**Batch T4**

**Practical No. 6**

**Title of Assignment :**

To design and implement a data warehouse for a customer order processing system in a company. [ Use MySQL Database ]

**Student Name: Parshwa Herwade**

**Student PRN: 22510064**

Specifications :

The target of data warehouse system is an enterprise that consists of a number of

stores located in different cities and states. Each store holds a variety of items in

various quantity. In addition, the enterprise keep the information of the customers.

There are two kinds of customers: walk-in led by tourism guide and mail-order by

post address inclusive. The city location of the customer, together with the data of the

customer’s first order, is stored by the existing system. Each customer lives in one

city only, and the enterprise will try to satisfy the customer’s order items by the

present stock in the city where the customer lives. Each customer order can be for any

quantity of any number of items, and each order is uniquely identified by an order

number.

The location of the stores is also recorded. Each store is located in one city, and there

can be many stores in the city. Each city has one headquarter for coordinating all of

its stores. The enterprise’s goal is to meet all of the customer’s requirements from

stores located in the customer’s city. If the requirement cannot be met, the company

will turn to the other cities where the item can be found if there is any.

Some processing information is important for the enterprise. For example, the total

quantity of item stored in each city. After every time an item is taken, the company

needs to know the total quantities of the item in all the stores in a city.

The relational schema of the enterprise’s current (operational) databases are:

Headquarter Database:

Relation Customer (Customer\_id, Customer\_name, City\_id, First\_order\_date)

Relation Walk-in\_customers (\*Customer\_id, tourism\_guide, Time)

Relation Mail\_order\_customers (\*Customer\_id, post\_address, Time)

Sales Databases:

Relation Headqarters (City\_id, City\_name, Headquarter\_addr, State, Time)

Relation Stores (Store\_id, \*City\_id, Phone, Time)

Relation Items (Item\_id, Description, Size, Weight, Unit\_price, Time)

Relation Stored\_items (\*Store\_id, \*Item\_id, Qantity\_held, Time)

Relation Order (Order\_no, Order\_date, Customer\_id)

Relation Ordered\_item (\*Order\_no, \*Item\_id, Quantity\_ordered, Ordered\_price,

Time)

Where underlined are primary key and “\*” prefixed are foreign keys.

Business requirements:

In order to meet users’ demand, the data warehouse system extracts data from the

existing two database into a data warehouse, and provides online analytical processing

with roll up, drill down, slice and dice features according to users selections based on

dimension tables to meet the user requirements.

Build data warehouse / OLAP which will answer the following queries :

1. Find all the stores along with city, state, phone, description, size, weight and

unit price that hold a particular item of stock.

2. Find all the orders along with customer name and order date that can be

fulfilled by a given store.

3. Find all stores along with city name and phone that hold items ordered by

given customer.

4. Find the headquarter address along with city and state of all stores that hold

stocks of an item above a particular level.

5. For each customer order, show the items ordered along with description, store

id and city name and the stores that hold the items.

6. Find the city and the state in which a given customer lives.

7. Find the stock level of a particular item in all stores in a particular city.

8. Find the items, quantity ordered, customer, store and city of an order.

9. Find the walk in customers, mail order customers and dual customers (both

walk-in and mail order).

Prepare report as follows :

1. Introduction – objective and scope of the project

2. Business requirement – application specification of the data warehousing for

the users.

3. Functional specification – input and output specification of the data

warehousing

4. Data Warehousing Design – stepwise procedure methodology of designing the

data warehousing including star schema.

5. Data cube implementation – computer automation of implementing the data

warehousing loading data into data cubes.

6. Observations :

a. Online analytical processing reports – invoke commands or panels to

generate OLAP reports.

b. Data verification – verify the OLAP reports source relational tables’ data

7. Conclusion.

**REPORT:**

1. Introduction

Objective:

The objective of this project is to design and implement a data warehouse for a customer order processing system within an enterprise that operates multiple stores across various cities and states. The data warehouse consolidates data from two operational databases (Headquarter and Sales) and supports online analytical processing (OLAP) operations such as roll-up, drill-down, slice, and dice to enable robust business reporting and decision making.

