Lab Manual

CS614 – Data Warehousing

Week No.	Lab Topic	Page No.
1	Installation of Microsoft SQL Server 2000	3
2	Normalization	18
3	De-Normalization	27
4	Process of Cube Creation	31
5	Dimension Modelling and Star Schema	50
6	Data Extraction	52
7	Basic Sorted Neighborhood Method (BSN)	63
8	Data Quality Rules	65
9	Key Range Partitioning	67
10	Indexing Technique	74
11	Nested Loop, Sort Merge, and Hash Join using SQL Server Query Analyzer	77

Lab 1

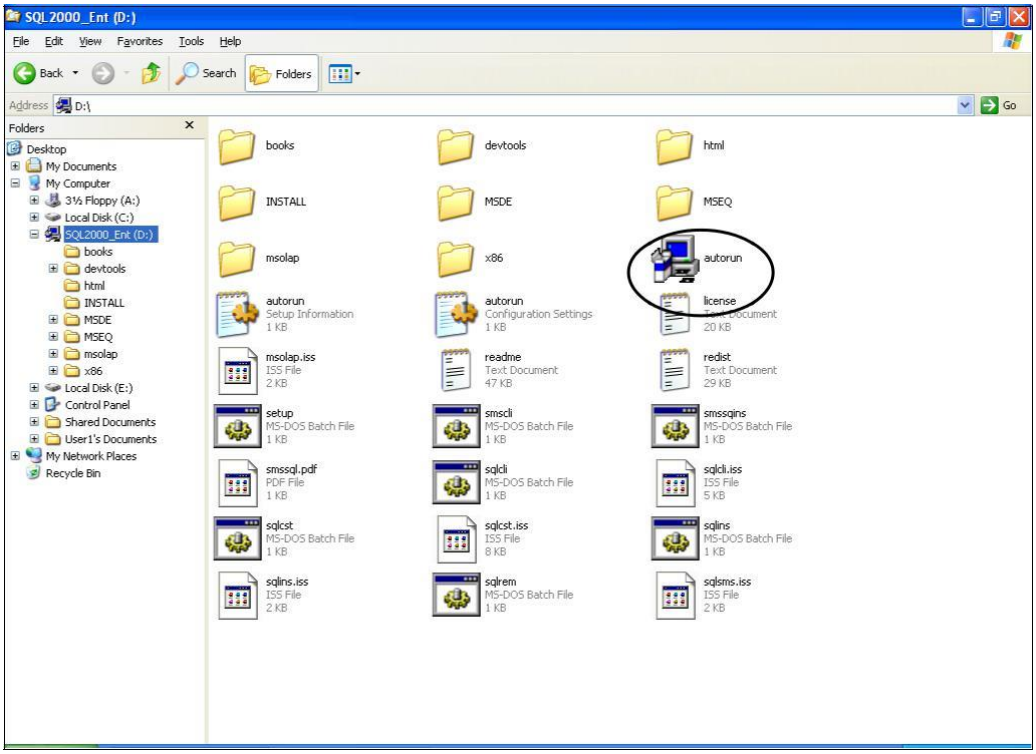
Installation of Microsoft SQL Server 2000

Important Instruction

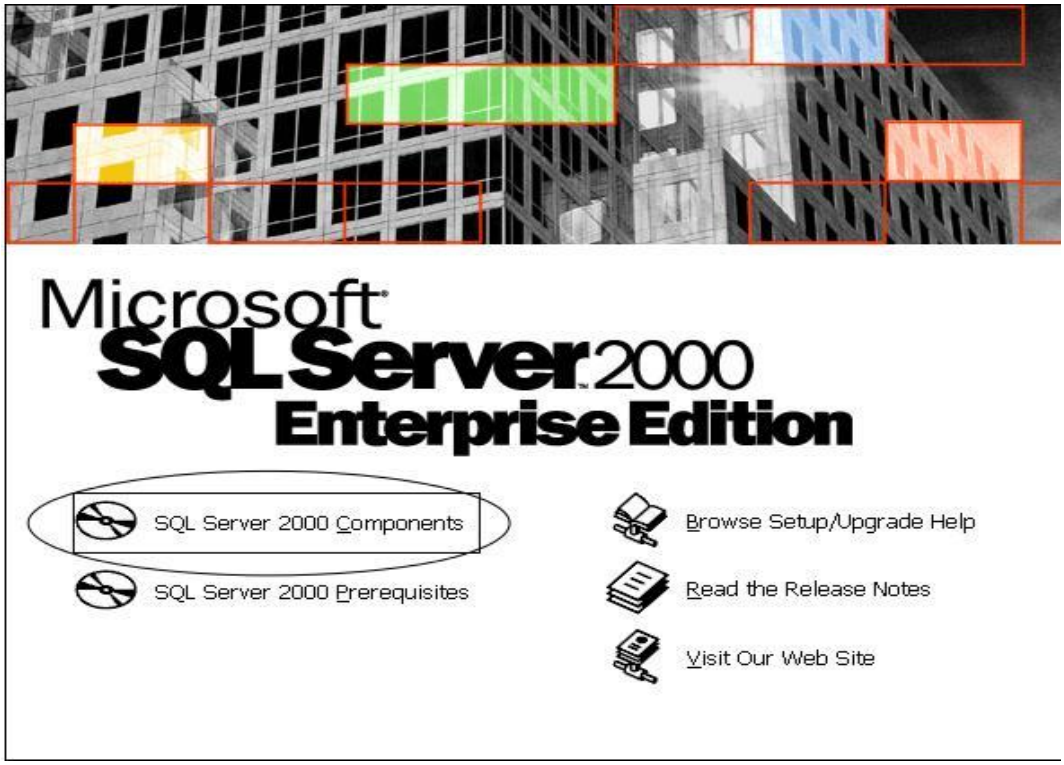
- ✓ User must have Administrators permissions in order to install Windows SQL Server 2000.
- ✓ Windows Server 2000 with SP3 or higher must be installed on the system.
- ✓ Close all programs running on computer before installation of SQL Server 2000.

Installation of SQL Server 2000 Step by Step:

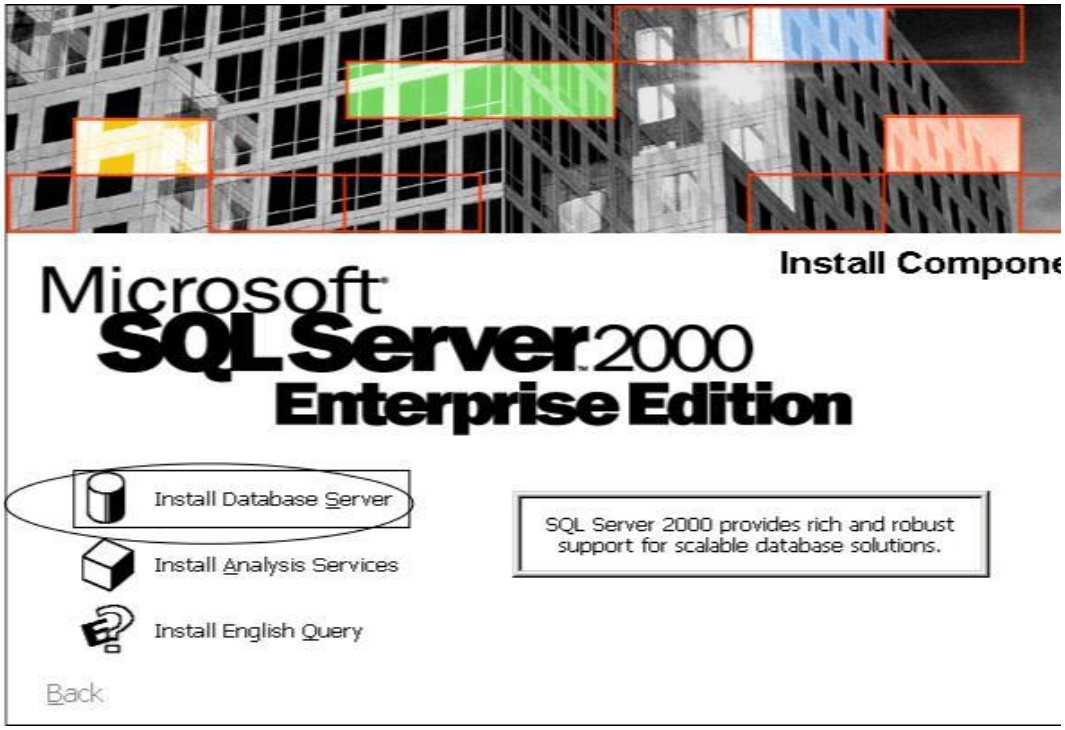
1. Open the folder/Drive containing set up of SQL Server 2000 and double click on “autorun”.



2. At the Microsoft SQL Server 2000 Enterprise Edition screen, press the “SQL Server 2000 Components” button.

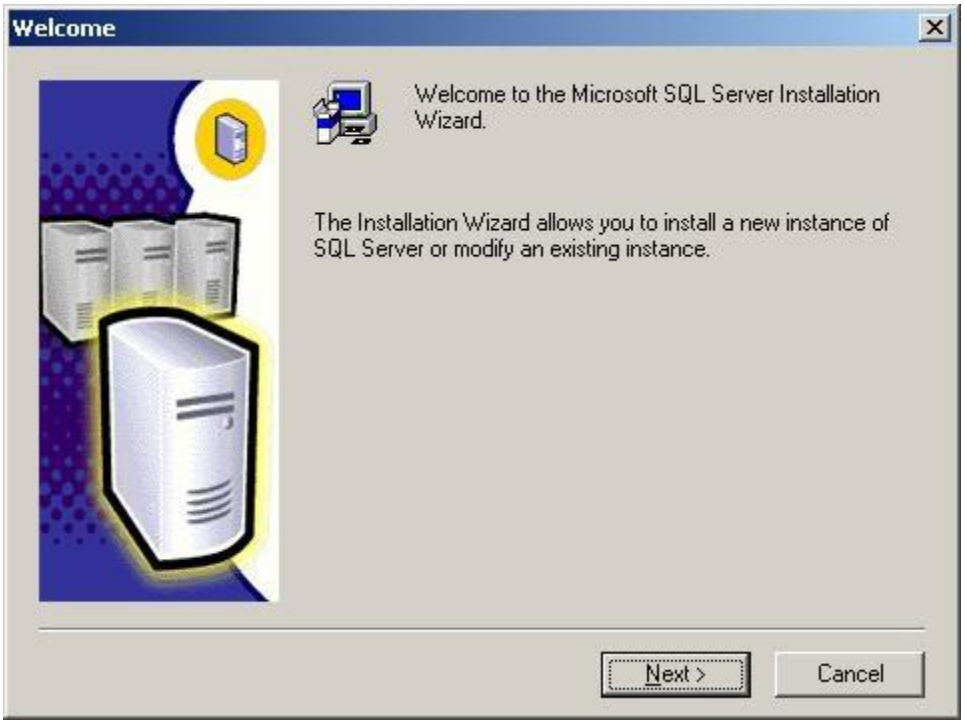


3. At the Microsoft SQL Server 2000 Enterprise Edition screen, press the “Install Database Server” button



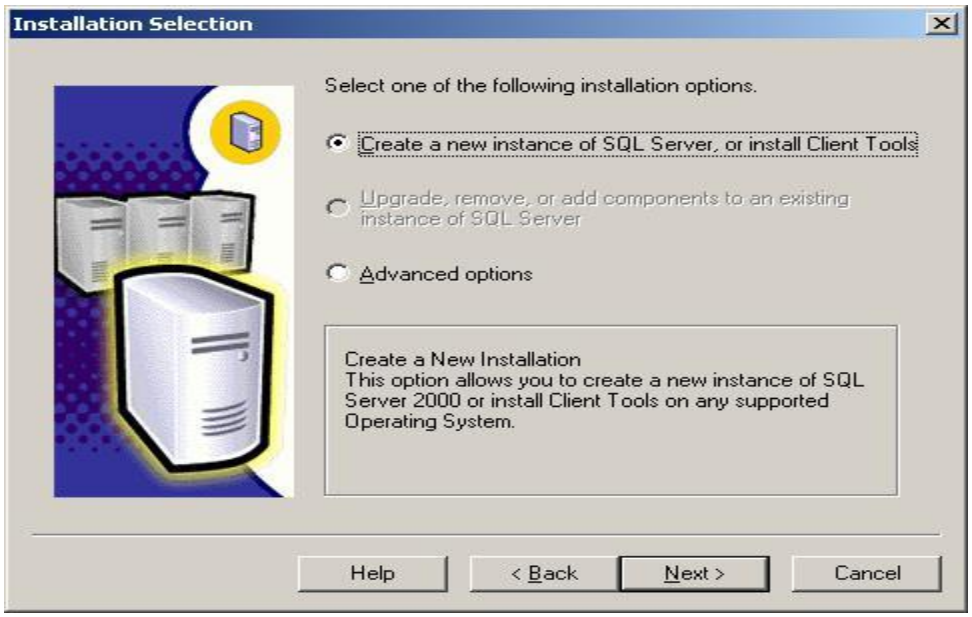
4. If Windows Server 2003 is installed you will see the SQL Server 2000 window. Press the Continue button.

5. You will be brought to the Welcome window to the Microsoft SQL Server Installation Wizard. Here press the Next button.



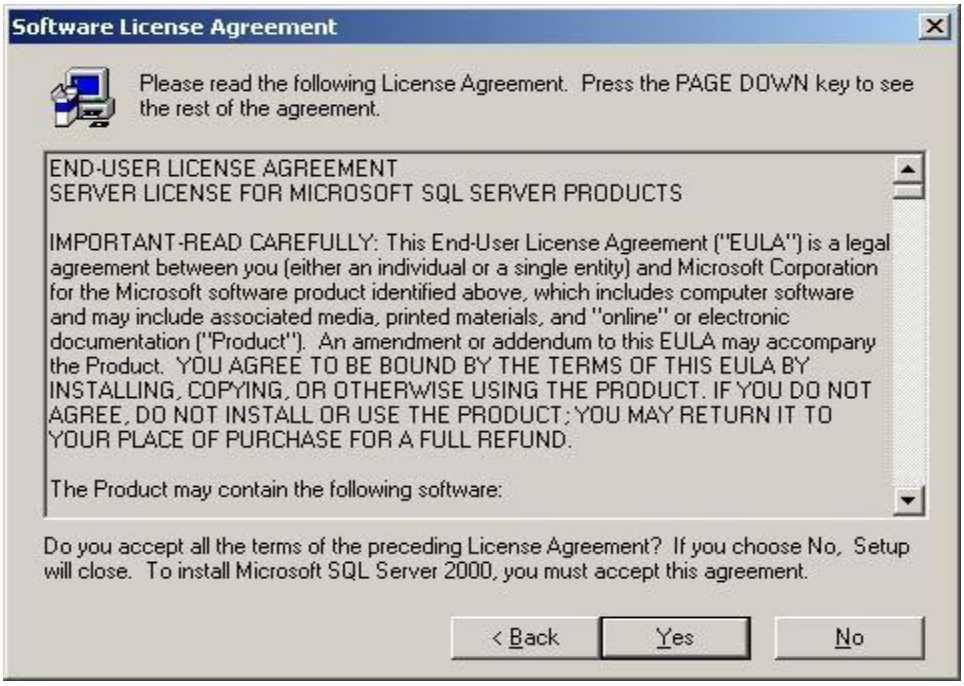
6. You will be brought to the Computer Name window. Since you are installing SQL 2000 Server on the Server computer, you will select the Local Computer and press the Next button.

7. You will be brought to the Installation Selection window. Select the “Create a new instance of SQL Server, or install Client Tools" radio button and press Next button.

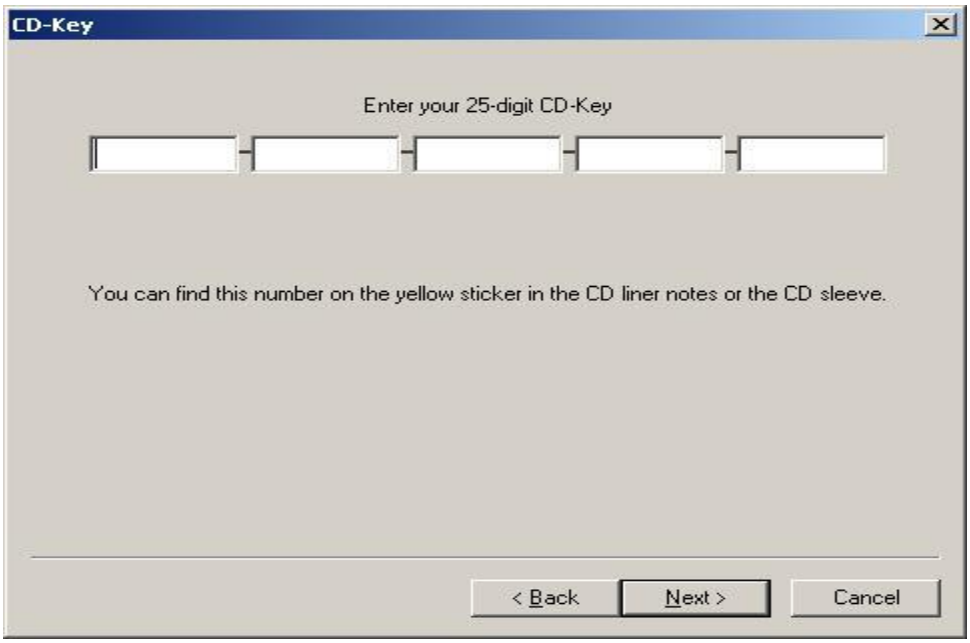


8. User will be brought to the User Information window. Enter the computer Name and Company for your system, press Next button.

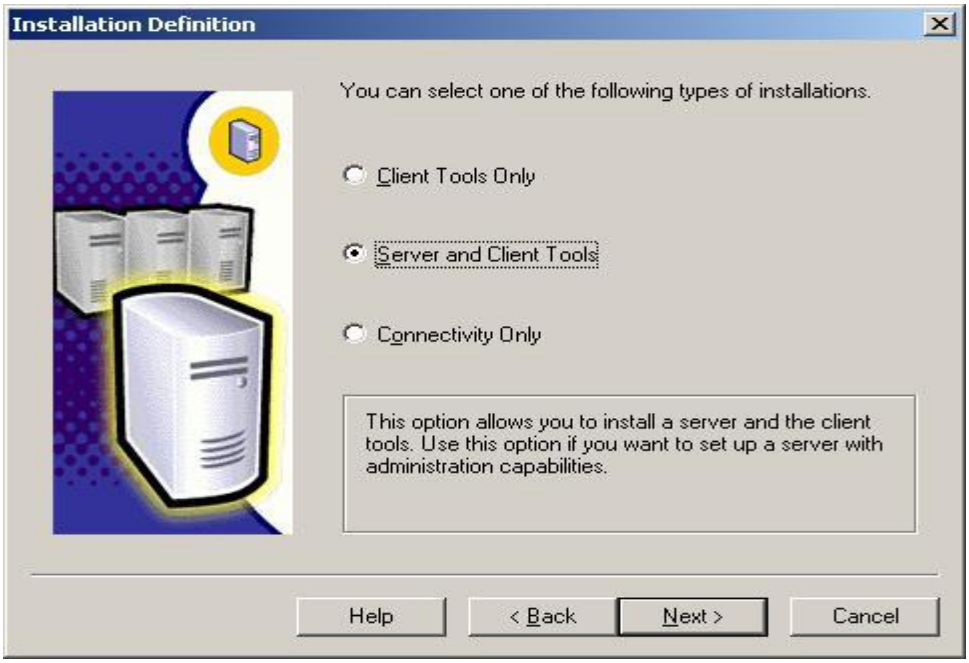
9. You will be brought to the Software License Agreement window. After reading the legal agreement, press the Yes button.



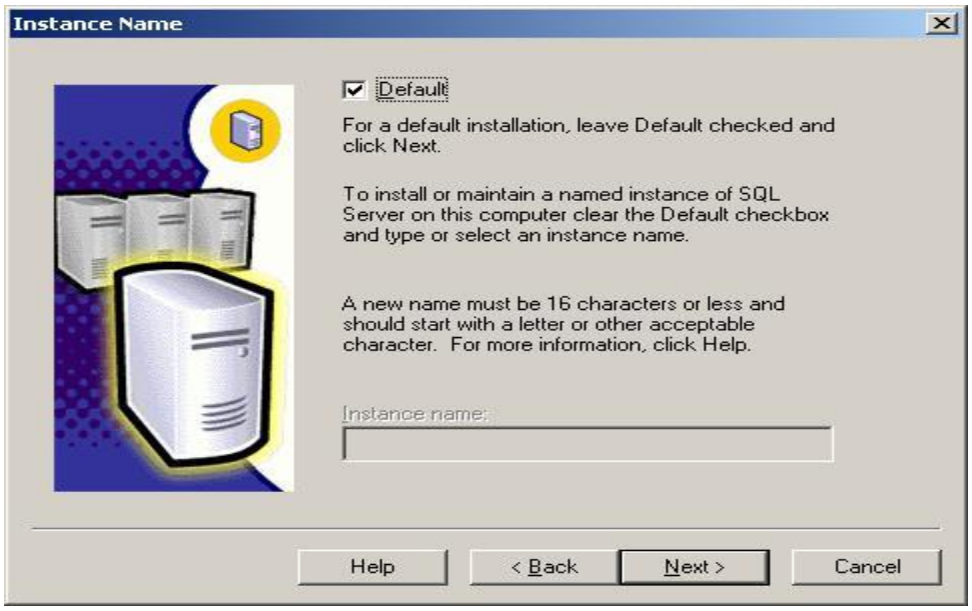
10. You will be brought to the registration key window. Enter the 25 digit registration Key and press the Next button.



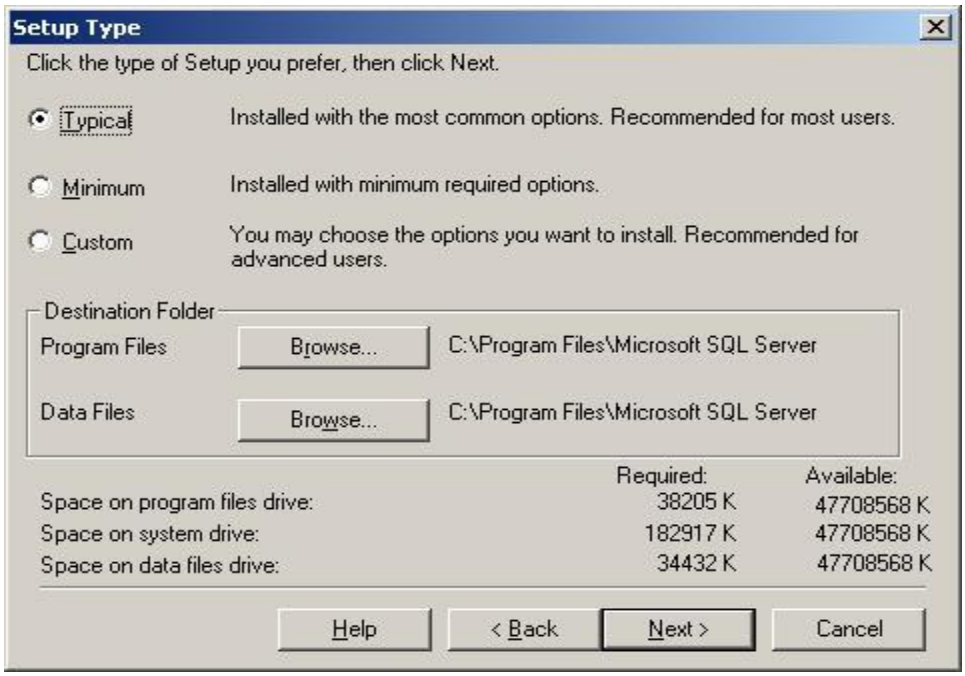
11. You will be brought to the Installation Definition window. Choose the "Serve and Client Tools" radio button and press the Next button.



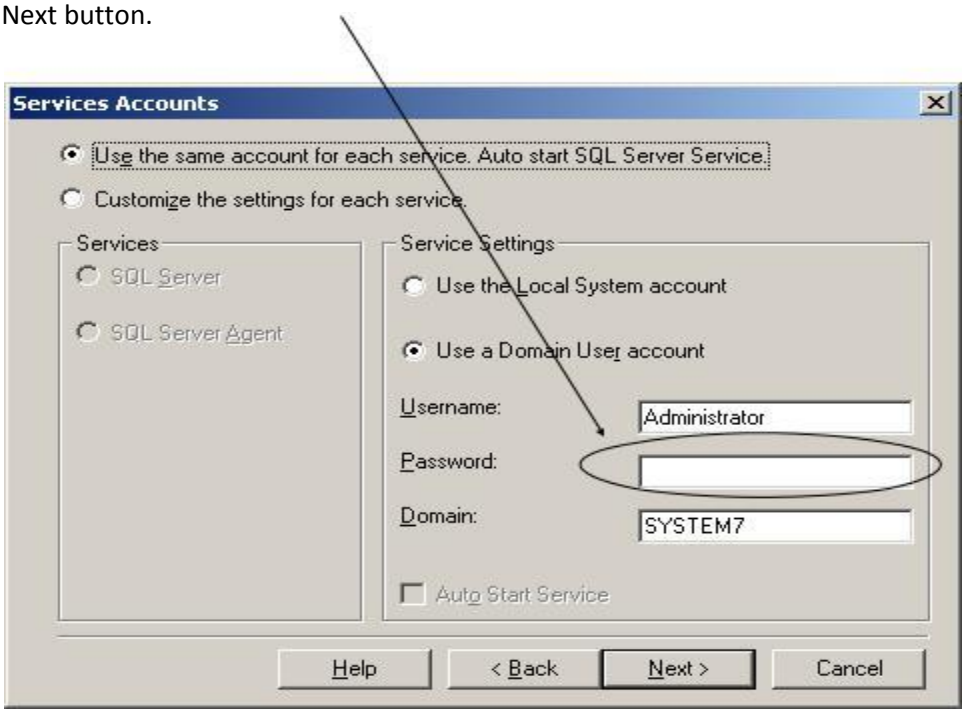
12. You will be brought to the Instance Name window. Her you have to press the Next button.



13. You will be brought to the Setup Type window. Of the type of Setup preferred, select Typical. And press the Next button.

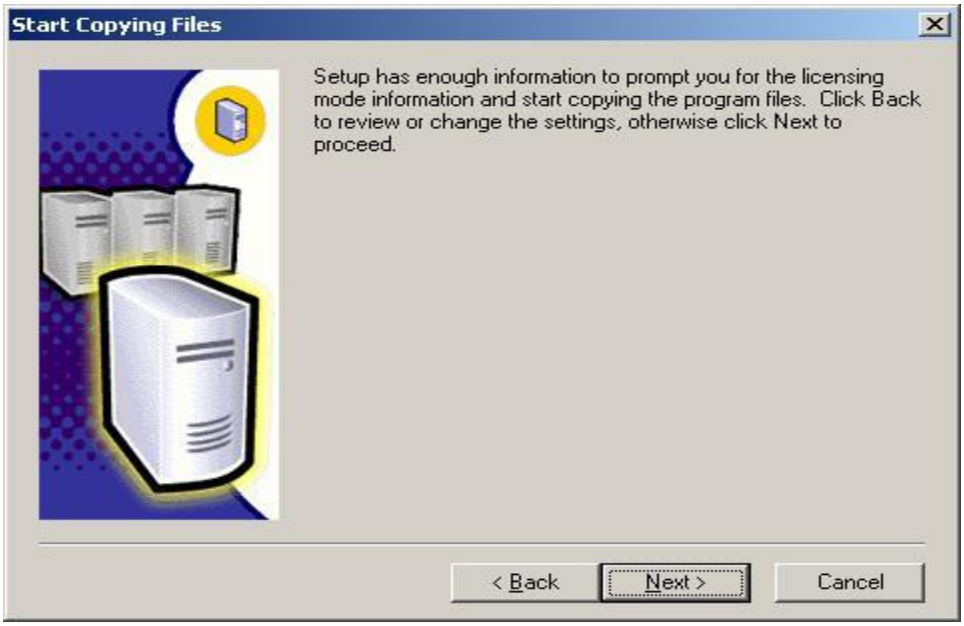


14. You will be brought to the Service Accounts window. Choose the ‘Use the same account for each service Auto start SQL Server Service.’ radio button. Enter Administrator’s Password and press the Next button.

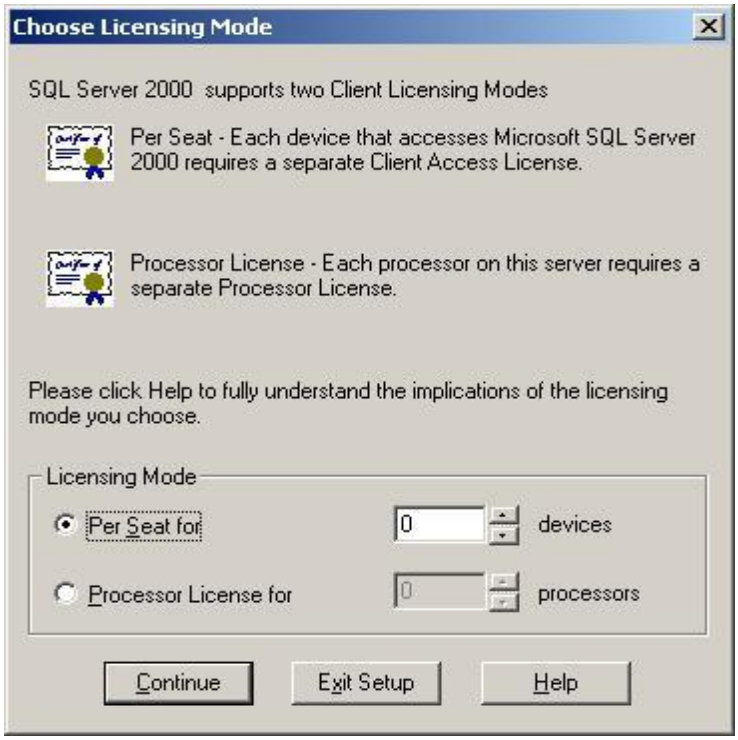


15. You will be brought to the Authentication Mode window. Select the ‘Mixed Mode (Windows Authentication and SQL Server Authentication) radio button. Enter your password in the Enter Password and Confirm Password fields and press the Next button.

