A Comprehensive Survey on Cloud Computing

Md. Imran Alam
School of Computer Engineering of Kiit University, Bhubaneswar, India
E-mail: imranirfan89@gmail.com

Manjusha Pandey
School of Computer Engineering of Kiit University, Bhubaneswar, India
E-mail: manjushapandey82@gmail.com

Siddharth S Rautaray
School of Computer Engineering of Kiit University, Bhubaneswar, India
E-mail: sr.rgpv@gmail.com

Abstract- Cloud Computing is an innovation ideas that helps in reducing the computing cost. Cloud Computing offers better computing through improved utilization and reduced administration and infrastructure costs. Cloud computing is the long-held dream of computing as a utility. Cloud Computing is the combination of Software as a Service (SaaS) and Utility Computing. Cloud computing shares characteristics with autonomic computing, peer to peer, grid computing, client server model, mainframe computer and utility computing. It has various open source resources which gives different platform for better computing utilization. Cloud computing are managed by Cloud Management tools, loaded and tested by various other software testing tools. Cloud computing modelling and simulation is done by CloudSim or SPECI or GroundSim or DCSim on the basis of testing benchmark. The application of Cloud Computing is discussed.

Index Terms- Cloud Computing, SaaS, Utility Computing, CloudSim, SPECI, GroundSim, DCSim

I. INTRODUCTION

Cloud computing describes about different types of computing concepts that involve a large number of computers connected through a real time communication network such as the Internet. Cloud computing is similar like distributed computing over a network which means it has ability to run a program on many connected computers at the same time. Cloud computing mainly relies on the sharing of resources for achieving coherence and economies of scale like utility computing. The foundation of cloud computing is the broader concept mainly converged infrastructure and shared services.

Buyya et al. have explained the cloud as follows: "Cloud is a parallel and distributed computing system consisting of a collection of inter connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service level agreement (SLA) established through negotiation between the service provider and consumers."

The cloud has reached into our daily life and led to a broader range of innovations, but people often misunderstand what cloud computing is. Built on many old IT technologies, cloud computing is actually an evolutionary approach that completely changes how computing services may produce, priced and delivered. It allows to access services that reside in a distant datacentre, other than local computers or other Internet-connected devices. Cloud services are charged based on the amount consumed by worldwide users. Such an idea of computing as a utility is a long-held dream in the computer corporate, but it is still immature until the advent of low-cost datacentres that will enable this dream to come true.

Datacentres, behaving as "cloud providers", is computing infrastructures which provide many kinds of agile and effective services to customers. A broad range of IT companies including Amazon, Cisco, Salesforce, Yahoo, Facebook, Microsoft and Google, Cisco have their own datacentres and provide pay-as-you-go cloud services. Two different but similar types of cloud service should be identified first. One is on-demand computing instance and the other is on-demand computing capacity.

The advancement of Cloud computing came up due to fast growing usage of internet among the citizenry. The cloud computing is not a totally new technology but it is essentially a journey through distributed, cluster, grid and now cloud computing. In fact, in surge of rapid usage of internet all over the globe, cloud computing has already been steered in the IT industry.

The paper contain analogous system, cloud computing basics, open source resource of cloud, cloud computing services, cloud management tools, cloud computing simulations and modelling. Even cloud testing benchmark and application of the cloud is discussed.

II. ANALOGOUS SYSTEM

Cloud computing has related characteristics with:
Autonomic Computing- Computer systems which is capable of managing itself means self-management.
Client Server model- Client server computing mainly refers to any distributed application that differentiate between service providers (servers) and service requesters (clients).

Grid Computing- A form of distributed and parallel computing where by a super and virtual information processing system is composed of a cluster of network loosely coupled computers acting in concert to perform very large tasks.

Mainframe Computer- Powerful computer used mainly by large organizations for critical applications, typically bulk data processing such as census, industry and consumer statics, police and secret intelligence services, enterprise resource planning and financial transaction processing.

Utility Computing- The [1] packaging of computing resources such as computation and storage as a metered service similar to a traditional public utility such as electricity.

Peer-to-peer- Distributed architecture without the need for central coordination, with participants being at the same both suppliers and consumers of resources.

III. CLOUD COMPUTING BASICS

Since 2007, the term Cloud has become one of the most buzz words in IT industry. Dozens of researchers seek to define cloud computing from different application aspects, but there is no exact definition on it. Among the various definitions, we choose three widely quoted as follows

- *I. Foster*: "A large-scale [2] distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over Internet."

As an academic representative, Foster focuses on various technological features that differentiate cloud computing from other distributed computing paradigms. For example, computing entities are virtualized and delivered as services, and these services are dynamically driven by economies of scale.

- *Gartner*: "A style of computing [2] where scalable and elastic IT capabilities are provided as a service to multiple external customers using Internet technologies."

Garter is an IT consulting company, so it examines the qualities of cloud clouding mostly from the point of view of the industry. Functional characteristics are emphasized in this definition, such as whether cloud computing is scalable, elastic, service offering and Internet based.

- *NIST*: "Cloud computing [2]is a model for enabling 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."

A. System Architecture

Clouds are usually referred to as a large pool of computing and storage resources which can be accessed via standard protocols with an abstract interface.
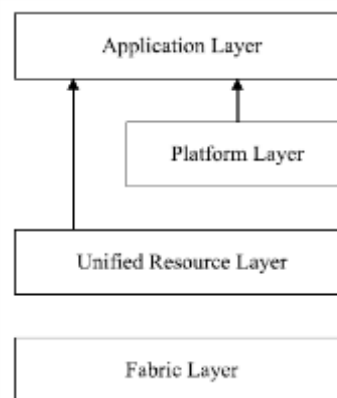

Fig. 1. System Architecture

The fabric layer contains the raw hardware level resources, such as compute resources, memory resources, and network resources. On the unified resource layer, resources have been virtualized so that they can be exposed to upper layer and end users as integrated resources. The platform layer adds to a collection of specialized tools, services and middleware on top of the unified resources to provide a development and deployment platform. The application layer includes the applications that would prevail in the clouds.

B. Characteristics of Cloud Computing

There are five necessary characteristics of Cloud Computing which explains their relation and difference from the traditional computing.

