DESIGN OF AN INTEGRATED AGRARIAN DATA DIMENSIONAL DATA WAREHOUSE

Awogbamila Y. and Adeyemo A. B.
Department of Computer Science
University of Ibadan
Ibadan, Nigeria
yetundekk@gmail.com
sesanadeyemo@gmail.com

ABSTRACT
The concept of the Data warehouse was developed to provide a single access point to data from a variety of sources. There is a need to have a single location for the storage and sharing of data that users can easily utilize to make effective and quality business decisions, rather than trying to traverse the multiple data sources that exist today. Although many frameworks have been developed to integrate these sources into a single database, a reliable framework has yet to be developed. A major hindrance to achieving a reliable warehouse is the poor quality of data obtained from the data transformation stage in the extract, transfer and load process. This poor quality of data contributes to inaccurate and unreliable results and if this data is used for decision making, unforeseen critical business errors can occur. This work reviews the data integration and transformation process in dimensional data warehouses and proposes a dual structure for data integration and metadata of multi-formatted data used for the design of dimensional data warehouse using Agrarian data collected from Ondo State, Nigeria as a case study.

Keywords: Data Warehouse, Data Integration, Metadata, Agrarian data.

1. INTRODUCTION

In the literature about Data Warehouses there are now many definitions of what data warehousing is, however Bill Inmon is cited very often and actually seems to be the father of the term. According to Inmon (1996), a Data Warehouse is a "subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management's decision support process." Also Kimbal (1996), concisely defines a Data warehouse as "a copy of transactional data specifically structure for query and analysis".

This definition by means of requirements specifies that:

i. The data warehouse provides *access to corporate or organizational data*
ii. The data in the data warehouse is *consistent*
iii. The data in the data warehouse can be separated and combined by means of every possible measure in a business (the classic *Slice* and *Dice* requirements)
iv. The data warehouse is not just a data, but also a set of tools to *query, analyze and present* information.
v. The data warehouse is a place where we *publish used data.*
vi. The quality of the data in the data warehouse is a *driver of business reengineering.*

Integration is one of the most important aspects of a Data Warehouse. When data passes from the sources of the application-oriented operational environment to the Data Warehouse, possible inconsistencies and redundancies should be resolved, so that the warehouse is able to provide an integrated and reconciled view of data of the organization.

To retrieve this information, there is need for pointers (metadata) to be integrated with the data, that will make it possible for users to interact with the Data Warehouse so that user can have a means of knowing how the tables are structured, what the precise definitions of the data are, where they originated from or how aged it is. (Diego Calvanese et al, 2001). Data is fed from multiple disparate sources into the Data Warehouse. As the data is fed it is converted, reformatted, summarized, and so forth. The result is that data once it resides in the data warehouse has a single physical corporate image. One simple example of lack of integration is data that is not encoded consistently, as shown by the encoding of gender. In one application, gender is encoded as *m* or *f*. In another, it is encoded as 0 or 1.

As data passes to the Data Warehouse, the applications' different values must be correctly deciphered and recoded with the proper value. This consideration of consistency applies to all application design issues, such as naming conventions, key structure, measurement of attributes, and physical characteristics of data. Some of the same data exists in various places with different names, some data is labeled the same way in different places, some data is all in the same place with the same name but reflects a different measurement, and so on. During the process of integration, many problems arise. (Kalinka Mihaylova Kaloyanova, 2005)

Modern integration applications involve a number of issues, including different kinds of data and data repositories, available resources at the site of the mediator (e.g., storage capacity), requirements on the integrated view (e.g., query response time and up-to-dateness). As a result, no single approach to supporting data integration can be universally applied. In essence, Extraction Transformation and Loading (ETL) applications display a lack of logical data independence: warehouse data can be transformed only during a physical load, not during querying. Among other problems, this prevents users from having different logical views of warehouse data. For example, once worldwide sales data is converted into dollars in the warehouse, French users cannot retrieve their local sales in Euros. The way out therefore is to integrate the multi-formatted data obtained from disparate sources with their respective metadata so as to ease data transformation and information retrieval. Since across multiple applications there is no application consistency in encoding, naming conventions, physical attributes and measurement of attributes, there is no single approach to supporting data integration that can be universally applied. To better understand the impact of those issues on data integration, this paper proposes an integration framework and algorithm that can be used to integrate the multi-formatted data obtained from disparate sources, using Agrarian data collected from Ondo State, Nigeria, as a case study. First a review of existing models of data warehouse architecture and their related ETL processes were carried out then the proposed framework which adopts a modified version of the one proposed by Inmon (2000) is presented. The data integration model was then simulated using Talend; an open source data integration tool.

2. DATA WAREHOUSE MODELS

Most existing data warehouses were designed using either the top down (Kimball, 1996) or bottom up (Inmon, 1992) approach. In the bottom-up approach, data marts are first created to provide reporting and analytical capabilities for specific business processes. When designed this way, the data warehouse is divided into tightly integrated logical components rather than a complex centralised model. In the other approach, called the top down approach, the data warehouse is designed using a normalized enterprise data model. In this approach the data warehouse is considered to be at the centre of a "Corporate Information Factory" (CIF), which provides a logical framework for delivering business intelligence.

