

Prototyping an Academic Data Warehouse: Case for a Public University in Kenya

Christopher A Moturi^{1*}, Alfred Emurugat²

^{1,2}*School of Computing and Informatics, University of Nairobi*

ABSTRACT

This study aimed at determining the gaps between university top decision makers in a public university in Kenya and 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 Fee Payment. 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

Research has shown that only a small fraction of data in an organization is timely available to decision makers (Hassan et al., 2013; Davenport, 2010). A set of significant new concepts and tools have evolved into new technology that makes it possible to address the problem by providing all the key people in the organization with access to whatever level of information needed for the organization to survive and prosper in an increasingly competitive world.

Data warehousing technology makes it easier for decision makers to analyze information in an organization without the help from Information Technology (IT) people. The main objective of this study was to determine the gaps between top decision makers and IT personnel in accessing, analyzing and reporting data. The other objectives were to determine whether there is an easier and quicker way to analyze data, design, and implement a solution to be used to turn analysis into a report with little or no help from IT department.

A data warehouse is a database containing data from multiple operational systems that has been consolidated, integrated, aggregated and structured so that it can be used to support the analysis and decision making process of an organization (Hobbs et al., 2011).

There has been a growing interest in Business Intelligence (BI) as a solution to reduced costs, improved service quality, and better decision-making processes (Foley and Guillemette, 2010). Zeng et al. (2006) define BI as the process of collection, treatment and diffusion of information that has an objective, the reduction of uncertainty in the making of all strategic decisions. BI is viewed as sets of powerful tools and approaches to improving

* Tel.: +xx xx 265xxxxx; fax: +xx aa 462xxxxx.
E-mail address: xyz@abc.com.

35 business executive decision-making, business operations, and increasing the value of the
36 enterprise. BI allows managers to make informed and intelligent decisions regarding the
37 functioning of their organization. BI is the set of techniques and tools for the transformation
38 of data into meaningful and useful information for business analysis purposes. The common
39 functions of business intelligence technologies are reporting, online analytical processing,
40 analytics, data mining, process mining, business performance management, text mining, and
41 predictive analytics.

42 The BI technology encompasses data warehousing, OLAP, and data mining. After
43 examining how well BI capabilities are leveraged to achieve BI success, Işık et al. (2013)
44 suggested that technological capabilities such as data quality, user access and the
45 integration of BI with other systems are necessary for BI success, regardless of the decision
46 environment. Tvrdíková (2007) describes the ability of BI tools to collect data from
47 heterogeneous source, use advance analytical methods and support multi user demands.
48 Herschel and Jones (2005) provided an analysis of the difference between business
49 intelligence BI and knowledge management (KM). They reported that BI focuses on explicit
50 knowledge while KM encompasses both tacit and explicit knowledge and BI should be
51 viewed as a subset of KM. Herschel and Jones (2005) provided an analysis of the difference
52 between business intelligence BI and knowledge management (KM). They reported that BI
53 focuses on explicit knowledge while KM encompasses both tacit and explicit knowledge and
54 BI should be viewed as a subset of KM. Albescu et al. (2008) reported that IT solutions are
55 designed to address the challenges of an organization using two different approaches:
56 structured data management (Business Intelligence) and unstructured content management
57 (Knowledge Management). They have proposed a design of data warehouse in attempt to
58 apply BI technology to economic environment analysis using public data sources.

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

65 The constituent college uses both operational systems and Enterprise Resource Planning
66 (ERPs) to manage its day-to-day operations. These systems include financial management,
67 human resources management, student information, library, examination management, and
68 course management. These applications typically run in an On-Line Transaction Processing
69 (OLTP) computing environment based on business functional areas.

70 However, the college lacks systems that will foster decision making. The majority of
71 businesses, including educational institutions and government agencies, realize the need for
72 easy and quick access to the right information (Akintola et al., 2011).

73 Different organizations face issues and challenges when implementing Data Warehouse
74 technologies. This includes management (management commitment and support, project
75 management, user involvement and participation, skills) and technological (selection of DW
76 architecture, creation of the enterprise schema, data integration and scalability, data quality,
77 design of human-computer interfaces, data mining, security and privacy risks, and networks
78 and telecommunication) (Al-Debei, 2011). However, Data Warehousing can yield real
79 dividends in terms of competitive advantage and business intelligence (Sharma and Jain,
80 2013; Alenazi et al., 2014). The university needs a data warehouse that can unlock its data
81 in various ERPs and other operational systems (Heise, 2005, Wierschem et al. 2003). A
82 university needs a different approach or technique to provide a centralized source of

information accessible across different academic units. The key stakeholders should quickly analyze and develop satisfactory solutions to their underlying problems as well as making better business decisions based on historical data available in the legacy databases. Research has shown universities can improving decision making in various areas using data warehousing and data mining. This includes student academic performance (Sreenivasarao and Yohannes, 2012), library management (Laitinen and Saarti, 2012), decision support systems (Gao and Chen 2015), and ETL (Extract, Transform and Load) process (Mekterović et al., 2009).

