NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 3, ISSUE2, Feb.-2016

USER STORAGE ALLOCATION IN VIRTUAL MACHINE USING MAP-REDUCED SYSTEM

Mr.Kuldeep Ambasana
Department of Computer Engg ,Pune University / S.G.R.E.F.College, A'nagar, India

Mr.Mahendra Dabhi
Department of Computer Engg ,Pune University / S.G.R.E.F.College, A'nagar,

Ms.Jyoti Girhe
Department of Computer Engg ,Pune University / S.G.R.E.F.College, A'nagar, India

Mr. Amol Chavan
Department of Computer Engg ,Pune University / S.G.R.E.F.College, A'nagar, India

ABSTRACT

The Data allocation paradigm has become very popular and useful tool since its introduction. Many large companies have found intuitive ways to incorporate simultaneous processing techniques into their current needs and operations. A driving force of the growth in the popularity of Map-Reduce is the need for a system to handle and process data. Map-Reduce is a system, which can handle varied quantities of data. With varying amounts of data getting created, there has been a need to increase the capacity of processing of current mechanisms. Due to these issues, many researchers in the past have tried to focus on making allocation of users to various virtual environments using various. More recently, the idea of storage allocation has become popular. The idea is to run virtual machines for storing varied capacity resources. Since these machines are virtual, we can spin up as many identical machines as the project calls in future scope. While this seems like a good fix to the heterogeneous Map-Reduce cluster problem, it leads itself to other issues that we will address. This paper will address a major issue in selecting virtual machines that will improve working for user allocation for storage of user data.

INTRODUCTION

OVERVIEW

Data storage is an essential need of many applications. Hence, analyzing data associated to various users according to their behaviour and identification of appropriate storage in virtual machines has to be optimized to facilitate better output. Map-Reduce allows users to handle a varying amount of data in a particular fashion, without worrying about the intricacies of parallelization. There are a couple of reasons why storage is the focus for this paper. First, it is the mechanism that most of the organizations for managing their user data efficiently. Second, since various behaviours vary across users their efficient identification and storage according to capability of virtual machine is required here.

AIM/MOTIVATION

We will take the approach of looking atthe problem as trying to minimize the execution time of allocation by selecting the best configuration to fit the user. This means we will need to understand what affects the storage behaviour of users. There is also need to allocate data resources dynamically based on application demands and support user storage requirements by optimizing the number of resources in use. By minimizing resource concentration, we can combine data of different types of users nicely and improve the overall utilization of resources. There is need for a set of heuristics that prevent overload in the system effectively. The proposed system focuses on catering these motives and enabling better utilization of resources.

LITERATURE SURVEY

Cloud computing allows business customers to scale up and down their resource usage based on needs. Many of the touted gains in the cloud model come from resource multiplexing through virtualization technology. There is need to allocate data resources dynamically based on application demands and support energy optimization as per green computing theories by optimizing the number of resources in use. By minimizing resource concentration, we can combine different types of workloads nicely and improve the overall utilization of resources. We develop a set of heuristics that prevent overload in the system effectively while saving energy used.

As per self-adaptive scheduling techniquesin*Computer andInformation Technology (CIT), 2010 IEEE 10th InternationalConference*, they discuss a self-adaptive scheduling algorithm which tries to classify the nodes to map-slow and reduce-slow nodes by using historical information. Whenever a system is found to be running slowly, the algorithm launches backup tasks accordingly on a faster node. It does not try to understand the user behaviour compatibility on a storage before a storage is allocated. And after storage is allocated on a particular machine, it has to be monitored for overload condition and re-trained to enhance decision making in future.

Server-storage virtualization: integration and load balancing in data centres has been presented in *Proceedings of the 2008ACM/IEEE conference on Supercomputing*. They present a design of storage virtualization along with the implementation of an end-to-end management layer facilitating storage access at organizational level instead of user level storage distribution. They also propose a novel scheme (similar to our vector model) to address the complexity introduced by the multidimensional nature of the loads on resources.

In *Proceedings of the 2009 conference on Hot topics in cloud computing*, some authors present an approach to maximize the utilization of each resource set which can optimize the performance, and also obtain potential energy savings by identifying the unused resources.

PROBLEM STATEMENT

Now-a-days big IT industries are using cloud computing techniques but, a policy issue remains as how to decide the mapping adaptively so that the resource demands of VMs are met.

This is challenging when the resource needs of VMs are dynamic due to the diverse set of users they serve and vary with time as the data grows and shrinks.

Cloud computing is a network-based environment that focuses on sharing resources. There is need of policies and procedures for governance and risk factors with respect to cloud resource management.

In this project, we will do the design and implementation of an automated resource management system that achieves a good balance between the two goals*:*

- Overload avoidance: To satisfy the resource needs of all VMs running. Otherwise, the storage is overloaded and can lead to degraded performance of its VMs.
- Optimized data storage: The number of VMs used should be minimized as long as they can still satisfy the needs of all users. Later on additional VMs can be identified to save.

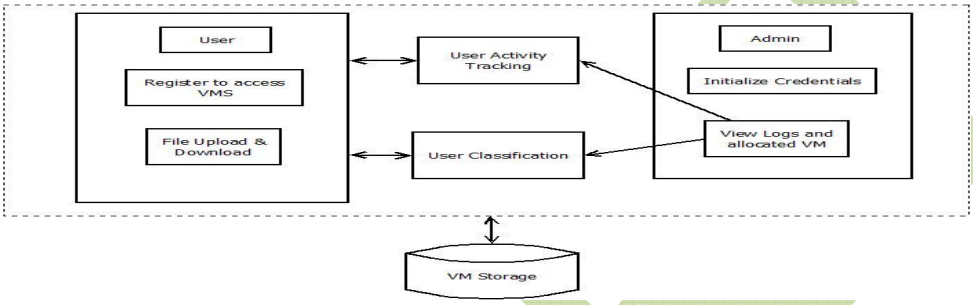

Fig: System Architecture diagram

CONCLUSION

There has been much work done on Map-Reduce; however,our work focuses more on how Map-Reduce behaves and how virtual machine selection can be done in our two experimental testbeds. In this paper, we offered a simple and effective way of selecting a cluster type for running a Hadoop Map-Reduce job in the cloud. Through experimentation, we showed how the system worked, and also showed the effectiveness of the system by running jobs in a real cloud environment. We have also looked at other works, and we hope to achieve similar results. We believe that our solution offers similar results for the classification of jobs, and is much simpler to use in comparison.

REFERENCES

[1]Adam Pasqua Blaisse, Zachary Andrew Wagner, and Jie Wu,*Selection of Virtual Machines Based on Classification of MapReduce Jobs,2015 IEEE 35th International Conference on Distributed Computing Systems Workshops.*

[2] RajkumarBuyya, Chee Shin Yeo, SrikumarVenugopal, James Broberg, and IvonaBrandic. *Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems, 25(6):599 – 616, 2009.*

[3] NiteshMaheshwari, RadheshyamNanduri, and VasudevaVarma. *Dynamic energy efficient data placement and cluster reconfiguration algorithm for mapreduce framework. FutureGeneration Computer Systems, 28(1):119 – 127, 2012.*

[4] Jeffrey Dean and Sanjay Ghemawat. Mapreduce: *simplified data processing on large clusters. Commun. ACM, 51(1):107– 113, 2008.*

[5] Jaideep Dhok, Nitesh Maheshwari, and Vasudeva Varma. *Learning based opportunistic admission control algorithm for mapreduce as a service. In ISEC '10: Proceedings of the3rd India software engineering conference, pages 153–160. ACM, 2010.*