INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 6 (2018) PAGES 4086 – 4090
RECEIVED : 25.10.2018 PUBLISHED : 29.11.2018
Nov 29, 2018

HANDLING DYNAMIC RESOURCE ALLOCATION OF HETEROGENEOUS WORKLOAD IN CLOUD

M.Mala M.E[1], K.Sankar M.TECH (P.hD) [2]

1. PG Scholar, Computer Science and Engineering, Vivekananda College of Engineering for Women, Tiruchengode, India.

2. Assistant Professor, Computer Science and Engineering, Vivekananda College of Engineering for Women, Tiruchengode, India.

ABSTRACT

Infrastructure-as-a-Service (IaaS) cloud technology has received much attention from users who require large amounts of computer resources. Current IaaS clouds provide resources in the form of virtual machines (VMs) with homogeneous resource configurations, where different types of resources in VMs have a similar amount of capacity in a physical machine (PM). However, most user jobs require different amounts for different resources. For example, high-performance computing jobs require more CPU cores, while large-scale computing applications require more memory. The Existing Heterogeneous Resource Allocation approach called Skewness - Avoidance Multi-Resource Allocation (SAMR) to distribute resources according to diversified requirements for different types of resources. We propose a cooperative resource provisioning solution, and take advantage of differences of heterogeneous workloads so as to decrease their peak resources consumption under competitive conditions for four typical heterogeneous workloads this technique reduces power consumption without affecting performance. Virtual Machine (VM) to increase the resource utilization. Moreover, the previous methods do not provide efficient resource allocation for heterogeneous jobs in current cloud systems and do not offer different SLO degrees for different job types to achieve higher resource utilization and lower SLO violation rate. Therefore, we propose a Customized Cooperative Resource Provisioning (CCRP) scheme for the heterogeneous jobs in clouds

1. INTRODUCTION

More and more computing and storage are moving from PC-like clients to data centers or (public) clouds, which are exemplified by typical services like EC2 and Google Apps. The shift toward server-side computing is driven primarily not only by user needs, such as ease of management (no need of configuration or backups) and ubiquity of access supported by browsers but also by the economies of scale provided by high-scale data centers , which is five to ten over small-scale deployments. However, high-scale data center cost is very high, e.g., it was reported in Amazon that the cost of a hosting data center with 15 megawatt (MW) power facility is high as $5.6 M per month. High data center cost puts a big burden on resource providers that provide both data center resources like power and cooling infrastructures and server or storage resources to hosted service providers, which directly provides services to end users, and hence how to lower data center cost is a very important and urgent issue.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 6 (2018) PAGES 4086 – 4090
RECEIVED : 25.10.2018 PUBLISHED : 29.11.2018
Nov 29, 2018

Our cooperative transmission protocol consists of two phases. In the *routing phase*, the initial path between the source and the sink nodes is discovered as an underlying "one-node-thick" path. Then, the path undergoes a thickening process in the "*recruiting-and-transmitting*" phase. In this phase, the nodes on the initial path become cluster heads, which recruit additional adjacent nodes from their neighborhood. Due to the fact that the cluster heads recruit nodes from their immediate neighborhood, the inter-clusters distances are significantly larger than the distances between nodes in the same cluster. Recruiting is done dynamically and per packet as the packet traverses the path.

2. LITERATURE SURVEY

Physical resources, such as a high-performance computing(HPC) cluster, provide static capacity yet experience dynamicload as demand fluctuates. As a result, resources maybe under-utilized during periods of low demand, with idlecycles drawing power and costing the organization money, orthey may be over-utilized during periods of high demand,resulting in increased queue wait times or crashing underthe load. In typical HPC environments, Grid technologies.

