**M.Sc C.S - I**

**SEM II**

**Journal**

|  |  |
| --- | --- |
| **Roll No.** | 003 |
| **Name** | PALAK RAWAL |
| **Subject** | BUSINESS INTELLIGENCE AND BIG DATA ANALYSIS |



CERTIFICATE

This is here to certify that Mr. **PALAK RAWAL**, Seat Number **003** of M.Sc. I Computer Science, has satisfactorily completed the required number of experiments prescribed by the **THAKUR COLLEGE OF SCIENCE & COMMERCE AUTONOMOUS COLLEGE, PERMANENTLY AFFILIATED TO UNIVERSITY OF MUMBAI** during the academic year 2021 – 2022.

Date: 18-03-2022

Place: Mumbai

Teacher In-Charge Head of Department

External Examiner

**INDEX**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **Experiments** | **Page**  **No** | **Date** |
| **1** | Do data pre-processing on data obtained from databases imported from external. | 01-08 | 07th Mar 2022 |
| **2** | Develop an application to implement defining subject areas, design of fact and dimension tables, data marts. | 09-15 | 21st Mar 2022 |
| **3** | Develop an application to implement OLAP, roll-up, drill-down, slice, and dice operations. | 16-18 | 24th Mar 2022 |
| **4** | Develop an application to construct a multidimensional data. | 19-24 | 24th Mar 2022 |
| **5** | Design and create cube by identifying measures and dimensions for star schema. | 25-33 | 04th April 2022 |
| **6** | Design and create cube by identifying measures and dimensions for snow flake schema. | 34-39 | 04th April 2022 |
| **7** | Create and uses excel pivot table report based on data cube operations. | 40-48 | 11th April 2022 |
| **8** | Prerequisites for Database applications. | 49-59 | 11th April 2022 |

|  |  |  |
| --- | --- | --- |
| **Fig. No.** | **List of Figures** | **Page No** |

**5.** Star Schema. **25**

**6.** Snowflake Schema. **34**

**7.** Data Cube **40**

**Practical 1**

**Aim: Do data preprocessing on data obtained from databases imported from external**

**sources..**

**Theory:**

**What is data preprocessing?**

Data preprocessing, a component of data preparation, describes any type of processing performed on raw data to prepare it for another data processing procedure. It has traditionally been an important preliminary step for the data mining process. More recently, data preprocessing techniques have been adapted for training machine learning models and AI models and for running inferences against them.

Data preprocessing transforms the data into a format that is more easily and effectively processed in data mining, machine learning and other data science tasks. The techniques are generally used at the earliest stages of the machine learning and AI development pipeline to ensure accurate results.

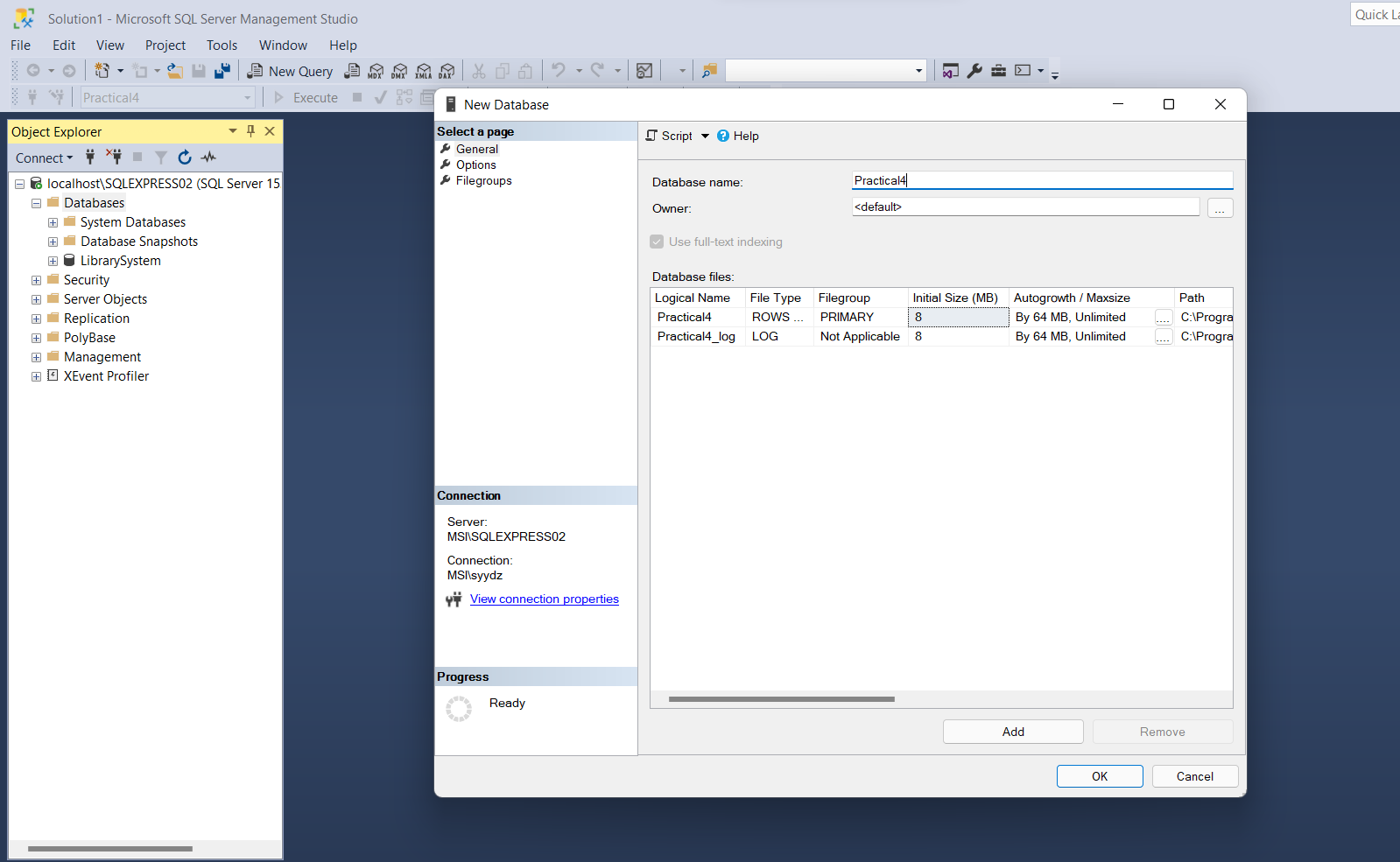
There are several different tools and methods used for preprocessing data, including the following:

* sampling, which selects a representative subset from a large population of data;
* transformation, which manipulates raw data to produce a single input;
* denoising, which removes noise from data;
* imputation, which synthesizes statistically relevant data for missing values;
* normalization, which organizes data for more efficient access; and
* feature extraction, which pulls out a relevant feature subset that is significant in a particular context.
* pulls out a relevant feature subset that is significant in a particular context.

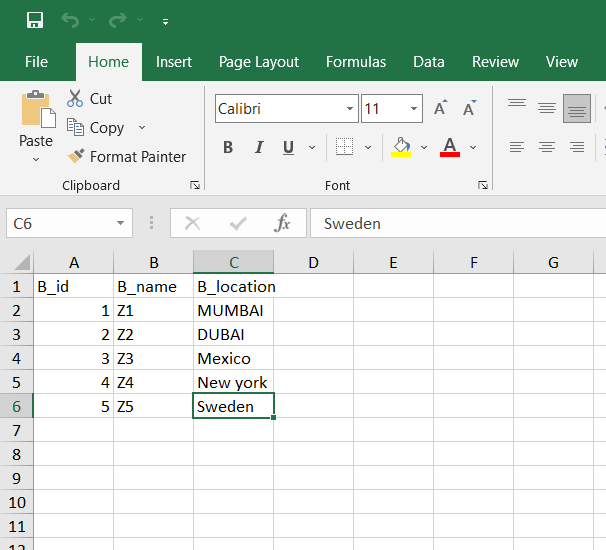
**#Steps:**

**1) Start Microsoft SQL server**

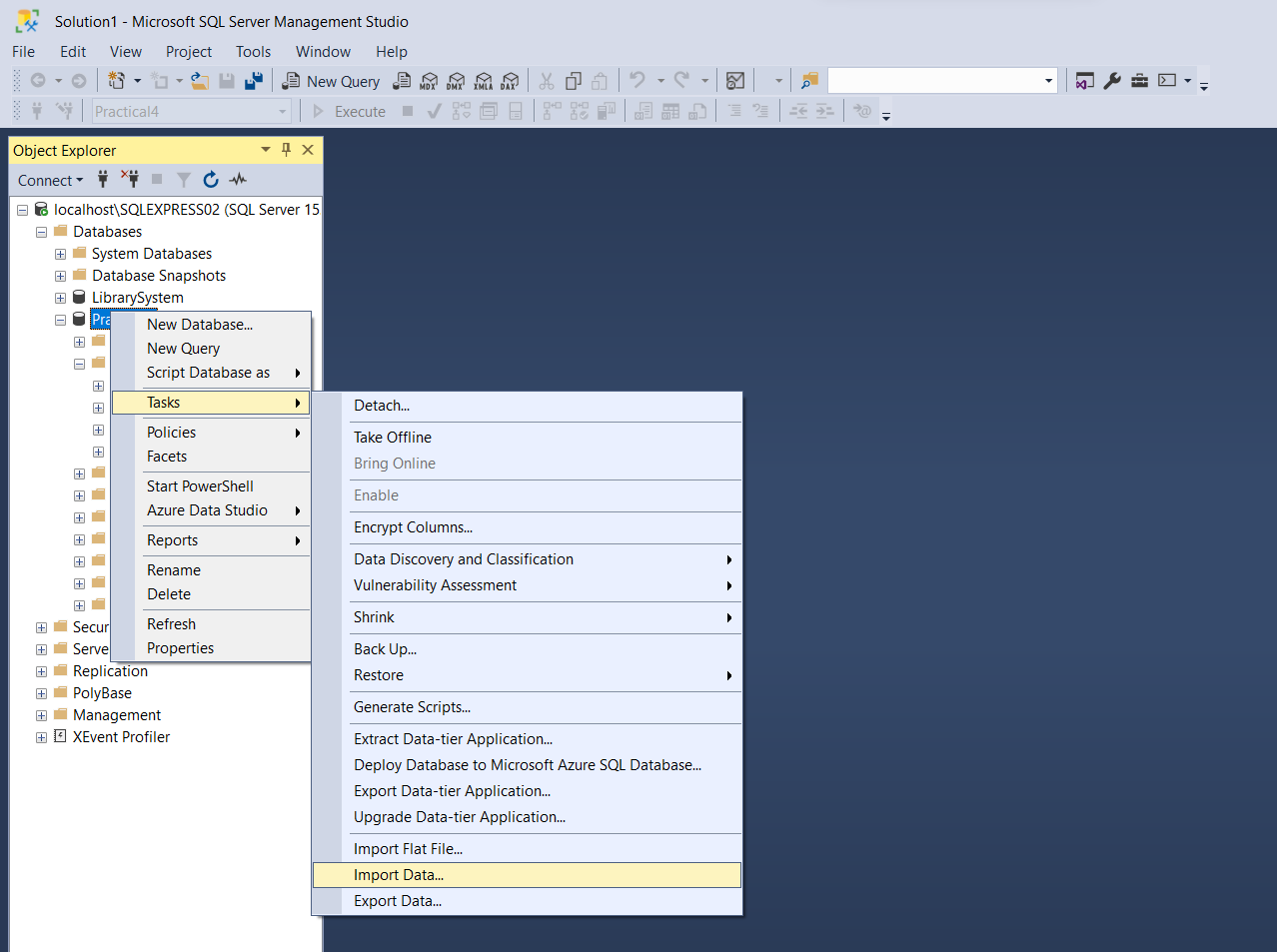
**2) Right Click On Databases > New Database > Give Database Name > Click Ok**

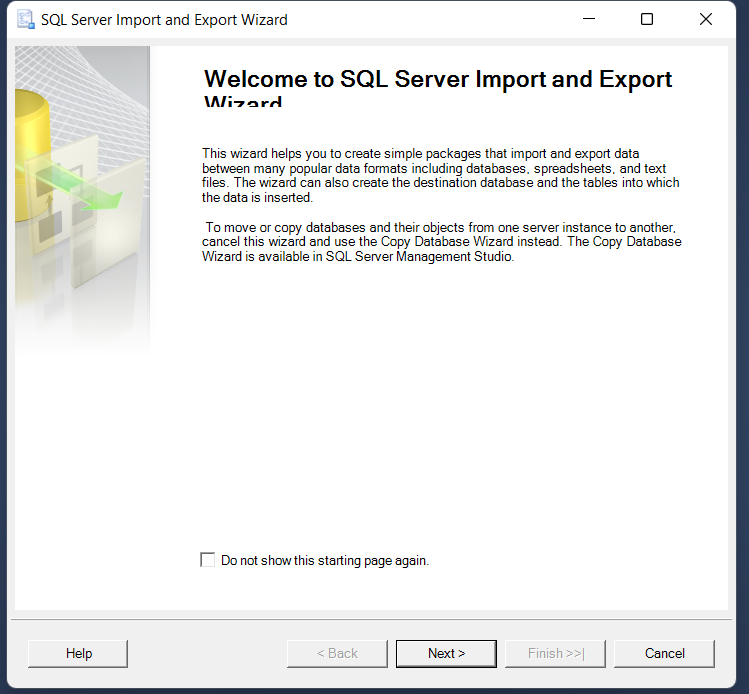
****

**3) Create a new Microsoft Excel 97-2003 Worksheet (.xls) with columns Branch id, Branch name, Branch location and enter some data into rows.**

****

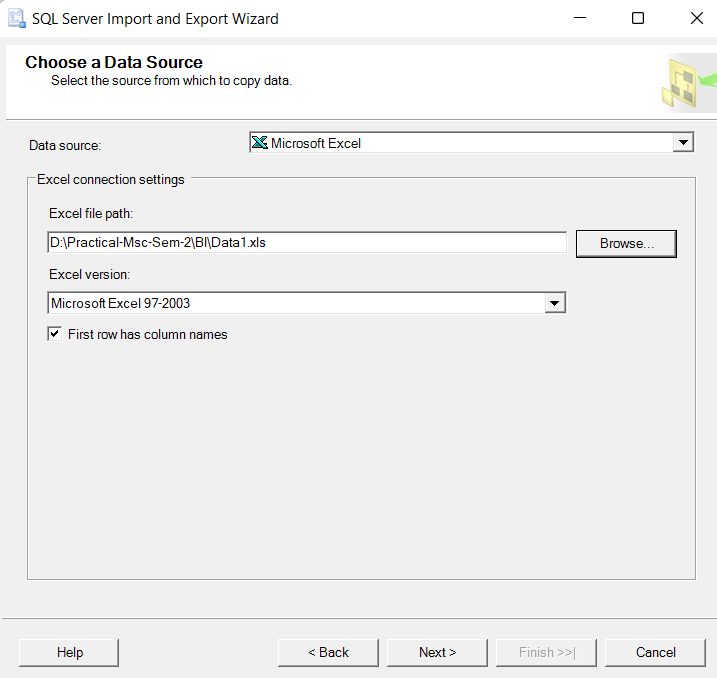
**4) Right Click On Practical4 > Tasks > Select: Import Data**

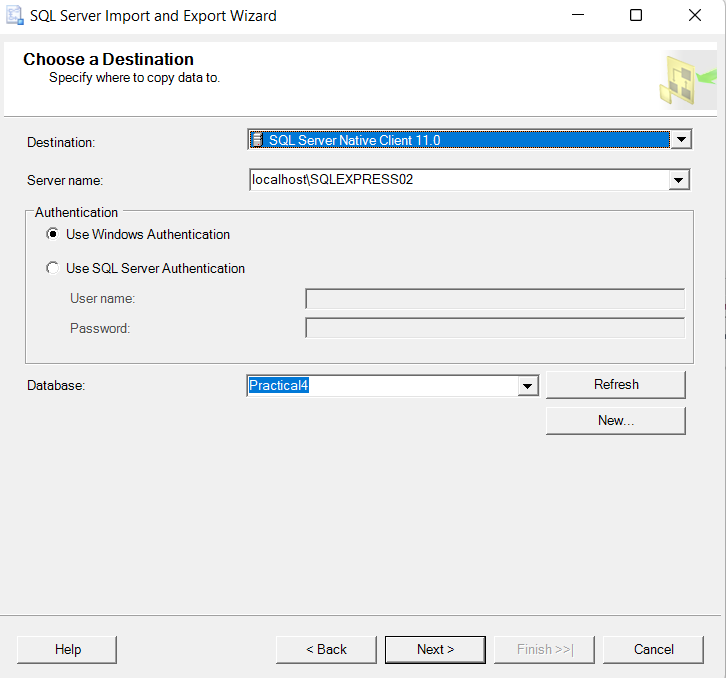
****

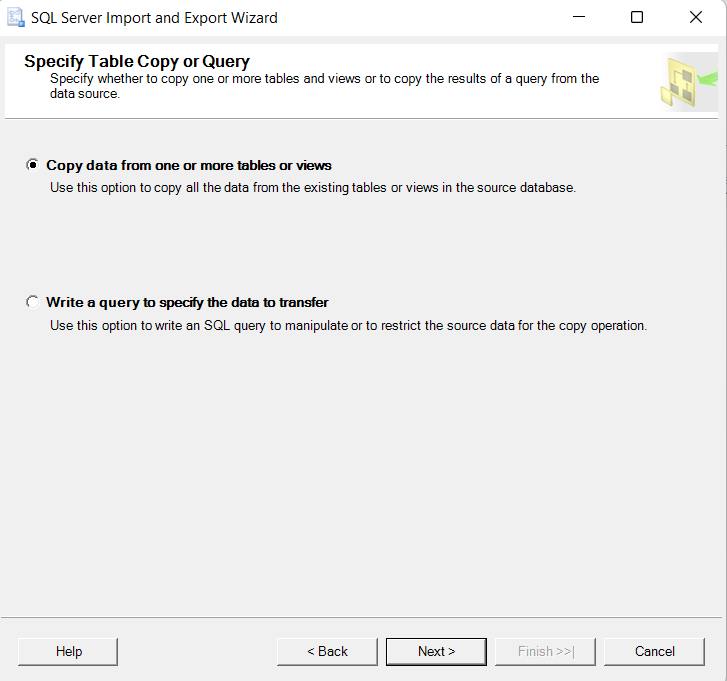
****

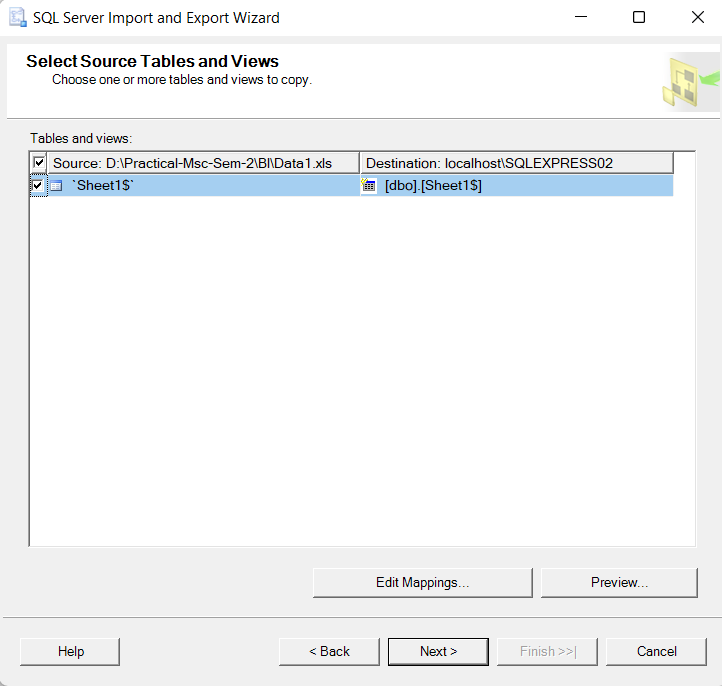
**5) Click on Next > Select Data Source as: Microsoft Excel > Give Excel File Path > Select Destination as: SQL Server Native Client 11.0 > Click Next > Select Sheet 1 > Click Next > Click Finish > Close the window after successful execution.**

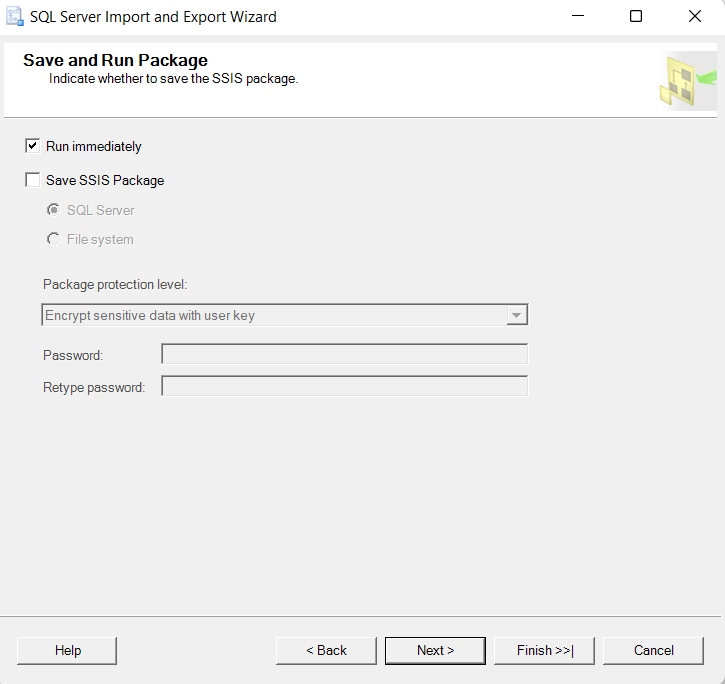
**Note: Close the excel file before using it as Source File**

