International Journal of Engineering Research And Management (IJERM)
ISSN : 2349- 2058, Volume-01, Issue-05, August 2014

A Survey report about power consumption in cloud computing

Apurva Saxena, Sunita Gond

Abstract— **Cloud computing is a internet based service that provide us services like platform(Paas),software(Saas), nfrastructure as a service(Iaas) and everyone wants to take benefit of that cloud services using the cloud computing. As usage of cloud going to increase day by day, it ultimately increases the data size and loaded records on cloud servers. As size of data in data centers increases then it leads to increase in power consumption also. There are many types of equipment in data center that are not directly taking part in running the applications of cloud but support various applications of cloud. So power consumption can be reducing if resources can be properly managed. Here we prepare a paper about resource management approach to reduce power consumption**

Index Terms— Cloud computing, Iaas, Paas, Saas, data centers

I. INTRODUCTION

The Internet is commonly visualized as clouds, hence the term "cloud computing" stands for computation done through the Internet. With Cloud Computing users can access database resources via the Internet from anywhere, any time, without worrying about any maintenance or management of resources like hardware and software.

Cloud computing is a term used to describe platform, infrastructure and application. A cloud computing platform dynamically provisions, configures, reconfigures, and de-provisions servers as needed. Servers in the cloud can be physical machines or virtual machines. Rightnow clouds also include other computing resources such as storage area networks (SANs), network equipment, firewall and other security devices.

Manuscript received Aug 09, 2014

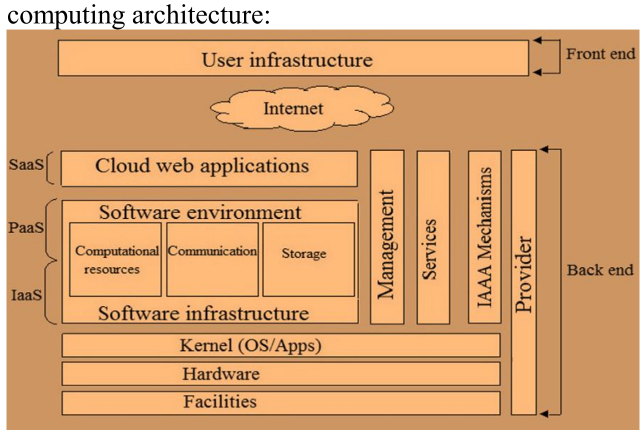
Apurva Saxena, Assistant Professor, Barkatullah University, Bhopal, India
Sunita Gond, Assistant Professor, Barkatullah University, Bhopal, India

a) Cloud computing architecture:-

The Cloud Computing architecture comprises of many cloud components, each of them is loosely coupled. We can broadly divide the cloud architecture into two parts:

1) Front End
2) Back End

The diagram shows the graphical view of cloud computing architecture:



Some Cloud Computing service providers: IBM, HP, Google, Microsoft, Amazon Web Services, Salesforce.com, NetSuite, VMware etc.

Examples of Cloud computing services include Google Docs, Office 365, Drop Box, SkyDrive etc.
Each of the ends is connected through a network, usually via Internet.

1. **Front End: -** Front End refers to the client part of cloud computing system. Without Graphical interface client cannot directly interact with any services of cloud. It consists of interfaces and applications that are required to access the cloud computing platforms, e.g., Web Browser.

2. **Back End: -** Back End refers to the cloud itself. It consists of all the resources required to provide cloud computing services. Data centers of cloud computing comprises of huge data storage, virtual machines, security mechanism, services, deployment models, servers, etc.

The services provided by cloud computing can be divided into three categories.

b) Service Models: - The service models are explained below:

Infrastructure as a service (IaaS): It involves the delivery of huge computing resources such as the capacity of storage, processing and network. It is the ability to remotely access computing resources. The major advantages of IaaS are pay per use, security, and reliability. IaaS is also known as hardware-as-a-service. An example of IaaS is the Amazon Elastic Compute Cloud (EC2).

Platform as a service (PaaS): supports a set of application programs interface to cloud applications. It has emerged due to the suboptimal nature of IaaS for cloud computing and the development of Web applications. Examples of PaaS are Google App Engine and Microsoft Azure [9].

Software as a service (SaaS): provides a service that is directly consumable by the end user. It is a software deployed over the Internet. This is a payas- you-go service. It seeks to replace the applications running on a PC. A typical example of SaaS is Salesforce.com [9].

c) Deployment Models:-

1. Private cloud: - The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises [10].

2. Community cloud: - The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises [10].

3. Public cloud: - The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider [10].

4.Hybrid cloud: - The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

II. Survey report about Energy conservation on cloud computing

Title: Risk and Energy Consumption Tradeoffs in Cloud Computing Service via Stochastic Optimization Models [1]
Author: Wang, Jue, and Siqian Shen

Technique Used: In this system stochastic integer programming models were used to optimize job scheduling and server consolidation. They developed three models to trade off risk for unsatisfied job requests and cost for energy consumption of machine. As far as uncertain job requests are considered this model vary QoS levels by either enforcing zero unsatisfied requests, or imposing a joint chance constraint to bound possible failures in a backlogging model with penalty for job incompleteness. The time based model minimizes the total energy consumption of all servers over time periods. In this backlogging of customer requests are not allowed along with optimizing active servers with their work load. A Job-based Model with Backlogging considers computational requests submitted by different customers through time. This model allows backlogging and jobs can be partitioned and scheduled on multiple servers in minimum time periods. A Backlogging Model with a Joint Chance Constraint used to investigate model variants of optimizing both sustainability and reliability in clouds.