Cloud computing holds a promise to deliver large-scaleutility computing services to a wide range of consumers inthe coming years. The model's attractiveness stems mainlyfrom the increased flexibility it is able to deliver through ondemandacquisition and release of IT resources. The past few years have seen a growing number ofscientific computing communities evaluating IaaS clouds forrunning scientific workloads. For example, several usershave tried virtual clusters built with resources leased fromcommercial clouds such as Amazon EC2.

The notion of Cloud computing has not only reshaped the fieldof distributed systems but also fundamentally changed howbusinesses utilize computing today. While Cloud computingprovides many advanced features, it still has some shortcomingssuch as the relatively high operating cost for both public andprivate Clouds. The area of Green computing is also becomingincreasingly important in a world with limited energy resourcesand an ever-rising demand for more computational power.

3. SYSTEM ANALYSIS

Cloud users rent VMsfrom IaaS public clouds to run their applications in a pay-asyou-gomanner. Cloud providers charge users according tothe resource amounts and running time of VMs. Cloud users submit their VM requests to the cloud data center accordingto their heterogeneous resource demands and choose theVM types that are most appropriate in terms of satisfyingthe user demands while minimizing the resource wastage.All VM requests are maintained by a schedulingqueue. According to the arrival ratesand service rates of requests, SAMR conducts resource predictionbased on a Markov Chain model periodically in everytime slot with a duration of t to satisfy the user experience interms of VM allocation delay. In VM schedulingphase during each time slot with the length t, cloudproviders allocate resources and host each VM into PMsusing SAMR allocation algorithm. Resource deployment means choosing provision and runtime management of softwareSo, the ultimate goal of the cloud user is minimize

the costs by leasing the resources and theto maximize the perspective of the cloud service providerbenefit by allocating resources efficiently. In order to reach the goal, the cloud user must ask cloud service provider a provision for the resources either static or dynamic, Virtual Machine (VM) to increase the resource utilization. A major benefit of using bytecode is porting. However, the overhead of interpretation means that interpreted programs almost always run more slowly than programs compiled to native executables would, and Java suffered a reputation for poor performance. This gap has been narrowed by a number of optimization techniques introduced in the more recent JVM implementations.

4. IMPLEMENTATION

To more accurately predict the execution time of jobs, we extract two types of features: job-related features and system-related features. We use the historical data to estimate the run time of jobs. We extract the numerical values of the features from the historical data for predicting the jobs' execution time .

bestpath

Field Name	Data Type
path	Text
cost	Text
count	Text
throughput	Text

Table.1.

In the historical data, we consider part of them as training data, and use part of the data as testing data. To improve the accuracy, we use the cross-validation to perform classification. We classify jobs into two types: short jobs and long jobs. We consider jobs with execution time no more than 10 minutes as short jobs [2], and we consider jobs with execution time more than 10 minutes as long jobs. To achieve high resource utilization, CCRP packs the complementary jobs belonging to the same type together and allocates the resource to the packed job. Below, we introduce the resource allocation algorithm. Database design is a collection of interactive data store. It is an effective method of defining, storing and retrieving the information in the database. The database design is independent of any relational database management system and it is a logical model. The logical design is mapped according to RDBMS used for implementation. The data contained in the database can be multiple application and users. It prevents the unauthorized from accessing data and ensures the privacy of data.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 6 (2018) PAGES 4086 – 4090
RECEIVED : 25.10.2018 PUBLISHED : 29.11.2018
Nov 29, 2018

