

A Study of Data Warehousing on Cloud Environment

Dr. J. R. Jeba ¹, D. S. Misbha ²Associate Professor and Head, Dept. of Computer Applications, Noorul Islam Centre for Higher Education,
Kanyakumari, Tamil Nadu, India ¹Assistant Professor, Dept. of Computer Applications, Nesamony Memorial Christian College, Tamil Nadu, India ²

ABSTRACT: Currently, the cloud environment has been receiving considerable excitement and concentration in the media and blogosphere. Cloud computing is an internet-based computing which supplies shared information, resources and software to processors and other appliances on demand. It is the major source for data sharing, storage and processing. Due to its high storage capacity and low cost it becomes more popular in IT's infrastructure. Data warehouse is a collection of data from multiple data sources which are nonvolatile, integrated, time variant and subject-oriented that supports decision-making process. This paper provides a detailed study about data warehousing in cloud environment. The application of data warehouse in cloud environment is the main objective of this study. Moreover, this study suggests the use of data warehousing over cloud environment and the opportunities of organizing data warehousing on the rising cloud computing environments.

KEYWORDS: Cloud Environment, Data Warehouse, Cloud Provider, Cloud Services, Cloud.

I. INTRODUCTION

Cloud computing has become known as a basic paradigm for information analysis, data processing and storage with the development of internet. Cloud environment allows on demand network access to a distributed collection of computing resources [1] [2]. The increasing amount of information and the multiplicity of new data sources have made novel challenges to manage the amount of information analysis becomes very ease.

Nowadays, a large amount of data is being used and thus all accommodation possibilities have to be made for this large amount of data. So a number of data centers are now making use of the cloud to have a pool of resources. For all Data Warehouse and online transaction processing workloads a distributed, normalized, and consolidated database architecture is an effectual strategy for data processing and analysis. Data warehousing [4] denotes a collection of databases that are used to store information and generate query by considering the important data across a whole enterprise. Data warehouse is a centralized storage scheme of organizing data from different sources.

II. CLOUD COMPUTING ENVIRONMENT

Cloud computing is a model that enables ubiquitous, on-demand access to a shared collection of customizable computing resources to afford actively compatible infrastructure for requests, information and file storage. [5] The cost of computation, application hosting, data storage and delivery has been reduced significantly with the evolution of cloud computing.

Cloud computing is an expedient approach to experience direct cost avail. Moreover, cloud computing has the possibility to convert a data center to an unpredictable priced environment from a resource-intensive set up.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

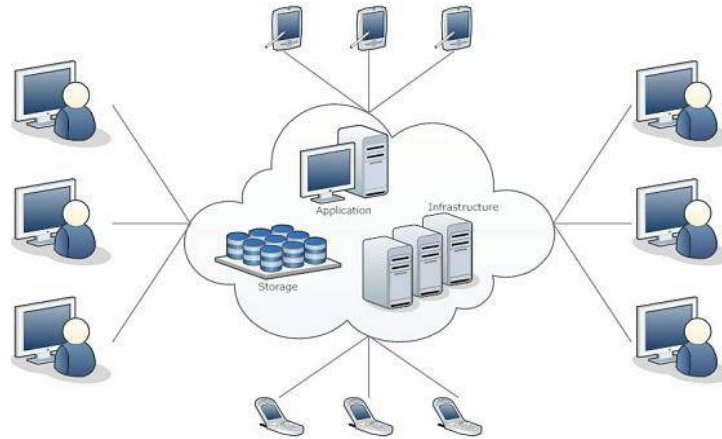


Fig. 1 Hypothetical View of Cloud Environment

Characteristics of Cloud Computing

Cloud is expected to have the following five characteristics that are described below. [6]

On-demand self-service: Without interacting with the service providers, a customer can independently provision computing proficiencies, such as system storage and server time.

Broad network access: Competences that are available over the network are accessed through typical mechanisms that confirm its use by diverse weak or strong client platforms.

Resource pooling: According to the needs of customer, they are dynamically allocated and deallocated with different physical and virtual resources. The cloud providers' computing resources are then shared to various customers' using a multi-tenant model. Due to location independence, the customer has no control or knowledge over the location of available resources.

