International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 07 | July -2017 www.irjet.net p-ISSN: 2395-0072

Cloud Computing: A Perspective on Next Basic Utility in IT World

Sonam Seth[1], Dr. (Mrs.) Nipur Singh[2]

[1]Sonam Seth, Research Scholar, Kanya Gurukul Campus, Dehradun (U.K.) India, [2]Dr. (Mrs.) Nipur Singh, Professor, Department of Computer Science, Kanya Gurukul Campus, Dehradun (U.K.), India,

---

Abstract- *Technology is growing rapidly. Cloud Computing is the most promising and latest technology in the IT era. In real world everyone want fast, secure and reliable access of data so cloud computing is the next basic need in the IT world. Cloud computing is the abstraction of web enabled computers, resources and services to utilize the resources optimally. Cloud is a network of virtually distributed resources and clusters. Virtualization of resources facilitates resource utilization on demand. The deployment of virtual clusters offer full connectivity to the virtual machines connected in different network. To deal with big data management cloud architecture is introduced that facilitate big data processing using big data analytics. Cloud computing is the advanced version of distributed computing and parallel processing.*

Keywords: *Cloud Computing, Distributed Computing, Virtualization, Virtual machine, Big Data.*

1 INTRODUCTION

In recent trend, it is common to access information over the internet independently without hosting the infrastructure. This infrastructure made up of data centres that are monitored and maintained by the cloud providers. Cloud is virtualization of resources and composition. Cloud Computing has been recognized as a model that provide infrastructure, platform and services. Each service is respectively called Infrastructure as Service (IaaS), Platform as Service (PaaS), and Software as Service (SaaS). "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. The cloud model is composed of five essential characteristics, three service models, and four deployment models" [1]. In business perspective cloud computing can be defined by 4E approach [2]. 4E approach is given by Ashok Soota:

i) Explore: Promises & challenges.
ii) Envision: How this can transform organization.
iii) Enable: Resources & skills.
iv) Execute: Design, development & operation of cloud.

Science Clouds projects were started by University of Chicago (UC) and University of Florida (UFL) as a product. The first cloud at the University of Chicago, became available on March 3, 2008, and was named "Nimbus" [4]. "A cloud is a type of parallel and distributed system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified resources based on service-level agreements established through negotiation between the service provider and consumers" [3]. A client/user requests a resource and if the request is authorized, a Virtual machine is deployed on host. The UFL cloud configuration contains an innovation: private IP addresses are used in deployed virtual machines and network virtualization is used to connect virtual machines to the client/owner machines. The Cloud computing models are based on virtualization of computing resources allowing customers to provision resources on-demand on a pay-as-you-go basis [3] to optimize performance evaluation parameters. Virtual machines (VMs) establish a development path for incorporating new functionality such as server consolidation, migration, and secure computing. "Cloud Computing is a computing technology that provide on demand reliable quality of service to end-users that optimize the usage of resources as well as cost of resources". Cloud computing is the latest computing technology that delivers IT resources as services in which users are free from the burden of the low-level implementation or system administration details.

Big Data Analytics (Technology) is used to store large amount of data (in terabytes) and handled by Relational Database Management System (RDBMS). Big Data Technology is more suitable to maintain data having high volume, high velocity and high variety that is need of recent trends [5]. The data is not the "stock" in a data warehouse but a continuous flow [6]. "Big data is defined as large amount of data which requires new technologies and architectures so that it becomes possible to extract value from it by capturing and analysis process" [7]. Big data technology helps to collect large data on cloud and cloud computing technology helps to provide this collected large data; Virtualization technology by creating virtual machines on hosts, is used to maximize the utilization of computing resources and manages memory.

In this paper we have discussed the view of cloud computing and cloud computing architecture, explore the key issues and challenges. The paper is divided into six sections. Cloud computing architecture is discussed in section 2, in section 3 deployment models are given, cloud computing challenges and issues are discussed in section 4, literature survey of some existing resource allocation

methods for identified challenged is explained in section 5, and conclusion is given in section 6.

2 CLOUD COMPUTING ARCHITECTURE

In paper [8] authors have proposed architecture of a data centre's resource management system where resource management is divided into local and global policies. At the local level, the system facilitates the guest OS's power management strategies. The global manager gets the information on the current resource allocation from the local managers and applies its policy to decide whether the VM placement needs to be adapted.

2.1 IaaS: This is the base layer of cloud stack. It works as a base for the other two layers, for their execution. Stack is based on Virtualization. Examples are Amazon, GoGrid, 3Tera etc.