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

The ad hoc analysis and reporting impossibilities significantly hinder the decision making process. This study therefore sought to show how BI can be the most efficient way to analyze data quickly and turn analysis into reports with little or no help from IT department.

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

2. METHODOLOGY

2.1 Research Design

A survey was conducted to ascertain the gaps between top decision makers in the college and IT department in accessing, analyzing and reporting data. A sample size of 97 employees was selected from the entire population of about 430 and issued with questionnaires with 74 questionnaires returned.

2.2 Survey Analysis

An analysis was done to determine the understanding and perception of the college users on several aspects of customer service on the following items: how long it takes to create reports, if reports are produced with the help of technician, access to all required data dynamically, dissimilar reports produced later, errors in the report due to source data, reports are not developed quickly and easily, users don't have direct access to operational data, overall, rating of the quality of data analysis and reporting.

The current situation is that the user interfaces do not allow decision makers to explore, analyze data, and create reports quickly or easily without the help from IT people. The majority (75.7%) of the decision makers depend on IT professionals in producing reports. The majority of the respondents (54.1%) were of the opinion that currently, typical reports takes at least a day to prepare. Many decision makers (48.6%) understand what the data means, but they do not have easy or dynamic access. Some of the reports (13.5%) prepared either contain errors or doctored to suit its purpose. A third (33.8%) of the respondents claims that similar reports are generated several months later against similar data. The quality of the reports produced is not up to standard. Only 21.7% of the respondents feel satisfied with the quality of data analysis and reporting. Majority (93.2%) of university decision makers are competent excel users although more than half have not used Excel as

131 a BI. The survey therefore established there are managerial gaps in the college that hinder
132 decision making process and that need to be addressed.

133

134 2.3 Data Warehouse Design

135 After literature survey, we considered data warehouse as the most valuable technology to be
136 used (Di Tria et al. 2012, Dell'Aquila et al., 2008, Dell'Aquila et al., 2007). This study
137 considered and applied star schema dimensional modeling techniques to design an
138 integrated data structure required in a data warehouse (Prasad, 2005). A data warehouse is
139 composed of subject-oriented data mart using one or more schemas to represent specific
140 business events or processes. A star schema contains a central fact table surrounded by
141 many dimensions tables. The design of dimension tables for the analytical needs of the
142 college is shown in Fig. 1, with the fee payment being fact table and the dimensions
143 comprising students, courses, employees (staff) and time.



144

145

Figure 1: Star Schema Design of the Dimension Tables

146

147 Since data was from disparate sources, transformation process was carried out which
148 involved conversion, summarization and filtering of data. Data Transformation Services
149 (DTS) tool was used to extract data from various sources including Course Management
150 Information System (CMIS) in MS SQL server, Student Management System (SMS) in MS
151 Access 2000, data files, external files and MS Excel spreadsheet (Fig. 2). The extracted data
152 was transformed into a format that can be loaded into the data warehouse. The staging area
153 is used to facilitate the transformation process. The transformed data was then loaded into
154 dimension tables before being loaded into the fact table that references the dimension.

