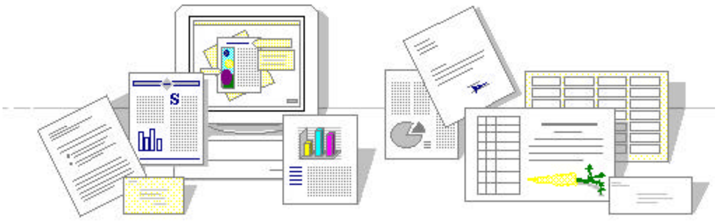


Data Warehousing, LANSA’s Practical Solution (2)

The term “Data Warehousing” has been around for just a few years, but the concept of data warehousing has been around for much longer, although under different names.

Data warehousing, as we know it today, is an evolution of data analysis techniques that we have been using for years. Data warehousing is a more formalised methodology of these techniques. For example, many sales analysis systems and executive information systems (EIS) get their data from summary files rather then operational transaction files. The method of using summary files instead of operational data is in essence what data warehousing is all about.

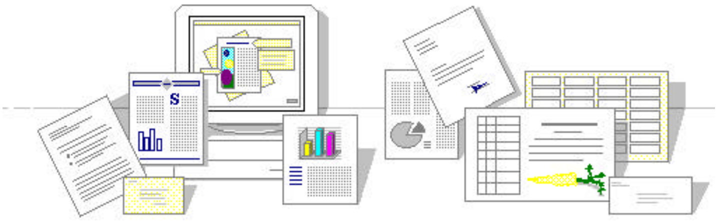
Some data warehousing tools neglect the importance of modelling and building a data warehouse and focus on the storage and retrieval of data only. These tools might have strong analytical facilities, but lack the qualities you need to build and maintain a corporate wide data warehouse. These tools belong on the PC rather than the host.

Your corporate wide (or division wide) data warehouse needs to be scalable, secure, open and, above all, suitable for publication.

- Scalable means that your data warehouse must be able to handle both a growing volume and variety of data and a growing number of users that can access it. Most companies prefer for this reason to store their corporate wide data warehouse in a relational database above a multi dimensional data base storage. (You can model your data dimensional and store it in a relational database. More about dimensional modelling techniques later.)
- Secure means that your data warehouse administrator can centrally control who is allowed to access what data and when.
- Open means that the data in your data warehouse is open to a wide range of query and other front end tools. For this reason a relational data base should be your first choice for a corporate wide data warehouse. The proprietary data storage structures that are used by some data analysis tools can be fed from this central data warehouse.
- Suitable for publication means that the data must be reliable, consistent , complete and well documented. The data must also be well layed out for retrieval by business users. Business users require easy data navigation and speed of access. The many layers of file join relationships in a production system make data access too complex and also too slow. A more suitable data structure for a data warehouse is a dimensional model.

Scalability, security and openness depend largely on your choice of platform, database and tools. To make data suitable for publication you need a combination of business skills, data modelling skills, tools to build your data warehouse and finally (the penthouse of the building) tools to access your data warehouse. No matter how glamorous and smart your query and data analysis tools are, if the underlying structure of your data warehouse is not well layed out, your decision support system (DSS) is doomed to fail.

This document will explain how to build a successful data warehouse on the AS/400 (a scalable, secure and open platform) using your business and modelling skills and LANSA as a tool to help you.



Data Warehousing, LANSA’s Practical Solution (2)

In this document we will have a look at the first 3 steps in your path to a successful data warehouse:

1. Justification of a data warehouse
2. Getting user information requirements
3. Designing the data warehouse

In a next document we will have a look at the further steps.

Justification of the data warehouse

In a nutshell:

- Get sponsorship from senior management
- Start with a small high impact pilot project
- Use the skills of an experienced business analysts with strong political skills as well.

Data warehousing is user driven requirement. It is most likely that the marketing or sales department initiates the data warehouse and not the IT department. Data warehouse consultants and solution providers are sometimes hired directly by one of these departments or by general management.

Operational systems are an easy to justify necessity and its benefits are often instantly measurable (for example faster invoicing and delivery, lower stock levels and faster debt collection). Data warehousing benefits are seldom instantly measurable and you need senior management’s sponsorship to achieve them. Typical data warehousing benefits are:

- Better business intelligence to enhance corporate strategy
- Greater insight in competitive positioning
- Enhanced customer service
- More cost effective decision making

Notice that they all support the corporate goals.

