1. Literature Review

The economics of the Cloud
Before we delve into how to architect an application for a cloud computing environment, we should explain why it is financially advantageous to do so.

The first clear advantage of using an existing cloud infrastructure is that you don’t have to make the capital investment yourself. Rather than expending the time and cost to build a data center, you use someone else’s investment and turn what would have been a massive capital outlay into a manageable variable cost.

In the pay-per-cycle model of cloud computing, you can start small and requisition more computer time as your business grows. This makes starting up inexpensive, and gives you time to build your on-demand business before investing in additional capacity. Instead of investing ahead of demand, you simply use and pay for exactly what you need when you need it.

Development time can also be a significant cost in creating an on-demand application environment. If you adopt a SaaS model, your entire application must be re-architected to support multi-tenancy. With cloud computing, the cost of a machine year in the Amazon EC2 cloud (~$880 annually) is much less than the cost of a fully-loaded developer (anywhere from $400-$1000 per day). This makes it a lot less expensive to scale up more virtual servers in the cloud than it is to spend even one day on development.

Finally, you can save money by designing your application with a simpler architecture ideal for cloud computing, which we’ll spend the rest of this paper discussing. A simpler architecture speeds time to market because it is easier to test, and you can eliminate some of the equipment and processes required to migrate an application from development into production. All the activities involved with development, test, QA and production can exist side-by-side in separate instances running in the cloud.

Architectural Considerations
Designing an application to run as a virtual appliance in a cloud computing environment is very different than designing it for an on-premise or SaaS deployment. We discuss the following considerations. To be successful in the cloud, your application must be designed to scale easily, tolerate failures and include management tools.
Scale

Cloud computing offers the potential for nearly unlimited scalability, as long as the application is designed to scale from the outset. The best way to ensure this is to follow some basic application design guidelines:

Start simple: Avoid complex design and performance enhancements or optimizations in favor of simplicity. It’s a good idea to start with the simplest application and rely on the scalability of the cloud to provide enough servers to ensure good application performance. Once you’ve gotten some traction and demand has grown, then you can focus on improving the efficiency of your application, which allows you to serve more users with the same number of servers or to reduce the number of servers while maintaining performance. Some common design techniques to improve performance include caching, server affinity, multi-threading and tight sharing of data, but they all make it more difficult to distribute your application across many servers. This is the reason you don’t want to introduce them at the outset and only consider them when you need to and can ensure that you are not breaking horizontal scalability.

Split application functions and couple loosely: Use separate systems for different pieces of application functionality and avoid synchronous connections between them. Again, as demand grows, you can scale each one independently instead of having to scale the entire application when you hit a bottleneck. The separation and reusability of functions inherent in SOA make it an ideal architecture for the cloud.

Network communication: Design the application to use network-based interfaces and not inter process communication or file-based communication paradigms. This allows you to effectively scale in the cloud because each piece of the application can be separated into distinct systems.

Consider the cluster: Rather than scale a single system up to serve all users, consider splitting your system into multiple smaller clusters, each serving a fraction of the application load. This is often called “sharding” and many web services can be split up along one dimension, often users or account. Requests can then be directed to the appropriate cluster based on some request attribute or users can be redirected to a specific cluster at login. To deploy a clustered system, determine the right collection of servers that yield efficient application performance, taking any needed functional redundancy into account; for example, 2 web, 4 application and 2 database servers. You can then scale the application by replicating the ideal cluster size and splitting the system load across the servers in the clusters.
The National Institute of Standards and Technology (NIST) defines cloud computing as “a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., network, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics and three service models and four deployment models”. Figure 2 shows the framework of the NIST definition of cloud computing:

**Cloud Computing: What It Is — And What It Isn't**

You know how traditional desktop computing works. The software programs you use are stored on each computer you own. The documents you create are stored on the computer on which they were created. And, although documents can be accessed from other computers on a network, they can’t be accessed by computers outside the network.
With cloud computing, the software programs you use aren't run from your personal computer, but are rather stored on servers housed elsewhere and accessed via the Internet. If your individual computer crashes, the software is still available for others to use.