155

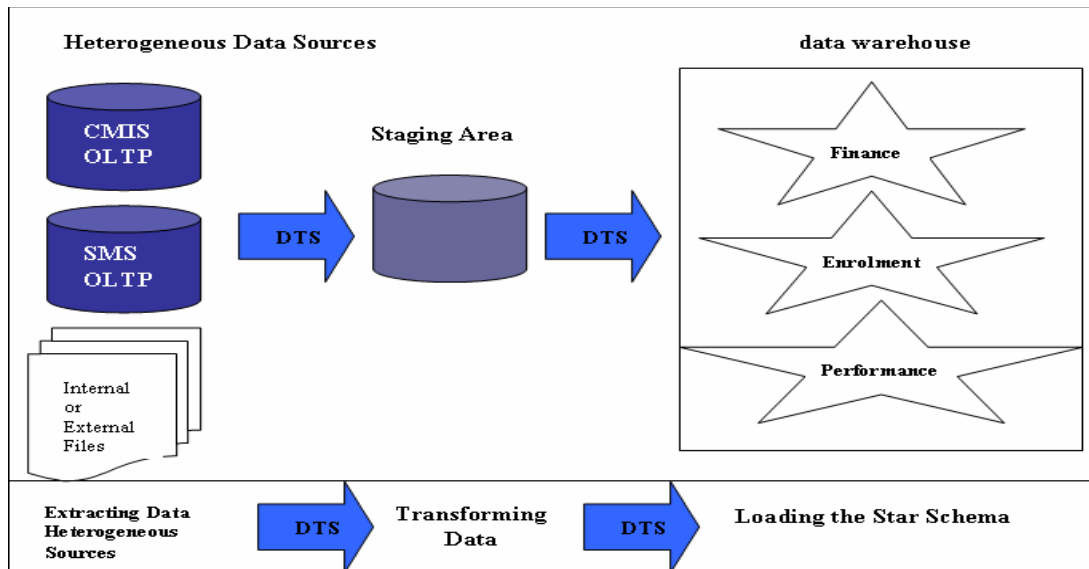


Figure 2: University Data Marts

A data mart is the access layer of the data warehouse environment that is used to get data out to the users. The data mart is a subset of the data warehouse that is usually oriented to a specific business line or community of workers. A cube is an array of data understood in terms of its dimensions. After implementation, cubes are the basis for an easy, but very rich analysis and reporting experience for knowledge workers.

Online Analytical Processing (OLAP) reporting based on Excel was identified as the most suitable Business Intelligence (BI) tool. OLAP is an approach to answering multi-dimensional analytical queries and is part of the broader category of business intelligence. Typical applications of OLAP include business reporting for management reporting and business process management. OLAP reporting tool based on MS Excel gives decision makers significant power and flexibility to access, analyze and reports data.

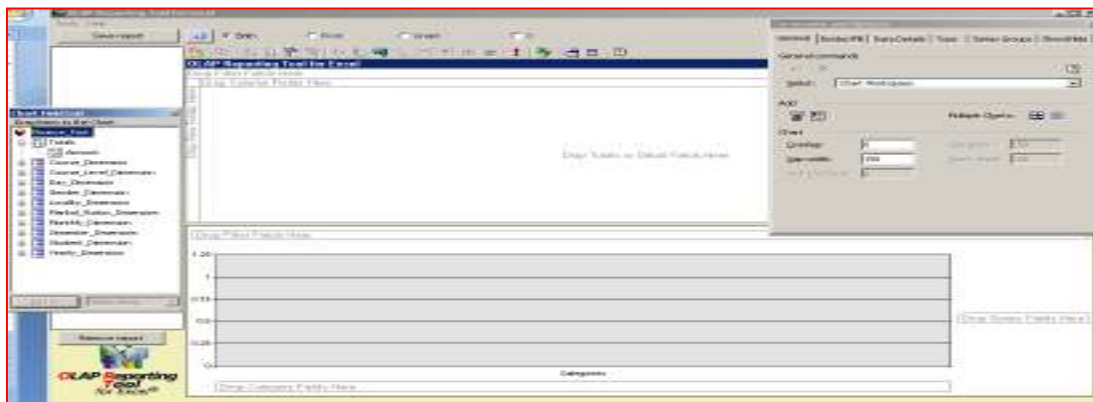
3. RESULTS AND DISCUSSION

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 single data mart – fee payment - that tracked revenue collected from students in various courses. The fee payment cube contained only one measure called total Amount - fee collection from students taking different courses between January, 2008 and December, 2011. The dimensions contained in the cube were students, course, employees and time. The time dimension consisted of a hierarchy by year, semester and month. 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.

187 To validate the data warehouse, a Business Intelligence (BI) prototype based on Excel was
 188 implemented. OLAP Reporting Tool for Excel is an example of Excel-based BI system as
 189 shown in Fig. 3. The working environment consists of a collection of integrated windows:
 190 OLAP Report Tool for Excel main window, Command and Option window and Chart Field
 191 List, designed to support a DSS; a working Screen section - designed for dropping different
 192 fields depending report to be created; and OLAP Reporting Tool for Excel that gets its data
 193 from the data warehouse.

