IMPROVING THE DATA TRANSMISSION SPEED IN CLOUD MIGRATION BY USING MAPREDUCE FOR BIGDATA

P.Naresh[1], P.Rajyalakshmi[2], Dr.Vempati Krishna[3], D.*Saidulu*[4]

[1]*Asst.Professor, Department of IT, Guru Nanak Institutions Technical Campus (A), Hyderabad*
[2]*Assistant.Professor, Dept of CSE, CMR Engineering College, Hyderabad*
[3] *Professor, Dept of CSE, TKR College of Engineering & Technology, Hyderabad*
[4]*Assoc.Professor, Department of IT, Guru Nanak Institutions Technical Campus (A), Hyderabad*

Abstract-Cloud acts as a data storage and also used for data transfer from one cloud to other. Here data exchange takes place among cloud centers of organizations. At each cloud center huge amount of data was stored, which interns hard to store and retrieve information from it. While migrating the data there are some issues like low data transfer rate, end to end latency issues and data storage issues will occur. As data was distributed among so many cloud centers from single source, will reduces the speed of migration. In distributed cloud computing it is very difficult to transfer the data fast and securely. This paper explores MapReduce within the distributed cloud architecture where MapReduce assists at each cloud. It strengthens the data migration process with the help of HDFS. Compared to existing cloud migration approach the proposed approach gives accurate results interns of speed, time and efficiency.

Keywords: cloud migration, MapReduce, HDFS

I. INTRODUCTION

Data migration refers to the process in which data transferring can be done one storage system of cloud center to other. There are many reasons leads to data migration [1] process from one cloud to another cloud to exchange information. It plays an essential role in overall processing of migrating on premises IT services to cloud computing environment. While migrating data through cloud we need to consider some important factors that will strengthen the migration process and it should be a secure, efficient, reliable and cost effective [16]. Today all organizations generate huge amount of data that will be transferred from one cloud center to another cloud center and it could store at the cloud server [2] by which end users can extract it for their needs.

And it is very essential to extract data from cloud which makes an immense pressure on business experts to maximize the value of what they are sending and receiving. We can choose several options for data transferring from local data center to the cloud. The cloud can be classified as public cloud, private cloud and hybrid [10] cloud depends on the usage and purpose.

A) Public cloud – in this the resources and data made available to all users as public at cloud server. In public clouds the data shared among organization cloud centers or among user and organization will be accessed at free of cost without any authentication.

B) Private cloud - These are the clouds which are owned by private organizations for their personal use to share secrete or working information. And also used for internal data storage private purpose and it is not available for public use [8]. The owned organization had whole rights on the data and the cloud center. In these types of clouds the sender uses different encryption techniques [12] to secure and the receiver had a key to decrypt it.

*C) Hybrid clouds-*These are the combination of both public and private clouds where the public cloud data can be used by private cloud of an organization if needed. This will need for some business organizations and in IT sector to share the basic information to others and also it hides secrete information which will maintained by privacy concerns.

***Infrastructure as a service*:** It provides direct access to the computing resources to perform some activity. These resources are accessed for exercise activities. Amazon EC2 uses IaaS to provide resources to end users.

Platform as a service: this service provides access to users to create, test and deploy applications. Google app uses Paas service model to develop web applications by the end users.
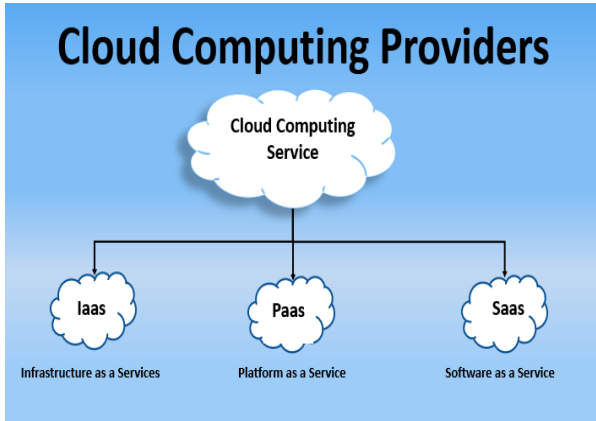
Fig 1: Cloud services

Software as a service*: majority of the software companies use this service to afford its applications to users and clients. The end users are going to use some applications which are located at cloud centers remotely without purchasing it. The companies grant some privileges to users.

