



Integration of Big Data in Banking Sector to Speed up the Analytical Process

¹Prof. Dr. P.K. Srimani, F.N.A.Sc. ²Prof. Rajasekharaiah K.M.

¹Former Chairman, Dept. of Computer Science & Maths, Bangalore University
Director, R & D, Bangalore University, Bangalore, India.

² Professor & HOD, Department of Computer Science and Engineering,
JnanaVikas Institute of Technology, Bangalore Mysore High Way, Bidadi, Bangalore,
Visvesvaraya Technological University (VTU), Belgaum, Karnataka, India.

Abstract: In banking area, we find Big Data which is scattered in different places or sources in heterogeneous format using different Databases or Files. Hence, it is very difficult to analyze the data fastly for making Decision Support System (DSS). In this paper, we have developed a High Level Design (HLD) of Data Warehouse system and making the whole process or the system automated using ETL (Extraction, Transformation and Loading) tools like IBM InfoSphere Information Server, PowerCenter Informatica etc/, In the first phase, Hadoop Data Warehouse is designed by integrating Big Data from various sources like Oracle DB's, DB2, Sybase, SAP, Data Marts, Flat Files, on WEB SPHERE etc. into a Warehouse in a single format and in one place. Hence, we use ETL tool – Informatica to integrate all banking data and also use “ERWIN” for warehouse design and “SQL LOADER” for fast data transfer. It can be operated on Windows and/or Unix O/s platform. In order to integrate all this data, initially we design a Multi-dimensional Modeling of Data (MDMD) by using Star Schema and Snow Flake Schema. Secondly, we pool all the data in one area called “Staging Area”, from this we make ETL process of all data into Data Warehouse.

Keywords: Hadoop Data Warehouse, heterogeneous data, Database files, Flat files, HLD, automated, ETL, Informatica, Web Sphere, Staging Area

I. INTRODUCTION:

In this paper, a detailed study of the banking system which uses OLTP (On Line Transaction Processing) for handling the day-to-day transactions and to generate the business analysis reports is made. The existing system provides limited options for analyst to generate reports for future business forecasting and also to develop business strategies. Further, these reports do not support system applications and thus cannot meet the requirements of the Bank to enhance their business objectives.

Currently the Big Data in the business is competitive in all directions vertically, horizontally and parallelly. The success of the banking sector or organizations depends on the effectiveness of the use of technology, tools and services in meeting the customer's requirements and their satisfaction.

Certain developmental activities in this direction move through a set of planned strategies consisting of establishment of clear objectives and goals, from the generation of ideas to concept development, service design, prototyping, service launch and customer feedback. As mentioned here some expert of literature exists in this direction but have served major drawbacks. Hence, the present study is carried out [1, 3, 10, 11].

II. OBJECTIVES AND GOALS:

A. Objectives:

Our research will dwell in the following area:

- Data Mining both from structured and unstructured data
- Mapping from heterogeneous sources of data through Staging Area into DWH
- Big Data integration and analytics to speed up the process for querying or report generation

B. Goals:

Our research goal is to create DWH using ETL tool – Informatica. This tool is used for analyze DW and provides us various reports of the Bank [2].

The results/solutions are compared with other business analytical tools and prove that the advantages in our solutions are the best to practice and to implement in all business enterprises.

III. PROBLEM DOMAIN:

Presently, the Big Data is scattered in various sources and also in different formats. We are facing the following problems –

- It is very difficult to analyze those data fastly
- Limited options for analysis
- Limited options for analyst to generate reports
- Reports taken are not sufficient or sometimes short falls for DSS like business forecasting and to develop business strategies
- Reports even do not support some system applications and can't meet the requirements of the Bank to enhance their business

IV. DESIGN PROCESS & DEPLOYMENT:

In Fig. 1, the technical diagram of a complex Data Warehouse Architecture (DWA) is presented, Implementations are done by using the following Hardware and Software's: [5, 6, 8, 9]

A Bank needs the development and design of an analytical **DWH** which is inextricably linked to various business needs. The various **design process** which involves are – [6, 9]