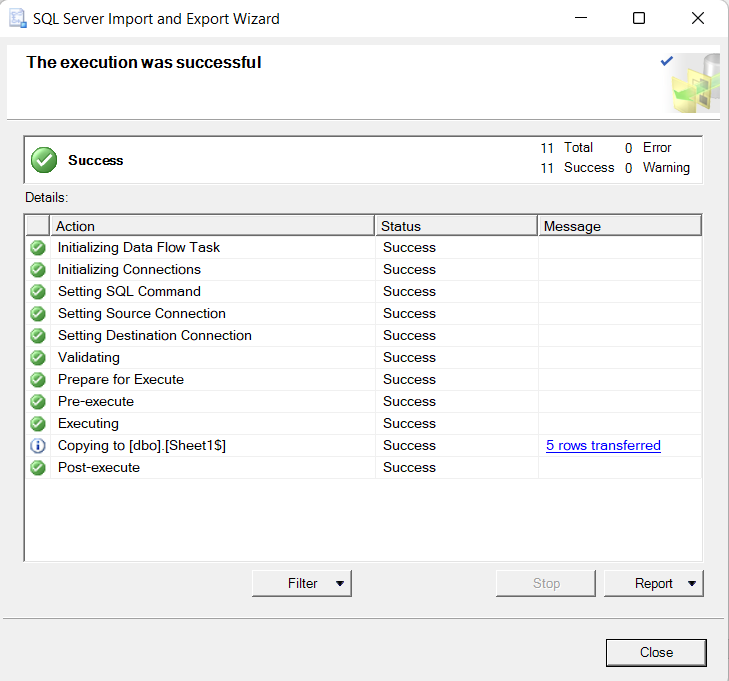
****

****

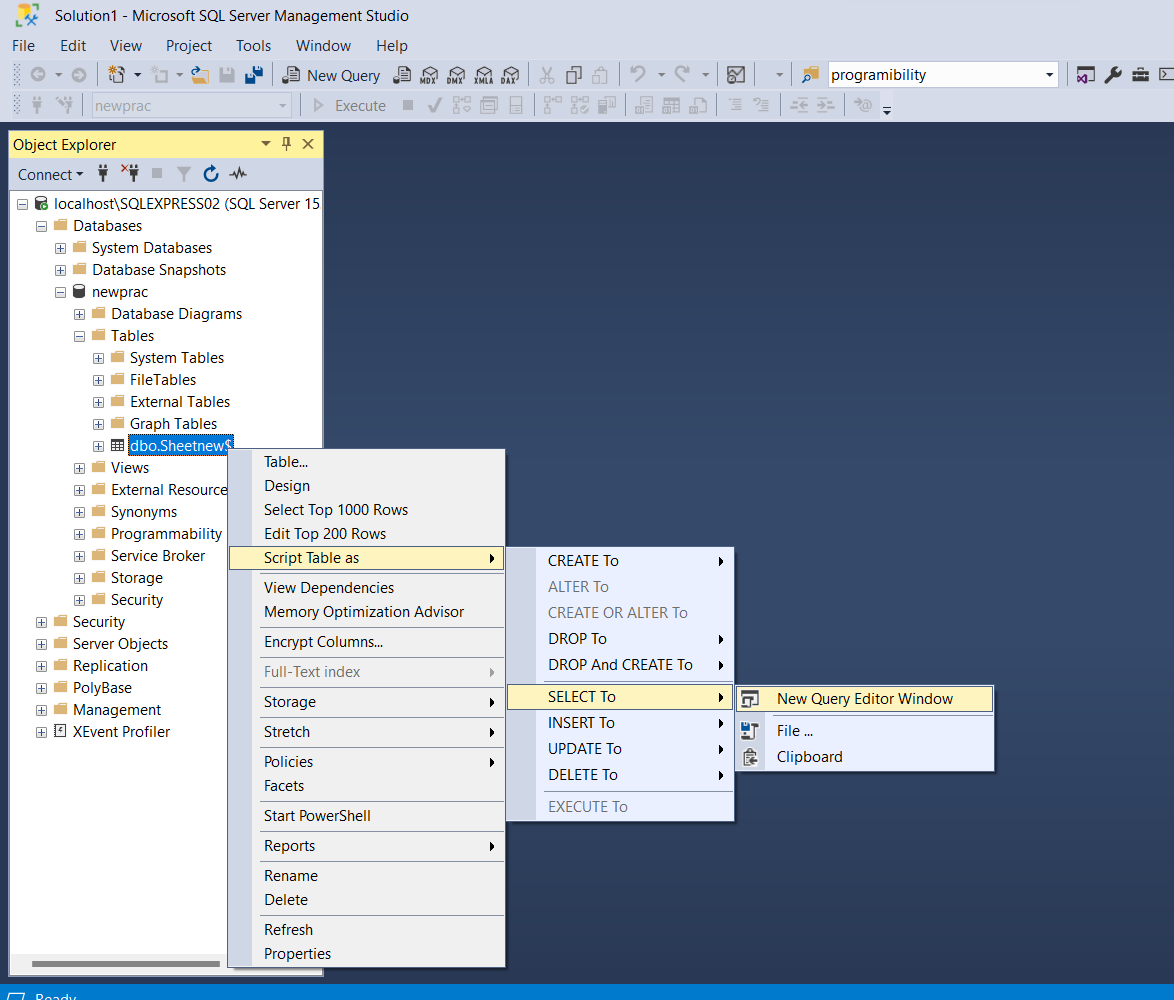
****

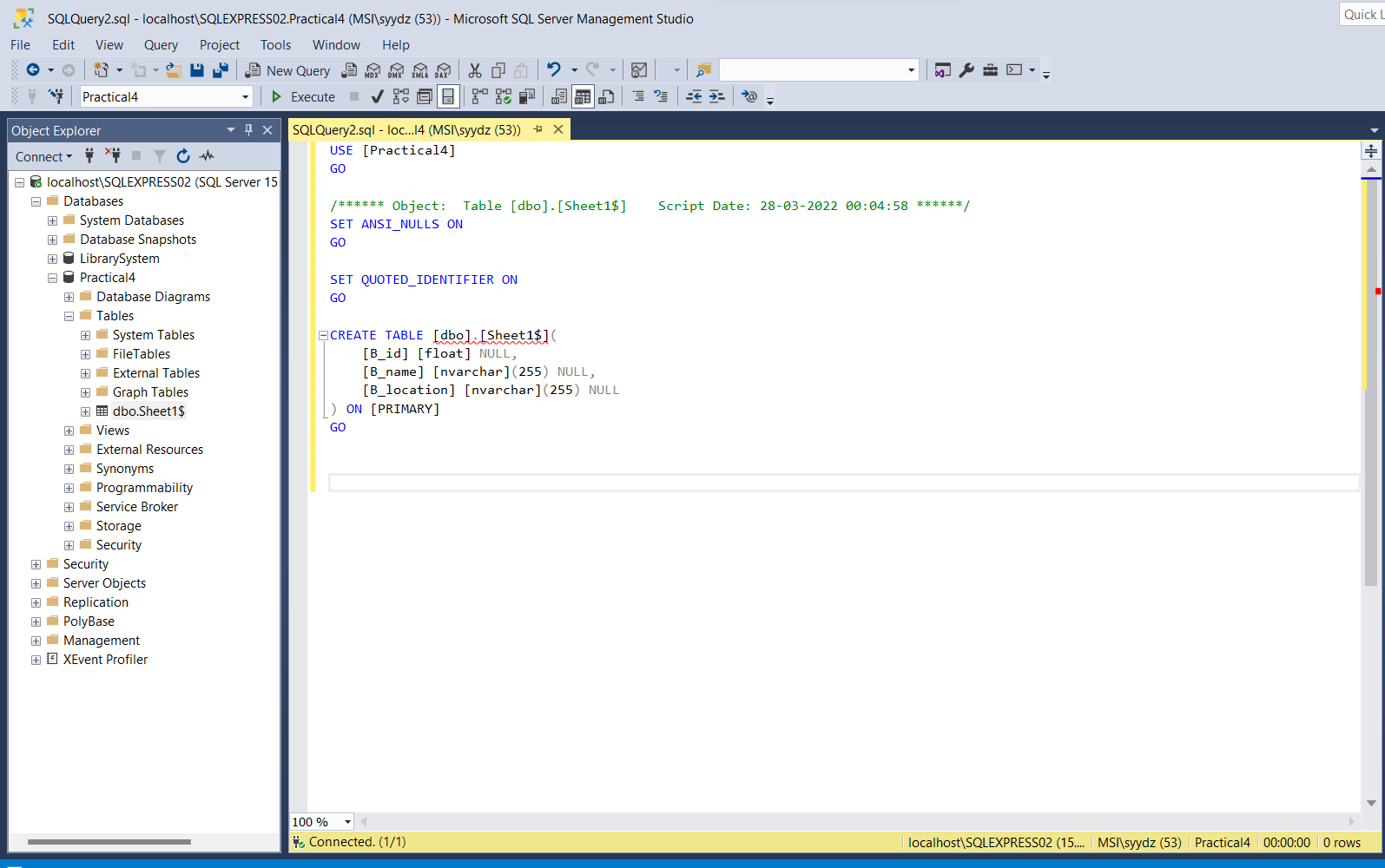
****

****

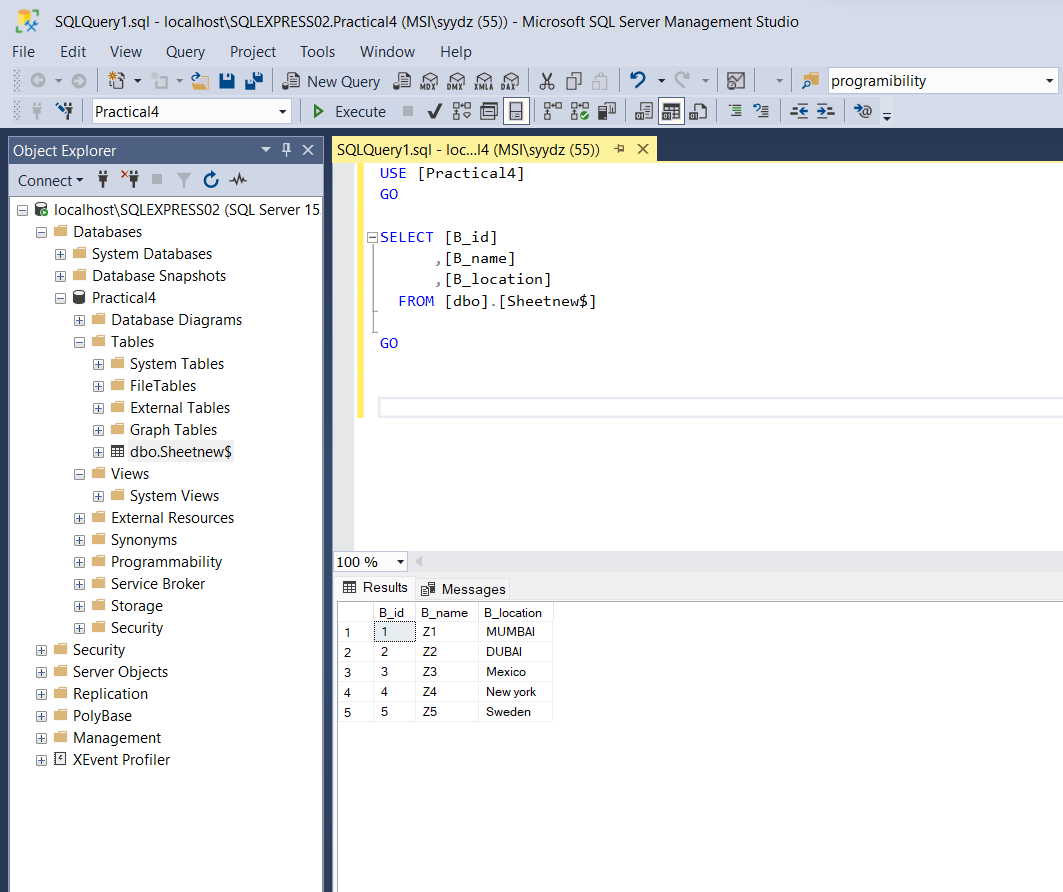
****

**6) Now expand the Tables Section > Right Click on dbo.Sheet$1 > Script table as > Select To > Click on New Query Editor Window.**

****

****

**7) Click on Execute**

****

**Conclusion: Data pre-processing on imported data from external sources implemented Successfully**

Practical 2

Aim: Develop an application to implement defining subject areas, design of fact and

dimension tables, data marts

Theory:

**Fact Table**

A Fact Table is a central table in a star schema of a data warehouse. It is an important concept required for Data Warehousing and BI . A fact table stores quantitative information for analysis and is often denormalized. A fact table works with dimension tables and it holds the data to be analyzed and a dimension table stores data about the ways in which the data can be analyzed.

Thus, a fact table consists of two types of columns. The foreign keys column allows to join with dimension tables and the measure columns contain the data that is being analyzed.

A dimension table is a table in a star schema of a data warehouse. A dimension table stores attributes, or dimensions, that describe the objects in a fact table.

**Dimension Table**

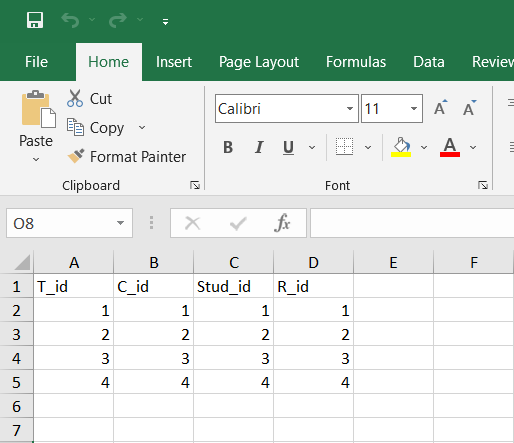
In data warehousing, a dimension is a collection of reference information about a measurable event. These events are known as facts and are stored in a fact table. Dimensions categorize and describe data warehouse facts and measures in ways that support meaningful answers to business questions. They form the very core of dimensional modeling.

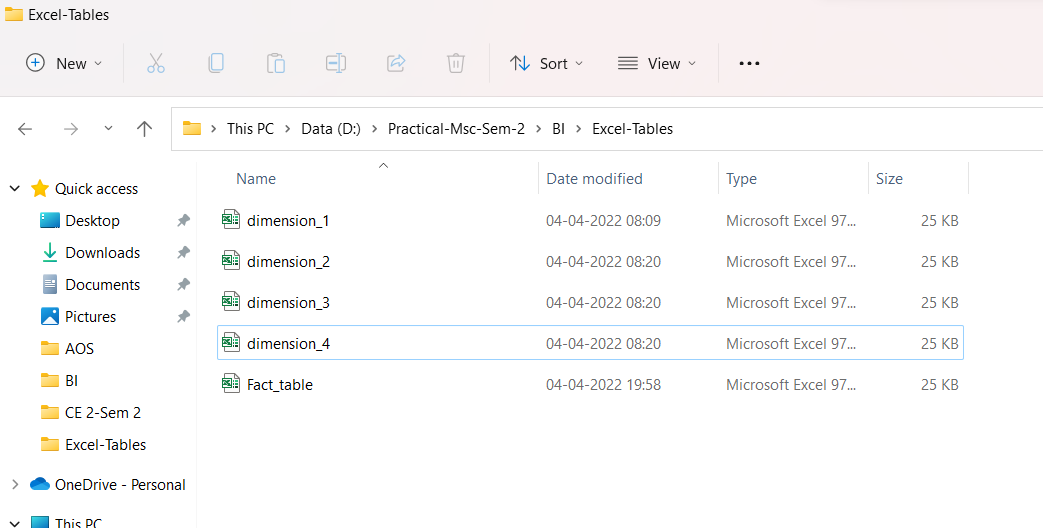
**Data Mart**

A data mart is a simple form of data warehouse focused on a single subject or line of business. With a data mart, teams can access data and gain insights faster, because they don’t have to spend time searching within a more complex data warehouse or manually aggregating data from different sources.

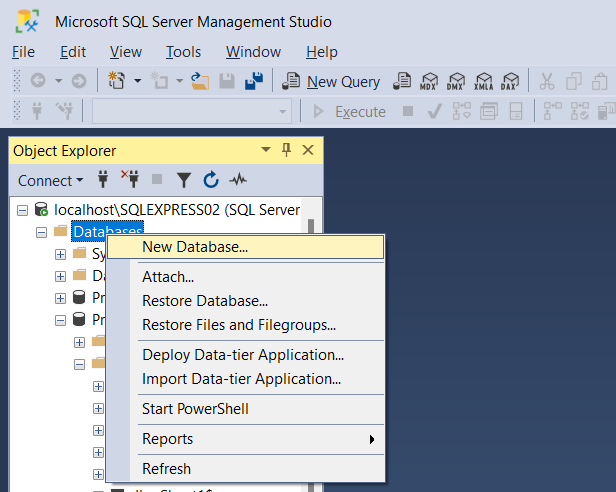
#Steps:

1) Create fact and dimensions tables in excel(.xls) (Change sheet name for every sheet)

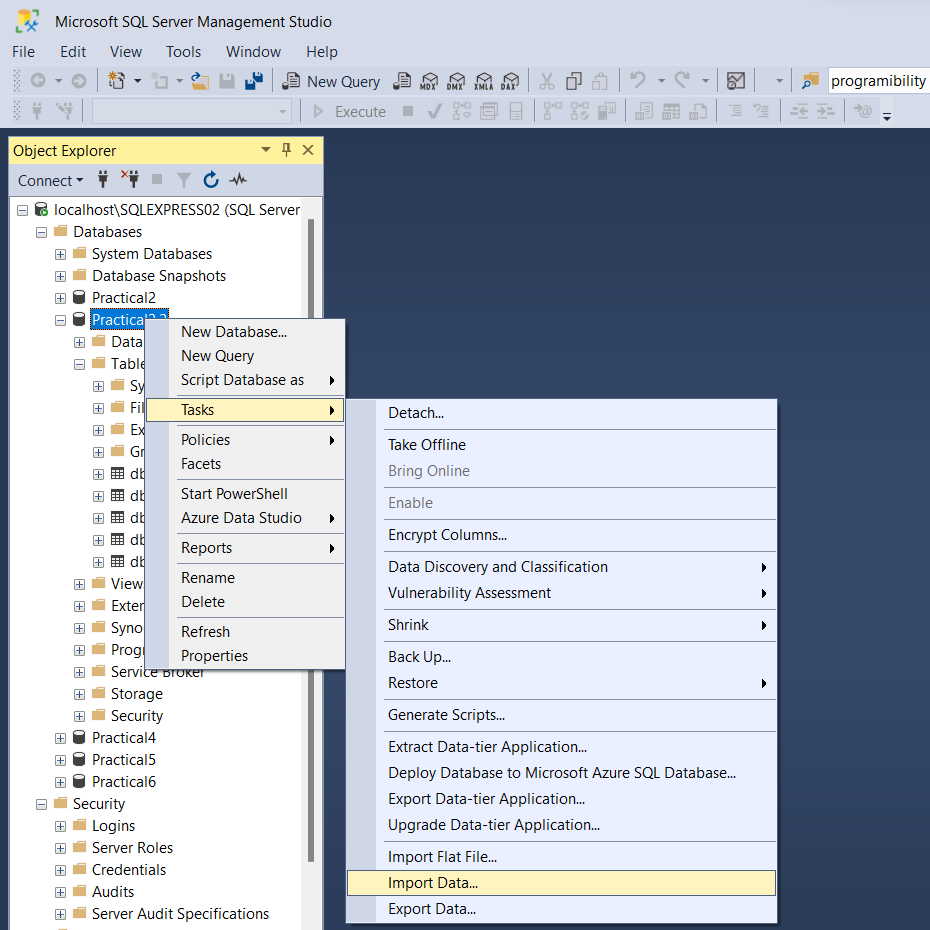


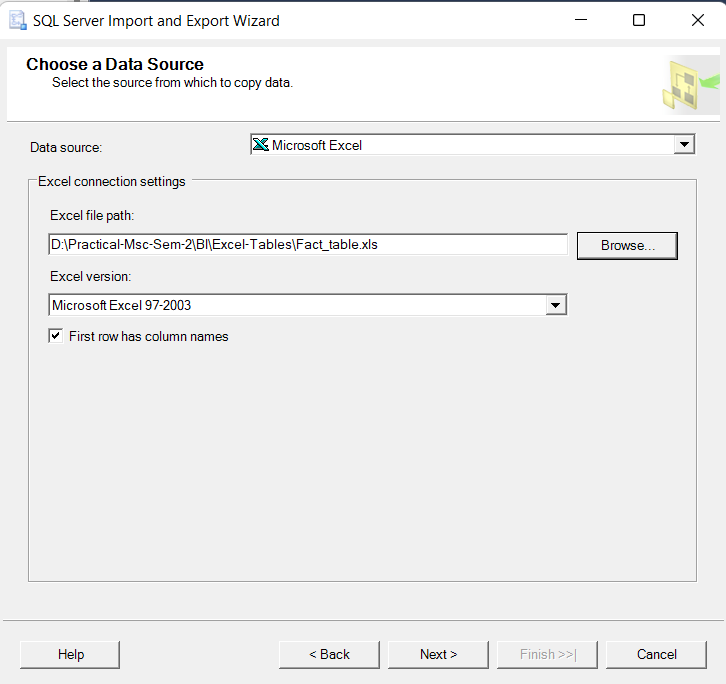


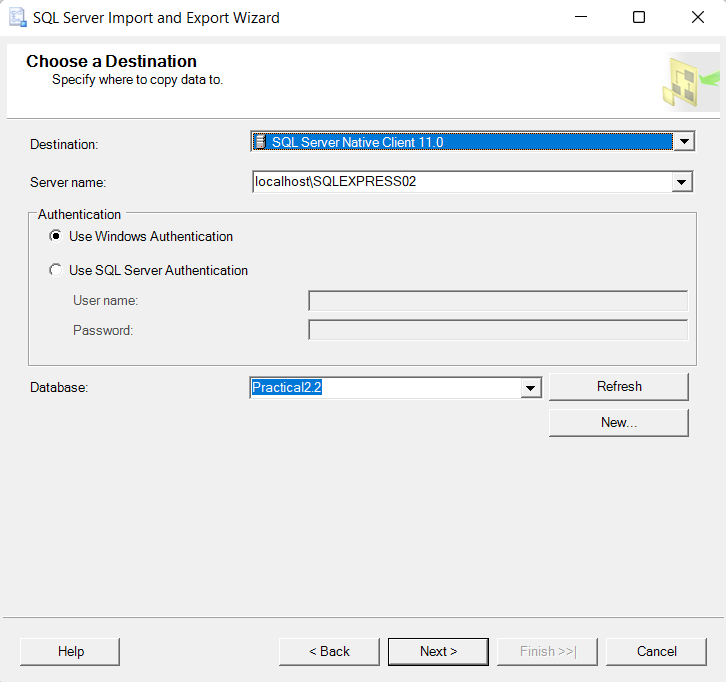
2) Create new database.



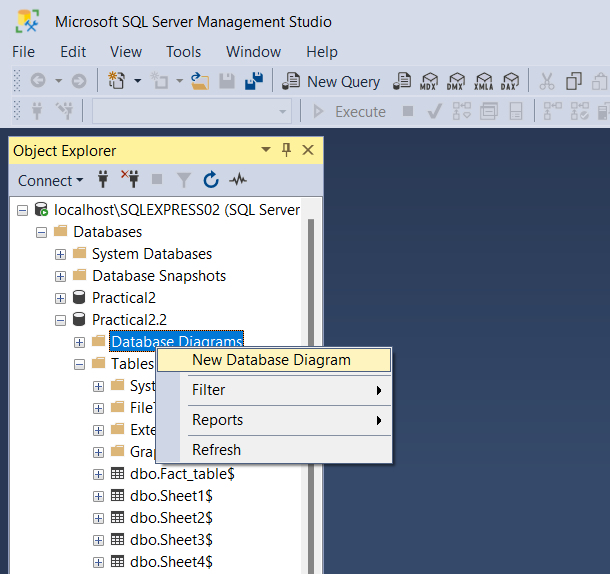
3) Import excel files > Choose data source > Choose destination source >Finish

