

Developing Business Intelligence Framework to Automate Data Mapping, Validation, and Data Loading from User Application

Muhammad Sohail, Khalid Khan, Rizwan Iqbal, and Noman Hasany

Abstract—In the last decade, Business Intelligence (BI) became a predictable technological advantage of the large enterprises which could afford to buy, implement and maintain BI solutions. Presently, small and medium size enterprises all over the globe have realized competitive and financial benefits of BI. However, limited IT budgets of small companies and BI's high total cost of ownership have created a gap between large and small enterprises where small enterprises do not become fortunate to avail the virtues of BI because of the affordability factor. This research proposes a BI framework which provides automated data mapping, loading from user application to BI framework and also validates it. The implementation of the proposed framework provides convenience of use and effective cost saving as it does not require any technical expertise and any non-technical user can get on with it very well.

Index Terms—Business intelligence, software as a service, small enterprise, SME

I. INTRODUCTION

Modern day enterprises rely on large amounts of data. The handling and conversion of this huge amount of data to knowledge is a highly critical job. In order to convert this data into knowledge and employing the knowledge to make wise decisions involves a combination of technologies and techniques like data warehousing, on-line analysis processing (OLAP) technologies, data mining and other tools. All of these technologies are coupled together to be called Business Intelligence.

The proposed framework for user application to BI framework makes simpler the process of data mapping, validating, and loading based on customer profile. The proposed framework helps in reducing time of data mapping, validating, and loading. It is also helpful for software BI vendors to increase customer base and strengthen relations with their customers. We also identified that there is no single solution available to fulfill all the requirements for on-demand business intelligence solution. This research follows Ralph Kimball who is considered as pioneers of data warehousing and Business Intelligence [1]. It is considered worth to discuss some important terms related to BI before proceeding to upcoming sections.

Manuscript received June 19, 2012; revised August 4, 2012.

Muhammad Sohail is with the BI Solutions working as a SAP Solution Consultant/Trainer. BI Solution is SAP BusinessObjects Partner's company and the largest enterprise solutions provider.

Khalid Khan is with the computer science at PAF-KIET, Pakistan.

Rizwan Iqbal is with the computer science at University Putra Malaysia.

Syed Noman Hasany is with the Qassim University, KSA and working as Assistant Professor.

A. Business Intelligence

There are many professional definitions of BI exist nevertheless none of them is a standard. "Business Intelligence is a term that refers to collecting, structuring, analyzing and leveraging of data to turn it into easy-to-understand information. The term dates back to October 1958 stated in an IBM Journal article by Hans Peter Luhn titled "A Business Intelligence System [2]." Another definition is "Business intelligence (BI) is a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions [3]. The goal of BI is to help decision-makers make more informed and better decisions to guide their business.

B. BI Terms

There are numerous BI terms uses in the market such as operational databases, OLTP, Data warehouse, data mart, OLAP and OLAP Server, Drill down/up, ad-hoc query and analysis, dimension, Fact/Measure, KPI, dashboard, and scorecard [4]-[7].

C. BI Tools

Business intelligence tools are a type of application software designed to report, analyze and present data. The tools generally read data that have been previously stored, often, though not necessarily, in a data warehouse or data mart [8]. Business intelligence tools can be broken down into three categories: 1) Query and reporting; 2) Online analytical processing (OLAP); and 3) Information mining [5]. There are numerous BI tools exist in the market for instance 1) SAP Business Objects; 2) IBM Cognos; 3) ProClarity; 4) and QlikView [9]-[12].

The remaining sections of the paper are as follows. Related Work is discussed in Section 2. In Section 3, we have discussed the proposed framework in detail. Section 4 covers the proposed framework implementation. Therefore Section 5 covers the conclusion and future work.

II. RELATED WORK

Business Intelligence is to get proper knowledge to make intelligent decisions but most BI systems are partial or local, which lack systematic information compilation and processing therefore compilation of extra information can lead to information and knowledge excess which make business intelligence difficult. Usually the business intelligence software scale is massive, development and

maintenance cycle is long, and costly. It is difficult for the small and medium-size companies to develop BI system. The problem is with data model blueprints and ontology style that face amalgamation of business intelligence systems based on different data sources and structure.

Integration framework for business intelligence system (BIS) and services resources based-ontology is proposed as well as data model prototypes and ontology method which can resolve the different data sources and structures integration dilemma [13]. BIS assimilation based on ontology is the central part of the architecture is the repository; this stores configuration information about the IT infrastructure, the metadata for all applications, projects, scenarios, and execution logs. Repositories can be installed on an OLTP relational database. This metadata is stored in a centralized metadata repository. There is only one master repository, which holds the following information: 1) safety information together with users, profiles and access privileges for the data integration units; 2) Topology information together with technologies, definitions of servers and schemas, contexts and languages; 3) Old edition of objects. The information controlled in the master repository is maintained with Topology Manager and Security Manager [13].

