International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622
National Conference on Advances in Engineering and Technology
(AET- 29th March 2014)

RESEARCH ARTICLE OPEN ACCESS

Survey of Adaptive and Dynamic Management of Cloud Datacenters

Kamali Gupta*, Vijay Kumar Katiyar**
*(Department of Computer Science & Engineering, GIMT, Kurukshetra Email: kamaligupta13@gmail.com)
** (Department of Computer Science & Engineering, M. M. University, Mullana Email: katiyarvk@mmumullana.org)

ABSTRACT
As cloud computing has emerged as an enabling technology that allows the Information Technology world to use the computer resources more efficiently and effectively such that the users have unlimited computing power at their disposals whenever required, so the cloud services has made it the best Information Technology solution. Additionally, it has increased the computational and storage capacity without investing in new infrastructure, training new personnel's or licensing new software. The concepts of virtualization, energy efficiency and resource provisioning have been recognized as the key techniques to enhance the scheduling services provided by the cloud. In this paper, we presented a summary of various research activities carried out for the effective management of cloud resources.
Keywords **–** Data Centers, virtualization, resource provisioning, energy efficiency, Infrastructure as a Service.

I. INTRODUCTION

Cloud computing is a service oriented paradigm that offers "everything as a service" over internet i.e. platform, infrastructure (server space) and services can be shared [1]. Cloud Computing is a term used to illustrate both a platform and type of application. As a platform it provides, configures and reconfigures servers, where the servers can be physical machines or virtual machines. On the other hand, Cloud Computing [2] describes applications that are extended to be accessible through the internet and for this purpose large data centers and powerful servers are used to host the web applications and web services.

It utilizes the techniques of virtualization and load balancing for increasing the cloud performance and complete utilization of resources. Other than these, it also makes use of technologies like distributed computing, networking, web services etc. Cloud computing is called 'cloud' since a cloud server can have any configuration and can be located anywhere in the world. Cloud services allow individuals and businesses to use software and hardware resources that are managed by third parties at the remote locations e.g. online file storage, social networking sites, operating webmail and online business applications [3].

Clouds are basically virtualized data centers and applications offered as service on a subscription basis. Web based companies (Amazon, eBay), hardware vendors (HP, IBM), telecom providers (AT&T, Verizon), and software firms (Oracle/Sun) are investing huge amount of capital in establishing huge data centers. Cloud computing emphasizes on pay per use economic model means customers pay for services on pay-per-use (or pay as you go) basis as per their requirement [4].

In a cloud computing environment, users can access the operational capability faster within internet application. The internet platform of cloud computing provides many applications for users like video, music etc.

Although cloud computing has been widely used, the research on resource management in cloud environment is still an early stage. The main objective of the research work is to investigate the relevant efficient and enhanced resource utilization approaches for cloud based system. Another focus of the work is to study the existing energy management techniques.

An overview of cloud environment is presented in this section. The rest sections of the paper are organized as follows: in section 2, three methods for improving the efficiency of cloud data centers are presented. Section 3 discusses the related work. Section 4 concludes with a summary of the research work.

II. SERVICES OF CLOUD COMPUTING

Cloud computing service models are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) [5]

- ***Software as a Service (SaaS)***

In this model, the service user only needs to access the service itself as a web application, and not the platform or the infrastructure the service is running on. Applications such as social media sites, office software's, and online games enrich the family of SaaS-based services.

- ***Platform as a Service (PaaS)***

Platform as a service (PaaS) is an entire infrastructure packaged that can be used to design and implement the applications and deploy them in a public or private cloud environment. Typical examples of PaaS are Google App Engine, Windows Azure, Engine Yard and Force.com.

- ***Infrastructure as a Service (IaaS)***

The Infrastructure as a Service is a provision model in which an organization outsources the equipment used to support operations, such as-storage, hardware resources, servers and networking components. The service provider himself owns the equipment and is responsible for housing and maintaining it. The client pays on a per-use basis. Infrastructure services are built on top of a standardized, secure, and scalable infrastructure. Some level of redundancy is required to be built into the infrastructure to ensure the high availability and elasticity of resources and it must be virtualized. Virtualized environments make use of server virtualization, typically from VMware [6], XEN as the basis of running services.