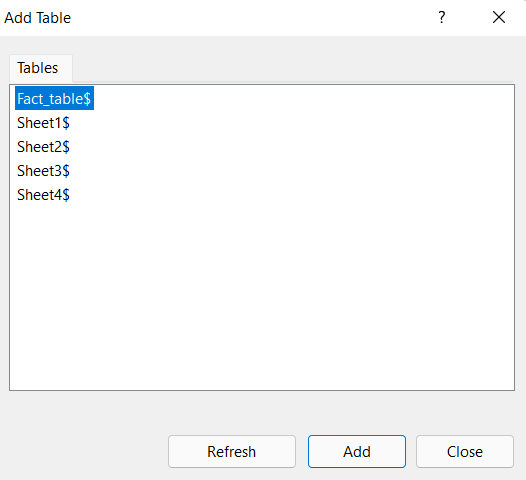




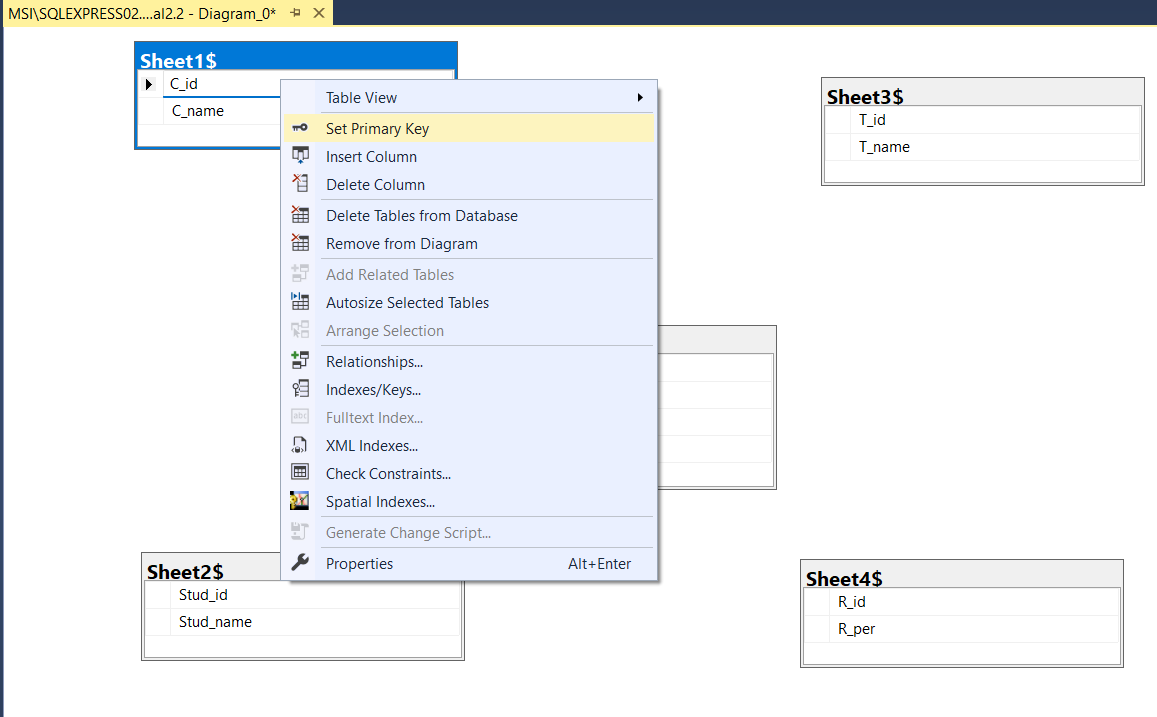


4) Create database diagrams

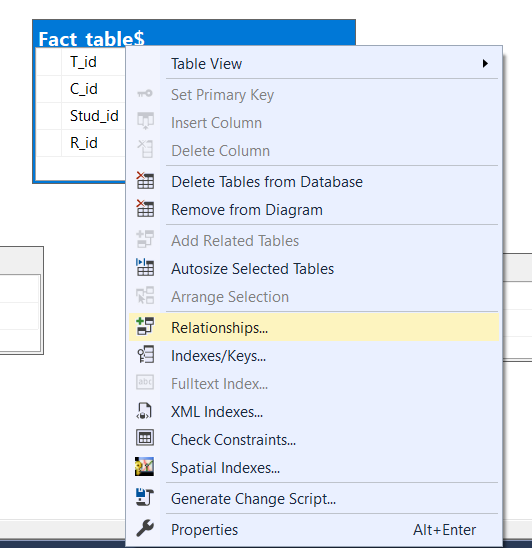


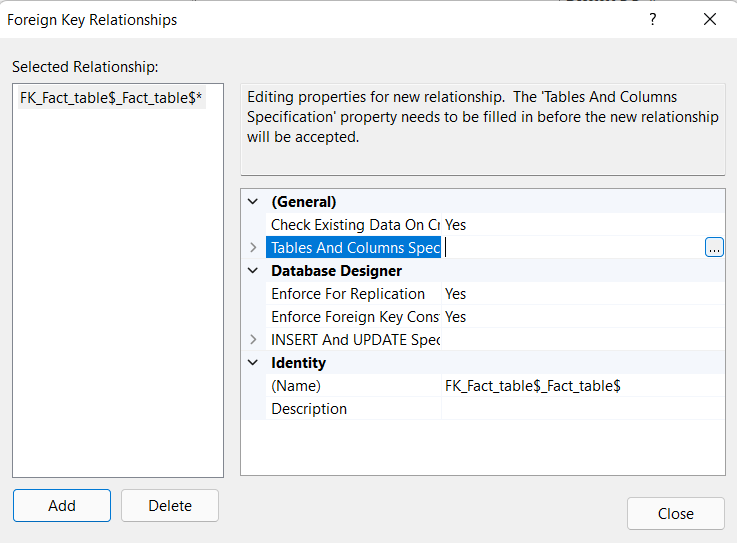


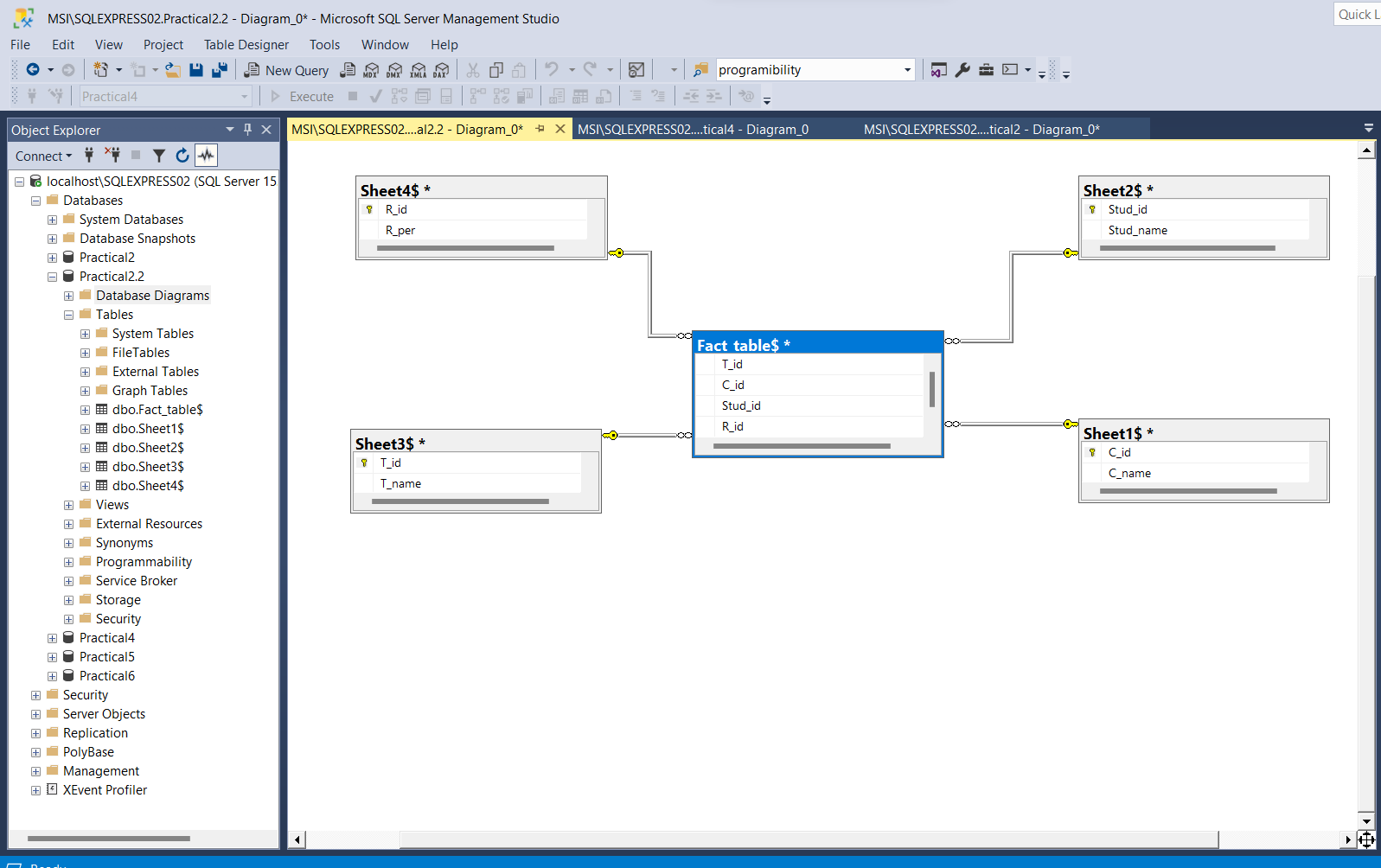
5) Set primary keys

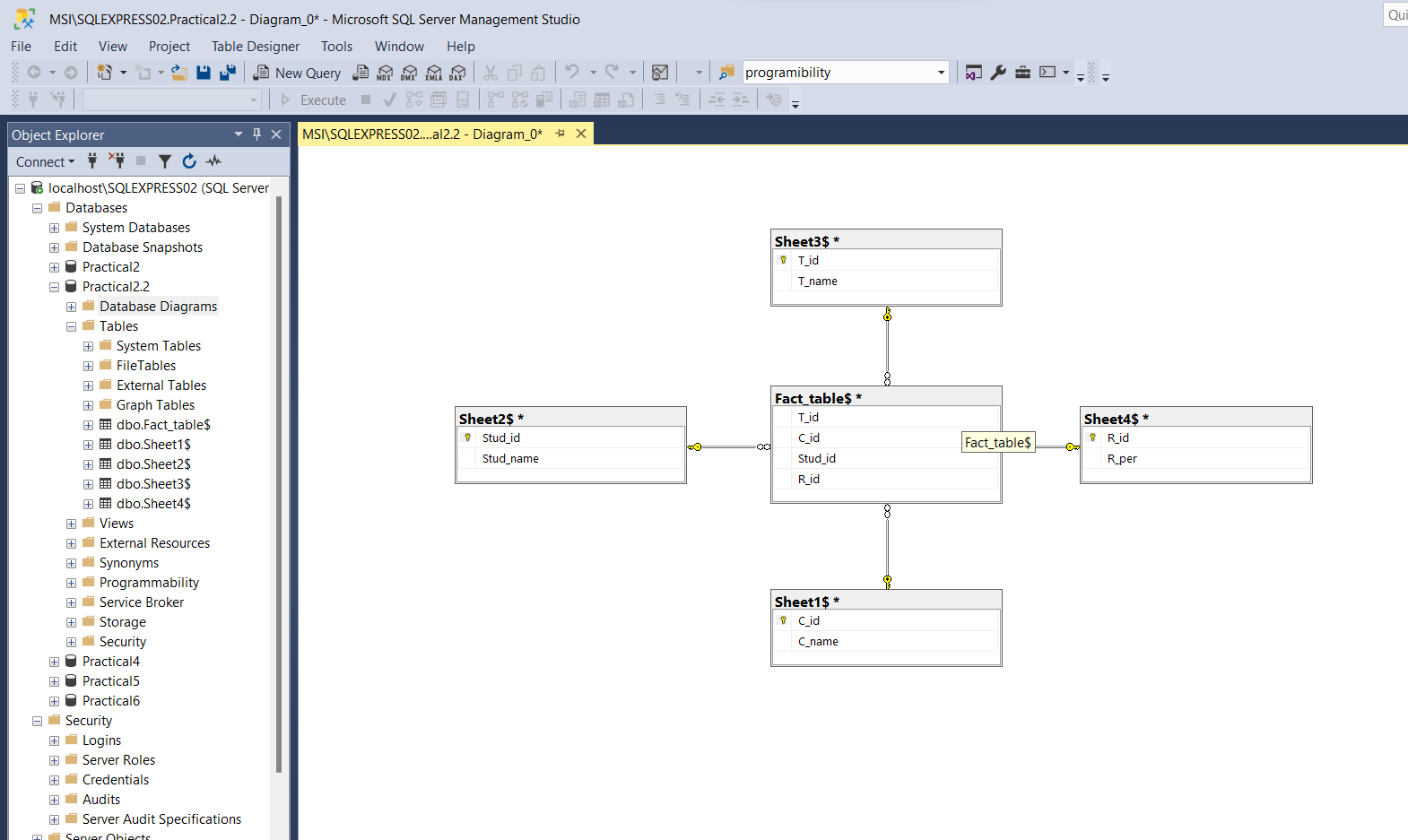


6) Create relationships for fact table









Conclusion: Successfully Implemented fact, dimension table and data marts in MS SQL.

Practical 3

Aim: Develop an application to implement OLAP, roll-up, drill-down, slice, and dice

operations.

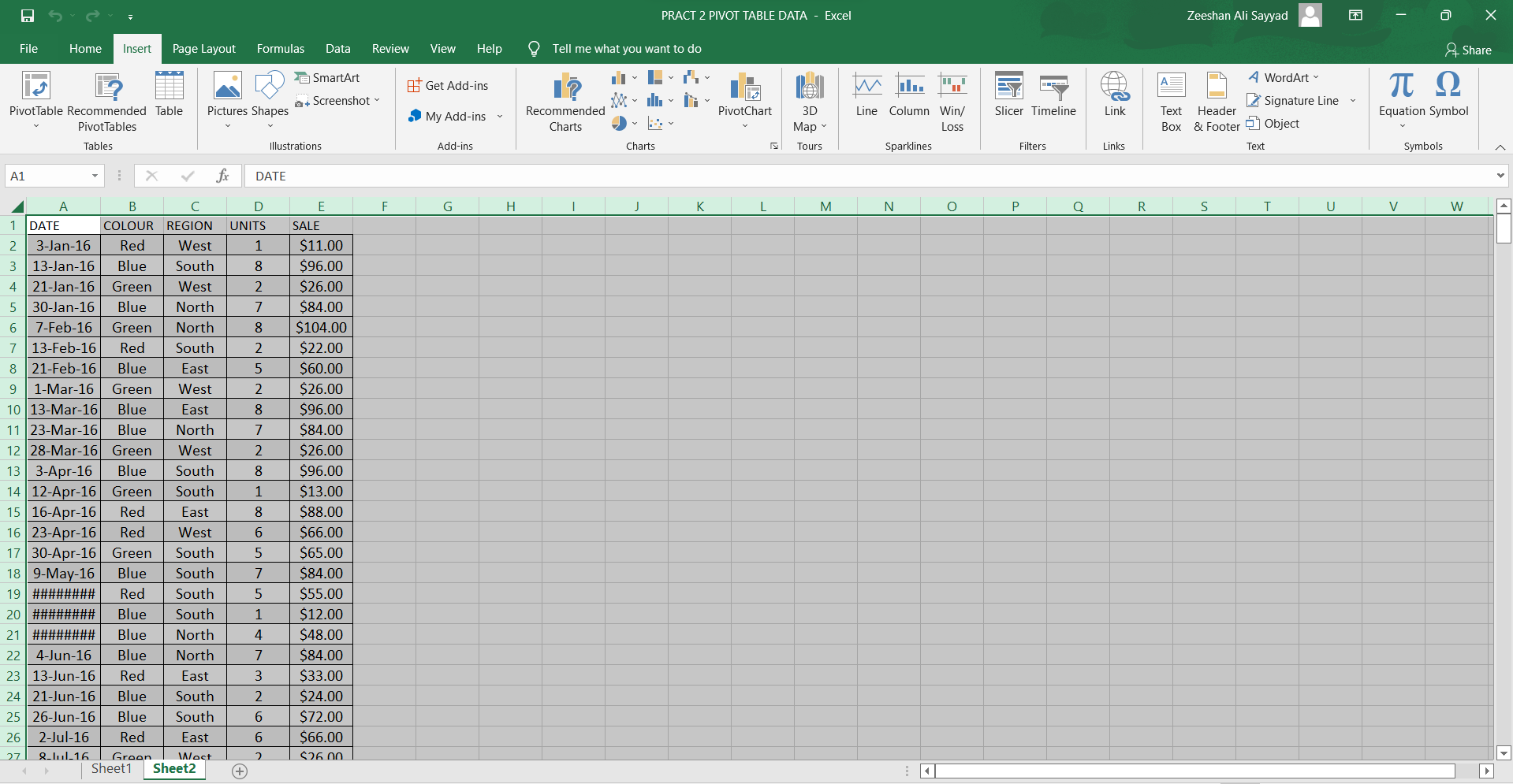
Theory:

A pivot table is a statistics tool that summarizes and reorganizes selected columns and rows of data in a spreadsheet or database table to obtain a desired report. The tool does not actually change the spreadsheet or database itself, it simply “pivots” or turns the data to view it from different perspectives.

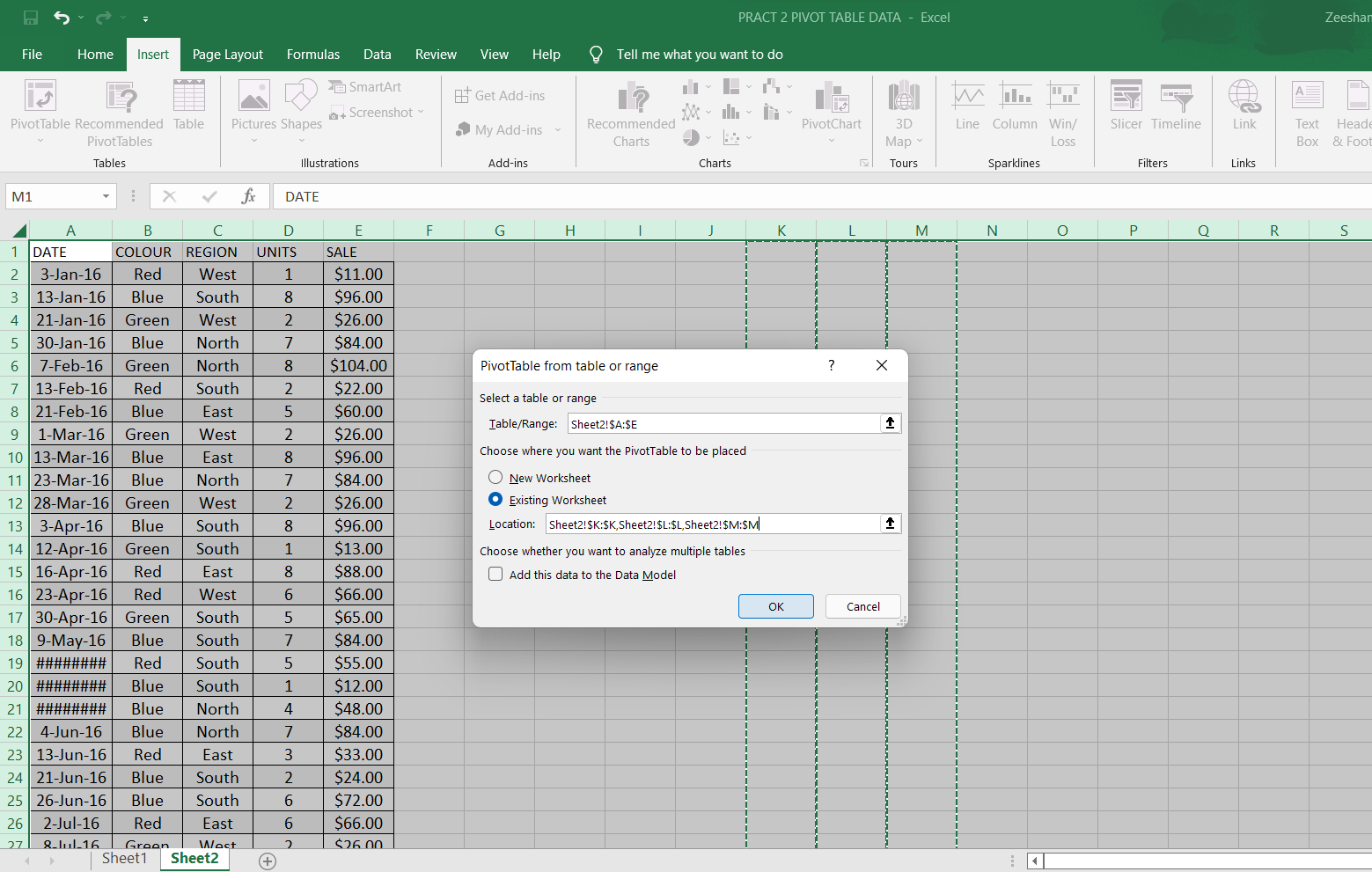
Pivot tables are especially useful with large amounts of data that would be time-consuming to calculate by hand. A few data processing functions a pivot table can perform include identifying sums, averages, ranges or outliers. The table then arranges this information in a simple, meaningful layout that draws attention to key values.

Steps:

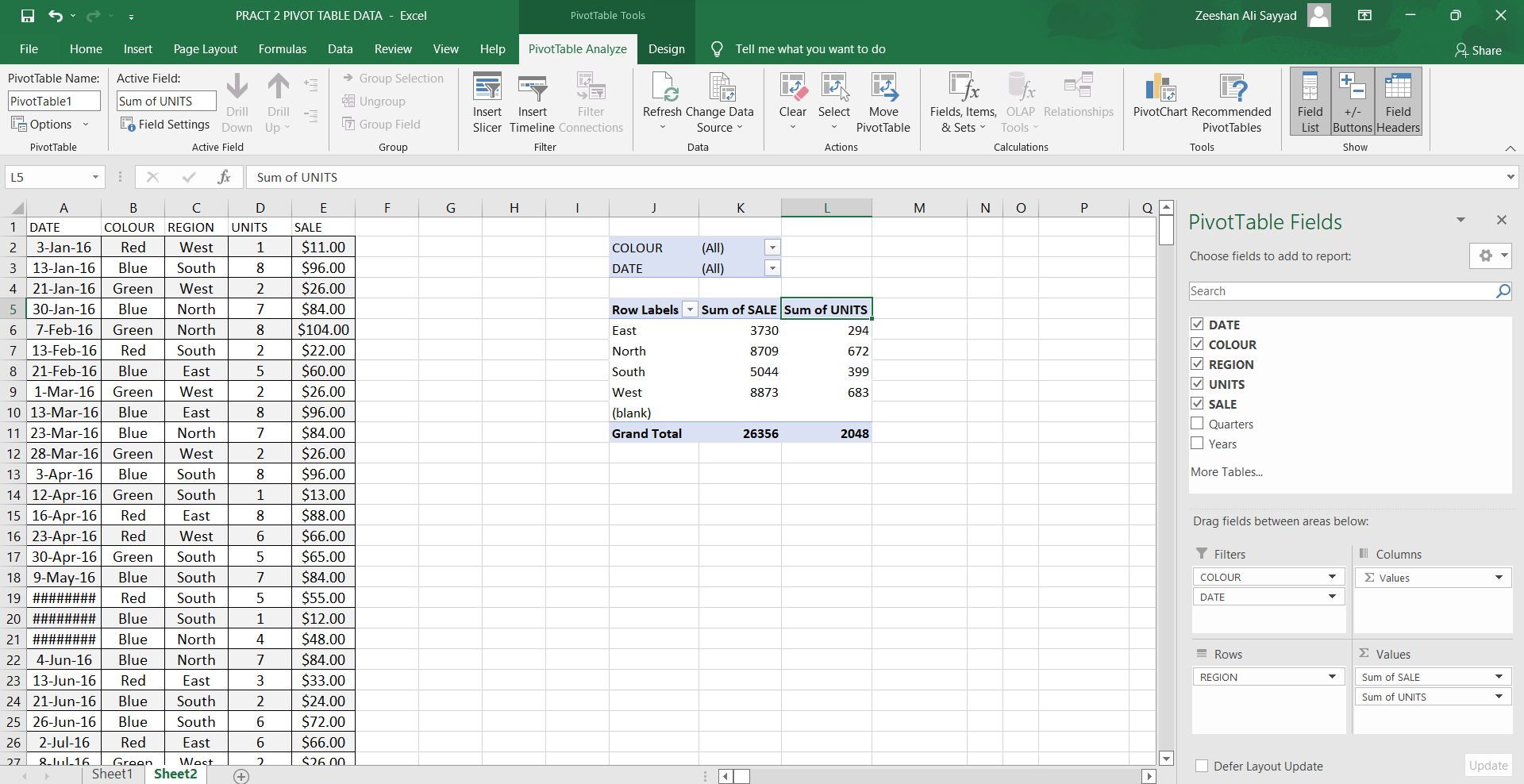
1) Select source data in excel sheet.



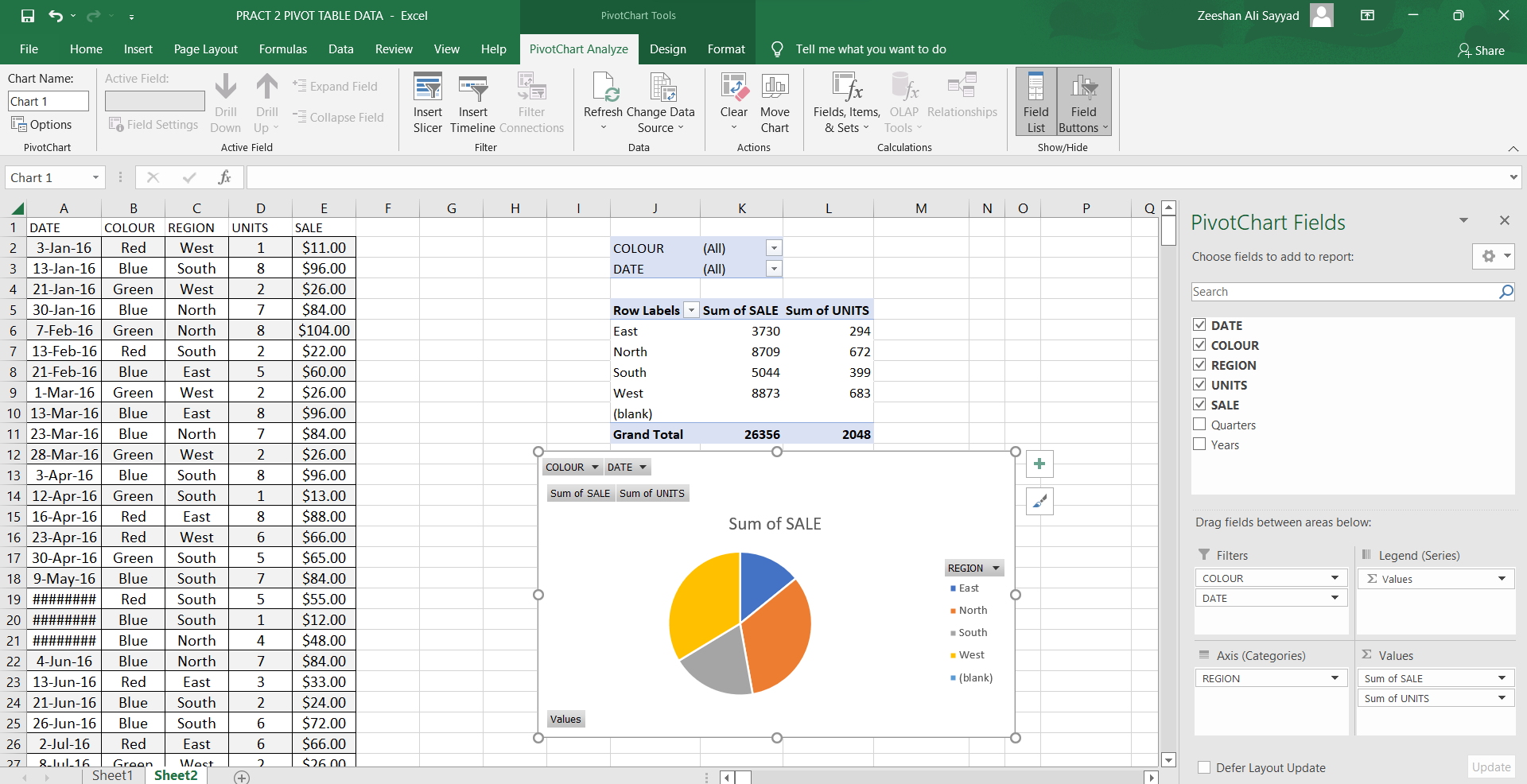
2) Click on Insert tab > Pivot Table > Select source and location > Click Ok



3) Now Drop PivotTable Fields into Filters, Columns, Rows, Values Respectively.



4) Insert Pie Chart or Bar graph.



Conclusion: Pivot table implemented successfully.

Practical 4

Aim: Develop an application to construct a multidimensional data.

Theory:

**What is Multi-Dimensional Data Model?**

A multidimensional model views data in the form of a data-cube. A data cube enables data to be modeled and viewed in multiple dimensions. It is defined by dimensions and facts.

The dimensions are the perspectives or entities concerning which an organization keeps records. For example, a shop may create a sales data warehouse to keep records of the store's sales for the dimension time, item, and location. These dimensions allow the save to keep track of things, for example, monthly sales of items and the locations at which the items were sold. Each dimension has a table related to it, called a dimensional table, which describes the dimension further. For example, a dimensional table for an item may contain the attributes item\_name, brand, and type.

A multidimensional data model is organized around a central theme, for example, sales. This theme is represented by a fact table. Facts are numerical measures. The fact table contains the names of the facts or measures of the related dimensional tables.

#Steps:

1) Create new database.

