IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

A Comparative Study of Clustering Techniques Used in Cloud Computing to Minimize the Adverse Environmental Impact

Renuka Raj [1*], Vivekanand[2], Sohit Shukla[3]

[1]Dept. of Information Technology Rajkiya Engineering College, AmbedkarNagar
[2]Dept. of Information Technology Rajkiya Engineering College, AmbedkarNagar
[3]Dept. of Information Technology Rajkiya Engineering College, AmbedkarNagar

Available online at: www.ijcseonline.org

Accepted: 22/May/2018, Published: 31/May/2018

Abstract- Cloud Computing is a computing paradigm where various tasks are assigned to a combination of connections, software and services that can be accessed by the user over a network. The research paper aims to reach a theoretical notion of sustainable development with proposing an incentive for reducing global warming through effective clustering techniques and methods. This paper is a comparative study on k-means clustering, map-reduce technique and resource clustering used in cloud computing. The focus of the paper is to suggest better methodology for handling the events of cloud computing and possibly reducing the similar types of events by clustering them. This approach might lead to the reduction of carbon-dioxide gas (which is a greenhouse gas) by less usage of servers in cloud data centers. With the advent of IT services in cloud computing energy consumption it is necessary for the developing technology to progress towards sustainable development rather thrashing and harnessing energy from every possible means.

Keywords- Cloud Computing, Clustering, Cloud Data centers, Clustering Algorithms, K-Means Clustering, Map-Reduce, Resource Identification and Clustering

I. INTRODUCTION

Cloud computing is the popular technology that has boomed the users' experience of services over Internet. Cloud Computing is the latest technology that delivers computing resources as a service such as infrastructure, storage, application, development platforms, software etc. Birman [1] has a notion for Cloud and network usage in it. There are also some environmental impacts [2] of cloud computing. Arif [3] has a valid description of Cloud Computing and its environmental impact. In this paper, we have tried to accomplish a comparative study on clustering methods used in Cloud computing so that a best suitable algorithm for efficient cloud computing can be possible without harming the environment. We have used the k-means clustering, Map-Reduce concept and resource clustering. K-means clustering is the process of segmenting a group of data points into a minor number of clusters. Map-reduce is a framework for processing parallelizable problems across the large datasets. A resource clustering represents the resource usage distribution for a group of cloud nodes with similar resource usage. The aim of the research in the broader view is to reduce the global warming by minimizing the carbon emission. It is well-known that with the continuous processing on servers the heat is emitted which is allowed to reduce with the usage of air-conditioned server-rooms, this one way or another results into increase of carbon dioxide concentration in the environment. Therefore, to build a clean and green environment with cloud computing many redundant events can be reduced or removed by applying various clustering techniques which may lead to lesser and lesser energy usage and waste.

The organization of remaining paper is as follows: Section II is literature review describing various services and approaches used in cloud computing. Section III is clustering in cloud whether section IV describes K-means clustering. Tech1 and Tech2 is described in section V and VI respectively. Section VII is a theoretical description of comparative study of all three approaches described in upper three sections. At last, conclusion is provided in section VIII.

II. LITERATURE REVIEW

Cloud Computing has its preliminary foundation over providing the hardware, which helps the cloud to run referred as the data centers. Cloud data centers can be

considered as a centralized repository for the storage and management of data. Within cloud data centers there are many cooling and electrical devices that do not directly provide services but are major contributors to the power consumption of a cloud data centers. The advantage of cloud is in terms of economic scale; when thousands of users share same facility the cost per user and server utilization is reasonable. More [4] provided the use of cloud computing for E-Governance services implementation. The services offered by cloud computing are Infrastructure-As-A-Service (IAAS), Platform-As-A-Service (PAAS), Software-As-A-Service (SAAS). Built on top of the data centers layers, IAAS layer provides the process of computing in infrastructure to the end users for example- storage capacity, CPU usage etc. PAAS allows the cloud user to deploy the application created by himself/herself on the cloud providers' infrastructures, in this way the user pays for platform software components which comprises its associated infrastructure cost such as operating systems, database, and middleware. SAAS allows the cloud user to access and use of providers' applications running in a cloud infrastructures, can be termed as a service on demand. Multi-tenancy is the core feature of SAAS. That means SAAS eliminates the need to install set up and run application on the individual computers. There are some more case studies over cloud computing [5], [6], [7] provided by many researchers. Gattulli [8] provided various routing algorithms for cloud computing.

