SUBSTANTIAL DATA IN CLOUD COMPUTING

Shweta Singhal[1] **Ankita Sharma[2]**

ABSTRACT
This paper describes how data mining is used in cloud computing. Basically Data Mining is a technique that is used for extracting useful information from raw or unused data. The integration of data mining techniques into normal day-to-day activities has become common place .This technique also helps business to run in a more efficient by lower costs. Data mining techniques and applications are very much needed in the cloud computing paradigm. The implementation of data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage.

KEYWORDS
Cloud Computing, OLAP, OLTP, Data Warehousing.

INTRODUCTION
The term Cloud computing is a basically a new way of referring to the sharing of computer resources and it is an alternative to having local servers handle applications. In this process of cloud computing a large number of computers are grouped together and also there is full utilization of resources which are on demand, and on pay per cycle basis. The end users are unaware of the location of the server .They just pin up their application and start the work. Cloud computing is fully enabled by virtualization technology (hypervisors) and virtual appliances. An application is known as virtual which contains all the application along with all the components which are required with a good operating system.

The main advantage of Cloud Computing is the flexibility which makes it distinguishable from grid or utility computing and Software As A Service (SaaS).This ability to launch new instances of the application with minimum labour and expense allows the application providers to:

a) Scale up and down rapidly
b) Recover from a failure
c) Bring up development or test instances
d) Roll out new versions to the customer base
e) Efficiently load test an application

BASIC ARCHITECTURE OF CLOUD COMPUTING
A cloud client consists of computer hardware and/or computer software that relies on cloud computing for application delivery.

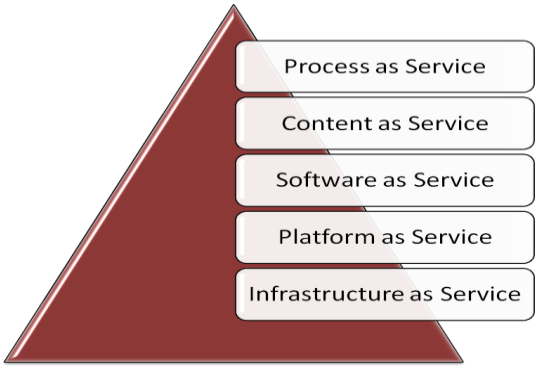


Figure 1: Layered Architecture of Cloud Computing

LAYERED ARCHITECTURE
The architecture or designing the layers of cloud computing can be done using several ways. A logical grouping of the layers depends on the services provided like hardware, networking infrastructure, database, storage, hardware and so on. The key features of cloud computing are kept in mind while designing the system. These are efficiency, availability and scalability.

While using Cloud Computing services, a workstation and the internet connection are required to connect to the "cloud" .Cloud Computing is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on a service-level agreement. According to this definition, Cloud Computing refers to the practice of delivering software and infrastructure as a service on a “pay as you go” basis. As shown in fig.1 the reference model of cloud computing consists of five layered which are described below:

1) APPLICATION
Cloud application services or "Software as a Service (SaaS)" eliminating the need to install and run the application on the customer's own computers and simplifying maintenance and support.

2) PLATFORM
Cloud platform services or "Platform as a Service (PaaS)" deliver a computing platform and/or solution stack as a service, often consuming cloud infrastructure and sustaining cloud applications. It facilitates deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers.

Copyright © IJICCT–JUN 2013;Vol 1,Issue 1;ISSN 2347-7202

This journal is cited as : **JIMS 8i-Int’l J. of Inf. Comm. & Computing Technology(IJICCT)**

3) INFRASTRUCTURE

Cloud infrastructure services or "Infrastructure as a Service (IaaS)" delivers computer infrastructure, typically a platform virtualization environment, as a service. Rather than purchasing servers, software, data center space or network equipment, clients instead buy those resources as a fully outsourced service.

4) SERVER

The server's layer consists of computer hardware and/or computer software products that are specifically designed for the delivery of cloud services, including multi-core processors, cloud-specific operating systems and combined offerings.

5) CLIENT

A cloud client consists of computer hardware and/or computer software that relies on cloud computing for application delivery. [1]

HOW TO MANAGE DATA IN CLOUD

The data Centers all around are dependent on cloud computing to give a flexible, scalable and efficient amount of data All the Data Centers basically are now dependent on cloud computing to deliver the flexibility, scalability, efficiency and speed to the data.

Implementation of the data in innovative ways helps to explore businesses unparalleled opportunities to understand and respond to the needs of a rapidly moving and changing market. Business analysts need access to a wide variety of data, in real time, for diverse uses such as standard and specialized reporting, online analytical processing(OLAP), advanced analytics such as data mining and data marts, and staging for detailed or real-time data sources. The strategic requirement for real-time data warehousing also adds online transaction processing (OLTP) workloads to the mix, furthering the strain on existing infrastructures.

In today's world now we see a large amount of data is being used and therefore all accommodation possibilities have to be made for this huge amount of data. So a number of data centers are now making use of the cloud or a cloud like model just to have a pool of resources. Rex Wang, vice president of product marketing at Oracle. "By pooling or sharing resources across multiple applications, cloud delivers the efficiency and increased flexibility data centers require to help the business be market-competitive. For data processing and analysis, having a shared, standardized, and consolidated database architecture for all DW and OLTP workloads is an effective strategy. [2]

DATA MINING IN CLOUD COMPUTING

Data Mining using cloud computing has become an area because it now covers almost all business and scientific computing.

