Gurudatt Kulkarni, Ramesh Sutar.Jayant Gambhir/ International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 1, Jan-Feb 2012, pp.117-125

"CLOUD COMPUTING-INFRASTRUCTURE AS SERVICE-AMAZON EC2"

Gurudatt Kulkarni[1], Ramesh Sutar [2] Jayant Gambhir [3]

1, 2,3 Lecturer in Marathwada Mitra Mandal's Polytechnic, Pune.
1, 2,3 ME Student Siddhant College of Engineering, Pune

ABSTRACT:-
Cloud computing is an increasingly popular paradigm for accessing computing resources. In practice, cloud service providers tend to offer services that can be grouped into three categories: software as a service, platform as a service, and infrastructure as a service. This paper discuss the characteristics and benefits of cloud computing. It proceeds to discuss the Infrastructure as a service (IaaS). This paper aims to provide a means of understanding and investigating IaaS. This paper also outlines the responsibilities of IaaS provider and the facilities to IaaS consumer. Amazon Elastic Cloud is one of the flexible clouds discussed in paper.

1.1WHAT IS A CLOUD COMPUTING?

Cloud computing is Internet("CLOUD") based development and use of computer technology ("COMPUTING").Cloud computing is a general term for anything that involves delivering hosted services over the Internet. It is used to describe both a platform and type of application. These cloud applications use large data centers and powerful servers that host Web applications and Web services. Anyone with a suitable Internet connection and a standard browser can access a cloud application. [1]

1.2BENEFITS OF CLOUD COMPUTING [2]

- Cloud technology is paid incrementally, saving organizations money.
- Organizations can store more data than on private computer systems.
- No longer do IT personnel need to worry about keeping software up to date.
- Cloud computing offers much more flexibility than past computing methods.
- Employees can access information wherever they are, rather than having to remain at their desks.
- No longer having to worry about constant server updates and other computing issues, government organizations will be free to concentrate on innovation.
- Decoupling and separation of the business service from the infrastructure needed to run it
- Flexibility to choose multiple vendors that provide reliable and scalable business services, development environments, and infrastructure that can be leveraged out of the box and billed on a metered basis—with no long term contracts

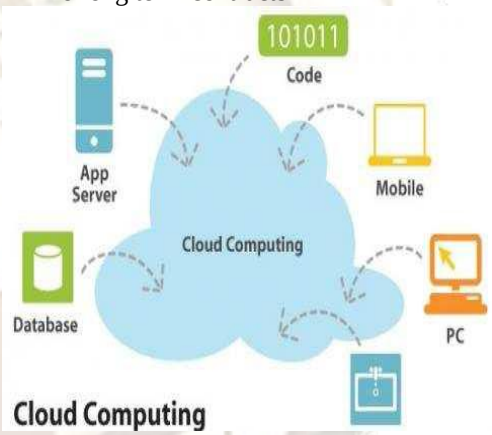


FIG:-1.1 GENERAL CLOUD DIAGRAM

1.2.1TECHNICAL BENEFITS OF CLOUD COMPUTING

Automation – "Scriptable infrastructure": You can create repeatable build and deployment systems by leveraging programmable (API-driven) infrastructure.

Auto-scaling: You can scale your applications up and down to match your unexpected demand without any human intervention. Auto-scaling encourages automation and drives more efficiency.

Proactive Scaling: Scale your application up and down to meet your anticipated demand with proper planning understanding of your traffic patterns so that you keep your costs low while scaling.

Improved Testability: Never run out of hardware for testing. Inject and automate testing at every stage during the development process. You can spawn up an "instant test lab" with pre-configured environments only for the duration of testing phase.

Disaster Recovery and Business Continuity: The cloud provides a lower cost option for maintaining a fleet of DR servers and data storage. With the cloud, you can take advantage of geo-distribution and replicate the environment in other location within minutes.

