Investigation on Efficient Management of workflows in cloud computing Environment

M.Sudha
Assistant Professor (Senior)
School of Information Technology & Engineering
VIT University
Vellore, Tamil Nadu, India

M.Monica
Assistant Professor
School of Computing Science & Engineering
VIT University
Vellore, Tamil Nadu, India

***Abstract*—Cloud computing is an on-demand service model often based on virtualization technique and this paper explores the use of cloud computing for scientific workflows, focusing on a widely used application. The approach is to evaluate from the point of view of a scientific workflow the tradeoffs between running in a local environment, if such is available, and running in a virtual environment via remote, wide-area network resource access. Our results show that a workflow with short job runtimes, the virtual environment can provide good compute time performance but it can suffer from resource scheduling delays and wide area communications**.

Keywords-work flow; virtualization;performance;WAN;

I. INTRODUCTION

The appetite for a new generation of network-based applications both for consumers e.g. Twitter, Face book, YouTube, Hulu, Animoto and for businesses e.g. Web Mail, Google Docs, Zoho, is driving the need to reorganize current datacenter infrastructure for massive scale. Another characteristic of the new web-based network applications is wildly fluctuating demand especially in the mass consumer market. This need is stretching current IT architectures to their limits in terms of the ability to ensure on-demand service availability, reliability, performance and security at a reasonable cost. While demand for network services is soaring, the economic pressure to do "more with less" is also rising. With virtualization technologies becoming more accepted, public and private cloud networks are emerging as an attractive means for sharing compute, storage and network resources amongst multiple service developers and service delivery applications. Such a sharing of resources immediately provides economies of scale through consolidation, energy savings and improved resource utilization. More importantly, the ability to dynamically reallocate resources using Virtualization technologies can help mitigate the need for additional investment in infrastructure to meet sudden spikes in demand by temporarily diverting existing resources from
Low-priority business applications to high priority business applications.

II. LITERATURE SUTVEY

We can conclude that the skyrocketing demand for a new generation of cloud-based consumer and business applications is driving the need for next generation of datacenters that must be massively scalable, efficient, agile, reliable and secure. The basic requirements for a cloud computing environment are In order to scale cloud services reliably to millions of service developers and billions of end users the next generation cloud computing and datacenter infrastructure will have to follow an evolution similar to the one that led to the creation of scalable telecommunication networks.

In the future network-based cloud service providers will leverage virtualization technologies to be able to allocate just the right levels of virtualized compute, network and storage resources to individual applications based on real-time business demand while also providing full service level assurance of availability, performance and security at a reasonable cost.

A key component Virtual Resource Mediation Layer (VRML) must be developed through industry collaboration to enable interoperability of various public and private clouds. This layer will form the basis for ensuring massive scalability of cloud infrastructure by enabling distributed service creation, service delivery and service assurance without any single vendor domination.

The next generation virtualization technologies must allow applications to dynamically access CPU, memory, bandwidth and storage (capacity, I/O and throughput) in a manner similar to that of the telecommunications 800 Service Call Model1 with one level of indirection and mediation.

III. SYSTEM MODEL

Cloud computing system model can be divided it into two sections: the front end and the back end. They connect to each other through a network, usually the Internet. The front end is the side the computer user, or client, sees. The back end is the "cloud" section of the system. The front end includes the client's computer (or computer network) and the application required to access the cloud computing system. Not all cloud computing systems have the same user interface. Services like Web-based e-mail programs leverage existing Web browsers like Internet Explorer or Firefox. Other systems have unique applications that provide network access to clients. Back end of the system are the various computers, servers and data storage systems that create the "cloud" of computing services. In theory, a cloud computing system could include practically any computer program you can imagine, from data processing to video games. Usually, each application will have its own dedicated server. A central server administers the system, monitoring traffic and client demands to ensure everything runs smoothly. It follows a set of rules called protocols and uses a special kind of software called middleware. Middleware allows networked computers to communicate with each other.

This model does not seem to address end-to-end management. Ultimately, the cloud service infrastructure must provide end-to-end service assurance to meet both service creation and service delivery platform user requirements. The service creators must be able to develop services rapidly using reusable and collaborating service components available globally. The infrastructure must also accommodate billions of users globally who will contribute to wildly fluctuating workloads.