16. You will be brought to the Start Copying Files window and press the Next button.



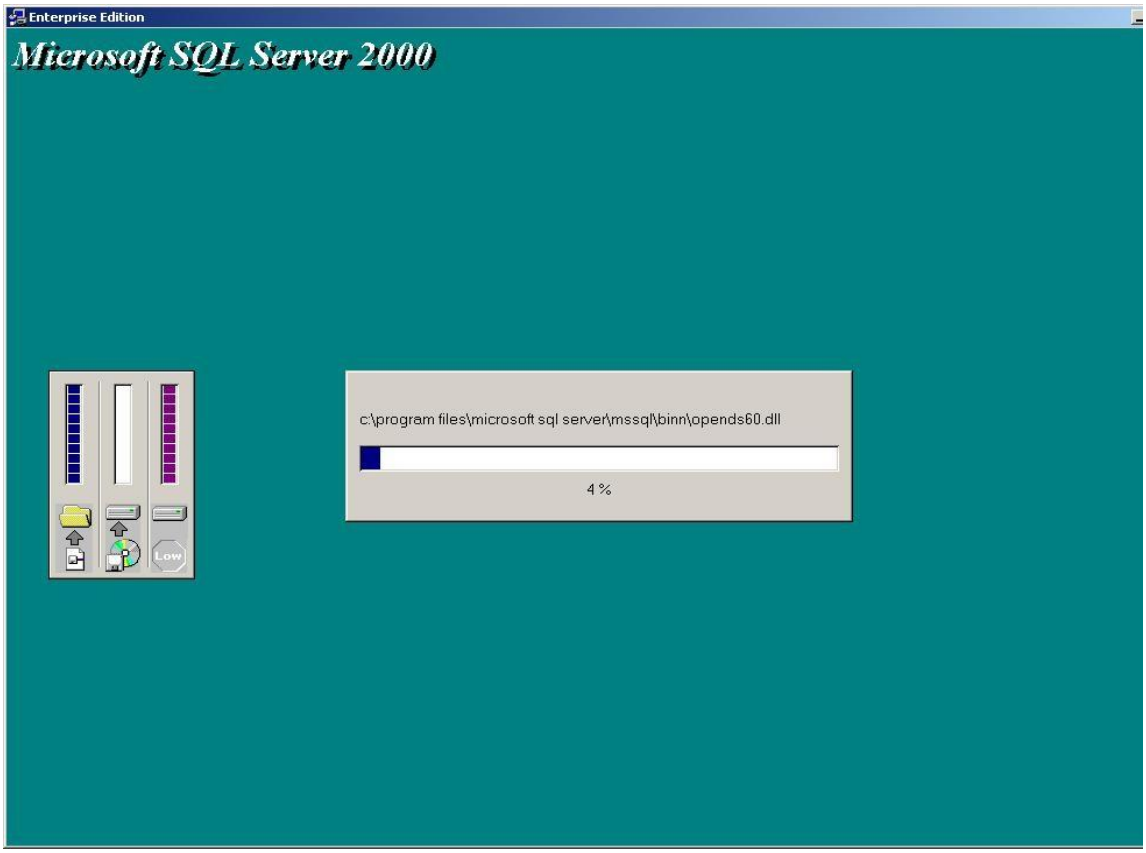
17. Once the file installation is complete, you will be brought to the Choose Licensing Mode. Select either Per Seat or Processor License. Press the Continue button.



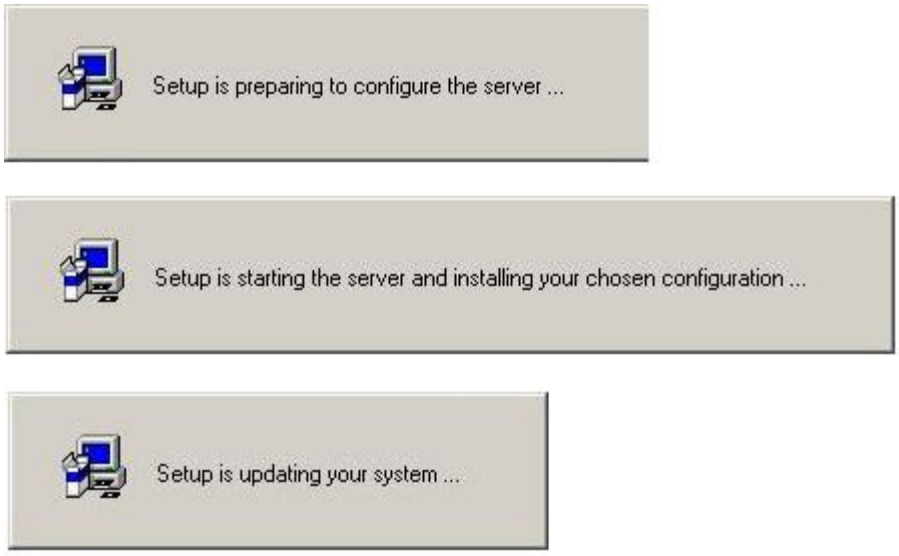
18. The System will display this window when loading the system.

19. The system will display this window when installing the Microsoft Full-Text Search Engine.

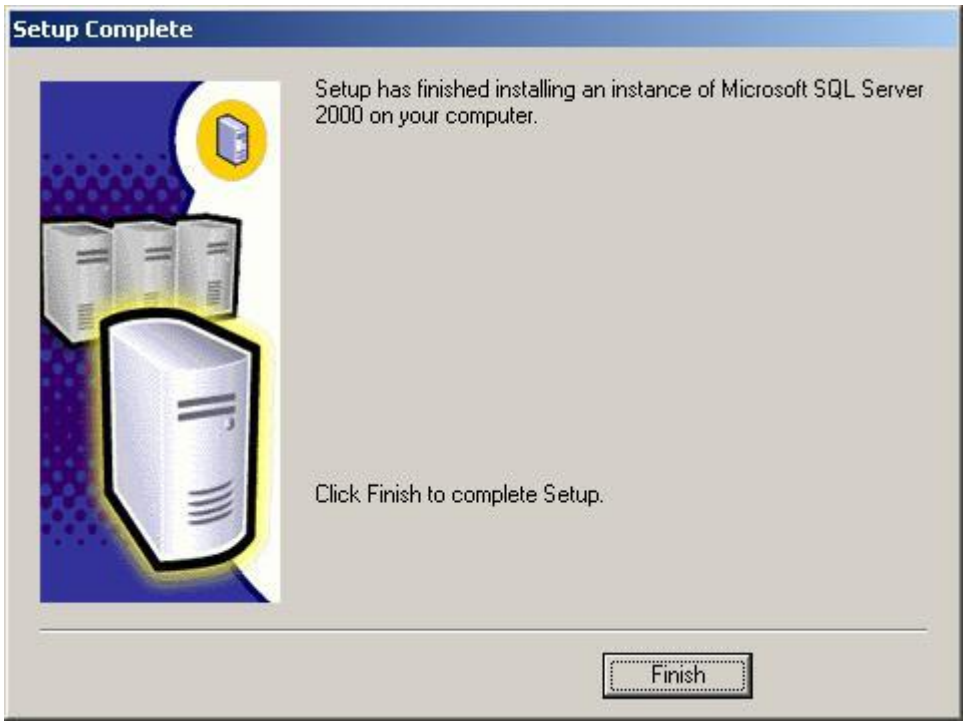
20. SQL Server 2000 progress screen will display.



21. Setup is preparing to configure the server and setup started the server and installation of selected configuration. After that setup updating window will display.

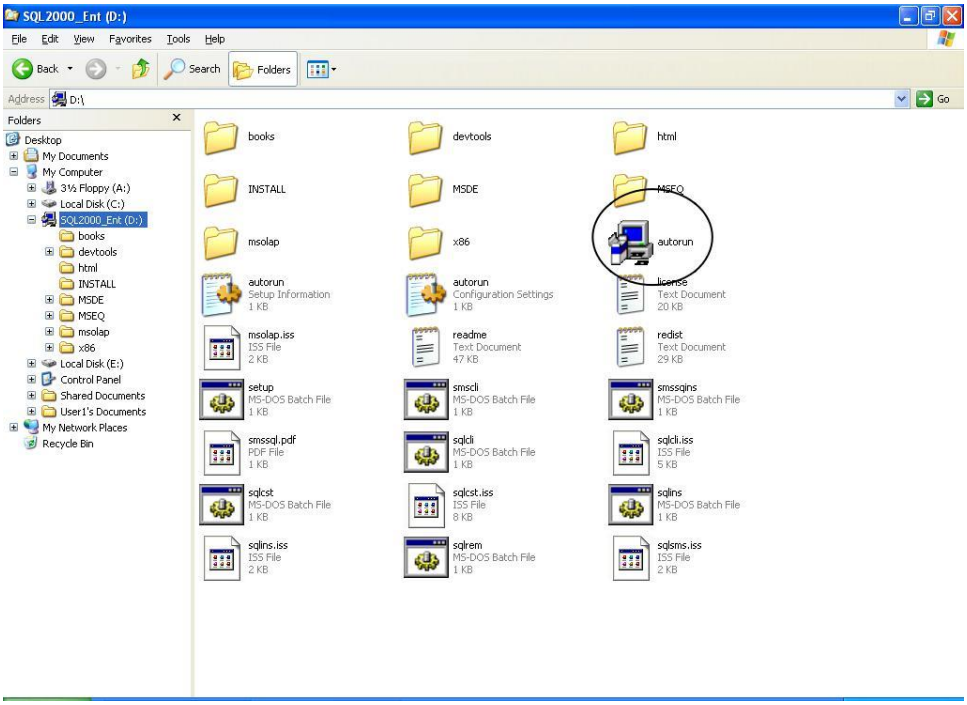


22. Once the setup is complete, you will be brought to the following window. Press the **Finish** button to complete the installation.

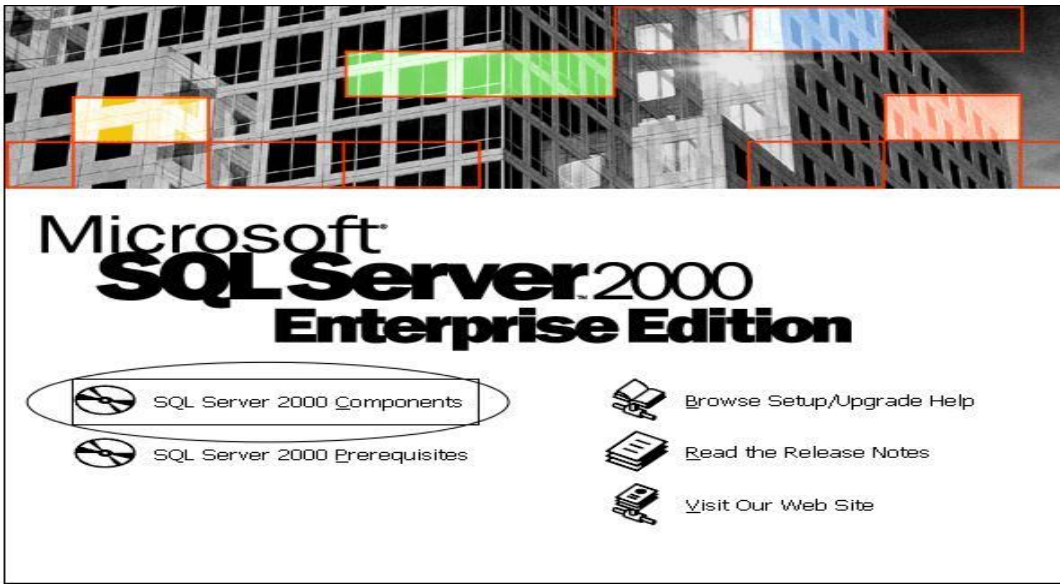


Installation of SQL Server 2000 Client:

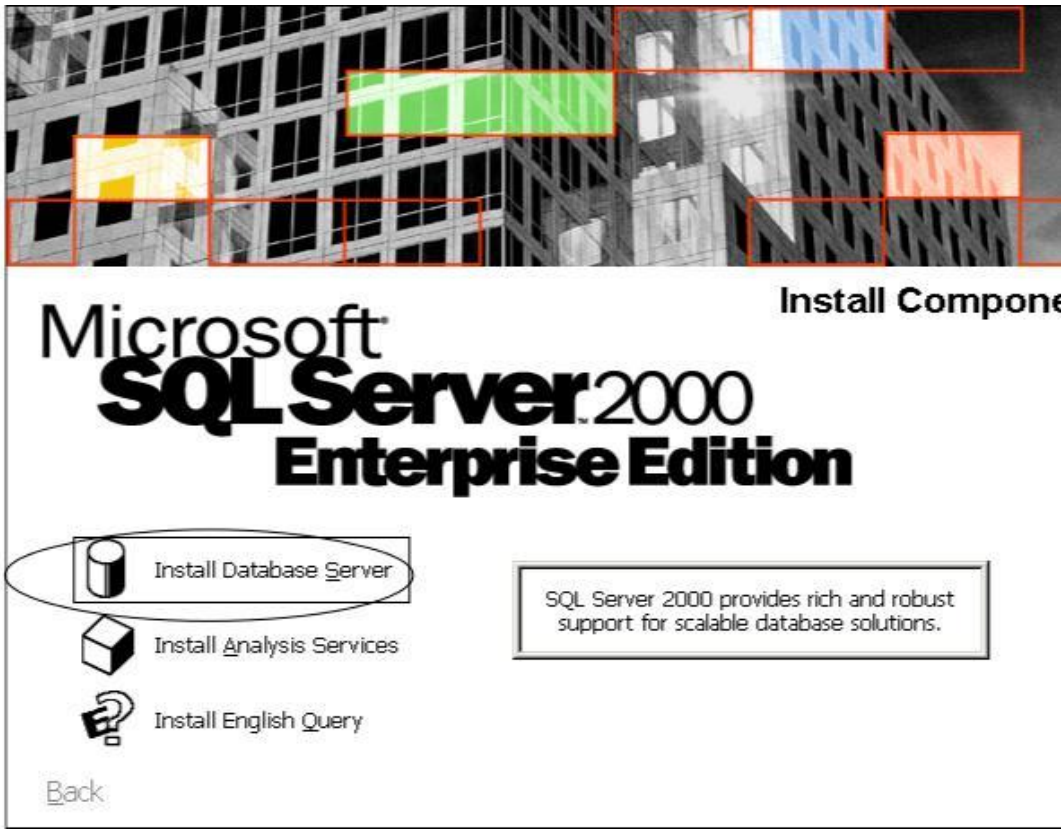
1. Open the Drive/Folder and click on “autorun”.



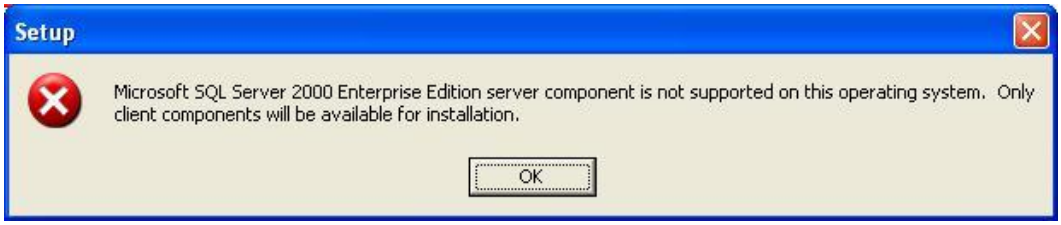
2. At the Microsoft SQL Server 2000 Enterprise Edition screen, press the SQL Server 2000 Components button.



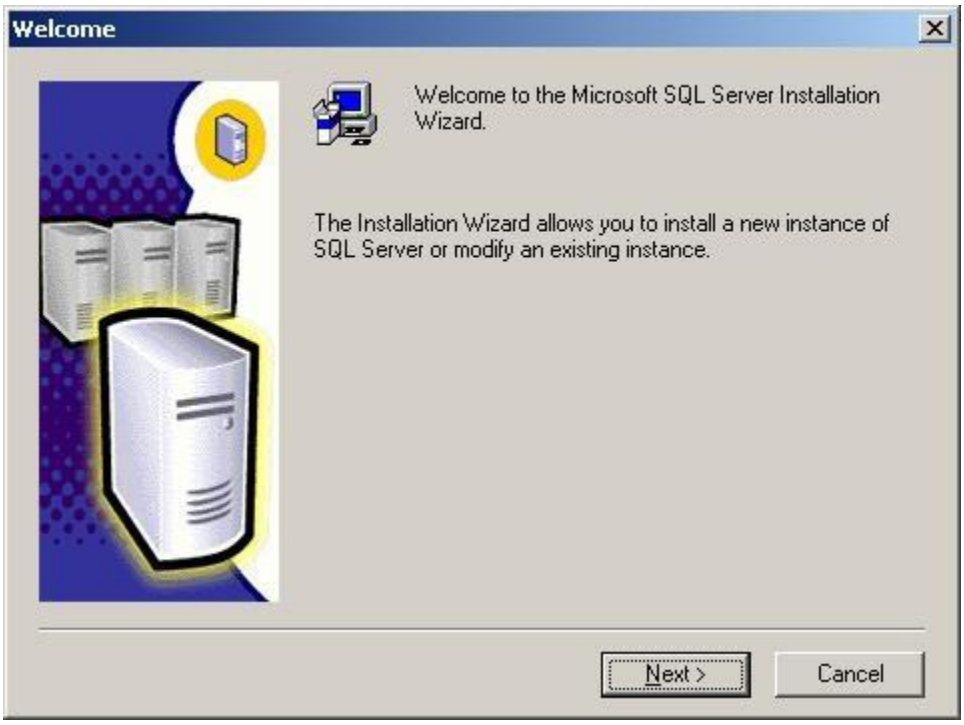
3. At the Microsoft SQL Server 2000 – Enterprise Edition – Install Components screen, press the Install Database Server button.



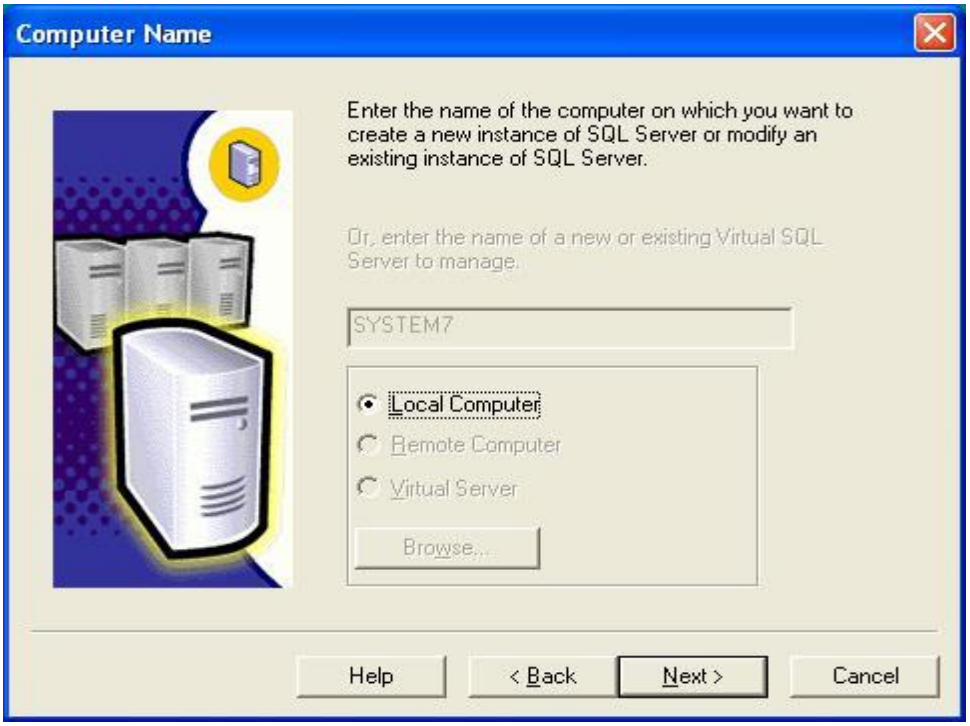
4. If Windows Server 2003 is installed, you will see the Setup window. Press the OK button.



5. You will be brought to the Welcome window to the Microsoft SQL Server Installation Wizard. Press the Next button.

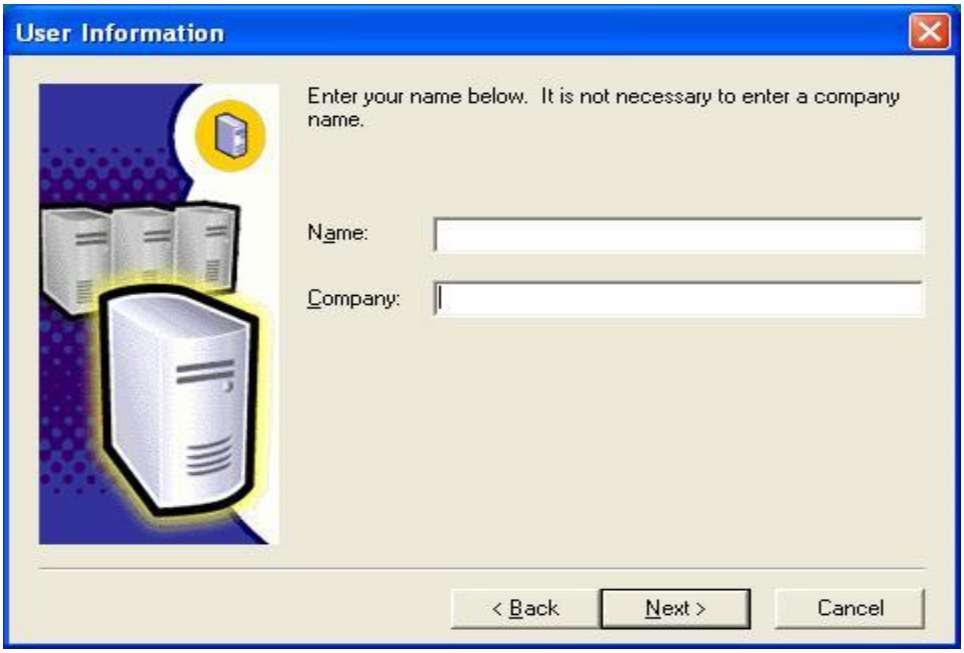


6. You will be brought to the Computer Name window. Since you are installing SQL 2000 Client, the Local Computer option will be the only selection. Press the Next button.

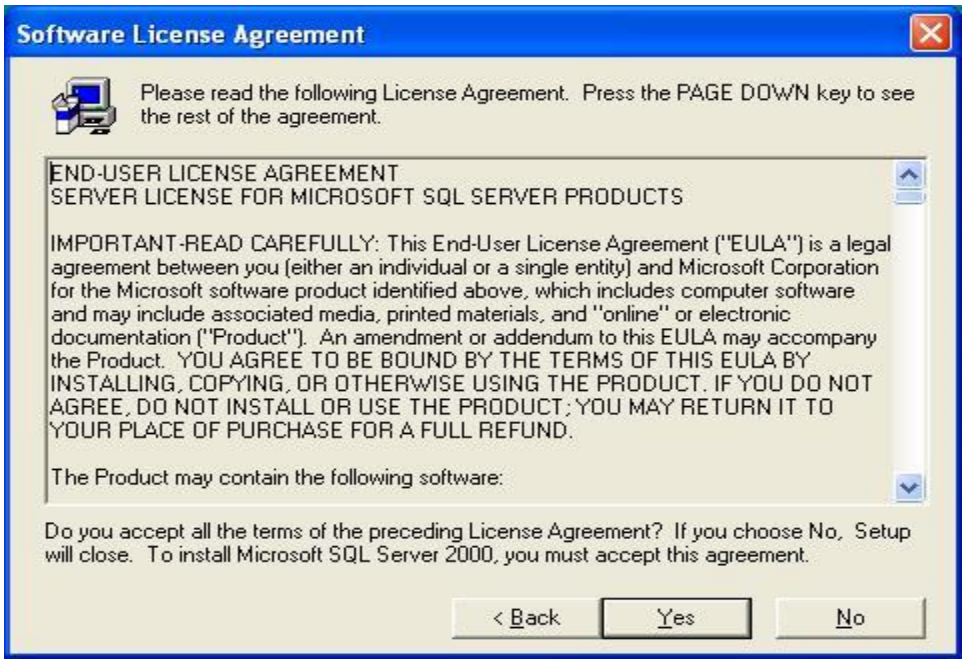


7. You will be brought to the Installation Selection window. The “Create a new instance of SQL Server, or install Client Tools" will be the only option available. Press the Next button.

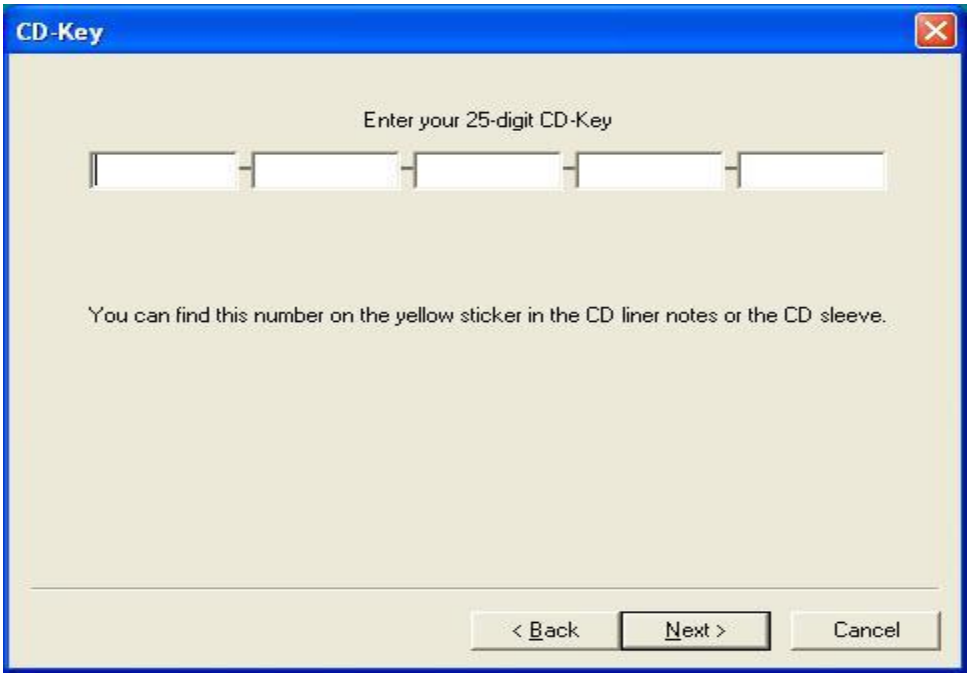
8. You will brought to the User Information window. Enter the computer Name and Company for your system. Press the Next button.



9. You will be brought to the Software License Agreement. After reading the legal agreement, press the Yes button.

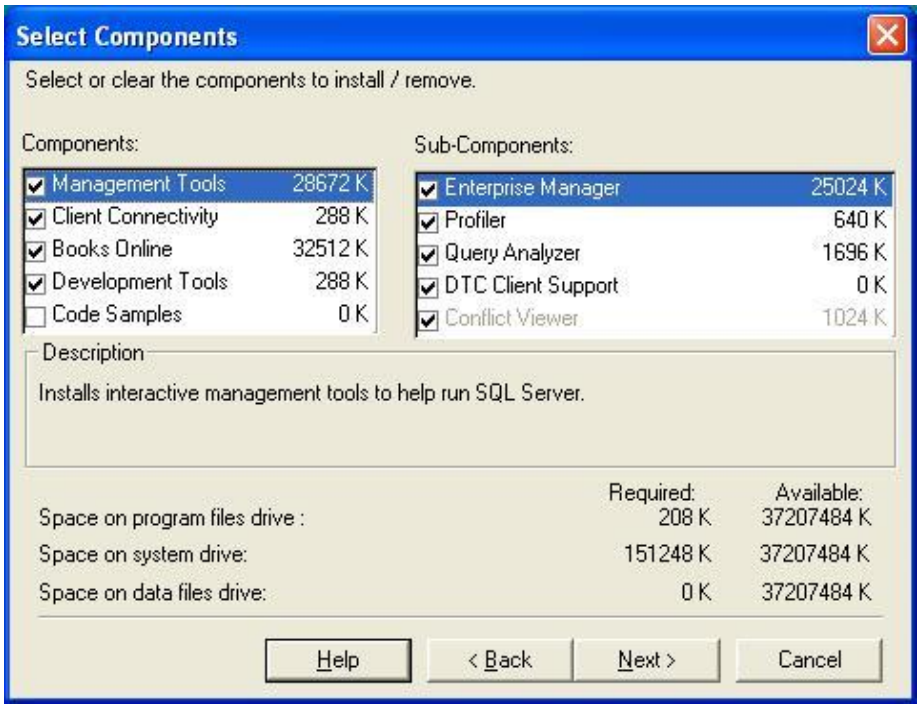


10. You will be brought to the Registration Key window. Enter the 25 digit CD-Key. Click on the Next button.

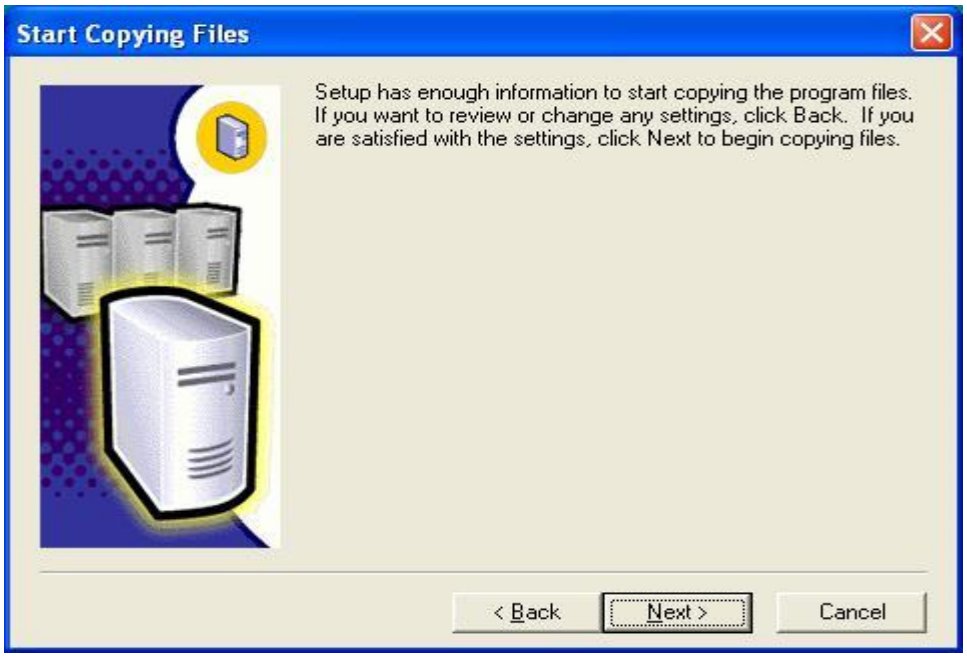


11. You will be brought to the Installation Definition window. Choose the "Client Tools Only" and then press the Next button.

