

The Advantage of Good Business Intelligence

By Peter B. Giblett

This is an excerpt from a book in-progress about successful BI implementation and is published here to give the reader clarity about my background...

The aim of most corporations is to run their organisations with minimal cost and use informed decisions in order to give them a competitive advantage. In all too many organisations data and information is not used to its best advantage. They have been leveraging operational data in order to make decisions but the process all too often is hit and miss and usually relies upon intensive manual labour to collate, sort, fix and summarise the data used. The process often takes most of the month to complete, with last month's month end reports being available during the last week of this month.

Ready access to information about an organisation's business, products and customers is a key element in supporting decision making. The creation of the data warehouse is an essential step in turning operational data into information. It is the Data Warehouse that is at the heart of the Business Intelligence provision.

At the very minimum the Data Warehouse provides a mechanism to collate, sort, validate fix and summarise data automatically, leaving those involved in manual processes to be reassigned to other duties. Computers have always been good at taking laborious monotonous tasks away from humans and doing them repetitively (shhhh! They can also do it much faster than the human can and are much more efficient).

However this is the minimum requirement, the Data Warehouse can provide so much more.

The Data Warehouse Defined

The Data Warehouse needs to be clearly distinguished from operational or on-line transaction processing (OLTP) systems. The readers should be familiar with OLTP systems. The Data Warehouse makes use of data in a different way to OLTP systems.

The Data Warehouse has been defined by Bill Inmon as:

A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process¹.

It was also defined by Ralf Kimball as:

a copy of transaction data specifically structured for query and analysis².

And by Price Waterhouse as:

An orderly and accessible repository of known facts...that is used as the basis for making better management decisions³.

It is *Subject Oriented* because it is classified and organised around subjects meaningful to an organisation. It has collected all data available about a subject from

all the operational systems. E.g. all sales, not just those captured via the e-business channel. The data from various systems is *integrated* into the Data Warehouse. It holds *time-variant* information and all facts/measurements will be time stamped either through their source transaction, or upon entry to the warehouse. Data can be both historical (e.g. 6 to 12 years old) or nearly current (e.g. yesterday, last week, last month). Data is *non-volatile* because it does not change in response to current operational activities and is based on a controlled data collection process. The data has not changed as soon as the report is created.

The Data Warehouse takes a *copy* of transactional data and *structures* it specifically to allow efficient *query and analysis* to take place throughout the warehouse in a uniform manner. This process will underpin corporate Business Intelligence (BI).

There are many ways, of course, to copy a transactional database, such as:

- To allow the management of a distributed database with each branch being responsible for its own data management with head-office systems being updated overnight.
- Replication to ensure that mission-critical database systems continue to operate in a 24x7 environment with the highest degree of availability possible. E.g. for use as a hot-standby.
- Replication for taking backups of critical systems, where the backup is completed off-line and subsequently re-synchronised.
- For inter-corporate communications. E.g. for automated inter-company purchase systems, or where data is sold.

It is clear that simple copies of databases do not amount to the creation of a Data Warehouse. This requires a restructuring process that reorganises data more effectively for analytical purposes and may include aggregation and summarisation of data. The Data Warehouse must not impact the operation or performance of the originating systems. The creation of a separate query and analysis database, which does not impinge on the day-to-day operational activities of the corporation, is one of the key reasons why it is necessary to build a Data Warehouse.

The Warehouse is *orderly* because it takes data from a number of diverse applications. It is methodically arranged with data being pulled into a structured data repository. This repository is *accessible* for use via reporting applications which will supply users with *facts* (or measurements) about their business, which will be used to support the management decision process.

Additionally Data Warehouses can provide the springboard for other management activities, as defined by Richard Tanler:

The data warehouse provides a single source of management information about the past, it should also provide information about the future... Bill Inmon's original definition... does not specify that the data warehouse is intended to be read only. Bill chose his words carefully in using non-volatile... Maintaining a forecast in the data warehouse requires periodic update of the tables or, more appropriately, creating new tables reflecting a new forecast version⁴.

It is clear that a Data Warehouse can ultimately do much more than simply report on past activities. It can provide the basis for future activities, such as budgeting and forecasting. World banks have, in the wake of September 11th 2001, had to take measures to block money laundering transactions from known terrorists, or their

financiers. Statutory bodies have also waded in with demands for greater compliance. The Sarbanes Oxley laws and the Basel II accords are examples of compliance requirements where applications will depend on a Data Warehouse as their data source.

Ultimately ***the Data Warehouse is a tool used by business to supply information necessary for informed decision making.*** Information is distinct from data. ***Information is data that has been intelligently processed via a Data Warehouse to allow it to be used for key business decision making.***

These characteristics are common to all data warehouses, but not all Data Warehouses are the same, they simply share common characteristics, both technically and in the business. These characteristics include the desire to use analysis of performance to drive the decision making process and identify areas to improve efficiency and gain a competitive advantage.

