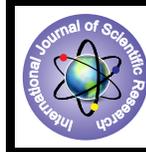


Cloud Computing: Energy Efficiency for Data Center Resources, Architectural Elements and Open Challenges



Engineering

KEYWORDS : Cloud computing Infrastructure , green cloud, data center, architectural elements and open challenges

Akshay Sharma

M.tech (C.S.E), Arni University, Kathgarh Indora

ABSTRACT

Cloud computing is offering utility-oriented IT services to users worldwide. Based on a pay-as-you-go model, it enables hosting of pervasive applications from consumer, scientific, and business domains. However, data centers hosting Cloud applications consume huge amounts of energy, contributing to high operational costs and carbon footprints to the environment. Therefore, we need Green Cloud computing solutions that can not only save energy for the environment but also reduce operational costs. This paper presents vision, challenges, and architectural elements for energy-efficient management of Cloud computing environments. We have validated our approach by conducting a set of rigorous performance evaluation study using the cloud Sim toolkit. Reducing energy consumption and emissions of carbon dioxide (CO₂) in data centers represent open challenges and driving the future research work for green data centers.

I. INTRODUCTION

Cloud computing (CC) is a promising area in distributed computing. This technology can achieve communication, storage, processing, high performance, hosting and services on demand to cloud customer. The growing of social applications and e-business need to increase the number of data centers. However, the combination of global warming and inconstant climate make the cost of energy a major challenge for the sustainability of the e-business [1]. Computer specialists expect that data centers technology is the optimal choices for next generation systems. Google managers and engineers expect that if energy consumed continues ascending, that drive to energy cost will be more than infrastructure cost. Energy consumption cost will influence to end user that pay as usage cloud resources and services. Also more power consumption requires more cooling, and that affects the environment in a negative way by producing more carbon dioxide (CO₂). Cloud computing is a binding form, for example, an IBM supercomputer consumes 20 MW which is almost equal to 22,000 US building energy consumption[2], and equivalent to 0.5% of the whole world's energy. Cloud computing deals and is compatible with different type of users demand and it has three types of access structure throw network, public cloud, private cloud and hybrid cloud. Cloud providers can offer access to any cloud services for cloud users through web interfaces, and they can manage and monitor computing resources such as hardware, software (application, operating systems), network and storage. Cloud computing provides three types of services, software (SaaS), platform (PaaS). Most of cloud operations, resources and physical locations cannot be seen or identified by users. Gartner's annual report acknowledged that the information and communication technology (ICT) costs on 2009 decreased by 5.2%, but the costs on 2010 have increased by 3.3% [3].

II. CLOUD COMPUTING INFRASTRUCTURE

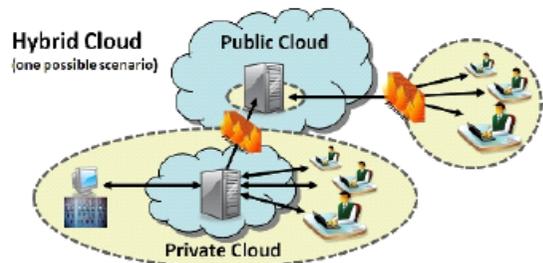


Fig2: Cloud computing infrastructure [7]

A. Private cloud

Also named company cloud, it is a marketing approach for a proprietary computing architecture that provides services like hosting, storage, servers and application. Private cloud can manage and control users because it serves a limited number of people behind a firewall. Private cloud has many features like virtualization and distributed computing power, that is has allowed company network and datacenter administrators to effectively meet the needs of their users inside the company. Organizations are moving to private cloud for many issues reduced energy cost, more IT performance and friendly business environment [10]

B. Public cloud

International Data Corporation (IDC) in annual report expect that Worldwide spending on public IT cloud services will be more than \$40 billion in 2013 and is expected to approach \$100 billion in 2016"[6]

C. Hybrid cloud

Fig2 show the structure of hybrid cloud structure. 2013 is the year that companies need to implement a hybrid cloud strategy that puts select workloads in the public cloud and keeps others in-local network[11]. The motivation of this review paper is to find gaps in previous study that concerned on green cloud to present these gaps to researchers for more development and support energy efficiency technique.