1.3 WHAT IS DRIVING CLOUD COMPUTING?

The CLOUD COMPUTING is driving in two types of categories as follows:

- Customer perspective
- Vendor perspective

1.3.1 CUSTOMER PERSPECTIVE:

- In one word: economics
- Faster, simpler, cheaper to use cloud computation.
- No upfront capital required for servers and storage.
- No ongoing for operational expenses for running datacenter.
- Application can be run from anywhere.

1.3.2 VENDOR PERSPECTIVE:

- Easier for application vendors to reach new customers.
- Lowest cost way of delivering and supporting applications.
- Ability to use commodity server and storage hardware.

1.4. TYPES OF SERVICES:

These services are broadly divided into three categories:

- Infrastructure-as-a-Service (**IaaS) [3]**
- Platform-as-a-Service **(PaaS)**
- Software-as-a-Service **(SaaS).**

1.4.1 INFRASTRUCTURE-AS-A-SERVICE (IAAS):

Infrastructure-as-a-Service (IaaS) provides virtual servers with unique IP addresses and blocks of storage on demand. Customers can pay for exactly the amount of service they use, like for electricity or water, this service is also called utility computing.

1.4.2 PLATFORM-AS-A-SERVICE (PAAS):[2]

Platform-as-a-Service (PaaS) is a set of software and development tools hosted on the provider's servers. Google Apps is one of the most famous Platform-as-a-Service providers. This is the idea that someone can provide the hardware (as in IaaS) plus a certain amount of application software - such as integration into a common set of programming functions or databases as a foundation upon which you can build your application. Platform as a Service (PaaS) is an application development and deployment platform delivered as a service to developers over the Web. It facilitates development and deployment of applications without the cost and complexity of buying and managing the underlying infrastructure, providing all of the facilities required to support the complete life cycle of building and delivering web applications and services entirely

1.4.3 SOFTWARE-AS-A-SERVICE (SAAS):[2]

Software-as-a-Service (SaaS) is the broadest market. In this case the provider allows the customer only to use its applications. The software interacts with the user through a user interface. These applications can be anything from web based email, to applications like Twitter or Last.fm. This is the idea that someone can offer you a hosted set of software (running on a platform and infrastructure) that you don't own but pay for some element of utilization - by the user, or some other kind of consumption basis. You don't have to

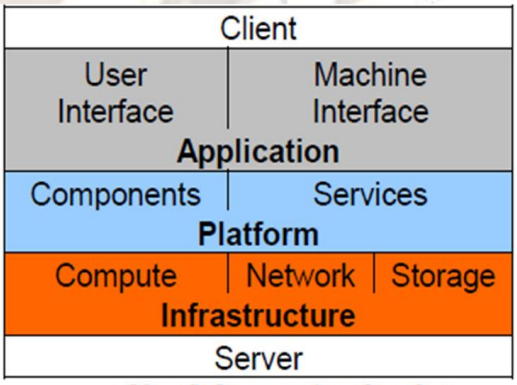


FIGURE 1.4.3 CLOUD COMPUTING STACK

do any development or programming, but you may need to come in and configure the (very flexible, configurable and sometimes customizable) software. You don't have to

purchase anything. You just pay for what you use. A SaaS provider typically hosts and manages a given application in their own data center and makes it available to multiple tenants and users over the Web. Some SaaS providers run on another cloud provider's PaaS or IaaS service offerings. Oracle CRM on Demand, Salesforce.com, and Netsuite are some of the well known SaaS examples.

1.4.4 TYPES BY VISIBILITY:

PUBLIC CLOUD:

In Public cloud or external cloud resources are dynamically provisioned on a fine-grained, self-service basis over the Internet, via web applications/web services, from an off-site third-party provider who shares resources and bills on a fine-grained utility computing basis.

HYBRID CLOUD:[3]

A hybrid cloud environment consisting of multiple internal and/or external providers. It can also describe configurations combining virtual and physical, collocated assets.