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



198

199 **Figure 3: OLAP Reporting Tool for Excel**

200 The results of the prototype implementation showed the growth in the revenue collected in
 201 the respective years. Revenue collection in 2009 and 2010 went up by 11.32% and 12.65%
 202 respectively compared to the previous respective years. However, there was sharp drop in
 203 2011, constituting 1.78% although still more revenue was collected compared to the
 204 previous year (Table 1, Fig. 4).

205 **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%

206

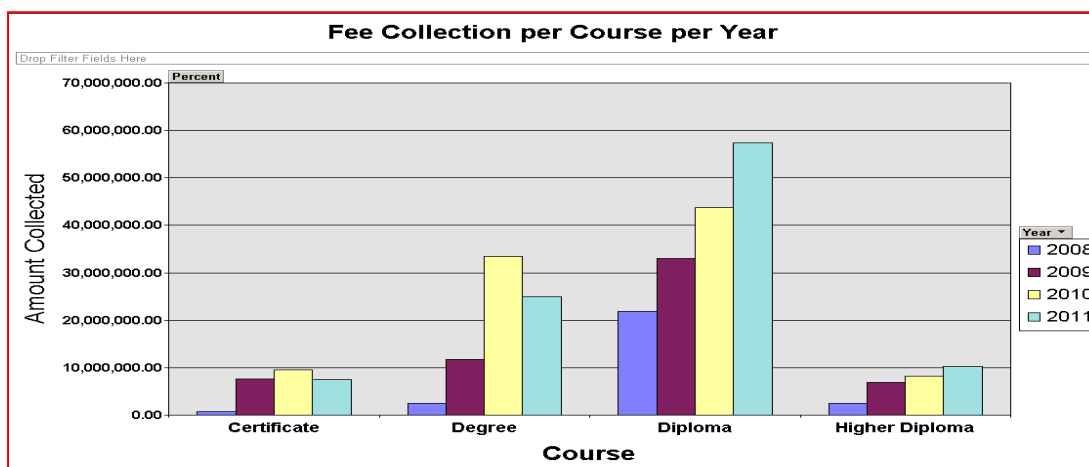


Figure 4: Fee Collection per course per year

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. 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.

Further results showed analysis of the fee collection by semester per year, fee collection per month year, fee collection per degree course name per year, most reliable and highly income generating course, and distribution of courses offered at the college.

Drill down was created to enable users retrieve from the source system. When a trend, an exception, or an error is identified, users can drill down to the transaction level to find the source data that generated this behavior. Users can navigate and analyze data from both types of databases in a single client environment.

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 at the college are adequate 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.

4. CONCLUSION

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

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

239 By successfully implementing a Business Intelligence (BI) prototype based on Excel, we
240 turned ordinary spreadsheets into a flexible, powerful, and inexpensive BI system. The
241 college management is provided with an integrated, user-friendly interface that will
242 dynamically allow access to required data, analyze it and transform results of the analysis
243 into a report with minimal intervention from IT department.

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

248
249 The newest approaches or techniques in data warehousing require the use of Big Data
250 Analytics (Picciano, 2012). Adoption of this approach will enable the college to analyze a mix
251 of structured, semi-structured and unstructured data in search of valuable business
252 information and insights.

253
254

255 **COMPETING INTERESTS**

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

257

258 **REFERENCES**

259 1. Akintola, K.G., Adetunmbi, A.O., Adeola, O.S. (2011). Building Data Warehousing and
260 Data Mining from Course Management Systems: A Case of Federal University of
261 Technology, Information Technology for People-Centred Development (ITePED 2011).
262 Available at www.ncs.org.ng

263 2. Albescu, F., Pugna, I., & Paraschiv, D. (2008). Business intelligence & knowledge
264 management-Technological support for strategic management in the knowledge based
265 economy. Journal of information economy, V, 8, 67-82.

266 3. Al-Debei, M. M. (2011). Data Warehouse as a Backbone for Business Intelligence:
267 Issues and Challenges. European Journal of Economics, Finance and Administrative
268 Sciences, 33, 153-166.

269 4. Alenazi, S. R. A., Gholamdin, A. G., Alanzi, A. R. A., & Al-Matari, E. M. (2014).
270 Prototyping Data Warehouse System for Ministry of Higher Education in Saudi Arabia.
271 Computer and Information Science, 7(4), 74 - 81

272 5. Davenport, T. H. (2010). Business intelligence and organizational decisions.
273 International Journal of Business Intelligence Research (IJBIR), 1(1), 1-12.

274 6. Dell'Aquila, C., Di Tria, F., Lefons, E., & Tangorra, F. (2008). Business Intelligence
275 Applications for University Decision Makers. WSEAS Transactions on Computers, 7(7),
276 pp. 1010-1019.