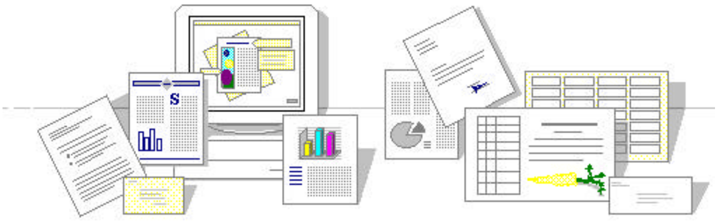
Data warehousing is about improving the score.
Operational systems are about keeping the score.

If senior management is not convinced that data warehousing will help to achieve the above benefits, than the company is not yet ready for data warehousing.

A few rules about the prove of concept in a pilot data warehousing project:

- The first DW project should **not** be too ambitious. Start small with a ‘data mart’ and expand from there. Data marts have higher and quicker return on investment.
- Do **not** try to give even better figures to the accountants. The pilot DW project should support an area where **better information and creativity can improve the results.** Almost always you can prove the data warehousing concept in a sales or customer focussed area.

Needless to say that you need business analysis skills in this step and quite some political skills as well. Have a business focus, not a technical focus.



Data Warehousing, LANSA’s Practical Solution (2)

Getting the User Requirements

In a nutshell:

- Have a formal interview process
- Make sure you are dealing with the right user(s)
- Use the skills of a business analysts, also must have interview skills and some political skills.

Use a formal Interview process to define user requirements more precisely.
A well executed interview process is essential for the success of the data warehouse. Have interview templates and checklists.

A series of interviews will be held with senior management, user management, end users of the data warehouse and the IT department.

Some data warehouses are based on the requirements of only two or three top executives. This is quite common. The right user will come with requests that support the company’s goals. Remember, DW is about improving the score, not about keeping the score.

Next to collecting user requirements, also have a look to what data is available in the operational systems. Don’t base your data warehouse model on what data is available in the operational systems. If some data is ‘missing’, suggest enhancements to the operational system or enter the data in a separate sub-system to complement the operational system. Some data maybe purchased from third parties.

A typical example of ‘missing’ data is information about sales campaigns (both of the company and its competitors). The start and end data of a sales campaign is usually totally unrelated to any reporting period in the operational system.

In this phase of the project you need a business analyst with sound interview skills. The business analyst must have a decision support system mind set. Some political skills are required as well. Again, you need a business focus not a technical focus.

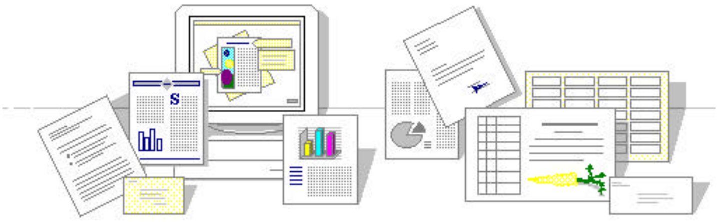
Designing the data warehouse

In a nutshell:

- Use dimensional modelling techniques
- Don’t just replicate the data from your operational system
- Use the skills of an experienced data modeller.

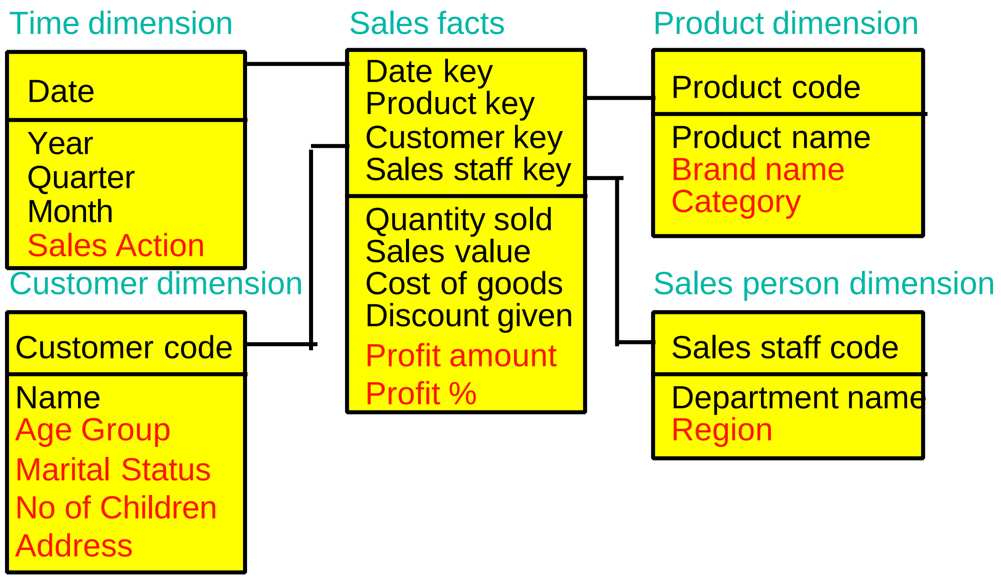
The user must be able to separate and combine the data in a data warehouse by means of every possible measure (dimension) in the business. This slicing and dicing of data must be easy and fast. The ease and speed of data access is very much related to the data model.