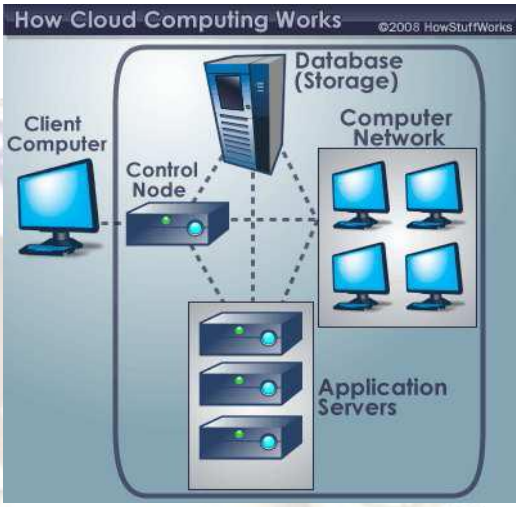
PRIVATE CLOUD:

Private cloud and internal cloud products claim to "deliver some benefits of cloud computing without the pitfalls", capitalizing on data security, corporate governance, and reliability concerns. They have been criticized on the basis that users "still have to buy, build, and manage them and as such do not benefit from lower up-front capital costs and less hands-on management, essentially"

1.5 HOW DOES CLOUD COMPUTING WORK?

Cloud computing aims to apply the power of supercomputer to problems like analyzing risk in financial portfolios, powering immersive computer games, in a way that users can tap through the Web. It does that by networking large groups of servers that often use low-cost consumer PC technology, with specialized connections to spread data-processing chores across them.

Soon instead of installing a suite of software for each computer, you'd only have to load one application. That application would allow workers to log into a Web-based service which hosts all the programs. Remote machines owned by another company would run everything from e-mail to word processing to complex data analysis programs. It's called cloud computing. The only thing the user's computer needs to be able to run is the cloud computing systems interface software, which can be as simple as a Web browser, and the cloud's network takes care of the rest. The software and storage for your account doesn't exist on your computer -- it's on the service's computer cloud.



Cloud computing providers deliver applications via the internet, which are accessed from a Web browser, while the business software and data are stored on servers at a remote location. In some cases, legacy applications (line of business applications that until now have been prevalent in thin client Windows computing) are delivered via a screen-sharing technology such as Citrix Xen App, while the computing resources are consolidated at a remote data center location; in other cases, entire business applications have been coded using web-based technologies such as AJAX.

2.0. INFRASTRUCTURES AS SERVICE[2]

Infrastructure as a Service (IaaS) is the delivery of hardware (server, storage and network), and associated software (operating systems virtualization technology, file system), as a service. It is an evolution of traditional hosting that does not require any long term commitment and allows users to provision resources on demand. Unlike PaaS services, the IaaS provider does very little management other than keep the data center operational and users must deploy and manage the software services themselves just the way they would in their own data center.

Amazon Web Services Elastic Compute Cloud (EC2) and Secure Storage Service (S3) are examples of IaaS offerings.

2.1UNDERSTANDING IAAS

Infrastructure as a Service is a form of hosting. It includes network access, routing services and storage. The IaaS provider will generally provide the hardware and administrative services needed to store applications and a platform for running applications. Scaling of bandwidth, memory and storage are generally included, and vendors compete on the performance and pricing offered on their dynamic services. The service provider owns the equipment and is responsible for housing, running and maintaining it. IaaS can be purchased with either a contract or on a pay-as-you-go basis. Characteristics and components of IaaS include:[3]

- Utility computing service and billing model.
- Automation of administrative tasks.
- Dynamic scaling.
- Desktop virtualization.
- Policy-based services.
- Internet connectivity.

