[NC-Rase 18]
DOI: 10.5281/zenodo.1483972

ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

PERFORMANCE ANALYSIS OF QOS USING DATA DISTRIBUTION TECHNIQUES IN CLOUD COMPUTING

Mrs. Snehal A.Narale*[1] & Dr.P.K.Butey[2]

[*1,2]Asst.Prof , Asso.Prof ,Head, Department of Computer Science ,Dharampeth M.P.Deo Memorial Science College Nagpur,India,Head ,Department of Computer Science,Kamla Nehru Mahavidyalaya, Nagpur, India

ABSTRACT

The biggest challenge in cloud data storage is customer's outsourced data. When the storage is outsourced in that case customer is depends only on single service provider and that fashion is not very much hopeful. Privacy and availability of data from cloud data storage is another issue of cloud computing. To overcome this problem to offer better privacy as well as reliability of cloud data can be accomplished by using data distribution techniques. In cloud computing to achieve data privacy the users' data is divided into no of pieces and distributed among the available service provider in such a way that a threshold number of SP s can take part in successful retrieval of the whole data block. In this paper proposed data distribution followed to distribute data in cloud computing which grips an economical distribution of data among the available SPs in the market, to provide customers with data availability as well as reliability. Data fragmentation plays an important role in data distribution.

Keywords: *Cloud computing, data distribution, reliability, availability, cloud analyst, ORT, DCPT.*

I. INTRODUCTION

Cloud computing is an emerging technology in the field of computer science. Cloud computing provides large infrastructure to the users on demand. The primary goal of cloud is to process large amount of data as per the requirement. Workload management is one of the challenge of cloud computing. Cloud always supports distributed system and the major role of the cloud computing is to distribute equal amount of load to each data center.

In cloud computing users can be retrieved Massive amount of data from anywhere at any place with location independency. Cloud offers various services to the users ubiquitously. One of the major service offered in cloud computing is cloud data storage. Customers use the benefit pay as on demand service to keep data in cloud storage for particular period of time and use it if it required geographically. Now a day's information technology shifted to pay per use service model and cloud computing exactly used this model. Cloud offers list of advantages to the users including resource provisioning, large computing capabilities, board heterogeneous network access, resource pooling and many more with measured services. Cloud analyst is GUI tools of cloudsim which is used for the simulation and modeling of data. In cloud analyst diverse scheduling algorithms and various service broker policies are used for testing various parameters like response time, data center processing time, grand total cost & etc.

For the processing of large dataset it required immense infrastructure. Cloud computing provides vast infrastructure to store and process Big data [1]. Formation of Vms cluster which can be provisioned on demand in cloud will help in the process of data.

Cloud service providers are separate market entities. Integrity & privacy of cloud data are the most critical issue of cloud computing and that need to be addressed. To address these issues of cloud computing different data distribution techniques are used. In this case customer divides data among several SP's to get availability of the data with better quality of service and reliability of the data. Two types of threat models are used in cloud computing. If a single failure occurs in sp's if will affect on availability of the data. Hence customers not use single service provider to store cloud data. Data is distributed in different providers by distributing techniques. A data distribution technique also helps in workload management by distributing small chunks into different VMs. This research paper

study QoS parameter like response time and data center processing by using data distribution techniques. It also studies many approaches of data distribution techniques like centralized, semi centralized and peer to peer. The solution is tested in cloud analyst Simulation environment. Customers' stored data at cloud service providers is vulnerable to various threats.

II. BACKGROUND OF PROBLEM

Data confidentiality, integrity and availability of the data are some major issues of cloud computing. In cloud computing customers can be stored their data on cloud storage .In that case only one service provider is there to manage customers requirements & the resources[1,2]. To provide better availability and reliability of the data, in cloud computing the enormous data is distributed amongst several service providers using various data distribution techniques and methods. Distribution of data will really help the customers to get privacy of the data with better quality of service. In this research paper researcher studied various distribution techniques which are used in cloud computing. Load balancing is equal distribution of load amongst various virtual machines. Data distribution techniques used various load balancing policies. Implementation of these techniques in cloud analyst really helps to reduce response time and data center processing time[5].