Figure 1 shows typical data warehouse architecture (Inmon 2000). It consists of four layers, going from left to right. Starting from the left hand side, we have the data source layer; which consists of operational sources. Operational sources are data sources which are necessary for the day to day operations of a business. They are usually relational databases, optimised for speed and data integrity. However, they may also include Flat files, Web pages, knowledge systems and legacy systems. Next, we have the data staging area, which is located between the operational sources and the warehouse. This temporary storage area has many uses. It can be used to keep data from different sources for processing at different times.

It can also be used to create free space on an operational database by storing its contents. ETL Process natively integrates heterogeneous sources optionally using the data-staging area. Another way, in which the data staging area is used, is for determining current changes in operational sources by comparing the versions of the data on both the operational source and the data staging area. After the data staging area, we have the warehouse itself. The data warehouse is essentially a database designed for the purpose of being analysed. It collects, organizes, and makes data in the form of metadata, raw data and summary data available for the purpose of analysis. The systems used to work with data derived from these systems are referred to as On-Line Analytical Processing systems. These systems represent the users and they are the last component of the data warehousing system

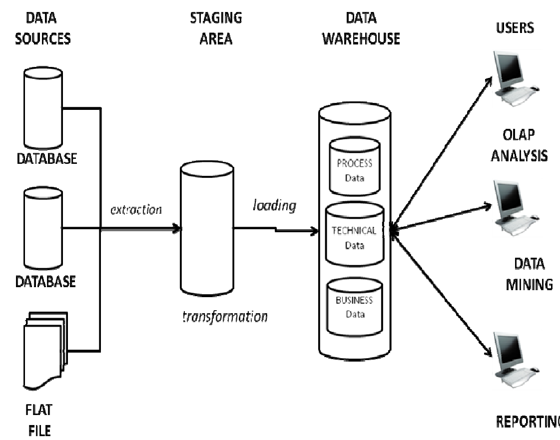

Figure 1: A typical Data Warehouse Architecture. (Inmon, 2000).

Most Data Warehouses operate by using a three step process known as the Extract Transfer and Load (ETL) framework. In the first step of this process, data is extracted from different internal and external sources in both structured and/or unstructured form. Plain queries are sent to the source systems, using native connections; message queuing, database middleware. At the next stage, the transfer stage, which occurs in the staging area, we have all the data in one database. This is achieved by structuring the operational source to conform to the data model used by the data warehouse system. This means that if an underlying data source is a relational database and the data warehouse is also relational then there might not be a need for any translation. But if the source is a flat file, then it must be structured to look like a relational table. Since all the data in the warehouse is in one table, it is possible to join and union tables derived from the data sources. It is also possible to filter and sort data using different attributes while pivoting to another structure and making calculations. However, the most important feature of this step is the ability to check and clean data if necessary. Finally the data from the data staging area is loaded into the warehouse for use by business intelligence.

2.1 ETL Processes Flow

Figure 2 presents the ETL process flow (Shaker H et al, 2011) There are various operational source systems that are developed for data entry such as data stored in Excel spread sheet, Access, Database etc. It is from here that the data warehouse gets populated. The ETL stages are:

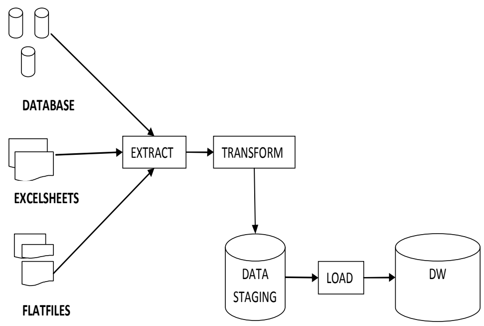

Figure 2: ETL Process Flow general framework (Shaker H et al, 2011)

i. Data Extraction: This is the first step in the process of integrating data into the data warehouse. It involves reaching and understanding the source data and coping the needed data for further manipulation (cleaning and transform).
ii. Transformation: Data transformation is the next step in the ETL processes. This allows cleaning and conforms the incoming data to gain an accurate data which is correct, complete, consistent and unambiguous. It defines the granularity of the fact tables, dimension tables.
iii. Data Staging Area: Is the Physical data store that contains all temporary tables created during the extraction process or resulted from the applied transformation process
iv. Loading: This is the last step of the ETL processes as shown in figure3.2. Once the mapping is complete, the multidimensional can then be loaded with the integrated data. It involves loading the integrated/ transformed data with their corresponding metadata into the target multidimensional data warehouse. This is then accessed by the end users and application system.

2.2 Agricultural Production In Ondo State

In Nigeria, agriculture is made up of forestry, livestock, fishing, food and cash crops such as yams, cassava, cola nut, maize, cocoa, groundnut and oil palm. The country is largely endowed with natural resources that are necessary for the development of agriculture. Such resources include abundant land supply, human and forestry resources. The country has a total land area of about 98.3 million hectares out of which 71.2 million hectares (72.4%) are cultivable but only 34.2 million hectares (34.8%) are under use (Daramola, 2004; Fasoranti, 2007). Agricultural production is still highly dominated by the small holder farming system. The farms are dominated by small scale farmers who are responsible for about 95% of total production (Fasoranti, 2007).

