

## Assignment 1, Cloud Computing

### Exercise 1: Understanding Cloud Computing Models

#### Questions:

- *What are the main differences between IaaS, PaaS, and SaaS?*

#### Infrastructure as a Service (IaaS)

IaaS gives users access to computing resources like servers, storage, and virtualization. The provider maintains the hardware, while customers are responsible for managing their software, such as OS, databases and apps.

#### Platform as a Service (PaaS)

PaaS provides the tools and environment needed to develop, test, and run applications. The cloud provider takes care of the infrastructure, including hardware and software.

#### Software as a Service (SaaS)

SaaS offers complete software applications that customers can access through the internet. These applications are fully managed by the cloud provider.

	Infrastructure as a Service (IaaS)	Platform as a Service (PaaS)	Software as a Service (SaaS)
Control	High control over Infrastructure. Customers manage OS, and networking infrastructure.	Less control over infrastructure, focus on development.	Minimal control over infrastructure. Provider manages everything.
Flexibility	High flexibility to configure network settings, storage and so on	Average flexibility, more focus on development	Low flexibility, customization is limited
Use Cases	Virtual machines	Testing and Deployment of Apps, apps hosting	CRM systems

**IaaS** provides the most control, taking care of things like the OS and apps, while the cloud provider manages the physical servers and hardware. **PaaS** simplifies things by handling most of the infrastructure, so customers can focus on building and running their apps without worrying about the backend. In **SaaS** the provider takes care of everything from infrastructure to software.

- *Which GCP services fall under each of these models?*

IaaS - Cloud Storage, Compute Engine

PaaS - Cloud SQL, App Engine

SaaS - BigQuery, Workspace

- *Provide a real-world example where each cloud service model might be the most appropriate choice.*

### **IaaS - Google Cloud Storage**

For example, we have a media company. So we need to store and retrieve a large amount of unstructured data, such as photos and videos.

And Google Cloud Storage is a fine scalable decision, as the data grows everyday, and would be good for media files, backups, and logs. It gives users control over how data is stored and organized, can be accessed from everywhere and supports various storage classes.

### **PaaS - Google Cloud SQL**

For instance, we have an e-commerce app and we need a relational db to store data on orders, user information and so on.

Google Cloud SQL is a satisfactory option for managing databases in the cloud because it's easy to use, can grow with your needs, and works well with other Google Cloud services. Also, it supports databases like MySQL, PostgreSQL, making it flexible for different apps. It means businesses can move their current databases to the cloud or create new ones using the database they prefer.

### **SaaS - Google BigQuery**

As an example, a healthcare company wants to analyze patient data from multiple sources, such as electronic devices, lab results, and insurance claims. And they need to identify trends in patient outcomes and improve treatment.

BigQuery can be a good decision because of serverless architecture, so the team can be more focused on analysis rather than system maintenance. Also, it is a good option for running complex queries in order to analyze data. And can be easily integrated with Visualization tools.

## **Exercise 2: Exploring Google Cloud Platform's Core Services**

### **Questions:**

- *What is the primary use case of Compute Engine?*

Google Compute Engine is mostly used for running VMs in the cloud. It is good for hosting websites, running apps, or performing computing tasks like machine learning.

- *How does Google Kubernetes Engine (GKE) simplify the management of containerized applications?*

For example, It automatically scales an app when traffic increases. So briefly it does lots of work on its own, like monitoring, setting up, and scaling.

- *What advantages does Cloud Storage offer for data management?*

Diversity of data types, starting from photos ending with datasets. Also it is scalable and reliable. Moreover, replication of data is available.

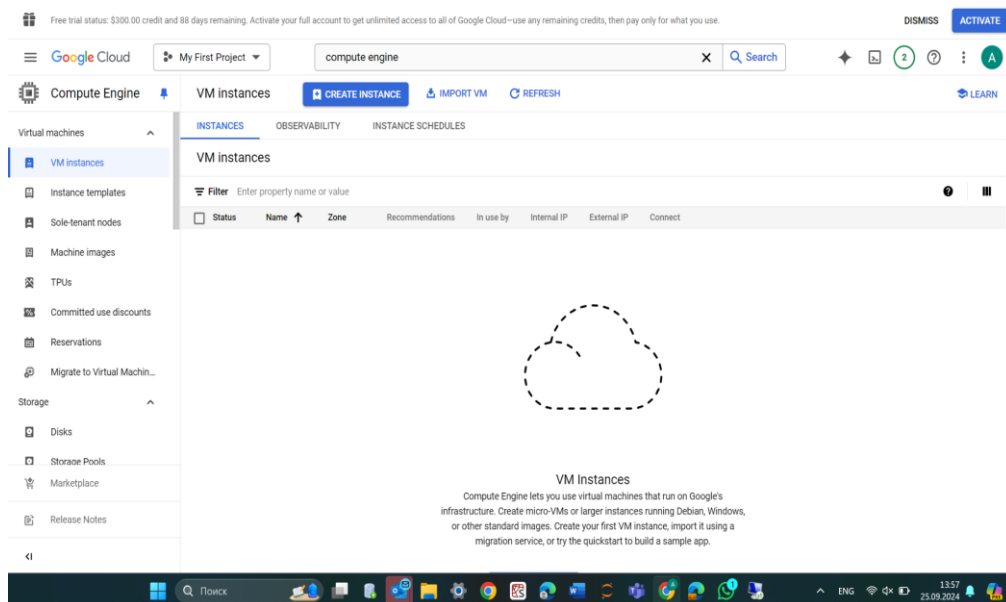
- *Why would a business choose BigQuery for their data analysis needs?*

It is a good option for the business when a large amount of data is needed to be analyzed. Complicated queries are run quickly, so it can be good for building reports.

