Study Material

On

Cloud Computing

Department of Computer Science & Engineering

CAPITAL ENGINEERING COLLEGE

Mahatapalla, Khordha, Bhubaneswar, Odisha: 752060

(Affiliated to Biju Patnaik University of Technology, Odisha and SCTE & VT, Odisha, Approved by AICTE, New Delhi and Recognised by Govt. of Odisha)

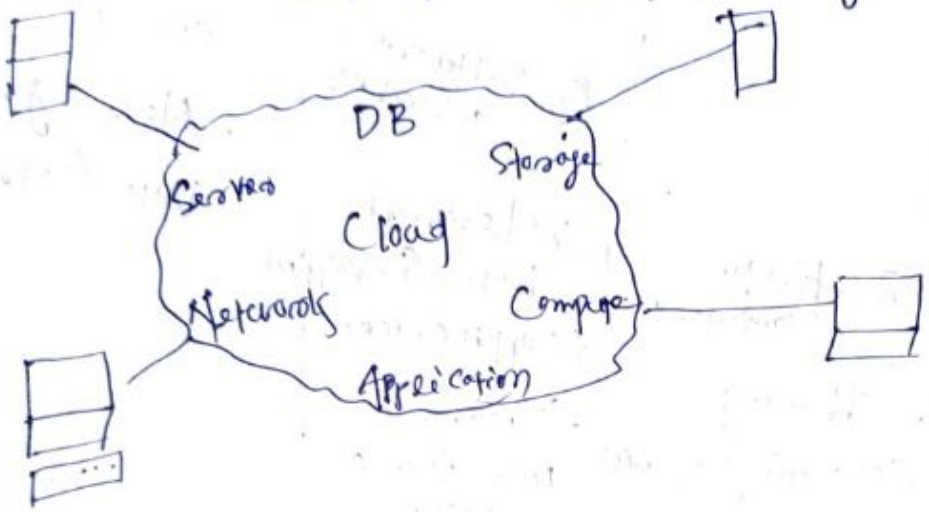
Chapter-1 (Introduction to Cloud Computing)

Cloud Computing

Cloud + Computing

↓ Internet

↓ Data processing or Data Storage



Defn

it is the delivery of on-demand computing services over the internet on a pay as you go basis.

Example

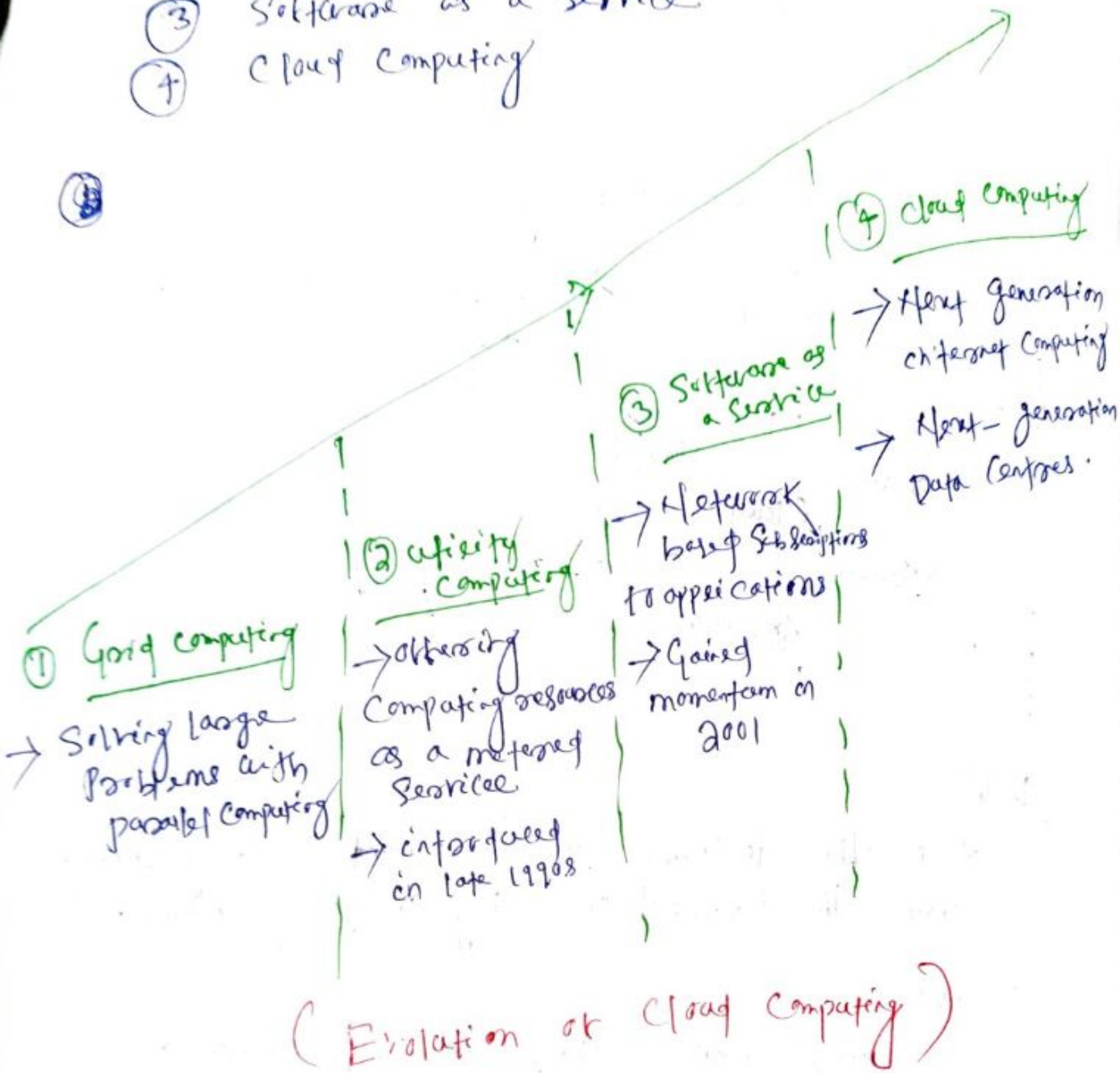
Gmail, Google Drive, YouTube, amazon drive

