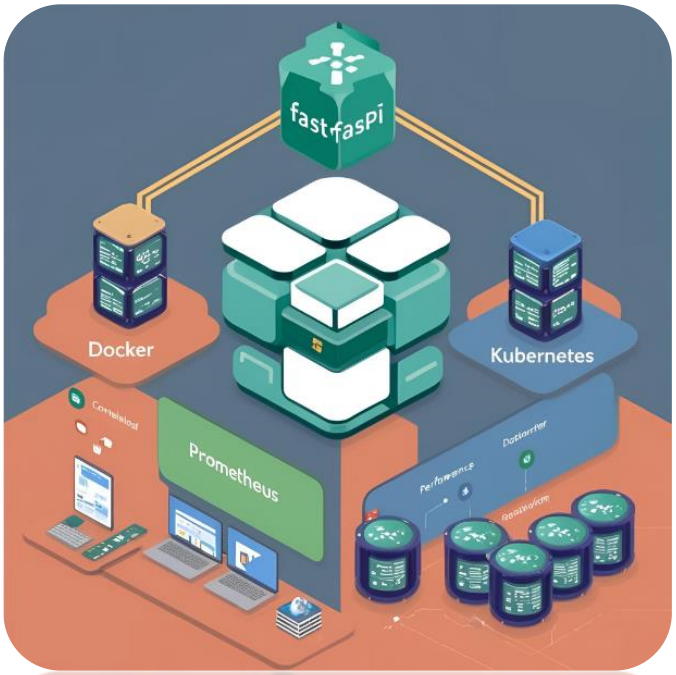


DEVOPS FOR CLOUD

Project Synopsis

Developed a backend application using FastAPI, containerized with Docker, deployed on Kubernetes, and monitored with Prometheus for performance tracking.

Kaushik Paul
2024mt03072@wilp.bits-pilani.ac.in

Contents

Kaushik Paul
2024MT03072@WILP.BITS-PILANI.AC.IN

Software and Platform Details

This section outlines the technologies, software, and platforms used throughout the completion of this assignment.

Component	Details
Operating System	Ubuntu 24.04 LTS (64-bit)
Python Version	Python 3.12.3
FastAPI Framework	FastAPI 0.115.12
ASGI Server	Uvicorn 0.34.2
Virtual Environment	venv (python -m venv .venv)
Containerization	Docker Engine (v26.1.3) installed via apt-get
Container Orchestration	Minikube v1.35.0 running Kubernetes v1.32.0
Monitoring Tool	Prometheus 3.3.1 deployed via Kubernetes YAML files
Code Editor	Visual Studio Code (VSCode) with Python and Docker extensions
Web Browser	Google Chrome
API Testing Tool	curl (command-line), browser (http://localhost:8000/get_info)
Version Control	Git + GitHub repository: link

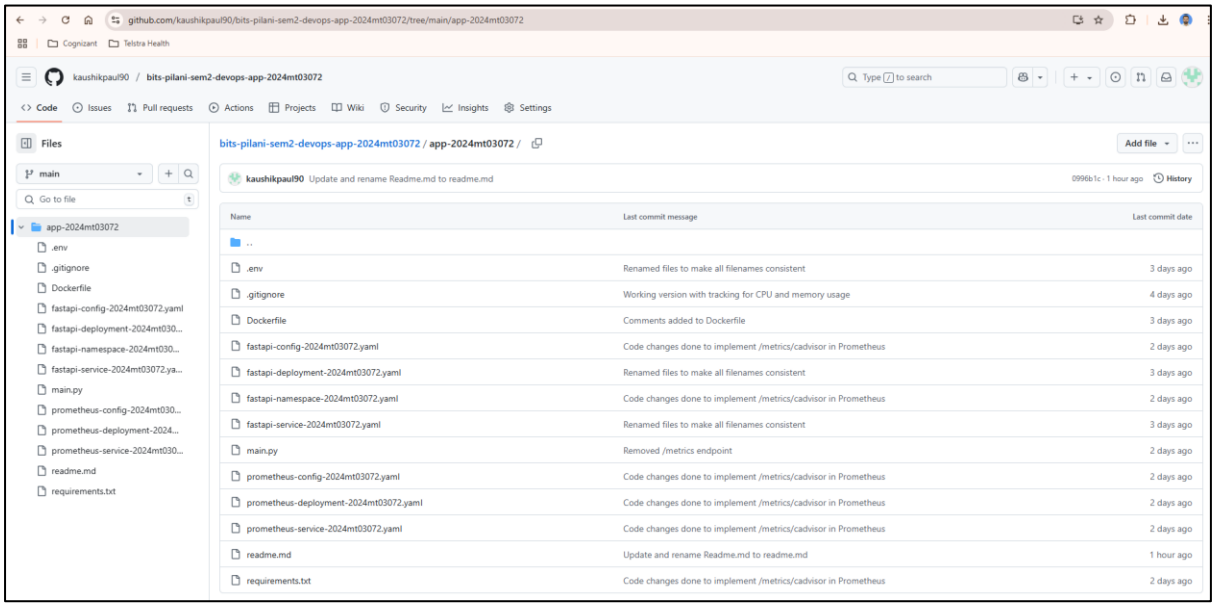
Note: Docker Desktop was not used. Instead, Docker was installed on a native Ubuntu Linux system, and Minikube was used for local Kubernetes deployment. All functional requirements remain the same.

Kaushik Paul
2024MT03072@WILP.BITS-PILANI.AC.IN

Project Repository

Github link → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/tree/main/app-2024mt03072

Screenshot



Task 1: Create the Backend Application using FastAPI

1. Project Setup

a. Creating a new directory named **app-2024mt03072** for the Python project FastAPI.

```bash
mkdir app-2024mt03072
```

b. Changing the current working directory to app-2024mt03072.

```bash
cd app-2024mt03072
```

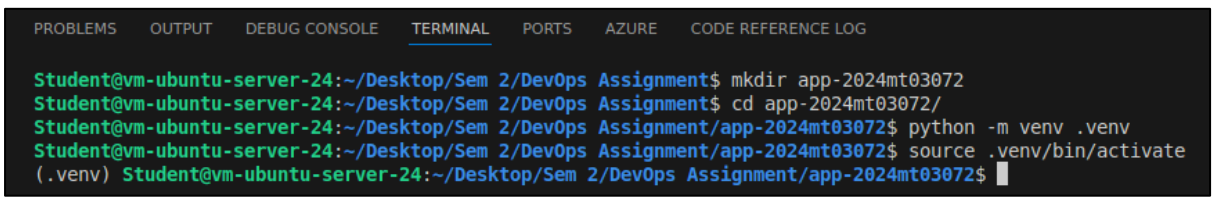
c. Creating a virtual environment for the project app-2024mt03072.

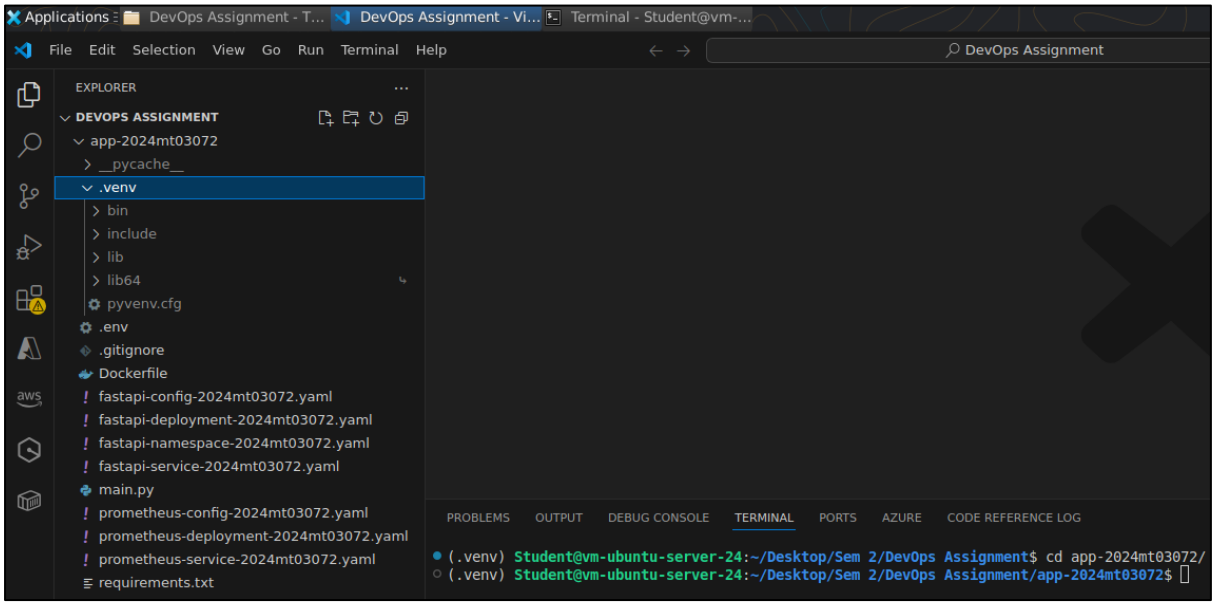
```bash
python -m venv .venv
```

d. Activating the Python virtual environment named ".venv".

```bash
source .venv/bin/activate
```

1.1. Screenshot





2. Dependency Installation

a. Adding required packages to requirements.txt file for installation.

File link → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/requirements.txt

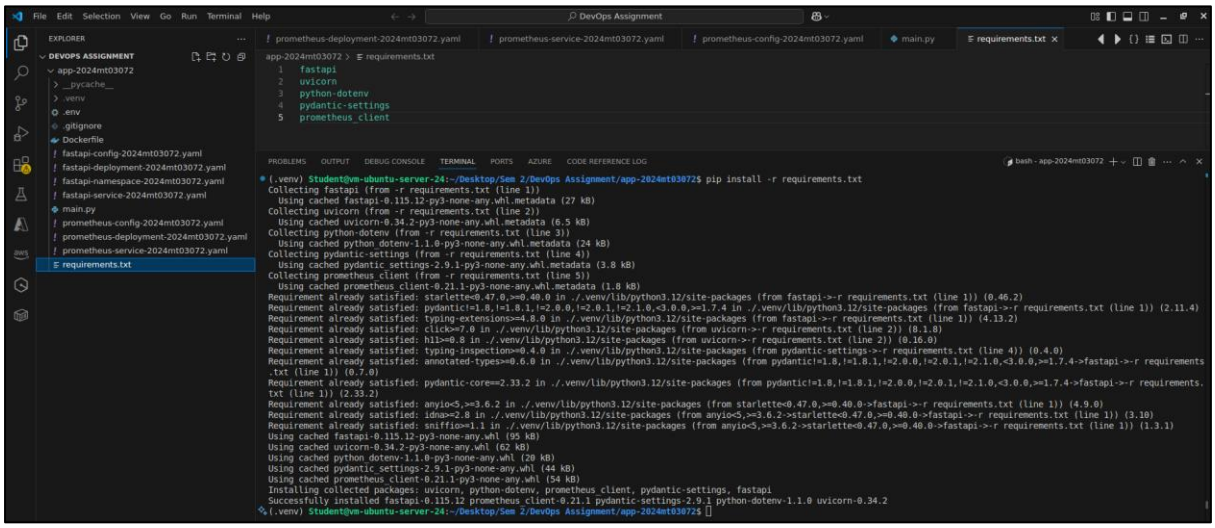
b. Installing the packages listed in requirements.txt using pip install.

```bash
pip install -r requirements.txt
```

Note:

- **fastapi**: A modern, high-performance web framework for building APIs with Python 3.7+ based on standard Python type hints.
- **python-dotenv**: A tool that loads environment variables from a .env file into the Python environment.
- **pydantic-settings**: A library to manage application settings using Pydantic models with support for environment variable loading.
- **uvicorn**: A lightning-fast ASGI server for serving FastAPI and other ASGI-compatible Python web apps.
- **prometheus_client**: A Python client for exposing metrics in a format Prometheus can scrape for monitoring and alerting.

