

# Cloud Application Development

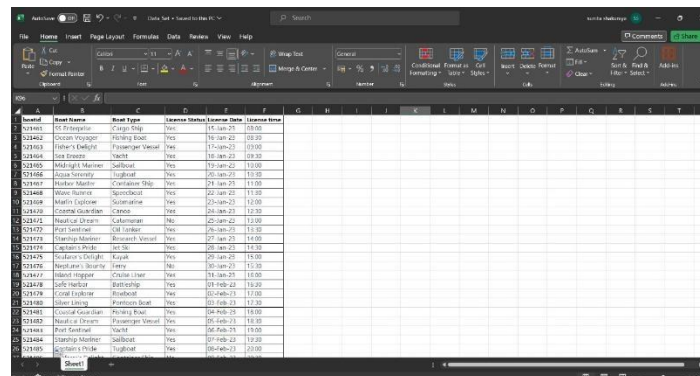
## Data Warehousing with IBM Cloud Db2 Warehouse

### Phase4

#### To Do:

*Building the data warehouse by implementing ETL processes and enabling data exploration. Implement ETL processes to extract, transform, and load data into the data warehouse. Enable data architects to explore and analyse data within Db2 Warehouse using SQL queries and analysis techniques.*

#### Implementation:



BoatID	Boat Name	Boat Type	License Status	License Date	License Time
123456	Blue Skimmer	Cargo Ship	Yes	15-Jan-23	10:00
123457	Orange Voyager	Fishing Boat	Yes	16-Jan-23	10:30
123458	Pinkie's Delight	Passenger Vessel	Yes	17-Jan-23	11:00
123459	Sea Explorer	Yacht	Yes	18-Jan-23	11:30
123460	Midnight Mariner	Selfboat	Yes	19-Jan-23	12:00
123461	Star Navigator	Lightboat	Yes	20-Jan-23	12:30
123462	Whisper Machine	Construction Ship	Yes	21-Jan-23	13:00
123463	Storm Navigator	Speedboat	Yes	22-Jan-23	13:30
123464	Marlin Explorer	Scuba diver	Yes	23-Jan-23	14:00
123465	Tranquil Explorer	Catamaran	Yes	24-Jan-23	14:30
123466	Harmonic Explorer	Catamaran	No	25-Jan-23	15:00
123467	Star Sailboat	Off Center	Yes	26-Jan-23	15:30
123468	Flowing Mariner	Research Vessel	Yes	27-Jan-23	16:00
123469	Captain's Ship	Jet Ski	Yes	28-Jan-23	16:30
123470	Velocity Explorer	kayak	Yes	29-Jan-23	17:00
123471	Neptune's Bounty	Tender	No	30-Jan-23	17:30
123472	Harbor Skipper	Cruise Boat	Yes	31-Jan-23	18:00
123473	Seaside Skipper	Boatbridge	Yes	01-Feb-23	18:30
123474	Grand Explorer	Powerboat	Yes	02-Feb-23	19:00
123475	Ocean Explorer	Powerboat	Yes	03-Feb-23	19:30
123476	Crystal Explorer	Private Boat	Yes	04-Feb-23	19:59
123477	Nautica Dream	Passenger Vessel	Yes	05-Feb-23	19:59
123478	Star Sailboat	Yacht	Yes	06-Feb-23	19:59
123479	Starship Mariner	Selfboat	Yes	07-Feb-23	19:59
123480	Seaside Skipper	Lightboat	Yes	08-Feb-23	20:00