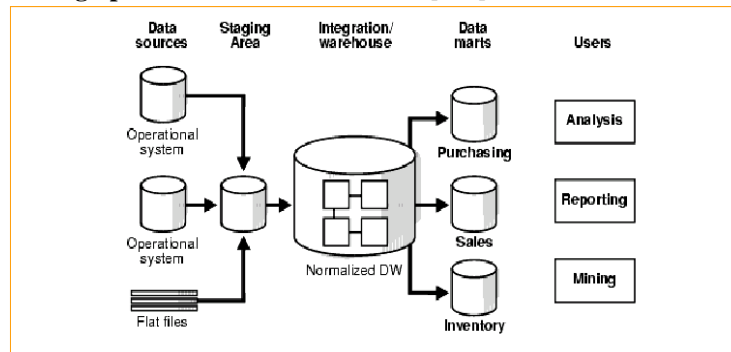


Figure 1: Technical Diagram of Complex Data Warehouse Architecture

1. OLTP – Transaction Processing

The input to the DWH (Data Warehouse) is from various sources like –

- Oracle DWH tables, dup files, data files etc
- Flat Files or Text Files, Excel Sheet etc.

2. CRS and SRS

(Customer and System Requirement Specifications)

3. ETL (Extraction Transformation & Loading) specifications

Involves Source Data to the Target Data

4. HLD – High Level Documents

Description of the tools used and naming conventions

5. DWH – Data Warehouse Design

It involves three phases of design –

- **Conceptual Design** - (Dimensions and Fact Tables)
- **Logical Design** - (Using Dimension Modeling Technique, Attributes and Constraints)
- **Physical Design** - (Data type, Data size, Data Tables and SQL statements)

6. Loading into DWH

(Loading all data from different sources into one storage area i.e. Staging Area into DWH and in one format to make query/retrieval of various reports easily)

7. Testing

(Next step is to test the loaded data by using Unit and System Testing)

Unit Testing is done by developer by writing SQL procedure or query.

System Testing is done by using Software Testing Tools.

8. Certification

(We have to complete **ETL** specifications with **mappings** done by developers. If our design meets the ETL specifications then it is implemented.)

9. Production Phase

(This is the final phase where in further enhancements are carried out depending upon the customer's need or requirements, after it is successful, full implementation will be done.) (See Fig. 2)

V. CASE STUDY of AFFIN BANK, MALAYSIA:

In our research, we implement Data Warehouse Architecture (DWA – Fig.2) which deals with heterogeneous data sets. In the first phase, we have created and designed the Data Warehouse, Dimensions and Fact tables. In the second phase, we are going to mapping with source and target data marts. The bank has a need for an

analytical data warehouse and a leading bank since from 30 years. Since it is fiancé cum facilitator banking company,

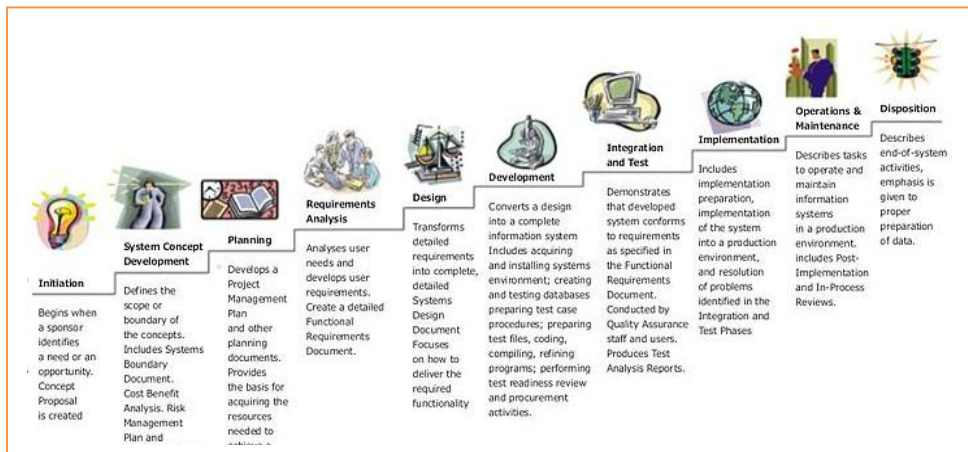


Figure 2: Systems Development Life Cycle Phases (SDLC)

it has to be linked with many kinds of business enhancements and competitive edge over business using Information Technology to make –