IaaS provides an environment for running user built virtualized systems in the cloud. Figure 2.1.1 illustrates how a virtual machine is built for an IaaS environment, uploaded to the environment, configured, and then deployed within the environment. Using this technique virtual machines are created on premise and loaded with all the software that will eventually run in the cloud. This includes custom built software as well as licensed software. After the virtual machine is built it is uploaded to the IaaS vendor's hosting environment where it can be configured to use the IaaS vendor's raw storage. Once configured, the virtual machine can be deployed and started via some form of automation which automatically finds available hardware to run the virtual machine. Once the virtual machine is started the IaaS vendor can ensure that the running virtual machine continues to look healthy as a whole. The computers needed to run the application and the raw storage that is needed by the application are owned and supported by the IaaS vendor. It is the responsibility of the customer to monitor all the custom built software and licensed software to insure that they are operating properly. IaaS is an option that is very flexible and is the best choice for moving applications to the cloud when there is no time to rework the application's code for a cloud environment.

2.2 IAAS PROVIDER AND CONSUMER

The key roles in a cloud environment include the service consumer and the service provider. The cloud service consumer needs a secure anytime anywhere access to low cost services that are flexible and easy to use. The biggest hurdle to adoption of cloud has to do with consumers discomfort in the following areas: security of service and the underlying data, service availability and reliability, service management to ensure service level agreements, ensuring control over access and policies, and the appropriate administration to facilitate flexible pricing structures. The service provider actually runs the service that the service consumer wants and was designed and developed by the service creator.

The cloud services creator needs tools and capabilities to offer differentiated services, offer incentives to ensure that consumers keep coming back to use the services, and the ability to change services on-demand to stay competitive and address threats. The service provider needs their IT resources integrated so their usage is optimized, the ability to add/remove resources on demand, a non-disruptive way to save money, and the means to charge for usage.

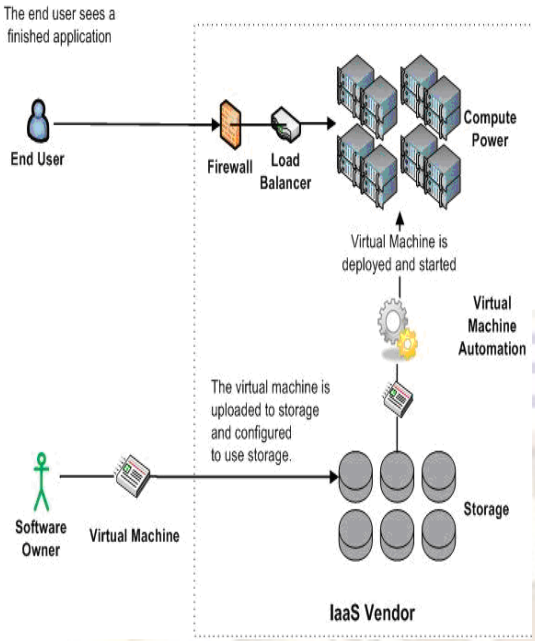


FIG 2.1.1 CLOUD-INFRASTRUCTURE AS SERVICE

2.2.1 CONSUMER'S VIEW ON IAAS

- Enable users to access applications from anywhere
- A modular system, which is flexible, scalable, virtualized and automated.
- Resilient and always available
- Enable to put applications and data on platform provisioning & maintenance by provider
- Own the hardware & nuances about provisioning & maintaining the OS & hygiene facts like space and power etc

2.2.2 PROVIDER'S VIEW ON IAAS

Provide virtual infrastructure Responsible for provisioning of space, power & cooling. Deploy web based applications to easily provision infrastructure for customer on demand. Responsible to provide load balancing services. Eases the process of cloning apps on additional infrastructure instances. Service level agreements with customers on "availability of infrastructure services" In a dense, shared, and pooled environment, the security of CPUs, data, and network is paramount. Account Management & Provisioning.

Table 1: **IaaS Summary**

Offering	Compute power, storage, and networking infrastructure. Some IaaS vendors may also provide Cloud Services.
Unit of deployment	Virtual Machine Image
Pricing structure	Compute usage per hour, data transfer in/out per GB, IO requests per million, storage per GB, data transfer in/out to storage per GB, data storage requests per thousand. All charges per billing period.
Customer	Software owner that would like an application hosted in the internet for their end users.
Examples	Amazon, GoGrid, and Rackspace.