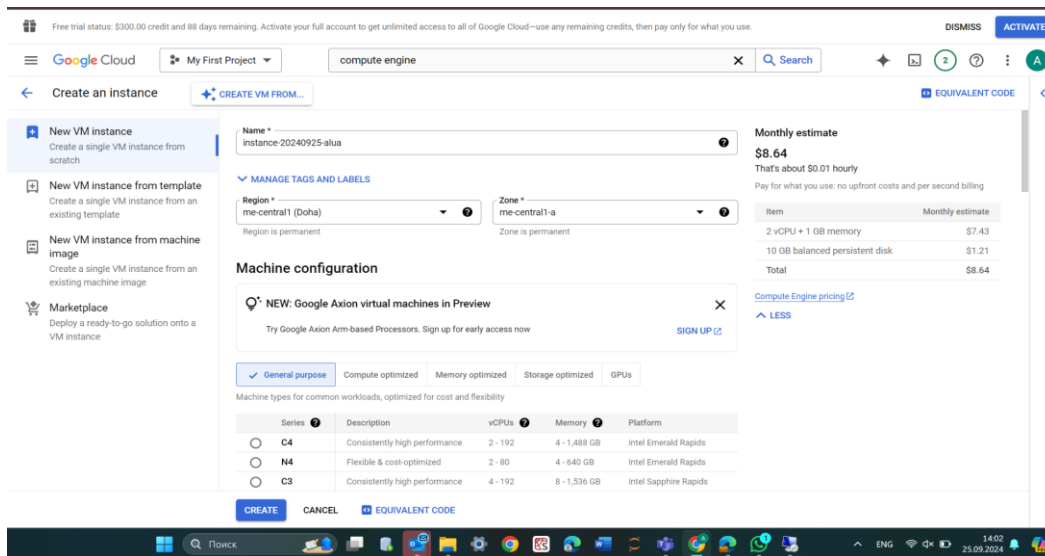
### Exercise 3: Creating and Managing Virtual Machines with Compute Engine

1. **Objective:** Learn how to create, manage, and interact with virtual machines (VMs) using Compute Engine.
2. **Steps:**
  - In the Google Cloud Console, navigate to Compute Engine and create a new VM instance.
  - Configure the VM with specific parameters, such as the machine type, region, and operating system.
  - Connect to the VM using SSH and install a basic web server (e.g., Apache or Nginx).
  - Stop, start, and delete the VM through the console.
3. **Questions:**
  - What steps did you follow to create the VM?

- 1) Press '**Create instance**' button



## 2) Edit configurations



Free trial status: \$300.00 credit and 88 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

Google Cloud My First Project compute engine

Create an instance CREATE VM FROM...

New VM instance  
Create a single VM instance from scratch

New VM instance from template  
Create a single VM instance from an existing template

New VM instance from machine image  
Create a single VM instance from an existing machine image

Marketplace  
Deploy a ready-to-go solution onto a VM instance

Name \*  
instance-20240925-alua

MANAGE TAGS AND LABELS

Region \*  
me-central1 (Doha)  
Region is permanent

Zone \*  
me-central1-a  
Zone is permanent

Monthly estimate  
\$8.64  
That's about \$0.01 hourly  
Pay for what you use: no upfront costs and per second billing

Item	Monthly estimate
2 vCPU + 1 GB memory	\$7.43
10 GB balanced persistent disk	\$1.21
Total	\$8.64

Compute Engine pricing [Compute Engine pricing](#)  
[LESS](#)

Machine configuration

NEW: Google Axion virtual machines in Preview  
Try Google Axion Arm-based Processors. Sign up for early access now [SIGN UP](#)

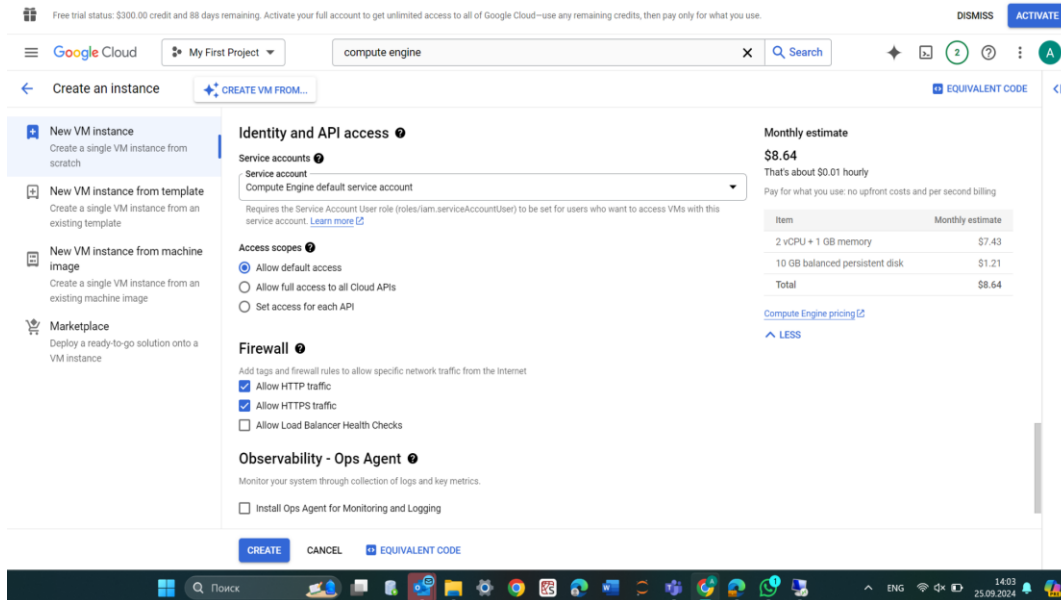
General purpose Compute optimized Memory optimized Storage optimized GPUs

Machine types for common workloads, optimized for cost and flexibility

Series	Description	vCPUs	Memory	Platform
<input type="radio"/> C4	Consistently high performance	2 - 192	4 - 1,488 GB	Intel Emerald Rapids
<input type="radio"/> N4	Flexible & cost-optimized	2 - 80	4 - 640 GB	Intel Emerald Rapids
<input type="radio"/> C3	Consistently high performance	4 - 192	8 - 1,536 GB	Intel Sapphire Rapids

CREATE CANCEL EQUIVALENT CODE

## 3) Allow HTTP, HTTPS traffic



Free trial status: \$300.00 credit and 88 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

Google Cloud My First Project compute engine

Create an instance CREATE VM FROM...

