About Topic Gateways

Topic Gateways are intended as a refresher or introduction to topics of interest to CIMA members. They include a basic definition, a brief overview and a fuller explanation of practical application. Finally they signpost some further resources for detailed understanding and research.

Topic Gateways are available electronically to CIMA members only in the CPD Centre on the CIMA website, along with a number of electronic resources.

About the Technical Information Service

CIMA supports its members and students with its Technical Information Service (TIS) for their work and CPD needs.

Our information specialists and accounting specialists work closely together to identify or create authoritative resources to help members resolve their work related information needs. Additionally, our accounting specialists can help CIMA members and students with the interpretation of guidance on financial reporting, financial management and performance management, as defined in the CIMA Official Terminology 2005 edition.

CIMA members and students should sign into My CIMA to access these services and resources.

The Chartered Institute of Management Accountants
26 Chapter Street
London SW1P 4NP
United Kingdom

T. +44 (0)20 7663 5441
F. +44 (0)20 7663 5442
E. tis@cimaglobal.com
www.cimaglobal.com
Data warehousing

Definition and concept

‘A data warehouse is a database in which information is held not for operational purposes, but to assist in analytical tasks such as the identification of new market segments. Data warehouses provide a repository for historical data, collecting, integrating and organising data from unintegrated application systems. The data stored in a data warehouse almost certainly comes from the operational environment, but is always physically separate from it.’

*CIMA Official Terminology, 2005*

Context

In the current CIMA syllabus students will learn and be examined on this topic in paper P6, Management Accounting Business Strategy. A study system for this paper is available from CIMA Publishing.

Related concepts

Data mining; business intelligence; database management systems (DBMS); relational databases; knowledge management.

Alternative concepts

Executive information systems (EIS); SEM.

Overview and comparison

The concept of data warehousing is aimed at addressing three sets of organisational requirements:

- the requirement for an organisation-wide view of information
- the requirement for the information systems department to manage organisational data effectively and efficiently
- the requirement to keep informational data separate
- from operational data to minimise the risk to operational data.

For many years, harnessing the value hidden in large quantities of data has been a high priority for big companies. It is considered that data harnessing will lead to better decision making and provide a competitive advantage.
A data warehouse is effectively a large relational database that collates a vast amount of data from many sources. It is also optimised for reporting and analysis. This is unlike operational data, where the databases are optimised for simplicity and data capture, for example, online transaction processes. In today’s ‘knowledge based’ company, information and its effective management have replaced tangible assets as the primary source of competitive advantage.

A data warehouse is made up of ‘bundles’ of data, each originating from a different source within the organisation. Most likely, each bundle of data originates from departmental databases accessed by staff within that department. This data’s use is primarily ‘operational’ and not ‘informational’. Once these bundles of operational data are combined, cleaned and configured, they form the corporate data warehouse.

In practice, data warehousing involves the storage of virtually all of the organisation’s data. This includes transactional data, master data (records about customers, staff, material, suppliers, etc.) and metadata. The latter comprises data about data, such as table names, dates of recent data loadings or the number of users logged on. Metadata can also hold data on what each table in the warehouse holds. This could include table column names and detailed descriptions of each table.

Essentially, a data warehouse forms the basis needed to achieve the overall benefits of advanced technology use. Data warehousing aims to optimise data management processes to support the exploitation of information creation.

At a basic level, every data warehouse comprises:

1. Data extraction and movement methods. This involves finding the source data from its current location in the business information systems/departmental databases. It also includes data extraction and preparing or ‘cleaning’ it for loading into the warehouse.

2. Connectivity software for client-server integration. These are front-end tools that can connect to a variety of back-end relational databases.

3. High performance relational database management systems (RDMSs). The RDMSs should be capable of handling large amounts of data, complex queries and analytic processing.
The term ‘data warehousing’ refers to the data importing and exporting components which are responsible for:

- accessing
- transforming
- distributing
- storing
- exporting the data and information.

From *Data Warehousing* (PDF 272KB), CIMA Technical Report

**Data warehousing in practice**

There are eight steps in the development of a data warehousing system:

**Step 1: The initial concept**

The most effective and efficient way to exploit business opportunities is via a data warehouse. Businesses may decide to invest in a data warehouse once they recognise this. They need to familiarise themselves with the capabilities and potential of the data warehouse as a business option. This step is taken prior to the more concrete evaluation of the financial (and other) costs and benefits of data warehousing in Step 2.
Step 2: Feasibility study

The primary objective here is to determine whether implementing a data warehouse would deliver a positive return on investment. The feasibility study should include an analysis of:

- business goals and objectives
- current data needs and usage that the warehouse will support
- examples of usage for each piece of data
- volume of each piece of data
- level of detail of each piece of data
- frequency with which data is required
- frequency with which data should be updated
- possible users, their location and level of expertise
- timeliness, or how current the information must be to be valuable to the decision making process.

Step 3: Designing the warehouse architecture

The design stage identifies facts about the data that needs to be stored in the warehouse (metadata) to support drill down (data mining) and drill across data inquiry. Design involves the following five steps in an iterative process:

1. Identification of the business area
2. Definition of the data content of each business area
3. Understanding of which data is important to decision making
4. Identification of the best data source
5. Development of a data warehouse model.

For each business subject area to be included in the warehouse, the system must consider the following information about the data:

The basic nature of the data. What type of data will be used? For example, will it be archival data or existing systems data?

Attribute mapping. This defines the data item characteristics or attributes. Agreement needs to be gained regarding what each piece of data is and how it should be displayed.
**The degree of summarisation.** Will single data items be mapped directly to the data warehouse, or will some degree of summarisation (e.g. monthly summaries) be performed?

**Frequency of transformation.** How often will the transformation of data occur?

**Occurrence mapping.** Will all or only selected source records result in the creation of a data warehouse record?

**Volume.** How much data will be maintained in the warehouse? How often will source records be scanned?

**Sequence.** What is the sequence of selected or summarised records passed to the warehouse?

**Step 4: Developing the warehouse**

The physical environment of a data warehouse is critical to its successful implementation. It is one of the biggest technological challenges of data warehousing projects. The architecture supporting the warehouse must provide sufficient information access speed for users. Such needs should be balanced with a hardware strategy that allows for anticipated growth in usage and data needs. The physical architecture of the warehouse includes all:

- hardware
- network configurations
- security requirements
- end user tools
- tangible pieces of the data warehouse.

**Step 5: Assembling and testing the warehouse**

Data from the existing sources may require (format) modification or cleansing (passing through a software filter). This is to eliminate inconsistencies or invalid (redundant and inconsistent information) before data can be placed in the data warehouse. The warehouse should be tested by the business users of the targeted data, who understand its usage and who have realistic expectations of inquiry results.
Step 6: Implementing the warehouse

All hardware must be in place before activating the data warehouse. Users must be trained in accessing their data. The software for cleansing, distributing and partitioning the data must be working accurately.

Step 7: Maintaining and enhancing the warehouse

It is both a business and a technological requirement to ensure that the data warehouse is adequately maintained. It is useful to collect and analyse information about the following to determine if any performance enhancements can or should be made:

- the nature of enquiries to the warehouse
- the data being accessed
- response times.

After a successful implementation, it is important to monitor and continually review future enhancement plans in the light of reported relationships, causalities and patterns of usage.

Step 8: Obsolescence

A data warehouse becomes obsolete when it no longer provides value to the business community it serves, or when maintenance costs exceed the return on investment. Out-dated technology should never be the reason behind obsolescence. Platforms can be changed for a price, and new tools purchased to provide better performance or usability.
Application

**Presentation of standard reports and graphics.** The data warehouse allows data from different transaction systems to be consolidated into the warehouse and used in reporting.

**Support of dimensional analysis and query reporting.** The data warehouse can be used to facilitate the comparison of results across different dimensional values, especially time periods.

**Creation of business intelligence.** The data warehouse is a critical element in the creation of business intelligence. This is through the use of analytical application systems (scenario modelling, planning and budgeting, data mining, balanced scorecard, etc).

**Uses of data warehouses**

Many organisations and industry sectors exploit their data warehouses most effectively by use of data mining techniques. Data mining may be defined as the process of selecting, exploring and modelling large amounts of data to reveal previously unrecognised patterns for business advantage.

Data mining allows companies to unlock the value of information and boost efficiency and productivity. Such tools can be applied to the information stored in the data warehouse. This enables the analysis of volumes of data too large and complex for cost-efficient analysis by the human brain. In turn, unanticipated relationships between apparently unconnected variables may be detected. The result is the creation of ‘new’ information which suggests unexplored but potentially profitable business growth.

Although the insurance and finance industries have already recognised the benefits of data mining, this technology can be effectively applied to many areas:

**Retail/marketing:**

- identifying buying patterns from customers
- finding associations among customer demographic characteristics
- predicting which customers will respond to mailings.
Banking:
• detecting fraudulent credit card use
• identifying ‘loyal’ customers
• predicting customers that are likely to change their credit card affiliation
• determining credit card spending by customer groups.

Insurance and healthcare:
• determining which medical procedures are claimed together (claims analysis)
• predicting which customers will buy new products
• identifying behaviour patterns of risky customers.

Transportation/logistics:
• determining the distributions schedules among outlets
• analysing loading patterns.

Medicine:
• characterising patient behaviour to predict office visits
• identifying successful medical therapies for different illnesses.

Reported benefits
There are a number of reported advantages of using a data warehouse, including:
• enhancing end user access to a wide variety of organisational data
• combining data from different sources in one place, thus making decision making more effective and efficient
• allowing for a better understanding of the business
• improved targeted marketing
• increased productivity and reduced technology costs
• increased data consistency
• improved data quality
• increased trust from users that they are viewing 'good data' that can be trusted.
In addition, the organisational decision makers are able to obtain trend reports in all areas of business such as sales by area or by product. This information may be of help in targeting future investment.

**Reported Drawbacks**

- Extracting, cleaning and loading of data can be very time consuming and costly.

- There may be problems of compatibility between the existing transaction processing systems.

- The scope of a data warehousing project might increase beyond the capability of the organisation.

- Security can be a very serious issue especially if the database is internet based.

**Further information**

**Articles**

Full text from Business Source Corporate available from:  
[www.cimaglobal.com/mycima](http://www.cimaglobal.com/mycima)  
[Accessed 14 November 2008]

Corbitt, T. *The power of data: mining and warehousing. Credit Management*, April 2006, pp 32-33


Murby, L. *Data warehousing: a new management opportunity for management accountants, CIMA Insight*, February 2003


**Articles**

Abstract only from Business Source Corporate available from: [www.cimaglobal.com/mycima](http://www.cimaglobal.com/mycima) [Accessed 14 November 2008]


**Other recommended articles**

Evans, H. and Ashworth, G. *Survey conclusion: wake up to the competition!* Management Accounting, May 1996, pp 16-18

Sheridan, T. *The changing shape of the finance function.* Management Accounting, February 1998, pp 18-20

**Books**


**Case studies**

Two case studies are cited in the CIMA Technical Briefing, Data warehousing: its role in improved organisational decision making. The first reviews how Capital One, the American credit card company, uses data mining within its global data warehouse. The second case study analyses how Wal-Mart uses data warehousing to create savings through more effective supply chain management.
A number of further case studies from a variety of industries are cited on the Data Mining Technologies website, including the utilities, telecommunications and biotechnology industries. Available from: www.data-mine.com [Accessed 14 November 2008]

Another useful website containing a number of case studies is the Computerworld web site. Available from: www.computerworld.com [Accessed 14 November 2008]

**CIMA publications**

Cobb, I. (2001). *Data warehousing and the management accountant.* (PDF 93KB). CIMA Research Executive Summary. London: CIMA. This executive summary documents a research project that sets out to explore the implications of data warehousing for management accountants and management accounting. The research suggests that management accountants are actively involved in shaping the direction of data warehousing in their organisations. Available from: www.cimaglobal.com/researchexecsummaries [Accessed 14 November 2008]


**Websites**

Computerworld  
Provides a good range of case studies highlighting success stories in business intelligence and data warehousing.  
Available from: [www.computerworld.com](http://www.computerworld.com)  
[Accessed 14 November 2008]

Datawarehousing.com  
An excellent glossary of terms for data warehousing, and includes company links, white papers and articles.  
Available from: [www.datawarehousing.com](http://www.datawarehousing.com)  
[Accessed 14 November 2008]

The Data Warehousing Information Center  
A collection of one practitioner’s essays on data warehousing, decision support and business intelligence. It is very useful to provide a simple introduction to data warehousing.  
Available from: [www.dwinfocenter.org](http://www.dwinfocenter.org)  
[Accessed 14 November 2008]

Kimball Group  
A collection of business articles sorted by year, including articles on data warehousing and data mining.  
Available from: [http://digbig.com/4xwpq](http://digbig.com/4xwpq)  
[Accessed 14 November 2008]