Rapid elasticity: In some cases, in order to quickly scale out and scale in, automatically the proficiencies are quickly and flexibly provisioned to the customers. The proficiencies can be purchased in any number at any time by the consumer since they are unlimitedly available.

Measured Service: Using a metering facility, cloud systems can automatically control and optimize resource usage based on the type of services at some level of abstraction. The utilization of resources can be controlled, observed and reported which allow both the provider and consumer to make use of their exploited services efficiently.

Service Models

The three categories of services offered by cloud providers are discussed below.

Software as a Service (SaaS): This model allows the customers to use the entire application of the providers running on the cloud environment. Companies such as Google, Microsoft, etc offer SaaS.

Platform as a Service (PaaS): This model deploys a layer of software or development environment on to the cloud infrastructure and offered as a service. Higher levels of service can be built over this layer. To meet the manageability and scalability requirements of the applications, a predetermined combination of OS and application servers are suggested by the PaaS providers.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

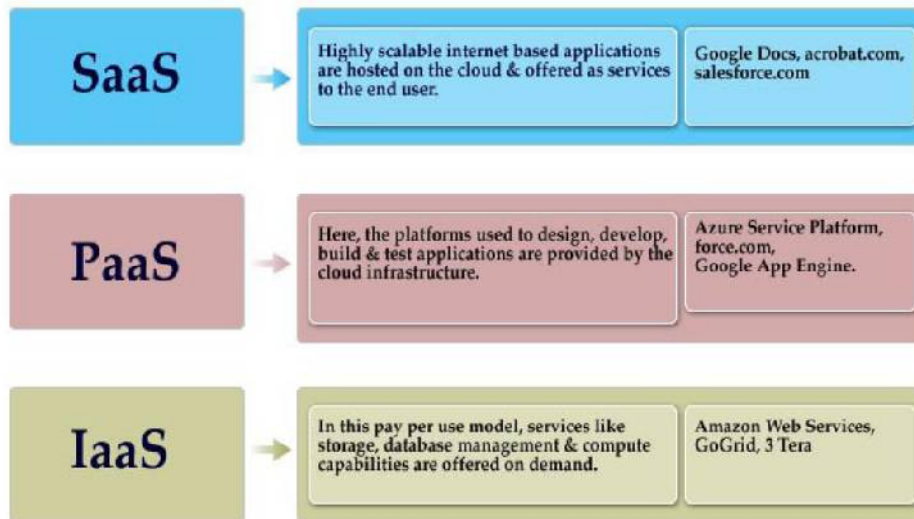


Fig. 2 Cloud Service Models

Infrastructure as a Service (IaaS): This model provides intrinsic storage and computing proficiencies to the customers as normalized services. The basic computing resources, storage, networking equipment, etc. are shared to customers to handle workloads. The customer can install software on the cloud infrastructure. 3 Tera, Amazon, etc are the examples of IaaS. [13]

Benefits of Cloud Services

Since cloud services can minimize the expenditure and difficulty of owning and operating computers and networks they are contemporary. As the cloud users avoid spending in IT infrastructure, they get benefits such as low up-front costs, quick return on investment, quick utilization, modification and flexible use. Additionally, cloud providers who are expert in a particular area can provide further services to an enterprise that cannot afford.

The users also get some other benefits such as efficiency, reliability, and scalability. Cloud computing is efficient in the sense that it permits enterprises to loosen resources to focus on modernization and product development. Since cloud allows the customers to access its applications and records from anywhere in the world, it is said to be reliable. They are accessed through internet.

One of the major benefits is that the individual data is securely stored in the cloud effectively. It also grants for improved inspection traces. Also, information in the cloud is stored efficiently without any loss.

Deployment of Cloud Services

Organizations are decided to organize their applications on Public, Private or Hybrid clouds.[7] Cloud coordinators can associate with a basic role in establishing the rigorous cloud path for each organization.

Public Cloud: As the infrastructure costs are spreading among a number of users, public clouds provide better frugality of expanse to customers. They offer each client a charming low-cost, “Pay-as-you-go” model. All customers share the comparable infrastructure collection with limited configuration, security, and accessibility differences.