III. PREVIOUS WORK

Swapnil V.Khedkar , A.D. Gawande[1] discussed an efficient data storage security in cloud service. The partitioning of data enables storing of the data in easy and effective manner. It also gives way for flexible access and there is less cost in data storage. The space and time is also effectively reduced during storage. Dynamic operation is another key concept where, encoding and decoding process secures data, when storing into cloud. Also the remote data integrity checking detects the threats and misbehaving server while storing the data in cloud ensuring data security.

Samip Raut, et.al [2] proposed A big data provisioning service has been presented that incorporate peer-to-peer data distribution techniques to speed up data loading into the VMs used for data processing. As soon as some data chunks start to be ready in the leaves of the tree, the topology evolves to a classic P2P mesh shape.

Suyog Ghodey [3] focused on the DROPS methodology, a cloud storage security scheme that collectively deals with the security and performance in terms of retrieval time. The performance of the DROPS methodology was compared with full-scale replication techniques. The results of the simulations revealed that the simultaneous focus on the security and performance, resulted in increased security level of data accompanied by a slight performance drop.

Samip Raut,et.al[4] discussed a concept of cloud computing and focus on the challenges of the load balancing. It also focus on the time requires to process the big data. The main aim of the study is to achieve qualitative analysis on previous VM load balancing algorithm and then implemented in CloudSim and java language along with the hadoop. Two scheduling algorithms are used to check performance of the system along with their scheduling criterion likes average response time, data centre derive.

G.Priyadharishini, et.al [5] proposed the DROPS methodology, a cloud storage security scheme that collectively deals with the security and performance in terms of retrieval time. The performance of the DROPS methodology was compared with full-scale replication techniques. The simulations resulted in increased security level of data accompanied by a slight performance drop.

B.AmarNadh Reddy,P.Raja Sekhar Reddy[6] proposed a different data fragmentation schemes for multi cloud storage in cloud computing, which seeks to provide each customer with reliability, availability and better cloud data storage decisions.

Aleksandar Hudic,et.al[7] focused on Cloud challenges regarding confidentiality, privacy, control and laws and legislation. With this it offers attractive and cost effective solutions to customers. This paper focuses on secure and confidential data outsourcing to Cloud environments by using fragmentation techniques and applying only minimal encryption to prevent data exposure. To distinguish the real contribution of the Cloud's features, to make a decent

[NC-Rase 18] ISSN 2348 – 8034
DOI: 10.5281/zenodo.1483972 Impact Factor- 5.070

thorough evaluation and to perform simulation of performance (e.g., availability, overhead costs, scalability, etc.).This experiment is run over a longer period of time.

D.Arun Shunmugam[8] studied about how the user has to register in cloud, for each registered user, a unique secret key is generated .The user when wants to upload the file, it gets splits into small chunks and for every upload of file a secret file key is also generated when user wants to download a file, they should enter a secret file key of their file, then splits chunks get merged and can download the file. This provides security at client level as well as in network level.

Dr S. Anitha Reddy [9] discussed about the Cloud computing is a revolutionary model that offers storage, servers and networks as a service. It enables users to outsource their data and trust the service provider that provides them with all the services and storage.This study aims on increasing the availability of services in cloud according to the classification and replication strategies that have been defined.

Rahul Rathore,et.al[10] should addressed on the algorithms which are being used for the selection of the service providers as per their priority. The higher priority will be served first in comparison to others user. To enhance the productivity of the service provider, it is necessary that it's delivered services to their customer on time with maximum throughput. The simulations demonstrate the higher priority user gets the services first.