New VM instance  
Create a single VM instance from scratch

New VM instance from template  
Create a single VM instance from an existing template

New VM instance from machine image  
Create a single VM instance from an existing machine image

Marketplace  
Deploy a ready-to-go solution onto a VM instance

Identity and API access

Service accounts  
Compute Engine default service account  
Requires the Service Account User role (roles/iam.serviceAccountUser) to be set for users who want to access VMs with this service account. [Learn more](#)

Access scopes  
☒ Allow default access  
☐ Allow full access to all Cloud APIs  
☐ Set access for each API

Firewall  
Add tags and firewall rules to allow specific network traffic from the Internet  
☒ Allow HTTP traffic  
☒ Allow HTTPS traffic  
☐ Allow Load Balancer Health Checks

Observability - Ops Agent  
Monitor your system through collection of logs and key metrics.  
☐ Install Ops Agent for Monitoring and Logging

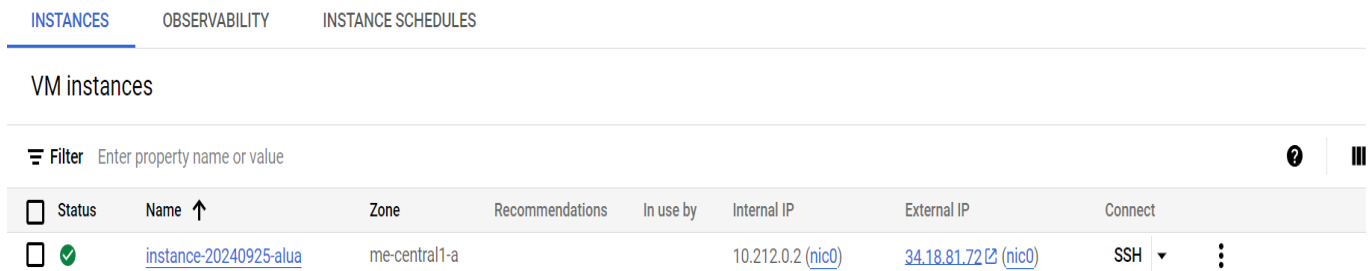
Monthly estimate  
\$8.64  
That's about \$0.01 hourly  
Pay for what you use: no upfront costs and per second billing

Item	Monthly estimate
2 vCPU + 1 GB memory	\$7.43
10 GB balanced persistent disk	\$1.21
Total	\$8.64

Compute Engine pricing [Compute Engine pricing](#)  
[LESS](#)

CREATE CANCEL EQUIVALENT CODE

## 4) Instance is created, and I'm pressing SSH



INSTANCES OBSERVABILITY INSTANCE SCHEDULES

VM instances

Filter Enter property name or value

Status	Name	Zone	Recommendations	In use by	Internal IP	External IP	Connect
<input checked="" type="checkbox"/>	<a href="#">instance-20240925-alua</a>	me-central1-a			10.212.0.2 (nic0)	<a href="#">34.18.81.72 (nic0)</a>	SSH

- How did you connect to the VM, and what commands did you use to install the web server?

After pressing to SSH, I install apache

```
SSH-in-browser
Linux instance-20240925-alua 6.1.0-25-cloud-amd64 #1 SMP PREEMPT_DYNAMIC Debian 6.1.106-3 (2024-08-26) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
onayeva_alua@instance-20240925-alua:~$ sudo apt install apache2 -y
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  apache2-bin apache2-data apache2-utils libapr1 libaprutil1 libaprutil1-dbd-sqlite3 libaprutil1-ldap
  libjansson4 liblua5.3-0 ssl-cert
Suggested packages:
  apache2-doc apache2-suexec-pristine | apache2-suexec-custom www-browser
The following NEW packages will be installed:
  apache2 apache2-bin apache2-data apache2-utils libapr1 libaprutil1 libaprutil1-dbd-sqlite3 libaprutil1-ldap
  libjansson4 liblua5.3-0 ssl-cert
0 upgraded, 11 newly installed, 0 to remove and 0 not upgraded.
Need to get 2378 kB of archives.
After this operation, 8468 kB of additional disk space will be used.
Get:1 file:/etc/apt/mirrors/debian.list Mirrorlist [30 B]
Get:2 https://deb.debian.org/debian bookworm/main amd64 libapr1 amd64 1.7.2-3 [102 kB]
Get:3 https://deb.debian.org/debian bookworm/main amd64 libaprutil1 amd64 1.6.3-1 [87.8 kB]
Get:4 https://deb.debian.org/debian bookworm/main amd64 libaprutil1-dbd-sqlite3 amd64 1.6.3-1 [13.6 kB]
Get:5 https://deb.debian.org/debian bookworm/main amd64 libaprutil1-ldap amd64 1.6.3-1 [11.8 kB]
Get:6 https://deb.debian.org/debian bookworm/main amd64 libjansson4 amd64 2.14-2 [40.8 kB]
Get:7 https://deb.debian.org/debian bookworm/main amd64 liblua5.3-0 amd64 5.3.6-2 [123 kB]
Get:8 https://deb.debian.org/debian bookworm/main amd64 apache2-bin amd64 2.4.62-1-deb12u1 [1385 kB]
Get:9 https://deb.debian.org/debian bookworm/main amd64 apache2-data all 2.4.62-1-deb12u1 [160 kB]
Get:10 https://deb.debian.org/debian bookworm/main amd64 apache2-utils amd64 2.4.62-1-deb12u1 [210 kB]
Get:11 https://deb.debian.org/debian bookworm/main amd64 apache2 amd64 2.4.62-1-deb12u1 [223 kB]
Get:12 https://deb.debian.org/debian bookworm/main amd64 ssl-cert all 1.1.2 [21.1 kB]
Fetched 2378 kB in 1s (2030 kB/s)
Preconfiguring packages ...
Selecting previously unselected package libapr1:amd64.
(Reading database ... 69885 files and directories currently installed.)
Preparing to unpack .../00-libapr1_1.7.2-3_amd64.deb ...
```