III. CHALLENGES OF CLOUD COMPUTING

The following are the challenges faced by cloud computing environment[7]:

- ***Security and Privacy***

It deals with securing the stored data and to monitor the use of the cloud by the service providers. This challenge can be addressed by storing the data into the organization itself and allowing it to be used in the cloud.

- ***Service Delivery and Billing***

The service level agreements (SLAs) of the provider are not adequate to guarantee the availability and scalability as it is difficult to assess the cost involved due to dynamic nature of services.

- ***Interoperability and Portability***

As the cloud environment is highly dynamic to user requests and due to the concept of virtualization, the leverage of migrating in and out of the resources and applications should be allowed.

- ***Reliability and Availability***

Cloud providers still lack in round-the-clock service which results in frequent outages. Therefore, it becomes important to monitor the service being provided using internal or third party tools.

- ***Automated service provisioning***

A key feature of cloud computing is elasticity; resources can be allocated or released automatically. So a strategy is required to use or release the resources of the cloud, by keeping the same performance as traditional systems and using optimal resources.

- ***Performance and Bandwidth Cost***

Businesses can save money on hardware but they have to spend more for the bandwidth. This can be low cost for smaller applications but can be significantly high for the data-intensive applications.

- ***Energy Cost***

Cloud infrastructure consumes enormous amounts of electrical energy resulting in high operating costs and carbon dioxide emissions [8].

- ***Virtual Machines Migration***

With virtualization technology, an entire machine can be taken as a file or set of files. To unload a heavily loaded physical machine, it is required to move a virtual machine between physical machines. The main objective is to distribute the load in a datacenter or set of datacenters. Then a strategy is required to dynamically distribute load when moving virtual machine to avoid bottlenecks in Cloud computing system.

IV. CLOUD DATACENTER RESOURCE MANAGEMENT

In order to improve the efficiency of cloud resources, most service providers are going to consolidate existing systems through virtualization [9]. Virtualization technology increases the energy efficiency by creating multiple virtual machine instances on a physical server thus improving the utilization of resources and increasing Return on Investment (ROI).The Virtual machines in a cloud infrastructure can be live migrated to another host in case user application needs more resources. The service providers of cloud monitor and predict the demand and thus allocate resources according to the demand. Those applications that require less number of resources can be consolidated on the same server. Datacenter always maintains the active servers according to the current demand which results in low

energy consumption than the conservative approach of over-provisioning [10].

Resource provisioning [11] plays an important role in ensuring that the service provider adequately accomplishes their obligations to customers in terms of Service Level Agreements (SLAs) while maximizing the utilization of underlying resources, and it requires two steps. In the first step, static planning is done in which the initial grouping of Virtual Machines (VM) takes place, then the classification of VMs is done and finally these are deployed onto a set of physical hosts. Second is dynamic resource provisioning in which allocation of additional resources, creation and migration of VMs takes place dynamically according to the varying workloads.

Energy efficiency is one of the main challenges that datacenters are facing nowadays. The rising energy cost is a highly potential threat as it increases the Total Cost of Ownership (TCO) and reduces the Return on Investment (ROI) of Cloud infrastructures. In cloud environment, the resource management should be energy-efficient as it reduces the cost of energy consumption of data center and the carbon dioxide footprint of a data center and increases the power efficiency at the architecture level [12]. Energy consumption at a data center is equal to total amount of energy consumed over a period of time. Energy management techniques employed at data centers can be static or dynamic. The static data management techniques are not suitable for responding to requests when workload changes abruptly. Dynamic energy management techniques configure the data centre at both hardware and software levels dynamically depending upon changing workload conditions [13].

