Information and Knowledge Management
ISSN 2224-5758 (Paper) ISSN 2224-896X (Online)
Vol.3, No.2, 2013

A Frame Work for Data Warehouse for Nigerian Police Force: A Way of Ensuring National Security

Georgina N. Obunadike [1*] Richard K. Tyokyaa [2]

1. Department of Mathematical Sciences and IT, Federal University, Dutsin-ma, Katsina State, Nigeria.
2. Department of Mathematical Sciences and IT, Federal University, Dutsin-ma, Katsina State, Nigeria.

* E-mail: nkoliobunadike@yahoo.com

Abstract

Security is presently a major challenge in Nigeria today, Nigerians and non- Nigerians are killed on daily basis and in their numbers. Since the advent of the present democratic dispensation, new forms of violent crimes have become common; these include kidnapping for ransom, pipeline vandalization, Boko Haram bombings, rape, political violence and more. Though the government claims to be on top of the situation, the problem persists. This work attempts to propose a frame work for integration of criminal databases from different state police databases to form a data warehouse for easy access and analysis of criminal data for necessary actions. Data integration involves combining data residing in different sources to form a data warehouse there by providing users with a unified view of these data. Data Integration Technologies have experienced explosive growth in the last few years, and data warehousing has played a major role in the integration process.

Key words: National Security, Data integration, Data Warehousing, Star Schema, Data Mart,

1. Introduction

The Nigerian society is getting more and more insecure, more people are getting into crimes and they are getting more ruthless, desperate and sophisticated. In Nigeria of today especially since the advent of the present democratic dispensation, new forms of violent crimes have become common; these include kidnapping for ransom, pipeline vandalization, Boko Haram bombings, rape, political violence and more. Public and private institutions are attacked and vandalized by gangs, even the United Nations building and the Nigerian Police Headquarters in Abuja were bombed and scores of people killed. Something needs to be done and that urgently. It is widely agreed that security should be the responsibility of all and sundry, not restricted to government, the Police force or security agencies (Otto and Ukpere, 2012).

This work attempts to propose integration of criminal databases from different state police databases to form a data warehouse for easy access and analysis of criminal data for necessary actions. Data integration involves combining data residing in different sources and providing users with a unified view of these data. Data Integration Technologies have experienced explosive growth in the last few years, and data warehousing has played a major role in the integration process. Data warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, and analyst) to make better and faster decisions. The goal of data warehouses is to make the right information available at the right time. It involves bringing together disparate data from throughout an organization for decision-support purposes (Surajit and Dayal, 1997).

The data warehouse concept was developed as IT professionals increasingly realized that the structure of data required for transaction reporting was significantly different from the structure required to analyze data. It was designed to contain summarized, historical views of data in production systems. This collection provides business users and decision-makers with a cross-functional, integrated, subject-oriented view of the enterprise. Data warehouse also known as OLAP (Online Analytical Processing), provides managerial users with meaningful views of past and present enterprise data. User-friendly formats, such as graphs and charts are frequently employed to quickly convey meaningful data relationships (Humphries, 1998).

It is not possible to anticipate the information requirements of decision-makers for the simple reason that their needs depend on the business situation that they face. Decision-makers need to review enterprise data from different dimensions and at different levels of detail to find the source of a business problem before they can attack it. They likewise need information for detecting business opportunities to exploit (Humphries, 1998).

Decision-makers also need to analyze trends in the performance of the enterprise. Rather than waiting for problems to present themselves, decision-makers need to proactively mobilize the resources of the enterprise in anticipation of a business situation. Since these information requirements cannot be anticipated, the decision-maker often resorts to reviewing predesigned inquiries or reports in an attempt to find or derive needed information. Alternatively, the IT professional is pressured to produce an ad hoc report from legacy systems as quickly as possible. If unlucky, the IT professional will find the data needed for the report are scattered throughout different legacy systems. An even unluckier may find that the processing required to produce the report will have a toll on the operations of the enterprise. These delays are not only frustrating both for the decision-maker and the IT professionals; they are dangerous for the enterprise. The information that eventually reaches the decision-maker may be inconsistent, inaccurate, or worse, obsolete. Decision support applications (or OLAP) that obtain data from the data warehouse are recommended for this particular need (Humphries, 1998).