So here I can stop, start and delete VM

VM instances

CREATE INSTANCE

IMPORT VM

REFRESH

LEARN

INSTANCES

OBSERVABILITY

INSTANCE SCHEDULES

X

1

▶ START / RESUME

■ STOP

|| SUSPEND

↺ RESET

👤 CREATE A GROUP BASED ON THIS VM

🗑️ DELETE

🏷️ LABELS

👤 PERMISSIONS

Filter

Enter property name or value

?

⋮

<input checked="" type="checkbox"/>	Status	Name	Zone	Recommendations	In use by	Internal IP	External IP	Connect
<input checked="" type="checkbox"/>		<a href="#">instance-20240925-alua</a>	me-central1-a			10.212.0.2 <a href="#">(nic0)</a>	34.18.81.72 <a href="#">(nic0)</a>	SSH <div>▼</div> <div>⋮</div>

Related actions

- What happens to the VM and its data when it is stopped versus when it is deleted?

So, when VM is stopped, we are not billed for CPU usage, but will be billed for boot disk. And resources will be available. When the VM is deleted, boot disk and all resources will be deleted and no longer available unless we have created a snapshot.

## Exercise 4: Deploying a Containerized Application on Google Kubernetes Engine (GKE)

1. **Objective:** Understand how to deploy and manage containerized applications using Google Kubernetes Engine.
2. **Steps:**
  - Create a simple Docker container for a web application.
  - Push the container image to Google Container Registry (GCR).
  - Create a GKE cluster in Google Cloud Console.
  - Deploy the containerized application to the GKE cluster.
  - Expose the application to the internet and verify its accessibility.

### 3. Questions:

- How did you create and push the Docker container to GCR?

Create **alua-project** directory:

```
Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to psyched-bonfire-436413-e2.
Use "gcloud config set project [PROJECT_ID]" to change to a different project.
onayeva_alua@cloudshell:~ (psyched-bonfire-436413-e2)$ mkdir alua-project
onayeva_alua@cloudshell:~ (psyched-bonfire-436413-e2)$ cd alua-project
onayeva_alua@cloudshell:~/alua-project (psyched-bonfire-436413-e2)$ nano app0.py
onayeva_alua@cloudshell:~/alua-project (psyched-bonfire-436413-e2)$
```

Simple Hello World Flask program:



```
GNU nano 2.9.3 app0.py
from flask import Flask
app = Flask(__name__)

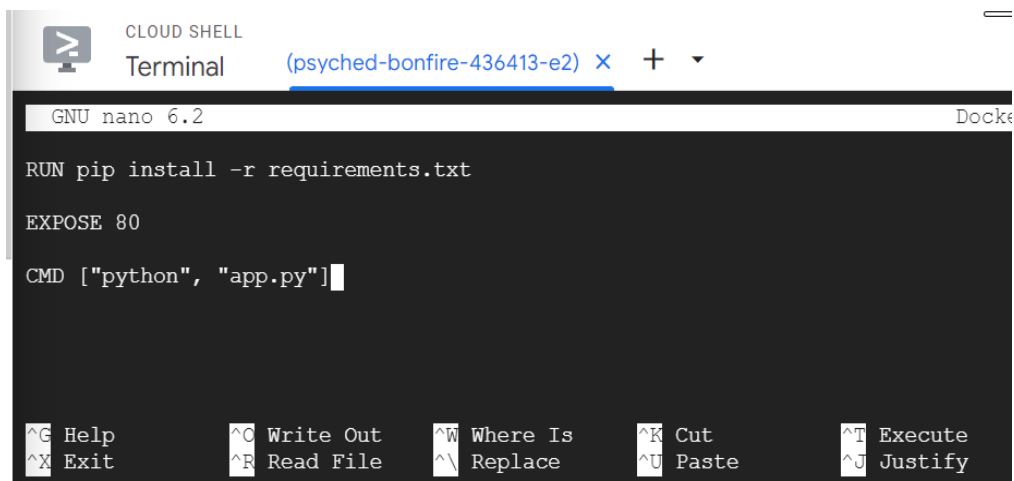
@app.route('/')
def hello_world():
    return 'Hello, World!'

if __name__ == '__main__':
    app.run(host='0.0.0.0', port=80)
```

```
onayeva_alua@cloudshell:~/alua-project (psyched-bonfire-436413-e2)$ nano app0.py
onayeva_alua@cloudshell:~/alua-project (psyched-bonfire-436413-e2)$ nano requirements.txt
onayeva_alua@cloudshell:~/alua-project (psyched-bonfire-436413-e2)$
```

