

Short Research Article**Data Warehousing for a Public University in Kenya****ABSTRACT**

This study aimed at determining the level of dependency between university top decision makers in a public university in Kenya with IT personnel in accessing, analyzing and reporting data. We sought to determine an easier and quicker way to analyze data, design, and implement a solution to turn analysis into a report with little or no help from IT department. A survey showed that the University's existing systems do not allow the top management to explore data, analyze data, and create quick reports. They depend on IT personnel to access data that they need to analyze and generate reports. As result the decision makers tend to have an uneasy relationship with IT people. The study successfully implemented a data warehouse database composed of a single multidimensional cube based on subject area of Finance. By implementing a Business Intelligence (BI) program based on Excel, we turned ordinary spreadsheets into a flexible, powerful, and inexpensive BI system that gives users significant power and flexibility.

Keywords: Data warehouse, University Data warehouse, University Data Mart, Business Intelligence

1. INTRODUCTION

In the last decade, the growth of Kenya's University education has been tremendous. This is as a result of the introduction of free primary education and subsidized secondary education. About 250,000 students graduate every year from Kenyan secondary schools. Out of these, approximately 81,000 qualify to join public universities but only 24,000 are admitted for regular programs and the remaining 56,000 absorbed through self-sponsored programs or admissions to private universities. This population posed steep competition for places in public universities. One of the government's responses has been to transform some colleges or polytechnics into universities in a bid to cater for the rising number of school leavers that are looking for higher education opportunities. Currently, there are 30 universities in Kenya, of which 7 are public and 23 private. The public universities have over 24 constituent colleges. One of the constituent colleges was used as case for this project.

The constituent college that was used in this study was founded in the late 1940s as a training institution for corporate students from East Africa Community. Over the period, the college has undergone upgrades, from a certificate college to a constituent college of a public university. The college has four faculties namely: Engineering, Computing and Information Technology, Media and Communication, and Business. The College offers various degrees, diplomas and certificates as well as short courses.

The constituent college uses both operational systems and ERPs to manage its day-to-day operations. These systems include financial management, human resources management, student information, library, examination management, and course management. These applications typically run in an on-line transaction processing (OLTP) computing environment based on business functional areas.

34 However, the college lacks systems that will foster decision making. The majority of
35 businesses, including educational institutions and government agencies, realize the need for
36 easy and quick access to the right information (Akintola et al., 2011). Data warehouses are
37 complex to implement but can yield real dividends in terms of competitive advantage
38 (Furrow, 2001). The university needs a data warehouse that can unlock its data in various
39 ERPs and other operational systems (Heise, 2005, Wierschem et al. 2003). A university
40 needs new approaches or techniques to provide a centralized source of information
41 accessible across different academic units. The key stakeholders should quickly analyze and
42 develop satisfactory solutions to their underlying problems as well as making better business
43 decisions based on historical data available in the legacy databases.

44 The university's existing systems do not store data in easily accessible data models that can
45 be easily transformed into complete, significant information to sustain sound decision-
46 making. The current user interface does not allow top management to explore, analyze data,
47 and create reports quickly without the help from IT staff. Moreover, reports produced are
48 filled with errors because the source data contained errors.

49 The ad hoc analysis and reporting impossibilities significantly hinder the decision making
50 process. This study therefore sought to establish the level of dependency between top
51 decision makers and IT personnel in accessing, analyzing and reporting data. It also sought
52 to determine the most valuable and most efficient way to analyze data quickly and turn
53 analysis into a report with little or no help from IT department.

54 The four research questions were as follows. What managerial gaps are there which hinder
55 decision making process in the constituent college? What is the most valuable technology
56 for the college users in accessing, analyzing and reporting data easily and quickly? How will
57 the technology be implemented to narrow the managerial gaps in the college? How can the
58 technology be optimized to provide timely and reliably information to the college?
59

60 **2. METHODOLOGY**

61 A survey was conducted to ascertain the level of dependency between top decision makers
62 in the college and IT department in accessing, analyzing and reporting data. The survey also
63 was meant to find out the level of competence in use of Ms Excel. A sample size of 97
64 members of staff was selected from the entire population of about 430 and issued with
65 questionnaires with 74 questionnaires returned. Majority of the respondents were of the
66 opinion that currently, typical reports takes at least a day to prepare. The users of the reports
67 hardly prepare them without the help of IT people. Many decision makers understand what
68 the data means, but they do not have easy or dynamic access to it. Some of the reports
69 prepared either contain errors or doctored to suit its purpose. A third of the staff claims that
70 similar reports are generated several months later against similar data. The quality of the
71 reports produced is not up to standard. Only 21.7% of the respondents feel satisfied with the
72 quality of data analysis and reporting. Majority of university staff are competent excel users
73 although more than half have not used Excel as an outstanding BI tool.