The same goes for the documents you create; they're stored on a collection of servers accessed via the Internet. Anyone with permission can not only access the documents, but can also edit and collaborate on those documents in real time—which is a real plus over the traditional desktop computing model.

To some, cloud computing might sound a little like network computing—but it isn't. With network computing, applications/documents are hosted on a single company's server and accessed over the company's network. Cloud computing is a lot bigger than that. It encompasses multiple companies, multiple servers, and multiple networks. Plus, unlike network computing, cloud services and storage are accessible from anywhere in the world over an Internet connection; with network computing, access is over the company's network only.

The key difference between network computing and cloud computing is the "cloud" itself. That cloud is a large group of interconnected computers that typically extends beyond a single company or enterprise. The applications and data served by the cloud are available to a broad group of users using different operating system platforms; access is via the Internet. Any authorized user can access these docs and apps from any computer over any Internet connection, using the common web browser. It isn't apparent (and, in most cases doesn't matter) whether cloud services are based on HTTP, HTML, XML, JavaScript, or other specific technologies; to the user, the technology and infrastructure behind the cloud is invisible.

In short, cloud computing enables a shift from the computer to the user, from applications to tasks, and from isolated data to data that can be accessed from anywhere and shared with anyone. The user no longer has to take on the task of data management; he doesn't even have to remember where the data is. All that matters is that the data is in the cloud, and thus immediately available to that user and to other authorized users.
Inside the Cloud: How Cloud Computing Works

In cloud computing, a network of computers functions as a single computer to serve data and applications to users over the Internet. The network exists in the "cloud" of IP addresses that we know as the Internet, offers massive computing power and storage capability, and enables wide-scale group collaboration.

What, exactly, is this "cloud"? Put simply, the cloud is a collection of computers and servers that are publicly accessible via the Internet. This hardware is typically owned and operated by a third party in one or more data center locations. The machines can run any combination of operating systems; it's the processing power of the machines that matter, not what their desktops look like.

Individual users connect to the cloud from their own personal computers or portable devices (such as Apple's iPhone), over the Internet. To these individual users, the cloud is seen as a single application, device, or document. The hardware in the cloud (and the operating system that manages the hardware connections) is invisible.

This cloud architecture is deceptively simple, although it does require some intelligent management to connect all those computers together and assign task processing to multitudes of users. Each cloud uses various monitoring and metering functions to track usage so that resources are apportioned and attributed to the proper user(s).

This automation of management tasks is key to the notion of cloud computing. The system isn't a cloud if it requires human management to allocate processes to resources. For a system to attain cloud status, manual management must be replaced by automated processes.

Understanding Cloud Storage

One of the primary uses of cloud computing is for offsite data storage. With cloud storage, data is stored on multiple third-party servers, rather than on the dedicated servers used in traditional networked data storage.

When storing data, the user sees a virtual server—that is, it appears as if the data is stored in a particular place with a specific name. But that place doesn't exist in reality; it's just a pseudonym
used to reference virtual space carved out of the cloud. In reality, the user's data could be stored on any one or more of the computers used to create the cloud. The actual storage location may differ from day to day or even minute to minute, as the cloud dynamically manages available storage space. But even though the location is virtual, the user sees a "static" location for his data—and can actually manage his storage space as if it were connected to his own PC.

Cloud storage has both financial and security advantages over traditional storage models. Financially, the cloud's virtual resources are typically cheaper than dedicated physical resources connected to a personal computer or network. As for security, data stored in the cloud is secure from accidental erasure or hardware crashes, because it is duplicated across multiple physical machines; because multiple copies of the data are kept continually, the cloud continues to function as normal even if one or more machines go offline. If one machine crashes, the data is duplicated on other machines in the cloud.

That said, the ultimate reliability and security of cloud storage is yet to be determined. While companies offering cloud services tout their security features, there are many stages in the process where security could be breached, just as with a traditional network. And access to your data could be compromised if the cloud storage provider has a service outage or goes out of business—both of which have happened in recent months. It may be too early in the adoption process to recommend that a company store its data exclusively in the cloud, without some sort of traditional physical backup.