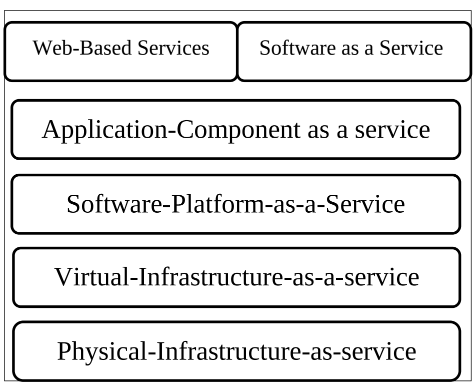


Figure 1. A Simple Cloud Architecture

Cloud computing represents a commercialization of these developments. Prior to cloud computing, acquiring such resources — the initial capital investment in purchasing the computers themselves and the significant resources devoted to maintaining the infrastructure — was an expensive and unlikely proposition for organizations and simply impossible for individuals. Now, cloud computing has the potential to benefit both providers and users. Cloud providers gain additional sources of revenue and are able to commercialize their large data centers and the expertise of large–scale data management. Overall cost is reduced through consolidation, while capital investment in physical infrastructure is amortized across many customers. Individual cloud users can store, access, and share information in previously inconceivable ways. Organizational cloud users no longer have to worry about purchasing, configuring, administering, and maintaining their own computing infrastructure, which allows them to focus on their core competencies.

IV. Problem Definition

Current cloud evolution is limited to the following three areas:

1) The Virtualization of servers, load balancers, and some server IP address management services.

2) The replacement of SAN infrastructure with large commodity server farms that support virtual applications using Direct Attached Storage (DAS) or File Systems (distributed or otherwise).

3) Efficient management of workflows in a cloud environment to allow fast scaling up and scaling down.

   a) Storing scalability/ compressibility options for every node in the workflow

   b) Input events and output events of every node in workflow

A. *Transparency*

a) An activity is a discrete step in a business process (workflow).
b) Activities range from calling a remote service to perform a task, e.g. calculating taxes, performing currency conversions, looking up inventory, to custom-defined services.
c) Actual Implementation" of services obscured an another version of virtualization
d) Transparent load-balancing and application delivery
e) Solution to be automated and integrated in workflow process.

B. *Scalability*

a) Scale up and build "mega data centers"
b) Need configuration or re-architecting
c) Potential of interrupting services is huge
d) Ability to transparently scale the service infrastructure and the solution
e) On-demand, real time scaling
f) Control node provides dynamic application scalability
g) Integration with virtualization solution or orchestration with workflow process to manage provisioning.

C. *Intelligent monitoring*

a) Control node – intelligent monitoring capabilities
b) Server overwhelming or application performance affected by network conditions – behavior outside accepted norms
c) More than knowing *when* a service in trouble *what* action should be taken
d) Example – application responding slowly, adjust application requests add more server if required.

D. *Capacity management*

a) From buckets to rivers
b) Constrained set of resources – predict peak usage and have in-house data centre to manage them
c) Not upper limit of computing power but speed at which new services can be provisioned and put into production
d) Initiate new system, transfer data, connect existing system, test combined system, manage complete life cycle

Cloud computing the creation of large data centers that can be dynamically provisioned, configured, and reconfigured to deliver services in a scalable manner and places enormous capacity and power in the hands of users. As an emerging new technology, however, cloud computing also raises significant questions about resources, economics, the environment, and the law. Many of these questions relate to geographical considerations related to the data centers that underlie the clouds: physical location, available resources, and jurisdiction.

V. OPERATION ASPECT OF THE WORK PROCEDURE

In this we have studied and analyzed the cloudsim architecture and studied how to simulate the cloudsim toolkit and have derived comparison graphs based on the detailed case study

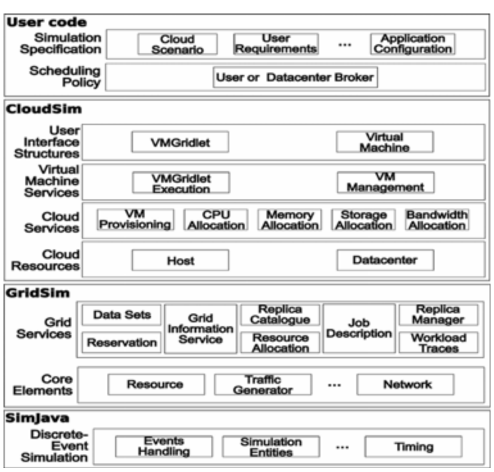


Figure 2. CloudSim Architecture

CloudSim: - CloudSim is a software framework and its architectural components are shown in the diagram given.

