International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 05
April 2017

Analysis of Cloud Computing Literature Using Clustering

Shirisha Getty
Department of CSE

ABSTRACT: The acceptance and rapid development of cloud computing in current years has led to a huge amount of periodicalscontaining the achieved data of this area of research. Due to the interdisciplinary nature and high relevance of cloudcomputing investigation, it turn out to beprogressively difficult or even impossible to comprehend the overall structure and developmentof this field without systematic approaches. The method to make the most of a various set of tasksfrom the available resources in cloud efficiently is proposed in this paper. For thisreservation cluster is presented, in which all the unscheduled tasks are located and a newmapping is done to minimalize both the resource usage and execution time.

KEYWORDS-cloud computing, cloud computing research,clustering, performanceenhancement

I. INTRODUCTION

For many computational applications, large numbers ofresources are needed and this utilization of resources is forlong period of time. Cloud computing, which is anemerging technology, provides the proper hosting ofresources by leasing them from huge data centers onlywhen they are needed. Cloud computing is replacing allexisting technologies by offering their customer to payonly what they use. For example, an organization can buyany software or service for required period of time on thecloud rather than to purchase a machine for that purpose. Itoffers infrastructure, platform, software and data asservices and these are subscription based services meanspay-as-you-go model (C. Pelletingeas, 2010 [4]). Theseservices are known as Infrastructure as a service (IaaS),platform as a service (PaaS), Software as a service (SaaS)and data as a service (DaaS) respectively. Infrastructure asa service ensures processing, storage, network and otherfundamental computing resources to the users. Examplesof IaaS based services are Amazon EC2, IBM`s Blue cloud,Eucalyptus, Rackspace Cloud etc. the platform as a servicegives a high level integrated environment to build, test,deploy and host customer created applications. Examplesof PaaS based services are Google App Engine, EngineYard, Heroku etc. Software as a service is a softwaredelivery model in that the applications are accessed bysimple interface like web browser over Internet. Examplesof SaaS based services are Web Mail, Google Docs,Facebook etc (G.Malathy et al [1]). Data as a serviceprovides an infrastructure for web scale data mining andknowledge discovery in order to empower the applicationsand services with intelligence.

Cloud computing models such as public, private,community and hybrid models can be implemented byusing virtualization. Virtualization is the virtual evaluationof computing elements like hardware, software, memory,storage, network and so on (C. Pelletingeas, 2010 [4]). Itallows the sharing of physical resources and higherutilization rate with optimal storage. It also reduces thepower consumption and hardware investment andimproves the system management without extra cost.Thus cloud is a package of services that offersinfrastructure, platform, software and data as services. Somany researches are being made for improving theseflavors of services. But the dark side of using thisvirtualization is degradation of performance due to extraoverhead. CPU usage, memory, storage and network arethe performance factors for cloud computing. Since fastaccessing of data and resources is highly demanded incloud environment. Any organization adopting cloudcomputing certainly expect the kind of enhancedperformance. But this performance is degraded due tolimited bandwidth, high response

time, inefficient CPU &memory utilization, scalability bottleneck and unnecessaryuse of data centers.

I/O virtualization poses a more difficult problem becauseI/O devices are shared among all virtual machines. Itrequires a privilege domain from guest VMs to access I/O.This intervention leads to longer I/O latency and higherCPU overhead due to context switches between the guestVMS and VMM (Virtual Machine Monitor). Performanceof cloud computing is also dependent on the underlyingcloud infrastructure. This work is aimed to addressdifferent issues that are responsible for improving theperformance of cloud computing.

II. RELATED WORKS

Qian etal [14] proposed using cloud resources for a class of adaptive programs, where applicationspecific flexibility in computation is needed with constant time-restrict and assistance budget. The adaptiveapplications are maximized with Quality of Service (QoS) very exactly and with the assistance of dynamically variousthe adaptive parameters the value of application-specific benefit characteristic is received. A multi-inputmulti-output feedback manipulate version primarily based dynamic resource provisioning set of rules is advanced thatadopts reinforcement studying to modify adaptive parameters to assure the most advantageous applicationapprovals in the time constraints.

Jaliya et al [9] proposed cloud technology Apache Hadoop and Microsoft DryadLINQ totwo bioinformatics applications with the scheduling of duties. The applications have pair clever Alucollection alignment utility and an Expressed Sequence Tag (EST) series meeting application.The performance of the generation is compared and analyzed with conventional implementation beneathvirtual and non-virtual hardware platforms.

