Information and Knowledge Management
ISSN 2224-5758 (Paper) ISSN 2224-896X (Online)
Vol.3, No.10, 2013
www.iiste.org

Security Architecture for Tanzania Higher Learning Institutions' Data Warehouse

Chande Kasita[1]*, Loserian .S. Laizer[1]

1. School of Mathematical, Computational and Communication Science and Engineering, Nelson Mandela African Institute of Science and Technology, Arusha, Tanzania, PO box 447, Arusha, Tanzania,

* E-mail: kasitac@nm-aist.ac.tz

Abstract

In this paper we developed security architecture for the higher learning institutions in Tanzania which considers security measures to be taken at different level of the higher learning institutions' data warehouse architecture. The primary objectives of the study was to identify security requirements of the higher learning institutions data warehouses and then study the existing security systems in and finally develop and architecture based on the requirements extracted from the study. The study was carried at three different universities in Tanzania by carrying out interviews, study of the existing systems in respective institutions and a literature review of the existing data warehouses systems and architectures. The result was the security requirements identified which lead to the development of the security architecture comprising security in source systems, data, and services to be offered by the DW, applications which use DW, networks and other physical infrastructure focusing on security controls like authentication, role-based access control, role separation of privileged users, storage of data, secure transfer of data, protective monitoring/ intrusion detection, penetration testing, trusted/secure endpoints and physical protection.

Keywords: Data warehouse, security architecture, higher learning institution.

1. INTRODUCTION

Higher learning institutions (HLIs) as the driver of development in society required to maintain accuracy decision making based on facts derived from the transaction and legacy systems and integrated in the repository which stores aggregate and historical which shows past business transactions and decisions. The repositories and associated technologies are known as data warehouse and is immerged as one of the important tool to facilitate decision making in academic institutions (Breiter. A & Light.D, 2006).

The quality of decision making and analysis performed based on data warehouse depends on the integrity, accuracy and consistence of the data in the data warehouse. Furthermore the HLIs transaction systems are residing in different institutions and departments which possess different security policies, legal jurisdiction and level of confidentiality of the data in question. For example, it is illegal to expose students' personal data in USA (ISACF (1998a)).

Also due to distribution nature and heterogeneity of the source systems in HLIs the vulnerability of the DWs systems from the source systems, staging areas, data transferring tools and the applications which make use of the DWs data become high.

The aim of this paper is to discuss the security requirements of the HLI DW and gives an overview of the security architecture model of the HLI DW. The model will be developed focusing on the five data warehouse goals which relates to security issues namely availability, confidentiality, integrity, Reliability, legal/compliance (FERPA,2013) and shows the structure, controls, tools (technologies and protocols) at different HLI DW layers which addresses the identified requirements.

2. DATA WAREHOUSING IN HIGHER LERANING INSTITUTIONS

Higher learning institutions have distinct information systems which support different types of decisions like administrative information systems, learning management systems and assessment information systems (Breiter. A& Light.D, 2006). To make fact based analysis and decision making accurate and efficient the DW technology can be used. The HLI DW consists of data, services, applications, user interfaces, user groups and the networks which connect different components (Bhanti.P. et al., 2011).

Fig 1. depicts the logical architecture of higher learning institutions' data warehouse with the data layer which describes the data to be logically visible and consolidated in their respective domain, the services and data integration layer which stresses the need for having consistence way of interacting with data across institutions and provide coherent way of sharing data and exposing services in consistent way whereas enabling services to be implemented in a variety of technologies. The architecture also consist of service layer which shows the separation of re-usable services form the application logic and categorized services in administrative services, academic services and core services which includes technical and basic services. Application layer shows the grouping of subject area specific applications and user interface layer exposes those applications with similar functions on single point of access in the form of portals, standalone GUI, web interface and mobile interface. The architecture also defined the user groups which enable to identify the requirements of the users and necessary access levels. The network and security layer extends throughout the architecture and provides connectivity and security services respectively.