- *On-demand-self-service*

The consumer can provision or un-provision the services when needed, without the human interaction with the service provider.

- *Broad Network Access*

It has capabilities over the network and access done through standard mechanism.

- *Resource Pooling*

The computing resources of the provider are pooled to serve multiple consumers which are using a multi-tenant model, with various physical and virtual resources dynamically assigned, depending on consumer demand.

- *Rapid Elasticity*

Services can be rapidly and elastically provisioned.

- *Measured Service*

Cloud Computing systems automatically check and optimize resource usage by providing a metering capability to the type of services (e.g. storage, bandwidth, processing, or active user accounts)

C. Cloud Deployment Model

Clouds are deployed in different modes, depending on the usage scopes. There are four primary cloud deployment models.

- *Public Cloud*
- Cloud infrastructure available for general public and general usages.
- *Private Cloud*
- The cloud available mainly for a single organization and provide security to its resources.

- *Community Cloud*
o Cloud infrastructure which can be used by several organizations in a private community and supports a specific community which shares resources.
- *Hybrid Cloud*
o It uses a combination of public cloud, private cloud and even local infrastructures, which is typical for most IT vendors.

D. Cloud Service Model

Cloud computing provides services, basically by three models that is software, platform and infrastructure.

a. Software as a Service

Software as a Service (SaaS) is a software delivery model in which applications are accessed by a simple interface such as a web browser over the Internet. The users are not concerned with the underlying cloud infrastructure and even not concern about network, operating systems, servers, platform and storage. This model also eliminates the needs to install and run the application on the local computers.

SaaS term is popularized by Salesforce.com that distributes business software on a subscription basis and not on a traditional on-premise basis. This is best known solution for its Customer Relationship Management (CRM). Now SaaS has now become a common delivery model for most business applications, including accounting, collaboration and management. Applications like social media, office software, and online games which enrich the family of SaaS-based services, for example like web Mail, Google Docs, Microsoft online, NetSuit, MMOG Games, Facebook and many more.

b. Platform as a Service

Platform as a Service (PaaS) offers a high-level integrated environment to build, test, deploy and host customer-created or acquired applications. Generally, developers may accept some restrictions on the type of software that can write in exchange for built-in application scalability. Customers of PaaS does not manage the underlying infrastructure as SaaS users do but control over the deployed applications and their hosting environment configurations.

PaaS offerings mainly aim at facilitating application development and related management topics. Some are destined to provide a generalized development environment and some only provide hosting-level services such as security and on-demand scalability. Some examples of PaaS are Google App Engine, Engine Yard, Force.com, Windows Azure, Heroku, and MTurk.

c. Infrastructure as a Service

Infrastructure as a Service (IaaS) provides processing, networks, storage and other fundamental computing resources to users. IaaS users can deploy arbitrary applications, software, operating systems on the infrastructure which is capable of scaling up and down dynamically.

IaaS user sends programs and related data, while the vendor's computer does the computational processing and returns the result. The infrastructure is virtualized, scalable, flexible and manageable to meet user requirements. Examples of IaaS include VPC, Amazon EC2, IBM Blue Cloud, Eucalyptus, Joyent, FlexiScale and Rackspace Cloud.

IV. OPEN SOURCE RESOURCES FOR CLOUD COMPUTING

Open source software has been on the rise as many concerns during the protracted economic downturn and one of the areas where it is starting to offer companies a lot of flexibility and cost savings is in cloud computing. Cloud deployments can save money, free businesses from vendor lock -ins that could really sting over time and offer flexible ways to unite public and private applications. The following are few top open source cloud applications, educational resources, services, funding options, general items of interest and more.

A. Eucalyptus

Ostatic spread news about US Santa Barbara's open source cloud project last year. It released an open source infrastructure for cloud computing on clusters that duplicates the functionality of Amazon's EC2. Eucalyptus mainly uses the Amazon command line tools. Start Up Eucalyptus Systems was launched this year with venture funding and the staff include original architects of the Eucalyptus project. The company released its first major update to the software framework, which is also powering the cloud features in the new version of Ubuntu Linux.

B. Red Hat's Cloud

Linux focused open source player Red Hat has been rapidly expanding its focus on cloud computing. Red Hat held its Open Source Cloud Computing Forum which includes a large number of presentations from movers and shakers focused on open source cloud initiatives. Novell is also an open source focused company that is increasingly focused on cloud computing.

C. Traffic Server

Yahoo moved its open source cloud computing initiatives up a notch with the donation of its Traffic Server product to the Apache Software Foundation. Traffic Server is used at Yahoo to manage its own traffic and it enables session management, configuration management, authentication, configuration management, routing and load balancing for entire cloud computing software stacks. It act as overlay to raw cloud computing services. Even allows IT administrators to allocate resources which includes handling thousands of virtualized services concurrently.

D. Cloudera

The open source Hadoop software framework is increasingly used in cloud computing deployments due to its flexibility to data intensive queries, cluster based, and other tasks. It's managed by the Apache Software Foundation and Yahoo has its own time tested Hadoop distribution. Cloudera can be promising startup focused on providing commercial support for Hadoop.

E. Puppet

Virtual servers are on the rise in cloud computing deployments and Reductive Labs open source software. It is built on the legacy of the Cfengine System and highly respected by many system administrators for managing them. It can be used to manage large numbers of systems or virtual machines through automated procedures, without having to manage a lot of complex scripting.

F. Enomaly

The Company's Elastic Computing Platform (ECP) has its roots in widely used Enomalism open source provisioning and management software. It is designed to take much of the complexity out of starting a cloud infrastructure. ECP is a programmable virtual cloud computing infrastructure for small, medium and large businesses.

G. Joyent

Joyent have Reasonably Smart, a fledgling open source cloud startup based on JavaScript and Git. Its cloud hosting infrastructure and cloud management software incorporate many open source tools for public and private clouds. The company can also help to optimize a speedy implementation of the open source MySQL database for cloud use.

H. Zoho

Many people use Zoho's huge suite of free, online applications which is competing with Google Docs. Zoho's core is entirely open source- a shining example of how SaaS solutions can work in harmony with open source.

