Data Warehousing based on a Cloud Computing Architecture

Walid Qassim Qwaider
Management Information Systems Dept. College of Science and Humanities
Majmaah University, Ghat. AL Majmaah,
Saudi Arabia

ABSTRACT

The paper explains the use of data storage based on cloud computing architecture. A Data Warehouse is a data warehouse center that stores and retrieves data continuously, and cloud computing is a new application model that works to deliver services over the Internet. It also eliminates the requirements for users to plan for provision and management allowing companies to start from small resources and only increase when there is a high demand for service.

This paper explains in detail about data warehouse based on cloud computing Architecture, the implementation of the data warehouse in cloud computing architecture is the primary objective of this study. Furthermore, this study provides the use of data warehousing based on cloud computing architecture and the opportunities of organizing and updating data warehousing on the rising cloud computing architecture.

Keywords

Cloud computing; Data Warehouse Architecture; Cloud Provider; Cloud Services

1. INTRODUCTION

The concept of cloud computing has recently emerged, which means, in general, the services that are made through devices and programs connected to a network of servers that carry data in a virtual cloud that ensures continuous, uninterrupted connectivity with different devices (computers, tablets, smartphones, etc.) And are therefore accessed from anywhere and at any time [1]. The IT sector has seen fierce competition between significant telecommunications and technology companies by offering their cloud computing services providing entrants to enterprises and businesses. Especially emerging ones, by providing cloud services to ready-made programs without the need to establish an IT infrastructure, leaving hardware and data management at the disposal of the organization [4].

Data Warehousing uses new technology systems to store large amounts of unmatched information for quick decision-making. These are called report systems versus production systems, including databases, and the architecture used is designed to extract knowledge to help make decisions. There is an increasing need for data warehouses in various large companies.

Therefore, the focus is more on collecting data stores while maintaining the unique rules. It is necessary to feed the stores periodically if there is an amendment or modification in the regulations and classify the input by material managing specific programs described Middleware that uses particular merge data in each item called Schema d integration determines how and how information is transferred [2].

The relationship between a data warehouse and cloud computing is an integrative relationship that involves data processing, data acquisition, storage, analysis, and retrieval. Cloud computing is a new technology approach while the data warehouse is a significant data phenomenon growing in a fast way. Large data technology allows users to process distributed queries from different sets of data and get results promptly [12]. The problem arises in how data warehousing is based on a cloud computing architecture. Data processing represents a new challenge in the adoption of a data warehouse on cloud computing architecture.

This paper is arranged as follows. Section II introduces cloud computing and types of cloud services, section III. Then, presents the data warehousing Architecture, the build phase of the data warehouse, advantages data warehouse. Section V is devoted to data warehouse-based cloud computing architecture, an aspect of data warehousing based on a cloud computing architecture, benefits of data warehousing-based cloud computing. Finally, the conclusions and future work.

2. CLOUD COMPUTING

Technical that depends on the transfer of processing and storage space of the computer to the so-called cloud, which is accessed through the Internet [1]. Also referred to as a term referring to Accessible via the Internet offering cheap and secure on-demand platforms that can be located quickly and conveniently [3].

The term cloud computing was inspired by the cloud code, which was often used to represent the Internet in maps and graphs. As with many other new technologies, it meant different things to different people. Many suppliers are motivated to escalate their product range [14].

The idea of cloud computing dates back to the 1960s, as John McCarthy expressed the view that" computing may be organized to become a public service one day [15].

However, cloud applications did not emerge until the early 2000s, when Microsoft expanded the concept of software use through the Web followed by many companies. Which launched some services that rely on this technology. Google launched services to take advantage of this technology alone, but launched 2009 an integrated operating system of computers working through the concept of cloud computing [14].

The definition that appears to be the collection of all these elements is the US National Institute of Standards. Cloud computing can also be defined as a model through which we can access the network from one side to another, with the participation of a group of computers, systems, applications, servers, and other resources, through which we can disseminate information and deliver it quickly and with minimal effort [5].