Ondo State is located in the South-Western part of Nigeria, bounded in the east by the Edo/Delta States, in the North by Ekiti State, in the West by Ogun / Oyo State and in the South by the Atlantic Ocean. The State which is fairly largely populated possesses what it takes to be a haven of rewarding commercial activities, and one of the economic backbones of the nation. Ondo State is predominantly an agricultural State with over 60% of its labour force deriving their income from farming. The State is richly blessed with varied and favourable ecological and climatological conditions with vegetation ranging from mangrove swamps of the southern coastal riverine areas through the rainforest of the midlands to the derived savannah in the northern part of the State. Thus the State can support the cultivation of a large variety of crops.

The State also has the longest coastline in the country which favours fishing activities in the riverine areas. Agriculture is the dominant occupation of the people of Ondo State providing income and employment opportunities for over seventy percent of the population. It also contributes well over seventy five per cent to the state's Gross Domestic Product (GDP). The main revenue yielding crops are cocoa, palm produce and timber. Ondo State is one of the most important timber producing states in Nigeria. Some of the hardwood species are Iroko, Mahogany, Obeche and Sapele wood. Since the introduction of cocoa in Nigeria around 1874, it has grown to be the fourth largest exporter in the world with production level reaching 385,000 metric tons per annum (Oluyole, 2005; Nwachukwu. et al., 2010).

In terms of capacity, Ondo state is rated as the largest cocoa producing state in Nigeria (Oluyole, 2005; Nwachukwu et a.l, 2010). Ondo State is one of the six coastal states in Nigeria. This unique situation presents a benefit of direct access to the rich resources of the Atlantic Ocean via the Ilaje/Eseodo, Okitipupa and Irele Local Government areas. In addition, the state is blessed with numerous inland waters that exude tremendous advantages of fish production. These are rivers, streams, reservoirs, dams and inan-rnade production ponds. Notable among these are rivers Ogbese, Owena, Oluwa, Ose, Ero, Oye, Usese, Oni, Oshunetc; the major dams include Ero, EgbeOwena, Ose/Owo, Awara, Itapaii and Owani. While activities along and 'vvithin the coastal fringe of Ilaje/Ese-Odo and Okitipupa seem to be organised and highly competitive, fish farming and aquaculture in the hinter-drylands is yet to show any traditional or scientific improvement (Akegbejo, 2004).

The oil palm (*Elaeisguineensis*) is one of the important economic crops in the tropics. The oil palm is a versatile tree crop with almost all parts of the tree being useful and of economic value. The principal product of oil palm is the palm fruit, which is processed to obtain three commercial products. The State also generates revenue from oil palm which is the second highest earner for the state through the sale of unrefined oil palm fruit and processed oil palm production which include palm oil, palm kernel oil and spices but there is presently no independent supply of oil palm fruits, therefore, oil palm processing companies are involved in oil palm cultivation (Anyanwu; Anyanwu and Anyanwu, 1982; Ibitoye, et a.l, 2011).

Ondo State is abundantly blessed with forest landscape that hosts different types of hardwood. Although a few sawmills are available in the State, they merely carry out minimal processing while the timber logs are carried outside the country for further processing hence; providing an opportunity in processing of timber locally into finished products such as furniture, paper and pulp (Mgbeje, 2004; Ibitoye, et al., 2011).

3. THE DATA INTEGRATION FRAMEWORK.

Figure 3 shows the modified data warehouse framework for integrating multi-formatted data. It differs from existing data warehouse systems in the design of the transformation and loading component of the data warehouse.

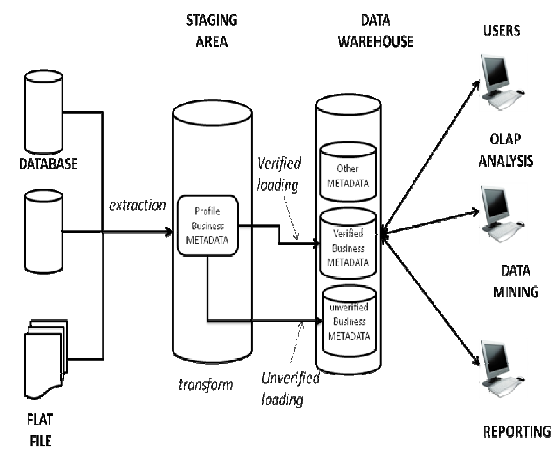

Figure 3: The proposed framework.

The first step in the process of integrating data into the warehouse, using the framework presented in figure 3 involves extracting source data from the operational sources. The source data is composed mainly of flat files. The next step is the transformation stage. At this stage, the extracted source files are cleaned and formatted to a warehouse compliant form. This occurs at the data staging area. However, no perfect data cleaning algorithm exists and so errors are inherent in the warehouse. The proposed framework handles this by profiling the data for correctness. If the transformed source file passes the test then it has metadata assigned to it signalling it as verified. If it fails, it has metadata created classifying it as unverified or unreliable. Table 1 shows the algorithm that implements the framework. The benefit of this framework is that users have an idea of the correctness of the data they are being presented with. The motivating idea is that end users, in this case, the decision makers, have a sense of certainty about the level of accuracy of the information presented to them. It is however important to point out that this framework does not preclude the need for more efficient data cleaning algorithms.

The algorithm begins by extracting source files from the data sources. Next it converts this file into a warehouse compliant form and then it cleans it .Subsequently, it extracts business metadata from the cleaned data .This cleaned data is then tested for correctness, if it passes, the data is then marked as correct and it is then loaded as verified data. On the other hand, if the data is incorrect, then it is marked as unverified and it is loaded as unverified data.