- Better decisions
- Dedicated better customer services
- Business Intelligence Analysis

Further, Bank offers the following additional services to their customers –

1. Offshore Finance
2. Commercial Fiancé
3. Trade Finance
4. Vehicle Fiancé
5. Housing Finance

The Bank have number of branches all over South Africa and searching for new business avenues, attracting more new investments and to increase number of customers by using various medias, promoting new finance schemes, implementing new business strategies and decisions. [7]

A. SCOPE:

The below Fig.4 describes the HLD – High Level Documents requirements of the Data Warehouse System. It is meant for use by the designers and developers and will be the basis for validating the final deliverables of the system.

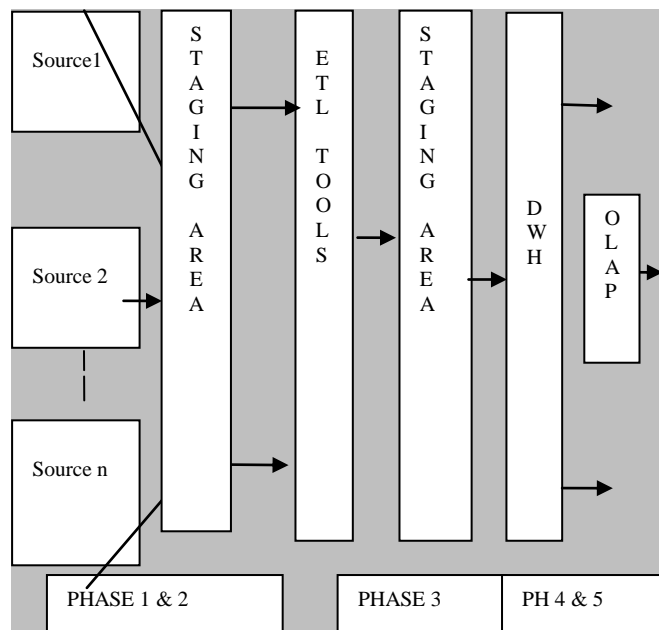


Figure 3: 5 Phases of Data Warehouse Architecture:

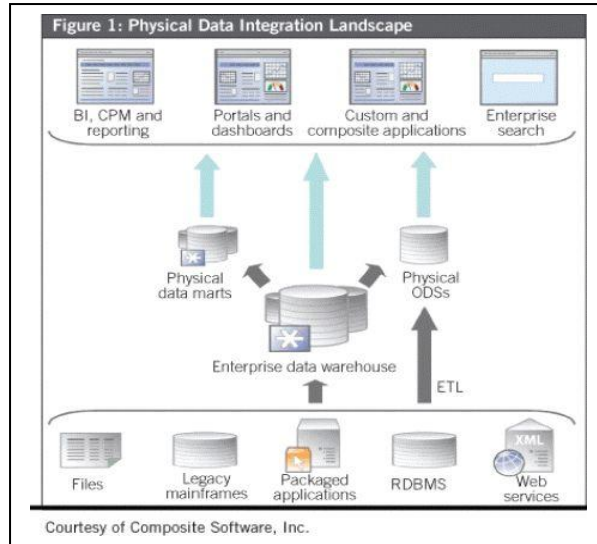


Figure 4: Physical Data Integration
VI. PROPOSED SOLUTIONS:

By considering all the above problems and reports, we are finding solutions as – [2, 5, 8]

- All these data is to be integrated in a single format and pooling in one place, (see Fig. 2 DWH - implementation) so that the mining will be effective and efficient report/information for making proper business analysis and decision making
- Using latest ETL technology tools as mentioned earlier for fast processing of data
- Mapping is done by using source and target data
- Whole process is made Automated by using the above ETL tool – Informatica Power Center Ver.9.0.
- Creating High Level Design (HLD) of DW System and making the whole process Automated
- Creating of Dimensions (MDDM)and Fact tables
- Using ‘ERWIN’ for DW design
- Using ‘SQL-LOADER’ for fast data transfer
- We design Multi Dimensional Modeling of Data (MDMD) in order to integrate all the data by using Star and Snowflake Schema
- After all the above operations is over we will pool all the data in an intermediate area called ‘STAGING AREA’ (Ref. Fig.3)
- Finally, from Staging Area, we are going to pool all data into DW by using ETL (See Figs. (3) to (6)).

