

Data Warehouse Newsletter #4

The Data Warehouse

The IT Managers Perspective

Version 1.0

16th October 2002

Author: Peter Nolan

pnolan@ozemail.com.au

Table of Contents

1. CHANGE CONTROL LOG	3
2. AUDIENCE	4
3. OVERVIEW	4
4. CONCLUSIONS	4
5. WHAT IS THE DATA WAREHOUSE ALL ABOUT?	5
5.1. What Do Your Business Managers Really Want?	6
5.2. Management Decision Making Process	7
5.3. The Role of the Data Warehouse	8
5.4. Fact Based Decision Making	9
5.5. A Way to Test New Ideas	9
6. WHAT IS A DATA WAREHOUSE?	10
6.1. What Does a Data Warehouse Look Like?	11

1. **CHANGE CONTROL LOG**

#	Date	Name	Description
1.0	16/10/02	P. Nolan	Initial publication to the web.

2. AUDIENCE

The intended audiences for this Newsletter are:

- IT Managers responsible for Data Warehouse initiatives.
- Database Administrators who are interested in how to design a Data Warehouse.

3. OVERVIEW

This Newsletter is an introductory overview of what the Data Warehouse is all about from the IT Managers perspective. Since it is for the IT Manager we have left the term Data Warehouse in this paper.

4. CONCLUSIONS

This Newsletter draws the following conclusions:

- A Data Warehouse has a defined, reliable and recognisable structure. It is not just 'a big database' nor any 'collection of data'. You should ask your vendor to explain how their offering fits into the industry accepted definition of a Data Warehouse.
- The purpose and role of a Data Warehouse is to be central to the management decision making process.

5. WHAT IS THE DATA WAREHOUSE ALL ABOUT?

You have probably been reading the IT press articles on Data Warehouses, hearing all sorts of vendors talk about Data Warehouse and in your mind you might be asking "What is this Data Warehouse thing all about?". I've been working on Data Warehouse projects since before it was called a "Data Warehouse". I can share with you, from eleven+ years of experience in Australia and around the world, that the Data Warehouse is all about "Supporting the Management Decision Making Process". It's that simple, and that complicated.

Why is "Supporting the Management Decision Making Process" complicated? Managing a company is all about having far more opportunities for investing (spending) money than there is money to invest. Where does management invest the money for best returns? Where do new ideas, with the possibility of improving profit, come from? How does management evaluate opportunities?

Today we use Internal Rate of Return, Payback Periods, Strategic Importance and other methods of evaluating and ranking projects to determine which will go ahead. No matter what method you choose to rank projects there is a common denominator for all methods.

DATA

What is your experience of having senior management demand data that comes from various sources so that they can quickly and efficiently evaluate three or four different opportunities for investment?

5.1. *What Do Your Business Managers Really Want?*

The first question we ask our business manager customers is “What do you really want?”. We invite you to consider this question and how it may apply to your organisation.

Over the years I have often found that what business managers really want is to make a difference to their business. Effective business managers are focused on profitability, both revenue and costs, and are most concerned with being able to manage and grow the business. Perhaps your business managers would provide answers to these questions including some statements like these.

I want:

- to be able to reasonably rank my investment opportunities and select those that are most likely to make the best return to my business.
- to know what is actually happening in my business.
- to know what my customers ‘look like’.
- to know what my customers are buying.
- to be selling my customers what they are likely to be interested in buying.
- to be able to ask new questions, questions I have only just thought of as possibly making a difference to the business.
- a source of new ideas. Ideas that can be tested quickly and simply and if they show promise I want to be able to implement them with a minimum of time and money.
- to know where my money is being spent.
- to know if my money is being spent most effectively.

