Database as a Service under Clustered Resources in Cloud Computing

S.Thirumurugan[1], K.Sankar[2]

[1]Vivekananda College, Department of Computer Science,
Chennai, Tamilnadu, India,
s.thiru.gan@gmail.com

[2]Vivekananda College, Department of Computer Science,
Chennai, Tamilnadu, India,
rkmvcks@gmail.com

Abstract*: *It is well known fact that there is a race ahead on optimum utilization of the resources in our hands. The cloud computing has emerged out on top of the existing networks to ensure resource utilization to the maximum possible level through share and conquer approach. In line with that objective of effective resource utilization, this paper proposes database as software resource to be made available for the users on pay and get service mode. This study recommends a clustered network as a residing place for the database to provide service on demand. Further, two models for cloud database with the advent of clustering is proposed in this study. This work also suggests the implementation procedure for providing database accessibility to the required users through clustered wireless network on cloud computing.

***Keywords*: Cloud database, Clustering, Virtual Machine, Cloud computing, Services.**

(Article history: Received: 30th June 2019 and accepted 27th November 2019)

I. INTRODUCTION

A. Cloud Computing

Cloud computing is one of the emerging technologies under utility computing. This pay and use model provides metered service to the end user. The term cloud is known as a bunch of servers which are communicated over the network. The term cloud computing is well defined as accessing the cloud over the network in order to fulfill users requirements. These requirements are generally known as services. These services are classified into these types: Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Everything as a Service (Xaas). Fig.1 shows services of the cloud. Cloud computing has many distinguished components such as virtual machines and load balancer.

Cloud Computing system is heavily trusting on the term virtualization, that improves the power efficiency of datacenters and empowers virtual machines to single physical server. It provides all services through the internet dynamically based on the users demand such as software, operating system, storage and processing units.

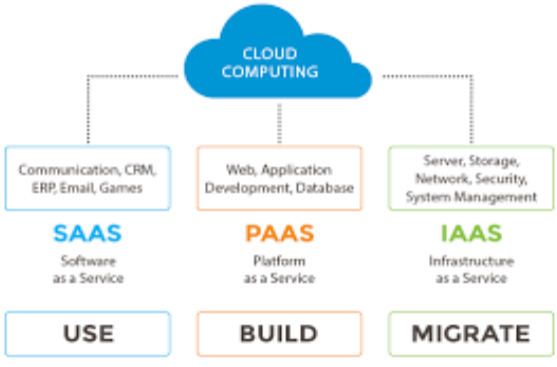

Fig. 1. Service Model

B. Clustering Mechanism

The clustered network shows maturity when it's service has been outstanding. The networks can be classified into clustered and nonclustered networks. The nonclustered network will need to face some issue when there is a linear growth happens over the network size. This growth has been measured by the number of nodes present in the network. Due to this continuous growth the resources consumption will also proportionately increases. This could device a major blockage in communication process. Thus, we may happen to face data loss since the resources are inadequate to do the indented service of the network. This nonclustered approach will burden the nodes of the network to hold and forward the oncoming packet. In order to get out of this disaster scenario the network with clustering approach has been introduced to boost up the functionality and reducing the overhead.

The cluster head should be known by the nodes which fall down to the same cluster. The cluster head takes the lead in sending the data towards destination. This will reduce the work burden of individual nodes. This clustering approach will speed up the communication process, reduces the bandwidth consumption and also makes the networks to carry out its functionality in an efficient manner. Fundamentally Clustering methods are separated into two broad categories namely Hard Clustering and Soft Clustering. In Hard Clustering, the element exists in only one cluster. It is also known as exclusive clustering. In Soft Clustering each element may exists in more than one group. This element is called as intersecting element since it exists as common element. This method of forming clusters is known as Overlapping Cluster technique. Fig.2 shows types of clustering.