II.RELATED WORK

There are three primary factors you should consider while doing the data migration [3]. First one is Type of workload, second one is size of data and third one is speed, by which we can estimate efficient migration. The best option for your precise data migration project depends upon how much data you need to move, how swiftly the migration must be accomplished [4], the types of workloads with that, and your security requirements.

Benefits of Data Migration
Some benefits reaped by enterprises from database migration solutions include:
- The process ensures comprehensive data integrity
- Reduces media and storage costs with significant improvements in ROI
- Minimizes disruption to daily business operations with minimal manual efforts
- Upgrades underlying applications and services while boosting efficiency and effectiveness
- Helps in scaling of resources to meet growing needs of business datasets

Some of data migration tools available are Veeam, Zetro, Rclone and Cyberduck .

At the same time we have to look at distributed cloud computing [5] approaches by which we can access shared and distributed resources over the distributed cloud

environment. The term distributed database refers to the data which is distributed among so many clients are users. That shared data can be utilized by all users dynamically and some organizations given privilege to users against adding data and manipulating it over the cloud.
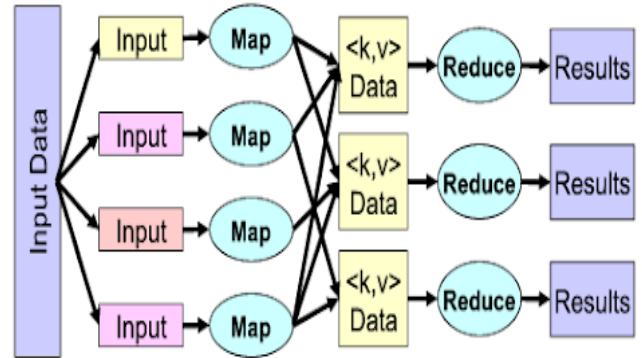

Fig 2: How MapReduce works

While doing these things on the cloud we may face some issues like slow data transfer rate, delays in handover and takes more time to migrate. So to overcome the discussed issues we may use MapReduce [12] of HDFS architecture for distributed data [15].

MapReduce – it is a model which used for parallel processing [6] and in distributed computing [13]. It consists two parts named map and reduce, map performs sorting and filtering functions where as reduce performs summary functions. It works on various tasks parallel and manages data transfer among various systems in the cloud architecture. This guaranties the some features of data consistency, redundancy and reliability. This system is fault tolerance and efficient for data transfer. The MapReduce [6][15] frame work operates on <key, value> pairs.

III.PROPOSED METHODOLOGY

To achieve the efficient data migration, the proposed approach maintains a well defined architecture with two clouds and a distributed cloud center [9]. The distributed cloud center stores all information of all formats like audio, image and text data which was distributed among cloud A and cloud B. In normal distributed cloud computing the needed information was migrated directly by using any suitable migration model.

While the information is migrating we can apply some security algorithms to secure the data. This can stops data threats from intruders. As going on the data sizes increases in the channel, the channel gets traffic and leads to decrease the fast in distributed computing [8]. Some files may contain multiple copies of same data and this will leads to wasting time, bandwidth and unwanted traffic. With this the migration process becomes downward. So to

overcome these obstacles the proposed architecture contains mapreduce [6][12] at each side. When we want to migrate data from cloud A, then it gets the needy data from cloud center and applies map-reduce functions. Here map process the data first to filter and sort it. At this stage maximum of unwanted and redundant data will remove by mapping functions. After the first stage the processed data enters to the reduce stage, where summary functions applied to summarize it. Next we can apply security algorithm to decrypt the data and then we are able to migrate data along with security key. At receiver cloud B receives it and decrypt with key. The same process works when cloud B wants to send data to cloud A.

IV.RESULTS-DISCUSSIONS

Initial experiments were made by considering small data for existing and proposed approaches over time taken to accomplish a particular process. The following explains the time vs data size.

Further coming to large data an analysis was done between earlier and proposed one by considering parameters time and size of data being migrated. At the same time the efficiency of the proposed system was observed and tabulated. Here migration also depends on the types of data (image data, audio/video, text data) being transferred.

