Experiment No.6
Data Visualization using Hive.
Date of Performance:
Date of Submission:

Aim: Data Visualization using Hive.

Theory:

Hive has a fascinating history related to the world's largest social networking site: Facebook. Facebook adopted the Hadoop framework to manage their big data. If you have read our previous blogs, you would know that big data is nothing but massive amounts of data that cannot be stored, processed, and analyzed by traditional systems.

Architecture of Hive

The architecture of the Hive is as shown below. We start with the Hive client, who could be the programmer who is proficient in SQL, to look up the data that is needed.

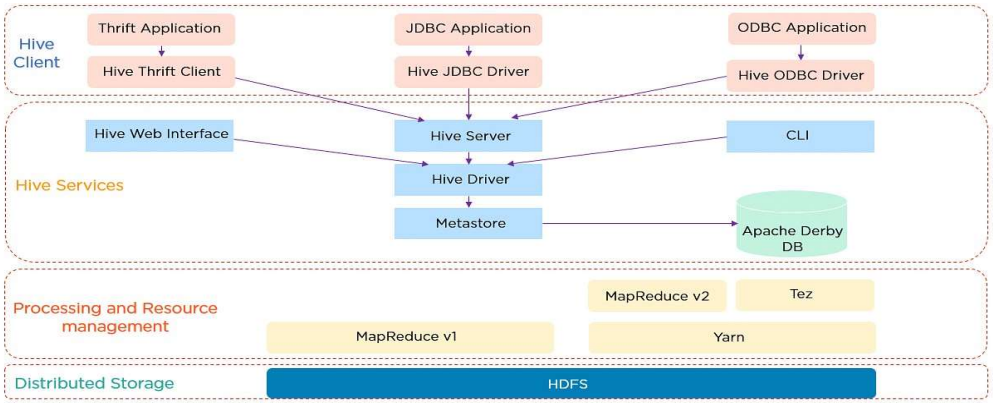


Fig: Architecture of Hive

The Hive client supports different types of client applications in different languages to perform queries. Thrift is a software framework. The Hive Server is based on Thrift, so it can serve requests from all of the programming languages that support Thrift.

The data flow in the following sequence:

1. We execute a query, which goes into the driver
2. Then the driver asks for the plan, which refers to the query execution
3. After this, the compiler gets the metadata from the metastore
4. The metastore responds with the metadata
5. The compiler gathers this information and sends the plan back to the driver
6. Now, the driver sends the execution plan to the execution engine
7. The execution engine acts as a bridge between the Hive and Hadoop to process the query
8. In addition to this, the execution engine also communicates bidirectionally with the metastore to perform various operations, such as create and drop tables
9. Finally, we have a bidirectional communication to fetch and send results back to the client.