Simjava-At the lowest layer is the SimJava discrete event simulation engine that implements the core functionalities required for higher-level simulation frameworks such as queuing and processing of events, creation of system components (services, host, data center, broker, virtual machines), communication between components, and management of the simulation clock.

Gridsim:- Grid sim toolkit that support high level software components for modeling multiple Grid infrastructures, including networks and associated traffic profiles, and fundamental Grid components such as the resources, data sets, workload traces, and information services.

The CloudSim is implemented at the next level by programmatically extending the core functionalities exposed by the GridSim layer. CloudSim:-It provides novel support for modeling and simulation of virtualized Cloud based data center environments such as dedicated management interfaces for VMs, memory, storage, and bandwidth.

CloudSim layer manages the instantiation and execution of core entities (VMs, hosts, data centers, application) during the simulation period.

This layer is capable of concurrently instantiating and transparently managing a large scale Cloud infrastructure consisting of thousands of system components. Finer details related to the fundamental classes of CloudSim.

Datacenter: - This class models the core infrastructure level services (hardware, software) offered by resource providers in a Cloud computing environment.

Datacenter Broker: - This class models a broker, which is responsible for mediating between users and service.

SANStorage: - This class models a storage area network that is commonly available to Cloud-based data centers for storing large chunks of data.

Virtual Machine:- This class models an instance of a VM, whose management during its life cycle is the responsibility of the Host component.

Cloudlet:-This class models the Cloud-based application services (content delivery, social networking, business workflow), which are commonly deployed in the data centers.

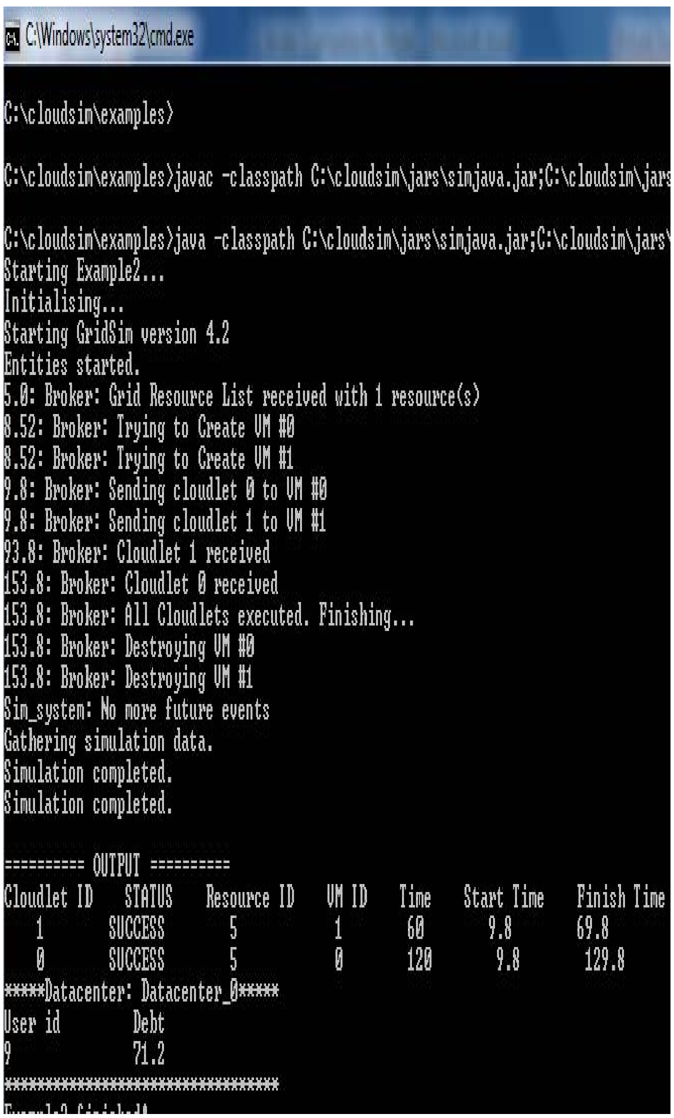
BW Provisioner. This is an abstract class that models the provisioning policy of bandwidth to VMs that are deployed on a Host component.

Memory Provisioner:-This is an abstract class that represents the provisioning policy for allocating memory to VMs.

VMM Allocation Policy: - This is an abstract class implemented by a Host component that models the policies (space-shared, time-shared) required for allocating processing power to VMs.

User Code: The top-most layer in the simulation stack is the user code that exposes configuration related functionalities for hosts (number of machines, their specification and so on, applications (number of tasks and their requirements), VMs, number of users and their application types, and broker scheduling policies.

