Chapter 1

Introduction

1. Background

In today's highly competitive business markets where survival of the fittest is considered the norm, business management seek tools and new strategies which allow them to stay one step ahead of their competitors. Decision support systems are yet just another tool identified by management teams in attaining this objective (Weber, 1982: 117). This evolving technology allows for the cultivation of information-sharing within organisations thereby enabling employees to solve dynamic problems and reduce costs (Warigon, 1998: 55).

The ability of decision support systems to be scaled down or up to meet their user's needs is another primary reason why decision support technology not only favours large conglomerates (Murphy, 1997: 1). Users processing a hundred to a thousand transactions a month derive just as much benefit from the decision support system as a multi-national processing millions of transactions a day. These systems may be heavily relied upon by management in deriving strategic and operational management decisions. Therefore both internal auditor and systems developer have significant responsibilities to ensure that management are aware of the consequences should continuity of operations or the quality of data be jeopardised (Curtis & Joshi, 1997: 40).

In a survey conducted by a South African computer magazine (Du Plessis, 1998: 1), data warehousing was identified as one of the top seventeen application and system areas which information technology specialists acknowledged were crucial to the survival of their organisations. In addition, another computer article (Anon, 1998: 22) cited the data warehouse market being worth $1.47 billion in 1996 with an estimated growth to 1998 of 28%. The article went on further to indicate that the increased demand was for applications that enabled organisations to access, analyse and report data.

Internal auditors seek to align themselves with the organisation's major objectives and focus on adding value to the business process. They therefore have a unique responsibility to focus on what management consider critical to the organisation's success and overall survival (Ridley, 1996: 24). If management considers the data warehouse environment to be a significant mainstay within their operations, internal auditors own the responsibility of ensuring that they are able to advise management concerning the internal control risks within such an environment.

2. Purpose and reason for study

2.1 Purpose of study

The purpose of this study is to identify the most pertinent internal control risks within the data warehouse environment. The study will also suggest suitable internal control considerations which the internal auditor can apply in assessing the extent of such internal control risks. Finally, the impact of future developments within the data warehouse environment on the assessment of internal control risks will be considered.

2.2 Reason for undertaking the study

Based on an extensive evaluation of audit resource materials, it would seem that little attention has been given to the data warehouse environment by the internal audit profession. There has also been a lack of focus on what impact this evolving technology will have on the assessment of internal controls[1]. The study will address internal control risks at the following stages:

- Development phase of the data warehouse.
- The established data warehouse environment.
- Dependent data mart.
- Distributed data warehouse environment.
- Future developments within the data warehouse environment.

[1] *Assumption based on the writer's personal evaluation of internal and external audit source materials and the results of extensive world wide web searches relating to the audit of the data warehouse environment.*

3. Defining the data warehouse environment and internal control risk

The following section provides an indication of how the data warehouse forms part of the decision support environment. It also describes the fundamental principles of the data warehouse. As an introduction to the remainder of the study, this section will conclude by identifying the overall internal control risks that can be expected within the data warehouse environment.

Since internal control risk is the central theme of this study, we will first give attention to defining internal control risk.

3.1 Defining internal control risk

Internal control risk is the risk that management's plans, organisation and associated procedures will not provide reasonable assurance that the organisation's goals and objectives will be achieved (IIA, 1983). We will also rely on COBIT's (ISACA, 1998) information criteria identified by the Information Systems Audit and Control Association as a method of categorising risk exposure to the organisation. These criteria are:

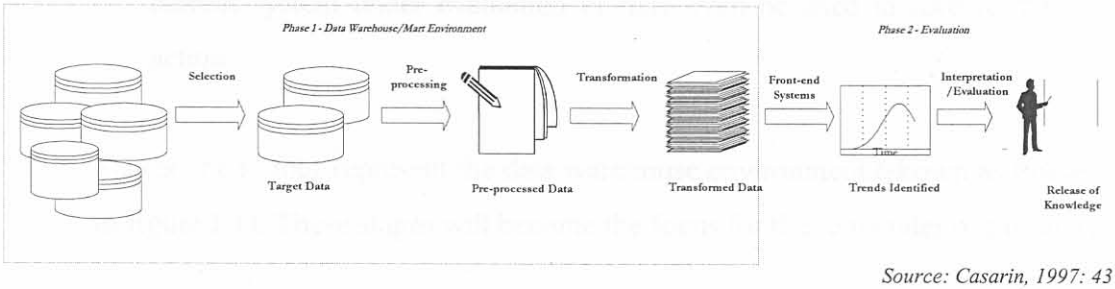
- *Effectiveness:* Information being relevant and pertinent to the business process as well as being delivered in a timely, correct, consistent and usable manner.
- *Efficiency:* Provision of information through the optimal (most productive and economical) use of resources.
- *Confidentiality:* Protection of sensitive information from unauthorised disclosure.
- *Integrity:* Accuracy and completeness of information as well as to the validity in accordance with business values and expectations.
- *Availability:* Information being available when required by the business process now and in the future. It also concerns the safeguarding of necessary resources and associated capabilities.
- *Compliance:* Complying with those laws regulations and contractual arrangements to which the business process is subject, i.e. externally imposed business criteria.