Table 1: The metadata loading and translation algorithm

FOR (I TO K) // where *i represents the i^{th} resource and k represents the $k^{th\ information}$ source.*

EXTRACT (I) // extract the source files from the operational source.

TRANSFORM (I) =>DATAWAREHOUSEFORM (I) // *transform the operational source to a data warehouse //compliant form.*

CLEAN (DATAWAREHOUSEFORM (I)) // *cleans the data*

EXTRACT_METADATA (DATAWAREHOUSEFORM (I))

FOR (METADATA (I) TO METADATA (K))

IF METADATA (I) =AGRARIAN_METADATA (I)

IF TEST (AGRARIAN_METADATA (I)) >THRESHOLD // *Test to see if its contents are reliable.*

THEN LOAD (AGRARIAN_METADATA (I)), METADATA_STORE (A))*//if it is reliable load its contents and tag it as unverified metadata A.*

ELSE LOAD (AGRARIAN_METADATA (I)), METADATA_STORE (B))*//if it is unreliable ,load its contents and tag it as unverified metadata B.*

To build the integrated data warehouse, the methodology proposed is similar to the one proposed by Inmon (1992). This methodology divides the implementation of the data warehouse into three stages namely: the prerequisite, the first and second iteration stages:

a. The Prerequisite Stage:
   i. Build the data model
   ii. Determine the technology required to build the database.
   iii. Decide on the size of the data warehouse.
   iv. Determine what type of information is required.

b. The First Iteration Stage:
   i. Establish the amount of data to be loaded.
   ii. Identify which of the components of the business will be implemented first.
   iii. Determine which subject area to implement first.
   iv. Create a physical database design for the warehouse

c. The Second Iteration Stage:
   i. Create a data profiling component for the data staging area.
   ii. Construct dual structures for the metadata, raw and summary data warehouse components.

In order to building an integrated agrarian data warehouse, a corporate data model is first developed. The corporate data model was built using agricultural production and sales data collected from Ondo State Government based agencies; Federal Government based agencies that have agricultural data on Ondo State, Non Governmental Organizations and on-line resources that have agricultural data on Ondo State of Nigeria. Figure 4, shows a high level view of the data model. The high level data model contains the major subject areas and the relationship existing between them.

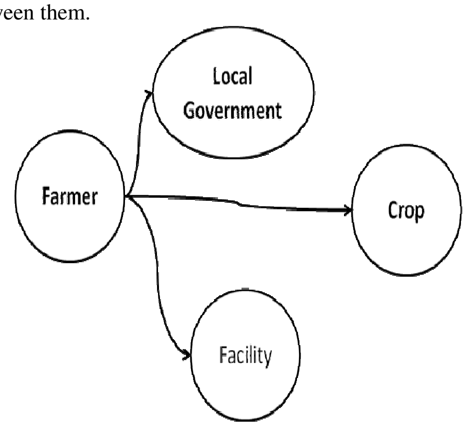

Figure 4: The high level data model of the framework.

Figure 4 also presents the four entities namely: Farmer, Crop, Local Government and Facility. There are direct relationships between the major entity farmer and each of the other three entities, local government, crop and facility. In larger warehouse designs, this high level view provides an overview of how the entities are connected and their relationships. Figure 5 shows the entity relationship mapping of the data model.

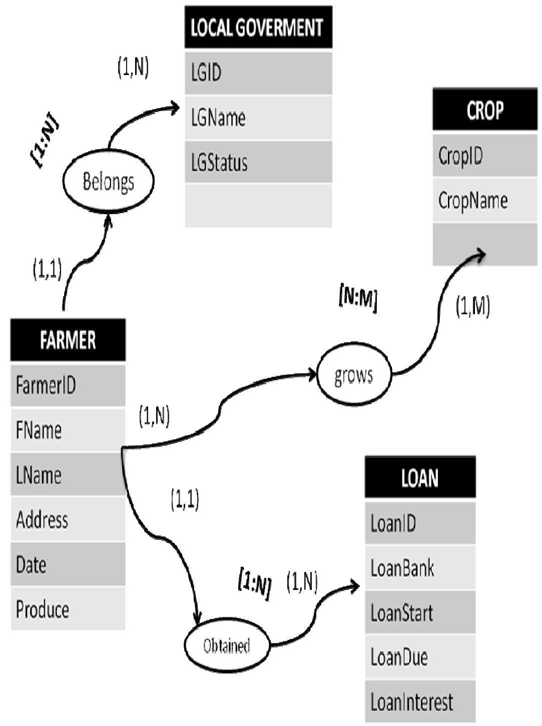

Figure 5: The Entity Relation Mapping (ERM) of the integrated data warehouse

Figure 5 shows the entity relation mapping of the integrated agricultural data warehouse. It is an abstract way to describe the warehouse database. The ER diagram has four tables; Farmer, Crop, Loan and Local Government. There are three connections namely: "grows", "obtained" and "belongs" with their corresponding relational cardinalities.