V. RELATED WORK

A new scheduling policy has been proposed in [14] which aim at managing data centers to optimize the provider's profit. Modeling virtualized data centers offers a lot of advantages such as resource management, reduced power consumption, heterogeneity management, and efficient utilization of under used nodes. The work also takes into account the outsourcing capabilities which make it possible to outsource the resources to third party IaaS (Infrastructure as a Service) service providers. A model is developed for describing a virtualized data centers and all decisions are taken from an economic point of view.

In [15], authors demonstrated that utilizing low power idle nodes is an immediate remedy to reduce data center power consumption. To quantify the difference in energy consumption caused exclusively by virtual machine schedulers, simulation are carried out. Besides demonstrating the inefficiency of wide-spread default schedulers, an optimized scheduler (Optsched) is developed and its performance is analyzed in terms of cumulative machine uptime.

A multivariate probabilistic model for improving resource utilization for cloud providers is presented [16]. The proposed algorithm selects suitable Physical Machines (PM) for VM re-allocation which is then used to generate a reconfiguration plan. Two heuristics metrics are also described which can be used in the algorithm to capture the multi-dimensional characteristics of VMs and PMs.

The major pitfalls in cloud computing is related to optimizing the resources being allocated. Due of the uniqueness of the model, the task of resource allocation is performed with the objective of minimizing the total cost. Its other challenges are meeting customer demands and application requirements. In [17], authors discussed various resource allocation strategies and their challenges in detail.

Most data center workload demands are very spiky in nature and often vary significantly in a day. As the resource availability in a data center is generally unpredictable due to the shared feature of the cloud resources and because of the stochastic nature of the workload, severe service level agreement (SLA) violations may occur frequently. To overcome this problem, anautonomic resource controller is proposed that dynamically controls the resource allocation for data center's virtual containers [18]. The controller has two parts: A resource modeler that models the non-linearity of the system by employing different Machine Learning techniques allowing the datacenter to allocate the appropriate resources and a resource fuzzy tuner that dynamically tunes the allocated resources using fuzzy logic to sustain the desired performance taking into consideration the enforcing of service differentiation among clients.

In [19], author provides an introduction to the technique of resource provisioning and power or thermal management problems in datacenters, and a review of strategies that maximize the datacenter energy efficiency subject to peak or total power consumption and thermal constraints, along with meeting service level agreements in terms of task throughput and/or response time.

A brief introduction of state-of-the art techniques and research related to power saving in the IaaS of a cloud computing system, which consumes a large part of total energy in a cloud computing environment is presented [20]. At the end, some feasible solutions for building green cloud

computing are proposed. The aim is to provide a better understanding of the design challenges of energy management in the IaaS of a cloud computing system.

Many market-based resource management strategies are been brought out to implement resource scheduling in cloud computing environment. A large number of consumers rely on cloud providers to supply computing service, so economic effectiveness is a crucial decisive factor for scheduling policy. An economic scheduling model with business parameters and a dynamic scheduling algorithm is presented [21], which makes a trade-off between economic effectiveness and performance. Based on the model and algorithm, market-oriented workflow management architecture for cloud is presented, in which QoS based resource allocation mechanism is introduced to meet different consumer's demands and improve scheduling efficiency.

The research work presented in [22] focuses on optimization of cloud system by lowering operation costs by maximizing energy efficiency while satisfying user deadlines that were defined in service level agreements. It has been discussed that the total energy consumption can be minimized by shutting down the servers which are not presently being used and balancing the resource utilization for all the active servers.

A Hierarchical Scheduling Algorithm (HAS) which aims at minimizing energy consumption of servers and a network device is proposed [23]. A DMNS (Dynamic Maximum Node Sorting) method is developed for optimizing the placement of applications on servers that are connected to a common switch. Then, in order to reduce the number of running servers, hierarchical crossing switch adjustment is done. This results in reduced data transfer and reduced number of servers that are required for processing.

The research work of [24] presents a novel management algorithm to perform the task of VM migration. It has been presented that for moving live sessions between servers; dynamic management of virtualized machines is done as it exploits the computing resources without considering the allocation of resources on local or remote servers. Innovative algorithms are presented for deciding when a physical host should migrate part of its loads, which part of load should be moved and where it should be moved. The focus in this work has also been on deciding when a dynamic redistribution of load is necessary.

