Chapter 3

Established Data Warehouse Environment

1. Introduction

The warehousing challenge is to technically capture, validate, integrate and transform data into meaningful information and then store that information into a data warehouse (Fryman, 1997: 46). In the preceding chapter we identified how the project team must develop suitable interfaces and integrate data warehouse applications needed to access data.

2. Aim

This chapter aims at identifying the internal control risks specific to the established data warehouse environment. The associated internal control considerations which may be applied in assessing the internal control risks are also discussed.

Results of the empirical survey conducted are provided at the end of the chapter.

3. Internal control risks and considerations within the established data warehouse environment

In this section we identify six internal control risks which may exist within an established data warehouse environment. Under each of the risks identified we provide a brief explanation of the risk and also indicate which of COBIT's information criteria, viz. effectiveness, integrity, availability, efficiency, confidentiality are affected.

The internal auditor is also provided with suitable internal control considerations which can be applied in assessing each of the internal control risks.

3.1 Inability to measure data quality and ensure satisfactory refreshing of data

3.1.1 Risk explanation

Without continuously monitoring data quality, management cannot ensure that data complies with approved management standards (Bohn, 1997: 1).

The refreshing of data within the data warehouse is fixed during the codification of the interfaces between the source system and data warehouse applications (Inmon, 1996: 280). If the refreshing rate of data it is not revisited with the user on a frequent basis, it is possible that such rates may become unsuitable in the future and result in users placing reliance on inaccurate data presented by the data warehouse.

According to COBIT's information criteria identified in chapter 1, the risk identified affects the effectiveness, integrity, availability and efficiency aspects of information (Curtis & Joshi, 1997: 40-43).

3.1.2 Internal control considerations

The following internal control considerations are applicable (ibid.):

- A data conversion plan should be developed.
- At a minimum, the data conversion plan addresses:
  i. The methodology applied in developing the data warehouse
  ii. Approved tolerance levels for errors in source data
  iii. The data standards and what data quality measures have been implemented.
- The management team should take steps to ensure that the transfer of data from the staging area to the final data warehouse is administered by the data administrator.
- All parties involved with the data warehouse should be familiar with the contents of the data conversion plan.
- The conversion plan should be updated with all new subject areas added to the environment.

- Tolerance levels for source data errors should be revisited regularly by the management team based on user feedback.
- Methods and monitoring procedures should be in place to assess the reliability and acceptability of data.
- Audit log issues recorded by data warehouses should be reviewed and significant anomalies followed up timeously.
- The time base applied in refreshing data should be determined based on a trade-off between timely data and the effective utilisation of information technology resources.

3.2 Not ensuring the completeness of data migrated to the data warehouse (Fryman, 1997: 46)

3.2.1 Risk explanation

In instances where management decides to include additional subject areas over time, ineffective project management and the lack of an approved development methodology will result in new subject areas not being included in the most efficient and effective manner.

Changes made to source systems without considering the data warehouse environment could affect the completeness of data migrated to such an environment (Inmon, 1996: 182). An ineffective communication process amongst the various Information Technology teams and end-users can result in these changes not being communicated effectively.

According to COBIT's information criteria identified in chapter 1, the risk identified affects the integrity aspect of information (Curtis & Joshi, 1997: 40-43).

3.2.2 Internal control considerations

The following internal control considerations are applicable (Fryman, 1997: 46):

- An approved data warehouse development methodology should be applied when adding new subject areas to the data warehouse environment (Inmon's development methodology detailed in chapter 2 is recommended).
- A proactive communication process should be in place to ensure that all changes made to source systems on which the data warehouse depends are reported timeously and to the correct personnel for action.

3.3 Ongoing availability of data warehouse operations cannot be ensured (Warigon, 1998: 55)

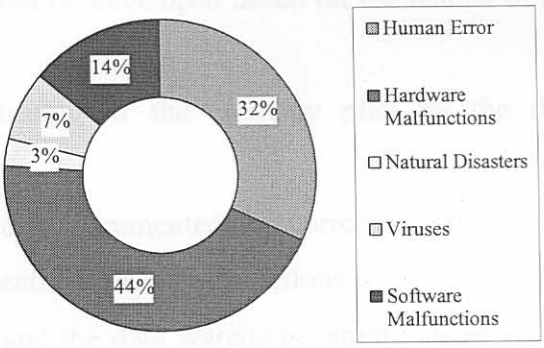
3.3.1 Risk explanation

A 1997 survey (Semer, 1998: 47) which tracked the major classifications of data loss among 50 000 organisations was conducted by Ontrack Data International Inc.. The survey indicated that 44% of data losses were caused by hardware or system malfunctions. Figure 3.1 provides more comprehensive results of the survey.

