RESEARCH ARTICLE OPEN ACCESS

Hashing Heterogeneity Aware Dynamic Capacity Provisioning in the Cloud

Vinoth Kumar.P[1], Sujitha.M[2]
[1]Assistant Professor, [2]M.Tech Student of CSE
SRM University, Ramapuram
Chennai

--*************************--

Abstract:

Data centers consume more amounts of energy in terms of power supply and cooling required processing the production machines. HHADC is a hopeful approach for reducing energy consumption by dynamically adjusting the number of working machines to match resource demands. However, existing solutions have not fully satisfied the heterogeneity of workload and machine hardware found in data center environment. The data centers often comprise capability of the heterogeneous machines with different configuration capacities and power consumption characteristics. The Memory details of both data center production machines and workloads will lead to sub-optimal power-savings and long delays, due to inconvenience between work load requirements and the resources offered by the production machines. To address this limitation, we create a HHADC, scheme for cloud data centers. First use the K-means clustering algorithm to divide processing load into distinct classes with same characteristics in terms of resource and performance needed. Then we have a technique HHADC technique that dynamically adjusting the number of machines to reduce the total power energy consumption and responding delay. HHADC can reduce energy by 25 percent compared to Harmony solutions.

Index Terms – **Hashing Heterogeneity, Power Supply, Dynamic capacity**.

--*************************--

I INTRODUCTION:

Data centers have recently gained popularity as a cost effective platform for deploying large scale service applications. Many data centers enjoy economies of scale by reducing long-term capital investments over a many number of machines; they also incur more energy costs in terms of power supply and cooling. At instance, it has been reported that energy consumption related costs account for approximately 12 percent of overall data center expenditures. Big companies like Google and Yahoo a 3 percent reduction in power energy cost can translate to over a million dollars in cost savings. Governmental agencies continue to implement to promote energy reduction. Reducing power energy consumption has become an important concern for today's data center connectors. One challenging technique is that Hashing Heterogeneity-Aware dynamic capacity provisioning (HHADC). This technique is to dynamically adjust the number of working machines in a data center in order to decrease power energy consumption while meeting the attain the upper workloads. Schedules the workload in data centers,shown for the level of the machines free space and occur due to power consumption and time delaying due to free space in server is scheduling delay.There is the scheduling delay for task allocating that is a major concern in data center environments for several reasons;

(i)A user might need to immediately increase an application to accommodate a surge in

demand and hence requires the resource request to be achieved as soon as possible.
(ii)Even for lower priority requests, long scheduling delay lead to starvation it hurt the performance of these applications. Even though switching off a large number of production machines can lead to high power energy savings, at the same time, it decrease service capacity and hence leads to high scheduling delay.

II RELATED WORK :

A. Task Allocation

In the task allocation module, firstly the user registers himself to the acquire the login credentials. After getting the login Credentials, the user is directed to the task allocation page in which there are two tasks by which they acts as a word load to separate the task. In one task, the multiple files are inputted as a input and generate a single file, by which it joins multiple files/images as a one file. In the second task, the url is taken as a input and saved as a picture format by which the word load is differentiated from the first task.

B. Task Status

Multi-linear In this module ,the status of the user file are verified by which the allocated file are reviewed / uploaded in the cloud environment .Firstly ,the files generated during the user task allocation are stored in the local server and will be in the waiting status for the storage in cloud . After the allocation done by the admin, the file will be allocated in the cloud .Then according to the users request; the file will be downloaded from the cloud if the file is lost in the local server.

C. Server Allocation

In this module, the admin allocates the heterogeneous data to the cloud through the type of data that the user uploads in the task panel. After uploading the data to the cloud, the admin segregates the type of data according to the content type and it will be allocated according to the free space which has high according meanwhile, other data centers will be in rest condition so that the performance and the cooling system will be managed properly which will extend the life period of the data center.

D. View Tasks

In this module, according to the task allocation, the user log will be saved which consists of the file the user has saved in the task allocation panel. The log will be saved according to the cloud allocation status of the system and randomly generates a name of the file according to the images or url type of the file. In the admin module, the admin views the remaining request which need to be allocated will be shown and the pending request will be cleared accordingly .Thus all these tasks are tabulated and shown according to the view of the user and also to the admin according to their request.

E. File Retrieval

In this module, the user will request for the file to be downloaded from the cloud. Thus the files will be stored in the local server as well as in the cloud. If the file in the local server is corrupted or lost accidently, the user will retrieve the files from the cloud .Thus according to the request of the user, the requested file will be directly downloaded to the user destination and can be accessed significantly.