Data warehousing technologies have been successfully deployed in many industries: manufacturing (for order shipment and customer support), retail (for user profiling and inventory management), financial services (for claims analysis, risk analysis, credit card analysis, and fraud detection), transportation (for fleet management), telecommunications (for call analysis and fraud detection), utilities (for power usage analysis), and healthcare (for outcomes analysis). The rest of this paper presents a roadmap of data warehousing technologies, focusing on the special requirements of data warehousing.

2. Data Warehouse Concept

Database computing has shifted its focus from operational to decisional concerns. The differences in operational and decisional information requirements presented new challenges that old computing practices could not meet. This change in computing focus has become the impetus for the development of data warehousing technologies (Maurice, 2002).

A Data Warehouse (DW) is defined as "a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management's decision-making process" (Batini et al, 1992). Data warehouses store huge amount of information from multiple data sources which is used for query and analysis. The data warehouse holds transformed and integrated enterprise-wide operational data appropriate for strategic decision-making, as shown in Figure 1. The data warehouse also contains data obtained from external sources, whenever this data is relevant to decision-making. Decision support applications analyze and make data warehouse information available in formats that are readily understandable by decision-makers.

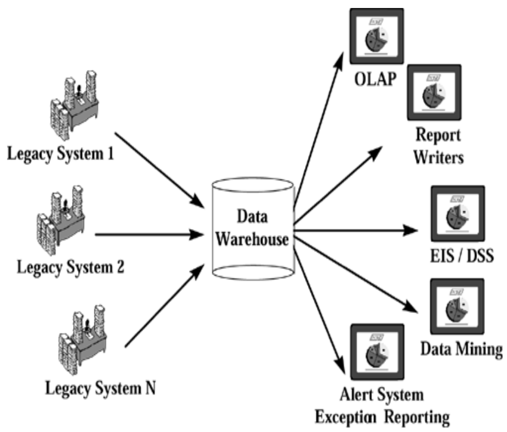

Figure 1: Data Warehouse as an integrated enterprise-wide operational data appropriate for strategic decision-making
Source: (Maurice, 2002)

Since these information requirements cannot be anticipated, operational systems (which correctly focus on recording and completing different types of business transactions) are unable to provide decision-makers with the information they need. As a result, business managers fall back on the time-consuming, and often frustrating, process of going through operational inquiries or reports already supported by operational systems in an attempt to find or derive the information they really need. Alternatively, IT professionals are pressured to produce an ad hoc report from the operational systems as quickly as possible (Maurice, 2002).

It will not be unusual for the IT professional to find that the data needed to produce the report are scattered throughout different operational systems and must first be carefully integrated. Worse, it's likely that the processing required to extract the data from each operational system will demand so much of the system resources that the IT professional must wait until nonoperational hours before running the queries required to produce the report. These delays are not only time-consuming and frustrating both for the IT professionals and the decision-makers they are dangerous for the enterprise. When the report is finally produced, the data may be inconsistent, inaccurate, or obsolete. There is also the very real possibility that this new report will trigger the request for another ad hoc report. Since a data warehouse maintains a copy of information from the source transaction systems. This provides the opportunity to:

- Maintain data history, even if the source transaction systems do not.
- Integrate data from multiple source systems, enabling a central view across the enterprise. This benefit is always valuable, but particularly so when the organization has grown by merger.
- Improve data quality, by providing consistent codes and descriptions, flagging or even fixing bad data.
- Present the organization's information consistently.
- Provide a single common data model for all data of interest regardless of the data's source.
- Restructure the data so that it makes sense to the business users.
- Restructure the data so that it delivers excellent query performance, even for complex analytic queries, without impacting the operational systems.
- Add value to operational business applications, notably customer relationship management (CRM) systems

3. DATA WAREHOUSE PLATFORM

A data warehouse platform consists of one or more hardware servers, an operating system, a database management system (DBMS), and data storage. These communicate via a LAN or WAN, although a multi-node data warehouse platform may have its own specialized network. Note that a data warehouse platform manages a data warehouse, defined as a collection of metadata, data model, and data content, designed for the purposes of reporting, analyzing information, and making decisions. But the data warehouse is not part of the platform per se (Russom, 2009).