3. TYPES OF CLOUD SERVICES

Cloud computing technology services are broadly divided into three services, and customers have the right to choose one or

more of the desired services [6] [13]. Figure 1 shows the cloud-computing services are:

1) *Software as a service (SaaS):* It means in the first place that the end-user will deal with the site for this purpose remotely over the Internet. A similar example is a remote service such as CRM or Customer Relationship Management, such as those provided by Salesforce, or billing software such as Invoicera or the data center displayed by Amazon Web Services.
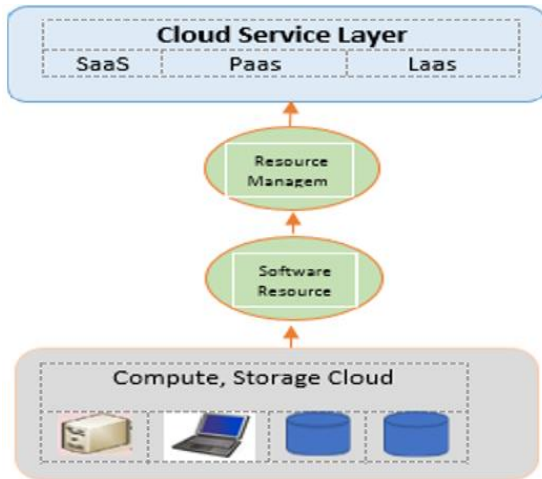

Fig 1. Cloud Computing Architecture

2) *Platform as a service (PaaS):* The best example of PaaS is the Google App Store. It is primarily aimed at the activities of developers who want to deploy their applications directly in the cloud server and do not much attention to the connection to the server infrastructure.

3) *Infrastructure as a service (IaaS):* It permits developers to take the highest level of direct interaction with the server infrastructure. It also allows them to deploy their applications in remote environments and to control them with greater powers remotely.

So far, the SaaS model is dominant in the current industry, and various forces in the market have discovered its capabilities and settings.

4. DATA WAREHOUSE ARCHITECTURE

It "Is the process of data collection from multiple, diverse, heterogeneous, time-changing and object-oriented data sources that support decision-making." [2], to assist in data access for time analysis, data extraction, processing, presentation, and representation in an appropriate format. As defined as "An integrated and differentiated time group and non-volatile data sets that rely on the actual implementation of a specific data set model that supports decisions and the storage of information that the Organization needs to make strategic decisions"[7]. As evidenced by the architecture of the data warehouse construction in Figure 2, which explains its general structure. The importance of data repositories has emerged with the presence of large companies. Each of these groups manages their databases (marketing, finance, management) [8].
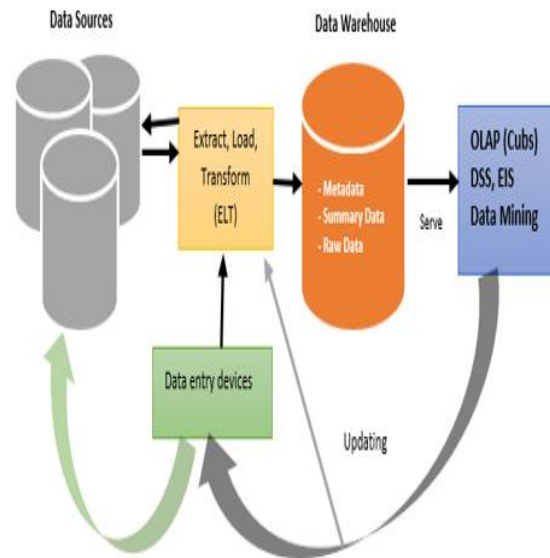

Fig 2. Data Warehouse Architecture

It can be said that data repositories are an extensive and broad database through which access to all information related to an institution or a company. The benefits of a data warehouse: The use of the data warehouse is one of the most prominent advantages, where users can access a significant amount of information. Data warehouses are characterized by allowing users to collect data from different locations in one place. Data repositories help create an area where you can explore, modify, and change stored data in your operating systems [7].

5. BUILD PHASE OF DATA WAREHOUSE

Building a data warehouse in an enterprise is not easy; it requires a professional team in process of analyzing systems and business, with expert technical skills in the group. These steps are as follows: [9]

1) Creating data preparation area: When creating a database, storage must be considered appropriate and usually broad to store all data sent by other systems to clean and modify the data before uploading it to the data warehouse. At this stage, the design of the database is to be consistent with the model of the data warehouse.

2) Creating a data warehouse in which a data warehouse is built, which has loaded after extraction and purification. The stores have always been designed to allow the existence of relationships of different dimensions.

3) Data warehouse fragmentation: Some companies divide the data warehouse into groups related to the data market, in particular, for example, data associated with the marketing department and another in the production or finance department. The division is based on the company's departments and functions in groups in the data market.