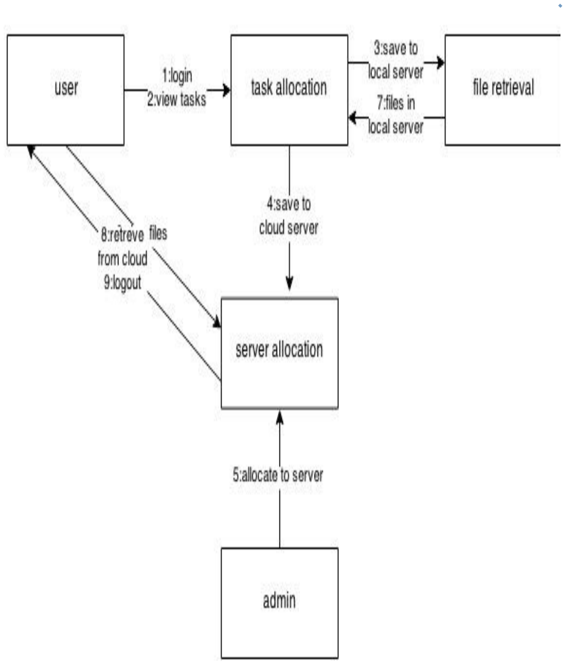

Fig 1 Workload Identification and separation

III SYSTEM MODELS AND PROBLEM STATEMENT :

A. Overview:-

The objectives of this maintenance work are to make sure that the system gets into work all time without any bug. Provision must be for environmental changes which may affect the computer or software system. This is called the maintenance of the system. Nowadays there is the rapid change in the software world. Due to this rapid change, the system should be capable of adapting these changes. In our project the process can be added without affecting other parts of the system.

Maintenance plays a vital role. The system liable to accept any modification after its implementation. This system has been designed to favor all new changes. Doing this will not affect the system's performance or its accuracy.

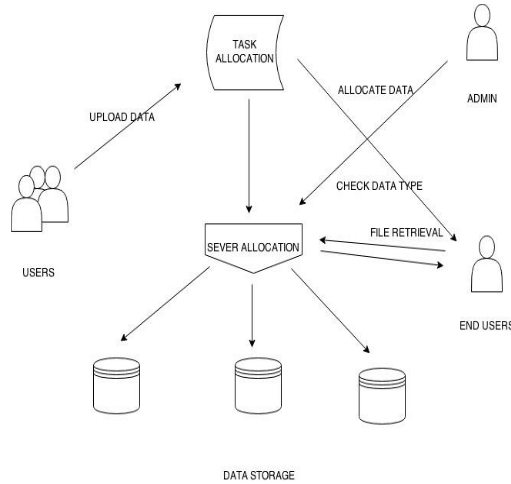

Fig 2 System Architecture

B. Problem Statement

Despite extensive studies of the problem, existing solutions have not fully considered the heterogeneity of both workload and machine hardware found in production environments .It can be easily integrated with existing scheduling algorithms ,variants of first-fit and best-fit algorithms and Open source platforms can adopt this mechanism by changing the scheduling policy to weight round-robin first fit and weight round-robin best fit ,respectively. The main benefit is its simplicity and practicality for deployment in existing systems. The existing work on this topic has not addressed a key challenge, which is the heterogeneity of workloads and physical machines.

i) Mishandling heterogeneity data.

ii) Improper workload maintanence.

iii) Machine hardware efficiency problems.

iv) No proper scheduling technique.

CONCLUSION: - Dynamic capacity provisioning (DCP) has become a promising solution for reducing energy consumption in data centers in recent years. However, existing work on this topic has not addressed a key challenge, which is the heterogeneity of workloads and physical machines. In this paper, we first provide a characterization of both workload and machine heterogeneity found in one of Google's production compute clusters. Then we present HHADC, Hashning heterogeneity-aware framework that dynamically adjusts the number of machines to strike a balance between energy savings and scheduling delay, while considering the reconfiguration cost. Through experiments using Google workload traces, we found HHADC yields large energy savings while significantly improving task scheduling delay

REFERENCES:-

[1] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. Time Series Analysis, Forecasting, and Control. Prentice–Hall, 1994.

[2] S. Boyd et al.Convex optimization. Cambridge University Press, 2004.

[3] C. Chekuri and S. Khanna. On multi–dimensional packing problems. In Symposium on Discrete algorithms, 1999.

[4] Y. Chen et al. Analysis and lessons from a publicly available Google cluster trace. Tech. Rep. UCB/EECS-2010-95, 2010.

[5] J. Diaz et al. A guide to concentration bounds. In Handbook on Randomized Computing, 2001.

[12] S. Ghemawat et al. The Google file system. In SOSP, 2003.

[6] D. Gross and C. Harris. Fundamentals of queueing theory. ISBN: 0-471-17083-6, pages 244–247, 1998

[7] A. K. Mishra, J. L. Hellerstein, W. Cirne, and C. R. Das. Towardscharacterizing cloud backend workloads: insights from Google compute clusters. SIGMETRICS Perform. Eval. Rev., 37, March 2010.

[8] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs. Cutting the electric bill for Internet–scale systems. In ACM SIGCOMM, 2009.

[9] C. Reiss, A. Tumanov, G. Ganger, R. Katz, and M. Kozuch. Heterogeneity and dynamicity of clouds at scale: Google trace analysis. In ACM Symposium on Cloud Comp., 2012.

[10] S. Ren et al. Provably–efficient job scheduling for energy.