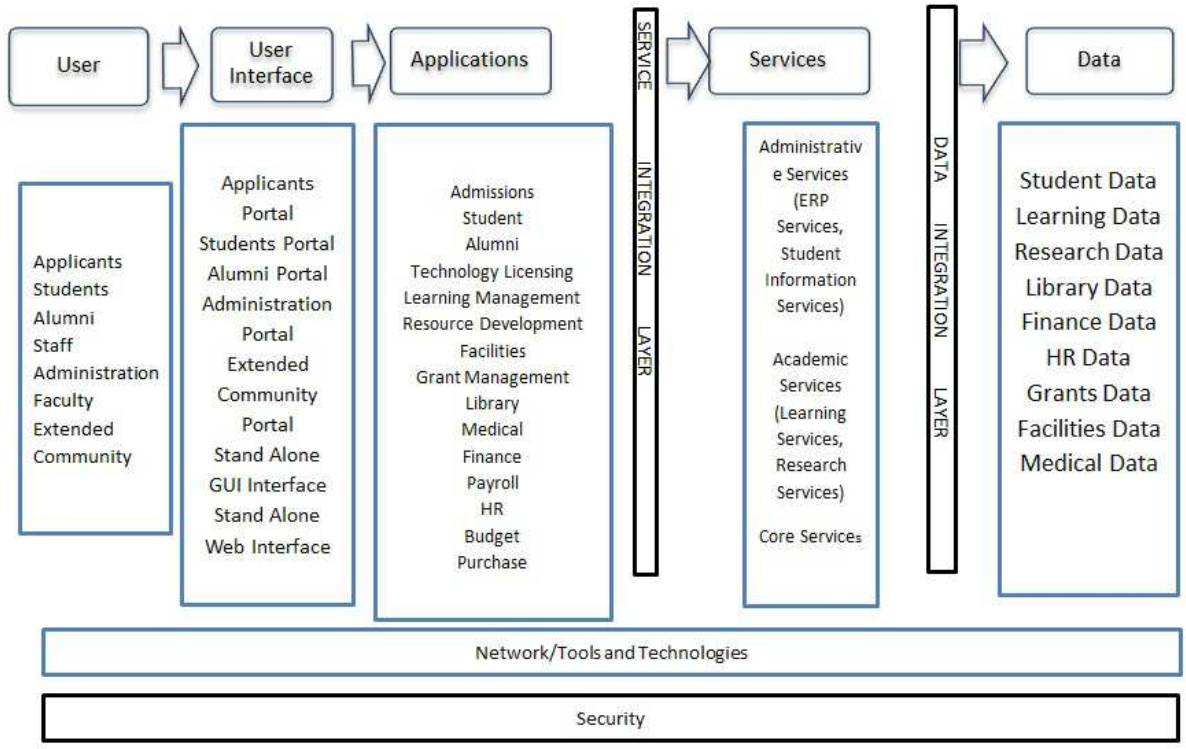

Fig 1. Higher Learning Institutions' Data warehouse architecture.

3. SECURITY IN HIGHER LEARNING INSTITUTIONS' DATA WAREHOUSE.

Owing to the structure, functions, infrastructure, people and processes existing in the higher learning institutions there must be a proper security model to protect the data warehouse architecture. Different authors have put forward the security architectures to be implemented in data warehouses (Bhanti.P et al., 2011),(Priebe.T & Pernul.G),(Katic.N et al),(Weaver.A.C,2007).

These approaches to security in DWs, focused, for instance, on access control and multilevel security Priebe.T & Pernul.G but neither of them considers security of the DW in all stages of the DW model from the source systems to ETL processes and finally to the applications tools like portals, OLAP and data mining tools. Moreover, other authors proposed the use of security model used in transactional databases, centered on tables, rows, and attributes(Kumar.C.S et al,2012), which inappropriate for DW and should be replaced by an ad hoc model centered on the main concepts of multidimensional modeling such as facts, dimensions, and measures data models.

Therefore we regard a multi-layers consideration to the security in higher learning institutions data warehouse as the best approach to address the security requirements and ensure optimal performance and security output.

