Cloud hosted business-data driven BI platforms

Cloud hosted Business-Data driven BI Platforms

Siddhesh Rajesh Inamdar

Rutgers University, Dept. of MSIS

Email : siddhesh.inamdar@rutgers.edu

Kemal Gursoy

Rutgers University, Dept. of MSIS

Email : kgursoy@business.rutgers.edu

Abstract

Business Intelligence in Platform as a Service (PaaS) for information analysis is increasingly being considered for its applications in the enterprises because of its advantages. It is widely used for Data Analysis, Customer Churn Prediction, etc. However, the challenges that the traditional BI platform faces includes the tremendous volume of data, high time and space complexity of algorithms and the incompatibility in the Integration to the BI tools. Thus instead of having traditional Data Warehouses for storage which require high maintenance and storages, companies are moving towards more efficient cloud based storages and platform simply because of the fact that they don't personally have to maintain the cloud and the computing speed in Cloud is much faster than traditional storages.

Keywords— Business Intelligence, Analysis, Data, Storage, Cloud Platforms, Data Warehouse

1. Introduction

In this paper, we consider the implement of a BI platform architecture which is extendable in the high level and can be easily customized and integrated, that we can add specified business behavior (program) into the platform according to our given scenario, which we call Business Driven. As a system, we discuss every part of the system, in the comparison of the traditional system. Furthermore, we apply cloud computing system into an application scenario that nearly meets real-world requirements of Health Insurance industry by employing a large volume of data obtained from the Healthcare operators, and the high efficiency of the system is demonstrated.

Business Intelligence (BI) has been widely used in enterprise in recent years. Business Intelligence as a software is used to collect information, then make an analysis, in order to provide a basement for decision-makers. It is an important application for enterprise to make a decision through the analysis and processing of information. For instance, Insurance companies use Enrollment data for the past months or years to analyze the trend and target next potential customers using data such as age group, sex, income, the user behavior to dig out potential customers, then make a marketing strategy for them in order to expand the user community.

However, as the exponential growth of information data from enterprise, traditional BI tools cannot resolve such massive amount of data in an acceptable period of time. On the other hand, with the complexity of enterprise business, there is much more requirement for the integration of BI software, and we need to modify BI tools according to our specified application because of the variable scenarios, but traditional BI software are quite difficult to integrate with each other, and costs much to modify the tools. [2] To solve these problems, we come up with BI platform based on cloud computing. The BI platform focuses on the mechanism of the work-flow for the integration of BI software, and dynamically deployment of plugins (BI tools). What's more, we parallel algorithms of BI software, for the purpose of processing large-scale dataset.

Business Intelligence solutions are the products to support this goal. [4] They provide ways to extract the data or information, structure, and analyze the data and summarize the results for decision makers. The necessary investments in Business Intelligence solutions are high

and it is not surprising that Cloud offerings in this area became important during the last years. Business Intelligence is one of the leading software categories used in the Cloud. The lower financial barrier makes Cloud based Business Intelligence also interesting for small and medium sized businesses, which were not able to afford them before. However, using Business Intelligence in the Cloud brings challenges regarding security and integration. The integration topic is mostly looked at with the question: how does the data that needs to be analyzed get into the BI solution? This article focuses also on integration, but looks at the integration of a BI solution with other Cloud products. While a simple upload of the data to be analyzed into a Cloud based BI solution is already possible, the visualization of the results and the interactive parts of the analysis are provided inside the Cloud based BI solution itself. Common usage scenarios of integrated Cloud applications however, demand an integration of the analysis into the user interface of other Cloud based application. After an initial introduction into Business Intelligence and Cloud computing, the interaction scenario of an integrated Cloud based BI solution gets defined. Based on this scenario, requirements are identified and needed interfaces are defined.

2. History and Evolution of Cloud

The evolution of cloud computing can be separated into three basic phases: 1. The Idea Phase - This phase incepted in the early 1960s with the emergence of utility and grid computing and lasted till pre-internet bubble era. Joseph Carl Robnett Licklider was the founder of cloud computing. 2. The Pre-cloud Phase- The pre-cloud phase originated in 1999 and extended to 2006. In this phase the internet as the mechanism to provide Application as Service. 3. The Cloud Phase- The much talked about real cloud phase started in the year 2007 when the classification of IaaS, PaaS, and SaaS got formalized.

