RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Artificial Intelligence and Machine Learning/ Artificial Intelligence and Machine Learning, VI-Semester

Open Elective AL604 (A) Cloud Computing

Course Objective: The objective of this course is to provide students with the comprehensive and in depth knowledge of Cloud Computing concepts, technologies, architecture and applications.

UNIT I Introduction of Grid and Cloud computing, characteristics, components, business and IT perspective, cloud services requirements, cloud models, Security in public model, public verses private clouds, Cloud computing platforms: Amazon EC2,Platform as Service: Google App Engine, Microsoft Azure, Utility Computing, Elastic Computing.

UNIT II Cloud services- SAAS, PAAS, IAAS, cloud design and implementation using SOA, conceptual cloud model, cloud stack, computing on demand, Information life cycle management, cloud analytics, information security, virtual desktop infrastructure, storage cloud.

UNIT III Virtualization technology: Definition, benefits, sensor virtualization, HVM, study of hypervisor, logical partitioning- LPAR, Storage virtualization, SAN, NAS, cloud server virtualization, virtualized data center.

UNIT IV Cloud security fundamentals, Vulnerability assessment tool for cloud, Privacy and Security in cloud, Cloud computing security architecture: Architectural Considerations- General Issues, Trusted Cloud computing, Secure Execution Environments and Communications, Micro-architectures; Identity Management and Access control-Identity management, Access control, Autonomic Security, Cloud computing security challenges: Virtualization security management-virtual threats, VM Security Recommendations, VM-Specific Security techniques, Secure Execution Environments and Communications in cloud.

UNIT V SOA and cloud, SOA and IAAS, cloud infrastructure benchmarks, OLAP, business intelligence, e-Business, ISV, Cloud performance monitoring commands, issues in cloud computing. QOS issues in cloud, mobile cloud computing, Inter cloud issues, Sky computing, Cloud Computing Platform, Xen Cloud Platform, Eucalyptus, OpenNebula, Nimbus, TPlatform, Apache Virtual Computing Lab (VCL), Anomaly Elastic Computing Platform.

References:

1. Dr.Kumar Saurabh, “Cloud Computing”, Wiley India.
2. Ronald Krutz and Russell Dean Vines, “Cloud Security”, Wiley-India.
3. Judith Hurwitz, R.Bloor, M.Kanfman, F.Halper, “Computing for Dummies”, Wiley India Edition.
4. Anthony T.Velte Toby J.Velte, “Cloud Computing – A Practical Approach”, TMH.
5. Barrie Sosinsky, ‘Cloud Computing Bible”, Wiley India.

Course Outcomes: After the completion of this course, the students will be able to:

1. Explain the core concepts of the cloud computing paradigm
2. Demonstrate knowledge of virtualization
3. Explain the core issues of cloud computing such as security, privacy, and interoperability.
4. Choose the appropriate technologies, algorithms, and approaches for the related issues.
5. Identify problems, and explain, analyze, and evaluate various cloud computing solutions

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Artificial Intelligence and Machine Learning/ Artificial Intelligence and Machine Learning, VI-Semester

Open Elective AL604 (B) Information Security & Management

UNIT-I

Introduction: Needs for Security; Basic security terminologies e.g. threats, vulnerability, exploit etc.; Security principles(CIA), authentication, nonrepudiation; security attacks and their classifications; Mathematical foundation - Prime Number; Modular Arithmetic; Fermat's and Euler's Theorem; The Euclidean Algorithms; The Chinese Remainder Theorem; Discrete logarithms**.**

UNIT-II

Symmetric Key Cryptography: Classical cryptography – substitution, transposition and their cryptanalysis; Symmetric Cryptography Algorithm – DES, 3DES, AES etc.; Modes of operation: ECB, CBC etc.; Cryptanalysis of Symmetric Key Ciphers: Linear Cryptanalysis, Differential Cryptanalysis.

UNIT-III

Asymmetric Key Cryptography: Key Distribution and Management, Diffie-Hellman Key Exchange algorithm; Asymmetric Key Cryptography Algorithm– RSA, ECC etc.; Various types of attacks on Cryptosystems.

UNIT-IV

Authentication & Integrity – MAC, Hash function, SHA, MD5, HMAC, Digital signature and authentication protocols; Authorization; Access control mechanism; X.509 Digital Certificate**.**

UNIT-V

E-mail, IP and Web Security: E-mail security – PGP, MIME, S/MIME; IP security protocols; Web security – TLS, SSL etc.; Secure Electronic Transaction(SET); Firewall and its types; Introduction to IDPS; Risk Management; Security Planning.

TEXT BOOKS RECOMMENDED:

1. Michael E. Whitman, Herbert J. Mattord, "Principles of Information Security", 6th Edition, Cengage Learning.

2. Stallings William, "Cryptography and Network Security - Principles and Practice", 7th Edition, Pearson.

REFERENCE BOOKS:

1. Roberta Bragge, Mark Rhodes, Keith Straggberg, "Network Security the Complete Reference", Tata McGraw Hill Publication,

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Artificial Intelligence and Machine Learning/ Artificial Intelligence and Machine Learning, VI-Semester

Open Elective AL604 (C) Intelligent Systems for Robotics

Course Objective:
The students will be able to understand the basic concepts and fundamentals of robotics. They will also be able to use AI in the field of robotics.
Detailed Contents:

Unit 1:
Introduction: Introduction to Robotics Fundamentals of Robotics, Robot Kinematics: Position Analysis, Dynamic Analysis and Forces, Robot Programming languages & systems: Introduction, the three levels of robot programming, requirements of a robot programming language, problems peculiar to robot programming languages.

Unit 2:
Need of AI in Robotics: History, state of the art, Need for AI in Robotics. Thinking and acting humanly, intelligent agents, structure of agents.

Unit 3:
Game Playing: AI and game playing, plausible move generator, static evaluation move generator, game playing strategies, problems in game playing.

Unit 4:
Robotics fundamentals: Robot Classification, Robot Specification, notation, kinematic representations and transformations, dynamics techniques; trajectory planning and control**.**

Unit 5:
Robotics and Its applications: DDD concept, Intelligent robots, Robot anatomy-Definition, law of robotics, History and Terminology of Robotics-Accuracy and repeatability of Robotics-Simple problems-Specifications of Robot-Speed of Robot, Robot joints and links-Robot classifications-Architecture of robotic systems-Robot Drive systems-Hydraulic, Pneumatic and Electric system

Suggested References:

1. Robotics, Vision and Control: Fundamental Algorithms in MATLAB, Peter Corke, Springer, 2011.
2. Robotics: Everything You Need to Know About Robotics from Beginner to Expert, Peter McKinnon, Createspace Independent Publishing Platform, 2016.
3. Introduction to AI Robotics, Second Edition, By Robin R. Murphy, MIT press, 2001.
4. Artificial Intelligence for Robotics: Build intelligent robots that perform human tasks using AI techniques, Francis X. Govers, Packt Publishers, 2018

H

H

Subject Notes
CS 8002 - Cloud Computing

Unit-1

Historical Development

"Cloud computing" concepts date back to the 1950s when large-scale mainframes were made available to schools and corporations. The mainframe's colossal hardware infrastructure was installed in what could literally be called a "server room" (since the room would generally only be able to hold a single mainframe), and multiple users were able to access the mainframe via "dumb terminals" – stations whose sole function was to facilitate access to the mainframes. Due to the cost of buying and maintaining mainframes, an organization wouldn't be able to afford a mainframe for each user, so it became practice to allow multiple users to share access to the same data storage layer and CPU power from any station. By enabling shared mainframe access, an organization would get a better return on its investment in this sophisticated piece of technology figure 1.1.

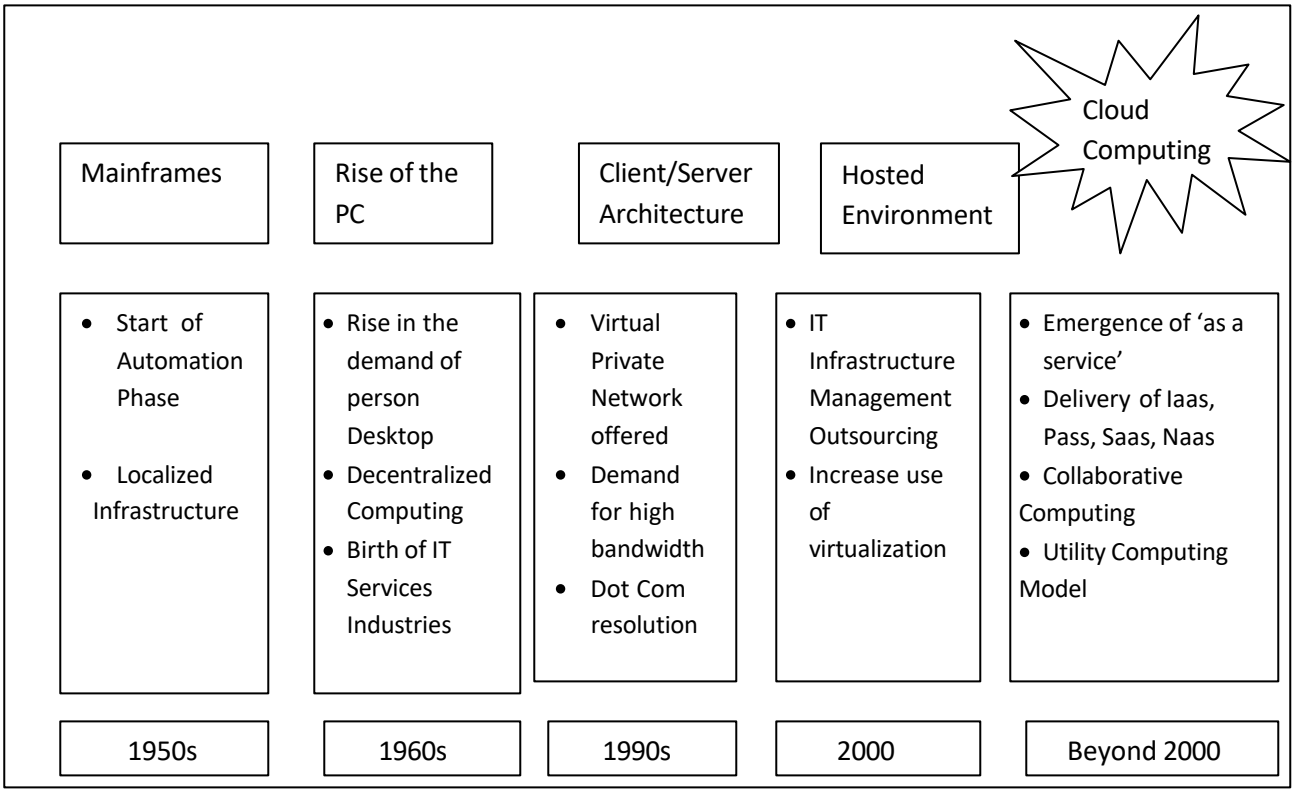

Figure: 1.1 History of Cloud Computing

A couple decades later in the 1970s, IBM released an operating system called VM that allowed admins on their System/370 mainframe systems to have multiple virtual systems, or "Virtual Machines" (VMs) on a single physical node. The VM operating system took the 1950s application of shared access of a mainframe to the next level by allowing multiple distinct compute environments to live in the same physical environment. Most of the basic functions of any virtualization software that you see nowadays can be traced back to this early VM OS: Every VM could run custom operating

systems or guest operating systems that had their "own" memory, CPU, and hard drives along with CD-ROMs, keyboards and networking, despite the fact that all of those resources would be shared. "Virtualization" became a technology driver, and it became a huge catalyst for some of the biggest evolutions in communications and computing.

In the 1990s, telecommunications companies that had historically only offered single dedicated point–to-point data connections started offering virtualized private network connections with the same service quality as their dedicated services at a reduced cost. Rather than building out physical infrastructure to allow for more users to have their own connections, telco companies were able to provide users with shared access to the same physical infrastructure. This change allowed the telcos to shift traffic as necessary to allow for better network balance and more control over bandwidth usage. Meanwhile, virtualization for PC-based systems started in earnest, and as the Internet became more accessible, the next logical step was to take virtualization online figure 1.2.

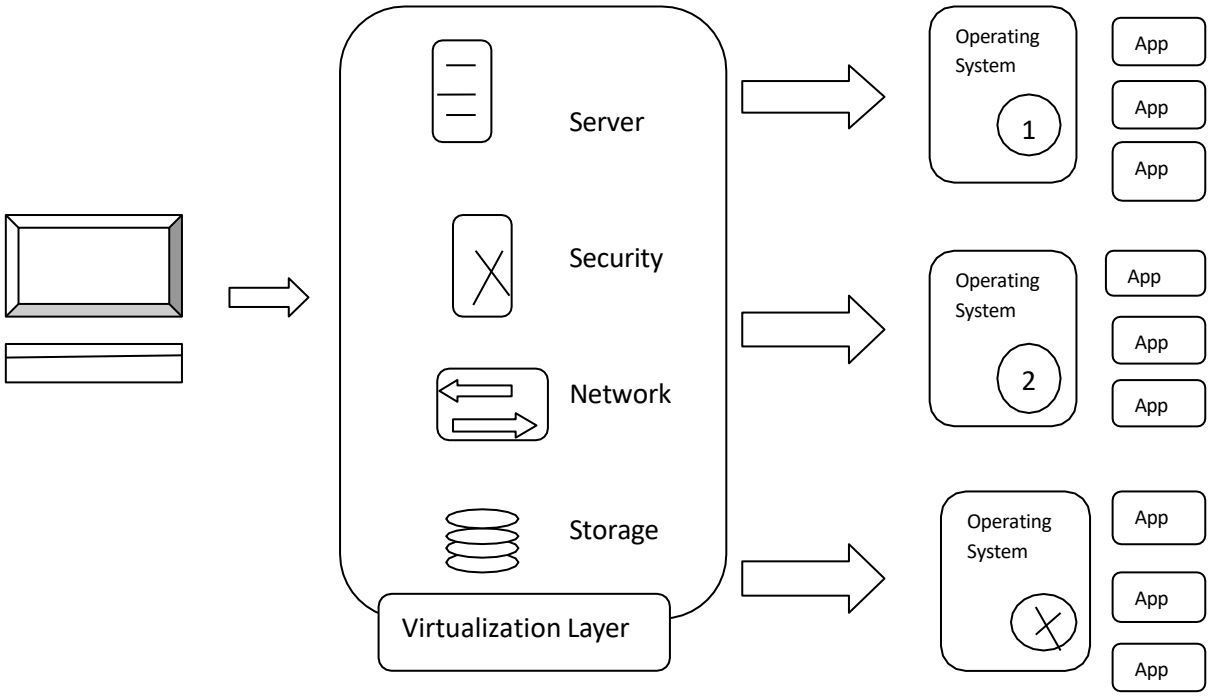

Figure: 1.2 Model Cloud Configuration

Vision of Cloud Computing

A cloud is simply a centralized technology platform which provides specific IT services to a selected range of users, offering the ability to login from anywhere, ideally from any device and over any connection, including the Internet.

Intercept IT believes that a true cloud computing service is one which removes the traditional barriers which exist between software applications, data and devices. In other words, it is the nirvana of computing from a user's perspective. no need to worry about location, device, or type of connection, all the data and the software applications required by the user are fully available and the experience remains consistent. The highest standards of data protection must be a given, whereby users do not have to think about protecting the integrity of the data they use and store.

Intercept IT provides a broad spectrum of both application delivery services to its clients, ranging from the design, implementation and management of private clouds, right through to the provision of hosted cloud solutions delivered via Intercept's own, cloud infrastructure.

Cloud Computing Principles

Here the top three key principles of cloud computing that struck me as making this topic very relevant, interesting and definitely the way of the future:

1: Abstraction

Cloud gives you just a few basic, but well- defined services and that's it. Take it or leave it. "Do you like our simple, RESTful foo interface? Fine, use it!", or: "Oh, you want your own special custom version? Sorry, we don't have it. Go away." It's that simple.

So the key point here is that well-defined abstraction layers between clouds and developers/users are the grease that lets both sides operate efficiently and completely independent of each other.

There are three layers of abstraction in clouds:

• Application as a Service (AaaS): This is what the end-user gets when they use a service like GMail, DropBox (please make an OpenSolaris version, thanks), the myriads of Facebook apps, SmugMug or even Adobe's online photoshop web service. AaaS services are very popular and there's really no reason to start a new application any other way today.

• Platform as a Service (PaaS): The abstraction layer here is some kind of developer environment, but the details of implementation (OS, Hardware, etc.) are completely hidden. You just get a programming language and some APIs/Libraries and off you go. This is what Zembly gives you (check it out and create your own Facebook app in minutes), or the Google App Engine. This is the development model of the future: Develop against the cloud, no need to know the details behind it.

• Infrastructure as a Service (IaaS): These are the Amazon S3s, EC2s, etc. and we recently introduced our own version of IaaS as the Sun Cloud (featuring open interfaces and a lot of Sun technology goodness under the hood.) In this model, you get access to a virtual server or virtual storage, treat them like real machines, but the physical details of what machine is in what rack or which disks you use are hidden from you.

2: Automation

Automation in the cloud means the developer/user is in complete, automatic control over their resources. No human interaction whatsoever, even from a developer/user perspective. Need more servers? Let the load-balancer tell the cloud how many more to provide. No need to wait for someone to unpack and cable your machine, no need to wait for your IT department to find the time to install. Everything is automatic.

Again, this is a win-win for both sides. While full automation reduces cost and complexity for the cloud provider, it puts the developer/user in control. Now you can reduce your time to market for your next rollout because you can do it yourself, fully automatic, and you don't need to call anybody, rely on someone else to set up stuff for you, or wait days until some minor hardware/software installation is completed.

3: Elasticity

In cloud computing, elasticity is defined as "the degree to which a system is able to adapt to workload changes by provisioning and de-provisioning resources in an autonomic manner, such that at each

point in time the available resources match the current demand as closely as possible". Elasticity is a defining characteristic that differentiates cloud computing from previously proposed computing paradigms, such as grid computing. The dynamic adaptation of capacity, e.g., by altering the use of computing resources, to meet a varying workload is called "elastic computing".

Example: Sun Cloud expands the business possibilities of the cloud model: You can choose to be the cloud (and we'll help you build it), you can choose to build the cloud (for others, out of our cloud components), you can build your own cloud (we'll help you build that, too) or you can just use it (the Sun Cloud). Just like we believe in open standards, we also believe in partnering, so no matter what your cloud business model is, Sun can help.

Characteristics of Cloud Computing as per NIST

Cloud technology is in the news quit often these days, but it still seems to be mysterious and confusing to the non-techie crowd. Cloud options are enticing various industries across the board, which is why it's important to know its essential characteristics as a software offering. Here are the five main characteristics that cloud computing offers businesses today.

1. On-demand capabilities:

A business will secure cloud-hosting services through a cloud host provider which could be your usual software vendor. You have access to your services and you have the power to change cloud services through an online control panel or directly with the provider. You can add or delete users and change storage networks and software as needed. Typically, you are billed with a monthly subscription or a pay-for-what-you-use scenario. Terms of subscriptions and payments will vary with each software provider.

2. Broad network access:

Your team can access business management solutions using their smart phones, tablets, laptops, and office computers. They can use these devices wherever they are located with a simple online access point. This mobility is particularly attractive for businesses so that during business hours or on off- times, employees can stay on top of projects, contracts, and customers whether they are on the road or in the office. Broad network access includes private clouds that operate within a company's firewall, public clouds, or a hybrid deployment.

3. Resource pooling:

The cloud enables your employees to enter and use data within the business management software hosted in the cloud at the same time, from any location, and at any time. This is an attractive feature for multiple business offices and field service or sales teams that are usually outside the office.

4. Rapid elasticity:

If anything, the cloud is flexible and scalable to suit your immediate business needs. You can quickly and easily add or remove users, software features, and other resources.

5. Measured service:

Going back to the affordable nature of the cloud, you only pay for what you use. You and your cloud provider can measure storage levels, processing, bandwidth, and the number of user accounts and you are billed appropriately. The amount of resources that you may use can be monitored and controlled from both your side and your cloud provider's side which provides transparency.

Cloud Computing Reference Model

•The NIST Cloud Computing Reference Architecture consists of five major actors. Each actor plays a role and performs a set of activities and functions. The reference architecture is presented as successive diagrams in increasing level of detail figure 1.3.

•Among the five actors, cloud brokers are optional, as cloud consumers may obtain service directly from a cloud provider.

1. Cloud Consumer:

Person or organization that maintains a business relationship with, and uses service from, Cloud Providers.

2. Cloud Provider:

Person, organization or entity responsible for making a service available to Cloud Consumers.

3. Cloud Auditor:

A party that can conduct independent assessment of cloud services, information system operations, performance and security of the cloud implementation.

4. Cloud Broker:

An entity manages the use, performance and delivery of cloud services, and negotiates relationships between Cloud Providers and Cloud Consumers.

5. Cloud Carrier:
The intermediary that provides connectivity and transport of cloud services from Cloud Providers to Cloud Consumers.

