1. Creation of a Data Warehouse.

¾ Build Data Warehouse/Data Mart (using open source tools like Pentaho Data Integration Tool, Pentaho Business Analytics; or other data warehouse tools like Microsoft-SSIS, Informatica, Business Objects,etc.,)

¾ Design multi-dimensional data models namely Star, Snowflake and Fact Constellation schemas for any one enterprise (ex. Banking, Insurance, Finance, Healthcare, manufacturing, Automobiles, sales etc).

¾ Write ETL scripts and implement using data warehouse tools.

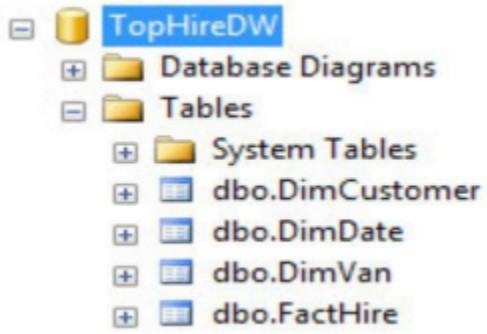
¾ Perform Various OLAP operations such slice, dice, roll up, drill up and pivot

---

A)Create the Data Warehouse

So now we are going to create the 3 dimension tables and 1 fact table in the data warehouse: DimDate, DimCustomer, DimVan and FactHire. We are going to populate the 3 dimensions but we'll leave the fact table empty. The purpose of this article is to show how to populate the fact table using SSIS.

First I'll show you how it looks when it's done:



Date Dimension:

	DateKey	Year	Month	Date	DateString
1	0	Unknown	Unknown	0001-01-01	Unknown
2	20060101	2006	2006-01	2006-01-01	2006-01-01
3	20060102	2006	2006-01	2006-01-02	2006-01-02
4	20060103	2006	2006-01	2006-01-03	2006-01-03
5	20060104	2006	2006-01	2006-01-04	2006-01-04
6	20060105	2006	2006-01	2006-01-05	2006-01-05

Customer Dimension:

	CustomerKey	CustomerId	CustomerName	DateOfBirth	Town	TelephoneNo	DrivingLicenceNo	Occupation
1	1	N01	Customer01	2000-01-01	Town01	Phone01	Licence01	Occupation01
2	2	N02	Customer02	2000-01-02	Town02	Phone02	Licence02	Occupation02
3	3	N03	Customer03	2000-01-03	Town03	Phone03	Licence03	Occupation03
4	4	N04	Customer04	2000-01-04	Town04	Phone04	Licence04	Occupation04
5	5	N05	Customer05	2000-01-05	Town05	Phone05	Licence05	Occupation05

Van Dimension:

	VanKey	RegNo	Make	Model	Year	Colour	CC	Class
1	1	Reg1	Make1	Model1	2009	White	2500	Medium
2	2	Reg10	Make10	Model10	2010	White	2500	Medium
3	3	Reg11	Make11	Model11	2011	White	3000	Large
4	4	Reg12	Make12	Model12	2008	White	2000	Small
5	5	Reg13	Make13	Model13	2009	Black	2500	Medium

B)multi-dimensional data models

Multidimensional schema is defined using Data Mining Query Language (DMQL). The two primitives, cube definition and dimension definition, can be used for defining the data warehouses

and data marts.

data models namely

1.Star Schema,

2.Snowflake Schema

3.Fact Constellation schemas.

C)ETL

ETL comes from Data Warehousing and stands for Extract-Transform-Load. ETL covers a process of how the data are loaded from the source system to the data warehouse. Extraction–transformation–loading (ETL) tools are pieces of software responsible for the extraction of data from several sources, its cleansing, customization, reformatting, integration, and insertion into a data warehouse.

Building the ETL process is potentially one of the biggest tasks of building a warehouse; it is complex, time consuming, and consumes most of data warehouse project‘s implementation efforts, costs, and resources.

Building a data warehouse requires focusing closely on understanding three main areas:

1. Source Area- The source area has standard models such as entity relationship diagram.

2. Destination Area- The destination area has standard models such as star schema.

3. Mapping Area- But the mapping area has not a standard model till now.

D) OLAP operations

Online Analytical Processing Server (OLAP) is based on the multidimensional data model. It allows managers, and analysts to get an insight of the information through fast, consistent, and interactive access to information.

OLAP operations in multidimensional data.

Here is the list of OLAP operations:

- Roll-up

- Drill-down
- Slice and dice
- Pivot (rotate)