2.1. Screenshot

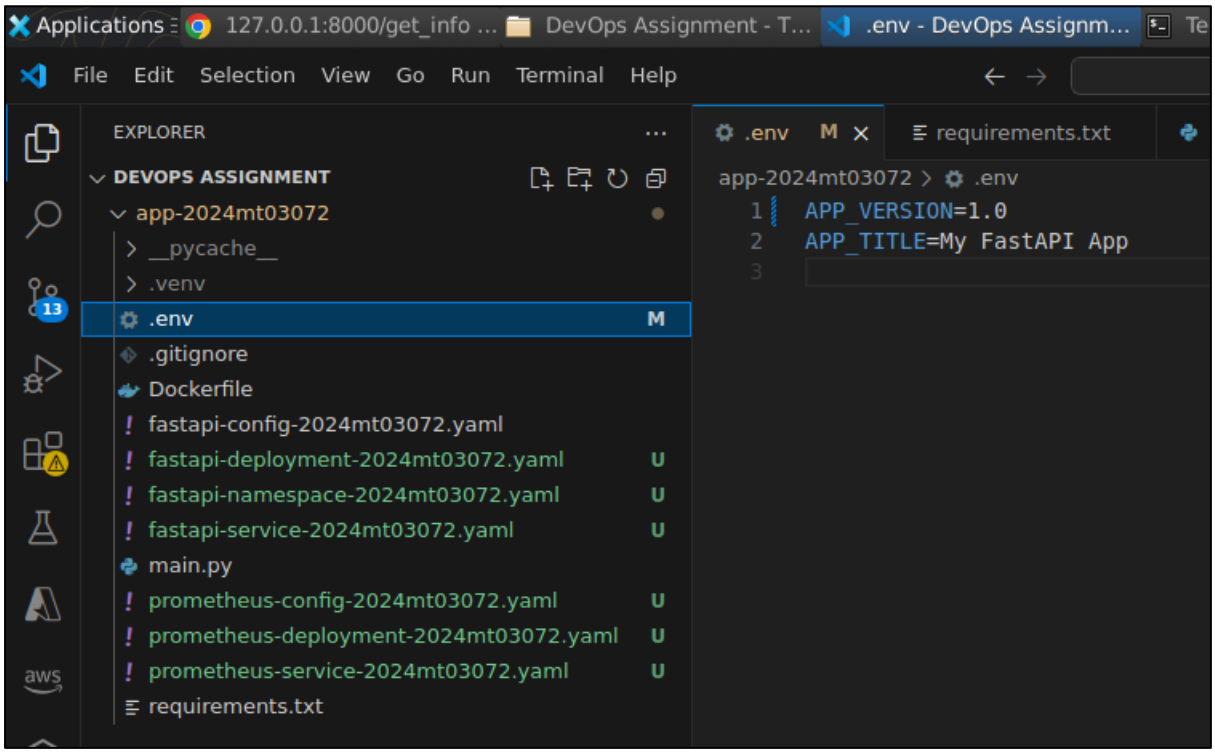


3. Environment Variables Setup

a. Creating a .env file in the project directory app-2024mt03072 to store the values of APP_VERSION and APP_TITLE.

File link → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/.env

3.1. Screenshot



Kaushik Paul
2024MT03072@WILP.BITS-PILANI.AC.IN

4. Creation of main.py

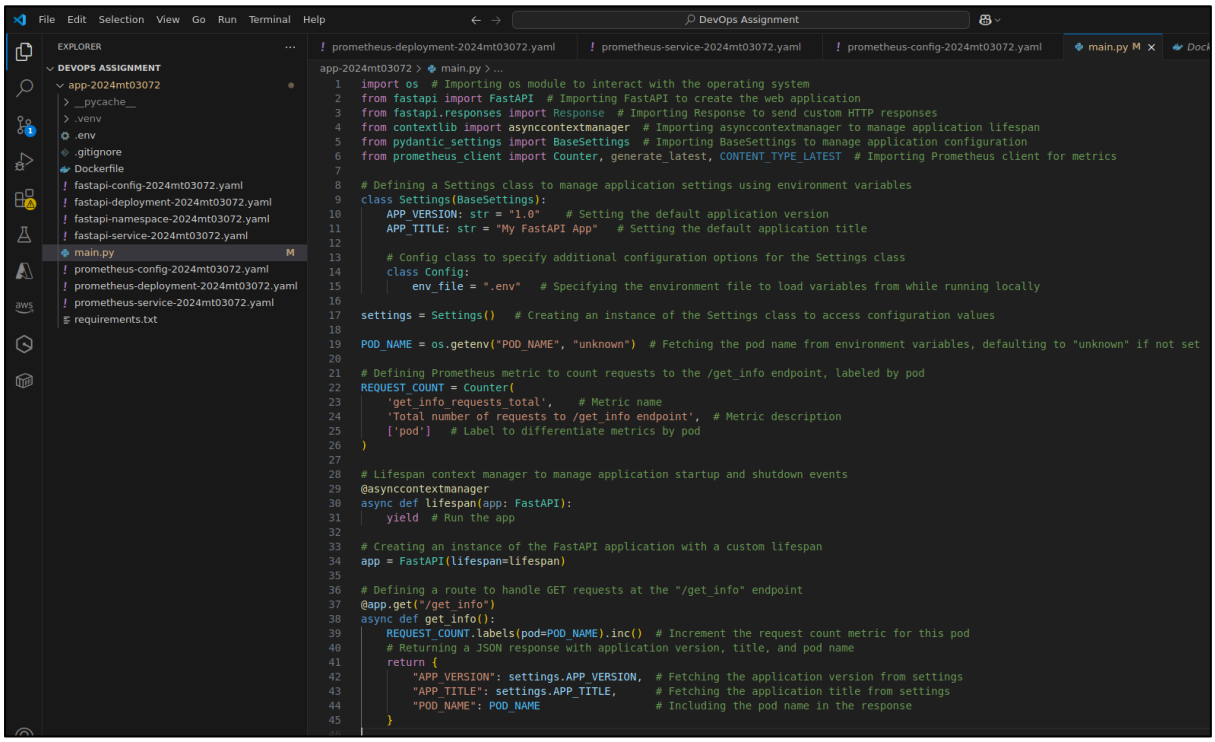
a. Creating Python file named main.py in the project directory app-2024mt03072.

File link → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/main.py

Note:

- The POD_NAME attribute has been included in the returned JSON object to indicate which pod replica is processing the request.
- When running the application locally or in a containerized environment, POD_NAME is displayed as “unknown”.
- When the application is deployed and browsed from Minikube cluster, it shows the actual pod replica name.

4.1. Screenshot



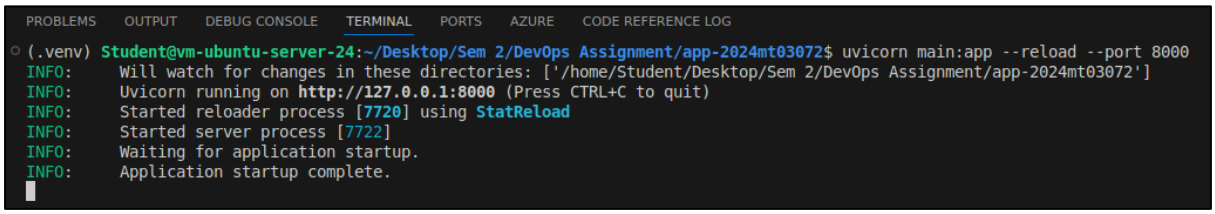
5. Testing FastAPI Application Locally

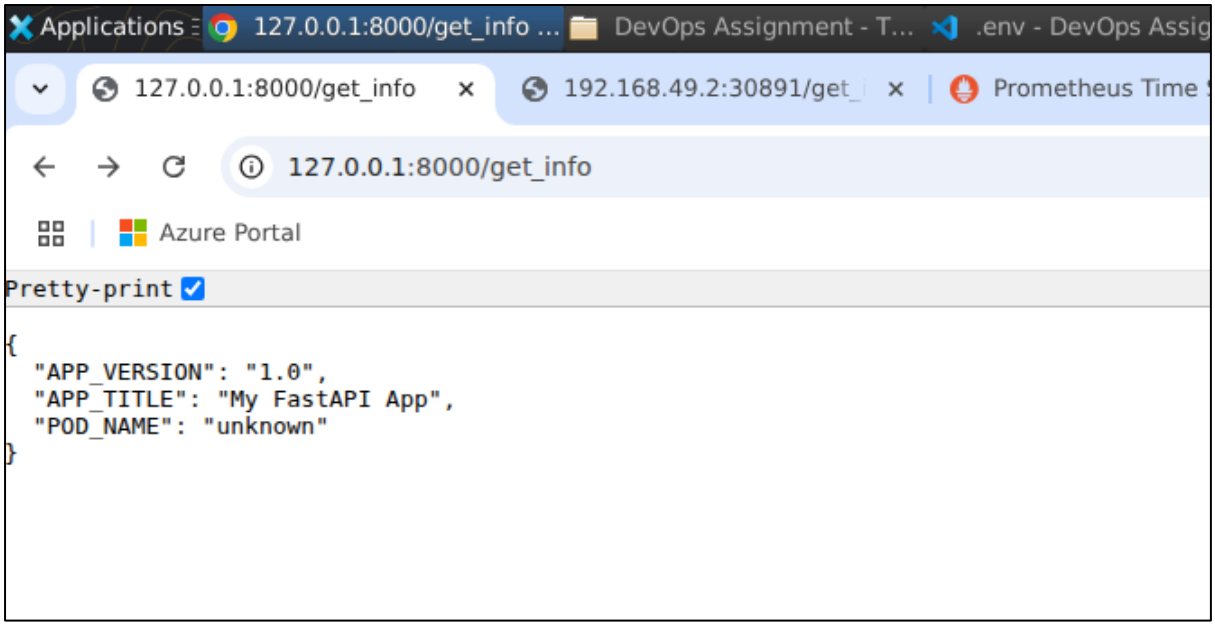
a. Running the application locally on Uvicorn.

```
bash

uvicorn main:app --reload --port 8000
```

b. Accessing the application via http://localhost:8000/get_info on Google Chrome.

