CHAPTER 1:

INTRODUCTION

In the information technology profession, we have seen few of the applications like order processing, general ledger, inventory, in-patient billing, checking accounts, insurance claims, and so on. These applications are very popular systems which run businesses. They process orders, maintain inventory, keep the accounting books, service the clients, receive payments, and process claims. Without these computer systems it is very difficult to survive in today’s modern generation.

As businesses grew more complex and complex, corporations spread gradually. The operational computer systems did provide information to run day-to day operations, but what the executives needed were different kinds of information that could be readily used to make strategic decisions. The operational systems, is important as they were not able to provide strategic information. Data warehousing is a new paradigm specifically intended to provide vital strategic information.

According to the author Bill Inmon the term ‘Data Warehouse ‘is defined as follows

“A data warehouse is a subject-oriented, integrated, time-variant and non volatile collection of data in support of management’s decision making process”
“Subject-oriented” here means that the data addresses a specific subject such as sales, inventory, etc. “Inventory” means that the data is obtained from heterogeneous sources. “Time variant” implies that the data is stored in such a way that when some data is changed, then that data has been changed is also stored. “Non-volatile” implies that data is never removed, i.e., historical data is also kept. So, whenever a change takes place in any of the field values of a table, the previous values also need to be kept in addition to the present values.

The Data warehouse obtains the data from various data sources. The data from these sources are then converted into a form suitable for data warehouse. This process is called Extraction Transformation and Loading (ETL) of data into target database as shown in Figure 1.
1.1 ETL introduction

ETL is an abbreviation for Extract, Transform, and load. ETL are the three database functions that are combined into one tool that automates the process to pull data out of one database and place it into another database [9, 10].

Extract -- the process of reading data from a specified source database and extracting a desired subset of data.

Transform -- the process of converting the extracted/ acquired data from its previous form into the form it needs to be in so that it can be placed into another database. Transformation occurs by using rules or lookup tables or by combining with other data.

Load -- The process of writing the data into the target database.

1.2 ETL Functional Elements

Regardless of how they are implemented, all ETL systems have a common purpose: they move data from one database to another [26]. Generally, ETL systems move data from OLTP systems to a data warehouse, but they can also be used to move data from one data warehouse to another. An ETL system consists of four distinct functional elements:

- Extraction
- Transformation
- Loading
- Meta data