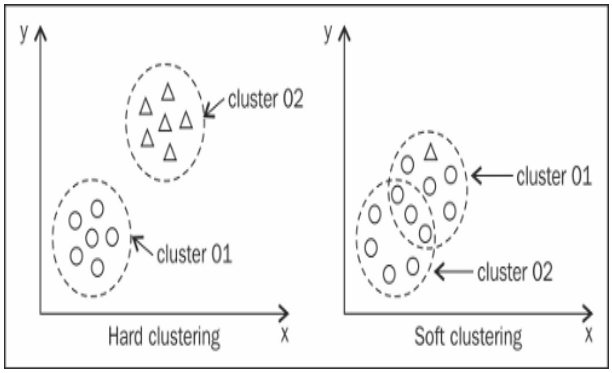

Fig. 2. Cluster Types

C. Clustering Mechanism in Cloud Computing

Cloud computing with the advent of clustering yields efficient results. This paper considers cloud database as a service under infrastructure based service to the users. This database takes clustered network as a platform to provide the service on demand. The forth coming part of this paper covers clustering mechanism on top of the cloud computing environment. Fig.3 shows the distributed database in cloud data centers. The database is distributed across the servers belong to the data centers.

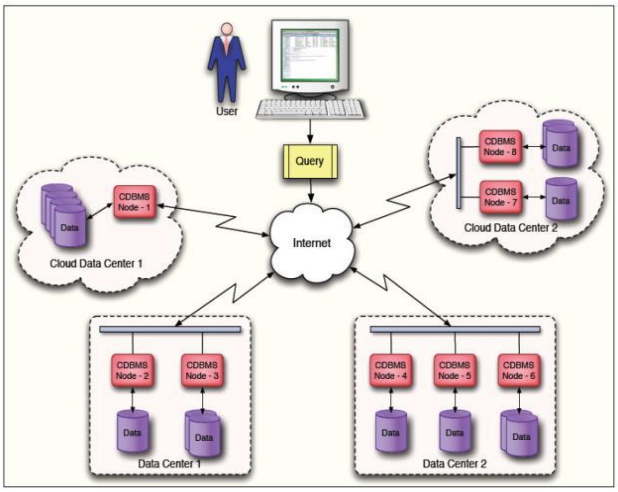

Fig. 3. Cloud with database service

The data distribution is based on the similarity nature of the data. The servers within the data center are clustered towards maintaining the database in efficient way.

II. Related Works

A. Cloud Computing

Among the cloud services, storage as a service holds key role in the IT era. Former to the concept of cloud computing important industrial data used to be deposited internally on the storage [1]. Right from music files to pictures and also sensitive documents, the cloud unnoticeably maintains all the files and folders and eliminates the need for ceaseless and expensive search for additional storage space. When there is enormous data, storage cloud eases by removing obsolete files to create space for the current data. As a result of this many enterprises have put themselves in the cloud environment for the storage service. These enterprises pay for the resources that they use in the cloud. Cloud storage is appropriate and economical. It works by storing the files on a server somewhere in the internet rather than on the local hard drive. This computing permits backing up, sync, and getting data from geographically differentiated locations as long as users have internet facility.

B. Clustering mechanism in Cloud Computing

The role of clustering in cloud computing has become an indispensable. Since the cloud should ensures reliability by guaranteeing the service availability to the maximum extent possible. In the non-clustered environment when one system goes down there should be an alternative to continue with the service to the end user. This may take a while to identify the system to be put in place of faulty system. But, In clustered environment, the resources are grouped to provide support when they are in need. This ensures availability of the data continuously irrespective of the failure in the resources. The authors Bhupendra panchal and R.K Kapoor [2] came with the clustering concept in cloud computing to improve the service availability percentage. This is achieved by replication of data on many servers exist within the cluster of the data center. This service could be further extended by grouping the data centers. This approach helps where one datacenter is not functioning then the service could be switched to other data center.

k-median clustering approach [3] is used to form the clusters in cloud to support the data management. In this approach, an average point of resources are identified and the resources closer to that average point will be put into form a cluster.

k-means [4] is One of the well-known clustering algorithms to support large datasets. In this approach, randomly points are selected then the objects which are closure will be put into create a cluster. The randomly chosen point is changed until the elected point covers maximum amount of objects into the cluster.