I. Globus Nimbus

This open source toolkit allows businesses to turn clusters into Infrastructure-as-a-Service (IaaS) clouds. The Amazon EC2 interface is carried over, but is not the only interface that can be selected.

J. Reservoir

This is a main European research initiative on virtualized infrastructure and cloud computing. It is far reaching project targeted to develop open source technology for cloud computing and help businesses avoid vendor lock in.

K. OpenNebula

The OpenNebula VM Manager is a core portion of Reservoir. It is an open source answer to the many virtual machine management offerings from proprietary players and interfaces easily with cloud infrastructure tools and services. It is an open source virtual infrastructure engine that enables the dynamic deployment and replacement of virtual machines on a pool of physical resources. It's good to see open source tools and resources competing in the cloud computing space. The end result should be more flexibility for organizations that want to customize their approaches. Open source cloud also have the potential to keep pricing for all competing services on a level playing field.

V. Cloud Computing Services

The three best cloud computing services are-

A. Heroku

It is an impressive cloud hosting platform that is both robust and easy to use. It has powerful features includes compatibility with every major Web language, faster application deployment, analytics and monitoring, an intuitive user interface, easy scale management and simple third party add on integration from outside vendors. It also provides a collection of tutorials and documentations for uncomplicated use.

B. Amazon Elastic Beanstalk

Amazon Elastic Beanstalk is an excellent cloud platform that provides a comprehensive set of cloud-computing tools. Its services include Amazon EC2 (Elastic Compute Cloud), a Linux or Windows server in cloud; Amazon S3 (Simple Storage Service), where files like documents, photos, videos etc. can be stored; Amazon RDS (Relational Database Service) for databases; and Amazon ELB (Elastic Load Balancing), which manages traffic in and out of cloud.

C. Windows Azure

Windows Azure, which is perfect for anyone who wants to source out all the administrative tasks associated with cloud computing. It offers a platform-as-a-service that handles everything from setting up and configuring to updating and maintaining your cloud server. It also support .NET applications. Windows Azure package has in-depth documentation and tutorials, email and phone support which makes it one of the most user-friendly cloud services. Cbeyond, a cloud services and communications provider for SMBs suggests of finding cloud-computing services that offer both cloud and network connectivity to describe logistics, save capital and avoid the headaches of dealing with multiple vendors.

VI. Cloud Management Tools

Cloud computing environments is managed by some form tool mainly called as Cloud Management Tools. This tools monitor, provision and tools that cross the divide between both. The main cloud infrastructure management products offer similar core features:

- Most support different cloud types (often referred to as hybrid clouds).
- Most support the on-the-fly creation and provisions of new objects and the destruction of unnecessary objects like servers, apps and storage.
- Most provide the usual types of reports on the status like response time, uptime, quota use etc. and have a dashboard that can be drilled into.

When compared on basis of those three criteria, few vendors only offer pervasive approaches in handling provisioning and managing metrics in hybrid environments: RightScale, Zeus, Kaavo, Scalr and Morph. There are also options offered by cloud vendors

themselves that meet the second and third criteria such as CloudWatch from Amazon Web Services.

The large companies known for their traditional data centre monitoring applications have been slow to hit the cloud market. Their products are rehashes of existing applications that do little in the way of providing more than alerting and reporting tools. CA is on an acquisition spree to fix this and just acquired 3Tera, a cloud provisioning player.

An example of the confusion in the industry is IBM's Tivoli product page for cloud computing. It is noticed that by clicking the Getting Started tab results in a 404 error. Nice work, IBM. Meanwhile, HP's OpenView (now called Operations Manager) can manage cloud based servers, but so far only manage like another server. BMC is working on a cloud management tool whereas it doesn't have anything beyond its normal products. These are the good infrastructure management and provisioning options available today.

A. RightScale

RightScale is better management tools for cloud environments. It also offer a free edition with limitations on features and capacity, designed for introduction of the product. RightScale's product is broken down into four parts:

- Cloud Management Environment
- Cloud-Ready [3] ServerTemplate and Best Practice Deployment Library
- Adaptable Automation Engine
- Multi-Cloud Engine

A fifth feature states that the "Readily Extensible Platform supports programmatic access to the functionality of the RightScale Platform."[3] When reviewing the product, these features aren't really separate from one another but make a nice integrated offering.

The management environment are the main interface and users can find with the software. It is designed to understand by user through the initial process of migrating to the cloud using their templates and library. The management environment is used for managing that environment, by continuing builds and ensuring resource availability. Here the automation engine comes into play: being able to quickly provision and put into operation additional capacity, [3] or remove that excess capacity as needed. There is also Multi-Cloud Engine which supports Amazon, Eucalyptus, GoGrid and Rackspace. It is also working on supporting the Chef open-source systems integration specifications. Chef is designed from the ground up for the cloud.

B. Kaavo

Kaavo plays in a very similar to RightScale. The product is mainly used for:

- Single-click deployment of complex multi-tier applications in the cloud (Dev, QA, Prod)
- Handling [3] demand bursts/variations by automatically adding/removing resources

- Run-time [3] management of application infrastructure in the cloud
- Encryption of persisting data in the cloud
- Automation [3] of workflows to handle run-time production exceptions without human intervention

The core of Kaavo's product is named IMOD. IMOD handles configuration, changes (adjustments in their terminology) and provisioning to the cloud environment and across multiple vendors in a hybrid model. Similar to all major CIM players, Kaavo's IMOD sits at the "top" of the stack, [3] managing the infrastructure and application layers.

One particular feature in IMOD is its multi-cloud, single system tool. For example, create a database backend in Rackspace while putting the presentation servers on Amazon. It supports Amazon and Rackspace in the public space and Eucalyptus in the private space. Though it should be noted that most cloud management can support Eucalyptus when it also support Amazon.

Both Kaavo and RightScale offer scheduled "ramp-ups" or "ramp-downs" (dynamic allocation based on demand) and monitoring tools to ensure that information and internal metrics (like SLAs) are transparently available. The dynamic allocation helps in meeting the demands of those SLAs. Both provide the ability to keep templates as well to ease the deployment of multi-tier systems.