4) Converting, purifying, and extracting data: Data have been extracted from several sources to the side where the data has been prepared, translated, and processed on request. We can also update, process, and merge data if necessary, as well as prospecting, data recovery, and deletion of irrelevant data.

5) Loading data in the data warehouse at that stage, the data preparation area is uploaded to the data warehouse, and the data is tested.

6) Creating OLAP & DSS applications at this stage, the forms of processing systems have been performed online

analytically. These applications display data in several dimensions and use sophisticated algorithms for data analysis, also at that stage using data mining tools.

6. DATA WAREHOUSING-BASED CLOUD COMPUTING WAREHOUSING

Data storage increases the value of cloud computing. The process of collecting, collecting, and retrieving data in large quantities is intense on the interface such as Excel. A separate area is configured in cloud computing to store data continuously and is referenced when needed from various data sources and provided as a service to users [10] [7].

The computing cloud addresses the problem of high data intensification because it will provide many advantages such as cost savings and incentives for companies and send them away when there is no need to use them. There is a set of standards in cloud computing that the user, can access to the network such as the use of systems of all kinds of applications that can be saved and used at the lowest cost and effort.

Suppose that the file related to the distribution of salaries of employees contains data and the administration wants to increase the wages of employees, for example, the same data, so we will help the computerized cloud to provide these sources and update them through access to the data of salaries of employees and processing and retrieval. Figure 3, explains the data storage integrated with cloud computing architecture.
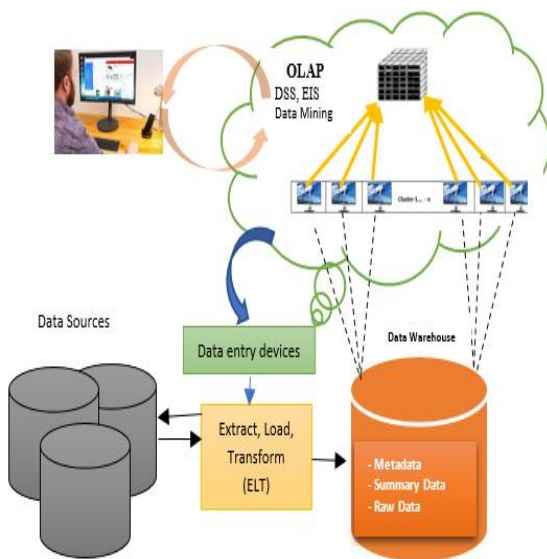

Fig 3. The Impact of Cloud Computing -Based Data Warehousing Architecture

7. BENEFITS OF DATA WAREHOUSING-BASED CLOUD COMPUTING

There are many advantages of cloud storage other than file storage service, including [11]:

1) Reduce costs: Cloud computing has freed you from spending your head on fixed assets. Cloud computing has reduced the price of the company's computers, helping you to pay for your services quickly, helping you save material costs and you can use them in other things within your company, as well as original setup and management.

2) High flexibility: Cloud-based services are ideal for companies as growth grows and demand fluctuates. If your business needs increase, it is easy to increase your cloud capacity. Similarly, if you need to reduce your cloud space again, your provider or service provider will cut it. New and this level of agility can give companies a real advantage over the solvent, and through the use of cloud computing, it is not surprising that IT managers and IT managers have arranged the "operational agility" of companies through their reliance on cloud computing.

3) Cloud computing helps you work from anywhere: You can do your work from anywhere you have Internet access so you can be able to do your own business, and most cloud services provide particular applications to work on mobile, that is, you are not specific to the computer you are working on in your office.

4) Collaboration: These techniques help your team to edit and share documents anytime, anywhere, making them able to do more together. Cloud-based applications give them a file that shows the division of work helps them make the required updates in real time and provides them a full vision of their cooperation with each other.

5) Retrieve and backup company data: Companies of any size must have the ability to recover intractable data from the company's data, but many small businesses lack the required cash and experience. Unfortunately, cloud computing now helps small and large companies to recover data According to the Aberdeen Group of Companies. Small businesses account for twice as many companies that implemented cloud-based backup and recovery solutions, saving time and avoiding much investment in fixed assets of the company it also led to increasing the efficiency and expertise of these companies as part of a package of cloud computing.