CREATE TABLE DIM\_customer

(

custid VARCHAR(6),

fname VARCHAR(30),

mname VARCHAR(30),

ltname VARCHAR(30),

city VARCHAR(15),

mobileno VARCHAR(10),

occupation VARCHAR(10),

dob DATE

)

CREATE TABLE DIM\_branch

(

bid VARCHAR(6),

bname VARCHAR(30),

bcity VARCHAR(30),

)

CREATE TABLE DIM\_account

(

acnumber VARCHAR(6),

custid VARCHAR(6),

bid VARCHAR(6),

opening\_balance VARCHAR(7),

aod DATE,

atype VARCHAR(10),

astatus VARCHAR(10)

)

CREATE TABLE DIM\_trandetails

(

tnumber VARCHAR(6),

acnumber VARCHAR(6),

dot DATE,

medium\_of\_transaction VARCHAR(20),

transaction\_type VARCHAR(20),

transaction\_amount VARCHAR(7)

)

CREATE TABLE DIM\_loan

(

loan\_id VARCHAR(10),

custid VARCHAR(6),

bid VARCHAR(6),

loan\_amount VARCHAR(7)

)

CREATE TABLE FACT\_BANK

(

custid VARCHAR(6),

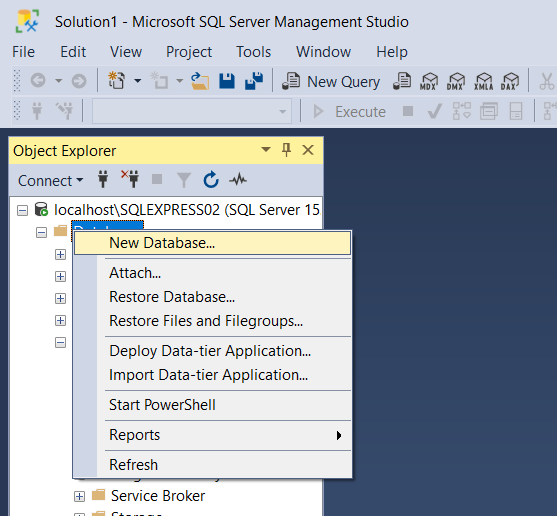
bid VARCHAR(6),

acnumber VARCHAR(6),

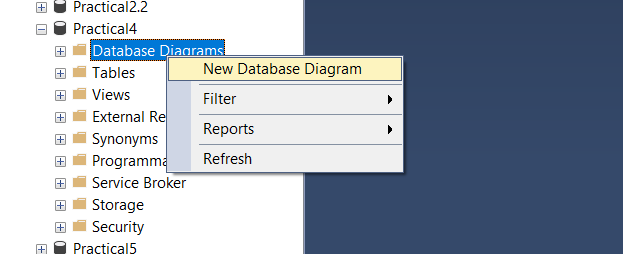
tnumber VARCHAR(6),

loan\_id VARCHAR(10)

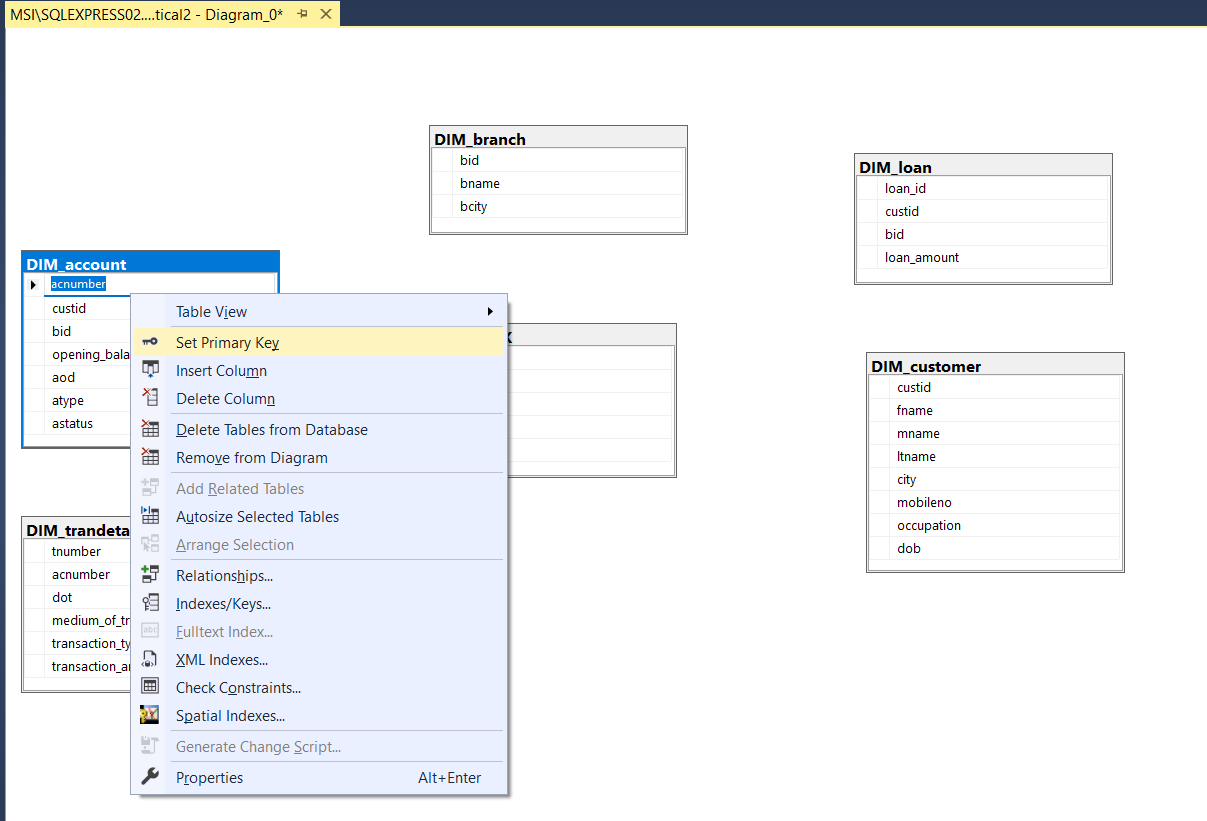
)



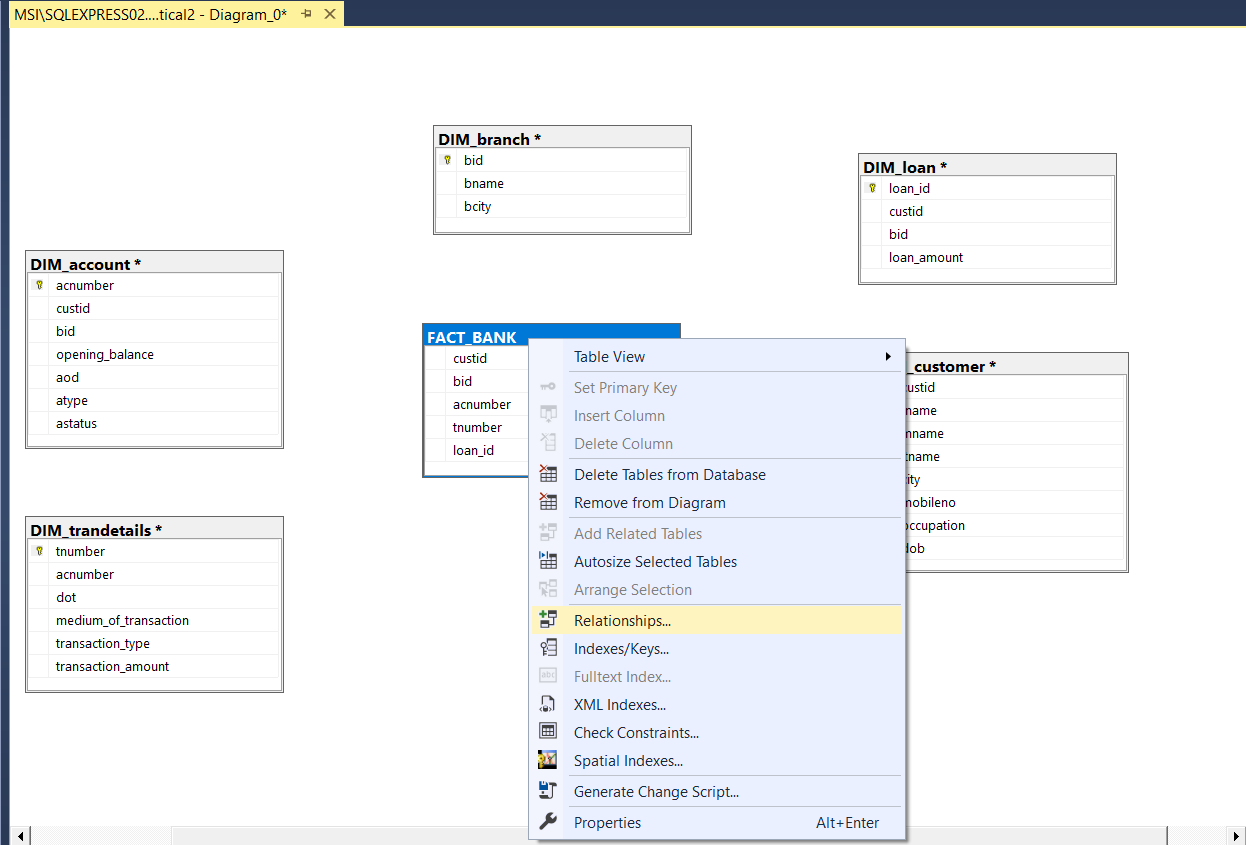
2) Create database diagram.

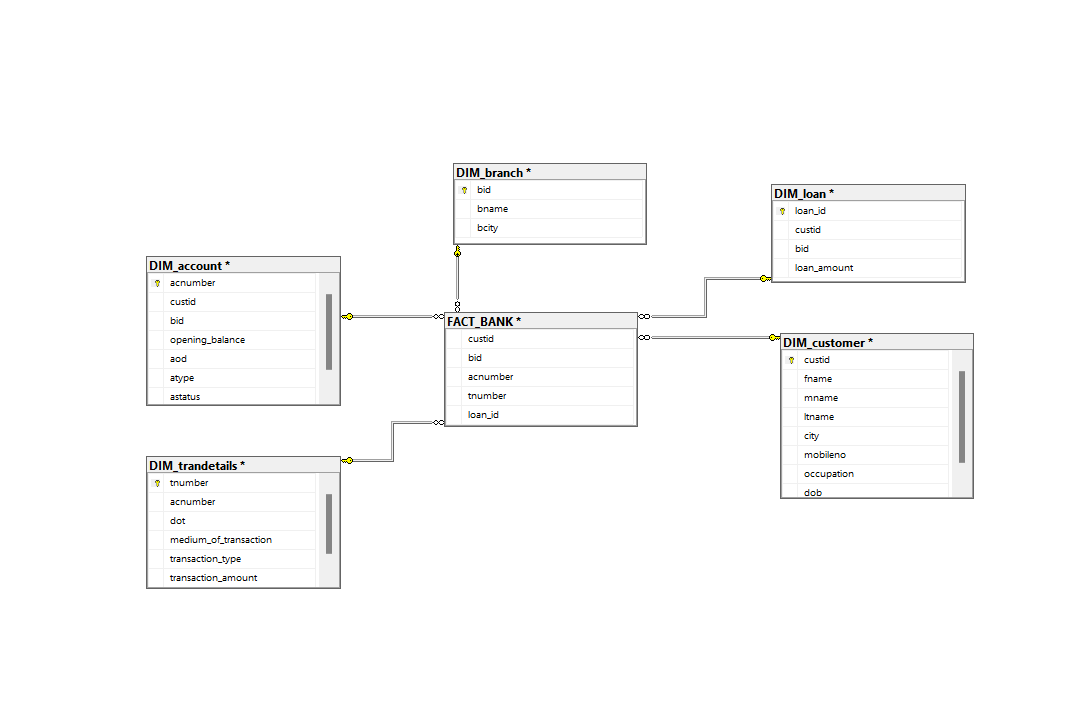


3) Set primary key for each table:

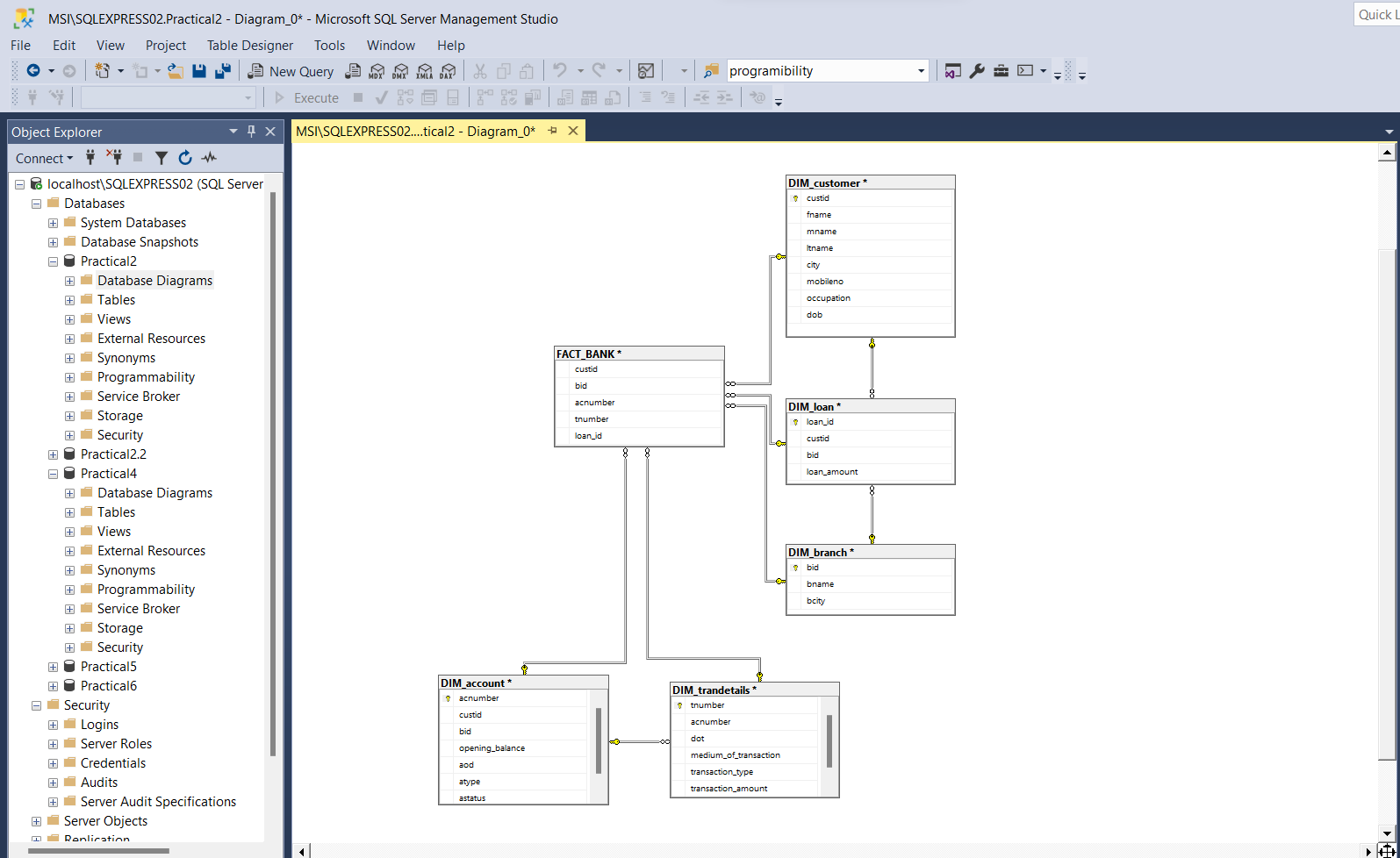


4)Set Relationships for fact table.





5) Set relationships of dimensional table with each other.



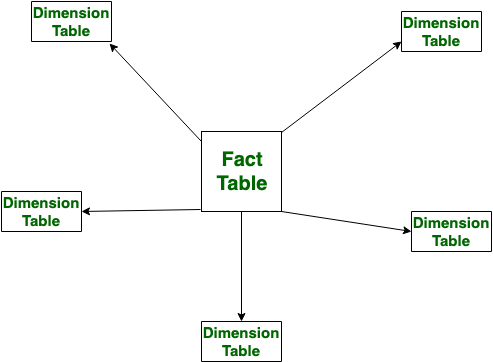
Conclusion: Multidimensional data constructed successfully.

Practical 5

Aim: Design and create cube by identifying measures and dimensions for star schema.

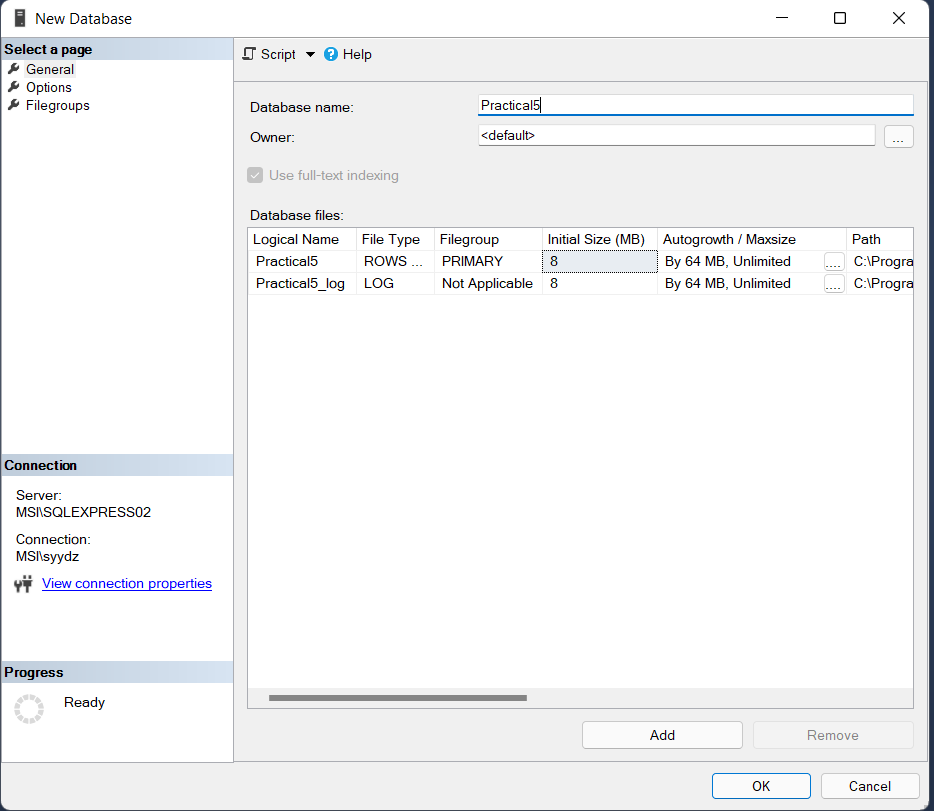
Theory:

**Star Schema:**  
Star schema is the type of multidimensional model which is used for data warehouse. In star schema, the fact tables and the dimension tables are contained. In this schema fewer foreign-key join is used. This schema forms a star with fact table and dimension tables.



#Steps

1) Create New Database



2) Right Click on Database > Script Database as > Create to > Select: New Query Editor Window.



3) Remove pre-written queries

4) Create tables:

CREATE TABLE DIM\_EMPLOYEE

(EMP\_ID INT,

EMP\_NAME VARCHAR(25))

CREATE TABLE DIM\_BRANCH

(BRANCH\_ID INT,

BRANCH\_NAME VARCHAR(25))

CREATE TABLE DIM\_PRODUCT

(PROD\_ID INT,

PROD\_NAME VARCHAR(25))

CREATE TABLE DIM\_CUSTOMER

(CUST\_ID INT,

CUST\_NAME VARCHAR(25))

CREATE TABLE FACT\_SHOP

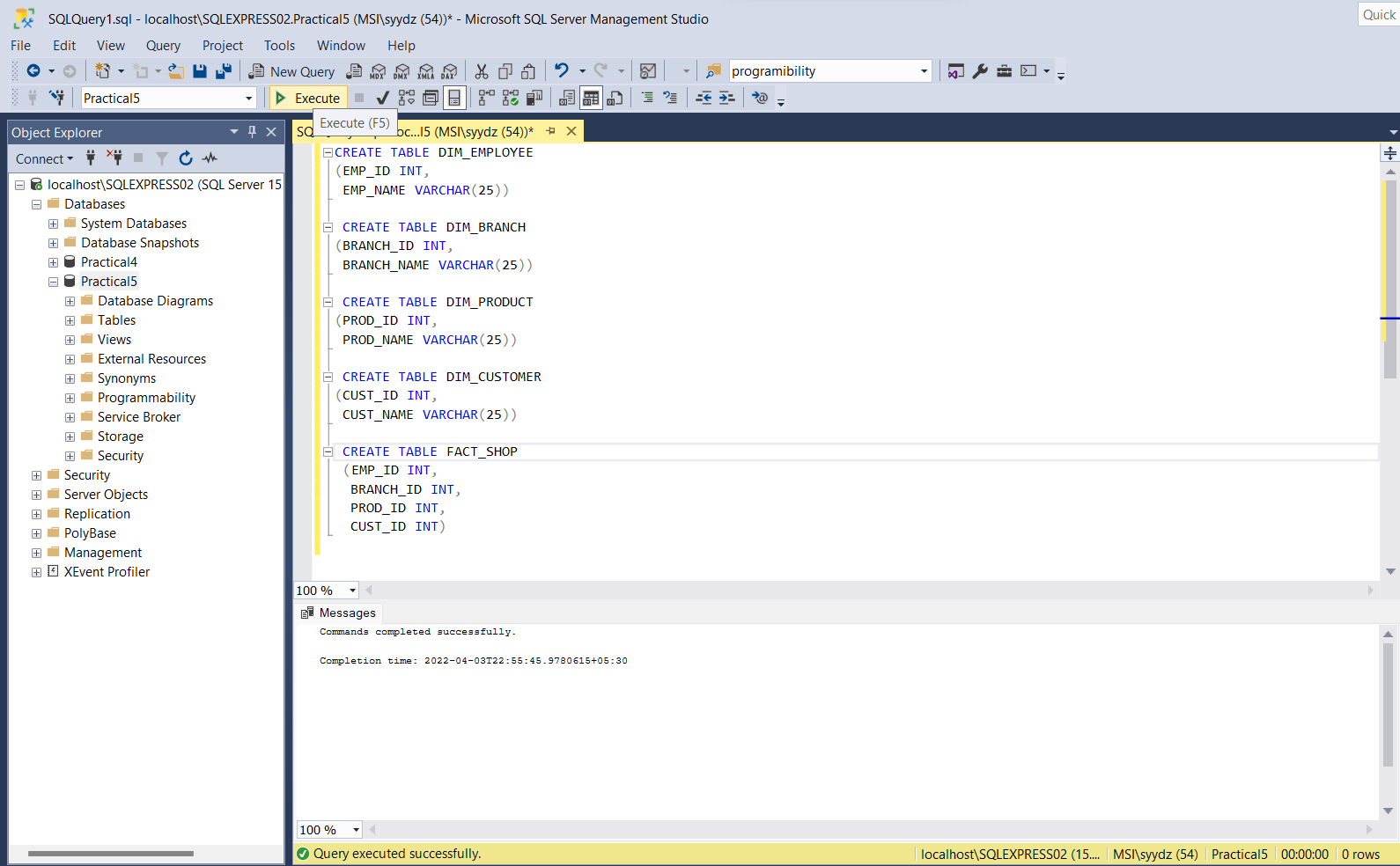
(EMP\_ID INT,

BRANCH\_ID INT,

PROD\_ID INT,

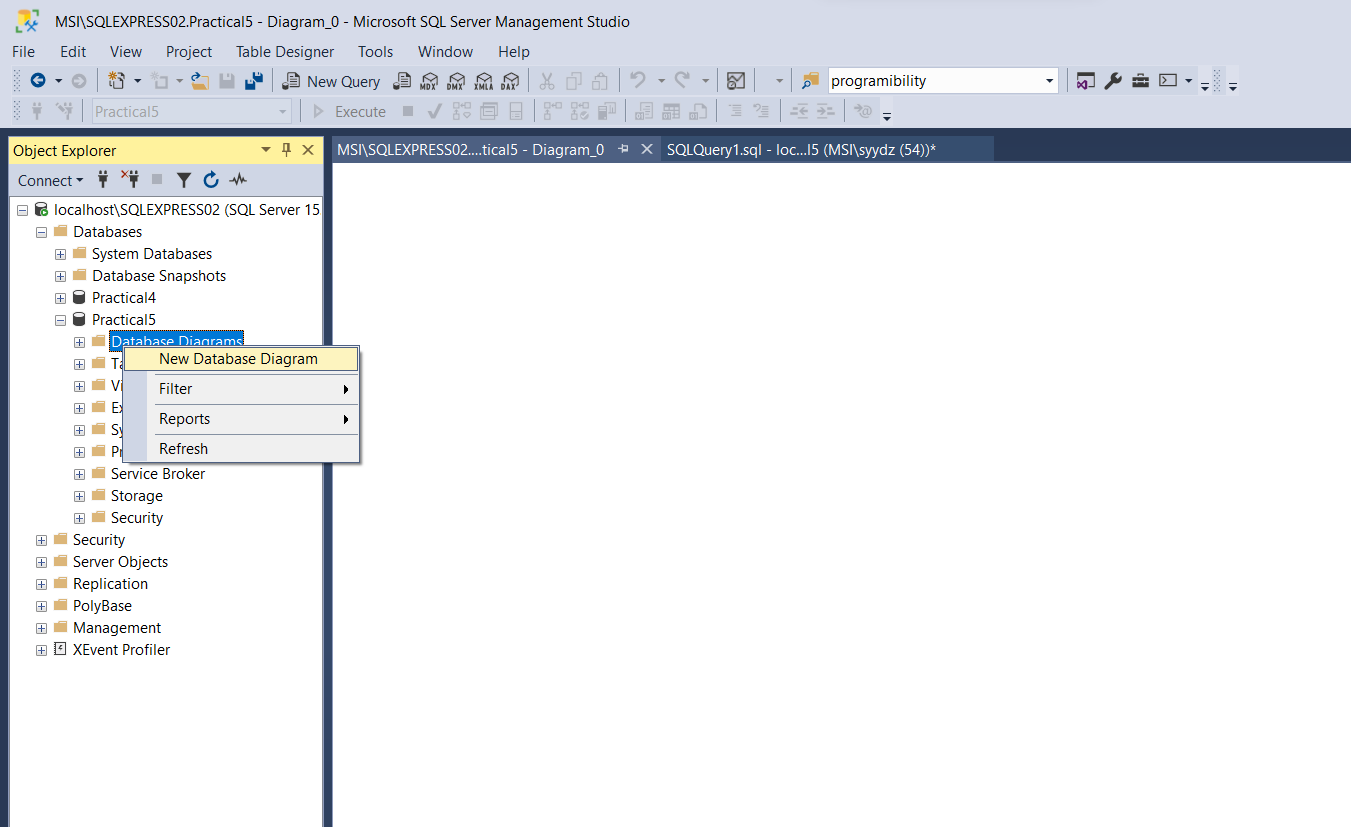
CUST\_ID INT)

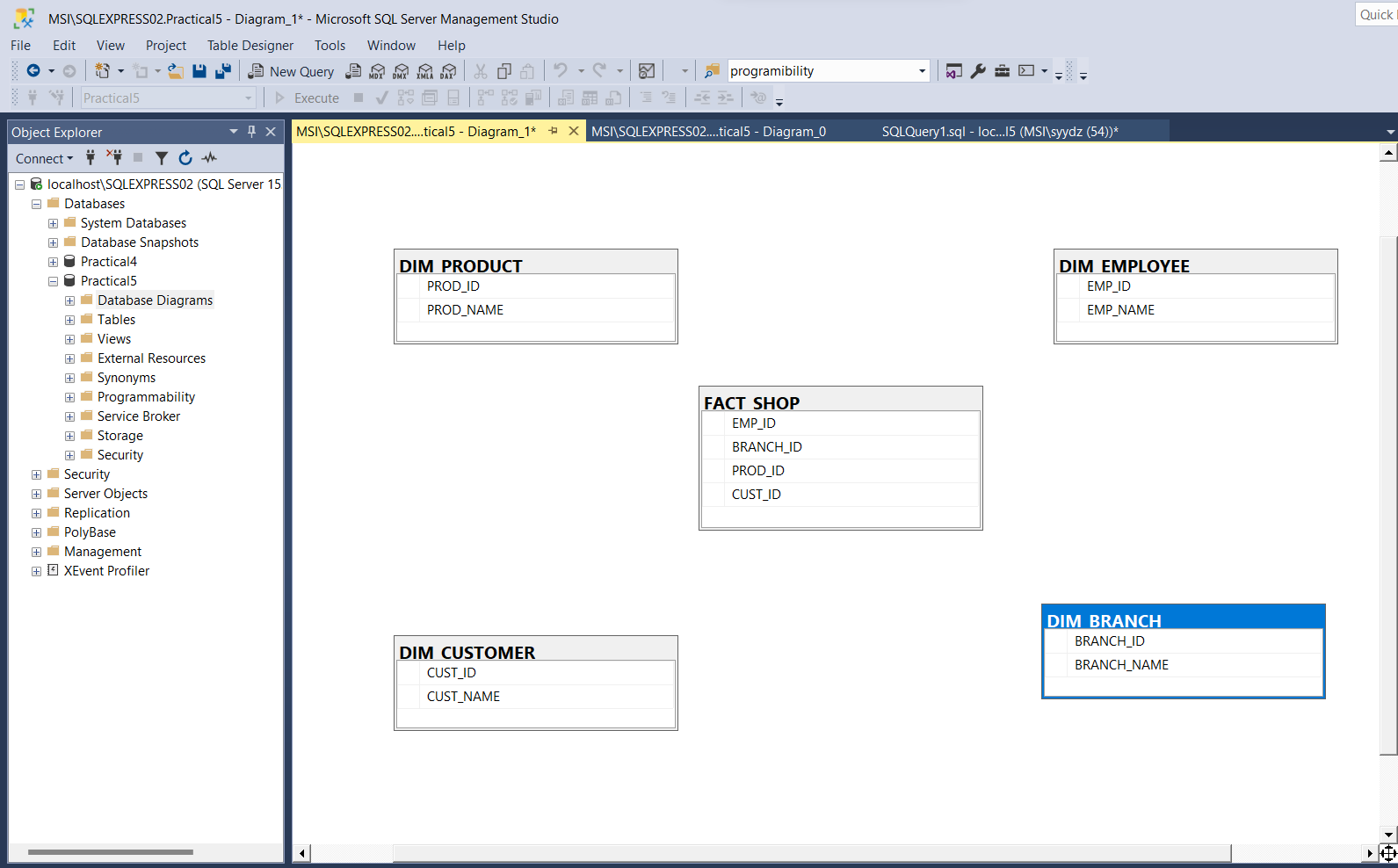
4) Execute queries



6) Create Database diagrams:

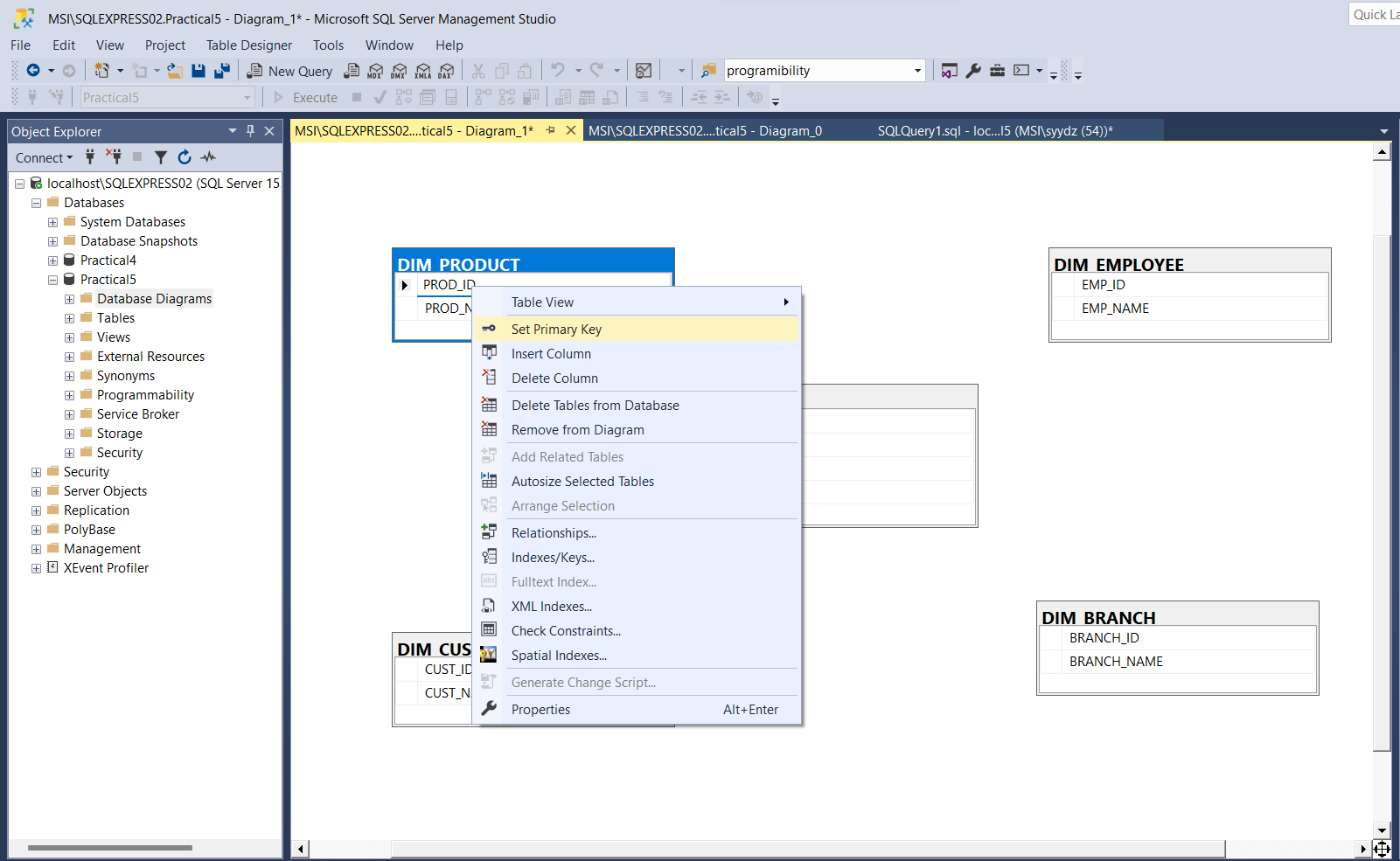
Right Click On Database Diagrams > New Database Diagram





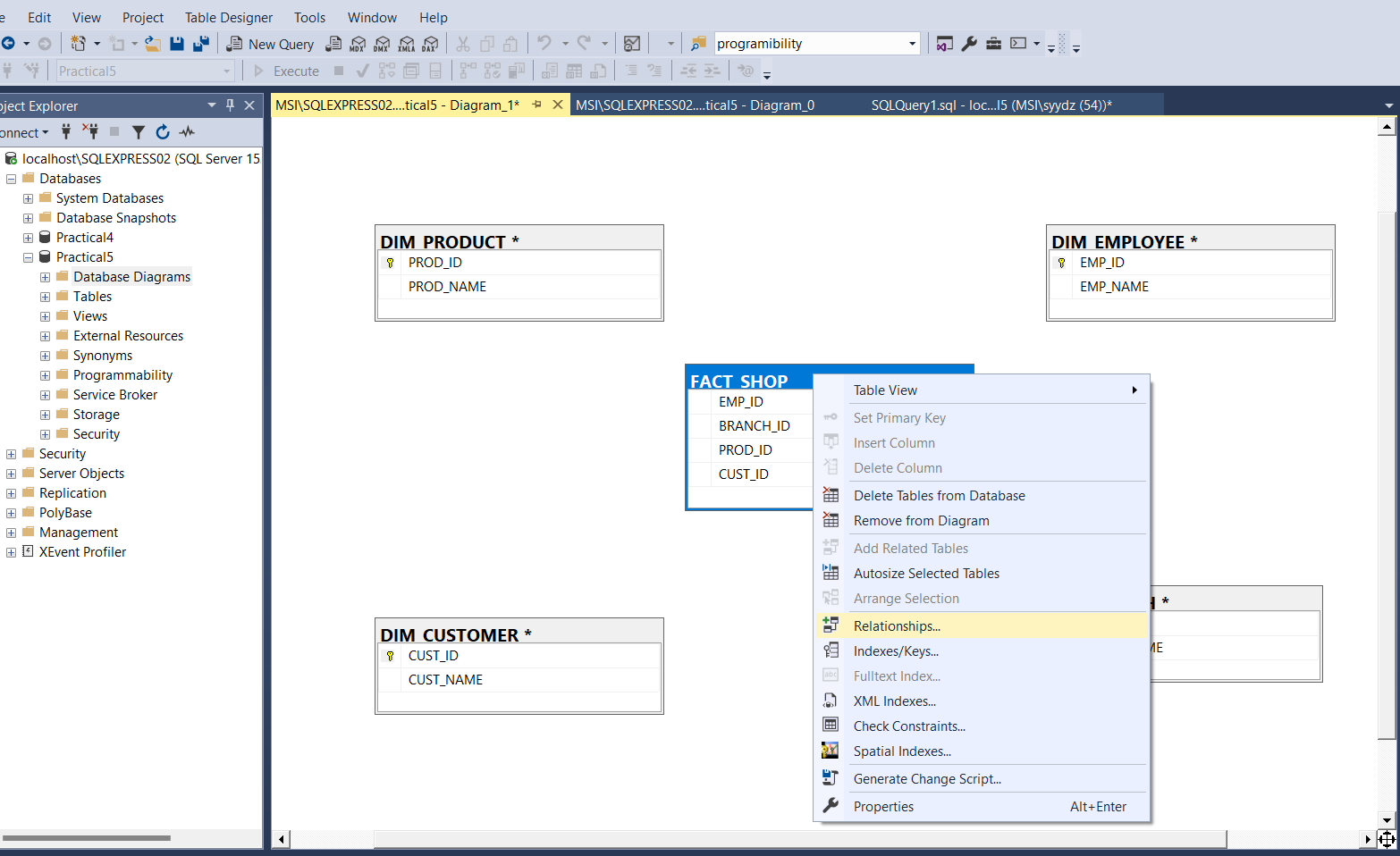
7) Set primary keys :

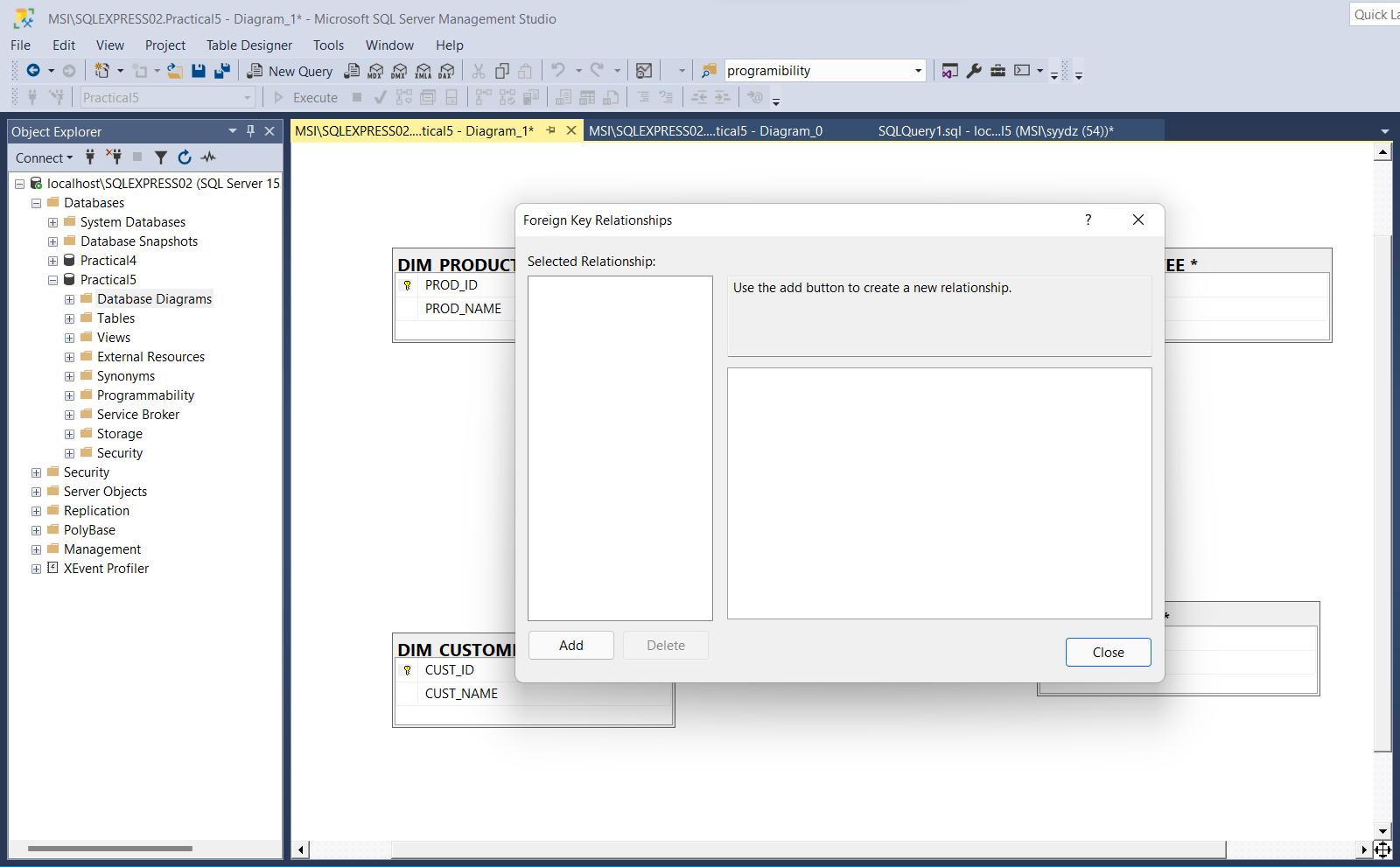
PROD\_ID, EMP\_ID, CUST\_ID, BRANCH\_ID



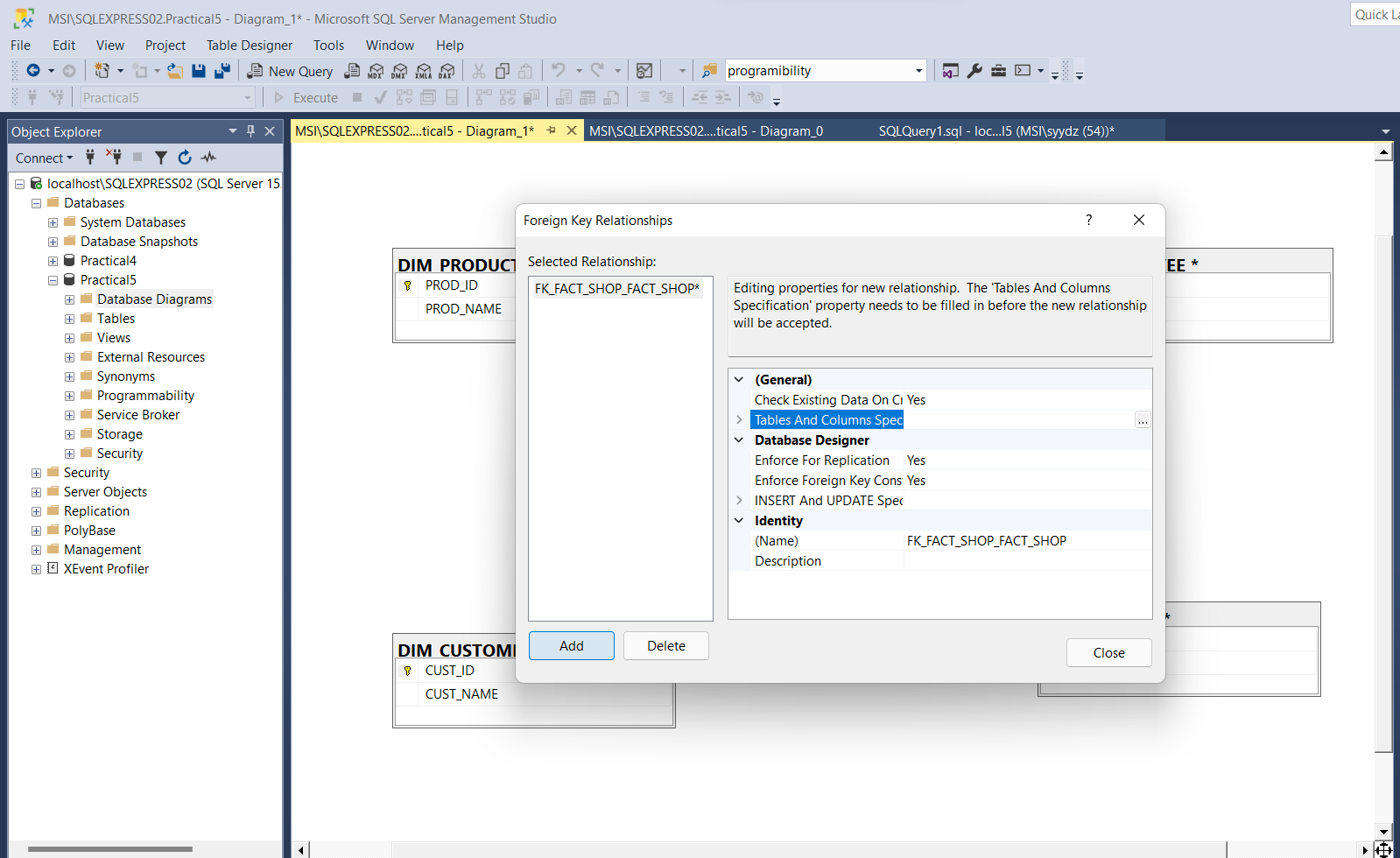
8) Create Relationships:

Right Click on Fact Shop table > Select Relationships > Add relations.

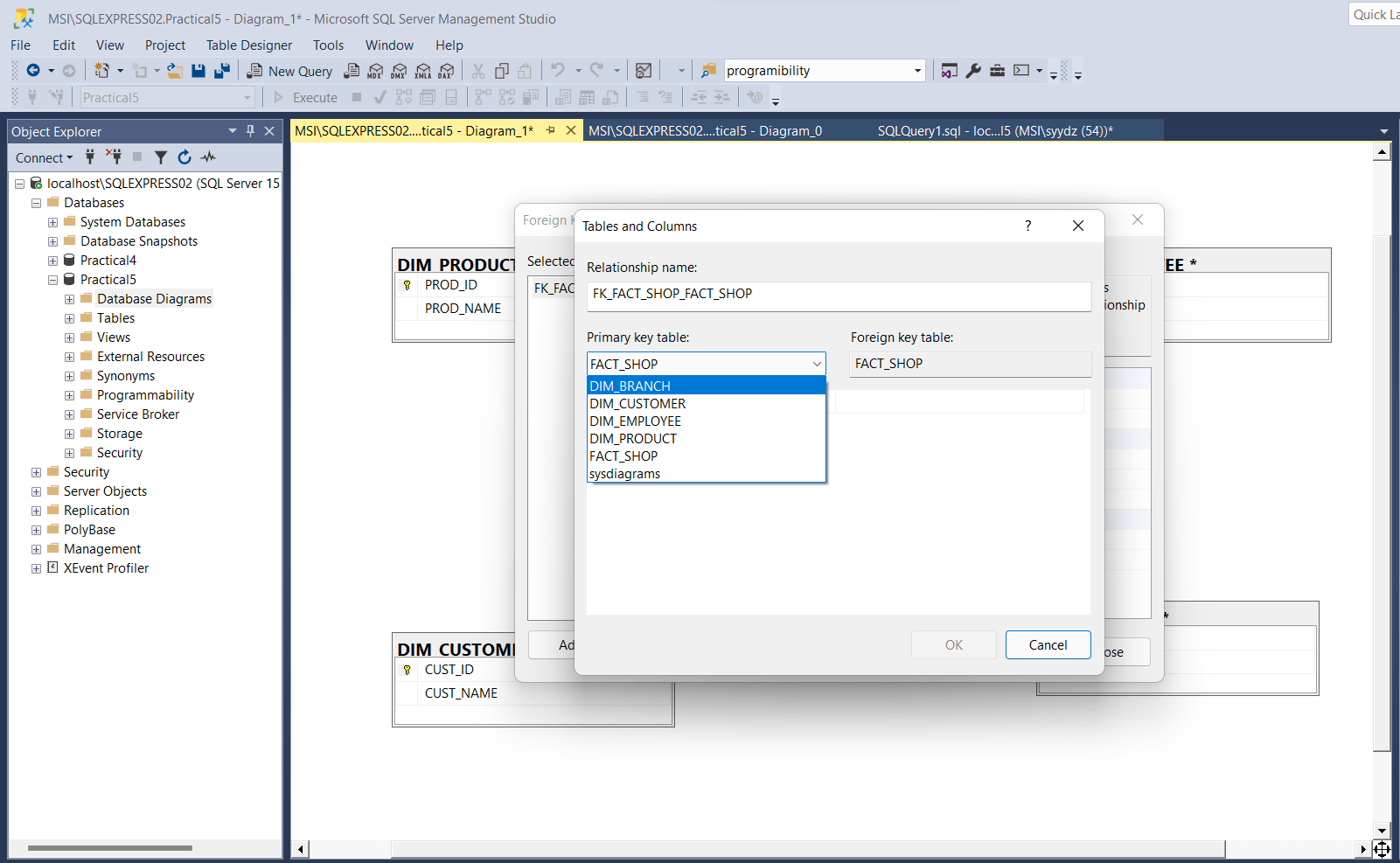


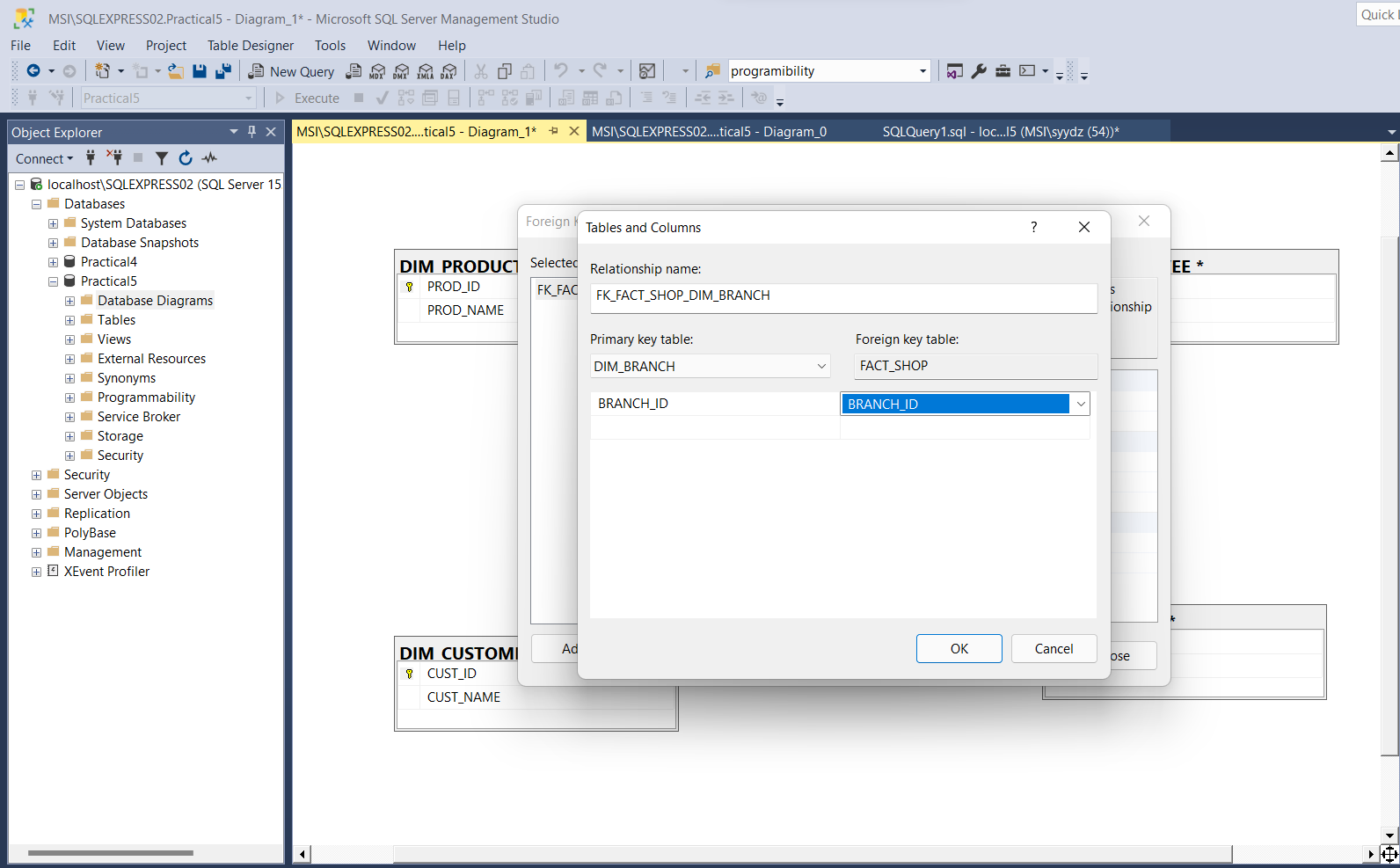


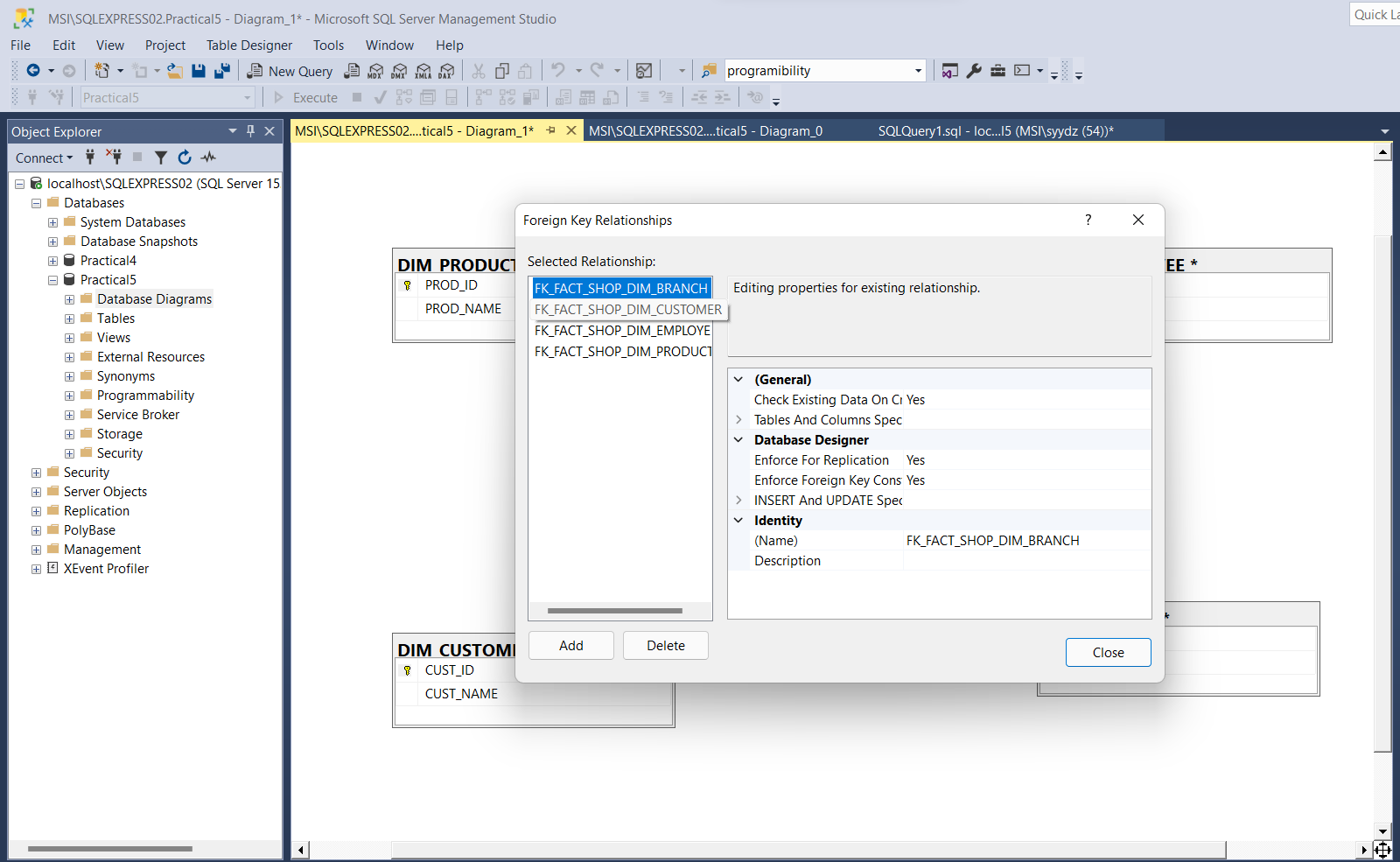
8) Click on Tables And Columns Sec( Expand …)