Scope:

Data Integration: Extract, transform, and load (ETL) data from existing operational systems.

Data Analysis: Provide a multidimensional view of orders, inventory, and customer details.

Reporting: Develop OLAP queries to address specific business questions.

Visualization: Use MySQL Workbench’s Visual Explain feature and result grids to analyze query performance and validate data.

2. Business Requirements

The enterprise has multiple stores in different cities and states with varied item stocks. It serves two types of customers—walk-in (via tourism guides) and mail-order (via postal addresses). Key business requirements include:

Inventory Management:

Identify which stores hold a particular item, along with details (city, state, phone, item description, size, weight, unit price).

Order Fulfillment:

Determine which orders (with customer name and order date) can be fulfilled by a given store.

Customer and Store Analysis:

Identify the stores that have supplied items ordered by a specific customer; determine a customer’s city and state.

Stock-Level Reporting:

Find headquarter information (address, city, state) for stores with inventory levels above a specified threshold.

Order Detail Reporting:

For every customer order, list the items ordered along with item descriptions, the associated store’s ID, and city.

Customer Classification:

Identify walk-in customers, mail-order customers, and those who are served by both channels.

OLAP Operations:

Support roll-up (aggregation by city/state), drill-down (detailed analysis at store/customer level), and slice/dice (filtering on dimensions) functionalities.

3. Functional Specification

Inputs:

Source Databases:

Headquarter Database:

Customer(Customer\_id, Customer\_name, City\_id, First\_order\_date)

Walk-in\_customers(\*Customer\_id, tourism\_guide, Time)

Mail\_order\_customers(\*Customer\_id, post\_address, Time)

Sales Database:

Headquarters(City\_id, City\_name, Headquarter\_addr, State, Time)

Stores(Store\_id, \*City\_id, Phone, Time)

Items(Item\_id, Description, Size, Weight, Unit\_price, Time)

Stored\_items(\*Store\_id, \*Item\_id, Quantity\_held, Time)

Order(Order\_no, Order\_date, Customer\_id)

Ordered\_item(\*Order\_no, \*Item\_id, Quantity\_ordered, Ordered\_price, Time)

Outputs:

OLAP reports and dashboards that answer:

Which stores (with city, state, and phone) hold a specific item and its details.

Which orders (with customer name and order date) are fulfillable by a given store.

Which stores (with city name and phone) hold items ordered by a given customer.

The headquarter address (with city and state) of stores holding stock above a threshold.

For each order, the ordered items with description, store ID, and city.

The city and state where a given customer lives.

The stock level of a specific item in all stores in a given city.

Detailed information (items, quantity ordered, customer, store, city) for an order.

A classification of customers by type (walk-in, mail-order, both).

4. Data Warehousing Design

A. ETL Process

Extract:

Retrieve data from both operational databases (Headquarter and Sales).

Transform:

Data Cleaning: Remove duplicates, handle missing values, and standardize formats.

Integration: Merge customer, order, and inventory data into a unified schema.