Issues: They incorporated backlogging and penalty cost for unfinished requests into their models. By incorporating stochastic processes techniques and queuing theories they try to examined dynamic decisions and large-scale data uncertainty, and employ scalable stochastic optimization tools to derive real-time operations in Clouds.

Title: PoWER: prediction of workload for energy efficient relocation of virtual machines [2]
Author: Qazi, Kashifuddin, Yang Li, and Andrew Sohn

Technique Used: PoWER tries to predict the behaviors and nature of cluster based on its history intelligently. On the basis of history VMs are distributed in clusters and turn off unused machines. This will lead a huge energy saving. This model is inspired with concepts of Chaos Theory that make it indifferent to the type of loads and inherent cycles to other current prediction algorithms. This model is tested on test bed cluster and analyzed accordingly.

Issues: Their analysis shows that PoWER is much effective than another FFT-based time series method in predicting VM loads and freeing resources on Physical Machines for test loads. PoWER has a prediction accuracy of 89% as opposed to an FFT-based method which has an accuracy of 75% these results were obtained when 48 VMs were involved in result analysis [2].

Title: An Energy-Efficient Self-Provisioning Approach for Cloud Resources Management [3]
Author: Hanen Chihi, Walid Chainbi, Khaled Ghedira

Technique Used: This work presented resources provisioning methods based on an unsupervised predictor model in the form of an unsupervised recurrent neural network. This is based on a self-organizing map. Another unique feature of this work is a resources administration strategy for energy saving in the cloud. In this work they considered hard real-time systems with two renewable and nonrenewable energy sources. The two dynamic voltage scaling controllers are used to reduce the cost and also minimize the energy attained from the latter source.

Issues: To show the efficiency of the controllers they also implemented a hypothetical clairvoyant controller. The result of this controller is give78% of the reference performance. They are also developing a self optimization scheduling application in the cloud.

Title: Low Latency Energy Efficient Communications in Global Scale Cloud Computing Systems [4]
Author: T.H. Szymanski

Technique Used: Resource Reservations in the Future Internet model explores the use of a recently-proposed Future-Internet network which can establish guaranteed-rate connections between data-centers over the cloud, with strict latency and energy-efficiency guarantees. In this model, nodes can signal the network to reserve a guaranteed-rate connection between the Ingress and Egress routers. A Maximum-Flow Minimum-Energy routing algorithm is used to route the VPN trunks when the network is provisioned. The MPI inter-processor communications between VMIs in remote data-centers are then multiplexed onto these high-capacity trunks. In Future-Internet network each "trunk" between data centers is a connection with a guaranteed data-rate. The large and variable queueing delays of traditional Best-Effort Internet links can be eliminated, and the latency over the cloud can be reduced to near-minimal values, i.e., the fiber latency.

Issues: In this work, it is shown that many highly-bursty ON/OFF processes can be multiplexed onto
a single high-capacity trunk, and achieve significantly improved resource utilization, latency and energy-efficiency. Each destination node has Traffic Playback Queue (TPQ) to regenerate the original bursty ON/OFF stream(s) at the destination data- centers with improved latency and energy-efficiency.

Title: Experimental Analysis of Task-based Energy Consumption in Cloud Computing Systems [5]
Author: Feifei Chen, John Grundy, Yun Yang, Jean-Guy Schneider and Qiang He

Technique Used: In this model, the total energy consumption of a cloud workload is divided into a fixed part and a dynamic part. The fixed part of energy consumption includes the energy consumption during idle time and energy consumption of the cooling system. They treated a single task as a unit and measured the energy consumed by the task under various system configurations and task workload. To operationalise this model, extensive experiments to profile the energy consumption in cloud computing systems based on three types of tasks: computation-intensive, data-intensive and communication intensive tasks. Currently, they are determining the functional shape of our energy consumption model based on the experimental results analysis. They compare the energy consumption of individual task against the energy consumption of composite load predicted by our model. For computation-intensive tasks, they are selecting CPU-intensive and memory-intensive applications which inherently have single threaded and multi threaded algorithms instead of developing a new application.

Issues: Understanding cloud system energy consumption dynamics is valuable for developing efficient energy-aware resource management techniques for green cloud computing. To integrate energy cost rate and an “energy dirtiness rate” into our energy consumption model to factor in the costs – monetary and environmental – of cloud energy generated by different resources. This enhanced energy cost model will be investigated in order to minimize total energy costs while meeting system performance needs.