A novel system called Entity Recognition in Context of Structured data (EROCS) is used for enhancing Business Intelligence by allowing OLAP tools to analyze structured and unstructured data in a collective manner. EROCS is quite competent in discovering relational entities unseen in unstructured text while management of unstructured data needs further research [14].

EROCS model enables the unstructured information to be correlated with the structured information and this enables cubes to be defined on aggregations of the concurrent data. Consequently current OLAP tools are limited to analyzing structured data. The knowledge base driving EROCS components hold repository of emails and a customer database that have records of all the clients of the enterprise, every email is given to different dictionary-based UIMA Annotators to explain the noun phrases appears in the document. The annotated text is then inputted in the EROCS module which uses the Entity Templates to recognize the best structured records to match with different parts in the text. The production of the EROCS component is a connection that is stored in the relation table. EROCS uses UIMA annotators to recognize the background information in the email, similar to customer, store name, and product particulars (i.e., noun expression, represent as XML annotations), and after that recognize the best operation in the relational database to matches this context [14].

A new business intelligence prototype namely BIaaS comprised a number of services and application to provide rapid development, lessen repair cost, investment price, decrease project risk, tiny development cycle and cheapest solutions. It is very valuable for the business intelligent development of small and medium-sized enterprise. BIaaS applications are mainly comparable to the other applications build using service-oriented design values. Using this architecture, the vendor's hosts numerous clients on a load-balanced farm of the same instance, with every client

data reserved separate, and with configurable metadata gives an exclusive user knowledge and characteristic set for each client. It is scalable to a large number of clients, because the number of servers and instances on the back end can be enhanced or reduced as essential to match demand, without requiring extra re-architecting of the application, and changes or fixes can be resolved to thousands of tenants as simply as a single tenant [15].

A systematic information collection method is designed based on extenics theory to describe information and give an information cube for collecting the integrated information. It helps to collect proper systematic data for business information and find more knowledge by extension transformations [16]. Initially, the information of both goals and conditions to build business intelligence target server are collected. Information for the pathway from condition to goal, all information will be combined with their nature of information. The pathway is called extenics. Extenics is a new discipline for dealing with contradiction with formalize model. Its theoretical pillars are basic element theory and extension set theory. The information for BI can be classified in the following stages: 1) Goal and the conditions; 2) the pathway of BI implementation; 3) Basic Element of Information; and 4) the conjugate nature of Basic Elements [16]. Additional research is needed on information processing with extension reasoning, transformation, intelligent knowledge management processes, and information cube process [16].

III. PROPOSED FRAMEWORK

The proposed BI framework facilitates automated data mapping, validating, and loading from user application. The proposed framework is highly customizable in nature capable of handling customer most critical BI requirements. The framework format consists of forms, database, reports, and dashboards. All information of reports and forms controls such as checkbox, labels, textbox, combo box, radio button, button, link button, tree view, timer, script manager, SQL data source, and conditions are on form. The proposed framework empowers to retrieve and store information required by the form and their controls. This is possible only if we have all the necessary information required by the forms and their controls.

We are covering pharmaceutical sales intelligence. All users information stores are included in a table with all of their business requirements by doing this we build a metadata for application in the database. Requirements, reports, KPI and dashboards are also reusable, for example customer business area, data source, data model names used on subject/business area form can also be used for other user. The client data will be authenticate based on the client profile available in the central metadata repository, by doing this, data will load at staging area and then finally in the target. After this, it will automatically start functioning for business intelligence.

Architecturally, this solution is alike to BIaaS [15] applications constructed using service-oriented design ethics. This framework for business intelligence application provides pre-built data mapping templates, databases, ODS,

data marts, data warehouse, and the KPIs for dashboard and scorecard for each of the business processes in a vertical fashion. These pre-built templates mitigate the risk from users not specifying the requirements completely. Next diagram (see Fig. 1) shows the operation of our model:

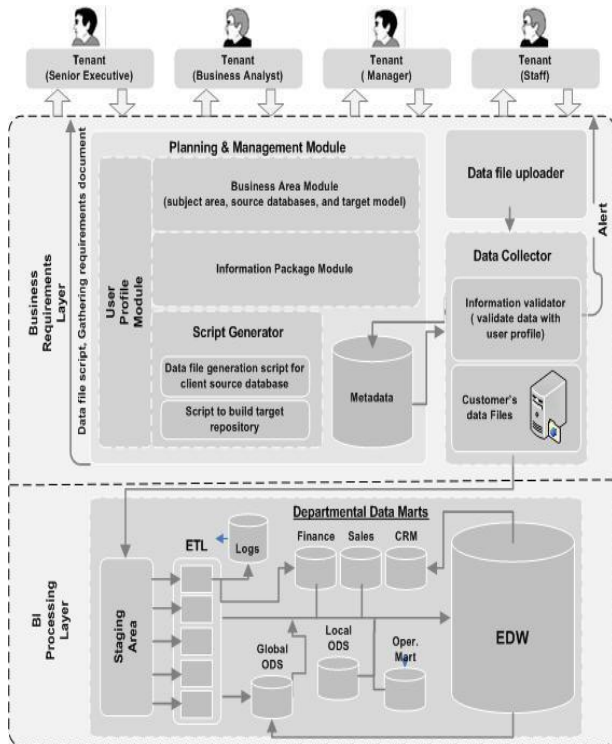


Fig. 1. BI framework to mechanize data mapping, validation, and loading from user application.

This model can be accessible not only from on-premises implementation, but also for cloud platform as well. The most significant distinction is the addition of planning & management module, script generator module, and data file up-loader which is accountable for managing application configuration for individual clients, data mapping, script generating for source database, validating, and file uploading.

The client can upload data file through this user interface, and as file will upload automatically the control will be transferred to data collector process and data file will be validated based on client profile. Once the file is validated successfully it is transferred at a particular location and afterwards it will be loaded at the staging area according to schedule.

A. Planning and Management Module

The processes are controlled by planning and management module which have three sub-modules i.e., user profile module, „business area module, data sources, and data warehouse type”, information package module, script generator routine, and metadata repository.

1) User profile module

This module is accountable for collection of client personal information. The information is stored in central repository. Depending on the content and the amount of information about the user, a business intelligence application facility will be modeled.

2) Business area module

This module is accountable to store client information

about their subject area, source database, and target model for example data warehouse requires, all the contents integrated with the user profile.

3) Information package module

We used information package technique that captures the dimensions, the attributes of the dimensions, and the facts or metrics for a specific subject-area. The first column headings classify the dimensions or fact with their attributes. Using this technique, we are mapping data, and generating two types of scripts 1) Data file generation script for client source database to generate data file for example SQL Server, Oracle, FoxPro etc.; and 2) Script to build target repository that generate dimensional model at target on the basis of user requirement.

B. BI Processing Layer

In BI processing, we make tight coupling between BI and data warehousing [2]. Generally there are multiple services including: 1) ETL tool and data warehouse; 2) Data modeling and reprocessing tool; 3) Meta data exchange; 4) Policy-making engine module rule and measure.

C. Data Presentation Layer

There is a variety of conventional ways to visualize data. In fact, there are much better, thoughtful, creative and absolutely fascinating ways to visualize data. In this service, data presentation layer provides pre-built report. These templates will reduce the time required to design and build the reports and analytics.

D. Data Visualization Layer

Without visualization, this information would be difficult to understand. Data visualization layer provides pre-built dashboards with the right visualizations. These templates will reduce the time required to design and build the dashboard and scorecard.

IV. CONCLUSIONS AND FUTURE WORK

We have focused to automate data mapping, validation, and loading from user application to BI framework. We have proposed a model with respect to customer profile based data mapping, validation and data loading solution which will help for Business intelligence's development for small and medium-sized enterprises. The future plan is to extend this existing proposed framework to support metadata based on auto generated ETL mapping, metrics, and project documentation.

REFERENCES

- [1] Kimball. On the use of conceptual technique to design data warehousing modeling using different approach. [Online]. Available: <http://www.ralphkimball.com/>
- [2] To understand SaaS BI. [Online]. Available: <http://www.dashboardinsight.com/articles/new-concepts-in-business-intelligence/saas-business-intelligence-hot-or-not.aspx>
- [3] What is Business Intelligence (BI)? [Online]. Available: <http://searchdatamanagement.techtarget.com/definition/business-intelligence>
- [4] To understand PKI term in business metric. [Online]. Available: <http://searchcrm.techtarget.com/definition/key-performance-indicator>
- [5] To understand numerous BI terms uses in the market. [Online]. Available: <http://www.redbooks.ibm.com/abstracts/sg245747.html?Open>