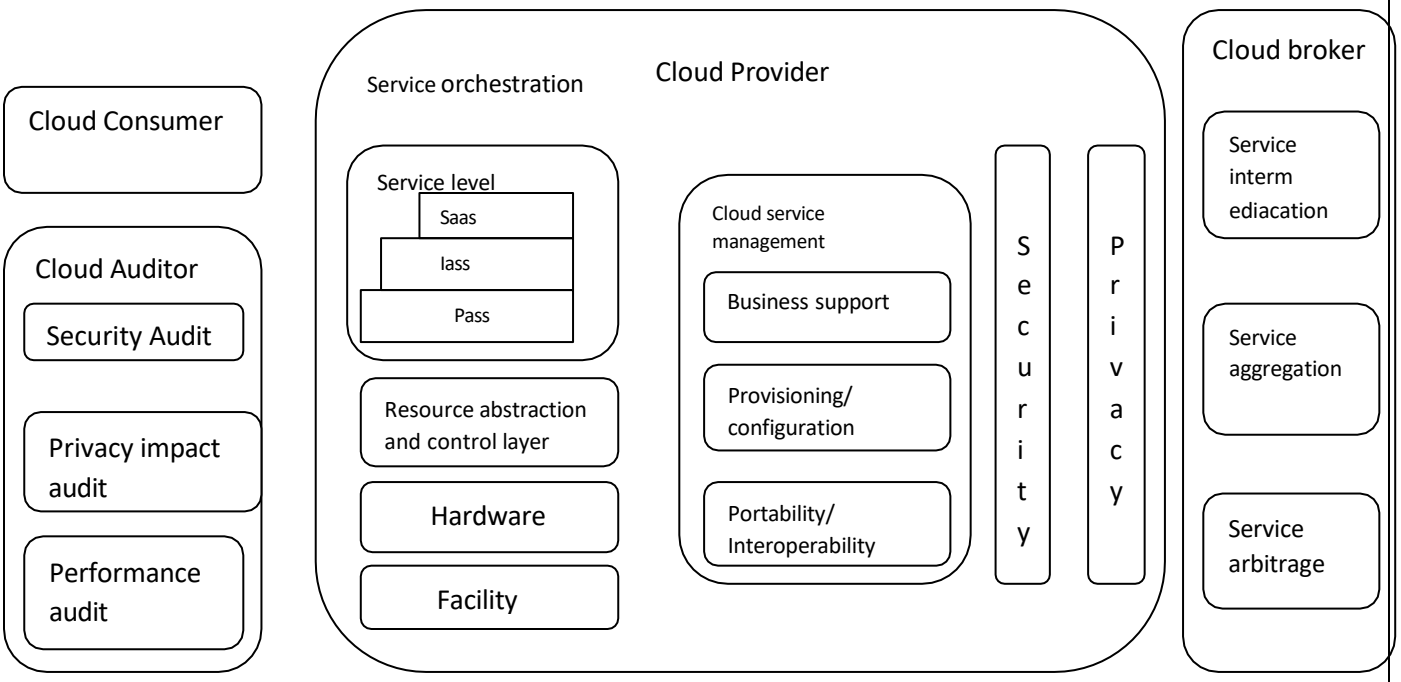

Figure: 1.3 NIST Cloud Reference Model

Cloud Computing Environments

More and more organizations are moving services, storage, email, collaboration and applications to the cloud. You need to decide whether to choose to support private, public or a hybrid cloud mix. What's the right mix of infrastructure (IaaS), platform (PaaS), and application (SaaS) environments for your organization? Where are the cost savings?

• **Private Cloud:** Which services require the most agility and speed? What's the right balance of standard service offerings that will drive the most business value? Do you need to build an internal shared-service center? How does a private cloud implementation impact your data center architecture?

• **Public Cloud:** Which applications are most likely to move to public cloud delivery models? Will your organization bypass your IT department and get its applications from the cloud via software-as-service (SaaS) for a monthly pay-per-user-per-month subscription pricing model?

• **Hybrid Cloud:** Is hybrid cloud really the future? What level of flexibility do you need to customize, manage and monitor your applications? How will the cloud services brokerage role define future IT organizations?

Cloud Services Requirements

Today the cloud services have several deficiencies - which from an enterprise perspective are the basic requirements for them to consider cloud services.

1. Availability - with loss less DR

As a cloud service provider, there will be enormous pressure to minimize costs by optimally utilizing the entire IT infrastructure. The traditional Active-Passive DR strategy is very expensive and cost inefficient. Instead, service providers will have to create an Active-Active disaster recovery mechanism - where more than one data center will be active at all times and ensures that the data and services can be accessed by the customer from either of the data centers seamlessly.

2. Portability of Data & Applications

As applications are being written on standard platforms - Java, PHP, Python, etc. It should be possible to move the customer owned applications from one service provider to another. Customers should also take care to use only the open standards and tools, and avoid vendor specific tools. Azure or Google services offers several tools/applications/utilities which are valuable - but it also creates a customer locking - as the customer who uses these vendors specific tools cannot migrate to another service provider without rewriting the applications.

3. Data Security

Security is the key concern for all customers - since the applications and the data is residing in the public cloud; it is the responsibility of the service provider for providing adequate security. In my opinion security for customer data/applications becomes a key differentiator when it comes to selecting the cloud service provider. When it comes to IT security, customers tend to view the cloud

service providers like they view banks. The service provider is totally responsible for user security, but there are certain responsibilities that the customer also needs to take.

4. Manageability

Managing the cloud infrastructure from the customer perspective must be under the control of the customer admin. Customers of Cloud services must be able to create new accounts, must be able to provision various services, do all the user account monitoring - monitoring for end user usage, SLA breaches, data usage monitoring etc. The end users would like to see the availability, performance and configuration/provisioning data for the set of infrastructure they are using in the cloud.

5. Elasticity

Customer on Cloud computing have a dynamic computing load. At times of high load, they need greater amount of computing resources available to them on demand, and when the work loads are low, the computing resources are released back to the cloud pool. Customer expects the service provider to charge them for what they have actually used in the process.

6. Federated System

Customers may have to buy services from several cloud service providers for various services - email from Google, online sales transaction services from Amazon and ERP from another vendor etc. In such cases customer want their cloud applications to interact with other services from several vendors to provide a seamless end to end IT services. In a federated environment there is potentially an infinite pool of resources. To build such a system, there should be inter-cloud framework agreements between multiple service providers, and adequate chargeback systems in place.

Cloud and dynamic infrastructure

For the architect employed with building out a cloud infrastructure, there are seven key requirements that need to be addressed when building their cloud strategy. These requirements include:

1. Heterogeneous Systems Support

Not only should cloud management solutions leverage the latest hardware, virtualization and software solutions, but they should also support a data center's existing infrastructure. While many of the early movers based their solutions on commodity and open source solutions like general x86 systems running open source Xen and distributions like CentOS, larger service providers and enterprises have requirements around both commodity and proprietary systems when building out their clouds. Additionally, cloud management providers must integrate with traditional IT systems in order to truly meet the requirements of the data center. Companies that don't support technologies from the likes of Cisco, Red Hat, NetApp, EMC, VMware and Microsoft will fall short in delivering a true cloud product that fits the needs of the data center.

2. Service Management

To productize the functionality of cloud computing, it is important that administrators have a simple tool for defining and metering service offerings. A service offering is a quantified set of services and applications that end users can consume through the provider — whether the cloud is private or

public. Service offerings should include resource guarantees, metering rules, resource management and billing cycles. The service management functionality should tie into the broader offering repository such that defined services can be quickly and easily deployed and managed by the end user.

3. Dynamic Workload and Resource Management

In order for a cloud to be truly on-demand and elastic while consistently able to meet consumer service level agreements (SLAs), the cloud must be workload- and resource- aware. Cloud computing raises the level of abstraction to make all components of the data center virtualized, not just compute and memory. Once abstracted and deployed, it is critical that management solutions have the ability to create policies around workload and data management to ensure that maximum efficiency and performance is delivered to the system running in the cloud. This becomes even more critical as systems hit peak demand. The system must be able to dynamically prioritize systems and resources on-the-fly based on business priorities of the various workloads to ensure that SLAs are met.

4. Reliability, Availability and Security

While the model and infrastructure for how IT services are delivered and consumed may have changed with cloud computing, it is still critical for these new solutions to support the same elements that have always been important for end users. Whether the cloud serves as a test bed for developers prototyping new services and applications or it is running the latest version of a popular social gaming application, users expect it to be functioning every minute of every day. To be fully reliable and available, the cloud needs to be able to continue to operate while data remains intact in the virtual data center regardless if a failure occurs in one or more components. Additionally, since most cloud architectures deal with shared resource pools across multiple groups both internal and external, security and multi-tenancy must be integrated into every aspect of an operational architecture and process. Services need to be able to provide access to only authorized users and in this shared resource pool model the users need to be able to trust that their data and applications are secure.

5. Integration with Data Center Management Tools

Many components of traditional data center management sill require some level of integration with new cloud management solutions even though the cloud is a new way of consuming IT. Within most data centers, a variety of tools are used for provisioning, customer care, billing, systems management, directory, security and much more. Cloud computing management solutions do not replace these tools and it is important that there are open application programming interfaces (APIs) that integrate into existing operation, administration, maintenance and provisioning systems (OAM&P) out of the box. These include both current virtualization tools from VMware and Citrix, but also the larger data center management tools from companies like IBM and HP.

6. Visibility and Reporting

The need to manage cloud services from a performance, service level, and reporting perspective becomes paramount to the success of the deployment of the service. Without strong visibility and reporting mechanisms the management of customer service levels, system performance, compliance and billing becomes increasingly difficult. Data center operations have the requirement of having real-time visibility and reporting capabilities within the cloud environment to ensure compliance, security, billing and charge backs as well as other instruments, which require high levels of granular visibility and reporting.

7. Administrator, Developer and End User Interfaces

One of the primary attributes and successes of existing cloud- based services on the market comes from the fact that self- service portals and deployment models shield the complexity of the cloud service from the end user. This helps by driving adoption and by decreasing operating costs as the majority of the management is offloaded to the end user. Within the self-service portal, the consumer of the service should be able to manage their own virtual data center, create and launch templates, manage their virtual storage, compute and network resources and access image libraries to get their services up and running quickly. Similarly, administrator interfaces must provide a single pane view into all of the physical resources, virtual machine instances, templates, service offerings, and multiple cloud users. On top of core interfaces, all of these features need to be interchangeable to developers and third parties through common APIs.

Cloud Adoption and rudiments

Cloud Adoption is a strategic move by organizations of reducing cost, mitigating risk and achieving scalability of data base capabilities. Cloud adoption may be up to various degrees in an organization, depending on the depth of adoption. In fact the depth of adoption yields insight into the maturity of best practices, enterprise-ready cloud services availability.

Organizations that go ahead with the strategic decision of adopting cloud-based technologies have to identify potential security thefts and controls, required to keep the data and applications in the cloud secured. Hence there is a need for compliance assessment during cloud adoption. The following measures are taken for compliance assessment to ensure security and accountability of data and applications in the cloud services:

- Matching the security requirements of the organization with the security capabilities of the cloud service provider
- Analyzing the security policies of the cloud service provider along with history of transparency and security related practices
- Proper understanding of the technical aspects of data and traffic flow
- Proper understanding and documentation of the roles and responsibilities of the cloud service provider
- Understanding of the certifications and compliances that can be leveraged from the cloud service provider

Key elements in adopting cloud:

Investment
One critical element of your business that requires a significant investment is your infrastructure. Looking to a cloud-based solution can allow you to focus your important resources on the business while leaving your infrastructure in the hands of an experienced team.

Scalability
As your business grows, the last thing you want to have to worry about is the ability of your infrastructure to scale. The key to finding and partnering with a cloud-based service is picking an organization that is devoted to improving their own capacity. Doing so will positively affect your ability to grow with limited interruption.

Predictable Expenses

Looking to a cloud-based service provider gives you the ability to predict and plan for a consistent expense. Even as you scale and grow, real-time insight allows you to precisely allocate funds towards growth initiatives.

Overview of Cloud Applications

1. ECG Analysis in Cloud

Our objective is to propose an architecturally generic Cloud-based system to accommodate multiple scenarios where patients need to be remotely monitored and recorded data must be analyzed by a computing system and become available to be visualized by specialists or by the patients themselves. Although our design and prototype are generic to accommodate several use cases, in this paper, we focus on one motivational case, namely: the monitoring of patients who suffer from cardiac arrhythmias, requiring continuous episode detection. Electrocardiogram (ECG) data from commodity wearable sensors are obtained in real-time and used to perform episode detection and classification figure 1.4.

The overall functionality of an ECG monitoring and analysis system involves the following steps:

1. A patient is equipped with a wireless ECG sensor attached to their body and a mobile device that is capable of communicating to the Internet;

2. The wireless ECG sensor module collects patient's data and forwards it the mobile device via Bluetooth without user intervention;

3. A client software in the mobile device transmits the data to the ECG analysis Web Service, which is hosted by a Cloud computing-based software stack. This communication can happen with a home wireless gate- way or directly via the mobile's data connectivity (e.g. mobile 3G network);

4. The analysis software carries out numerous computations over the received data taking the reference from the existing demographic data, and the patient's historic data. Computations concern comparison, classification, and systematic diagnosis of heartbeats, which can be time-consuming when done for long time periods for large number of users;

5. The software then appends the latest results to the patient's historic record maintained in private and secure Cloud-based storage, so that authenticated users can access it anytime from anywhere. Physicians will later interpret the features extracted from the ECG wave- form and decide whether the heartbeat belongs to the normal (healthy) sinus rhythm or to an appropriate class of arrhythmia;

6. The diagnosis results are disseminated to the patient's mobile device and/or monitor, their doctor and/or emergency services at predefined intervals;

7. The monitoring and computing processes are repeated according to user's preference, which may be hourly or daily over a long period of time.

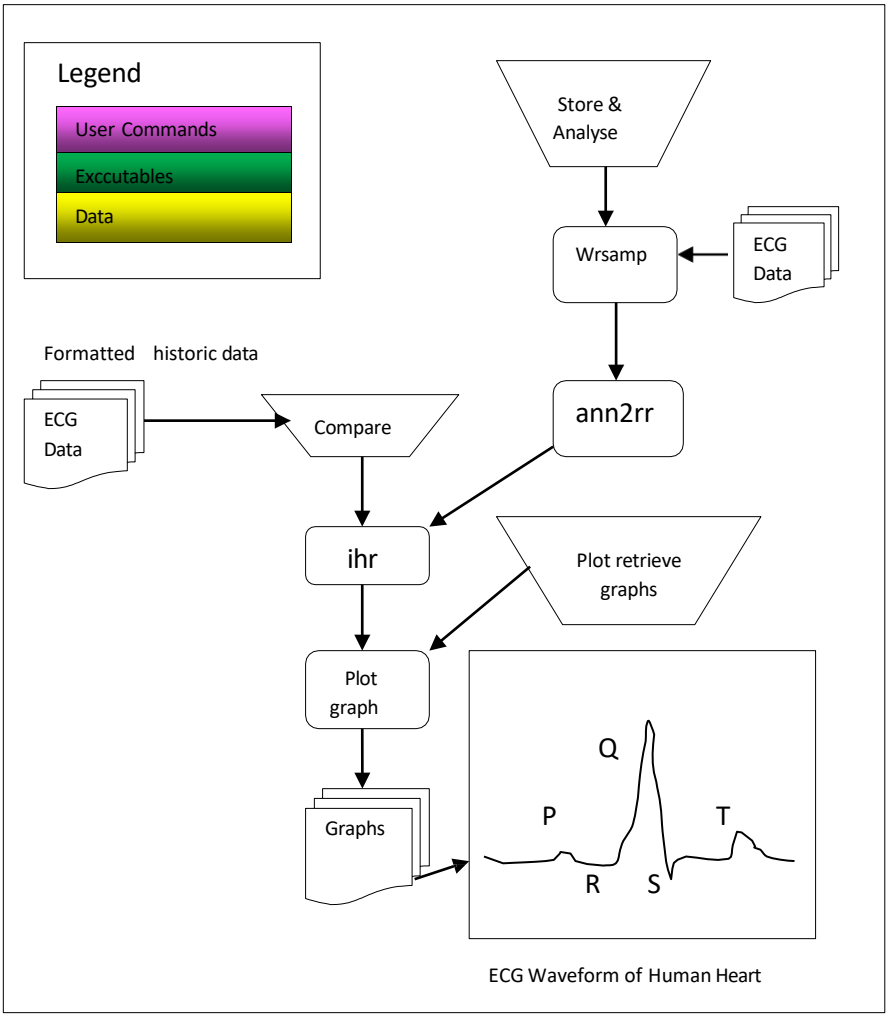

Figure: 1.4 Steps of ECG Cloud Analysis

Task Name	Description
wrsamp	Reads the raw data to produce a binary file with specified sample frequency, gain, format, etc
ann2rr	Creates an annotation file from ECG data; creates RR interval series from ECG annotation files
ihr	Reads an annotation file and produces an instantaneous heart rate signal
Plot Graph	Plots the graphs of the heart rate signal, RR interval, etc.

Table 1.1: ECG Component Analysis

2. Protein structure prediction

Protein structure prediction is one of the most important and yet difficult processes for modern computational biology and structural bioinformatics. The practical role of protein structure prediction becomes even more important in the face of dynamically growing number of protein sequences obtained through the translation of DNA sequences coming from large-scale sequencing projects. Protein structure prediction refers to the computational procedure that delivers a three-dimensional structure of a protein based on its amino acid sequence. Modeling complex protein structures is still

challenging. We have designed and developed scalable system, called Cloud4PSP, which enables predictions of 3D protein structures with the use of Warecki-Znamirowski method in Microsoft Azure public cloud. Scalability of the system is provided by Microsoft Azure cloud platform. The platform allows creating and publishing user's applications that can be scaled up (vertical scalability) or scaled out (horizontal scalability).

Case Study 1: Protein Disorder in Completely Sequenced Organisms

The goal of this study is to collect evidence for three hypotheses on protein disorder: (1) it is more useful to picture disorder as a distinct phenomenon than as an extreme example of protein flexibility; (2) there are many very different flavors of protein disorder, but it is advantageous to recognize just two main types, namely, well structured and disordered; (3) nature uses protein disorder as a tool to adapt to different environments. We predicted protein disorder both on an in-house compute grid and on a compute grid manually setup in the OpenNebula cloud service provided by the CSC Finland.

Data and tool (the PPMI) images for grid nodes in the cloud were downloaded from http://predictprotein.org/. The PPMI image was extended with a grid client, and a separate machine instance was used as grid master. PredictProtein for the local grid was installed from the main Debian repository. Required databases (28 GB) were included on a data disk image for cloud machine instances. Input to PredictProtein jobs consisted of protein sequences (in total less than 1 GB). Grid job submissions to the local and the cloud grid were manually adjusted according to available resources. Over 9 million disorder predictions were made over the course of the past few years.

Case Study 2: Comprehensive In Silico Mutagenesis of Human Proteome

This project aims at providing information about the functional effect of every possible point mutation in all human proteins, that is, for the replacement of 19*N amino acids for a protein with N residues. Overall, this generated 300 million human sequence variants (point mutants). The method SNAP predicted the effect of each variant, that is, each “nonsynonymous single nucleotide polymorphisms” (nsSNPs) upon protein function. These predictions are useful for several reasons.

First, the study of all possible mutations in human will provide the background against which we can assess the effect of mutations that are actually observed between people. This is crucial for both the advance toward personalized medicine and health and the understanding of human diversity and variation. Second, our computation provides quick “look-up” answers available for all the important variants that are observed and implied in important phenotypes. The only way to cover those lookups is by precomputing all the possible changes. SNAP can take advantage of PredictProtein results for faster processing. With the PredictProtein packages presented here, a solution was built in the form of a public Amazon Machine Image (AMI, ami-3f5f8156) that allows running PredictProtein on the Amazon Elastic Compute Cloud (EC2). We extended an Ubuntu-based StarCluster AMI with PredictProtein and its required databases (28 GB). Because every protein can be computed independently, we formed a grid job out of each protein and used the Grid Engine (GE) to distribute work on the machine instances.

We used StarCluster to automate grid setup on the EC2. Because a lot of CPU power was needed, the “Cluster Compute Eight Extra Large Instance” was chosen. This instance type is especially crafted for big data with a lot of CPU power. One instance has 60.5 GB memory, 88 EC2 Compute Units (2x Intel Xeon E5-2670, eight-core-architecture “Sandy Bridge”), and 3370 GB instance storage. The sequence variants were analyzed based on the human reference proteome from the National Center for Biotechnology Information (build 37.3, proteins, 21MB). We processed 29,036 sequences with 16,618,608 residues. This amounted to predicting the functional effect of 315,753,552 individual amino

acid changes.

3. Gene Expression Data Analysis

Gene-expression profiling using DNA microarrays can analyze multiple gene markers simultaneously. Consequently, it is widely used for cancer prediction. Informative genes can be extracted for predicting cancer/non-cancer class or type of diagnosis. The former is more interesting for biologists due to the fact that distinguishing sub category of a cancer is a difficult task. Moreover, the accuracy of diagnosis at early stages is vital, while most cancer treatments like chemotherapy kill both cancer and non cancer cells, and seriously weaken the human defense system. And most of these drugs have both long-term and short-term side effects.

Classification methods either result in the identification of simple rules for class discovery or the identification of highly related genes to a specific cancer. Recently, there have been a number of investigations for class discovery of gene expression data sets using machine learning techniques: Decision Tree [3, 11], Support Vector Machines (SVM) [4, 14] and k-Nearest Neighbor (k-NN) [2]. How- ever, gene expression data sets have a unique characteristic: they have high-dimensional features with few samples (also known as"the curse of dimensionality"). Typically, ma- chine learning methods cannot avoid the over-fitting problem in this situation. Additionally, when the search space is vast, most common machine learning techniques could not find a suitable solution in a reasonable time frame.

XCS overview

The eXtended Classifier system (XCS) [20] is the most successful learning classifier systems based on an accuracy model. Figure 1.5, describes the general architecture of the XCS model. XCS maintains a population of classifiers and each classifier consist of a condition-action-prediction rule, which maps input features to the output signal (or class).

A ternary representation of the form 0, 1, # (where # is don't care) for the condition and 0, 1 for the action can be used. In addition, real encoding can also be used to accurately describe the environment states. Input, in the form of data instances (a vector of features or genes), is passed to the XCS. A match set [M] is created consisting of rules (classifiers) that can be "triggered" by the given data instance. A covering operator is used to create new matching classifiers when [M] is empty. A prediction array is calculated for [M] that contains an estimation of the corresponding rewards for each of the possible actions. Based on the values in the prediction array, an action, a (the out- put signal), is selected. In response to a, the reinforcement mechanism is invoked and the prediction, p, prediction error, E, accuracy, k, and fitness, F, of the classifier are up- dated.