Dockerfile

```
onayeva_alua@cloudshell:~/alua-project (psyched-bonfire-436413-e2)$ nano Dockerfile
onayeva_alua@cloudshell:~/alua-project (psyched-bonfire-436413-e2)$
```



```
CLOUD SHELL
Terminal (psyched-bonfire-436413-e2) x + v
GNU nano 6.2 Dockerfile
RUN pip install -r requirements.txt
EXPOSE 80
CMD ["python", "app.py"]
```

Structure of alua-project:

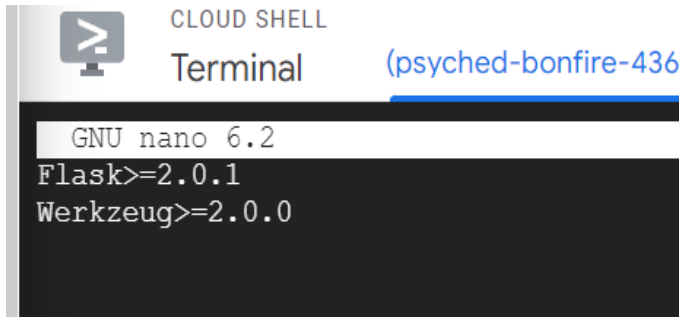
- app.py
- requirements.txt
- dockerfile

## Building docker image

**docker build -t alua-project .**

```
onayeva_alua@cloudshell:~/alua-project (psyched-bonfire-436413-e2)$ docker build -t alua-project .
[+] Building 13.1s (9/9) FINISHED
=> [internal] load build definition from Dockerfile
=> => transferring dockerfile: 159B
=> [internal] load metadata for docker.io/library/python:3.8-slim
=> [internal] load .dockerignore
=> => transferring context: 2B
```

Requirements.txt



Docker container is running the app, but it's serving on port inside the container

**docker run -p 8080:80 alua-project**

```
onayeva_alua@cloudshell:~/alua-project (psyched-bonfire-436413-e2)$ docker run -p 8080:80 alua-project
* Serving Flask app 'app'
* Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on all addresses (0.0.0.0)
* Running on http://127.0.0.1:80
* Running on http://172.17.0.2:80
Press CTRL+C to quit
^Conayeva_alua@cloudshell:~/alua-project (psyched-bonfire-436413-e2)$
```

project number (later i don't need it)

```
onayeva_alua@cloudshell:~ (psyched-bonfire-436413-e2)$ gcloud projects list
PROJECT_ID: psyched-bonfire-436413-e2
NAME: My First Project
PROJECT_NUMBER: 545668553047
```

Need to enable this API, to create artifacts like images and etc.



Artifact Registry API

[Google Enterprise API](#)

ENABLE

[TRY THIS API](#)

We need project number not project id.

**docker tag alua-project gcr.io/psyched-bonfire-436413-e2/alua-project:v1** (Tagging Docker image with the Google Container Registry (GCR) format)

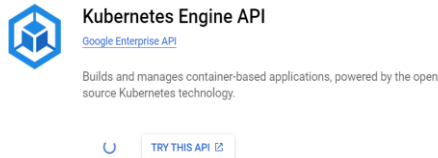
**gcloud auth configure-docker** (configure Docker to authenticate with GCR)

**docker push gcr.io/psyched-bonfire-436413-e2/alua-project:v1** (pushing Docker)

```
onayeva_alua@cloudshell:~ (psyched-bonfire-436413-e2) $ docker tag alua-project gcr.io/psyched-bonfire-436413-e2/alua-project:vl
onayeva_alua@cloudshell:~ (psyched-bonfire-436413-e2) $ docker push gcr.io/psyched-bonfire-436413-e2/alua-project:vl
The push refers to repository [gcr.io/psyched-bonfire-436413-e2/alua-project]
596575d77cb5: Pushed
aa77a101547b: Pushed
ec6b39214c20: Pushed
de06647e8a8c: Layer already exists
837964438a9e: Layer already exists
8ffef0ea5e5c: Layer already exists
8e2ab394fabf: Layer already exists
```

- What steps were involved in setting up the GKE cluster?

## Enable Kubernetes Engine API.



## We need to create a cluster

### Cluster basics

Create an Autopilot cluster by specifying a name and region. After the cluster is created, you can deploy your workload through Kubernetes and we'll take care of the rest, including:

- ✓ **Nodes:** Automated node provisioning, scaling, and maintenance
- ✓ **Networking:** VPC-native traffic routing for public or private clusters
- ✓ **Security:** Shielded GKE Nodes and Workload Identity
- ✓ **Telemetry:** Cloud Operations logging and monitoring

**Name**  
alua-cluster

Cluster names must start with a lowercase letter followed by up to 39 lowercase letters, numbers, or hyphens. They can't end with a hyphen. You cannot change the cluster's name once it's created.

**Region**  
us-central1

The regional location in which your cluster's control plane and nodes are located. You cannot change the cluster's region once it's created.