Determining the size of the data warehouse is more of an art than a science (Inmon, 1992; 2000). If the size of the data warehouse is too small then the performance of the system will be poor. On the other hand if the size of the system is huge, then the costs of maintaining and running the data warehouse will be exorbitant. The information requirements of the data warehouse can be determined by considering the different sources such as spread sheets and reports. Spread sheets are easy and inexpensive sources of information; however spreadsheets, being fluid can change with no documentation unlike a relational database that logs changes. The agrarian data obtained were mainly sourced from spread sheet files (Table 2).

Table 2: A sample spread sheet, showing cocoa graded figures for 2010

S/N	AREA OFFICES	JAN.	FEB.	MARCH	APRIL	MAY	JUNE
1	AKURE N.	94	54	-	-	-	-
2	AKURE S.	614.5	350	470	551.1	109	46
3	ALADE	922	4565	105	106	110	111
4	IDANRE	3453	1959	416	394	581	471
5	ONDO WEST	1159	672	175	211.5	399.5	5605
6	ONDO EAST		371	77	98	82	101
7	ORE	618	408	140	298	422	451
8	ONIPARAGA	107	52	12	20	25	33
9	OWO	1205	563	724	403	323.5	115
10	OSE	247	102	121	65	34	21
11	AKOKO N.E	52.5	23	7	15.5	4	-
12	AKOKO N.W	-	-	-	-	-	-
13	AKOKO S.E	75	25	30	8	19	-
14	AKOKO S.W	88	130	29	15	-	2
15	ILARA	442	145	138	50	21	29
16	ILE-OLUJI	421.5	218	10	24	78	89.5
17	OKITIPUPA	-	-	-	-	-	-
18	IRELE	46	25	-	13	35	58
		10,059.50	5,553.50	2,454	2272.1	2243	2088

The following options apply to the building of the first iteration of a data warehouse: whether to implement a single subject area, implement a subset of a single subject area, or implement a subset of multiple subject areas. Also the quantity and variety of data to be loaded is considered. Since the recommended first user of the first iteration will be any decision maker overseeing the agricultural ministry, the functional components proposed for the implementation were the classical functional areas that data warehousing has proven to be effective such as finance, accounting, sales, production and marketing data.

Production data was used and the subject area “Farmer” was selected as the starting point. For the physical design of the database, Key attributes were identified and specified based on the end users understanding of the data, the format of the data as it resides in the operational sources, the different data formats that were to be consolidated, and the volume of data to be created and a unit of time was decided on. The design proposes at least two data sources: the operational data and the changed data sources. The system of record designed ensured that it would be able to accurately capture data from both sources (figure 6).

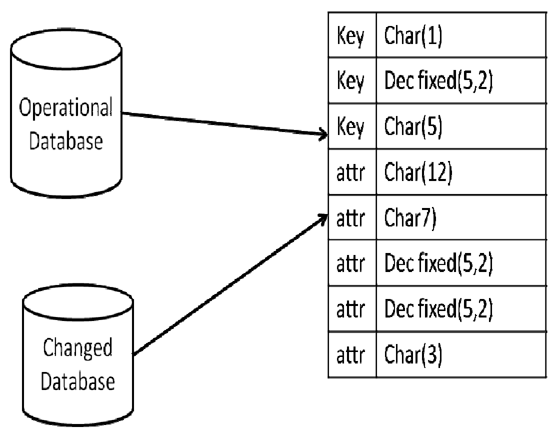

Figure 6: The two sources of data for the system of record (Inmon, 2000).

The changed database represents data that has been gleaned from the updates that have occurred to an operational database since the last time a refreshment of the data warehouse was accomplished. It is used where there are very large files that undergo very little change from one refreshment of the data warehouse to the next, for example, in the case of agricultural implements. Transformation rules were applied to the data from the operational sources. The metadata extracted from this transformation was divided into verified and unverified metadata (figure 7).

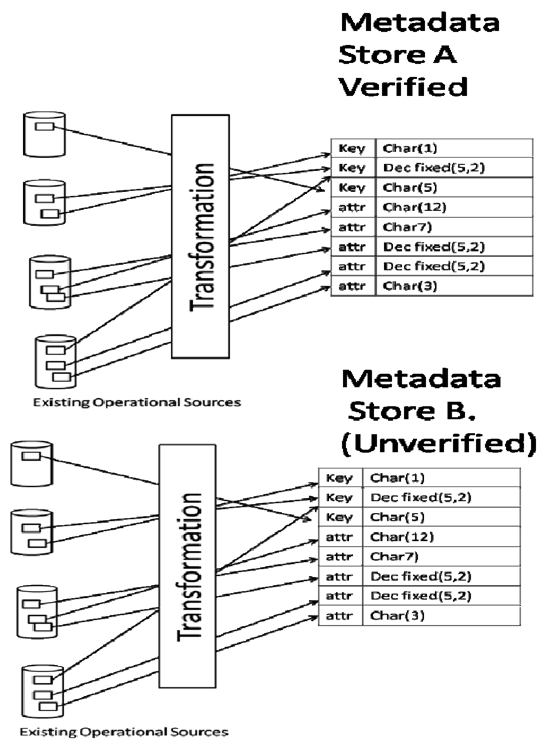

Figure 7: The verified and unverified metadata stores.

Once the transformations information had been specified, it was transferred to the dual metadata infrastructure that sits above the data warehouse. The metadata contains a record of what the source data is, what transformations and summarizations had occurred, and what the description of the data warehouse looks like. The unverified metadata contains data that is yet to be properly cleaned. This process is then repeated for the raw data and the summary data.