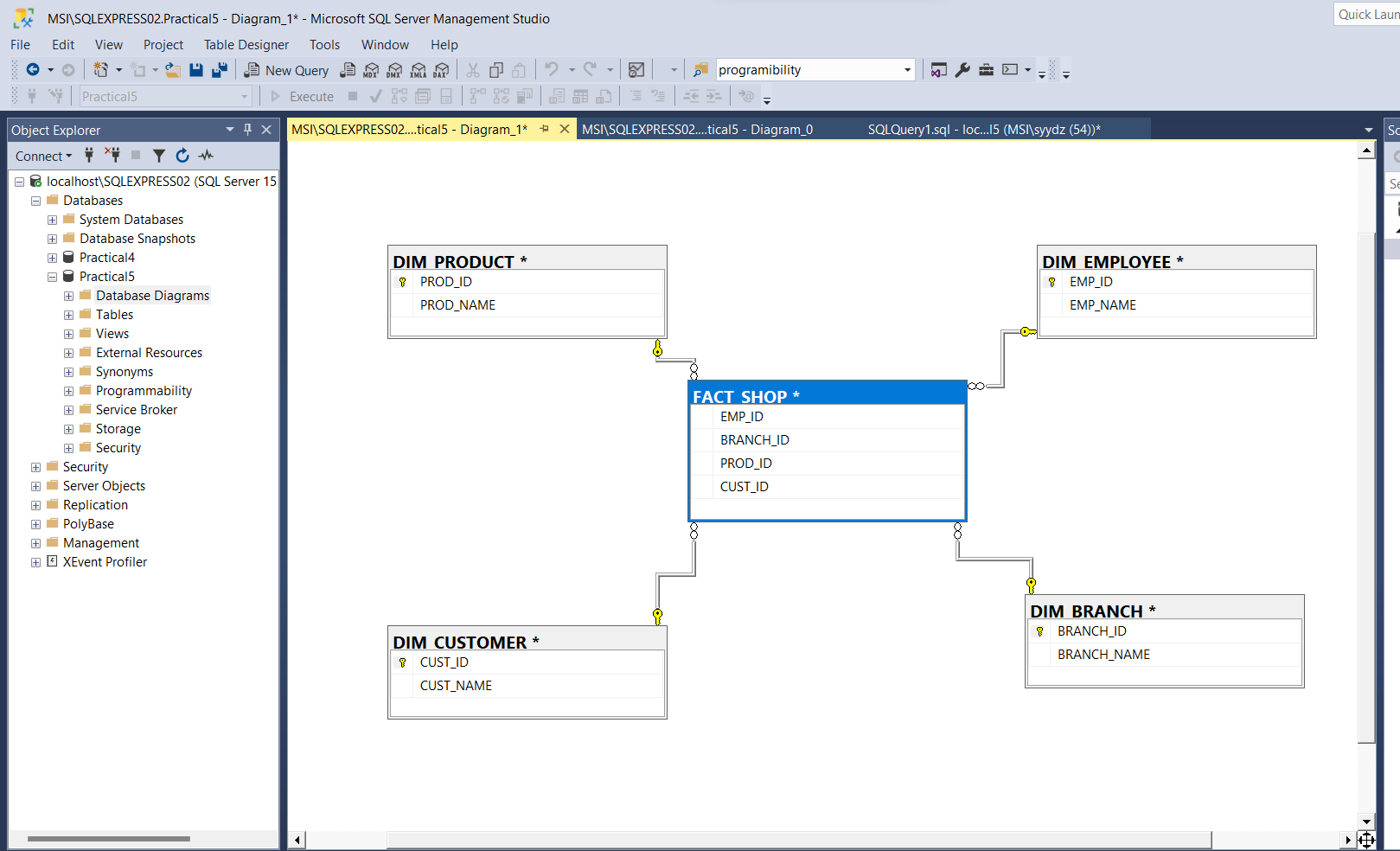
9) Select Primary key tables for every relation

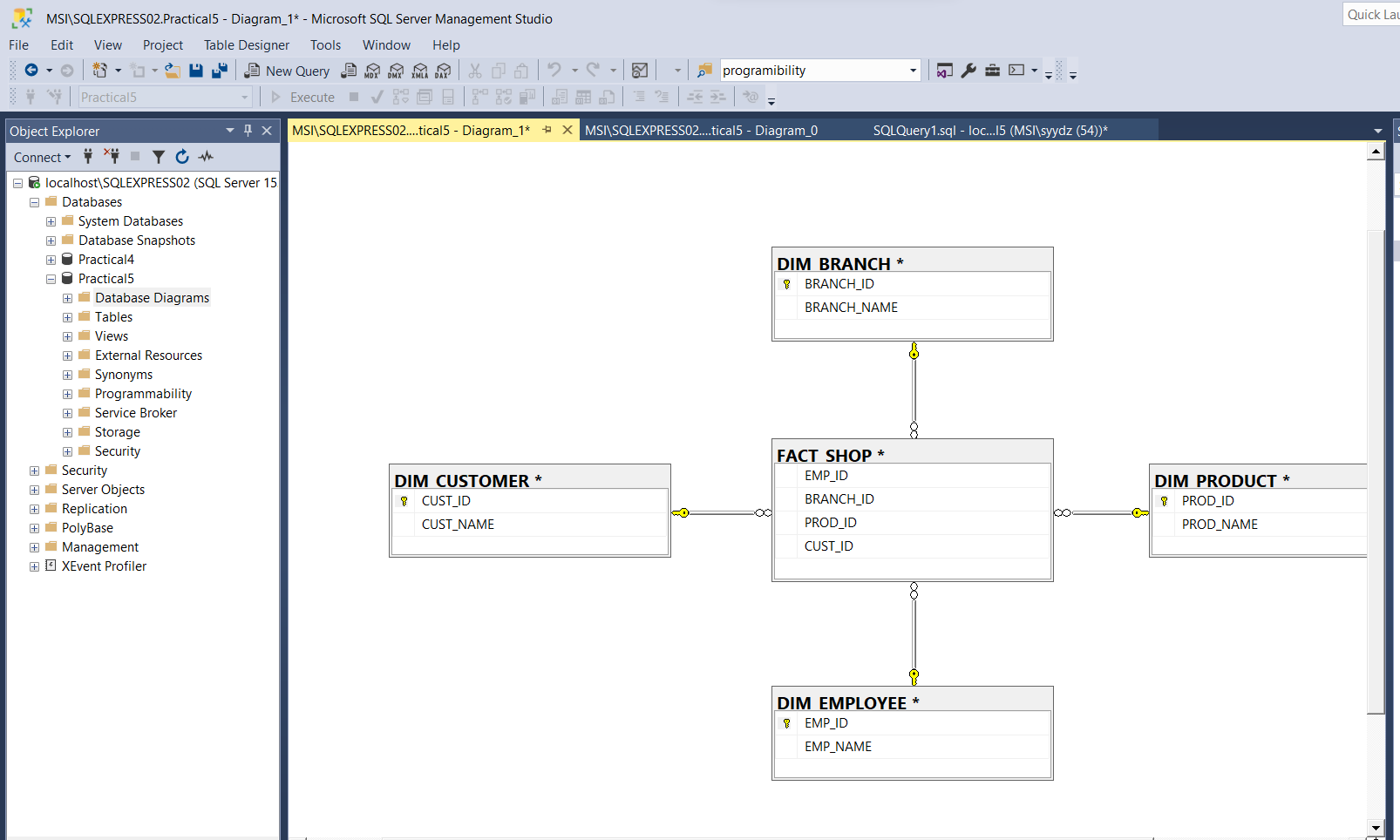






10) Close Relationships window





Conclusion: Star schema executed successfully.

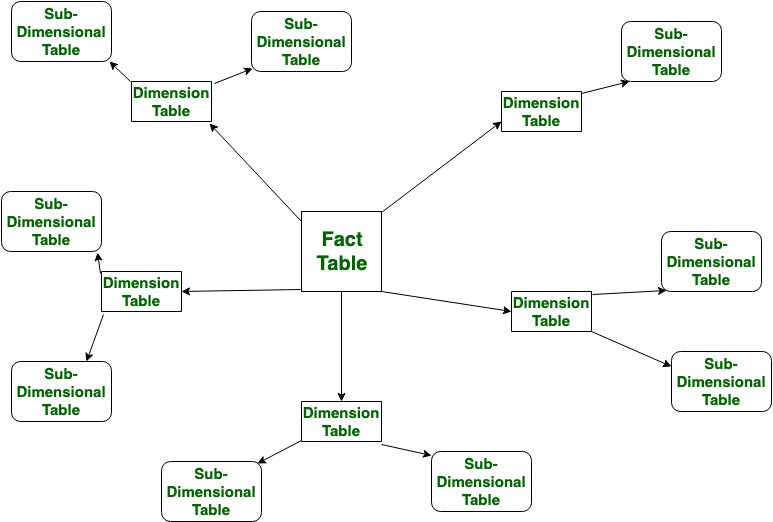
Practical 6

Aim: Design and create cube by identifying measures and dimensions for snow flake

schema.

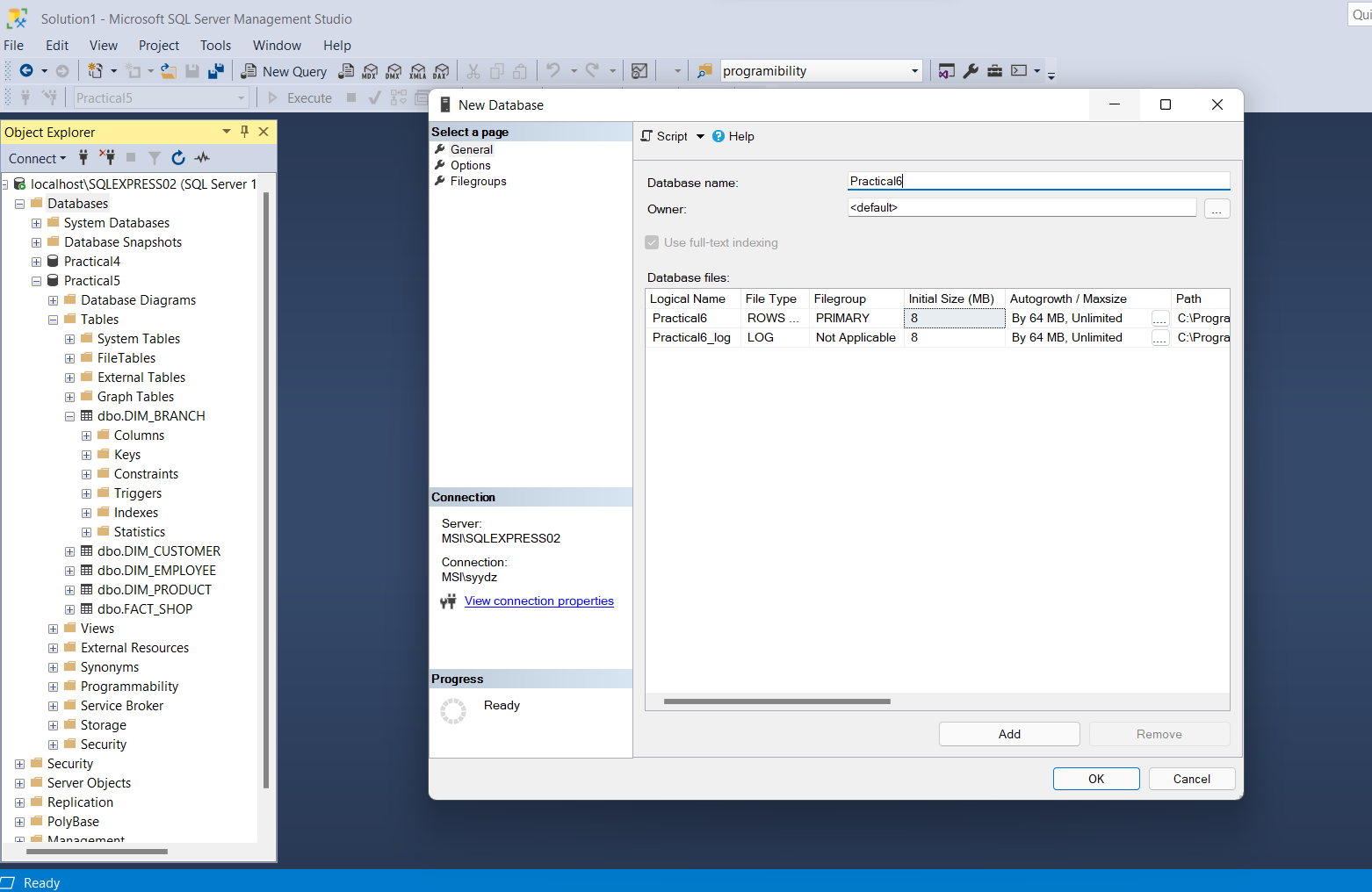
Theory:

**Snowflake Schema:**  
Snowflake Schema is also the type of multidimensional model which is used for data warehouse. In snowflake schema, The fact tables, dimension tables as well as sub dimension tables are contained. This schema forms a snowflake with fact tables, dimension tables as well as sub-dimension tables.

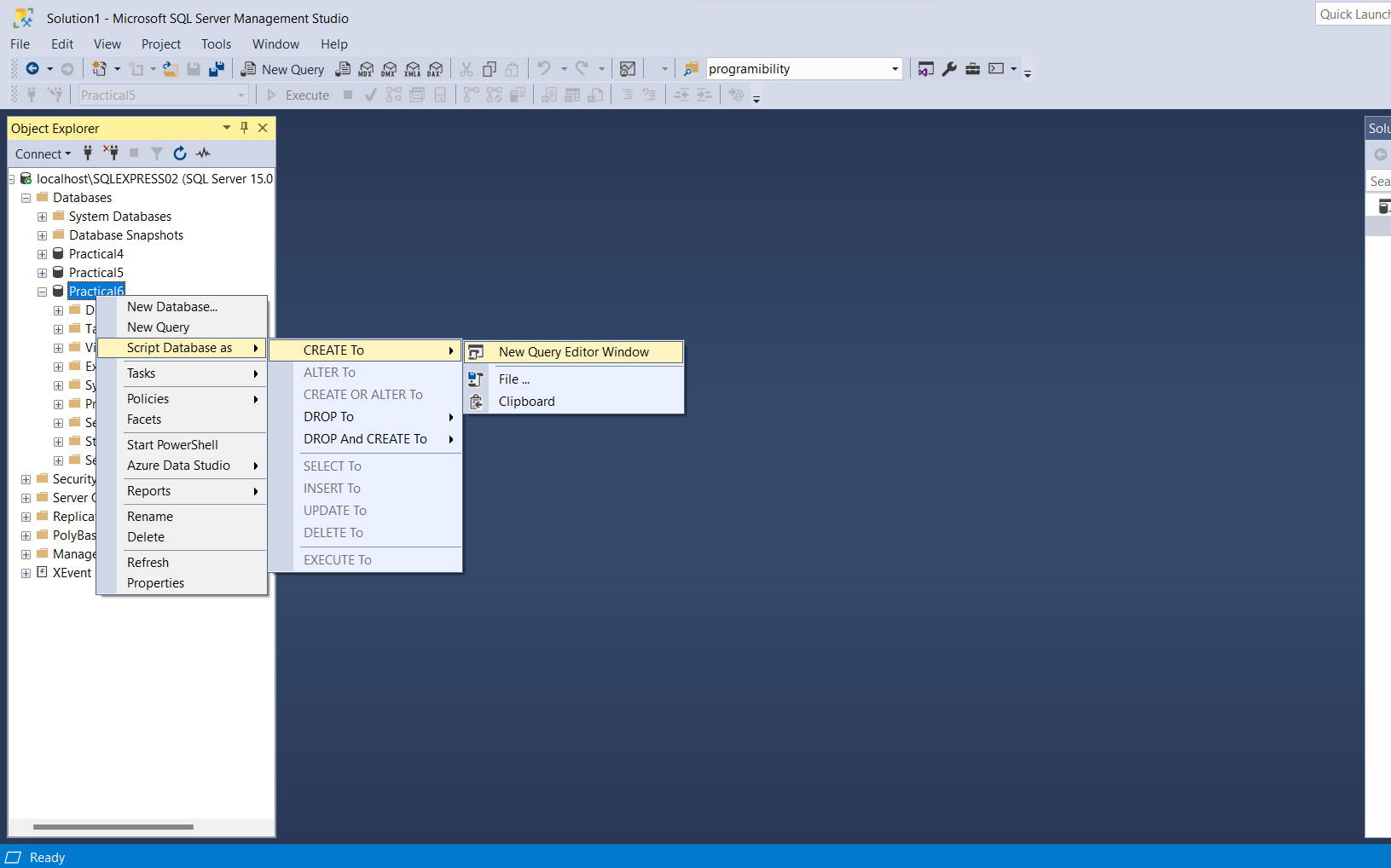


Steps:

1) Create database.



2) Script Database as > Create To > New Query Window



3) Create tables:

CREATE TABLE DIM\_EMPLOYEE

(EMP\_ID INT,

EMP\_NAME VARCHAR(25))

CREATE TABLE DIM\_BRANCH

(BRANCH\_ID INT,

BRANCH\_NAME VARCHAR(25))

CREATE TABLE DIM\_PRODUCT

(PROD\_ID INT,

PROD\_NAME VARCHAR(25),

BRAND\_ID INT)

CREATE TABLE DIM\_BRAND

(BRAND\_ID INT,

BRAND\_NAME VARCHAR(25))

CREATE TABLE DIM\_CUSTOMER

(CUST\_ID INT,

CUST\_NAME VARCHAR(25))

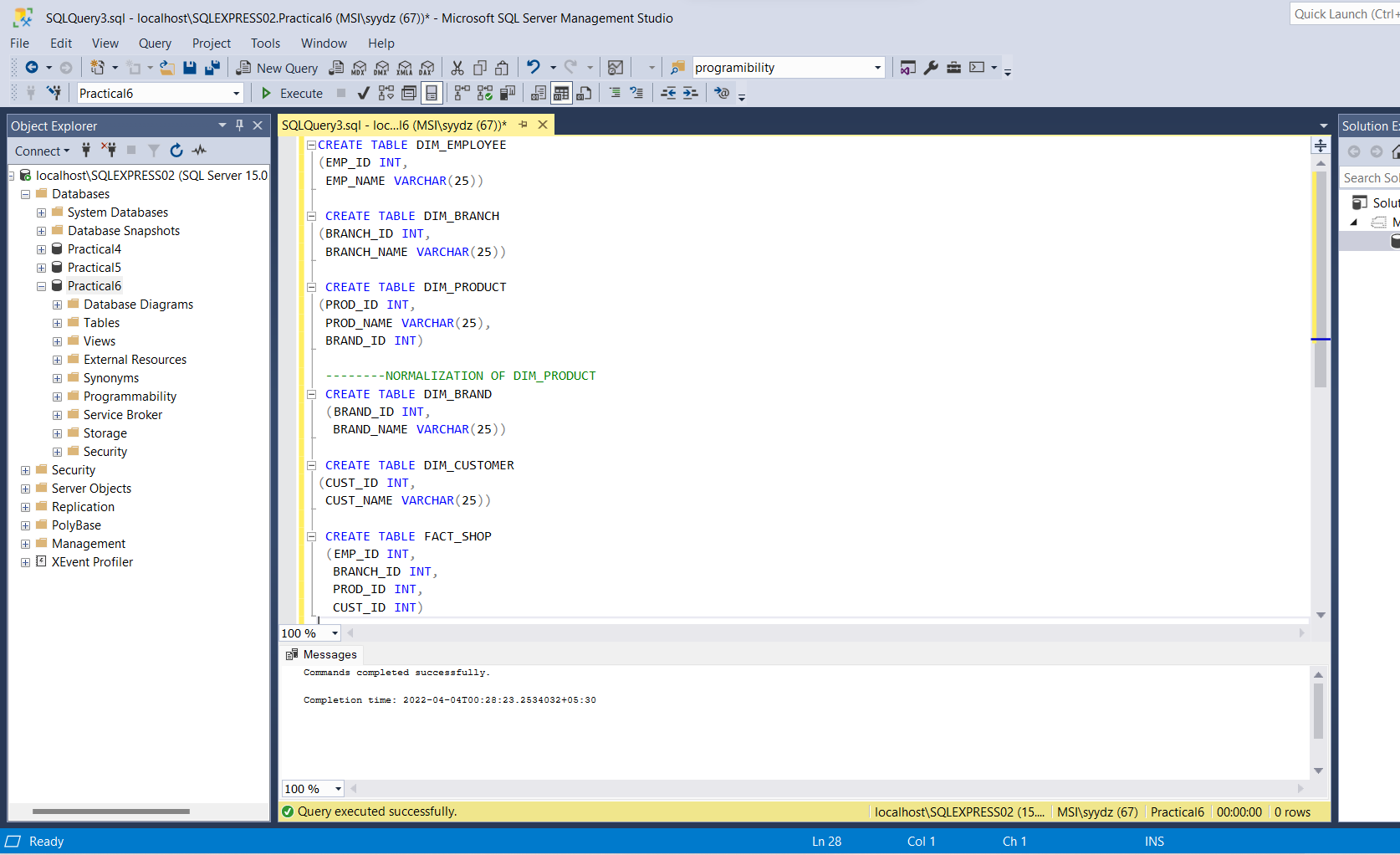
CREATE TABLE FACT\_SHOP

(EMP\_ID INT,

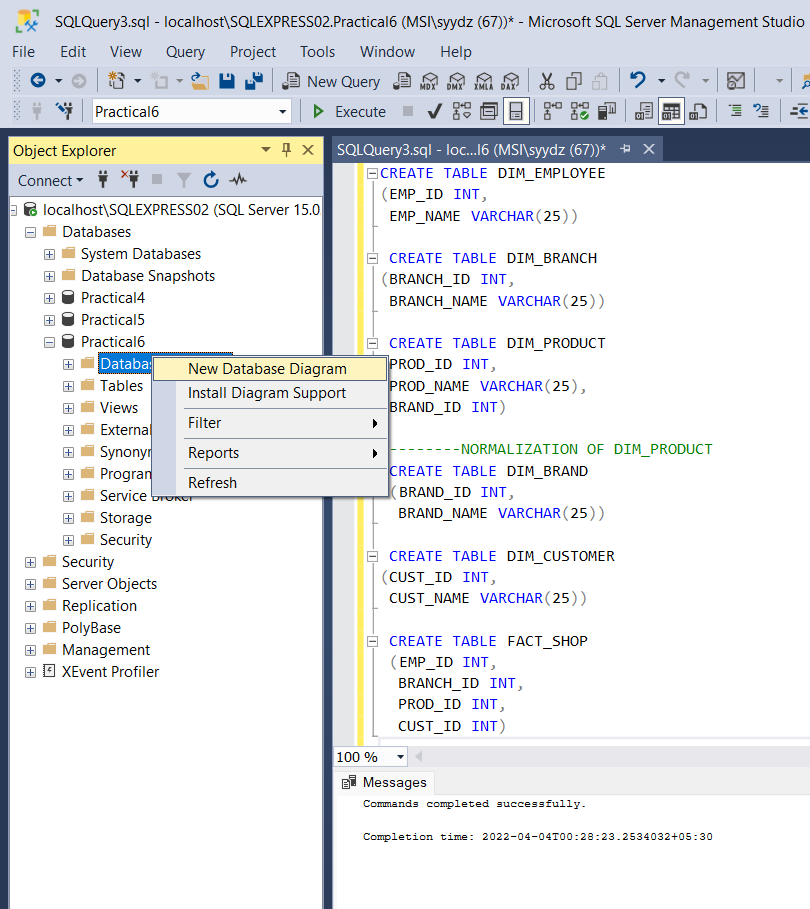
BRANCH\_ID INT,

PROD\_ID INT,

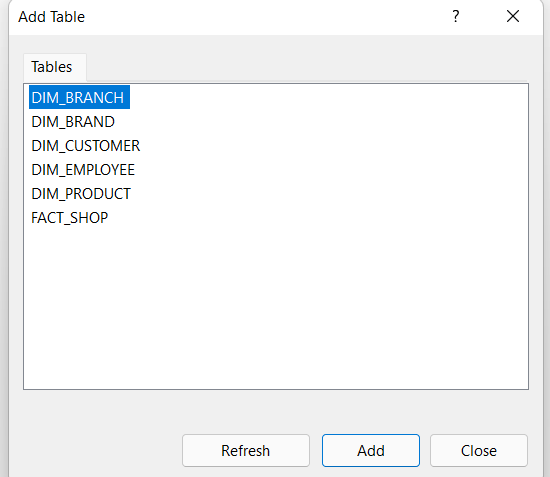
CUST\_ID INT)



