

Cloud Business Intelligence Survey

Mai Kasem

Information Systems Department
Faculty of Computers and Information
Cairo University, Egypt

Ehab E. Hassanein

Information Systems Department
Faculty of Computers and Information
Cairo University, Egypt

ABSTRACT

Organizations are always seeking to become more intelligent and achieve a greater competitive advantage by using Business Intelligence (BI) solutions. The new phenomenon of business intelligence is how businesses handle data. One of the instruments that can make the BI tools more accessible is Cloud Computing. This paper outlines some important definitions of Cloud Computing and Business Intelligence, the importance of business intelligence and some problems in traditional business intelligence solutions. Also, the benefits and challenges of Cloud Business Intelligence is discussed with a comparison between public and private cloud and how to choose the appropriate Cloud Business Intelligence for an organization.

Keywords

Cloud Computing, Business Intelligence, Cloud Business Intelligence.

1. INTRODUCTION

Over recent years, the business landscape has witnessed rapid evolution. Each organization has tended to become more scalable, flexible and intelligent, using new Business Intelligence (BI) solutions. For businesses to make better decisions and take more appropriate action, it is important to apply data analysis techniques to their information. This will help them to define a strategy to improve their business, and identify the issues that could affect their business development in order to become a dynamic business that can meet today's challenges [1]. Using business intelligence increases the knowledge of the business environment and helps in making better plans for the future.

Nowadays, cloud computing is considered one of the most important technologies, and many experts expect that cloud computing will have a great impact on information technology (IT) processes and the IT marketplace. The cloud provides flexibility, scalability and enables organizations to react faster to the needs of their business. One of the primary benefits of the cloud is supporting organizations with business agility that enables them to respond quickly and effectively to the ever changing business environment [5].

Cloud computing affects most current industries, including having a significant impact on the business intelligence industry. BI helps organizations in analyzing their data and turning it into valuable business information that could help them in making better decisions [3]. Organizations tend to invest more in BI solutions based on cloud computing, called Cloud BI or Software as a Service BI, because investing in traditional BI solutions become unpractical and unattractive [4]. Although using BI on the cloud has a number of benefits, it also comes with a number of risks. It is critical to have a thorough understanding of the nature of cloud computing, and

know the benefits and risks of using BI tools on the cloud [19].

2. DEFINITIONS

2.1 Cloud Computing

Gartner defines cloud computing as “a style of computing where massively scalable IT-related capabilities are provided ‘as a service’ using Internet technologies to multiple external customers” [6].

The National Institute of Standards and Technology (NIST) defines cloud computing as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” [7].

B. Furht describes cloud computing as “a new style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet” [8].

G. Yuvraj and R. Vijay propose a new definition for cloud computing: “Clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load(scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay- per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs.” [2].

2.2 Business Intelligence

Gartner defines business intelligence as “an umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance” [9].

Moss, L. T., and Atre, S. define it as follows: “BI is neither a product nor a system. It is an architecture and a collection of integrated operational as well as decision-support applications and databases that provide the business community easy access to business data” [10].

B. Patrick and C. Bob define business intelligence as “getting the right information, to the right decision makers, at the right time” [11]. It provides businesses with a solution to access and analyze data sources and get useful information to make informed and intelligent business decisions [1].

2.3 Cloud Business Intelligence

Cloud BI is a revolutionary concept of delivering business intelligence capabilities as a service using cloud based architecture that comes at a lower cost yet has faster

deployment and flexibility [2]. "Software as a Service business intelligence (SaaS BI) is a delivery model for business intelligence in which applications are typically deployed outside of a company's firewall at a hosted location and accessed by an end user with a secure Internet connection." [12].

3. HISTORY OF BUSINESS INTELLIGENCE

In a 1958 article, IBM researcher Hans Peter Luhn used the term business intelligence. He defined intelligence as: "the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal." [20]. Business Intelligence was born in the early 90s to help organizations in analyzing data in order to understand the situation of their business and to make better decisions [13]. In the mid 90s, BI became an important topic for the academic world, and ten years of research managed to transform a bundle of naive techniques into a well-founded approach to information extraction and processing [13].

4. WHY BI IS CRITICAL TO BUSINESS DECISION MAKING

The problems that face organizations without BI solutions:

- Generating reports takes a lot of time and effort, and requires examining unstructured data like working on spreadsheets and text files.
- Lack of consistency and accuracy of the manually generated reports.
- Including and relating different types of data manually into one single report is very prone to errors, especially if data relates to different departments.
- An organization's data can only be accessed via individual communication, which may lead to inflexibility and confusion.