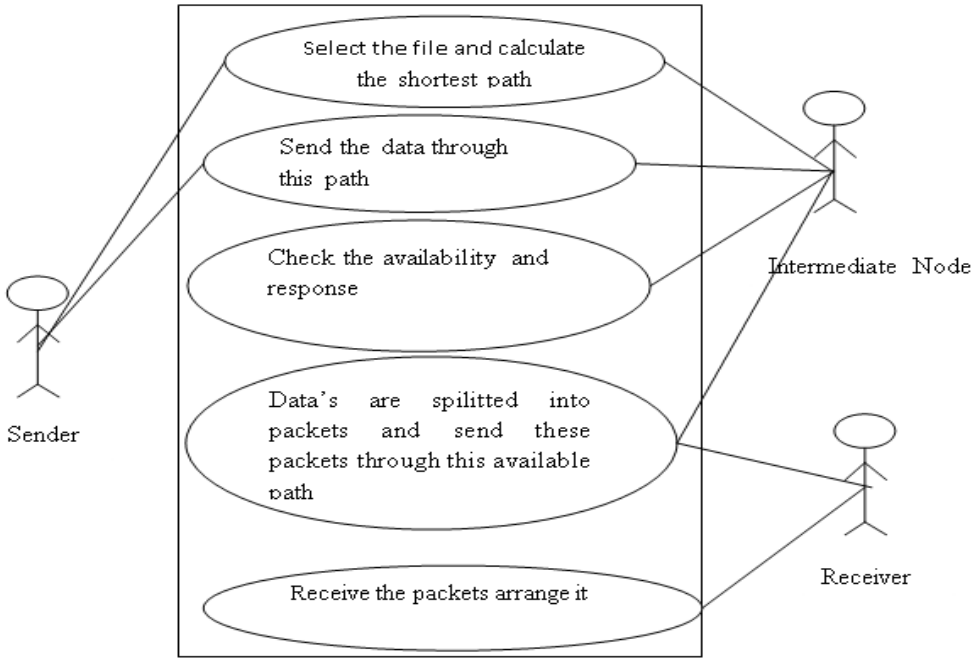
systemdetails

Field Name	Data Type
systemname	Text
ipaddress	Text
port	Text
status	Text
password	Text

Table.2

5. ANALYSIS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.



The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML. In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information. A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes

operate with one another and in what sorder. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams. Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

CONCLUSION

In this paper, we have discussed about resource management in general, the existing resource allocation and monitoring strategies from the current research works. In this paper, we propose customized cooperative resource provisioning scheme (CCRP) in clouds to increase the resource utilization and reduce SLO violation rate by customizing SLO availability and offering different degrees of SLO availability for different jobs types. This paper has summarized different method (algorithms technique) and theory which being used to formulate framework and model, derived to provide a better resource allocation and monitoring process in terms of a better performance, competitive and efficiency to meet the required SLA, improved the resource performance and lowered the power consumption. We hope this paper will motivate researchers to explore and formulate a new mechanism to solve issues in allocating and monitoring resources in cloud computing.

REFERENCES

[1] S. Genaud and J. Gossa, “Cost-wait trade-offs in client-side resource provisioning with elastic clouds,” in Proc. IEEE Int. Conf. Cloud Comput., 2011, pp. 1–8.
[2] E. Michon, J. Gossa, S. Genaud, et al., “Free elasticity and free cpu power for scientific workloads on IaaS clouds,” in Proc. IEEE 18th Int. Conf. Parallel Distrib. Syst, 2012, pp. 85–92. [3] P. Marshall, H. Tufo, and K. Keahey, “Provisioning policies for elastic computing environments,” in Proc. IEEE 26th Int. Parallel Distrib. Process. Symp. Workshops PhD Forum, 2012, pp. 1085–1094.
[4] L. Wang, J. Zhan, W. Shi, and Y. Liang, “In cloud, can scientific communities benefit from the economies of scale?” IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 2, pp. 296–303, Feb. 2012. [5] R. V. den Bossche, K. Vanmechelen, and J. Broeckhove, “Costoptimal scheduling in hybrid IaaS clouds for deadline constrained workloads,” in IEEE 3rd Int. Conf. Cloud Comput., 2010, pp. 228–235.
[6] M. Malawski, G. Juve, E. Deelman, and J. Nabrzyski, “Cost-and deadline-constrained provisioning for scientific workflow ensembles in IaaS clouds,” in Proc. Int. Conf. High Perform. Comput., Netw., Storage Anal., 2012, pp. 22.