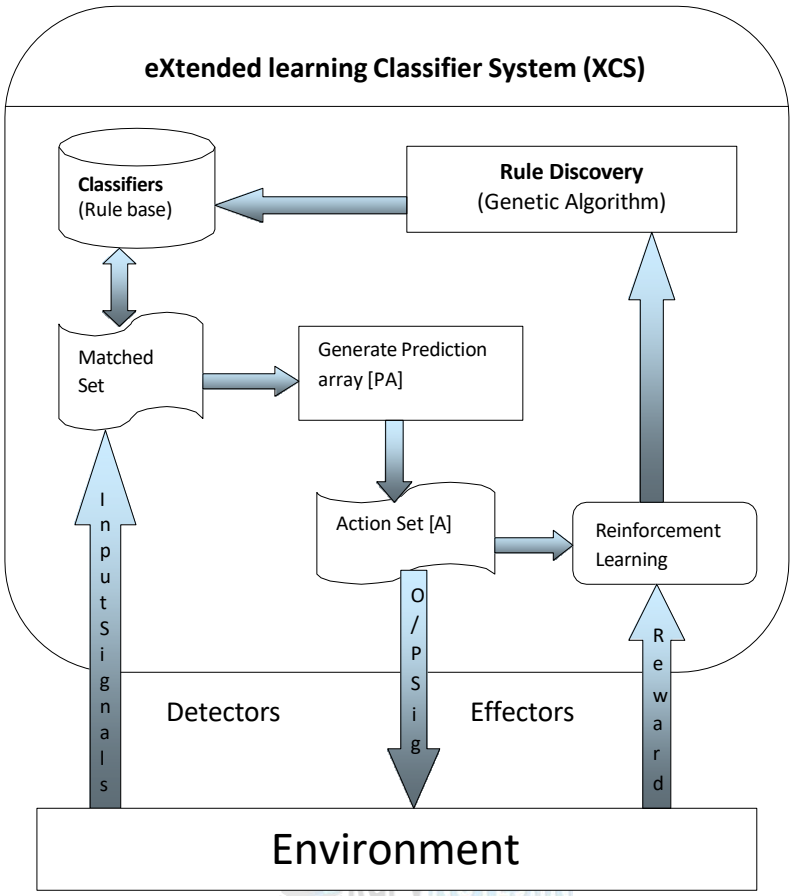

Figure: 1.5 XCS Architecture

4. Satellite Image Processing

PCI Geomatics is at the forefront of one of the most challenging applications of Cloud Computing: high volume processing of Earth Observation imagery. To play a significant role in the emerging high speed computing industry as it relates to imagery, PCI has invested significantly in developing technology that is innovative, nimble, portable, scalable, automated, and most of all optimized for high speed performance.

Cloud Computing Features	Benefit/Challenge
Reduced cost	**Benefit:** Optimize the costs by paying only for what an organization uses (Large capital expenses are not required, pay incrementally)
Increased storage	**Challenge**: Transferring large amounts of imagery is time consuming and costly. I/O is a specific issue that is more challenging to EO industry.
Highly automated	**Benefit:** Software can be maintained automatically. Cloud services include guaranteed machine up times and redundancy.

Flexibility	**Benefit:** Ability to add/remove computing resources as required, often automatically based on demand.
More mobility	**Benefit:** Simplified, easily accessible management consoles that can be managed/ viewed from anywhere.
Allows IT to shift focus	**Benefit:** No longer having to worry about constant server updates and other computing issues – IT can focus on innovation.

Table 1.2: Cloud Computing Featured

- **System Deployment**

The GXL system can be deployed on the Cloud much in the same way a non-cloud based GXL system. The main difference is that there is no need to purchase any physical hardware in order to configure a GXL system that can achieve the stated throughput requirements. PCI's non- Cloud based GXL systems are typically deployed on desktop or rack mounted systems, which include a certain set of hardware specifications, including (PCI would specify which hardware to purchase):

- CPU / GPU
- RAM
- Disk Storage
- UPS
- File Server
- Network Switch
- Operating System

- **Large volume processing on the Cloud**

PCI Geomatics has successfully deployed its GXL System to the Amazon Cloud to process large volumes of imagery for one of its customers. The following terminology is key to understanding the GXL Cloud system and how it is deployed to the Amazon Cloud.

Gluster:Main Data Repository (where data is stored, accessed by GXL system)

License Server: Central node which contains all s/w licenses, dbase for GXL, QA tools

Processing Nodes: Cloud based instances (virtual machines) that get allocated for processing, on demand

S3:Amazon Simple Storage Service – used for data storage (in the case of GXL, Gluster is the preferred method over S3 for data storage, due to more efficient handling of I/O

EC2: Elastic Computing – management console within Amazon Cloud Services for adding/removing computing resources

Instance: A virtual machine – Amazon provides standard configurations that range in processing capability (i.e. Micro, Small, Large, Extra Large)

5. CRM and ERP

Customer relationship management applications were among the first solutions businesses decided to migrate to cloud-based environment. Now, more companies are doing the same with ERP software, according to a recent TechTarget report. The news source explained that moving these apps to the cloud is driven by organizational demands and the cloud is an ideal choice to support such initiatives because of its cost-effectiveness.

Other industry professionals believe CRM and ERP business solutions are even more effective in a cloud-based setting. Mark Walker, technical services vice president at Scribe Software, said companies with cloud-based apps can improve the pacing of how they run their operations, the news source reported. Walker also said that organizations with these systems in place can provide better customer service by having data available to them immediately.

Business 2 Community's Thomas Stone also believes that ERP and CRM applications are given new life in cloud-based environments. The writer explained, however, that large ERP, CRM and HR deployments may still need to run on-site, but information from these platforms can be placed in the cloud for data analysis. Other apps created for customers are also ideal candidates for the cloud.

Some businesses are eager to launch most or all of their enterprise applications in the cloud simultaneously. This approach may work for smaller businesses, but according to Stone, medium and large organizations may want to implement their applications over time to determine their effectiveness. The main goal of migrating to the cloud should be to implement apps with minimal impact.

"There are many ways to use cloud computing and it's important to understand where the biggest benefits may be gained," Stone explained. "Portions of the infrastructure may be moved to the cloud over time allowing you to begin to see a return on your investment early in the restructuring of your computing resources."

But yet some of us need to care what's obscured; the folks tasked with building out a cloud environment need to know what's hidden in the cloud in order to build out an infrastructure that will support such a dynamic, elastic environment.

It is the obscuring of the infrastructure that makes cloud seems so simple. Because we're hiding all the moving parts that need to work in concert to achieve such a fluid environment it appears as if all you need is virtualization and voila! The rest will take care of itself.

But without a dynamic infrastructure supporting all the virtualized applications and, in many cases, infrastructure such an environment is exceedingly difficult to build.

Social networking

Social network platforms have rapidly changed the way that people communicate and interact. They have enabled the participation in digital communities as well as the representation, documentation and exploration of social relationships.

Social networks help boost internet usability by storing heavy multimedia content in cloud storage systems. Videos and photographs are the most popular content on social media, which essentially use up the maximum space allocated to them. They have the capacity to slow down applications and servers with all of their resource demands. Cloud computing vendors such as Salesforce and Amazon nowadays provide varied services including Customer Relationship Management (CRM) and Enterprise

Resource Planning (ERP). As they deliver these things through cloud servers, **clients can use the flexibility and scalability of the system** without purchasing standalone software or hardware.

Apart from data storage, the social networks are now also using clouds for various other tasks. For example, this can be **ideal for big data analytics.** One of the benefits of using cloud systems is that **users can access vast amount of structured and even non-structured data** easily. Just take a look at the much-improved analytics provided by sites like Facebook, especially for its business users.

Another way cloud computing becomes helpful is by **reducing the cost of data backup and recovery in case of a disaster.** If the data is only stored in one central location, it becomes much riskier. If something happens there, it is almost impossible to recover the data. But through cloud they remain accessible through shared resources across the globe. This is especially useful for social networks as the store personal data of its users, and so cannot afford to lose even one part of it.

Overall, it can be said that cloud computing has several usages, and some of them are still being discovered. For instance, in the near future, **personal secure clouds are likely to gain ground.** New age social networks and messaging apps such as Snapchat thrive on privacy and they will eventually utilize such resources to offer a more secure and faster service to its users.

Cloud Computing and Social Network Sites are among some of the most controversially discussed developments in recent years. The opportunities of using powerful computing resources on demand via the web are considered as a possible driver for the growth of the world economy. Cloud is a reality and will remain the most distinguished technological breakthrough, transforming the way business is done.

Web Resources:

http://computer.howstuffworks.com/cloud-computing/cloud-computing2.htm

https://www.ibm.com/blogs/cloud-computing/2014/02/top-7-most-common-uses-of-cloud-computing/

http://www.pewinternet.org/2008/09/12/use-of-cloud-computing-applications-and-services/

https://www.hindawi.com/journals/bmri/2013/398968/

http://zti.polsl.pl/w3/dmrozek/science/cloud4psp.htm

Subject Notes
CS 8002 - Cloud Computing

Unit-2

Cloud Computing Reference Model:

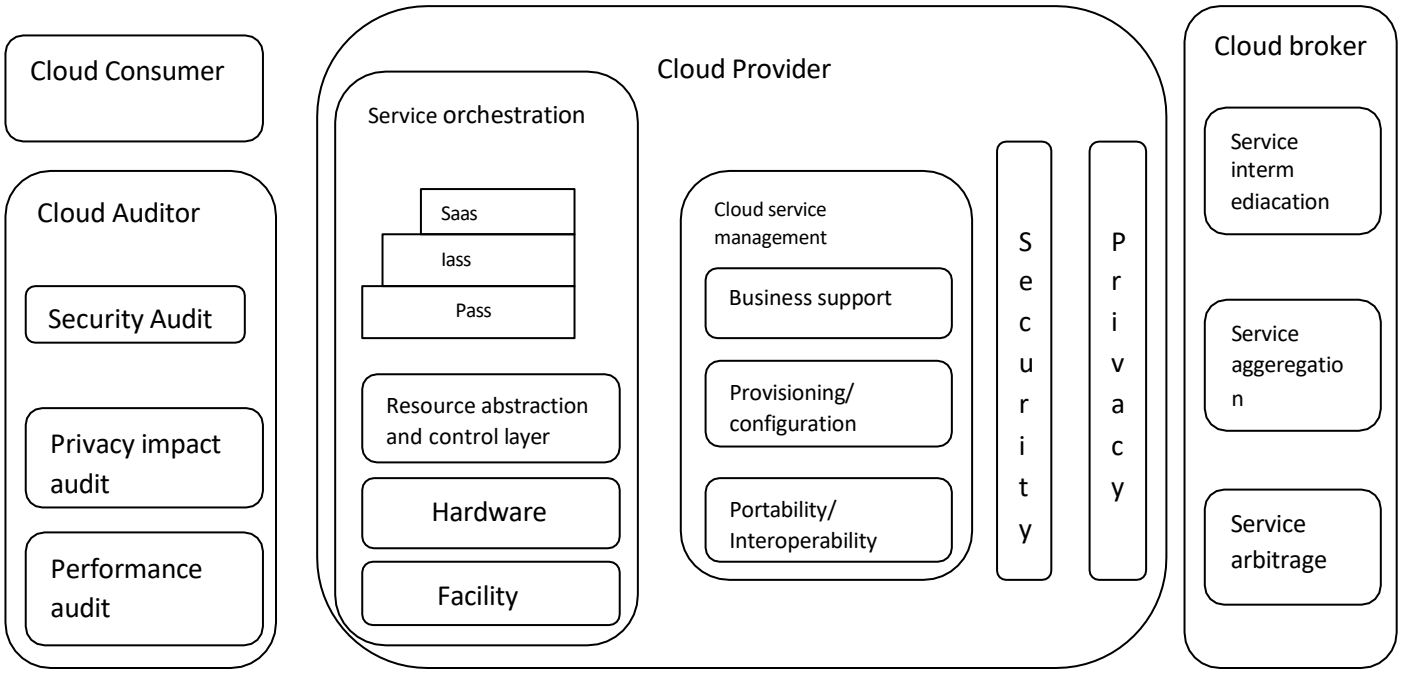

Figure: 2.1 Cloud Computing Reference Model

The cloud ecosystem for beginners: A data center view

From a cloud service provider point of view, data center giants like IBM, Google, Microsoft, and Amazon that have massive infrastructure at their disposal, use technologies such as virtualization, service-oriented architecture to "rent out" infrastructure to small and medium businesses (SMB) that appear to constitute a fairly large chunk of the customers.

Similarly, from a cloud consumer point of view, smaller businesses can reduce up-front infrastructure capital and maintenance costs by using the infrastructure (compute, memory, and storage) offered by the cloud providers. This can also reduce or keep their in-house infrastructure footprint or inventory under control.

IaaS forms the primary service delivery model, the others being software (SaaS) and platform services (PaaS). The primary use of IaaS is to run development and test, and production workloads and applications. These workloads run on the machines residing in the cloud data centers. These public cloud data centers reside in remote locations. From a cloud service consumer perspective, the user gets direct access to a machine in the cloud as if it were in the user's own backyard, however

connected through the Internet using a remote connection (remote desktop connection or through SSH). The machine is characterized by set of resources (CPU, memory, and storage), operating system and software, which are requested as per requirements by the user. The users may similarly use the SaaS and PaaS models to use readily available software, and develop applications on platforms respectively.

Some examples are as follows:

- IaaS providers: IBM SmartCloud Enterprise, Amazon Elastic Compute Cloud (EC2), RackSpace Hosting, Microsoft
- SaaS providers: Google, Microsoft, SalesForce, Yahoo
- PaaS providers: Google, Microsoft, TIBCO, VMware, Zoho

Now we shift our attention to how a cloud service provider delivers these services. First, we need to understand how virtualization acts as a key driver for cloud computing.

Hardware virtualization is a technology that enables the creation of multiple abstract (virtual) machines on the underlying physical hardware (bare metal). Every virtual machine (VM) has a set of resources (CPU, memory, storage), which forms a subset of the parent physical machine resources. You can assign resources to your VM based on your requirements. This also means that multiple VMs when packed together on a single piece of hardware helps us achieve server consolidation (optimally packing multiple VMs) thereby reducing server sprawl. Server sprawl was observed when companies used the traditional model of deploying a single heavy stand-alone application per physical server (1:1 — one application to one server). This, over the years, has resulted in increased capital and operational costs. Virtualization helps in consolidating multiple applications and aims to achieve optimum utilization of a physical hardware's underlying resources. From a cloud service provider's angle, physical machines in the data center are virtualized so as to deliver infrastructure resources to customers via virtual machines. Read more about hardware virtualization figure 2.1.

A cloud is a virtualized data center to achieve the following objectives:

- Elasticity: Ability to scale virtual machines resources up or down
- On-demand usage: Ability to add or delete computing power (CPU, memory) and storage according to demand
- Pay-per-use: Pay only for what you use
- Multitenancy: Ability to have multiple customers access their servers in the data center in an isolated manner

Let's look at the components that make up a "cloud." To understand this section better, think from the perspective of a cloud service provider so as to understand the components required to deliver cloud services. This perspective throws light on the data center, giving you an insight into how a cloud data center is structured. Two important terms in this context are management (managed-from) environment and managed (managed-to) environment. These terms inexplicitly describe the roles of a service provider and the service consumer.

The management environment is the central nervous system equivalent of the cloud; it manages the cloud infrastructure. This environment manages the infrastructure that is dedicated to the customers. The environment consists of components required to effectively deliver services to consumers. The various services offered span from image management and provisioning of machines to billing, accounting, metering, and more. The environment is characterized by hardware and software

components; realized by powerful computer servers, high speed network, and storage components. The cloud management system (CMS) forms the heart of the management environment along with the hardware components.

The managed environment is composed of physical servers and in turn the virtual servers that are "managed-by" the management environment. The servers in the managed environment belong to a customer pool; where customers or users can create virtual servers on-demand and scale up/down as needed. These virtual servers are deployed from the pool of available physical servers.

In short, the management environment controls and processes all incoming requests to create, destroy, manage, and monitor virtual machines and storage devices. In the context of a public cloud, the users get direct access to the VMs created in the "managed" environment, through the Internet. They can access the machines after they are provisioned by the management layer.

User makes a request to create a VM by logging onto the cloud portal.

The request is intercepted by the request manager and is forwarded to the management environment.

The management environment, on receiving the request, interprets it and applies to it provisioning logic to create a VM from the set of available physical servers.

External storage is attached to the VM from a storage area network (SAN) store during provisioning in addition to the local storage.

After the VM is provisioned and ready to use, the user is notified of this information and finally gains total control of the VM. The user can access this VM through the public Internet because the VM has a public IP address.

Similar to this workflow, users can decommission and manage their servers according to their needs. They can also create new images, store snapshots of their system, and so on.

Smart Ways Cloud Computing Can Help Organizations to Become Eco-Friendly

Tons of electronic waste ends up in the landfills all over the world, poisoning the earth and polluting our ecosystem – every single day. At the same time, a lot of power is used to feed the insatiable needs of the IT industry. Well, in the current scenario, where every business organization is doing all it can to reduce carbon footprints, cloud computing seems like a great option. It is perhaps, one of the most modern and advanced approaches to going green without compromising on business performance. Even the United Nations endorses it as an effective measure in reducing the world's carbon footprint. By moving over to the cloud, any company, big or small, can do their bit and making the planet greener and at cheaper cost.

1. By enabling remote working

While setting up a new plant in a remote location, a company would have to deploy their own experts. This means, exponential cost in flying the required personnel to the field and back till the work is complete. Since, airplanes burn tonnes of greenhouse gases, frequent flying is not the best way to go green. All this unnecessary travelling can be avoided by investing in cloud computing, which will allow your employees to work from anywhere.

2. By reducing the usage of paper

In a traditional office setup, large amount of paper is wasted in excessive documentation and printing.

But, with the help of cloud computing, an organization can save a lot of trees from being cut down. In a virtual office, each and every piece of information is accessible via internet at any time. So, there is no need for printed reports, memos or brochures. And, a paperless office is a sure way to go green.

3. By lowering power consumption

A study conducted in the United States, in 2007, reported that up to 60 per cent of office PCs are left on overnight, resulting in 19.82 billion kWh5 of electricity wastage and 14.4 million tonnes of associated carbon dioxide emissions. With cloud computing, any business can save a lot on the utility bills, as there is no need to invest in hardware and other systems that would need round-the- clock maintenance. This would make a huge difference in the power consumed by the company.

4. By cutting down the office space requirement

The adoption of cloud computing would mean that all your employees have remote access to all the data. More than one employee can work on the same file simultaneously making it possible for

them to communicate and share information. Thus, a company can save huge amount of investment in procuring a large office space and maintaining large servers.

5. By ensuring efficient management of resources

For most businesses, utilizing their servers to full capacity is not always possible. When there is a peak in the data load, in-house data centers would need extra servers to handle it efficiently. And, when there would a fall, then the additional servers would become idle. This is neither good for the organization nor for the environment. A cloud service provider has a large number of clients and is fully equipped to handle such issues. So, peaks of data load can be easily handled by allocating resources where needed. This would mean fewer machines, and less energy consumption.

Companies that adopt cloud computing could lead the way in making the IT industry more sustainable and significantly greener.

The delivery of dynamic, cloud-based infrastructure, platform and application services doesn't occur in a vacuum. In addition to best practices for effective administration of all the elements associated with cloud service delivery, cloud service management and cloud monitoring tools enable providers to keep up with the continually shifting capacity demands of a highly elastic environment.

Cloud monitoring and cloud service management tools allow cloud providers to ensure optimal performance, continuity and efficiency in virtualized, on-demand environments. These tools -- software that manages and monitors networks, systems and applications -- enable cloud providers not just to guarantee performance, but also to better orchestrate and automate provisioning of resources.

Cloud monitoring tools, specifically, enable cloud providers to track the performance, continuity and security of all of the components that support service delivery: the hardware, software and services in the data center and throughout the network infrastructure.

Through successful cloud service management and monitoring, cloud providers can use service quality to differentiate themselves in what remains a crowded and noisy marketplace.

Through successful cloud service management and monitoring, cloud providers can use service quality to differentiate themselves in what remains a crowded and noisy marketplace. Effective cloud service management also helps lower the risk of frequent cloud outages that can jeopardize security systems. Using these tools also supports greater operational efficiency, helping cloud providers minimize costs

and maximize profit margins. However, achieving these goals can be difficult in a complex virtual delivery environment where visibility and control are limits.

Cloud Service Management

Cloud service management shares some basic principles with traditional IT service management (ITSM). Cloud management tools help providers administrate the systems and applications that facilitate the on-demand service delivery model. The goal of these practices is to improve the efficiency of the cloud environment and achieve a high level of customer satisfaction.

Essentially, cloud service management takes the customer perspective as the measure of service assurance and manages all the individual IT resources in a way that will support that. This involves adjusting the operations and policies, as necessary, of all the assets in the virtual environment that support and affect the on-demand service delivery model. Such assets include servers, software and services that provide access and connectivity to these cloud services.

The core elements of cloud service management mirror those of traditional ITSM -- including cloud service-level agreement (SLA) management, cloud capacity management, availability management and billing -- and are applied to administrate a cloud delivery environment in a systemic way. These processes are supported with tools that track provisioning and change management, configuration management, release management, incident management, performance management and service continuity. Customers are supported directly and indirectly through a help desk function. Cloud service management is complemented by monitoring software that tracks operational information and feeds that data to the appropriate management resource.

Given the elastic, highly virtualized nature of cloud environments, there are some key differences in approaches to cloud service management and conventional IT service management. The two disciplines have different objectives, requiring tools that emphasize their individual requirements. Whereas the goals of traditional ITSM are effective SLA management, improved performance and streamlined billing, the goal of cloud service management is to orchestrate resources for fast provisioning, effective capacity management and ongoing service stability. Automation is vital to ensure efficiency and reduce costs.

Cloud service management shares all of the obstacles to managing any IT environment -- event correlation, incident prioritization, capacity management and performance management -- plus the unique challenges of a dynamic virtual environment. Visibility remains a common challenge in managing highly elastic and complex virtual systems that function at a tremendous scale.

Despite the challenges, cloud providers must implement management processes and use best practices to optimize efficiency, improve performance and, ultimately, maximize customer satisfaction. The highly competitive nature of the cloud market requires providers to focus on delivery in order to not just survive, but to thrive.

Types of cloud computing

- **Public Clouds**

A public cloud is basically the internet. Service providers use the internet to make resources, such as applications (also known as Software-as-a-service) and storage, available to the general public, or on a ‘public cloud. Examples of public clouds include Amazon Elastic Compute Cloud (EC2), IBM’s Blue Cloud, Sun Cloud, Google AppEngine and Windows Azure Services Platform.

For users, these types of clouds will provide the best economies of scale, are inexpensive to set-up because hardware, application and bandwidth costs are covered by the provider. It's a pay-per-usage model and the only costs incurred are based on the capacity that is used.

There are some limitations, however; the public cloud may not be the right fit for every organization. The model can limit configuration, security, and SLA specificity, making it less-than-ideal for services using sensitive data that is subject to compliancy regulations.

- **Private Clouds**

Private clouds are data center architectures owned by a single company that provides flexibility, scalability, provisioning, automation and monitoring. The goal of a private cloud is not sell "as-a-service" offerings to external customers but instead to gain the benefits of cloud architecture without giving up the control of maintaining your own data center.

Private clouds can be expensive with typically modest economies of scale. This is usually not an option for the average Small-to-Medium sized business and is most typically put to use by large enterprises. Private clouds are driven by concerns around security and compliance, and keeping assets within the firewall.