	On-Line Transaction Processing	Data Warehouse
Purpose	To run day-to-day operational activities of an organisation. E.g. to enter or modify purchase orders for a customer.	A copy of transaction data used for management and statutory reporting, information retrieval and analysis.
Structure	May be held on relational and non-relational databases.	A corporate relational data repository used for reporting. Normally based on a relational database.
Data Model	Normally defined as entity relationship models. Mainly normalised, although elements will be de-normalised when performance demands.	Defined as a star-join schema, a denormalised structure to support separate definition of dimensions and factual data.
Access	Via OLTP applications having SQL or middleware or proprietary data interface.	Via analytical applications that enable access to the Data Warehouse or subject-area Data Mart.
Type of data	Data that runs day-to-day activities within the business	Information used to analyse the business and make informed decisions.
Condition of data	Relates to current activities only, subject to change. Old data to be archived and removed.	Non-volatile, historical, and time variant. Historical data may be summarised.

Figure 2.1 – Comparing OLTP and the Data Warehouse

It is useful to understand the similarities and differences between OLTP and Data Warehouse Systems this can be understood from *Figure 2.1* that shows the purpose of each database type, how it is organised, the type of access that applies, and the data held in the database.

On-Line Transactional Processing

On-Line Transactional Processing (OLTP) systems (also called operational systems) are the backbone to corporate operational activities. These are systems that enable the processing of customer orders, generation of relevant invoices and control manufacturing and shipping procedures. Such systems are focused on fulfilling the day-to-day tasks of the business.

Operational systems are supported by a range of database types and filing systems. *Figure 2.2* provides some examples of non-database file types. Non-database filing systems are found in use for either older mainframe applications or with PC based data files or those stored on Local Area Networks. In certain organisations mainframe non-relational data files will be the primary source of transactional data, in other organisations they may not exist. Some of these filing systems, such as ISAM, have developed their data access capabilities and are considered by some now to be a database.

VSAM	Mainframe - Virtual Storage Access Method
ISAM	Mainframe - Indexed Sequential Access Method
RDE	Rapid Data Entry file.
XLS	PC/LAN - Microsoft Excel spreadsheet
CSV	PC/LAN/Other - Comma Separated Value file

There is a wide assumption that all databases are relational in nature. In fact many

IMS	A mainframe based hierarchical database
IDMS	Database provided by CA to support Cobol applications
INFOS	Data General database based on AOS/VS
MF-COBOL	Microfocus COBOL
DBM	A Hash table filing system on Unix

older systems were designed using network or hierarchical databases. It is not necessary to understand how these databases are structured, it is simply sufficient to recognise that they may contain vital corporate data. *Figure 2.3* provides some examples of non-relational databases. As with non-database filing systems these databases lie at the heart of many older legacy business applications.

Relational databases provide the data component layer for the majority of modern applications. For many organisations there is a fundamental architectural principle that all new applications shall interface with a relational database. *Figure 2.4* shows the leading relational databases, at the time of writing. Many applications are database independent (although they may have a preferred database).

Oracle	The leading relational database provided on multiple platforms.
DB/2	IBM's contender, provided on a wide range of platforms as DB/2-UDB.
SQL/Server	Microsoft's contender. This is a Windows/Intel database.
Sybase	A former leading contender.

Most of the relational databases also provide object support, that enables object oriented languages to connect to them effectively. Relational databases share a common data access language, SQL. SQL enables the database and application layers to remain independent.

There are also a new-wave of OLTP applications that store their data within Object databases. They are designed to work well with object programming languages such as C++ and Java. An ODBMS should be used when there is a Business need for high performance and optimisation associated with complex operational data (including for example multi-media data) containing a lack of unique identifiers plus a large number of many-to-many relationships. They are often used in real-time event driven applications. *Figure 2.5* shows some of the leading object databases. The issues surrounding their adoption as a mainstream technology based on the fact that here is no universally adopted standard for object databases and their lack of a

CA Jasmine	Features neural-networking and visualization technology
Versant	Object model with tight binding between the application object programming language and the underlying object database.
Objectivity	A multi-threaded database engine that claims exceptional performance and scalability
Gemstone	A database interface for Smalltalk, and offers integration with Oracle.

common access standard, unlike SQL for relational databases, places reliance on the use of host languages.

Each of these operational data storage methods share one essential ingredient, they are all capable of storing and supplying data. Thus they are all potential data sources to the data warehouse.

Please note that this section does not include multi-dimensional databases, which are one of the mechanisms for delivering data from the Data Warehouse onto the user desktop.

Data Warehouses

Data Warehouses are capable of holding granular and historic data. In the Data Warehouse context granular data is the lowest level of detail retained within the warehouse. Granularity is discussed in detail in later chapters. Historic data is held in a data warehouse for two primary reasons:

1. To build a deeper understanding of the functioning of the corporation over a long time span. This may include the ability to forecast future events of the basis of past history, or
2. To meet statutory demands.

Data Warehouses can become eternally large. Until the advent of the Warehouse the terms terabyte and petabyte had no practical usage. They were term only known to academics and research scientists. Data Warehouse adoption is leading corporations to adopt large servers with the capability to store very large volumes of data for long periods of time.

Normally historic data is summarised according to business rules. E.g. monthly Sales Volumes are retained, not every order and order line. The amount of historic data that must be kept will depend on specific business requirements.