C. Zeus

Zeus was famous for its rock-solid Web server, one that didn't have a great deal of market share but did cause a bunch of rapid fans and top-tier customers. Zeus with Apache and to a lesser-extent IIS which dominates the market had glut of load balancers. It took its expertise in the application server space and produced the Application Delivery Controller piece of the Zeus Traffic Controller. It is using traditional load balancing tools to test availability and spontaneously generate or destroy additional instances in the cloud and provides on-the-fly provisioning. It currently supports this on Rackspace and Amazon platforms.

D. Scalr

Scalr is a small project hosted on Google Code and Scalr.net that creates dynamic clusters, as same like Kaavo and RightScale on the Amazon platform. It supports triggers upsizing and downsizing based on traffic demands, snapshots (which can be shared, incidentally, a very cool feature). It is custom building of images for each server or server-type. Scalr does not support the wide number of platforms, applications, operating systems and databases. It is attached to the traditional expanded-LAMP architecture ([3] LAMP plus Ruby, Tomcat etc.)

E. Morph

It is not a true management platform. MSP-minded Morph products offers similar functionality in its own private space. Morph CloudServer is a newer product fills the management and provisioning space as an appliance. [3] It is aimed at the enterprise seeking to

deploy a private cloud. Morph CloudServer a top tier product is based on the IBM BladeCenter and supports hundreds of virtual machines.

The core is an Ubuntu Linux operating system and the Eucalyptus cloud computing platform. It is mainly aimed to manage service provider market. It allows for the creation of private clouds and the dynamic provisioning within those closed clouds. Later, it made quite a splash and bears watching mostly because of its open-source roots and participation in open-cloud systems.

F. CloudWatch

Amazon's CloudWatch works on Amazon's platform only. It limits its overall usefulness as it cannot be a hybrid cloud management tool. Since Amazon's Elastic Compute Cloud (EC2) is the biggest platform and still bears mentioning.

CloudWatch for EC2 supports dynamic provisioning (called auto-scaling), load balancing and monitoring, all managed through a central management console used by Amazon Web Services. Its biggest advantage is that no additional software is required to install and no additional website to access applications through. The product is clearly not for enterprises that need hybrid support which exclusively use Amazon should know that it is as robust and functional as the other market players.

VII. Software Testing Tools to Test Cloud Applications

New Open Source Testing Tools are emerging which can be deployed, managed and tested for the latest Cloud Computing Software applications. With its dynamic scalability, flexibility and virtualized resources are provided as a service, Cloud Computing [4] is seen as the dawn of a new era for application services. With examples of general purpose applications Google Documents, Zoho, Buzzword and Flickr that use Cloud Computing Technology can be seen as the most viable option for application development and deployment. Various IT Giants such as Microsoft, Amazon and Google and they all vying for a spot within the Cloud Computing space for Cloud based Software Applications currently available. Even in the near future Software Testing appear to be the current favoured use of Cloud environments.

A recent survey by Evans Data, an independent research firm that conducts periodic surveys of developers had found that about those who is using cloud facilities to run applications, 49.8% said they were doing so experimentally or for prototyping; 28.6% for non-critical business applications and 21.6% of business critical applications. They understand Cloud environments as being "better for testing because they can be set and torn down rapidly, sometimes at less expense than on-premise facilities".
The question to answer then is what Software testing tools are available to assist developers and Quality Assurance individuals in their application development and testing processes. Software Testing tools that are

used for testing of conventional applications are of little use when applied to Cloud Testing as there is a need for tools to allow Software developers and Tester to [4] analyse the network, desktop and implications of changes within the Cloud.

A growing variety of Cloud based Open Source Software Testing Tools are being published. Cloud Tools for example is a set of tools for deploying, [4] managing and testing Java EE applications on Amazon's Elastic Computing Cloud (EC2). Containing three main parts, which includes machine images that can be configured to run on Tomcat and Maven & Grails plug-in this is an amazing tool to use for Open source cloud software testing.

PushToTest TestMaker is a distributed test environment that can run tests on test equipment, [4] or in a Cloud Computing environment. It introduces specific commands to support automatic Cloud Testing services. Cloud Tools and PushToTest Test Maker represent examples of products that will help shape the future of robust Cloud based Software Testing Applications. Though the technology is in its early stage, various testing tools are emerging that can provide assistance in cloud based software testing. Other popular testing tools include:

A. SoapUI

It is a functional testing tool used for testing web services and APIs. Even provide extensive support for various kinds of protocols such as SOAP, REST, JMS, HTTP and JDBC. The reason for using SoapUI, is that it's easy to get started with. It does not requires any programming knowledge, but is nonetheless very powerful and has extensive scripting support for all kinds of advance use cases.

B. LoadUI

LoadUI [4] is the one of the world's most downloaded load testing software. It's free, open-source and easily extendable using Groovy. It can also integrate with SoapUI, letting to leverage existing functional tests when load testing. It helps in finding answer about the questions: "[4] Does it perform? Does it scale? What causes the bottlenecks? Is the experience consistent for users on different continents? ".

C. TestMaker

TestMaker is a distributed test environment. It runs the tests on the test equipment, in a Cloud Computing environment, or both. It introduces specific commands to support automatic Cloud Testing. For example, [4] identify a cloud testing service like Amazon EC2 in a TestScenario. TestMaker creates the TestNodes in EC2 instances [4], runs the test, retrieves the results, and puts down the EC2 instances. All in a 'lights out' manner for full Cloud Test automation.

VIII. Cloud Computing Load Test and Performance Monitoring Companies

Test Perspective is a cost-effective and completely self-service By Test Perspective the most realistic load tests on-demand can be executed and immediate feedback can be received by modifications made on the Web site.

A. Soasta

Is available as an on demand service in the cloud or virtual appliance or as a physical. SOASTA CloudTest is an integration of test design, monitoring, and reporting. It offers everything for testing and delivering high quality Web applications and services on an affordable price.

B. Monitis

Monitis has its Universal Cloud Monitoring Framework. It can now sync to other Cloud computing providers very quickly – [5] from Rackspace, GoGrid, Softlayer, and more. Monitis' Universal Cloud Monitoring Framework will automate monitoring in highly dynamic cloud environments, where [5] customers' servers maybe added and terminated according to the load by management software or manually.