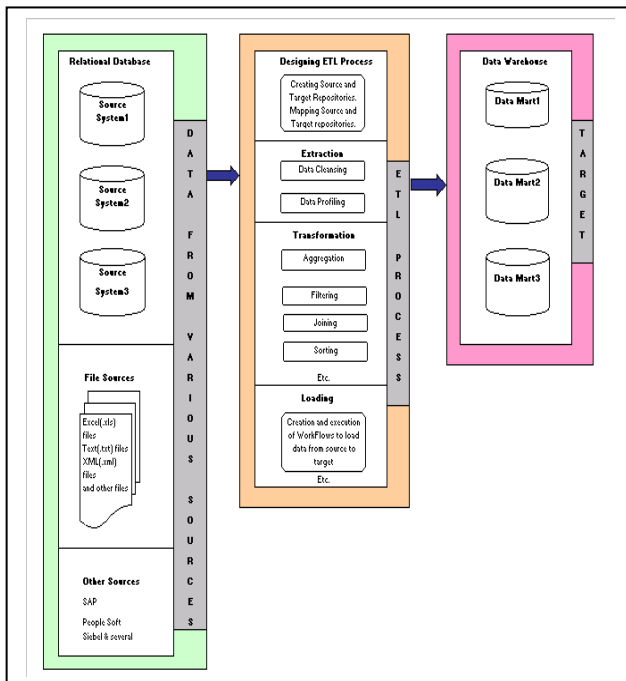


Figure5: ETL Process

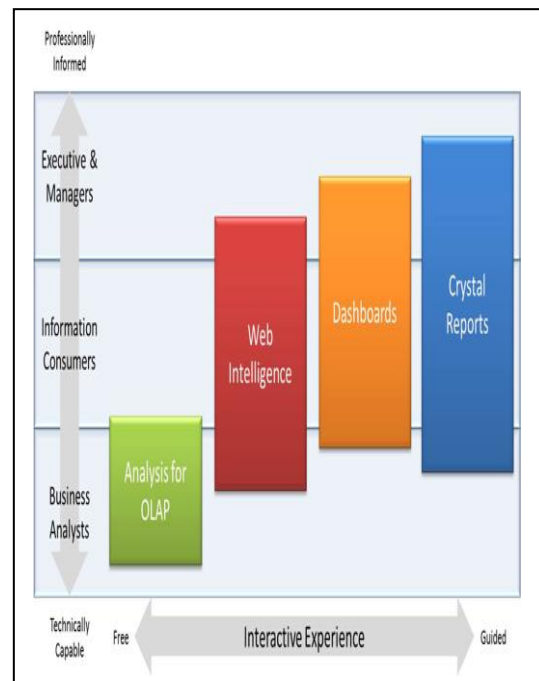


Figure 6: Stakeholders who uses reports

VII. ADVANTAGES

The following are the advantages of the system over the existing system of the Bank:-

- User friendly, easy to handle and flexible in all reports
- Cost is reduced by using this tool and also saves lot of time
- Uncovering those details which are lacking right before
- Validations are made throughout the entire process to avoid occurrence of errors
- Error handling and exceptions are made easy by redirecting to a particular box by naming its path.
- Error handling with error descriptions are also populated in the system
-

VIII. APPLICATIONS

The applications are widely used in banking sector and the following are various **stakeholders** –

- Business Analysts and Executives
- Senior Managers
- Top level and Middle level management people to take DSS in their business
- **Management Information System** tool –
- ✓ To make forecasting of business
- ✓ To analyze trend identification
- ✓ To make market analysis
- ✓ To make competitive business edge in the market
- ✓ To create global market
- Also supports **OLAP** applications and to generate various reports
-

IX. CONCLUSION

In this paper, we **concluded** that the Data Integration of Banking Finance System is successfully designed, developed, tested and implemented with case study. Care is taken for data validation check at each level of data flow. Further, the Software is friendly, menu driven, easy accessible and maintainable.