It is important to understand that the Data Warehouse is NOT an operational archive. It is specifically formatted for business intelligence, and may not retain all values present in operational systems. It retains a history, but the purpose of this is to facilitate comparative analysis over time. It should be the primary source for all business intelligence activities within the organisation. The Data within the Warehouse is an important corporate resource that must be effectively managed.

When the Data Warehouse should be used for Reporting

Whilst there are many demands for reports to be produced within day-to-day business activities the majority of these requirements should be met through the operational reports available within the applications such a Oracle Applications, SAP etc. The Data Warehouse is aimed at supporting management decisions within the company and should only be used when one or more of the following conditions are met:

- ❑ When regulatory reporting is demanded.
- ❑ For cross business reporting

- ❑ Where the subject matter is to be included within KPI measures
- ❑ Where historical snapshots are required
- ❑ Time series analysis is required

The warehouse will become the source for all business intelligence activities within the organisation. As shown in *Figure 2.1* the data warehouse gathers a copy of transaction data used for generating management and statutory reporting and information retrieval and analysis.

The equations shown in *Diagram 2.6* are the common business drivers behind the information gathering exercise that is at the heart of creating the Data Warehouse:



and

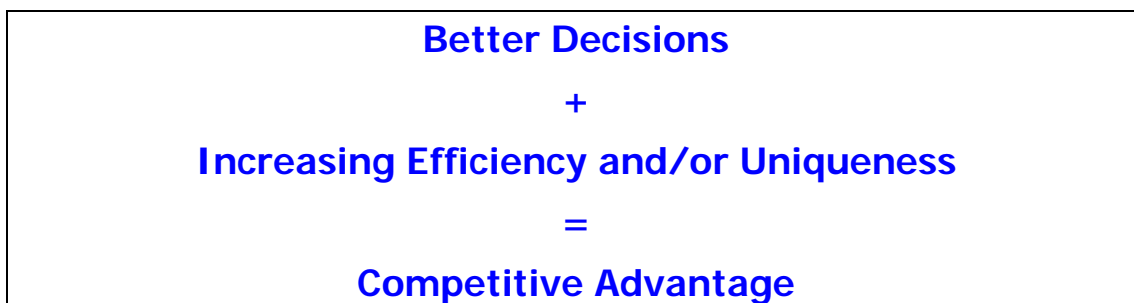


Diagram 2.6 – Business drivers for the Data Warehouse

The aim of the warehouse is to empower the business community with mission critical information at the time when decisions need to be taken. Each of the elements described in this diagram are discussed throughout the remainder of this chapter

Information

Information is a competitive resource that is in increasing demand in the business environment today.

Information is made up from data, but data is not information, Information, when looked at in relation to a Data Warehouse is much more than a collection of data. Information has three sub-components, the past, the present, and the future.

The task of turning data into information relies on the following primary steps

- ❑ Extract relevant data from all relevant systems
- ❑ Transport data to the warehouse platform
- ❑ Stock source data tables
- ❑ Validate source data
- ❑ Identify timeliness of source data

- ❑ Assign consistent, repeatable keys
- ❑ Update dimensional attributes
- ❑ Update base measure attributes
- ❑ (Calculate derived measure attributes)
- ❑ Aggregate data using dimensional structure and business rules

Information is therefore processed data, associated with the structure of the organisation, its products etc.

One of the keys to generating information is tying data to a timeline. All data items will relate to a time period. Legacy systems cannot be relied upon to record when events happened; it is therefore the duty of the Data Warehouse to identify this. Consider the questions posed in *Diagrams 2.1* and *2.2*, these are typical problems faced.

Within the context of the Data Warehouse project it is necessary to recognise that information is a resource that must be managed effectively. Data cannot be an afterthought; it is the foundation for the project and must be effectively managed. This is the basis for the role of Information Resource Management (IRM) within the team. Data quality and integrity are also central to Information Resource Management.

The objectives of IRM are:

- ❑ Planning and managing the organisation's data.
- ❑ Documenting the data requirements and business rules necessary to support the information needs of the organisation.
- ❑ Defining an information architecture as part of the corporate IT architecture.
- ❑ Establishing an organisational structure to ensure effective data, quality and integrity management.

Diagrams 2.7 and *2.8* show some problem scenario's concerning periodic (or time based) information. These are both based on real world scenarios encountered by the author.