- *Reliability*: Provision of appropriate information for management to operate the entity and for management to exercise its financial and compliance reporting responsibilities.

3.2 Interaction with the decision support environment

It is important that the internal auditor understands how the decision support environment is structured so that he/she will be able to effectively identify and assess internal control risk.

Figure 1.1 - Overview of the knowledge discovery in databases process



Source: Casarin, 1997: 43

A process referred to as the Knowledge Discovery in Databases was developed to fully explain this environment. Figure 1.1 above provides a brief overview of the Knowledge Discovery in Databases process (Fayyad, 1996: 27-34).The Knowledge Discovery in Databases process is split into five individual stages (Casarin, 1997: 43-46):

- *Selection*

  This stage involves identifying data which is needed by management to aid in the decision making process. It also identifies what decisions will be made in utilising this data.

- *Pre-processing*

  Once data is identified, the cleaning of data is performed. This is done to remove irregularities or inconsistencies which may render the data unreliable for use by management. Possible irregularities could include, missing data fields, invalid characters loaded and even duplicate records.

- *Transformation*

Data is reduced to a uniform source which is consistent in all aspects and which is considered reliable in making management decisions. This stage usually involves relocating cleansed data to a separate database.

- *Information access layer systems*
  These systems extract the uniform data and present it in a format which will aid users in either confirming or disproving their hypothesis. A large number of systems exist, each providing their own unique capabilities.
- *Interpretation and evaluation*
  Interpreting presented results creates a higher level of user knowledge. Interpreted results or newly gained knowledge can be incorporated into the current system under evaluation or may even be used to take formalised action.

Stages one to four represent the data warehouse environment (shown as Phase 1 in figure 1.1). These stages will become the focus for the remainder of the study.

3.3 The data warehouse and its concepts

A data warehouse is an architecture for organising information systems. This technology has stemmed from repeated attempts by various researchers and organisations to provide businesses with flexible, effective and efficient means of obtaining sets of data. These sets have come to represent one of the organisations most critical and valuable assets (Gupta, 1997: 15). W.H. Inmon, acknowledged as the father of the data warehouse concept, defined a data warehouse as a "subject-oriented, integrated, time variant, non-volatile collection of data in support of management's decision making process" (Inmon, 1996: 33).

This collection of data could contain both highly detailed and summarised historical data relating to various processes within the organisation. Data is stored in time frames (e.g. month, years, etc.), so that trends such as market inclinations can be spotted and production accordingly amended. The following key concepts are raised in light of Inmon's definition of a data warehouse (Inmon, 1996: 33-37):

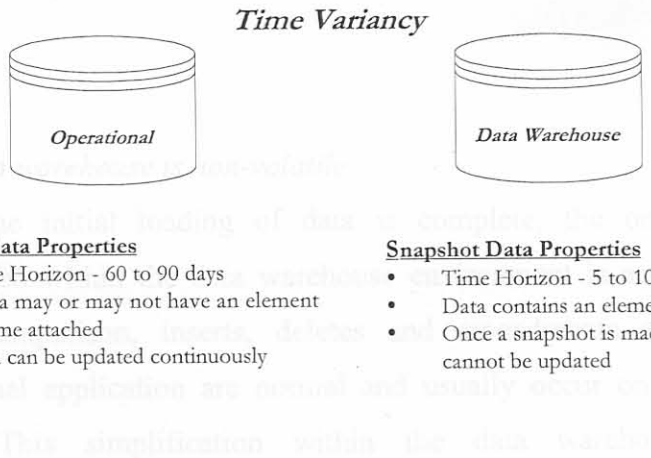
- *The data warehouse is subject orientated*

  Data stored within a data warehouse is physically separated from its source. The source which is operational in nature could include data relating to anything such as aviation records, hamburger sales or even telephone calls placed by subscribers. By nature the data warehouse supports the organisation's core functions (i.e. what the business does to derive income), and is not suited to provide information on ancillary processes within the organisation. Because of this orientation, the design and implementation of the data warehouse is affected most greatly by the major core functions and to a lesser degree by smaller, less prevalent subject areas.

