RESEARCH ARTICLE OPEN ACCESS

Distributed Large Dataset Deployment with Improved Load Balancing and Performance

Siddharth Bhandari*
**SATI, Vidisha. M.P. India*

ABSTRACT
Cloud computing is a prototype for permitting universal, appropriate, on-demand network access. Cloud is a method of computing where enormously scalable IT-enabled proficiencies are delivered 'as a service' using Internet tools to multiple outdoor clients. Virtualization is the establishment of a virtual form of something such as computing device or server, an operating system, or network devices and storage device. The different names for cloud data management are DaaS Data as a service, Cloud Storage, and DBaaS Database as a service. Cloud storage permits users to store data, information in documents formats. iCloud, Google drive, Drop box, etc. are most common and widespread cloud storage methods. The main challenges connected with cloud database are fault tolerance, scalability, data consistency, high availability and integrity, confidentiality and many more. Load balancing improves the performance of the data center. We propose an architecture which provides load balancing to the cloud database. We introduced a load balancing server which calculates the load of the data center using our proposed algorithm and distributes the data accordingly to the different data centers. Experimental results showed that it also improve the performance of the cloud system.

Keywords: Cloud datacenter, data distribution, load balancing, system performance, virtualization,

I. INTRODUCTION

Cloud computing is a prototype for permitting universal, appropriate, on-demand network access. Key circulation and key storing are more challenging issue in the cloud database. Cloud database should provision of cloud database and old-style relational databases for extensive satisfactoriness. The other challenges are data reliability and truthfulness, confidentiality and many more. Improving the privacy of data and information stored in cloud computing databases signifies an important involvement to the acceptance of the cloud computing as the fifth usefulness because it addresses most user apprehensions. Cloud services, similar to many other services, are perishable in nature and cannot be stored for future sale. The cloud database as a service is an innovative prototype that can support numerous Internet-based applications.

Virtualization is the establishment of a virtual form of something such as computing device or server, an operating system, or network devices and storage device. Storage Virtualization Storage virtualization is the virtualization in which storage devices are virtualized. It delivers a way for numerous consumers or applications to use storage starved of being anxious with where or how that hard disk storage is physically located. SAN are main example of storage virtualization. The various advantages of storage virtualization are cost of operation, resource optimization, improved performance, and increased availability.

The potential challenges associated with cloud database are scalability, high accessibility and error acceptance, data consistency and integrity, confidentiality and many more. We propose an architecture which provides load balancing to the cloud database. We introduced a load balancing server which calculates the load of the data center using our proposed algorithm and distributes the data accordingly to the different data centers.

The paper is structured as given below. Section 2 provides the contextual, related work and literature review relevant for the context. Section 3 provides the proposed methodology, proposed algorithm and description of proposed methodology. Section 4 represents the implementation of proposed methodology, discussion on simulation Results and performance analysis of simulation results. Section 5 concludes the thesis with a summary of the main findings concluding remarks, limitation discussion and future research directions.

II. LITERATURE SURVEY

Cloud computing can be defined as new computational capabilities that focus on both academia and industry. Cloud computing is the outcome of development and acceptance of current technologies and prototypes.

Public cloud are present everywhere in the world. Virtual systems can be introduced on cloud datacenter on request to critical situation data after inserting data into the virtual machines. Scaling is one of the main factor deciding the performance of the cloud datacenter. If load is more on virtual machines then it takes more time to load data on

datacenter. It will degrade the performance of the cloud system.

This cloud computing model is consists of five important features, three service prototypes, and four deployment models [8]. Enormous progression in digital information and data, better broadband conveniences, altering data storage necessities, and Cloud computing led to the appearance of cloud databases. Fundamentally, there are two foremost participants in the Cloud Computing environments. The Cloud providers also called service producers and Cloud customers also called service consumers or clients are the main pillars in cloud database. Cloud clients can be either application /software service providers. A cloud provider is a vendor or company that provides economically effective cloud services using the hardware and software. Enormous progression in digital information and data, better broadband conveniences, altering data storage necessities, and Cloud computing led to the appearance of cloud databases.