Kriti Srivastava [5] recommended the functionality of agglomerative hierarchical clustering algorithm to facilitate the benefits such as scalability, elasticity and handling large datasets.

A.Mahendiran et al [6] implemented k-means clustering approach in Google Cloud using Google App Engine with cloud SQL. This helps to analyze large dataset and also make the storage cost very low. The results of this work show that clustering performs well in the cloud, since cloud is holding the large data sets and they are regularly accessed as per user requirements.

Michael Shindler et al [7] have proposed fast and accurate k-means algorithm as a solution where the data is too big in size and must be accessed sequentially. The well-known BigCross dataset and census 1990 dataset were used for implementing this algorithm.

In parallel computing the role of clustering is very essential. It makes the integrated single system image to carry out the task. Virtual machines are used to devise the clusters to support fault tolerance of the computing [8]. The

cluster creation could be carried at levels and based on various measures. There are clusters which exist due to similarity nature of the data which is stored in the database. These clusters could be optimized with the help of evaluation algorithms [9].

C. Cloud database

Authors Harrison and Babak discusses various cloud databases [10] possible on cloud. These databases are distributed and run on several servers which are geographically distributed across the world. There are various challenges such consistency, persistence and reliability should be maintained in cloud database.

In Cloud computing, database-as-a-Service relieves the burden of user from administrating and monitoring the database. At the user end, there is no need of database administrator since the cloud service provider takes the responsibility of the managing the database [11]. This database supports scalability and security when it is enacted through proper Service Level Agreement (SLA) between the beneficiary and the Cloud Service Provider (CSP).

Real DBaaS will help the user to get what they want rather than worrying about how they are maintained under the cloud environment. This also ensures seamless scalability, maintenance, updation, upgradation, back-up and fault tolerance. This assurance will not hinder the activities of the developer or end user who gets the service by pay mode. In order to deliver a complete DBaaS solution across large number of customers, the cloud service providers need to bring in a high-degree of automation such as automation of elastic scale-out based on certain business rules. For instance, the quality of service in accordance with service level agreement may insist on limited users to access the database. If the users are increasing, then instance of database has to be created automatically through multi-tenancy service of cloud computing. The cloud service provider should have the ability to automate the creation and configuration of database instances [12]. To automate the database service, the service providers have to make use of Application Programming Interface (API). The DBaaS solution must provide the necessary API for the service provider to automatize the system [13]. The core purpose of DBaas is taking burden out of customer side and providing the instances in accordance with the demand. These instances could be available for service with the help of virtual machines. Virtualization allows cloud to provide the various services to customers while automating much of the customary pain of buying, installing, configuring and managing these resources. Now, database virtualization is doing the same thing for the cloud database and it is being provided as Database as a Service (DBaaS).

III. CLOUD DATABASE AS SERVICE

A. Clustering Mechanism on Servers

Fig.4 shows the model proposed for the clustered cloud database as a service. In this approach, the database is distributed among the servers and belongs to the datacenter. The clustering mechanism with in the data center elects the clusterhead to maintain the distribution details of the database. This would ease the service on client request. The clusterhead contains the distribution details of the database for the user applications. The User needs to pay and utilize the database which resides at the server nodes. The clusterhead is responsible for serving the members belong to the cluster where it exists. The databases are grouped to form cloud logically for providing the service on request. The member nodes just need to run the applications and can get the storage service from the cloud.