BI Solutions can provide the following features:

- Automating reports and solution generation, saving time, effort and reducing human errors.
- Giving the ability to build and run complex reports and make strategic decisions.
- Extracting data from multiple sources and formats, and generating reports easily at any time when needed in a single informative view.
- Tracking an organization's revenue and plan for future growth.
- Tracking an organization's progress and identifying areas that need further enhancement by analyzing historical data.
- Finding opportunities to make the organization better by analyzing the current and future needs, and helping in setting realistic goals for the organization.
- Reducing wasted time and resources spent in correcting wrong manual reports and decisions made based on wrong data.
- Helping organizations to take advantage of trends and new opportunities as they come up, and making accurate predictions about where the business is going.

- Getting more visibility on organizational behaviour and turning that into valuable knowledge.
- Improving an organization's decision making by building a big informative picture of the existing data.

5. BUSINESS INTELLIGENCE ON THE CLOUD

Based on Gartner's list of CIOs technology priority, Business Intelligence topped the list from 2007 to 2009, ranked 5th in 2010, and again took top rank in 2011-12; also Cloud Computing held 1st rank in 2010 and has remained in the top 5 during the last 5 years [2].

5.1 Problems in traditional BI:

Cost: Organizations need to pay for hardware, maintenance, software licensing and installation, which are very expensive.

Setup: To run the BI application, the organization has to set up software and system platforms that could be complex.

Usability: Most traditional BI applications are too difficult for most users to operate.

Mobile Accessibility: Traditional BI has limited access via the browsers running on mobile devices.

Reporting: Traditional BI tools are not designed for cooperative reporting; cooperation can only be obtained by emailing the reports via the organization managers then starting a discussion over mail threads. This could cause redundant data and conflict in decisions.

Feedback: With traditional BI, it is difficult to measure the usability or conduct effective auditing.

5.2 Using BI as a Service

Gartner research [21] mentions some considerations that must be taken into account when deciding to use BI as a service. Firstly, the organization needs to know whether there is a predefined BI as a service that fits its requirements or whether it will need some customization. When the service contract has expired, the organization has to make the decision about signing another rental contract or having the BI application shut off. Also there are some hidden fees that could be added to the service offering, such as fees for additional information sources or fees for integration services to third-party applications. Most BI as a service applications are hosted applications, which means data will not be on-site. This implies that the organization needs to know its clients' concerns, as they might have serious concerns or contractual terms about storing their information on a third-party server. The organization has to define clear process integrity that ensures that the business process delivers the expected outcome. Since most of the business processes reside outside the organization, ensuring process integrity will become more difficult. Finally, the outages of the internet must be taken into account because, if the internet is down with a hosted BI as a service solution, the application may also be down.

5.3 BI Phases

5.3.1 Collecting data

The first and most important stage in Business Intelligence is collecting data from business data sources. Organizations have a lot of data stored in various databases across all their departments, which may seem unrelated.

5.3.2 Convert business data to information

Business intelligence is responsible for gathering and converting the data into information and helping the decision makers to find the important pieces of information and the relationships between them.

5.3.3 Linking databases

Data is integrated and linked from one or more disparate sources in the data warehouse. Data warehouses store current as well as historical data. Often BI applications use data gathered from a data warehouse; however, not all business intelligence applications require a data warehouse.

5.3.4 Performing Queries

Organizations can make a good business analysis based on multilayer queries. With the queries results, the organization can know its business needs and can plan for future growth.

5.3.5 Reports creation

The final step is acting on the collected data, and deciding how to represent it to the organization's managers. Representing data through reports enables decision makers to see the big picture of their business needs and make better decisions.

6. BI COMPONENTS

All business intelligence systems require specific components to produce business intelligence [14]. Some of these components are data warehouses, ETL (extract, transform, and load, three database functions that are combined into one tool to pull data out of one database and place it into another database) tools, multidimensional analysis, data mining, and visualization [15].

6.1 Data warehouses

A data warehouse is a database used for reporting and data analysis. It is a central repository of data which is created by integrating data from one or more disparate sources. Data warehouses are used for creating trending reports for senior management reports, such as annual and quarterly comparisons [16].

6.2 ETL tools

ETL tools and processes are responsible for the extraction of data from one or many source systems, as they transform data from many different formats into a common format and then load that data into a data warehouse [14].