IV. DATADISTRIBUTI.ON TECHNIQUES

Data distribution is a technique of distributing chunks of data over numerous provisioned VM's. In the process of data distribution one major factor that has to be taken in concern and that is load balancing. Load balancing factor partition whole data into equal provisioned VM's. There are various approaches which are being used for performing data distribution among provisioned VMs [3]. Some of them are as follows:

1) Centralized Approach

At the initial stage all the VM are connected to central repository. As VM are connected to central repository after the process of booting VM's get the required data. In this case central repository used to download the required dataset by VMs one of the major limitation of this approach is the connection will drop if request are run in parallel [3]. Due to this sometime the bottleneck problem will appear.

2) Semi-Centralized Approach

To overcome the problem of connection drop if VM requested in parallel to central server, semi centralized approaches potentially reduce the networking infrastructure stress. Due to this approach it would be possible to share the dataset across different machines in the data centre. By this VMs do not get the same shared at the same time. The limitation of this approach is when the datasets change over the time [4].

3) Hierarchical Approach

In semi centralized approach it is difficult to maintain the data if new data is continuously added. In hierarchical approach, there is build a relay tree where data not gets from the original store by VMs, but data is get from parent node in the hierarchy. In this way all VMs will access the central server to fetch data, and again this fetch data is provide to other VMs and so on. The limitation of this approach is that it cannot provide fault tolerance during the transfer [4].

4) P2P Approach

In this approach each machine acts as server as well as client also. One more thing that has to be mention for this approach is that the data should be more manageable. Hierarchical approach requires more synchronization. Low latency should be provided by the data center for accessing the VMs. No Firewall or NAT issues and no ISP traffic shaping to deliver a P2P delivery approach for big data in the data centre.

V. DOD & R (DISTRIBUTION OF DATA & REPLICATION) METHOD

When any data owner wants to send some file to the cloud server to do this procedure first user must registered & then send some input data to the cloud service provider. To get the better availability, reliability and accessibility of the data, it broken down into chunks. Apply some load balancing policies like RR, Throttled and equally spread

ISSN 2348 – 8034
Impact Factor- 5.070

current execution load policy along with service broker policy on the given partitioned data. Now distributed data is loaded to VM and then it produces some output. The data is partitioned using fragmentation (horizontal & vertical) and replication. Fragmentation used for data reliability, availability of the data on the cloud. Replication is used to make the copy of the data and it is distributed to number of the users. It will help to reduce the workload of the main server. Replication decreases the chance of data loss, increases the performance, availability, & reliability too.

- Input some data.
- Partition data into small chunks/parts.
- Apply load balancing policy and service broker policy to the partitioned data.
- Distribute load to the VM.
- Use Hadoop/Map Reduce.
- Get the output.

VI. EXPERIMENTAL SETUP AND RESULTS

The experiment is evaluated using GUI of cloudsim simulator Cloud Analyst. Cloud analyst is used for simulation and modeling of the data. While using cloudsim the user has to face the problem. Cloud sim done all its work programmatically. To overcome this problem cloud analyst provide a platform to the user which allow the user to do all simulation and modeling of the data with some change in parameters simply and rapidly. It also allows setting location of the users that generating application and also location of data centers. In Cloud Analyst various configuration parameter can be set like number of users, number of VMs, number of processors, network bandwidth, amount of storage and other parameters. Based on parameters tool computes the simulation result and show them in a graphical form. The result consists of response time, processing time, and cost. Fig 1 describes different no. of regions and data center used in cloud analyst. Fig 2 shows main, data center and advance configuration which is used in cloud analyst by setting the parameters.

