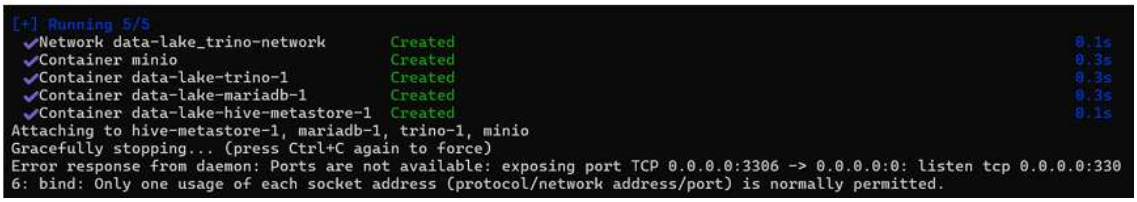
Ex No: 5 Date: 24/9/25	Hive, Trino and Minio

Step 1: Set Up Docker Environment

- Install Docker and Docker Compose on your machine.
- Create a Docker Compose file defining containers for:
- Minio (object storage)
- Trino (distributed query engine)
- MariaDB (metadata or transactional database)
- Hive Metastore (metadata service)
- Configure network settings to allow containers to communicate.

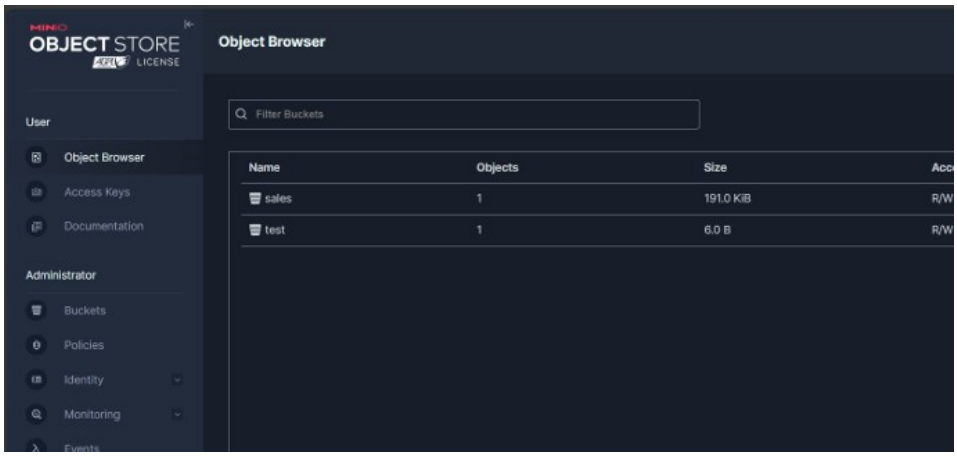
Step 2: Start the Docker Containers

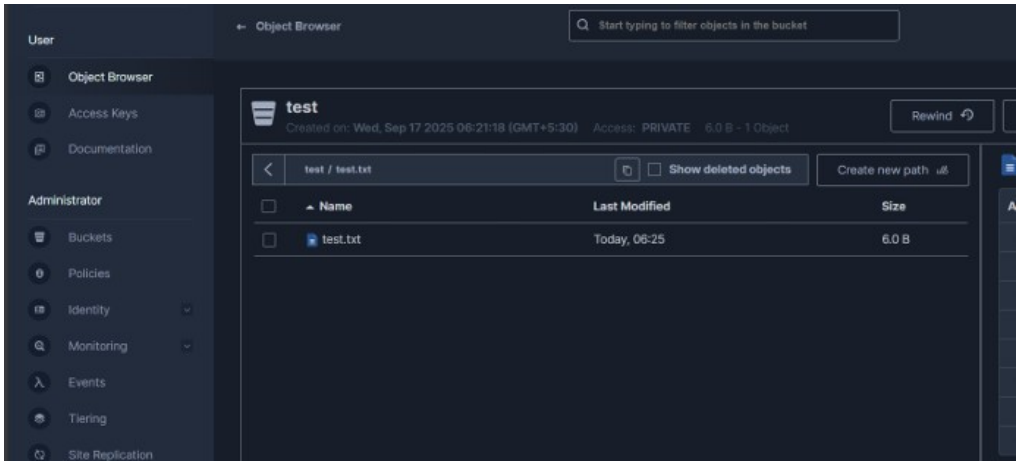
- Run docker-compose up -d to start all the containers.
- Ensure containers start successfully and no port conflicts occur (address port 3306 conflicts if any).
- Use docker ps to see running containers.