12. You will be brought to the Select Components window– press on the Next button for default values.

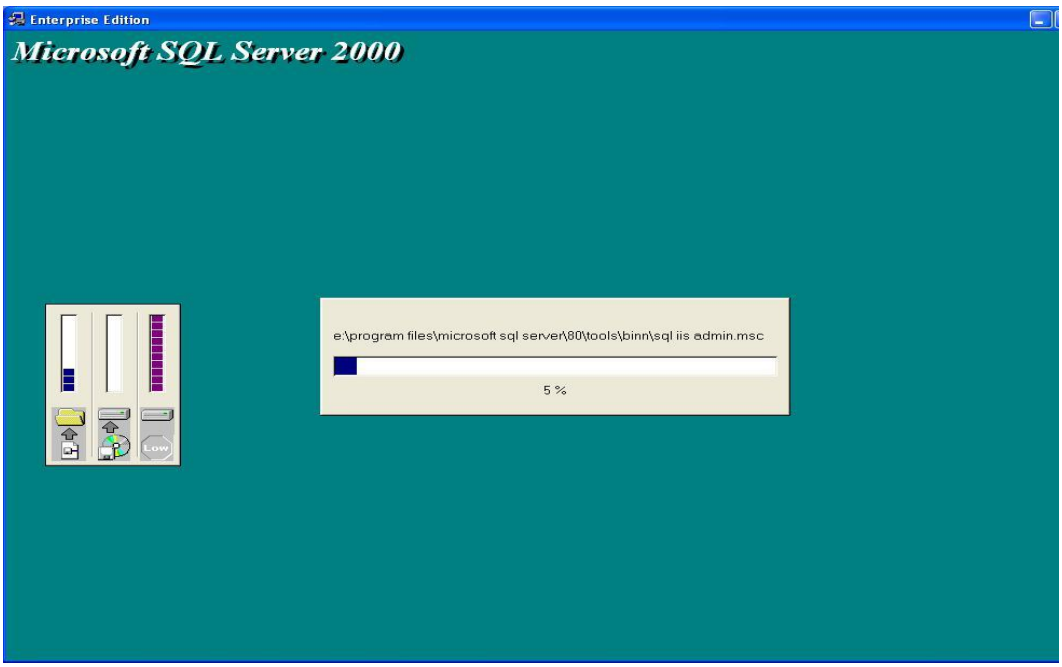


13. At the "Start Copying Files" window, press the Next button.



14. The system will display this window when loading the system.

15. SQL Server 2000 progress screen will display.



16. Once the setup is done, you will see the following window. Press the Finish button to complete the installation.

Mechanism to Conduct Lab:

Students and teacher communicate through Adobe Connect.

Lab 2

NORMALIZATION

Consider the following table

Project ID	Project Name	Project Budget	EmpID	Emp Name	Hourly Rate
1001	Pakistan International Airlines Database	1 billion	101 102 103	Sana Ali Hasan	60000 80000 45000
1002	NADRA database	20 million	111 112	Amir Umer	90000 80000

Following functional dependences exist.

1. ProjectID->ProjectName, Project Budget
2. EmpID->EmpoyeeName, HourlyRate
3. ProjectID, EmpID->ProjectName, Project Budget , EmpName, HourlyRate

Normalize above given table into first and second normal form.

Solution:

First Normal Form: Remove repeating groups

Project ID	Project Name	Project Budget	EmpID	Emp Name	Hourly Rate
1001	Pakistan International Airlines Database	1 billion	101	Sana	60000
1001	Pakistan International Airlines Database	1 billion	102	Ali	80000
1001	Pakistan International Airlines Database	1 billion	103	Hasan	45000
1002	NADRA database	20 million	111	Amir	90000
1002	NADRA database	20 million	112	Umer	80000

Second Normal Form: Remove Partial Dependencies

The above table is not in the second normal form since there exists the partial dependency through the FDs 1, 2 and 3. To bring it into second normal form, we will decompose the table into the following tables:

Employee Table

EmpID	Emp Name	Hourly Rate
101	Sana	60000
102	Ali	80000
103	Hasan	45000
111	Amir	90000
112	Umer	80000

Project Table

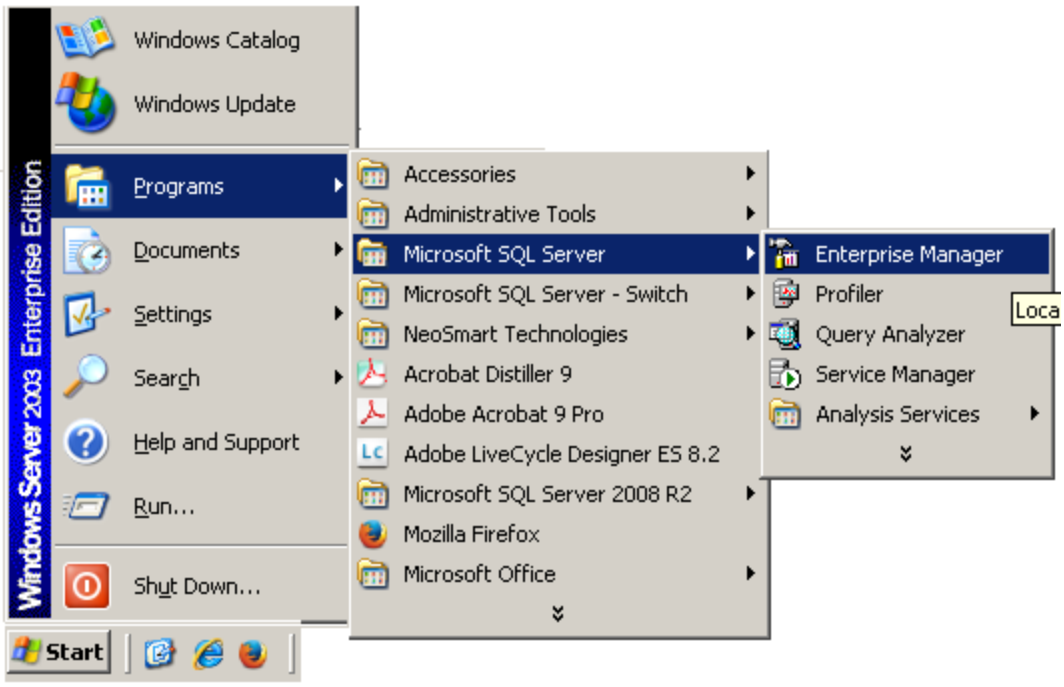
Project ID	Project Name	Project Budget
1001	Pakistan International Airlines Database	1 billion
1001	Pakistan International Airlines Database	1 billion
1001	Pakistan International Airlines Database	1 billion
1002	NADRA database	20 million
1002	NADRA database	20 million

Employee-Project Table

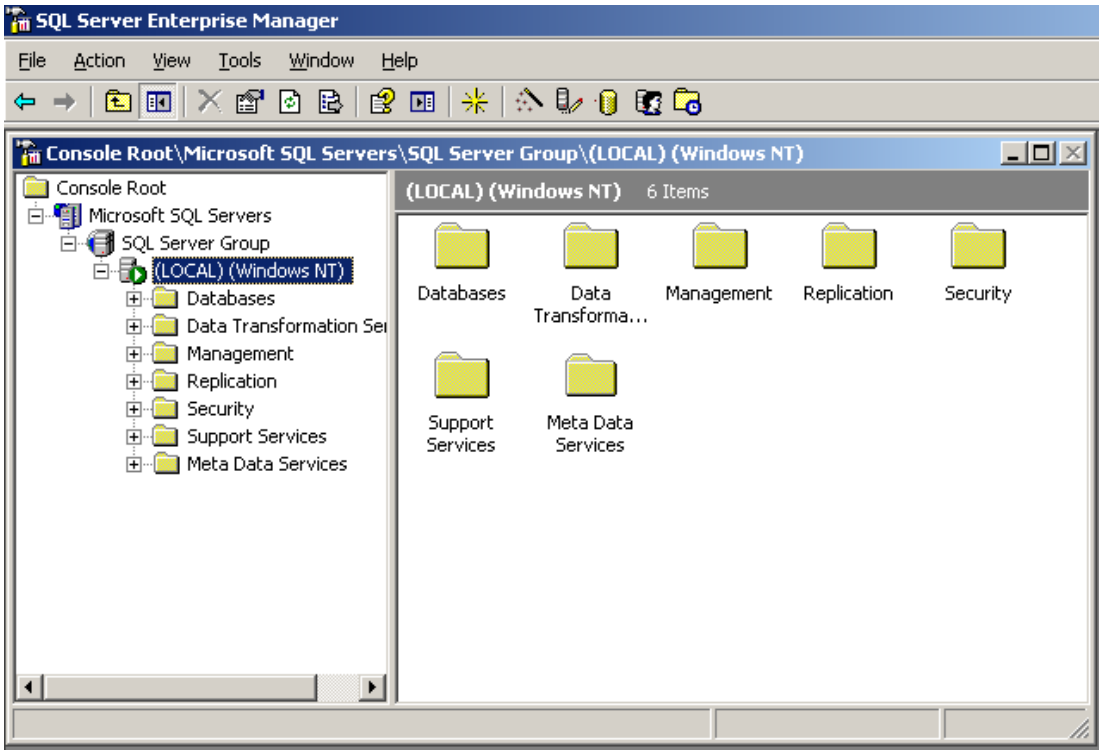
Project ID	EmpID
1001	101
1001	102
1001	103
1002	111
1002	112

Now, above tables are in second normal form. Create above all tables in SQL server enterprise manager in normalized form. The procedure of table creation is given below.

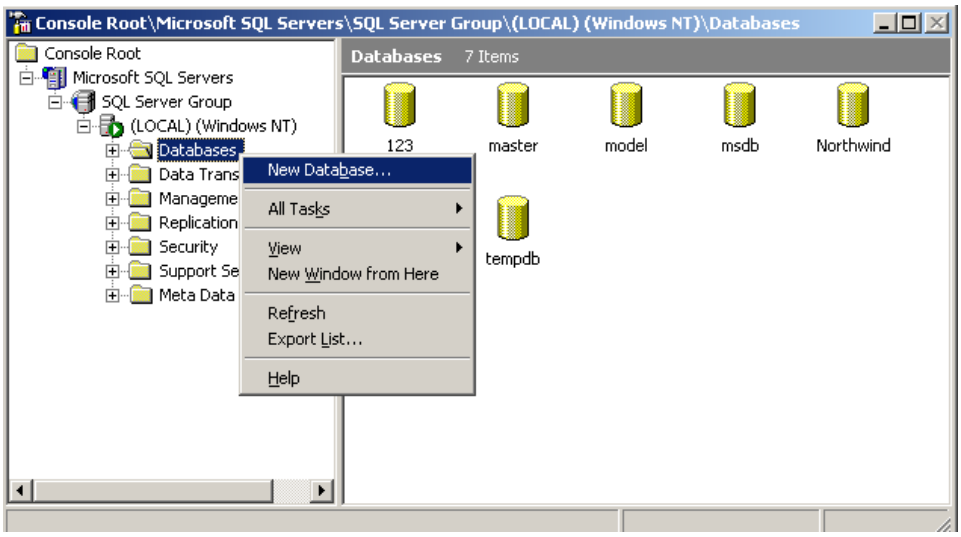
1. Open SQL Server Enterprise Manager by clicking on **Start menu->Programs->Microsoft SQL Server->Enterprise Manager**.



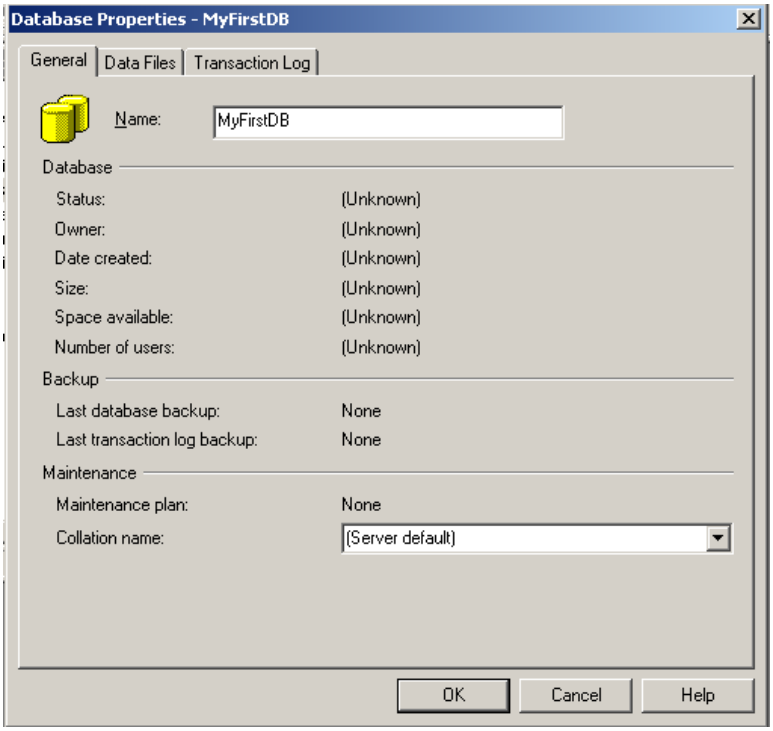
2. Click server node



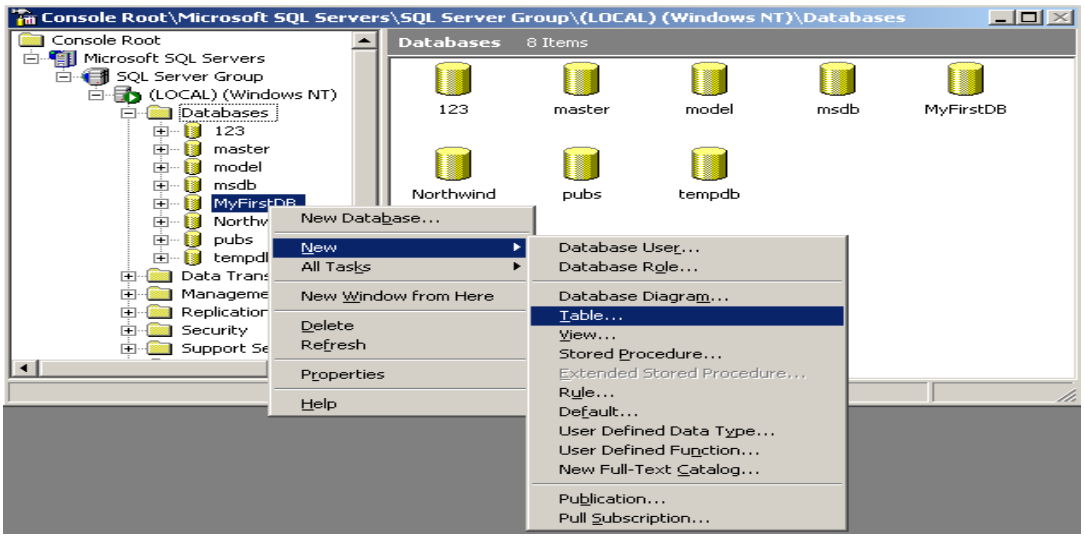
3. Right Click on **Databases** node and click on **New Database…**



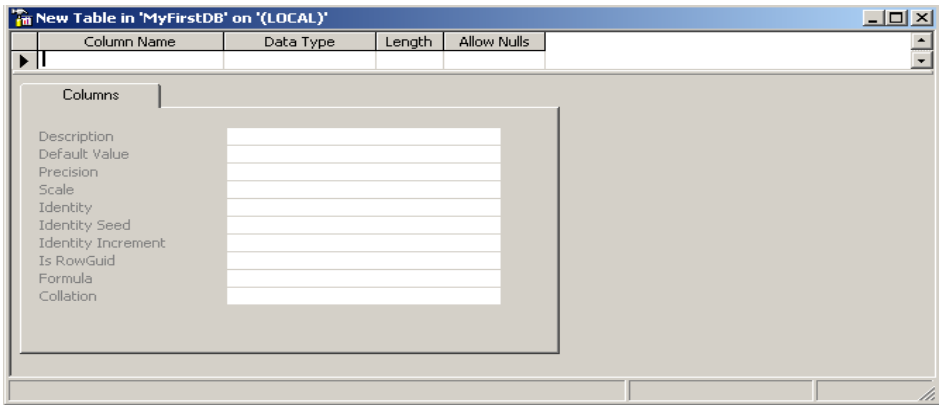
4. Type MyFirstDB in **Name** as Database Name and click oK.



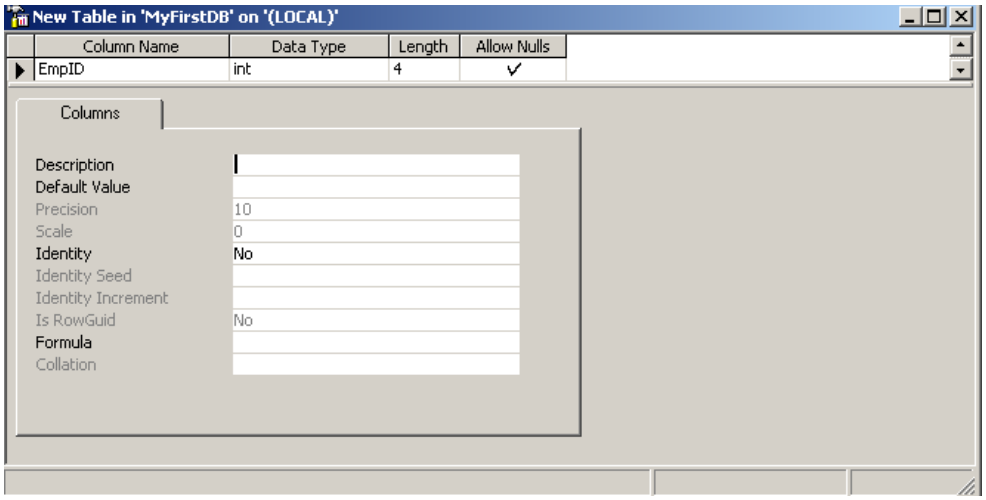
5. Expand **Databases** node, right click on your newly created databases **MyFirstDB,** click on **New** and then click on **Table**.



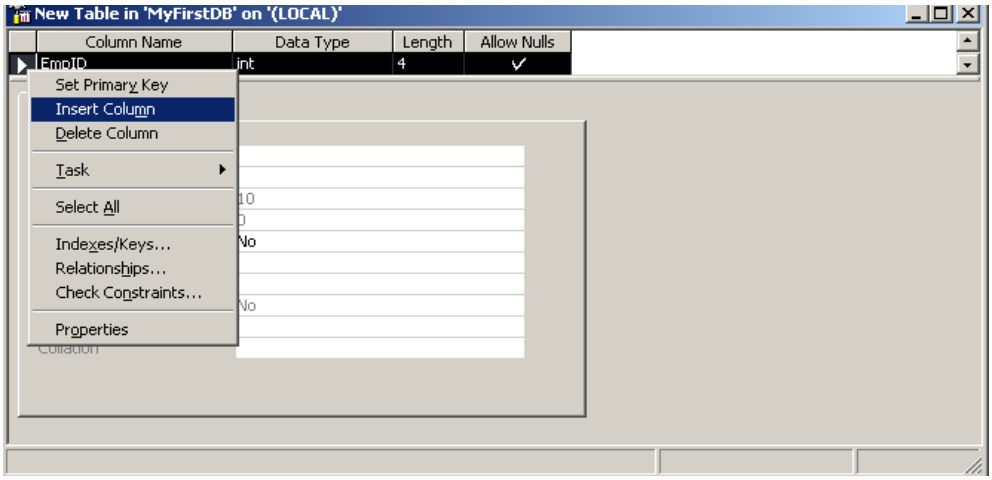
6. The following window will open.



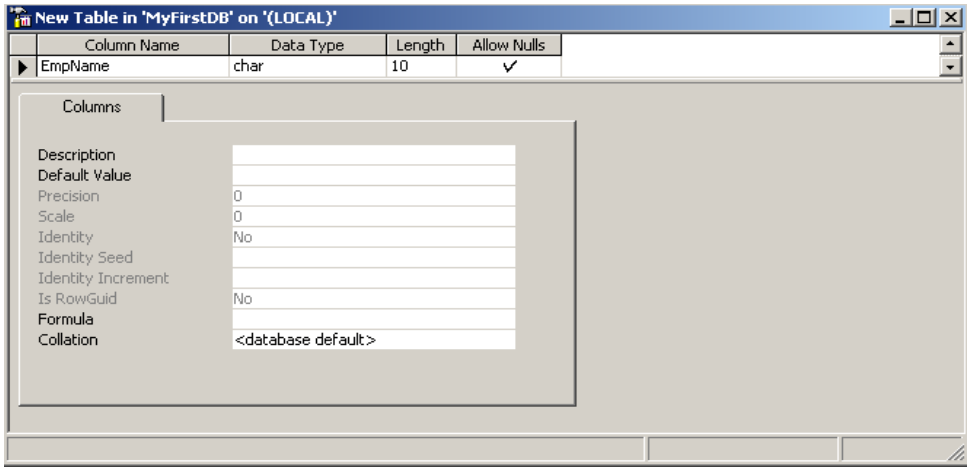
7. Type Column Name and select data type. See following figure in which we have created variable EmpID.

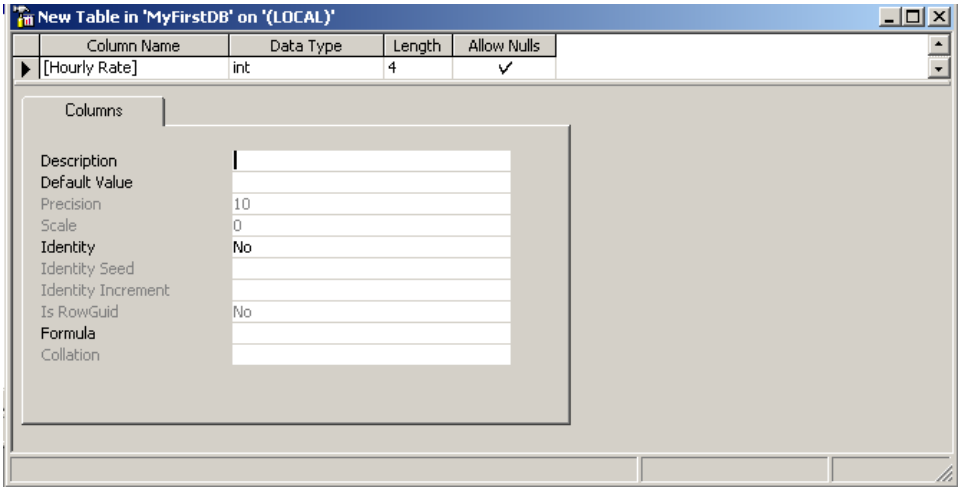


8. Right Click on column row and select option **Insert Column**.

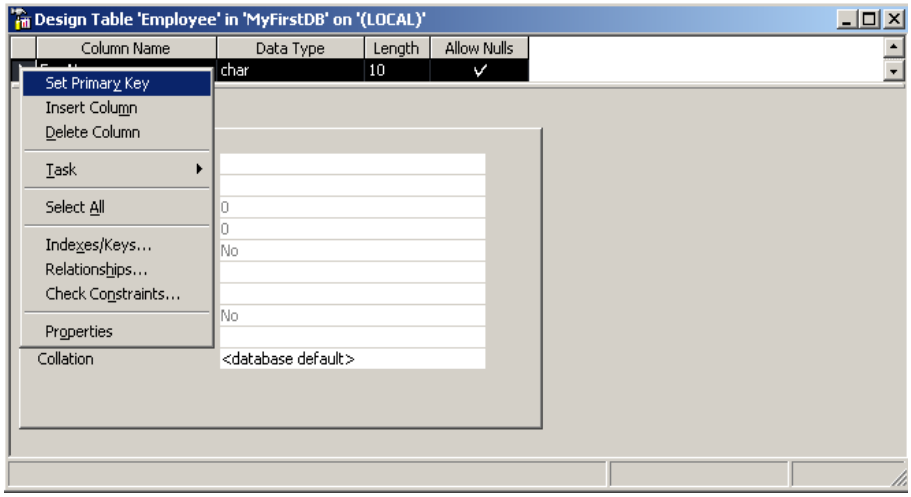


9. Similarly create other variables, EmpName and Hourly Rate

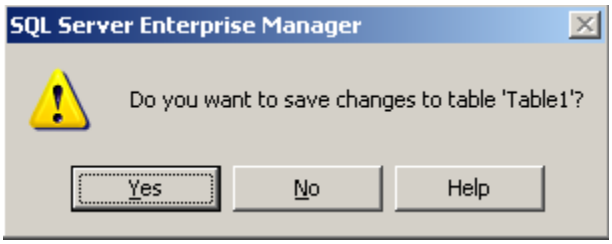




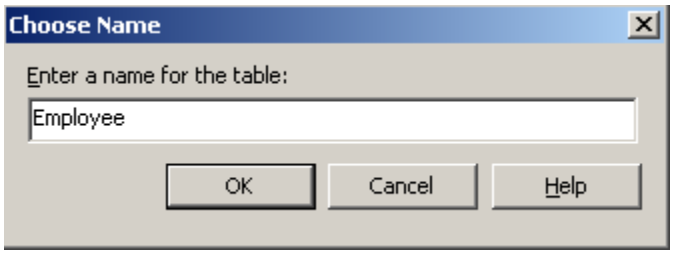
10. To set a column as primary key, right click on column name and click **Set Primary Key**.



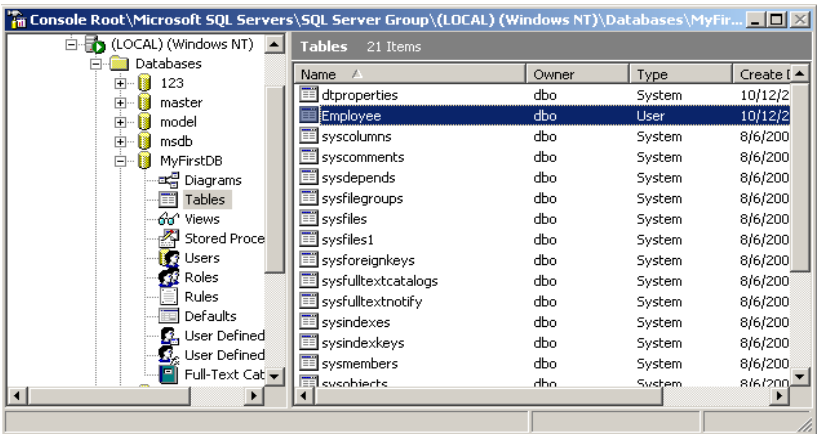
11. Click close icon, the **SQL Server Enterprise Manager** will open. Click **Yes**.



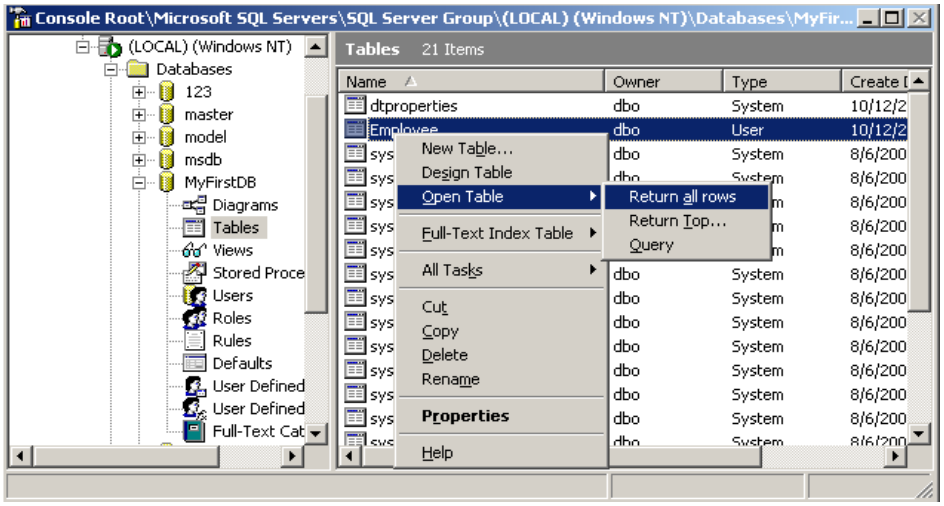
12. Type Employee in variable name. Click OK.



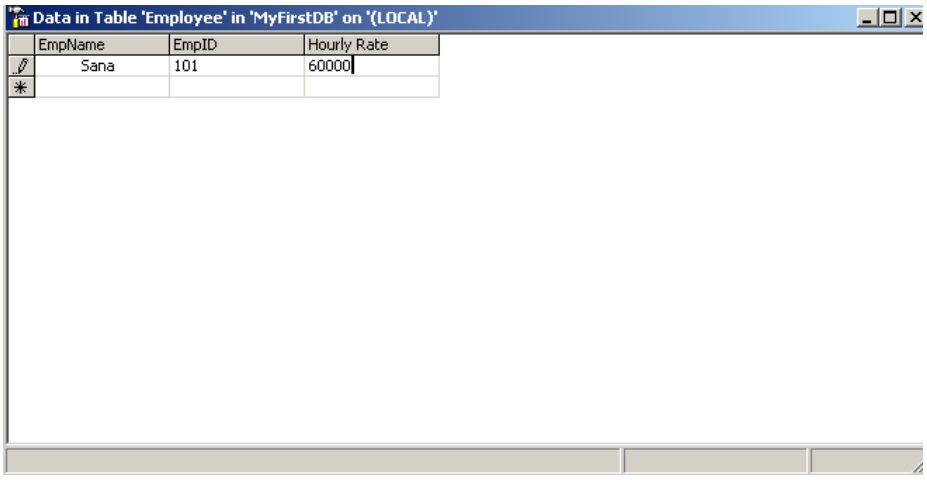
13. See in the following figure the new created table Employee. It is added in Tables list.



14. Right click on table name, select **Open Table** options and then **Return all rows** option.



15. Enter data in table manually or using some application.



Follow above method for all tables.

Mechanism to Conduct Lab:

Students and teacher communicate through Adobe Connect.

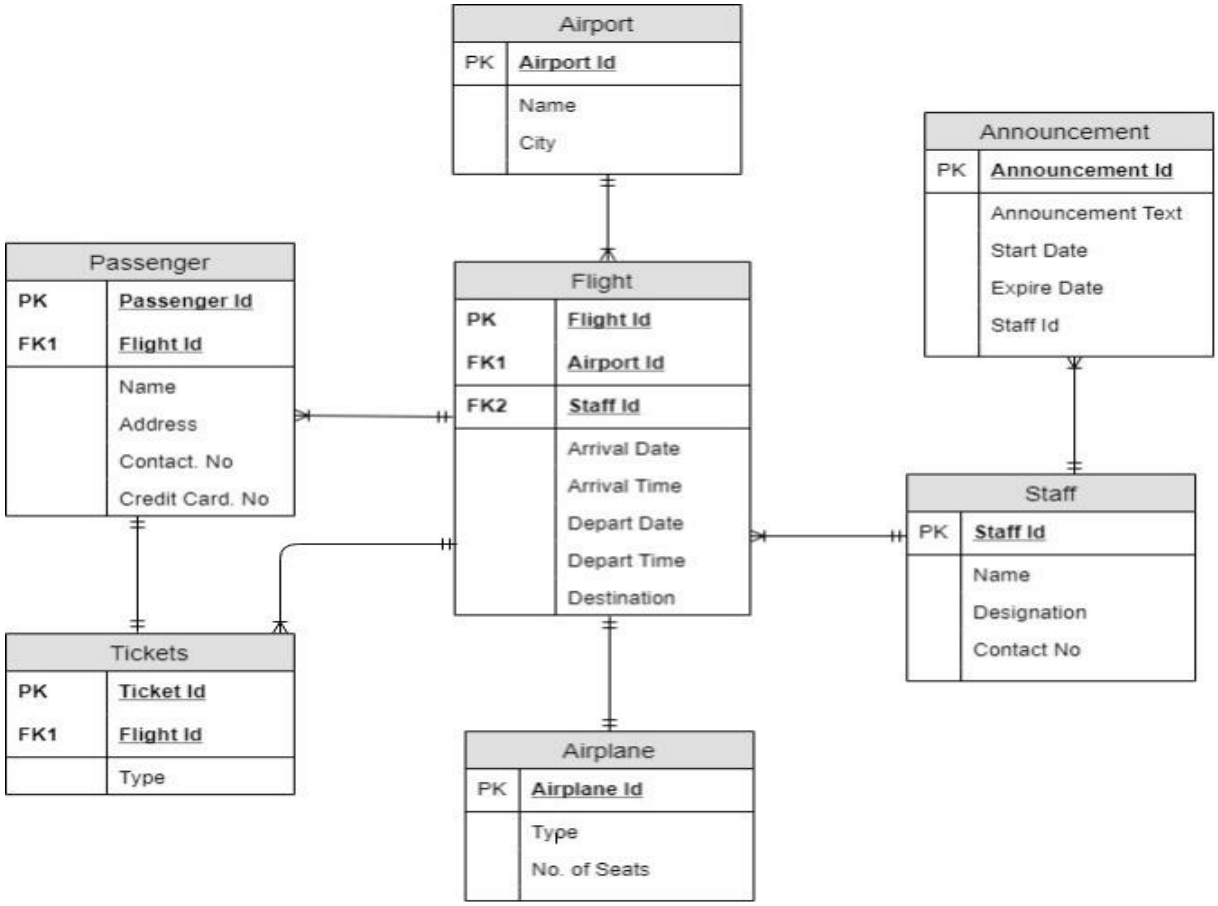
Lab 3

DENORMALIZATION

Scenario

"Pak Airline" is an airliner reservation company, which is operating in more than 10 countries. They have developed the airline reservation system to avoid the errors faced in manual system. The staff of the airline use airline reservation system form the tasks such as flight scheduling, ticket reservation, announcements in automated way. Similarly, users/passengers can search for flight schedule according to date and time and fare details. The staff of the airline can manage the reservation systems by flight rout, runway details, flight scheduling and reservation.