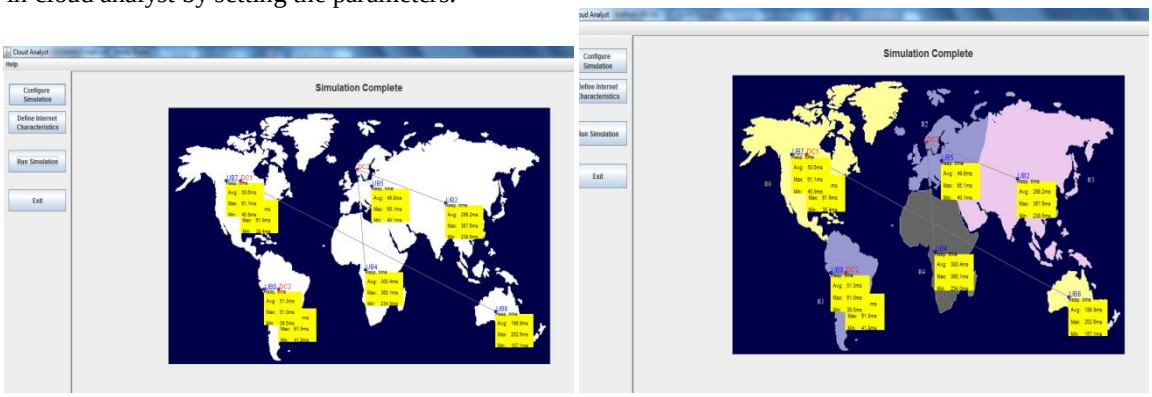
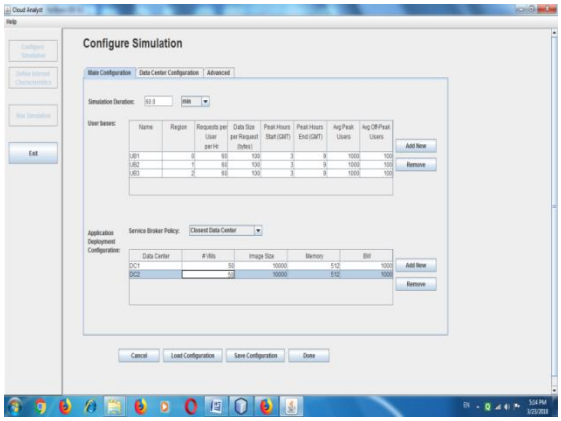
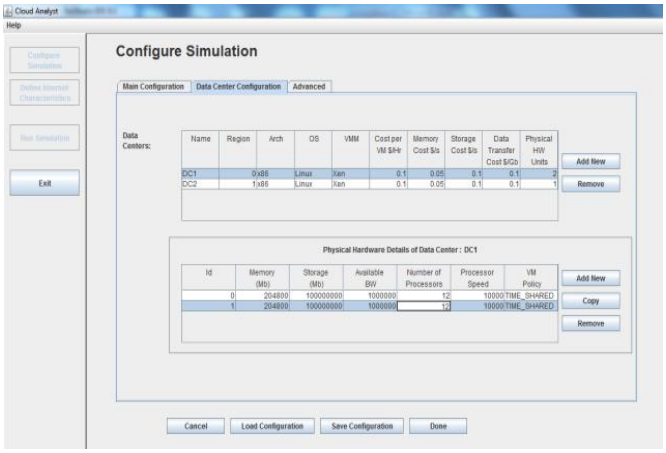