**Service Models:**

Once a cloud is established, how its cloud computing services are deployed in terms of business models can differ depending on requirements. The primary service models being deployed are commonly known as:
Software as a Service (SaaS) — Consumers purchase the ability to access and use an application or service that is hosted in the cloud. A benchmark example of this is Salesforce.com, as discussed previously, where necessary information for the interaction between the consumer and the service is hosted as part of the service in the cloud.

Platform as a Service (PaaS) — Consumers purchase access to the platforms, enabling them to deploy their own software and applications in the cloud. The operating systems and network access are not managed by the consumer, and there might be constraints as to which applications can be deployed.

Infrastructure as a Service (IaaS) — Consumers control and manage the systems in terms of the operating systems, applications, storage, and network connectivity, but do not themselves control the cloud infrastructure.

Deployment Model

Deploying cloud computing can differ depending on requirements, and the following four deployment models have been identified, each with specific characteristics that support the needs of the services and users of the clouds in particular ways (see Figure).
• Private Cloud — The cloud infrastructure has been deployed, and is maintained and operated for a specific organization. The operation may be in-house or with a third party on the premises.

There are two variations of private clouds:

• **On-Premise Private Cloud:** This format, also known as an “internal cloud,” is hosted within an organization’s own data centre. It provides a more standardized process and protection, but is often limited in size and scalability. Also, a firm’s IT department would incur the capital and operational costs for the physical resources with this model. On-premise private clouds are best used for applications that require complete control and configurability of the infrastructure and security.

• **Externally-Hosted Private Cloud:** This private cloud model is hosted by an external cloud computing provider. The service provider facilitates an exclusive cloud environment with full guarantee of privacy. This format is recommended for organizations that prefer not to use a public cloud infrastructure due to the risks associated with the sharing of physical resources.
Some of the characteristics of private cloud are:

1) *Enhanced Security Measures:* Security has become one of the primary concerns for many organizations especially for financial institutions. Let’s take a bank or a mortgage company, the confidentiality and security of their critical data is the utmost concern. Virtual private cloud computing comes equipped with a customizable and thorough firewall and a plethora of security tools which ensure maximum protection against unauthorized use, hacking and other such malicious attempts.

2) *Dedicated Resources:* The essence of private cloud is “no compromise”. As a subscriber to private cloud computing an enterprise has its own dedicated resources such as processor time and data buses which ensure optimum performance.

3) *Greater Customization:* Private cloud services are acquiescent and customizable so they can be molded to suit the exact requirements of an enterprise. This in turns bestows the enterprise with more control over their data

- Community Cloud — The cloud infrastructure is shared among a number of organizations with similar interests and requirements. This may help limit the capital expenditure costs for its establishment as the costs are shared among the organizations. The operation may be in-house or with a third party on the premises.

- Public Cloud — The cloud infrastructure is available to the public on a commercial basis by a cloud service provider. This enables a consumer to develop and deploy a service in the cloud with very little financial outlay compared to the capital expenditure requirements normally associated with other deployment options.

The public cloud, offer applications, storage and other services to the general public by a service provider. This is based on “pay-as-you-go” model. A public cloud is constructed with a view to offer unlimited storage space and increased bandwidth via Internet to all businesses. Public clouds are owned, hosted and operated by third-party service providers. A public cloud caters to all kind of requirements from small, medium or big businesses. A public cloud is the most
simplest to setup as it liberates that subscriber from woes of hardware, application and bandwidth expenses. Enterprises pay for only those condiments which they are utilizing. Users have to pay a monthly bill for public cloud services. Public cloud functions on the prime principle of storage demand scalability, which means it requires no hardware device. Popular examples of public clouds include Amazon Elastic Cloud Compute, Google App Engine, Blue Cloud by IBM and Azure services Platform by Windows.

Public cloud caters to four basic characteristics that are as follows:

**Flexible and Elastic Environment:** Public clouds like Google App engine or Amazon elastic cloud compute offers its users highly flexible cloud environment. They enable users to share and store data as per their personal capacities. They can decide what to share and what not to share with their clients.