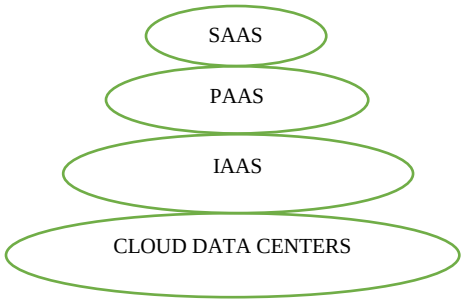

Fig 1: Foundation stones of cloud Computing

III. CLUSTERING

The term 'Clustering' or 'Cluster Analysis' can be understood as the grouping of similar types of items in one set referred as Cluster. Many clustering algorithms [9], [10], [11], [12], [13] have been provided by many researchers. From the perspective of data mining it is one of the fundamental concepts to grasp. Various algorithms are used to find the significant characteristics in the items that can allow them to form a cluster. But the notion of 'cluster' cannot be defined efficiently which is the main reason for proposing various algorithms. In cloud computing the aim for forming the cluster of various events could not be fulfilled by using only one algorithm so we have performed a comparative study for the best suitable algorithm. The K-means clustering, resource identification and clustering and Map-Reduce technique has helped the comparative study to reach its desired goal.

IV. K-MEANS CLUSTERING

K-means clustering divides n-observation into k- clusters in which each observation lies in a cluster with the nearest mean, serving as a blueprint of the cluster. It is often evident that k-mean clustering is well suited for large data sets. As cloud computing deals with massive databases k-means clustering is an efficient approach for handling the data in the form of clusters. K-means clustering allows a suitable structure for unstructured data in cloud computing. An extension of K-means is provided by Singh [13] using encryption.

V. MAP-REDUCE

Map Reduce model is a framework for processing parallelizable events across the large datasets. Kim [14] has power efficient map-reduce description for integrated workload GPU In cloud computing it allows to run particular computation based on client query on the closest nodes (or computers) where data are stored in the form of clusters. Further these mapped tasks at each node are reduced. Hence this validates thesis approach by its name as map reduce. There are various algorithms to facilitate this task such as the canopy method, CPCLUSTER Map-reduce algorithm and many more.

VI. RESOURCE IDENTIFICATION AND CLUSTERING

A cluster of resources determines the resource usage distribution for a group of cloud nodes or cloudlets with similar resource usages. This type of cluster requires two subordinate techniques to capture the temporal detail of cloud nodes. For performing this task we use Histograms so that this statistical information of the uses of resource capacities and aggregation can be achieved for compact representation of a set of similar cloudlets for scalability. In cloud computing this technique will allow the events to group under the optimum usage of resources. Malathy [9] has a valid description for performance improvement in cloud computing using resource clustering.

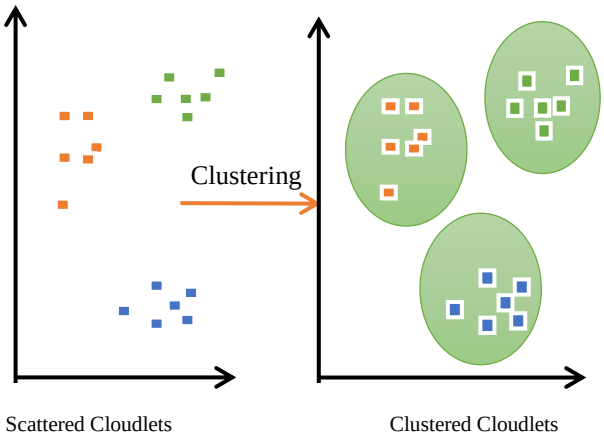

Fig 2: Clustering phenomenon in Resource Clustering

VII. A COMPARATIVE STUDY

- On the basis of theoretical analysis it is evident that the K-means clustering has an ease implementation and has high speed performance in cloud computing technology. The Hadoop map-reduce uses storage method known as distributed file system, which basically deals with large volume of data that is unstructured. Hadoop map reduce takes minutes to process terabytes of data and hours for petabytes of data in cloud computing. In the resource clustering the process of cluster formation for cloud nodes is swift.
- K-means clustering has a measurable and efficient data collection result. Hadoop Map-reduce has an ability to store as well as distribute large datasets across plenty of servers. Resource clustering finds the group of cloudlets which allows them to find a common requirement, this in some way or the other leads to optimum utilization of resources for cloud computing.
- Selection of optimum number of cluster in K-means clustering is difficult. When the data is sent to a specific node over entire cloud network the very same set of data is also forwarded to the other numerous nodes that make up the cloud. Thus, if there is any failure that affects a particular node for Hadoop map-reduce there will always be other copies that can still be accessed whenever the need may arise, in totality Hadoop map-reduce assures the availability of data. In resource clustering, choice of an appropriate value for the count of clusters needs to be viewed seriously if the number of clusters is too small the resource cluster representative will improperly represent their constituent cloudlets due to high heterogeneity on the same side if the cluster is too large then the technique for aggregation will not be suitable due to large overhead in transmission over cloud network.
- Selection of initial centroid in K-mean clustering is random for cloud computing. Parallel processing is the heart of Hadoop Map-reduce technique. The resource clustering algorithm is capable to handle multi-element data in the cloud environment. This algorithm groups the cloudlets together that have similar histogram distribution.
- K-means clustering in the global cloud environment fails to work well due to difficult predictive k-value. Hadoop Map-reduce has an access to various new sources of data and also operates on different types of data. This gives it a power to identify the task effectively and efficiently. The histogram representation of resource clustering gives the information about the percentiles for a particular resource capacity along with preservative norms of all usage data.
- One of the bigger issue with k-means clustering is that different initials partition can result in different final clusters. Hadoop map reduce works with Hadoop distributed file system which allows the users to operate on data stored in the cloud along with parallel processing with multiple processors that take divided tasks in such a way that they run an entire program in less time. Disseminating all the details over the cloud network creates an addition overhead for the network traffic making the system infeasible.