6) Automatic software updates: One of the best and most beautiful features of cloud computing is that it provides your servers in a place where your company does not exist and is out of sight and beyond the boundaries of your company. Service providers have delivered a package of regular software updates - including security updates and helping you cloud compute also maintain your company's system. Cloud computing lets you focus on things that are important to your company, such as expanding your business.

7) Cloud computing increases the control of documents within the company: The majority of employees and partners within the company collaborate on document monitoring, but cloud computing has provided many of these procedures. In the past, workers had to send files back and forth as attachments in an email to work on them one user at a time. Eventually, cloud computing made it possible to work on the same company data from multiple employees simultaneously and from anywhere in the world, and cloud computing reduced the conflicting chaos around file content, regarding formats and titles.

8) Increase security to save files within your company: The loss of your laptop will result in a loss of data that could lead to a significant work problem that could lead to the loss of a lot of money. Cloud technologies provide greater security when these positions are stored because the data is stored in the cloud, and you can access it no matter what happens on your device. You can also erase the data on the remote computer so that it does not fall into the wrong hands.

8. CONCLUSION AND FUTURE WORK

The study based on data storage based on cloud computing architecture, research has shown that the adoption of data warehousing in the cloud computing architecture has many benefits such as high flexibility, fast information retrieval, reliability, and low costs. The data storage system in cloud computing is parallel, distributed, and very accurate. This paper presents the importance of building and data warehousing based on cloud computing architecture, and that the relationship between extensive data and cloud computing is an integrative relationship. Demand for the development of data warehousing tools is increasing with the relationship between the computerized cloud and the data warehouse becoming more stringent. Therefore, in future work, they propose a methodology to improve the efficiency of Cloud computing and semantic web technologies architecture.

9. REFERENCES

[1] Aishwarya S. Patil. (2015). A review of DM-based Cloud Computing, International Journal of Research In Science & Engineering (IJRSE), Volume: 1 Special Issue: 1.

[2] Jeba, J. R., Misbha D. S. (2016). A Study of DW on Cloud Environment, International Journal of Innovative Research in Science (IJIRS), Engineering and Technology - ISO 3297: 2007 Certified Organization Vol. 5, Issue 7.

[3] Ren Ying, Lv Hong. (2016). DM Based on CCT, MATEC Web of Conferences 61, 07015, DOI: 10.1051/matecconf/20166107015.

[4] Creeger M. (2009). CTO Roundtable: Cloud Computing Communications (CCC) of the ACM, vol. 52, no. 8, pp. 50-56.

[5] Mell, Peter; Grance, Timotht.(2011). The NIST definition of Cloud Computing (CC): Recommendations of the National Institute of Standards and Technology.NIST National Institute of Standards and Technology.

[6] Patidar A., Patidar V. (2015). Analysis of Solving E-Learning Problems (ASELP) using Cloud Computing. International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 11.

[7] Bharathi, K. Kala; K. Sandhya Sree. (2015). Recent Developments in Data Warehouse and Data Mining In Cloud Computing, International Journal of Computer Science Engineering and Technology(IJCSET), Vol 5, Issue 2,31-34.

[8] William H. Inmon. (2013). Book, "Building the Data Warehouse." Page. 1-30.

[9] Parek, Arpit. (2013). Introduction to Data Warehouse with OLTP and OLAP, International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume2 Issue 8, Page No. 2569-2573.

[10] Vaibhav, Gandhi; Jignesh, Prajapati; and Pinesh, Darji. (2012). Cloud Computing with Data Warehousing and Analysis Market June 2009, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 1, Issue 3.

[11] Verma, Hemlata. (2013). Data-warehousing on Cloud Computing. IJARCET, Volume 2, Issue 2.

[12] Sonia, O., S., Alba, C. L. (2017). D W and significant data integration, International Journal of Computer Science & Information Technology (IJCSIT) Vol 9, No 2.

[13] Nabeel, Z. A., Sufian, M. K. (2017).CC and Big Data is there a Relation between the Two: A Study. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, pp. 6970-6982.

[14] Vojtěch Uhlíř. (2017). Network-based cloud is benchmarking. Master Thesis, Czech Technical University in Prague Faculty of Electrical Engineering.

[15] C.P.Chandgude. (2017). The core of cloud computing. Int. Journal of Engineering Research and Application: ISSN: 2248-9622, Vol. 7, Issue 4, (Part -3), pp.76-81.