Private Cloud: Private clouds are built solely for a specific enterprise. The main aim of private cloud is to address information security concerns that is not provided by public cloud and offers great control over private cloud. The private clouds are differed based on the two types: On-premise Private Cloud and Externally Hosted Private Cloud. The cloud hosted within one’s own data center is called on-premise private cloud or internal clouds. Even though this model offers a more systematized process and security, it is confined in extent and scalability aspects. The enterprises

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

that need complete control and customizing of the infrastructure and security use on-premise private cloud. The externally hosted Private Cloud is hosted superficially with a cloud provider. In this, the provider facilitates privileged cloud surroundings with full privacy assurance.

Hybrid Cloud: It is a combination of both private and public cloud models. In order to improve the elasticity of cloud computing, service providers utilize third party in a complete or incomplete manner in a hybrid cloud. The Hybrid cloud environment is capable of providing on-demand, externally provisioned level.

III. DATA WAREHOUSE

To maintain decision-making within an organization, a variety of resources were integrated and aggregated by data warehouses. To support decision making, a data warehouse stores data that is used for organizes tasks. This is called online-analytical processing (OLAP). The data extracted from operational databases are restored in larger time intervals since OLAP tasks are independent on the latest updates. Thus, a data warehouse dissociates the input from operational databases and output to the OLAP system. The data warehouse is organized by a collection of augmented views called data marts. Data warehouse systems require different databases than standard databases. The data warehousing systems are used for various analytics concerning huge data (OLAP) while the use of database systems are to perform small transactions (OLTP or on-line transaction processing), [8].

Data warehouse is a collection of data from multiple data sources which are nonvolatile, integrated, time variant and subject- oriented that supports decision-making process. Therefore the enterprises make tactical decisions based on the stored data. The data warehouses are superior to operational database since they contain integrated data. The queries on data warehouses are friendly and complicated and can access millions of records. Apart from that, the queries on data warehouses carry out a lot of scans, aggregates and joins. Industries such as manufacturing for order shipment and customer support, utilities for power usage analysis, trade for user profiling and inventory management, monetary services for claims analysis, credit card analysis, and scam detection, transportation for fleet (navy) management, healthcare for outcomes analysis, risk analysis, and telecommunications for call analysis and fraud detection [9] are successfully organizing data warehousing technologies. The following characteristics of data are available in a data warehouse.

- Subject oriented: The data are sensibly systematized around major subjects of the enterprise. E.g., storing of data regarding the number of customers, sales, or items produced on everyday operations in an enterprise.
- Integrated: All informations regarding the subjects of the enterprise are combined and analyzed together.
- Time variant: Chronological or historical data are maintained in detail. E.g., data of sales in last 3 years, etc.
- Non-volatile: The data are read only. No updation or changes can be made by users.

A data warehouse represents data from operational systems, but it is physically separated from operational systems and serves for different purposes. Operational systems are used for transaction processing and they have their own databases. A data warehouse is used to support decision-making and has its own database. After the creation of data warehouse, the SQL queries are generated by accessing the data in the warehouse. "Data warehousing" is used to elucidate the creation, use, maintenance, and continuous updation of the data in the warehouse.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

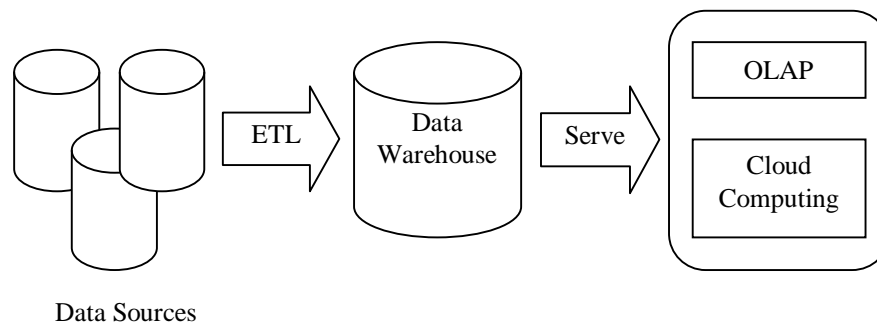


Fig. 3 General Structure of a Data Warehouse

The data in a data warehouse are generally OLTP systems that are approached from the data sources. All data from an enterprise or business can collaborate in the data warehouse component. Before entering a data into the data warehouse component it must be updated. Thus the abnormalities like dissimilarities and replicas were eliminated from the data that runs the data warehouse system as desired. Performing less transformation of data will lead to difficulties in querying the data [10]. After the data pre-processing is completed, the data is loaded into the data warehouse. Finally, the data is organized for usage by OLAP and cloud computing tools.