C. Cloudslueth

It monitors the performance of own cloud application in real time. Firstly by submitting the URL and it will monitor the cloud application's live performance and availability measurements. It is owned and operated by Compuware.

D. BrowserMob

Its monitoring service uses real web browsers to simulate a user interacting with particular website., despite being similarly priced to simplistic "ping" style monitoring services that only check individual URLs [5]. The use of real browsers allows to check that individual objects are responding and also to functionally verify that the website is working from end-to-end. [5] This includes ensuring that *3rd party components* such as advertisements, gadgets, analytics, and Content Delivery Networks don't adversely affect the website.

E. CloudTools

Tools for deploying Java Spring and Grails applications in the Cloud [5]. It is a set of tools for deploying, managing and testing Java EE applications on Amazon's Elastic Computing Cloud (EC2) and VMware environments.

F. GFI

GFI MAX RemoteManagement makes it easy to monitor the clients' servers, raid arrays, remote offices, ISP connection, bandwidth usage and their website. The key parts of their network infrastructure can be monitored and proactively taking care of the clients, keeps hardware costs to a minimum and preventing costly downtime.

G. LoadStorm

Load testing allows web developers to recognize how their applications respond under heavy volumes of HTTP traffic. It put massive cloud resources in the hands of web developers for enabling them to improve the performance of their web applications.

H. CloudHarmony

CloudHarmony is still in beta. However their extensive and continuously updated benchmarks can be used for viewing and comparing performance metrics from various cloud providers and services.

I. InterMapper

Amazon Web Services (AWS) delivers your business-critical services and InterMapper makes sure your services are continually operationally. Amazon Web Services' (AWS) CloudWatch technology collects performance metrics for the virtual servers running in its cloud. InterMapper [5] queries these metrics to monitor the status of an Amazon ec2 Instance.

IX. CLOUD COMPUTING MODELLING AND SIMULATIONS

Cloud simulators are required for cloud system testing to decrease the complexity and separate quality concerns. They enable performance analysts to analyse system behaviour by focusing on quality issues of specific component under different scenarios. Some of the published cloud computing simulators for evaluating cloud computing system performance are described briefly in this section.

A CloudSim

ICloudSim is a simulation application which enables seamless modelling, simulation, and experimentation of cloud computing and the application services, proposed due to the problem that existing distributed system simulators were not applicable to the cloud computing environment. It had been mentioned that users could analyse specific system problems through CloudSim, by not considering the low level details related to the Cloud-based infrastructures and services.
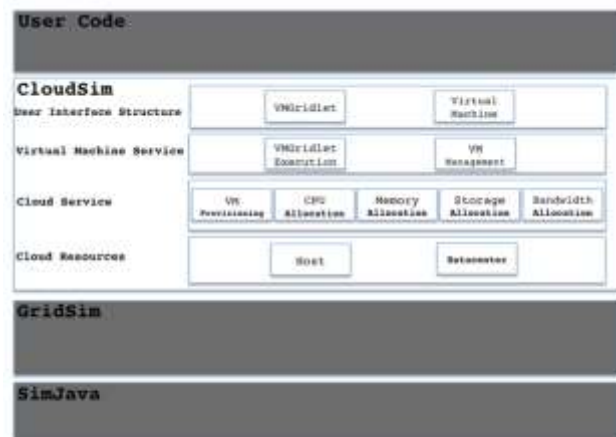
Fig. 2. Layered CloudSim Architecture

a. CloudAnalyst: CloudAnalyst was derived from CloudSim. This simulator can be applied to examining the behaviour of the large scaled Internet application in a cloud environment and separating the simulation experimentation exercise from a programming exercise.

It also enables a modeller to repeatedly perform simulations and to conduct a series of simulation experiments with slight parameter variations in a quick and easy manner.

b. *GreenCloud:* GreenCloud is an improvement in CloudSim to prove the green cloud computing approach. The lack of detailed simulators, and having no provisioning system for analysing energy efficiency of the clouds was the motivation behind GreenCloud development to interact and measure cloud performance. It's an advanced packet-level simulator with concentration on cloud communication. It provides a detailed fine-grained modelling of the energy consumed by the data centre IT equipment such as computing servers, communication links and network switches. It is considered as an extension of the network simulator Ns2.

c. *EMUSIM*: EMUSIM is an integrated architecture proposed to anticipate service's behaviour on cloud platforms to a higher standard build on Automated Emulation Framework (AEF) for emulation and CloudSim for simulation.

d. *NetworkCloudSim*: It is an extension of CloudSim as a simulation framework that supports generalized applications such as high performance computing applications, e-commerce and workflows besides real cloud data centres modelling.
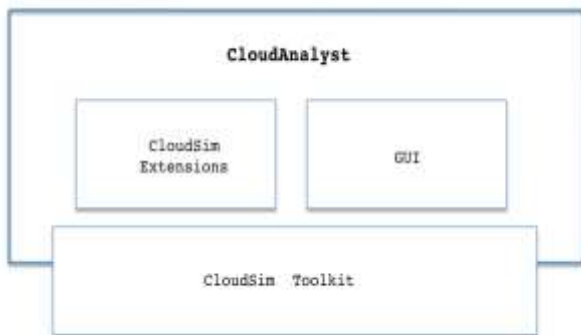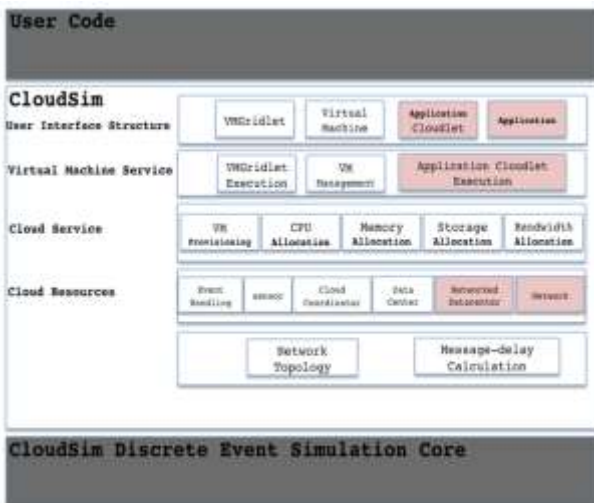

Fig. 3. Architecture of CloudAnalyst

Fig. 4. Architecture of CloudSim based NetworkCloudSim

B. SPECI

The size of data centres that provide cloud computing services is increasing and some middleware properties that manage these data centres will not scale linearly with the number of components.