According to the Disaster Recovery Journal (Semer, 1998: 42), organisations suffered significant disaster-related costs in 1997:

- Each on-line outage averaged four hours and cost American companies and average of $329 000 in lost revenues and productivity.
- For each hour of unscheduled downtime, 355 worker hours were lost.
- Major businesses lost 38.1 million work hours, or $444 million in wages annually.

Figure 3.1 - Source of data loss



Source: Semer, 1998: 47

It is apparent from the above mentioned statistics that a significant risk is faced by organisations should mission critical systems become unavailable. Although the data warehouse environment is only one source providing information to the user, the statistics provide an indication of the potential losses which can be incurred. Without consistent and supported data warehouse services, the end user may be unable to make informed management decisions.

According to COBIT's information criteria identified in chapter 1, the risk identified affects the availability aspect of information.

3.3.2 Internal control considerations

The following internal control considerations are applicable (ISACA, 1998):

- The data warehouse environment should be included in the business's overall continuity planning process.
- The data warehouse environment should be considered in the business impact analysis. This analysis identifies and inventories mission critical resources; quantifies the costs associated with failure to transact business due to the loss of resources; and estimates the downtime the organisation can bear while those resources are being restored.
- If the data warehouse environment is identified as one of the mission critical applications, management should identify and weight the risks specific to the data warehouse.
- Recovery plans and procedures should be developed based on the results of the initial business impact analysis.
- Routine training and simulation testing of the recovery plan by the data warehouse users should be conducted.
- Results of recovery plan tests should be documented and corrective action taken to address significant weaknesses identified during simulations.
- Interfaces between source systems and the data warehouse environment should be inventoried to ensure synchronisation of data during backup and recovery.
- Backup and recovery procedures should be fully documented, understood, accessible, enforced and tested regularly (Clark, Holloway & List: 114-115).

- The back-up and recovery procedures should take into account possible error conditions which could be encountered during the back-up and recovery process and provides suitable troubleshooting guidelines (ibid.)

3.4 *Overall data warehouse administration becomes ineffective and inefficient* (Warigon, 1998: 59)

3.4.1 *Risk explanation*

By not effectively monitoring the data warehouse environment, it can become unwieldy. In many instances, ineffective and inefficient data warehouses are caused by not regularly archiving outdated data and by not executing frequent capacity planning measures. The ultimate effect of not performing these activities are increased annual storage, processing and operating costs (ibid.).

Routine archiving of data involves the rolling up of outdated data to higher levels of summary (Inmon, 1996:69). This rolling of data can either be by means of transferring data from one level of the data warehouse architecture to another or, retaining data within a high-performance storage medium.

According to COBIT's information criteria identified in chapter 1, the risk identified affects the efficiency, effectiveness and availability aspects of information.

3.4.2 *Internal control considerations*

The following internal control considerations are applicable:

- Formalised assessments should be conducted in conjunction with the end user as a means of identifying data elements whose probability of access is close to zero (Inmon, 1996: 306). Factors which should be considered before data is archived (Inmon, 1998: 4):

  i. Time

  The project team must consider whether there is a probability that once archived that this data will be needed by the end user again. The costs of restoring the data can sometimes exceed the cost of retaining it within the data warehouse environment.

ii. Classes of data

Ascertain what classes of data are most frequently used in queries and whether this pattern will change in the foreseeable future.

iii. Level of detail

Ascertain what level of detail is most commonly utilised by the end user and whether this pattern will change in the foreseeable future.

iv. Strategic importance of data

Although certain classes of data may not be accessed on a frequent basis, there is a probability that the class may be strategically important. In such instances, it is recommended that the data remain within the data warehouse environment.

- Maintenance and data management policies should be developed which clearly stipulate the methods and time frames which should be applied in phasing out unnecessary data classes (Zicker, 1998: 1).
- Management should consider scheduling and monitoring software which can simplify the tracking of outdated data and provide automated archiving functionality (ibid.).
- A data administrator should be employed or assigned the responsibility of monitoring and administering the archiving of data (Curtis & Joshi, 1997: 40-43).
- Statistics on performance, capacity, and availability should be provided (including historical versus forecast performance variance explanations) for the data warehouse on a regular basis (ISACA, 1998).
- Performance reporting information to users regarding usage and availability should take place (such reporting should include capacity, workload scheduling and trends) (ibid.).
- The package provider or data warehouse project team should be requested to give assurances that data warehouse applications will be able to manage growth in processing rates (Curtis & Joshi, 1997: 40-43).
- As part of the post-implementation phase of the system development life cycle, criteria should be included to determine the future growth and changes to performance expectations (ISACA, 1998).

