Available at: www.researchpublications.org

Analysis & Architecture of Cloud Computing

Prof.Parag.P.Gudadhe
Lecturer in Elect & Telecomm
VYWS Polytechnic Badnera
parag_gudadhe@rediff.com

Prof.Mrs.M.S.Deshmukh
H.O.D Elect & Telecomm
VYWS Polytechnic Badnera
madhurisdeshmukh@gmail.com

Prof.Milind.R.Dhande
Lecturer in Computer Tech
VYWS Polytechnic Badnera
milinddhande8852@rediffmail.com

ABSTRACT

A computing platform for the next generation of the internet. The paper defines clouds explain the business benefits of cloud computing and out line cloud architecture with its major components. This paper will discover how a business can used cloud computing to further innovation and reduced its cost. Introduction enterprises strive to reduce computing costs. Many start by consolading their IT operation & later introducing virtualization technologies. Cloud computing takes this steps to a new level & allows an organization to further reduce cost through improved utilization, reduced administration and infrastructure costs, and faster deployment cycles. The cloud is a next generation platform that provides dynamic resource pools, virtualization, and high availability. Cloud computing describes both a platform and a type of application. A cloud computing platform dynamically provisions, configure, reconfigures. Cloud applications are applications that are extended to be accessible through the internet. These cloud applications use large data centers and powerful servers that host web applications and web services.

Keywords

Cloud Technology, Cloud Computing, Cloud Services, Xen Hypervisor.

1. INTRODUCTION

Cloud computing infrastructure accelerates and fasters the adoption of innovations Enterprises are increasingly making innovation their highest priority. They realize the need to seek new ideas and unlock new sources of value. Cloud computing enables innovation. It alleviates the need of innovators to find resources to develop, test, and make their innovations available to the user community. Innovators are free to focus on the innovation rather than the logistics of finding and managing resources that enable the innovation. Cloud computing helps leverage innovation as early as possible to deliver business value. Fostering innovation requires unprecedented flexibility and responsiveness. The enterprise should provide an ecosystem where innovators are not hindered by excessive processes, rules, and resource constraints. In this context, a cloud computing service is a necessity. It comprises an automated framework that can deliver standardized services quickly and cheaply. Cloud computing is a term used to describe both a platform and type of application. A cloud computing platform dynamically provisions, configures, reconfigures, and deprovisions servers as needed. Servers in the cloud can be physical machines or virtual machines. Advanced clouds typically include other computing resources such as storage area networks (SANs), network equipment, firewall and other security devices. Cloud computing also describes applications that are extended to be accessible through the Internet. 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.

A cloud is a collection of virtualized computer resources. What happens when applications and services are moved into the internet "cloud." Cloud computing is not something that suddenly appeared overnight; in some form it may trace back to a time when computer systems remotely time-shared computing resources and applications. More currently though ,cloud computing refers to the many different types of services and applications being delivered in the internet cloud, and the fact that, in many cases, the devices used to access these services and applications do not require any special applications. Many companies are delivering services from the cloud.

2. BENEFITS AND CHARACTERISTICS

Cloud computing infrastructures can allow enterprises to achieve more efficient use of their IT hardware and software investments. They do this by breaking down the physical barriers inherent in isolated systems, and automating the management of the group of systems as a single entity. Cloud computing is an example of an ultimately virtualized system, and a natural evolution for data centers that employ automated systems management, workload balancing, and virtualization technologies. A cloud infrastructure can be a cost efficient model for delivering information services, reducing IT management complexity, promoting innovation, and increasing responsiveness through real time workload balancing. The Cloud makes it possible to launch Web 2.0 applications quickly and to scale up applications as much as needed when needed. The platform supports traditional Java and Linux, Apache, MySQL, PHP (LAMP) stack-based applications as well as new architectures such as Map Reduce and the File System, which provide a means to scale applications across thousands of servers instantly architecture[1].

Some characteristics of cloud are as follows -

Shared Infrastructure- Uses a virtualized software model, enabling the sharing of physical services, storage, and networking capabilities. The cloud infrastructure, regardless of deployment model, seeks to make the most of the available infrastructure across a number of users.

Dynamic Provisioning- Allows for the provision of services based on current demand requirements. This is done automatically using software automation, enabling the expansion and contraction of service capability, as needed. This dynamic scaling needs to be done while maintaining high levels of reliability and security.

Network Access - Needs to be accessed across the internet from a broad range of devices such as PCs, laptops, and mobile devices, using standards-based APIs (for example, ones based on HTTP). Deployments of services in the cloud include everything from using business applications to the latest application on the newest smart phones.