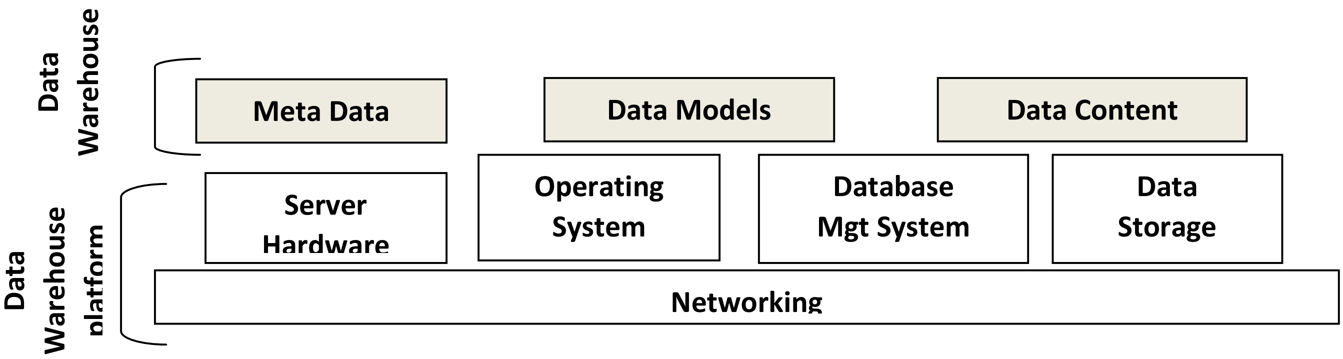

Figure 2: A data warehouse platform manages a data warehouse, but the two are separate
Source: (Russom, 2009).

4. Data Warehousing Architecture

Data warehouse architecture includes tools for extracting data from multiple operational databases and external sources; for cleaning, transforming and integrating this data; for loading data into the data warehouse; and for periodically refreshing the warehouse to reflect updates at the sources and to purge data from the warehouse, perhaps onto slower archival storage. In addition to the main warehouse, there may be several departmental data marts. Data the warehouse and data marts is stored and managed by one or more warehouse servers, which present multidimensional views of data to a variety of front end tools: query tools, report writers, analysis tools, and data mining tools. Finally, there is a repository for storing and managing metadata, and tools for monitoring and administering the warehousing system. Figure 3: shows a typical data warehousing architecture.