Ticket reservation system of the Pak Airline provides the information about schedule of flights, availability of seats, flight number and destination. For reservation of ticket user have to provide its personal information such as name, age, address etc. For payment purpose user will provide credit card number and bank details. Moreover, information about flight number, date of departure, no. of tickets to be booked is also required for confirmation of ticket. Following is the ERD of above airline reservation system.



Question Statement:

You are required to de-normalize above tables using table pre-joining technique according the relationship between entities. Carefully identify all tables with such relationships on which Pre-joining technique can be applied.

Solution:

There are One-to-Many relationships in following entities:

1. Staff-Announcement
2. Airport-Flight
3. Flight-Passenger
4. Staff-Flight
5. Flight-Ticket

As Pre-joining De-normalization technique is based on 1-many relationship so the De-normalization will be performed on following tables.

Airport Table

Airport Id	Name	City

Flight Table

Flight Id	Arrival date	Arrival time	Depart Date	Depart Time	Destination	Airport Id	Staff Id

Staff Table

Staff Id	Name	Designation	Contact No

Announcement Table

Announcement Id	Announcement Text	Start Date	Expire Date	Staff Id

Passenger Table

Passenger Id	Name	Contact No	Address	Email	Credit Cr.No	Flight Id

Ticket Table

Ticket Id	Type	Flight Id

Merge Staff and announcement table as there is one to many relationship which is the requirement of Pre-joining.

Staff-Announcement

Staff Id	Name	Designation	Contact No	Announcement Id	Text	Start Date	Expire Date

Airport-Flight Table

Airport Id	Name	City	Flight Id	Arrival date	Arrival time	Depart Date	Depart Time	Destination

Flight –Passenger Table

Flight Id	Arrival date	Arrival time	Depart Date	Depart Time	Destination	Passenger Id	Name	Contact No	Address	Email	Credit Cr.No	Airport Id	Staff Id

Staff-Flight Table

Staff Id	Name	Designation	Contact No	Flight Id	Arrival date	Arrival time	Depart Date	Depart Time	Destination

Flight-Ticket Table

Flight Id	Arrival date	Arrival time	Depart Date	Depart Time	Destination	Airport Id	Staff Id	Ticket Id	Type

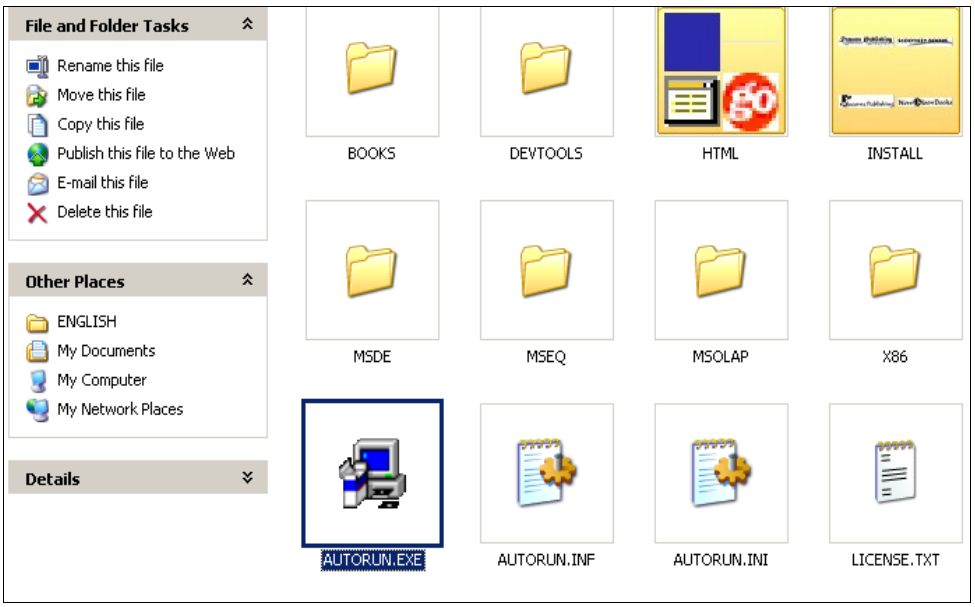
Mechanism to Conduct Lab:

Students and teacher communicate through Adobe Connect.

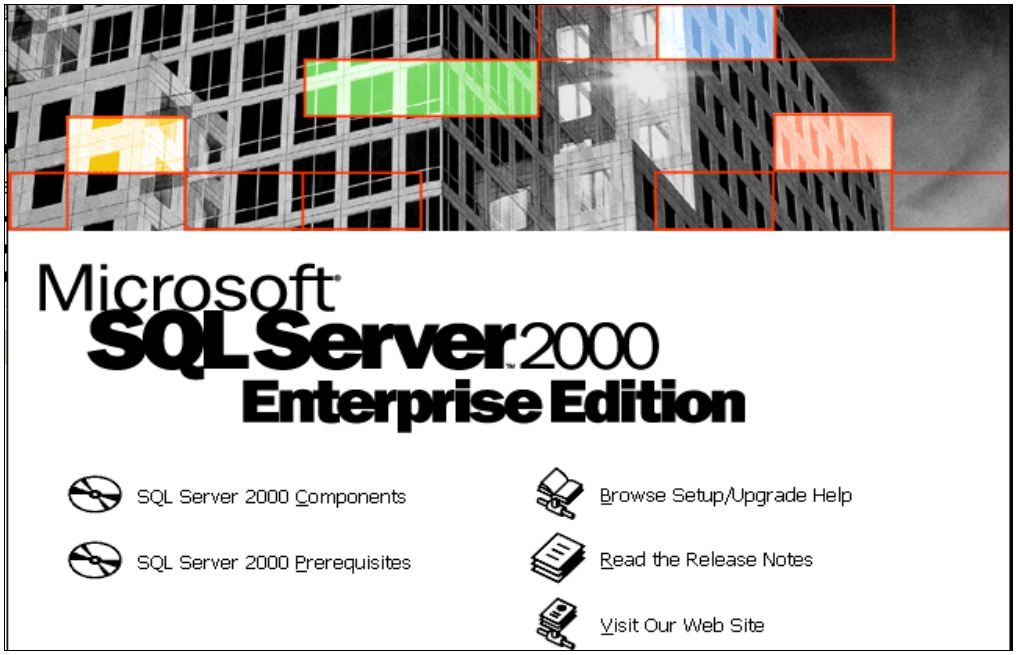
Lab 4 Part I

Installation Guide for MS SQL Server 2000 Analysis Services

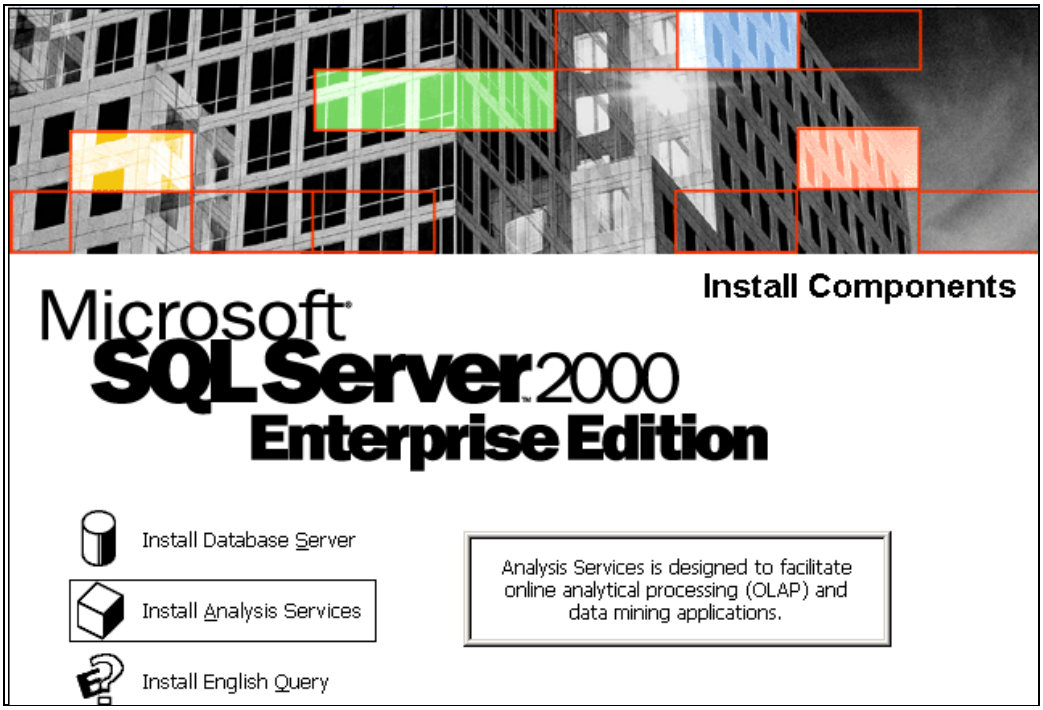
1. Follow all instructions in Lab lecture 1 section 3.2 “Installing Microsoft SQL Server 2000” and then Double Click on “AutoRun.Exe” icon.



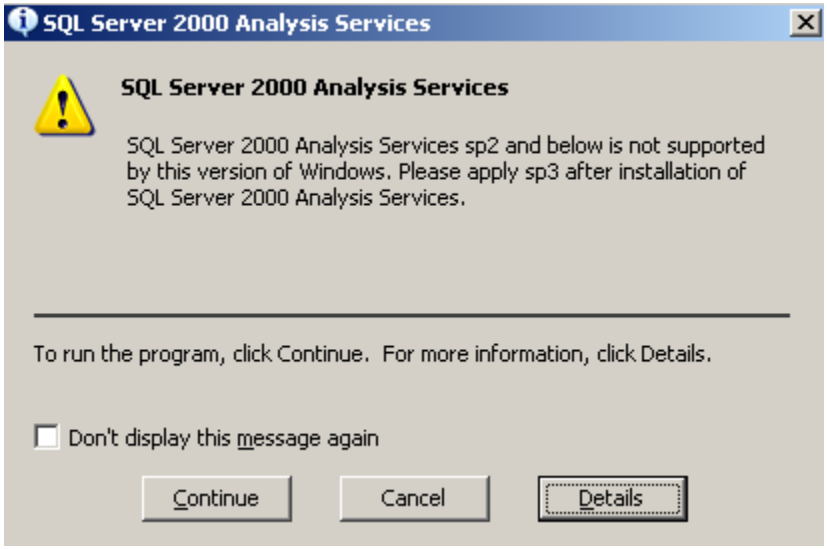
2. The Microsoft SQL Server 2000 Enterprise Edition screen will appear. Now, press the SQL Server 2000 Components button.



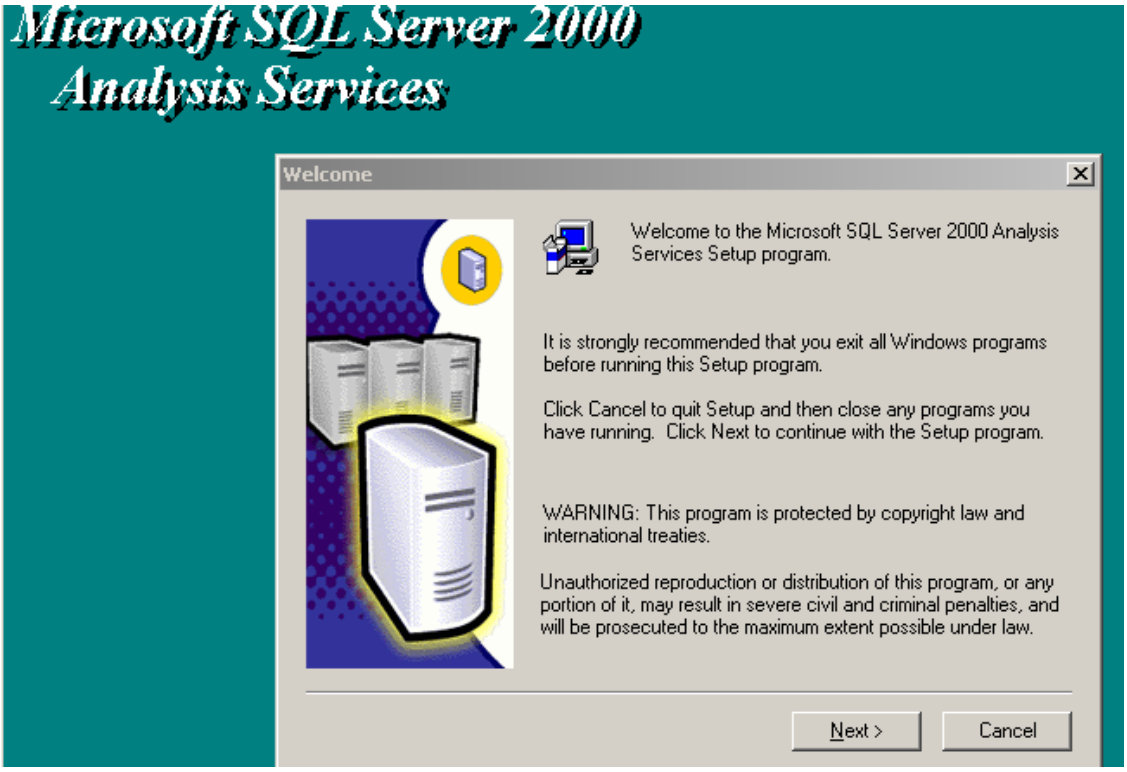
3. Now, Press “Install Analysis Services” button.



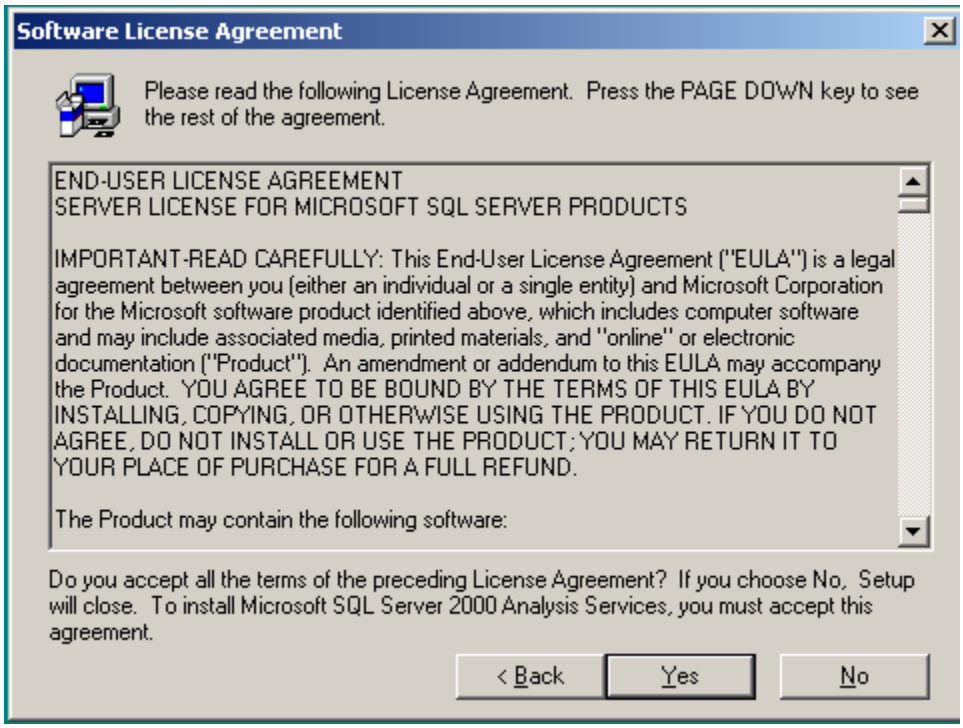
4. The following window will appear. Click “Continue” button.



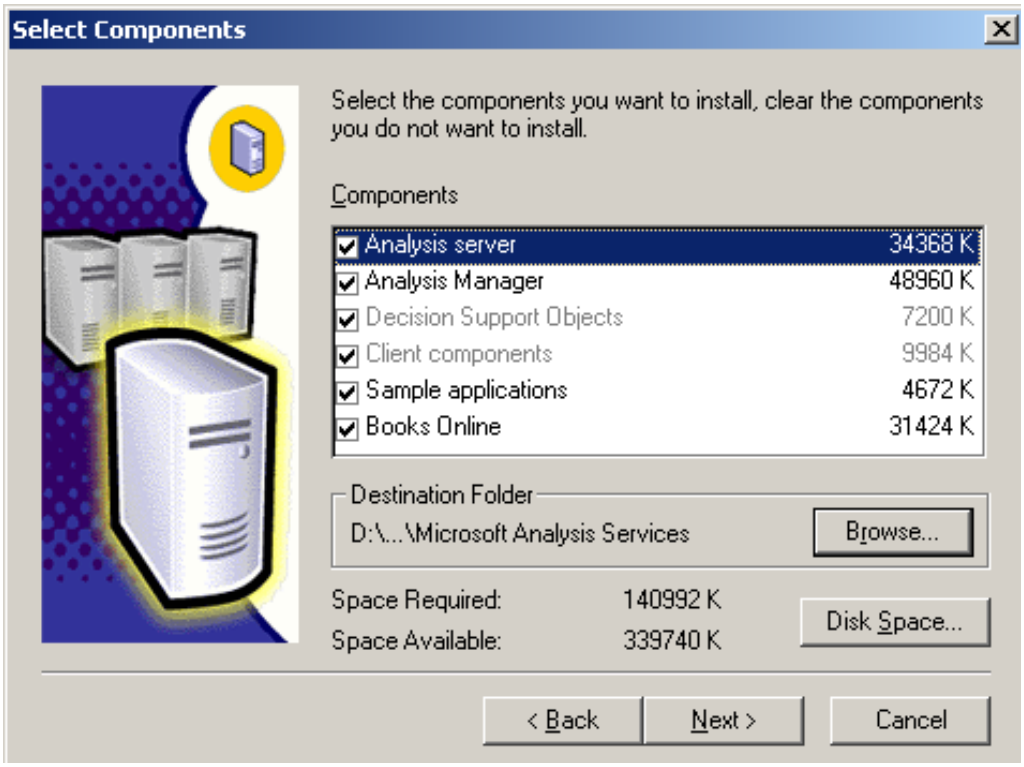
5. From Microsoft SQL Server 2000 Analysis Services screen, Click Next button.



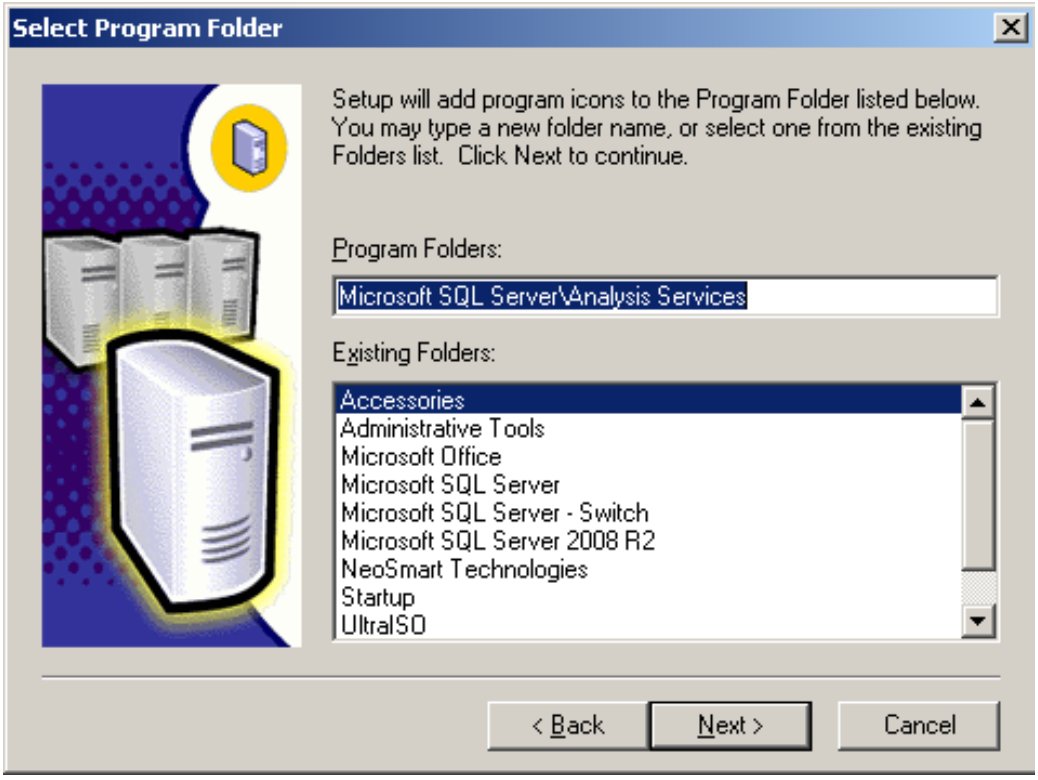
6. Click “Yes” from following “Software License Agreement” window.



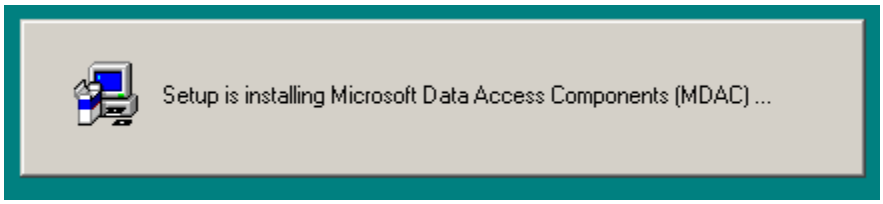
7. Select Component as selected in following “Select Components” window.



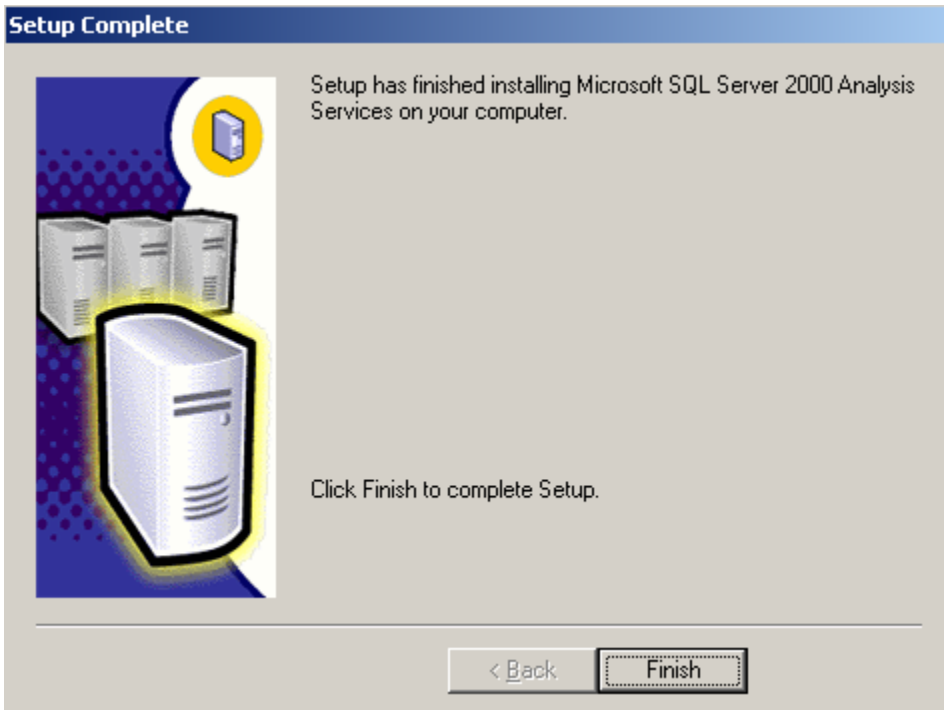
8. Select Program Folder and click “Next” button.



9. The following window will appear.

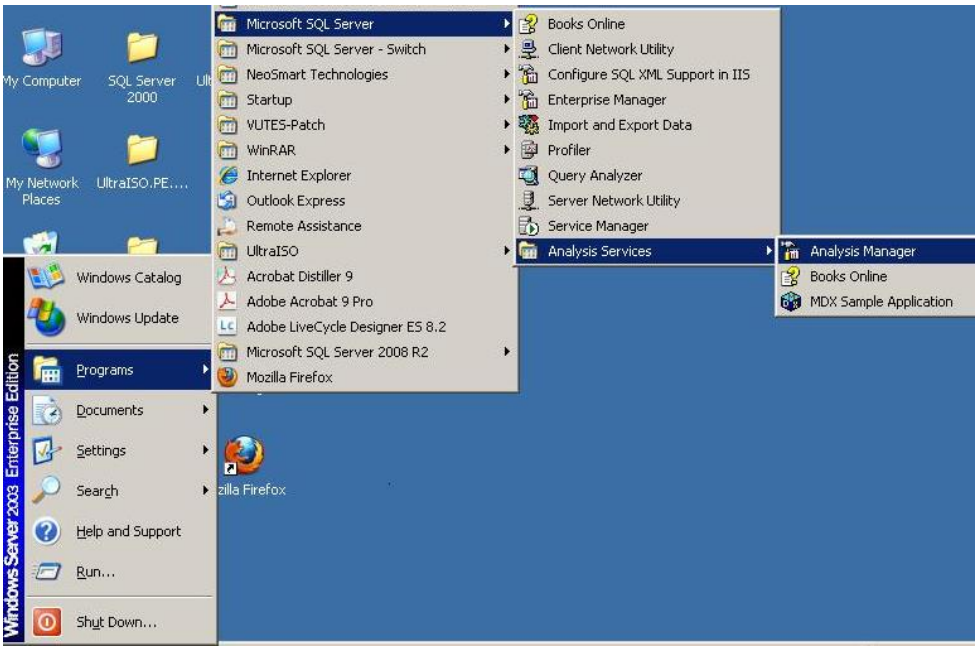


10. Click “Finish” button to complete setup.

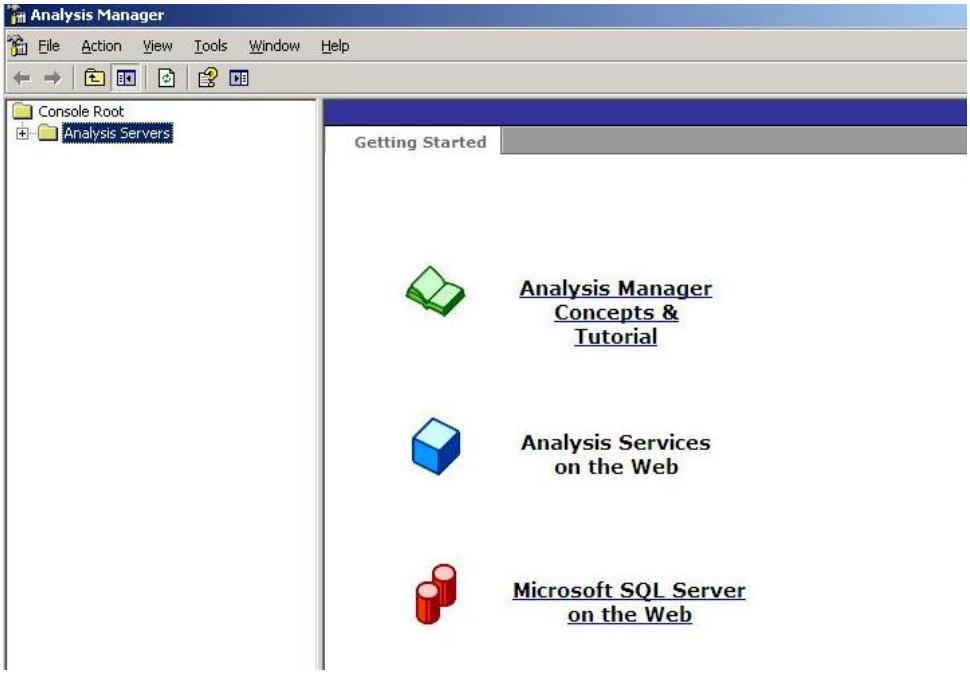


Working with Analysis Manager

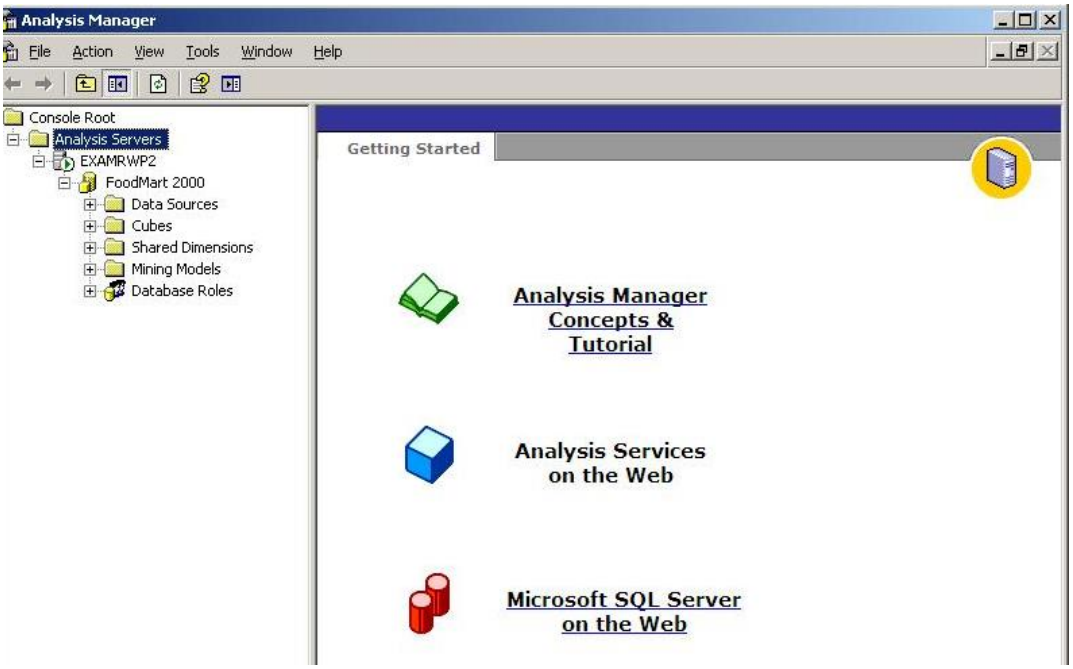
1. Open **Analysis Manager** by clicking Programs->Microsoft SQL Server->Analysis Services->Analysis Manager**.**



2. The following **Analysis Manager** Window will open. To get help on working with **Analysis Manager**, click on **Analysis Manager Concepts & Tutorial.**



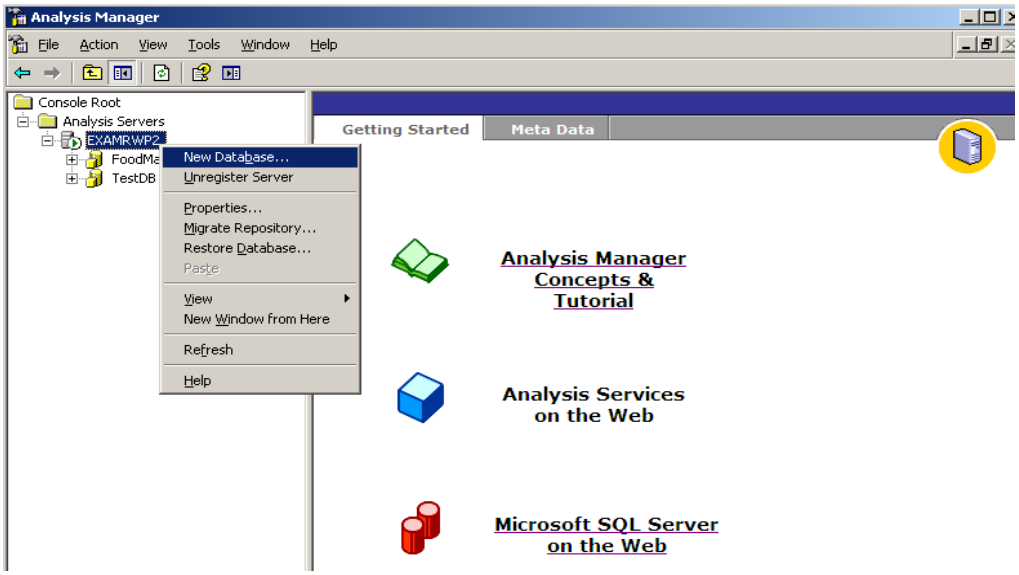
3. In the Analysis Manager tree view, expand **Analysis Servers**.