Managed Metering - Uses metering for managing and optimizing the service and to provide reporting and billing information .In this way, consumers are billed for services according to how much they have actually used during the billing period .In short, cloud computing allows for the sharing and scalable deployment of services, as needed, from almost any location, and for which the customer can be billed based on actual usage.

3. SERVICE AND DEPLOYMENT MODELS

3.1 Service Models

Once a cloud is established, how its cloud computing services are deployed in terms of business models can differ depending on requirements.

Software as a Service- 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. Also, Microsoft is expanding its involvement in this area, and as part of the cloud computing option for Microsoft® Office 2010, its Office Web Apps are available to Office volume licensing customers and Office Web App subscriptions through its cloud-based Online Services.

Platform as a Service- Consumers purchase access to the platforms, enabling them to organize 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- Consumers control and manage the systems in terms of the operating systems, applications, storage, and network connectivity, but do not they control the cloud infrastructure. Also known are the various subsets of these models that may be related to a particular industry or market.

Communications as a Service (CaaS) - One such subset model used to describe hosted IP telephony services. Along with the move to CaaS is a shift to more IP-centric communications and more SIP trucking deployments. With IP and SIP in place, it can be as easy to have the PBX in the cloud as it is to have it on the premise use a 9-point Times Roman font, or other Roman font with serifs, as close as possible in appearance to Times Roman in which these guidelines have been set. The goal is to have a 9-point text, as you see here. Please use sans-serif or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

3.2 Deployment Models

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

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.

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.

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.

4. TECHNOLOGIES

4.1 Cloud Computing Application Architecture

This gives the basic architecture of a cloud computing application. We know that cloud computing is the shift of computing to a host of hardware infrastructure that is distributed in the cloud. The commodity hardware infrastructure consists of the various low cost data servers that are connected to the system and provide their storage and processing and other computing resources to the application. Cloud computing involves running applications on virtual servers that are allocated on this distributed hardware infrastructure available in the cloud. These virtual servers are made in such a way that the different service level agreements and reliability issues are met. There may be multiple instances of the same virtual server accessing the different parts of the hardware infrastructure available. This is to make sure that there are multiple copies of the applications which are ready to take over on another one in case of failure. The virtual server distributes the processing between the infrastructure and the computing is done and the result returned. There will be a workload distribution management system, also known as the grid engine, for managing the different requests coming to the virtual servers. This engine will take care of the creation of multiple copies and also the preservation of integrity of the data that is stored in the infrastructure. This will also adjust itself such that even on heavier load, the processing is completed as per the requirements. The different workload management systems are hidden from the users. For the user, the processing is done and the result is obtained. There is no question of where it was done and how it was done. The users are billed based on the usage of the system - as said before - the commodity is now cycles and bytes. The billing is usually on the basis of usage per CPU per hour or GB data transfer per hour. Cloud computing makes use of a large physical resource pool in the cloud. As said above, cloud computing services and applications make use of virtual server instances built upon this resource pool. There are two applications which help in managing the server instances, the resources and also the management of the resources by these virtual server instances.

4.2 Xen Hypervisor

Xen is an open-source, which makes it possible to run many instances of an operating system or indeed different operating systems in parallel on a single machine (or host). Xen is the only type-1 hypervisor that is available as open source. Xen is used as the basis for a number of different commercial and open source applications, such as: server virtualization, Infrastructure as a Service (IaaS), desktop virtualization, security applications, embedded and hardware appliances. Xen enables users to increase server utilization, consolidate server farms, reduce complexity, and decrease total cost of ownership which provides an abstraction layer between the hardware and the virtual OS so that the distribution of the resources and the processing is well managed. Another application that is widely used is the Enomalism server management system which is used for management of the infrastructure platform. When Xen is used for virtualization of the servers over the infrastructure, a thin software layer known as the Xen hypervisor is inserted between the server hardware and the operating system. This provides an abstraction layer that allows each physical server to run one or more virtual servers, effectively decoupling the operating system and its applications from the underlying physical server. The Xen hypervisor is a unique open source technology, developed collaboratively by the Xen community and engineers at over 20 of the most innovative data center solution vendors, including MD, Cisco, Dell, HP, IBM, Intel, Mellanox, Network Appliance, Novell, Red Hat, SGI, Sun, Unisys, Veritas, Voltaire, and Citrix. Xen is licensed under the GNU General Public License (GPL2) and is available at no charge in both source and object format. The Xen hypervisor is also exceptionally use that translates to extremely low overhead and near-native performance for guests. Xen re-uses existing device drivers (both closed and open source) from Linux, making device management easy. More over Xen is robust to device driver failure and protects both guests and the hypervisor from faulty or malicious drivers The Enomalism virtualized server management system is a complete virtual server infrastructure platform. Enomalism helps in an effective management of the resources. Enomalism can be used to tap into the cloud just as you would into a remote server. It brings together all the features such as deployment planning, load balancing, resource monitoring, etc. Enomalism is an open source application. It has very simple and easy to use web based user interface. It has a module architecture which allows for the creation of additional system add-ons and plugins. It supports one click deployment of distributed or replicated applications on a global basis. It supports the management of various virtual environments including KVM/Qemu, Amazon EC2 and Xen, penVZ, Linux Containers, Virtual Box. It has fine grained user permissions and access privileges.

5. MAP REDUCTION

Map Reduce is a software framework developed at Google in 2003 to support parallel computations over large (multiple pita byte) data sets on clusters of commodity computers. This framework is largely taken from map and reduces functions commonly used in functional programming, although the actual semantics of the framework are not the same. It is a programming model and an associated implementation for processing and generating large data sets. Many of the real world tasks are expressible in this model. Map Reduce implementations have been written in C++, Java and other languages. Programs written in this functional style are automatically parallelized and executed on the cloud. The run-time system takes care of the details of partitioning the input data, scheduling the program execution across a set of machines, handling machine failures, and managing the required inter-machine communication. This allows programmers without any experience with parallel and distributed systems to easily utilize the resources of a largely distributed system. The computation takes a set of input key/value pairs, and produces a set of output key/value pairs. The user of the Map Reduce library expresses the computation as two functions: Map and Reduce. Map, written by the user, takes an input pair and produces a set of intermediate key/value pairs. The Map Reduce library groups together all intermediate values associated with the same intermediate key I and passes them to the Reduce The Reduce function, also written by the user, accepts an intermediate key I and a set of values for that key. It merges together these values to form a possibly smaller set of values. Typically just zero or one output value is produced per Reduce invocation. The intermediate values are supplied to the user reduce function via an iterator. This allows us to handle lists of values that are too large to fit in memory.

Map Reduce achieves reliability by parceling out a number of operations on the set of data to each node in the network; each node is expected to report back periodically with completed work and status updates. If a node falls silent for longer than that interval, the master node records the node as dead, and sends out the node assigned work to other nodes. Individual operations use atomic operations for naming file outputs as a double check to ensure that there are not parallel conflicting threads running; when files are renamed, it is possible to also copy them to another name in addition to the name of the task (allowing for side-effects).

5.1 Google File Systems

Google File System (GFS) is a scalable distributed file system developed by Google for data intensive applications. It is designed to provide efficient, reliable access to data using large clusters of commodity hardware. It provides fault tolerance while running on inexpensive commodity hardware, and it delivers high aggregate performance to a large number of clients. Files are divided into chunks of 64 megabytes, which are only extremely rarely overwritten, or shrunk; files are usually appended to or read. It is also designed and optimized to run on computing clusters, the nodes of which consist of cheap, commodity computers, which means precautions must be taken against the high failure rate of individual nodes and the subsequent data loss. Other design decisions select for high data throughputs, even when it comes at the cost of latency. The nodes are divided into two types: one Master node and a large number of Chunk servers. Chunk servers store the data files, with each individual file broken up into fixed size chunks (hence the name) of about 64 megabytes, similar to clusters or sectors in regular file systems. Each chunk is assigned a unique 64-bit label, and logical mappings of files to constituent chunks are maintained. Each chunk is replicated several times throughout the network, with the minimum being three, but even more for files that have high demand or need more redundancy. The Master server doesn't usually store the actual chunks, but rather all the metadata associated with the chunks, such as the tables mapping the 64-bit labels to chunk locations and the files they make up, the locations of the copies of the chunks, what processes are reading or writing to a particular chunk, or taking a snapshot of the chunk pursuant to replicating it (usually at the instigation of the Master server, when, due to node failures, the number of copies of a chunk has fallen beneath the set number). All this metadata is kept current by the Master server periodically receiving updates from each chunk server (Heart-beat messages). Permissions for modifications are handled by a system of time-limited, expiring leases, where the Master server grants permission to a process for a finite period of time during which no other process will be granted permission by the Master server to modify the chunk. The modified chunk server, which is always the primary chunk holder, then propagates the changes to the chunk servers with the backup copies. The changes are not saved until all chunk servers a acknowledge, thus guaranteeing the completion and atomicity of the operation. Programs access the chunks by first querying the Master server for the locations of the desired chunks; if the chunks are not being operated on (if there are no outstanding leases), the Master replies with the locations, and he program then contacts and receives the data from the chunk server directly. As opposed to many file systems, it not implemented in the kernel of an Operating System but accessed through a library to avoid overhead.