6.3 Multidimensional analysis

OLAP, Online Analytical Processing, is an example of multidimensional analysis. It allows users to analyze database information from multiple database systems at one time [17]. While relational databases are considered to be two-dimensional, OLAP data is multidimensional, which means that it enables users to analyze multidimensional data interactively from multiple perspectives.

6.4 Data mining

Data mining techniques are designed to identify relationships and rules within a data warehouse, and then create a report of these relationships and rules [14].

6.5 Visualization

It is concerned with presenting data in a pictorial or graphical format [18, 23]. The main goal of data visualization is to communicate information clearly and effectively through graphical presentation. As more and more data is collected and analyzed, decision makers at all levels welcome it; data

visualization software enables them to see analytical results presented visually, communicate concepts and hypotheses to others, and even predict the future [18].

7. BUSINESS INTELLIGENCE ARCHITECTURE

The basic architecture needed to run a business intelligence solution in the cloud is illustrated in Fig 1 [2]. The lower layers are formed by hardware and software systems. The minimum elements that have to be offered by the cloud computing provider are [2]:

Hardware: This refers to processing, storage, and networks. An important aspect is the processing speed of the hardware on which the data will be physically stored.

Software: This refers to the operating systems and drivers required to handle the hardware.

Data integration: This refers to the tools needed to perform the ETL and data cleansing processes.

Database: This refers to the relational or multidimensional database systems that contain the data.

Data warehousing tools: They are the set of applications that allow the creation and maintenance of the data warehouse.

BI tools: They are the set of front-end applications that read and analyze data that has been previously stored in a data warehouse.

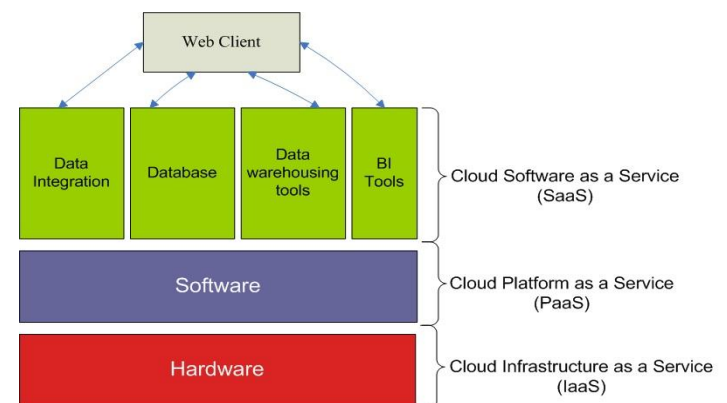


Fig 1: BI on the Cloud Architecture

8. BENEFITS OF BUSINESS INTELLIGENCE ON THE CLOUD

There are many benefits of Business Intelligence on the Cloud. Listed below are some of its benefits.

Lower Cost: In BI on the cloud, you only pay for what you need, which reduces the total cost of your BI system through reducing overhead resources. The cloud also reduces maintenance and administration costs.

Scalability: Cloud BI systems have greater scalability than hosting and operating them locally. It enables the organization to react faster and more efficiently to the business needs, making it easier to upscale or downscale the needed resources.

Flexibility: Since Cloud BI systems are not dependent on local hardware resources, they provide more flexibility to the users. Users can access analyzed information and reports anytime and from anywhere.

Disaster recovery: Cloud systems provide the ability of backing up data offsite in multiple locations, which allows for disaster recovery. This makes the cloud BI solution more reliable than the on-site BI solution.

9. CHALLENGES OF BUSINESS INTELLIGENCE ON THE CLOUD

However, while there are many benefits of using cloud-based business intelligence, there are also many risks. The following is a discussion of the cloud-based BI challenges.

Security: Concerning the risks of Cloud Computing integration, approximately 75% of the Chief Information Officers and IT specialists consider security as the number one risk [4]. With cloud computing, data is stored and accessed via the internet. For some businesses, it is essential to keep the data on the premises as their data is confidential. There are several solutions to secure data including encryption, but it is the organization's responsibility to encrypt the data appropriately on the cloud. Although the virtualization process is essential in any cloud technology, it might cause highly technical security breaches as the data will be stored forever on virtual hardware even when its index is deleted [19].

Data Recovery: Since the data is saved on the cloud, there will be a probability of data violation, the recovery process will be difficult and it will take a long time due to dependency on third parties.

Availability: The cloud BI availability depends on the third party's server availability. If the server is down, the cloud BI users could lose control of their data.

Incompatibility: One of the most important features of BI is the ability to import and export data from various data sources to be reused with other enterprise applications. The incompatibility with other enterprise applications can occur because of the separation of BI tools from other departments of the organization.