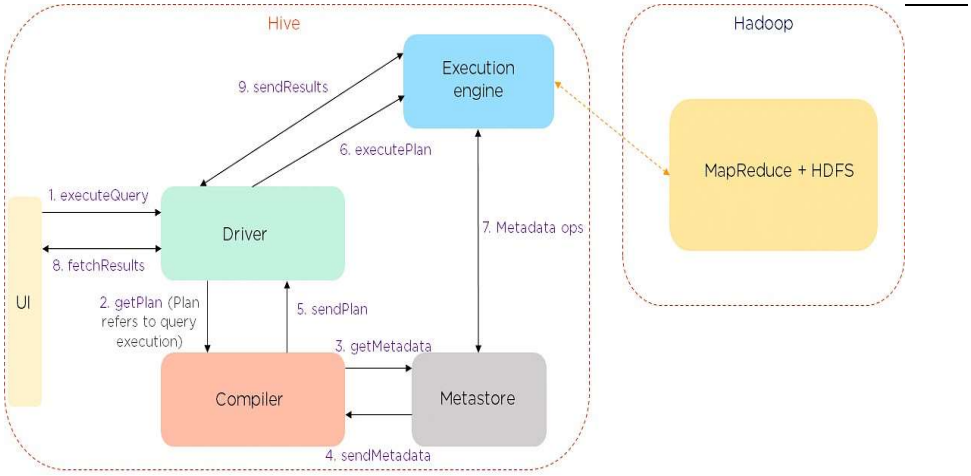
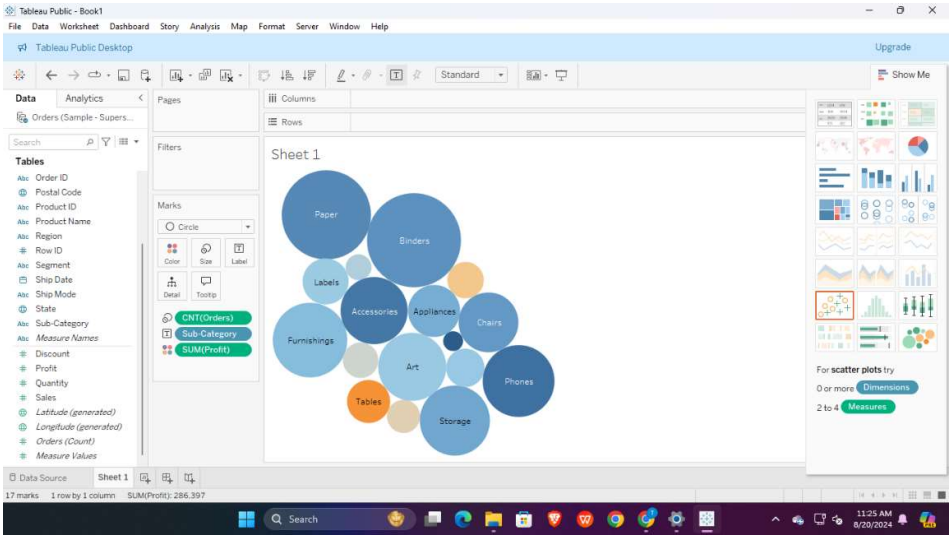


Fig: Data flow in Hive

Implementation:



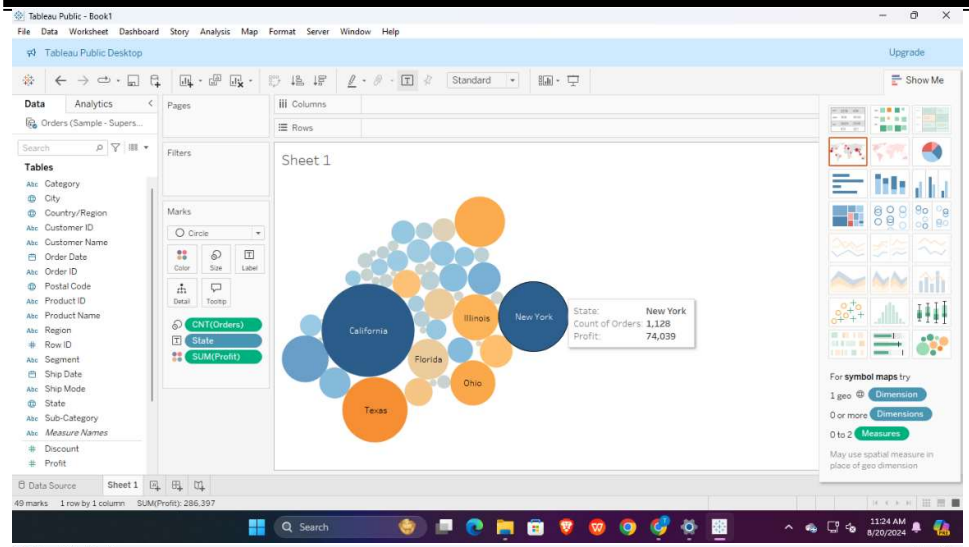


Tableau Public - Book1
Sheet 1
California
Illinois
New York
Florida
Ohio
Texas
State: New York
Count of Orders: 1,128
Profit: 74,039