- **Hybrid Clouds**

By using a Hybrid approach, companies can maintain control of an internally managed private cloud while relying on the public cloud as needed. For instance during peak periods individual applications, or portions of applications can be migrated to the Public Cloud. This will also be beneficial during predictable outages: hurricane warnings, scheduled maintenance windows, rolling brown/blackouts.

The ability to maintain an off-premise disaster recovery site for most organizations is impossible due to cost. While there are lower cost solutions and alternatives the lower down the spectrum an organization gets, the capability to recover data quickly reduces. Cloud based Disaster Recovery (DR)/Business Continuity (BC) services allow organizations to contract failover out to a Managed Services Provider that maintains multi-tenant infrastructure for DR/BC, and specializes in getting business back online quickly.

Models of cloud computing

- **Infrastructure as a Service (IaaS):**

Infrastructure as a Service, sometimes abbreviated as IaaS, contains the basic building blocks for cloud IT and typically provide access to networking features, computers (virtual or on dedicated hardware), and data storage space. Infrastructure as a Service provides you with the highest level of flexibility and management control over your IT resources and is most similar to existing IT resources that many IT departments and developers are familiar with today.

- **Platform as a Service (PaaS):**

Platforms as a service remove the need for organizations to manage the underlying infrastructure (usually hardware and operating systems) and allow you to focus on the deployment and management of your applications. This helps you be more efficient as you don't need to worry about resource procurement, capacity planning, software maintenance, patching, or any of the other undifferentiated heavy lifting involved in running your application.

- **Software as a Service (SaaS):**

Software as a Service provides you with a completed product that is run and managed by the service provider. In most cases, people referring to Software as a Service are referring to end-user applications. With a SaaS offering you do not have to think about how the service is maintained or how the underlying infrastructure is managed; you only need to think about how you will use that particular piece software. A common example of a SaaS application is web-based email where you can send and receive email without having to manage feature additions to the email product or maintaining the servers and operating systems that the email program is running on.

Cloud Eco Systems

Tons of electronic waste ends up in the landfills all over the world, poisoning the earth and polluting our ecosystem – every single day. At the same time, a lot of power is used to feed the insatiable needs of the IT industry. Well, in the current scenario, where every business organization is doing all it can to reduce carbon footprints, cloud computing seems like a great option. It is perhaps, one of the most modern and advanced approaches to going green without compromising on business performance. Even the United Nations endorses it as an effective measure in reducing the world's carbon footprint. By moving over to the cloud, any company, big or small, can do their bit and making the planet greener and at cheaper cost.

1. By enabling remote working
While setting up a new plant in a remote location, a company would have to deploy their own experts. This means, exponential cost in flying the required personnel to the field and back till the work is complete. Since, airplanes burn tonnes of greenhouse gases, frequent flying is not the best way to go green. All this unnecessary travelling can be avoided by investing in cloud computing, which will allow your employees to work from anywhere.

2. By reducing the usage of paper
In a traditional office setup, large amount of paper is wasted in excessive documentation and printing. But, with the help of cloud computing, an organization can save a lot of trees from being cut down. In a virtual office, each and every piece of information is accessible via internet at any time. So, there is no need for printed reports, memos or brochures. And, a paperless office is a sure way to go green.

3. By lowering power consumption
A study conducted in the United States, in 2007, reported that up to 60 per cent of office PCs are left on overnight, resulting in 19.82 billion kWh5 of electricity wastage and 14.4 million tonnes of associated carbon dioxide emissions. With cloud computing, any business can save a lot on the utility bills, as there is no need to invest in hardware and other systems that would need round-the-clock maintenance. This would make a huge difference in the power consumed by the company.

4. By cutting down the office space requirement
The adoption of cloud computing would mean that all your employees have remote access to all the data. More than one employee can work on the same file simultaneously making it possible for them to communicate and share information. Thus, a company can save huge amount of investment in procuring a large office space and maintaining large servers.

5. By ensuring efficient management of resources
For most businesses, utilizing their servers to full capacity is not always possible. When there is a peak in the data load, in-house data centers would need extra servers to handle it efficiently. And, when

there would a fall, then the additional servers would become idle. This is neither good for the organization nor for the environment. A cloud service provider has a large number of clients and is fully equipped to handle such issues. So, peaks of data load can be easily handled by allocating resources where needed. This would mean fewer machines, and less energy consumption.

Companies that adopt cloud computing could lead the way in making the IT industry more sustainable and significantly greener.

Cloud Analytics

Information is a strategic asset. Companies acknowledge that value and are collecting huge volumes of data, from all possible sources. But very few companies can leverage that data to their competitive advantage. Challenges range from data accuracy and completeness to speed and complexity of implementing analytics.

An even bigger issue is that, once implemented, analytics remains so disconnected from operations that it is almost irrelevant. The insights revealed are generally at an aggregate level and provide information that is merely “good to know” and seldom actionable by operational teams.

Today, cloud and mobile technologies are providing enterprises of all sizes with opportunities to use big data and analytics to make better, data-driven decisions. New-generation platforms (cloud, big data, and analytics) bring analytics and operational applications together to deliver demonstrable ROI.

Cloud computing allows organizations to consolidate data from all sources, across all communication channels, and do it at a big data scale. Without cloud, collecting data from all internal applications, social networks, devices, and data subscriptions would be cost prohibitive for most organizations. On-premise big data deployments could involve significant operational risks and expensive infrastructure. The ongoing maintenance of on-premise systems itself would be daunting enough to discourage many organizations.

Let’s consider some of the advantages that cloud offers over on-premise data analytics implementations.

- **Robust Data Foundation**

Bringing together reliable data for analytics has always been a challenge. Analytics are not accurate if data is scattered, stale, and incomplete. Many of your applications and data sources, such as social and third-party data subscriptions, are in the cloud. In this environment, creating an on-premise data store is less than optimal. A cloud-based data management platform makes it easier for companies to blend data from all such sources and helps match, merge, and clean data. Real-time access to social and third-party data sources and real-time data stewardship enabled by cloud solutions keeps your data current, complete, and clean.

- **Fast Time to Value**

A modern data-management platform brings together master data management and big data analytics capabilities in the cloud so that business can create data-driven applications using the reliable data with relevant insights. The principal advantage of this unified cloud platform is faster time-to-value, keeping up with the pace of business. Whenever there is a need for a new, data-driven decision

management application, you can create one in the cloud quickly. There is no need to setup infrastructure (hardware, operating systems, databases, application servers, analytics), create new integrations, or define data models or data uploads. In the cloud, everything is already set up and available. Use cases are limited only by your imagination. Sales operation teams can create better alignments and account planning applications, marketing teams can create segmentation for campaign planning, contact centers can uncover up-sell and cross-sell opportunities, and strategy groups can simulate pre and post-merger scenarios.

- **Improved Collaboration**

On-premise and disconnected systems make it tedious to develop analytical models collaboratively and to share the insights. Team members use emails and printouts to discuss ideas and consolidate feedback manually. Development takes time; many inputs are lost, and many members with valuable ideas are never included. The situation is even more complicated in globally distributed teams. Teams cannot learn from each other, and they spend expensive resources duplicating analytics already performed by others.

In cloud-based big data analytics, groups collaborate on data incubation and analytics design and share insights across departments, across the globe. Insights are available in real time and, when included within operational applications, are actionable immediately. For example, marketing findings are not locked in marketing systems, but shared with all customer-facing teams. The information gathered by sales in the field is not left in spreadsheets, but is fed back to marketing, in a closed-loop, to improve the customer experience.

- **Quicker Adoption**

On-premise applications historically have seen slow adoption rates. Even after investment in training and skills development, utilization remains low and many applications are reduced to shelfware.

Built on the self-learning paradigm and user experience similar to consumer applications, cloud-based applications are easy to use and promote fast adoption. The cloud facilitates democratization of analytics across the organization, increasing the access and utilization. When insights from cloud-based analytics are presented within online operational applications, adoption improves even further. Users do not have to create one-off reports or log into separate systems to “run analytics.” It is just available within the current task. Data-driven applications in the cloud can be readily accessible to everyone from any place, any time, on any device.

- **Scalability and Elasticity**

Another big benefit of analytics in the cloud is on-demand computational power. Whether it is a Fortune 500 company or small to medium business, they can access similar analytic resources. With on-premise installations, there is always a risk of over-spending or underestimating the computing needs. Adding servers is not easy, and reducing them is equally agonizing.

Elasticity in cloud computing has taken that uncertainty out of the equation. With cloud technologies, you can start small and expand as your business needs grow, and scale back if your strategy changes. You can access higher compute power on demand if you are running complex analysis, and scale back once you are back on a routine.

- **Lower Total Cost of Ownership**

Companies are painfully aware of maintenance, upgrades, and migrations required by on-premise analytics platforms. Every 18 months or so, there is a massive effort to upgrade to a newer version. Not only is this costly, it affects business continuity. Not every new feature is backward compatible; businesses often end up struggling to redesign reports, redefine analysis, and redo integrations.

With cloud-based modern data management platforms with big data analytics, applications are always current. There are no upgrade issues, and enabling new capabilities requires minimal IT intervention. Companies can enjoy new features multiple times a year without big investments or downtime.

Reliable data is the foundation of analytics. If the data is not correct, complete, or current, you cannot expect much from the analytics. Cloud-based data management as a service helps organizations to blend master data and big data across all domains and formats, from all internal, third-party, and social media sources, to form a complete view of the business. This union of data, operations, and analytics, in a closed-loop, provides an unprecedented level of agility, collaboration, and responsiveness. All made possible by cloud technologies.

Cloud Interoperability & Standards

Interoperability in cloud computing has the same problems as interoperability between homegrown applications and commercial software, or between different commercial software. These problems in the cloud are widely recognized but standards bodies turned to them only recently. The world of cloud is a bit wilder than the world of applications because clouds currently offer an opportunity for cloud providers to lock in new, and not technically savvy, business customers.

- **Avoid synchronous communications between clouds**

Try to avoid synchronous communication between clouds as much as possible. Engage an acquire-store-resend model. You will pay some performance penalties but avoid binding the life cycles of your applications with the life cycles of SaaS. The goal of this preference is to achieve a loose coupling between clouds while maintaining some resiliency to changes in connectivity and location.

- **Monitor the connections**

Monitor connections in the integration hub at all available levels. Reserve a mechanism for an automated acquiring of lost connections. Compare monitored metrics against your contract (SLA) with the SaaS provider, and act actively on any discrepancies.

- **Pay attention to the interactions**

Like in any service-oriented ecosystem, put the maximum attention on semantics and ontologies of operations and data involved in the interactions between clouds. You will see that the informational aspects, not only formats, are crucial for all emerging cloud standards. Information "translation" in the cloud integration hub is a must-have feature.

- **Minimize the interactions**

Keep the number of interactions between clouds to the bare minimum but use coarse-grained interfaces. Watch for an availability of such interfaces in the selected SaaS and create a service layer on the top of your own applications if they do not support such granularity.

- **REST is best**

In line with the last note, try to model the application landscape in both clouds as a landscape of resources. This will allow you to minimize data volumes moved between clouds and construct a RESTful interaction solution. However, do not discard simple XML over HTTP and similar cheap mechanisms. More sophisticated integration is available through innovative services like cloud integration from Boomi or Cast Iron, which allow the Internet to be seen as an enterprise service bus (ESB) of sorts.

- **Do it yourself with security**

Do not trust any declarations of SaaS providers regarding security. Protect your channel to SaaS from the integration hub (this is one of the major roles of having such hub) with all security means your corporate policies specify. If your applications are deployed in another cloud, the communication channel with this one has to be equally protected.

Cloud Business Process Management

Business Process Management and the automation it delivers are key to operational efficiency. But, BPM also drives more growth, manages governance, risk, and compliance, and improves customer and employee engagement. What's not to love?

- **Design**

Model processes that drive powerful apps with a low-code process modeler driven by BPMN-notation.

- **Execute**

Create process automation that can transform your business. Eliminate reliance on old-school paper forms and speed operations.

- **Manage**

Dynamic Business Rules embed process automation in your apps, ensuring consistency and reinforcing organizational structure.

- **Optimize**

Automated testing, predictive analytics, dynamic reporting, business activity monitoring, and more keep your processes continually optimized and your apps in tip-top shape.

Advantages of Cloud-based BPM

Cloud-based BPM lets you "test the waters." Even if you are 100% convinced that adopting BPM will be a good move for your business (and we're 100% it will be), using BPM in the cloud allows you to try

it without making an all-in commitment. If you find managing business processes manually and on-premises works better for you, you haven't made significant investments that cannot then be recouped.

Cloud-based BPM saves you money. Because you will be using BPM software as a service (Saas) that is delivered from the cloud, you will not be building a large and complicated IT infrastructure. No infrastructure to build also means no infrastructure to maintain. And you "pay as you go" – per-use or subscription pricing takes the place of a sizeable up-front investment.

Cloud-based BPM saves you time. The lack of any large internal infrastructure means you'll be able to roll out business process management in your organization quite rapidly. This faster time to market boosts investor confidence. And because cloud-based apps and data are easier to coordinate, the processes you manage will be more efficient.

Cloud-based BPM encourages collaboration. Cloud computing erases borders that once hindered teamwork. A survey from RW3 CultureWizard finds that 41% of all corporate teams never meet in person, and 48% of respondents report that over half of their corporate teams include members in other countries. Understanding business processes that are managed from the cloud results in understanding how this kind of borderless productivity is now possible.

Cloud-based BPM equips you to go mobile. Mobile technology is transforming the way we work. In 2015, about 35% of workforce members worldwide were working on-the-go, from mobile devices—and they weren't just checking email or managing their calendars. More and more, they were managing business processes. Market research company Technavio predicts mobile BPM will grow at an annual rate of almost 21% between now and 2020.

Testing the Cloud: Definitions, Requirements, and Solutions

The virtualized data center, whether within the enterprise or located at a cloud service provider, must be properly provisioned in order to provide the necessary functions and performance of cloud-based applications. Testing of cloud services has some familiar aspects and some new challenges. Even though they will be used in a cloud environment, the basic components that populate the data center need to be tested for functionality, performance, and security. This is complemented with testing of the data center and end- to-end services.

At the network interconnectivity infrastructure level, testing must validate:

- Routers
- Switches, including fibre channel forwarders
- Application delivery platforms
- Voice over IP (VoIP) gateways

At the server and storage infrastructure level, testing must validate:

- Data center capacity
- Data center networks
- Storage systems

- Converged network adapters

At the virtualization level, testing must validate:

- Virtual hosts
- Video head ends
- VM instantiation and movement

At the security infrastructure level, testing must validate:

- Firewalls
- Intrusion Prevention Systems (IPS)
- VPN gateways

Network Interconnectivity Infrastructure Level

- Routers
- Switches, including fibre channel forwarders
- Application delivery platforms
- Voice over IP (VoIP) gateways

Each of the networking components used within the data center must be thoroughly tested for conformance to standards, functionality, interoperability, performance, and security before deployment. This type of testing is the bread and butter of network testing companies such as Ixia.

Ixia's test solutions cover the wide range of data center network testing. Ixia's chassis house up to 12 interface cards, which include Ethernet speeds from 10Mbps to 100Gbps; high-density solutions for 1Gbps and 10Gbps are available. Direct fibre channel interfaces are used for storage area network (SAN) testing. Each test port is backed by substantial special-purpose traffic generation hardware, and substantial compute power and memory.

Security Testing

Network security in a cloud environment is particularly important. Classical data centers can secure their facilities through the "front door" that connected them to the Internet or other corporate sites. Not so in a cloud environment. Each cloud computing and storage component can be located at a different physical location and connected over the Internet or private networks. Each of the connections is a potential security risk.

A number of dedicated security appliances are in widespread use, protecting enterprises and data centers worldwide. The culmination of the development of these devices is

the unified threat management (UTM) system that encompasses the roles of firewalls, intrusion prevention systems, anti-virus, anti-spam, and data loss prevention.

Virtual security applications are becoming widespread in the cloud environment. These software-only, VM-aware implementations of security functions are distributed between components of cloud applications. They serve to protect each component from other traffic on shared networks and other

VMs on virtualized servers.

Regardless of whether they are physical or virtual and where they are placed in the data center, security mechanisms must be tested thoroughly in three dimensions:

- **Effectiveness** – do the security mechanisms effectively defend against the attacks they were designed to prevent?

- **Accuracy** – does it produce any false positives?

- **Performance** – do the security mechanisms pass an acceptable amount of traffic? The last category is extremely important. Security devices have a difficult job to do watching all traffic on high speed links, inspecting for malware, fending off denial of service attacks, etc. They must be able to find and prevent attacks when processing large amounts of traffic. Likewise, they must pass an acceptable amount of "normal" when under heavy attack. A security device that cannot prevent penetration when under full load is easily defeated. A security device that blocks critical business applications when under attack has effectively been defeated.

Testing of network security devices requires a number of techniques, which will be discussed in the next few sections:

- Known vulnerability testing

- Distributed denial of service

Each cloud computing and storage component can be located at a different physical location and connectd over the Internet or private networks. Each of the connections is a potential security risk.

- Line-rate multiplay traffic

- Encrypted traffic

- Data leakage testing

- Known Vulnerability Testing

Known vulnerability testing is the cornerstone of network security device testing. Attacks are mounted against security mechanisms by using a large database of known malware, intrusions, and other attacks. A number of organizations exist to maintain this list. One leading organization is the U.S. National Vulnerability Database maintained by the National Institute of Standards and Technology (NIST). The Mitre Corporation provides access to this database, called the CVE—Common Vulnerabilities and Exposures. As of May 2010, more than 42,000 vulnerabilities are listed, with more than 15 added on a daily basis.

Proper security testing requires that a number of known vulnerabilities be applied to security devices at a significant percentage of line-rate. The device under test (DUT) should properly reject all such attacks, while maintaining a reasonable rate of transmission of "good" communications.

In addition, known vulnerabilities must be applied using the wide variety of evasion techniques. The combination of thousands of known vulnerabilities and dozens of evasion techniques requires that a subset of all possibilities be used for testing. Test tools offer representative samples, including special cases for newly published vulnerabilities.

- Distributed Denial of Service

Denial of service attacks often use large numbers of computers that have been taken over by hackers. Those computers, called "zombies", use dozens of attack techniques designed to overload network and security devices. This type of testing requires test equipment capable of simulating thousands of computers.

Devices must be tested to ensure that none of the denial of service attacks, singly or in combination, is able to disable the device. In addition, the ability of the DUT to accept new connections and provide an acceptable level of performance must be measured.

Testing and the Cloud

While many companies are approaching cloud computing with cautious optimism, testing appears to be one area where they are willing to be more adventurous. There are several factors that account for this openness toward testing in the cloud:

• Testing is a periodic activity and requires new environments to be set up for each project. Test labs in companies typically sit idle for longer periods, consuming capital, power and space. Approximately 50% to 70% of the technology infrastructure earmarked for testing is underutilized, according to both anecdotal and published reports.

• Testing is considered an important but non-business-critical activity. Moving testing

to the cloud is seen as a safe bet because it doesn't include sensitive corporate data and has minimal impact on the organization's busi- ness-as-usual activities.

• Applications are increasingly becoming dynamic, complex, distributed and compo-nent-based, creating a multiplicity of new challenges for testing teams. For instance, mobile and Web applications must be tested for multiple operating systems and updates, multiple browser platforms and versions, different types of hardware and a large number of concurrent users to understand their performance in real-time. The conventional approach of manually creating in-house testing environments that fully mirror these complexities and multiplicities consumes huge capital and resources.

Operational Challenges

Despite the bright upside, cloud-based testing has its limitations, too. Organizations must contend with a different set of challenges in their quest to reap cloud's benefits.

• **Lack of standards**: Presently, there are no universal/standard solutions to integrate public

cloud resources with user companies' internal data center resources. Public cloud providers have their own architecture, operating models and pricing mechanisms and offer very little interoperability. This poses a big challenge for companies when they need to switch vendors.

• **Security in the public cloud:** Security in the public cloud is still a major concern, and encryption techniques currently available today are considered insufficient. Procedures are being developed to improve security and performance in the public cloud. For instance, service providers are developing virtual private clouds and client partitions. The main cause for concern is that the data may be stored in a remote location beyond an organization's legal and regulatory jurisdiction.

• **SLAs**: Terms and conditions of cloud service are sometimes hard to understand, misleading and biased toward the vendor. Such areas include clauses governing data integrity, data preservation, data location and transfer, according to a study by The Center for Commercial Law

Studies at Queen Mary, University of London 2010. Companies would do well to be diligent and proactive in sorting through these issues with their vendors.

• **Infrastructure**: Some cloud providers offer only limited types of configurations, technology, servers and storage, networking and bandwidth, making it difficult to create real-time test environments.

• **Usage**: Improper usage of cloud-based test environments can increase costs. Even though some vendors offer pay-as-you-go cloud- based testing services, this approach can be expensive or out of sync with requirements, particularly if user estimates are too conservative or wildly overblown. Companies that apply pay-as-they-go approaches must first perfect their cost models or apply process-driven estimates rather than utilizing projections that are unsupported by data.

• **Planning**: Testing teams should rigorously plan their test environments, from utilization periods through disassembly. They should also be aware of the associated expenses, such as cost of encrypting data, before putting testing in a cloud environment, since these requirements will consume additional CPU and memory. It's important to monitor utilization of cloud resources to avoid over-usage and over-payment.

• **Performance**: As public clouds are shared by numerous users, there may be cases where a company might have to wait for the required bandwidth. There may also be cases where a service provider may suddenly announce disruption of service due to a maintenance window or network outage. Some of these issues can be resolved by working closely and proactively with the service provider.

Why is Cloud Testing Important?

Comparing with current software testing, cloud-based testing has several unique advantages listed below.

- Reduce costs by leveraging with computing resources in clouds – This refers to effectively using virtualized resources and shared cloud infrastructure to eliminate required computer resources and licensed software costs in a test laboratory.
- Take the advantage of on-demand test services (by a third-party) to conduct large-scale and effective real-time online validation for internet- based software in clouds.
- Easily leverage scalable cloud system infrastructure to test and evaluate system (SaaS/ Cloud/Application) performance and scalability.

IBM reported the experience on cloud testing in small business division, where a flexible and cost-efficient cloud-based development and testing environment is implemented, and cloud testing has demonstrated the following major benefits in [19].

- Reduce its capital and licensing expenses as much as 50% to 75% using virtualized resources.
- Reduce operating and labor costs as much as 30% to 50% by automating development and testing resource provisioning and configuration.
- Shorten its development and testing setup time from weeks to minutes.
- Improve product quality and reduce the detected defects by as much as 15% to 30%.
- Help to accelerate cloud computing initiatives with IBMCloudBurst™ implemented through

QuickStart services.