3.1 Security Restrictions in HLI-DW

Higher learning institutions' data warehouses are characterized by high data volume growth rate (Yanosky.R, 2009), high data quality requirements, large number of users who requires access to wide range of data for analytical purposes and open nature to make it easily accessible. These characteristics pose differing insight to the security requirements as many of the security measures to the data warehouse in higher learning institutions results in compromising performance of the data warehouse (Yanosky.R, 2009).
Due to sensitivity of the data and the overall data warehouse's integrity, availability, confidentiality legal and compliance requirements it is worth taking the security measures apart from the performance bottleneck which may arise.

3.2. Security requirements in HLI-DW

The HLI DW contains all of the users' credentials data, institutions administrative, academic and financial data, and other information which should be protected from hackers who would try to infiltrate the system. In addition to that data are transferred from different systems and required to reach destinations unaltered so as to guarantee the integrity of the data and consequently the quality of the decision made through stored data and information. Hence the employment of various security measures like Secure Socket Layer (SSL) encryption, digital certificate to give security to involved institutions, firewalls, physical security and security policies becomes inevitable in the HLI DW. These security controls aimed at addressing fundamental security requirements for HLI DW namely legal and compliance requirements, integrity, confidentiality, availability, disaster recovery and business continuity (Kimball.R and Ross.M, 2000).

3.2.1. Legal/Compliance Requirements:

It is important to comply with regulatory, contractual and legal obligation in ensuring security in higher learning institutions data warehouse as it helps to abide with copyright and patent laws, data protections acts, computer misuse acts, human rights act and any relevant legislations and by so doing prevents institutions to suffer from any breach of existing legal and compliance obligations.
The following are some of the requirements:

- Policy to govern legally sensitive data.
- Identification and sorting of data subjected to legal restrictions.
- Policy to dictate the handling of storage, access and maintenance of data.
- Policy to govern the limit of analyses to be performed on the particular data.
- Policy which states the existing legal, contractual and compliance obligations existing in the particular institutions' jurisdiction and statement on which data is localized and which are to be shared among institutions.

These requirements will ensure data warehouse users and data custodian are aware on the existing compliance and legal controls to be followed and protect data, infrastructure and institutions as a whole.

3.2.2. Integrity requirements:

The data integrity of DW ensures the accuracy of the data. Maintenance of integrity of DW helps guarantee that any reporting or analysis out of the DW provides an accurate representation of the data relating to the higher learning institutions. All parties involved have an important role in this goal. The DW should enforce referential integrity to its dimensional model as one of integrity requirements (Kimball.R and Ross.M, 2000).

3.2.3. Availability requirements:

- There must be a maximum availability of the higher learning institutions data warehouse during core service hours. The DW is expected to be of use to all higher learning institutions, research communities and other stakeholders. In the country like Tanzania core service hours can be defined as 08:00 to 16:00 Monday to Friday and lower service levels could be provided outside of these hours.
- There should be a pre-planned downtime and kept as minimum as possible to avoid interference with operations and use of DW services. Planned downtime should be provided in Service Level Agreement (SLA) and people/organ responsible for network and system availability.
- Batch update periods uploads of data to the DW should not affect the availability of the service.
- The HLI DW should be continually monitored to immediately identify disruption in service availability, such that actions can be taken to rectify the problem and restore services.
- Regular reporting of DW availability will be required in support of the service. It is anticipated that this

reporting will occur in a number of forms and frequencies.
- The service should not be affected by the loading of data into the service or the execution of long running queries.