- *The data warehouse has integrated components*

  All data stored must be integrated to ensure the success of the data warehouse. Naming conventions, measurements of variables and encoding structures must be consistent and no variations should be able to be entered into the data warehouse's repository. An example of the possible variations which could result in incorrect source data include the use of "M" and "F" to denote gender, while in other source applications the use of "x" or "y" may be applied. Integration affects almost every aspect of the data warehouse development. As we will see later, this issue has caused major concern not only for the internal auditor but also senior management. They run the risk of placing reliance on faulty information and trends.

Figure 1.2 - The issue of time variancy

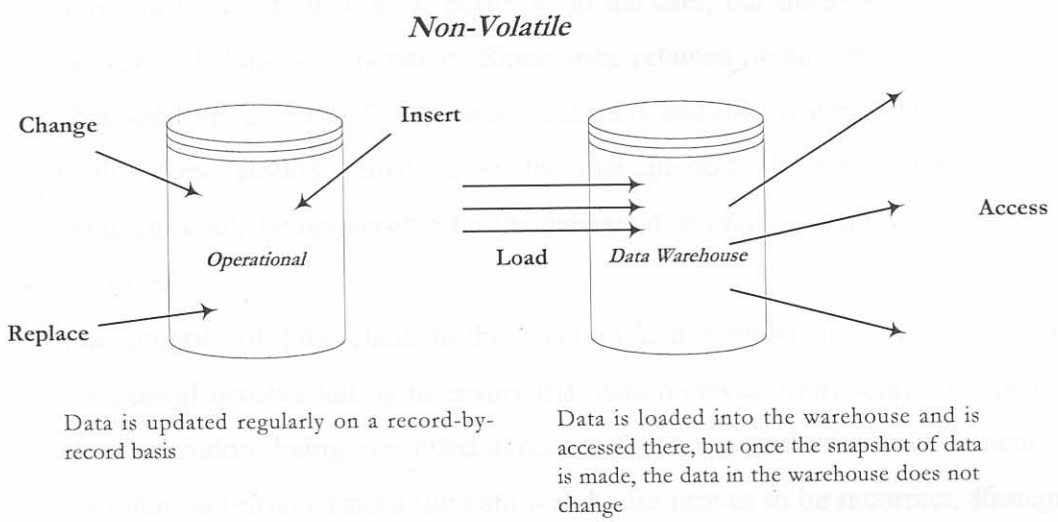


Source: Inmon, 1996: 37

- *The data warehouse is time variant*

  As detailed in Figure 1.2, the data retained within a data warehouse spans over a long time horizon - anywhere from five to ten years. Conversely in traditional applications, current values cover nothing more than sixty to ninety days (this is considered sound design for applications). Every group of data retained within the data warehouse includes some element of time, such as day, week, month, etc. This time element may even be implicit in nature. Many of the information technology professionals refer to the data warehouse as a long series of snapshots. This is based on the fact that these snapshots cannot be changed once correctly embedded within the repository (in certain instances changes to data correctly stored may be unethical).

Figure 1.3 - The issue of non-volatility



Source: Inmon, 1996: 36

- *The data warehouse is non-volatile*

  Once the initial loading of data is complete, the only other operation performed within the data warehouse environment is access to the data. In stark comparison, inserts, deletes and amendments to data within the traditional application are normal and usually occur on a record-by-record basis. This simplification within the data warehouse expedites the development process in that control over changes to data is no factor. This simplification also diminishes the need for highly complex technology

supporting backup and recovery that would usually be the case in traditional applications (figure 1.3 illustrates this).

3.4 *General internal control risks within the data warehouse environment*

According to the COBIT control criteria identified above, the major internal control risks which could be faced by an organisation employing a data warehouse, could have material effects on the successful operation of the organisation. An indication of the major internal control risks faced by an organisation embarking on an implementation or who might currently possess such an environment (Curtis & Joshi, 1997: 40-44):

- *Effectiveness*

  Effectiveness is arguably the most important criteria for a data warehouse. Not only must the information be pertinent to the user, but timeous and consistent delivery of data is imperative. Since data retained within the warehouse is obtained from a multitude of sources, untimely and inconsistent delivery of data could skew results relied upon by management. Resultant misinformed decisions could be responsible for the demise of an organisation.