[NEXT: FLEET REGISTRATION](#) [RESET SETTINGS](#)

## Cluster is created.

OVERVIEW

OBSERVABILITY

COST OPTIMIZATION

Filter

Enter property name or value

<input type="checkbox"/> Status	Name	Location	Number of nodes	Total vCPUs
<input checked="" type="checkbox"/>	<a href="#">alua-cluster</a>	us-central1		0

## gcloud container clusters get-credentials alua-cluster --zone us-central1 --project psyched-bonfire-436413-e2

## Set up VM to use the GKE cluster

```
ire-436413-e2cloudshell:~ (psyched-bonfire-436413-e2) $ gcloud container clusters get-credentials alua-cluster --zone us-central1 --project psyched-bonfire-436413-e2
Fetching cluster endpoint and auth data.
kubeconfig entry generated for alua-cluster.
onayeva_alua@cloudshell:~ (psyched-bonfire-436413-e2) $
```



## Deployment configuration

```
Once finished, enter the verification code provided in your browser: 4/0AQ1Ed8zqE_57H8CzDoBJP07Ik0WLokE8FZf48QWtzNZ3hFo0IByuXVttcRF8Lr7

GNU nano 6.2 deployment.yaml *
apiVersion: apps/v1
kind: Deployment
metadata:
  name: web-app
spec:
  replicas: 3
  selector:
    matchLabels:
      app: web-app
  template:
    metadata:
      labels:
        app: web-app
    spec:
      containers:
        - name: web-app
          image: gcr.io/[YOUR_PROJECT_ID]/my-web-app:v1
          ports:
            - containerPort: 80

^G Help      ^C Write Out  ^W Where Is   ^K Cut        ^T Execute    ^G Location
^X Exit      ^R Read File  ^N Replace    ^U Paste      ^J Justify    ^_ Go To Line
```

Firstly, I got an error that I have insufficient privileges, while deploying.  
But when I changed my role to Kubernetes Engine Admin, everything worked.

Filter Enter property name or value						?	
<input type="checkbox"/> Type	Principal	Name	Role	Security insights			
<input type="checkbox"/>	545668553047-compute@developer.gserviceaccount.com	Compute Engine default service account	Editor				
<input type="checkbox"/>	onayeva.alua@gmail.com	Alua Onayeva	Editor Kubernetes Engine Admin Owner				

## kubectl apply -f deployment.yaml (App deploy)

```
onayeva_alua@cloudshell:~ (psyched-bonfire-436413-e2) $ nano deployment.yaml
onayeva_alua@cloudshell:~ (psyched-bonfire-436413-e2) $ kubectl apply -f deployment.yaml
- containerPort: 80

Warning: autopilot-default-resources-mutator:Autopilot updated Deployment default/web-app: defaulted unspecified 'cpu' resource for containers [web-app] (see http://g.co/gke/autopilot-defaults).
deployment.apps/web-app created
onayeva_alua@cloudshell:~ (psyched-bonfire-436413-e2) $
```

- How did you verify that your application was successfully deployed and accessible?

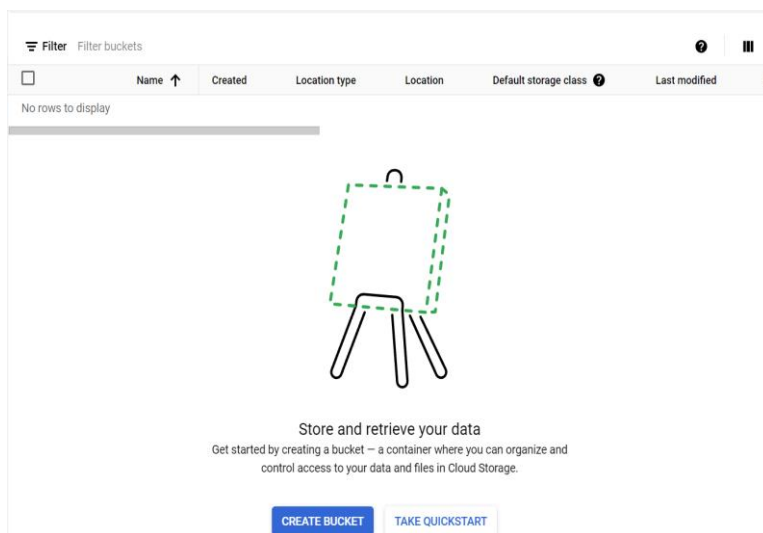
## kubectl get deployments

```
onayeva_alua@cloudshell:~ (psyched-bonfire-436413-e2) $ kubectl get deployments
NAME      READY   UP-TO-DATE   AVAILABLE   AGE
web-app   0/3     3             0           55s
onayeva_alua@cloudshell:~ (psyched-bonfire-436413-e2) $
```

## Exercise 5: Storing and Accessing Data in Google Cloud Storage

1. **Objective:** Learn how to store, manage, and access data using Google Cloud Storage.
2. **Steps:**
  - Create a new Cloud Storage bucket in the Google Cloud Console.
  - Upload various types of files (e.g., text, images, videos) to the bucket.
  - Set access permissions for the bucket and test public and private access to the files.
  - Use the Cloud Console to download, move, and delete files in the bucket.
3. **Questions:**
  - *How do you create a Cloud Storage bucket, and what options are available during setup?*

Press **'Create bucket'** button



So, we can choose the name, region, and other parameters for the bucket.

- **Choose where to store your data**

Location: us (multiple regions in United States)  
Location type: Multi-region

- **Choose a storage class for your data**

Default storage class: Standard

- **Choose how to control access to objects**

Public access prevention: On  
Access control: Uniform

- **Choose how to protect object data**

Soft delete policy: Default  
Object versioning: Disabled  
Bucket retention policy: Disabled  
Object retention: Disabled  
Encryption type: Google-managed

- *What are the differences between setting a bucket to public versus private?*

In public anyone with the link can access files. In private only users with permissions can access files.