Forms of Cloud-Based Software Testing

There are four different forms of cloud-based software testing. Each of them has different focuses and objectives.

- **Testing a SaaS in a cloud** – It assures the quality of a SaaS in a cloud based on its functional and non-functional service requirements.

- **Testing of a cloud** – It validates the quality of a cloud from an external view based on the provided cloud specified capabilities and service features. Cloud and SaaS vendors as well as end users are interested in carrying on this type of testing.

- **Testing inside a cloud** - It checks the quality of a cloud from an internal view based on the internal infrastructures of a cloud and specified cloud capabilities. Only cloud vendors can perform this type of testing since they have accesses to internal infrastructures and connections between its internal SaaS(s) and automatic capabilities, security, management and monitor.

- **Testing over clouds** – It tests cloud-based service applications over clouds, including private, public, and hybrid clouds based on system- level application service requirements and specifications. This usually is performed by the cloud-based application system providers.

Test Type	**Testing focuses**	**Cloud/SaaS-Oriented Testing inside a Cloud**	**Online Application-Based Testing on a Cloud**	**Cloud-Based Application Testing over Clouds**
Service Function Testing	GUI-based and API- based service functions	Testing SaaS/Cloud-based service functions inside a cloud	Testing online-based application service functions on a cloud	Testing cloud-based application service functions over a cloud
Integration Testing	SaaS interactions and Cloud connections	Vendor-specific component and service integration inside a private/public cloud	Integration between online clients and back-end servers on a cloud	- End-to-end application integration over clouds - Integration with legacy systems over clouds
API and Connect ivity Testing	API interfaces and connectivity protocols (HTTPS, REST, SOAP, RMI)	SaaS/Cloud API & connectivity testing in a cloud	Testing user-centered service APIs and connectivity on a cloud	Testing application service APIs and connectivity over clouds
Performance & Scalability Testing	Performance and scalability based on a SLA	SaaS/Cloud performance and scalability testing in a cloud based on the given SLA	User-oriented application performance and scalability testing on a cloud	End-to-end system-level performance and scalability inside/on/over cloud based on a given SLA

Security Testing	SaaS/Application data, processes, functions, and user privacy	SaaS/Cloud security features and user privacy in a cloud	User-oriented security and privacy on a cloud	System-level end-to-end security over clouds
Interoperability & Compatibility Testing	Validate different client interfaces and technologies and diverse compatibilities on different platforms and browsers	Testing Cloud/ SaaS compatibility, connectivity protocols and UI/client technologies inside a cloud	Testing user-centered interoperability, compatibility of platforms/ OS/browsers, and client technologies on a cloud	Testing application compatibility, end-to- end interoperability and application connectivity to legacy systems over clouds
Regression Testing	Changed & impacted SaaS/Cloud service features and related APIs/	Cloud/SaaS-oriented regression testing inside a cloud	User-centered re-validation on a cloud	End-to-end application system regression over clouds

Table: 2.1 Forms of Cloud Based Testing

New Requirements and Features in Cloud Testing

There are four new requirements and features in cloud testing.

- **Cloud-based testing environment** – This refers to use a selected cloud infrastructure (or platform) as a base to form a test bed equipped with diverse and scalable computing resources, system infrastructures, and licensed tools, which are allocated using auto-provision based on static/ dynamic requests. Both virtual and physical computing resources can be included and deployed inside.

- **Service-level-agreements (SLAs)** – In cloud computing, all clouds, SaaS, and applications usually provide diverse services to their end users and customers with well-defined service-level-agreement. Naturally, these agreements will become a part of testing and quality assurance requirements, such as system reliability, availability, security, and performance.

- **Price models and service billing** – Since utility computing is one of basic concepts and features in cloud computing, so price models and utility billing becomes basic parts and service for testing as a service. In other words, required computing resources and infrastructures (including tools), and testing task services will be charged based on pre-defined cost models and

- **Large-scale cloud-based data and traffic simulation** - Applying and simulating large-scale online user accesses and traffic data (or messages) in connectivity interfaces is necessary in cloud testing, particularly in system-level function validation and performance testing.

	Internet-Based Software Testing (i.e. Distributed/Web-Based System Infrastructure)	**Cloud-Based Software Testing**

Primary Testing Objectives	- Assure the quality of system functions and performance based on the given specifications - Check usability, compatibility, interoperability.	- Assure the quality of functions and performance of SaaS , Clouds, and applications by leveraging a cloud environment - Assure the quality of cloud elasticity & scalability based a SLA
Testing as a service	- In-house internal software testing as engineering tasks	- Real-time on-demand testing service offered by a third-party - Online testing service based on a pre-defined SLA
Testing and Execution Time	- Offline test execution in a test lab. - Testing a product before its delivery	- On-demand test execution by third-parties; - Online test execution in a public cloud; - Offline test execution in a private cloud
Testing Environment	- A pre-fixed and configured test environment in a test lab. with purchased hardware and/or software	- An open public test environment with diverse computing resources - A scalable private test environment in a test lab.
Testing Costs	- Required hardware costs and software (license) costs - Engineering costs in a test process	- Based on a pre-defined service-level-agreement (SLA) - TaaS and Cloud testing service costs (pay-as-you-test) - Engineering costs in SaaS/Cloud/application vendors
Test Simulation	- Simulated online user access - Simulated online traffic data	- Virtual/online user access simulation - Virtual/online traffic data simulation
Function Validation	- Validating component functions and system functions as well as service features	- SaaS/Cloud service functions, end-to-end application functions - Leveraged functions with legacy systems
Integration Testing	- Function-based integration - Component-based integration - Architecture-based integration - Interface/connection integration	- SaaS-based integration in a cloud - SaaS integration between clouds - Application-oriented end-to-end integration over clouds - Enterprise-oriented application integration between SaaS/Cloud and with legacy systems
Security Testing	Aim to the following targets: - Function-based security features - User privacy - Client/server access security - Process access security - Data/message integrity	Aim to the following targets: - SaaS/Cloud security features, including monitor and measurement - User privacy in diverse web clients - End-to-end application security over clouds - SaaS/Cloud API and connectivity security - Security testing with virtual/real-time tests in a vendor's cloud
Scalability & Performance Testing	- Performed a fixed test environment - Apply simulated user access, ,messages, and test data - Online monitor and evaluation	- Performed in a scalable test environment based on a SLA - Apply both virtual and real-time online test data - Online monitor, validation, and measurement

Table: 2.2 Comparison of Cloud Testing

A comparison view about cloud testing products, solutions and services from four major players:

	PushtoTest http://www.pushtotest.com/	**Cloud Testing http://www.cloudtesting.com/**	**SOASTA http://www.soasta.com/**	**iTKO http://www.itko.com/**
Testing Products	- Test Maker	- Cross browser testing - Website archiving - Function testing	CloudTest™ supports test recording, editing, assembly, monitoring, and integrated Analytics	LISA product suite: LISA VirtualizeTM LISA Test LISA Validate LISA Pathfinder
Test Services	- PushTotest, TestOnDemand	- Function test service - Cross browser testing - Website archiving service	- CloudTest On-Demand - CloudTest Appliances	Education and consulting service on cloud-based application testing
Function Testing	Web-Based, RIA-based, and SOA-based function testing	Script-based function testing support for testers, developers, and website managers	Visual and UI-based function testing	- Automatic GUI testing, codeless testing, continuous and event-based validation
Test Development	- Record/playback, unit test, - object/component oriented test development	- Script-based test development - Web-based record/replay	- Test editing and test assembly with tools - Visual web-based record, capture, filtering and automated test clip creation	- Virtualized and codeless test development - Build executable tests by integrating with existing test repositories.

Table: 2.3 Comparison of Cloud testing products

Fault Tolerance Techniques and Comparative Implementation in Cloud Computing

Fault tolerance aim to achieve robustness and dependability in any system. Based on fault tolerance policies and techniques we can classify this technique into 2 types: proactive and reactive. The Proactive fault tolerance policy is to avoid recovery from fault, errors and failure by predicting them and proactively replace the suspected component means detect the problem before it actually come. Reactive fault tolerance policies reduce the effort of failures when the failure effectively occurs. These can be further classified into two sub-techniques error processing and fault treatment. Error processing aims at removing errors from the computational state. Fault treatment aims at preventing faults from being re- activated.

Fault tolerance is carried out by error processing which have two constituent phases. The phases are “effective error processing” which aimed at bringing the effective error back to a latent state, if possible before occurrence of a failure and “latent error processing” aimed at ensuring that the error does not become effective again.

EXISTING FAULT TOLERANCE TECHNIQUES IN CLUD COMPUTING

Various fault tolerance techniques are currently prevalent in clouds:

- **Check pointing**

It is an efficient task level fault tolerance technique for long running and big applications .In this scenario after doing every change in system a check pointing is done. When a task fails, rather than from the beginning it is allowed to be restarted that job from the recently checked pointed state.

- **Job Migration**

Some time it happened that due to some reason a job can- not be completely executed on a particular machine. At the time of failure of any task, task can be migrated to another machine. Using HA-Proxy job migration can be implemented.

- **Replication**

Replication means copy. Various tasks are replicated and they are run on different resources, for the successful execution and for getting the desired result. Using tools like HA-Proxy, Hadoop and AmazonEc2 replication can be implemented.

- **Self- Healing**

A big task can divided into parts .This Multiplication is done for better performance. When various instances of an application are running on various virtual machines, it automatically handles failure of application instances.

- **Safety-bag checks**

In this case the blocking of commands is done which are not meeting the safety properties.

- **S-Guard**

It is less turbulent to normal stream processing. S- Guard is based on rollback recovery. S-Guard can be implemented in HADOOP, Amazon EC2.

- **Retry**

In this case we implement a task again and gain. It is the simplest technique that retries the failed task on the same resource.

Virtual Desktop Infrastructure

As the size of your enterprise increases, so does the scope of its technical and network need? Something as seemingly simple as applying the latest OS hotfixes, or ensuring that virus definitions are up to date, can quickly turn into a tedious mess when the task must be performed on the hundreds or thousands of computers within your organization.

- **VDI Allows One to Manage Many**

A virtual desktop infrastructure (VDI) environment allows your company's information technology pros to centrally manage thin client machines, leading to a mutually beneficial experience for both end-users and IT admins.

- **What is VDI?**

Sometimes referred to as desktop virtualization, virtual desktop infrastructure or VDI is a computing model that adds a layer of virtualization between the server and the desktop PCs. By installing this virtualization in place of a more traditional operating system, network administrators can provide end users with 'access anywhere' capabilities and a familiar desktop experience, while simultaneously heightening data security throughout the organization.

Some IT professionals associate the acronym VDI with VMware VDI, an integrated desktop virtualization solution. VMware VDI is considered the industry standard virtualization platform;

VDI Provides Greater Security, Seamless User Experience Superior data security: Because VDI hosts the desktop image in the data center, organizations keep sensitive data safe in the corporate data center—not on the end-user's machine which can be lost, stolen, or even destroyed. VDI effectively reduces the risks inherent in every aspect of the user environment.

More productive end-users: With VDI, the end-user experience remains familiar. Their desktop looks just like their desktop and their thin client machine performs just like the desktop PC they've grown comfortable with and accustomed to. With virtual desktop infrastructure, there are no expensive training seminars to host and no increase in tech support issues or calls. End- user satisfaction is actually increased because they have greater control over the applications and settings that their work requires.

Other Benefits of VDI

- Desktops can be set up in minutes, not hours
- Client PCs are more energy efficient and longer lasting than traditional desktop computers
- IT costs are reduced due to a fewer tech support issues
- Compatibility issues, especially with single-user software, are lessened
- Data security is increased

Web Resources:

http://www.asigra.com/blog/cloud-types-private-public-and-hybrid

http://www.globaldots.com/cloud-computing-types-of-cloud/

https://fedoraproject.org/wiki/OpenStack

https://aws.amazon.com/types-of-cloud-computing/

Subject Notes
CS 8002 - Cloud Computing

Unit-3

Cloud Management

Following tasks are the main components of cloud management:

- **Task Resubmission**

A job may fail now whenever a failed task is detected, In this case at runtime the task is resubmitted either to the same or to a different resource for execution.

- **Timing check**

This is done by watch dog. This is a supervision technique with time of critical function.

- **Rescue workflow**

This technique allows the workflow to persist until it becomes unimaginable to move forward without catering the failed task.

- **Software Rejuvenation**

It is a technique that designs the system for periodic reboots. It restarts the system with clean state and helps to fresh start.

- **Preemptive Migration**

Preemptive Migration count on a feedback-loop control mechanism. The application is constantly monitored and analyzed.

- **Masking**

After employment of error recovery the new state needs to be identified as a transformed state. Now if this process applied systematically even in the absence of effective error provide the user error masking.

- **Reconfiguration**

In this procedure we eliminate the faulty component from the system.

- **Resource Co-allocation**

This is the process of allocating resources for further execution of task.

- **User specific (defined) exception handling**

In this case user defines the particular treatment for a task on its failure.

Several models are implemented based on these types of techniques. Table summarized the Comparison among various models based on protection against the type of fault, and procedure.

“AFTRC” a fault tolerance model for real time cloud computing based on the fact that a real time system can take advantage the computing capacity, and scalable virtualized environment of cloud computing for better implement of real time application. In this proposed model the system tolerates the fault proactively and makes the diction on the basis of reliability of the processing nodes [9].

“LLFT” is a propose model which contains a low latency fault tolerance (LLFT) middleware for providing fault tolerance for distributed applications deployed within the cloud computing environment as a service offered by the owners of the cloud. This model is based on the fact that one of the main challenges of cloud computing is to ensure that the application which are running on the cloud without a hiatus in the service they provided to the user. This middleware replicates application by the using of semi-active replication or semi-passive replication process to protect the application against various types of faults.

“FTWS” is a proposed model which contains a fault tolerant work flow scheduling algorithm for providing fault tolerance by using replication and resubmission of tasks based on the priority of the tasks in a heuristic metric. This model is based on the fact that work flow is a set of tasks processed in some order based on data and control dependency. Scheduling the workflow included with the task failure consideration in a cloud environment is very challenging. FTWS replicates and schedule the tasks to meet the deadline.

“FTM” is a proposed model to overcome the limitation of existing methodologies of the on-demand service. To achieve the reliability and resilience they propose an innovative perspective on creating and managing fault tolerance .By this particular methodology user can specify and apply the desire level of fault tolerance without requiring any knowledge about its implementation. FTM architecture this can primarily be viewed as an assemblage of several web services components, each with a specific functionality.

”Candy” is a component base availability modeling frame work, which constructs a comprehensive availability model semi automatically from system specification describe by systems modeling language. This model is based on the fact that high availability assurance of cloud service is one of the main characteristic of cloud service and also one of the main critical and challenging issues for cloud service provider.

“Vega-warden” is a uniform user management system which supplies a global user space for different virtual infrastructure and application services in cloud computing environment. This model is constructed for virtual cluster base cloud computing environment to overcome the 2 problems: usability and security arise from sharing of infrastructure.

"FT-Cloud" is a component ranking based frame work and its architecture for building cloud application. FT-Cloud employs the component invocation structure and frequency for identify the component. There is an algorithm to automatically determine fault tolerance stately.

"Magi-Cube" a high reliable and low redundancy storage architecture for cloud computing. The build the system on the top of HDFS and use it as a storage system for file read /write and metadata management. They also built a file scripting and repair component to work in the back ground independently. This model based on the fact that high reliability and performance and low cost (space) are the 3 conflicting component of storage system. To provide these facilities to a particular model Magi cube is proposed.

Model no	**Model name**	**Protection against Type of fault**	**Applied procedure for tolerate the fault**
M1	AFTRC	Reliability	1. Delete node depending on their reliability 2. Back word recovery with the help of check pointing
M2	LLFT	Crash-cost, trimming fault	Replication.
M3	FTWS	Dead line of work flow	Replication and resubmission of jobs
M4	FTM	Reliability, availability, on demand service	Replication users application and in the case of replica failure use algorithm like gossip based protocol.
M5	CANDY	Availability	1. It assembles the model components generated from IBD and STM according to allocation notation. 2. Then activity SNR is synchronized to system SRN by identifying the relationship between action in activity SNR and state transition in system SRN.
M6	VEGA-WARDEN	Usability, security, scaling	1. Two layer authentication and standard technical solution for the application.
M7	FT-CLOUD	Reliability, crash and value fault	1. Significant component is determined based on the ranking. 2. Optimal ft technique is determined.
M8	MAGI-CUBE	Performanc e, reliability, low storage cost	1. Source file is encoded in then splits to save as a cluster. 2. File recovery procedure is triggered is the original file is lost.

Table: 3.1 Comparison among various models based on protection against the type of fault, and procedure

Resiliency

Resilient computing is a form of failover that distributes redundant implementations of IT resources across physical locations. IT resources can be pre-configured so that if one becomes deficient, processing is automatically handed over to another redundant implementation. Within cloud computing, the characteristic of resiliency can refer to redundant IT resources within the same cloud (but in different physical locations) or across multiple clouds. Cloud consumers can increase both the reliability and availability of their applications by leveraging the resiliency of cloud-based IT resources.

- Resiliency is the capacity to rapidly adapt and respond to risks, as well as opportunities.
- This maintains continuous business operations that support growth and operate in potential adverse conditions.
- The reach and range step of the assessment process examines business driven, data-driven and event -driven risks.
- The resiliency blueprint includes different layers- facilities, technology, applications and data, processes
- The framework enables people to examine business, understand what areas of vulnerabilities that might exist and quickly pinpoint areas of concern and help them understand what actions they can take to reduce the risk associated with those areas.

The framework combines multiple parts to mitigate risks and improve business resilience

- From a facilities perspective, you may need to implement power protection
- from security perspective- to protect applications and data
- From process perspective- you may implement identification and documentation of most critical business process
- From organizational perspective- geographical diversity, backup of workstation data
- From strategy and vision perspective, you would want to have a crisis management

Provisioning

Cloud provisioning is the allocation of a cloud provider's resources to a customer.

When a cloud provider accepts a request from a customer, it must create the appropriate number of virtual machines (VMs) and allocate resources to support them. The process is conducted in several different ways: advance provisioning, dynamic provisioning and user self-provisioning. In this context, the term provisioning simply means “to provide."

With advance provisioning, the customer contracts with the provider for services and the provider prepares the appropriate resources in advance of start of service. The customer is

charged a flat fee or is billed on a monthly basis.

With dynamic provisioning, the provider allocates more resources as they are needed and removes them when they are not. The customer is billed on a pay-per-use basis. When dynamic provisioning is used to create a hybrid cloud, it is sometimes referred to as cloud bursting.

With user self-provisioning (also known as cloud self-service), the customer purchases resources from the cloud provider through a web form, creating a customer account and paying for resources with a credit card. The provider's resources are available for customer use within hours, if not minutes.

Provisioning process is a service that uses group of compliant processes called “solution Realization”

- Provisioned products are servers built with all the software and infrastructure required to support a business application.
- Standard solutions are defined so that standard workflows can be derived
- Server hardware is assembled, cabled and connected to the network and SAN before work orders are released.

Asset management

Cloud computing offers the potential to transform asset managers’ technology ecosystem. However, CIOs will need to consider a number of criteria to determine whether cloud solutions fit into their future plans and strategies.

Asset managers have traditionally developed custom technology or implemented a vendor solution. This decision was primarily influenced by the importance of the business function, the ability of the internal technology team to support it and the availability of mature products in the space.

Recently, however, the advent of cloud computing has added a completely new dimension to this decision-making process. It has opened up avenues to host custom-developed applications on third-party-managed platforms and created opportunities to use software as a service (SaaS).

According to a recent Confluence survey, over a four-year period, cloud solutions could lower the total cost of ownership by 55 percent. Numbers like these are making CIOs around the world take notice and realize that cloud computing provides an opportunity to shift to an entirely new technology operating model. With this shift, IT can move from managing applications on an internal infrastructure to managing the integration of different cloud services, platforms and cloud-based solutions. And while the potential benefits are clear, firms must conduct proper due diligence and understand the impact before making the move.

Asset management and change management interact regularly. The asset management strategy includes

- Software packaging
- Incident management
- Pool Management
- Release management
- configuration management
- Systems management
- Operational readiness Management
- Backup management

Concept of Map Reduce

The MapReduce framework has two parts:

1. A function called "Map," which allows different points of the distributed cluster to distribute their work
2. A function called "Reduce," which is designed to reduce the final form of the clusters' results into one output

The main advantage of the MapReduce framework is its fault tolerance, where periodic reports from each node in the cluster are expected when work is completed.

A task is transferred from one node to another. If the master node notices that a node has been silent for a longer interval than expected, the main node performs the reassignment process to the frozen/delayed task.

The MapReduce framework is inspired by the "Map" and "Reduce" functions used in functional programming. Computational processing occurs on data stored in a file system or within a database, which takes a set of input key values and produces a set of output key values.

Each day, numerous MapReduce programs and MapReduce jobs are executed on Google's clusters. Programs are automatically parallelized and executed on a large cluster of commodity machines. The runtime system deals with partitioning the input data, scheduling the program's execution across a set of machines, machine failure handling and managing required inter machine communication. Programmers without any experience with parallel and distributed systems can easily use the resources of a large distributed system.

MapReduce is used in distributed grep, distributed sort, Web link-graph reversal, Web access log stats, document clustering, machine learning and statistical machine translation.

Cloud Governance

This is a brief summary of some Cloud Computing governance issues:

Technical Issues

Determine how the Cloud Provider::

- Supports change management
- Provides for high-availability
- Provides for redundancy and failover (if any)
- Provides for security related to the Internet
- Provides for physical security

Legal Issues

It is important to determine what needs to be in a contract with your Cloud provider. Things to consider:

- Service standards to be maintained
- Retention of rights to your data
- Legal jurisdiction where the data center is located
- Privacy laws where the data center is located
- Liability of data breaches
- Policies and procedures related to providing digital forensics data in the event of any legal dispute, cyber attack, or data breach.
- Notification of changes when they occur at the data center
- Disaster recovery
- Remedies for various possible problems
- Details for what occurs at the beginning and end of the contract period

Business Issues

Your business relationship with a Cloud provider should involve:

- The Cloud provider's reputation
- Financial stability of the Cloud provider
- The length of time the Cloud provider has been in business
- Management practices for the data center

Cloud Backup

Get maneuverability by replacing expensive backup equipment and complicated media outsourcing by cheap backup in the cloud. With our **Cloud Backup Service**, we provide cheap and reliable backup storage of AWS in your data center or corporate network. You can continue to use your existing Enterprise Backup Software. We take care of all the technology and

guarantee the performance and security of cloud backups. While your data remains in Germany. The best: We adapt the cloud backup as your needs grow over time.

