Data Warehouse

Raphael Javaux

December 2011

Abstract

This article introduces data warehousing and shows the differences with traditional database management. Data warehousing and data mining are the techniques of extraction of correlations and facts in corporations data and take part in the business intelligence process.

Contents

1 Introduction

As of today, more and more information can be stored and modern computing power enable complicated statistics variables to be calculated and analysed.

Today's corporations maintain huge data about their activities. This large bank of information is a real forgotten treasure for most companies.

Most of time, all the knowledge a firm needs for marketing and economical strategy can be found in its operations history.

How to make this product more attractive ? What is the best period to launch an advertisement campaign for a given product ? These are examples of questions for which the data warehouse field is able to respond.

The purpose of this paper is to introduce the reader to data warehouse. There is no prerequisite, excepts basics in databases management.

2 Definition

Since their early days, computers were build to manage and analyse large amounts data.

By automatising information handling, these machines were quickly proved required for any knowledge driven company.

By years, firms have accumulated and kept digital materials for about everything in touch with their business, even when data were no longer required for every day operations.

Quickly, computer scientists have seen the benefits of these "old" information for affair decisions and have been looking for new techniques for storage, centralization and extraction of facts, with success. First data warehouses were born.

Because of information scattering and the really different requirements of facts extraction (which is called *data mining*), databases in use for statistics are not the same as *operational databases* (databases in use in business parts of the company). These databases are called *data warehouse* and are only used for data mining.

3 Comparing Operational Databases and Data Warehouse

Data warehouses don't meet the same needs as operational databases.

Conception   Both systems are not accessed the same way. Data warehouses users will **ask for statistics about data** (using dimensions and grains[1]) while operational databases users will **ask for data** (using relations between records).

Thus, DBMS software in use for operational databases are not the same as systems in use in data warehouse. Though Oracle have some extensions for data-mining, MySQL doesn't. Software designed for data warehouse exist in both proprietary and free software worlds.

Performance   Data warehouses don't have to give the same performance level. For example, an operational database must be able to answer to hundreds of concurrent requests while a data warehouse will only receive a few queries per day.

Security   Whereas in a operational database, security is the top priority, this is not so important in a data warehouse. A lot of critical operations related to operational databases, like financial transactions, don't exist in a data warehouse.

Storage requirement   Because data warehouses need to store a large history of operations, they expect more devices for storage.

Real time   Data warehouses don't need to be immediately updated with new data. Most of the time, a data warehouse is synchronised with an operational database a few times per week.

---

[1]**Dimensions** are the different facts available in a data warehouse, for example *sales* or *users registrations*. **Grains** are the levels of detail of a given dimension, for example *sales per week* or *sales per day*

4   Conclusion

Whereas data warehouse systems store structured data, just like traditional relational database systems, they are really different by design.

The way of working with decisional data is really different from the way of working with operational data. Both approach require divergent infrastructures, software and skills.