Figure 1 DATA CREATION

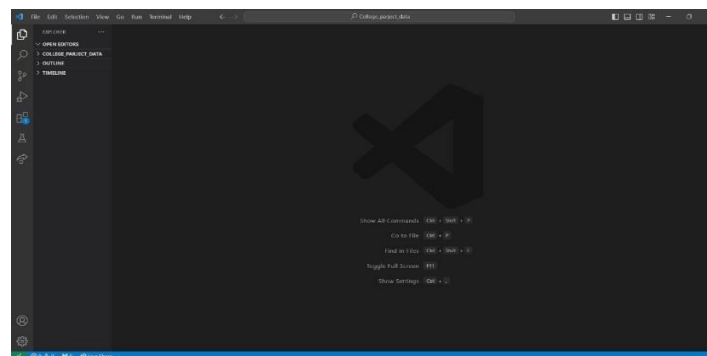


Figure 2 OPEN VISUAL STUDIO

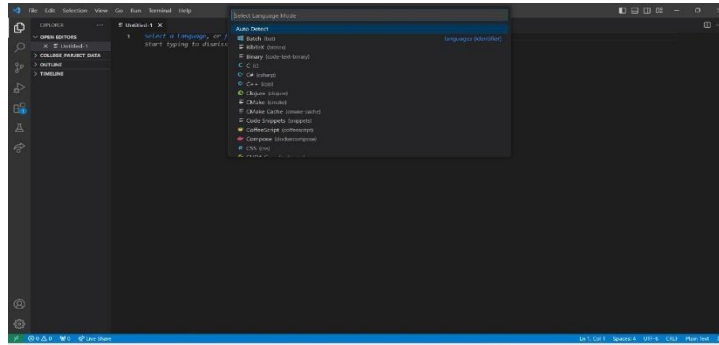


Figure 2.1

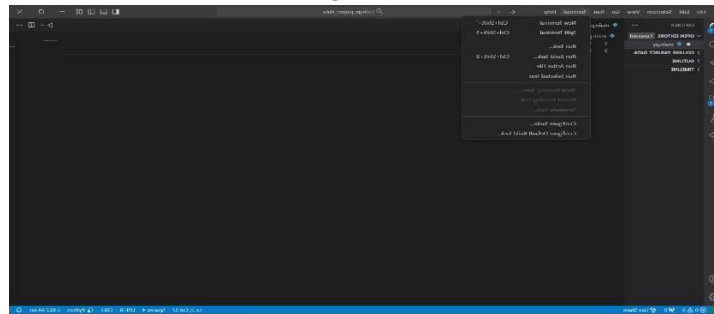


Figure 2.2

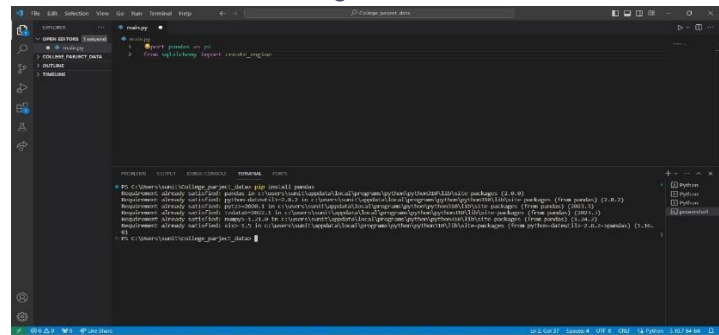


Figure 2.3

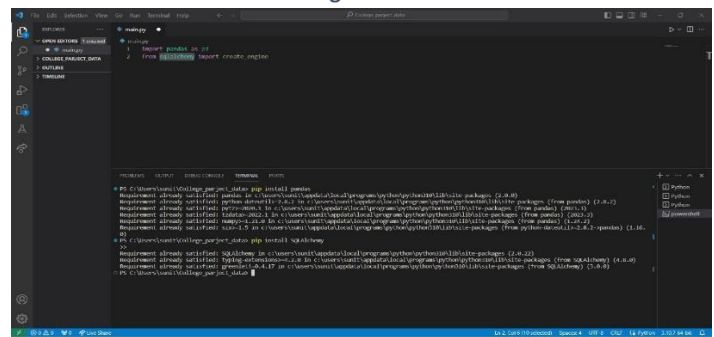


Figure 2.4

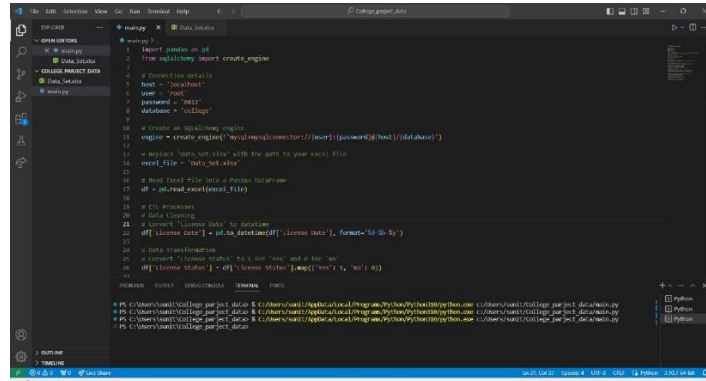


Figure 2.5 PANDAS CODING

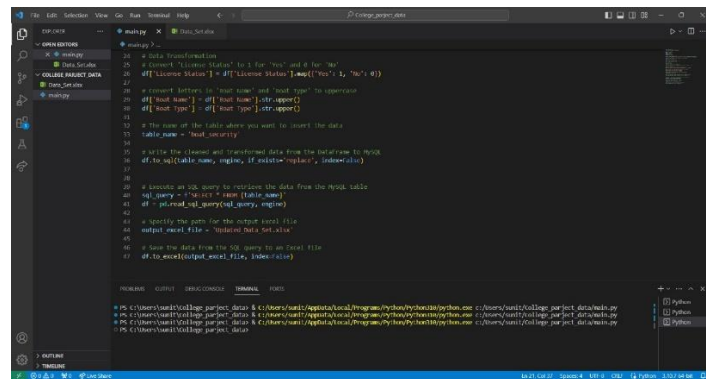


Figure 2.6 PANDAS CODING END

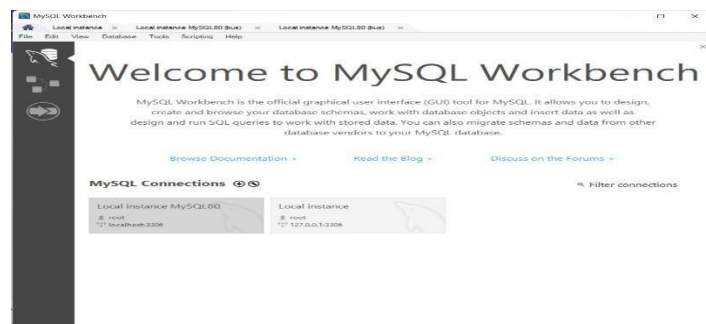


Figure 3 OPEN MYSQL

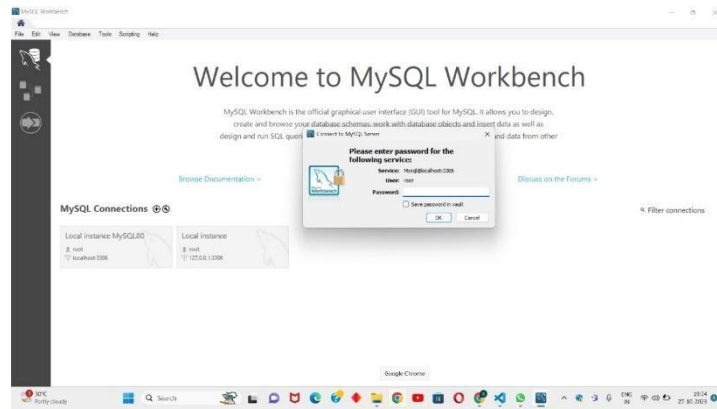


Figure 3.1

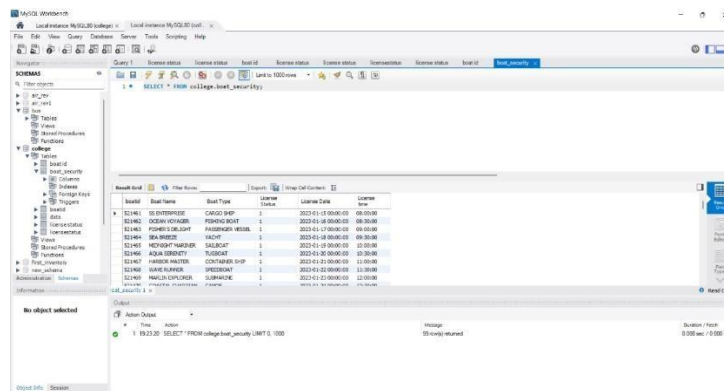


Figure 3.2 AFTER RUNNING THE CODE IN MYSQL SERVER

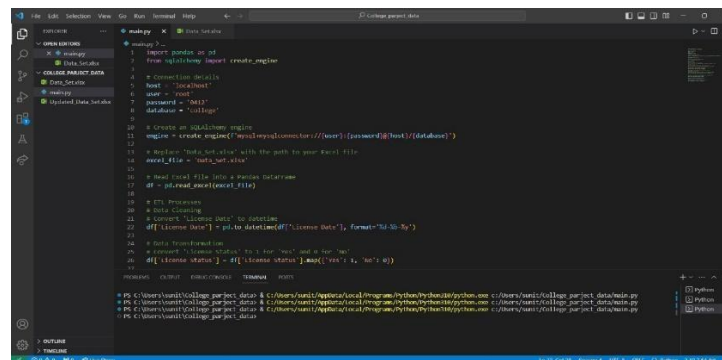


Figure 3.3 UPDATED DATA FILE GENERATED (SEE LEFT SIDE)

```

import pandas as pd
import sqlalchemy as sa

# Connect to the database
engine = sa.create_engine('mysql://root:root@localhost:3306/warehouse')

# Query the data
sql_query = 'SELECT * FROM test_data'
df = pd.read_sql_query(sql_query, engine)

# The name of the table where you want to insert the data
table_name = 'test_data'

# Write the (columns and transformed data from the database to) to a local file
df.to_excel('output_data.xlsx', index=False)

# Save the data from the local file to a local file
df.to_excel('output_data.xlsx', index=False)

```

Figure 3.4 ENDING

Rowid	Boat Name	Boat Type	Cruise Start	Cruise Date	Cruise Name
1	123456 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
2	123457 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
3	123458 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
4	123459 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
5	123460 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
6	123461 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
7	123462 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
8	123463 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
9	123464 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
10	123465 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
11	123466 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