Fig 1: Different No of Regions & Data Center



ISSN 2348 – 8034
Impact Factor- 5.070



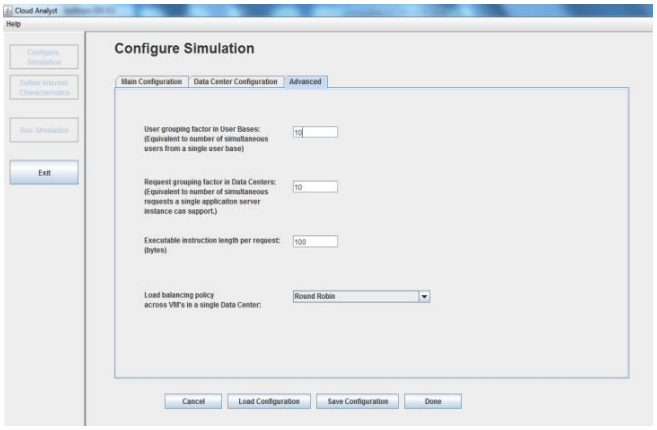


Fig 2 : Configuarion With Main ,Data Center and Advanced

VII. RESULT & ANALYSIS USING SYMMETRIC APPROACH

The experiment is evaluated by using two different approaches. First approach is consider by using three cases for different no. of processors like 4,8 and 12 for data center processing time and overall response time . Table 1 depicts result analysis of these attributes of DCPT and ORT using various use cases with 4 no. of processors with 8 processors & 50,100,200 VM and using 12 processors. Tables1 evaluate result for symmetric approach.

Table 1 : Comparative Analysis of ORT & DCPT using Symmetric Approach of the Processors

No of VM	No of Processor	ORT (ms)	DCPT(ms)
50	4	132.52	0.77
	8	132.37	0.62
	12	132.31	0.55
100	4	132.92	1.17
	8	132.67	0.87
	12	132.51	0.76
200	4	133.7	1.96
	8	133.17	1.4
	12	132.98	1.21

ISSN 2348 – 8034
Impact Factor- 5.070

Fig 3 depicts comparative analysis of ORT (overall response time) & DCPT (data center processing time) using different no. of virtual machine (50,100& 200) with 4,8,12 processors. In this symmetric approach of processor is used. Similarly Fig 4, 5 and 6 also represents comparative analysis of ORT & DCPT independently for 50 VM,100 VM & 200 VM with 4,8&12 processors each respectively.

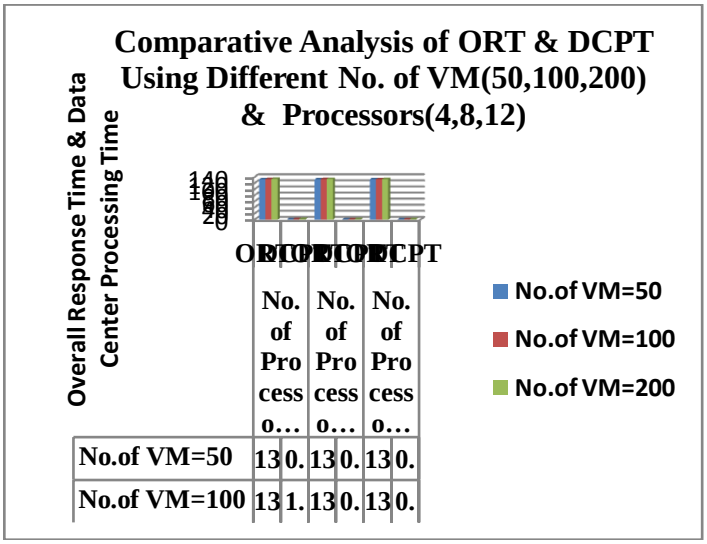


Fig 3: Comparative Analysis of ORT & DCPT (Symmetric Approach)