Data warehousing systems have proficiencies to supervise and manage the data warehouse [8]. The change in real-time workload must be detected during execution. Thus workload management is an important factor for data warehousing in the cloud environment.

IV. DATA WAREHOUSING ON CLOUD ENVIRONMENT

Data warehousing elevates the bar on cloud computing. Cloud computing is the hostile of traditional data warehousing. In cloud computing, data are location free, translucent and functional, but in data warehouse, the centralized and persistent data are stored. Run-time metadata is needed to register and access data sources as a service. In the conflict between computing power and increased data, large amount of data begins to overfeed behind the I/O subsystems with limited bandwidth.

Cloud computing is substantially for storing and accessing of applications from the processor, whereas data warehousing refers to the integration of different databases from an enterprise. Data warehouse is used to store data and generate query regarding the necessary information. Both cloud computing and data warehouse helps to access data, save downloads and revise data. One of the benefits is that the data on cloud can be recycled, reused, reduced and recovered [11]. An enterprise will lose some of its control when it tries to move its data from a data warehouse system to the cloud. This leads to security and trust problems. Sensitive data from the data warehouse can be left out of the analysis reducing security risk.

V. CONCLUSION

Data warehousing over the cloud environment has been studied and from the study it was understood that the data warehouse system over the cloud environment has advantages such as flexibility, scalability, dependability and reduced costs. The capabilities of the data warehousing on cloud computing is parallel, disseminated and high. High Security issues are considered in the judgment of moving a data from data warehouse to a cloud. Hence, with the increasing of data and growing of cloud environment, the future scope is to develop an energy-aware scheduling algorithm to minimize peak power consumption in job execution in real-time cloud environment.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 7, July 2016

REFERENCES

- [1] Veena Goswami, Sudhansu Shekhar Patra, and G. B. Mund, "Optimal Management of Cloud Centers with Different Arrival Modes for Cloud Computing Environment", International Journal of Cloud Applications and Computing", Volume 2 :Issue 3, pages 86-97, 2012.
- [2] Vaibhav C. Gandhi, Jignesh A. Prajapati and Pinesh A. Darji, "Cloud Computing with Data Warehousing", International Journal of Emerging Trends & Technology in Computer Science, Volume 1: Issue 3, Page 72,73, ISSN 2278-6856, September – October 2012.
- [3] William H. Inmon, Book, "Building the Data Warehouse", Page. 1-30.
- [4] K. Kala Bharathi and K. Sandhya Sree, "Recent Developments on Data Warehouse and Data Mining In Cloud Computing", International Journal of Computer Science Engineering and Technology (IJCSET), Vol 5, Issue 2, pp.31-34, Feb 2015.
- [5] Mell, P. and Grance, T., "The NIST Definition of Cloud Computing", Recommendations of the National Institute of Standards and Technology, NIST, 2011.
- [6] Robert Vrbić, "Data Mining and Cloud Computing", Journal of Information Technology and Applications, Vol 2, December 2012.
- [7] S. Chaudhuri, U. Dayal, "An Overview of Data Warehousing and OLAP Technology", In ACM Sigmod record, 1997.
- [8] Inmon, Bill and chuck Kelley, "The twelve rules of data warehouse for a c/s world", Data Management Review, Vol.4, Issue 5, pp.6-16, May 1994.
- [9] E. Rahm, H. H. Do, "Problems and Current Approaches", IEEE Data Engineering Bulletin, 2000.
- [10] Vaibhav C. Gandhi, Jignesh A. Prajapati and Pinesh A. Darji, "Cloud Computing with data warehousing and Analysis Market June 2009", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 1, Issue 3, September – October 2012.