The research work presented in [25] explores the approaches for modeling, simulation or prototype implementation to help researchers to develop and evaluate their technical solutions. A comparison of various cloud simulation software's is presented. The research work also presents a survey of various cloud testbeds (Amazon EC2, Amazon S3, Google App Engine, Google Apps and Windows Azure) and the services that they offer.

The research work investigated the process of allocating virtual machines to the requesting jobs in a way that maximizes the resource utilization [26]. An improved genetic algorithm is used for automated scheduling policy.

A Pre-emptive online task scheduling algorithm [27] is presented which aims to provide a solution for online scheduling problems being faced by real-time tasks in an IaaS model. The ultimate goal is to maximize the total resource utility and efficiency. The generated results presented in the work shows an improvement over the existing Earliest Deadline First scheduling algorithm.

In virtual desktop cloud computing, client applications are executed in virtual desktops on remote servers. Its advantage can be measured in terms of usability and resource utilization; however, handling a large amount of users in the most efficient manner poses important challenges. In [28], authors introduced an optimization to increase the average utilization on a single host. It has been shown that the proposed overbooking approach together with an advanced scheduler can increase the average utilization of the resources. A cost-based allocation algorithm has been presented that aims to maximize the quality of the service both for the customers and the service provider. To further optimize the quality of the service, a reallocation algorithm has been proposed to rebalance the virtual desktops among the available hosts after a busy period. The last optimization presented in this paper concern on optimization of the energy consumption by dynamically adapting the amount of powered-on hosts to the actual system load.

VI. CONCLUSION

This paper discusses the mechanisms through which the efficiency of cloud data centers can be enhanced. The techniques of virtualization, energy consumption and resource management are discussed in detail. Several algorithms have been discussed in order to improve the total energy consumption in a data center. Additionally, many other authors proposed several algorithms to implement the techniques of virtualization in an efficient manner in order to increase the number of tasks executed by a cloud environment while increasing resource consumption i.e. overall goal is to allocate resources while minimizing energy consumption in a dynamic cloud environment.

REFERENCES

[1] S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang and A. Ghalasi, "Cloud Computing-The business perspective", *Decision Support Systems, 51(1), 2011, 176-189.*

[2] The NIST Definition of Cloud Computing, http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

[3] I. Brandic, "Towards self-manageable cloud services", *Proceeding of IEEE International Conference on Computer Software and Applications*, 2009, 128-133.

[4] M. Alhamad, T. Dillon and E. Chang, "A survey on SLA and performance measurement in cloud computing", *Lecture notes on Computer Science, Berlin Heidelberg: Springer, 2011, 469-477.*

[5] IBM, "Fundamentals of Cloud Computing", *Instructor Guide: ERC 1.0,2010, 17-23.*

[6] M. Rosenblum, "VMware's Virtual Platform", *Proceedings of Hot Chips, 1999, 185-196.*

[7] Q. Zhang, L. Cheng and R. Boutaba, "Cloud computing: state-of-the-art and research challenges", *Journal of Internet Services and Applications, 1(1),* 2010, *7-18.*

[8] F. Chen, J. Schneider, Y. Yang, J. Grundy and Q. He, "An energy consumption model and analysis tool for cloud computing environments", *International workshop on Green and Sustainable Software, 2012, 45-50.*

[9] A. K. Das, T. Adhikary, C. Hong, "An Intelligent approach for Virtual Machine and QoS Provisioning in Cloud Computing", *Proceeding of IEEE International Conference on Information Networking, 2013, 462-467.*

[10] S. Garg and R. Buyya, *"*Green Cloud Computing and Environment Sustainability*", Harnessing Green IT: Principles and Practices, Wiley Press, 2012, 315-340.*

[11] R. Urgaonkar, U. Kozat, K. Igarashi and M. Neely, "Dynamic resource allocation and power management in virtualized data centers", *Proceeing of IEEE Network Operations and Management Symposium*, 2010, 479-486.