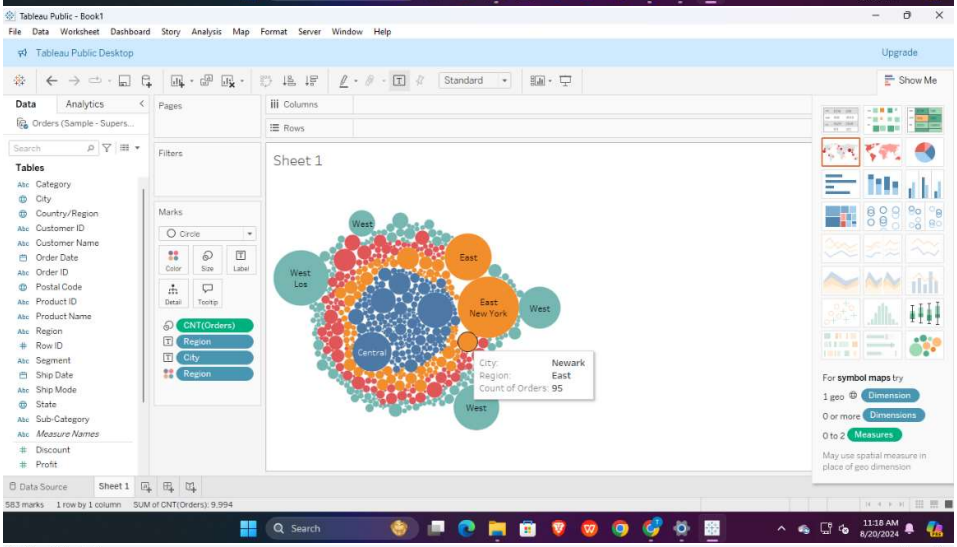


Tableau Public - Book1
Sheet 1
West
East
West Los
East New York
West
Central
City: Newark
Region: East
Count of Orders: 95