74 The results of the survey revealed that majority of the staff depend on IT professionals in
75 producing reports. The survey also showed that a substantial number of respondents do not
76 have easy access to the data and reports are not developed quickly. There are therefore
77 managerial gaps in the college that need to be addressed.

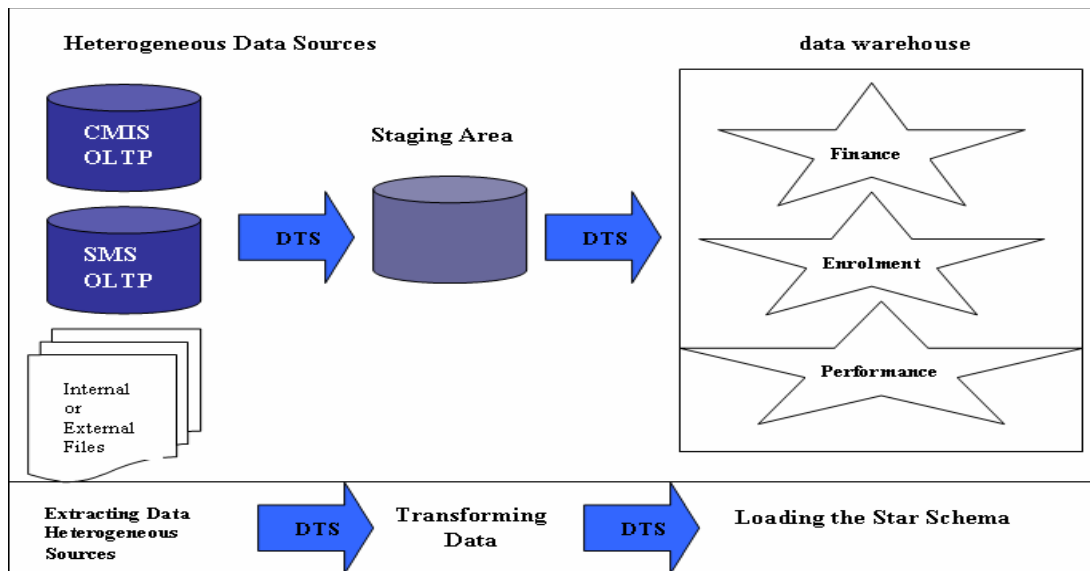
78 After literature survey was carried out, we considered data warehouse as the most valuable
79 technology to be used (Di Tria et al. 2012, Dell'Aquila et al., 2008, Dell'Aquila et al., 2007).
80 This study considered and applied Star Schema (Fig. 1) dimensional modeling techniques to
81 design an integrated data structure required in a data warehouse (Prasad, 2005). Data

82 warehouse composed of subject-oriented data mart using one or more schemas to represent
83 specific business events or processes. A star schema contained a central fact table
84 surrounded by many dimensions tables.
85



86
87
88
89
90
91
92
93
94
95
96
97
Figure 1: Star Schema

Since data was from disparate sources, transformation process was carried out which involved conversion, summarization and filtering of data. Data Transformation Services (DTS) tool was used to extract data from varied sources for instance CMIS OLTP system resided in Ms SQL server, SMS database that resided in Ms Access 2000, data files, external files and Ms Excel spreadsheet (Fig. 2). The extracted data was transformed into a format that can be loaded into the data warehouse. The staging area is used to facilitate the transformation process. The transformed data was then loaded into dimension tables before being loaded into the fact table that references the dimension.



98
99
100
101
102
103
104
105
106
Figure 2: University Business Lines

A subject-area or business line is also referred to as data mart, fact table and cube. After implementation, cubes are the basis for an easy, but very rich analysis and reporting experience for knowledge workers. In practice however, many business intelligence (BI) systems, even the ones utilizing data warehouse technologies, struggle with a low user adoption.

OLAP Reporting based on Excel was identified as the most suitable Business Intelligence tool. OLAP Reporting tool based on Excel gives users and more specifically decision makers significant power and flexibility to access, analyze and reports data.

3. RESULTS AND DISCUSSION

Data stored in a data warehouse is typically extracted from operational systems and possibly from data sources external to the institution with batch processes during off-peak operational hours (Velicanu and Matei, 2007). Data that has been extracted is filtered to eliminate any unwanted data items and transformed to meet the data quality, security and usability requirements. It is then loaded into an appropriate data warehouse data mart to be accessed by end users who are charged with the responsibility of making informed decisions.

The proposed data warehousing technology consolidated essential historical data from the underlying college disparate databases. The source data was identified and organized into main subjects such as Finance, Enrolment and Performance. Each of this subject area constituted a multidimensional-cube in the resultant data warehouse database.