Title: Green cloud computing: Balancing energy in processing, storage, and transport [6]
Author: Baliga, Jayant, Robert WA Ayre, Kerry Hinton, and RodneyS Tucker.

Technique Used: in this work they explore the balance between server energy consumption along with network energy consumption and end-user energy consumption to present a fuller assessment of the benefits of cloud computing. They also presented a survey on energy consumption in cloud computing along with the energy consumption of conventional computing is the energy consumed when the same task is carried out on a standard consumer personal computer (PC) that is connected to the Internet but does not utilize cloud computing.

Issues: the number of file downloads per hour increases, the energy consumption in transport grows and storage as a service consumes more power than storage on local hard disk drives. The energy savings from cloud storage are minimal. According to result energy consumption in transport and switching can be a significant percentage of total energy consumption in cloud computing. Cloud computing can enable more energy-efficient use of computing power, especially when the users' predominant computing tasks are of low intensity or arise infrequently.

Title: Energy-Optimal Mobile Cloud Computing under Stochastic Wireless Channel [7]
Author: Zhang, Weiwen, Yonggang Wen, Kyle Guan, Dan Kilper, Haiyun Luo, and D. Wu.

Technique Used: in this work they adopted a canonical model that captures the essentials of a typical mobile application. Specifically, a mobile application is abstracted into a profile with two parameters, including:
Input data size L: the number of data bits as the input to the application;
Application completion deadline T: the delay deadline before which the application should be completed.
They investigate the problem of minimizing the energy consumption for executing an application in the mobile device. Since the energy consumed by CPU is much larger than the energy consumed by memory and screen, they only consider the computation energy of executing the application on mobile device. They also adopted a probabilistic performance requirement.

Issues: The additional computation energy is negligible when the application completion probability is very close to 1. As a result, the application completion probability is assumed to be very close to 1. In this work, they consider the problem of data scheduling to minimize the transmission energy under cloud execution, with a deadline constraint.

Title: Low-Emissions Routing for Cloud Computing in IP-over-WDM Networks with Data Centers [8]
Author: Gattulli, Mirko, Massimo Tornatore, Riccardo Fiandra, and Achille Pattavina

Technique Used: In this work various routing algorithms that are implemented are proposed. The routing algorithms provides the shortest path from source and the destination , but when these routing algorithms are implemented in the cloud environment for the access of the data centers are proposed.

Issues: in this model configurations with larger transport energy requirements (such as IP basic and IP-over-SDH) benefit less from renewable-energy-aware algorithms can be implemented in the future.

References

[1]Wang, Jue, and Siqian Shen "Risk and Energy Consumption Tradeoffs in Cloud Computing Service via Stochastic Optimization Models", In 2012 IEEE Fifth International Conference on Utility and Cloud Computing (UCC-2012), pp. 239-246, 2012.

[2]Qazi, Kashifuddin, Yang Li, and Andrew Sohn. "PoWER: prediction of workload for energy efficient relocation of virtual machines", In Proceedings of the 4th annual Symposium on Cloud Computing, article no. 31, ACM, 2013.

[3]Mohaqeqi, Morteza, Mehdi Kargahi, and Maryam Dehghan "Adaptive scheduling of real-time systems cosupplied by renewable and nonrenewable energy sources", ACM Transactions on Embedded Computing Systems (TECS), vol. 13, no. 1s, article no. 36, 2013.

[4]Szymanski, Ted H. "Low latency energy efficient communications in global-scale cloud computing systems." In Proceedings of the 2013 workshop on Energy efficient high performance parallel and distributed computing, pp. 13-22. ACM, 2013.

[5]Chen, Feifei, John Grundy, Yun Yang, Jean-Guy Schneider, and Qiang He. "Experimental analysis of task-based energy consumption in cloud computing systems." In Proceedings of the ACM/SPEC international conference on International conference on performance engineering, pp. 295-306. ACM, 2013.

[6]Baliga, Jayant, Robert WA Ayre, Kerry Hinton, and RodneyS Tucker. "Green cloud computing: Balancing energy in processing, storage, and transport." Proceedings of the IEEE 99, no. 1, pp. 149-167, 2011.

[7]Zhang, Weiwen, Yonggang Wen, Kyle Guan, Dan Kilper, Haiyun Luo, and D. Wu. "Energy-Optimal Mobile Cloud Computing under Stochastic Wireless Channel", pp.1-13, 2013.

[8]Gattulli, Mirko, Massimo Tornatore, Riccardo Fiandra, and Achille Pattavina. "Low-Emissions Routing for Cloud Computing in IP-over-WDM Networks with Data Centers." IEEE Journal on Selected Areas in Communications, vol. 32, no. 1, pp. 28-38, 2014.

[9]Matthew N.O. Sadiku, Sarhan M. Musa, and OMonowo D. MoMOh, Cloud computing: Opportunities and challenges, IEEE potentials, pp. 34-36, January/February 2014.

[10] NIST Cloud Computing Standards Roadmap, July 2013.