3.5 *Data warehouse access is not restricted to authorised users* (Warigon, 1998: 55)

3.5.1 Risk explanation

Unauthorised access to data retained within the data warehouse can result in significant losses to the organisation (Warigon: 1998: 55-60). These threats can be caused by accidental or malicious attacks from employees. Outside threats can be caused by competitors. The result of such unauthorised access could be negative publicity for the organisation and a loss of continuity of data warehouse operations. Management will need to identifying security vulnerabilities which could negatively impact the organisation's image. As part of this assessment, physical security risks should also be considered. (ibid.).

According to COBIT's information criteria identified in chapter 1, the risk identified affects the confidentiality aspect of information.

3.5.2 Internal control considerations

The following internal control considerations are applicable:

- Management should classify data to ensure that the application of security resources is optimised and that different protective measures are used for different categories (Warigon: 1998: 55-60). Classifications of data could include public, moderately sensitive and highly sensitive data.
- Project management should quantify the value of data requiring protection. The criteria which could be used in determining the value of data requiring protection can include:
  - i. The cost to reconstruct the data should a disaster occur.
  - ii. The cost of restoring the integrity of violated data.
  - iii. The inability to obtain data timeously thereby preventing informed decisions being made.

iv. Costs of litigation should customer's data be erroneously or intentionally exposed to unauthorised sources.

- Vulnerabilities to the data warehouse should be identified, evaluated and documented.
- A security policy should be developed. The policy should include (ISACA, 1998):
  i. Identification of security roles.
  ii. Security validation.
  iii. Documented proof of management support and commitment.
  iv. Access philosophy.
  v. Access authorisation procedures.
  vi. Annual reviews of access authorisation.
  vii. Password standards identified.
  viii. Security awareness drives.
  ix. The role of the security administrator defined.
- Confidentiality and intellectual rights agreements should be in place for all data warehouse users (ISACA, 1998).
- Each user should be defined to the database with a unique user identification.
- Passwords should be assigned to each user and the system pre-empts personnel to change theirs every thirty days.
- All access privileges should be approved by the data owner.
- Access rights should be revisited on a monthly basis to ensure that all terminated or transferred personnel are removed from the access rights to the system.
- Improved control is realised when management consider encrypting data. This should however only be considered in cases where data is extremely confidential. Encryption is costly and may prove cumbersome since the algorithms consume large portions of the central processing unit's resources (Warigon: 1998: 55-60).
- Ultimately, the management team should have selected the most cost effective security measures, i.e. the costs of protecting the data does not exceed the maximum monetary amount that the loss of data would represent (ISACA, 1998).

3.6 *Ongoing risk assessments over the data warehouse environment are not conducted* (Warigon, 1998: 55)

3.6.1 Risk explanation

Cost-effective measures are required to address the most significant risks within the data warehouse environment. Organisations should be focusing on ways to limit costs and only secure mission critical assets (ibid.). This valuable information can only be obtained by performing, and revisiting risk assessments over information technology environments such as the data warehouse. These assessments will allow management to identify how critical the risks are within the data warehouse environment and thereby apply the limited resources in the most effective and efficient manner.

Ultimately, if organisations do not perform risk assessments on a frequent basis, the effective and efficient utilisation of resources cannot be ensured.
According to COBIT's information criteria identified in chapter 1, the risk identified affects the effectiveness, integrity, efficiency and reliability aspects of information.

3.6.2 Internal control considerations

The following internal control considerations are applicable (ISACA, 1998):

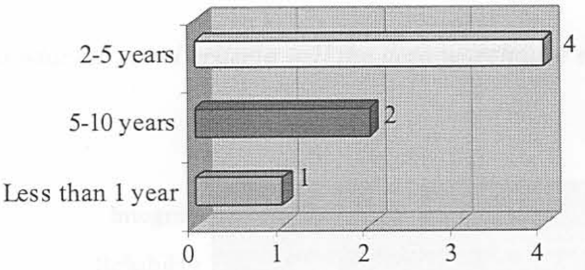
- Management should have a program in place to identify additional data types which could provide additional benefit to the user if migrated to the data warehouse environment.
- Overall warehouse administration should meet user's expectations (e.g. conduct user satisfaction surveys).
- Regular benchmarking with similar facilities should take place to ensure best practice.
- Data quality control should be revisited to ensure that the number of instances of contaminated data going undetected is reduced.
- Company wide business reviews should be performed to identify, investigate and resolve data elements that are not within quality standards.