5.1. Screenshot





6. Challenges Faced

a. No challenges faced while creating the backend application using FastAPI.

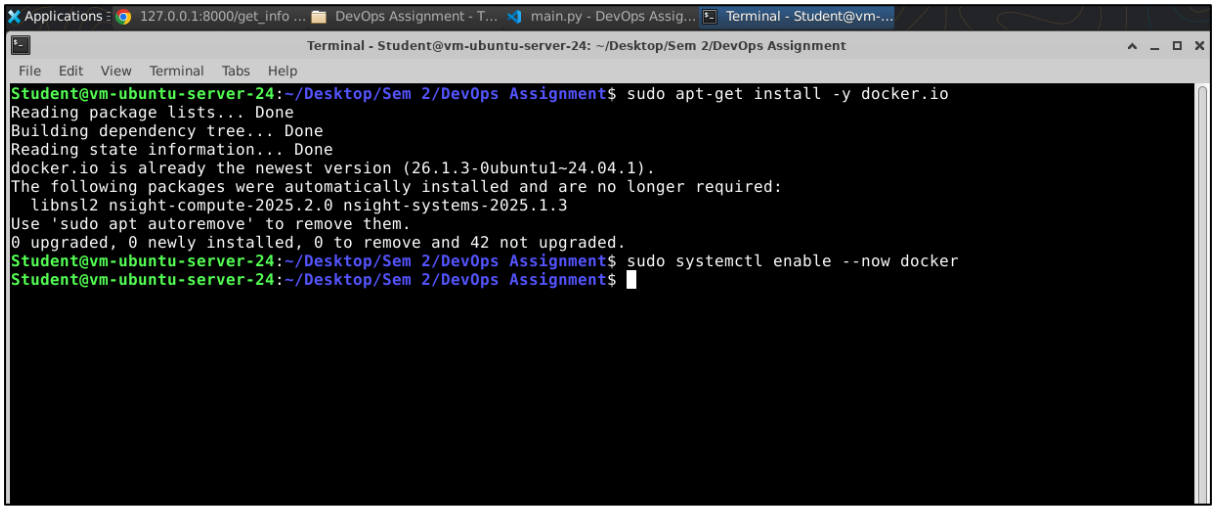
Task 2: Dockerize the Backend Application

1. Docker Setup

a. Installing and enabling Docker on Linux OS.

```
bash

sudo apt-get install -y docker.io
sudo systemctl enable --now docker
```

1.1. Screenshot

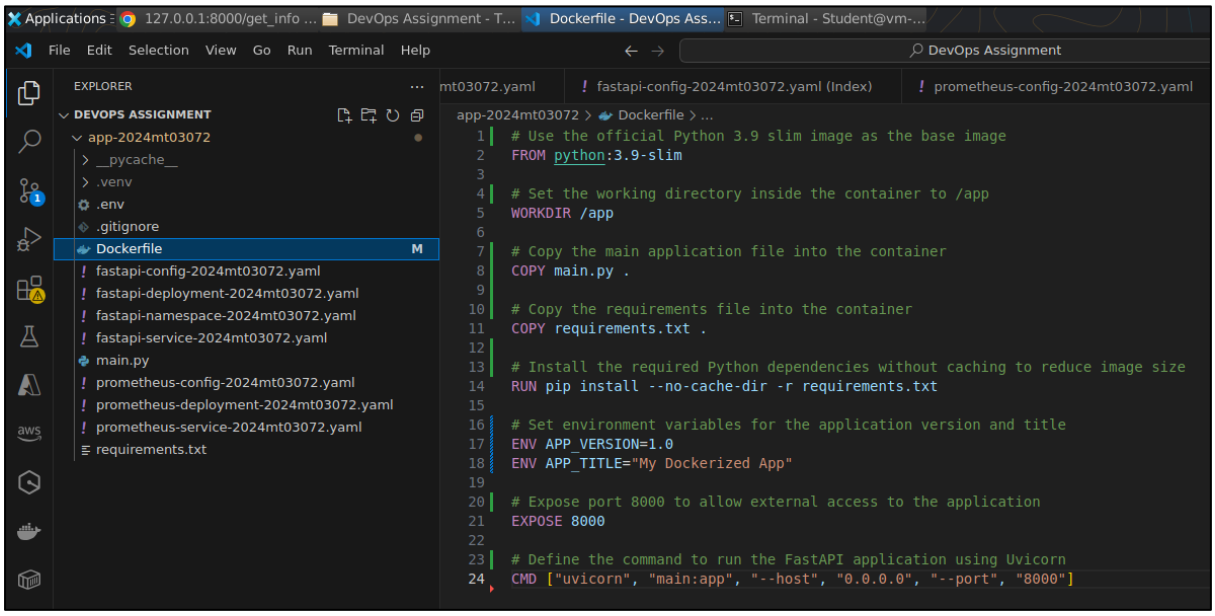


2. Dockerfile Creation

a. Creating a file named Dockerfile inside the project directory app-2024mt03072.

File link → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/Dockerfile

2.1. Screenshot

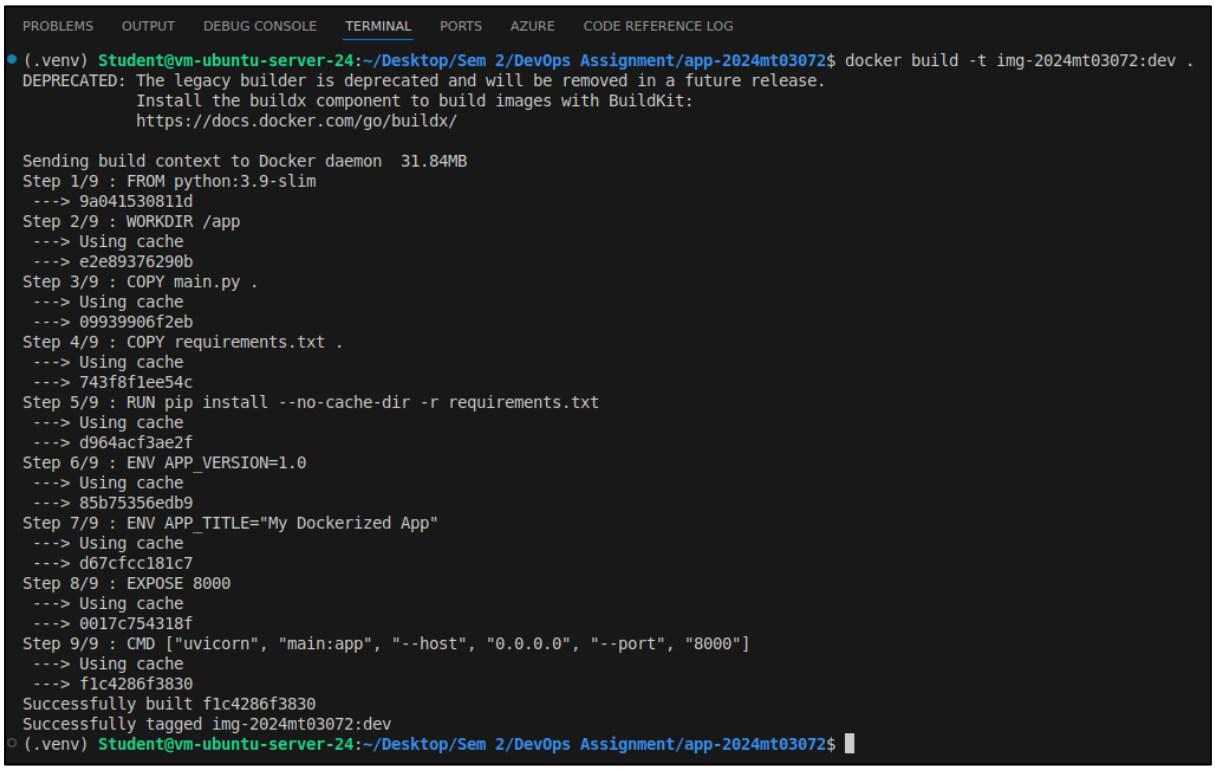


3. Docker Image Build

a. Building the Docker image using Dockerfile. Docker image is named as **img-2024mt03072** and tag is set to **dev**.

```bash
docker build -t img-2024mt03072:dev .
```

3.1. Screenshot

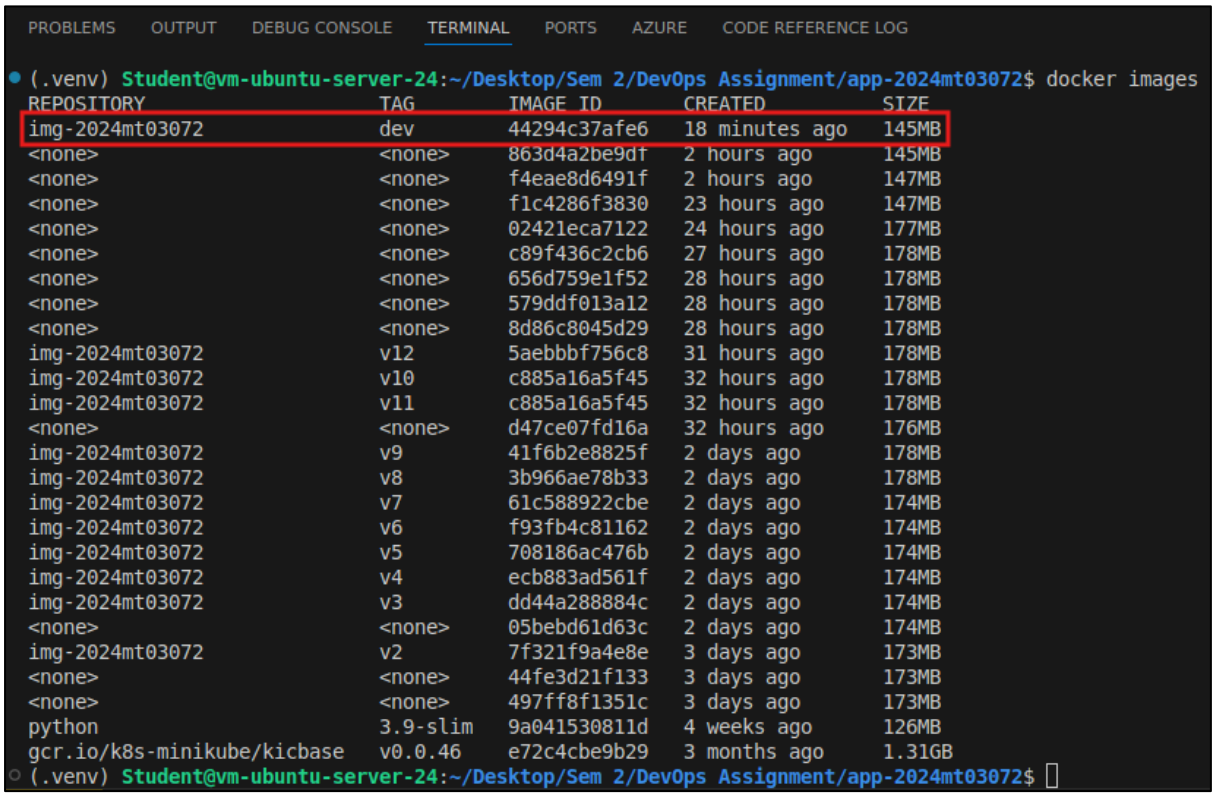


4. Docker Image Verification

a. Verifying the Docker image creation.

```
bash

docker images
```

4.1. Screenshot



5. Challenges Faced

a. No challenges faced while dockerizing the backend application.