FIGURE 2.2.1 IaaS SUMMARY

an Infrastructure as a Service (IaaS) offering provides solid cost savings because the infrastructure associated with providing compute power, storage, and networking does not need to be purchased and maintained by the customer. These assets are the responsibility of the IaaS vendor and customers are only charged for what they use when they use it. Table 1 summarizes Infrastructure as a Service [9]. IaaS is also a flexible offering that often appeals to infrastructure architects. Infrastructure architects like IaaS because it provides an infrastructure based approach to outsourcing datacenter workloads to the Cloud. If an application can be virtualized it can be uploaded to an IaaS environment and run.

3.0. AMAZON ELASTIC COMPUTE CLOUD (AMAZON EC2)

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers. Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment. Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change. Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use. Amazon

EC2 provides developers the tools to build failure resilient applications and isolate themselves from common failure scenarios.

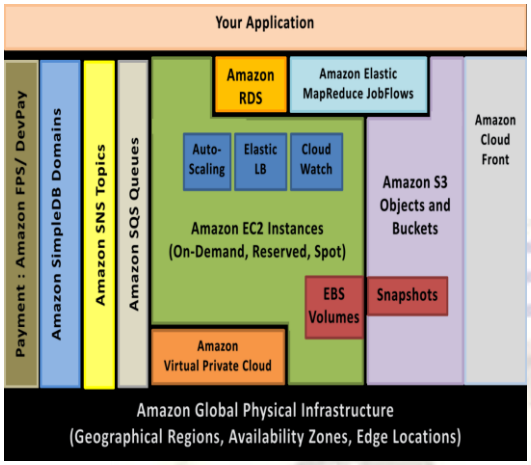


FIGURE 3.1.AMAZON WEB SERVICES

3.1AMAZONEC2 FUNCTIONALITY [4]

Amazon EC2 presents a true virtual computing environment, allowing you to use web service interfaces to launch instances with a variety of operating systems, load them with your custom application environment, manage your network's access permissions, and run your image using as many or few systems as you desire.

To use Amazon EC2, you simply:

✓ Select a pre-configured, template image to get up and running immediately. Or create an Amazon Machine Image (AMI) containing your applications, libraries, data, and associated configuration settings.

✓ Configure security and network access on your Amazon EC2 instance.

✓ Choose which instance type(s) and operating system you want, then start, terminate, and monitor as many instances of your AMI as needed, using the web service APIs or the variety of management tools provided.

✓ Determine whether you want to run in multiple locations, utilize static IP endpoints, or attach persistent block storage to your instances.

✓ Pay only for the resources that you actually consume, like instance-hours or data transfer.

3.2. SERVICES

3.2.1 ELASTIC – Amazon EC2 enables you to increase or decrease capacity within minutes, not hours or days. You can commission one, hundreds or even thousands of server instances simultaneously. Of course, because this is all controlled with web service APIs, your application can automatically scale itself up and down depending on its needs.
You have the choice of multiple instance types, operating systems, and software packages. Amazon EC2 allows you to select a configuration of memory, CPU, instance storage, and the boot partition size that is optimal for your choice of operating system and application.

3.2.2 RELIABLE – Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned. The service runs within Amazon's proven network infrastructure and datacenters. The Amazon EC2 Service Level Agreement commitment is 99.95% availability for each Amazon EC2 Region.[8]

3.2.3 SECURE – Amazon EC2 provides numerous mechanisms for securing your computer resources. Amazon EC2 includes web service interfaces to configure firewall settings that control network access to and between groups of instances. When launching Amazon EC2 resources within Amazon Virtual Private Cloud (Amazon VPC), you can isolate your compute instances by specifying the IP range you wish to use, and connect to your existing IT infrastructure using industry-standard encrypted IPsec VPN.

3.2.4 INEXPENSIVE – Amazon EC2 passes on to you the financial benefits of Amazon's scale. You pay a very low rate for the compute capacity you actually consume. See Amazon EC2 Instance Purchasing Options for a more detailed description.