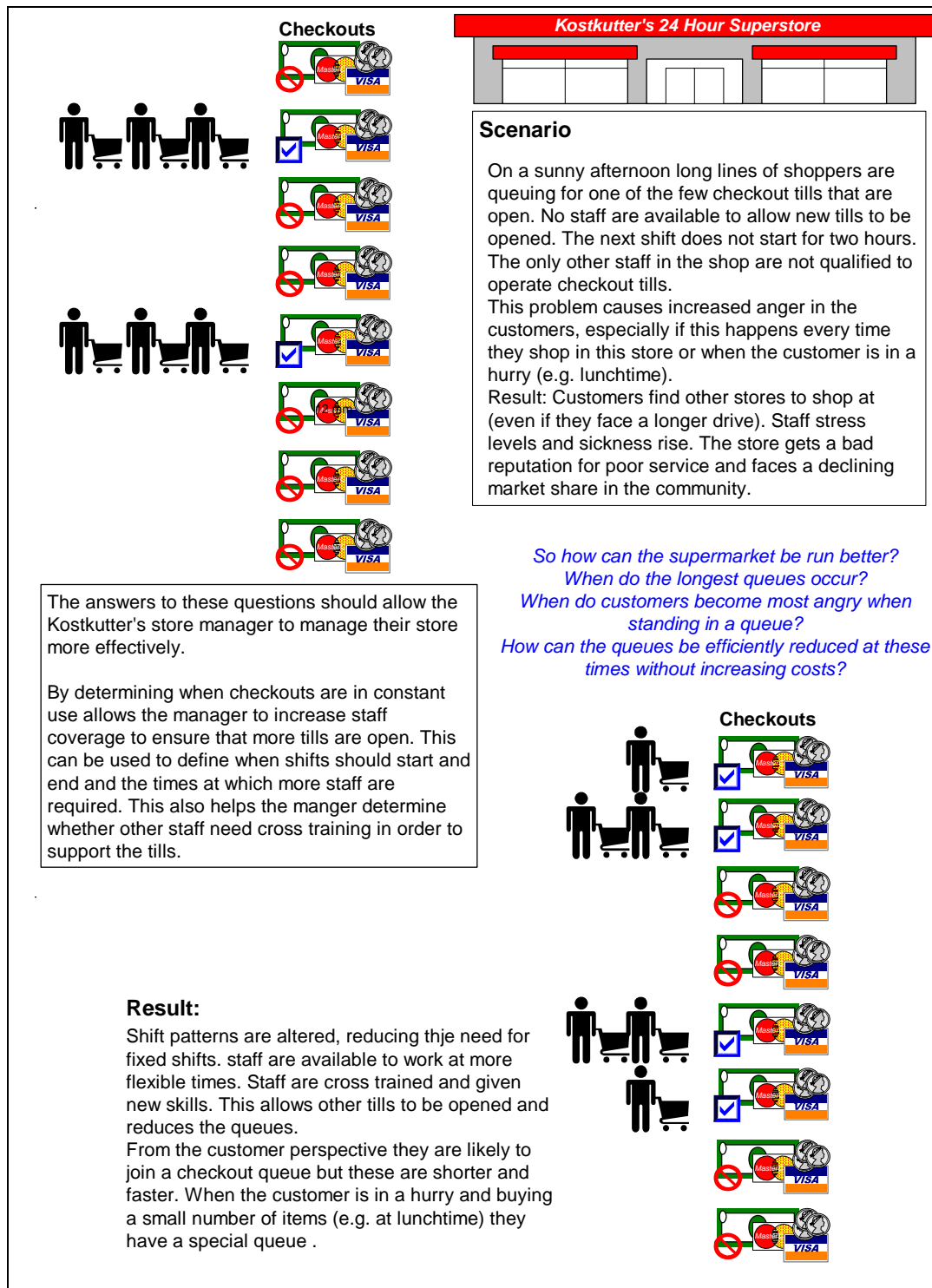


Diagram 2.7 – Kostkutter's 24 Hour Supermarket

Kostkutters 24 Hour Supermarket required a range of information about supermarket queues. Certain data they could ascertain from their own in-store data, e.g. frequency of use of checkout's, the number of items each purchaser bought, Whether the transaction was paid by cash, cheque, or credit/charge card. Other data, for example the mood of the shoppers would only be available through the use of surveys.