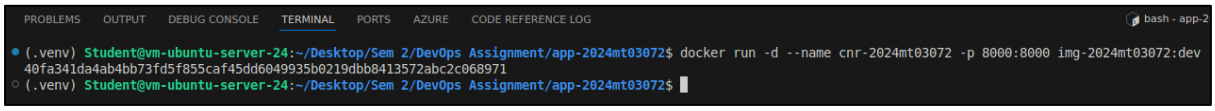
Task 3: Run the Docker Container

1. Docker Container Creation

a. Creating Docker container using the built image img-2024mt03072. Container is named as **cnr-2024mt03072:dev**.

```
bash

docker run -d --name cnr-2024mt03072 -p 8000:8000 img-2024mt03072:dev
```

1.1. Screenshot

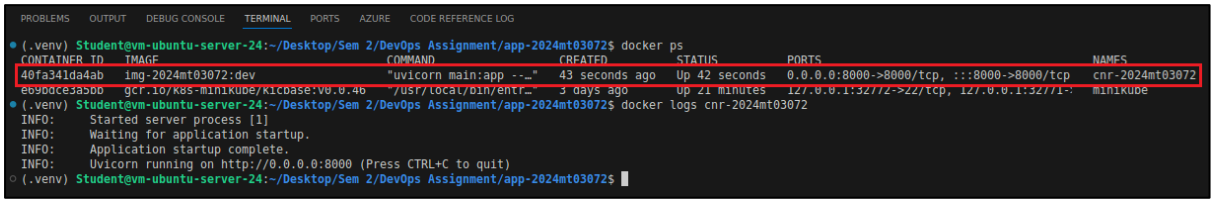


2. Docker Container Verification

a. Verifying the created Docker container.

```
bash

docker ps
docker logs cnr-2024mt03072
```

2.1. Screenshot



3. Testing the Containerized Application

a. Accessing the containerized application via http://localhost:8000/get_info on Google Chrome.

3.1. Screenshot

4. Challenges Faced

a. No challenges faced while running the dockerized backend application over Google Chrome.

Kaushik Paul
2024MT03072@WILP.BITS-PILANI.AC.IN

Task 4: Deploy Docker Image to Kubernetes Cluster

1. Minikube Setup

a. Downloading Minikube.

```
bash

curl -LO
https://storage.googleapis.com/minikube/releases/latest/minikube-
linux-amd64
```

b. Installing Minikube.

```
bash

sudo install minikube-linux-amd64 /usr/local/bin/minikube
```

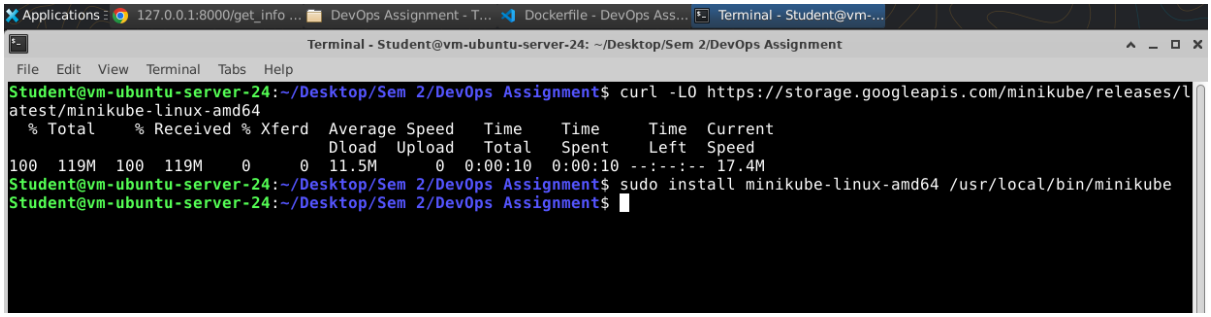
c. Starting Minikube.

```
bash

minikube start
```

d. Verifying if the Minikube cluster is running properly.

```
bash

minikube status
```

1.1. Screenshot



```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS  AZURE  CODE REFERENCE LOG

(.venv) Student@vm-ubuntu-server-24:~/Desktop/Sem 2/DevOps Assignment/app-2024mt03072$ minikube start
😄  minikube v1.35.0 on Ubuntu 24.04
✨  Using the docker driver based on existing profile
👍  Starting "minikube" primary control-plane node in "minikube" cluster
🚜  Pulling base image v0.0.46 ...
🔄  Updating the running docker "minikube" container ...
🐳  Preparing Kubernetes v1.32.0 on Docker 27.4.1 ...
    ▪ kubelet.read-only-port=10255
🔎  Verifying Kubernetes components...
    ▪ Using image gcr.io/k8s-minikube/storage-provisioner:v5
🌟  Enabled addons: storage-provisioner, default-storageclass
💡  kubectl not found. If you need it, try: 'minikube kubectl -- get pods -A'
🏄  Done! kubectl is now configured to use "minikube" cluster and "default" namespace by default
(.venv) Student@vm-ubuntu-server-24:~/Desktop/Sem 2/DevOps Assignment/app-2024mt03072$ minikube status
minikube
type: Control Plane
host: Running
kubelet: Running
apiserver: Running
kubeconfig: Configured

(.venv) Student@vm-ubuntu-server-24:~/Desktop/Sem 2/DevOps Assignment/app-2024mt03072$
```

2. Loading Docker Image into Minikube

a. Loading the image img-2024mt03072 into Minikube cluster.

```
bash

minikube image load img-2024mt03072:dev
```

2.1. Screenshot

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS  AZURE  CODE REFERENCE LOG

(.venv) Student@vm-ubuntu-server-24:~/Desktop/Sem 2/DevOps Assignment/app-2024mt03072$ minikube image load img-2024mt03072:dev
(.venv) Student@vm-ubuntu-server-24:~/Desktop/Sem 2/DevOps Assignment/app-2024mt03072$
```

3. Kubernetes Namespace Creation

a. Creating Kubernetes Namespace **ns-fastapi-2024mt03072** using the Namespace configuration file **fastapi-namespace-2024mt03072.yaml**.
File → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/fastapi-namespace-2024mt03072.yaml.

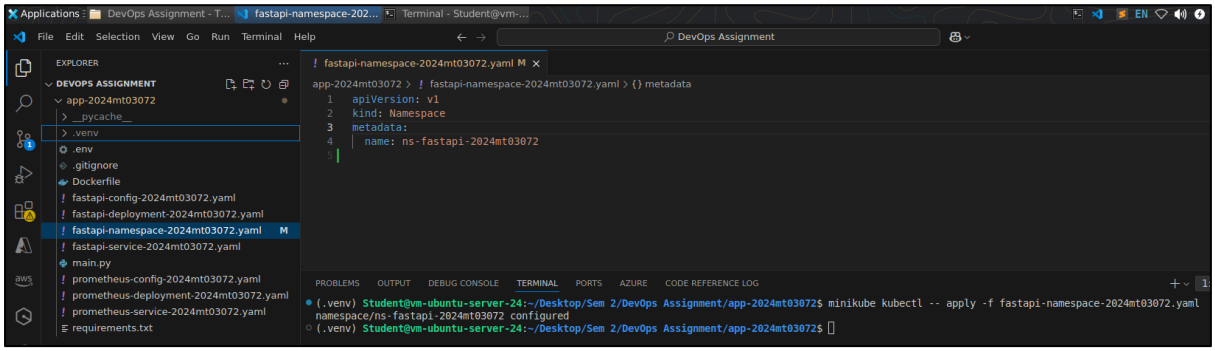
b. Applying the Kubernetes Namespace.

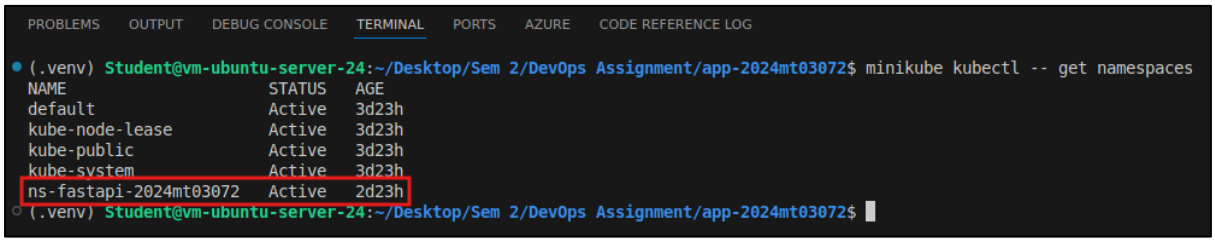
```
bash

minikube kubectl -- apply -f fastapi-namespace-2024mt03072.yaml
```

c. Verifying the creation of the Kubernetes Namespace ns-fastapi-2024mt03072.