VI. SIMULATION RESULTS



```
C:\Windows\system32\cmd.exe

C:\cloudsim\examples>

C:\cloudsim\examples>javac -classpath C:\cloudsim\jars\simjava.jar;C:\cloudsim\jars

C:\cloudsim\examples>java -classpath C:\cloudsim\jars\simjava.jar;C:\cloudsim\jars\
Starting Example2...
Initialising...
Starting GridSim version 4.2
Entities started.
5.0: Broker: Grid Resource List received with 1 resource(s)
8.52: Broker: Trying to Create VM #0
8.52: Broker: Trying to Create VM #1
9.8: Broker: Sending cloudlet 0 to VM #0
9.8: Broker: Sending cloudlet 1 to VM #1
93.8: Broker: Cloudlet 1 received
153.8: Broker: Cloudlet 0 received
153.8: Broker: All Cloudlets executed. Finishing...
153.8: Broker: Destroying VM #0
153.8: Broker: Destroying VM #1
Sim_system: No more future events
Gathering simulation data.
Simulation completed.
Simulation completed.

========= OUTPUT =========
Cloudlet ID    STATUS    Resource ID    VM ID    Time    Start Time    Finish Time
    1          SUCCESS       5            1       60        9.8          69.8
    0          SUCCESS       5            0       120       9.8          129.8
*****Datacenter: Datacenter_0*****
User id        Debt
9              71.2
********************************
```

A. *Functionality of CloudSim Toolkit*

CloudSim toolkit covers most of the activities taking place within a Data Center in detail.

a) Simulating Data Center hardware definition in terms of physical machines composed of processors, storage devices, memory and internal bandwidth
b) Simulating virtual machine specification, creation and destruction
c) The management of virtual machines, allocation of physical hardware resources for the operation of virtual machines based on different policies (e.g. time-shared and spaceshared)
d) Simulating the execution of user programs or requests (Cloudlet/Gridlet) on the virtual machines
e) These features are directly used in the Cloud Analyst.

B. *Comparison Graphs Based on Case Study*

Comparing the results for different input and output and we have obtained the following result for efficient management of workflows using different parameters like time, speed and finally inferring the performance and scalability in different environments.