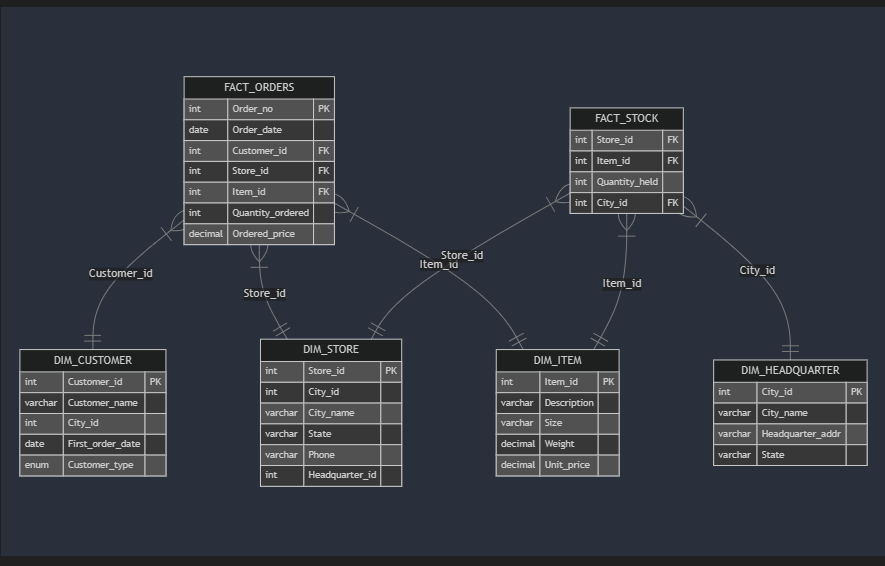
Derived Attributes: Create fields such as Customer\_type (with values like 'Walk-in', 'Mail-order', 'Both').

Load:

Populate the new data warehouse tables with the cleaned and transformed data.

B. Star Schema Design

The star schema centers on fact tables that capture transactional data, linked to dimension tables containing descriptive attributes.



**(FOR OLAP QUERIES REFER THE SQL FILE)**

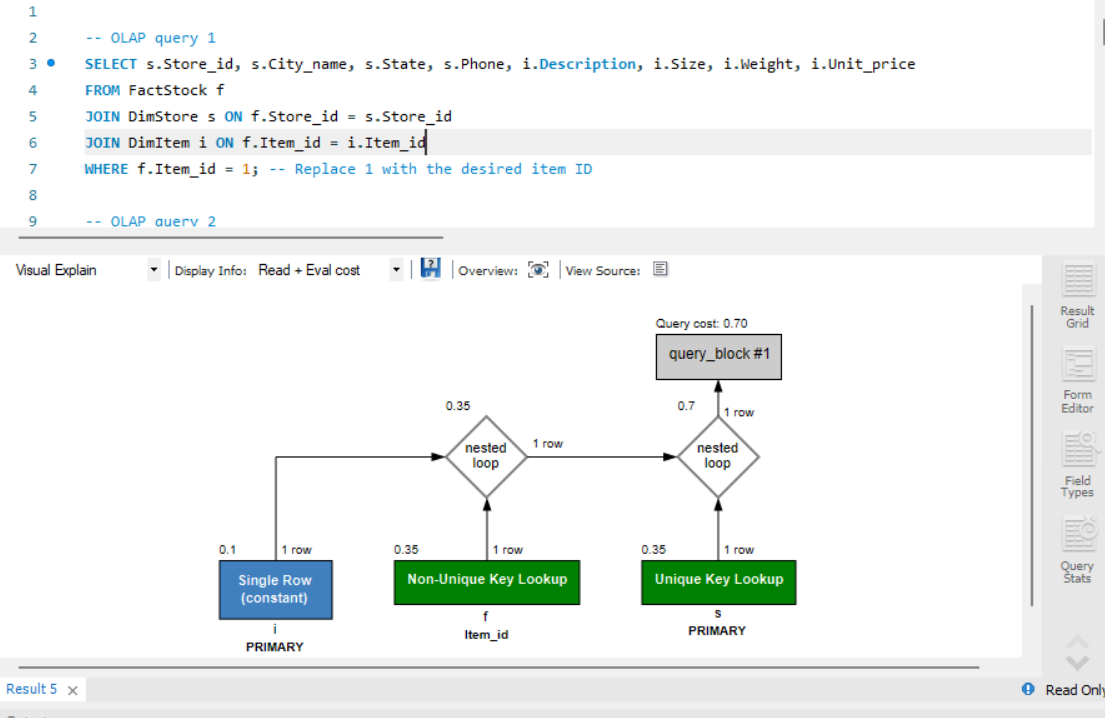
5. Data Cube Implementation (OLAP)