4. MODEL SIMULATION

The data integration was simulated using an open source data integration tool: Talend open studio for data integration. Figure 8 shows the corporate data model used for the simulation.

The agrarian data warehouse data model presented in figure 8 is represented in terms of building blocks like the operational sources, data staging area and data warehouse and their connections. It shows the data flow from extracted sources, like the “Agric CSV file” , “Agric Excel File” , “Farmer” and “ State” files to the data staging area where they are cleaned and transformed. The transformed data can then be loaded into the agrarian data warehouse along with its corresponding metadata.

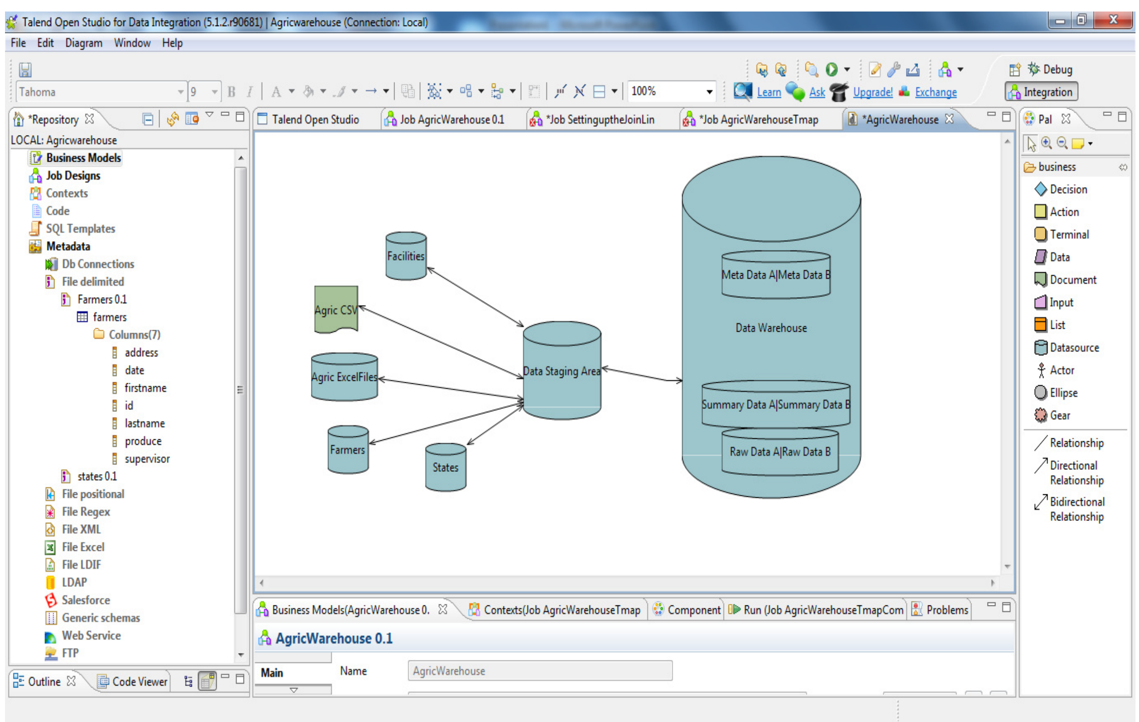

Figure 8: The data model of the agrarian data warehouse

4.1 Developing the metadata model

The design stores the metadata which is derived and transferred from transactions in the data warehouse. If this metadata is inaccurate, wrong or nonexistent then it would be impossible to make changes in the data warehouse that will match recent changes in transaction operations. Without metadata it would be difficult to compare data coming from a variety of external sources and processes, because; there wouldn't be any means of identifying these sources and processes in the context of the warehouse. If a data warehouse doesn't have metadata, it would have to employ adhoc systems which are very expensive to use, maintain and integrate. Furthermore business intelligence tools depend on metadata to integrate data from different sources. Without this metadata information, these tools will store different redundant versions of the same data repeatedly in the warehouse. Metadata from different external and internal sources are usually shared by different transformation algorithms in the data storage phase of the warehouse. This enables these algorithms to combine data from different sources efficiently and accurately.

The first step in creating the metadata model was to create a new job to build the model for a delimited CSV file representing farmers in local governments of Ondo state. Table 2 presents a sample of the farmer metadata data

Table 2: Sample of the farmer metadata data

ID	Firstname	Lastname	Date	Vol Of Cocoa Produced(Mt)	LG
1	Kunle	Owumi	20/01/2005	55239	Owo
2	Tolu	Babarinde	19/05/2006	78148	Akure south
3	Dipo	Owumi	28/09/2000	77912	Akure north
4	Agbeja	Fawole	27/11/2001	76036	Idanre
5	Adewale	Oshiomole	30/09/2004	56895	Idanre
6	Tolu	Oniped e	04/09/2001	26144	Owo

The farmer's metadata schema created is shown in figure 9.