Step 3: Set Up Minio Object Storage

- Access Minio web UI via port 9000.
- Create buckets like "sales" and "test".
- Upload your dataset files (CSV, Parquet, etc.) into buckets.





Step 4: Configure Hive Metastore

- Configure Hive Metastore to connect with MariaDB for metadata management.
- Define necessary schemas and tables in Hive pointing to data locations in Minio.

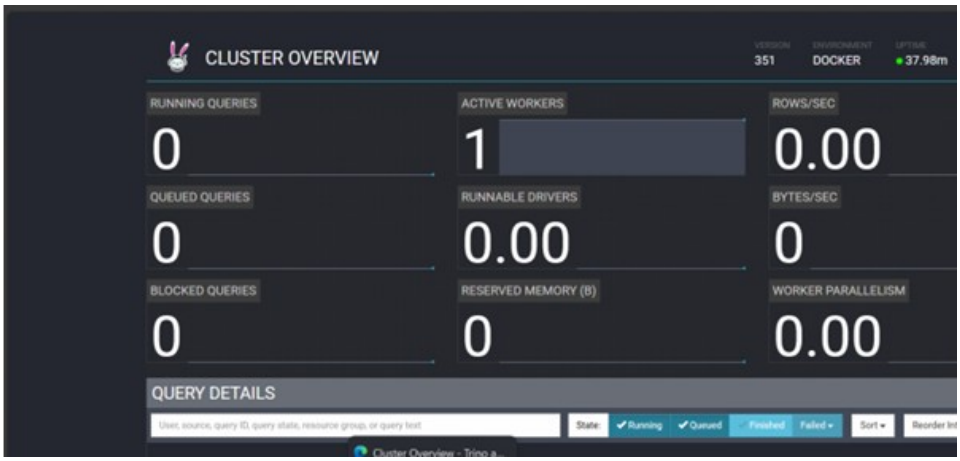
Step 5: Configure Trino

- Configure Trino catalog to connect to Minio's data using an object storage connector (e.g., Hive or Iceberg connector).
- Configure Trino to connect to Hive Metastore and MariaDB for metadata.

Configure the catalog properties for accessing Minio buckets and schema metadata.