```
bash

minikube kubectl -- get namespaces
```

3.1. Screenshot





4. Kubernetes ConfigMap Creation

a. Creating Kubernetes ConfigMap **config-2024mt03072** using the ConfigMap configuration file **fastapi-config-2024mt03072.yaml**.

File → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/fastapi-config-2024mt03072.yaml.

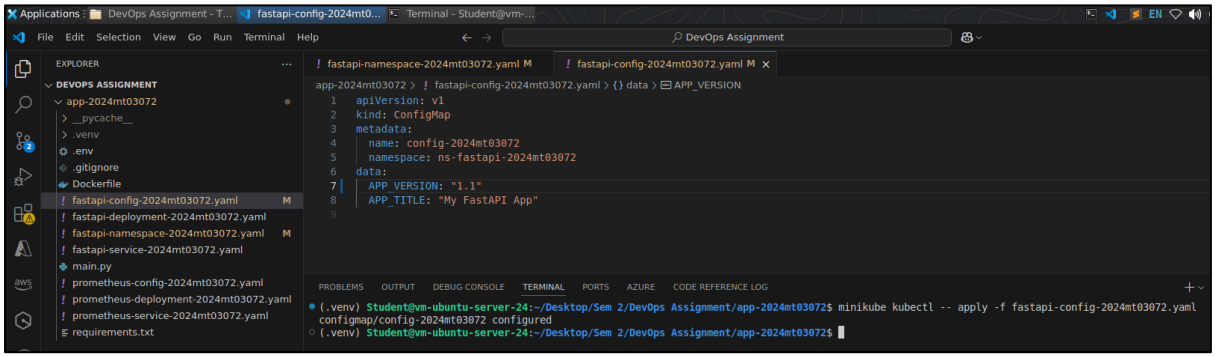
b. Applying the Kubernetes ConfigMap.

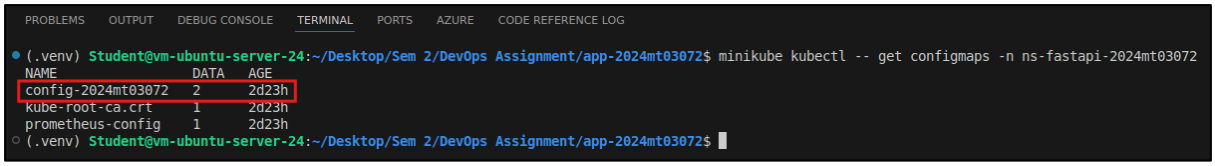
```
bash

minikube kubectl -- apply -f fastapi-config-2024mt03072.yaml
```

c. Verifying the creation of the Kubernetes ConfigMap config-2024mt03072.

```
bash

minikube kubectl -- get configmaps -n ns-fastapi-2024mt03072
```

4.1. Screenshot





5. Kubernetes Deployment Creation

a. Creating Kubernetes Deployment **fastapi-deployment** using the Deployment configuration file **fastapi-deployment-2024mt03072.yaml**.
File → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/fastapi-deployment-2024mt03072.yaml.

b. It creates 2 replicas of the FastAPI application.

c. It reads the APP_VERSION and APP_TITLE values from Kubernetes ConfigMap named fastapi-config-2024mt03072.yaml.

d. Applying the Kubernetes Deployment.

```bash
minikube kubectl -- apply -f fastapi-deployment-2024mt03072.yaml
```

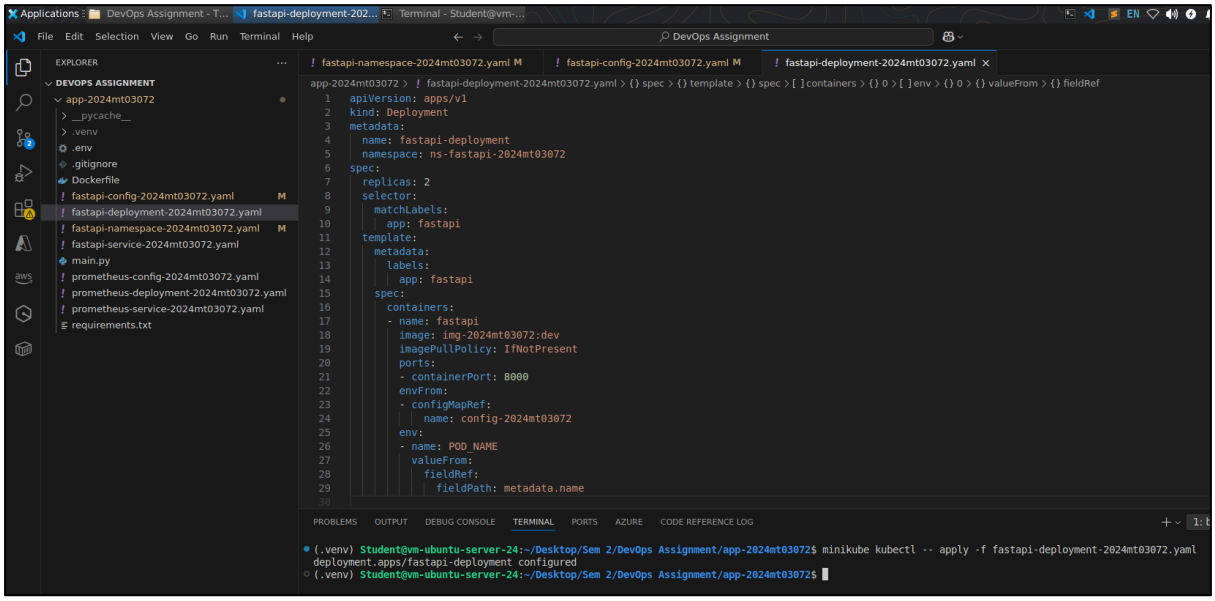
e. Verifying the creation of the Kubernetes Deployment fastapi-deployment.

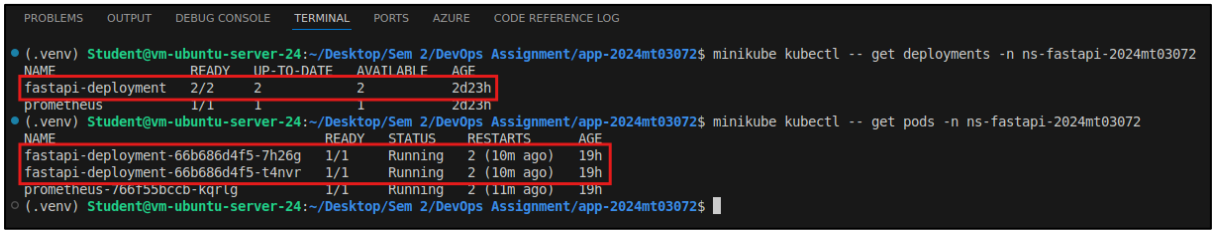
```bash
minikube kubectl -- get deployments -n ns-fastapi-2024mt03072
```

f. Listing all pods running inside Kubernetes Namespace ns-fastapi-2024mt03072.

```bash
minikube kubectl -- get pods -n ns-fastapi-2024mt03072
```

5.1. Screenshot





6. Challenges Faced

a. Pods not reflecting the latest implementation:

- Setting **imagePullPolicy** to **IfNotPresent** in fastapi-namespace-2024mt03072.yaml resolved the issue.

b. Adding new image tag each time while building the image and updating the same in fastapi-namespace-2024mt03072.yaml:

- Using fixed image tag “dev” during image build and adding it to the image name in fastapi-namespace-2024mt03072.yaml only once resolved the issue.

Task 5: Configure Networking with Load Balancer in Kubernetes Cluster

1. Kubernetes Service Creation

a. Creating Kubernetes Service **fastapi-service** using the Service configuration file **fastapi-service-2024mt03072.yaml**.

File → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/fastapi-service-2024mt03072.yaml.

b. It creates a Load Balancer to balance requests across both replicas.

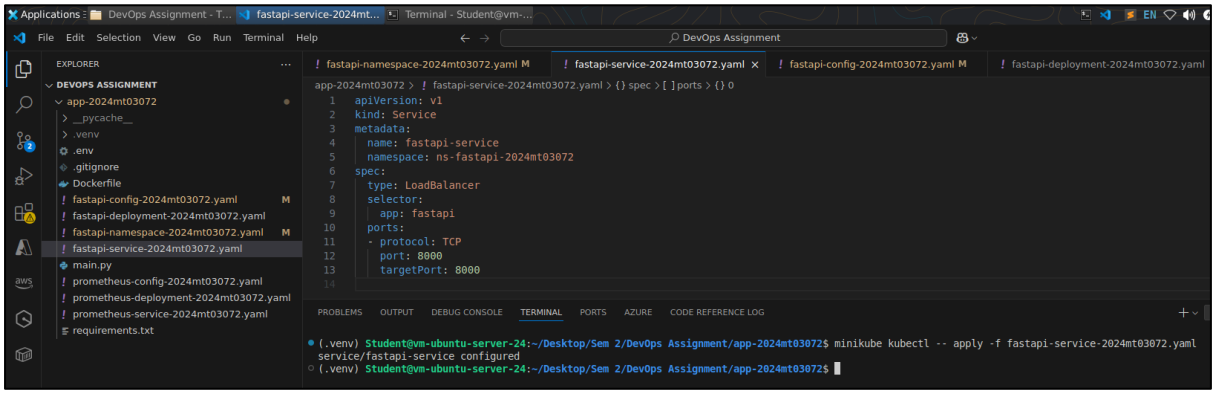
c. Applying the Kubernetes Service.

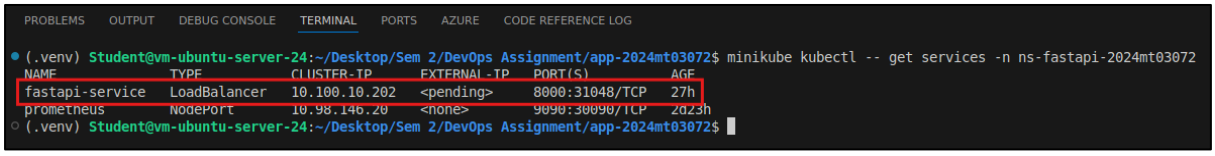
```bash
minikube kubectl -- apply -f fastapi-service-2024mt03072.yaml
```

d. Verifying the creation of the Kubernetes Service fastapi-service.

```bash
minikube kubectl -- get services -n ns-fastapi-2024mt03072
```

1.1. Screenshot



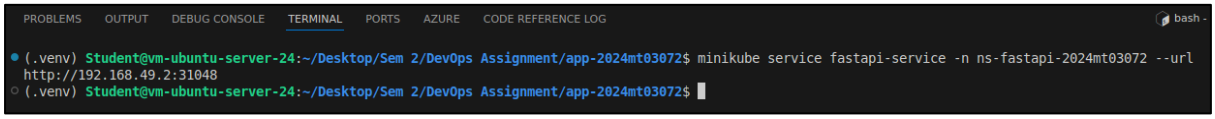


2. Obtaining Access URL

a. Retrieving the URL to access FastAPI application from browser.