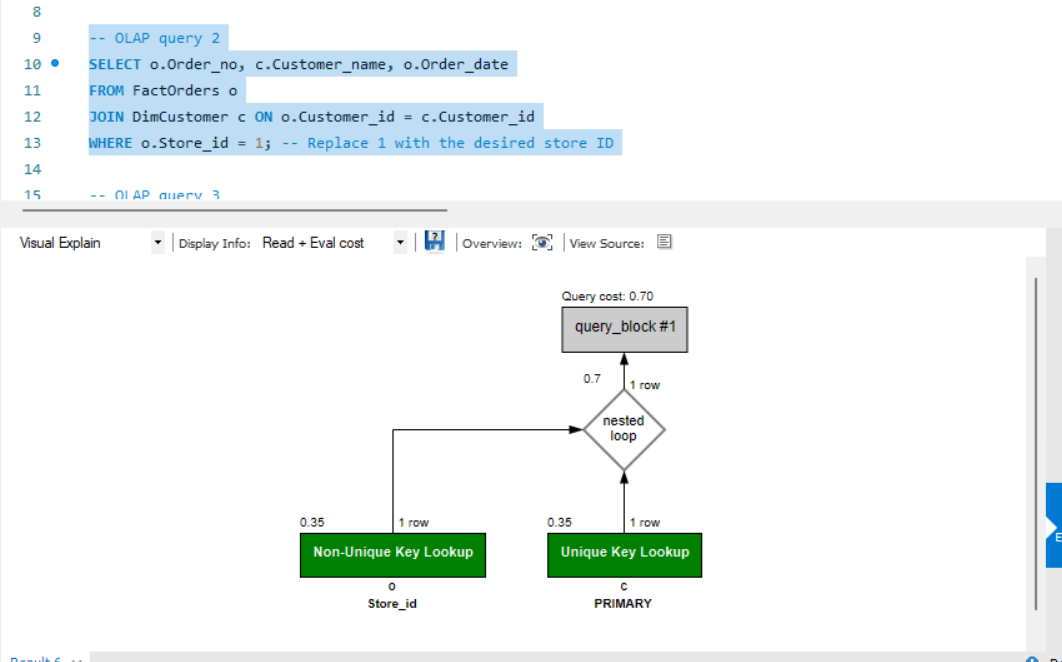
After loading your data into the star schema, you can implement OLAP operations using SQL queries to generate data cubes. This process involves:

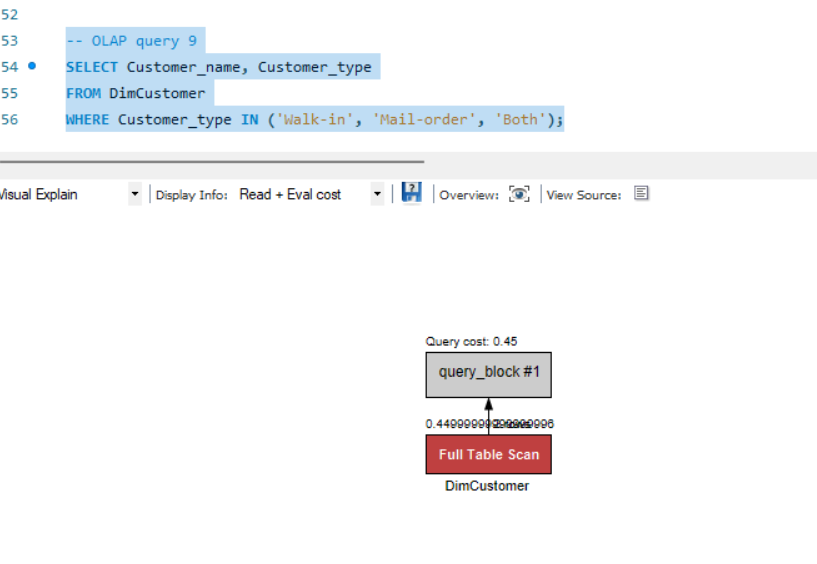
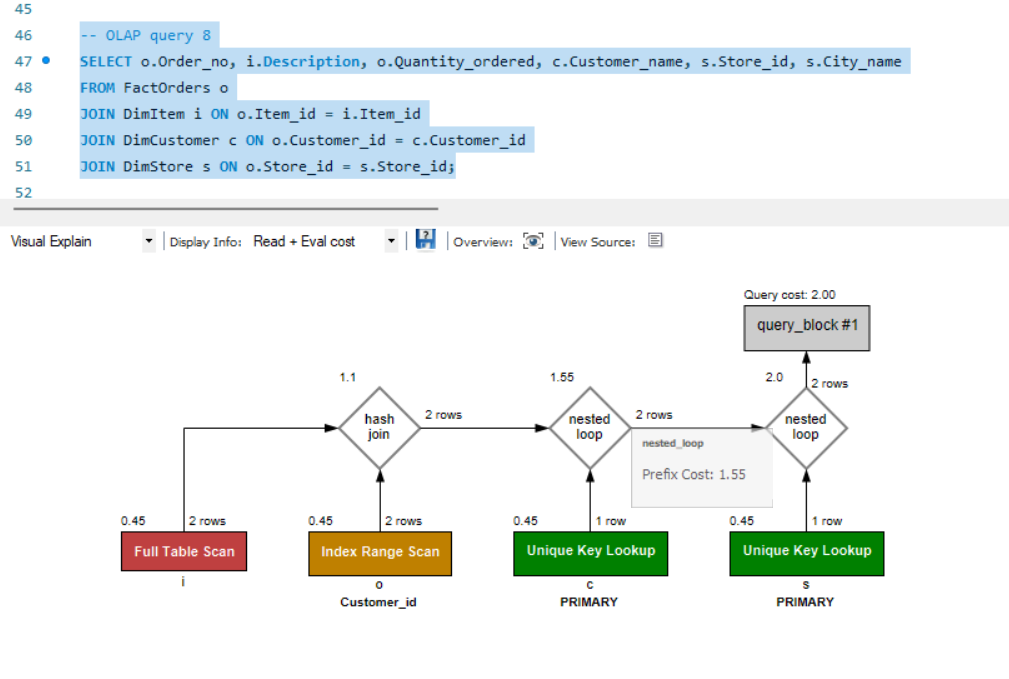
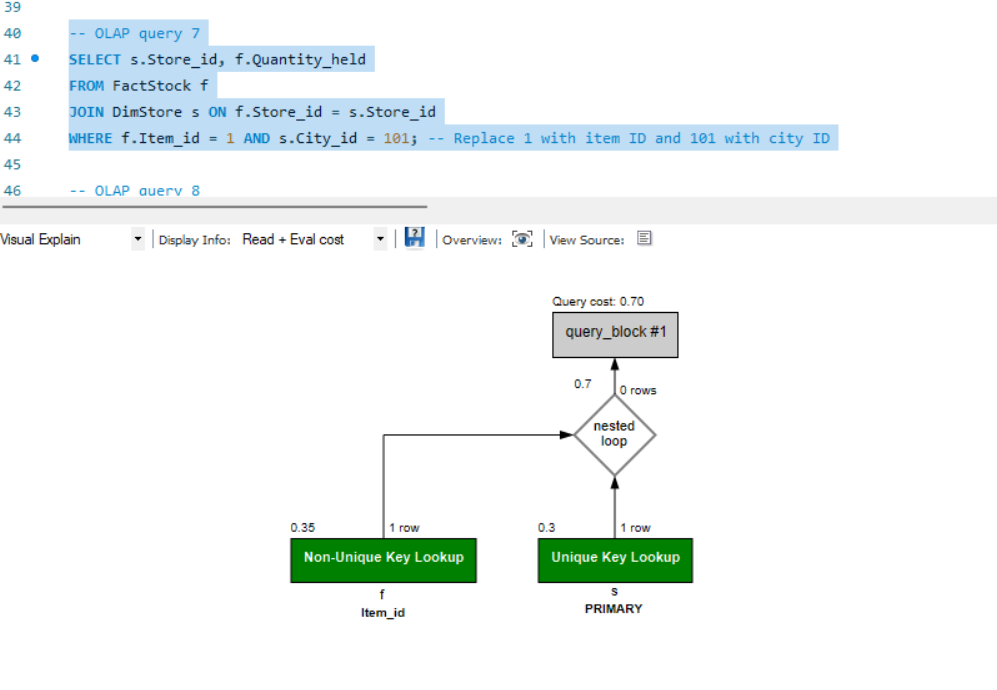