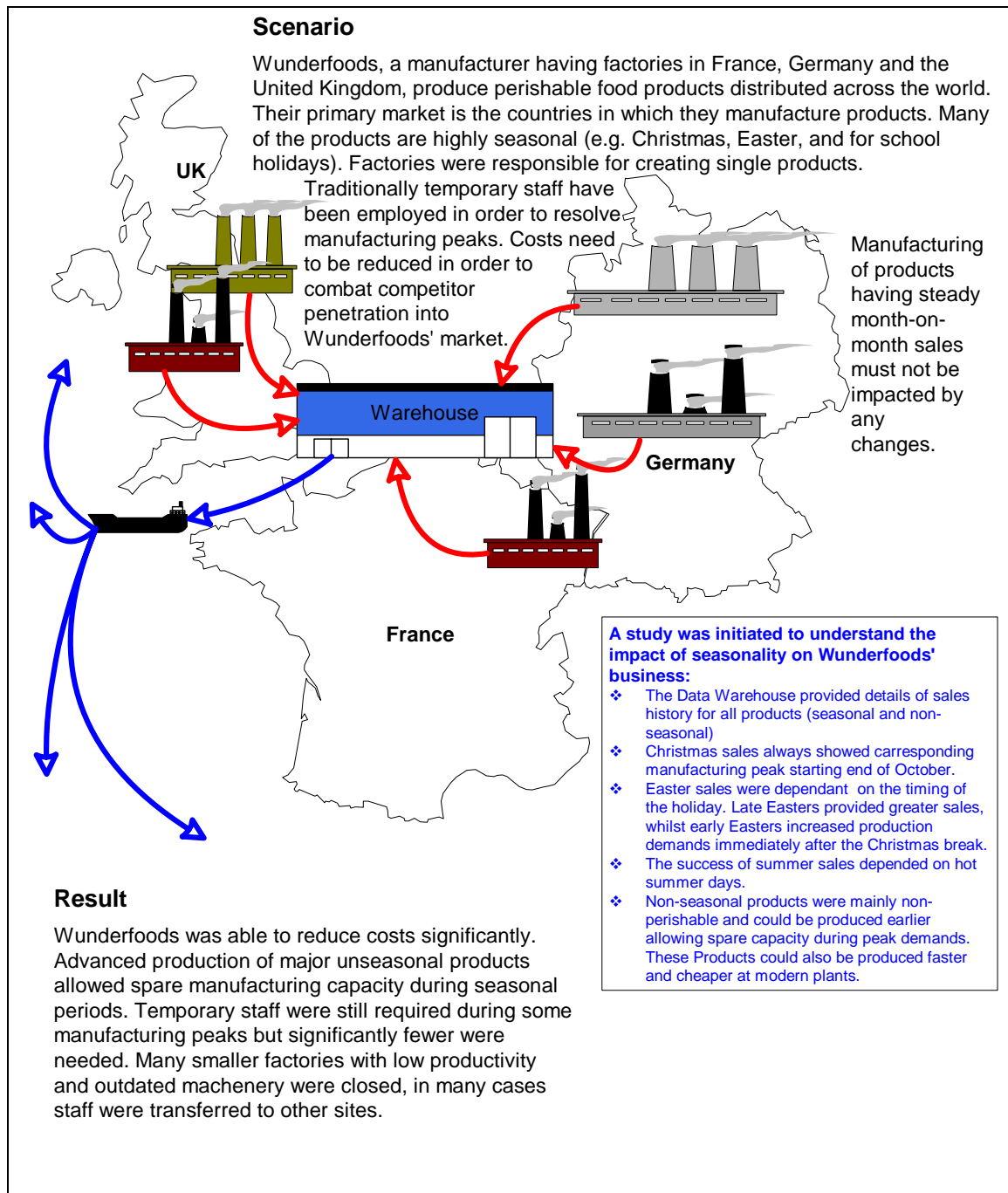


Diagram 2.8 – Wunderfoods Manufacturing

A component of turning data into information is the association of measurements to time. From the examples shown in *Diagrams 2.7* and *2.8* it is possible to see how time periods are an important factor in the decision making process. With the supermarket it is the time of day and day of the week that matters, whereas for manufacturing annual seasonality is important. Companies operating in the same industrial sector may have common characteristics.

Historic Information

Historical information could be seen as the total of all history (potentially including yesterday). Normally historical information will be seen as data that is not part of any

current time period. Historic data is normally heavily summarised, and is rarely kept at a granular level.

The amount of historic data to be kept will depend on the business being analysed. Some industries, like Banks, are governed by strict statutory regulations enforced either by the national governments or by international organisations. Before starting the Data Warehouse project it is essential to have an idea about how much history is required.

Today's Information

As soon as the project starts processing data from source systems it will become obvious that the warehouse rarely stores any information about today. If today's information were actually stored within the Data Warehouse then it would be subject to change during the continuing business day.

In Data Warehouse terms today's information is the current period, the current day, week, month, quarter, year etc. The term "Current Period" has special significance within many OLAP reporting tools and can be used in order to determine reporting periods.

The current period is the most recent period within the Data Warehouse. In designing the database it will be necessary to decide on levels of granularity and the longevity of the data stored.

Future Information

So we have built our Data Warehouse and filled it with current and historic data, so where do we get future information from? The Data Warehouse is the natural home of forecasting and budgeting applications.

Analytics & Business Intelligence

The population of the Data Warehouse is simply the enabler to provide analytical information, the Starting point for Business Intelligence (BI). Analysis is not simply the production of weekly or monthly reports. It is at a very minimum taking reporting outputs, identifying trends and measurements of key business importance.

Analysis is defined in the Oxford Dictionary of English as:

Detailed examination of the elements or structure of something, typically as the basis for discussion or interpretation.

and:

The systematic and critical assessment by an organisation of a feature of a product to ensure that its cost is no greater than necessary to carry out its functions.

Within the context of the business intelligence application analysis is the discovery of key performance indicators through the use of the information stored within the Data Warehouse.

Analytical use of a Data Warehouse relies upon the use of Business Intelligence or On-Line Analytical Processing (OLAP) tools. These are reporting tools with the ability to navigate through the dimensional structure and show the impact of important measures contained within it. Analysis of the information contained within the

warehouse will contribute to the decision making process of the organisation. This is supported by the ability to drill down for more detail or drill up to higher summary levels.

Data Mining also aids analysis of the data. Data mining tools are a specialist type of analytical tool that provides the capability of applying mathematical or statistical processes for the purpose of extracting hidden knowledge within large data sets. Data mining techniques provide information inherent through patterns or relations within the data. This data can be used to make predictions about future behaviour based upon learning algorithms. Data Mining will be applicable once the organisation has built up sufficient validated history within its Enterprise-wide Data Warehouse.

BI Supports Better Decisions

Improving the decision making process is always the goal of any corporation. The term "Better Decisions" in the context of *Diagram 2.6 (repeated below)* means using the information contained within our Data Warehouse and applying relevant analysis in order to make informed decisions about the business. The decision is of course still a human action, what differs is that the person making the decision is making decisions based on facts, not on guess work.



The types of decisions made will differ according to the perspective of the decision maker. *Diagram 2.9* looks at the differences between operational, tactical and strategic decision making. Each has a place within the corporate decision structure and must be supported by the Data Warehouse.

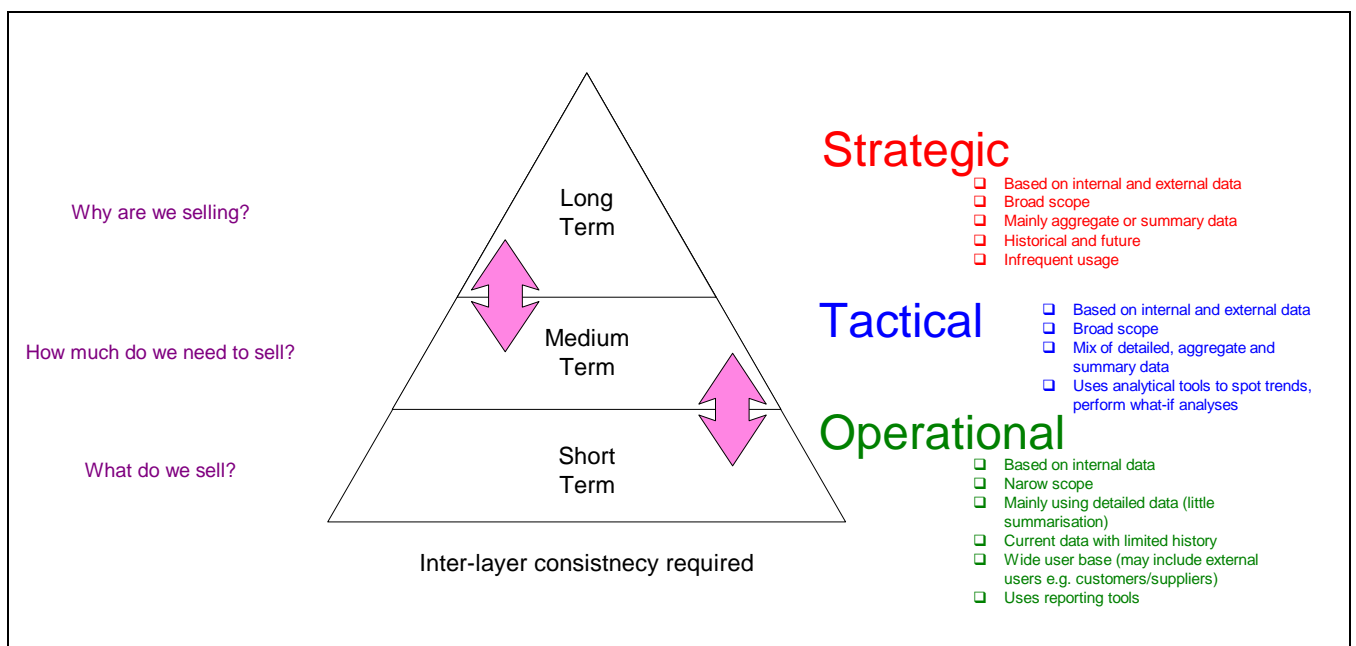


Diagram 2.9 – Operational, Tactical and Strategic Decision Making

Operational decisions are based on effective and efficient use of facilities and resources to carry out activities within budgetary constraints. Operational decisions are based on questions that include: What do we sell? Where is it sold? Who did we sell it to? The answers to these questions will generally be based on detailed (highly granular) data. Operational decisions are by their nature responsive to a changing environment. There will be a need for periodic reporting, such as the 'Monthly Sales Report by Salesman' but this is accompanied by ad-hoc reporting with new reports generated when they are required. Today's demands will be different than next month's queries. Operational reporting tends to have limited summarisation.

Tactical decisions are based on acquisition of resources, tactics the establishment and monitoring of budgets. They deal primarily with past events. Analytical tools are required to understand and act on the day-to-day changes in market demand, variations in environmental factors that may be unique to a specific industry or area of competition. Much of this type of analysis will be trend (or time series) based with a need to monitor how factors change over time, particularly with reference to industrial norms or in relation of specific competitors.

Strategic decisions are based on corporate goals, policies and determination of organizational objectives. Strategic decision making is about answering the following questions: What is the overall health of our business (now and in the future)? Are all the disparate parts of our organisation working towards the same goals? How many new customers will we have next quarter/year? How many will we lose to the competition? What new products/services will we introduce? What impact will technology, government regulations and customer patterns and preferences have on our industry?

A Data Warehouse will provide the information that enables these questions to be answered via quantitative methods and analysis. Often this is provided by a combination of internal data and external research. The business will take the information contained within the Data Warehouse, analyse it, and use it to implement new goals, which in-turn require monitoring and further modification. This is a performance cycle that requires continuous monitoring for sustained success.

Increased Efficiency

Increasing efficiency has been a long term corporate goal. All pay negotiations taking place today focus on increasing productivity, yet in the past productivity has been difficult to measure and has been done very much on a hit-and-miss basis. The use of data within a Data Warehouse provides corporations with the opportunity to measure organisational activities more effectively, especially when combined with applications such as Activity Based Costing (ABC).

Activity Based Costing is discussed as an application later within this book.

A few areas where corporate efficiency may apply:

- Responding to budgetary constraints
- Increased import pressure as well know international corporations enter the local market
- Increased competition within the marketplace. This may stimulate productivity improvements
- Effects of privatisation on publicly owned corporations.

There are many theories about increasing corporate efficiency. It is not the place of this book to comment on those, it will be for the reader to put into practice whatever method they deem applicable. The author's is simply identifying the role of the Data Warehouse in driving corporate efficiency on the basis of measured facts and not on guesswork.

Uniqueness

Uniqueness is a rare business quality. The majority of businesses start life with the perception of having a unique product or service. But as the company grows their product or service becomes less distinctive. Sometimes the company name will also have a special place in the heart of many of its customers, which may extend the distinctiveness of the product or service offered.

However as we know business conditions change, particularly after patents and other intellectual property rights expire. Competitors enter the market with similar products or ones that are better than the original. Maintaining uniqueness is a difficult challenge, but a task that is made easier through the introduction of informed decision making.

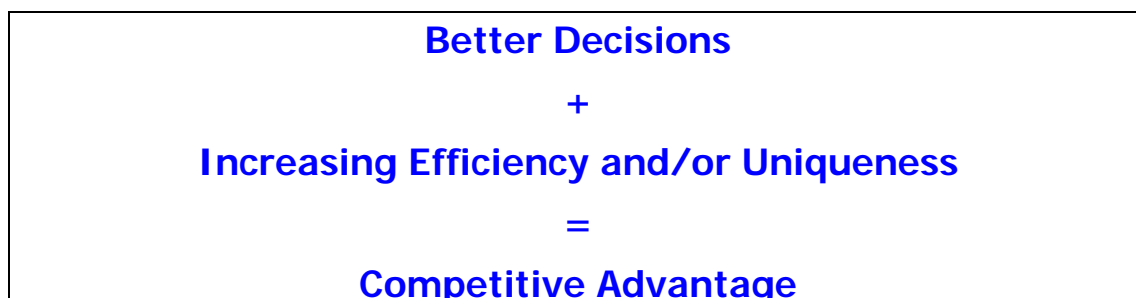
Even where products/services retain their uniqueness improvements to efficiency are equally important in building a competitive advantage.

Competitive Advantage

Corporations are faced with the constant challenge of reinventing themselves due to either corporate takeovers or adapting to the competitive market place. In order to react to these needs management need to:

- Have an in-depth knowledge of their company's operations
- Know how key business factors affect the organisation
- Know how business factors change over time
- Understand and analyse their organisation's performance relative to their key competitors or industry benchmarks.

In order to adapt corporations will have already been using Data Warehousing and Decision Support technology to measure what they are doing and will know the importance of corporate information. Gaining a competitive advantage (as shown in *Diagram 2.6* and repeated here to aid the reader) requires using the knowledge gained from the Data Warehouse, identifying where processes can be made more efficient or unique in order to gain a competitive advantage.



Key business factors will differ from industry to industry from sector to sector from corporation to corporation but the need to measure and quantify results is constant in every organisation.

Flexible solutions are necessary, to allow the company to re-invent itself on an ongoing basis in order to beat the competition. Corporations are therefore becoming more flexible with a cycle of change involving:

- Performance assessment based on information & analysis.
- Informed decision making.
- Training of managers, team leaders and team members to successfully implement new policies and working practices.
- Ownership of business events (including data quality)
- Identification and implementation of efficiencies within the business.
- Identification and enhancement of uniqueness.
- Continuous improvement to ensure the competitive advantage is maintained.

A Data Warehouse will provide quantitative methods and analysis to enable informed decision making. The business must take the information contained within the Data Warehouse, analyse it, and identify increased efficiencies within corporate processes of increase the uniqueness of the product/service offered. Alongside the corporate change is the need to ensure that staff at all levels are involved in the change process and understand the reasons for it.

Industry and Statutory Demands

In addition to the corporate reasons for creating a Data Warehouse, outlined earlier within the chapter, Governments and many industry sectors bodies are demanding tighter controls over what reporting is available and how data should be made publicly available. Additionally public companies are required to report profit and loss figures to the relevant stock-exchange within a limited time-frame at the end of each trading period. The Data Warehouse can aid with each of these reporting requirements, but in order to achieve these demands it must be flexible and provide alternative reporting structures.

The Data Warehouse is therefore required to solve internal as well as external reporting requirements. Thus we are immediately aware of multiple reporting structures that are required to work with the same data set. These demands often cause problems in a poorly designed Warehouse. The corporation must be flexible in its reporting methods to meet each of these demands.

To build a separate warehouse to meet each of these demands would waste development time and storage resources. The design of the warehouse will need to take each of these needs into consideration. The following sub-sections shows a few examples from a few industries. This is intended a flavour of things to look out for. Inevitably this list will be out-of-date as before this book is published. It is not possible to provide a full list of regulatory demands and you will find that these change as you are working on your project. The reader will have to research the applicable rules during the course of their own project, and make sure they keep up-to-date.

General Demands

Recent high profile cases involving accounting fraud have led to tougher reporting controls for companies listed on stock exchanges. In the USA this includes senior company officers taking responsibility for company reports combined with tighter disclosure deadlines. Corporations must therefore be in a position to speed of the processing of periodic (e.g. half-year and financial year) reports. A Data Warehouse integrated with a general ledger system provides the means of achieving this goal in a fast and accurate manner.

Various Wall Street accounting scandals led to the introduction of the Sarbanes-Oxley Act during 2002. This demands more open accounting procedures that are visible to both investors and regulators. This applies to all public companies that are traded on American stock markets. A recent survey indicates that 85% of US firms believe that this act will require changes to their IT infrastructure. If the Corporate Data Warehouse is at the heart of corporate reporting then the impact of these changes should be significantly reduced, if not eliminated.

Government bodies collect statistics about company performance this includes monthly turnover, employment, investment and other measures. In the UK its National Statistics Office requires information to be returned by the 7th of the following month (e.g. January's data must be returned by the 7th of February) or the company may face financial penalties. The corporate Data Warehouse is the natural home for such figures, which can be available as soon as month end processing has completed.

Finance

Anti-Money Laundering regulations, were adopted in the aftermath of September 11th 2001 by both the USA and European Union. These regulations are intended to make it more difficult for terrorist organisations to move money through the banking network. The measures were proposed by European Union Finance and Economics Ministers and the G-7 Finance Ministers. This is managed through the Financial Action Task Force (FATF) which has proposed forty recommendations for adoption by banks. Any automated solution demands the implementation of both a CRM and a customer focused Data Warehouse in order to trace suspect transactions.

The new Basel II Capital Accord that comes into effect during 2006 requires the linking of banking systems and reporting systems in order to manage risk. This is intended to address types of risk. Their aim is to motivate banks to continuously improve their risk management and measurement capabilities and avail themselves of the more risk-sensitive methodologies and thus more accurate capital requirements. Although the focus is primarily on internationally active banks this is likely to have general applicability. Reporting systems will almost inevitably have to change to accommodate the accord. Between two and five years of history will be required for full compliance.

Insurance

Insurance is a heavily regulated industry. The author is aware of a large volume of regulations within almost every area of statutory control, including international rules. Most states have regulators demanding information relating to directly written insurance, reinsurance relating to types of loss and claims. Some bodies require accounting for past losses and risk associate with late claims. I hope the reader will

by now have started to think about how Data Warehousing applications can aid the solution.

Conclusion

We started this chapter asking why it is necessary to build a Data Warehouse. This chapter was aimed to provide some food-for-thought, and not an exhaustive list of reasons to build the Warehouse. As you read further this book will discuss justifying the project, organisation of the project team, the solution development life-cycle, and adherence to legal rules and then the management of the project itself.

To answer the question of why it is necessary to build a Data Warehouse one should take a simplistic approach. Without discussing how a Data Warehouse will be structured and what tools to be used to create and manage it. There is one primary reason for creating a Data Warehouse, that is to separate reporting databases from transaction databases.

The separation of the two types of database allows each database to be managed differently. Transaction databases require the ability to rapidly insert and update small groups of records in response to user activity. Reporting databases need to read data, sometimes in large volumes and present it in the form of a 'report' to the user. The 'report' produced may be either output on paper or on screen.

Each type of database needs to be optimised to ensure its performance is adequate for the task at hand. For the operational system this requires rapid writing of a set of data (e.g. the purchase order and all of the products purchased). Database Administrators manage these databases to ensure performance criteria are adhered to, including the removal (archiving) of outdated data.

For the reporting database there is the need to collect data then load it. Data can be collected from multiple data sources. Part of the loading process may require its restructuring into a common format consistent with reporting needs or a corporate structure. Data is time-stamped, either at source or as part of the loading process, to allow monitoring of 'things' over time. Knowing past trends can aid the prediction of future trends through budgets and forecasts. Overall the key advantage of the Data Warehouse will be in its aid to the corporate decision making process allowing decisions to be based on measurable facts, and not 'gut-feeling'.

Information must drive the future of corporations. The Data Warehouse contains the information at the heart of the decision cycle.

¹ Bill Inmon, Building the Data Warehouse – Originally published in 1990 by John Wiley & Sons Inc.

² Ralph Kimball, The Data Warehouse Toolkit – 1996 John Wiley & Sons Inc.

³ Price Waterhouse, Data Warehousing (an overview of the benefits and the system management methodology – 1997).

⁴ Richard Tanler, The Internet Data Warehouse – 1997 John Wiley & Sons Inc.