Seokho et al [15] proposed a service-levelagreement whilemaking reservations for cloud offerings. The provided multi-issue negotiation mechanism supportseach charge and time-slot negotiations among cloud retailers and tradeoff between rate and time-slotutilities. The agents make more than one proposals in a negotiation spherical to generate aggregated utility withvariations in cloud agents and time-slot utilities.

Qian et al [13] proposed a framework the usage of a measurement module in each guest virtualsystem to degree each running executable in that digital device. The size moduletransfers the measured values to the trusted digital device via a widespread inter-virtual systemcommunication mechanism. This additionally stores the values in a measurement desk and the device extendsthe values right into a distinct platform configuration sign in. A memory watcher module is also added tomake sure process trustworthiness. Lei et al [11] proposed a public cloud utilization model for small-tomedium scale scientific groups to make use of elastic resources on a public cloud web site. Also,carried out an modern device named DawningCloud, at the middle of which a lightweight providercontrol layers going for walks on top of a common control service framework. The system has beenevaluated and determined that DawningCloud saves the resource intake to a maximum quantity.

Zibin et al [21] proposed a component rating framework, named FTCloud with two ratingalgorithms for building fault-tolerant cloud packages. The first algorithm employs issueinvocation structures and invocation frequencies for making extensive factor ranking and 2dset of rules systematically fuses the machine structure records as well as the application designer'sexpertise to identify the extensive additives in a cloud application. After this an finest faulttolerance method for the sizeable cloud additives is automatically determined.

Ganesh et al [7]investigated the usage of a divisible load paradigm to design green strategies to decrease the generalprocessing time for acting massive-scale polynomial product computations in compute

cloudenvironments. For publish-processing a compute cloud gadget with the aid allocator distributing thewhole load to a hard and fast of digital CPU times is processed. Finally thru simulation the performanceof the strategy is quantified.

Daniel et al [4] mentioned the opportunities and demanding situations for green parallel data processingin clouds and provided the challenge named Nephele. Nephele is the primary records processing framework toexplicitly take advantage of the dynamic resource allocation provided by each undertaking scheduling and execution.Particular responsibilities of a processing process can be assigned to diverse forms of virtual machines which can beinstantiated routinely and concluded in the course of the job execution.

Hong-Ha et al [8] taken into consideration theproblem of scheduling lightpaths and computing resources for sliding grid demands in Wave DivisionMultiplexing (WDM). On every call for a joint scheduling set of rules decides the begin time, reserve anquantity of computing sources and provide a number one lightpath. For obtaining an Integer LinearProgramming (ILP) components is developed and to reap scalability heuristic algorithms primarily based onjoint aid scheduling is used.

Khawar et al [10] proposed a pilot activity idea that has shrewd information reuse and process executionstrategies to minimize the scheduling, queuing, execution and information get admission to latencies. By this approach,considerable upgrades inside the average turnaround time of a workflow can be carried out. This isevaluated the usage of CMS Tier0 information processing workflow, and then in a controlled environment.

Min et al[12] developed an availability-driven scheduling scheme that improves the actual-time Directed AcyclicGraph (DAG), iteratively with the aid of allocating two copies of one conversation assignment to 2 disjointlightpaths for statistics switch even as pleasant utility cut-off date necessities. Analysis demonstratedthe effectiveness and feasibility of the proposed scheduling scheme.

Thomas et al [17] added a version for estimating the enterprise effect of operational threatas a result of changes. The version takes into account the community of dependencies between systemand services, probabilistic trade-associated downtime, uncertainty in enterprise manner call for, andnumerous infrastructural characteristics. The version is evaluated using simulations based totally on the economicstatistics.

Xiao et al [18] proposed a communication-conscious load-balancing method that is able toimproving the performance of communication-intensive packages with the aid of increasing the effectiveusage of networks in cluster environments. Also a conduct version for parallel applications isadded with the load-balancing technique with large necessities of community, CPU, memory and diskI/O assets.

Young et al [19] investigated the hassle of scheduling workflow programs on grids andoffers a unique scheduling algorithm for the minimization of utility of entirety time. Theperformance of grid assets modifications dynamically and the accurate estimation of overall performance istough, and the proposed rescheduling method deal the unforeseen performance fluctuationssuccessfully.

Dharma et al [5] proposed a data replication algorithm that isn't only a provabletheoretical performance assure, but also may be implemented in dispensed manner. This is based ona polynomial time centralized replication algorithm that reduces the full statistics file get admission to postpone with the assistance of atleast half of of that reduced via the most reliable replication solution.