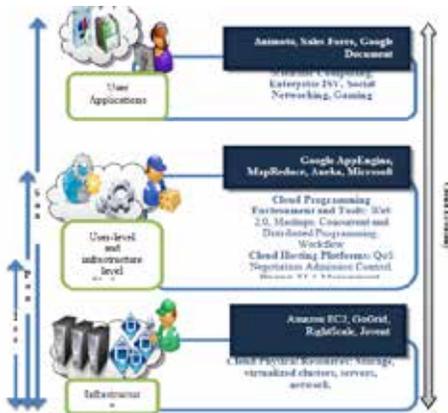


Fig1: Cloud computing layers [9].

Table 3.2: The Distribution of Power in data center

Cost	Component	Sub-Components
45%	Servers	CPU, memory, storage systems
25%	Infrastructure	Power distribution and cooling
15%	Power draw	Electrical utility costs
15%	Network	Links, transit, equipment

II. GREEN CLOUD COMPUTING

The worldwide agitation to achieve ecological, business and environmental sustainability is starting to redraw industrial landscape, table 1: show that the consumption of cloud components. The current status of global warming, ecological deterioration and the severity of its potential consequences explain the overwhelming popularity of environmental initiatives across the world. Environmental impact of Information Technology (IT) under the banner of "Green IT" has started being discussed by academia, media and government. Since 2007 when the Environmental Protection Agency (EPA) submitted a report to the US Congress [2] about the expected energy consumption of data centers. The main objective of Green IT is to increase energy efficiency and reduce CO2emissions [13]. To make cloud greener there are two ways, first improve energy efficiency of cloud, second use clean energy supply. Table 2 show the energy consumption in different level in cloud computing. Power Usage Effectiveness (PUE) measures if the data center is green or not by measure the efficiency of data center [14].

PUE = Total Facility Power / IT Equipment Power (1)

Total facility power= power + cooling IT equipment power= server power +storage power + communication power

Green computing or green IT, refers to environmentally sustainable computing or IT. It is "the study and practice of designing, manufacturing, using, and disposing of computers, servers, and associated subsystems—such as monitors, printers, storage devices, and networking and communications systems—efficiently and effectively with minimal or no impact on the environment. Green IT also strives to achieve economic viability and improved system performance and use, while abiding by our social and ethical responsibilities. Thus, green IT includes the dimensions of environmental sustainability, the economics of energy efficiency, and the total cost of ownership, which includes the cost of disposal and recycling.

III. DATA CENTER

Data center is a cornerstone of the infrastructure of cloud computing approach by which a variety of information technology (IT) services are built. They extend the ability of centralized repository for computing, hosting, storage, management, monitoring, networking and deployment of data. With the rapid increase in using data centers, there is a continuous increase in the energy consumption[14]. Data center beside consumed energy also produces carbon dioxide and that riddled with IT inefficiencies. Data center major components are thousands of servers; however these servers consumed huge energy without performing useful work. There are many models to support energy efficiency in data center; the most important model is virtualization. Cloud computing supports virtualization like resources (i.e. computes, storage, and network capacity). The most basic one is at Virtual Machine (VM) level where different applications can be executed within their containers or operating systems running on the same machine hardware In data centers, the number of physical machines can be reduced using virtualization by consolidating virtual appliances onto shared servers. This can help to improve the efficiency of IT systems. The advantages are simple, it allows multiple virtual machines to be run on a single physical machine in order to provide more capability and increase the utilization level of the hardware. It always increases efficiency; it allows you to do more work with less IT equipment [2].

IV. GREEN CLOUD COMPUTING

The worldwide agitation to achieve ecological, business and environmental sustainability is starting to redraw industrial landscape, table 1: show that the consumption of cloud components. The current status of global warming, ecological

deterioration and the severity of its potential consequences explain the overwhelming popularity of environmental initiatives across the world. Environmental impact of Information Technology (IT) under the banner of "Green IT" has started being discussed by academia, media and government. Since 2007 when the Environmental Protection Agency (EPA) submitted a report to the US Congress [2] about the expected energy consumption of data centers. The main objective of Green IT is to increase energy efficiency and reduce CO2emissions [13]. To make cloud greener there are two ways, first improve energy efficiency of cloud, second use clean energy supply. Table 2 show the energy consumption in different level in cloud computing. Power Usage Effectiveness (PUE) measures if the data center is green or not by measure the efficiency of data center [14]. PUE = Total Facility Power / IT Equipment Power (1)

Cloud computing has different techniques to solve energy-efficient problem to minimize the impact of Cloud computing on the environment

V. Green Cloud Architectural Elements

The aim of this paper is to addresses the problem of enabling energy-efficient resource allocation, hence leading to Green Cloud computing data centers, to satisfy competing applications' demand for computing services and save energy. Figure 3 shows the high-level architecture for supporting energy-efficient service allocation in Green Cloud computing infrastructure. There are basically four main entities involved:

a) Consumers/Brokers:

Cloud consumers or their brokers submit service requests from anywhere in the world to the Cloud. It is important to notice that there can be a difference between Cloud consumers and users of deployed services.