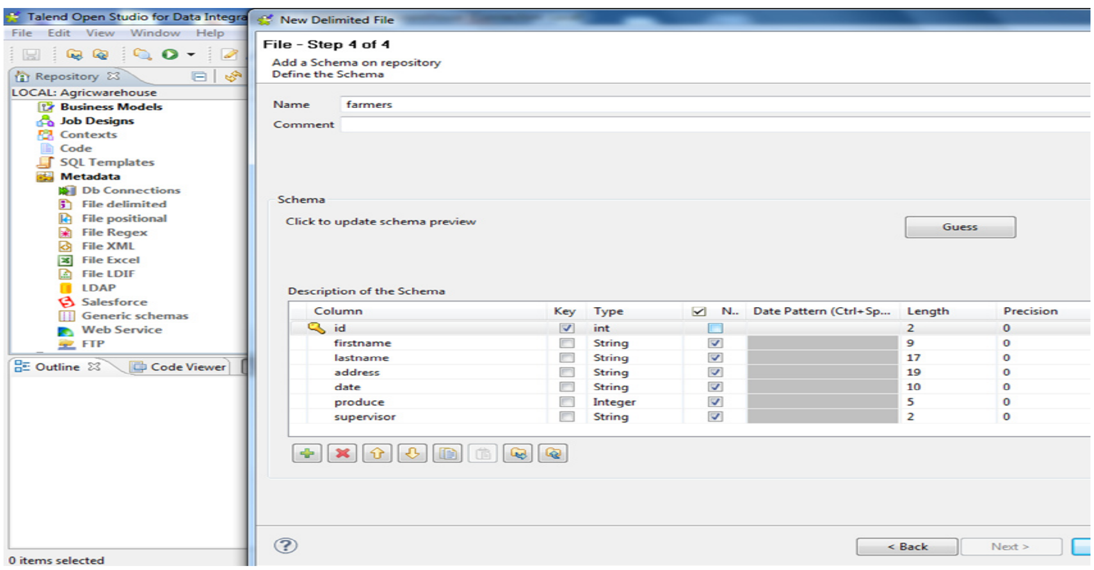

Figure 9: The farmer's metadata schema.

The farmer's metadata schema consists of the following attributes: Firstname, Lastname, Address, Date and Volume of Produce (Metric tonnes). This procedure was repeated for all the other metadata schemas that were created.

4.2 Data transformation: Operational source to Warehouse database.

The flat file datasheets were then converted into a relational database table. A connection was created between the farmer's metadata file and the output file on the job screen. This connection serves as the link between the farmer's CSV (flat) file and the Output file.

This connection creates the means of communication. Any processing that generates errors or malfunctions is automatically sent to the log file. Also any other component connected to this output file and has the required permission can receive or access the information collected in the output file. When the transformation script is executed the data from the farmers metadata file is transformed into a warehouse compliant form as shown in figure 10.

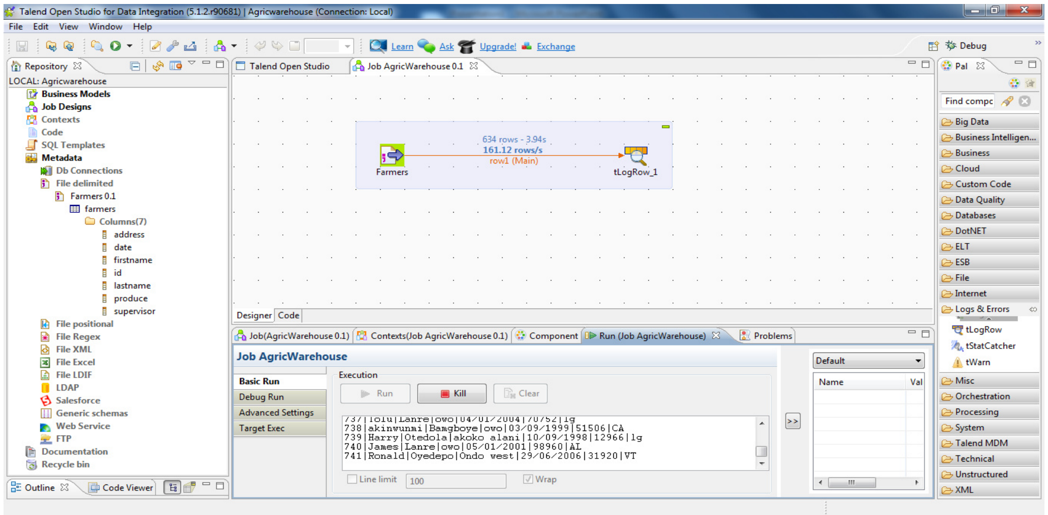

Figure 10: Running the connection between the farmer's metadata file and the output file

Data transformation: Two operational sources merged in the warehouse database.
Two operational sources can also be joined through a mapping function to create a single data warehouse file. Considering the two input files representing the farmers file and the states file shown schematically in figure 11. The state and farmer schemas specify, based on the files we have, the facts that can enter the database, or those of interest to the possible end-users. A transformation function is used to transfer the input files into a form compatible with the data warehouse and a connection is the formed between them to enable the mapping of the Farmer and State files to the new farmer local government file in the data warehouse as shown in figure 12.