III. CLOUD TECHNOLOGY APPROACHES

The pseudo code for the proposed algorithm is described below. Consider a network G(V, E)with a

set of N cloudlets in a cloud system. The objective is to compute the tasks within the cloudsystem efficiently. That is the tasks should be completed with the available cloudlets in the cloudsystem.

Pseudo Code for the Proposed Method

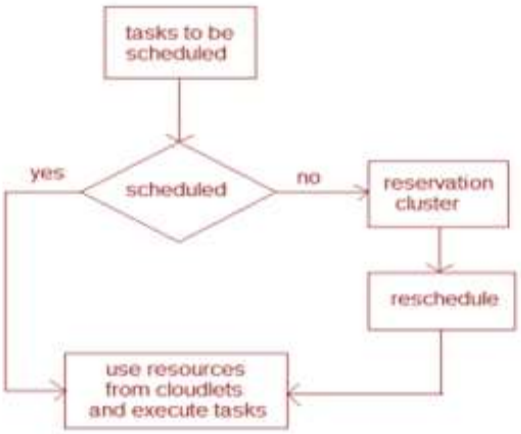

Figure 2: Reservation cluster-based cloud computing approach

Input: For an application G(V, E) with N cloudlets in a cloud system

Output: Tasks scheduled within the cloud system

1. Compute the availability of cloudlets
2. Initialize resource usage and execution time to infinity
3. Generate a random schedule on the tasks
4. Share the task to each cloudlet
5. If all the tasks are scheduled
6. Select the best resource usages and execution time
7. Endif
8. Place the unscheduled tasks on the reservation cluster
9. Find the number of unscheduled task
10. If the reservation cluster size is not sufficient
11. Increment the size of reservation cluster
12. Endif
13. Schedule the tasks in the reservation cluster at a time
14. Compute the tasks
15. Obtain the resource usage and execution time
16. End the process

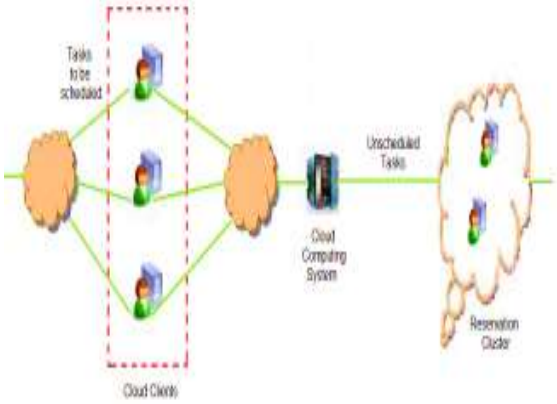

Figure 3: Structure of reservation cluster-based cloud computing approach

IV. CONCLUSION

Recital of virtualization with cloud computing is aforemost issue to be investigated. Poor performance candeficiencythe interest of clients. Clustering and caching are theproposed methodologies for cultivating the performance inthis work.This paper discourses a reservation cluster-based cloud computing system. Cluster is formed by theunscheduled tasks, and a rescheduling is done for the cloudlets inside the cluster without any iteration.Since there is no iteration, the computation time for the task to be executed with the resources will bereduced.

REFERENCES

[1] G. Malathy, Rm. Somasundaram, "PerformanceEnhancement in Cloud Computing using ReservationCluster", European Journal of Scientific Research, ISSN1450-216X Vol. 86 No 3 September, 2012, pp.394-401.

[2] Donglai Zhang, Paul Coddington and AndrewWendelborn," Improving Data Transfer Performance ofWeb Service Workflows in the Cloud Environment", Int. J.Computational Science and Engineering, Vol. 1, No. 1/1,2012.

[3] Wei Huang ,Jiuxing Liu, BulentAbali and Dhabaleswar K.Panda, "A Case for High Performance Computing withVirtual Machines", ICS '06 Proceedings of the 20th annualinternational conference on supercomputing, pages 125-134.

[4] C. Pelletingeas, "Performance Evaluation of Virtualizationwith Cloud Computing", MSc Advanced Networking, 2010.

[5] Dharma TejaNukarapu, Bin Tang, Liqiang Wang, and Shiyong Lu, "Data Replication in DataIntensive Scientific Applications with Performance Guarantee", IEEE Transactions on Paralleland Distributed Systems, Vol. 22, No. 8, pp. 1299- 1306, August 2011.

[6] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud Computing and Grid Computing 360 degreecompared", Proceedings of Grid Computing Environments Workshop (GCE'08), pages 1-10,2008.