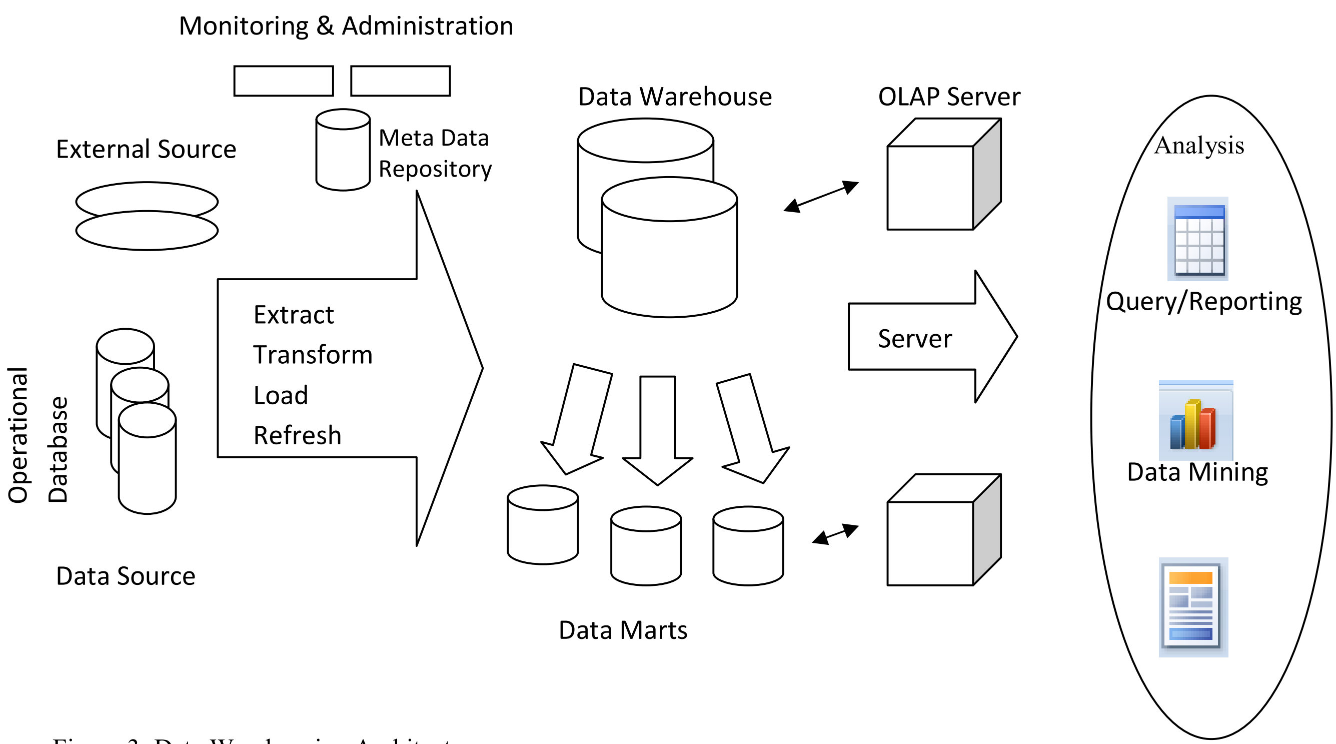

Figure 3: Data Warehousing Architecture
Source: (Surajit and Dayal, 1997).

5. Data Warehousing Methodology

Data warehousing has been cited as the highest-priority post-millennium project of more than half of IT executives. A large number of data warehousing methodologies and tools are available to support the growing market. However, with so many methodologies to choose from, a major concern for many firms is which one to employ in a given data warehousing project (Arun and Atish, 2005). In this article, we propose a top-down design approach of data warehousing.

Online transaction processing (OLTP) systems are useful for addressing the operational data needs of a firm. However, they are not well suited for supporting decision-support queries or business questions that managers typically need to address. Such questions involve analytics including aggregation, drilldown, and slicing/dicing of data, which are best supported by online analytical processing (OLAP) systems. Data warehouses support OLAP applications by storing and maintaining data in multidimensional format. Data in an OLAP warehouse is extracted and loaded from multiple OLTP data sources (including DB2, Oracle, IMS databases, and flat files) using Extract, Transfer, and Load (ETL) tools. The warehouse is located in a presentation server.

It can span enterprise wide data needs or can be a collection of "conforming" data marts (Kimball and Ross, 2002). Data marts (subsets of data warehouses) are conformed by following a standard set of attribute declarations called a data warehouse bus. The data warehouse uses a metadata repository to integrate all of its components. The metadata stores definitions of the source data, data models for target databases, and transformation rules that convert source into target data. The concepts of time variance and non volatility are essential for a data warehouse (Inmon, 2002).

Inmon emphasized the importance of cross-functional slices of data drawn from multiple sources to support a of needs (Inmon, 2002); the foundation of his subject-oriented design was an enterprise data model.

Kimball introduced the notion of dimensional modeling marts (Kimball and Ross, 2002), which addresses the gap between relational databases and multidimensional databases needed for OLAP tasks. These different definitions and concepts gave rise to an array of data warehousing methodologies and technologies. Data warehousing methodologies share a common set of tasks which includes business requirements analysis, data design, architecture design, implementation, and deployment which represents the top-down design approach of data warehousing (Kimbal et al, 1998).