2.2 PaaS: This layer provides the platform that is development environment upon which other applications run. Examples are LAMP platform, Google's App Engine, Force.com etc.

2.3 SaaS: In this layer or model, a complete application is offered to the user, as a service on demand. A single instance of the service runs on the cloud and multiple end users are served. Examples are Google, Microsoft, and Salesforce etc.

3 DEPLOYMENT MODELS

3.1 Private Cloud: Private Cloud is used by one organization. The cloud infrastructure is operated only for an organization. It may be managed by the organization or a third party. Private Cloud is used by one organization. Services are paid.

3.2 Public Cloud: Mega-scale cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services [9]. Public cloud is used by general people. Public clouds are owned by large organization such as Google, Amazon, and Microsoft.

3.3 Community Cloud: Community clouds are shared by more than one organization. Services are based on pay per use. These types of clouds are setup for specific purpose; especially for research purpose. Example is NASA etc.

3.4 Hybrid Cloud: The cloud infrastructure is a composition of two or more clouds (private or public) that work together by using virtualization technology that enables data and application portability. In recent trend of research Hybrid clouds are used. These clouds are combination of other clouds.

4 CHALLENGES WITH CLOUD COMPUTING

In cloud computing framework scheduling of tasks with QoS constraints is a challenging technical problem. Dynamic resource provisioning for Big data application scheduling is a challenge in modern high performance computing systems. A key challenge for these systems is to provision shared resources on demand to meet QoS. Cloud computing is based on virtualization and distributed computing to support cost-efficient usage of computing resources, focuses on resource scalability and on demand services. Traditional data-centre oriented models are converted into distributed clouds with a loosely coupled network that offers enhanced communication and computational services to end-users with quality of service (QoS) requirements [10].

4.1 Resource Allocation, Scheduling and Optimization Issue

Resource allocation indicates that the resources are allocated to end users on-demand. Resources are distributed among various ports to fulfil their requests. Virtual resource model for virtualization of resources increases the utilization of resources optimally and describes the execution time.Virtualization of resources can conquer some limitations and allow on-demand creation/deployment of multiple isolated virtual networks that enable the creation of virtual private clusters on a per user basis. The virtualization technology allows Cloud providers to create multiple virtual machine (VMs) instances on a single physical server, and the utilization of resources increases and increases the return on investment [8].

4.2 Cost Optimization Issue

Cost is calculated in two aspects: computing cost and communication cost. Computing cost is the cost associated with resource computing capacity and communication cost is the data transfer cost. There are two types of computing resources: on-demand instances and reserved instances. On-demand instances are paid only when utilized and they are useful to satisfy dynamic demand. While reserved instances are paid for a certain time period and are independent of usage.

4.3 Processing Time and speed

Another issue is to maximize the throughput in less time with high speed. Performance is designed to an application's capabilities within the cloud infrastructure itself. Limited bandwidth, disk space, memory, and CPU cycles.

4.4 Memory management, Storage

Memory management is one of the main challenge in cloud computing. Capacity of the cloud computing systems can vary using cache memory by applying virtualization concept.

4.5 Dynamic Load Balancing, Scalability

Load balancing is a technique that provide maximum throughput with minimum response time. Load balancing is dividing the load among all servers, so the requests are serviced without any delay with load balancing. Load balancing is used to distribute a larger load to smaller processing nodes for enhancing the overall performance of system. Under provisioning and over provisioning are also two major issues in load balancing.

4.6 Security Issue

There are three types of security concern: Physical, Operational and Programmatic security. Security is one of the major issue which reduces the growth of cloud computing and complications with data privacy and data protection like virus in the system. In cloud computing system privacy of data (content) can be handled by the feature 'obfuscation', where this is possible otherwise simple Encryption-Decryption techniques can also be used. The obfuscation method uses a key which is chosen by the user and known by the privacy manager, but which is not communicated to the service provider. Thus the service provider is not able to de-obfuscate the user's data. This reduces the risk of unauthorized access of data on cloud.

4.7 Fault tolerance and Reliability

Fault tolerance is one of the major issue in cloud computing. Fault tolerance techniques are in use during the procurement, or development of the software. Performance efficient resource management strategies that can be applied in a virtualized data centre by a Cloud provider (e.g. Amazon EC2).

4.8 QoS (Quality of Service)

QoS is the combined effort of service performance, which determines the degree of satisfaction of a user for the service. Managing the QoS parameters on the resource provider's side such as price and load is the recent challenge in Cloud Computing. QoS comprises computation time, execution price, packet loss rate, throughput, and reliability [11].