This complete Cloud Backup Service consists of part services that you can make use of individually:

- **Consultation and Implementation:** Determination of performance and quantity requirement based on Recovery Point Objectives (RPOs), recovery-time objectives (RTOs) and configuring and tuning Enterprise Backup Software of leading manufacturers for the utilization of backup targets in the Cloud.
- **Gateway Realization and Operation:** Installation of backup gateways from leading vendors including NetApp Steel gates, Ctera, Veeam and AWS Storage Gateway with corresponding encryption and redundancy configurations. Operation of the gateways includes maintenance, updates and troubleshooting.
- **Automation of the Backup Lifecycle Management:** Cloud Backup is usually done on backup gateways in the data center. From there the backups are automatically replicated on durable cloud storage by AWS. Prior to transfer data is de-duplicated, compressed and encrypted with user-defined keys. This ensures that, just in case, foreign intelligence services only get to see encrypted data. Lifecycle policies control how long backup data is stored in the data center, for fast recovery. Policies also regulate when backups are moved from durable cloud storage to an even more favorable long-term archive in the cloud and when backups are deleted, for example, at the end of the retention period.

Cloud Disaster Recovery

Get maneuverability by shifting the provision of expensive recovery infrastructure to the cloud. We provide proven recovery scenarios of your applications to secure Virtual Private Clouds (VPCs) with our **Cloud Recovery Service** in all available AWS regions worldwide. Your regular cloud backups guarantee the most current data state. We take care of the adjustments of the complete scenario, guarantee recovery point objectives (RPOs) and recovery time objectives (RTOs), and the safety and ease of use of the Cloud Recovery environment. The best: We adapt the Cloud Recovery if your application environments or your disaster recovery requirements change over time. This complete Cloud Recovery Service consists of part services that you can make use of individually:

- **Cloud Recovery Consultation and Implementation:** Analysis of Cloud Recovery scenarios, determining the resource and performance requirements, design and implementation of recovery scripts.
- **Automated Application Tests:** Regular automated checking the serviceability of the Cloud Recovery scenarios.

The Cloud Recovery Services - whether booked individually or as a whole - can be used for example for:

- 2-stage Cloud Recovery for all application environments of your company: while the restore takes place into the data center, quickly restore to a temporary cloud environment.
- Forks of the production environment for testing, simulations, expert opinions or reproduction of transactions outside the production environment.
- Recovery environments for critical communications infrastructure such as email, important documents (SharePoint, file servers), chat and video / audio conferencing for board and Communications Department, to other regions and continents, if necessary.

Virtualization Technology

Virtualization is the process of converting a physical IT resource into a virtual IT resource. Most types of IT resources can be virtualized, including:

- **Servers** - A physical server can be abstracted into a virtual server.
- **Storage** - A physical storage device can be abstracted into a virtual storage device or a virtual disk.
- **Network** - Physical routers and switches can be abstracted into logical network fabrics, such as VLANs.

- **Power** - A physical UPS and power distribution units can be abstracted into what are commonly referred to as virtual UPSs.

This section focuses on the creation and deployment of virtual servers through server virtualization technology.

The terms virtual server and virtual machine (VM) are used synonymously throughout this book. The first step in creating a new virtual server through virtualization software is the allocation of physical IT resources, followed by the installation of an operating system. Virtual servers use their own guest operating systems, which are independent of the operating system in which they were created.

Both the guest operating system and the application software running on the virtual server are unaware of the virtualization process, meaning these virtualized IT resources are installed and executed as if they were running on a separate physical server. This uniformity of execution that allows programs to run on physical systems as they would on virtual systems is a vital characteristic of virtualization. Guest operating systems typically require seamless usage of software products and applications that do not need to be customized, configured, or patched in order to run in a virtualized environment.

Virtualization software runs on a physical server called a host or physical host, whose underlying

hardware is made accessible by the virtualization software. The virtualization software functionality encompasses system services that are specifically related to virtual machine management and not normally found on standard operating systems. This is why this software is sometimes referred to as a virtual machine manager or a virtual machine monitor (VMM), but most commonly known as a hypervisor.

This section covers the following topics:

- Hardware Independence
- Server Consolidation
- Resource Replication
- Hardware-based and Operating System-based Virtualization
- Virtualization Operation and Management
- Technical and Business Considerations

Hardware Independence

The installation of an operating system's configuration and application software in a unique IT hardware platform results in many software-hardware dependencies. In a non-virtualized environment, the operating system is configured for specific hardware models and requires reconfiguration if these IT resources need to be modified.

Virtualization is a conversion process that translates unique IT hardware into emulated and standardized software-based copies. Through hardware independence, virtual servers can easily be moved to another virtualization host, automatically resolving multiple hardware-software incompatibility issues. As a result, cloning and manipulating virtual IT resources is much easier than duplicating physical hardware. The architectural models explored in Part III of this book provide numerous examples of this.

Server Consolidation

The coordination function that is provided by the virtualization software allows multiple virtual servers to be simultaneously created in the same virtualization host. Virtualization technology enables different virtual servers to share one physical server. This process is called server consolidation and is commonly used to increase hardware utilization, load balancing, and optimization of available IT resources. The resulting flexibility is such that different virtual servers can run different guest operating systems on the same host.

These features directly support common cloud computing features, such as on-demand usage, resource pooling, elasticity, scalability, and resiliency.

Resource Replication

Virtual servers are created as virtual disk images that contain binary ?le copies of hard disk content. These virtual disk images are accessible to the host's operating system, meaning simple file operations, such as copy, move, and paste, can be used to replicate, migrate, and back up the virtual server. This ease of manipulation and replication is one of the most salient features of virtualization technology as it enables:

- The creation of standardized virtual machine images commonly configured to include virtual hardware capabilities, guest operating systems, and additional application software, for pre-packaging in virtual disk images in support of instantaneous deployment.

- Increased agility in the migration and deployment of a virtual machine's new instances by being able to rapidly scale out and up.

- The ability to roll back, which is the instantaneous creation of VM snapshots by saving the state of the virtual server's memory and hard disk image to a host-based file. (Operators can easily revert to these snapshots and restore the virtual machine to its prior state.)

- The support of business continuity with efficient backup and restoration procedures, as well as the creation of multiple instances of critical IT resources and applications.

Operating System-Based Virtualization

Operating system-based virtualization is the installation of virtualization software in a pre-existing operating system, which is called the host operating system. For example, a user whose workstation has a specific version of Windows installed decides it wants to generate virtual machines. It installs the virtualization software into its host operating system like any other program and uses this application to generate and operate one or more virtual machine. This user needs to use its virtualization software to enable direct access to any of the generated virtual machines. Since the host operating system can provide hardware devices with the necessary support, operating system virtualization can rectify hardware compatibility issues even if the hardware driver is unavailable to the virtualization software.

Hardware independence that is enabled by virtualization allows hardware IT resources to be more flexibly used. For example, let's take a scenario in which the host operating system has the software necessary for controlling five network adapters that are available to the physical computer. The virtualization software can make the five network adapters available to the virtual machine, even if the virtualized operating system is usually incapable of physically housing five network adapters.

Virtualization software translates hardware IT resources that require unique software for operation into virtualized IT resources that are compatible with a range of operating systems. Since the host operating system is a complete operating system in itself, many operating system-based services that are available as organizational management and administration tools can be used to manage the virtualization host.

Examples of such services include:

- Backup and Recovery
- Integration to Directory Services
- Security Management

Operating system-based virtualization can introduce demands and issues related to performance overhead, such as:

- The host operating system consumes CPU, memory, and other hardware IT resources.
- Hardware-related calls from guest operating systems need to traverse several layers to and from the hardware, which decreases overall performance.
- Licenses are usually required for host operating systems, in addition to individual licenses for each of their guest operating systems.

A concern with operating system-based virtualization is the processing overhead required to run the virtualization software and host operating systems. Implementing a virtualization layer will negatively affect overall system performance. Estimating, monitoring, and managing the resulting impact can be challenging because it requires expertise in system workloads, software and hardware environments, and sophisticated monitoring tools.

VMware Virtualization

Server Virtualization

The architecture of today's x86 servers allows them to run only one operating system at a time. Server virtualization unlocks the traditional one-to-one architecture of x86 servers by abstracting the operating system and applications from the physical hardware, enabling a more cost-efficient, agile and simplified server environment. Using server virtualization, multiple operating systems can run on a single physical server as virtual machines, each with access to the underlying server's computing resources.

Server virtualization unleashes the potential of today's powerful x86 servers. Most servers operate less than 15 percent of capacity; not only is this highly inefficient, it also introduces server sprawl and complexity.

VMware vSphere offers a complete server virtualization platform that delivers:

- 80 percent greater utilization of server resources
- Up to 50 percent savings in capital and operating costs
- 10:1 or better server consolidation ratio

Network Virtualization

Network virtualization is the complete reproduction of a physical network in software. Virtual

networks offer the same features and guarantees of a physical network, yet they deliver the operational benefits and hardware independence of virtualization—rapid provisioning, no disruptive deployment, automated maintenance and support for both legacy and new applications.

Network virtualization presents logical networking devices and services—logical ports, switches, routers, firewalls, load balancers, VPNs and more—to connected workloads. Applications run on the virtual network exactly the same as if on a physical network.

You can create a highly scalable network fabric that provides greater levels operational efficiency and agility, faster provisioning, troubleshooting and cloning, with monitoring, QoS, and security all backed by VMware network virtualization software.

VMware NSX™ will be the world's leading network and security virtualization platform providing a full-service, programmatic and mobile virtual network for virtual machines, deployed on top of any general purpose IP network hardware.

The VMware NSX platform brings together the best of Nicira NVP and VMware vCloud® Networking and Security™ (vCNS) into one unified platform. VMware NSX exposes a complete suite of simplified logical networking elements and services including logical switches, routers, firewalls, load balancers, VPN, QoS, monitoring and security.

Desktop Virtualization

Deploying desktops as a managed service gives you the opportunity to respond quicker to changing needs and opportunities. You can reduce costs and increase service by quickly and easily delivering virtualized desktops and applications to branch offices, outsourced and offshore employees and mobile workers on iPad and Android tablets.

VMware desktop solutions are scalable, consistent, fully secure and highly available to ensure maximum uptime and productivity.

- Streamline deployment and management by delivering desktops as a service.
- Provide secure remote access to teleworkers and temporary workers without sacrificing performance.

Application Virtualization

Organizations are increasingly virtualizing more of their Tier 1 mission-critical business applications and platforms, such as databases, ERP, CRM, email, collaboration, Java middleware, business intelligence and many others.

In order to maintain the required levels of QoS and SLA for these Tier 1 business applications in virtual environments, IT organizations must focus equally on the virtualization components of the project and on the robust management and monitoring of virtualized business applications, as well as on maintaining corporate guidelines for business continuity and disaster recovery.

These virtualized applications simply run better and provide high availability, disaster recovery, speed and agility as well as cloud-readiness. With the VMware Tier 1 Application Virtualization solution built on VMware vCloud® Suite™, you can enhance the quality of IT services delivered, while simplifying your infrastructure, maximizing efficiency and eliminating costly over-provisioning.

Storage Virtualization

Storage virtualization is part of the software-defined storage layer that must offer improvements in performance and space efficiency without requiring the purchase of additional storage hardware.

It must enable rapid provisioning so that high-performance, space-efficient storage can be spun up as fast as a VM can be spun up today. It must offer a VM-centric storage management model that is intuitive for virtual administrators who are taking on more of the storage management tasks in virtual environments. And it must integrate with the hypervisor platform to leverage familiar, native workflows.

VMware storage virtualization is a combination of capabilities that provide an abstraction layer for physical storage resources to be addressed, managed and optimized in a virtualization deployment.

Storage virtualization technology provides a fundamentally better way to manage storage resources for your virtual infrastructure, giving your organization the ability to:

- Significantly improve storage resource utilization and flexibility
- Simplify OS patching and driver requirements, regardless of storage topology
- Increase application uptime and simplify day-to-day operations
- Leverage and complement your existing storage infrastructure

Block level storage

Anyone who has used a Storage Area Network (SAN) has probably used block level storage before. Block level storage presents itself to servers using industry standard Fibre Channel and iSCSI connectivity mechanisms. In its most basic form, think of block level storage as a hard drive in a server except the hard drive happens to be installed in a remote chassis and is accessible using Fibre Channel or iSCSI.

When it comes to flexibility and versatility, you can't beat block level storage. In a block level storage device, raw storage volumes are created, and then the server-based operating system connects to these volumes and uses them as individual hard drives. This makes block level storage usable for almost any kind of application, including file storage, database storage, virtual machine file system (VMFS) volumes, and more. You can place any kind of file system on block level storage. So, if you're running Windows, your volumes will be formatted with NTFS; VMware servers will use VMFS.

File level storage devices are often used to share files with users. By creating a block-based volume and then installing an operating system and attaching to that volume, you can share files out using that native operating system. Remember, when you use a block-based volume, you're basically using a blank hard drive with which you can do anything.

When it comes to backup, many storage devices include replication-type capabilities, but you still need to think about how to protect your workloads. With this type of storage, it's not unusual for an organization to be able to use operating system native backup tools or third-party backup tools such as Data Protection Manager (DPM) to back up files. Since the storage looks and acts like a normal hard drive, special backup steps don't need to be taken.

With regard to management complexity, block-based storage devices tend to be more complex than their file-based counterparts; this is the tradeoff you get for the added flexibility. Block storage device administrators must:

- Carefully manage and dole out storage on a per server basis.
- Manage storage protection levels (i.e., RAID).
- Track storage device performance to ensure that performance continues to meet server and application needs.
- Manage and monitor the storage communications infrastructure (generally iSCSI or Fibre Channel).

From a use case standpoint, there are a lot of applications that make use of this block-level shared storage, including:

- Databases. This is especially true when you want to cluster databases, since clustered databases need shared storage.
- Exchange. Although Microsoft has made massive improvements to Exchange, the company still does not support file level or network-based (as in, CIFS or NFS) storage. Only block level storage is supported.
- VMware. Although VMware can use file level storage via Network File System (NFS), it's very common to deploy VMware servers that use shared VMFS volumes on block level storage.
- Server boot. With the right kind of storage device, servers can be configured to boot from block level storage.

File level storage

Although block level storage is extremely flexible, nothing beats the simplicity of file level storage when all that's needed is a place to dump raw files. After all, simply having a centralized, highly available, and accessible place to store files and folders remains the most critical need in many organizations. These file level devices -- usually Network Attached Storage (NAS) devices provide a lot of space at what is generally a lower cost than block level storage.

File level storage is usually accessible using common file level protocols such as SMB/CIFS (Windows) and NFS (Linux, VMware). In the block level world, you need to create a volume, deploy an OS, and then attach to the created volume; in the file level world, the storage device handles the files and folders on the device. This also means that, in many cases, the file level storage device or NAS needs to handle user access control and permissions assignment. Some devices will integrate into existing authentication and security systems.

On the backup front, file level storage devices sometimes require special handling since they might run non-standard operating systems, so keep that in mind if you decide to go the file level route.

With the caveat that you may need to take some steps with regard to authentication, permissions, and backup, file level-only devices are usually easier to set up than block level devices. In many cases, the process can be as simple as walking through a short configuration tool and moving forward.

If you're looking for storage that screams -- that is, if you need high levels of storage performance -- be very careful with the file level option. In most cases, if you need high levels of performance, you should look at the block level options. Block level devices are generally configurable for capacity and performance. Although file-level devices do have a performance component, capacity is usually the bigger consideration.

File level use cases are generally:

- Mass file storage. When your users simply need a place to store files, file-level devices can make a lot of sense.

- VMware (think NFS). VMware hosts can connect to storage presented via NFS in addition to using block level storage.

The block and file worlds are converging. Some new storage devices include both block and file level capabilities. So if you are torn about whether to go with block or file, a hybrid/converged device might fit your needs.

Hypervisor Virtualization Software

A hypervisor, also known as a virtual machine manager/monitor (VMM), is computer hardware platform virtualization software that allows several operating systems to share a single hardware host. Each operating system appears to have the host's processor, memory, and resources to it. Instead, the hypervisor is controlling the host processor and resources, distributing what is needed to each operating system in turn and ensuring that the guest operating systems/virtual machines are unable to disrupt each other.

The term 'hypervisor' originated in IBM's CP-370 reimplementation of CP-67 for the System/370, released in 1972 as VM/370. The term 'hypervisor call' refers to the paravirtualization interface, by which a guest operating system accesses services directly from the higher-level control program. This is the same concept as making a supervisor call to the same level operating system.

Hypervisor Classifications

Hypervisors are classified into two types:

- **Bare Metal/Native Hypervisors**

Software systems that run directly on the host's software as a hardware control and guest operating system monitor. A guest operating system thus runs on another level above the hypervisor. This is the classic implementation of virtual machine architectures.

A variation of this is embedding the hypervisor in the firmware of the platform, as is done in the case of Hitachi's Virtage hypervisor and VMware ESXi. See below definition.

- **Embedded/Host Hypervisors**

Software applications that run within a conventional operating system environment. Considering the hypervisor layer being a distinct software layer, guest operating systems thus run at the third level above the hardware.

Comparison between VLAN and VSAN

S.No.	VLAN(Virtual Local Area Network)	VSAN(Virtual Storage Area Network)
1	VLAN is a network technology used to logically separate large broadcast domains using layer 2 devices.	VSAN is a logical partition in a storage area network.
2	It divides the network into different virtual sub-networks reduces unnecessary traffic and improve performance.	VSANs allow traffic to be isolated within specific portions of a storage area network.
3	VLANs are implemented to achieve scalability, security and ease of network management.	The use of multiple VSAN's can make a system easier to configure and scale out.
4	VLAN's can quickly adapt to change in network requirements and relocation of workstations and server nodes.	In this subscribers can be added or relocated without the need for changing the physical layout.
5	The purpose of implementing a VLAN is to improve the performance of a network or apply appropriate security features.	The VSANs minimizes the total system's vulnerability, security is improved. VSANs also offer the possibility of data redundancy, minimizing the risk of catastrophic data loss.

Table: 3.2 Comparison between VLAN & SAN

Web Resources:

https://hortonworks.com/

https://www.itgovernance.co.uk/cloud-governance

http://www.cloudcomputingpatterns.org/map_reduce/

http://searchstorage.techtarget.com/definition/virtual-storage-area-network

http://www.netmagicsolutions.com/disaster-recovery-on-cloud

Subject Notes
CS 8002 - Cloud Computing

Unit-4

Cloud Security Fundamentals

Cloud evolution can be considered synonymous to banking system evolution. Earlier people used to keep all their money, movable assets (precious metals, stones etc.) in their personal possessions and even in underground lockers as they thought that depositing their hard earned money with bank can be disastrous. Banking system evolved over the period of time. Legal and security process compliances protected by Law played a big role in making banking and financial systems trustworthy. Now, people hardly keep any cash with them. Most of us carry plastic money and transact digitally. Cloud computing is also evolving the same way.

Robust cloud architecture with strong security implementation at all layers in the stack powered with legal compliances and government protection is the key to cloud security. As Banks didn't vanish despite frauds, thefts and malpractices, cloud security is going to get evolved but as much faster rate. Digital world has zero tolerance for waiting! Evolution is natural and is bound to happen.

Cloud is complex and hence security measures are not simple too. Cloud needs to be secured at all layers in its stack. Let's briefly look into major areas.

At infrastructure level: A sysadmin of the cloud provider can attack the systems since he/she has got all the admin rights. With root privileges at each machine, the sysadmin can install or execute all sorts of software to perform an attack. Furthermore, with physical access to the machine, a sysadmin can perform more sophisticated attacks like cold boot attacks and even tamper with the hardware.

Protection measures:

1. No single person should accumulate all these privileges.

2. Provider should deploy stringent security devices, restricted access control policies, and surveillance mechanisms to protect the physical integrity of the hardware.

3. Thus, we assume that, by enforcing a security processes, the provider itself can prevent attacks that require physical access to the machines.

4. The only way a sysadmin would be able to gain physical access to a node running a costumer's VM is by diverting this VM to a machine under his/her control, located outside the IaaS's security perimeter. Therefore, the cloud computing platform must be able to confine the VM execution inside the perimeter, and guarantee that at any point a sysadmin with root privileges remotely logged to a machine hosting a VM cannot access its memory.

5. TCG (trusted computing group), a consortium of industry leader to identify and implement security measures at infrastructure level proposes a set of hardware and software technologies to enable the construction of trusted platforms suggests use of "remote attestation" (a mechanism to detect changes to the user's computers by authorized parties).

At Platform level:

Security model at this level relies more on the provider to maintain data integrity and availability. Platform must take care of following security aspects:

1. Integrity
2. Confidentiality
3. Authentication
4. Defense against intrusion and DDoS attack
5. SLA

At Application level:

The following key security elements should be carefully considered as an integral part of the

SaaS application development and deployment process:

1. SaaS deployment model
2. Data security
3. Network security
4. Regulatory compliance
5. Data segregation
6. Availability
7. Backup/Recovery Procedure
8. Identity management and sign-on process

Most of the above are provided by PaaS and hence optimal utilization of PaaS in modeling SaaS

is very important.

Some of the steps which can be taken to make SaaS secured are:

- Secure Product Engineering
- Secure Deployment
- Governance and Regulatory Compliance Audits
- Third-Party SaaS Security Assessment

At Data level:

Apart from securing data from corruption and losses by implementing data protection mechanism at infrastructure level, one needs to also make sure that sensitive data is encrypted during transit and at rest.

Apart from all the above measures, stringent security process implementation should also be part of making cloud secure. Periodic audits should happen. Governing security laws should be amended with advent in technologies, ethical hacking and vulnerability testing should be performed to make sure the cloud is secure across all layers.

Cloud Security Devices

It doesn't matter what size you are when it comes to protecting your network. Big company, small company, and startup: Hackers will still want your information and they'll still stealthily poke holes in your network wherever they can.

You need to get security measures in place and fast.

That's why "security as a service" companies have become vital for anyone looking to deploy security for everything from documents to your entire business.

Security as a service can be loosely described as a "software as a service" security tool that doesn't require any on-premise hardware or software distribution. Unlike older security tools, like anti-virus software that needs to be installed on every single computer on your network, it's almost plugged and play — you click a button (and likely put in some credit card information) and sudd enly you've got major security resources at your fingertips.

These security services aren't the same as an on-premise firewall that watches the network from a physical appliance attached in your data center. But these products promise to protect you from malware, help you keep track of who signs into your network, monitor all your other cloud applications such as Salesforce and Google Docs, and more.