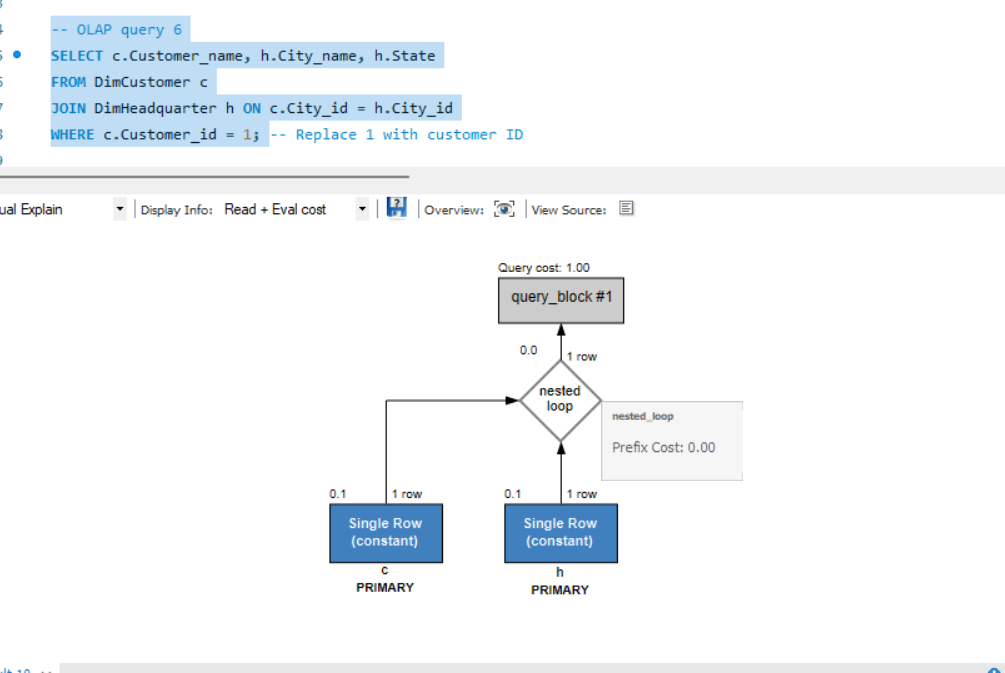
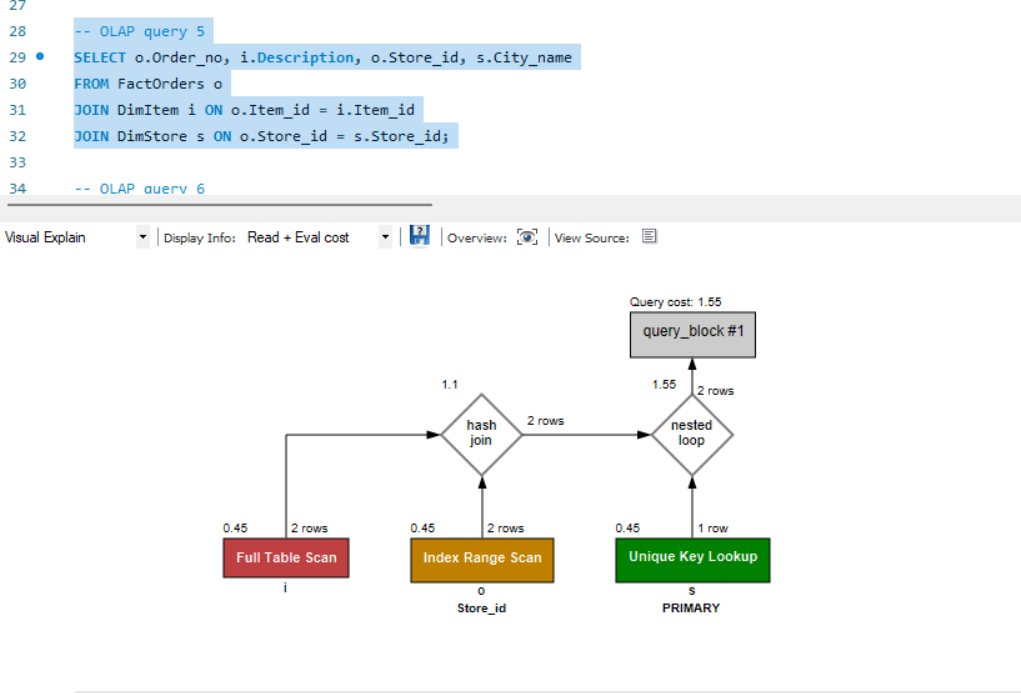
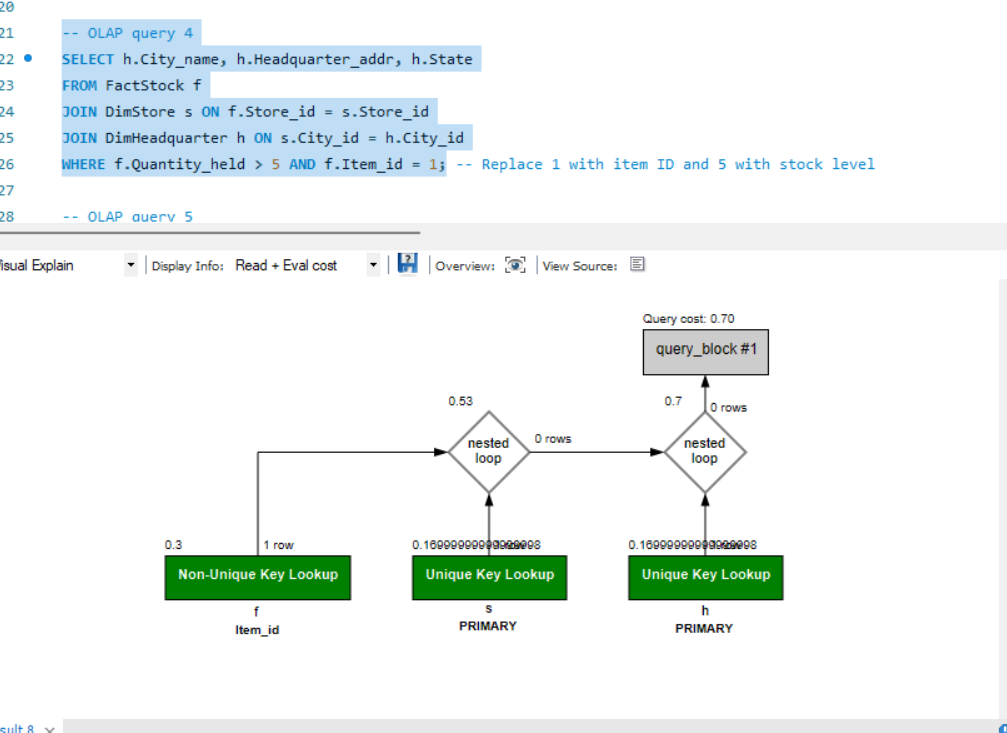
