

14244925



INTERNAL CONTROL RISKS WITHIN THE DATA WAREHOUSE ENVIRONMENT

by

SEAN PAUL DE LA ROSA

Submitted in fulfillment of the requirements for the degree

**MAGISTER COMMERCII
(COMPUTER AUDITING)**

in the

Faculty of Economic and Management Sciences

at the

UNIVERSITY OF PRETORIA

PRETORIA

NOVEMBER 1999

AKADEMIESE INLIGTINGSDIENS UNIVERSITEIT VAN PRETORIA	
- 5 JUN 2000	
Klasnommer:	658-40380285574
Aerwinnommer:	1145 23991

DE LA ROSA

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my study leader, Prof. Dr. H. de Jager for his ongoing support and assistance during this project.

The research is dedicated to my parents, Steve and Peggy de la Rosa for affirming my ambitions and their unending words of encouragement. All the honour and glory for both the preparation and successful completion of this work goes to my Lord and Saviour Jesus Christ who gave me the wisdom, understanding and endurance.

Thanks Tommy.

JOHANNESBURG

NOVEMBER 1999

TABLE OF CONTENTS

1.1 Background	1
1.2 Objectives and scope of the study	1
1.3 Organization of the study	1
1.4 Summary	1

2. Background

2.1 Objectives and scope of the study

2.1.1 Objectives of the study

2.1.2 Scope of the study

“He that will not apply new remedies must expect new evils, for time is the greatest innovator.”

2.1.3 Justification of the study

2.1.4 The data warehouse and its evolution

2.1.5 General architectural considerations within the data warehouse environment

2.1.6 Summary

Francis Bacon 1561 - 1626

3. Summary

“The significant challenges we face today cannot be resolved by the same level of thinking that created them.”

4. Conclusion

Albert Einstein 1879 - 1955

Chapter 2: Data Warehouse Development

2.1 Introduction	12
2.2 Aim	12
2.3 Why a different system development life cycle suits for data warehouses	12
2.4 Interface development	12
2.4.1 Strategic planning	14
2.4.2 Data model analysis	14
2.4.3 Breadth analysis	16
2.4.4 Technical assessment	20
2.4.5 Technical environment preparation	21
2.4.6 Subject area analysis	24
2.4.7 Data warehouse design	25
2.4.8 Source systems analysis	27
2.4.9 Interface specifications and population	30

TABLE OF CONTENTS

<i>Summary</i>	<i>i</i>
<i>Opsomming</i>	<i>iii</i>
<i>List of figures</i>	<i>v</i>
Chapter 1: Introduction	1
1. <i>Background</i>	1
2. <i>Purpose and reason for study</i>	2
2.1 Purpose of study	2
2.2 Reason for undertaking study	2
3. <i>Defining the data warehouse environment and internal control risk</i>	3
3.1 Defining internal control risk	3
3.2 Interaction with the decision support environment	4
3.3 The data warehouse and its concepts	5
3.4 General internal control risks within the data warehouse environment	8
4. <i>Research methodology</i>	9
5. <i>Presentation approach</i>	9
6. <i>Summary</i>	11
7. <i>Conclusion</i>	11
Chapter 2: Data Warehouse Development	12
1. <i>Introduction</i>	12
2. <i>Aim</i>	12
3. <i>Why a different system development life cycle exists for data warehouses</i>	13
4. <i>Interface development</i>	13
4.1 Strategic planning	14
4.2 Data model analysis	18
4.3 Breadbox analysis	20
4.4 Technical assessment	22
4.5 Technical environment preparation	23
4.6 Subject area analysis	24
4.7 Data warehouse design	25
4.8 Source systems analysis	27
4.9 Interface specifications and population	30

<hr/>	
Chapter 2: Data Warehouse Development (continued)	33
<hr/>	
5. <i>Data warehouse package and vendor evaluation</i>	33
6. <i>A South African perspective on the audit of developing data warehouse environments</i>	35
7. <i>Summary</i>	40
8. <i>Conclusion</i>	40
<hr/>	
Chapter 3: Established Data Warehouse Environment	41
<hr/>	
1. <i>Introduction</i>	41
2. <i>Aim</i>	41
3. <i>Internal control risks and considerations within the established data warehouse environment</i>	41
3.1 Inability to measure data quality and ensure satisfactory refreshing of data	42
3.2 Not ensuring the completeness of data migrated to the data warehouse	43
3.3 Ongoing availability of data warehouse operations cannot be ensured	44
3.4 Overall data warehouse administration becomes ineffective and inefficient	46
3.5 Data warehouse access is not restricted to authorised users	48
3.6 Ongoing risk assessments over the data warehouse environment are not conducted	50
4. <i>A South African perspective on the audit of established data warehouse environments</i>	51
5. <i>Summary</i>	55
6. <i>Conclusion</i>	55
<hr/>	
Chapter 4: Dependant Data Mart Environment	56
<hr/>	
1. <i>Introduction</i>	56
2. <i>Aim</i>	57
3. <i>Understanding the dependant data mart environment</i>	57
3.1 Background	57
3.2 Development considerations	57
3.3 Closed-loop business performance management	77
3.4 Increased access to data warehouse information	77
3.5 Removal of source data quality problems automatically	78
3.6 Re-engineering the development technology	78
3.7 Transferring of report and query functionality	79