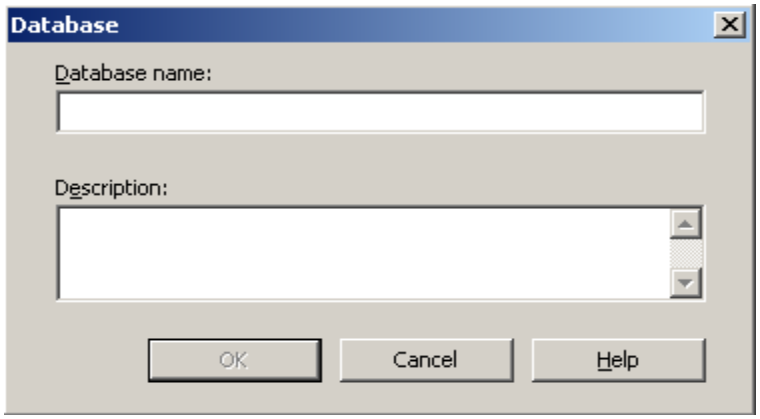
4. In above screen, **FoodMart2000** under the server node is the sample database available in Analysis Manager.

5. **To create a new database:**

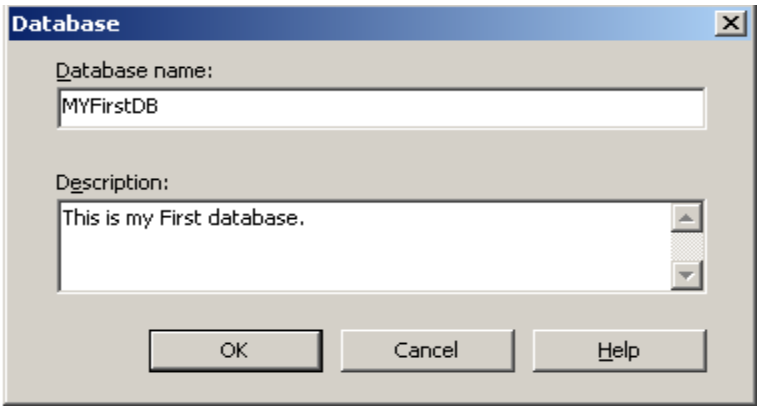
1. Click the name of your server (e.g. E**XAMRWP2**), a connection with the Analysis server will be established.
2. Right-click your server's name, and then click **New Database**.



3. The following database dialog box will open.



4. In the **Database** dialog box, in the **Database name** box, enter "MYFirstDB", and in **Description** box, enter "This is my Fisrt database.", then click **OK**.



5. In the Analysis Manager tree pane, expand the server, and then expand the **MYFirstDB** database that you just created.
6. Your new **MYFirstDB** database contains the following items:

- Data Sources
- Cubes
- Shared Dimensions
- Mining Models
- Database Roles

Lab Exercise: Explore sample database **FoodMart2000,** all of its tables, tables' schema and browse their data.

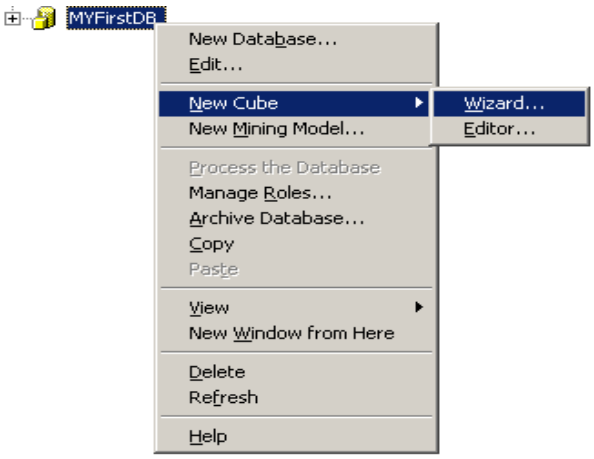
Mechanism to Conduct Lab:

Students and teacher communicate through Skype/Adobe Connect. Students perform the task using the following simulator:

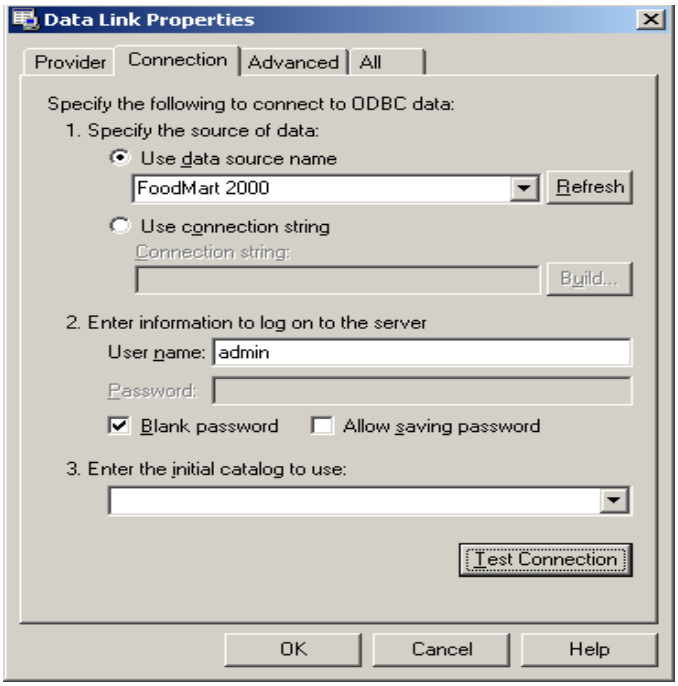
Lab 4 Part II

CREATING AND PROCESSING A CUBE

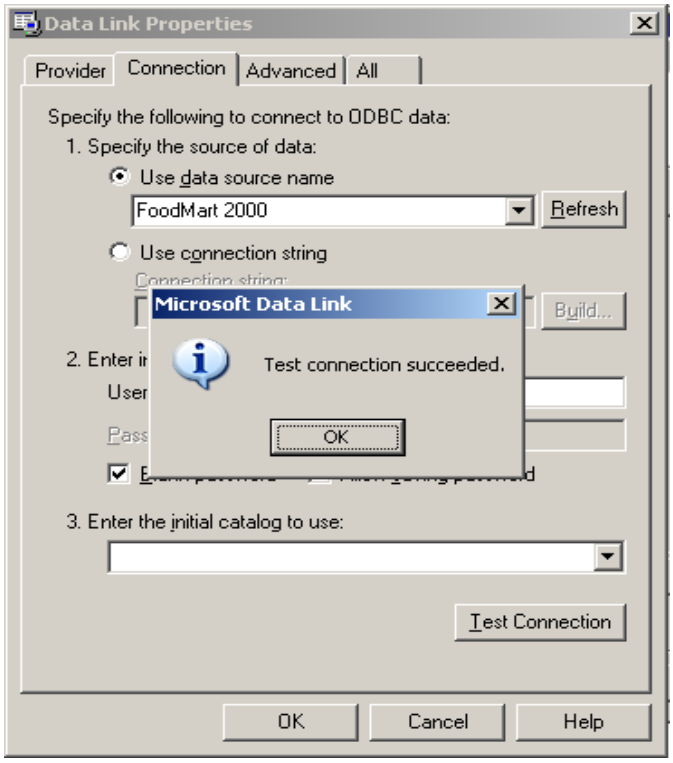
1. Click on newly created database **"MYFirstDB",** from its drop down list, select **New Cube,** then select **Wizard** option**.**



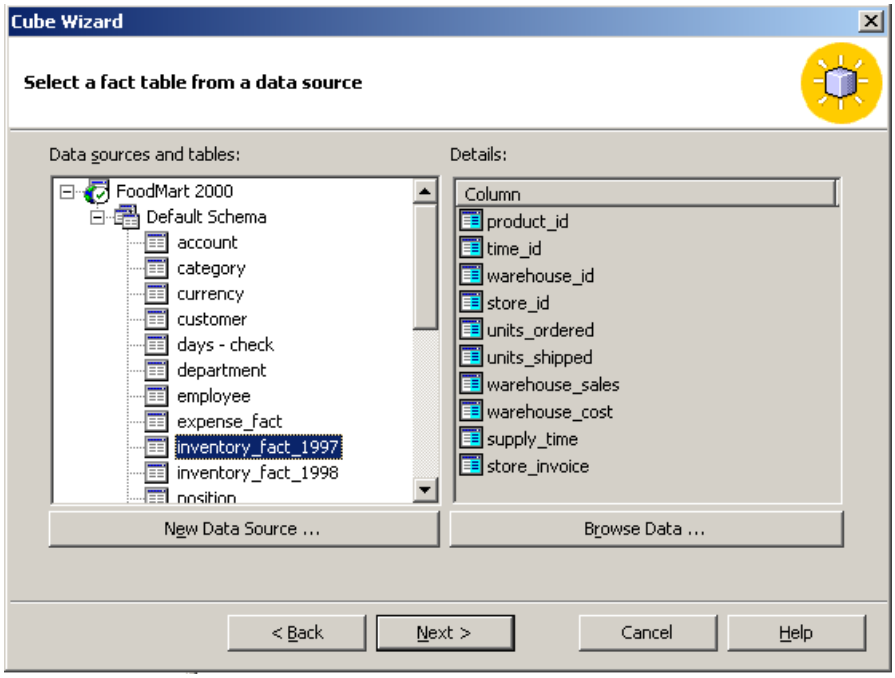
2. The following **Data Link Properties** window will open. Select option **Use data source name** and then select **FoodMart2000** from its dropdown list.
3. Write **Admin** as **User name** and check **Blank Password** option.
4. After this, Click Test Connection button.



5. You will get message “Test Connection Succeeded”. Click OK, then again click OK from Data Link Properties.

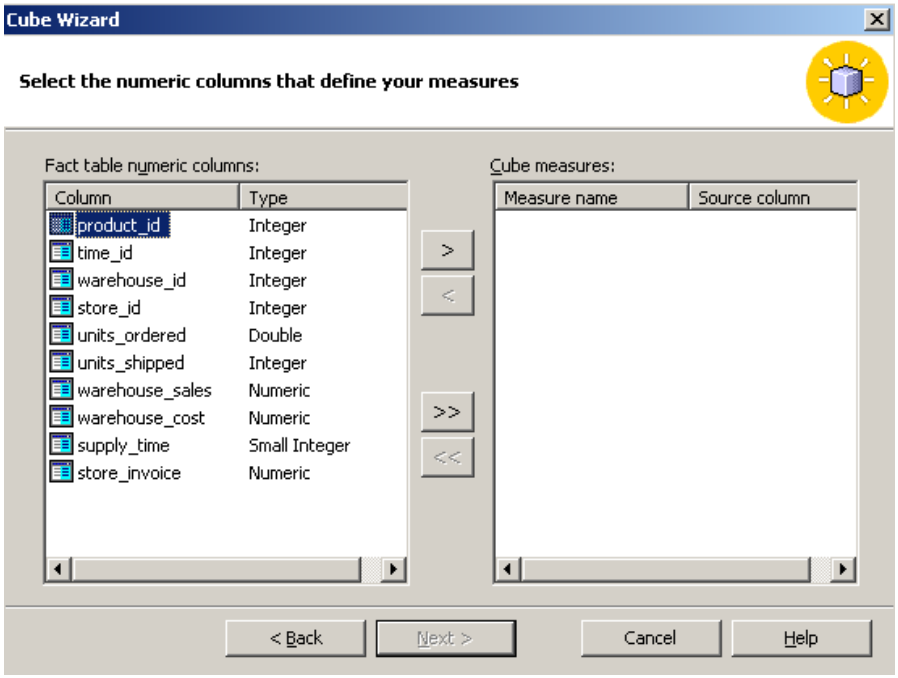


6. Cube Wizard window will show different tables which are created in FoodMart2000 database. Select **inventory_fact_1997** table.

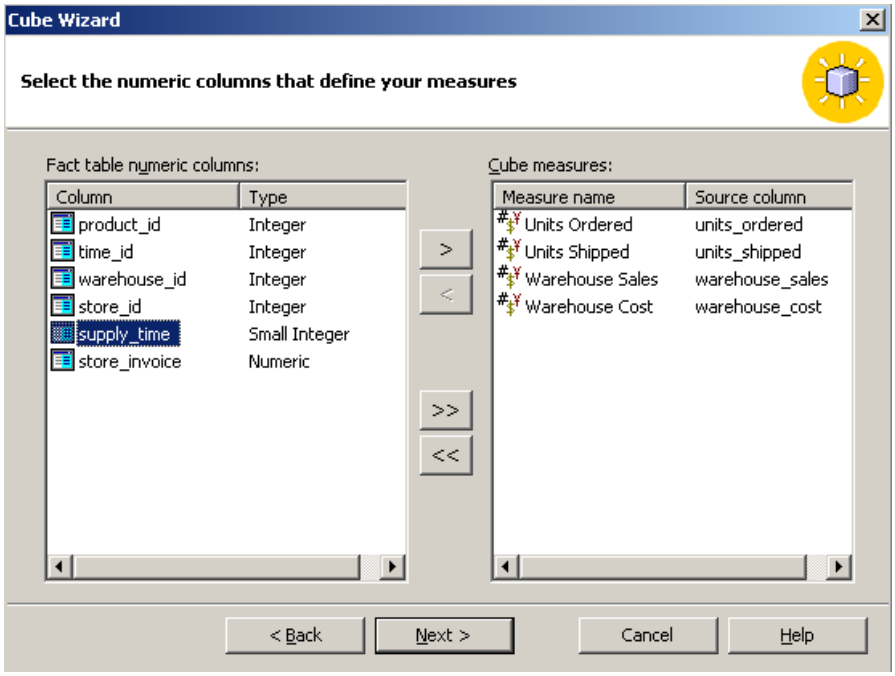


7. You can browse data of selected table by clicking on Browse Data button (See above window).

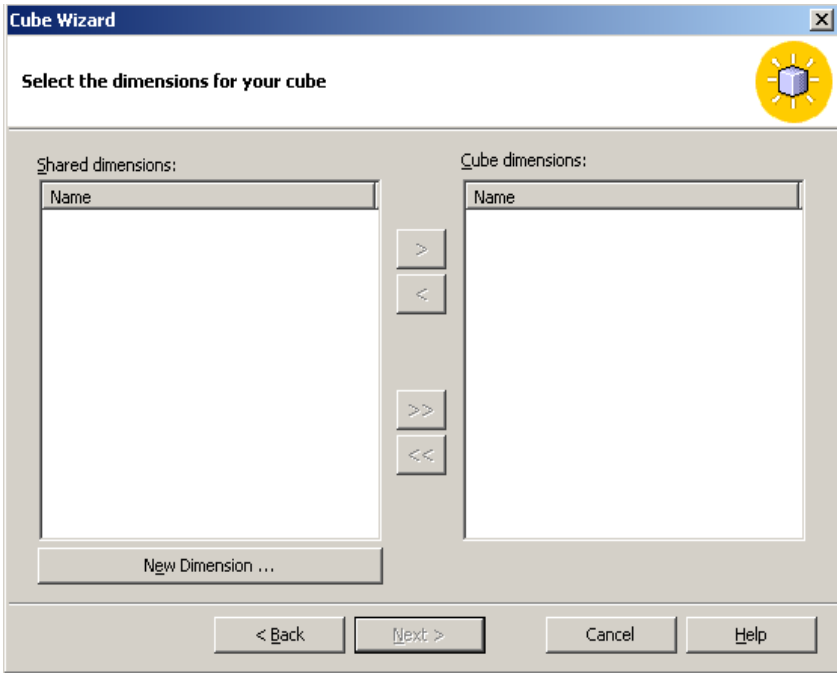
8. Now you will define your cube measures. Select table numeric columns to define cube measures.



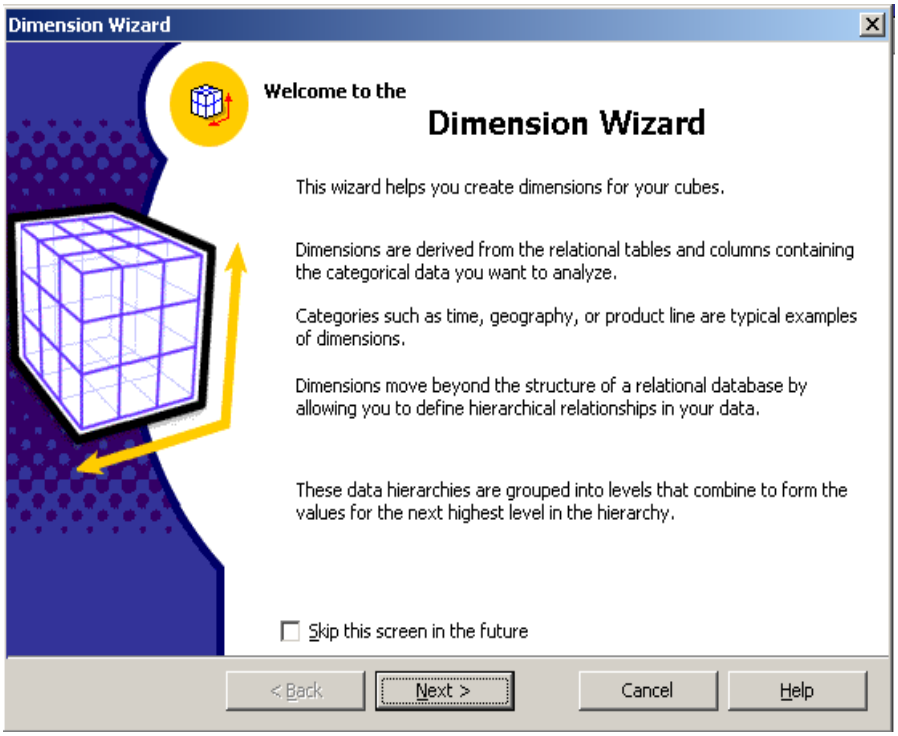
9. Select e.g. Units Ordered, Links Shipped,, Warehouse Sales, and Warehouse Cost. Then click on Next button.



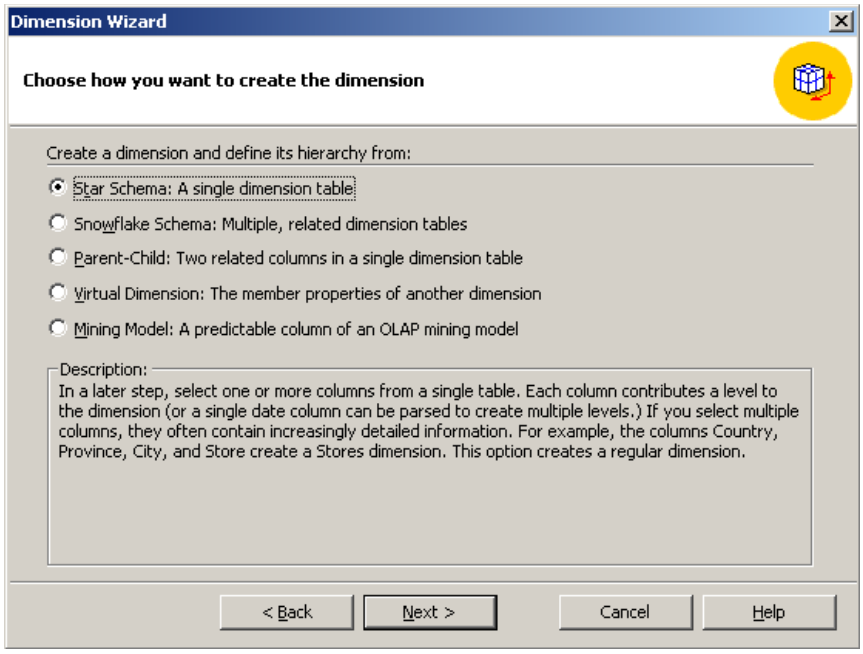
10. Now, you have to select cube dimension. From **Cube Wizard** window, click on **New Dimension.**



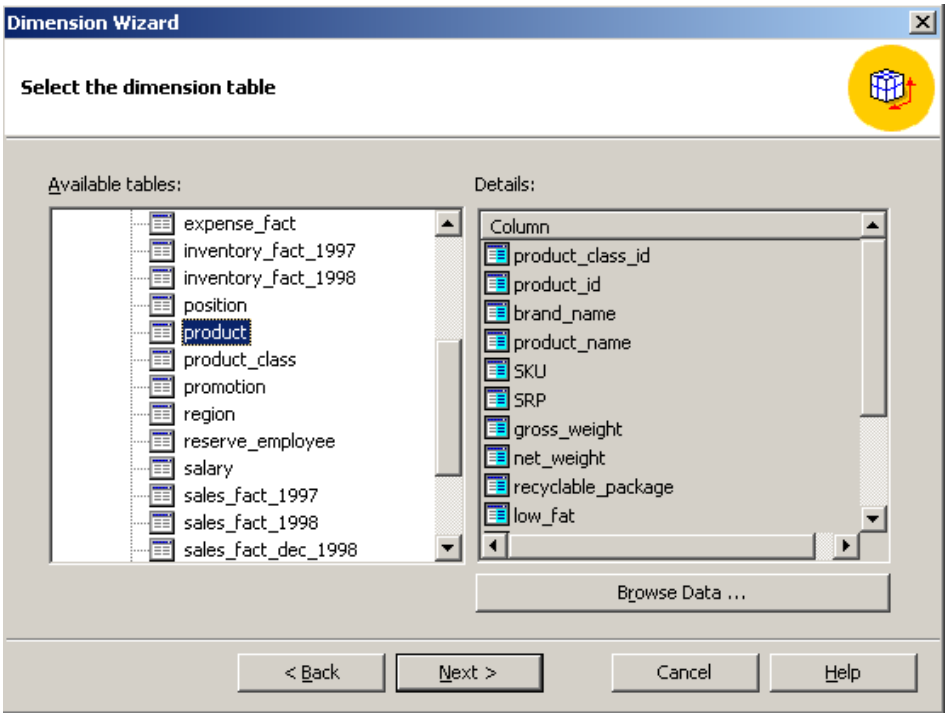
11. Click on **Next** button.



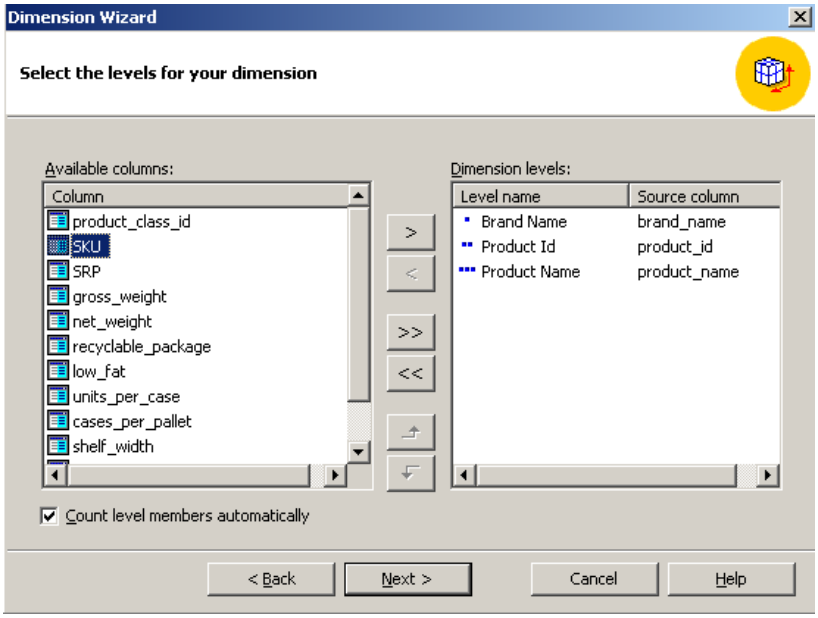
13. Select **Star Schema: A single dimension tables** option**.** You can use some other option too. Click **Next**.



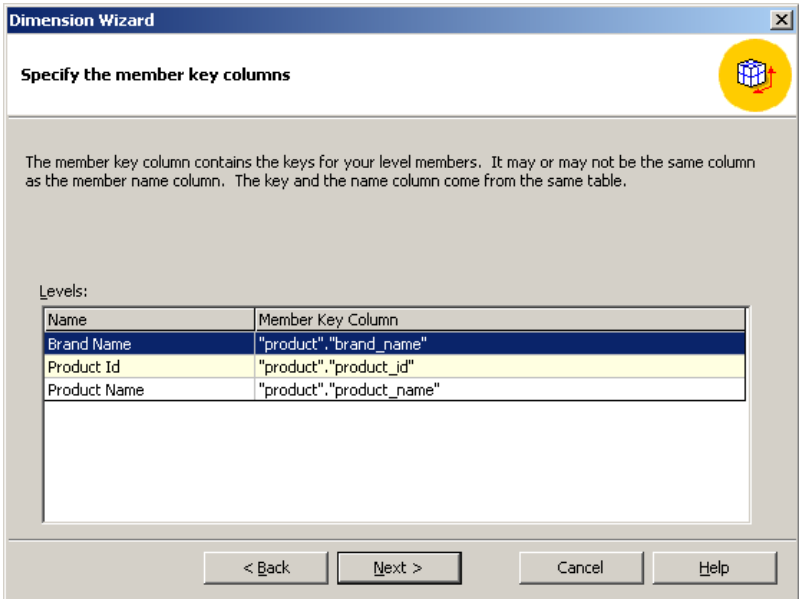
14. Now, you have to select Dimension table. In this lab, we have selected **product** table. You can choose any table of your choice. Click **Next**.



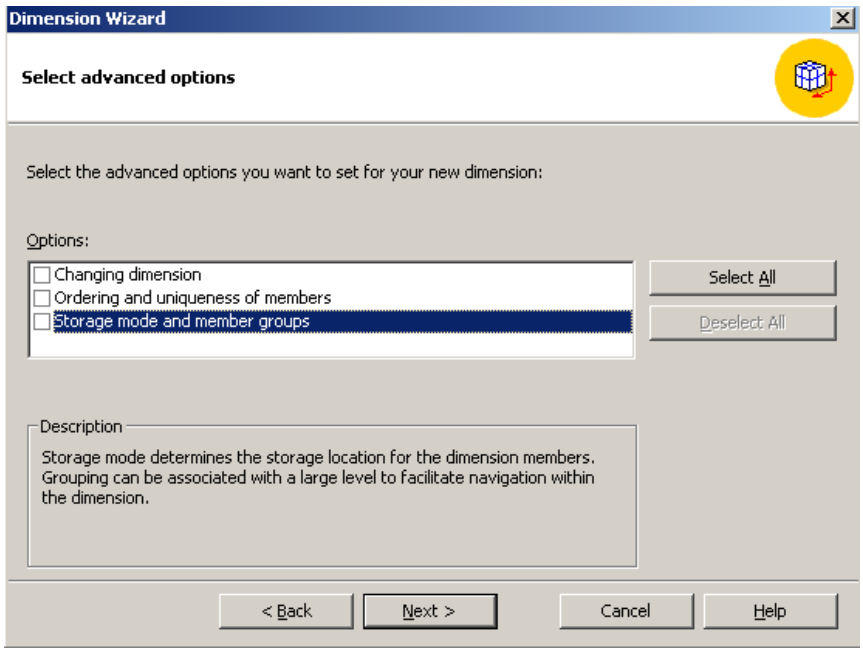
15. Now, select **level of your dimension**. Here in this example, we have selected BrandName, ProductId and ProductName as shown in following Figure. You can also select any other columns. Then, click **Next**.