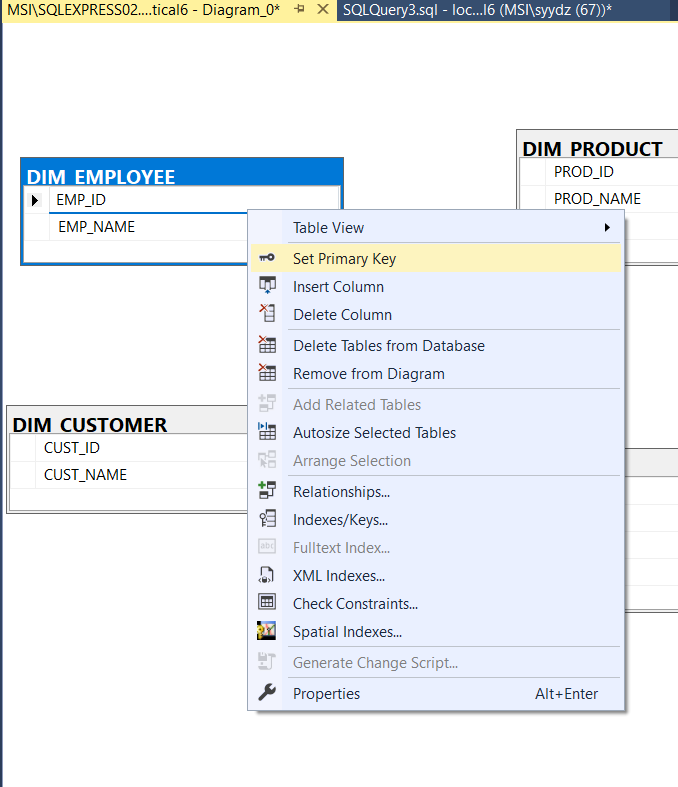
4) Create New database diagram



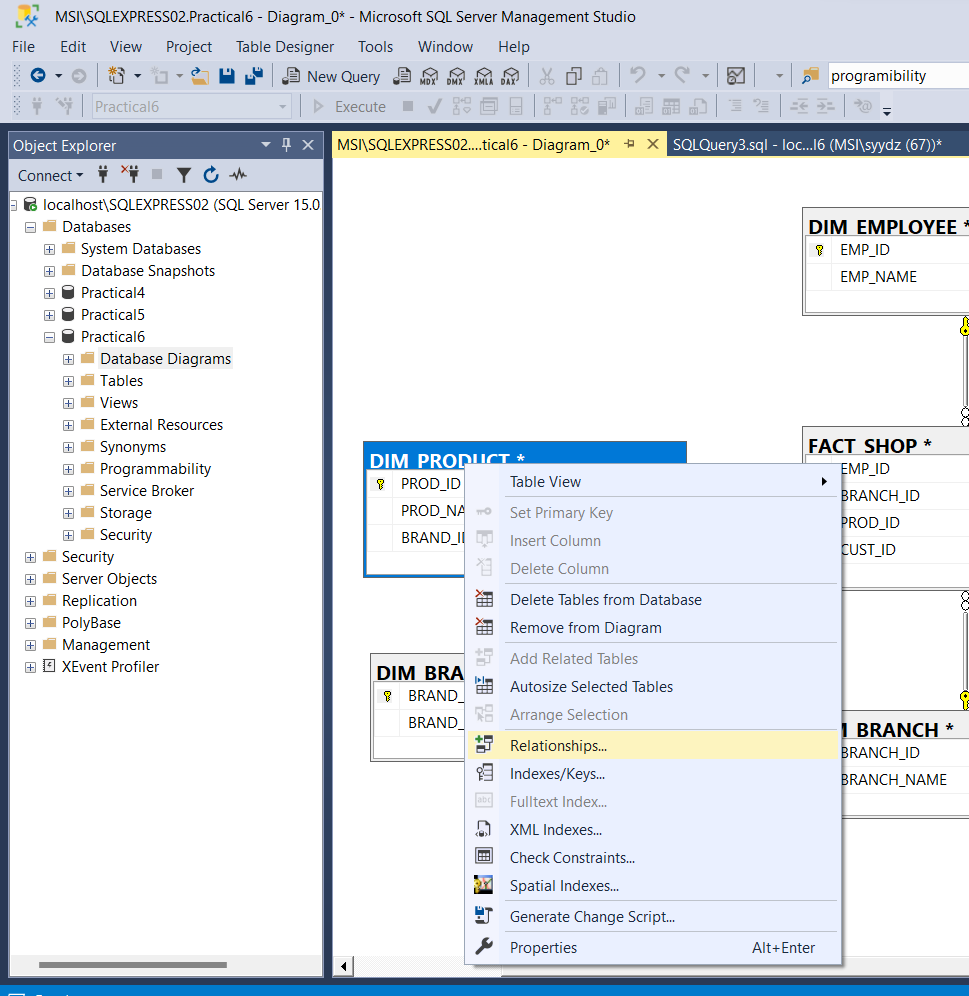
5) Add tables



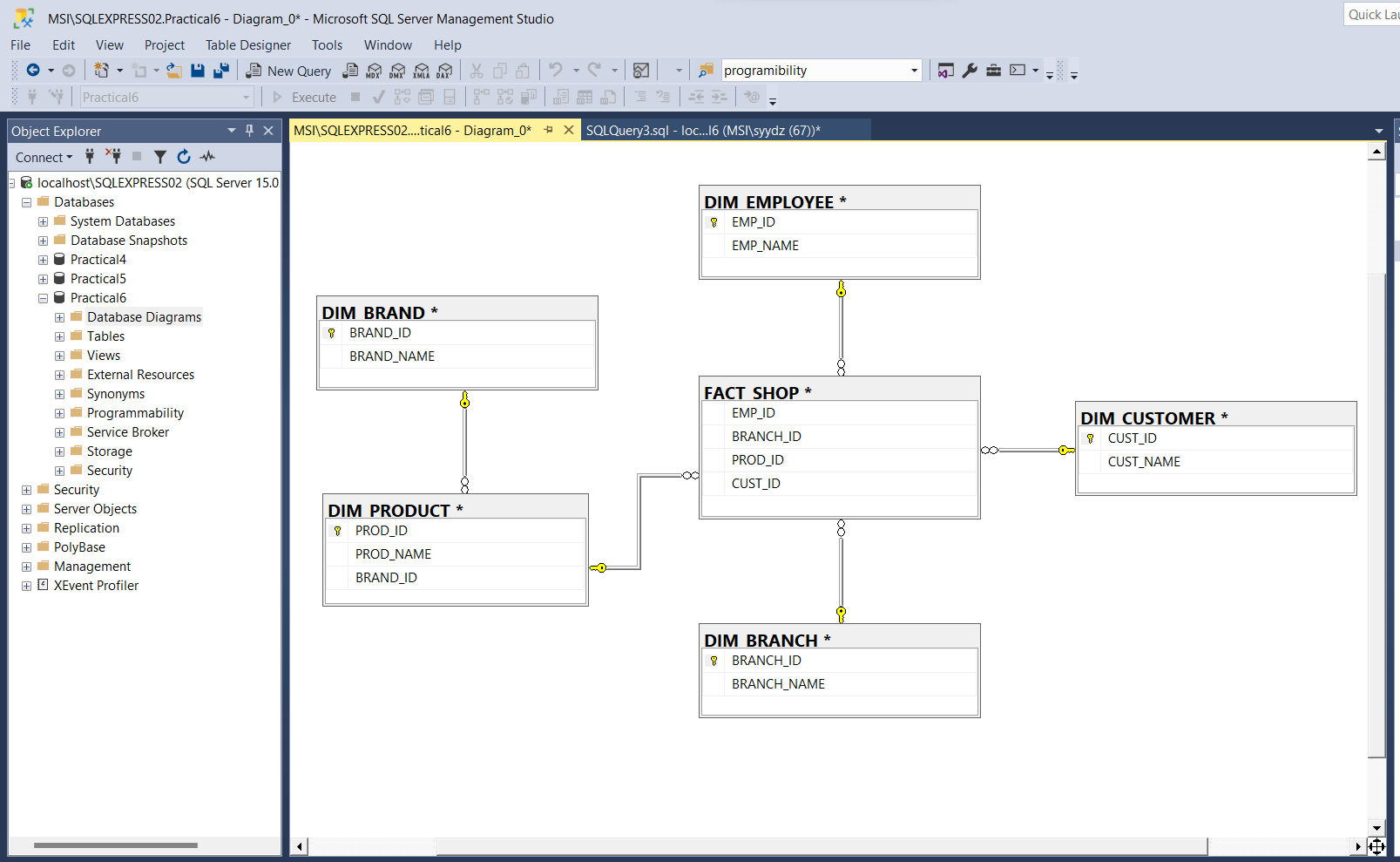
6) Set Primary key for all tables:



7) Add relationship of tables with fact shop.



8) Add relationship of product and brand .



Conclusion: Snowflake schema implemented successfully.

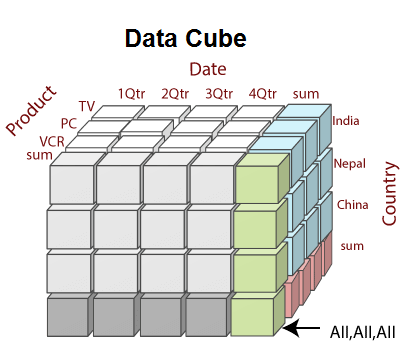
**Practical 7**

Aim: Create and uses excel pivot table report based on data cube operations.

Theory:

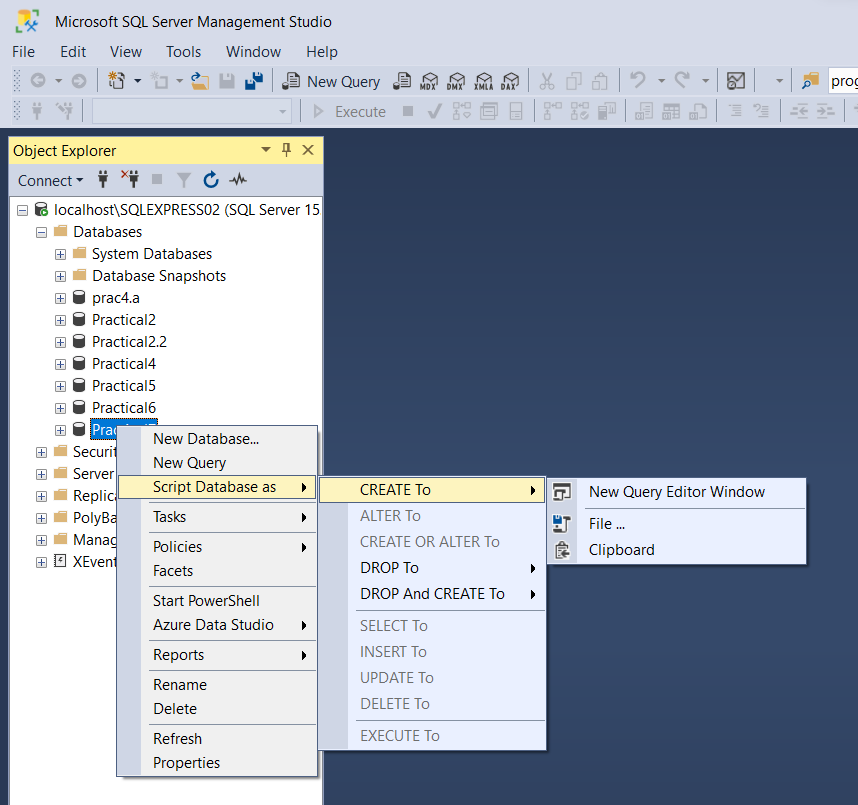
**What is Data Cube?**

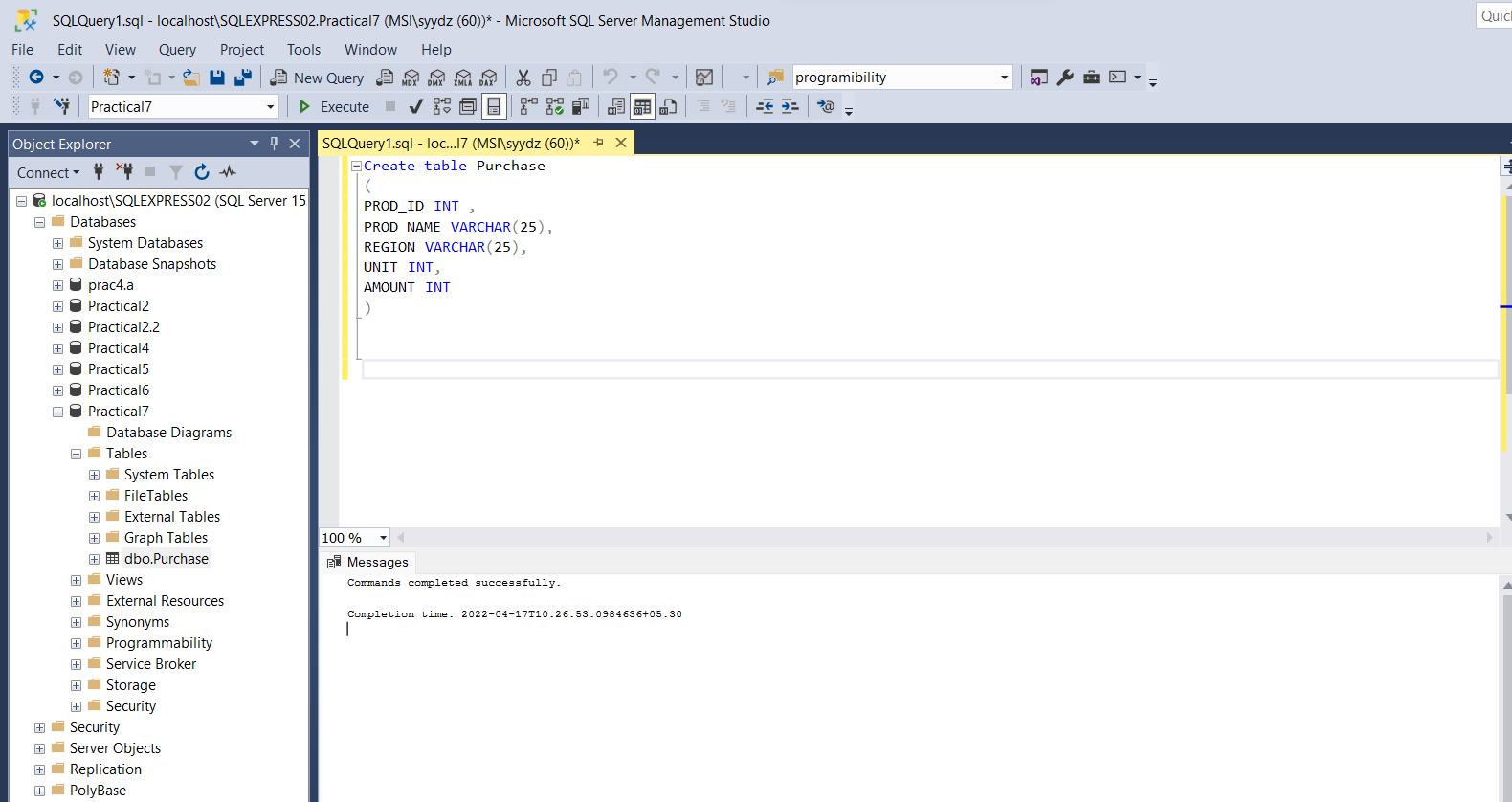
When data is grouped or combined in multidimensional matrices called Data Cubes. The data cube method has a few alternative names or a few variants, such as "Multidimensional databases," "materialized views," and "OLAP (On-Line Analytical Processing)."

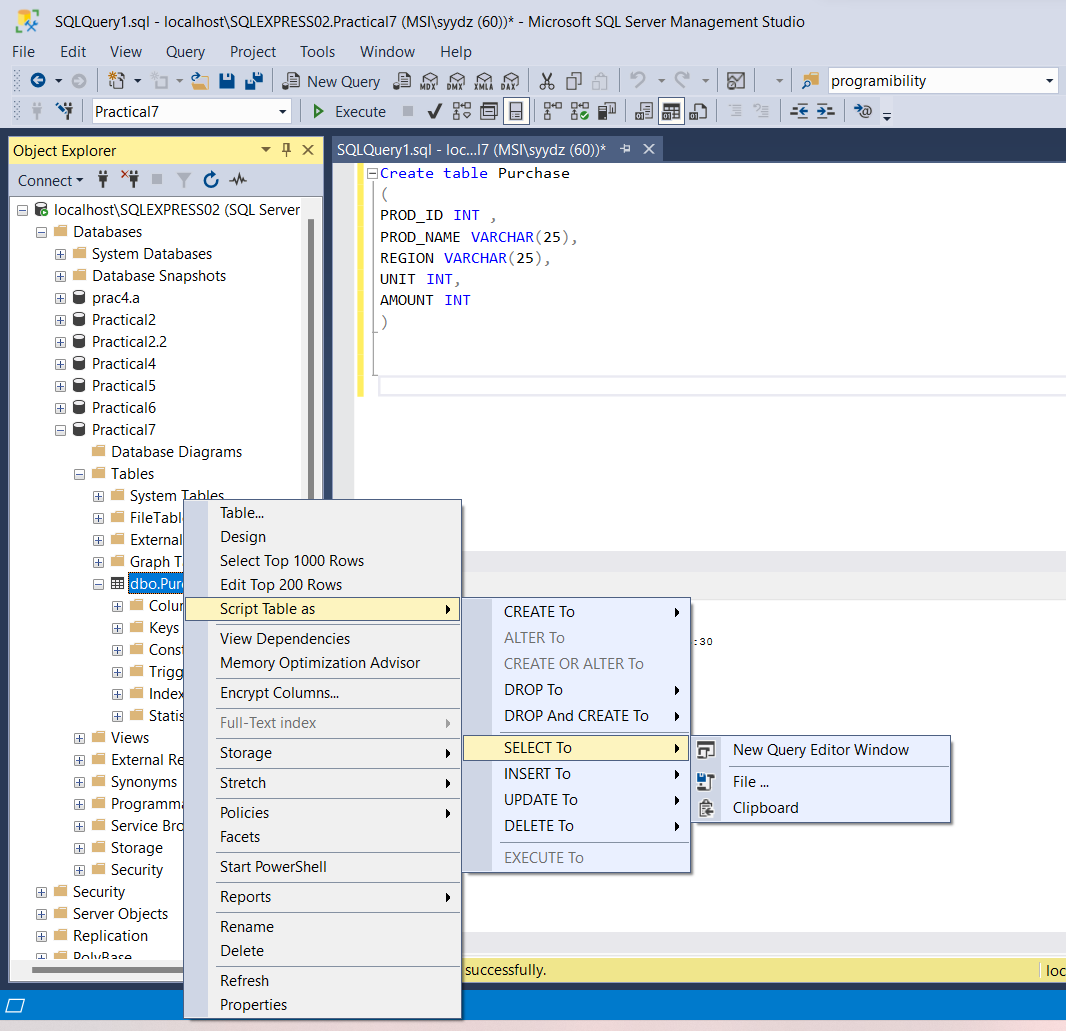


Steps:

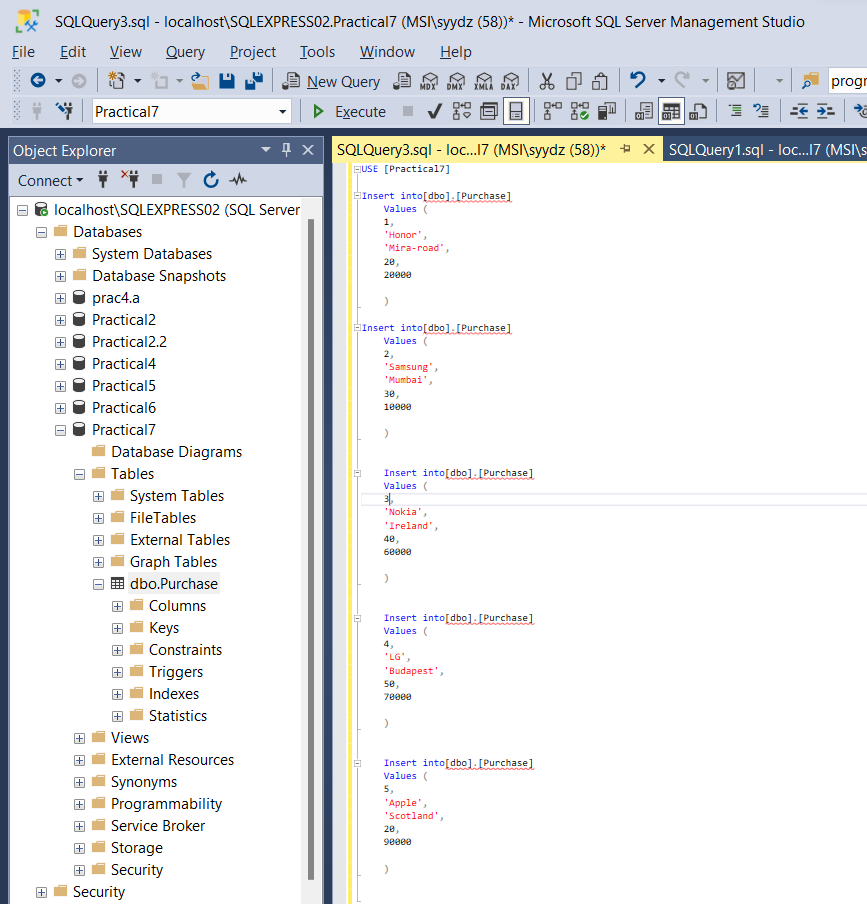
1) Create new database

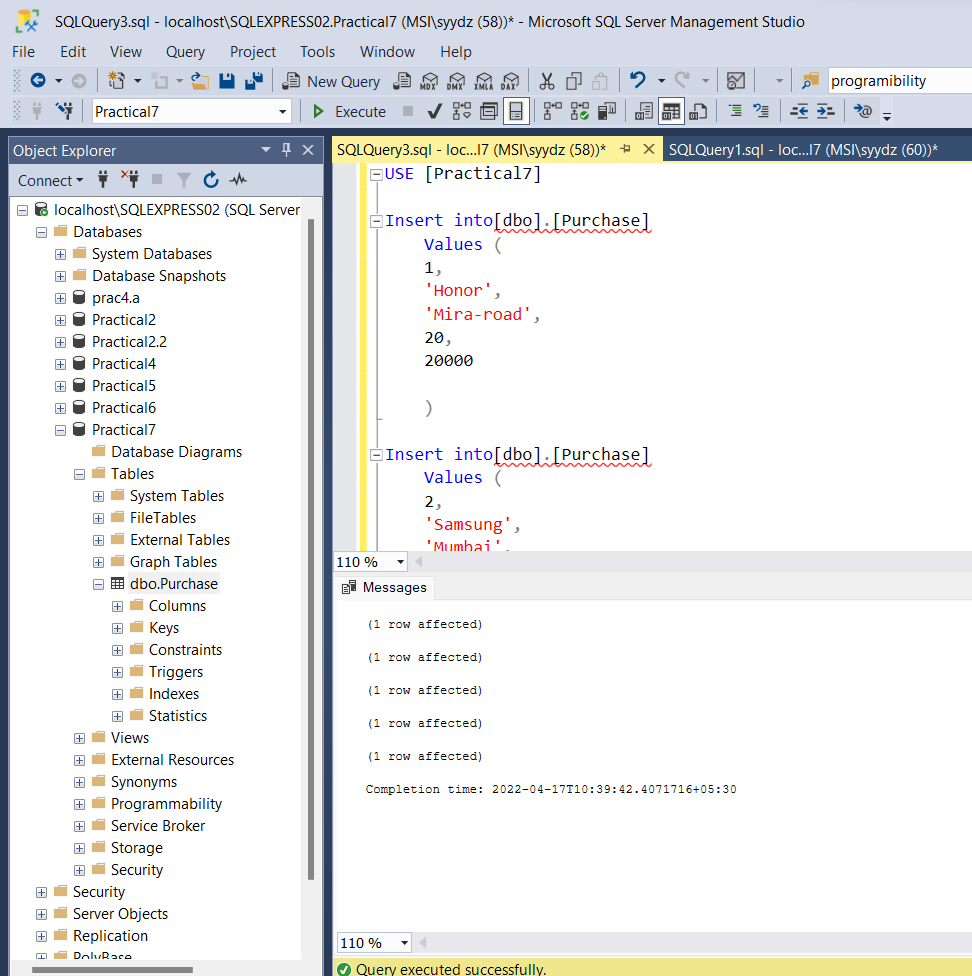




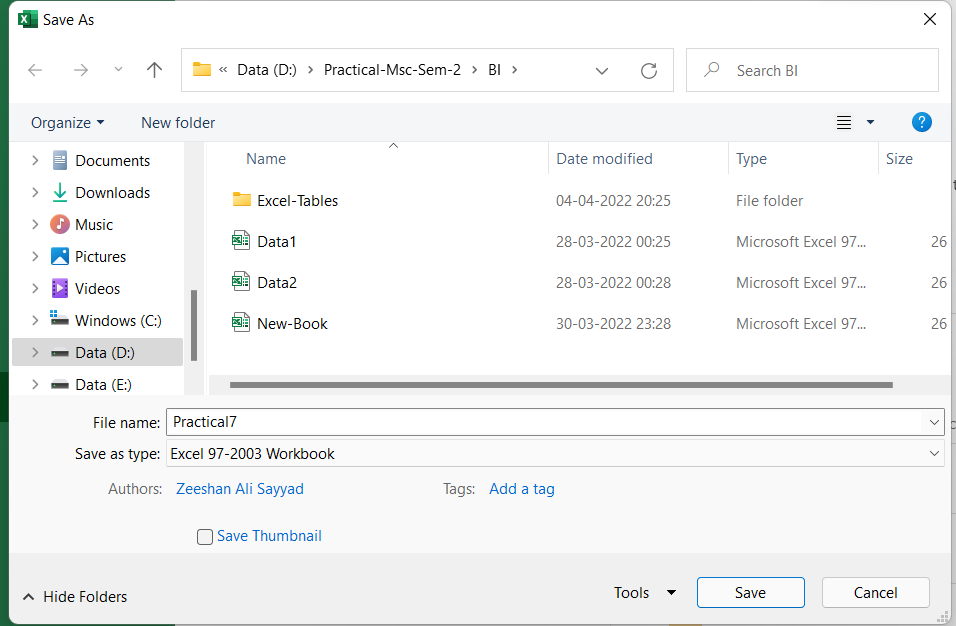


2) Insert data in table

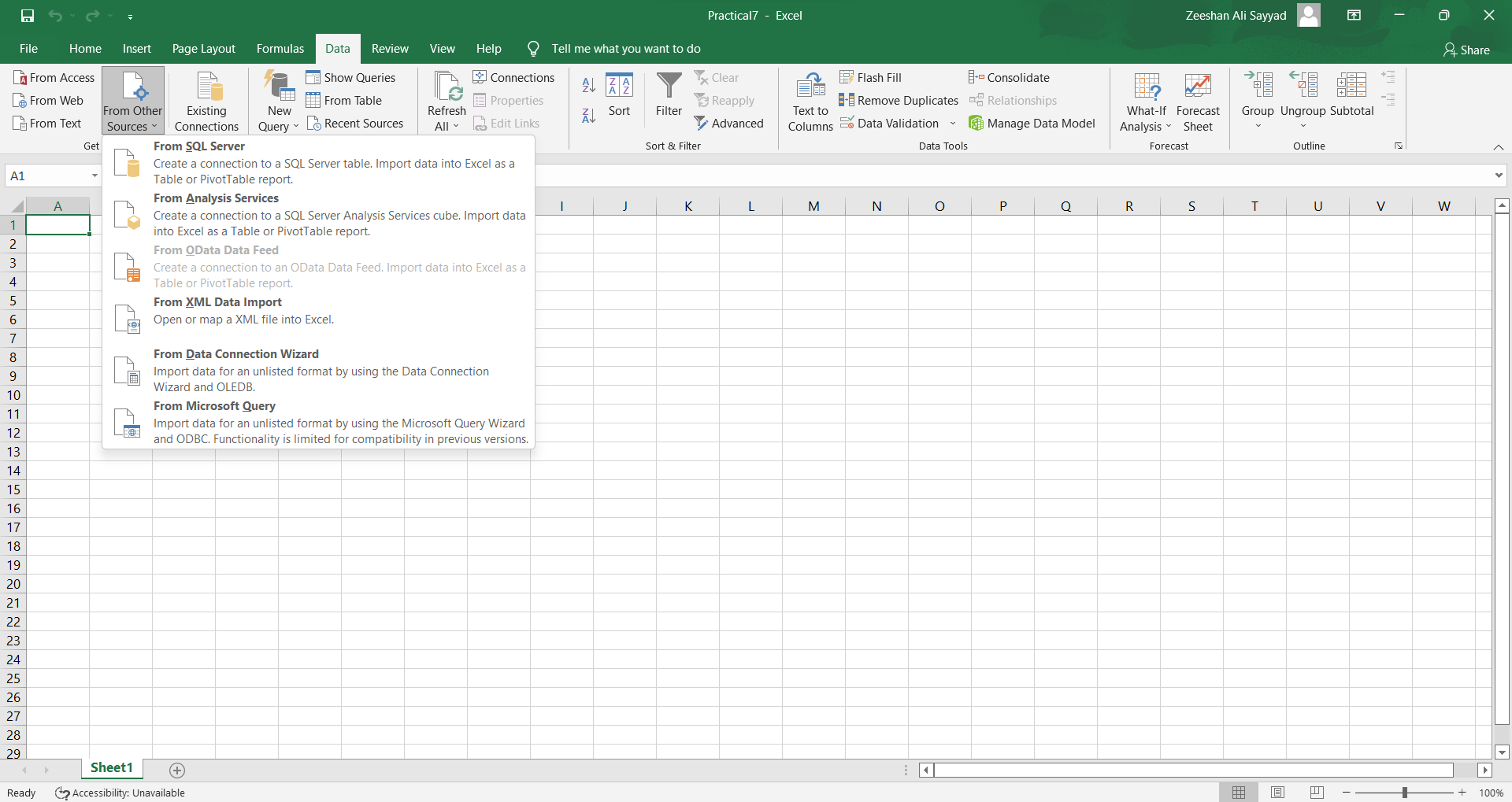




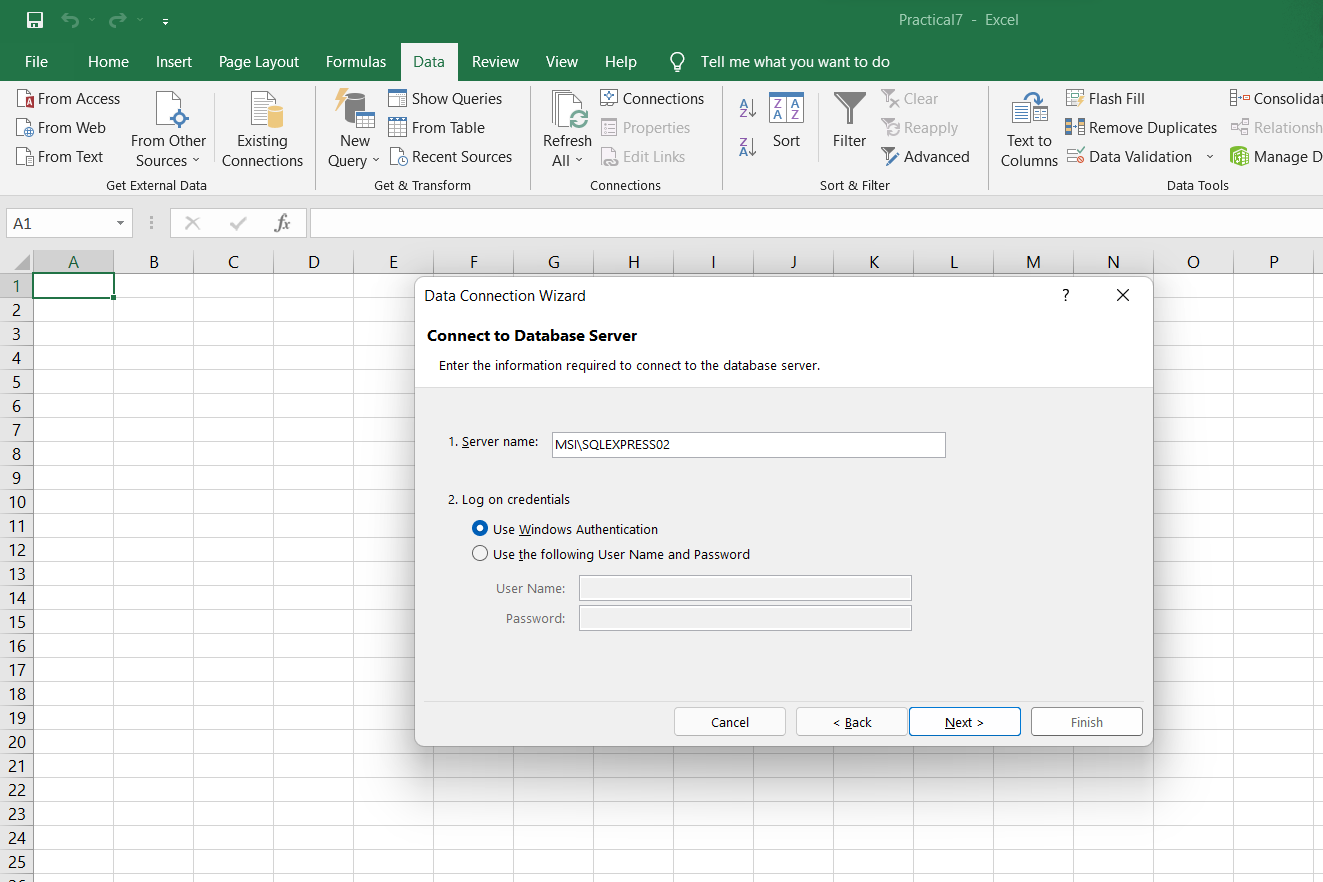
3) Create a new Excel File



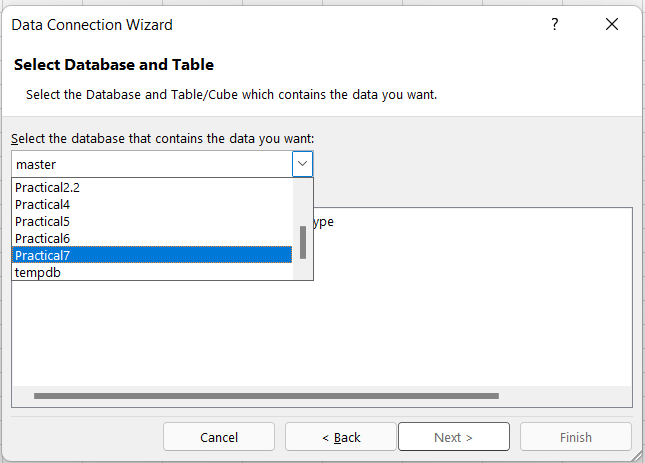
4) Go to Data > From Other Sources> Click on From SQL Server



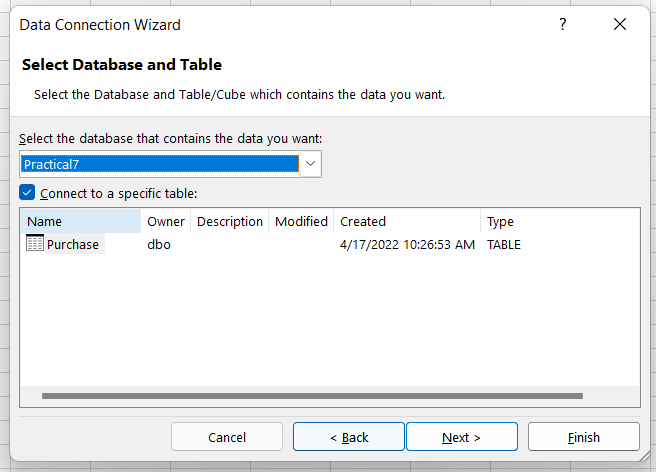
5) Connect to database using server name.



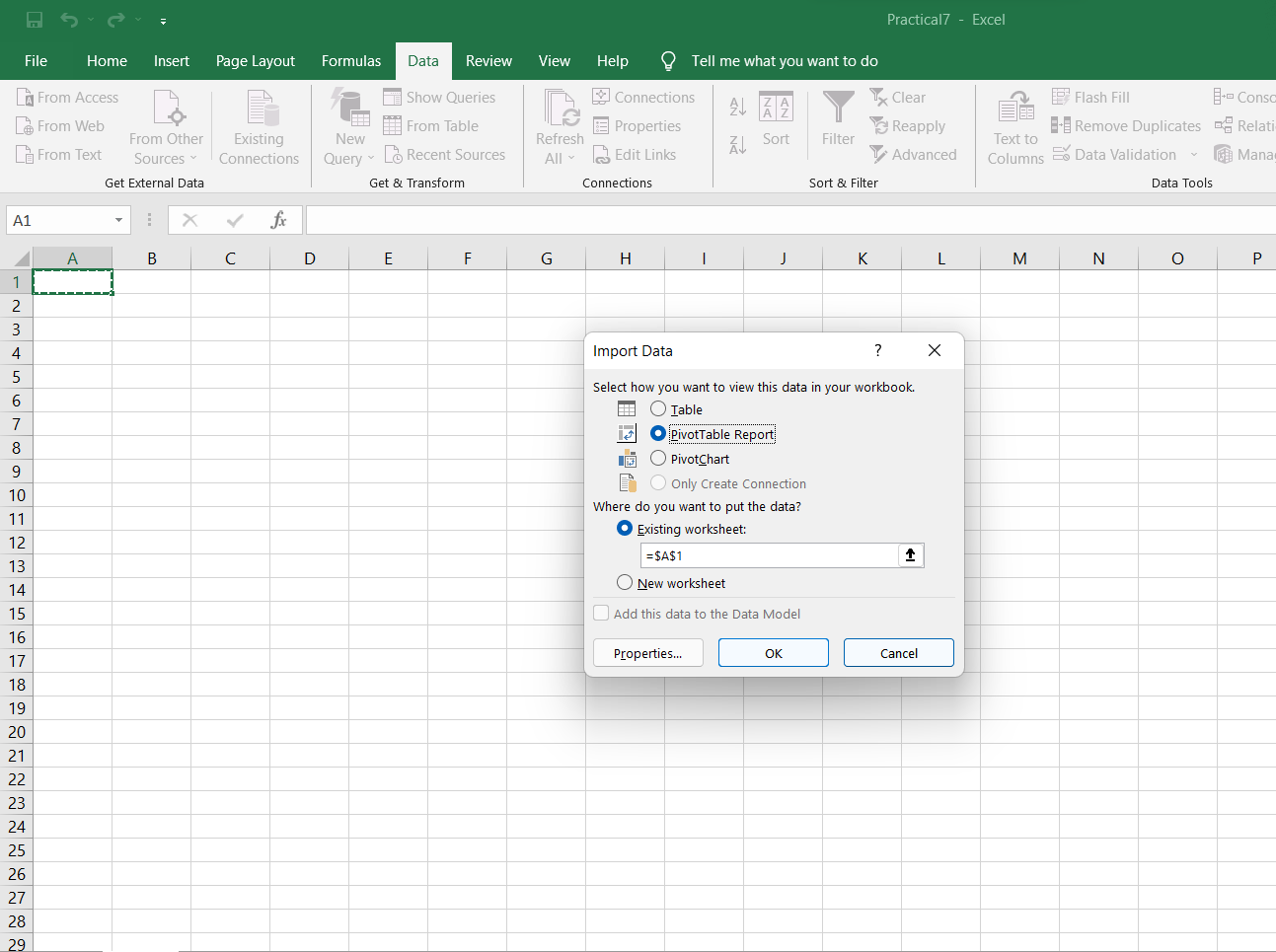
6) Select database and table we want to connect to.

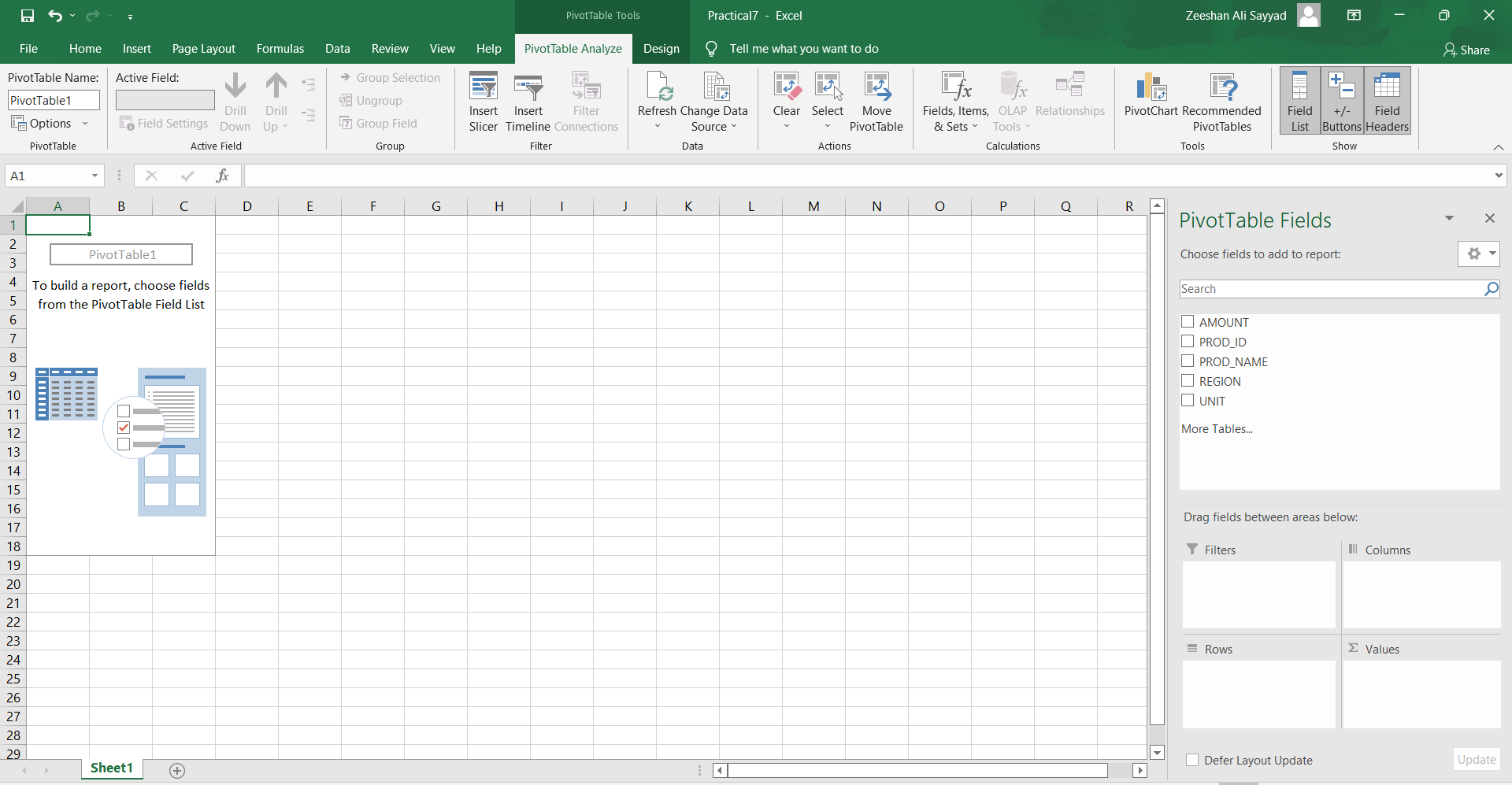


7) Click on Next.

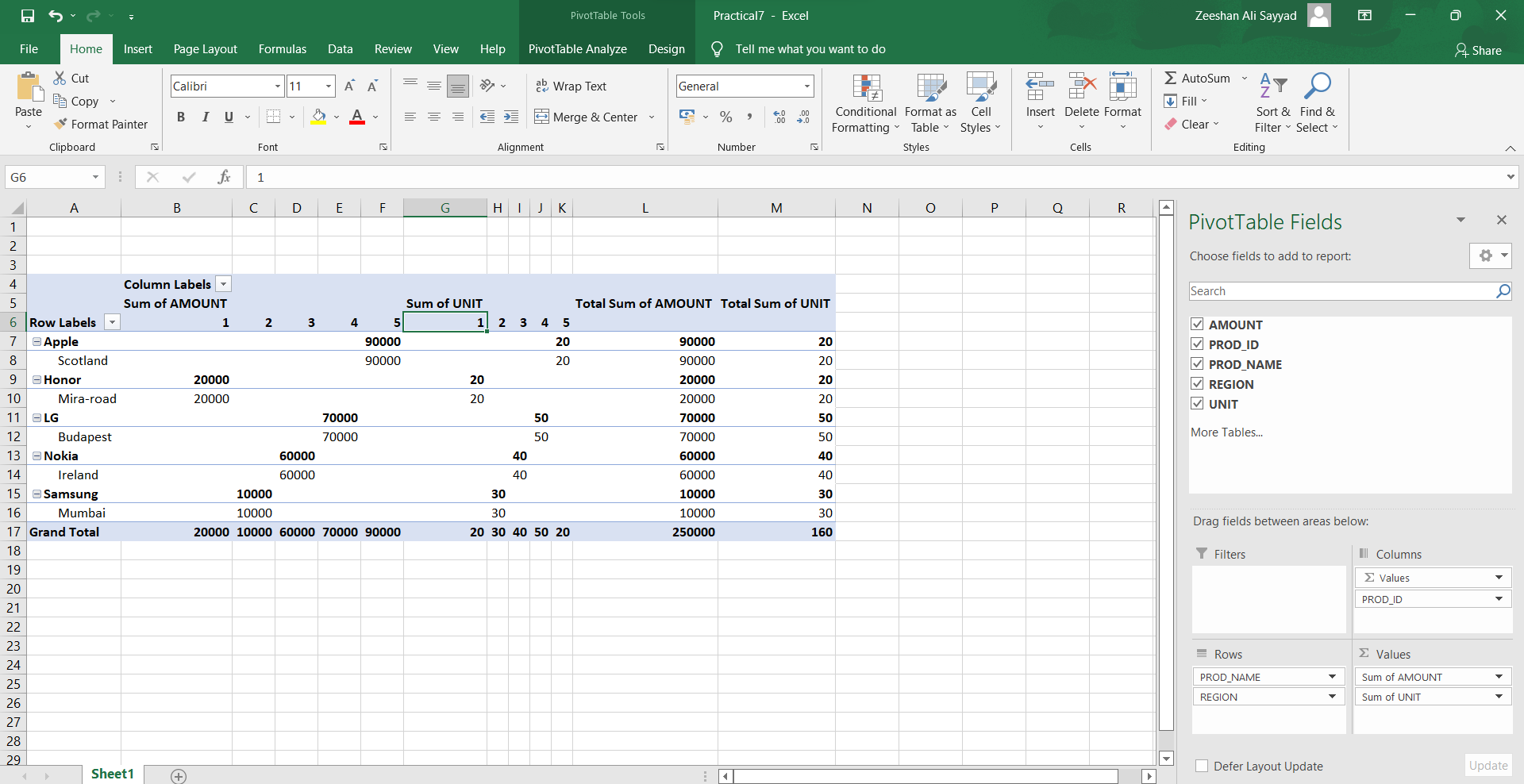


8) Select PivotTable report





9) Arrange fields PivotTable Fields respectively.



Conclusion: Successfully implemented pivot table report based on data cube operations

**Practical 8**

**Aim: Prerequisites for Database applications.**

**Queries:**

**Creating employee table:**

**SQL>**create table employee(

empid number(3) primary key,

empname varchar2(25),

age number(3));

**Creating client table:**

**SQL>**create table client(

clientid number(3) primary key,

clientname varchar(25),

location varchar2(20));

**Creating project table:**

**SQL>**create table project(

projid number(5) primary key,

empid number(3),

projname varchar2(25),

clientid number(3));

**Inserting records in table employee:**

**SQL>**insert into employee values(1,'Vardhan',22);

**SQL>**insert into employee values(2,'Khushi',30);

**SQL>**insert into employee values(3,'Simran',40);

**SQL>**insert into employee values(4,'Chetan',20);

**SQL>**insert into employee values(5,'Swati',27);

**SQL>**insert into employee values(10,'Vedant',33);

**SQL>**insert into employee values(15,'Gauri',35);

**SQL>**insert into employee values(16,'ABC',36);

**SQL>**insert into employee values(17,'PQR',37);

**Inserting records in table client:**

**SQL>**insert into client values(33,'will','mumbai');

**SQL>**insert into client values(34,'jones','bangalore');

**SQL>**insert into client values(35,'deep','kolkata');

**SQL>**insert into client values(36,'smith','delhi');

**SQL>**insert into client values(37,'marin','mumbai');

**SQL>**insert into client values(38,'jason','mumbai');

**Inserting records in table project:**

**SQL>**insert into project values(111,1,'abc',33);

**SQL>**insert into project values(222,2,'def',34);

**SQL>**insert into project values(333,3,'xyz',35);

**SQL>**insert into project values(444,4,'pqr',36);

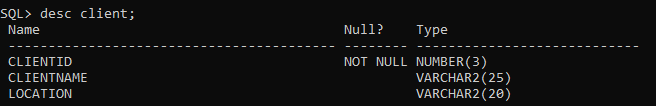
**SQL>**insert into project values(555,5,'mnq',37);

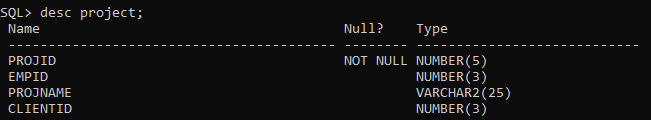
**SQL>**insert into project values(666,6,'efg',38);

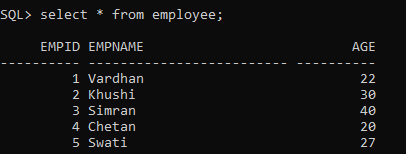
**SQL>**insert into project values(777,7,'lmn',39);

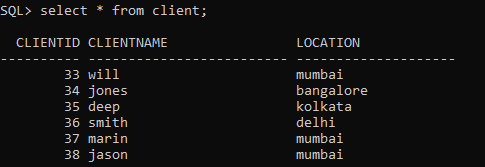
**SQL>**insert into project values(888,8,'ppd',40);



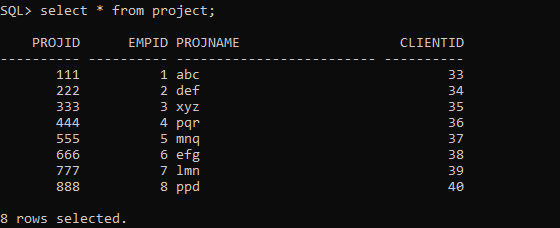


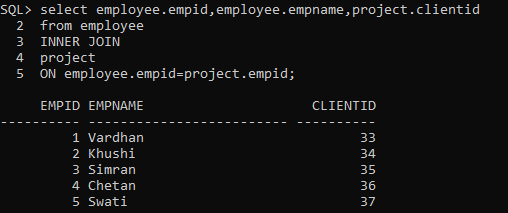


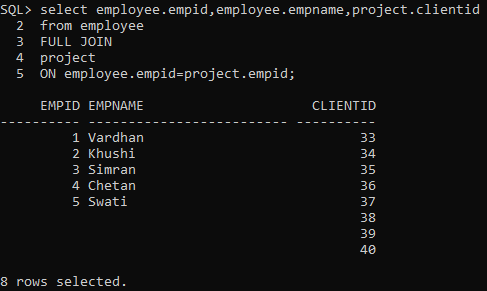


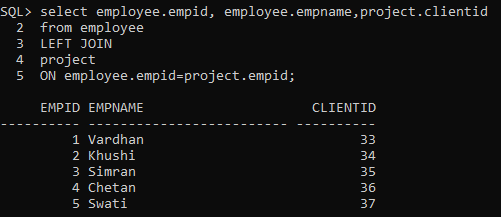


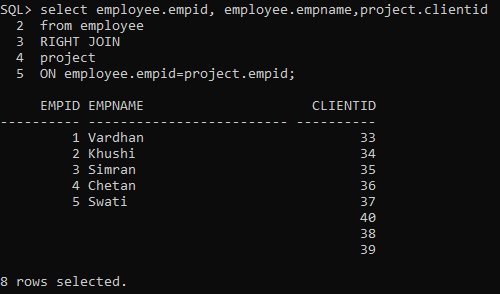
G:\6.PNG



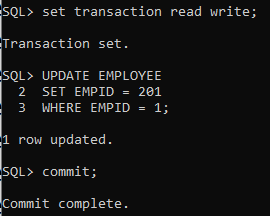


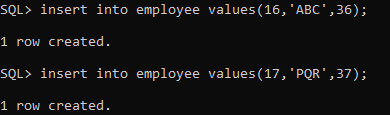




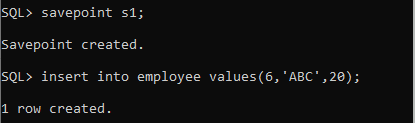


**Transactions**

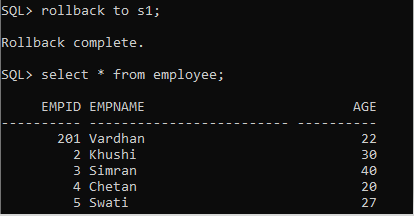




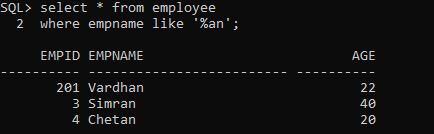
**Savepoint**

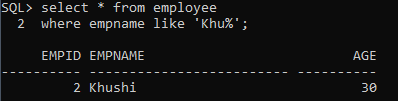


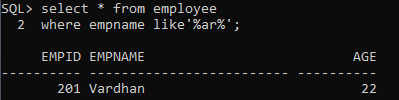
**Rollback**



**Wild cards**







**Distinct**