3.2.4. Confidentiality requirements:

Information confidentiality is a significant security issue. Loss of confidentiality of information, especially information of which you are the custodian rather than the owner e.g. institutional financial information can cause problems like loss of money through frauds (as a result of compromised identity information), embarrassment (through malicious publication of lost / stolen information), legal challenges (through breach of contractual clauses) and breaking the law or regulations (where data protection is mandatory).Therefore confidentiality requirement is important to ensure disclosure of the confidential data.
Some of the confidential requirements include:

- The HLI DW must implement authentication technique so as to ensure only authentic users are accessing the DW.
- The DW must implement role based access control to limit access to applications and functions defined by users' roles and permissions.
- There should be a role separation of privileged users for different aspects of the system e.g. standard users, administrators etc.
- Storage of confidential data in DW must be physically protected against theft and be stored in encrypted format or environment that is physically secure against other forms of unauthorized access.
- Data transferred over non-secure networks must be encrypted using an appropriate standard, e.g. Secure Sockets Layer (SSL).
- Data transferred by removable media must equally be encrypted in transit. Additionally, the media must be protected from theft whilst they are in transit and processes implemented to ensure guaranteed delivery.
- The HLI DW must be protected by an intrusion detection system.
- Regular penetration testing must be undertaken, with appropriate corrective action taken to improve the security of the HLI DW following the results of the testing.
- Access to the HLI DW for data exchange (upload or extraction), management or general purposes must be achieved from trusted endpoints.

3.2.5. Disaster recovery/business continuity requirements:

There must be business continuity plan to ensure that whenever a DW suffers any natural or man- made disaster the system recovers in time and provides the service as previously was. The following issues must be clarified in ensuring business continuity and disaster recovery:

- What is the mean time to recover (MTTR) in the event of loss?
- What are measures put in place to ensure zero data loss any event?
- Which activities are to take place to manage the service in the event of a disaster leading to loss or potential loss of service?
- Which supporting processes and procedures e.g. escalation and invoking the disaster recovery plan should be taken in case of any disaster.

4. HLI-DW SECURITY ARCHITECTURE

We developed HLI DW security architecture by Identifying data and nature of data to be protected type of threats to which each type of data is exposed e.g. Accidental corruption or destruction, deliberate corruption or destruction, hacking and viruses. We then identify the threats to physical security like theft of components, unauthorized physical access and physical theft of users and people who may be source of these threats e.g. IT and non-IT personnel and people outside institutions.
We finally designed an architecture which shows areas and objects which require protection by logical or physical security to encounter any possible threats. Fig 2.Depicts the security architecture which shows different controls to be employed in different stages of HLI DW.
The architecture is described by area to be secured and associated technologies, protocol or procedure/policy to ensure security of respective protected objects or security goal(s).
Our architecture is based on centralized/consolidated implementation model as it is easier to implement and uses lower cost while providing higher level of security (Kimball.R and Ross.M, 2000).