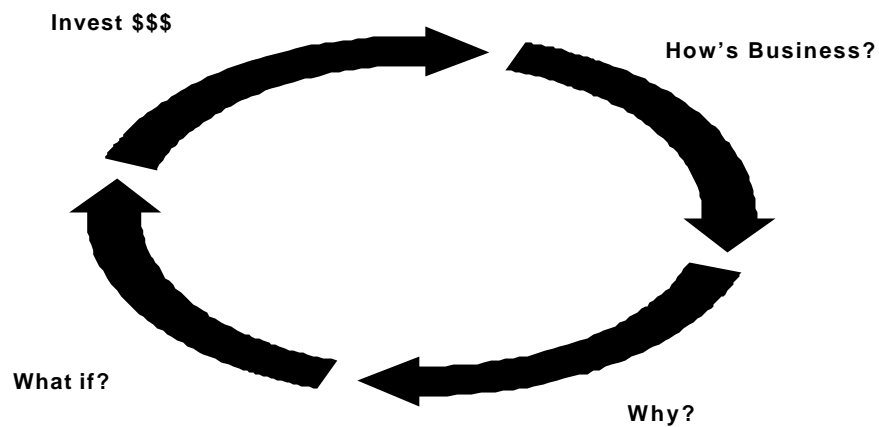
I don’t want:

- ‘sanitised’ reports of what people wish was happening in my business.
- five different reports from five different areas, all of which disagree.
- to take so long to get any information that by the time I make any changes it takes 12 months to find out if they are working.
- new ideas stifled by an analysis, prioritisation and funding process such that it takes 12-24 months for a new idea to be implemented.

If your business managers wish list looks like this, you are not alone. Many managers in many companies want to be able to manage the business with information about what is actually happening. Managers want to reduce the cycle time of decision making.

Providing this information is what the Data Warehouse is all about.

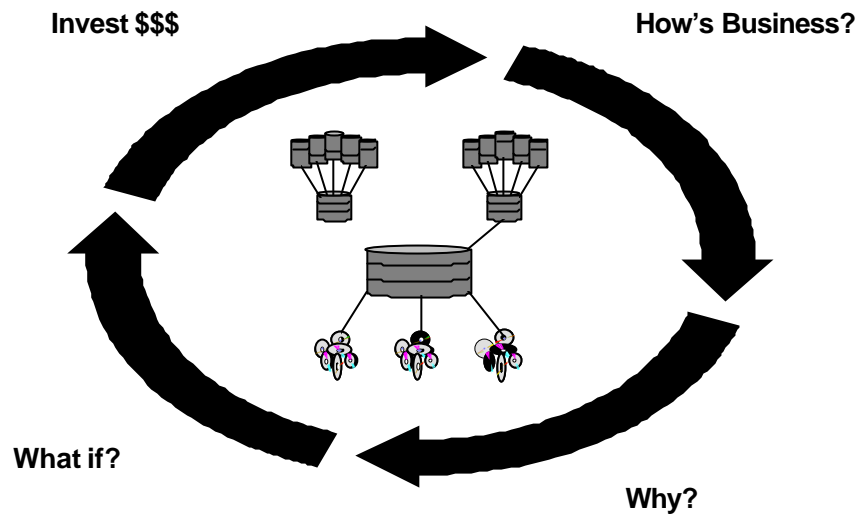
5.2. Management Decision Making Process



The management decision making process could be depicted as above:

1. How's Business?
The first and eternal question, continually on the mind of the astute business manager.
2. Why?
So why is the business like it is? What is driving the business? What is internal? What is external? There are many questions that come under the heading of "Why is the business the way it is?"
3. What if?
What if the business didn't have to be the way it is? What if we were able to do something other than what was going to happen? What if we did something different? What might be possible then?
4. Invest \$\$\$
On the basis of doing something different the business manager invests some money in a project to make a change to the business. Which leads to the eternal questions of "How did that work out?". That is...
5. How's Business?

5.3. The Role of the Data Warehouse



The role of the Data Warehouse is to be central to the management decision making process. The data by which management answers the questions:

- How's business?
- Why?
- What If?

is stored in the Data Warehouse.

We invite you to consider the question:

"How long does it take to go around this cycle once in my business?"

The reason we invite you to consider this question is because this cycle time has another name. The other name for this cycle time is "**Competitive Advantage**". For example, if it takes a competitor 24 months to go through this cycle to launch a new product and your company can do it in 6 months, who wins?

5.4. Fact Based Decision Making

The Data Warehouse brings to your business the possibility of Fact Based Decision Making. You will have the opportunity to provide a system that will enable business managers to make decisions based on what is actually happening in the business. Your business managers will be able to look at trends over time and you will be able to make predictions based on solid historical information.