This research integrated and consolidated data into a multidimensional (OLAP) database. The OLAP database contained a single data mart – fee payment data mart that tracked revenue collected from students in various courses. The fact table contained numeric metrics at the intersection of many business entities. The fee payment fact table contained data summarized by student, course, staff and time dimension tables. The dimension tables provided context of revenue collected from students in various courses. The design of the dimension tables targeted analytical needs of the business users by presenting usable, descriptive information that is easy for users to browse.

OLAP Reporting Tool for Excel is an example of Excel-based BI system as shown in Fig. 3. The working environment consists of a collection of integrated windows: OLAP Report Tool for Excel main window, Command and Option window and Chart Field List, designed to support a DSS; a working Screen section - designed for dropping different fields depending report to be created; and OLAP Reporting Tool for Excel that gets its data from the data warehouse.

This tool can be easily deployed and give users an easy access to data from virtually any sources. An Excel spreadsheet can pull data from a cube in the BI database; send it to the add-in; receive the results; process them in the spreadsheet, and then write the results back to the BI database.

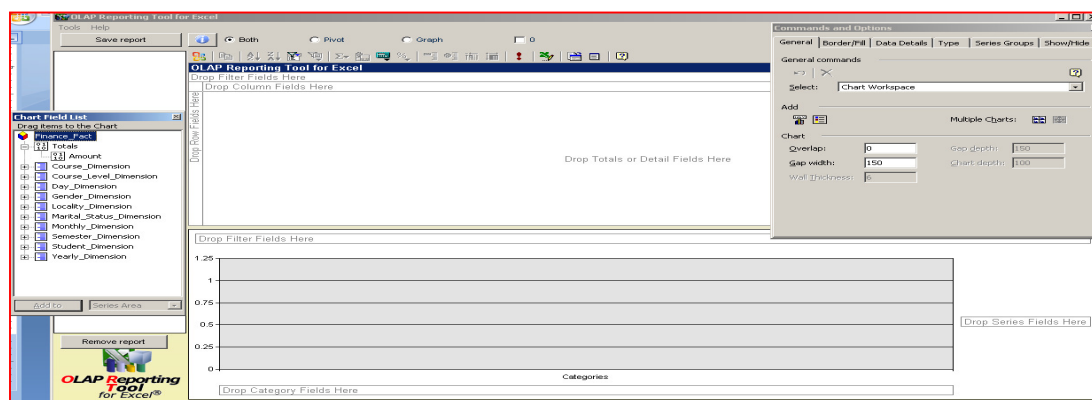


Figure 3: OLAP Reporting Tool for Excel

With BI for Excel, workbooks are dynamically linked to BI data. Better yet, it is easy to turn the analysis into a report as shown in table 1 and Fig. 4. Excel users merely needs to error-check a workbook, clean it up, and then distribute the results, without involving the IT Department. This tool is used to connect to Ms Analysis Services Server by specifying server name.

Table 1: Fee Collection per Course per Year

	Certificate		Degree		Diploma		Higher Diploma	
	Amount	Percent	Amount	Percent	Amount	Percent	Amount	Percent
2008	748,567	2.95%	2,448,740	3.37%	21,853,923.	14.01%	2,424,550	8.72%
2009	7,593,106	29.88%	11,811,355	16.25%	33,046,219	21.18%	6,931,565	24.94%
2010	9,545,938	37.56%	33,530,481	46.13%	43,745,542	28.04%	8,212,114	29.54%
2011	7,526,186	29.61%	24,898,093	34.25%	57,391,060	36.78%	10,227,920	36.80%
Total	25,413,797	100.00%	72,688,669	100.00%	156,036,744	100.00%	27,796,149	100.00%