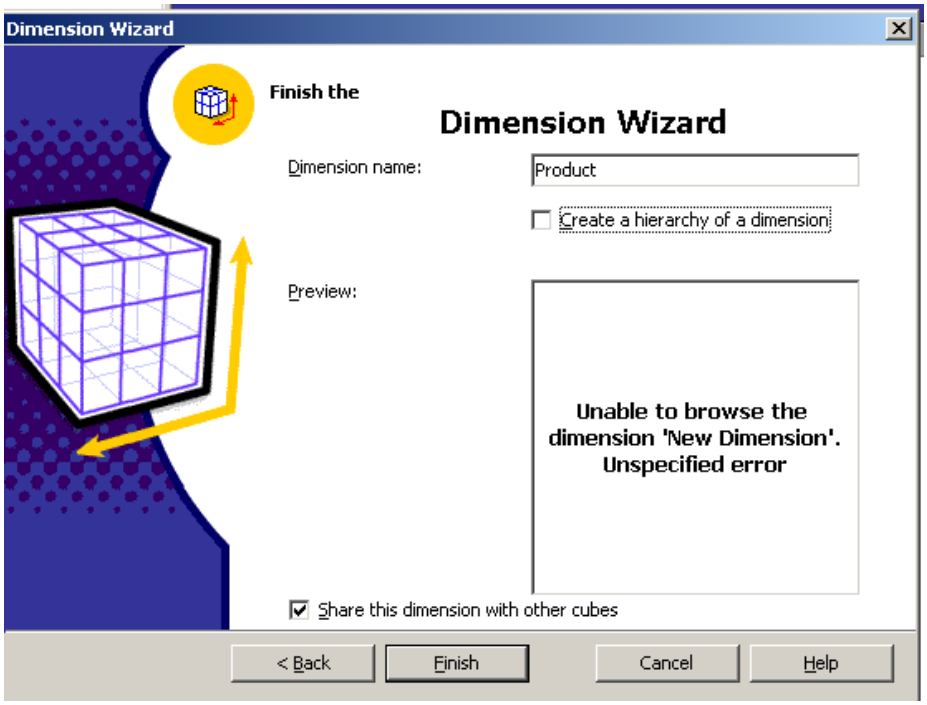
16. Now, select **member key columns** and Click **Next**.



17. Click **Next**.

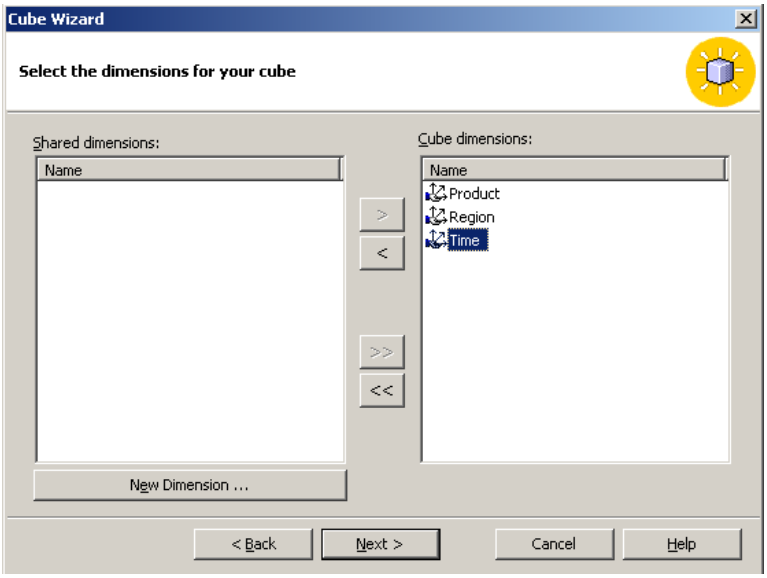


18. Click **Next**.

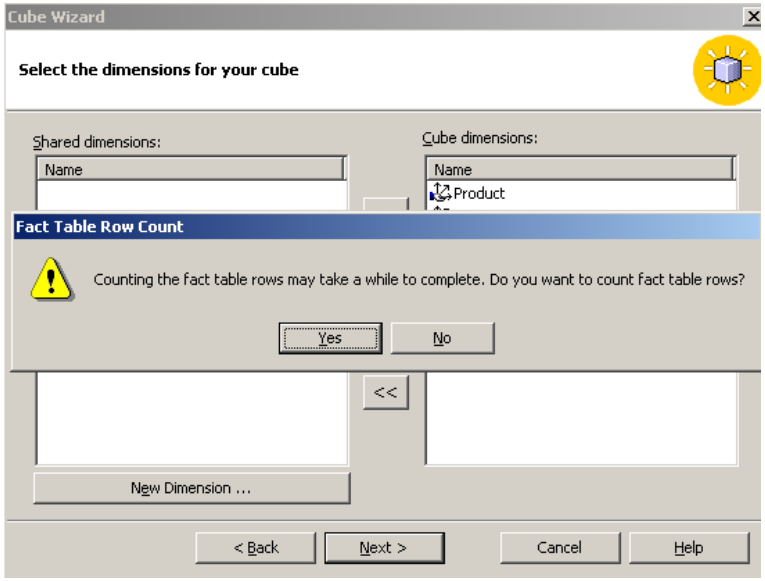


19. Type Product in **Dimension Name** and click **Finish**.

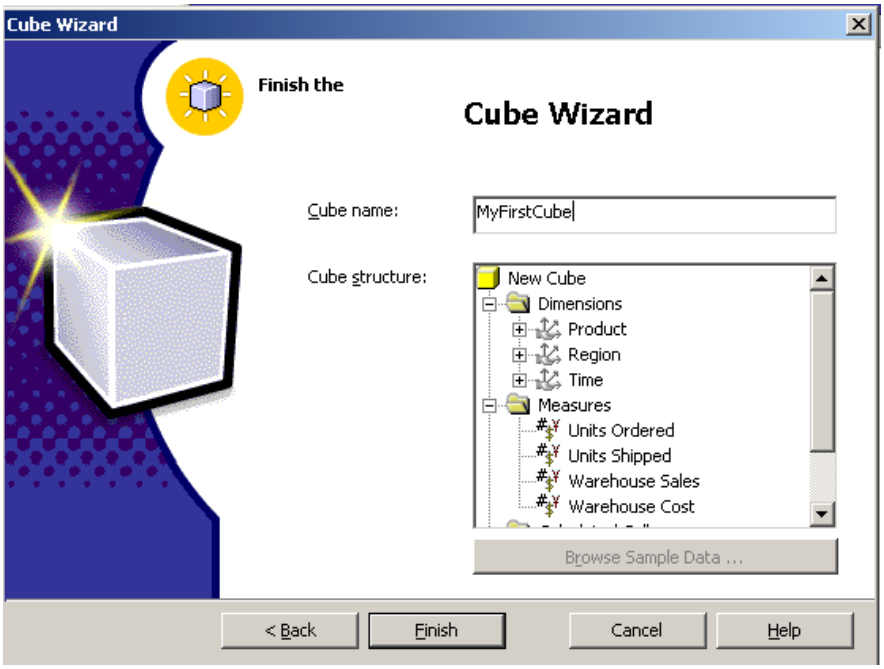
20. By following same steps, add two more dimensions Product, Region and time.



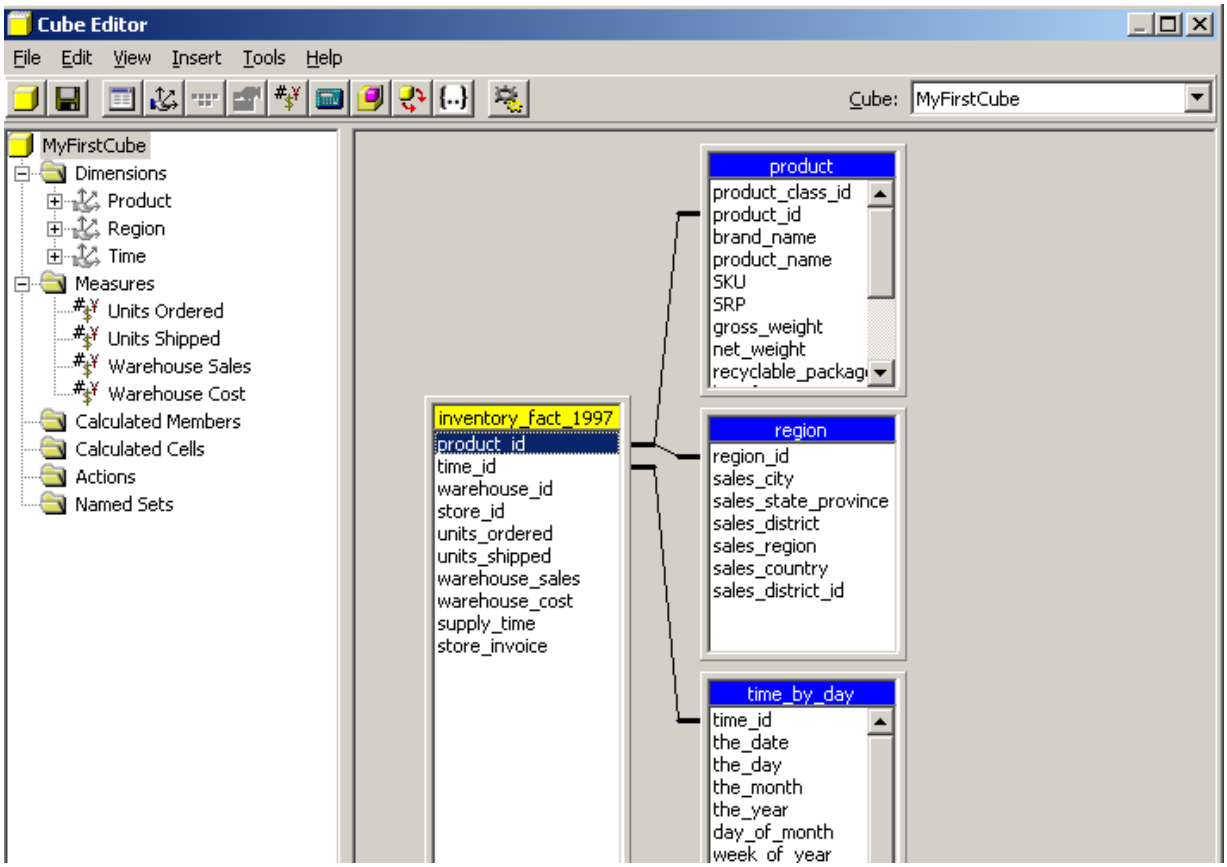
21. Click **Next**. Then, **Cube Wizard** window will open.



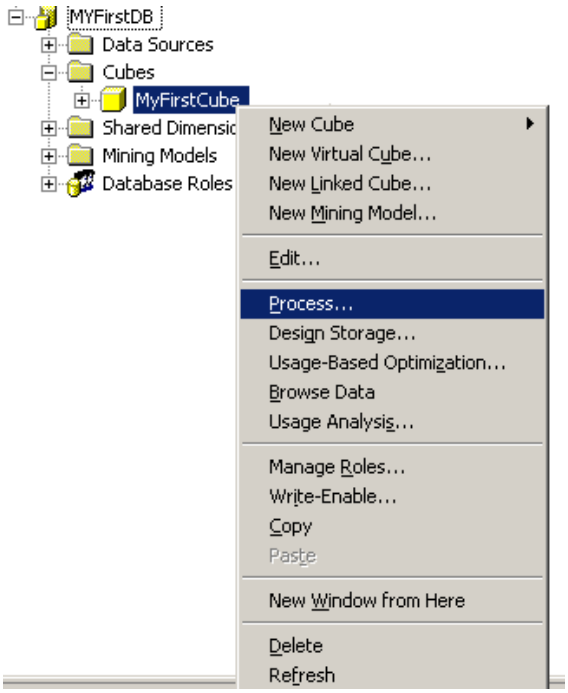
22. Click **Yes** in **Fact Table Row Count** window. Cube Wizard window will open. Type MyFirstCube in **Cube Name** and then Click **Finish**.



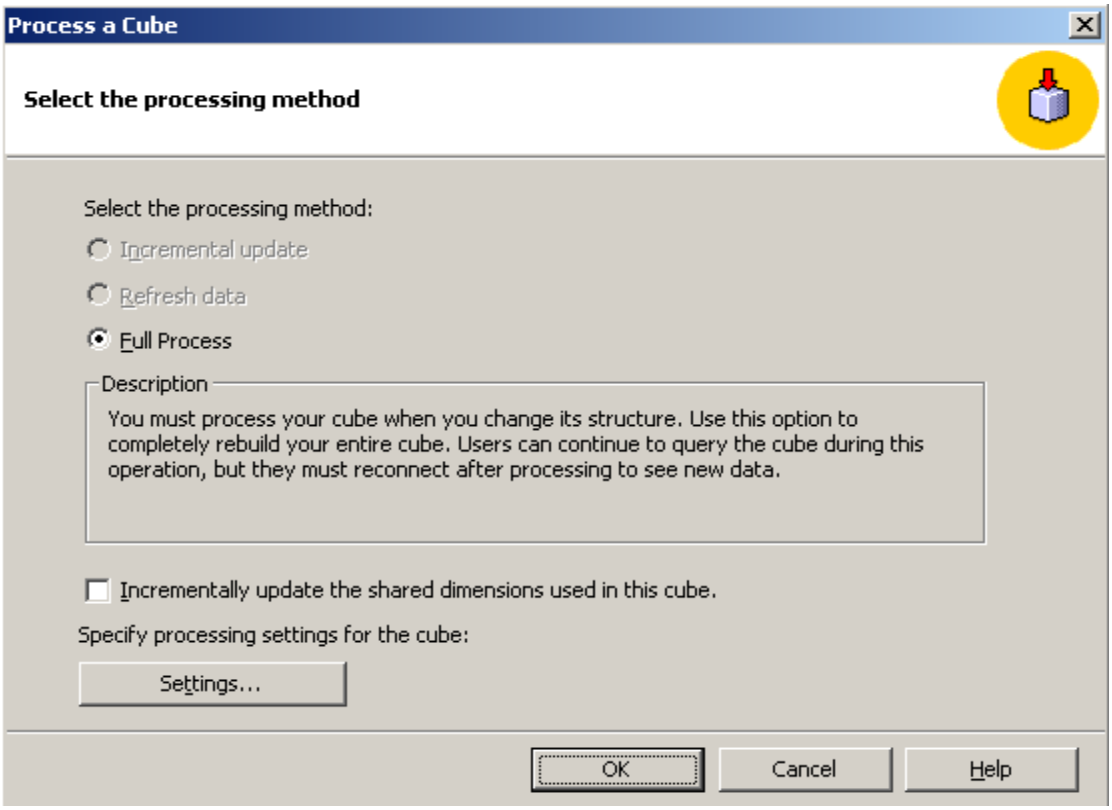
23. **Cube Edi**tor Window will open which shows Star Schema of your cube.



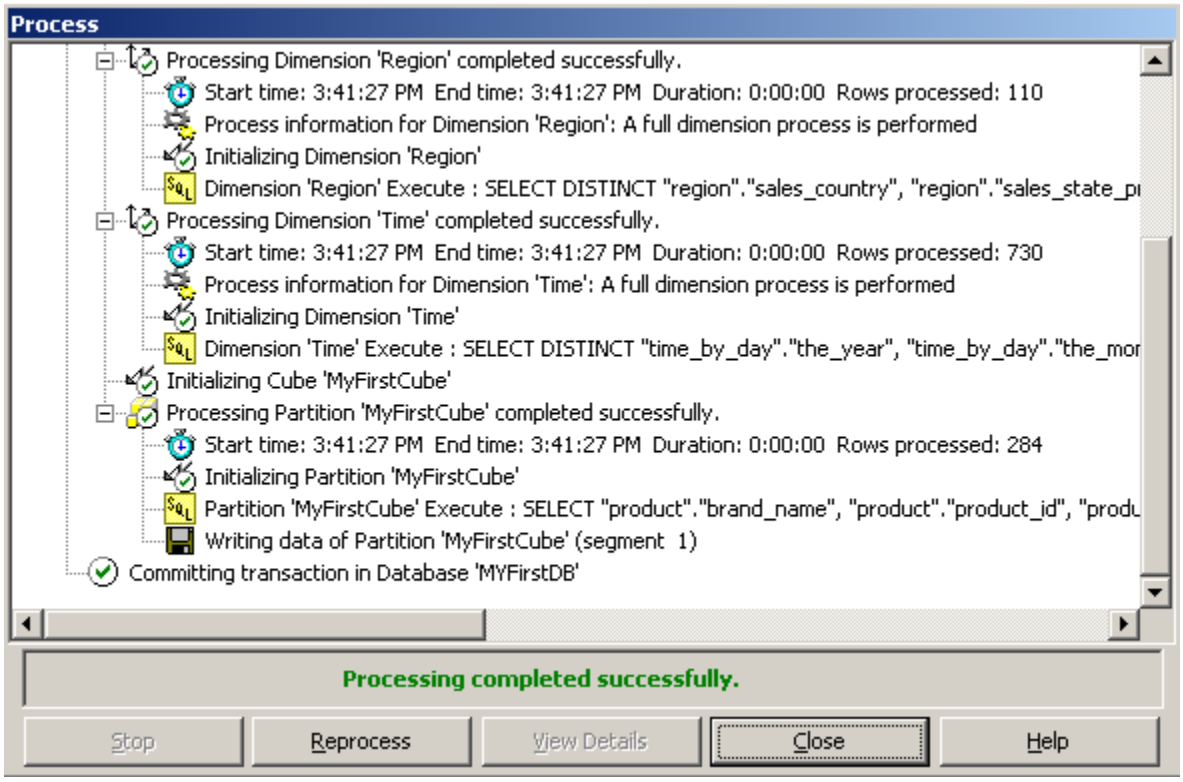
24. To process cube, click on cube name **MyFirstCube->Process.**



25. Select the processing method and click **Next**.



26. Cube processing completed.



Lab Exercise: Right Click on cube name and select Browse Data option to view cube data. Perform different cube operation on this data e.g. drill down, roll up etc.

Mechanism to Conduct Lab:

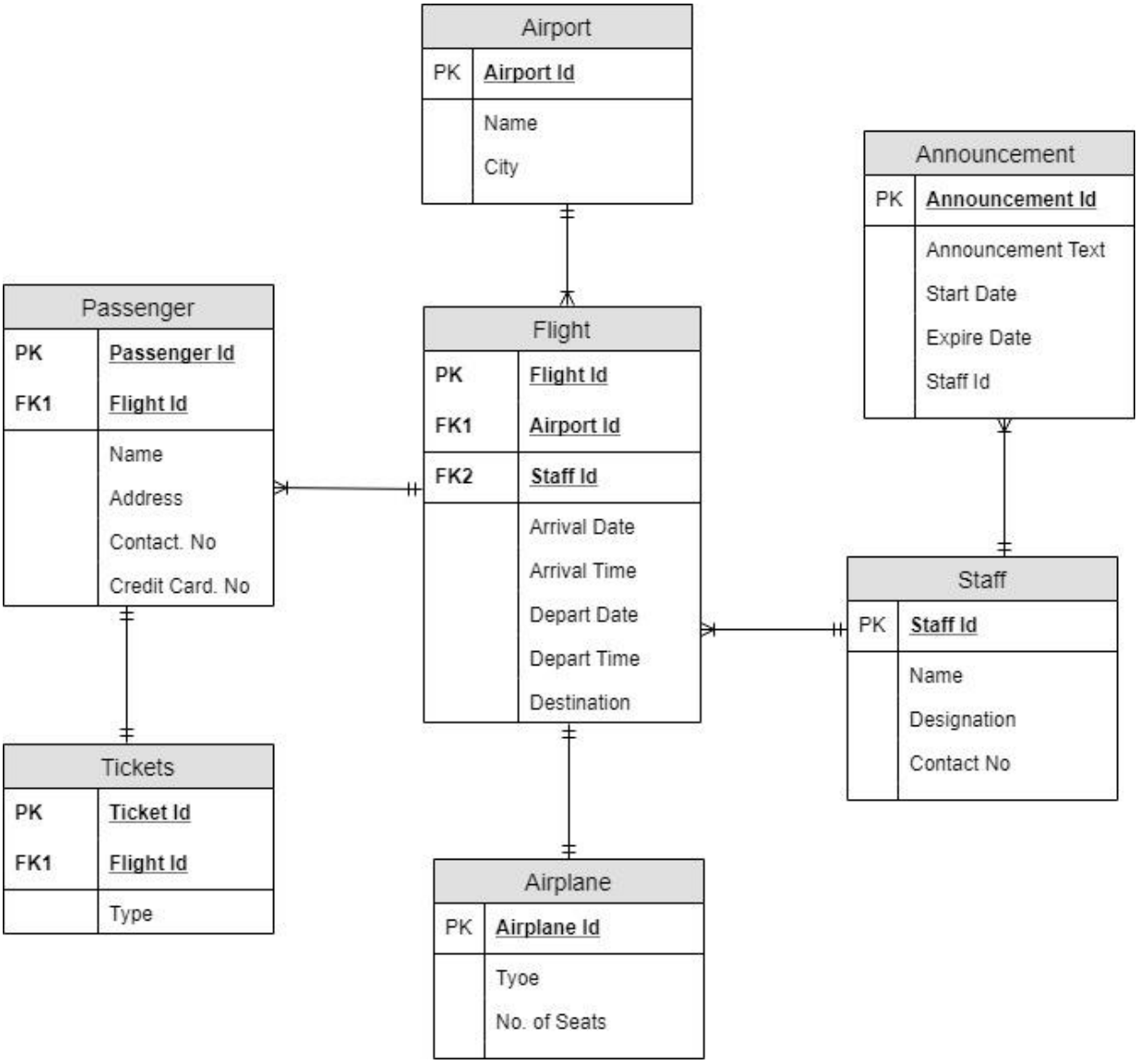
Students and teacher communicate through Adobe Connect.

Lab 5

STAR SCHEMA

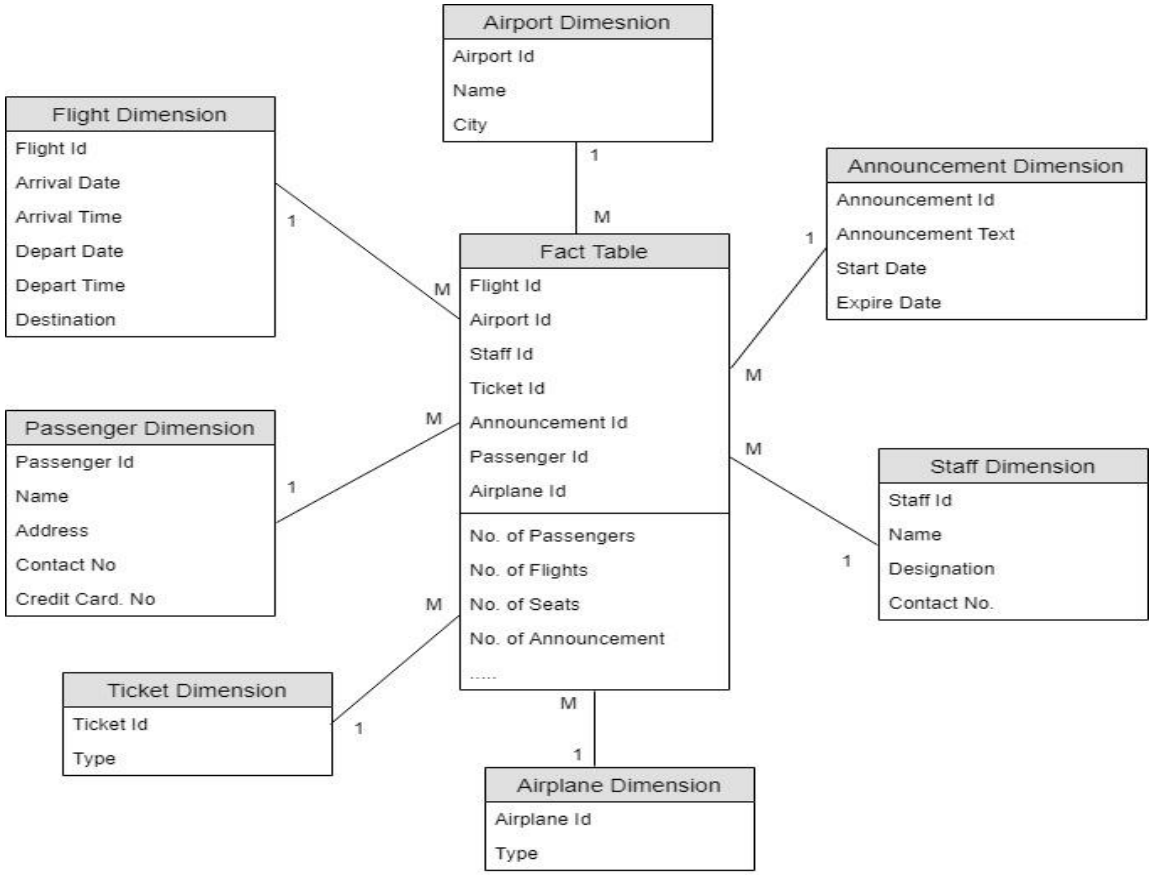
Question Statement:

Following is an **Entity Relationship Diagram (ERD)** of Airline Reservation System, you have to design **Star Schema** using all steps of dimension modelling in any drawing tool e.g. MS Visio etc.



Solution:

Following are steps of dimensional modeling.
Step-1: Choose the Business Process
The business process is “Air Ticket Reservation System”.
Step-2: Choose the Grain
Grain represents the atomic level of information required from business process and it is termed as unit of analyses. The grain statement is **“Total number of passengers arrives/depart in each flight by the airline.”**
Step-3: Choose the Facts
Facts are numeric, continuously valued and additive. The fact in our case is “**Total number of passengers in an airplane**”.
Step-4: Choose the Dimensions
The dimensions will be “Airport”, “Flight”, “Staff”, “Announcement”, “Passenger”, “Airplane”, and “Tickets”.
Star Schema:



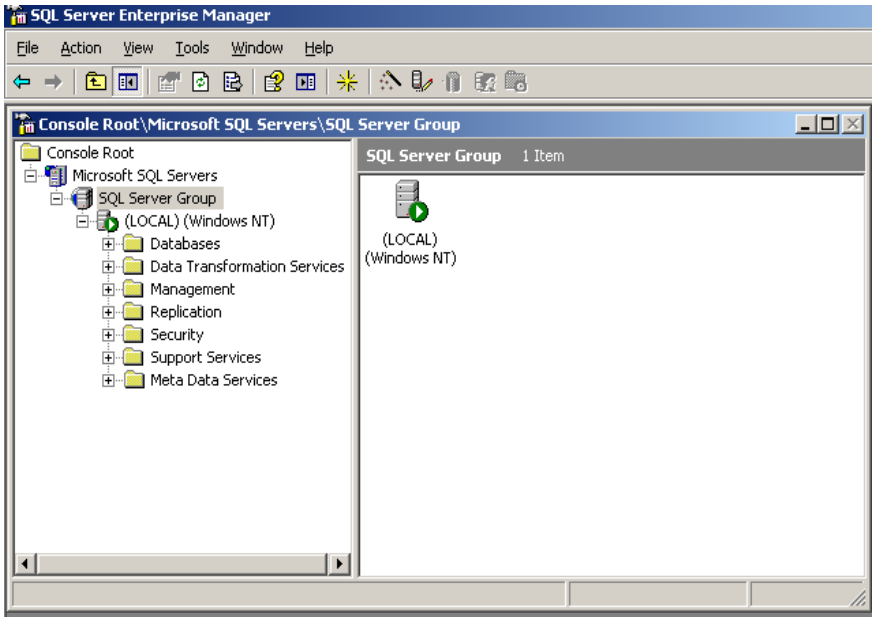
Mechanism to Conduct Lab:

Students and teacher communicate through Adobe Connect Session.

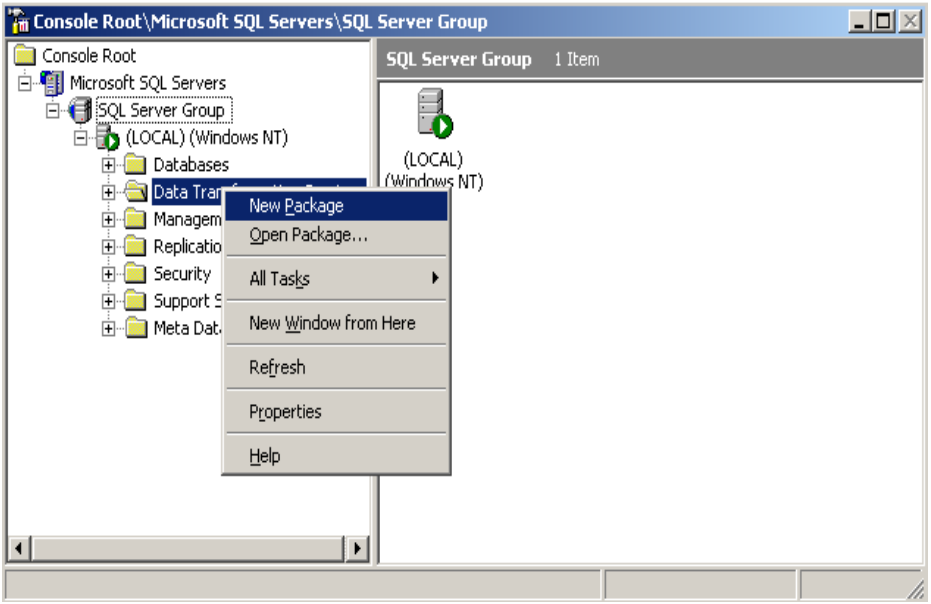
Lab 6

DATA EXTRACTION

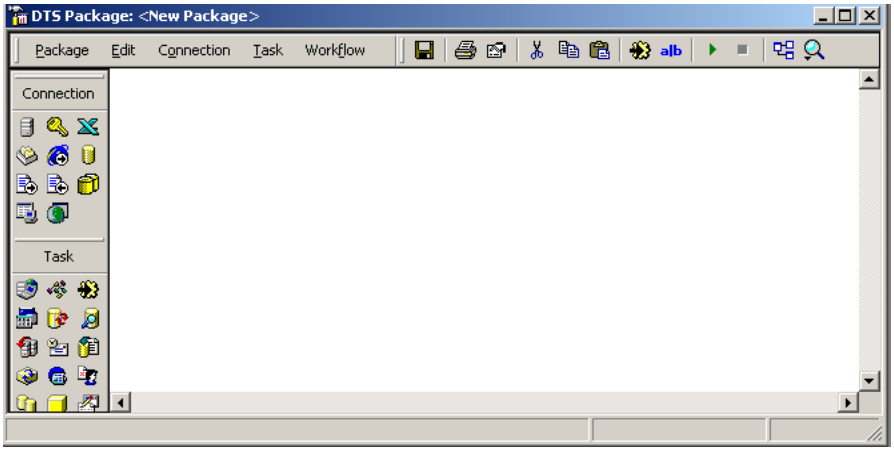
1. Open **SQL Server Enterprise Manager** and click on **SQL Server Group**.