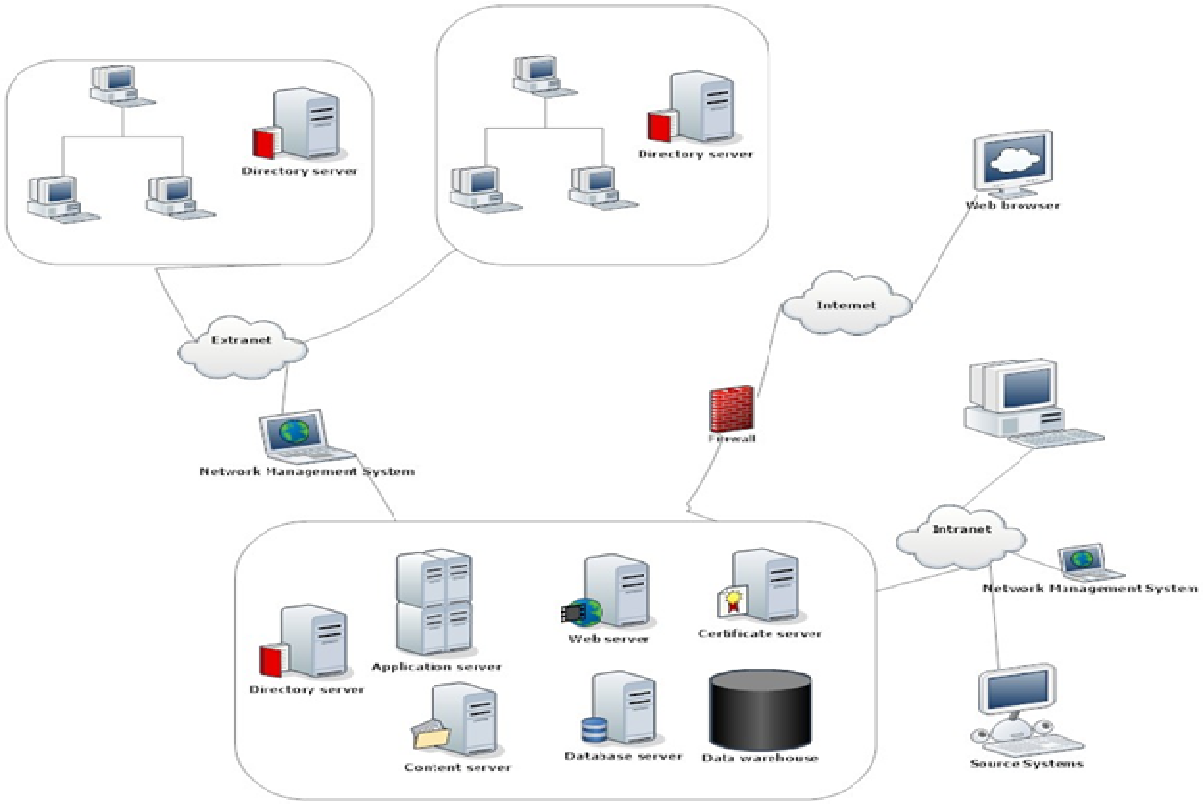

Fig 2.Higher Learning Institutions DW security architecture

4.1 Security in data storage and networks

Security of data in higher learning institution data warehouse consists of flexible multi-layer security model so that the failure of one mechanism does not lead to the security breach of entire storage system. The storage security in HLIDW converges the storage system, storage networks, security disciplines, technologies and methodologies necessary to protect stored information assets of the higher learning institutions. The HLI DW network infrastructure comprises of storage area network, intranet whereby local users of DW get access, extranet to connect different institutions and the internet to make DW accessible through the web.

4.1.1. Storage System Security: Applications, databases, file systems and server operating systems must be secured to prevent unauthorized or disruptive access to your stored data. Storage system based on volume or logical unit number mapping and masking must be implemented in HLIDW. The logical controls for protecting storage systems include authorization, authentication, disk level encryption and passwords, firewalls, running antispyware and virus-detection programs on server, regular change of key-code or door-lock combinations. The architecture suggest that the access to the data should be controlled at the database level using Virtual Private Database(VPD) and not at the application level so as to minimize implementation cost and the possible security breach through SQL*Plus or direct querying(An Oracle White Paper,2005).

4.1.2. Storage Resource Management : The DW must be implemented in a software (DBMS) which Securely provisioning, monitoring, tuning, re-allocating, and controlling the storage resources so that data may be stored and retrieved smoothly and securely to avoid any performance or security bottleneck.

4.1.3. Data In-Flight: Due to distribution nature of higher learning institutions data are subjected to frequent movement and subsequently exposure to threat on transit. So in order to protect confidentiality, integrity and/or availability of data as they are transferred across the storage network, the LAN, extranet and the WAN we propose the use of security certificate key, encryption techniques (e.g. IPsec) and Virtual Private Network on extranet environment.

4.1.4. Data At-Rest: Protecting the confidentiality, integrity and/or availability of data residing on servers, storage arrays, NAS appliances, tape libraries, and other media (especially tape).We propose the use of encryption on the storage media by using encryption server for the easy management of encryption and avoid performance hindrance which might be caused by using DBMS managed encryption (A White Paper for Developers e-Business Managers and IT (2002)).

4.2. *Security in DW user interface.*
The users of HLI DW include applicants, students, alumni, faculty, administrator, researchers from different institutions and extended general community who uses different apparatus to get access to data and applications of the DW. The interfaces used to gain access to data and applications includes applicants' portal, students' portal, alumni portal, administration portal, extended community portal, standalone GUI Interface and standalone web interface .These interfaces has to be secured so as to prevent and compromise to the security requirements.
Securing Portals, Standalone GUI and web interface:
To secure portals we must make sure the user or software request for service or application originates from the legitimate requestor of the particular resource. We propose the use of three techniques to secure HLIDW portals namely authentication, authorization and federation.