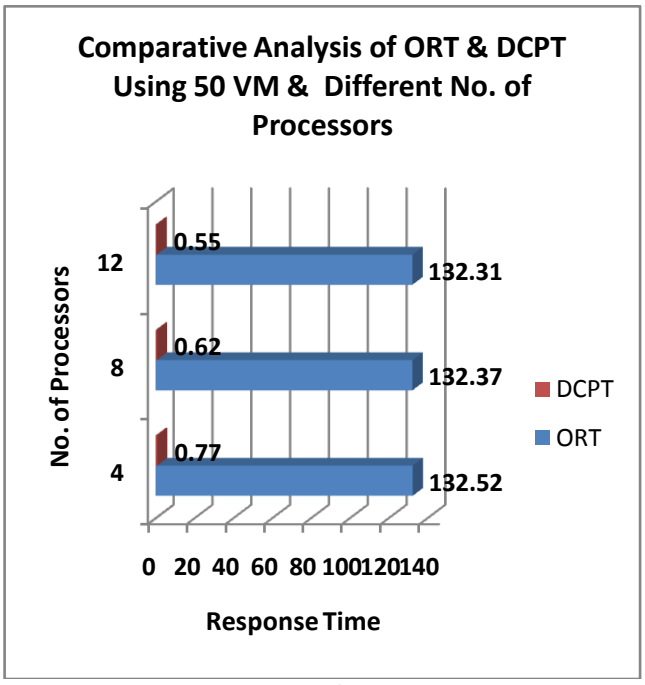


Fig 4 : Comparative Analysis of ORT & DCPT using 50 VM

ISSN 2348 – 8034
Impact Factor- 5.070

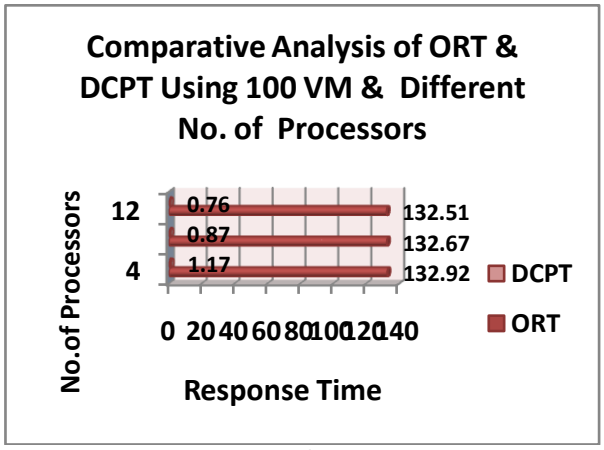


Fig 5: Comparative Analysis of ORT & DCPT using 100 VM

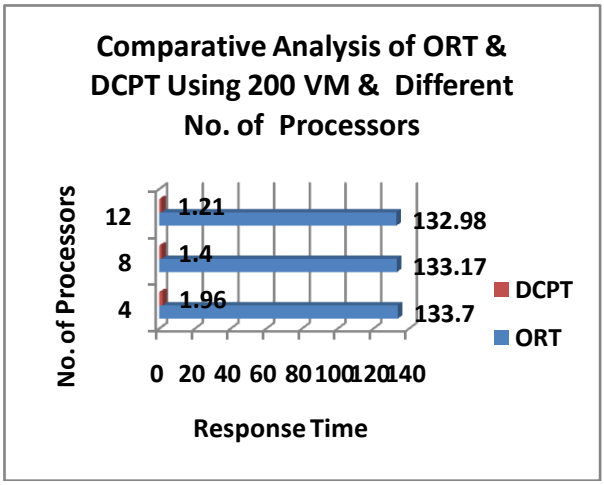


Fig 6: Comparative Analysis of ORT & DCPT using 200 VM

VIII. CONCLUSION

This paper studied various issues of cloud computing and its various challenges. The thing should be heighted in this paper is issues like data integrity, privacy, availability and reliability. All these issues are related to cloud data storage from users perceptive. Data distribution is the proposed solution for these problem .Fragmentation (horizontal, vertical) and distribution of data over the service providers for cloud users are implemented to get the reliability and availability of the data. This paper presents various approaches of Data distribution techniques and implementation of these techniques in cloud analyst (GUI of Cloudsim). Load balancing will help in Distribution of equal no of load to the multiple cloud service providers. In cloud analyst various load balancing policies are used like RR, Throttled and equal distribution of load. Various load balancing algorithms followed by Map reduce aims to achieve qualitative analysis as well as help to avert overloading of the server which degrades the performance and response time will also be improved.