- Mechanisms should be in place to record and monitor the data warehouse's capturing of data and the quality of such data over time.
- Frequent security assessments should be undertaken (Warigon, 1998: 60). Evaluations should be conducted continuously to determine whether security measures and controls are:
  i. Simplistic and straightforward.
  ii. Carefully monitored.
  iii. Do not hamper authorised users from performing their duties effectively and efficiently.
  iv. Are easily adaptable to necessary changes in control standards.

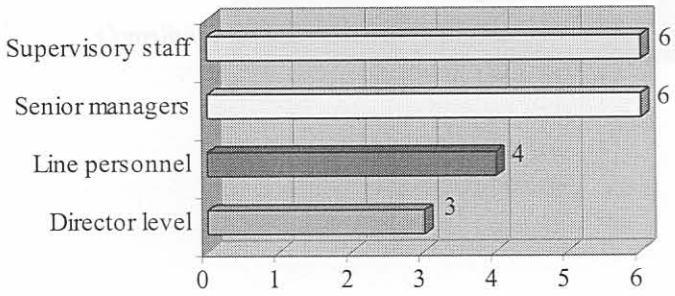
4. A South African perspective on the audit of established data warehouse environments

The results of the local survey are featured below. The results relate specifically to the internal control risks within the established data warehouse environment:

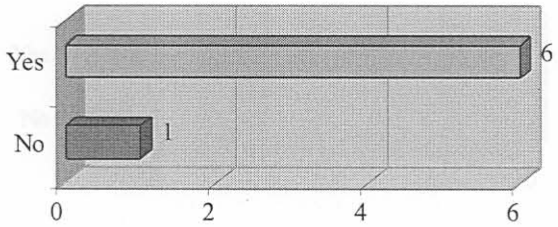
1. How long has the organisation had a data warehouse?



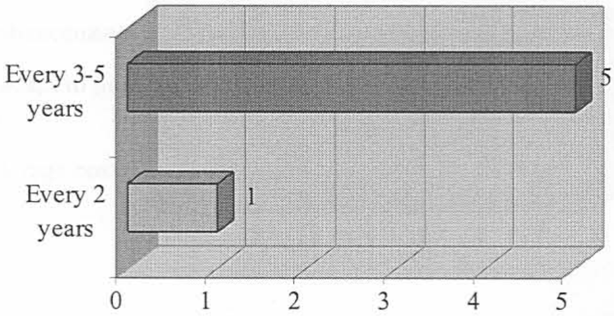
2. Which level of staff utilise the data warehouse structure?



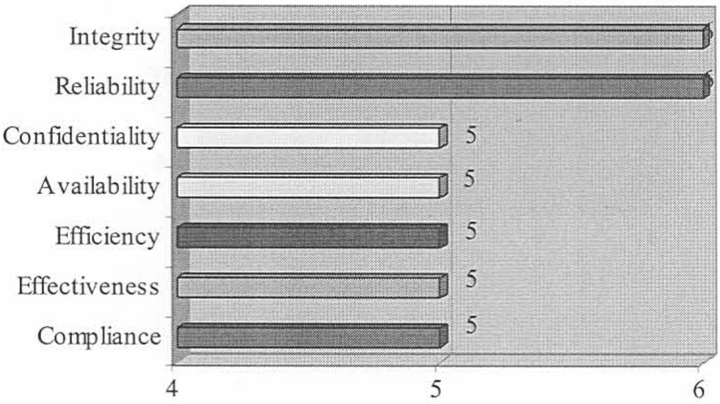
3. Is the data warehouse environment identified as an application reviewed by the internal audit team on a periodic basis?



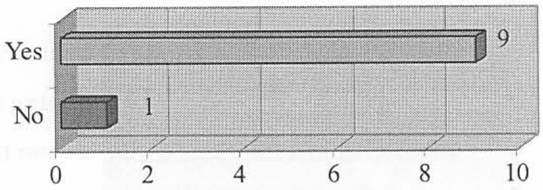
4. If it is audited, how frequently will the data warehouse environment be reviewed by internal audit?