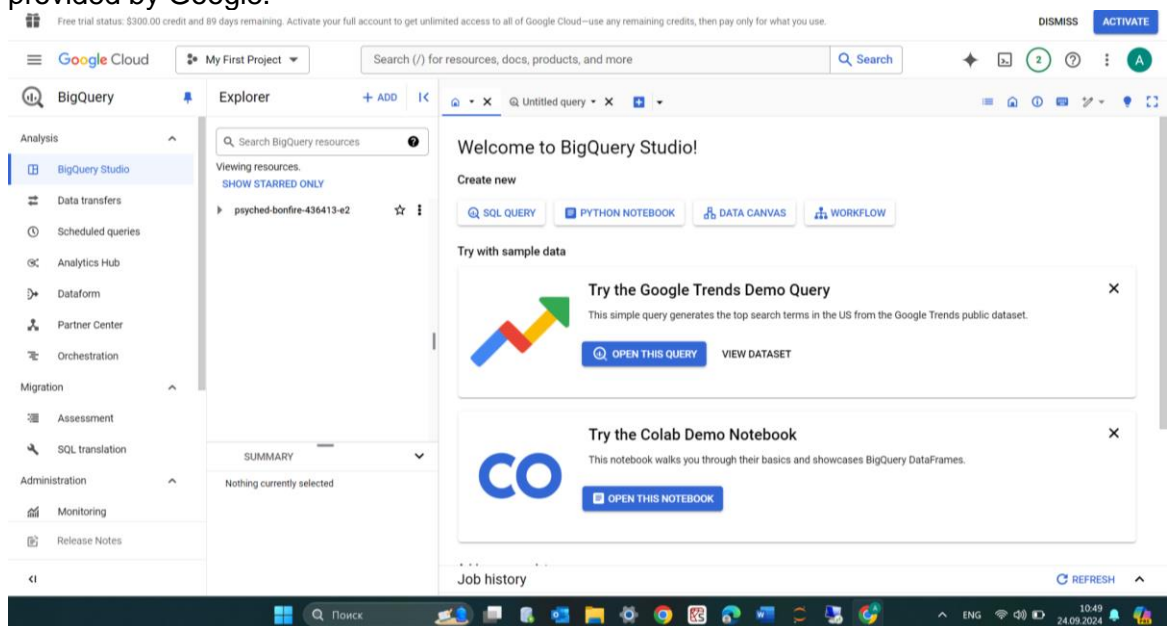
- *How can you manage access permissions for individual files in a bucket?*

So here is my bucket, and in access control I got 'Uniform'. If I edit this part I can set ACL (Access Control List), and grant specific users with permissions.

Filter Filter buckets <span>?</span> <span>⋮</span>							
<input type="checkbox"/>	Name <span>↑</span>	class <span>?</span>	Last modified	Public access <span>?</span>	Access control <span>?</span>	Protection <span>?</span>	Hierarchy
<input type="checkbox"/>	<a href="#">alua-bucket</a>		Sep 25, 2024, 6:42:41 PM	Not public	Uniform	Soft Delete	Not ena <span>⋮</span>

## Exercise 6: Analyzing Data with BigQuery

1. **Objective:** Perform data analysis tasks using BigQuery.
  2. **Steps:**
    - Access BigQuery in the Google Cloud Console.
    - Create a dataset and table by importing a sample dataset provided by Google.
    - Write and execute SQL queries to perform basic data analysis, such as filtering, aggregation, and sorting.
    - Visualize the results using Google Data Studio or another visualization tool.
  3. **Questions:**
    - *What steps did you take to create a dataset and table in BigQuery?*
1. Open BigQuery and choose 'View Dataset' option, in order to work with datasets provided by Google.



2. There is a list of datasets

Free trial status: \$300.00 credit and 89 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use. DISMISS ACTIVATE

Google Cloud My First Project Search (/) for resources, docs, products, and more Search

Explorer + ADD baseball

baseball

Dataset info EDIT DETAILS

Dataset ID: bigquery-public-data.baseball

Created: Oct 18, 2016, 11:08:24PM UTC+6

Default table expiration: Never

Last modified: Sep 20, 2022, 12:42:20PM UTC+5

Data location: US

Description: Overview: This public data includes pitch-by-pitch data for Major League Baseball (MLB) games in 2016. This dataset contains the following tables: games\_wide (every pitch, steal, or lineup event for each at bat in the 2016 regular season), games\_post\_wide (every pitch, steal, or lineup event for each at-bat in the 2016 post season), and schedules (the schedule for every team in the regular season). The schemas for the games\_wide and games\_post\_wide tables are identical. With this data you can effectively replay a game and rebuild basic statistics for players and teams.

Update frequency: Historic (none)

Dataset source: SportRadar

Terms of use: Copyright Sportradar LLC. Access to data is intended solely for internal research and testing purposes, and is not to be used for any business or commercial purpose. Data are not to be exploited in any manner without express approval from Sportradar. Display of data must include the phrase, "Data provided by Sportradar LLC," and be hyperlinked to [www.sportradar.com](https://console.cloud.google.com/marketplace/details/sportradar-public-data/mlb-pitch-by-pitch)

See the GCP Marketplace listing for more details and sample queries: <https://console.cloud.google.com/marketplace/details/sportradar-public-data/mlb-pitch-by-pitch>

Default collation

Job history REFRESH

### 3. To create table we can select 'Create Dataset' button

Create table ×

Source

Create table from  
Empty table

Destination

Project \*  
bigquery-public-data BROWSE

Dataset \*  
baseball

Table \*  
baseball

Maximum name size is 1,024 UTF-8 bytes. Unicode letters, marks, numbers, connectors, dashes, and spaces are allowed.

Table type  
Native table

Schema

☐ Edit as text

Partition and cluster settings

Partitioning  
No partitioning

CREATE TABLE CANCEL

- *How did you write and execute SQL queries in BigQuery?*

Open a query, and write a select, I use a public dataset already provided by Google. Dataset about NYC taxi rides.