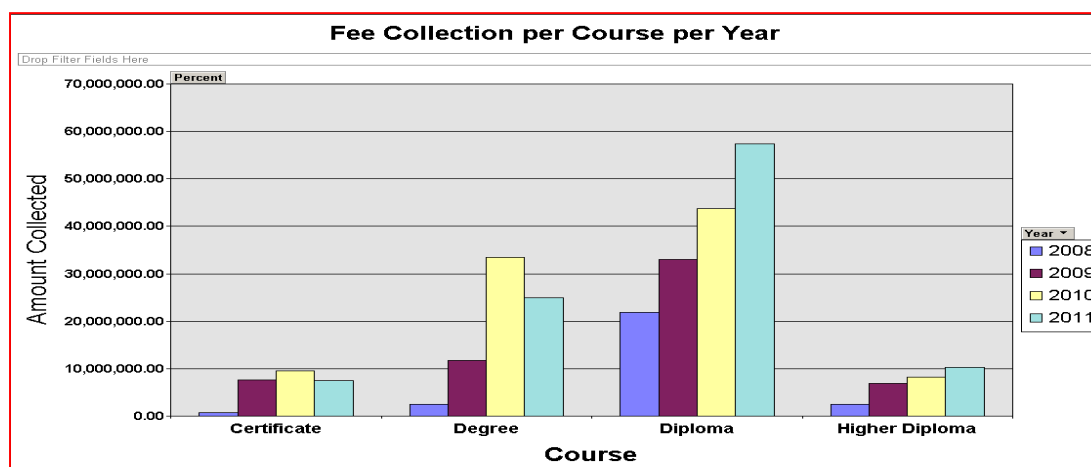


Figure 4: Fee Collection per course per year

The BI result shows the college top management that there has been a constant increase in revenue collection from diploma and higher diploma courses throughout the four year period. This therefore means that diploma courses are the key income generators to the university. It is important also to note that both certificate and degree courses showed increase in revenue in 2009 and 2010 but a drop in 2011. The management team therefore should explore this target population of large percentage of students who do not qualify for degree programs but have the potential to continue pursuing after completing secondary education. The college courses should be restructured to consider offering courses that can generate more revenue.

The college should implement a data warehouse using existing resources to address managerial gaps that hinder decision making process. The study showed that existing resources or equipments, the prevailing environment, and manpower the college can meet 80% of the required resources to realize data warehousing system. The issue of cost should not arise but instead complexity involved in implementing a data warehouse may be looked into and be varied.

166 **4. CONCLUSION**

167 This research sought to establish the level of dependency between top management in a
168 public university in Kenya and IT personnel in accessing, analyzing and reporting data. We
169 sought to determine the most valuable way to analyze data quickly and turn analysis into
170 reports with little or no help from IT department.

171 The study successfully implemented a data warehouse database composed of a single
172 multidimensional cube based on subject area: Finance. This database brought together data
173 from multiple sources providing principals and decision makers of the college greater insight
174 into the college financial performance.

175 By successfully implementing a Business Intelligence (BI) program based on Excel, we
176 turned ordinary spreadsheets into a flexible, powerful, and inexpensive BI system. An Excel-
177 friendly OLAP programs gives users significant power and flexibility, while eliminating many
178 problems associated with writing values to cells.

179 The college management should be provided with an integrated, user-friendly interface that
180 will dynamically allow them to access required data, analyze it and transform results of the
181 analysis into a report with minimal intervention from IT department.

182 The study can be extended by including a larger number of data warehouse schemata such
183 as the fact constellation schema, multi star schema, starER schema. The consistency of the
184 data warehouse system performance can be further investigated using larger datasets for
185 the dimension tables and with different aggregation levels.

186 **COMPETING INTERESTS**

187 The authors would like to declare that there are no competing interests in this paper.

188 **REFERENCES**

- 189 1. Akintola, K.G., Adetunmbi, A.O., Adeola, O.S. (2011). Building data warehousing
190 and data mining from course management systems: a case of Federal University of
191 Technology, Information Technology for People-Centred Development (ITePED 2011).
192 Available at www.ncs.org.ng
- 193 2. Dell'Aquila, C., Di Tria, F., Lefons, E., & Tangorra, F. (2008). Business intelligence
194 applications for university decision makers. WSEAS Transactions on Computers, 7(7), pp.
195 1010-1019.
- 196 3. Dell'Aquila, C., Di Tria, F., Lefons, E., & Tangorra, F. (2007). An academic data
197 warehouse. In Proceedings of the 7th WSEAS International Conference on Applied
198 Informatics and Communications, pp. 229-235.
- 199 4. Di Tria, F., Lefons, E., & Tangorra, F. (2012). Research Data Mart in an Academic
200 System. In Engineering and Technology (S-CET), 2012 Spring Congress on (pp. 1-5). IEEE.
- 201 5. Furlow, G., (2001). The Case for Building a Data Warehouse, IT Professional 3.4, pp
202 31 – 34
- 203 6. Gallegos, F. (1999). Data warehousing: A strategic approach. Information Strategy:
204 The Executive's Journal, 16(1), pp. 41-47.
- 205 7. Heise, D. L. (2005). Data warehousing and decision making in higher education in
206 the United States, PhD Dissertation, Andrews University.
- 207
- 208

- 209 8. Microsoft. Populating a Data warehouse with Microsoft SQL Server 2000 Data
210 Transformation Services, 2092A
- 211 9. Prasad, K.V.K.K., (2005). Data warehouse development tools. Dreamtech Press
- 212 10. Weisensee, D., Matthews, E., & Mcinnis, A. (2005). Implementing Data
213 Warehousing and Business Intelligence at McMaster University Using the SAS Intelligence
214 Value Chain. Apr, 10, 10.
- 215 11. Wierschem, D., McMillen, J., & McBroom, R., (2003). What academia can gain from
216 building a data warehouse, Educause Quarterly, No. 1, 41 - 46
- 217 12. Velicanu, M., & Matei, G. (2007). Building a Data Warehouse step by step.
218 Informatica Economica, 11(2), 83-89.
219