Simulation Program for Elastic Cloud Infrastructures (SPECI) is a simulation tool, which allows analysing large data centre behaviour under the size and design policy of the middleware as inputs. It is composed of two packages namely data centre layout and topology and the components for experiment execution and measuring.

C. GroundSim

GroundSim is an event-based simulator that needs just one simulation thread for scientific applications on grid and cloud environments. It is mainly concentrated on the IaaS, but it is easily extendable to support additional models like PaaS, or cloud storage.

More investigation was carried out in order to allow the user to simulate their experiments from the same environment used for real applications, by integrating GroudSim into the ASKALON environment.

D. DCSim

DataCenter Simulator is concentrated on virtualized data centres, offering IaaS to Multiple tenants, in order to achieve a simulator to evaluate and develop data centre management techniques.
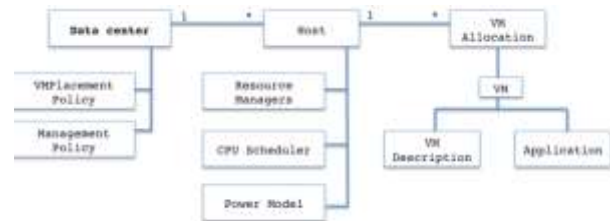

Fig. 5. Architecture of DCSim

X. CLOUD TESTING BENCHMARK

There is no single or ideal approach for cloud testing. [6] This is basically due to the fact that there exists different factors such as cloud architecture design, non-functional and compliance requirements, etc., can be taken into account to ensure successful and complete testing when an organization starts cloud testing. Some common benchmarks developed to support cloud testing are briefly presented below.

A. YCSB

A key design goal of "Yahoo! Cloud Serving Benchmark's tool is extensibility; it is designed to be extensible and portable to mixed clouds to provide a comparison between cloud storage systems. [6] Since this benchmark is under an open source license, others are able to use and extend the tool, and contribute new workloads and database interfaces. It can be used to measure the performance of several cloud systems, [6] and it is intended to deal with various quality concerns such as performance, scalability, availability and replication.

Performance examines the response time with increasing throughput until database saturation. Scaling tests how increasing in the number of machines affects system performance [6]. Scale-up and speed-up are used as scaling metrics. A workload generator, which defines YCSB tool, utilizes user-defined workload descriptions, and a standard workload package, which is a collection of programs representing typical cloud operations.

B. Enhanced TPC-W

The architecture of TPC-benchmarks and its metrics are designed for transactional database systems therefore they are not suitable for cloud systems. Hence the new benchmark system specifically for cloud scalable, pay-as-you-go and fault-tolerance testing and evaluation was suggested. The benchmark defines web interactions as benchmark drivers with the usage of e-commerce scenarios. Scalability, cost, fault tolerance and peaks are defined as new metrics for cloud storage system evaluation.

- *Scalability*: Cloud services are expected to scale linearly with a constant cost per web interaction. It's been proposed that the deviations of response time for the perfect linear scale can be measured by using correlation coefficient, R^2.
- *Fault tolerance*: Since hardware failures are common in IaaS, the metric is determined to analyse the potential of cloud self-healing. The recoverability [6] of failures in a period of time is defined as the ratio between WIPS, Web Interactions per Second, in RT, real-time, and Issued WIPS.
- *Cost*: Cloud Performance [6] economy is measured by $/WIPS used by conventional TPC-W benchmark.
- *Peaks*: This metric [6] is to measure how well a cloud can adapt to peak loads, scale-up and scale-down. The adaptability is defined by the ratio between WIPS in RT and Issued WIPS.

C. CloudStone

TCloudstone is an open source project to deal with the new system architectures' performance characteristics. It is used for comparing across different development stacks using social applications and automated load generators. The metric [6] dollars-per-user-per-month regarding to realistic usage and cost is used to analyse cloud performance.

D. MalStone

MalStone is mainly designed for cloud middleware performance testing for data intensive computing by Open Data Group. MalGen is developed to generate synthetic log-entity files that are used for testing inputs. It can generate tens of billions of events on cloud with over 100 nodes.

MalGen, an open source software package, generates the data for MalStone, synthetic log-entity files to test inputs. MalGen [6] generates large, distributed data sets suitable for testing and benchmarking software designed for data intensive computing.

XI. APPLICATION OF CLOUD

Cloud computing has gained huge popularity in the industry due to its ability to host applications for which the services can be delivered to consumers rapidly at minimal cost. Applications from a range of domains, from science to engineering, gaming, and social networking, are considered.

A. Scientific Applications

Scientific applications are a sector that is increasingly using cloud computing systems and technologies. The immediate benefit seen by researchers and academics are the potentially infinite availability of computing resources and storage at sustainable prices compared to a complete in-house deployment. Cloud computing systems meet the needs of different types of applications in the scientific domain: high-performance computing (HPC) applications, high-throughput computing (HTC) applications, and data-intensive applications. The opportunity to use cloud resources is even more appealing because minimal changes need to be made to existing applications in order to leverage cloud resources.

The most relevant option is IaaS solutions, which offer the optimal environment for running bag-of-tasks applications and workflows. PaaS solutions have been considered as well. They allow scientists to explore new programming models for tack- ling computationally challenging problems. Applications have been redesigned and implemented on top of cloud programming application models and platforms to leverage their unique capabilities. Problems that require a higher degree of flexibility in terms of structuring of their computation model can leverage platforms such as Aneka, which supports MapReduce and other programming models.

B. Healthcare: ECG analysis in the cloud

Healthcare is a domain in which computer technology has found several and diverse applications: from supporting the business functions to assisting scientists in developing solutions to cure diseases. An important application is the use of cloud technologies to support doctors in providing more effective diagnostic processes.