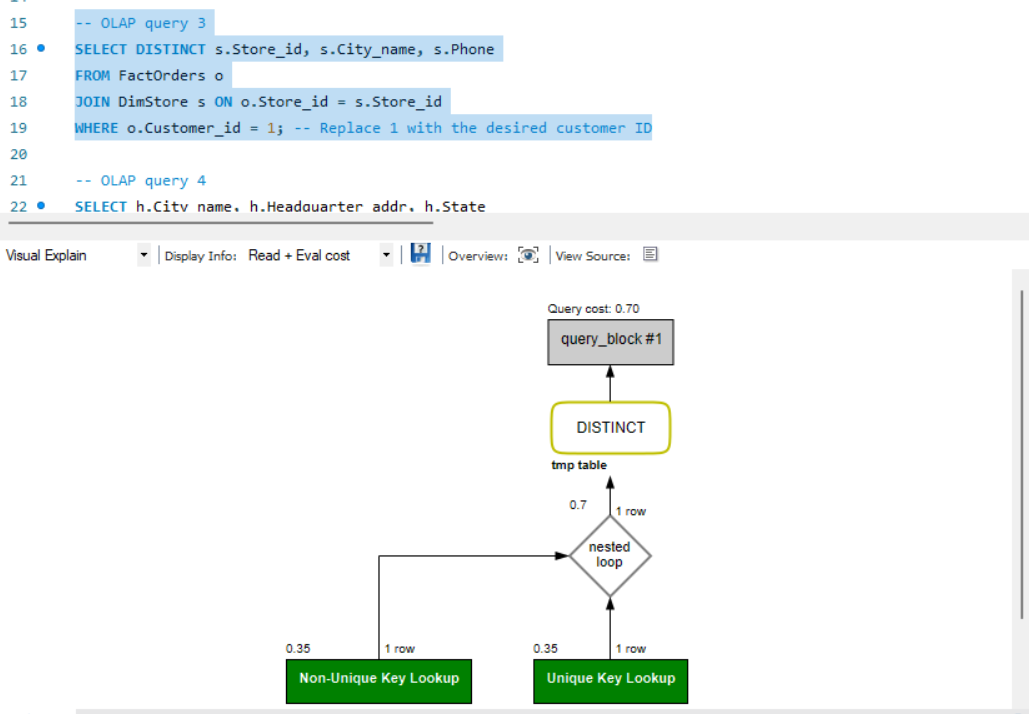
Roll-Up: Aggregating data to a higher-level summary (e.g., total orders by city or state).