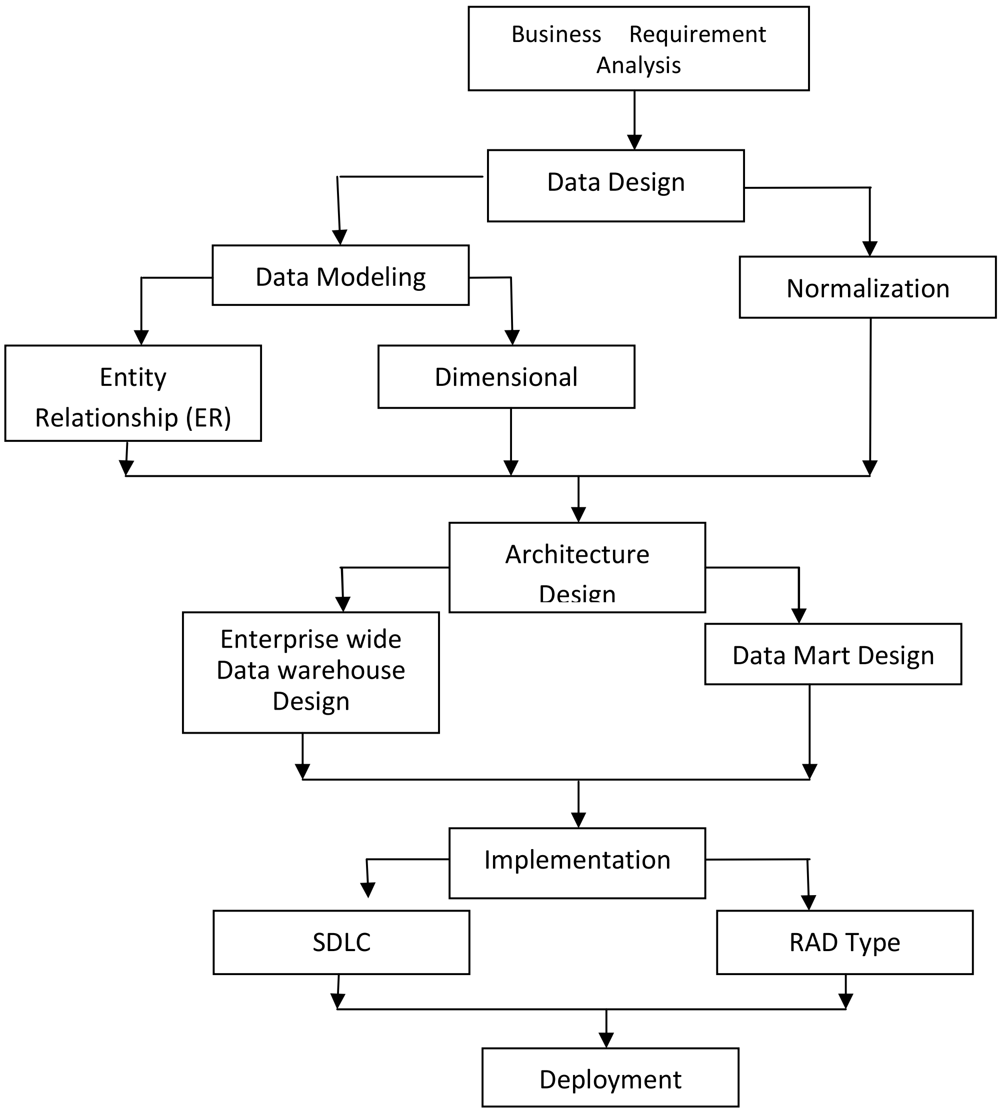

Figure 4: Flowchart for the proposed Top Down Design approach

For business requirements analysis, techniques such as interviews, brainstorming, and JAD sessions are used to elicit requirements. Business requirements analysis is used to elicit the business questions from the intended users of the data warehouse. Business questions are decision support or analytic questions that managers typically pose project (Arun and Atish, 2005). After all the business questions are elicited, they are prioritized by asking the users to rate the questions, or by estimating the risk associated with the solutions needed for the questions. Next, a very high-level conceptual model (also known as the subject-area data model) of the solution for each of the business questions is created. The conceptual model serves as the blueprint for the data requirements of an organization.