There are three main types of Cloud deployments:

1. Public Cloud – which generally means that is open for public use. Users can store data on it without having the privacy that a secure network provides. This type of Cloud can also be offered as free to use. The main concern is privacy because providers don't guarantee it.
2. Private Cloud – this is an infrastructure that is used by one user/company. This can reside whether internally the company building (data center) or externally (provider). This is a highly secured Cloud and is preferred by most companies because, when is resided externally, maintenance can be offered by the provider. Another advantage is the resources allocated to a single client. By doing this, the infrastructure is custom built to offer great performance and stability.
3. Hybrid Cloud – this is a combination of private and public Cloud. By doing this, users can specify which data resides on public Cloud and which on private. The advantage of using Hybrid Cloud is that the expenses can be reduced. The data is aggregated from both sources in order to provide the results that users needed.

The history of cloud computing has witnessed some very interesting breakthroughs launched by some of the leading computer/web organizations of the world. As the quantity of data produced every day goes on increasing, the technology to compute it and to store it should evolve too, this need of storing and high level computing led to the popularity and growth of cloud computing. And as we know, raw data is useless unless knowledge is bought out of it. Thus Business Intelligence, Analytics and Cloud we merged to help organizations grow by making sense out of raw data. Prediction, analytics, 5G, Extra layer of security etc are the advantages cloud brings to the table.

2.1 Traditional BI Flow using Warehousing

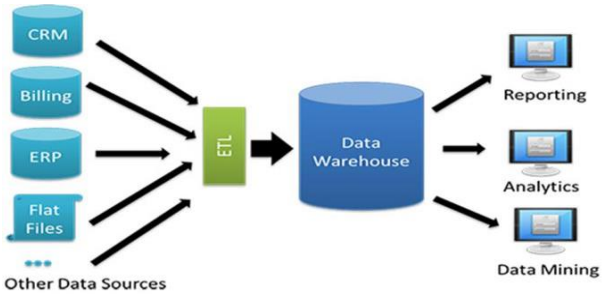

Figure 2.1

Traditionally, preparing data for analytics involved a time-consuming and complex process known as extract, transform and load (ETL), which required specific tools and specialized skills.

[1] In the ETL process, the IT team **extracts** data from one or more sources, typically in a scheduled batch job, **transforms** it into a normalized format, and **loads** it into a centralized data warehouse. The warehouse stores the business's historical data in ways that make it easy to generate standardized reports and dashboards that can be used throughout the organization.

Data warehouses remain a viable strategy for supporting analytics. Perhaps their greatest benefit resides in the fact that the data models that govern how they're loaded into a data warehouse are centrally controlled by IT.

For instance, once the business settles on the attributes that meaningfully define a high-value customer for its purposes (i.e. at least one purchase within the last six months, purchase over X dollars etc.), IT uses these attributes to normalize data and to set up standardized reports that track the business's key performance indicators (KPIs).

However, there are also three major issues with ETL and data warehousing:

1. **Data warehouses are inflexible.** Once the data model has been defined and the data has been loaded into the warehouse, the paths for analyzing the data get frozen into place, limiting the number of potential insights that can be derived from it.

2. **Only IT can prepare data.** ETL tools aren't designed for business users, even business analysts. Moreover, since one of the points of the data warehouse is to normalize the organization's data in standardized formats, it makes organizational sense to task a single unit with preparing it.

3. **Data warehouses emphasize reporting over ad hoc exploration.** Data warehouses are architected to support scheduled reports or real-time dashboards created by the IT team. The rigidly structured ways in which they store data lend themselves to reports and dashboards that track pre-defined KPIs, but aren't as well suited to exploratory analysis. For instance, the particular data attributes in which a business analyst is interested in order to answer a one-time inquiry from the CEO may not have been considered when the data was transformed into a normalized format and loaded into the warehouse.