```
bash

minikube service fastapi-service -n ns-fastapi-2024mt03072 --url
```

2.1. Screenshot

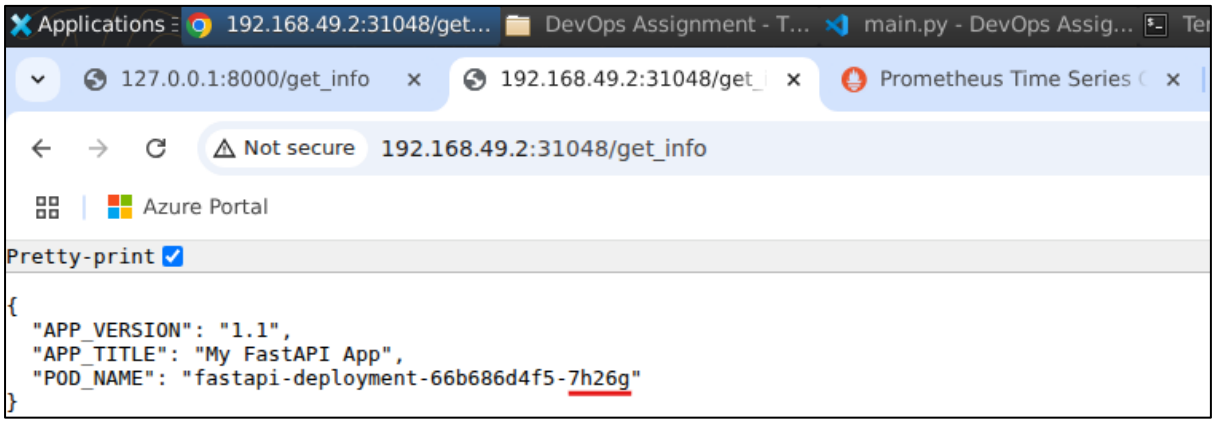


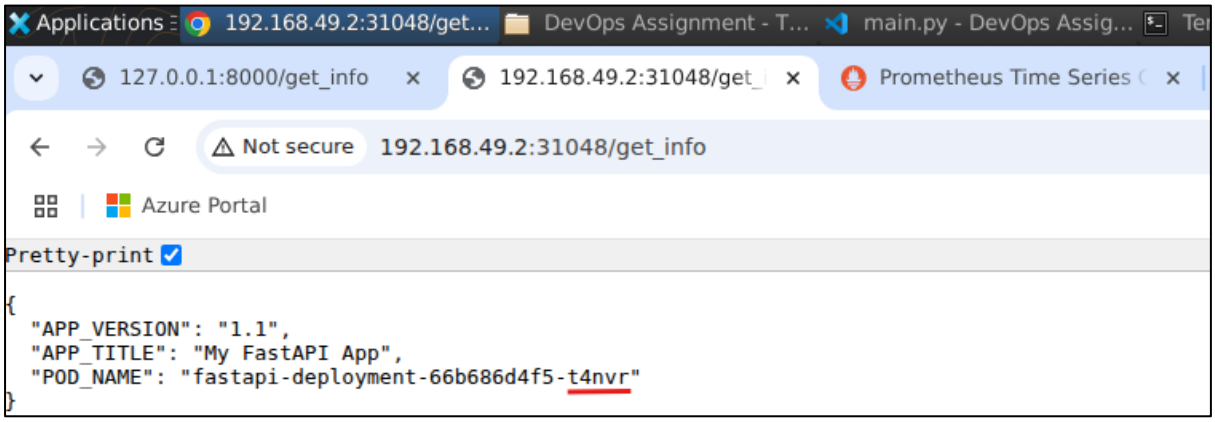
3. Accessing FastAPI Application

a. Accessing the application via http://192.168.49.2:31048/get_info on Google Chrome.

b. The fact that requests are getting routed to both the replicas is evident from the POD_NAME in the JSON object.

3.1. Screenshot





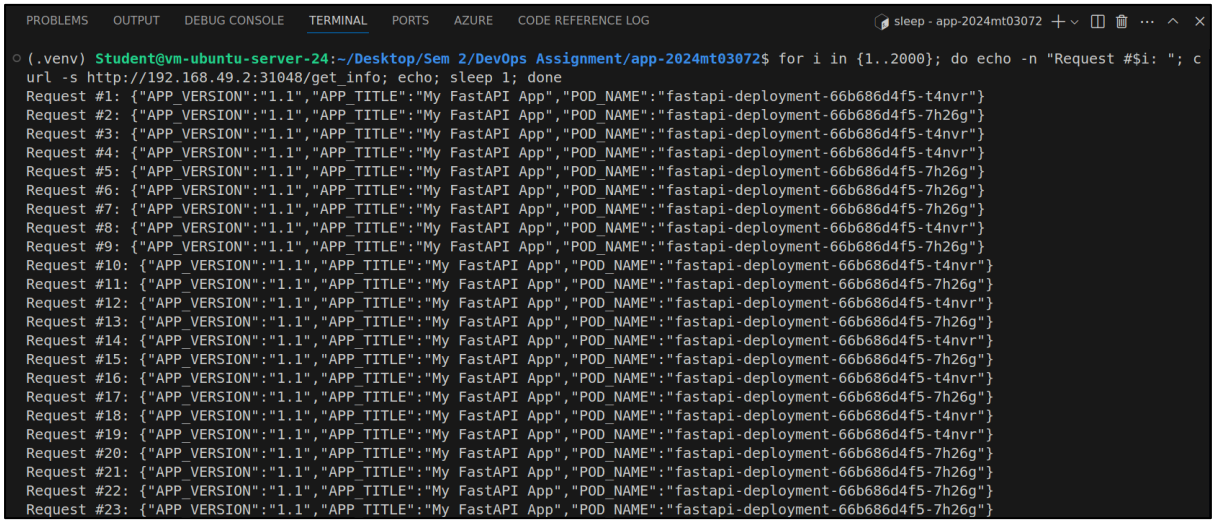
4. Load Balancer Testing

a. Using cURL to make several requests to /get_info endpoint.

b. The output clearly shows that requests are getting routed to both the replicas.

```bash
for i in {1..2000}; do echo -n "Request #$i: "; curl -s
http://192.168.49.2:31048/get_info; echo; sleep 1; done
```

4.1. Screenshot



5. Challenges Faced

a. No challenges faced while running the dockerized backend application over Google Chrome.

Task 6: Configure Prometheus for Metrics Collection

1. Prometheus Setup

a. Adding the package **prometheus_client** to requirements.txt file.

File link → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/requirements.txt

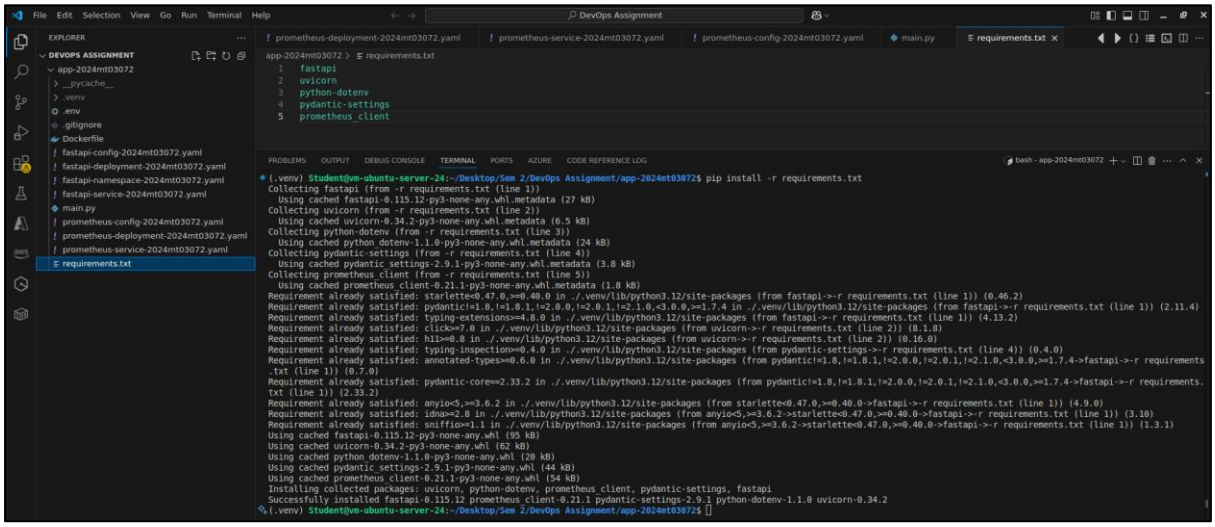
b. Installing prometheus_client by executing requirements.txt using pip install.

```bash
pip install -r requirements.txt
```

Note:

- **prometheus_client**: A Python client for exposing metrics in a format Prometheus can scrape for monitoring and alerting.

1.1. Screenshot



2. Integrating Prometheus in FastAPI Application

a. Integrating Prometheus in the FastAPI application to collect metrics by importing Prometheus client in main.py file.

File link → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/main.py

3. Prometheus ConfigMap Creation

a. Creating Prometheus ConfigMap **prometheus-config** using the ConfigMap configuration file **prometheus-config-2024mt03072.yaml**.
File → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/prometheus-config-2024mt03072.yaml.

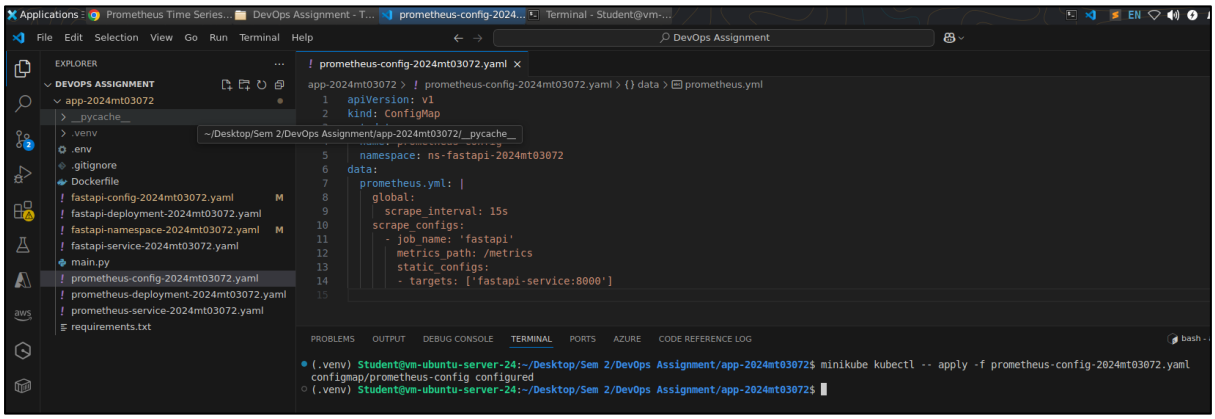
b. Applying the Prometheus ConfigMap.

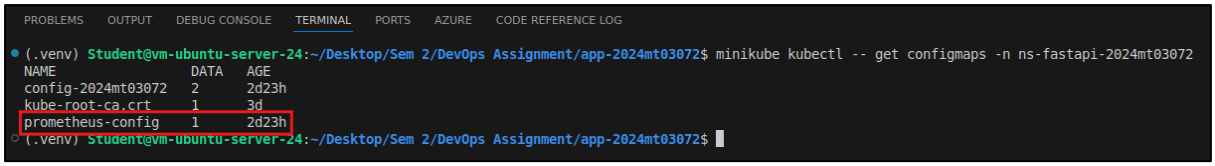
```bash
minikube kubectl -- apply -f prometheus-config-2024mt03072.yaml
```

c. Verifying the creation of the Prometheus ConfigMap prometheus-config.