Drill-Down: Breaking the aggregate view into detailed data (e.g., orders by individual store).

Slice and Dice: Filtering on specific dimensions (e.g., retrieving data for a particular item, store, or customer).







6. Observations

A. OLAP Reports

Execution:

Each OLAP query has been designed to pull multidimensional insights. Running these queries in MySQL Workbench returns result sets that can be used for reporting and further analysis.

Visualization:

Use MySQL Workbench’s “Visual Explain” feature to view the execution plan. This helps in understanding how MySQL processes each query and highlights if a full table scan or an index lookup is used. For instance, you may see an output like:

B. Data Verification

Reconciliation:

Verify that the numbers returned by OLAP queries match the source data in the operational databases. For example, cross-check order totals or stock levels.

Screenshots:

Capture screenshots of both the result grids and the Visual Explain output to include in your final report. This documentation demonstrates that the ETL and OLAP operations have been correctly implemented.

7. Conclusion

In this project, we designed and implemented a data warehouse for a customer order processing system using MySQL. The solution involved:

ETL Process: Extracting, cleaning, integrating, and loading data from separate operational databases.

Star Schema Design: Creating a dimensional model (fact and dimension tables) to support OLAP queries.

OLAP Implementation: Writing queries to support roll-up, drill-down, and slice/dice operations.

Visualization & Optimization: Using MySQL Workbench’s Visual Explain feature to review query performance and document our work.

This system now enables the enterprise to analyze customer orders, inventory levels, and store performance from multiple dimensions, thereby supporting effective business decision-making.

7. References

Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Wiley.

Inmon, W. H. (2005). Building the Data Warehouse. John Wiley & Sons.