Free trial status: \$300.00 credit and 89 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use. DISMISS ACTIVATE

Google Cloud My First Project Search (/) for resources, docs, products, and more Search

Explorer + ADD IK

Search BigQuery resources

Viewing resources. SHOW STARRED ONLY

- austin\_311 ☆
- austin\_bikeshare ☆
- austin\_crime ☆
- austin\_incidents ☆
- austin\_waste ☆
- baseball ☆
- bbc\_news ☆
- bigqueryml\_ncaa ☆
- bitcoin\_blockchain ☆
- blackhole\_database ☆
- blockchain\_analytics\_et... ☆

SUMMARY

Nothing currently selected

Untitled query RUN SAVE DOWNLOAD SHARE SCHEDULE OPEN IN MORE Query cancelled.

```

1 SELECT *
2 FROM `bigquery-public-data.new_york_taxi_trips.tlc_yellow_trips_2020`;
3

```

Query results SAVE RESULTS EXPLORE DATA

Row	mta_tax	tip_amount	tolls_amount	imp_surcharge	airport_fee	total_amount	pickup_location_id
1	0.5	0.5	2	0	0.3	16.3	230
2	0.5	0.5	2.66	0	0.3	15.96	209
3	0.5	0.5	2.66	0	0.3	15.96	162
4	0.5	0.5	3.76	0	0.3	22.56	144
5	0.5	0.5	1	0	0.3	13.8	262

Results per page: 50 1 – 50 of 24649081 IK < > IK

Job history REFRESH

- What insights were you able to derive from the data analysis?

Average fare and tip amounts for different payment methods.

Free trial status: \$300.00 credit and 89 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use. DISMISS ACTIVATE

Google Cloud My First Project Search (/) for resources, docs, products, and more Search

Explorer + ADD IK

Search BigQuery resources

Viewing resources. SHOW STARRED ONLY

- austin\_311 ☆
- austin\_bikeshare ☆
- austin\_crime ☆
- austin\_incidents ☆
- austin\_waste ☆
- baseball ☆
- bbc\_news ☆
- bigqueryml\_ncaa ☆
- bitcoin\_blockchain ☆
- blackhole\_database ☆
- blockchain\_analytics\_et... ☆

SUMMARY

Nothing currently selected

Untitled query RUN SAVE DOWNLOAD SHARE SCHEDULE OPEN IN MORE This query will process 822.75 M...

```

1 SELECT
2   payment_type,
3   AVG(fare_amount) AS avg_fare,
4   AVG(tip_amount) AS avg_tip
5 FROM
6   `bigquery-public-data.new_york_taxi_trips.tlc_yellow_trips_2020`
7 GROUP BY
8   payment_type
9 ORDER BY
10  payment_type;
11

```

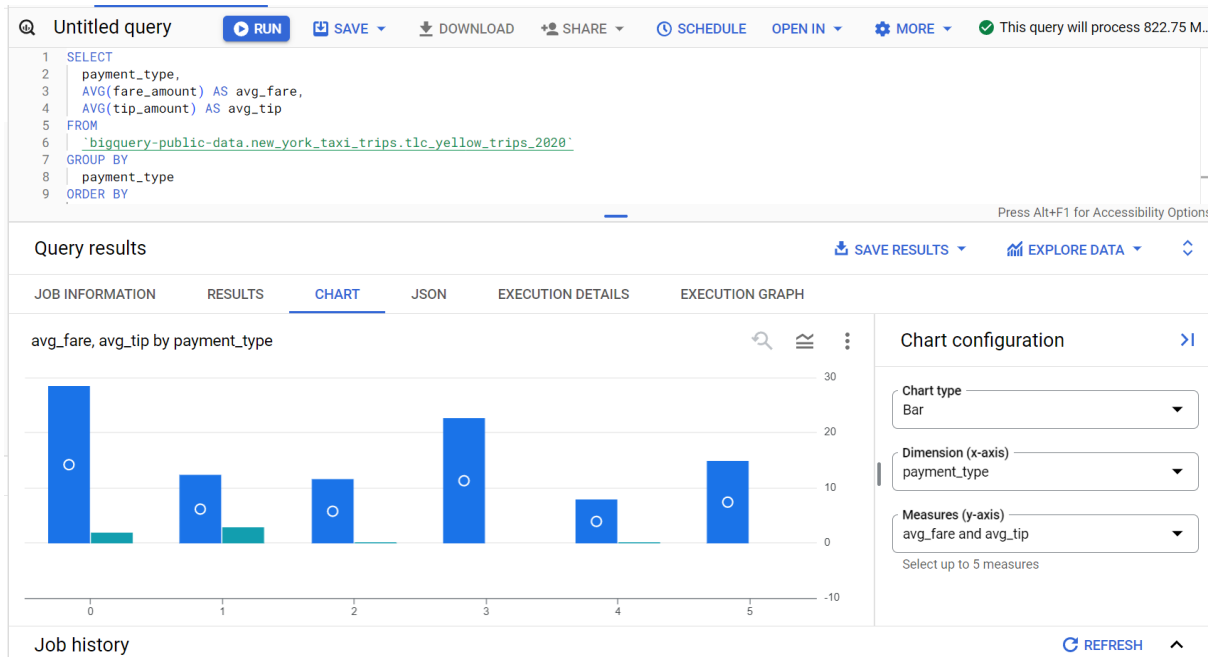
Query results SAVE RESULTS EXPLORE DATA

Row	payment_type	avg_fare	avg_tip
1	0	28.359941368	1.758334969
2	1	12.239391793	2.856962972
3	2	11.646761444	0.000347372
4	3	22.679912794	-0.002991037
5	4	7.837719541	0.009820798
6	5	14.9	0

Results per page: 50 1 – 6 of 6 IK < > IK

Job history REFRESH

Also, we can visualize it.



I want to see the distribution of trips amount by each day of the week. So as it can be seen, Friday is the day with the most taxi trips.