Regardless of the above revealed service models, cloud services can be set up in four ways depending upon the consumers' necessities. Hybrid Cloud, Public Cloud, Community Cloud[9], and Private Cloud. Cloud computing arrangement, prepared accessible only to a particular client and managed either by the association itself or third party service provider is called private cloud. A cloud association is provided to many clients and is accomplished by a third party is called a public cloud. Organization communal by numerous establishments for a shared cause for particular community is called community cloud computing. An arrangement of two or more cloud distribution models is called hybrid cloud. Cloud database is formed above the service provider location. So security should be very much needed in the cloud computing database. The client has to secure data from the others as well as from cloud service provider. Most of the method regarding encryption for cloud-based services are unsuitable to the database model.

Virtualization is the establishment of a virtual form of something such as computing device or server, an operating system, or network devices and storage device. Virtualization has develop a practical requirement these days, and the tendency is ongoing for a decent motive because when applied, it delivers many profits such as the following: Decrease in operational costs and capital, Energy savings for a greener environment, Hard-to-find person's resource savings

Access to storage resources, network, server and on demand, Physical space decrease Server virtualization also referred as hardware virtualization. In hardware virtualization many operating system run on a single hardware. The hypervisor software are used for creating server virtualization. The advantages of server virtualization are Management, partitioning, flexibility, and encapsulation.

Storage Virtualization Storage virtualization is the virtualization in which storage devices are virtualized. It delivers a way for numerous consumers or applications to use storage starved of being anxious with where or how that hard disk storage is physically located. SAN are main example of storage virtualization.

The various advantages of storage virtualization are cost of operation, resource optimization, improved performance, and increased availability.

Network Virtualization Network virtualization is the virtualization in which network resources are virtualized with the help of virtualization services. A VLAN is example of network virtualization.

Service Virtualization Service virtualization in virtualized data bases denotes to the facilities such as firewall servers, load balancing, security etc.

III. PROPOSED WORK

Our architecture consist of 3 layers. Clients, Middleware load balancing servers, Cloud Database servers. Middleware load balancing servers provides load balancing according to tasks and scalability of distributed cloud service. Java platform will be used for the implementation of the algorithm and Oracle 11g server is used as the back-end. Windows operating system is considered well in the security point of view thus windows operating system is used. We will get the mechanism by which we can transfer the data to the right virtual machine. Our system will decrease the setup time (VM creation, software configuration and VM population with data) of virtual clusters for data processing in the cloud.

Steps involved in our proposed work

1. Setup all the data centers for storing large database.
2. Start the load balancing server for proper data distribution and tasks distribution.
3. The client request to the data centers for dataset using application.
4. The load balancing server take the request and according to the load calculations from different servers it will transfer the load to the appropriate server for better performance.

Load balancing algorithm
Initialize all the server allocation status to AVAILABLE in the state list
Initialize hash map with no entries
While(new request are received by the Intermediate server)
Do

Intermediate server queue the requests
Intermediate server removes a request from the beginning of the queue
If(data structure contain any entry of a datacenter server equivalent to the existing demanding user
&& server allocation status == AVAILABLE) then
The database server is modified to the requesting user
Else
Allocate a datacenter server to the requesting user using RR method or procedure
Allocate data center CPU to every process in round robin fashion, according to given time quantum only for one time.
After completion of step 1 process are arranged in increasing order or their remaining burst time in the ready queue.
New priorities are assigned according to the remaining burst time of processes; the process with shortest remaining burst time is assigned with highest priority.
The processes are executed according to the new priorities based on the remaining bursts time.
Modified the record of the user and the database server in the record data structure and the state list
End if
End loop
End
The algorithm steps are as follows.

The list of available servers are stored in AVAILABLE server list. AVAILABLE is an array having list of all initialize datacenter server. Initialize hash map with no entry in it. A client request to the data centers for required information. The load balancing server acts as an intermediate server. It accepts the request from the client, checks the load from different data centers. It also checks availability status of the data center server. If server is available then the request is transferred to the server and task will be completed.