4.9 SLA (Service Level Agreement)

In the consideration of profit of both parties service providers and consumers, SLA based scheduling in cloud computing is the major challenge in recent trend to optimize the response time, throughput and QoS.

5 LITERATURE SURVEY OF CHALLENGES WITH CLOUD COMPUTING ALONG WITH RESOURCE SCHEDULING ALGORITHMS

Following table shows challenges with cloud computing and resource scheduling algorithms to handle key issues:

Table 1: Literature survey of some Resource Scheduling Algorithms to handle Key Issues

Key Issues	Model Used	Outcomes	Limitations
Cost and Time	Pre-emptable shortest job next scheduling Algorithm(PSJN) [12]	Cost and improved response and execution time.	Need to improve to handle under-provisioning and over-provisioning.
Makespan	User priority guided Min-Min scheduling Algorithm [13]	Average SLA.	Need to improve SLA based parameters.
Makespan, Economic Cost, Energy Consumption, Reliability	Scheduling with Genetic [14]	Consume more energy and achieve higher level of load balancing.	Rescheduling of unexecuted task is required to minimize the computation cost.
Resource utilization, Time	A Particle Swarm Optimization based Heuristic for Scheduling [14]	Task scheduling	Lack of both reliability and resource availability criteria.
Process completion Time	Improved Particle Swarm Optimization [15]	Average lower SLA and average completion time.	Need to improve SLA.
Cost, Performance	Improved Cost-Based Algorithm for Task Scheduling [16]	Improves the computation and communication ratio.	Need to minimize the execution time that minimizes the makespan.
Makespan, Load Balance	Heuristic based strategy list scheduling [14]	Optimization of load balancing and reduced makespan.	Need to optimize scheduling algorithm.
Resource allocation based on load balance	Meta Heuristic Techniques like GA, ACO, PSO [17]	Reduce the power consumption and execution time.	Can achieve more optimized performance and optimal use of resources.
Cost, Virtualization, Time	Cost-Effective Virtual Machine Allocation Algorithm within Execution Time Bound [18][19]	A two-step heuristic scheduling method has been used to maximize the resource	Need to optimize QoS.

		utilization. Delay time and makespan are considered to reduce.	
Cost and time	Dynamic resource provisioning techniques [19][20]	Time and dynamic load assignment.	Need to optimize the result by modifying the model in the consideration of dynamic nature of users.
Scheduling, VM management	Dynamic Provisioning Dynamic Scheduling [19][21]	Better performance of VM.	Need to design efficient algorithm to handle unpredictable workloads optimally.

6 CONCLUSION

Cloud computing is the technology which enables the user to access resources using front end machines, there is no need to install any software. In this paper authors have discussed the concepts and definitions of cloud computing and cloud computing systems with the key issues or challenges and related existing some models to handle key issues of cloud systems. Also authors have discussed the architecture models of cloud computing paradigm. As clouds are designed to provide services to external users, providers need to be compensated for sharing their resources and capabilities. This field is growing up hence further research in this field is required in the direction of interaction protocols to support interoperability between different cloud service providers and optimized models to handle key issues.

REFERENCES

[1]. Peter Mell, Timothy Grance, "The NIST Definition of Cloud Computing", NIST Special Publication 800-145, 2011, Pages 7.

[2]. Pankaj Arora, Raj Bivani, Salil Dave, "Cloud Powering an Enterprise", TMH Edition, 2011, Pages 146.

[3]. Buyya R, Yeo CS, Venugopal S, Broberg J, Brandic I, "Cloud Computing and emerging IT Platforms: Vision, hype and reality for delivering computing as the 5th utility ", Future Generation Computer Systems, Science Direct, Vol. 25, No. 6, 2009, Pages 599-616.

[4]. K. Keahey, R. Figueiredo, J.Forstes, T. Freeman, M. Tsugawa, "Science Clouds: Early Experiences in Cloud Computing for Scientific Applications", 2008.

[5]. Marko Grobelnik, "Big Data Tutorial", 2012.

[6]. Thomas H. Davenport, Paul Barth and Randy Bean, "How Big Data is Different: Technology, Data & Analytics, Analytics & Performance, IT Strategy", Magzine: Fall 2012 Opinion & Analysis, 2012, Pages 43-46.

[7]. Katal A., Wazid M., Goudar R.H., "Big Data: Issues, Challenges, tools, and Good Practices", Contemporary Computing (IC3), Sixth International Conference, ISBN 978-1-4799-0190-6, INSPEC Access No. 13797966, 2013, Pages 404-409.