2.2 Proposed model

Business intelligence (BI) has been historically one of the most resource intensive applications. It comprises a number of data warehouses created by fetching decision-support data from organization wide databases. The data warehouses are updated at frequent intervals through appropriate queries executed on the business processing and transactional databases online analytical processing (OLAP) is the

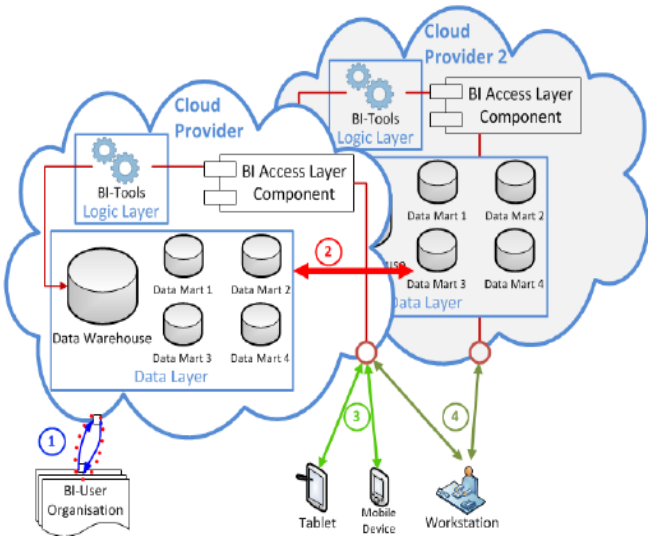

Figure 2.2

User-end interface of BI that is designed to present multi-dimensional graphical reports to the end users. OLAP employs data cubes formed as a result of multidimensional queries run on an array of data warehouses. Furthermore, an OLAP application fetches data from the data warehouses, organizes them in highly complex multidimensional data cubes, and presents to the users through user defined and configured GUI dashboards. BI and OLAP framework has a high business utility, because it helps in locating and eliminating or solving business process deficiencies, inefficient process steps and waste process steps. [3] A BI and OLAP framework is expected to provide timely, accurate, organized and integrated information to business decision makers. Despite of excellent business utility of BI and OLAP framework, many business owners were compelled to look for its alternative because of uncontrolled increase in computing and storage resource requirements in self hosted environments. At some stage, the cost of maintaining and upgrading the BI and OLAP framework becomes unjustified for a business. However, the unique selling points of cloud computing offer exactly what businesses need to successfully run BI and OLAP frameworks-unlimited resources, resource elasticity (resources on demand), moderate usage costs, high uptime and availability, high security, no hassles of upgrading and maintaining loads of servers and databases, and so on.

Hence, it is hereby argued that cloud computing has the potential to offer a new lease of life on BI and OLAP framework. Moreover, it is also argued that cloud computing can extend the power of BI and OLAP to small- and medium-scale businesses, which could not have afforded the framework in self-hosted IT infrastructures. However, it is important to establish a framework for implementing BI and OLAP on cloud computing platform.

BI in the cloud is a game changing phase of IT, as it makes BI finally affordable and accessible as compared to traditional BI. On the cloud, the matrices in the OLAP cubes can be formed using the web data warehousing concept making use of XML data files using DTD (document type definition) described XML programming language. The data structures in the cubes are formed using the DTD parsed XML files [11]. The DTD format helps an XML file to exhibit relational properties of a conventional database. This is what enables the OLAP cubes stored on the cloud making use of XML data files following DTD structures (called web cubes). This also helps the BI system to make use of web services components thus ensuring better performance on the cloud. The entire OLAP framework comprising the dashboards and the data analytics layer can be hosted as SaaS. The BI and OLAP framework software platforms available for cloud hosting are SAP, IBM Cognos and Web-Sphere Dashboards dashboard, Oracle business objects and Salesforce.com. The integration of data warehouses (XML based) and OLTP/DSS databases can be hosted on PaaS. The underlying servers and databases can be hosted on IaaS mode of cloud hosting. For optimum performance on the cloud, the servers and database arrays should be implemented as a massively parallel system capable of processing large scale parallel queries. The databases on the cloud needs to be implemented in the form of a massively parallel system to support high demand elasticity of BI and OLAP framework. [9] A centralized schema object may be designed to hold the details and privileges of all tenants on the cloud. Each schema object holding the data files may be massively partitioned such that each partition can be held by a separate server on a large scale server array. The IaaS provider should be capable of rapid expansion of the server array making use of virtualized array expansion. In this way, it may be possible to serve one partition through more than one servers that can enhance the performance of BI. The IaaS provider should keep a close watch on both load distribution and response time patterns and make effective network changes to ensure that the network load is also distributed evenly. The OLAP application hosted on the cloud may not be web services compatible. For example, consider a cloud-based E-commerce insurance industry having data stored on cloud, to create analysis or visualizations for predictions on enrollment data related to enrollees or even insurance agents data, an OLAP cube is supposed to be maintained having the flexibility of addition or removal or data from different sources, basically having ETL capabilities. This ETL feature will be the backend of the BI application. The front end initiate with the user selecting one of the cubes created in the backend ETL and then creating visualizations according to the requirements and within his access of rights to the data. [8] To make an OLAP application compatible to web services architecture, the SaaS provider may allow the creation of an intermediate layer to host a dependency graph that helps in dropping the attributes not needed in the finalized XML data cube.