The data design task includes data modeling and normalization. The two most popular data modeling techniques for data warehousing are Entity-Relational and Dimensional modeling. The Entity-Relational modeling follows the standard OLTP database design process, starting with a conceptual entity-relationship (ER) design, translating the ER schema into a relational schema, and then normalizing the relational schema.

A dimensional model is composed of a fact table and several dimension tables (Kimbal and Ross, 2002). A fact table is a specialized relation with a multi-attribute key and contains attributes whose values are generally numeric and additive. A dimension table has a single attribute primary key (usually surrogate) that corresponds exactly to one of the attributes of the multi-attribute key of the fact table. The characteristic star-like structure of the physical representation of a dimensional model is called a star join schema, or simply a star schema. A dimensional model can be extended to a snowflake schema, by removing the low cardinality attributes in the dimensions and placing them in separate tables, which are linked back into the dimension table with artificial keys (Kimbal et al, 1998). Dimensional modeling is a data modeling techniques that have gained popularity and acceptance for data warehouse design (Arun and Atish, 2005).

Figure 4 presents a star schema for the proposed data warehouse design. It is made up of a suspect fact table and several dimension tables, the schema analyzes suspect by offenses, dates, place, time, age, courts, prison and informant.

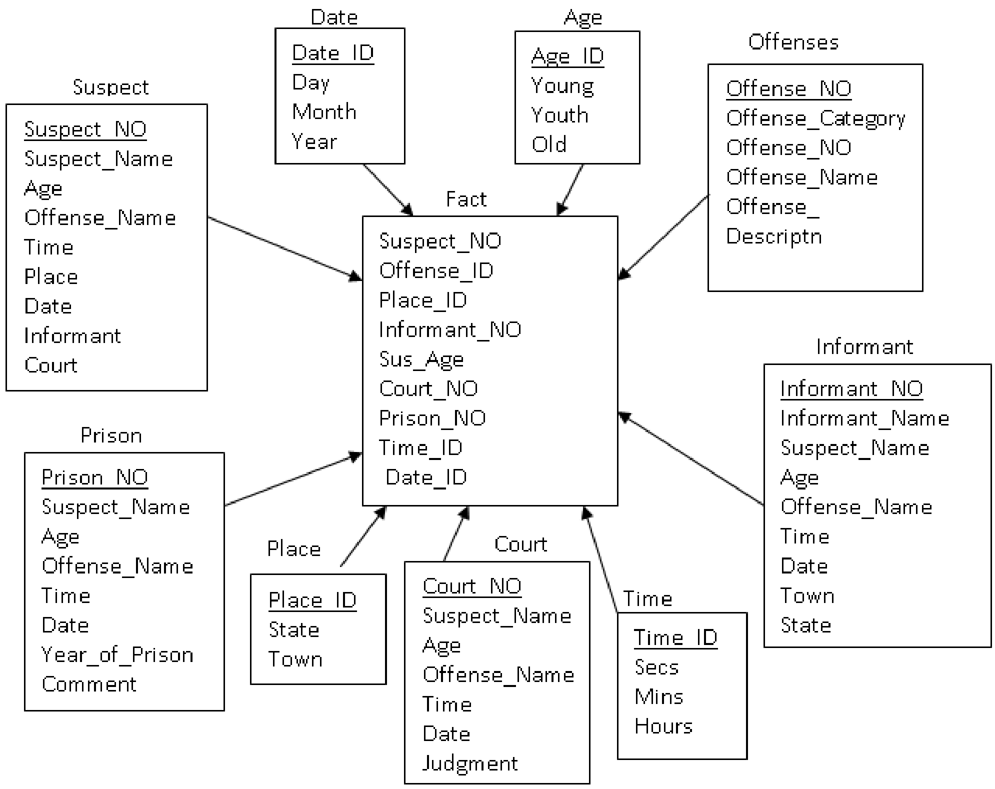

Figure 4: A Star Schema for the Proposed Data Warehouse Design for Nigerian Police Force

Figure 4: A Star Schema for the Proposed Data Warehouse Design for Nigerian Police Force