Chapter 4: Dependant Data Mart Environment (continued)

4. <i>Internal control risks and considerations within the data mart environment</i>	58
4.1 A lack of sufficient response time monitoring on a periodic basis	59
4.2 Transfer of data from the organisation wide data warehouse to the data mart is not controlled	60
5. <i>Summary</i>	61
6. <i>Conclusion</i>	61

Chapter 5: Distributed Data Warehouse Environment

1. <i>Introduction</i>	63
2. <i>Aim</i>	63
3. <i>Understanding the distributed data warehouse environment</i>	64
3.1 Background	64
3.2 Development considerations	65
3.3 Access and security considerations	66
4. <i>Internal control risks and considerations within the distributed data warehouse environment</i>	67
4.1 Distributed data warehouse access is not restricted to authorised users	68
4.2 Ongoing availability of the distributed data warehouse operations cannot be ensured	70
4.3 Efficiency of processing within the distributed data warehouse is not maximised	71
5. <i>A South African perspective on the distributed data warehouse environment</i>	73
6. <i>Summary</i>	74
7. <i>Conclusion</i>	75

Sources Referenced

Chapter 6: Future Developments and Trends

1. <i>Introduction</i>	76
2. <i>Aim</i>	76
3. <i>Future developments and the effects on internal control risk</i>	
3.1 Closed-loop business performance management	77
3.2 Increased access to data warehouse information	77
3.3 Removal of source data quality problems automatically	78
3.4 Re-engineering the development methodology	78
3.5 Transferring of report and query functionality	79

Chapter 6: Future Developments and Trends (continued)

4. <i>Utilisation of data warehouse technology by the internal auditor</i>	79
4.2 Background	79
4.2 Defining data mining	81
4.3 Process followed in utilising data warehouse technology	81
5. <i>A South African perspective on the future of data warehouse technology</i>	82
6. <i>Summary</i>	84
7. <i>Conclusion</i>	85

Chapter 7: Conclusion

1. <i>Summary</i>	86
2. <i>Further areas of research</i>	87
3. <i>Conclusion</i>	88

Annexures

<i>Annexure 1: Data warehousing survey questionnaire</i>	90
<i>Annexure 2: Strategic data warehouse checklist</i>	97
<i>Annexure 3: Roles and responsibilities of the data administrator</i>	102
<i>Annexure 4: Vendor prescreening and application selection questionnaire</i>	107

Glossary

Source References

management to be assured that the environment is well controlled. Management also require reliable and accurate information to be provided timely.

The internal audit profession strives to provide management with suitable suggestions as to how internal control risks can be identified and managed within the various high risk areas of their organisations. A study of the sufficiency of audit resources relating to the data warehouse was conducted. It is evident that the internal audit profession has been slow in identifying the specific internal control risks relating to such an

SUMMARY

INTERNAL CONTROL RISKS WITHIN THE DATA WAREHOUSE ENVIRONMENT

by

SEAN PAUL DE LA ROSA

LEADER : **PROF. DR. H DE JAGER**
FACULTY : **ECONOMIC AND MANAGEMENT SCIENCES**
DEPARTMENT : **SCHOOL OF ACCOUNTANCY**
DEGREE : **MCOM (COMPUTER AUDITING)**

Decision support systems enable management to improve both their strategic and short term decision making processes. In most cases, decision support systems rely on historical transactional data to provide trends and statistics which management teams can base their decision making process on. An example of a decision support system is the data warehouse.

The data warehouse environment has received increased attention over the years. The augmented importance of this evolving technology has emphasised the need for management to be assured that the environment is well controlled. Management also require reliable and accurate information to be provided timeously.

The internal audit profession strives to provide management with suitable suggestions as to how internal control risks can be identified and managed within the various high risk areas of their organisations. A study of the sufficiency of audit resources relating to the data warehouse was conducted. It is evident that the internal audit profession has been slow in identifying the specific internal control risks relating to such an

environment. With this in mind, we undertook a study to determine the nature of the specific internal control risks within the data warehouse environment.

The overall purpose of the study was to provide the internal auditor with an understanding of the internal control risks within the various components of the data warehouse. Suitable internal control considerations to assess the extent of such risks were also outlined.

The study also provides a brief insight into what future trends and developments can be expected within the data warehouse environment. The effect that such enhancements could have on the internal auditor's assessment of internal control risk were also noted.

The study concludes by providing four general recommendations which should minimise internal control risks to a satisfactory level. Provided that these recommendations are considered, both management and the internal auditor will achieve maximum benefit from this environment.