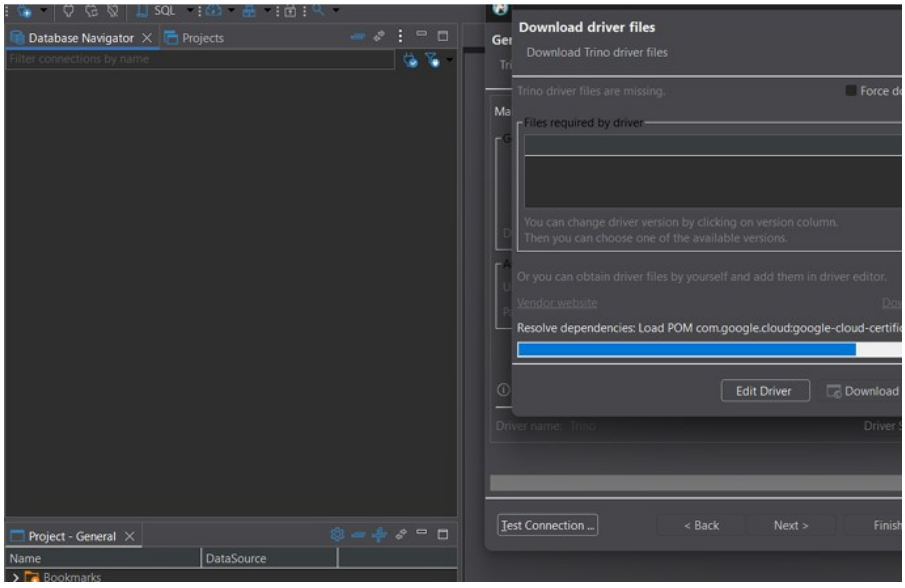
Step 6: Query Data Using Trino

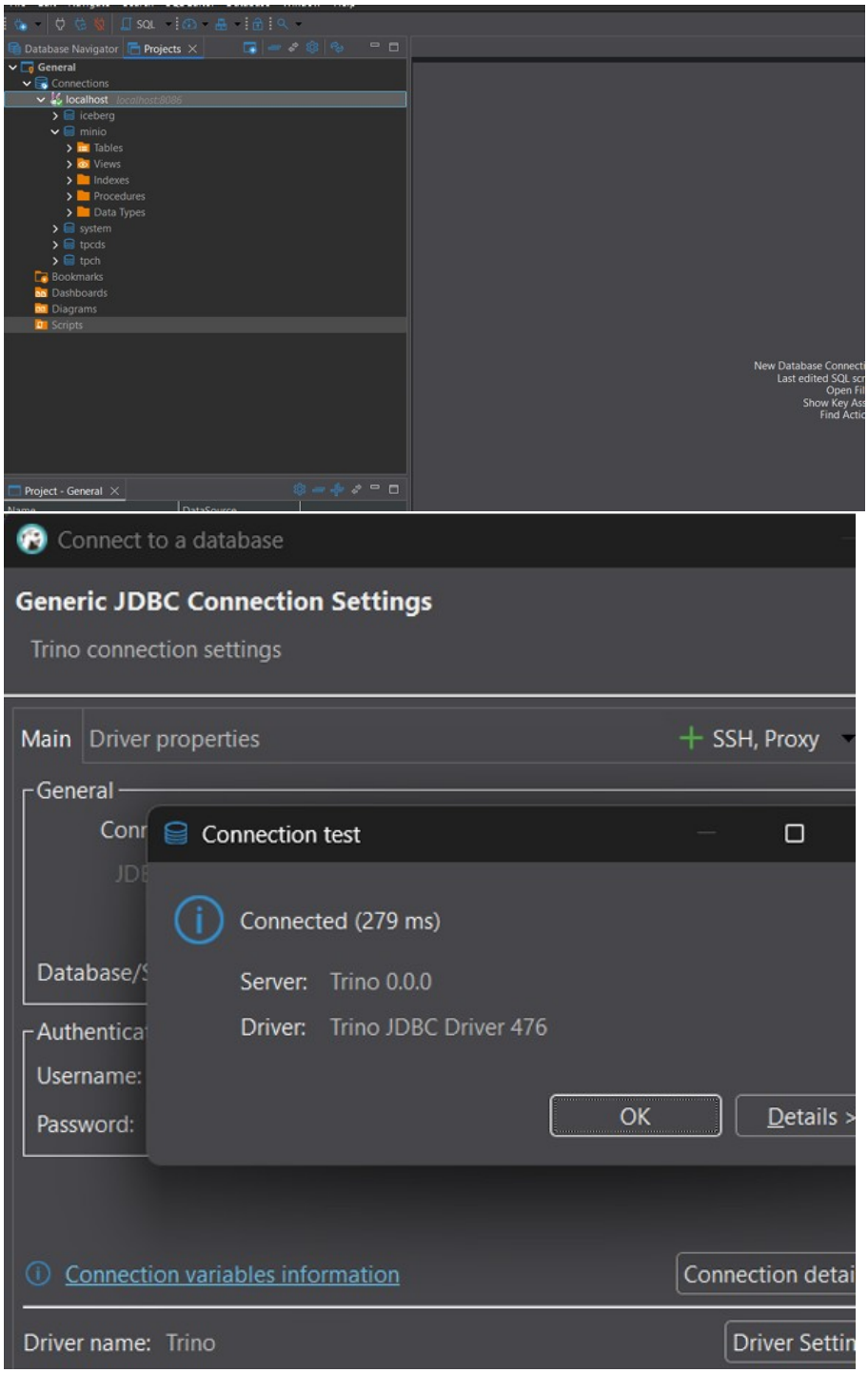
- Use Trino CLI or a SQL client connected to Trino (e.g., DBeaver).
- Run SQL queries to explore data, e.g.:
- Analyze sales, filter by category, country, or other attributes.



Step 7: Manage and Monitor

- Monitor container logs and performance.
- Manage data buckets and metadata schemas.
- Scale or update containers as needed.





This setup creates a distributed data lake solution with a scalable query engine, enabling complex queries over large datasets stored in object storage with managed metadata.