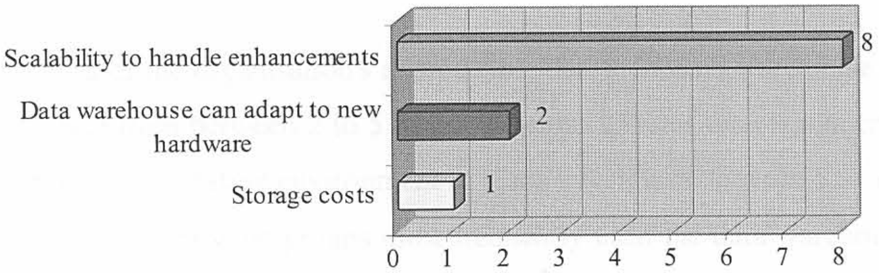
5. According to which control criteria will the data warehouse environment will be reviewed?



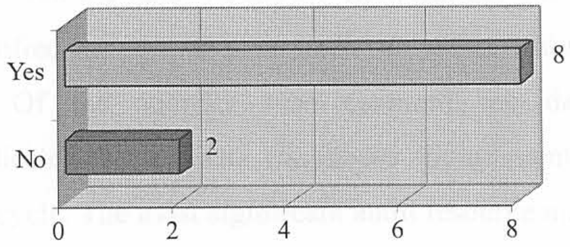
6. Was capacity management identified as part of the audit approach?



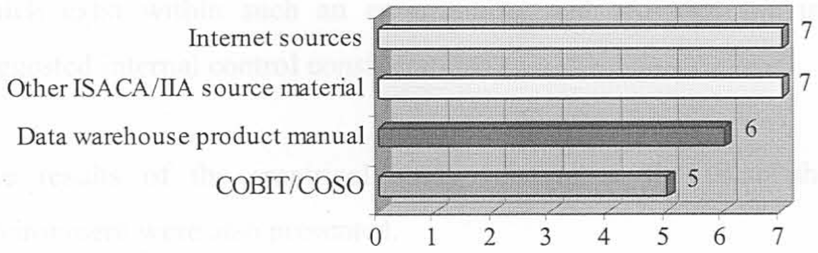
7. If capacity planning was considered, for which of the following reasons was it considered?



8. Will/has the data warehouse environment been included in the organisation's continuity plans/efforts?



9. Which audit resource materials were used in formulating a suitable audit approach and program?



Based on the above mentioned responses, the most significant findings raised included:

- 57% of the organisation's approached, had their data warehouse environments in place from between 2 to 5 years. Whereas 25% of the respondents indicated that they had a mature environment in place which was between 5 to 10 years old.
- When asked what groups most frequently used the data warehouse environment, 63% of the respondents indicated that the data warehouse was applied by senior management and supervisory staff. Only 15% indicated that the data warehouse environment was being used at director level. This strongly indicates that localised data warehouses are focused on middle management level (i.e. medium term planning) as opposed to the strategists within each of the organisations.
- Internal auditors stated that the data warehouse environment was an application which had been identified as part of their audit universe. Only 1 respondent indicated otherwise. Of the auditors who assessed the data warehouse environment, 83% indicated that the data warehouse environment was reviewed on a 3 to 5 year audit cycle. The most significant audit resource materials used in preparing a suitable audit approach were obtained from internet sources and other professional body materials.

5. Summary

In this chapter we identified the internal control risks which may exist within the established data warehouse environment. We also identified six internal control risks which exist within such an environment and provided the internal auditor with suggested internal control considerations.

The results of the empirical study relating to the established data warehouse environment were also presented.

6. Conclusion

Significant internal control risks are not only found in an environment where a data warehouse is being developed but also in an established data warehouse environment.

The internal auditor's primary aim is to ensure that management attain their primary goals and objectives. He/she should therefore ensure that data warehouse reviews are all founded both on management's assessment of how critical the data warehouse environment is and the risk assessment model adopted by the internal audit department. This approach will ensure that the internal auditor evaluates the data warehouse on a basis commensurate with the overall risk the organisation could be exposed to.

In reviewing the results of the empirical study, we can say that the most significant audit resources used in preparing the internal auditor's approach were obtained from internet resources. It appears that the data warehouse environment is in actual fact a relatively new audit cycle with limited resource materials available to the internal auditor.