Small businesses can benefit from this kind of distribution model because it doesn't require a big IT or security teams to get it up and running. Of course, you're trusting a lot of your security to another company, but in reality these security-focused third parties have more resources (read: time and money) to focus on security than you do.

So what are the best security-as-a-service products out there? We talked to experts in the security community to compile this initial list of the top-tier providers.

Here are our top 6 in no particular order:

- **VentureBeat**

It is researching cloud platforms and we're looking for your help. We're starting with marketing specifically marketing automation. Help us by filling out a survey, and you'll get the full report when it's complete.

- **Qualys**

Qualys secures your devices and web apps, while helping you remain compliant through i ts cloud-only solution — no hardware or software required. The company analyzes threat information to make sure nothing gets in your system. If some malware already happens to be there, it will give you the steps to fix the problem. Beyond that, Qualys will verify that the issue has been fixed. It scans any and all web apps you use for vulnerabilities as well, keeping your data safe while you head out in the wonderful world of SaaS, IaaS, and PaaS. In the future, Qualys plans to create a cloud-only firewall to even further protect your websites from harm.

- **Proofpoint**

When we talk about attack vectors — holes in the network where bad guys can get in — email pops out as one of the weakest links. Proofpoint focuses specifically on email, with cloud-only services tailored to both enterprises and small to medium sized businesses. Not only does it make sure none of the bad stuff gets in, but it also protects any outgoing data. Proofpoint further promises that while it stores that data to prevent data loss, it does not have the keys to decrypt any of the information.

- **Zscaler**

Zscaler calls its product the "Direct to Cloud Network," and like many of these products, boasts that it's much easier to deploy and can be much more cost efficient than traditional appliance security. The company's products protect you from advanced persistent threats by monitoring all the traffic that comes in and out of your network as a kind of "checkpost in the cloud." But you don't have to filter all that traffic in from one central point. You can monitor specific, local networks as well given the flexibility of the cloud. Zscaler also protects iOS and Android devices within your company, which can then be monitored through its special mobile online dashboard.

- **CipherCloud**

CipherCloud is here to secure all those other "as a service" products you use, such as Salesforce, Chatter, Box, Office 365, Gmail, Amazon Web Services, and more. It promises to protect that prized company data you're just giving away to these services, as well as your communications, and more. It does this through many of the means we've already seen including encryption, traffic monitoring, anti-virus scans, and more. It also provides mobile security support.

- **DocTrackr**

DocTrackr is a security layer that sits on top of file sharing services such as Box and Microsoft Sharepoint. It is built on the idea that once you send a document out of your system, it is truly out of your hands: People can save it, change it, send it, and more and you've lost control of it. DocTrackr aims to stop that from happening. It lets you set user privileges for each person you share a document with. It further tracks everyone who opens the file, so you know who's looking at your stuff — and you can even pull documents back, effectively "unsharing" them, if you want.

Secure Cloud Software Requirements

1. **Authentication**: The process of providing identity is called authentication. Most computer system uses a user ID and password combination for identity and authentication. You identity yourself using a user ID and authenticate your identity with a password. Let's look at some examples of authentication from everyday life: at an automatic bank machine, you identify yourself using bank card, when you use a credit card etc.

2. **Single sing on**:Single sing on (SSO) is a session/user authentication process that permits a user to enter one name and password in order to access multiple applications. The process authenticates the user for all the applications they have been given rights to and eliminates further prompts when they switch applications during a particular session.

3. **Delegation**: If a computer user temporarily hands over his authorizations to another user then this process is called delegation. There are two classes of delegation.

Delegation at authentication level and delegation at access control level

4. **Confidentiality**: confidentiality assures you that cannot be viewed by unauthorized people. The confidentiality service protects system data and information from unauthorized disclosure. When data leave one extreme of a system such as client's computer in a network, it ventures out into a non-trusting environment. So, the recipient of data may not fully trust that no third party like a cryptanalysis or a man-in-the middle has eavesdropped on the data.

5. **Integrity**: It assures you that data has not changed without your knowledge (the information cannot be altered in storage or transit sender and receiver without the alteration being detected).The integrity can be used in reference to proper functioning of a network, system, or application.

6. **Non-repudiation**: Repudiation is the denial by one of the entities involved in a communication of having participated in all or part of the communication. Such denial can be prevented by non-repudiation. Non repudiation allows an exchange of data between two parties in such a way that the parties cannot subsequently deny their participation in the exchange.

7. **Privacy**: Internet privacy involves the desire or mandate of personal privacy concerning transaction or transmission of data via the internet. It also involves the exercise of control over the type and amount of information revealed about person on the internet and who mat access said information personal information should be managed as part of the data use organization. It should be manage from the time the information is conceived through to its final disposition.

8. **Trust**: Organization's belief in the reliability, truth, ability, or strength of someone or something. Trust revolves around 'assurance' and confidence that people, data entities, information or processes will function or behave in expected ways. Trust may be human to human, machine to machine, human to machine or machine to human. At a deeper level, trust might be regarded as a consequence of progress towards security or privacy objectives.

9. **Policy**: The term policy is high-level requirement that specify which access is managed and who, under what circumstances, may access what information. A security policy should fulfill many purposes. It should protect people and information, and set the rules for expected behavior by users, system administrators, management, and security personnel.

10. **Authorization**: Authorization is the act of checking to see if a user has the proper permission to access a particular file or perform a particular action. It enables us to determine exactly what a user is allowed to do. Authorization typically implemented through the use of access control. While determining what access will be provided to the parties to whom we have provided authorized access, there is an important concept we should consider, called the principle of least privilege.

11. **Accounting**: accounting services keep track of usage of services by other services/ users so that they can be charged accordingly.

12. **Audit**: Audit services keep track of security related events.

Cloud Computing Security Challenges

Clouds are everywhere these days. They are often cheaper, more powerful, compatible with single sign-on (SSO) and often accessible via a Web browser. There are four main types of clouds: on-

premises, or clouds hosted by the customer; off-premises, or clouds hosted by a vendor; dedicated, which are clouds used only for a particular tenant; and shared, a cloud where resources are spread among many tenants.

These categories are more descriptive than public and private clouds. There are also virtual machine-based clouds where several separate computing environments can be used, versus bare-metal, where each compute node is a separate physical machine.

Threats to the Cloud

The first and most dangerous threat in any IT system is the insider threat. It's especially hard to defend against because users, and particularly administrators, have usually been granted some degree of trust. Technological countermeasures can usually be circumvented if the user has the right level of access. This is why it is critical for organizations to have an efficient off boarding process so that disgruntled released employees do not have access to the systems.

Side-channel threats occur when an attacker has the ability to obtain information from another tenant's node by measuring some side effect of the system's use. These have been popularized in the research community but, to IBM X-Force's knowledge; have not been seen in the real world.

Perhaps the most dangerous real-world threat is the loss of authority over the cloud control interface. We aren't talking about the provisioning portal but rather the administrative interface of your enterprise's cloud. Think of it as a control console for your cloud nodes.

In the right situation, this can lead to a complete loss of integrity, confidentiality and availability. Note that the attack here is against the interface's Web server, or a cross-site scripting (XSS) or cross-site request forging (CSRF) attack against the administrator's Web browser.

Make sure the interface's Web server is up to date and that the interface does not have any XSS or CSRF vulnerabilities. These are just good security practices in general and are not unique to the cloud. If you use SSO, be sure your security assertion markup language (SAML) implementation follows the recommended specification.

Additionally, use two-factor authentication. Note that this is good practice for restricting access to any sensitive servers and data.

Additional Risks to Cloud Environments

There is a somewhat rare attack called virtual host confusion. It is often seen with content delivery networks and shared platform-as-a-service (PaaS) clouds. This attack can allow for server impersonation under the right circumstances. Once again, the X-Force team is not aware of this being exploited in the wild. For more information, read the paper "Network-based Origin Confusion Attacks against HTTPS Virtual Hosting."

This attack is from the same group that identified Logjam, FREAK, SLOTH and others. To prevent this attack, never use certificates for more than one domain. Avoid using wildcard certificates and carefully configure TLS caching and ticketing parameters to be different for every Web server. Finally, make sure your domain fallback page is an error page.

Shared data and computations on shared (typically off-premises) clouds can be exposed in the right circumstances. This particularly applies to MapReduce operations. To prevent this leakage, consider dedicated clouds, where there is a lesser chance of malicious actors having a presence.

Never make the mistake of assuming that on-premises or dedicated clouds need not be secured according to industry best practices. These clouds are often considered a more valuable target by attackers.

Finally, there is shadow IT, or the inability of IT to monitor the activities of the user. This happens when the user's client is connected to a cloud with an encrypted connection. In that case, the user can interact with the cloud and perhaps perform unauthorized actions. To combat this, consider federating. Monitor your logs to see which applications are in use and use a proxy to intercept cloud traffic. You can also use an analytics engine and create relevant rules at your endpoint device.

Overcoming Challenges

In general, what can be done to improve cloud security? Always follow the best security practices whether you are a tenant or a provider, such as tracking new vulnerabilities and attacks against components of your cloud. If you are a cloud provider, do background research on entities that wish to join your environment.

If you are a tenant, always understand your cloud model and compensate for any weaknesses inherent in that type. Be sure to support TLS 1.2 access. This ensures stronger cryptography and is the latest secure protocol for connections to Web servers.

Both providers and tenants should institute regular vulnerability scanning as frequently as is feasible. They should also lock IP addresses so only authorized networks are able to access your cloud or site. If this is not possible as a provider, then be sure to employ strong authentication and access controls.

As a provider, make logs relevant to your tenants available. This complements the tenant's own logging.

As a tenant, make sure all software is up to date. PaaS providers need to do the same with their environments. In one of the most important measures, tenants must encrypt data. This is critical for data protection, but be sure to implement cryptography correctly. There are solutions available to minimize the ciphertext reduplication problem.

Virtualization Security in Cloud Computing

2011 ended with the popularization of an idea: Bringing VMs (virtual machines) onto the cloud. Recent years have seen great advancements in both cloud computing and virtualization On one hand there is the ability to pool various resources to provide software-as-a-service, infrastructure-as-a-service and platform-as-a-service. At its most basic, this is what describes cloud computing. On the other hand, we have virtual machines that provide agility, flexibility, and scalability to the cloud resources by allowing the vendors to copy, move, and manipulate their VMs at will. The term *virtual machine* essentially describes sharing the resources of one single physical computer into various computers within itself. *VMware* and *virtual box* are very commonly used virtual systems on desktops. Cloud computing effectively stands for many computers pretending to be one computing environment. Obviously, cloud computing would have many virtualized systems to maximize resources.

Keeping this information in mind, we can now look into the security issues that arise within a cloud-computing scenario. As more and more organizations follow the "Into the Cloud" concept, malicious hackers keep finding ways to get their hands on valuable information by manipulating safeguards and breaching the security layers (if any) of cloud environments. One issue is that the cloud -computing scenario is not as transparent as it claims to be. The service user has no clue about how his information is processed and stored. In addition, the service user cannot directly control the flow of data/information storage and processing. The service provider usually is not aware of the details of the service running on his or her environment. Thus, possible attacks on the cloud-computing environment can be classified in to:

1. **Resource attacks:**

   These kinds of attacks include manipulating the available resources into mounting a large-scale botnet attack. These kinds of attacks target either cloud providers or service providers.

2. **Data attacks**: These kinds of attacks include unauthorized modification of sensitive data at nodes, or performing configuration changes to enable a sniffing attack via a specific device etc. These attacks are focused on cloud providers, service providers, and also on service users.
3. **Denial of Service attacks**: The creation of a new virtual machine is not a difficult task, and thus, creating rogue VMs and allocating huge spaces for them can lead to a Denial of Service attack for service providers when they opt to create a new VM on the cloud. This kind of attack is generally called virtual machine sprawling.
4. **Backdoor:** Another threat on a virtual environment empowered by cloud computing is the use of backdoor VMs that leak sensitive information and can destroy data privacy.
5. Having virtual machines would indirectly allow anyone with access to the host disk files of the VM to take a snapshot or illegal copy of the whole System. This can lead to corporate espionage and piracy of legitimate products.

With so many obvious security issues (and a lot more can be added to the list), we need to enumerate some steps that can be used to secure virtualization in cloud computing.

The most neglected aspect of any organization is its physical security. An advanced social engineer can take advantage of weak physical-security policies an organization has put in place. Thus, it's important to have a consistent, context-aware security policy when it comes to controlling access to a data center. Traffic between the virtual machines needs to be monitored closely by using at least a few standard monitoring tools.

After thoroughly enhancing physical security, it's time to check security on the inside. A well-configured gateway should be able to enforce security when any virtual machine is reconfigured, migrated, or added. This will help prevent VM sprawls and rogue VMs. Another approach that might help enhance internal security is the use of third-party validation checks, preformed in accordance with security standards.

Cloud Security Architecture

Architecting appropriate security controls that protect the CIA of information in the cloud can mitigate cloud security threats. Security controls can be delivered as a service (Security-as-a-Service) by the provider or by the enterprise or by a 3rd party provider. Security architectural patterns are typically expressed from the point of security controls (safeguards) – technology and processes. These security

controls and the service location (enterprise, cloud provider, 3rd party) should be highlighted in the security patterns.

Security architecture patterns serve as the North Star and can accelerate application migration to clouds while managing the security risks. In addition, cloud security architecture patterns should highlight the trust boundary between various services and components deployed at cloud services. These patterns should also point out standard interfaces, security protocols (SSL, TLS, IPSEC, LDAPS, SFTP, SSH, SCP, SAML, OAuth, Tacacs, OCSP, etc.) and mechanisms available for authentication, token management, authorization, encryption methods (hash, symmetric, asymmetric), encryption algorithms (Triple DES, 128-bit AES, Blowfish, RSA, etc.), security event logging, source-of-truth for policies and user attributes and coupling models (tight or loose).Finally the patterns should be leveraged to create security checklists that need to be automated by configuration management tools like puppet.

In general, patterns should highlight the following attributes (but not limited to) for each of the security services consumed by the cloud application figure 4.1:

Logical location – Native to cloud service, in-house, third party cloud. The location may have an implication on the performance, availability, firewall policy as well as governance of the service.

Protocol – What protocol(s) are used to invoke the service? For example REST with X.509 certificates for service requests.

Service function – What is the function of the service? For example encryption of the artifact, logging, authentication and machine finger printing.

Input/Output – What are the inputs, including methods to the controls, and outputs from the security service? For example, Input = XML doc and Output =XML doc with encrypted attributes.

Control description – What security control does the security service offer? For example, protection of information confidentiality at rest, authentication of user and authentication of application.

Actor – Who are the users of this service? For example, End point, End user, Enterprise administrator, IT auditor and Architect.

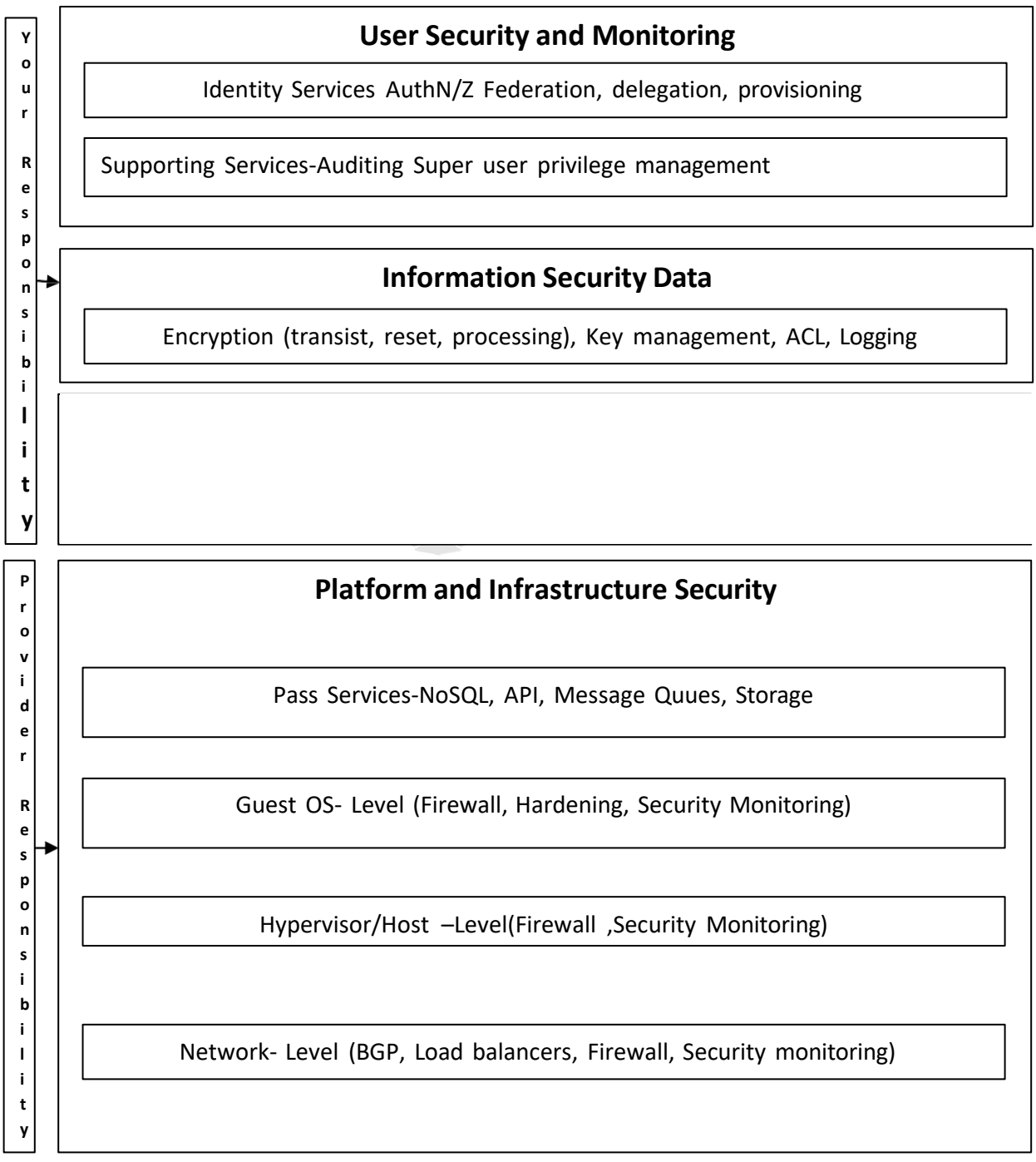

Figure: 4.1 Cloud Security Architecture

Web Resources:

https://aws.amazon.com/security/introduction-to-cloud-security/

https://www.fortinet.com/solutions/enterprise-midsize-business/cloud-security.html

https://www.solutionary.com/managed-security-services/cloud-security/

http://www.csoonline.com/article/3053159/security/cloud-security-challenges.html

Subject Notes
CS 8002 - Cloud Computing

Unit 5

Market Based Management of Clouds

Cloud marketing is the process of an organizations efforts to market their goods and services online through integrated digital experiences, by which they are specialized for every single end user. The aim is to use advertising methods to target customers via online applications through social media websites such as Facebook, Twitter and various online portals to target consumers. Cloud marketing platforms are supported with third parties who maintain the platform. Cloud Marketing was established by a company in Raleigh, North Carolina calling it SharedVue Marketing Technologies. The marketers targeting clients need to ensure their material is compatible with all electronic media devices.

Advantages of Cloud marketing for business

- **Cost Effectiveness**

Cloud marketing enables a business to decrease the cost of marketing distribution materials; these include sending catalogues and magazines to consumers as digital media, allowing a business to send promotional content through digital formats, which enables a faster and cheaper approach to target consumers. The cost reduces the printing costs and increases efficiency using online material accessible continuously.

- **Customization**

Customization allows a business to creatively use interactive means to create an relevant and effective advertising approach when targeting a consumer. Customization includes social media sites such as Facebook to customize pages to send to friends or the public over the internet. Marketers can combine data through third party data sources, including email and surveys, to visualize the consumer's experience.

- **Time**

Time is vital for targeting customers. Advertising using traditional methods, such as posters and surveys, have limited time before they often become invalid. Cloud marketing enables a business to produce advertising when required. The material can easily be removed and if a campaign or season is over, the material can be erased from the internet or adapted to enhance material to the end user, linking with the customization element to ensure the marketing material is fit for its purpose, and delivered at the correct time.

Disadvantages of Cloud Marketing for business

- **User Experience**

When a company markets their products and services to a consumer, the end consumer is not able to

touch or physically manage the product or service. The experience could potentially lay of customers that have been targeted, if the businesses efforts have not satisfied the consumers decision to buy the merchandise. The material content would vary on the device, as compatibility and operating systems will affect the material content being delivered.

- **Fraudulent Material**

Internet fraud has grown rapidly globally, faster than the internet. More and more fraudulent criminals can send promotional pop ups in the form of online advertising on the World Wide Web to attract web traffic to display promotional content. The malware attacks can lay off customers responding to marketing material posted to their devices. Labor MP Chris Evans said: 'Copycat websites are a part of a growing industry which exists purely to trick the public out of their hard-earned money.

- **Digital Divide**

Digital divide is the partition between a given population within their use of information technology. This can be due to factors including:

- Geographic
- Cultural
- Economic growth
- Democracy
- Disabilities

This limits a business's performance to market their goods and services globally to new locations if there is limited access to information technology in certain locations. The segment of consumers would be unable to experience and view online marketing methods from a business or resources resulting in adopted a traditional method of leaflets and bill boards known as direct marketing.

- **Cloud Marketing Plan Strategy**

Strategy is the direction of action which will achieve a goal or objective. The strategy for cloud marketing is broken down into 4 key elements.

- **Establishing the goals**

The first steps into cloud marketing include finding the objective or goal for the marketing project. The proposer would need to clearly state the objectives, which can be retained in quantitative or qualitative data. By establishing the goal and objectives of the marketing campaign, this limits the plan being deployed haphazardly.

- **Development**

The development stage is where the marketing team creates the graphics and media material. The web development team find a method to post the material onto the world wide web or online source. The marketing ad would need to meet its main objective and purpose, the development team will need to develop and plan to make the material visually appealing.

- **Maintenance**

The maintenance step will require updating whilst the material is online. it will require continuous upkeep. Cloud marketing techniques include regular updating to ensure they are reaching their end user and have a valid subject. Marketing members are responsible for moderating any discussion

boards and keeping content updated increasing the validity.

- **Evaluation**

Throughout the duration of the marketing material, the message would need to be evaluated to determine how successful it has been to the end user. The outcome should be established in the strategy allowing the marketer to adapt and increase the overall efficiency of the cloud marketing method.