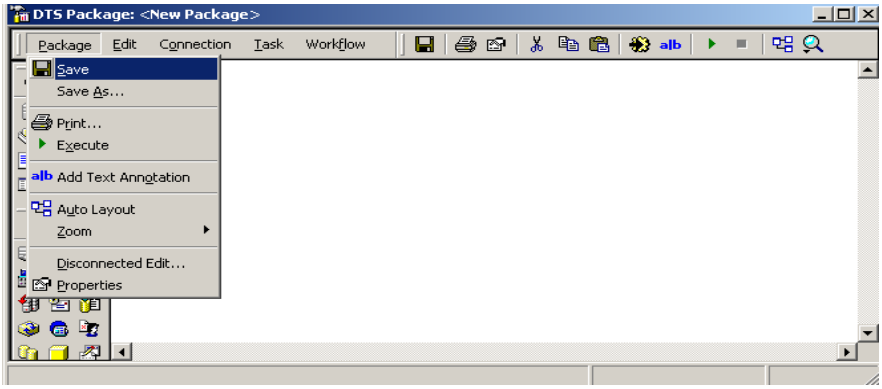
2. Right Click on **Data Transformation Services** node and click **New Package.**



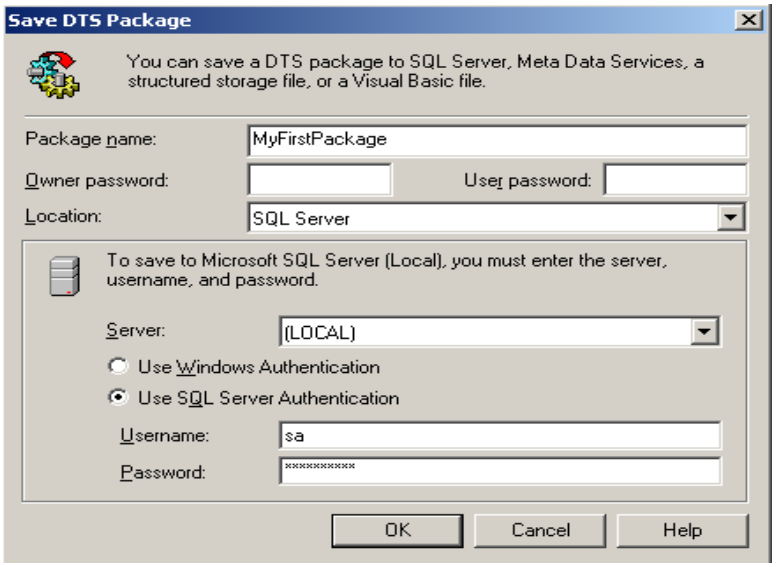
1. The following window will open.



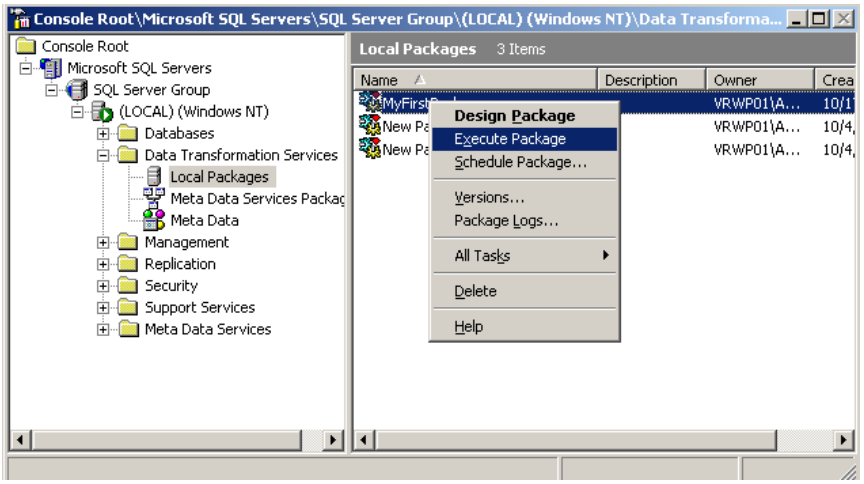
2. Click on **Package** menu and click **Save** to save the package.



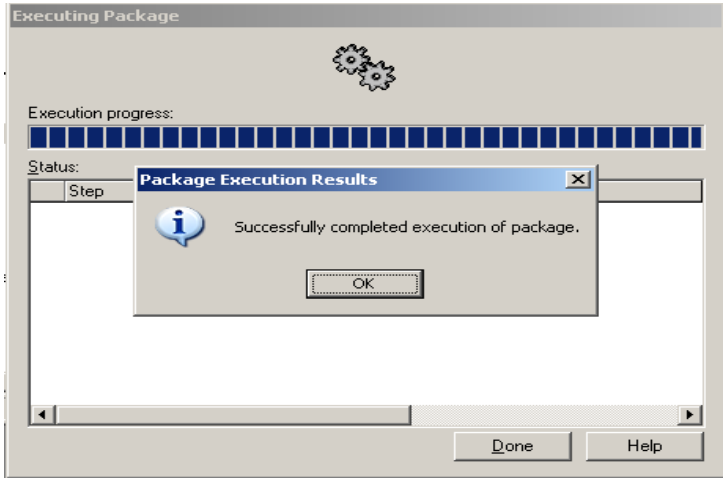
3. Enter **MyFirstPackage in** Package name and Click OK.



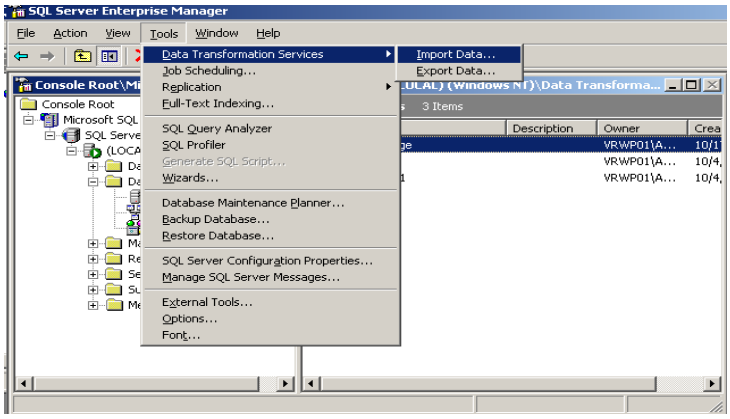
4. Right click on **MyFirstPackage** and select execute package option.



5. Click OK. Then Click Done.

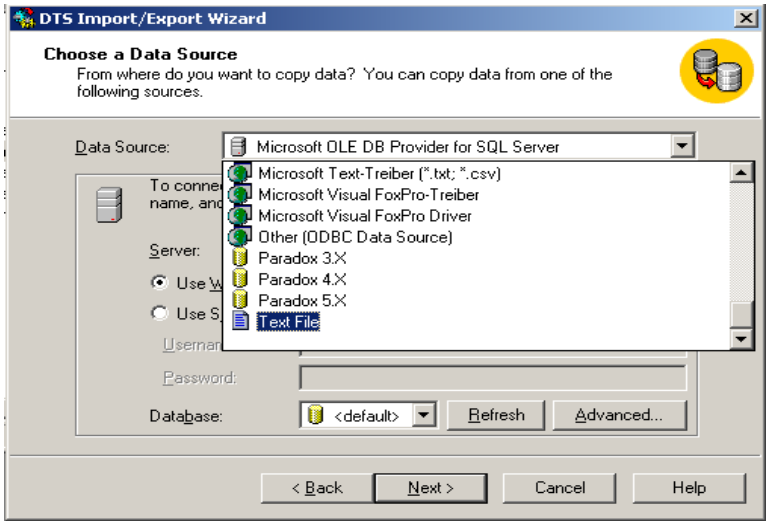


6. Click on **Tools Menu->Data Transformation Services->Import Data**.

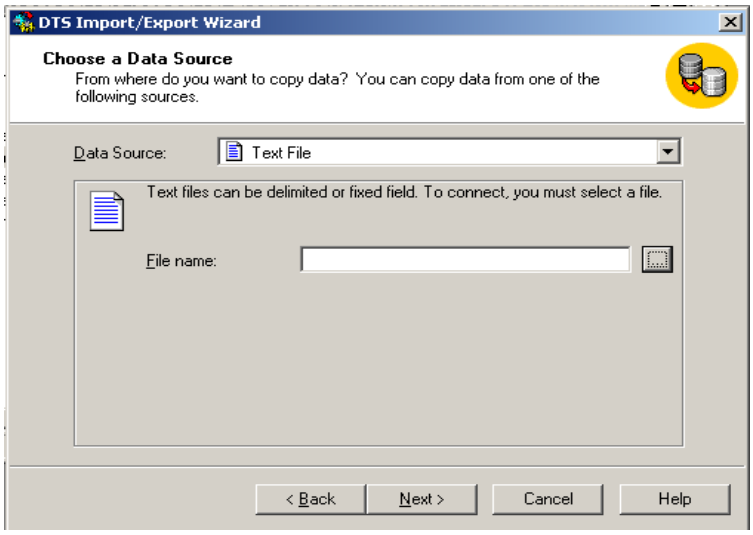


7. Click Next.

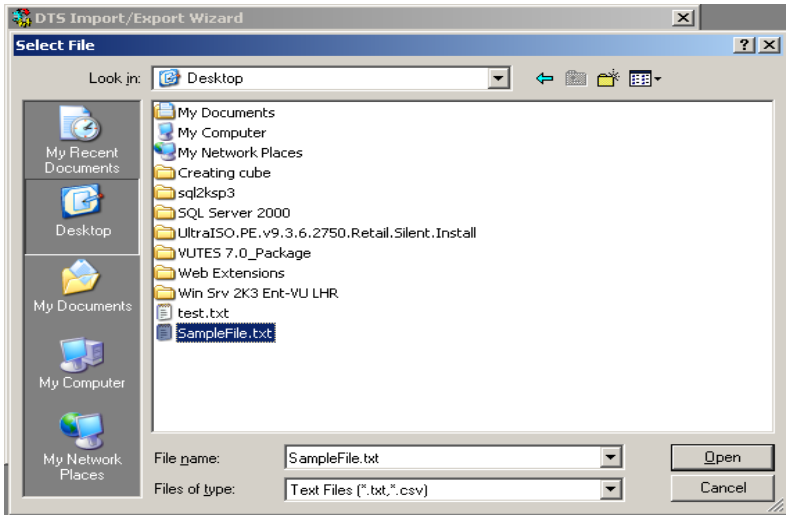
8. Select **Text File** as data source and click Next.



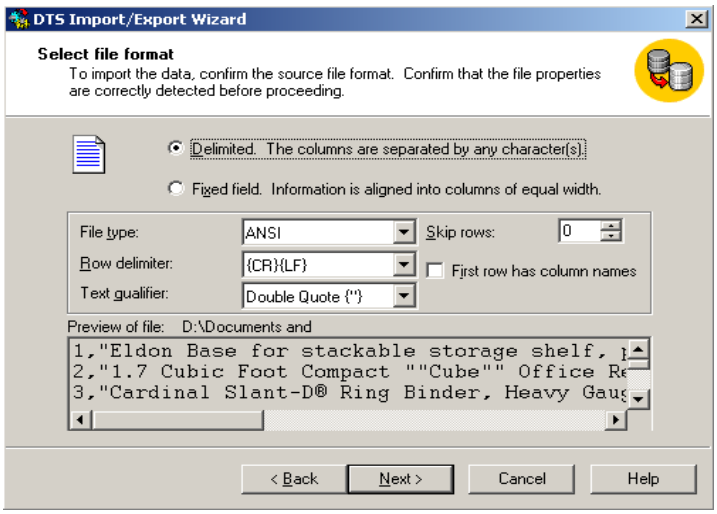
9. Select text file which you want to import and then click Open.



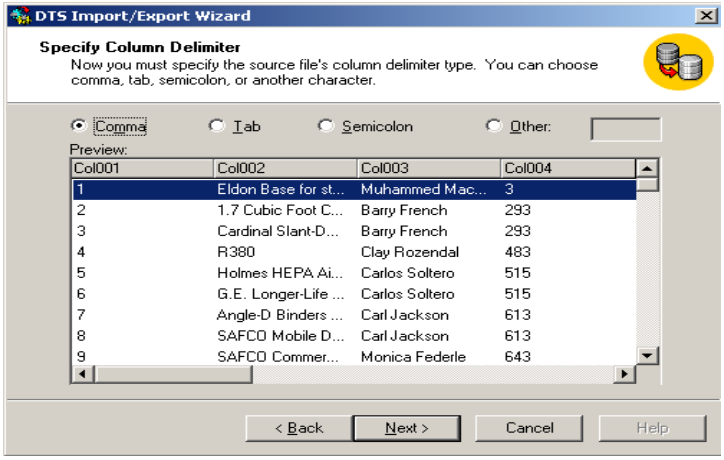
10. You can create a text on your system containing sample data. You can enter data about customers or students etc. Select the file and click Open. Then click **Next.**



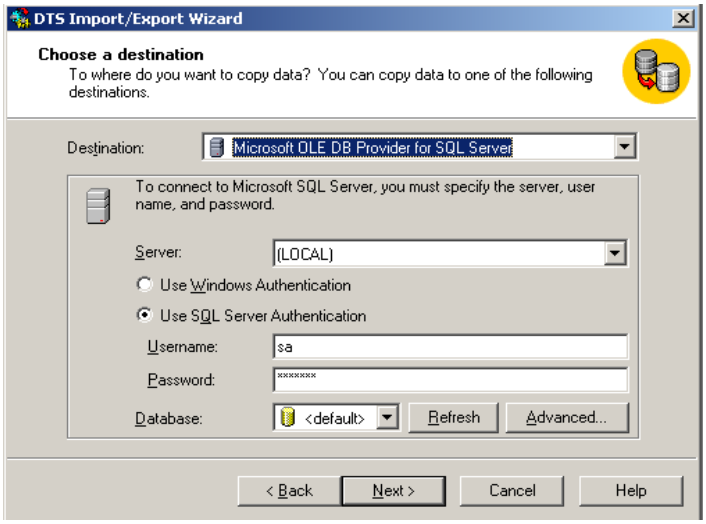
11. Click Next.



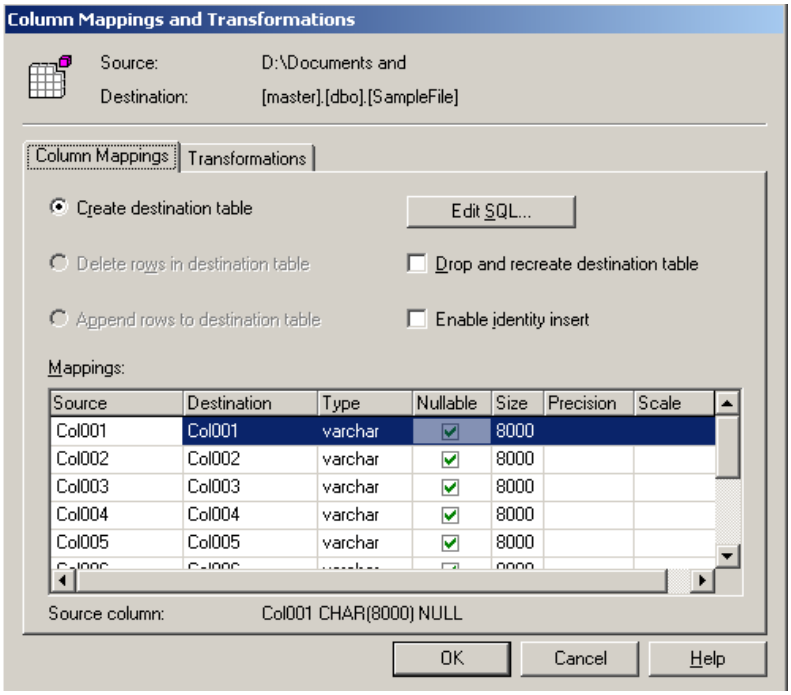
12. Click Next. Select Column Delimiter which in this case is comma. Click Next.

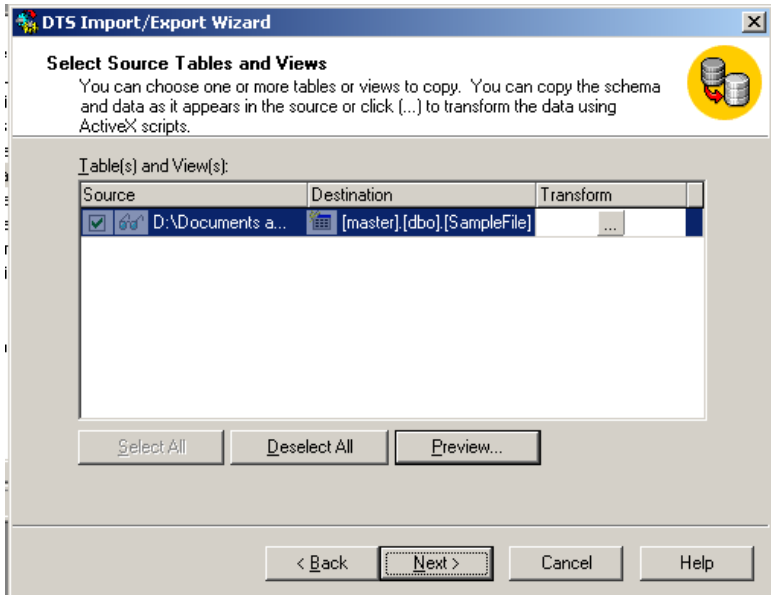


13. Choose destination and click Next.

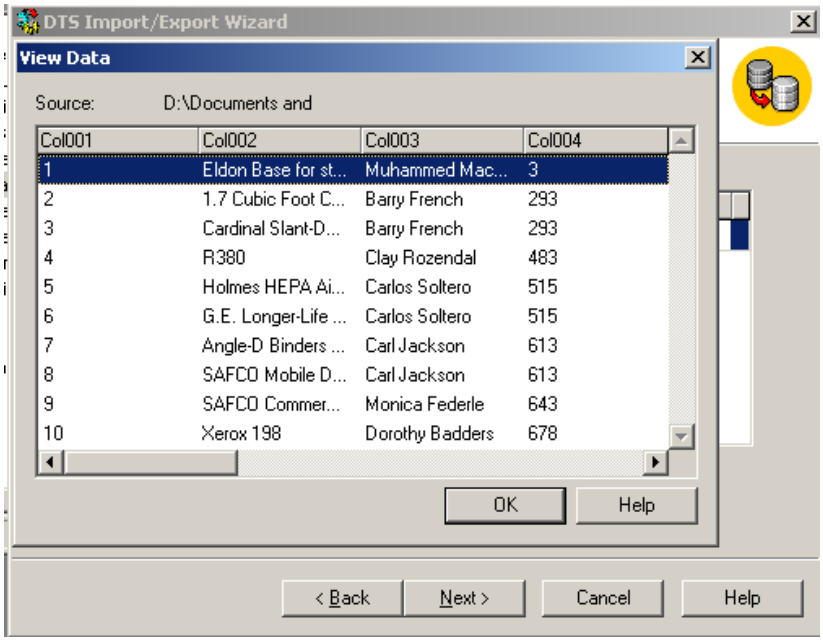


14. Select **Create destination table** option and create OK.

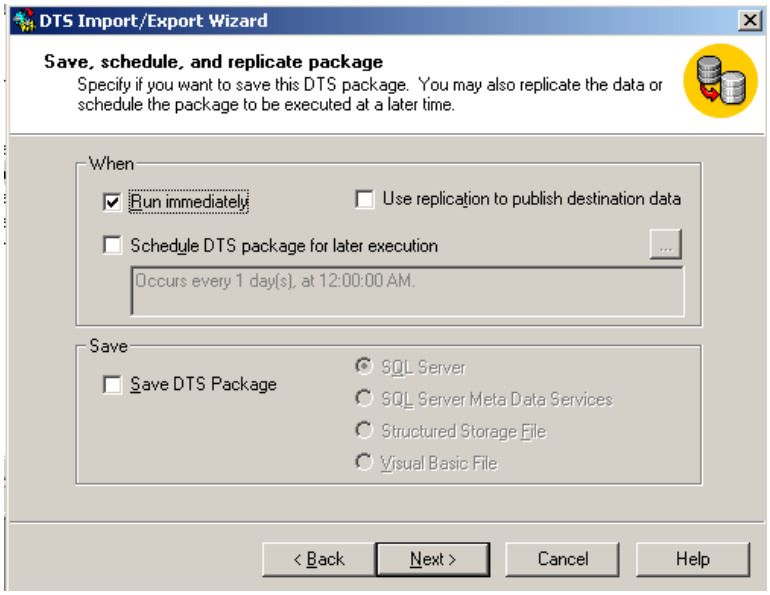




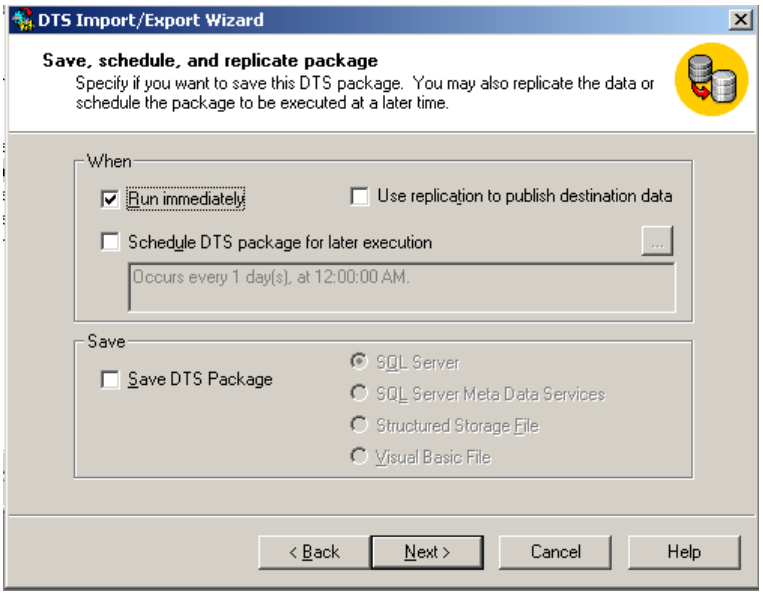
15. Click **Preview….** to view data.



16. Click Next.

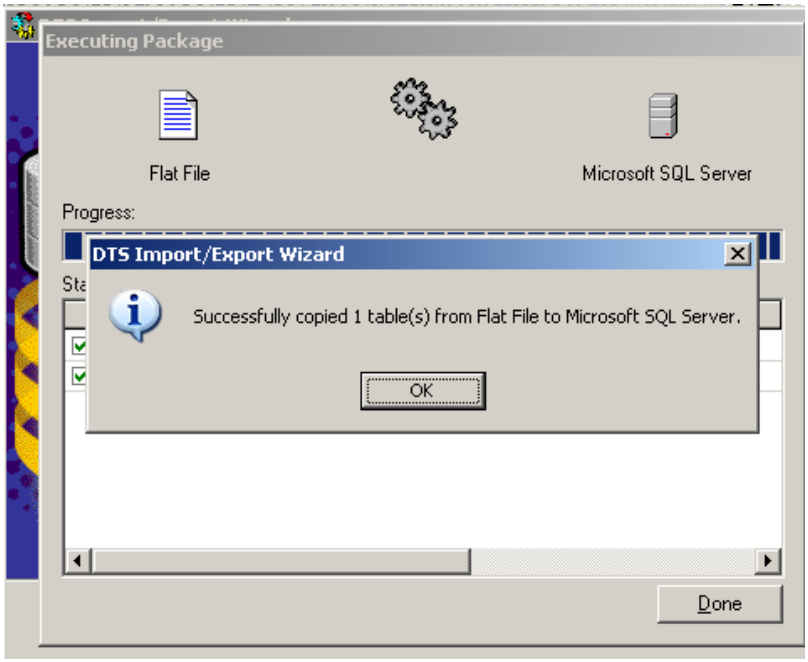


17. Select **Run immediately** and click Next.

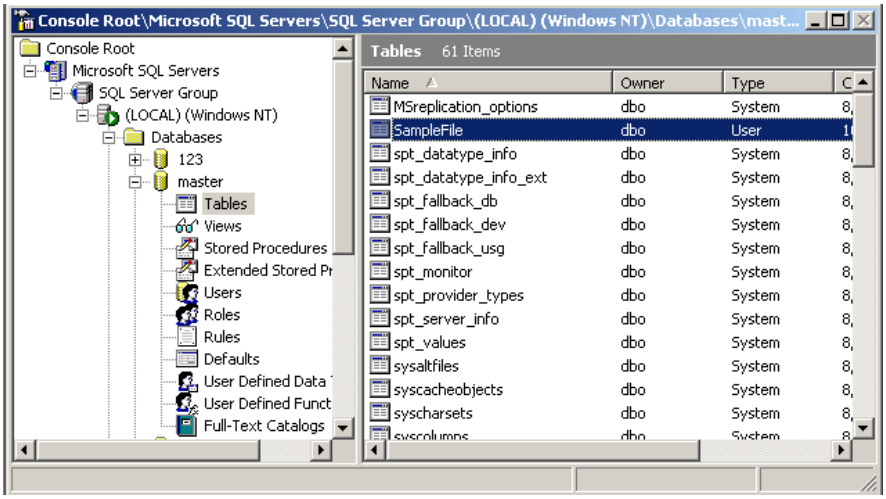


18. Click Finish.

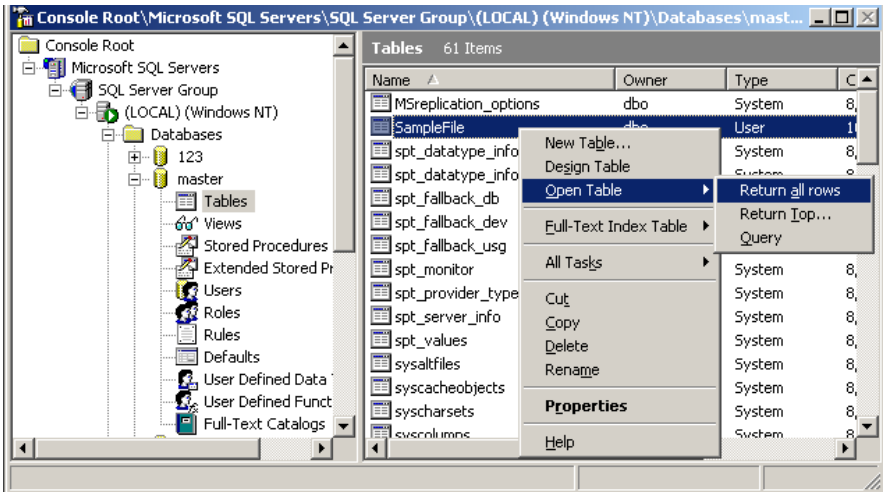
19. Click OK. Then Click Done.



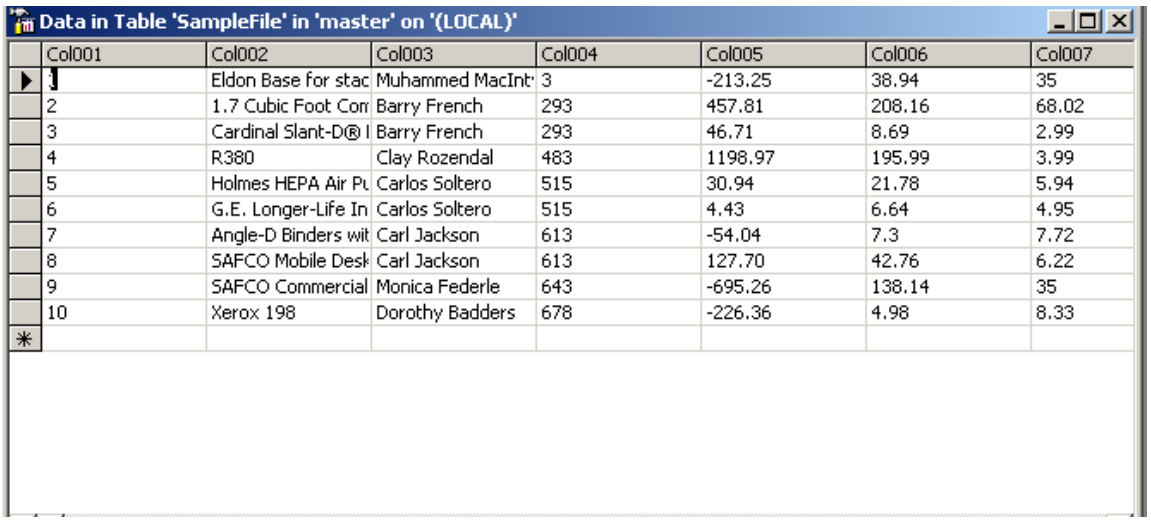
20. View your table in master table node.



21. To view data, right click on table SampleFile, click on Open Table->Return all rows.



22. This is data of table **SampleFile**.



	Col001	Col002	Col003	Col004	Col005	Col006	Col007
▶	1	Eldon Base for stac	Muhammed MacInt	3	-213.25	38.94	35
	2	1.7 Cubic Foot Com	Barry French	293	457.81	208.16	68.02
	3	Cardinal Slant-D® I	Barry French	293	46.71	8.69	2.99
	4	R380	Clay Rozendal	483	1198.97	195.99	3.99
	5	Holmes HEPA Air Pu	Carlos Soltero	515	30.94	21.78	5.94
	6	G.E. Longer-Life In	Carlos Soltero	515	4.43	6.64	4.95
	7	Angle-D Binders wit	Carl Jackson	613	-54.04	7.3	7.72
	8	SAFCO Mobile Desk	Carl Jackson	613	127.70	42.76	6.22
	9	SAFCO Commercial	Monica Federle	643	-695.26	138.14	35
	10	Xerox 198	Dorothy Badders	678	-226.36	4.98	8.33
*							

Mechanism to Conduct Lab:

Students and teacher communicate through Adobe Connect.

Lab 7

BASIC SORTED NEIGHBORHOOD (BSN) METHOD

PatientID	Patient Name	Patient Contact No	Address
F101	Sana	051-123456	Islamabad
M102	Ali	051-123456	Lahore
F103	Aliya		Rawalpindi
F104	Hina		Faisalabad
F103	Aliaa		Rawalpindi
F101	Sanah		Islamabad
M105	Hassan		Karachi

Question:

Consider the above table and apply all three steps (**Create key, Sort the data, Merge)** of Basic Sorted Neighborhood (BSN) method to find out the duplicate records in the table. Records will be considered duplicate if the value of column "PatientID" is same in this table data.

The steps which you have to follow are:

Key:

Key will consist of one character from "**PatientID**", then first three characters from **"Patient Name"** and then first two characters from **"Address"** column.

Step 1: Create key

In step-1, you will create the key according to the rules as mentioned above against each record. Add an extra column at the end of the table to show the new key created against each record.

Step 2: Sort the data

In step-2, you will sort the record on the basis of key which you created in step-1.

Step 3: Merge

In step-3, consider the window size (w) equal to two (2) and identify the similar records on the basis of sorted key.

Solution:

Step 1: Create key

Key is created in first step as per rules given in question statement.

PatientID	Patient Name	Patient Contact No	Address	Key
F01	Sana	051-123456	Islamabad	FSanIs
M01	Ali	051-123456	Lahore	MAliLa
F02	Aliya		Rawalpindi	FAliRa
F03	Hina		Faisalabad	FHinFa
F02	Aliaa		Rawalpindi	FAliRa
F01	Sanah		Islamabad	FSanIs
M02	Hassan		Karachi	MHasKa

Step 2: Sort the data

Now sort the records based on the above created keys

PatientID	Patient Name	Patient Contact No	Address	Key
F02	Aaliya		Rawalpindi	FAalRa
F02	Aaliaa		Rawalpindi	FAalRa
F03	Hina		Faisalabad	FHinFa
F01	Sana	051-123456	Islamabad	FSanIs
F01	Sanah		Islamabad	FSanIs
M01	Ali	051-123456	Lahore	MAliLa
M02	Hassan		Karachi	MHasKa

Step 3: Merge

Duplicate/Identical keys

PatientID	Patient Name	Patient Contact No	Address	Key
F02	Aaliya		Rawalpindi	FAalRa
F02	Aaliaa		Rawalpindi	FAalRa

F01	Sana	051-123456	Islamabad	FSanIs
F01	Sanah		Islamabad	FSanIs

Mechanism to Conduct Lab:

Students and teacher communicate through Adobe Connect.

