

Indian Institute of Technology Jodhpur

Fundamentals of Distributed Systems

Assignment– 1

Student Name: Avinash Kumar

Roll Number: G24AI2025.

Q.2 Dynamic Load Balancing for a Smart Grid.

Introduction:

In a future powered by electric vehicles (EVs), the demand on charging infrastructure will grow exponentially. Imagine dozens—or hundreds—of EVs trying to charge simultaneously across a city. If all requests are routed without intelligence, some substations may become overwhelmed while others remain underused. This can lead to grid instability, longer wait times, and inefficient resource utilization.

This project tackles that very challenge by simulating a smart, load-aware EV charging system. It acts like a traffic controller for electricity—monitoring real-time load across substations and ensuring that each new request is routed to the least loaded substation.

What makes this system intelligent is:

- Its use of live monitoring (via Prometheus),

- A custom-built load balancer that makes real-time routing decisions,
- And a visual dashboard (Grafana) to observe how load changes as requests flow in.

The entire architecture is written in Python, containerized using Docker, and orchestrated with Docker Compose to mirror a true distributed, scalable system.

2.Objective:

Build a system that can intelligently balance EV charging requests in real time, so no single substation gets overloaded.

But this wasn't just about writing some routing logic. I wanted to simulate a real-world grid-like environment where:

- Charging stations (substations) experience real-time usage variations.
- A central decision-making unit (load balancer) polls and reacts dynamically.
- The system is observable, meaning we can see exactly how traffic is flowing, how load is shifting, and whether the logic is working under pressure.

The vision was to create an end-to-end microservice architecture where:

- Each service had a clear, independent role,
- The system could handle rush hours,
- And the entire infrastructure was modular and scalable through Docker.

3. System Architecture

The architecture is made up of six distinct services working together:

Component	Purpose
charge_request_service	Entry point for all charging requests
load_balancer	Acts as the brain: polls loads and routes requests intelligently
substation_service	Simulates real-world charging and exposes real-time load metrics
load_tester	Generates a flood of charging requests to simulate traffic surges
prometheus	Continuously scrapes metrics from substations for monitoring
grafana	Provides real-time visualization of how the system is performing

Each substation works as an independent microservice. When a request arrives:

1. It goes to the **charge_request_service**.
2. The request is then forwarded to the **load balancer**.
3. The balancer consults all **substation metrics** and routes the request to the **least busy** one.

4. Prometheus scrapes the latest loads.
5. Grafana visualizes the whole system in real time.

4. Screenshots.

The below screenshot shows

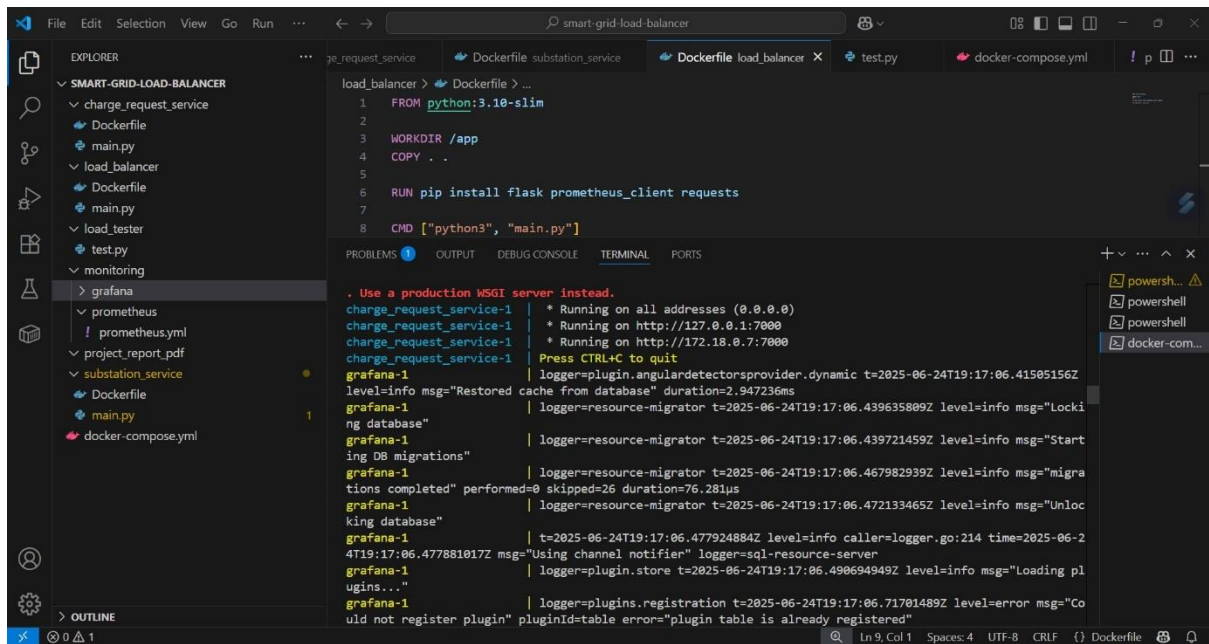
The left pane reveals a well-organized folder structure with clearly separated services: `charge_request_service`, `load_balancer`, `substation_service`, and `monitoring`.