Federated Clouds/Inter Cloud

Cloud Federation refers to the unionization of software, infrastructure and platform services from disparate networks that can be accessed by a client via the internet. The federation of cloud resources is facilitated through network gateways that connect public or external clouds, private or internal clouds (owned by a single entity) and/or community clouds (owned by several cooperating entities); creating a hybrid cloud computing environment. It is important to note that federated cloud computing services still rely on the existence of physical data centers.

- **Cloud Federation Benefits**

The federation of cloud resources allows clients to optimize enterprise IT service delivery. The federation of cloud resources allows a client to choose the best cloud services provider, in terms of flexibility, cost and availability of services, to meet a particular business or technological need within their organization. Federation across different cloud resource pools allows applications to run in the most appropriate infrastructure environments. The federation of cloud resources also allows an enterprise to distribute workloads around the globe, move data between disparate networks and implement innovative security models for user access to cloud resources.

- **Cloud Federation Implementation**

One weakness that exists in the federation of cloud resources is the difficulty in brokering connectivity between a client and a given external cloud provider, as they each possess their own unique network addressing scheme. To resolve this issue, cloud providers must grant clients the permission to specify an addressing scheme for each server the cloud provider has extended to the internet. This provides customers with the ability to access cloud services without the need for reconfiguration when using resources from different service providers. Cloud federation can also be implemented behind a firewall, providing clients with a menu of cloud services provided by one or more trusted entities.

Cloud Federation Stack

With the aid of Cloud computing technology, businesses and institutions make compute resources available to customers and partners to create more capable, scalable, flexible, and cost effective environments for application development and hosting. Cloud computing continues the trend started with on-demand, strategic outsourcing, and grid computing, to provideIT resources as a standardized commodity, targeting real-time delivery of infrastructure and platform services. A next stepin this evolution is to have cooperating providers of Cloud services in which a customer request submitted to one Cloudprovider is fulfilled by another, under mediation of a brokering structure. This latter idea invokes a federation ofCloud domains providing a service analogous to that of interoperating grid resources created for a similar goal by researchinstitutions using grid brokers in the grid computing framework.

To illustrate how this works, consider a business providing a SaaS offering from a private or public Cloud. Users submit requests to the application layer which assesses if sufficient local resources are available to service the requests within a specified time. If the application layer cannot meet its service goals it can optionally fulfill the requests through an independent SaaS layer provider of the same service as indicated by the horizontal (federation. Results are returned to the user as if locally produced by the application executing in Cloud A. Federation at the SaaS layer is analogous to the use in traditional business of 'sub' or 'peer' contractors who supply equivalent final parts or services to the primary provider facilitating elasticity to support a dynamic market. While this approach is common in industry sectors that produce goods or services such as manufacturing or publishing, it is not as common in software due to lack of standard interfaces and insufficient market forces to motivate sharing at the service layer. An application layer under stress also has a second option to increase capacity through delegation. In this service abstraction, the application layer works together with its underlying layers to provide the required computing needs. In delegation, the application layer asks the PaaS layer in the local Cloud for additional resources. The request for more resources may befulfilled in multiple ways depending on availability in the current Cloud. The PaaS layer can delegate to the local IaaS layer a request for more raw virtual machines and then provision the necessary platform software. If sufficient resources are not available locally the PaaS layer can attempt to acquire them from another Cloud in the federation through brokering at the PaaS layer.

In a typical scenario, the PaaS layer represents executing middleware such as web application containers and other application execution platforms, or distributed data applications. Here a more general view of federation is needed in which these support programs and environments form the federations between the Clouds in a way that isolates them from the underlying infrastructure layer. Some current middleware products, such as web application servers (e.g., IBM WebSphere Application Server or Oracle Fusion Middleware), provide isolation or lightweight virtualization from the underlying hardware and allow applications to dynamically expand across machines increasing capacity.

While attractive from a business perspective, this federated Cloud model requires new technologies to work efficiently.Because it is a layered model, an important part of the design is to maintain isolation of concerns between layers. For example, the SaaS application delivers a result to the customer in a certain response time. It is aware of the aggregate processing and network transmissions necessary to meet the delivery time. But the application does not need to know the details of the underlying infrastructure. Thus, it is necessary to translate requirements at the application to those understood by the PaaS and IaaS layers. This is accomplished through empirical modeling and experiments that map metrics of application performance such as response time onto the middleware and compute resource requirements understood by the PaaS or IaaS layer.

One challenge to making the operation of delegation work is to introduce a standardized form of expressing inter-layermappings. Some work along this line is contained in the manifestapproach used by the Reservoir project. A related issue is how to choose between delegation and federation when both options are available. Selection criteria suchas the mapping of performance metrics may be combined with policies as discussed in Sections 2 and 5. Another challenge is defining the protocols and policies for the inter-Cloud brokering required to join each layer in a federation. Considers brokering at different Cloud service layers and then proceeds to the inner workings and policy issues by whichbrokers expose and share Cloud services and resources.

Third Party Cloud Services: Google App Engine

Google App Engine is an application hosting and development platform that powers everything from enterprise web applications to mobile games, using the same infrastructure that powers Google's

global-scale web applications. Developers know that time-to-market is critical to success, and with Google App Engine's simple development, robust APIs and worry-free hosting, you can accelerate your application development and take advantage of simple scalability as the application grows. With support for Python, Java, and Go, you don't have to change the way you work. Your application can take advantage of powerful APIs, High Replication data storage, and a completely hands-free hosting environment that automatically scales to meet any demand, whether you're serving several users or several million.

Google App Engine makes it easy to take your app ideas to the next level.

- **Quick to start**

With no software or hardware to buy and maintain, you can prototype and deploy applications to your users in a matter of hours.

- **Simple to use**

Google App Engine includes the tools you need to create, test, launch, and update your apps.

- **Rich set of APIs**

Build feature-rich services faster with Google App Engine's easy-to-use APIs.

- **Immediate scalability**

There's almost no limit to how high or how quickly your app can scale.

- **Pay for what you use**

Get started without any upfront costs with App Engine's free tier and pay only for the resources you use as your application grows.

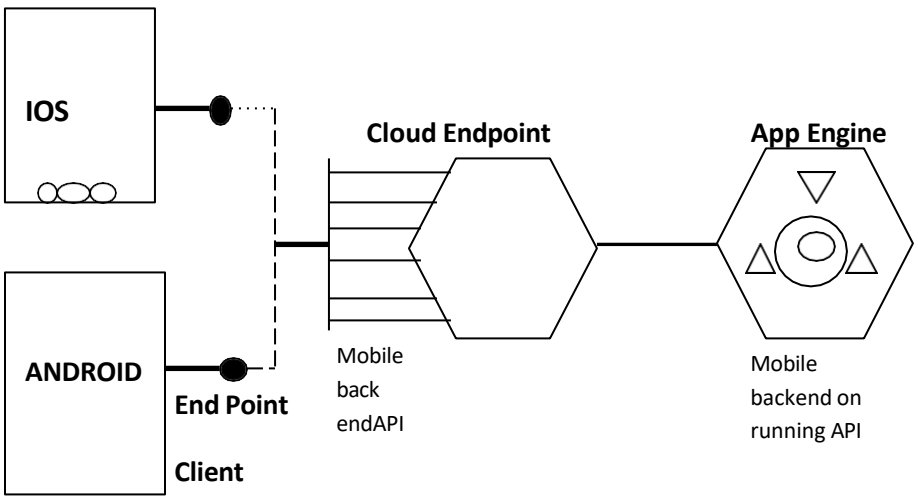

Figure: 5.1 Google App Engine Model

Microsoft Azure

Microsoft Azure is Microsoft's application platform for the public cloud. The goal of this article is to give you a foundation for understanding the fundamentals of Azure, even if you don't know anything about cloud computing.

- **The Components of Azure**

Azure groups services into categories in the Management Portal and on various visual aids like the What Is Azure Info graphic. The Management Portal is what you use to manage most (but not all) services in Azure.

This article will use a different organization to talk about services based on similar function, and to call out important sub-services that are part of larger ones.

- **Management Portal**

Azure has a web interface called the Management Portal that allows administrators to access and administer most, but not all Azure features. Microsoft typically releases the newer UI portal in beta before retiring an older one. The newer one is called the "Azure Preview Portal".

There is typically a long overlap when both portals are active. While core services will appear in both portals, not all functionality may be available in both. Newer services may show up in the newer portal first and older services and functionality may only exist in the older one. The message here is that if you don't find something in the older portal, check the newer one and vice-versa.

- **Compute**

One of the most basic things a cloud platform does is execute applications. Each of the Azure compute models has its own role to play.

You can use these technologies separately or combine them as needed to create the right foundation for your application. The approach you choose depends on what problems you're trying to solve.

The ability to create a virtual machine on demand, whether from a standard image or from one you supply, can be very useful. This approach, commonly known as Infrastructure as a Service (IaaS), is what Azure Virtual Machines provides.

To create a VM, you specify which VHD to use and the VM's size. You then pay for the time that the VM is running. You pay by the minute and only while it's running, though there is a minimal storage charge for keeping the VHD available. Azure offers a gallery of stock VHDs (called "images") that contain a bootable operating system to start from. These include Microsoft and partner options, such as Windows Server and Linux, SQL Server, Oracle and many more. You're free to create VHDs and images, and then upload them yourself. You can even upload VHDs that contain only data and then access them from your running VMs.

Wherever the VHD comes from, you can persistently store any changes made while a VM is running. The next time you create a VM from that VHD, things pick up where you left off. The VHDs that back the Virtual Machines are stored in Azure Storage blobs, which we talk about later. That means you get redundancy to ensure your VMs won’t disappear due to hardware and disk failures. It's also possible to copy the changed VHD out of Azure, then run it locally.

Your application runs within one or more Virtual Machines, depending on how you created it before or decide to create it from scratch now.

This quite general approach to cloud computing can be used to address many different problems

Hadoop Introduction

Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models. A Hadoop frame-worked application works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

Hadoop Architecture

Hadoop framework includes following four modules:

- Hadoop Common: These are Java libraries and utilities required by other Hadoop modules. These libraries provide filesystem and OS level abstractions and contains the necessary Java files and scripts required to start Hadoop.
- Hadoop YARN: This is a framework for job scheduling and cluster resource management.
- Hadoop Distributed File System (HDFS™): A distributed file system that provides high-throughput access to application data.
- Hadoop MapReduce: This is YARN-based system for parallel processing of large data sets.

We can use following diagram to depict these four components available in Hadoop framework figure 5.2.

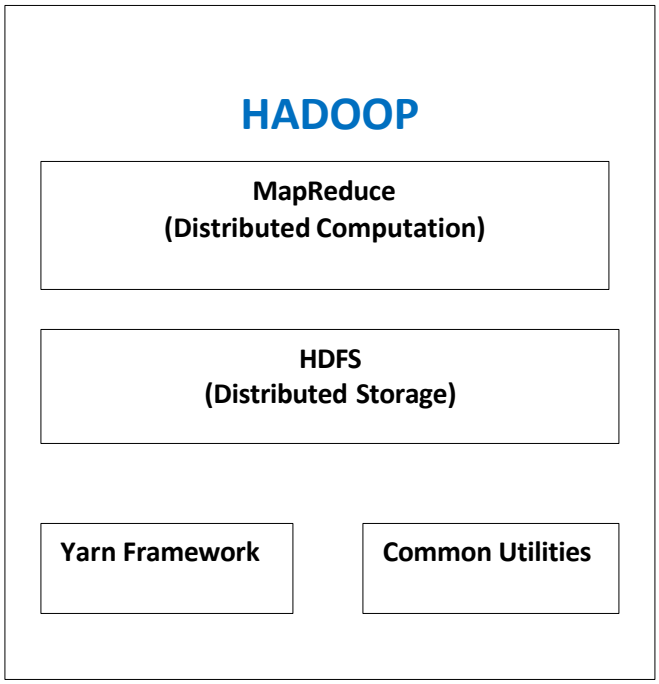

Figure: 5.2 Hadoop Model

Since 2012, the term "Hadoop" often refers not just to the base modules mentioned above but also to

the collection of additional software packages that can be installed on top of or alongside Hadoop, such as Apache Pig, Apache Hive, Apache HBase, Apache Spark etc.

MapReduce

Hadoop MapReduce is a software framework for easily writing applications which process big amounts of data in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner.

The term MapReduce actually refers to the following two different tasks that Hadoop programs perform:

- The Map Task: This is the first task, which takes input data and converts it into a set of data, where individual elements are broken down into tuples (key/value pairs).
- The Reduce Task: This task takes the output from a map task as input and combines those data tuples into a smaller set of tuples. The reduce task is always performed after the map task.

Typically both the input and the output are stored in a file-system. The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks.

The MapReduce framework consists of a single master JobTracker and one slave TaskTracker per cluster-node. The master is responsible for resource management, tracking resource consumption/availability and scheduling the jobs component tasks on the slaves, monitoring them and re-executing the failed tasks. The slaves TaskTracker execute the tasks as directed by the master and provide task-status information to the master periodically.

The JobTracker is a single point of failure for the Hadoop MapReduce service which means if JobTracker goes down, all running jobs are halted.

Hadoop Distributed File System

Hadoop can work directly with any mountable distributed file system such as Local FS, HFTP FS, S3 FS, and others, but the most common file system used by Hadoop is the Hadoop Distributed File System (HDFS).

The Hadoop Distributed File System (HDFS) is based on the Google File System (GFS) and provides a distributed file system that is designed to run on large clusters (thousands of computers) of small computer machines in a reliable, fault-tolerant manner.

HDFS uses a master/slave architecture where master consists of a single **NameNode** that manages the file system metadata and one or more slave **DataNodes** that store the actual data.

A file in an HDFS namespace is split into several blocks and those blocks are stored in a set of DataNodes. The NameNode determines the mapping of blocks to the DataNodes. The DataNodes takes care of read and write operation with the file system. They also take care of block creation, deletion

and replication based on instruction given by NameNode.

HDFS provides a shell like any other file system and a list of commands are available to interact with the file system. These shell commands will be covered in a separate chapter along with appropriate examples.

How Does Hadoop Work?

- **Stage 1**

A user/application can submit a job to the Hadoop (a hadoop job client) for required process by specifying the following items:

1. The location of the input and output files in the distributed file system.
2. The java classes in the form of jar file containing the implementation of map and reduce functions.
3. The job configuration by setting different parameters specific to the job.

- **Stage 2**

The Hadoop job client then submits the job (jar/executable etc) and configuration to the JobTracker which then assumes the responsibility of distributing the software/configuration to the slaves, scheduling tasks and monitoring them, providing status and diagnostic information to the job-client.

- **Stage 3**

The TaskTrackers on different nodes execute the task as per MapReduce implementation and output of the reduce function is stored into the output files on the file system.

Advantages of Hadoop

- Hadoop framework allows the user to quickly write and test distributed systems. It is efficient, and it automatic distributes the data and work across the machines and in turn, utilizes the underlying parallelism of the CPU cores.
- Hadoop does not rely on hardware to provide fault-tolerance and high availability (FTHA), rather Hadoop library itself has been designed to detect and handle failures at the application layer.
- Servers can be added or removed from the cluster dynamically and Hadoop continues to operate without interruption.
- Another big advantage of Hadoop is that apart from being open source, it is compatible on all the platforms since it is Java based.

Amazon EC2

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud. It is designed to make web-scale cloud 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 them from common failure scenarios.

EC2 Benefits

- **Elastic Web-Scale Computing**

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. Because this is all controlled with web service APIs, your application can automatically scale itself up and down depending on its needs.

- **Completely Controlled**

You have complete control of your instances including root access and the ability to interact with them as you would any machine. You can stop any instance while retaining the data on the boot partition, and then subsequently restart the same instance using web service APIs. Instances can be rebooted remotely using web service APIs, and you also have access to their console output.

- **Flexible Cloud Hosting Services**

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. For example, choice of operating systems includes numerous Linux distributions and Microsoft Windows Server.

- **Integrated**

Amazon EC2 is integrated with most AWS services such as Amazon Simple Storage Service (Amazon S3), Amazon Relational Database Service (Amazon RDS), and Amazon Virtual Private Cloud (Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

- **Secure**

Cloud security at AWS is the highest priority. As an AWS customer, you will benefit from a data center and network architecture built to meet the requirements of the most security-sensitive organizations. Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality for your computing resources.

- **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 data

centers. The Amazon EC2 Service Level Agreement commitment is 99.95% availability for each Amazon EC2 Region.

Aneka

Aneka is a platform and a framework for developing distributed applications on the Cloud. It harnesses the spare CPU cycles of a heterogeneous network of desktop PCs and servers or datacenters on demand. Aneka provides developers with a rich set of APIs for transparently exploiting such resources and expressing the business logic of applications by using the preferred programming abstractions. System administrators can leverage on a collection of tools to monitor and control the deployed infrastructure. This can be a public cloud available to anyone through the Internet, or a private cloud constituted by a set of nodes with restricted access figure 5.3.

The Aneka based computing cloud is a collection of physical and virtualized resources connected through a network, which are either the Internet or a private intranet. Each of these resources hosts an instance of the Aneka Container representing the runtime environment where the distributed applications are executed. The container provides the basic management features of the single node and leverages all the other operations on the services that it is hosting. The services are broken up into fabric, foundation, and execution services. Fabric services directly interact with the node through the Platform Abstraction Layer (PAL) and perform hardware profiling and dynamic resource provisioning. Foundation services identify the core system of the Aneka middleware, providing a set of basic features to enable Aneka containers to perform specialized and specific sets of tasks. Execution services directly deal with the scheduling and execution of applications in the Cloud.

One of the key features of Aneka is the ability of providing different ways for expressing distributed applications by offering different programming models; execution services are mostly concerned with providing the middleware with an implementation for these models. Additional services such as persistence and security are transversal to the entire stack of services that are hosted by the Container. At the application level, a set of different components and tools are provided to: 1) simplify the development of applications (SDK); 2) porting existing applications to the Cloud; and 3) monitoring and managing the Aneka Cloud.

A common deployment of Aneka is presented at the side. An Aneka based Cloud is constituted by a set of interconnected resources that are dynamically modified according to the user needs by using resource virtualization or by harnessing the spare CPU cycles of desktop machines. If the deployment identifies a private Cloud all the resources are in house, for example within the enterprise. This deployment is extended by adding publicly available resources on demand or by interacting with other Aneka public clouds providing computing resources connected over the Internet.

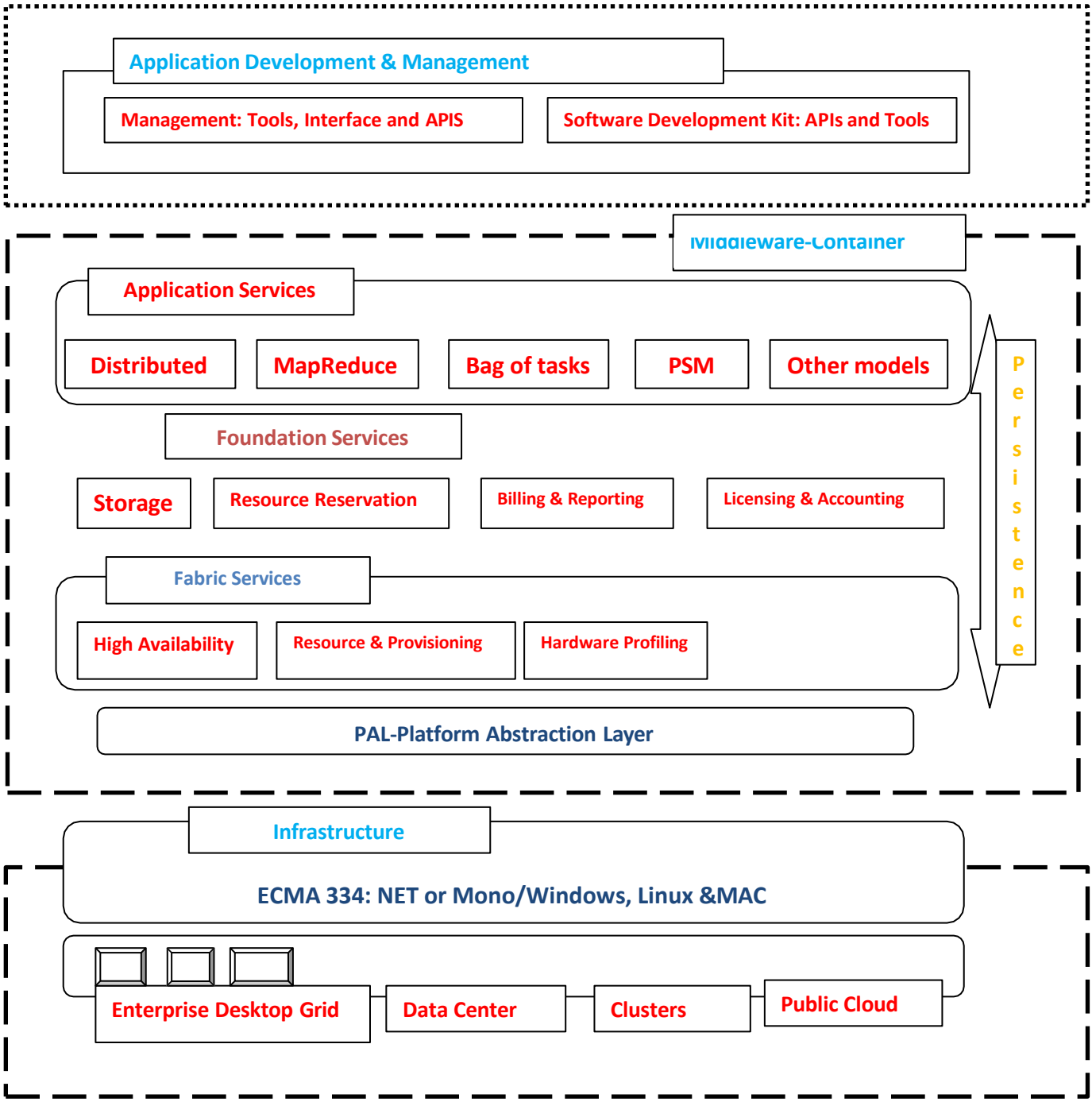

Figure: 5.3 Aneka Model

Web Resources:

- https://aws.amazon.com/security/introduction-to-cloud-security/
- https://www.cloudera.com/
- https://owncloud.org/
- https://www.fortinet.com/solutions/enterprise-midsize-business/cloud-security.html
- https://www.solutionary.com/managed-security-services/cloud-security/
- https://www.tibco.com/blog/2012/12/20/federated-cloud/