This research paper proposed a different data distribution techniques like centralized, semi centralized, hierarchical and peer to peer approaches used in cloud computing for cloud storage which seeks to provide each customer with reliability, availability and better cloud data storage decisions. [6] In this paper the results are evaluated using two approaches first is Symmetric approach and another one is Asymmetric approach.

[NC-Rase 18] ISSN 2348 – 8034
DOI: 10.5281/zenodo.1483972 Impact Factor- 5.070

IX. FUTURE SCOPE

\Future work is planned to provide higher level of security for outsourced in cloud computing. Experiment is done by using simulation of different scheduling algorithms for executing user request in a cloud environment. Every algorithm is observed and their scheduling criterion likes average response time, data centre derive. Increasing availability of the services in cloud computing to the classification and replication strategies. Future work save the time and resource utilization in updating and downloading the file while distributing the fragments to VM. More security is provided to the network level as well as client level.

REFERENCES

1. *Swapnil V.Khedkar , A.D.Gawande, Data Partitioning Technique to Improve Cloud Data Storage Security, (IJCSIT) International Journal of Computer Science and Information Technologies, ISSN : 0975-9646, Vol. 5 (3) , 2014,pp 3347-3350,*
2. *Samip Raut, 2Kamlesh Jaiswal, 3Vaibhav Kale, 4Akshay Mote, 5Soudamini Pawar, 6Hema Kolla, Survey on Data Distribution Handling Techniques on Cloud International Journal of Recent Advances in Engineering & Technology (IJRAET), ISSN (Online): 2347 - 2812, Volume-4, Issue -7, 2016,pp 54-57.*
3. *Suyog Ghodey1, Niranjan Dandekar2, Sagar Kate3, Vikrant Bhalerao4, Shubham Bhange5, Alka Londhe, Increasing Security and Performance of Cloud Storage Using Data Division and Replication Strategies,9th international conference on recent development in engineering science ,humanities and management,ISBN : 978-93-86171-88-7 , 2017,pp325-332.*
4. *Samip Raut, Kamlesh Jaiswal, Vaibhav Kale, Akshay Mote, Ms. Soudamini Pawar and Mrs. Suvarna Kadam, data distribution handling on cloud for deployment of big data, International Journal on Cloud Computing: Services and Architecture (IJCCSA) Vol. 6, No. 3, June 2016,pp 15-22.DOI: 10.5121/ijccsa.2016.6302 15*
5. *G. Priya dharishini[1], A. Sairamya[2] ,Mrs.W.Mercy[3],Data Fragmentation In Cloud For Optimal Performance And Security, International Journal of Advanced Research in Computer Science Engineering and Information Technology Volume: 4, Issue: 3,Special Issue: 1 ,Mar,2016 ,ISSN_NO: 2321-3337 ISR Journals, pp 615-619.*
6. *B.AmarNadh Reddy, P.Raja Sekhar Reddy, Effective Data Distribution Techniques for Multi-Cloud Storage in Cloud Computing International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 5, September- October 2012, pp.1130-1134 1130.*
7. *Aleksandar Hudic ,Shareeful Islam,et.al, Data Confidentiality using Fragmentation in Cloud Computing, International Journal of Pervasive Computing and Communication, research gate , March 2013,pp1-11.*
8. *D. Arun Shunmugam[1], Dr.S.Singaravelan[2],P. Archana[3]. Dr.K.Ruba Soundar [4] ,P.Archana [5], P.Ponesakki[6], Fragmentation and Replication of Data in Cloud Security,SSRG international journal of computer science and engineering,ICCREST,2017,ISSN : 2348-9-8387,pp 52-60.*
9. *Dr.S. Anitha Reddy,et.al , Data Distribution Methods in Cloud Computing, International Journal of Engineering Research & Technology (IJERT) IJERTIJERT ISSN: 2278-0181, IJERTV3IS090398 www.ijert.orgVol. 3 Issue 9, September- 2014,PP 531-534.*