Architecture is a blueprint that allows communication, planning, maintenance, learning, and reuse. It includes different areas such as data design, technical design, hardware and software infrastructure design. The architecture design philosophy has its origins in the schema design strategy of OLTP databases. Several strategies for schema design exist, such as top-down, bottom-up, inside-out, and mixed (Batini et al, 1992). The data warehouse architecture design philosophies can be broadly classified into enterprise wide data warehouse design and data mart design (Hackey, 1997). The data mart design, espoused by Kimball et al (1998) follows the mixed (top-down as well as bottom-up) strategy of data design. The goal is to create individual data marts in a bottom- up fashion, but in conformance with a skeleton schema known as the "data warehouse bus." The data warehouse for the entire organization is the union of those conformed data marts.

Data warehouse implementation activities include data sourcing, data staging (ETL), and development of decision support oriented end-user applications. These activities depend on two things—data quality management and metadata management (Inmon,2000). As data is gathered from multiple, heterogeneous OLTP sources, data quality management is a very important issue. A data warehouse generates much more metadata than a traditional DBMS. Data warehouse metadata includes definitions of conformed dimensions and conformed facts, data cleansing specifications, DBMS load scripts, data transform runtime logs, and other types of metadata (Kimbal et al, 1998). Because of the size of metadata, every data warehouse should be equipped with some type of metadata management tool.

The deployment task focuses on solution integration, data warehouse tuning, and data warehouse maintenance. Although solution integration and data warehouse tuning are essential, maintenance is cited as one of the leading causes of data warehouse failures. Warehouses fail because they do not meet the needs of the business, or are too difficult or expensive to change with the evolving needs of the business. Due to increased end-user enhancements, repeated schema changes, and other factors, a data warehouse usually goes through several versions.

6. Summary

The interview technique was use to elicit the required information from the Nigerian Police Force that supposed to be keeping the criminal data. It was observed in the course of the interview that as a Nation we don't have a database for criminal activities not to talk of integrating the database. That is why we have proposed the frame work to serve as a guideline for designing a data warehouse for this Nation to aid in monitoring of criminal activities in this Nation and for better analytics and decision making in order to ensure national security. After all the required information was elicited from the Nigeria Police Force it was used to design a star schema for the proposed data warehouse.

7. Conclusion

The successful implementation of data warehousing will creates new possibilities for tackling crime in this Nation. Applications that previously were not feasible due to the lack of integrated data will now be possible. Since the data warehouse will be designed to store historical data, it will be an ideal technology for analyzing criminal trends in this nation. Warehouse users can produce reports that compare current crime rate to historical figures. Their analysis may highlight trends that will reveal some major issues that affect our national security. Such performance trend analysis capabilities are crucial for the success of planning for the so called national security.

References

Arun S. and Atish P. S. (2005) "A Comparism of Data Warehousing Methodologies, ACM,Vol. 48, No. 3, pp. 79-84.

Batini, C., Ceri, S., and Navathe, S.K. (1992) "Conceptual Database Design: An Enity- Relationship Approach". Benjamin/Cummings, Redwood City, CA

Hackney, D. (1997) "Understanding and Implementing Successful Data Marts", Addison-Wesley, Reading, MA.

Humphries, H. (1998) "Data Warehousing Architecture and Implementation", First Edition, Prentice Hall, USA.

Inmon, W.H. (2002) "Building the Data Warehouse", 3rd edition", Wiley, New York.

Inmon, W. (2000) "Metadata in the data warehouse", White Paper, www.inmoncif.com/library/whiteprs/earlywp/ttmeta.pdf.

Kimball, R. and Ross, M.(2002) "The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling", 2nd edition, Wiley, New York.

Kimball, R., Reeves, L., Ross, M., and Thronthwaite, W. (1998) " Data Warehouse Lifecycle Toolkit", Wiley, New York.

Russom P. (2009) " Next Generation Data Warehouse Platforms", TDWI Report, www.tdwi.org

Surajit C. and Dayal U. (1997) " An Overview of Data Warehousing and OLAP Technologies" , ACM Sigmod

Impact of Information Literacy Skills on Academic Staff Research