One might say that business managers have always been performing Fact Based Decision Making. After all, the IT department supplies the facts and the managers make decisions based on those facts. However, facts are valid at one point in time, and to get a corporate-wide set of figures valid at one point in time typically takes significant effort. There is evidence of this in the closing of accounts and creation of the end of year statements for companies which, in many cases, takes weeks or months. Thus business managers have been managing with facts that usually represent how the business was quite some time ago and the more numerous the source systems for those facts the less well integrated and reliable those facts have been.

Given recent events in the US (Enron, Worldcom etc.) one might ask whether decisions were made on anything even closely related to 'the facts'. Certainly a lot of shareholders made buying decisions based on anything but the facts.

A well designed Data Warehouse usually contains 'facts' as at the end of the previous business day. This is a significant improvement for almost any Bank or Insurance company.

5.5. A Way to Test New Ideas

In any business there is more opportunity in a new idea than an old idea repeated. So, where do new ideas come from? How does one test a new idea to see if it has potential?

One of the major uses of the Data Warehouse is for business analysts to go "Data Mining" in the Data Warehouse, testing their new ideas. Business analysts make comments such as "Nine times out of ten when I test an idea against the Data Warehouse I come up with a lump of coal. The idea, though it looked good, would not produce enough profit. One time out of ten I come up with a diamond. The idea would produce significant profit." A 'diamond' could be classed as an idea that generates more profit than the entire Data Warehouse project cost. The fastest payback period I have heard of for a Data Warehouse is two weeks.

Without a Data Warehouse no-one tests new ideas against what is actually happening because it is too hard and too time consuming. The cost of the nine fruitless efforts discourages people from looking long enough to find the diamond.

Make no mistake, there are diamonds in your data, just waiting to be discovered. The Data Warehouse will make those diamonds visible to your talented business analysts.

6. WHAT IS A DATA WAREHOUSE?

This section defines in some detail what is a Data Warehouse. At the end of this Newsletter you will be well armed to question your vendor as to how they are going to support the various features of a Data Warehouse.

It is important to define what a Data Warehouse is. Despite good definitions being available for some years now every vendor has rebranded anything remotely to do with 'data' as a component of the Data Warehouse, the Decision Support System or, the latest thing, the Analytical Application. This is so much so that almost any collection of data is called a Data Warehouse. Nothing could be further from the truth. A Data Warehouse has a well defined, reliable and identifiable structure. Indeed there is a well accepted industry definition of the Data Warehouse.

A Data Warehouse can be defined as:

- Time Variant
- Subject Oriented
- Integrated
- Non Volatile

"A Data Warehouse is a collection of data in support of the management decision making process."¹

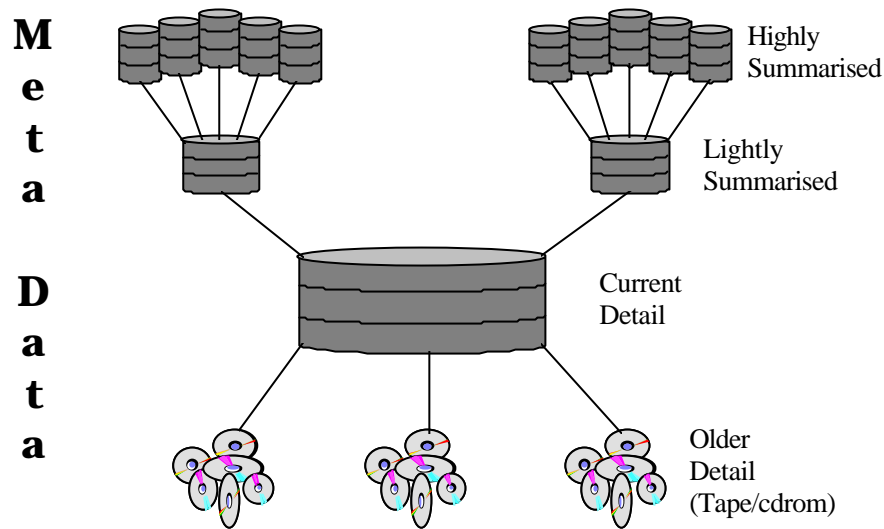
The terms are defined as follows:

1. **Time Variant:** This means that the Data Warehouse contains information over a series of time periods. For example, days, weeks, months years. A specific example would be a Data Warehouse can tell you your interest earned this year versus last year, or this month versus this month last year. The Data Warehouse can tell you your interest earned from the very high company level all the way down to individual products and customers.
2. **Subject Oriented:** This means that the Data Warehouse is organised around the various real subjects of the business as defined by strategic marketing thinking. The most fundamental question, "Who bought/used, what, when, where and why?" reveals the natural subjects of the business such as Customers, Products, Time, Geographies. The why question is often more difficult to deal with. In a Data Warehouse the question, "How many customers own a category of product (loan) over time?" is a trivial question. Yet this is a very difficult question for most banks to answer if there are different loan systems.
3. **Integrated:** Most banks have many different operational systems. If the different operational systems manage different products it is very difficult to get an integrated view of the customer. Many banks are striving to have one view of the customer and many industry analysts are saying that Relationship Banking and Customer Centric Banking is a must. A Data Warehouse, when all relevant source systems are fed into it, contains data integrated from many different systems and provides the ability for 'one view of the customer'.
4. **Non-Volatile:** Most computer systems are 'volatile'. That is, they are changed by the transactions that occur in the bank. For example, withdrawing money changes your account balance, changing your address changes the address field on the transaction system database. A Data Warehouse usually does not change data in tables. Generally it writes an entirely new record so that one can tell the value of the data in the record both before and after the change. Let's take account balance for example. Account balances are constantly changing and your account system can only tell you, "What is the total of account balances, right now." In a Data Warehouse you can say, "What was the total account balance at the close of business yesterday, and the day before, and what was the percentage change?". In fact you could ask, "Give me a trend line for the last 24 months by month, week, day of account balances." Because the data is never changed it is available until it is deleted from the Data Warehouse.

¹ Building the Data Warehouse by W. H. Inmon.

6.1. What Does a Data Warehouse Look Like?

A widely used diagram of the database for the decision support system is as follows. A Data Warehouse has data that can be represented as follows:



Specifically, a Data Warehouse:

- Stores highly summarised information which can be used to support requirements such as Executive Information Systems. The highly summarised information provides consistent fast response times for well defined reports such as monitoring reports and charts. A perfect example would be the amount of interest earned/paid by a specific demographic group of customers over a period of two to three years. If the detail level of data were used, the computer would have to read every single interest transaction and add them together to produce a chart with 24 data points on it. If the data was presummarised then far fewer records would be read. Resource and response times can be reduced over 1000 fold using such summarisation techniques. This makes trend analysis a far more cost viable activity.
- Stores lightly summarised information which can be used for analysis that requires more detail than reading the highly summarised information.
- Maintains a copy of Current Detail Data which also includes data which is historical and not yet archived. In my designs, the length of time the data remains in the current detail level is optionally defined by the customer.
- Stores archival data on tape or CD-ROM so that old data is safe, yet easily accessible.
- Stores multiple levels of data that will allow the users to look at long terms trends and then to drill down to the detailed transactions that make up the trends in a cost efficient and structured way. This is a key feature of the Data Warehouse.
- Always receives detailed transactions and then summarises those transactions internally to the Data Warehouse. Thus the figures in the Data Warehouse summary levels always match the details due to the fact that the Data Warehouse is responsible for all summaries.

-
- Does not necessarily store all data on the one physical machine. Should you wish to distribute information from the central Data Warehouse to distributed Data Warehouses this distribution can happen automatically and seamlessly. Specifically the summarised data may reside on a separate physical machine to the vast store of detailed data.
 - Contains a metadata database which describes the data that is available within the Data Warehouse and the processes that were performed to create the data in the Data Warehouse.

In summary, when you are considering a Data Warehouse you should satisfy yourself that the Vendor is talking about a database that satisfies the four key definitions of a Data Warehouse, has multiple levels of summarisation and has user defined data names that can be used in the query tools.