VIII. CONCLUSION

With the analysis and thorough study on K-means clustering, map-reduce concept and resource clustering an optimum utilization of resources, parallel to lesser usage of processers is required which in directly or indirectly leads to greater server usage. The aim of our research has reached its theoretical conclusion that with the lesser usage of servers or optimum usage of servers can be possible with the usage of Hadoop map-reduce concept it can allow. Hadoop map-reduce manages the tasks and events of Cloud Computing network effectively and efficiently. The optimum utilization of servers energy will allow lesser Carbon-dioxide emission that can save the environment from heating up or more precisely ‘global-warming’ could be prevented. With the increasing advent of cloud computing technology we have aimed to reach a sustainable development by contributing lesser harm from the cloud computing world.

REFERENCES

[1.] K. Birman, "Networks and Cloud", CS5412 Spring (Cloud Computing: Birman), 2015.
[2.] R. K. Trivedi, R. Sharma, "Case Study on Environmental Impact of Cloud Computing", IOSR-JCE e-ISSN: 2278-0661, p-ISSN: 2278-8727 Volume 16, Issue 2, Ver. VI, PP 81-86, 2014.
[3.] M. Arif, T. Mahmood, "Cloud Computing and its Environmental Effects", International Journal of Grid Distribution Computing Vol.8, No.1, pp.279-286, 2015.
[4.] A. More, P. Kanungo, "Use of Cloud Computing for Implementation of e-Governance Services", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.3, pp.115-118, 2017.
[5.] Y. G. Patil, P. S. Deshmukh, "A Review: Mobile Cloud Computing: Its Challenges and Security", Vol.06, Issue.01, pp.11-13, 2018.
[6.] M. K. Saggi, A. S. Bhatia, "A Review on Mobile Cloud Computing: Issues, Challenges and Solutions", International Journal of Advanced Research in Computer and Communication Engineering, 2015.
[7.] R. V. Dharmadhikari, S. S. Turambekar, S. C. Dolli, P K Akulwar, "Cloud Computing: Data Storage Protocols and Security Techniques", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.2, pp.113-118, 2018.
[8.] M. Gattulli, M. Tornatore, R. Fiandra, A. Pattavina, "Low-Carbon Routing Algorithms for Cloud Computing Services in IP-over-WDM Networks", IEEE ICC Optical Network and Systems, 2012.
[9.] Malathy, G. R. Somasundaram, K. Duraiswamy, "Performance Improvement in Cloud Computing Using Resource Clustering", Journal of Computer Science 9 (6): 671-677, ISSN: 1549-3636, 2013.
[10.] D.K. Sharma, S.K. Dhurandher, A. Kumar, A. Kumar, A.K. Jha, "Cloud Computing based Routing Protocol for Infrastructure-based Opportunistic Networks", CAITFS Division of Information Technology, 2016.
[11.] S.N. Bushra, A.C. Sekar, "An Efficient Clustering Method for Incremental Cloud Data", IJARCSSE ISSN: 2277128X, 2014.
[12.] E. Sarkar, C.H Sekhar, "Organizing Data in Cloud using Clustering Approach", International Journal of Scientific & Engineering Research, Volume 5, Issue 5, 2014.
[13.] I. Singh, P. Dwivedi, T. Gupta, P. G. Shynu, "Enhanced K-means clustering with encryption on cloud", IOP Conf. Series: Materials Science and Engineering 263, 042057, 14th ICSET, 2017.
[14.] S.Y Kim, J. Bottleson, J. Jin, P. Bindu, S.C. Sakhare, J.S Spisak, "Power Efficient Map Reduce Workload Acceleration using Integrated GPU", IEEE First International Conference on Big Data Computing Service and Applications, 2015.

AUTHORS PROFILE

Renuka Raj- She is pursuing B.Tech. from REC Ambedkarnagar affiliated to AKTU in Information Technology. She has completed her 12th from St. James Senior Secondary School. Her field of interest are Cloud Computing, Software Testing and Data Mining.

Vivekanand- He is pursuing B.Tech. from REC Ambedkarnagar affiliated to AKTU in Information Technology. He has completed his 12th from Sarvoday Inter College Belahi. His field of interest are Cloud Computing, Data Analysis and Data Interpretation.

Sohit Shukla- He is an Assistant Professor at REC Ambedkarnagar affiliated to AKTU. He has completed his M.Tech. from Amity University in Computer Science and B.Tech. from Feroze Gandhi Institute of Engineering & Technology. His field of interest are Cloud Computing, Software Engineering and Data Mining. He has 5 years of teaching experience and 4 years of research experience.