Lab 8

DATA QUALITY RULES

"Pak Airline" is running in more than 15 countries and maintaining its data in a database. They have also started using data warehouse. They have different policies for running their business. One of their policies is that, if the airline itself cancels the flight due to weather or any other technical reason, then they have to payback 100% amount to customers and if flight is canceled by passenger then 40% amount will be returned. The Payback codes 'FP' or 'PP' are put in the payment claims to identify whether the claim is for full or partial payment respectively. Another policy is that allowed luggage weight per passenger is 25 kg.

There exists some specific data problems which are linked to business rules, and then generic and specific rule sets are developed for measuring how good the data is within an information system.
These rule sets shows the data quality metrics in order to judge conformance of data according to these business rules.

Note: Here, Assume, amount of ticket is Rs. 15,000 per passenger.

Question:

Considering above scenario, apply data quality rules on the historical data problems faced by "Pak Airline". For this, fill the following table (Rule Type, Generic Rule Set and Specific Rule Set) against each historical data problems as given below.

Historical Data Problem	Rule Type	Generic Rule Set	Specific Rule Set
Payback amount is less than 100% when flight is canceled due to weather or other technical reasons.			
The field "reason_cancel" were often left blank.			
The luggage per passenger exceeds 25 kg.			
The payback code in payback claim is not **'FP'** or **'PP'** sometimes.			

Solution:

Historical Data Problem	Rule Type	Generic Rule Set	Specific Rule Set
Payback amount is less than 100% when flight is canceled due to weather or other technical reasons.	Business Rule	If total payback amount is less than Rs. 15,000 and reason of flight cancelation is weather or technical, then, error.	Select total_payback_amount from Claim where total_payback_amount <15000 and reason_cancel='weather' or reason_cancel='technical';
The field "reason_cancel" were often left blank.	Null Constraints	If reason_cancel is blank or null, then, error	Select reason_cancel from Claim where reason_cancel= ' ' or reason_cancel =NULL;
The luggage per passenger exceeds 25 kg.	Operational Rule	If luggage_weight is greater than 25kg, then, error.	Select luggage_weight from Passenger where luggage_weight>25; Note: Assume that weight in field luggage_weight is input as kg.
The payback code in payback claim is not **'FP'** or **'PP'** sometimes.	Domain Validation	If payback_code is not 'FP' or 'PP', then, error.	Select payback_code from Claim where NOT (payback_code= 'FP' or 'PP');

Mechanism to Conduct Lab:

Students and teacher communicate through Adobe Connect.

Lab 9

KEY RANGE PARTITIONING

Order ID	Order Date	Ship Date	Customer ID	Customer Name	State	Region	Product ID
CA-2016-152156	11/8/2016	11/11/2016	CG-12520	Claire Gute	Kentucky	South	FUR-BO-10001798
CA-2016-152156	11/8/2016	11/11/2016	CG-12520	Claire Gute	Kentucky	South	FUR-CH-10000454
CA-2016-138688	6/12/2016	6/16/2016	DV-13045	Darrin Van Huff	California	West	OFF-LA-10000240
US-2015-108966	10/11/2015	10/18/2015	SO-20335	Sean O'Donnell	Florida	South	FUR-TA-10000577
US-2015-108966	10/11/2015	10/18/2015	SO-20335	Sean O'Donnell	Florida	South	OFF-ST-10000760
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	FUR-FU-10001487
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	OFF-AR-10002833
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	TEC-PH-10002275
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	OFF-BI-10003910
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	OFF-AP-10002892
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	FUR-TA-10001539
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	TEC-PH-10002033
CA-2017-114412	4/15/2017	4/20/2017	AA-10480	Andrew Allen	North Carolina	South	OFF-PA-10002365
CA-2016-161389	12/5/2016	12/10/2016	IM-15070	Irene Maddox	Washington	West	OFF-BI-10003656
US-2015-118983	11/22/2015	11/26/2015	HP-14815	Harold Pawlan	Texas	Central	OFF-AP-10002311
US-2015-118983	11/22/2015	11/26/2015	HP-14815	Harold Pawlan	Texas	Central	OFF-BI-10000756
CA-2014-105893	11/11/2014	11/18/2014	PK-19075	Pete Kriz	Wisconsin	Central	OFF-ST-10004186
CA-2014-167164	5/13/2014	5/15/2014	AG-10270	Alejandro Grove	Utah	West	OFF-ST-10000107
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	West	OFF-AR-10003056
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	West	TEC-PH-10001949
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	West	OFF-BI-10002215
CA-2016-137330	12/9/2016	12/13/2016	KB-16585	Ken Black	Nebraska	Central	OFF-AR-10000246
CA-2016-137330	12/9/2016	12/13/2016	KB-16585	Ken Black	Nebraska	Central	OFF-AP-10001492
US-2017-156909	7/16/2017	7/18/2017	SF-20065	Sandra Flanagan	Pennsylvania	East	FUR-CH-10002774

PART I

Consider the subset of dataset data taken from https://community.tableau.com/docs/DOC-1236.

You are required to perform range partitioning on this data into yearly partitions of ship date.

Solution:

Partition 1 of year 2014:

Order ID	Order Date	Ship Date	Customer ID	Customer Name	State	Region	Product ID
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	FUR-FU-10001487
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	OFF-AR-10002833
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	TEC-PH-10002275
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	OFF-BI-10003910
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	OFF-AP-10002892
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	FUR-TA-10001539
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	West	TEC-PH-10002033
CA-2014-105893	11/11/2014	11/18/2014	PK-19075	Pete Kriz	Wisconsin	Central	OFF-ST-10004186
CA-2014-167164	5/13/2014	5/15/2014	AG-10270	Alejandro Grove	Utah	West	OFF-ST-10000107
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	West	OFF-AR-10003056
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	West	TEC-PH-10001949
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	West	OFF-BI-10002215

Partition 2 of year 2015:

Order ID	Order Date	Ship Date	Customer ID	Customer Name	State	Region	Product ID
US-2015-108966	10/11/2015	10/18/2015	SO-20335	Sean O'Donnell	Florida	South	FUR-TA-10000577
US-2015-108966	10/11/2015	10/18/2015	SO-20335	Sean O'Donnell	Florida	South	OFF-ST-10000760
US-2015-118983	11/22/2015	11/26/2015	HP-14815	Harold Pawlan	Texas	Central	OFF-AP-10002311
US-2015-118983	11/22/2015	11/26/2015	HP-14815	Harold Pawlan	Texas	Central	OFF-BI-10000756

Partition 3 of year 2016:

Order ID	Order Date	Ship Date	Customer ID	Customer Name	State	Region	Product ID
CA-2016-152156	11/8/2016	11/11/2016	CG-12520	Claire Gute	Kentucky	South	FUR-BO-10001798
CA-2016-152156	11/8/2016	11/11/2016	CG-12520	Claire Gute	Kentucky	South	FUR-CH-10000454
CA-2016-138688	6/12/2016	6/16/2016	DV-13045	Darrin Van Huff	California	West	OFF-LA-10000240
CA-2016-161389	12/5/2016	12/10/2016	IM-15070	Irene Maddox	Washington	West	OFF-BI-10003656
CA-2016-137330	12/9/2016	12/13/2016	KB-16585	Ken Black	Nebraska	Central	OFF-AR-10000246
CA-2016-137330	12/9/2016	12/13/2016	KB-16585	Ken Black	Nebraska	Central	OFF-AP-10001492

Order ID	Order Date	Ship Date	Customer ID	Customer Name	State	Region	Product ID
CA-2017-114412	4/15/2017	4/20/2017	AA-10480	Andrew Allen	North Carolina	South	OFF-PA-10002365
US-2017-156909	7/16/2017	7/18/2017	SF-20065	Sandra Flanagan	Pennsylvania	East	FUR-CH-10002774

PART II

Perform list portioning by Partitioning by following sales table by region. You can use values of region from table given in **Question 1**.

Solution:

First you will define values for region partitions.

Region partitions will be **East**, **West**, **South** and **Central Partition**.

List of East partition values: Pennsylvania

List of West Partition Values: Utah, California, Washington

List of South Partition Values: Florida, Kentucky, North Carolina

List of Central Partition Values: Wisconsin, Texas, Nebraska

South Partition:

Order ID	Order Date	Ship Date	Customer ID	Customer Name	State	Product ID
CA-2016-152156	11/8/2016	11/11/2016	CG-12520	Claire Gute	Kentucky	FUR-BO-10001798
CA-2016-152156	11/8/2016	11/11/2016	CG-12520	Claire Gute	Kentucky	FUR-CH-10000454
US-2015-108966	10/11/2015	10/18/2015	SO-20335	Sean O'Donnell	Florida	FUR-TA-10000577
US-2015-108966	10/11/2015	10/18/2015	SO-20335	Sean O'Donnell	Florida	OFF-ST-10000760
CA-2017-114412	4/15/2017	4/20/2017	AA-10480	Andrew Allen	North Carolina	OFF-PA-10002365

Central Partition:

Order ID	Order Date	Ship Date	Customer ID	Customer Name	State	Product ID
CA-2016-152156	11/8/2016	11/11/2016	CG-12520	Claire Gute	Kentucky	FUR-BO-10001798
CA-2016-152156	11/8/2016	11/11/2016	CG-12520	Claire Gute	Kentucky	FUR-CH-10000454
CA-2016-138688	6/12/2016	6/16/2016	DV-13045	Darrin Van Huff	California	OFF-LA-10000240
US-2015-108966	10/11/2015	10/18/2015	SO-20335	Sean O'Donnell	Florida	FUR-TA-10000577
US-2015-108966	10/11/2015	10/18/2015	SO-20335	Sean O'Donnell	Florida	OFF-ST-10000760
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	FUR-FU-10001487
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	OFF-AR-10002833
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	TEC-PH-10002275
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	OFF-BI-10003910
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	OFF-AP-10002892
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	FUR-TA-10001539
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	TEC-PH-10002033
CA-2017-114412	4/15/2017	4/20/2017	AA-10480	Andrew Allen	North Carolina	OFF-PA-10002365
CA-2016-161389	12/5/2016	12/10/2016	IM-15070	Irene Maddox	Washington	OFF-BI-10003656
US-2015-118983	11/22/2015	11/26/2015	HP-14815	Harold Pawlan	Texas	OFF-AP-10002311

US-2015-118983	11/22/2015	11/26/2015	HP-14815	Harold Pawlan	Texas	OFF-BI-10000756
CA-2014-105893	11/11/2014	11/18/2014	PK-19075	Pete Kriz	Wisconsin	OFF-ST-10004186
CA-2014-167164	5/13/2014	5/15/2014	AG-10270	Alejandro Grove	Utah	OFF-ST-10000107
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	OFF-AR-10003056
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	TEC-PH-10001949
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	OFF-BI-10002215
CA-2016-137330	12/9/2016	12/13/2016	KB-16585	Ken Black	Nebraska	OFF-AR-10000246
CA-2016-137330	12/9/2016	12/13/2016	KB-16585	Ken Black	Nebraska	OFF-AP-10001492
US-2017-156909	7/16/2017	7/18/2017	SF-20065	Sandra Flanagan	Pennsylvania	FUR-CH-10002774

Order ID	Order Date	Ship Date	Customer ID	Customer Name	State	Region	Product ID
US-2017-156909	7/16/2017	7/18/2017	SF-20065	Sandra Flanagan	Pennsylvania	East	FUR-CH-10002774

West Partition:

Order ID	Order Date	Ship Date	Customer ID	Customer Name	State	Product ID
CA-2016-138688	6/12/2016	6/16/2016	DV-13045	Darrin Van Huff	California	OFF-LA-10000240

CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	FUR-FU-10001487
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	OFF-AR-10002833
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	TEC-PH-10002275
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	OFF-BI-10003910
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	OFF-AP-10002892
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	FUR-TA-10001539
CA-2014-115812	6/9/2014	6/14/2014	BH-11710	Brosina Hoffman	California	TEC-PH-10002033
CA-2016-161389	12/5/2016	12/10/2016	IM-15070	Irene Maddox	Washington	OFF-BI-10003656
CA-2014-167164	5/13/2014	5/15/2014	AG-10270	Alejandro Grove	Utah	OFF-ST-10000107
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	OFF-AR-10003056
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	TEC-PH-10001949
CA-2014-143336	8/27/2014	9/1/2014	ZD-21925	Zuschuss Donatelli	California	OFF-BI-10002215

East Partition:

Order ID	Order Date	Ship Date	Customer ID	Customer Name	State	Product ID
US-2015-118983	11/22/2015	11/26/2015	HP-14815	Harold Pawlan	Texas	OFF-AP-10002311
US-2015-118983	11/22/2015	11/26/2015	HP-14815	Harold Pawlan	Texas	OFF-BI-10000756
CA-2014-105893	11/11/2014	11/18/2014	PK-19075	Pete Kriz	Wisconsin	OFF-ST-10004186
CA-2016-137330	12/9/2016	12/13/2016	KB-16585	Ken Black	Nebraska	OFF-AR-10000246
CA-2016-137330	12/9/2016	12/13/2016	KB-16585	Ken Black	Nebraska	OFF-AP-10001492

Mechanism to Conduct Lab:

Students and teacher communicate through Skype/Adobe Connect. Students perform the task using the following simulator:

Lab 10

CLUSTER INDEX

"Pak Airline" is an airliner reservation company, which is operating in more than 10 countries. They have developed the airline reservation system to avoid the errors faced in manual system. The staff of the airline use airline reservation system form the tasks such as flight scheduling, ticket reservation, announcements in automated way. Similarly, users/passengers can search for flight schedule according to date and time and fare details. The staff of the airline can manage the reservation systems by flight rout, runway details, flight scheduling and reservation.

Ticket reservation system of the Pak Airline provides the information about schedule of flights, availability of seats, flight number and destination. For reservation of ticket user have to provide its personal information such as name, age, address etc. For payment purpose user will provide credit card number and bank details. Moreover, information about flight number, date of departure, no. of tickets to be booked is also required for confirmation of ticket.

Question Statement:

You are required to create Cluster Index based on destination city, Designation of staff, and address of passenger. For this consider the following tables and apply Cluster indexing technique on required columns.

Aiport-Flight Table

Airport Id	**Flight Id**	**Arrival date**	**Arrival time**	**Depart Date**	**Depart Time**	**Destination**	**Aiport Name**	**City**
AP1	Pk01	15-10-2018	12:30 PM	16-10-2018	10:00 AM	United Kingdom	Allama Iqbal Airport	Lahore
AP2	Pk02	19-11-2018	10:00 AM	19-11-2018	04:00 PM	Dubai	Baynazir Airport	Rawalpindi
AP3	Bg01	16-10-2018	12:30 PM	16-10-2018	06:30 PM	KSA	New Islamabad Airport	Islamabad
Ap4	Pk03	17-10-2018	09:00 PM	17-10-2018	04:00 PM	United Kingdom	Jinnah Airport	Karachi

Ap5	Bg02	25-10-2018	08:00 AM	26-10-018	12:00 PM	USA	New Lahore Airport	Lahore
Ap6	Bg03	19-11-2018	10:00 AM	19-11-2018	04:00 PM	Dubai	International Quetta Airport	Quetta

Staff Table

Staff Id	**Name**	**Designation**	**Contact No**
St01	Micheal	Operation Agent	0092-345-7865439
St02	Jackob	Flight Attendant	0062-876-0987654
St03	Alfard	Avionic Technician	0092-321-9865321
St04	Jackson	Operation Agent	0072-098-7854321
St05	Joseph	Flight Dispatcher	0092-333-9054213
St06	Joliana	Passenger Service Agent	0062-900-6789012
St07	Thomsan	Flight Attendant	0052-321-9084563

Passenger Table

Passenger Id	**Name**	**Contact No**	**Address**	**Email**	**Credit Cr.No**
Ps01	Julia	0092-345-7865439	Islamabad	julia@gmail.com	123-987
Ps02	Alexandra	0062-876-0987654	Dubai	alex@gmail.com	324-908
Ps03	Robert	0092-321-9865321	London	robert@live.com	457-975
Ps04	Alaf	0072-098-7854321	Islamabad	alaf@yahoo.com	345-075
Ps05	Julia Sanf	0092-333-9054213	New York	jausanf@gmail.com	123-890
Ps06	Charistea	0062-900-6789012	London	charist@live.com	768-054

Solution:

Cluster index on **Destination** column

Airport Id	**Flight Id**	**Arrival date**	**Arrival time**	**Depart Date**	**Depart Time**	**Destination**	**Aiport Name**	**City**
AP2	Pk02	19-11-2018	10:00 AM	19-11-2018	4:00 PM	Dubai	Baynazir Airport	Rawalpindi
Ap6	Bg03	19-11-2018	10:00 AM	19-11-2018	4:00 PM	Dubai	International Quetta Airport	Quetta
AP3	Bg01	16-10-2018	12:30 PM	16-10-2018	16-10-2018	KSA	New Islamabad Airport	Islamabad
AP1	Pk01	15-10-2018	12:30 PM	16-10-2018	10:00 AM	United Kingdom	Allama Iqbal Airport	Lahore
Ap4	Pk03	17-10-2018	9:00 PM	17-10-2018	4:00 PM	United Kingdom	Jinnah Airport	Karachi

Ap5	Bg02	25-10-2018	8:00 AM	26-10-018	12:00 PM	USA	New Lahore Airport	Lahore

Cluster index on **Designation** column

Staff Id	**Name**	**Designation**	**Contact No**
St03	Alfard	Avionic Technician	0092-321-9865321
St02	Jackob	Flight Attendant	0062-876-0987654
St07	Thomsan	Flight Attendant	0052-321-9084563
St05	Joseph	Flight Dispatcher	0092-333-9054213
St01	Micheal	Operation Agent	0092-345-7865439
St04	Jackson	Operation Agent	0072-098-7854321
St06	Joliana	Passenger Service Agent	0062-900-6789012

Cluster index on **Address** column

Passenger Id	**Name**	**Contact No**	**Address**	**Email**	**Credit Cr.No**
Ps01	Julia	0092-345-7865439	Islamabad	julia@gmail.com	123-987
Ps04	Alaf	0072-098-7854321	Islamabad	alaf@yahoo.com	345-075
Ps02	Alexandra	0062-876-0987654	Dubai	alex@gmail.com	324-908
Ps03	Robert	0092-321-9865321	London	robert@live.com	457-975
Ps06	Charistea	0062-900-6789012	London	charist@live.com	768-054
Ps05	Julia Sanf	0092-333-9054213	New York	jausanf@gmail.com	123-890

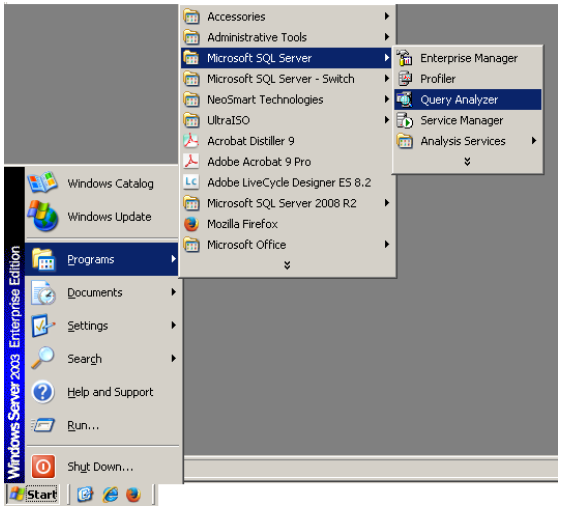
Mechanism to Conduct Lab:

Students and teacher communicate through Skype/Adobe Connect. Students perform the task using the following simulator:

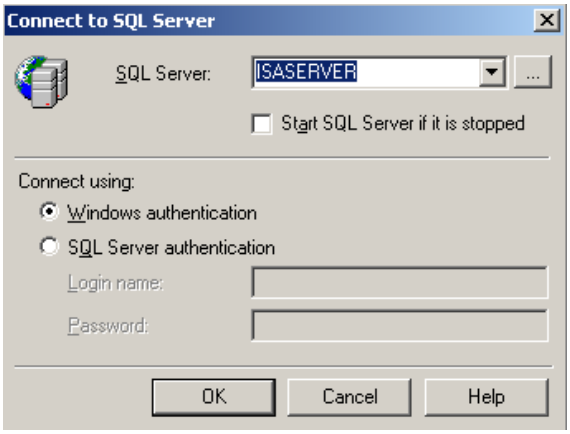
Lab 11

Nested Loop, Sort Merge, and Hash Join using SQL Server Query Analyzer

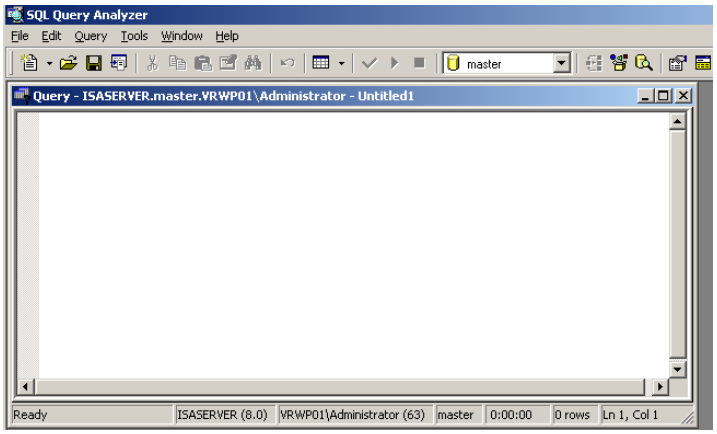
1. Open SQL Query Analyzer by clicking Programs->Microsoft SQL Server and Query Analyzer.



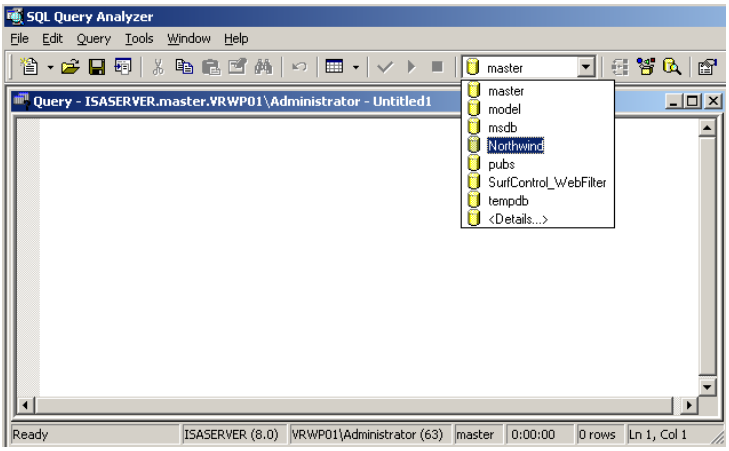
2. Select SQL Server **ISASERVER** and click oK.



3. The following SQL Query Analyzer window will open. In this window, you will write SQL queries.



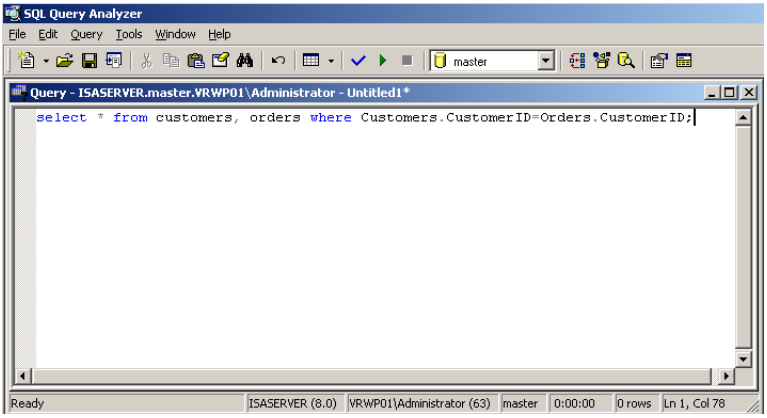
4. First select sample database **Northwind.**



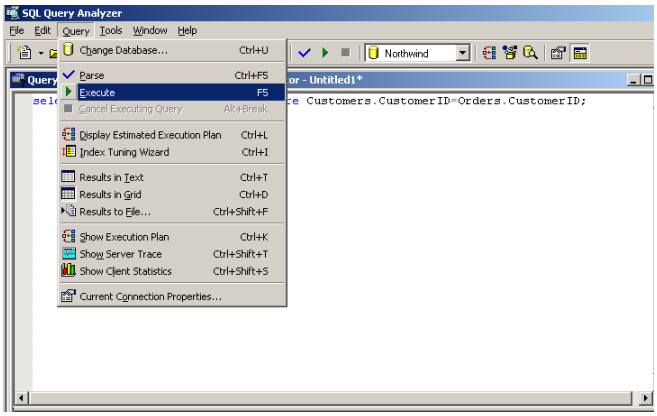
5. Write following SQL query in query analyzer window.

select * from customers, orders where Customers.CustomerID=Orders.CustomerID;

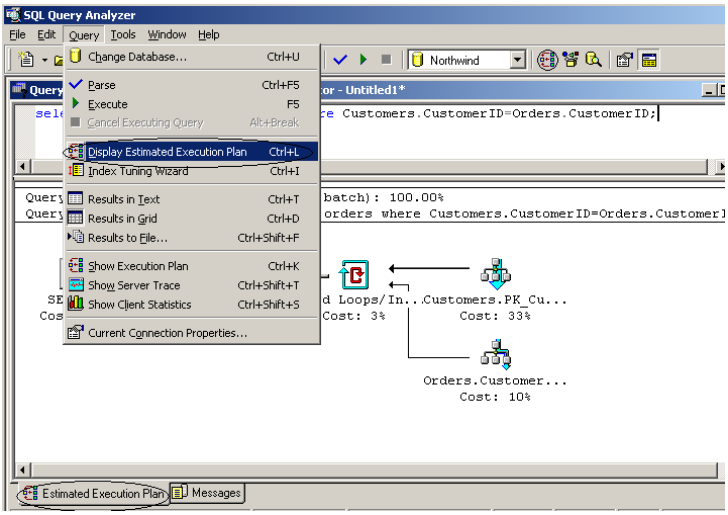
Here customers and orders are table of **Northwind** database**.**



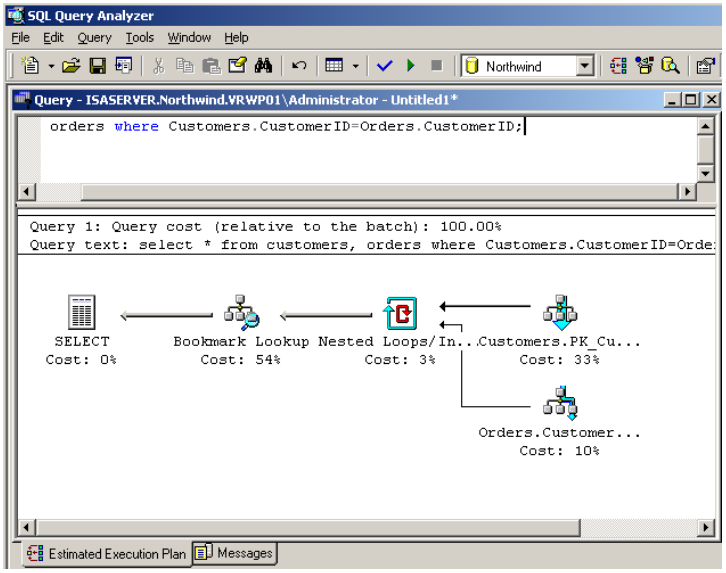
6. Execute SQL query by clicking **Query** menu and select **Execute** option. You can also click execute button available on standard toolbar.



7. Click on **Query** menu and select option **Display Estimated Execution Plan** to see execution plan of query. Estimated execution plan can also be viewed by using highlighted options in following figure.



8. By default, Nested loop join is performed, you are required to analyze execution plan.

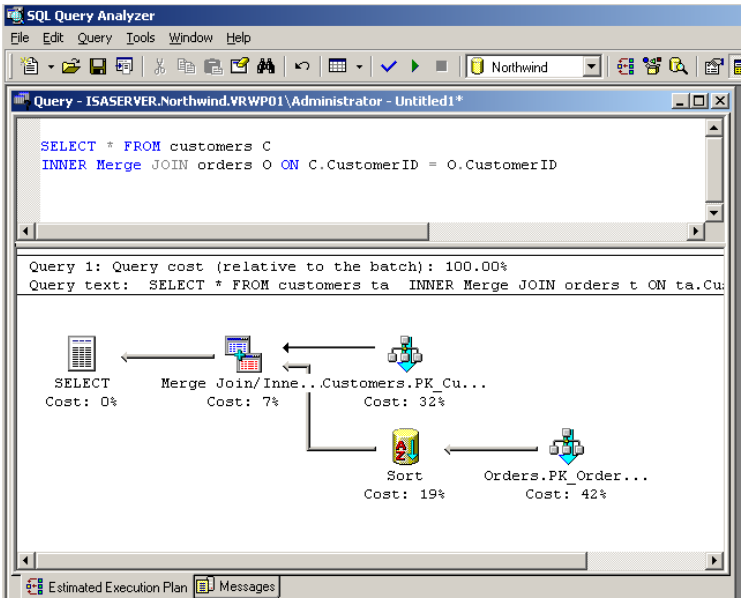


Merge Join Query:

For Merge Join, write following query in SQL Query Analyzer window.

SELECT * FROM customers C

INNER Merge JOIN orders O ON C.CustomerID = O.CustomerID

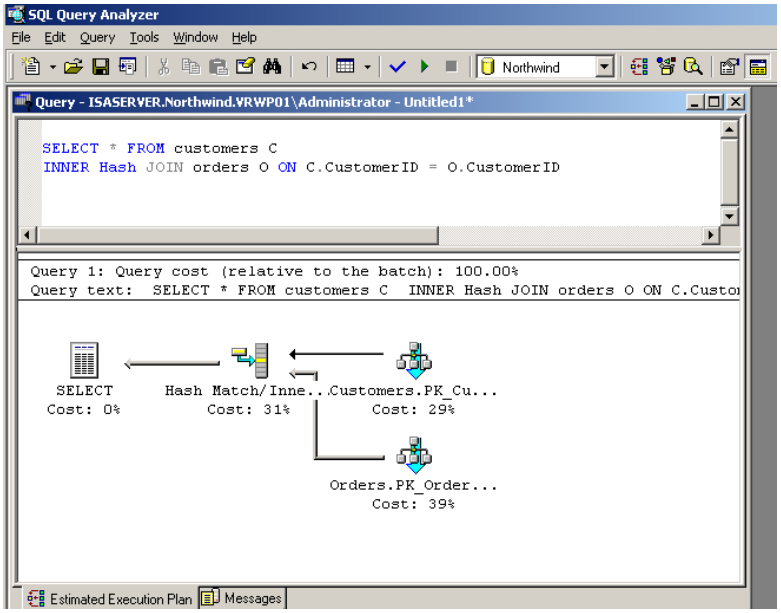


Hash Join Query:

For Merge Join, write following query in SQL Query Analyzer window.

SELECT * FROM customers C

INNER Hash JOIN orders O ON C.CustomerID = O.CustomerID



Lab Exercise: You are required to analyze Nested Loop Join, Merge Join and Hash Join for the given query in terms of which one is efficient in terms of execution time.

Mechanism to Conduct Lab:

Students and teacher communicate through Skype/Adobe Connect. Students perform the task using the following simulator: