Unit III
Introduction to Cloud Computing

Abhirup Khanna
Department of Virtualization

Introduction to Cloud Computing

- 3.1 Virtualization and cloud
- 3.2 Overview to Cloud Computing
- 3.3 History of Cloud Computing
- 3.4 Anatomy of Cloud
- 4.4.1Provisioning and Configuration
- 4.4.2 Monitoring and Optimization
- 4.4.3 Metering and Chargeback
- 4.4.4 Orchestration
- 4.4.5 CMDB
- 4.4.6 CLM
- 4.4.7 SC

Overview to Cloud Computing…

- Cloud provides a mechanism for delivering IT as a service to business with the following characteristics:
- ***You do not need to own a datacenter to use hardware or software services***
- ***You have an option to subscribe or unsubscribe to services delivered in the cloud.***
- ***The cloud service provider will provide you a SLA (Service Level Agreement) for the services defined in the service catalog.***
- ***The cloud services could be accessed using a standard interface***

Contd…..

- ***The cloud services are available on demand***
- ***It can scale up or down based on your usage. The excess capacity goes into a general pool and used by someone else***
- ***Monitoring allows for more granular optimization of usage of resources in the cloud***
- ***The subscriber gets charged for the services subscribed and the periodic usage***
- ***A large part of cloud operations is automated. The process is known as orchestration.***

Service Framework for Cloud…

- Central to cloud technology is the information technology service framework – i.e. delivery of information technology as a service.
- Cloud ties business need to a technology definition.
- With Cloud, it becomes easier to convert a business requirement directly to a technology requirement.
- A **Service Catalog** defines what a particular instance of cloud is designed to provide to end-users.
- A **configuration management system** provides a central repository for all the configuration items present in the cloud.
- In short, *Cloud defines a framework to deliver IT as a service in the most efficient and the fastest way possible without a need to actually own the resources required to achieve the same, all along providing a level of transparency and monitoring not feasible in earlier paradigms.*

Virtualization and Cloud……….

- Virtualization and Cloud are sometimes confused as one technology in certain scenarios.
- Virtualization and Cloud are two distinct technologies.
- However, in certain aspects these are related to each other.

The overlapping aspects of Virtualization and Cloud….

- Virtualization is only an enabler for Cloud.
- Virtualization provides the required infrastructure flexibility to cloud by virtualizing the resources which allows for easy provisioning and management of these resources across hardware pools
- Virtualization is only one of the eight major building blocks of Cloud.
- Virtualization is most relevant to IaaS (Infrastructure as a Service).
- However, PaaS(Platform as a Service) and SaaS (Software as a Service) could largely be achieved without virtualization.
- Increasingly, PaaS and SaaS units are being packaged as VMs for easy deployment.
- The real difference between Cloud and Virtualization comes from the business aspects of Cloud and the Service management framework built into Cloud.

Contd……..

- Virtualization is a technology without no direct business benefits other than saving cost to business.
- On the other hand, Cloud connects a business requirement to technology and can be used to map and fulfill a business requirement.
- A virtualized infrastructure has limited automation built-in.
- However, Cloud based infrastructure is built around complete automation and orchestration.
- The elasticity and the flexibility in the Cloud are built on the foundation of Virtualization.
- **It is recommended and an industry practice to first consolidate a physical infrastructure using Virtualization before moving in for a Cloud Transformation**.

Contd….

- Virtualization and Cloud both brings saving to the business due to consolidation and the flexibility and increased manageability.

Areas of Savings…….

Parameter	Virtualization	Cloud
Utilization	Typically 60-70%	Typically 60-70%. However, it is possible to share resources across pools (physical hosts) and maintain the utilization levels. Spare capacity available on demand
Provisioning	Manual, 1 day	Automatic, On-demand within minutes or few hours
Monitoring	Comparative ease in monitoring using automated tools. However, need manual intervention to take care of any failures	Typically automated. No manual intervention required due to advanced orchestration capabilities built into the cloud
Sizing	Easier to resize. However, manual intervention required to resize for new requirements	On-demand automatic rescaling of the resources
Staff for Administration	Reduced number of Full Time employees	Typical reduction in number of employees required to manage the infrastructure. On an average, FTE reduction is generally in the range of 1 administrator for 400 cloud instances
Cost	Initial hardware cost reduced due to sharing of hardware assets and due to increased utilization	In most cases, the initial hardware cost is almost negligible. There'll be running cost on a monthly basis.
Optimization	Easy to share resources and re-balance loads on the virtual machines on the same host. However, re-balancing across physical hosts require advanced features and planned downtime	Easy to share resources and re-balance loads across resource pools. No manual intervention required to move resources or resize resources for an application.