Table 1: Migration time for different data formats

Data size(Mb)	Image Data Time(sec)		Audio MP3 Data Time(sec)		Text Data Time(sec)	
	Normal Sharing	Proposed Sharing	Normal Sharing	Proposed Sharing	Normal Sharing	Proposed Sharing
5	2.18	1.86	1.95	1.74	1.73	1.41
10	3.99	2.45	2.43	2.01	2.14	1.96
15	7.93	6.73	6.88	5.86	5.35	4.79
20	8.42	8.01	7.96	7.25	7.01	6.86

V.CONCLUSION

In this paper we discussed various migration methods for distributed cloud. To strengthen the migration process by means of speed, accuracy and efficiency a distributed and parallel MapReduce architecture was explored. The proposed MapReduce parallel processing frame work reviewed by considering different kinds of data with varying sizes. A comparison was made between data

sizes and time taken to migrate over the cloud. An experimental analysis also done for normal cloud computing and the distributed one with proposed approach. The MapReduce frame work improves the performance of distributed sharing. Compared to existing cloud migration approach the proposed approach gives accurate results interns of speed, time and efficiency. Further investigations will be done to implement an incremental migration process, to achieve efficiency.

References

[1]. Jamshidi, P., Ahmad, A., & Pahl, C. 2014, Cloud Migration Research: A Systematic Review, Transactions on Cloud Computing, doi:10.1109/TCC.2013.10.

[2]. Gholami, M. F., Daneshgar, F., Low, G., & Beydoun, G, (2016, October). Cloud migration process - A survey, evaluation framework, and open. The Journal of Systems & Software. 31-69. doi:10.1016/j.jss.2016.06.068.

[3]. Khan, N., & Al-Yasiri, A. 2016. Framework for cloud computing adoption: A road map for SMEs to cloud migration. 5. doi:10.5121/ijccsa.2015.5601.

[4]. Bazi, H. R., Hassanzadeh, A., & Moeini, A. 2017. A comprehensive framework for cloud computing migration using Meta-synthesis approach. 87-105, doi:10.1016/j.jss.2017.02.04.

[5]. T. Kakantousis, I. Boutsis, V. Kalogeraki, D. Gunopulos, G. Gasparis, and A. Dou, "Misco: A system for data analysis applications on networks of smartphones using mapreduce", in Proc. Mobile Data Management, 2012.

[6]. F. Marozzo, D. Talia, and P. Trunfio, "P2p-mapreduce: Parallel data processing in dynamic cloud environments," J , Comput. Syst. Sci., 2012.

[7]. K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system", in Proc. of MSST, 2010.

[8]. P Naresh, et al. "Implementation of Map Reduce Based Clustering for Large Database in Cloud", Journal for Innovative Development in Pharmaceutical and Technical Science, vol 1 issue 1,2018.

[9]. P. Basanta-Val and M. Garcia-Valls, "A distributed real-time java-centric architecture for industrial systems", IEEE Transactions on Industrial Informatics, vol. 10, no. 1, pp. 27--34, 2014.

[10]. T. Laszewski and P. Nauduri, "Migrating applications to the cloud," in Migrating to the cloud: Oracle client/server modernization. Elsevier, 2011, ch. 8, pp. 181–208.

[11]. Cores, M. Rodr´ıguez, P. Gonz´alez, and M. J. Mart´ın, "Reducing the overhead of an mpi application-level migration approach," Parallel Computing, vol. 54, pp. 72–82, 2016.

[12]. J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," Communications of the ACM, vol. 51, no. 1, pp. 107–113, 2008.

[13]. L. F. Bittencourt, A. Goldman, E. R. Madeira, N. L. da Fonseca, and R. Sakellariou, "Scheduling in distributed systems: A cloud computing perspective," Computer Science Review, vol. 30, pp. 31–54, 2018.

[14]. L. Zhang, C. Wu, Z. Li, C. Guo, M. Chen, and F. C. Lau, "Moving big data to the cloud: An online cost-minimizing approach," IEEE Journal on Selected Areas in Communications, vol. 31, no. 12, pp. 2710–2721, 2013.

[15]. P.Naresh,et.al., "Implementation of Map Reduce Based Clustering for Large Database in Cloud", Journal For Innovative Development in Pharmaceutical and Technical Science,vol.1,pp 1-4,2018.

[16]. H. Li, G. Zhu, C. Cui, H. Tang, Y. Dou, and C. He, "Energy-efficient migration and consolidation algorithm of virtual machines in data centers for cloud computing," Computing, Springer, vol. 98, no. 3, pp. 303–317, 2016.