277 7. Dell'Aquila, C., Di Tria, F., Lefons, E., & Tangorra, F. (2007). An Academic Data
278 Warehouse. In Proceedings of the 7th WSEAS International Conference on Applied
279 Informatics and Communications, pp. 229-235.

280 8. Di Tria, F., Lefons, E., & Tangorra, F. (2012). Research Data Mart in an Academic
281 System. In Engineering and Technology (S-CET), 2012 Spring Congress on (pp. 1-5).
282 IEEE.

- 283 9. Foley, É., & Guillemette, M. G. (2010). What is business intelligence? *International*
284 *Journal of Business Intelligence Research (IJBIR)*, 1(4), 1-28.
- 285 10. Gao, L., & Chen, Y. (2015). Application Research of University Decision Support System
286 Based on Data Warehouse. In *Computer Science and its Applications* (pp. 631-637).
287 Springer Berlin Heidelberg.
- 288 11. Hasan, Y., Shamsuddin, A., & Aziati, N. (2013). The Impact of Management Information
289 Systems Adoption in Managerial Decision Making: A Review. *Management*, 8(4), 010-
290 017
- 291 12. Heise, D. L. (2005). Data warehousing and decision making in higher education in the
292 United States, PhD Dissertation, Andrews University.
- 293 13. Herschel, R. T., & Jones, N. E. (2005). Knowledge management and business
294 intelligence: the importance of integration. *Journal of Knowledge Management*, 9(4), 45-
295 55.
- 296 14. Hobbs, L., Hillson, S., Lawande, S., & Smith, P. (2011). *Oracle 10g Data Warehousing*.
297 Digital Press.
- 298 15. Işık, Ö., Jones, M. C., & Sidorova, A. (2013). Business intelligence success: The roles of
299 BI capabilities and decision environments. *Information & Management*, 50(1), 13-23.
- 300 16. Laitinen, M., & Saarti, J. (2012). A model for a library-management toolbox: Data
301 warehousing as a tool for filtering and analyzing statistical information from multiple
302 sources. *Library Management*, 33(4/5), 253-260.
- 303 17. Mekterović, I., Brkić, L., & Baranović, M. (2009). Improving the ETL process and
304 maintenance of higher education information system data warehouse. *WSEAS*
305 *transactions on computers*, 8(10), 1681-1690.
- 306 18. Picciano, A. G. (2012). The Evolution of Big Data and Learning Analytics in American
307 Higher Education. *Journal of Asynchronous Learning Networks*, 16(3), 9-20.
- 308 19. Piedade, M. B., & Santos, M. Y. (2011). Semana da Escola de Engenharia October 24-
309 27, 2011 Business Intelligence *in Higher Education Promoting the students success
310 with a SRM system.
- 311 20. Prasad, K.V.K.K., (2005). Data warehouse development tools. Dreamtech Press
- 312 21. Sharma, S., & Jain, R. (2013). Enhancing Business Intelligence using Data
313 Warehousing: A Multi Case Analysis. *International Journal*, 1(7).
- 314 22. Sreenivasarao, V., & Yohannes, C. G. (2012). Improving Academic Performance of
315 Students of Defence University Based on Data Warehousing and Data mining. *Global*
316 *Journal of Computer Science and Technology*, 12(2).
- 317 23. Tvrdikova, M. (2007). Support of decision making by business intelligence tools. In
318 *Computer Information Systems and Industrial Management Applications*, 2007.
319 *CISIM'07. 6th International Conference on* (pp. 364-368). IEEE.

- 320 24. Weisensee, D., Matthews, E., & Mcinnis, A. (2005). Implementing Data Warehousing
321 and Business Intelligence at McMaster University Using the SAS Intelligence Value
322 Chain. Apr, 10, 10.
- 323 25. Wierschem, D., McMillen, J., & McBroom, R., (2003). What academia can gain from
324 building a data warehouse, Educause Quarterly, No. 1, 41 - 46
- 325 26. Velicanu, M., & Matei, G. (2007). Building a Data Warehouse Step by Step. Informatica
326 Economica, 11(2), 83-89.
- 327 27. Zeng, L., Xu, L., Shi, Z., Wang, M., & Wu, W. (2006, October). Techniques, process, and
328 enterprise solutions of business intelligence. In Systems, Man and Cybernetics, 2006.
329 SMC'06. IEEE International Conference on (Vol. 6, pp. 4722-4726). IEEE.