Anatomy of Cloud…….has 8 Components

- 1. Provisioning and Configuration Module
- 2. Monitoring and Optimization
- 3. Metering and Chargeback
- 4. IT Service Management
- 5. Orchestration
- 6. CMDB (Configuration Management Database)
- 7. Cloud Lifecycle Management Layer
- 8. Service Catalog

The components….

- Each component serves a specific function in Cloud.
- **Not all components are present** in every Cloud solution available in the market.
- Cloud vendors may choose to implement a subset of features to suit their end-user requirements.
- The **choice of components** may also depend on the Delivery model.
- Some of these components **may also be partially implemented** depending on the deployment model used for Cloud.

Function of each cloud module…

- □ **Provisioning and Configuration :**
- Provisioning and Configuration layer forms the lowest layer of cloud and typically reside on bare hardware (as firmware) or on the top of the hypervisor layer.
- The function of this layer is to abstract the underlying hardware and provide a standard mechanism to spawn instance of virtual machines on demand.
- The layer also handles the post-configuration of the operating systems and applications residing on the VM.
- □ **Monitoring and Optimization :** This layer handles all the monitoring of all server, storage, networking and application components in Cloud.
- Based on this statistics, it could perform routine functions that optimize the behavior of the infrastructure components and provide relevant data to the cloud administrator to further optimize the configuration for maximum utilization and performance.
- □ **Metering and Chargeback :** This layer provides functions to measure the usage of resources in Cloud.
- The metering module collects all the utilization data per domain per user. The module gives the Cloud administrator enough data to measure ongoing utilization of resources and to create invoices based on the usage on a periodic basis.

Contd…

- ☐ **Orchestration :** Orchestration is central to Cloud operations. Orchestration converts requests from the
- Service management layer and the monitoring, chargeback modules to appropriate action items which are then submitted to provisioning and configuration module for final closure.
- Orchestration updates the CMDB (Configuration Management Database) in the process.
- ☐ **Configuration Management Database (CMDB):** CMDB is a central configuration repository wherein all the meta-data and configuration of different modules, resources is kept and updated on a real-time basis.
- The repository can then be accessed using standard protocols like SOAP by third party software and integration components.
- All updates to CMDB happen in real-time as requests get processed in Cloud.

Components detailed……..

- **Cloud Lifecycle Management Layer (CLM)**

The Layer handles the coordination of all other layers in Cloud. All requests internal and external are addressed to the CLM layer first.

CLM may internally route requests and actions to other layers for further processing.

- **Service Catalog (SC)**
- Service Catalog is central to the definition of Cloud.
- SC defines what kind of services the cloud is capable of providing and at what cost to the end-user.
- SC is the first thing that is drafted before a Cloud is architected.
- The Service Management Layer consults SC before it processes any request for a new resource.

Benefits of Cloud

- Cloud provides tangible business benefits to business.
- It saves cost to the business thus improving the bottom line.
- It also adds value to the existing business processes by incorporating new functions for increasing efficiency, flexibility, manageability and improved transparency.

Benefits……….

- □ **Provisioning** is automated and on-demand and can be done on a self-service basis.
- The provisioning typically takes from few seconds to few hours.
- Also, the demand for resources can be estimated well in advance to plan for procurement of hardware.
- □ **Utilization** typically is around 60-70%.
- The freed up resources goes into a pool which can be assigned transparently to other users.
- It is also possible to downscale and upscale based on demand.
- □ **Elasticity** allows scaling up and down on-demand.

Contd….

- □ **Scalability is on-demand**. Capacity can be planned and operational expenses can be fine-tuned to meet the current demand.
- □ **Availability:** Typically the VM instances are not tied to any particular hardware. Rather they are designed to run over a range of hardware.
- Hence, it is possible to restart the instance on secondary hardware if the primary hardware fails.
- This happens automatically and transparently.
- □ **Capital Expenditure:** Depending on the cloud deployment, a customer saves about 40% in upfront capital expenditure required to procure hardware.
- The operational expense can be further fine-tuned based on demand, thus resulting in much higher savings.

Contd……

- Chargeback allows for more granular monitoring of usage of resources in terms of cost.
- This opens up avenues for further optimization.
- **Monitoring** in cloud allows for further optimization of resources for maximum utilization and reduced wastage. This results in higher savings over time.
- Based on the cloud implementation, the savings in Cloud may vary.
- There are multiple factors that may impact the features of Cloud and the savings and benefits inherent in Cloud.

History of Cloud computing

Cloud History (Cloud Roots)

clusters, grids

Cloud roots can be tracked by observing the advancement of several technologies, especially in **hardware** (*virtualization, multi-core chips*), **Internet technologies** (*Web services, service-oriented architectures, Web 2.0*), **distributed computing** (), and **systems management** (*autonomic computing, data center automation*).

Contd….

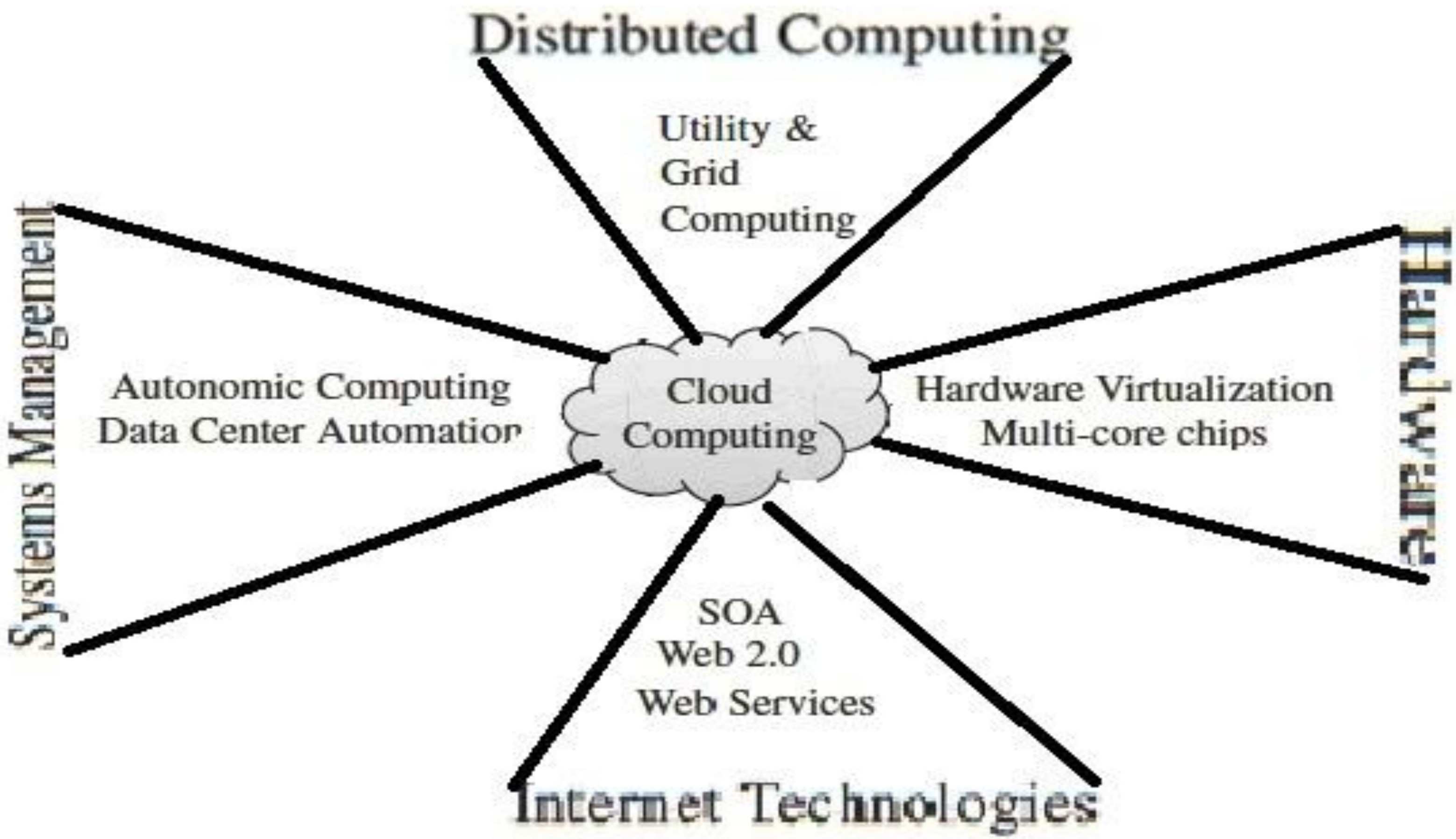

FIGURE 1

Internet Technologies (SOA, Web Services, Web 2.0, and Mashups)

Service-oriented architecture (SOA) is largely seen as the harbinger of cloud computing.

The main aim of Cloud computing is to deliver information technology as a service.

An important part of achieving this is to build applications around the concept of service.

Rapid prototyping of services on cloud require re-use of independent components that have no dependence on the underlying operating systems or the hardware.

Plus, these components must use standard interfaces that expose the protocols and functions that can be used to glue them into a cloud application.

SOA provides architecture to define, design service oriented applications that fit the needs of cloud framework.

The maturity of Web services has enabled the creation of powerful services that can be accessed on-demand, in a uniform way.

Whereas some Web services are published with the aim of serving end-user applications, their real power resides in its interface being accessible by other services.

Contd….

- **An enterprise application,** that follows the Service-oriented architecture (SOA) paradigm, is nothing but a collection of services that together perform complex business logic.
- This concept of joining services initially focused on the enterprise Web, but gained popularity in the consumer domain as well, specially with the birth of Web 2.0.
- In Web 2.0, information and services may be programmatically added together, pretending as building blocks of complex compositions, called service **mashups.**
- Many service providers, such as Amazon, Facebook, Google, make their service Application Programming Interfaces (APIs) publicly accessible using standard protocols such as SOAP and REST .
- Resulting in, one can put an idea of a fully functional Web application into practice just by attaching pieces with few lines of code.

Contd.........

- Services such as user authentication, e-mail, payroll management, and calendars are examples of building blocks that can be reused and combined in a business solution in the situation wherein a single, ready-made system does not provide all those features.
- Many building blocks and solutions are now available in public marketplaces.

- For example, Programmable Web is a public repository of service Application Programming Interfaces(APIs) and mashups presently more than thousands of APIs and mashups are available.
- Popular APIs such as YouTube, Amazon eCommerce, Google Maps, Flickr and Twitter, when clubbed together, produce a variety of interesting solutions, from searching video game retailers to weather maps.
- Salesforce.com also offers AppExchange, which enables the sharing of solutions developed by third-party developers on top of Salesforce.com

Distributed computing (Utility and Grid Computing)

- Grid computing provides the combination of distributed resources and transparently access to distributed resources.
- Production grids namely Enabling Grids for E-Science in Europe (EGEE) and TeraGrid, attempt to achieve the sharing of compute and storage resources distributed across different administrative domains, with their main aim in speeding up a wide range of scientific applications, such as climate modeling, drug design and protein analysis.
- A main aspect of the grid vision realization has been in building standard Web services-based protocols that allow distributed resources to be "discovered, accessed, allocated, monitored, accounted for and billed for, etc., and in general managed as a single virtual system."
- The Open Grid Services Architecture (OGSA) addresses this need for standardization by defining a set of main capabilities and behaviors that address main concerns in grid systems.
- The major drawback, which is disappointing, when using grids is the availability of resources with diverse software configurations, including disparate operating systems (OSs), libraries, compilers, runtime environments, and so on.
- At the same time, user applications would often run only on specially customized environments. Thus, a portability barrier has often been present on most grid infrastructures, preventing users to adopt grids as utility computing environments.

Contd…..

- Virtualization technology (VT) has been identified as the perfect fit to issues that have drawbacks when using grids, such as hosting many different software applications on a single physical platform.
- In this way, some research projects targeted at evolving grids to support an additional layer to virtualize computation, storage, and network resources.
- With increasing popularity and usage, large grid installations have encountered new problems, such as excessive spikes in demand for resources along with strategic and negative behavior by users.
- Initially, grid resource management techniques did not ensure fair and equitable access to resources in many systems.
- Traditional metrics (throughput, waiting time, and slowdown) failed to capture the more skillful requirements of users.
- There were no real advantages for users to be flexible about resource requirements or job deadlines, nor provisions to accommodate users with urgent work.

Contd…..

- Basically in utility computing environments, users assign a "utility" value to their jobs, where utility is a fixed or time-varying valuation that captures various QoS constraints (deadline, importance, satisfaction).
- The valuation is the amount, users are willing to pay a service provider to satisfy their demands.
- The service providers than attempts to maximize their own utility, where mentioned utility may directly complement with their profit.
- Service providers can opt to prioritize high yielding user jobs, leading to a situation where shared systems are viewed as a marketplace, where users compete for resources based on the perceived utility or value of there jobs.
- Further to this information and comparison of these utility computing environments are available in an extensive survey of these platforms.