IV. IMPLEMENTATION

Java platform is used for the implementation of the algorithm and Oracle 11g server is used as the back-end. Windows operating system is considered good in the security point of view. The experiments are carried out in lab, which provides us with a set of machines in a controlled environment. Each client machine runs the Java client prototype of our architecture on a Intel PIV machine having a single 3 GHz processor, 2 Giga Byte of Random Access Memory and two 7200 RPM 500 GB SCSI disks. The database server is Oracle 11g running on Intel machine having a PIV 2.4 GHz processor, 4GB of RAM and a 7,200 RPM 500 GB SATA disk. The dataset used in implementation is college database. The database consists of students, faculty and company related records. The middleware server is mainly used for load balancing of datacenter.

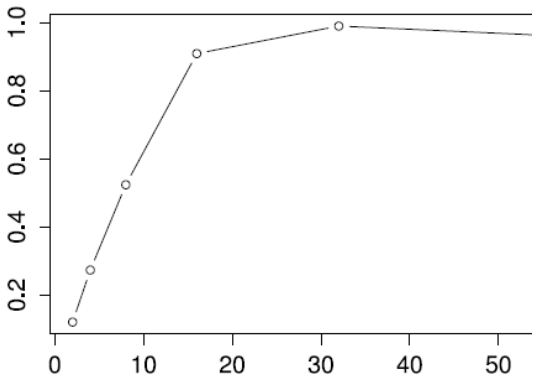

Fig. 1 Transfer rate Vs Intermediate machines

Fig above shows the effect of changing the intermediate machines for large data set. As intermediate machine increases the data transfer rate is also increases. Due to our proposed algorithm the load of the datacenter is distributed to other datacenter for better performance.

V. CONCLUSION AND FUTURE WORK

Our proposed work will transfer the data to the right virtual datacenter with data load balancing. The potential challenges associated with cloud database are scalability, high availability and fault tolerance, data consistency and integrity, confidentiality and many more. We introduced a load balancing server which calculates the load of the data center using our proposed algorithm and distributes the data accordingly to the different data centers. Experimental results showed that it also improve the performance of the cloud system. We proposed an architecture which provides load balancing to the cloud database.

Our future work may include load balancing with security. In future work we are planning to detect the data leak from datacenter.

REFERENCES

[1]. Luis M. Vaquero, Antonio Celorio, Felix Cuadrado, and Ruben Cuevas, Deploying Large-Scale Datasets on-Demand in the Cloud: Treats and Tricks on Data Distribution, IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 3, NO. 2, APRIL/JUNE 2015, pp 132-137
[2]. I. Foster and C. Kesselman, The Grid 2: Blueprint for a New Computing Infrastructure, San Francisco. CA, USA: Morgan Kaufmann Publishers Inc., 2003.

[3]. J. Dean and S. Ghemawat. (2008, Jan.). Map reduce: Simplified data processing on large clusters. Commun. ACM, vol. 51, no. 1, pp. 107–113.
[4]. S. Loughran, J.AlcarazCalero,A. Farrell, J.Kirschnick, and J.Guijarro, "Dynamic cloud deployment of a map reduce architecture," ,IEEEInternetComput.,vol.16,no.6,pp.40–50,Nov.2012.
[5]. Z. Khayyat, K. Awara, A. Alonazi, H. Jamjoom, D. Williams, and P. Kalnis. (2013). Mizan: A system for dynamic load balancing in large-scale graph processing. In Proceedings of the 8th ACM European Conference on Computer Systems, ser. Euro Sys '13, New York, NY, USA: ACM, pp. 169–182.
[6]. L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner. (2008, Dec.). A break in the clouds: Towards a cloud definition. SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50–55.
[7]. K. Andreev and H. R€acke. (2004). Balanced graph partitioning. Proceedings of the Sixteenth Annual ACM Symposium on Parallelism in Algorithms and Architectures, ser. SPAA '04. New York, NY, USA: ACM, pp. 120–124
[8]. G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski. (2010). Pregel: A system for large-scale graph processing. in Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, ser. SIGMOD '10, New York, NY, USA: ACM, pp. 135–146.
[9]. X. Yang and G. de Veciana, "Service capacity of peer to peer networks," in Proc. 23rd Annu. Joint Conf. IEEE Comput. Commun. Soc., vol. 4, Mar. 2004, pp. 2242–2252 vol. 4.
[10]. J. Turnbull. (2007). Pulling strings with Puppet : Configuration Management Made Easy, ser. First Press, Berkeley, CA, USA: A press.