4.2.1 Authentication: To ensure the authenticity of the software processes we suggest the use of digital signatures and for human users can be attained through the use of web services-security tokens (WS-security) for XML services or any other legacy techniques such as passwords and biometric techniques like fingerprints, iris scans, and signature recognition, digital techniques such as e-tokens and Radio Frequency Identification (RFID), plus a two-factor technique that requires both a password and a random number that changes once per minute (Weaver.A.C,2007).This will enable features of the interface meant to a particular users are not exposed to illegitimate users.

4.2.2. Authorization: User interfaces should provide tool for granting of a right or privilege, which enables a subject to have legitimate access to a system's objects or a system itself. Subjects represent active entities usually acting on behalf of principals such as users whereas objects are considered to be of passive nature such as, for instance, a database table, view, or any other object that can be created within the system. There must be a policy in place to ensure that privileges granted at user interface controls only simple access issues.

CONCLUSIONS

The model presented seeks to promote the consideration of taking appropriate measures to secure higher learning institutions data warehouses at all levels of the data warehouse architecture. It also presents the essence of using multi-layered security approach so as to create backup line of defense upon the failure of one line of defense. The paper also presents the security requirements of the higher learning institutions which are geared to address confidentiality, integrity and availability of data as the main security goals. Furthermore the architecture points out some of the security controls which can be implemented at different levels of the higher learning institutions' data warehouses.

ACKNOWLEDGEMENT

The authors want to acknowledge the support provided by Nelson Mandela Africa Institute of Science and Technology for providing financial and material support for the research undertaking. He also acknowledges the guidance and comments provided by Mr.Lozerian Laizer and Professor Irina Zlotnikova which helped to shape the idea into real research work.

REFERENCES

A White Paper for Developers e-Business Managers and IT, “Securing Data at Rest: Developing a Database Encryption Strategy”, 2002.

An Oracle White Paper, “Security and the Data Warehouse”, April 2005.

Bhanti.P, Kaushal.U, Pandey.A,” E-Governance in Higher Education: Concept and Role of Data Warehousing Techniques”, International Journal of Computer Applications, Volume 18– No.1, March 2011

Breiter. A, Light.D, “Data for School Improvement: Factors for designing effective information systems to support decision-making in schools”, Educational Technology and Society, vol.9, no.3, p. 206-217, 2006.

Edge.I.E, “Employ Five Fundamental Principles to Produce a SOLID, Secure Network”, Information Security Journal: A Global Perspective archive, Volume 19 Issue 3, January 2010, Pages 153-159

ISACF (1998a), “COBIT Audit guidelines”. Information Systems Audit and Control Foundation. Rolling Meadows. IL, USA.

Katic.N, Quirchmayr.G, Schiefer.J, Stolba.M, Tjoa.A.M “A Prototype Model for Data Warehouse Security Based on Metadata”

Kimball.R and Ross.M,” The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)”, New York: John Wiley & Sons, 2000.

Kumar.C.S, Seetha.J, Vinotha.S.R,” Security Implications of Distributed Database Management System Models”, International Journal of Soft Computing And Software Engineering (JSCSE), Vol.2, o.11, 2012, Nov 25, 2012

Priebe.T, Pernul.G, “Towards OLAP security design: Survey and research issues.” In Proceedings of the ACM International Workshop on Data Warehousing and OLAP (pp. 33-40).Washington, DC

The Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. & 1232g; 34 CFR Part 99). Retrieved on 25th July, 2013 from http://www.ed.gov/policy/gen/guid/fpco/ferpa/index.html

Weaver.A.C,” A Security Architecture for Data Privacy and Security”, 2007.

Yanosky.R, Educause Centre for Applied Research, “Key findings: Institutional Data Management in Higher Education”, 2009.