Introduction

Nowadays most companies have huge operational databases which provide a large wealth of useful information, for example analysis, in order to take decisions faster and better. Information provided can support:

- demand evolution;
- critical business area identification;
- budgeting and management transparency;
- winning stragegies.

All aspects previously reported are known as *business intelligence* whose purpose is transform data into actionable information, looking at:

- user's requirements;
- hardware and software infrastructures.

Applications supported are very different:

- manifacturing companies;
- distribution;
- financial services and insurance;
- telecommunications;
- public service;
- health.

Data warehouse's database is exclusively use to decision support and it is separated from company operational database. For the first one, data is:

- integrated;
- time dependent;
- applied on a specific subject.

In data management, instead, data should have the following problems:

- missing information (history which is fundamental in data warehouse);
- consolidation;
- low quality, due to inconsistency problems.

Company data warehouse contains all information on the company business so to design the structure and implement it takes a long time. Data mart, instead, focuses on a given subject so information is only a subset. Implementation is faster but the design is a very critical task because subsequent data mart integration problems have to be avoided. Two kind of architectures can be used:

. dependent;

. independent.

The first one is fed by the company data warehouse while the second one is fed directly by the source.

ETL (Extraction Transformation Loading) represent the following tasks:

. data extraction from the source;

. data cleaning (possible errors, missing or duplicate data);

. transformation and conversion;

. data loading;

. periodical refresh.

Usually data is represented as a cube which can have three or more dimensions; cells at dimension intersections are measures on which analysis are performed. Analysis can be:

. OLAP;

. data mining techniques.

The first approach supports different types of aggregation function, average for example, while analysis done with data mining techniques are several types with significant algorithmic contribution, such as:

. classification and regression;

. association rules;

. clustering.

All tools allow the user give several presentation of data processed by a query and provide data exploration by means of progressive incremental steps.

Servers for data warehouse are:

. ROLAP (Relational OLAP);

- MOLAP (Multidimensional OLAP);
- HOLAP (Hybrid OLAP).

The relational representation is the star model, in which:

- numerical measures (measures whose attribute domain is numerical) are stored in the fact table;
- dimensions are charactherized by many descriptive attributes and describe the context of each measure in the fact table.

There are different types of metadata, according to the purpose:

- metadata for data transformation and loading;
- metadata for data management;
- metadata for query management.