- *Integrity*

  The integrity of data relates to the accuracy and completeness of information. This could involve failing to ensure that data received from source systems is cleansed before being permitted access to the data warehouse environment. If information retrieved from the data warehouse proves to be incorrect, strategic management decisions derived from such inaccurate data could also negatively impact the organisation's operations.

- *Availability*

  Since data warehouses usually retain unbelievably high volumes of data which are combined to present consolidated results, ineffective storage could result in management not being able to obtain data which can be effectively analysed. A similar scenario is true if one considers that availability is also impacted by ineffective backup and comprehensive disaster recovery procedures.

- *Efficiency*

  As management seek to minimise costs and reach optimal return on investment, warehouses can become uneconomical over time if not effectively

monitored for efficiency. Inefficient data warehouses storing needless data can adversely impact the ability of the database resources to make information available in the form and time frame needed by the user.

- *Confidentiality*

  The data warehouse is a clearly categorised repository of data, usually storing information by degree of importance. Intruders therefore find this environment irresistible. Leaking of past patents, company strategies and other sensitive information could leave an organisation doomed. Physical security risks may also negatively impact the data warehouse environment.

- *Reliability of information*

  Ineffective planning of the data warehouse could leave management with a model which does not supply the information required for improved management decision making. Not only would this result in a waste of valuable company resources, but will also deter future management from ever considering data warehousing as a tool which could improve their strategic thinking.

4. Research methodology

The research methodology applied in this study consisted mainly of an understanding of literature in the fields of internal auditing and systems and applications development.

To identify the most pressing issues within the data warehouse environment, an empirical study has also been conducted.

5. Presentation approach

The study shall comprise seven chapters which are described as follows:

- *Chapter 1: Introduction*

  The need, purpose and research methodology for the study is sketched briefly. Internal control risks, purpose, research and presentation approaches have also

been formalised. The reader is introduced to the data warehouse concept and its unique components.

- *Chapter 2: Data warehouse development*

  Internal control risks specific to the development and implementation of a data warehouse are identified. The system development methodology utilised in this development are commented on. The chapter provides the internal auditor with suitable internal control considerations which could be used to assess internal control risks. The results of the empirical study specific to data warehouse development are also presented.

- *Chapter 3: Established data warehouse environment*

  Given the nature of an established data warehouse environment, consideration will be given to highlighting the internal control risks specific to such an environment. The chapter will conclude by discussing internal control considerations which can be applied in assessing such an environment. The results of the empirical study specific to established data warehouse environment are also presented.

- *Chapter 4: Dependent data mart environment*

  The dependent data mart is a sub-environment of the data warehouse. Attention will be given to providing the internal auditor with an understanding of how the dependent data mart relies on the existing data warehouse. It also outlines what control risks exist within such an environment. The chapter will detail internal control considerations which the internal auditor can apply in ascertaining whether suitable control measures are in place to mitigate significant exposures.

- *Chapter 5: Distributed data warehouse environment*

  Internal control risks specific to the distributed data warehouse environment will be discussed. This will provide the internal auditor with in insight on how to ensure integrity of data across open communication mediums. Appropriate internal control considerations and the results of the empirical study are also provided.

- *Chapter 6: Future developments and trends*

  The chapter includes a brief insight into future developments expected within the data warehouse environment and what affect these changes could have on the internal auditor's assessment of internal control risk.

The chapter will also detail the advantages internal auditors can realise in utilising data warehouse technology as part of other audit reviews performed. The chapter will conclude by comparing how important South African internal auditors regard the data warehouse environment to other technologies in the future.

- *Chapter 7: Conclusion*
  Conclusions arising from this study are elaborated on and overall recommendations are provided. Finally, further areas of research are proposed.

6. Summary

In this chapter we identify the purpose of the study as being the identification of the most pertinent internal control risks within the data warehouse environment. This study will also provide the internal auditor with suitable internal control considerations on how to assess internal control risk.

The chapter also defines the core components of the data warehouse environment and its association with decision support systems. Finally, a brief summation of the key internal control risk areas which could exist within the data warehouse environment has been presented.

7. Conclusion

Since internal auditors have a responsibility of ensuring that management implement controls which allow them to attain their overriding goals and objectives, internal auditors should be aware of the risks and necessary control mechanisms needed to mitigate exposures within the data warehouse environment.

It is clear from statistics presented in the background section of this chapter, that the prevalence of data warehouses will dramatically increase in time to come. It is therefore imperative that the data warehouse environment be considered as an integral part of the internal auditor's universe and should be consistently monitored by management for possible control weaknesses.