[12] A. Beloglazov and R. Buyya, "Energy efficient resource management in virtualized cloud data centers", *Proceeding of IEEE International Conference on Cluster, Cloud and Grid Computing, 2010, 826-831.*

[13] J. Yuan,"A Novel Energy Efficient Algorithm for Cloud Resource Management", *International Journal of Knowledge and Language Processing, 4(2), 2013, 12-22.*

[14] I. Goiri, J. Berral, J. Fito, F. Julia, R. Nou, J. Guitart, R. Gavalda and J.Tores, "Energy-efficient and multifaceted resource management for profit-driven virtualized Data Centers", *Proceeding of International Conference on Future Generation Computer Systems, 2012, 718-731.*

[15] T. Knauth and C. Fetzer, "Energy-aware scheduling for infrastructure clouds", *Proceeing of IEEE International Conference on Cloud Computing Technology and Science, 2012, 58-65*.

[16] S. He, L. Guo, M. Ghanem and Y. Guo, "Improving Resource Utilization in the Cloud Environment using Multivariate Probabilistic Models", *Proceeing of IEEE International Conference on Cloud Computing, 2012, 574-581.*

[17] V. Vinothina, R. Sridaran and P. Ganapathi, "A Survey on Resource Allocation Strategies in Cloud Computing", *International Journal of Advanced Computer Science and Applications, 3(6), 2012, 97-104.*

[18] N. Elprince, "Autonomous resource provision in virtual data centers", *International Symposium on Integrated Network Management, 2013, 1365-1371.*

[19] M. Pedram, "Energy-Efficient Datacenters", *Proceeding of IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems, 31(10), 2012, 1465-1484.*

[20] S. Jing, S. Ali, K. She and Y. Zhong, "State-of-the-art research study for green cloud computing", *Journal of Supercomputing, 65(1), 2013, 445-468.*

[21] X. Shi and Y. Zhao, "Dynamic Resource Scheduling and Workflow Management in Cloud Computing", *Workshop on Web Information Systems Engineering, Lecture Notes in Computer Science, 6724, 2011, 440-448.*

[22] Y. Gao, Y. Wang, S. Gupta and M. Pedram, "An Energy and Deadline Aware Resource Provisioning, Scheduling and Optimization Framework for Cloud Systems", *Proceeding of IEEE International Conference on Hardware/Software Codesign and System Synthesis, 2013, 1-10.*

[23] G. Wen, J. Hong, C. Xu, P. Balaji, S. Feng and P. Jiang, "Energy-aware Hierarchical Scheduling of Applications in Large Scale

Data Centers", *Proceeding of IEEE International Conference on Cloud and Service Computing, 2011, 158-165*.

[24] X. Wang, B. Wang and J. Huang, "Cloud computing and its key techniques", *Proceeding of IEEE International Conference on Computer Science and Automation Engineering, 2011, 404-410.*

[25] G. Sakellari and G. Loukas, "A survey of mathematical models, simulation approaches and testbeds used for research in cloud computing", *Journal of Simulation Modeling Practice and Theory, ELSEVIER, 2013, 1-12*.

[26] H. Zhong, K. Tao and X. Zhang, "An Approach to Optimized Resource Scheduling Algorithm for Open-source Cloud Systems, *Proceeding of 5th IEEE Annual ChinaGrid Conference, 2010, 124-129*.

[27] R. Santhosh and T. Ravichandaran, "Pre-emptive Scheduling of On-line Real Time Services With Task Migration for Cloud Computing", *Proceeding of IEEE International Conference on Pattern Recognition, Informatics and Mobile Engineering, 2013, 271-275*.

[28] L. Deboosere, B. Vankeisbilck, P. Simoens, F. Turck, B. Dhoedt and P. Demeester, "Efficient resource management for virtual desktop cloud computing", *Journal of Supercomputing, 62(2), 2012, 741-767.*