X. FUTURE ENHANCEMENT

Future enhancements can be done to control data redundancy, data independence, data accuracy and integrity and also recovery from failure.

REFERENCES

- [1]. Inmon W.H. "Building the Data Warehouse", Second Edition, J Wiley and Sons, New York, 1996
- [2]. B de ville (2001), "Microsoft Data Mining :Integrated Business Intelligence for E-Commerce and knowledge Management". Boston: Digital press.
- [3]. Frawley W Piatetsky –Shapiro G and Matheus C, "Knowledge Discovery in Databases" An overview". AI Magazine, Fall 1992, pgs 213-228
- [4]. Integrate the Insight An oracle approach to integrate the big data and white paper. 5. 2012" IBM Global Training outlook " March 2012. http://www.research.ibm.com/files/pdfs/goto_booklets_executive_review_march_12.pdf.
- [5]. "Data warehousing Life cycle and ETL tool kit. Informatica Guide Ralph Kimball
- [6]. D Pyle (2003) "Business Modeling and Data mining" Morgan Kaufmann, an Francisco, CA
- [7]. Barry D Data Warehouse from architecture to implementation Addison Wesley 1997.
- [8]. Krulj D "Design and implementation of data warehouse systems .M.Sc. Thesis, Faculty of Organizational sciences, Belgrade 2003.
- [9]. Lohr, Steve. "The Age of Big Data" "New York Times. 11 Feb 2012. http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=2&pagewanted=all
- [10]. Manyika, James, Michel Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Reixburg and Angela H. Byers. "Big data: The net frontier for innovations, competition and productivity c Kinsey Global Institute (2011) 1-137 May 2011.
- [11]. Boyd, Dana and Crawford, Kate "Six Provocations for Big Data" Working Paper –Oxford Internet Institute 21 Sept. 2011 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431 Boyd, Dana and Crawford, Kate.
- [12]. Bohanec, M (2001) What is Decision Support? Proceedings Information Society IS-2001: Data Mining and Decision Support in action! (pp 86-89), Ljubljana, Slovenia
- [13]. Bajec, M & Krisper, M (2005) .A Methodology and Tool Support for Managing Business Rules in Organizations, Information Systems, 30, 423-443
- [14]. Holsheimer, M (1999) data mining by Business Users :Integrating Data Mining in Business Process. Proceedings International Conference on Knowledge Discovery and Data Mining KDD-99(p.p 266-291), San Diego USA: ACM.

ACKNOWLEDGEMENT

One of the author's Mr. Rajasekharaiah K.M. thanks Ms. Chhaya Dule, Asst. Prof. Jyothy Institute of Technology, Bangalore for her valuable suggestions.

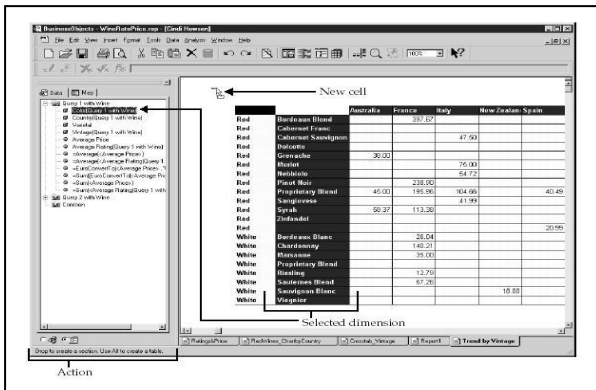
AUTHOR:

Presently Mr. **Rajasekharaiah K.M.** is working as Professor & HOD Department of Computer Science & Engineering, Jnana Vikas Institute of Technology, Bangalore. He has done M.Tech. in Computer Science & Engg. M.Sc. Information Technology, M.Phil. in Computer Science, and PGDIT from reputed Universities, India. He is having 30+ years of total experience including 16 years of Industrial experiences. He is a Life fellow Member of Indian Society for Technical Education (ISTE), New Delhi. He is presently pursuing the doctoral degree in the Branch of Computer Science & Engineering, in the domain area of Data Mining & Warehousing.