Green Resource Allocator:

Acts as the interface between the Cloud infrastructure and consumers. It requires the interaction of the following components to support energy-efficient resource management:

1) Green Negotiator:

Negotiates with the consumers/brokers to finalize the SLA with specified prices and penalties (for violations of SLA) between the Cloud provider and consumer depending on the consumer's QoS requirements and energy saving schemes.

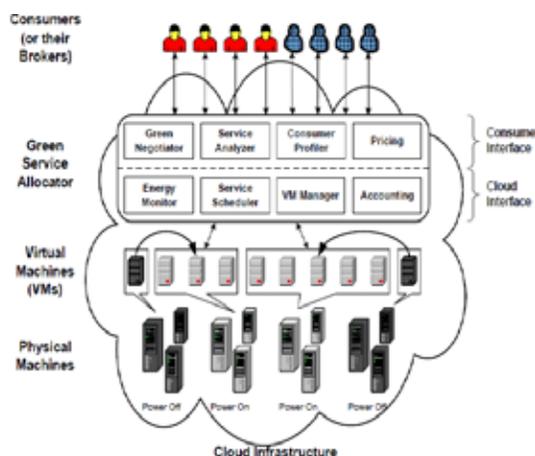


Figure 1: High-level system architectural framework

Fig :-3 High Level system architectural Framework.

2) Service Analyser:

Interprets and analyses the service requirements of a sub-

mitted request before deciding whether to accept or reject it. Hence, it needs the latest load and energy information from VM Manager and Energy Monitor respectively.

3) Consumer Profiler:

Gathers specific characteristics of consumers so that important consumers can be granted special privileges and prioritised over other consumers.

4) Pricing:

Decides how service requests are charged to manage the supply and demand of computing resources and facilitate in prioritising service allocations effectively.

5) Energy Monitor:

Observes and determines which physical machines to power on/off.

6) Service Scheduler:

Assigns requests to VMs and determines resource entitlements for allocated VMs. It also decides when VMs are to be added or removed to meet demand.

7) VM Manager:

Keeps track of the availability of VMs and their resource entitlements. It is also in charge of migrating VMs across physical machines.

8) Accounting:

Maintains the actual usage of resources by requests to compute usage costs. Historical usage information can also be used to improve service allocation decisions.

a) VMs:

Multiple VMs can be dynamically started and stopped on a single physical machine to meet accepted requests, hence providing maximum flexibility to configure various partitions of resources on the same physical machine to different specific requirements of service requests. Multiple VMs can also concurrently run applications based on different operating system environments on a single physical machine

VI. Open Challenges

In this section, we identify key open problems that can be addressed at the level of management of system resources. Virtualization technologies, which Cloud computing environments heavily rely on, provide the ability to transfer VMs between physical nodes using live or offline migration. This enables the technique of dynamic consolidation of VMs to a minimal number of nodes according to current resource requirements.

6.1 Energy-aware Dynamic Resource Allocation :-

Recent developments in virtualization have resulted in its proliferation of usage across data centers. By supporting the movement of VMs between physical nodes, it enables dynamic migration of VMs according to QoS requirements. When VMs do not use all provided resources, they can be logically resized and consolidated on a minimal number of physical nodes, while idle nodes can be switched off. Currently, resource allocation in a Cloud data center aims to provide high performance while meeting SLA, without a focus on allocating VMs to minimize energy consumption.

6.2 QoS-based Resource Selection and Provisioning :-

Data center resources may deliver different levels of performance to their clients; hence, QoS-aware resource selection plays an important role in Cloud computing. Additionally, Cloud applications can present varying workloads. It is therefore essential to carry out a study of Cloud services and their

workloads in order to identify common behaviors, patterns, and explore load forecasting approaches that can potentially lead to more efficient resource provisioning and consequent energy efficiency. In this context, we will research sample applications and correlations between workloads, and attempt to build performance models that can help explore the trade-offs between QoS and energy saving. Further, we will investigate a new online approach to the consolidation strategy of a data center that allows a reduction in the number of active nodes required to process a variable workload without degrading the offered service level.