Challenges of data transfer: If an organization has a lot of data, moving it over the internet in a reasonable time will be a challenge. If the network is slow, moving data will take a lot of time. The organization also may have some concerns about the confidentiality of its data, so the process of moving data needs to be more secure.

Performance: If storage resources are separate from server resources, there may be considerable latency in accessing data, especially when accessing large amounts of data [2].

Choosing the right vendor: The organization needs to know more about the vendor's offerings and choose the most suitable one for its business. This process may be difficult as there are many cloud BI vendors with many offers.

Total Cost: It is difficult to calculate the whole budget of the needed resources as their costs are variable.

10. COMPARISON BETWEEN PUBLIC AND PRIVATE CLOUD FOR BI

To make the best decision for a business, the differences between public and private cloud need to be clarified. Table 1 shows the strong and weak points in both of them.

Table 1: Public vs. Private Cloud

Features	Public Cloud	Private Cloud
Dedication	Shared	Dedicated
Pricing	Variable (Pay per use)	Fixed
Uses	Variable or low workload	High workload
Control	Low	High
Security	Low	High
Performance	Low	High
Scalability	Yes, servers can be automatically scaled according to the workload	Yes, servers and storage can be added and scaled based on need
Maintenance	No maintenance	IT expertise needed for maintenance
Time saving	Yes	No

Public cloud is a shared cloud, so a careful choice is required to avoid the noisy neighbours effect from sharing. In public cloud, a pay-as-you-go model can be used, which means you pay only for what you need. This is preferred for variable or low workload business like development and testing websites or pay-per-use applications. In public cloud, organizations have little control of their resources as all the resources will be on the cloud not on-site. Security is an outstanding question in cloud systems especially in the public shared cloud, because all the data is stored and accessed via the internet. Public cloud does not require maintenance or need time changing or updating the software, because the software is not on-site and because the cloud provider is responsible for it. This means that internal IT employees are not required for maintaining the servers, which translates into saving time and resources.

There are many Cloud service providers (such as Amazon, Google, Informatica, etc.). Amazon, GoGrid, Google, Sun Microsystems, and Rackspace do not provide Cloud BI; Informatica provides only data integration of BI in Cloud using Amazon EC2 cloud; IBM has private BI Cloud called Blue Insight; RightScale provides BI Cloud using data processing and business intelligence tools, partnering with database firm Vertica, open source business intelligence maker, Jaspersoft, and data integration specialist, Talend. [22].

Here is a comparison between BI cloud providers with their features [20]:

RightScale: It is a public BI cloud, which means that it is publically available not just to their organization. It is not open source which means that the source code of BI cloud is not available. It is a full BI cloud which means that all functions of Business Intelligence are addressed. Common functions of business intelligence technologies are reporting, online analytical processing, analytics, benchmarking, predictive analytics and data mining. The data management technique that is used by it is Specialized Analytic Databases. It does not support forecasting; however, it is scalable and flexible.

IBM (Blue Insight): It is a private BI cloud and is not open source. The data management technique that is used by it has more than a petabyte of data storage. It not only supports forecasting, but it is also scalable and flexible.

Salesforce.com (Sales Cloud 2): It is a public BI cloud and is not open source. The data management technique it uses is Automated Data Management. It supports forecasting but it is not flexible and it has low scalability.

Salesforce.com (Service Cloud 2): It is a public BI cloud and is not an open source. It is not using any data management technique. It not only support forecasting, but it is also not flexible and it has low scalability.

Informatica: It is a public BI cloud and is not open source. The data management technique that is used by it is Data Migration, Replication and Archiving. It not only supports forecasting but it is also not flexible and it has low scalability.

This comparison concludes that RightScale BI Cloud is better among available Public BI Clouds whereas Private BI Cloud has been implemented only by IBM [20].

11. CONCLUSION AND FUTURE WORK

Cloud computing plays an important role in the future of Business Intelligence. Business intelligence in the cloud has been developed in order to enhance flexibility of implementation, availability, scalability and increased performance of business intelligence software. This paper discussed the importance of business intelligence for decision making and mentioned some considerations that must be taken into account before choosing between BI as-a-service offerings. It also illustrated how BI works and showed BI components and architecture. Also, the benefits and challenges of Cloud BI were discussed, the differences between public and private cloud outlined, and how to choose the best one for an organization.