[8]. Anton Beloglazov, Rajkumar Buyya, "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machine in cloud data centers", Concurrancy and Computation: Practice and Experience 2012, Vol. 24, No. 13, 2011, Pages 1397-1420.

[9]. Fang Liu, Jin Tong, Jian Mao, Robert Bohn, John Messina, Lee Bedger, and Dawn Leaf, "NIST Cloud Computing Reference Architecture" (Special Publication 500-292), ACM, ISBN 1478168021 9781478168027, 2011, Pages 34.

[10]. Papagianni, C, Leivadeas, A., Papavassiliou, S. Maglaris, V., "On the optimal allocation of virtual resources in cloud computing networks", IEEE, Vol. 62, No. 6, ISSN 0018-9340, INSPEC Access No. 13475363, 2013, Pages 1060-1071.

[11]. Xiaonian Wu, Mengqing Deng, Runlian Zhang, Bing Zeng, Shengyuan Zhou, "A Task Scheduling Algorithm based on QoS-Driven in Cloud Computing", International Conference on Information Technology and Quantitative Management(ITQM2013), ScienceDirect, Procedia Computer Science, Vol. 17, 2013, Pages 1162-1169.

[12]. Nishant S.Sanghani, R.J. Khimani, K.K. Sutaria, Pooja P. Vasani, "Pre-emptable Shortest Job Next Scheduling in Private Cloud Computing", Journal of Information, Knowledge and research in Computer Engineering, Vol. 2, No. 2, ISSN 0975-6760, 2013, Pages 385-388.

[13]. Huankai Chen, Frank Wang, Na Helian, Gbola Akanmu, "User Priority Guided Min-Min Scheduling Algorithm for Load balancing in Cloud Computing", National Conference on Parallel Computing Technologies, IEEE, ISBN 978-1-4799-1589-7, INSPEC Access No. 13822968, 2013, Pages 1-8.

[14]. P. Kowsik, K. Rajakumari, "A Comparative Study on Various Scheduling Algorithms in Cloud Environment", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, No. 11, ISSN(online) 2320-9801, ISSN(print) 2320-9798, 2014, Pages 6494-6500.

[15]. Hao Yuan, Changbing Li, Maokang Du, "Optimal Virtual Machine Resources Scheduling based on Improved Particle Swarm Optimization in Cloud Computing", Journal of Software, Vol. 9, No. 3, 2014.

[16]. S. Selvarani, G. Sudha Sadhasivam, "Improved Cost-Based Algorithm for Task Scheduling in Cloud Computing", Computational Intelligence and Computing Research (ICCIC) International Conference, IEEE, E-ISBN 978-1-4244-5967-4, ISBN 978-1-4244-5965-0, INSPEC Access No. 11822736, 2010, Pages 1-5.

[17]. Ritu Kapur and Maitreyee Dutta,"Review of various Load Balancing and Green Computing Techniques in Cloud", Journal of Basic and Applied Engineering Research, Vol. 2, No. 2, ISSN(online) 2350-0255, ISSN(print) 2350-0077, 2015, Pages 122-127.

[18]. Zhu, M., Wu, Q., Zhao, Y., "A Cost Effective Scheduling Algorithm for Scientific Workflows in Clouds", 31st International Performance Computing and Communications Conference (IPCCC), Proceeding of IEEE, ISSN 1097-2641, ISBN 978-1-4673-4881-2, INSPEC Access No. 13228073, 2012, Pages 256-265.

[19]. Ehab Nabiel Alkhanak, Sai Peck Lee, Reza Rezaei, Reza Meimandi Parizi, "Cost Optimization approaches for scientific workflow scheduling in cloud and grid computing: A Review, Classification, and Open Issues", Journal of Systems and Software, Science Direct, Vol. 113, 2015-16, Pages 1-26.

[20]. Simon Ostermann, Radu Prodan, and Thomas Fahringer,"Dynamic Cloud Provisioning for Scientific Grid Workflows", Proceedings of 2010 11th IEEE/ACM International Conference on Grid Computing, IEEE, ISSN 2152-1085, ISBN 978-1-4244-9347-0, INSPEC Access No. 11763399, 2010, Pages 97-104.

[21]. Maciej Malawski, Gideon Juve, Ewa Deelman and, Jarek Nabrzyski, "Cost and Deadline Constrained Provisioning for Scientific Workflow Ensembles in IaaS Clouds", In Proceedings of International Conference on High Performance Computing, Networking, Storage and Analysis, ACM, ISBN 1467308048, 2012.