5.2 Hadoop

Hadoop is a framework for running applications on large cluster built of commodity hardware. The Hadoop framework transparently provides applications both reliability and data motion. Hadoop implements the computation paradigm named Map Reduce which was explained above. The application is divided into many small fragments of work, each of which may be executed or re-executed on any node in the cluster. In addition, it provides a distributed file system that stores data on the compute nodes, providing very high aggregate bandwidth across the cluster. Both Map Reduce and the distributed file system are designed so that the node failures are automatically handled by the framework. Hadoop has been implemented making use of Java. In Hadoop, the combination of the entire JAR files and classed needed to run a Map Reduce program is called a job. All of these components are themselves collected into a JAR which is usually referred to as the job file. To execute a job, it is submitted to a job Tracker and then executed. Tasks in each phase are executed in a fault tolerant manner. If node(s) fail in the middle of a computation the tasks assigned to them are re-distributed among there maiming nodes. Since we are using Map Reduce, having many map and reduce tasks enables good load balancing and allows failed tasks to be re-run with smaller runtime over head. The Hadoop Map Reduce framework has master/slave architecture. It has a single master server or a job Tracker and several slave servers or task Trackers, one per node in the cluster. The job Tracker is the point of interaction between the users and the framework. Users submit jobs to the job Tracker, which puts them in a queue of pending jobs and executes them on a first come first-serve basis. The job Tracker manages the assignment of Map Reduce jobs to the task Trackers. The task Trackers execute tasks upon instruction from the job Tracker and also handle data motion between the map and reduce phases of the Map Reduce job. Hadoop is a framework which has received a wide industry adoption. Hadoop is used along with other cloud computing technologies like the Amazon services so as to make better use of the resources. There are many instances where Hadoop has been used. Amazon makes use of Hadoop for processing millions of sessions which it uses for analytics. This is made use of in a cluster which has about 1 to 100 nodes. Facebook uses Hadoop to store copies of internal logs and dimension data sources a use it as a source for reporting/analytics and machine learning. The New York Times made use of Hadoop for large scale image conversions. Yahoo uses Hadoop to support research for advertisement systems and web searching tools. They also use it to do scaling tests to support development of Hadoop.

6. CLOUD COMPUTING SERVICES

Even though cloud computing is a pretty new technology, there are many companies offering cloud computing services. Different companies like Amazon, Google, Yahoo, IBM and Microsoft are all players in the cloud computing services industry. But Amazon is the pioneer in the cloud computing industry with services like EC2 (Elastic Compute Cloud) and S3(Simple Storage Service) dominating the industry. Amazon has an expertise in this industry and has a small advantage over the others because of this. Microsoft has good knowledge of the fundamentals of cloud science and is building massive data centers. IBM, the king of business computing and traditional supercomputers, teams up with Google to get a foothold in the clouds. Google is far and away the leader in cloud computing with the company itself built from the ground up on hardware.

6.1 Amazon Web Services

The Amazon Web Services is the set of cloud computing services offered by Amazon. It involves four different services. They are Elastic Compute Cloud (EC2), Simple Storage Service (S3), Simple Queue Service (SQS) and Simple Database Service (SDB).

6.2 Elastic Compute Cloud (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. It provides on-demand processing power. Amazon EC2â„¢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 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. Amazon EC2 presents a true virtual computing environment, allowing you to use web service interfaces to requisition machines for use, load them with your custom application environment, manage your network access permissions, and run your image using as many or few systems as you desire. To set up an Amazon EC2 node we have to create an EC2 node configuration which consists of all our applications, libraries, data and associated configuration settings. This configuration is then saved as an AMI (Amazon Machine Image). There are also several stock instances of Amazon AMIs available which can be customized and used. We can then start, terminate and monitor as many instances of the AMI as needed. Amazon EC2 enables you to increase or decrease capacity with in minutes. You can commission one, hundreds or even thousands of server instances simultaneously. Thus the applications can automatically scale it self up and down depending on its needs. You have root access to each one, and you can interact with them as you would any machine. You have the choice of several instance types, allowing you to select a configuration of memory, PU, and instance storage that is optimal for your application. Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and reliably commissioned. Amazon EC2 provides web service

interfaces to configure firewall settings that control network access to and between groups of instances. You will be charged at the end of each month for your EC2 resources actually consumed. So charging will be based on the actual usage of the resources.