```bash
minikube kubectl -- get configmaps -n ns-fastapi-2024mt03072
```

3.1. Screenshot





4. Prometheus Deployment Creation

a. Creating Prometheus Deployment **prometheus** using the Deployment configuration file **prometheus-deployment-2024mt03072.yaml**.

File → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/prometheus-deployment-2024mt03072.yaml.

b. Applying the Prometheus Deployment.

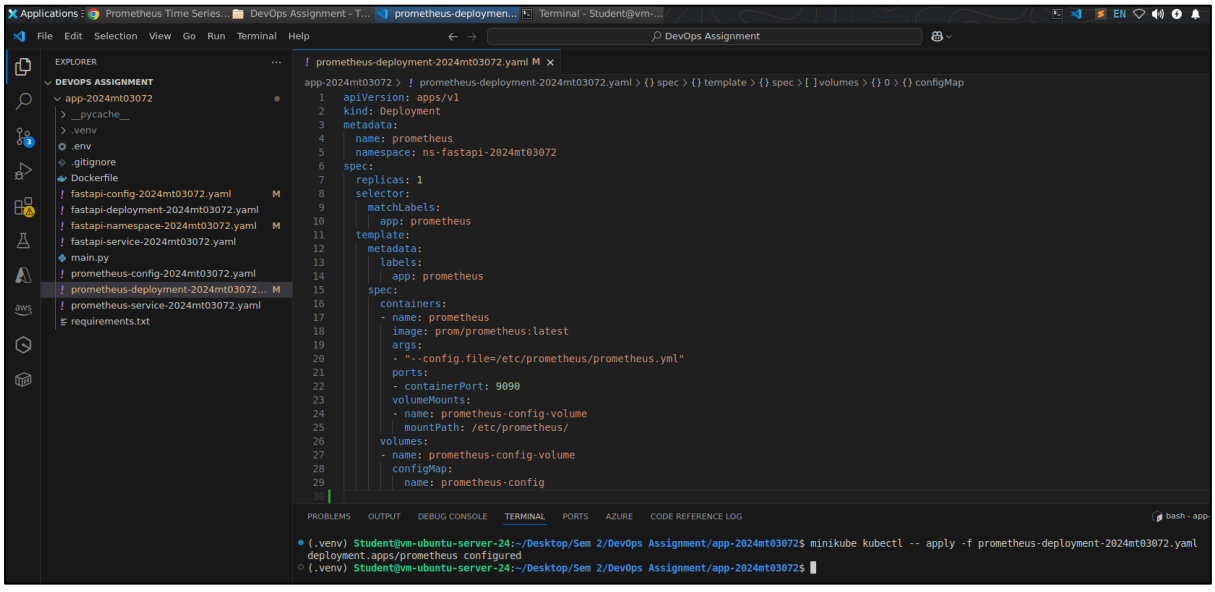
```bash
minikube kubectl -- apply -f prometheus-deployment-2024mt03072.yaml
```

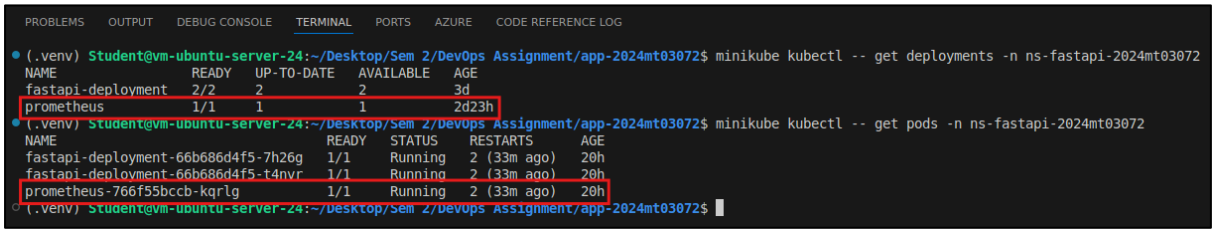
c. Verifying the creation of the Prometheus Deployment prometheus.

```bash
minikube kubectl -- get deployments -n ns-fastapi-2024mt03072
```

d. Listing all pods running inside Kubernetes Namespace ns-fastapi-2024mt03072.

```bash
minikube kubectl -- get pods -n ns-fastapi-2024mt03072
```





5. Prometheus Service Creation

a. Creating Prometheus Service **prometheus** using the Service configuration file **prometheus-service-2024mt03072.yaml**.

File → https://github.com/kaushikpaul90/bits-pilani-sem2-devops-app-2024mt03072/blob/main/app-2024mt03072/prometheus-service-2024mt03072.yaml.

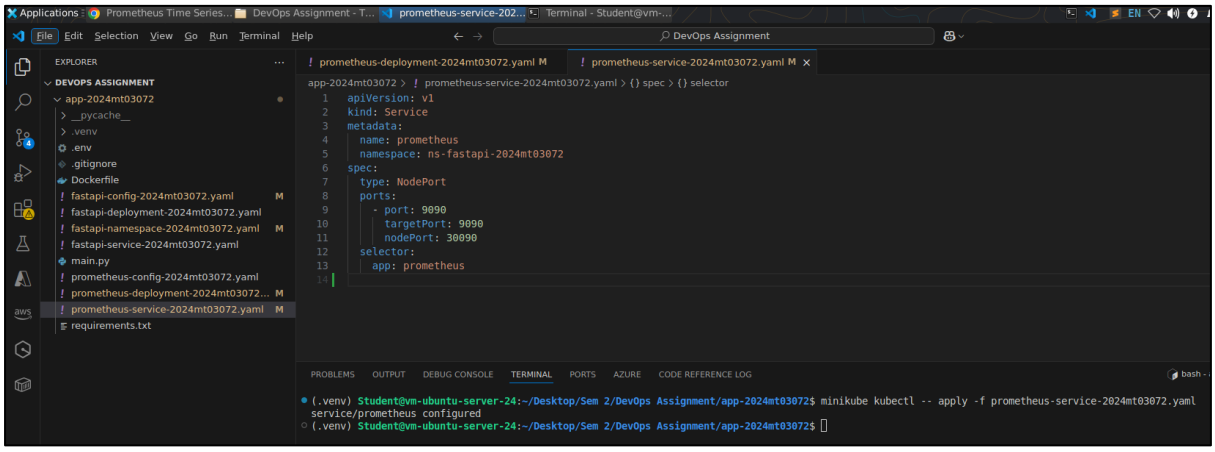
b. Applying the Kubernetes Service.

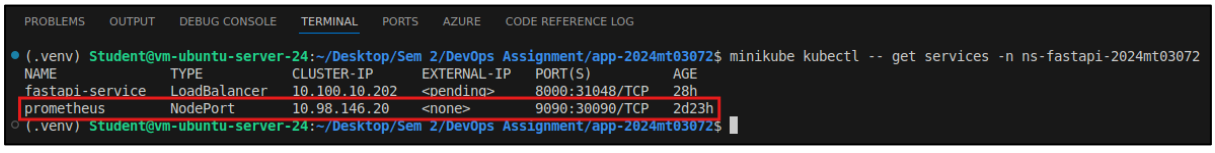
```bash
minikube kubectl -- apply -f prometheus-service-2024mt03072.yaml
```

c. Verifying the creation of the Kubernetes Service prometheus.

```bash
minikube kubectl -- get services -n ns-fastapi-2024mt03072
```

5.1. Screenshot



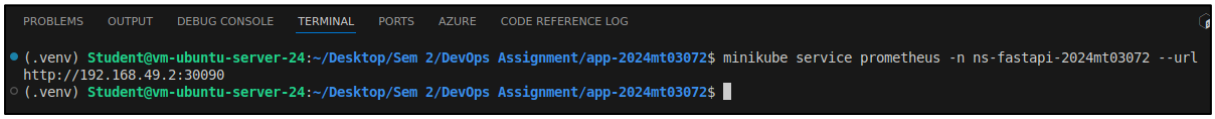


6. Obtaining Prometheus Access URL

a. Retrieving the URL to access Prometheus UI from browser.