On-Demand Instances – On-Demand Instances let you pay for compute capacity by the hour with no long-term commitments. This frees you from the costs and complexities of planning, purchasing, and

maintaining hardware and transforms what are commonly large fixed costs into much smaller variable costs. On-Demand Instances also remove the need to buy "safety net" capacity to handle periodic traffic spikes.

Reserved Instances – Reserved Instances give you the option to make a low, one-time payment for each instance you want to reserve and in turn receive a significant discount on the hourly usage charge for that instance. After the one-time payment for an instance, that instance is reserved for you, and you have no further obligation; you may choose to run that instance for the discounted usage rate for the duration of your term, or when you do not use the instance, you will not pay usage charges on it.

Spot Instances – Spot Instances allow customers to bid on unused Amazon EC2 capacity and run those instances for as long as their bid exceeds the current Spot Price. The Spot Price changes periodically based on supply and demand, and customers whose bids meet or exceed it gain access to the available Spot Instances. If you have flexibility in when your applications can run, Spot Instances can significantly lower your Amazon EC2 costs.

4.0 FEATURES [4]

Amazon EC2 provides a number of powerful features for building scalable, failure resilient, enterprise class applications, including:

4.1. Amazon Elastic Block Store – Amazon Elastic Block Store (EBS) offers persistent storage for Amazon EC2 instances. Amazon EBS volumes provide off-instance storage that persists independently from the life of an instance. Amazon EBS volumes are highly available, highly reliable volumes that can be leveraged as an Amazon EC2 instance's boot partition or attached to a running Amazon EC2 instance as a standard block device. When used as a boot partition, Amazon EC2 instances can be stopped and subsequently restarted, enabling you to only pay for the storage resources used while maintaining your instance's state. Amazon EBS volumes offer greatly improved durability over local Amazon EC2 instance stores, as Amazon EBS volumes are automatically replicated on the backend (in a single Availability Zone).

4.2 Multiple Locations – Amazon EC2 provides the ability to place instances in multiple locations. Amazon EC2 locations are composed of Regions and Availability Zones. Availability Zones are distinct locations that are engineered to be insulated from failures in other Availability Zones and provide inexpensive, low latency network connectivity to other Availability Zones in the same Region. By launching instances in separate Availability Zones, you can protect your applications from failure of a single location. Regions consist of one or more Availability Zones, are geographically dispersed, and will be in separate geographic areas or countries. The Amazon EC2 Service Level Agreement commitment is 99.95% availability for each Amazon EC2 Region.

4.3. Elastic IP Addresses – Elastic IP addresses are static IP addresses designed for dynamic cloud computing. An Elastic IP address is associated with your account not a particular instance, and you control that address until you choose to explicitly release it. Unlike traditional static IP addresses, however, Elastic IP addresses allow you to mask instance or Availability Zone failures by programmatically remapping your public IP addresses to any instance in your account. Rather than waiting on a data technician to reconfigure or replace your host, or waiting for DNS to propagate to all of your customers, Amazon EC2 enables you to engineer around problems with your instance or software by quickly remapping your Elastic IP address to a replacement instance.

4.4 Amazon Virtual Private Cloud – Amazon VPC is a secure and seamless bridge between a company's existing IT infrastructure and the AWS cloud. Amazon VPC enables enterprises to connect their existing infrastructure to a set of isolated AWS compute resources via a Virtual Private Network (VPN) connection, and to extend their existing management capabilities such as security services, firewalls, and intrusion detection systems to include their AWS resources.

4.5. Amazon Cloud Watch – Amazon Cloud Watch is a web service that provides monitoring for AWS cloud resources and applications, starting with Amazon EC2. It provides you with visibility into resource utilization, operational performance, and

overall demand patterns—including metrics such as CPU utilization, disk reads and writes, and network traffic.

4.6 Auto Scaling – Auto Scaling allows you to automatically scale your Amazon EC2 capacity up or down according to conditions you define. Auto Scaling is enabled by Amazon Cloud Watch and available at no additional charge beyond Amazon Cloud Watch fees.