The capillary development of Internet connectivity and its accessibility from any device at any time has made cloud technologies an attractive option for developing health-monitoring systems. ECG data analysis and monitoring constitute a case that naturally fits into this scenario. ECG is the electrical manifestation of the contractile activity of the heart's myocardium. This activity produces a specific waveform that is repeated over time and that represents the heartbeat. The analysis of the shape of the ECG waveform is used to identify arrhythmias and is the most common way to detect heart disease. Cloud computing technologies allow the remote monitoring of a patient's heartbeat data, data analysis in minimal time, and the notification of first-aid personnel and doctors should these data reveal potentially dangerous conditions. [7] This way a patient at risk can be constantly monitored without going to a hospital for ECG analysis.

C. Biology: Protein Structure Prediction

Applications in biology often require high computing capabilities and often operate on large data- sets that cause extensive I/O operations. Because of these requirements, biology applications have often made extensive use of supercomputing and cluster computing infrastructures. Similar capabilities can be leveraged on demand using cloud computing technologies in a more dynamic fashion, thus opening new opportunities for Bioinformatics applications [7].

Protein structure prediction is a computationally intensive task that is fundamental to different types of research in the life sciences. Among [7] these is the design of new drugs for the treatment of diseases. The geometric [7] structure of a protein cannot be directly inferred from the sequence of genes that compose its structure, but it is the result of complex computations aimed at identifying the structure that minimizes the required energy. This task requires the investigation of a space with a massive number of states, consequently creating a large number of computations for each of these states. The computational power required for protein structure prediction can now be acquired on demand, without owning a cluster or navigating the bureaucracy to get access to parallel and distributed computing facilities. [7] Cloud computing grants access to such capacity on a pay-per-use basis. This concept is distinctive of cloud technologies and constitutes a strategic advantage when applications are offered and delivered as a service.

D. Biology: Gene Expression Data Analysis for Cancer Diagnosis

Gene expression profiling is the measurement of the expression levels of thousands of genes at once. It is used to understand the biological processes that are triggered by medical treatment at a cellular level. [7] Together with protein structure prediction, this activity is a fundamental component of drug design, [7] since it allows scientists to identify the effects of a specific treatment. Another important application of gene expression profiling is cancer diagnosis and treatment.

This problem is often approached with learning classifiers, which generate a population of condition-action rules that guide the classification process. Among these, the eXtended Classifier System (XCS) has been successfully utilized for classifying large datasets in the bioinformatics and computer science domains. A variation of this algorithm, CoXCS, has proven to be effective in these conditions. Cloud-CoXCS is a cloud-based implementation of CoXCS that leverages Aneka to solve the classification problems in parallel and compose their outcomes.

E. Geoscience: Satellite Image Processing

Geoscience applications collect, produce, and analyse massive amounts of geospatial and nonspatial data. [7] As the technology progresses and our planet become more instrumented (i.e., through the deployment of sensors and satellites for monitoring), the volume of data that needs to be processed increases significantly. In particular, the geographic information system (GIS) is a major element of Geoscience applications. GIS applications capture, store, manage, analyse, manipulate, and present all types of geographically referenced data. [7] This type of information is now becoming increasingly relevant to a wide variety of application domains: from advanced farming to civil security and natural resources management. As a result, a considerable amount of Geo-referenced data is ingested into computer systems for further processing and analysis. Cloud computing is an attractive option for executing these demanding tasks and extracting meaningful information to support decision makers.

F. Business and consumer applications

The business and consumer sector is the one that probably benefits the most from cloud computing technologies. On one hand, the opportunity to transform capital costs into operational costs makes clouds an attractive option for all enterprises that are IT-centric. On the other hand, the sense of ubiquity that the cloud offers for accessing data and services makes it interesting for end users as well. The combination of all these elements has made cloud computing the preferred technology for a wide range of applications, from CRM and ERP systems to productivity and social-networking applications.

CRM and ERP: Customer relationship management (CRM) and enterprise resource planning (ERP) applications are market segments that are flourishing in the cloud, with CRM applications the more mature of the two. Cloud CRM applications constitute a great opportunity for small enterprises and start-ups to have fully functional CRM software without large up-front costs and by paying subscriptions. ERP solutions in the cloud are less mature and have to compete with well-established in-house solutions. ERP systems integrate several aspects of an enterprise: finance and accounting, human resources, supply chain management, manufacturing, project management, and CRM. Their goal is to provide a uniform view and access to all operations that need to be performed to sustain a complex organization.

Salesforce.com is probably the most popular and developed CRM solution available today. The application provides customizable CRM solutions that can be integrated with additional features developed by third parties. Salesforce.com is based on the Force.com cloud development plat- form. This represents scalable and high-performance middleware executing all the operations of all Salesforce.com applications [7].

Microsoft Dynamics CRM is the solution implemented by Microsoft for customer relationship management. Dynamics CRM is available either for installation on the enterprise's premises or as an online solution priced as a monthly per-user subscription. NetSuite provides a collection of applications that help customers manage every aspect of the business enterprise. Its offering is divided into three major products: NetSuite Global ERP, NetSuite Global CRM1, and NetSuite Global ECommerce. Moreover, an all-in-one solution: NetSuite One World, integrates all three products together.

Productivity: Productivity applications replicate in the cloud some of the most common tasks that we are used to performing on our desktop: from document storage to office automation and complete desktop environments hosted in the cloud. One of the core features of cloud computing is available anywhere, at any time, and from any Internet-connected device. Therefore, document storage constitutes a natural application for such technology. Online storage solutions preceded cloud computing, but they never became popular.

Perhaps the most popular solution for online document storage is Dropbox, an online application that allows users to synchronize any file across any platform and any device in a seamless manner. Dropbox provides users with a free amount of storage that is accessible through the abstraction of a folder. Another interesting application in this area is iCloud, a cloud-based document-sharing application provided by Apple to synchronize iOS-based devices in a completely transparent manner. There are other solutions for online document sharing, such as Windows Live, Amazon Cloud Drive and CloudMe.

Google Docs is a SaaS application that delivers the basic office automation capabilities with support for collaborative editing over the Web. The application is executed on top of the Google distribute computing infrastructure, which allows the system to dynamically scale according to the number of users using the service. Google Docs allows users to create and edit text documents, forms, spreadsheets, presentations and drawings. It aims to replace desktop products such as Microsoft Office and OpenOffice and provide similar interface and functionality as a cloud service.