The Dockerfile for `load_balancer` is visible, specifying Python 3.10 and dependencies like `flask`, `prometheus_client`, and `requests`.

The terminal confirms that:

- `charge_request_service` is live and listening on port 7000 (internally and externally).
- `grafana-1` is booting successfully and migrating the internal database.

Prominent log messages show that Grafana is restoring from cache, migrating plugins, and initializing correctly.



The below screenshot says that

The terminal logs are from substation1-1.

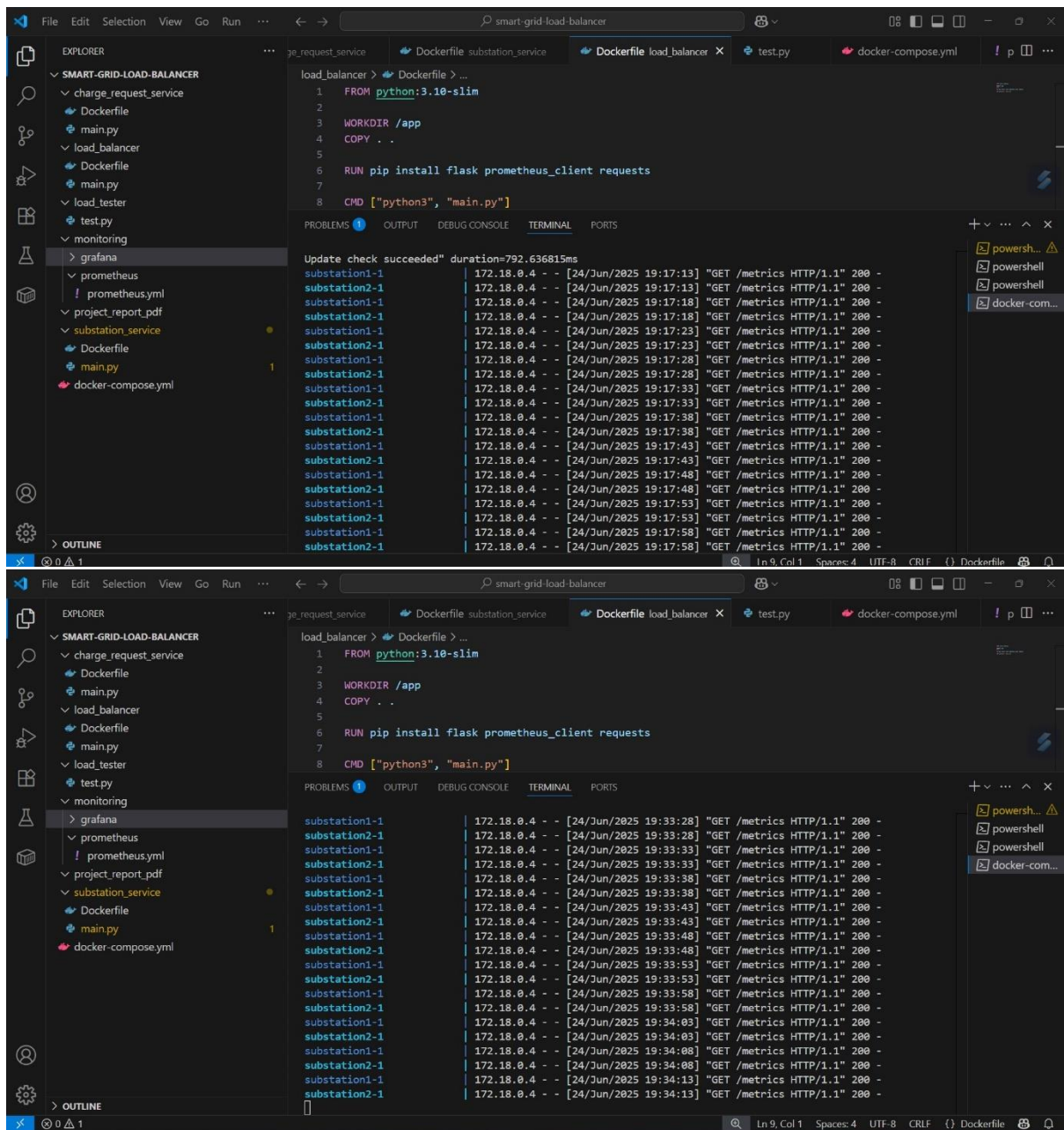
Prometheus is sending repeated GET /metrics requests.

