_model`

OPTIONS(

model\_type='logistic\_reg',

input\_label\_cols=['purchase']

) AS

SELECT

age,

income,

browsing\_time,

purchase

FROM

`my\_dataset.customer\_data`;

* + - **Evaluate Model**

SELECT

\*

FROM

ML.EVALUATE(MODEL `my\_dataset.purchase\_classification\_model`)

* + - **Predict Using Model**

SELECT

customer\_id,

predicted\_purchase

FROM

ML.PREDICT(MODEL `my\_dataset.purchase\_classification\_model`,

(

SELECT

customer\_id,

age,

income,

browsing\_time

FROM

`my\_dataset.new\_customer\_data`

));

* + **Regression**
    - **Create Model**

CREATE OR REPLACE MODEL `my\_dataset.house\_price\_regression\_model`

OPTIONS(

model\_type='linear\_reg',

input\_label\_cols=['price']

) AS

SELECT

square\_footage,

num\_bedrooms,

num\_bathrooms,

location,

price

FROM

`my\_dataset.house\_data`;

* + - **Evaluate**

SELECT

\*

FROM

ML.EVALUATE(MODEL `my\_dataset.house\_price\_regression\_model`)

* + - **Predict**
* SELECT
* house\_id,
* predicted\_price
* FROM
* ML.PREDICT(MODEL `my\_dataset.house\_price\_regression\_model`,
* (
* SELECT
* house\_id,
* square\_footage,
* num\_bedrooms,
* num\_bathrooms,
* location
* FROM
* `my\_dataset.new\_house\_data`
* ));
  + **Forecasting**
  + **Clustering**
  + **List all the models**
* SELECT
* model\_id,
* model\_type,
* creation\_time,
* last\_modified\_time
* FROM
* ML.MODELS();
  + **Deleting Model**
* DROP MODEL `my\_dataset.my\_model`;
* **Other features:**
  + Dataform: It helps to orchestrate and manage data pipelines, transformations, and workflows in a more structured and reproducible manner.
  + Orchestration
  + SQL translation
  + Monitoring

**Container Registry:**

Google Cloud Container Registry (GCR) is a fully managed Docker container registry that makes it easy for developers to store, manage, and deploy Docker container images. It integrates seamlessly with Google Cloud Platform (GCP) services and tools, providing a secure and scalable solution for storing and distributing Docker images.

**Artifact Registry:**