Bestuursinligting- en verspreidingsprosedure maak dit moontlik vir bestuur om beide hul
sinnelike en kragtige beslutfunksies te verbeter. In die meeste gevalle
maak die bestuursinligting- en verspreidingsprosedure gebruik van
bestuursinligting- en verspreidingsprosedure om te versek dat die
bestuursinligting- en verspreidingsprosedure korrek en akkuraat is.
Bestuursinligting- en verspreidingsprosedure is die belangrikste
bestanddeel van die bestuursinligting- en verspreidingsprosedure.

Die belangrikste ontwikkeling in die afgelope paar jaar toekoms is
Die toenemende belangrikheid van hierdie ontwikkelende tegnologie het die belangrikheid
beklemtoon dat bestuurders moet weet dat die omgewing goed beheer word.
Bestuursinligting- en verspreidingsprosedure is die belangrikste
bestanddeel van die bestuursinligting- en verspreidingsprosedure.

Die interne audit professionele streef om te versek dat die bestuursinligting- en verspreidingsprosedure
waarop interne beheer streef getoets word en bestuur kan word, met inagneming van
die verskillende hoë risiko-areas in hul organisasie- strukture op 'n stelsel- en

OPSOMMING

INTERNE BEHEER RISIKO'S IN DIE INLIGTINGSTOOROMGEWING

deur

SEAN PAUL DE LA ROSA

LEIER : **PROF. DR. H DE JAGER**
FAKULTEIT : **EKONOMIESE EN BESTUURSWETENSKAPPE**
DEPARTEMENT : **SKOOL VIR REKENMEESTERSOPLEIDING**
GRAAD : **MCOM (REKENAAR OUTIDEERING)**

Besluitneming-ondersteuningstelsels maak dit moontlik vir bestuur om beide hul strategiese en kort termyn besluitnemingsprosesse te verbeter. In die meeste gevalle, maak die besluitneming-ondersteuningstelsels staat op histories transaksiedata om bestuurspanne van tendense en statistieke te voorsien waarop hulle hul besluitnemingsprosesse kan baseer. 'n Voorbeeld van 'n besluitneming-ondersteuningstelsel is die inligtingstoor.

Die inligtingstooromgewing het oor die afgelope paar jaar toenemende aandag geniet. Die toenemende belangrikheid van hierdie ontwikkelende tegnologie het die behoefte beklemtoon dat bestuurslui seker moet wees dat die omgewing goed beheer word. Bestuur verlang verder om betyds van betroubare en akkurate inligting voorsien te word.

Die interne oudit professie streef om bestuurslui met gepaste metodes te voorsien waarop interne beheer risikos geïdentifiseer en bestuur kan word, met inagnome van die verskillende hoë risiko-areas in hul organisasie. Gebaseer op 'n studie oor die

beskikbaarheid van oudit hulpbronne wat verband hou met die inligtingsstoor, blyk dit die professie traag is met die identifisering van spesifieke interne beheer risiko's wat verband hou met so 'n omgewing. Met hierdie waarneming in gedagte, is 'n studie onderneem met die doel om die kenmerke van die spesifieke interne beheer risiko's binne die inligtingsstoor omgewing te bepaal.

Die algehele doelwit van die studie was om die interne ouditeur se kennis en begrip uit te brei aangaande interne beheer risiko's binne die verskillende komponente van die inligtingstoor. Gepaste interne beheer-oorwegings om die erns van sulke risiko's te evalueer, is ook uitgestip.

Die studie verskaf ook 'n kort insig oor watter toekomstige tendense en verwickelinge vermag kan word binne die inligtingsstooromgewing, asook die effek wat sulke verbeteringe kan hê op die interne ouditeur se evaluasie van interne beheer risiko's.

Die studie sluit af deur vier algemene aanbevelings te maak wat sal verseker dat interne beheer risiko's sal verminder tot 'n meer aanvaarbare vlak. Indien hierdie aanbevelings oorweeg word, sal beide bestuur en die interne ouditeur maksimum voordeel uit hierdie omgewing kan put.

Chapter 10 Future Developments and Trends

Figure 10.1: Summary of the study process

LIST OF FIGURES

Chapter 1: Introduction	1
<i>Figure 1.1: Overview of the knowledge discovery in databases process</i>	4
<i>Figure 1.2: The issue of time variancy</i>	6
<i>Figure 1.3: The issue of non-volatility</i>	7
Chapter 2: Data Warehouse Development	12
<i>Figure 2.1: Traditional system development life cycle</i>	13
<i>Figure 2.2: System development life cycle for the data warehouse</i>	14
<i>Figure 2.3: Controlled data conversion process</i>	29
Chapter 3: Established Data Warehouse Environment	41
<i>Figure 3.1: Source of data loss</i>	44
Chapter 5: Distributed Data Warehouse Environment	63
<i>Figure 5.1: The global data warehouse structure</i>	64
Chapter 6: Future Developments and Trends	76
<i>Figure 6.1: SEMMA data mining process</i>	81