Social networking applications have grown considerably in the last few years to become the most active sites on the Web. To sustain their traffic and serve millions of users seamlessly, services such as Twitter and Facebook have leveraged cloud computing technologies. The possibility of continuously adding capacity while systems are running is the most attractive feature for social networks, which constantly increase their user base.

Media applications are a niche that has taken a considerable advantage from leveraging cloud computing technologies. In particular, video-processing operations, such as encoding, transcoding, com- position, and rendering, are good candidates for a cloud-based environment. These are computationally intensive tasks that can be easily offloaded to cloud computing infrastructures. Animoto is perhaps the most popular example of media applications on the cloud. Video encoding and transcoding are operations that can greatly benefit from using cloud technologies.

XII. Conclusion

Cloud computing is the use of computing resources that are delivered as a service over a network. It shares characteristics with autonomic computing, grid computing, client server model, mainframe computer, utility computing and peer to peer. The detailed study of cloud computing basics like deployment model, system architecture, cloud services and types of cloud has been done. The open source resources of cloud computing has been studied. The management tools, load tools and test tools are found for cloud computing.

References

[1] Dorian Minarolli and Bernd Freisleben, Uitlity–based Resource Allocations for virtual machines in cloud computing, IEEE, 2011

[2] J. Geelan, "Twenty-one experts define cloud computing," Cloud Computing Journal, vol.4, pp. 1-5, 2009.

[3] Cloud Management Tools. http://searchcloudcomputing.techtarget.com/report/Cloud-management-tools-guide-for-beginners

[4] Software testing tools to test cloud computing applications. http://soa.dzone.com/articles/software-testing-tools-test

[5] Cloud computing load test and performance monitoring companies. http://www.cloudtweaks.com/2010/08/top-10-cloud-computing-load-test-and-performance-monitoring-companies/

[6] A survey on cloud computing simulations and cloud testing. http://www.cse.wustl.edu/~jain/cse567-13/index.html

[7] Mastering Cloud Computing: Cloud Foundation and Applications Programming. Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi. MK Publisher Edition 2013.

[8] Chandrashekhar S. Pawar and R. B. Wagh. A review of resource allocation policies in cloud computing. World Journal of Science and Technology 2012 Vol. 2 Issue 3, ISSN: 2231-2587, Page 165-167.

[9] Sandeep Tayal. Tasks Scheduling optimization for the Cloud Computing Systems. IJAEST Vol. 5 Issue 2, Page 111-115, 2011.

[10] Monika Choudhary, Sateesh Kumar Peddoju. IJERA Vol. 2 Issue 3, May- Jun 2012, Page 2564-2568, ISSN: 2248-9622.

[11] Open source resource for cloud computing. http://gigaom.com/2009/11/06/10-top-open-source-resources-for-cloud-computing/

[12] Best Cloud computing services. http://www.businessnewsdaily.com/4427-cloud-computing-small-business.html

[13] Open source cloud computing testing tools. http://www.cloudcomputingexpo.com/node/2625872

[14] Aoun R., Doumith E.A., Gagnaire M., "Resource Provisioning for Enriched Services in Cloud Environment," IEEE Second International Conference on Cloud Computing Technology and Science, Feb. 2011, pp. 296-303.

[15] "Adaptive Management of Virtualized Resources in Cloud Computing Using Feedback Control," in First International Conference on Information Science and Engineering, April 2010, pp. 99-102.

[16] V. Vinothina, Dr. R. Shridaran, and Dr. Padmavathi Ganpathi, A survey on resource allocation strategies in cloud computing, International Journal of Advanced Computer Science and Applications, 3(6):97--104,2012

[17] Atsuo Inomata, Taiki Morikawa, Minoru Ikebe and Md. Mizanur Rahman, Proposal and Evaluation of Dynamic Resource Allocation Method Based on the Load Of VMs on IaaS, IEEE, 2010.

[18] Q. Cao, B. Wei and W. M. Gong, "An optimized algorithm for task scheduling based on activity based costing in

cloud computing," In International Conference on eSciences 2009, pp. 1-3.

[19] Sumit Goyal,"Public vs Private vs Hybrid vs Community - Cloud Computing: A Critical Review", IJCNIS, vol.6, no.3, pp.20-29, 2014. DOI: 10.5815/ijcnis.2014.03.03

[20] Bahman Rashidi,Mohsen Sharifi,Talieh Jafari,"A Survey on Interoperability in the Cloud Computing Environments", IJMECS, vol.5, no.6, pp.17-23, 2013.DOI: 10.5815/ijmecs.2013.06.03

[21] Ronak Patel, Sanjay Patel. Survey on Resource Allocation Strategies in Cloud Computing. IJERT Vol. 2 Issue 2 Feb 2013, ISSN: 2278-0181.

Authors' Profiles

Md. Imran Alam received his B.E degree in the year 2011 from GTEC, Vellore affiliated to Anna University, Chennai, and currently pursuing M.Tech in Computer Science Engineering from KIIT University, Bhubaneswar. His research area includes Distributed Database systems, Task Scheduling, Cloud Computing and big data.

Siddharth Swarup Rautaray is presently working as an Assistant Professor in School of Computer Engineering, KIIT University, Bhubaneswar, Odisha. He is pursuing his Ph.D. work in the Department of Information Technology at Indian Institute of Information Technology, Allahabad, India. His current research interests include Human–Computer Interactions, User Interface Design and Computer Vision Image Processing. He has published around 15 research papers in different international and national journals and conferences. He is a member of the IEEE Computer Society, ACM SIGCHI and CSI.

Manjusha Pandey is presently working as an Assistant Professor in the School of Computer Engineering, Kalinga Institute of Industrial Technology, and Bhubaneswar. She is pursuing her PhD from Indian Institute of Information Technology, Allahabad. She has more than 10 research publications to her credit in journals and conferences of repute. Her research interest areas include Wireless Sensor Network, Security and Privacy in Wireless Sensor Network, Human Computer Interaction and Computer Networks.