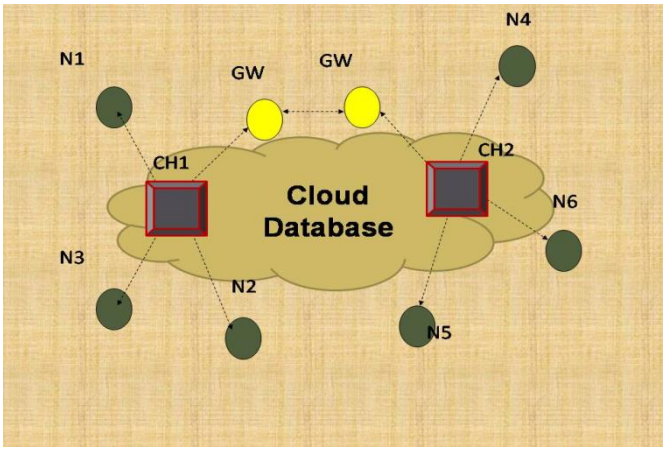

Fig.4. Model of Cloud database with clustering of Servers

Cluster Formation and CH Election Algorithm

CH_1 & CH_2	:	Server nodes to Store the database distribution details among the Servers.
N_1……. N_6	:	Server nodes in data center.
T	:	Cluster Size based on number of nodes.
C_1 & C_2	:	Cluster contains server nodes.
DC	:	Data center contains N nodes
NC	:	Number of clusters.

1. T = 3, j = 1;
2. For i = 1 to NC
3. For m = 1 to T
   $C_i = N_j$; { Add server node to C_i }
   j = j + 1;
4. End for
5. End for
6. For i = 1 to NC
7. CH_i = Node closure to All members of C_i
8. End for

The above algorithm shows the cluster formation and clusterhead election. Threshold (T) decides the size of cluster in terms of the number of server nodes. The cluster C_i is set contains the server nodes.

$$\begin{cases} C1 \in Ni & i = 1,2,3 \\ C2 \in Ni & i = 4,5,6 \end{cases}$$

$$\begin{cases} CH1 \in C1 & i = 1,2,3 \;\; CH1 \; is \; closure \; for \; \forall \; C1 \\ CH2 \in C2 & i = 4,5,6 \; CH2 \; is \; closure \; for \; \forall \; C2 \end{cases}$$

The clusters are formed and clusterhead is chosen based on their close proximity. The data distribution details are maintained in clusterhead.

B. Clustering Mechanism on Virtual Machines

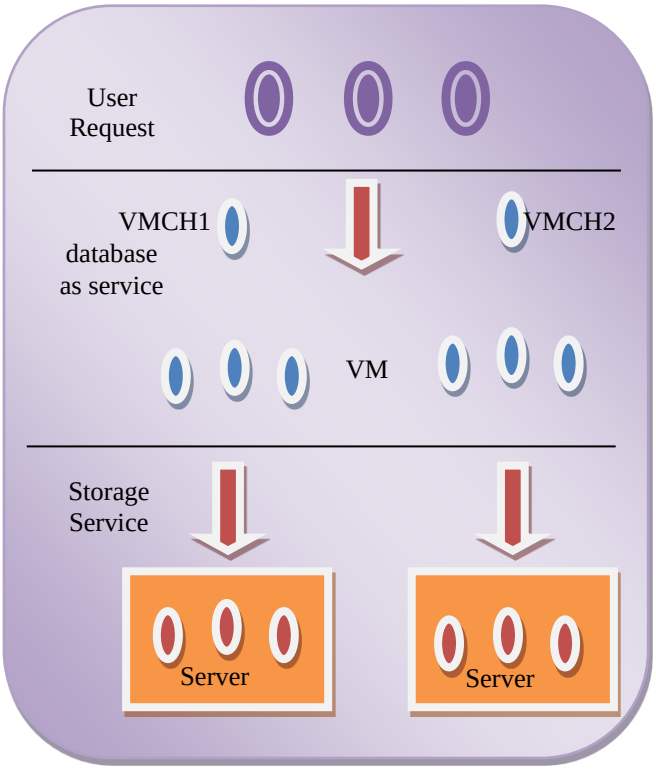

Fig.5 Model of database service with clustering of VMs

Fig.5 shows proposed model of database service with advent of clustering in cloud computing. The virtual machine is called as VMN(virtual machine member) and clusterhead is known VMCH(virtual machine clusterhead). The database is distributed among the VMNs pertaining to the server of datacenter. The groups of VMNs are called as cluster where each group will have VMCH as representation VM node to maintain the data distribution details of VMs.