```bash
minikube service prometheus -n ns-fastapi-2024mt03072 --url
```

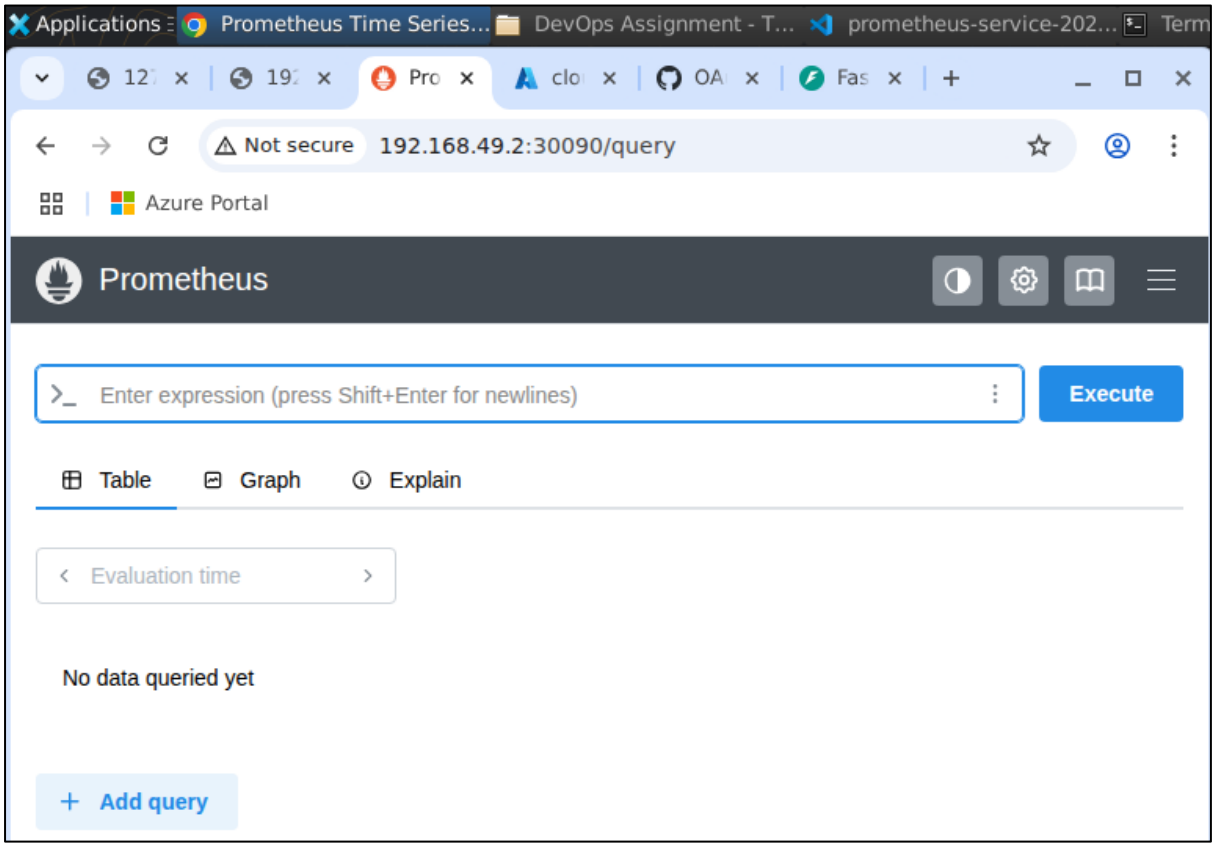
6.1. Screenshot



7. Accessing Prometheus UI

a. Accessing Prometheus UI via http://192.168.49.2:30090 on Google Chrome.

7.1. Screenshot



8. Prometheus Metrics Collection

a. Querying the number of requests received by the /get_info endpoint using metrics **get_info_requests_total**.

b. Querying CPU usage for each replica using metrics **container_cpu_usage_seconds_total**.

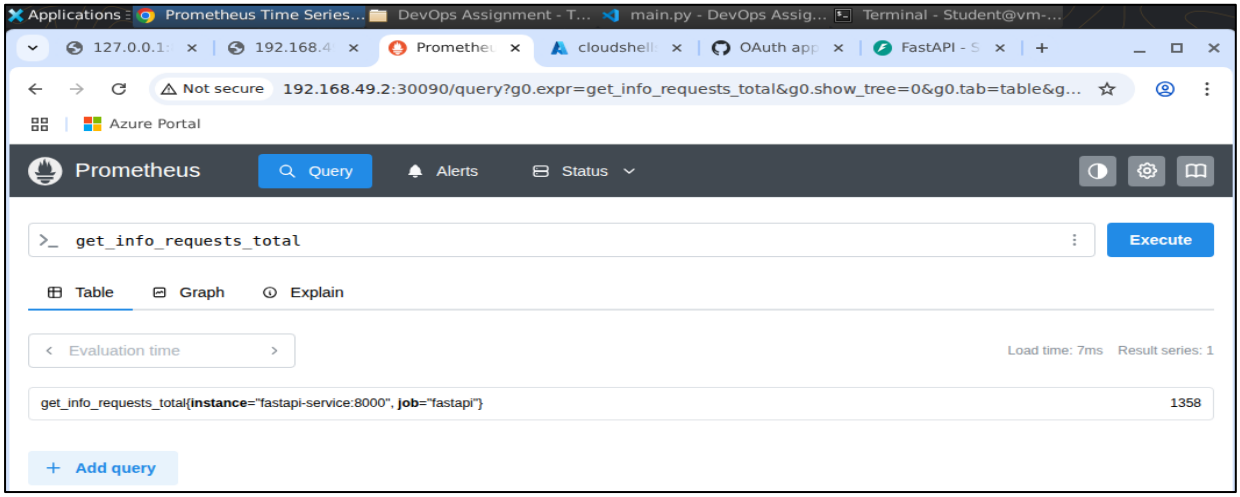
- **container_cpu_usage_seconds_total{namespace="ns-fastapi-2024mt03072"}** → This Prometheus query returns the total cumulative CPU

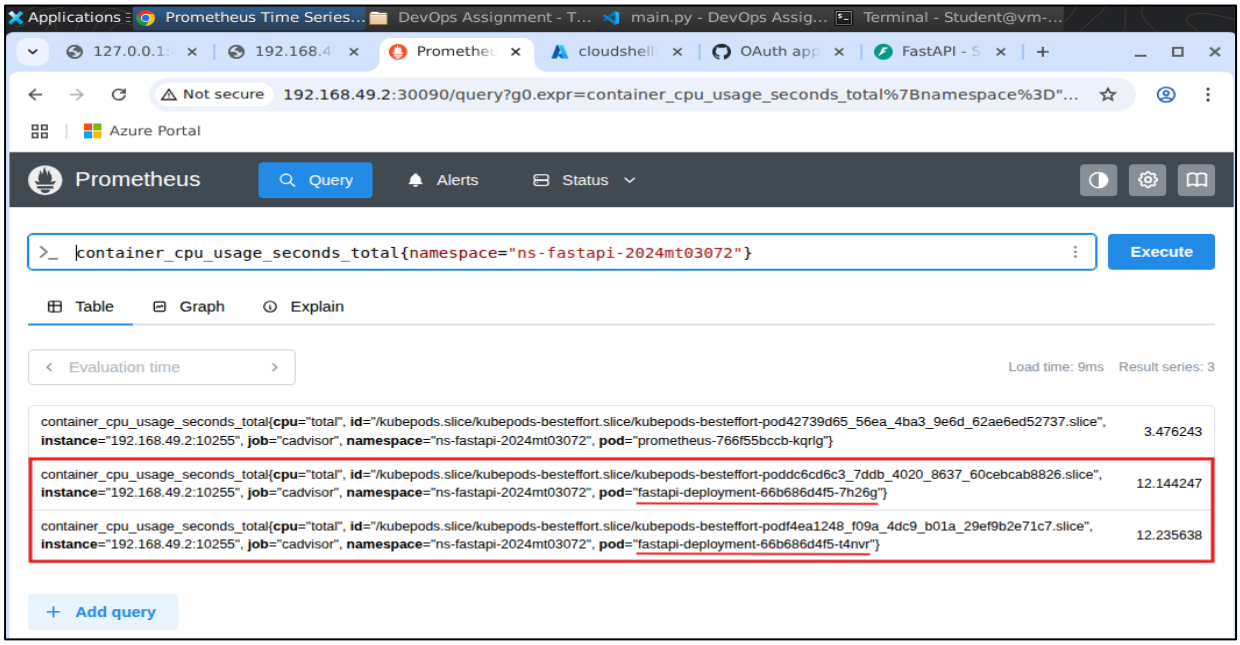
time (in seconds) consumed by each pod replica in the Kubernetes namespace ns-fastapi-2024mt03072 on the Minikube cluster.

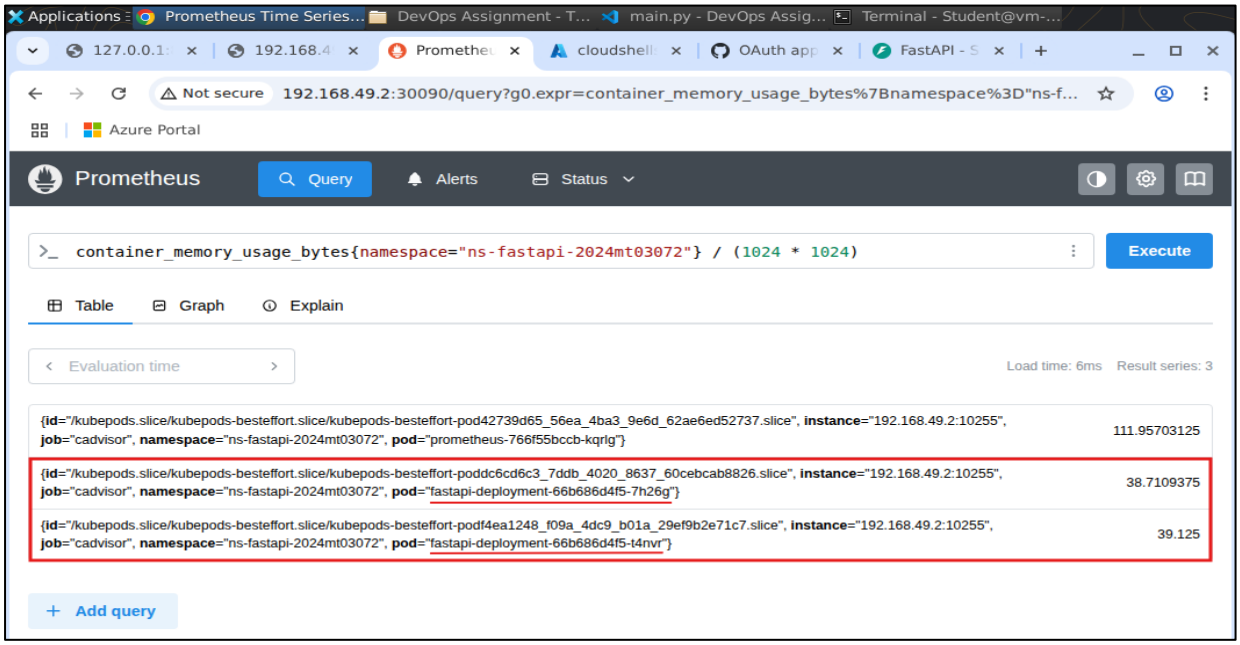
c. Querying memory usage for each replica using metrics **container_memory_usage_bytes**.

- **container_memory_usage_bytes{namespace="ns-fastapi-2024mt03072"} / (1024 * 1024)** → This Prometheus query returns the memory usage (in Mega Bytes) of each pod replica running in the Kubernetes namespace ns-fastapi-2024mt03072 on the Minikube cluster.

8.1. Screenshot







9. Challenges Faced

a. Metric Discovery Complexity:

- Identifying the correct Prometheus metrics (container_cpu_usage_seconds_total, container_memory_usage_bytes) to monitor CPU/memory usage per pod.
- After lot of research and analysis, I came across the metrics container_cpu_usage_seconds_total and container_memory_usage_bytes to monitor CPU and memory usage per pod respectively.

b. Scraping Configuration:

- Even after identifying the metrics, Prometheus was not able to collect them without proper scraping configuration.
- Adding cAdvisor job to prometheus-config-2024mt03072.yaml exposed container-level metrics for Prometheus to scrape, finally resolving the issue.

Conclusion

This assignment provided a comprehensive, hands-on experience with the complete DevOps lifecycle for a cloud-native backend application. Starting from the development of a FastAPI-based microservice, the project covered containerization with Docker, orchestrated deployment using Kubernetes, and advanced monitoring with Prometheus.

Key learnings included:

- **Environment Management:** Leveraging environment variables and ConfigMaps for flexible configuration across local and cloud environments.
- **Containerization:** Building and running Docker images and containers, ensuring portability and consistency.
- **Kubernetes Orchestration:** Deploying scalable, highly available applications with replica management, namespace isolation, and service exposure for load balancing.
- **Monitoring and Observability:** Integrating Prometheus to collect and visualize application and infrastructure metrics and overcoming the challenge of identifying and scraping the correct metrics for resource usage and request tracking.
- **Troubleshooting:** Addressing common issues such as image update propagation, service exposure, and metric visibility, which are crucial skills for real-world DevOps scenarios.

Overall, this assignment not only solidified theoretical concepts but also enhanced practical skills in deploying, scaling, and monitoring modern cloud applications. The experience gained through iterative debugging and configuration will be invaluable for future projects in cloud-native DevOps and site reliability engineering.