12	123467 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
13	123468 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
14	123469 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
15	123470 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
16	123471 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
17	123472 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
18	123473 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
19	123474 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
20	123475 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
21	123476 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
22	123477 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
23	123478 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
24	123479 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
25	123480 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
26	123481 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00
27	123482 DEEP SEA DIVER	CRUISE SHIP	1	2023-01-15 00:00:00	00:00:00

Figure 4 UPDATED DATA AFTER RUNNING ETL PROCESS OVER THA DATA USING PANDAS

## Project Goals:

### Building the Data Warehouse:

The primary goal was to create a data warehouse infrastructure using IBM Db2 Warehouse.

### Implementing ETL Processes:

We aimed to establish efficient ETL processes to extract, transform, and load data into the data warehouse.

**Eg :**

Extract data from a source (e.g., CSV file)

```
INSERT INTO TargetTable (Column1, Column2, Column3)
```

```
SELECT SourceColumn1, SourceColumn2, SourceColumn3
```

```
FROM SourceCSV;
```

### **Enabling Data Exploration:**

The project aimed to provide data architects with the tools and capabilities to explore and analyze data within Db2 Warehouse using SQL queries and analysis techniques.

### **Basic SQL Query:**

Retrieve data from a table

```
SELECT Column1, Column2
```

```
FROM WarehouseTable
```

```
WHERE Condition = 'Value';
```

### **Joining Tables:**

Join multiple tables for more complex queries

```
SELECT W.ColumnA, T.ColumnX
```

```
FROM WarehouseTable W
```