Cluster Formation and VMCH Election Algorithm

$VMCH_1$ & $VMCH_2$: VM nodes to Store the database distribution details among the Servers.
VMN_1……. VMN_6 : VM nodes runs under server.
T : Cluster Size based on number of nodes.
C_1 & C_2 : Cluster contains VM nodes.
NC : Number of clusters.

1. T = 3, j = 1;
2. For i = 1 to NC
3. For m = 1 to T
   C_i = VMN_j ; { Add VMNi node to C_i }
   j = j + 1;
4. End for
5. End for
6. For i = 1 to NC
7. $VMCH_i$ = Node closure to All members of C_i
8. End for

The above algorithm shows the cluster formation and clusterhead election. Threshold (T) decides the size of cluster in terms of the number of server nodes. The cluster C_i is set contains the VM nodes.

$$\begin{cases} C1 \in VMNi & i = 1,2,3 \\ C2 \in VMNi & i = 4,5,6 \end{cases}$$

$$\begin{cases} VMCH1 \in C1 & i = 1,2,3 \ VMCH1 \ is \ closure \ for \ \forall \ C1 \\ VMCH2 \in C2 & i = 4,5,6 \ VMCH2 \ is \ closure \ for \ \forall \ C2 \end{cases}$$

The clusters are formed and clusterhead is chosen based on their close proximity. The data distribution details are maintained in VMCH (clusterhead). The performance of proposed model increases the overall utilization of the resources and reduces the burden of the member nodes in the data centre.

C. Comparison of proposed models

The clustering machanism which includes server node as member of cluster may involve more than one data centers. This will cover huge space when clustering is concerned on cloud computing. The clustering machanism which includes VM as member node will reside within single data center. By means, the coverage is for small area when compared with cluster of server nodes. Futher, The grouping of VMs are easier than grouping of server nodes.

D. Merits and Demerits of clustering in cloud computing

Advantages

- The clustering Mechanism reduces the overhead of the network traffic.
- Cloud on top of the clustered networks further reduces the resources requirements.
- The users can able to access in wireless ambience to get the service from the cloud. This also reduces the installation cost of the network to be setup for the cloud computing.
- As and when required the services can be expanded without or minimum hurdles.
- The network may contain IPv4 or IPv6 configured nodes.
- Under IPv4 network CH node will run DHCPv4 server to configure the nodes.
- Under IPv6 network CH node and members will do self-configuration without any conflict.

- The clustering improves the cloud computing by providing a network with less traffic and consistency in connectivity.

Disadvantages

- It is highly need to provide a secured and well authorized cloud service.
- The service has been confined by the transmission and reception capacity of the transreceivers.

E. Security in cloud computing

In cloud computing system, the behaviour of service provider and subscriber is determined on a predefined Service Level Agreement (SLA) [14][15]. The validity of this system is determined by the correctness of the SLA, and to what extend the subscriber complies with it. Any breach into this agreement is considered as a violation. Therefore, a robust technique to detect this violation is required in order to build a trusted computing environment. However, the complicated nature of service provision chains in cloud computing makes detecting and preventing the violation a significant challenge. Accountability is one of the mechanisms that are used to provide a trustworthy computing environment by improving the data protection at different levels. Accountability is defined as a commitment of the organization for accepting actor of the host for the personal data that are entrusted in the computing environment from the collection time until when it is destroyed. This commitment also involves presenting a proper remedy to any failure. The notion of accountability has been established by the Organization for Economic Cooperation and Development (OECD). The main components of the accountability are: Responsibility, Transparency, Assurance, and Remediation. She also discussed that the accountability should be retrospective and prospective by extending the security rules to prevent the fault from happening and takes an action if happen.