"Cloud computing denotes the new trend in Internet services that rely on clouds of servers to handle tasks. Data mining in cloud computing is the process of extracting structured information from unstructured or semi-structured web data sources.

The data mining in Cloud Computing allows organizations to centralize the management of software and data storage, with assurance of efficient, reliable and secure services for their users. [3].

CLOUD COMPUTING MEETS DATA WAREHOUSING

The effect of data warehousing will be firstly it will raise the bar. Secondly it will also help in pushing the pendulum back in the direction of data marts.. Thirdly, it will also help to lower down the hype generated in the press.

Initially, it raises the bar on cloud. It has capabilities such as aggregation, roll up and some query related intensive operations and therefore may be exposed at the interface whether as Excel-like functions or actual API calls. Cloud computing is the opposite of traditional data warehousing. Also Cloud computing requires data to be location independent, transparent and shippable, whereas the data warehouse is a centralized, persistent data store. In the race between computing power and the explosion of data, large volumes of data continue to be stuffed behind I/O subsystems with limited bandwidth. Still, with cloud computing (as with web services), the service, not the database, is the primary data integration method.

Secondly data warehousing in the cloud will push the pendulum back in the direction of data marts because it is hard to image anyone moving an exiting multiterabyte data warehouse to the cloud. Therefore such kind of databases will be exposed to intra-enterprise corporate clouds, so the database will need to be web service friendly. In any case, it is easy to imagine setting up a new ad hoc analytic app based on an existing infrastructure and a data pull of modest size. This will address the problem of data mart proliferation since it will make clear the cost and provide incentives for the business to throw it away when it is no longer needed.

Thirdly, the inevitable hype around cloud computing will get a good dose of reality when it confronts the realities of data warehousing. Questions that a client surely needs to ask are: If

I want to host the data myself, is there a tool to move it? Since this might be special project, how much does it cost? What are the constraints on tariffs (costs)? It is a familiar and proven revenue model to give away the razor and charge a little bit extra for the razor blade. Technology lock-in! It is an easy prediction to make that something like that will occur once the

Copyright © IJICCT–JUN 2013;Vol 1,Issue 1;ISSN 2347-7202

This journal is cited as : **JIMS 8i-Int'l J. of Inf. Comm. & Computing Technology(IJICCT)**

computing model has been demonstrated to be scalable, reliable and popular.

Companies need only pay for the storage they actually use as it is also possible for companies by utilizing actual virtual storage features like thin provisioning. [3]

Companies do not need to install physical storage devices in their own datacenter or offices, but the fact that storage has to be placed anywhere stays the same (maybe localization costs are lower in offshore locations).

Storage maintenance tasks, such as backup, data replication, and purchasing additional storage devices are offloaded to the responsibility of a service provider, allowing organizations to focus on their core business, but the fact stays the same that someone has to pay for the administrative effort for these tasks.

COMPUTING A SOLUTION

To derive benefits from this combination, the IT organizations must look for a solution that combines a wide array of features. Below given are the features which are required:

a) A single integrated platform that combines servers, low-latency storage, high-performance networking, and intelligent software for rapid deployment and scalability as well as easy interoperability with minimal performance tuning requirements.

b) Extreme performance for all workloads including data warehousing, OLTP and mixed workloads, to ensure fast access to larger volumes of data so more data is available for more purposes in less time.

c) An archival mode optimized for space, to maximize server and storage space and deliver significant storage capacity at extremely low cost, lowering the total cost of ownership (TCO) on existing hardware and limiting the size and frequency of future storage purchases.

d) Ease of implementation, to maximize return on investment (ROI) by freeing IT resources to focus on strategic efforts rather than database maintenance.

e) A resource manager to manage workloads, to allow IT to set a minimum level of input and output for each database and the workloads within them, making it easier to guarantee service levels and ensure consistent performance.

CONCLUSION

Therefore finally we conclude that Data Mining technologies using Cloud computing has become an important part of the business in order to increase the flexibility and scalability of the business. Also using cloud we can also predict the growth and future of the projects in business. A large number if IT organizations have adopted a number to techniques to gain out benefits out of this combination i.e. Data mining in cloud.

This paper also further gives an overview of how the huge amount of data is accommodated in the cloud. Also by integrating the data mining tools and techniques the use of cloud has become a compulsion now. So storing large volumes of data in cloud is definitely been beneficial to the business of organizations.

REFERENCES

[1] "Introduction to Cloud Computing Architecture", White Paper, 1st Edition, June 2009, Sun Microsystems.
[2] "Oracle In The Cloud", Oracle Datasheet, www.oracle.com
[3] http://en.wikipedia.org/wiki/Cloud_storage
[4] "Data mining in Cloud Computing" , R.Ş. Petre, Database Systems Journal vol. III, no. 3/2012, PP 67-71
[5] cloud-computing-guide-1561727 "Data Center Best Practices: Managing Data With Cloud Computing, White paper, www.oracle.com
[6] "A Review on Data mining from Past to the Future", Venkatadri.M, Dr. Lokanatha C. Reddy, International Journal of Computer Applications (0975 – 8887), Volume 15– No.7, February 2011, pp. 19-22

Copyright © IJICCT–JUN 2013;Vol 1,Issue 1;ISSN 2347-7202

This journal is cited as : **JIMS 8i-Int'l J. of Inf. Comm. & Computing Technology(IJICCT)**