4.7 Elastic Load Balancing – Elastic Load Balancing automatically distributes incoming application traffic across multiple Amazon EC2 instances. It enables you to achieve even greater fault tolerance in your applications, seamlessly providing the amount of load balancing capacity needed in response to incoming application traffic. Elastic Load Balancing detects unhealthy instances within a pool and automatically reroutes traffic to healthy instances until the unhealthy instances have been restored. You can enable Elastic Load Balancing within a single Availability Zone or across multiple zones for even more consistent application performance. Amazon Cloud Watch can be used to capture a specific

4.8 High Performance Computing (HPC) Clusters – Customers with complex computational workloads such as tightly coupled parallel processes, or with applications sensitive to network performance, can achieve the same high compute and network performance provided by custom-built infrastructure while benefiting from the elasticity, flexibility and cost advantages of Amazon EC2. Cluster Compute and Cluster GPU Instances have been specifically engineered to provide high-performance network capability and can be programmatically launched into clusters – allowing applications to get the low-latency network performance required for tightly coupled, node-to-node communication.

4.9 VM Import – VM Import enables you to easily import virtual machine images from your existing environment to Amazon EC2 instances. VM Import allows you to leverage your existing investments in the virtual machines that you have built to meet your IT security, configuration management, and compliance requirements by seamlessly bringing those virtual machines into Amazon EC2 as ready-to-use instances.

7.0 CONCLUSION

In emphasizing the cost and performance benefits of the cloud, some fundamental security problems have receded into the background and been left unresolved. Several critical pieces of technology, such as a solution for federated trust, are not yet fully realized, impinging on successful deployments. Determining the security of complex computer systems is also a long-standing security problem that overshadows large scale computing in general. Attaining the high assurance qualities in implementations has been an elusive goal of computer security researchers and practitioners, and is also a work in progress for cloud computing. Security of the cloud infrastructure relies on trusted computing and cryptography. Organizational data must be protected in a manner consistent with policies, whether in the organization's computing center or the cloud. No standard service contract exists that covers the ranges of cloud services available and the needs of different organizations. Having a list of common outsourcing provisions, such as privacy and security standards, regulatory and compliance issues, service level requirements and penalties, change management processes, continuity of service provisions, and termination rights, provides a useful starting point. The migration to a cloud computing environment is in many ways an exercise in risk management. Both qualitative and quantitative factors apply in an analysis. The risks must be carefully balanced against the available safeguards and expected benefits, with the understanding that accountability for security remains with the organization. Too many controls can be inefficient and ineffective, if the benefits outweigh the costs and associated risks. An appropriate balance between the strength of controls and the relative risk associated with particular programs and operations must be ensured.

8.0 REFERENCES

[1] Sushil Bhardwaj, Leena Jain, Sandeep Jain, "Cloud computing: a study of Infrastructure as a service (Iaas)" IJEIT 2010, 2(1), pp 1-4.

[2] L. Wang, G. Laszewski, M. Kunze and J. Tao, "Cloud computing: a perspective study", *J New Generation Computing*, 2010, pp 1-11.

[3] M. Dodani, "The Silver Lining of Cloud Computing", J. Object Technology, vol. 8(2), pp 29-38, 2009.

[4] P. Mell and T. Grance, "Cloud Computing Definition", National Institute of Standards and Technology, Version 15, 10-7-09.

[5] Craig Balding, "Assessing the Security Benefits of Cloud Computing", http://cloudsecurity.org/blog/2008/07/21/assessing-the-security-benefits-of-cloud-computing.html

[6] Sujay. R "Hybrid Cloud: A New Era", International Journal of Computer Science and Technology, ISSN: 2229-4333(Print) ISSN :0976 -8491 (Online), June 2011.

[7] Rupinder Kaur, "Cloud computing", International Journal of Computer Science and Technology, IJCST Vol. 2, Issue 3, September 2011.