The entity-relationship data modelling technique might be very efficient for production systems with a high number of transactions, but it is not a suitable method for designing a data warehouse. The many layers of file join relationships in a production system make data



access too complex for business users and also too slow for queries. A more suitable data structure for a data warehouse is a dimensional model, also called star join schema.

A dimensional data model is much flatter than an entity relationship model. In a dimensional data model you need rarely more than one level of a file join to get to the data. Typically a dimensional model consists of one fact table that contains the data and a number of surrounding dimensions, hence the name star join schema.



Time dimension

Date

Year
Quarter
Month
Sales Action

Customer dimension

Customer code

Name
Age Group
Marital Status
No of Children
Address

Sales facts

Date key
Product key
Customer key
Sales staff key

Quantity sold
Sales value
Cost of goods
Discount given
Profit amount
Profit %

Product dimension

Product code

Product name
Brand name
Category

Sales person dimension

Sales staff code

Department name
Region

In an operational system to list sales/order details with customer data and product category description, you would usually have to

1. Join from the order detail to the order header file
2. Join from the order header to the customer file to get the current customer data
3. Join from the order detail to the product file
4. Join from the product file to the product category file to get the category description.

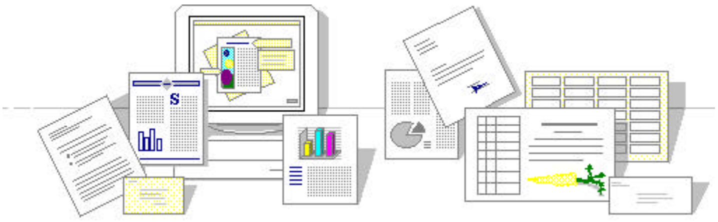
In the above dimensional model you would need only two joins:

1. From the fact table to the customer dimension (usually containing historic data).
2. From the fact table to the product dimension.

Also notice that a data warehouse, is likely to contain:

- Derived information such as profit amount and profit % instead of just the sales and cost values.
- Historic information as it was at the time of a transaction. For example, to examine the purchase behaviour of customers you will probably be more interested in the demographic data of a customer (age, # of children, address, etc.) at the moment of purchase, than his/her most current address.
- Summaries/aggregates by several levels. If business users require sales data summarized by, for example, product category, region and sales campaign, than it is quite common to keep that data summarized at all required levels.

Don’t try to save disk space when you are designing a data warehouse. Model for easy data navigation and speed of retrieval.



Data Warehousing, LANSA’s Practical Solution (2)

Please be aware that a multi dimensional data storage facility (usually presented by a cube and offered by data analysis tools like PowerPlay) is not directly related to a dimensional data model.

Multi dimensional data storage facilities provide optimized proprietary data retrieval methods that are only available to the tool it self. When a dimensional data model/star join schema is implemented in a relational database, it can be accessed with multiple tools. You can then load some of the data of your central data warehouse into the proprietary data storage of a data analysis tool. These data analysis tools are complimentary to your central data warehouse and usually PC based.

There is much more to be said about dimensional data modelling. To make your self familiar with dimensional modelling techniques, read :

- **“**The Data Warehouse Toolkit”, by Ralph Kimball, ISBN 0-471-15337-0
- DBMS monthly magazine, Data Warehouse architect column by Ralph Kimball
- See Web address http://www.dbmsmag.com/artindx3.html#A000314.

In a next data warehousing paper, we will discuss the further steps of data warehousing:

4. Building a data warehouse
5. Implementing a data warehouse
6. Reviewing a data warehouse