6.3 Optimisation of Virtual Network Topologies :-

In virtualized data centers VMs often communicate between each other, establishing virtual network topologies. However, due to VM migrations or non-optimized allocation, the communicating VMs may end up hosted on logically distant physical nodes providing costly data transfer between each other. If the communicating VMs are allocated to the hosts in different racks or enclosures, the network communication may involve network switches that consume significant amount of power. To eliminate this data transfer overhead and minimize power consumption, it is necessary to observe the communication between VMs and place them on the same or closely located nodes. To provide effective reallocations, we will develop power consumption models of the network devices and estimate the cost of data transfer depending on the traffic volume.

6.4 Efficient Consolidation of VMs for Managing Heterogeneous Workloads:-

Cloud infrastructure services provide users with the ability to provision virtual machines and allocate any kind of applications on them. This leads to the fact that different types of applications (e.g., enterprise, scientific, and social network applications) can be allocated on one physical computer node. However, it is not obvious how these applications can influence each other, as they can be data, network or compute intensive thus creating variable or static load on the resources. The problem is to determine what kind of applications can be allocated to a single host that will provide the most efficient overall usage of the resources. Current approaches to energy efficient consolidation of VMs in data centers do not investigate the problem of combining different types of workload. A compute intensive (scientific) application can be effectively combined with a web-application (file server), as the former mostly relies on CPU performance, whereas the latter utilizes disk storage and network bandwidth. We will investigate which particular kind of applications can be effectively combined and what parameters influence the efficiency; and develop resource allocation algorithms for managing them. This knowledge can be applied to energy efficient resource management strategies in data centers to achieve more optimal allocation of resources and, therefore, improve utilization of resources and reduce energy consumption.

VII. CONCLUSION AND FUTURE WORK

As a result of the investigated literature review, we concluded that the previous techniques and approaches lack several features like QoS and performance against energy efficiency. Additionally, the time complexity and the reduction of the energy consumption are not highly effective. Based on our literature review we realized that the lack of and need for an integrated data center energy efficiency framework which consider the social network applications as a vital related factor in elevating energy consumption, as well as high potential for energy efficiency. The framework provides a platform on top of which the Green Cloud could be built. The framework practices from Energy Aware Computing will improve the efficiency of Cloud systems and their data centers and Clouds

themselves will produce naturally efficient and focused centers of computation, advancing the pursuit of green computing. The required integrated data center energy efficiency framework should be also applicable in different types of data centers including public, private and hybrid. The existence of such framework will offer a great powerful capability to deal with service levels and resources management. The required data center Energy Efficiency framework will offer improved in scalability, elasticity, simplicity for management, delivery of cloud services and better reduction in data centers energy consumption taking into consideration the QoS for the user services.

REFERENCE

- [1] Mell, P. and T. Grance. The NIST Definition of Cloud Computing, 2009. | [2] James W. Smith, Green Cloud A literature review of Energy-Aware Computing, 2011. | [3] Gartner, Inc. (NYSE: IT), annual report, 2012. <http://www.gartner.com/DisplayDocument?id=914826>. | [4] Ian Foster, Cloud Computing and Grid Computing 360-Degree Compared, 2008. | [5] Chase, J.S., Anderson, D.C., Thakar, P.N., Vahdat, A.M., and Doyle, R.P. 2001. Managing energy and server resources in hosting centers. | [6] Chunmei Chi, The Trend of Cloud Computing in China, 2011. | [7] Raj kumar Buyya, Energy-Efficient Management of Data Center Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges, 2010. | [8] Hybrid Clouds for Real-Time Systems, 2012. <http://real-timecloud.com/hybrid-cloud/hybrid-clouds-for-real-time-systems/>. | [9] Chris Morris, Cloud Services and Technologies End-User Survey, 2011. | [10] Christine Burns, Hybrid clouds, cloud brokers, big data and software-defined networking (SDN) predicted to be the major trends (SDN) predicted to be the major trends in cloud computing in 2013, 2012. | [11] Anton Beloglazov, A Survey on Power Management Solutions for Individual Systems and Cloud, 2010. | [12] AsgharSabbaghi, Green Information Technology and Sustainability: A Conceptual Taxonomy, 2012. | [13] Nguyen Quang Hung, Performance constraint and power-aware allocation for user requests in virtual computing, 2011. | [14] MueenUddin, Green Information Technology (IT) framework for energy efficient data centers using virtualization, 2012. |