The HTTP responses are all 200 OK, indicating successful retrieval.

Its also validating

- The observability pipeline is functional.
- Prometheus is collecting real-time load data from each substation.
- The load balancer and Grafana can now use this data to:
 - Make routing decisions
 - Display live dashboards

Without this metric flow, the system would be blind to substation states. This screenshot proves full integration between application logic and monitoring stack.



The below screenshot says

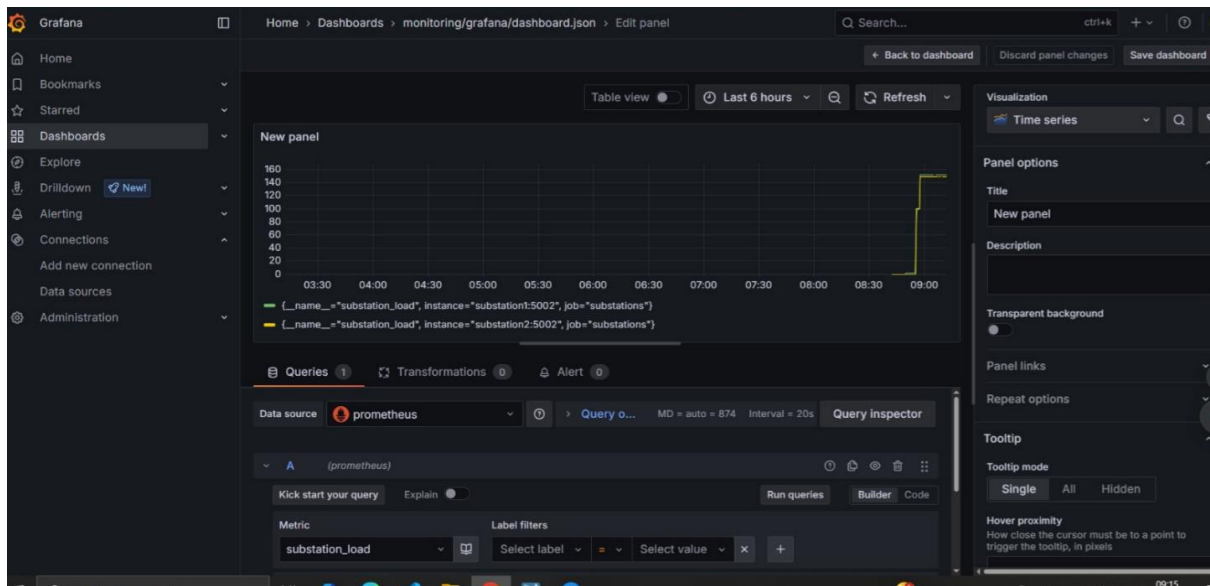
The query `substation_load` (bottom section) is fetching live metrics from Prometheus.

The line graph shows a **step-wise increase in load**, as charging requests are distributed across substations (substation1 and substation2).

Two colored lines:

- **Yellow:** Substation 1
- **Green:** Substation 2

You can clearly observe when a **rush of requests** was received, leading to sharp upward curves.



Video Link :

<https://drive.google.com/file/d/14-8PPYa8LtkCgItV0L9jo-QbGuwOAQqV/view?usp=sharing>