```
INNER JOIN AnotherTable T ON W.ID = T.ID;
```

### **Aggregation and Analysis:**

Perform aggregate functions for analysis

```
SELECT Year, SUM(Sales) AS TotalSales
```

```
FROM SalesData
```

```
GROUP BY Year;
```

## Project Milestones and Achievements

### 1. Data Warehouse Implementation

Successfully deployed IBM Db2 Warehouse, providing a scalable platform for data storage and management.

### 2. ETL Process Implementation

Designed and implemented ETL processes that automate data extraction from various sources, perform necessary transformations, and load data into the warehouse.

Achieved data integration across different systems, ensuring a unified and consistent data source.

### 3. Enabling Data Exploration

Provided data architects with access to Db2 Warehouse, including necessary permissions and credentials.

Facilitated the use of SQL queries and data analysis techniques, empowering data architects to explore the data effectively.

## Conclusion

In conclusion, our hands-on experience with IBM cloud, DB2 and Watson Studio has been both enlightening and informative. These platforms have offered us a valuable glimpse into the world of cloud computing and data science. However, it is worth noting that we encountered certain limitation in performing ETL (Extract, Transform, Load) processes over cloud services.

However, it has become evident that proceeding with ETL (Extract, Transform, Load) processes over cloud services presents certain limitations and challenges. As a results, we have made the strategic decision to transition towards using Pandas and MySQL for our ETL processes. This shift allows us to work with greater flexibility and control, ensuring seamless data transformations and management. By combining the capabilities of Pandas and MySQL, we are better equipped to tackle the intricacies of data processing and analysis, setting the stage for future development and data-driven projects. This decision marks a pivotal moment in out journey, as we adapt and evolve in response to the ever-changing landscape of data science and technology.