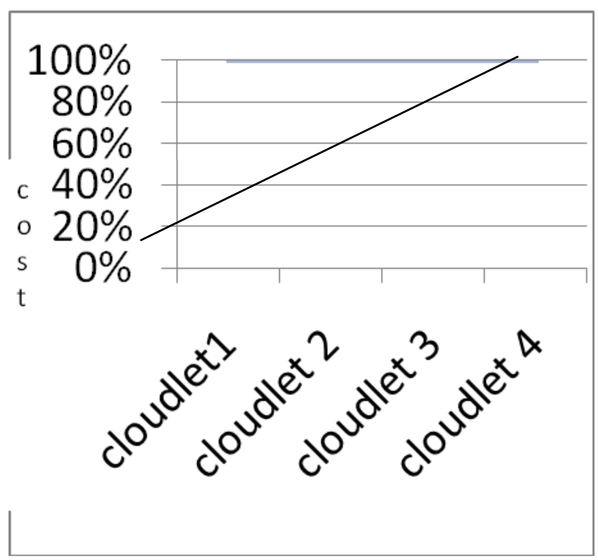


Figure 3. Graph between cloudlet and cost

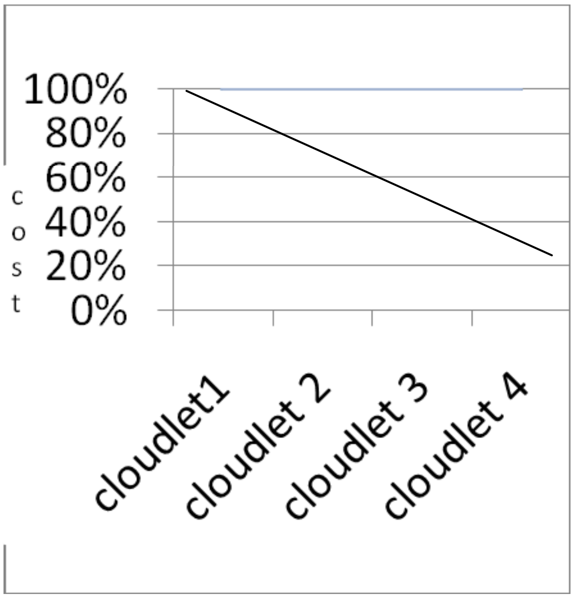


Figure 4.Graph Between Cloudlet & Average response time

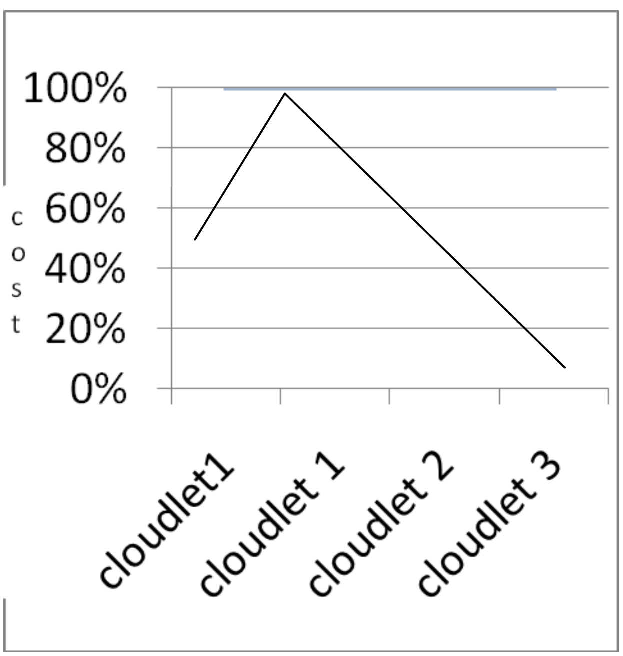


Figure 5. Average time spent on processing

REFERENCES

[1] Christina Hoffa, Gaurang Mehta, Timothy Freeman, Ewa Deelman, Kate Keahey, Bruce Berriman, John Good (2005) "On the Use of Cloud Computing for Scientific WorkflowsSudipto guha"

[2] David R. Cheriton(2007)," CloudWF: A Computational Workflow System for Clouds Based on Hadoop."

[3] Dr. Rao Mikkilineni & Vijay SarathyNan (2008) "Cloud Computing and the Lessons from the Past"

[4] D.J. Solove, "A Taxonomy of Privacy", University of Pennyslavania Law Review, vol 154, January 2006, p. 477. http://papers.ssrn.com/sol3/papers.cfm

[5] Guy Bunker, Farnam Jahanian, Aad van Moorsel, Joseph Weinman," Dependability in the cloud: Challenges and opportunities", IEEE 2009

[6] Ivona Brandic" Towards Self manageable Cloud services" 0730-3157/09, 2009 IEEE

[7] John Harauz, Lori M. Kaufman, Bruce Potter, "Data security in the world of cloud computing ", 2009 IEEE CO Published by the IEEE Computer and Reliability Societies

[8] Lizhe Wang, Jie Tao, Marcel Kunze , Alvaro Canales Castellanos, David Kramer, Wolfgang Karl "scientific Cloud computing: early Definition and Experience", 2008 IEEE

[9] Lijun Mei, W.K.Chan, T.H.Tse, "A Tale of Clouds: Paradigm comparisons and some thoughts on research issues", 2008 IEEE Asia-Pacific Services Computing Conference

[10] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia. *Above the Clouds: A Berkeley View of Cloud Computing*. University of California at Berkley, USA. Technical Rep UCB/EECS-2009-28, 2009.

[11] Meiko Jensen, Jorg Schwenk, Nils Gruschka, Luigi Lo Iacono" On technical security issues in Cloud computing" 978-0-7695-3840-2/092009, IEEE Computer Society

[12] Rajkumar Buyya1, Rajiv Ranjan2 and Rodrigo N. Calheiros "Modeling and simulation of scalable cloud computing Environments and the CloudSim Toolkit: challenges and opportunities" 978-1-4244-4907-1/09, 2009 IEEE

AUTHORS PROFILE

M.Sudha received the Bachelor of Engineering in Computer Science in 2000 from Madras University and Masters of Engineering in 2006 from Anna University and pursuing PhD in Computer Science at VIT. She is as Assistant Professor Senior in school of Information Technology and Engineering at VIT. Her research interests include Cloud computing, Distributed Systems and security architectures for cloud.

M.Monica received the Bachelor of Technology in Information Technology in 2007 from Anna University and Masters of Engineering in Computer Science in 2009 from Anna University; she is as Assistant Professor in the School of Computing Sciences and Engineering at VIT, Her research interests include Distributed Computing and operating systems.