6.3 Simple Storage Service

S3 or Simple Storage Service offers cloud computing storage service. It offers services for storage of data in the cloud. It provides a high-availability large-store database. It provides a simple SQL-like language. It has been designed for interactive online use. S3 is storage for the Internet. It is designed to make web-scale computing easier for developers. S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from any where on the web. It gives any developer access to the same highly scalable, reliable, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites. Amazon S3 allows write, read and delete of objects containing from 1 byte to 5 gigabytes of data each. The number of objects that you can store is unlimited. Each object is stored in a bucket and retrieved via a unique developer-assigned key. A bucket can be located anywhere in Europe or the Americas but can be accessed from anywhere. Authentication mechanisms are provided to ensure that the data is kept secure from unauthorized access. Objects can be made private or public, and rights can be granted to specific users for particular objects. Also the S3 service also works with a pay only for what you use method of payment.

6.4 Simple Queue Service

Amazon Simple Queue Service (SQS) offers a reliable, highly scalable, hosted queue for storing messages as they travel between computers. By using SQS, developers can simply move data between distributed components of their applications that perform different tasks, without losing messages or requiring each component to be always available With SQS, developers can create an unlimited number of SQS queues, each of which can send and receive an unlimited number of messages can be retained in a queue for up to 4 days. It is simple, reliable, secure and scalable.

6.5 Simple Database Service

Amazon Simple DB is a web service for running queries on structured data in real time. This service works in close conjunction with the Amazon S3 and EC2, collectively providing the ability to store, process and query data sets in the cloud. These services are designed to make web-scale computing easier and more cost-effective to developers. Traditionally, this type of functionality is accomplished with a clustered relational database, which requires a sizable upfront investment and often requires a DBA to maintain and administer them. Amazon SDB provides all these without the operational complexity. It requires no schema, automatically indexes your data and provides a simple API for storage and access. Developers gain access to the different functionalities from within the Amazon proven computing environment and are able to scale instantly and need to pay only for what they use.

6.6 Dynamic Web Serving

With full support for common web technologies persistent storage with queries, sorting and transactions. Automatic scaling and load balancing APIs for authenticating users and sending email using Google accounts fully featured local development environment that simulates Google App Engine on your computer Google App Engine applications are implemented using the Python programming language. The run time environment includes the full Python language and most of the Python standard library. Applications run in a secure environment that provides limited access to the provides limited access to the underlying operating system. These limitations allow App Engine to distribute web requests for the application across multiple servers, and start and stop servers to meet traffic demands. App Engine includes a service API for integrating with Google Accounts.Your application can allow a user to sign in with a Google account, and access the email address and displayable name associated with the account. Using Google Accounts lets the user start using your application faster, because the user may not need to create a new account. It also saves you the effort of implementing a user account system just for your application App Engine provides a variety of services that enable you to perform common operations when managing your application. The following APIs are provided to access these services: Applications can access resources on the Internet, such as web services or there data, using App Engine URL fetch service. Applications can send email messages using App Engine mail service. The mail service uses Google infrastructure to send email messages. The Image service lets your application manipulate images. With this API, you can resize, crop, rotate and flip images in JPEG and PNG formats. In theory, Google claims App Engine can scale nicely. But Google currently places a limit of 5 million hits per month on each application. This limit nullifies App Engine scalability, because any small, dedicated server can have this performance. Google will eventually allow webmasters to go beyond this limit.

7. CONCLUSION

Cloud computing is a powerful new abstraction for large scale data processing systems which is scalable, reliable and available. In cloud computing, there are large self-managed server pools available which reduces the overhead and eliminates management headache. Cloud computing services can also grow and shrink according to need. Cloud computing is particularly valuable to small and medium businesses, where effective and affordable IT tools are critical to helping them become more productive without spending lots of money on in-house resources and technical equipment. Also it is a new emerging architecture needed to expand the Internet to become the computing platform of the future.

8. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed directly or indirectly in development of the paper.

REFERENCES

[1] White Papers on Cloud
http://www.dialogic.com/~/media/products/docs/whitepapers/12023-cloud-computing-wp.pdf

[2] Application and Architecture of Cloud Computing
http://www.slideshare.net/whitepapers/application-architecture-for-cloud-computing

[3] Xen Hypervisor
http://xen.org/products/xenhyp.html

[4]Google File System by Jeffrey Dean and Sanjay]Ghemawat
http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en//archive/mapreduce-osdi04.pdf

[5] Cloud Services
http://aws.amazon.com/documentation/ec2