Pros and cons regarding the integration of a BI solution into Cloud

Pros:

1. Scalability and elasticity;
2. Reduced costs;
3. Ease of use and access;
4. Cloud relational database;
5. Availability;
6. Hardware maintenance.

Cons:

1. Privacy;
2. Government Regulations

3. *Key Challenges Faced*

1. Compliance of the BI application with web services architectural standards (and the standards

defined by the SaaS or PaaS provider, like Google Apps standards)

2. Deployment of massively parallel data-warehousing system with evenly distributed query load and even patterns of response times from all database servers. The IaaS provider should effectively use the virtualized server array management and expansion to meet the resources on demand.
3. The network architecture should be designed in such a way that the query load can be evenly distributed among the servers in an array. This will ensure even query processing response times by the servers in an array. If the server array employs storage area networking for storing the XML data files and the OLAP cubes, the data fetching from various storage devices should again be evenly distributed by virtue of appropriate network connections.

4. *Conclusion*

With Cloud based application getting more important for companies, the integration between Cloud based services becomes more important. As integration of business applications was always a challenge inside companies, it is not less challenging over company boundaries. Using the example of Cloud based Business Intelligence, it is shown how relevant this integration is and that BI solutions currently offered in the Cloud are only providing a one-way integration of data via ETL processes into the Cloud based BI solution. However, more and more Cloud based applications demand a full integration where also the results of the analysis done in the BI solutions are provided back via interfaces into the user interface of the Cloud based application. This technical concept shows how this interface can be provided and also provides an example of this ongoing communication. Such a concept of a Cloud based BI solution providing its services could even have a different revenue model, with a usage/request based payment.

5. *References*

[1] https://panoply.io/data-warehouse-guide/data-warehouse-architecture-traditional-vs-cloud/

[2] Business Intelligence as a Service for Cloudbased Applications Volker Herwig Faculty of Building Technologies and Informatics, Postfach 45 01 55, 99051 Erfurt, Germany, volker.herwig@fh-erfurt.de, http://www.fh-erfurt.de

[3] Cloud Computing and Business Intelligence Alexandru Adrian TOLE

Romanian – American University, Bucharest, Romania

[4] DESIGN AND IMPLEMENTATION OF BUSINESS-DRIVEN BI PLATFORM BASED ON CLOUD COMPUTING Bin Wu, Lei Qin

[5] Frolick, M.N. and T.R. Ariyachandra, Business performance management: One truth. Information Systems Management, 2006.

[6] Khajeh-Hosseini, A., D. Greenwood, and We. Sommerville. Cloud migration: A case study of migrating an enterprise it system to iaas. in Cloud Computing, 2010 IEEE 3rd International Conference on. 2010. IEEE. [13] Glonet, GloNet Platform Design. Online: https://sites.google.com/site/glonetproject/8-download/GloNet-D3.1.ppdf, Jun 2012.

[7] Zimmer, M., H. Baars, and H. Kemper. The Impact of Agility Requirements on Business Intelligence Architectures. in System Science (HICSS), 2012 45th Hawaii International Conference on. 2012. IEEE.

[8] Baars, H. and H.G. Kemper, Ubiquitous Computing–an Application Domain for Business Intelligence in the Cloud? 2011.

[9] Thomson, W.J.J. and J.S. van der Walt, Business intelligence in the cloud South African Journal of Information Management, 2010. 12(1).

[10] Linder, J.C., Transformational outsourcing. MIT Sloan Management Review, 2004. 45(2): p. 52-58.

[11] Qie, L. and H. Baars, Die Cloud als neuer Ansatz zur Erhöhung der BIAgilität? BI Spektrum, 2012. 7(2): p. 26-29.