Although there are many benefits of Cloud BI, there are some challenges that still need more research. Public cloud BI has some security and privacy risks, as all services and data reside on the cloud and can only be accessed via the internet. It is important to find reasonable ways to mitigate these risks and to find fast and secure ways for data transfer via the internet. Also, more research is needed to allow the accurate calculation of the needed resources' budget, and to handle the variable cost problem.

12. REFERENCES

- [1] Thompson, W.J.J. & Van der Walt, J.S., 2010, "Business intelligence in the cloud", SA Journal of Information Management 12(1), Art. #445, 5 pages. DOI: 10.4102/sajim. v12i1.445.
- [2] Yuvraj Singh Gurjar & Vijay Singh Rathore, 2013, "Cloud Business Intelligence – Is What Business Need Today", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-6.
- [3] "Business Intelligence in the Cloud", <http://cloudcomputingtopics.com/2011/09/business-intelligence-in-the-cloud/>, Requested in December 2013.
- [4] Marinela MIRCEA, Bogdan GHILIC – MICU and Marian STOICA, 2011, "COMBINING BUSINESS INTELLIGENCE WITH CLOUD COMPUTING TO DELIVERY AGILITY IN ACTUAL ECONOMY", Economic Computation & Economic Cybernetics Studies & Research, Vol. 45 Issue 1, p1.
- [5] Bharat Chandra and Meena Iyer, 2010, "BI in the Cloud: Defining the Architecture for Quick Wins", SETLabs Infosys Technologies Limited, Vol. 8, No 1.
- [6] Stamford, Conn., 2008, "Gartner Says Cloud Computing Will Be As Influential As E-business", <http://www.gartner.com/newsroom/id/707508>, Requested in January 2014.
- [7] Peter Mell and Timothy Grance, 2011, "The Nist Definition of Cloud Computing". National Institute of Standards and Technology Special Publication, 800 - 145.
- [8] Borko Furht, 2010, "Cloud computing fundamentals", In Handbook of cloud computing (pp. 3-19). Springer US.
- [9] "Business Intelligence (BI)", <http://www.gartner.com/it-glossary/business-intelligence-bi/>, Requested in January 2014.
- [10] Moss, L. T., & Atre, S., 2003, "Business intelligence roadmap: the complete project lifecycle for decision-support applications", Addison-Wesley Professional.
- [11] Patrick Baesel and Bob Croker, 2010, "Business Intelligence as a Shared Service", Oracle Corporation World Headquarters 500 Oracle Parkway Redwood Shores, CA 94065 U.S.A.
- [12] "Software as a Service BI (SaaS BI)", <http://searchbusinessanalytics.techtarget.com/definition/Software-as-a-Service-BI-SaaS-BI>, Requested in February 2014.
- [13] M. Golfarelli, S. Rizzi, and I. Cella, 2004, "Beyond datawarehousing: What's next in business intelligence?" In Proc. DOLAP, pages 1–6.
- [14] John Lloyd, 2011, "Identifying Key Components of Business Intelligence Systems and Their Role in Managerial Decision making", Master of Science, University of Oregon Applied Information Management Program.
- [15] Maciej Pondel, 2013, "Business Intelligence as a service in a cloud environment", Proceedings of the 2013 Federated Conference on Computer Science and Information Systems pp. 1269–1271.
- [16] "Data warehouse", http://en.wikipedia.org/wiki/Data_warehouse, Requested in January 2014.
- [17] "What is OLAP (Online Analytical Processing)?" <http://www.techterms.com/definition/olap>, Requested in January 2014.
- [18] "Data Visualization", http://www.sas.com/en_us/insights/big-data/data-visualization.html, Requested in January 2014.
- [19] Christina Tamer, Mary Kiley, Noushin Ashrafi & Jean-Pierre Kuilboer, "RISKS AND BENEFITS OF BUSINESS INTELLIGENCE IN THE CLOUD", University of Massachusetts Boston, Management Science and Information Systems Department, 100 Morrissey Blvd, Boston, MA, 02125 United States of America.
- [20] Sonam Chawla, 2011, "Design and Development of Business Intelligence Portal in Cloud Computing", Master of Engineering in Software Engineering,

Computer Science and Engineering Department, Thapar University.

- [21] Bill Hostmann, 2009, “Business Intelligence as a Service: Findings and Recommendations”, Gartner, Inc., ID Number: G00164653.

- [22] Carl Brooks, 2009, “Cloud BI from RightScale and Jaspersoft is priced for penny pinchers but lacks polish”, Requested in January 2014.

- [23] “What is Data Visualization”, <http://www.slideshare.net/QueryIO/hadoop-big-data-visualization>, Requested in January 2014.