- [6] To search BI terms and tools. [Online]. Available: <http://en.wikipedia.org/>
- [7] To understand; how BI tool work. [Online]. Available: <http://www.qlikview.com/>
- [8] To understand different general categories of business intelligence tools. [Online]. Available: http://en.wikipedia.org/wiki/Business_intelligence_tools
- [9] Business intelligence tool. [Online]. Available: <http://www.sap.com/solutions/analytics/business-intelligence/index.epx>
- [10] To understand the traditional BI capabilities with planning, modeling, and analytics, and more. [Online]. Available: <http://www-01.ibm.com/software/analytics/cognos/business-intelligence/>
- [11] To understand the Microsoft BI tool. [Online]. Available: <http://www.microsoft.com/india/bi/products/proclarity/proclarity-overview.aspx>
- [12] What is Business Intelligence (BI)? [Online]. Available: <http://searchdatamanagement.techtarget.com/definition/business-intelligence>
- [13] X. Xu and H. F. Xu, "Developing a framework for business intelligence systems integration based on ontology," *Presented at the 2009 International Conference on Networking and Digital Society*, Guiyang China, 2009.
- [14] M. Bhide, V. Chakravarthy, A. Gupta, H. Gupta, M. Mohania, K. Puniyani, P. Roy, S. Roy, and V. Sengar, "Enhanced business intelligence using EROCS," *Presented at the ICDE 2008*, New Delhi India, 2008.
- [15] T. Fu, "Research on business intelligence pattern based on the baaS," *Presented at the International Symposium on Intelligent Information Technology Application Workshops*, China, 2008.
- [16] X. S. Li, H. L. Qu, Z. X. Zhu, and Y. S. Han, "A systematic information collection method for business intelligence," *Presented at the 2009 International Conference on Electronic Commerce and Business Intelligence*, China, 2009.



Muhammad Sohail received his M.S/MPhil in computer science, software engineering from Karachi Institute of Economics and Technology University, Karachi, Pakistan in 2011; the EMS in computer science, software engineering from Preston University, Karachi, Pakistan in 2002; the M.A in economics from Shah Abdul Latif University, Khairpur, Pakistan in 1993. Currently, he is an employee of BI Solutions working as a SAP Solution Consultant/Trainer. BI Solution is SAP BusinessObjects Partner's company and the largest enterprise solutions provider. He has over 14 years of experience in information technology industry; and worked for well known companies like SAP, NCR, Teradata, Oracle, Info Group and many more. He has been worked on local and international projects using SAP, Teradata, Oracle, and Microsoft technologies; and involved in development of enterprise Business Intelligence, Data Warehousing, ERP development and customized business application solutions. His research interests are in the areas of Self-service business intelligence and data warehousing, distributed systems, Intelligent and Multi Agent Systems, e-Learning and related issues and evolutionary computing. Mr. Sohail has received certificate of Merit and Cash Award from RASG Consulting on successfully

development of ERP & FAMS system; and Certificate of Achievement Award from NCR Corporation on NADRA project.



Khalid Khan is PhD Candidate in computer science at PAF-KIET, Pakistan; the MBA in business administration from PAF-KIET; the MSCS in computer science from SZABIST, Pakistan; and the MCS in computer science from Karachi University, Pakistan. Currently, He is an employee of PAF-Karachi Institute of Economics and Technology and working as Director in PAF- and Chief Executive Officer (CEO) in Institute of Enterprise Information Systems. He has worked as Corporate Trainer in SBP, UBL, ABL banks; IT Consultant for Pakistan Steel Mills; Assistant Professor for Usman Institute of Technology, Hamdard University from August 2002 to May 2005, and Software Engineer for IT Works from January 2001 to July 2002. His research interests are in the areas of Intelligent and Multi Agent Systems, Security and Management Issues in Next Generation Networks, Retransmission Issues in Wireless Networks, and E-Learning and related issues



Rizwan Iqbal is PhD Candidate in computer science at University Putra Malaysia, the MS in computer science from PAF-KIET, Pakistan; and the BS in computer science from Syed University Of Engineering And Technology, Pakistan. He has worked as consultant, Corporate Trainer, programmer and faculty member in Karachi Institute of Management Sciences (KIMS). He has conducted different Workshops and Presentations in different universities on natural language interfaces to Ontology, information retrieval and knowledge management. He has internationally traveled for research work in China, Malaysia, and Singapore. Mr. Rizwan has international publications 1) A Negation Query Engine for Complex Query Transformations. *Pertanika Journal of Science and Technology (JST)*; 2) A. Negation query handling engine for natural language interfaces to ontology, *IEEE CAMP'12*; and 3) Fourth International Conference on Machine Vision (ICMV 2011): Machine Vision, Image Processing, and Pattern Analysis, Dec 2011, Singapore, Proc. SPIE 8349, 834903 (2011); doi:10.1117/12.920081.



Dr. Syed Noman Hasany has PhD in Computer Systems Engineering from University Putra Malaysia in 2010; the M.E in Computer Systems Engineering from NED University of Engineering and Technology in 2002; the B.E in Computer Systems Engineering from NED University of Engineering and Technology Pakistan in 1995. Currently, He is an employee of Qassim University, KSA and working as Assistant Professor. He has supervised both bachelor and masters students in final year projects in CS and IT. He has worked as Assistant Professor at Sir Syed University of Engineering and Technology, Pakistan.