IV. CONCLUSION

The advent of clustered networks can be further improved while this has been served as a place for the cloud computing. This service on demand can be expanded as the number of cluster increases. This computing eliminates the need of storage at the user side and also reduces the overhead involved in management of the data storage. Thus, this approach makes the user free from caring about the data and enables them to put their focus towards other resources maintenance. This method of sharing the global storage at the pay mode gives an understanding to the end user that as long as the data needs to be retained which could be possible by payment. This also creates a room for integrating multiple services to share the database. In this study, two models are proposed to achieve database under cloud as service.

REFERENCES

[1] Naskar Ankita, Mrs. R. Mishra Monika , “using cloud computing to provide data mining services”, International journal of engineering and computer science, Vol. 2, No. 3, pp. 545-550, Mar. 2013.

[2] Bhupendra Panchal, R.K Kapoor, “Performance Enhancement of cloud computing with clustering”, International journal of engineering and advanced technology, Vol. 2, No.5, pp. 37-40, Jun. 2013.

[3] Kashish Ara Shakil, Manasaf Alam, “data management in cloud based environment using k median clustering technique”, International journal of computer Applications, 4th International IT Summit Confluence 2013- The Next Generation Information Technology Summit”, pp. 8 -11. 2013.

[4] A.Mahendiran, N.Saravanan, N.Venkata Subramanian and Sairam, “Implementation of K-means Clustering in cloud computing environment” , Research journal of applied sciences, engineering and technology, Vol. 4, No. 10, May. 2012.

[5] Kriti Srivastava, R. Shah, D. Valia, and H. Swaminarayan, “Data Mining Using Hierarchical Agglomerative Clustering Algorithm in Distributed Cloud Computing Environment”, International Journal of Computer Theory and Engineering, Vol. 5, No. 3, pp. 45-49, Jun. 2013.

[6] A. Mahendiran, N. Saravanan, N. Venkata Subramanian and N. Sairam, “Implementation of K-Means Clustering in Cloud Computing Environment” , Research Journal of Applied Sciences, Engineering and Technology, Vol. 4, No. 10, pp. 1391-1394, Jan. 2012..

[7] Michael Shindler, Alex Wong and , and Adam Meyerson “Fast and Accurate k-means For Large Datasets” , NIPS'11 Proceedings of the 24th International Conference on Neural Information Processing Systems, pp. 2375-2383, 2011.

[8] Priti Kumari⇑, Parmeet KaurA, “survey of fault tolerance in cloud computing”, Journal of King Saud University –Computer and Information Sciences, Sept. 2018.

[9] Deepak Ahlawat and Deepali Gupta, “An Enhanced Mechanism of Big Data Clustering In Cloud”, International Journal of Advanced Studies of Scientific Research, pp.91-96, 2018.

[10] Harrison John Bhatti , Babak Bashari Rad , “Databases in Cloud Computing: A Literature Review”, I.J. Information Technology and Computer Science, Vol.4, 2017.

[11] http://www.scaledb.com/dbaas-database-as-a-service.php

[12] http://en.wikipedia.org/wiki/Cloud_database

[13] http://en.wikipedia.org/wiki/Cloud Computing

[14] RONG, C., NGUYEN, S. T. & JAATUN, M. G., “Beyond lightning: A survey on security challenges in cloud computing”. Computers & Electrical Engineering, 39, 47-54, 2013.

[15] Muhammad Imran Tariq, “Agent Based Information Security Framework for Hybrid Cloud”, KSII Transactions on Internet and Information Systems, 13(1),406-434 , 2019.

AUTHOR PROFILE

Dr. S.Thirmurugan
Vivekananda College, Assistant Professor, Department of Computer Science. He has 14 publications in his credit and holding around 16 years of experience in teaching and industry. His focus includes wireless network, computing and sensor networks.

Mr.K.Sankar
Vivekananda College, Assistant Professor, Department of Computer Science. He has 2 publications in his credit and holding around 20 years of experience in teaching. His focus includes advanced operating system, data structure and algorithms.