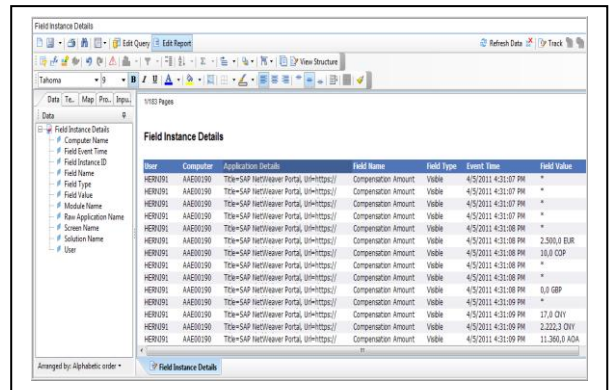


He has research publications in reputed national and international journals. His other area of interests are DBMS, Software Engg., Software Architecture, Computer Networks, Programming Languages, Data Structures and Mobile Computing. He is also a resource scholar for other Engineering Colleges/University

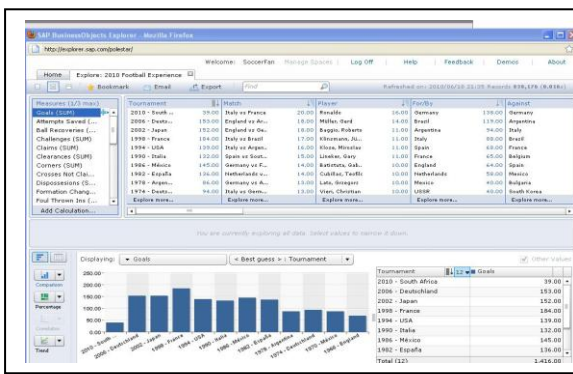
Screen Shots, Reports and Dashboard Snapshots



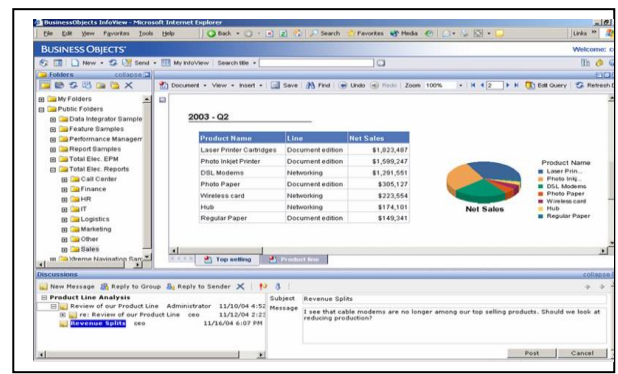
Report: 1



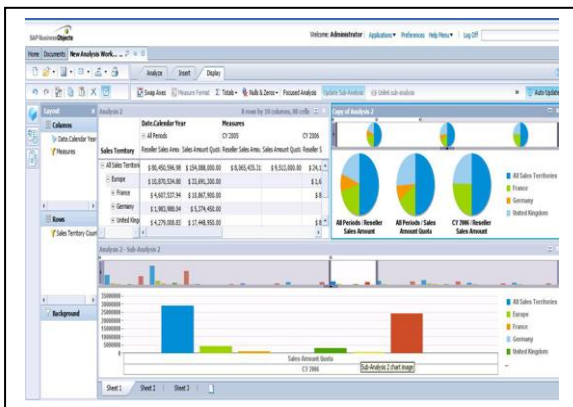
Report: 2



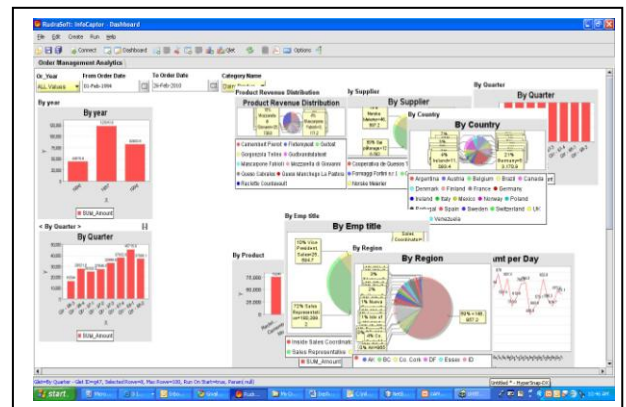
Report: 3



Report: 4



Report: 5



Report: 6