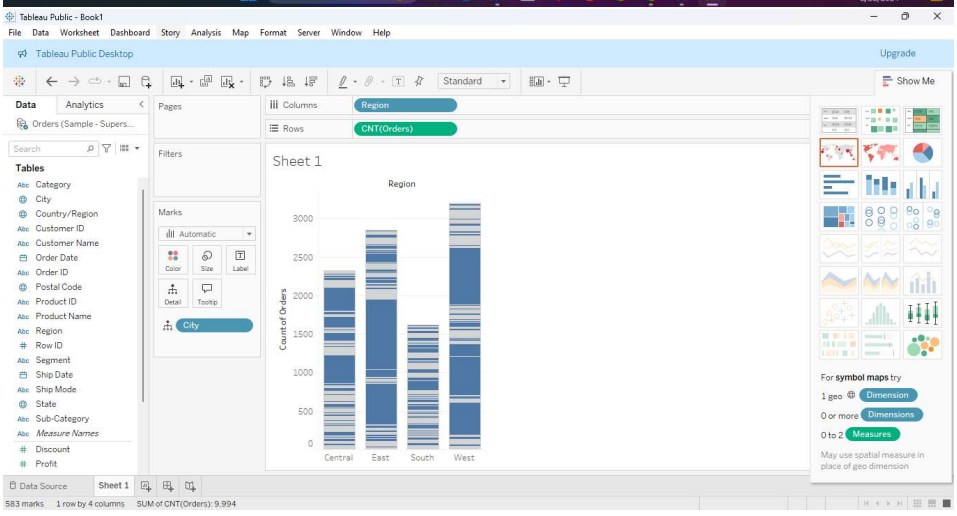
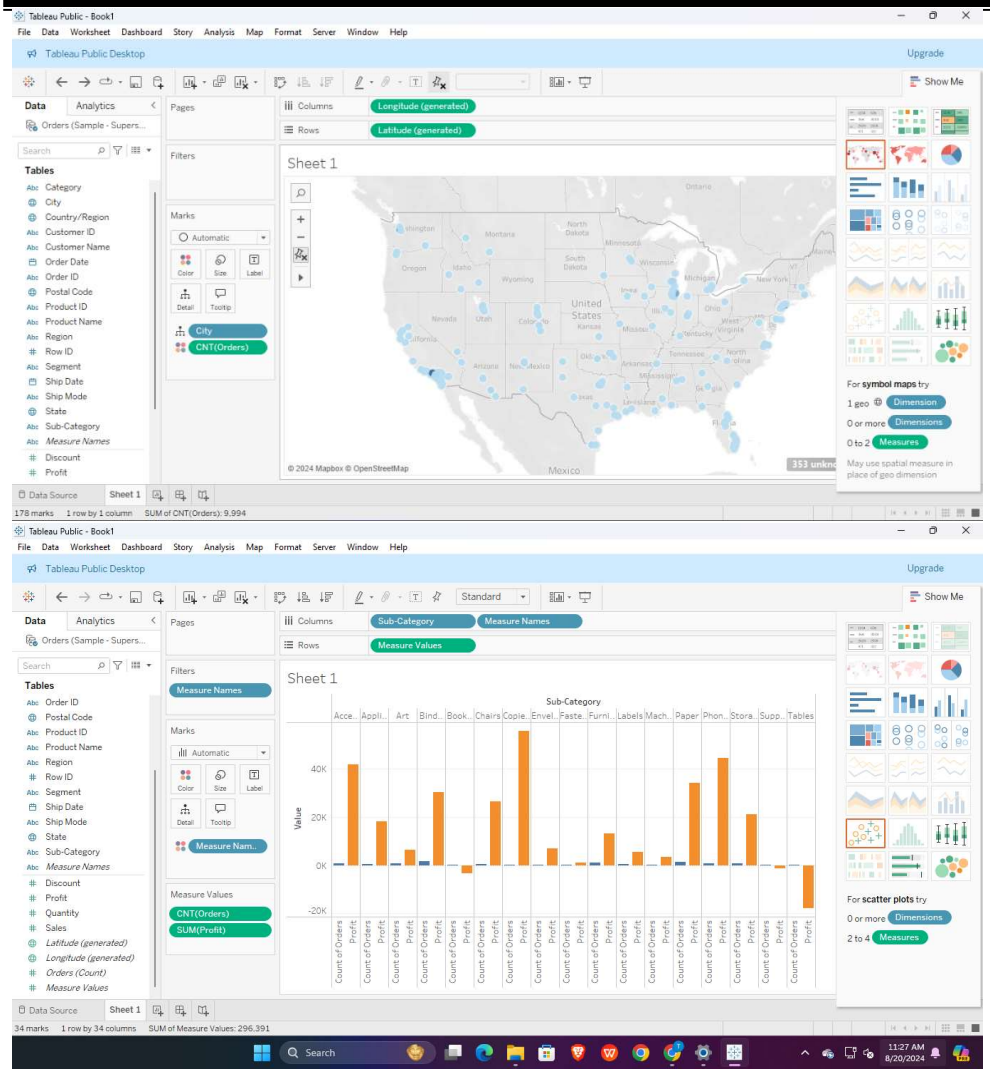


Tableau Public - Book1
Sheet 1
Region
Count of Orders
Central
East
South
West



Conclusion:

Explain Datatypes in Hive

Hive supports a variety of data types that are categorized into primitive and complex types. **Primitive data types** include numeric types like `TINYINT`, `SMALLINT`, `INT`, `BIGINT`, `FLOAT`, `DOUBLE`, and `DECIMAL`, which handle different kinds of numbers. String types include `STRING`, `VARCHAR`, and `CHAR`, which represent variable or fixed-length character data. Date and time types such as `DATE`, `TIMESTAMP`, and `INTERVAL` deal with temporal data, while `BOOLEAN` represents true/false values and `BINARY` is used for binary data. **Complex data types** allow for more structured data, including `ARRAY` for ordered collections, `MAP` for key-value pairs, `STRUCT` for grouping fields, and `UNIONTYPE` for handling values that can be of different types. These data types enable Hive to efficiently store and process structured data in distributed environments, making it a powerful tool for large-scale data management and analysis.