1) **Freedom of Self-Service:** Public clouds encourage users to create a cloud on their own without taking anyone's help. These are pre-configured clouds existing on Internet. The only thing businesses that wish to opt for public cloud need to do is to visit public cloud portals and get started with it. You don’t have to depend upon on any third party help to create or run this type of cloud. It will be managed and handled by you as you will be the prime proprietor of it.

2) **Pay for what you use:** This particular characteristic enables cloud technology more accessible for businesses to work in a synchronized fashion. The more you use cloud services, better will be the future business prospects. However, payment is charged on the basis of cloud services used by users.

3) **Availability and Reliability:** Yet another feature of public cloud is that, it is available to all and believes in agility. You can catch up with your work any time you wish and from any corner of the globe. Not only users become more independent in running important business tasks but also more efficient in strengthening customer relations across the globe.
Hybrid Cloud — The cloud infrastructure consists of a number of clouds of any type, but the clouds have the ability through their interfaces to allow data and/or applications to be moved from one cloud to another. This can be a combination of private and public clouds that support the requirement to retain some data in an organization, and also the need to offer services in the cloud.

Some of the characteristics of hybrid clouds are:

1) **Optimal utilization:** The available server resources in typical data centers are actually used from 5-20%. This is because of peak loads which are ten times higher than that of the average load. Hence, servers are mostly idle - generating unnecessary costs. Hybrid clouds can increase server utilization by scaling out to public resources to handle crowds.

2) **Data centre consolidation:** Instead of providing the capacity to cope for worst-case scenarios, a private cloud only needs resources in average cases. The option to burst out allows server consolidation and hence resulting in reduction of operating costs. In particular, this includes the costs for hardware, power, cooling, maintenance, and administration.

3) **Risk transfer:** The companies themselves are responsible for keeping up and running their data centre and private cloud. The public cloud provider has to ensure a high uptime for their service.
Using a hybrid cloud model, “the risk of misestimating workload is shifted to the cloud vendor from the service operator”. Most of the cloud providers have service level agreements which ensure an uptime of more than 99.9% per year, i.e., a downtime of max. 9 hours per year.

4) **Availability:** The high availability in the corporate data centre is difficult and expensive, because it requires redundancy, backups, and geographic dissemination. Especially in companies where IT is not the core business, the expertise in this area is rather limited. In a hybrid cloud environment, the public cloud can scale up or take over operations completely if the company’s data centre is unavailable due to failures or Distributed Denial of Service (DDoS) attacks.

**The attributes of cloud computing are:**

1) **Service Based:** Consumer concerns are abstracted from provider concerns through service interfaces that are well-defined. The interfaces hide the implementation details and enable a completely automated response by the provider of the service to the consumer of the service.

2) **Scalable and Elastic:** The service can scale capacity up or down as the consumer demands at the speed of full automation (which may be seconds for some services and hours for others). Elasticity is a trait of shared pools of resources.

3) **Shared:** Services share a pool of resources to build economies of scale. IT resources are used with maximum efficiency. The underlying infrastructure, software or platforms are shared among the consumers of the service (usually unknown to the consumers). This enables unused resources to serve multiple needs for multiple consumers, all working at the same time.

4) **Metered by Use:** Services are tracked with usage metrics to enable multiple payment models. The service provider has a usage accounting model for measuring the use of the services, which could then be used to create different pricing plans and models. These may include pay-as-you-go plans, subscriptions, fixed plans and even free plans. The implied payment plans will be based on usage, not on the cost of the equipment.
5) Uses Internet Technologies: The service is delivered using Internet identifiers, formats and protocols, such as URLs, HTTP, IP and representational state transfer Web-oriented architect.

Benefits of cloud computing:

The most frequently cited benefits of cloud computing are:

- It is agile, with ease and speed of deployment
- Its cost is use-based, and will likely be reduced
- In-house IT costs are reduced
- Capital investment is reduced
- The latest technology is always delivered
- The use of standard technology is encouraged and facilitated

As an application moves to the cloud, the access to it becomes more simple and ubiquitous. Low cost ultra light devices and inexpensive hand held devices build on latest operating systems such as android provide access to the internet, the number and types of tasks taking advantage of the new technology will increase by several orders of magnitude, going far beyond the comparatively modest list of things that we use computers and the Internet for today.