Figure 11: Schematic diagram of two operational sources transformation process

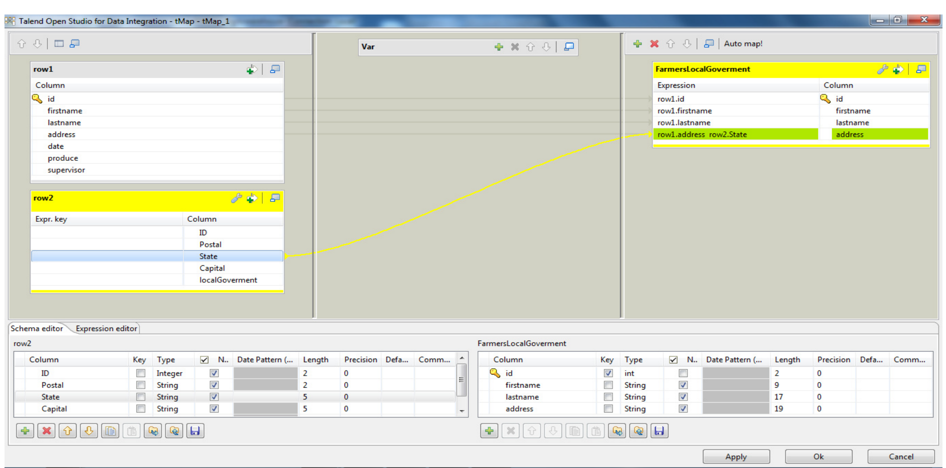

Figure 12: The mapping of the Farmer and State files to the new farmer local government file in the data warehouse.

5. CONCLUSION

In this study, a modified version of the framework proposed by Inmon (2000) was used for the design of an integrated data, dimensional data warehouse for agricultural data. The design intends to reduce the effect of inefficient data cleaning algorithms by layering the internal architecture of the data warehouse. The data integration model was simulated using an open source data integration tool: Talend open studio for data integration. The simulation showed the design of the metadata repositories for the operational sources used by the ETL tool; the process of transforming the input operational sources into the integrated data format and finally, the process of combining the operational source files. The work continues with the development of a prototype customized data integration tool for the data warehouse implementation.

REFERENCES

1. Anyawu, A.C; Anyawu, B.O and Anyawu,V.A (1982). *A Textbook of Agriculture for School Certificate*. 4th Edition, Nsukka, Nigeria: Africana Educational Publishers Ltd.
2. Daramola, A.G., (2004) “Competitiveness of Nigeria Agriculture in a Global Economy: Any Dividends of Democracy?” Inaugural Lecture Series 36 delivered at the Federal University of Technology, Akure, 2nd March, pp: 1.
3. Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Daniele Nardi, Riccardo Rosseti, (2001) Data Integration In Data Warehousing International Journal of Cooperative Information Systems Vol. 10, No. 3 237-271 ©World Scientific Publishing Co.
4. Fasoranti Olayiwola Olujenyo (2007) The Determinants of Agricultural Production and Profitability in Akoko Land, Ondo-State, Nigeria Department of Sociology, Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria.
5. Ibitoye, O. O., Akinsorotan A.O., Meludu N.T. And B.O Ibitoye (2011) Factors Affecting Oil Palm Production in Ondo State Of Nigeria Department of Agricultural Extension and Rural Development, University of Ibadan, Nigeria. Journal of Agriculture and Social Research (JASR) Vol. 11
6. Inmon W.H (1992) Building the Data Warehouse QED Technical Pub. Group, Jan 1 Copyright by William H. Inmon, . New York: Wiley.
7. Inmon W.H. (1996) Building the Data Warehouse, Second Edition. New York: John Wiley & Sons.
8. Inmon W. H. (2000), Building The Data Warehouse: Getting Started © Copyright by William H. Inmon, all rights reserved.
9. Inmon W. H. (2000), “Metadata in the Data Warehouse” © Copyright by H. Inmon.
10. Inmon, W.H. (2005). Building the Data Warehouse.Forth Edition.Wiley. Indianapolis. © Copyright by H. Inmon, all rights reserved.
11. Kimball, Ralph (1996). The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses.John Wiley & Sons.
12. Kimball, R. and Ross, M. (2002).The Data Warehouse Toolkit. Second Edition. Wiley Computer Publishing. New York.
13. Kalinka Mihaylova Kaloyanova, (2005) Improving Data Integration for Datawarehouse: A Data Mining Approach “St. Kliment Ohridski” University of Sofia, Faculty of Mathematics and Informatics Sofia 1164, Bulgaria
14. Mgbeje, (2004). Raw Materials Research and Development Council (RMRDC), Abuja. Report on Survey for Selected Agricultural Raw Materials in Nigeria. http://www.rmrdc.gov.ng Retrieved on 22/12/06.
15. Nwachukwu, Ifeanyi Ndubuto, Nnanna Agwu, Jude Nwaru and George Imonikhe (2010). Competitiveness And Determinants Of Cocoa Export From Nigeria Report and Opinion ;2(7):51-]. (ISSN: 1553-9873).
16. Oluyole, K.A. (2005).Evaluation of the Economics of Post-Harvest Processing of Cocoa in Cross River State, Nigeria. Journal of Agriculture, Forestry and the Social Sciences. 3 (2): 58-64.
17. Shaker H. Ali El-Sappagh, Abdeltawab M. Ahmed Hendawi , Ali Hamed El Bastawissy (2011). A proposed model for Data warehouse ETL processes. Journal of King Saud University – Computer and Information Sciences 23, 91–104. © 2011 King Saud University. Production and hosting by Elsevier B.V. All rights reserved
18. Yemi Akegbejo (2004) Samson Practical suggestions for the commercialisation of fishery in Ondo state. Federal College of Education, Abeokuta Nigeria