[7] Ganesh NeelakantaIyer, BharadwajVeeravalli, and Sakthi Ganesh Krishnamoorthy, "OnHandling Large-Scale Polynomial Multiplications in Compute Cloud Environments usingDivisible Load Paradigm", IEEE Transactions on Aerospace and Electronic Systems, Vol. 48,No. 1, pp. 820-831, January 2012.

[8] Hong-Ha Nguyen, Mohan Gurusamy, and Luying Zhou, "Scheduling Network and ComputingResources for Sliding Demands in Optical Grids", Journal of Lightwave Technology, Vol. 27,No. 12, pp. 1827-1836, June 15 2009.

[9] JaliyaEkanayake, ThilinaGunarathne, and Judy Qiu, "Cloud Technologies for BioinformaticsApplications", IEEE Transactions on Parallel and Distributed Systems, Vol. 22, No. 6, pp. 998-1011, June 2011.

[10] KhawarHasham, Antonio DelgodoPeris, AshiqAnjum, Dave Evans, Stephen Gowdy, Jose M.Hernandesz, Eduardo Huedo, Dirk Hufnagel, Frank van Lingen, Richard McClatchey, andSimon Metson, "CMS Workflow Execution Using Intelligent Job Scheduling and Data AccessStrategies, IEEE Transactions on Nuclear Science, Vol. 58, No. 3, pp. 1221-1232, June 2011.

[11] Lei Wang, Jianfeng Zhan, Weisong Shi, and Yi Liang, "In Cloud, Can Scientific CommunitiesBenefit from the Economies of Scale", IEEE Transactions on Parallel and Distributed Systems,Vol. 23, No. 2, pp. 296-303, February 2012.

[12] Min Zhu, Wei Guo, Shilin Xiao, Anne Wei, Yaohui Jin, Weisheng Hu, and Benoit Geller,"Availability-Driven Scheduling for Real-Time Directed Acyclic Graph Applications in OpticalGrids", Journal of Optical Communication Network, Vol. 2, No. 7, pp. 469-480, July 2010.

[13] Qian Liu, Chuliang Weng, Minglu Li, and Yuan Luo, "An In-VM Measuring Framework forIncreasing Virtual Machine Security in Clouds", IEEE Security and Privacy, Vol. 8, No. 6, pp.56-62, Nov.-Dec. 2010.

[14] Qian Zhu, and GaganAgrawal, "Resource Provisioning with Budget Constraints for AdaptiveApplications in Cloud Environments", IEEE Transactions on Services Computing, 27December 2011, DOI: 10.1109/TSC.2011.61.

[15] Seokho Son, and KwangMongSim, "A Price and Time-Slot Negotiation Mechanism for CloudService Reservations", IEEE Transaction on Systems, Man, and Cybernetics – Part B:Cybernetics, Vol. 42, No. 3, pp. 713-728, June 2012.

[16] B. Sotomayor, R. Santiago Montero, I. Martin Llorente, and I. Foster, "Virtual InfrastructureManagement in Private and Hybrid Clouds", IEEE Internet Computing, Vol. 13, No. 5, pp. 14-22, September/October 2009.

[17] Thomas Setzer, Kamal Bhattacharya, and Heiko Ludwig, "Change Scheduling based onBusiness Impact Analysis of Change-Related Risk", IEEE

Transactions on Network ServiceManagement, Vol. 7, No. 1, pp. 58-71, March 2010.

[18] Xiao Qin, Hong Jiang, Adam Manzanares, XiaojunRuan, and Shu Yin, "CommunicationAware Load Balancing for Parallel Applications on Clusters", IEEE Transactions onComputers, Vol. 59, No. 1, pp. 42-52, January 2010.

[19] Young Choon Lee, RikySubrata, and Albert Y. Zomaya, "On the Performance of a DualObjective Optimization Model for Workflow Applications on Grid Platforms", IEEETransactions on Parallel and Distributed Systems, Vol. 20, No. 9, pp. 1273-1284, September2009.

[20] S. Zhong, and J. Ghosh, "A comparative study of generative models for document clustering",Proceedings of the SDM Workshop on Clustering High Dimensional Data and Its Applications,2003.

[21] Zibin Zheng, Tom Chao Zhou, Michael R. Lyu, and Irwin King, "Component Ranking forFault-Tolerant Cloud Applications", IEEE Transactions on Services Computing, ISSN: 1939-1374, DOI: 10.1109/TSC.2011.42.