1.1 Historical development / Evolution of Cloud Computing

Cloud Computing is not latest technology. Cloud Computing has evolved (developed gradually) through a number of phases which includes

1. Grid computing
2. utility computing

(3) Software as a Service
(4) Cloud Computing



(Evolution of Cloud Computing)

1.2 Vision of Cloud Computing

In simplest terms, Cloud computing means storing and accessing the data and programs on remote servers that are hosted on the internet instead of computer's hard drive or local server. Cloud computing is also referred as internet based computing

These are following vision of Cloud Computing

1. Cloud computing provides the facility to provision virtual hardware, runtime environment and services to a person having money.
2. These all things can be used as long as they are needed by the user.
3. The whole collection of computing system is transformed into collection of utilities, which can be provisioned and composed together to deploy systems in ~~hours~~ hours rather than days, with no maintenance cost.
4. The long term vision of Cloud computing is that IT services are traded as utilities in an open market without technological and legal barriers.
5. A cloud provider can also become a consumer of a competition services in order to fulfil its promises to customers
6. A cloud provider can also be a buyer of a competitive service to fulfill its promises to customers.

1.3 Characteristics of Cloud Computing

(1) On demand Self-Service : Computing capabilities such as network storage can be set-up whenever required without requiring any human interaction.

(2) Broad network access : All capabilities are available over a network can be accessible from anywhere by means of any client platforms (e.g mobile phones, tablets, laptops and workstations)

(3) Resource pooling : Cloud Service provider pools his computers resources to multiple consumers with different Physical resources dynamically assigned according to Consumer demand (e.g Storage, processing, memory and network bandwidth)

(4) Rapid elasticity : Capabilities can be elastically set-up and scale rapidly according to consumer demand.

(5) Measured Service : Resource usage can be monitored, controlled and reported, providing transparency for both the provider and Consumer.

(6) No maintenance / easy maintenance

(7) Security : Copy of our data on various Services. ek 1 fails, data is safe on the other.

1.4 Cloud Computing Reference Model

Three types of cloud computing reference models are

1. IaaS (Infrastructure as a Service)
2. Paas (platform as a Service)
3. Saas (Software as a Service)

(1) IaaS (Infrastructure as a Service)

→ Virtualized Servers
→ Storage and networking

Examples

Amazon EC2, S3, Rightscale, VCloud

(2) (platform as a Service) → Paas

→ Runtime environment for applications.
→ Development and data processing Platforms

Examples: Windows Azure, Hadoop, Google AppEngine

(3) Software as a Service - Saas

→ End user applications
→ Scientific applications
→ Office Automation, Photo editing
→ Social networking

Examples

Google Documents, Facebook, Flickr.

1.5 Cloud Computing Environment

1. Application development
2. Infrastructure and System development
3. Computing platforms and technologies

(1) Application development

Applications that leverage cloud computing benefit from its capability to dynamically scale on demand

Example: Web applications.

Scientific applications can require huge computing capacity to perform large-scale experiments once in a while. So it is not feasible to buy the infrastructure supporting them. In this case, Cloud computing can be the solution

(2) Infrastructure and System development

→ Infrastructure as a Service Solutions provide the capabilities to add and remove resources.

→ Platform as a Service Solutions control the provisioning process and lease of resources. These can be scaled completely transparent to developers or subject to fine control.

(3) Computing platforms and technologies

→ **Amazon Web Services (AWS)** – AWS is a cloud computing platform that provides customers with a wide array of cloud services.

→ **Google AppEngine** : it is a platform as a service and cloud computing platform for developing and hosting web applications in Google-managed data centers.

→ **Microsoft Azure** : it is microsoft's public cloud computing platform. it provides a range of cloud services, including those for compute, analytics, storage and networking

→ **Hadoop** : Hadoop is a java-based framework used to manipulate data in the cloud or on premises. Hadoop can be installed on cloud servers to manage big data where as cloud alone can not manage data without hadoop in it.

1.6 Cloud Services Requirements

1. Efficiency / cost reduction : By using cloud infrastructure, you don't have spend huge amounts of money on purchasing and maintaining equipment.

2. Data Security : Cloud offers many advanced security features that guarantee that data is securely stored and handled. Cloud storage providers implement baseline protections for their platforms and the data they process, such as authentication, access control and encryption.

3. Scalability : Different companies have different IT needs - a large enterprise of 1000+ employees won't have the same IT requirements as a start-up. Using cloud is a great solution because it enables enterprise to efficiently and quickly scale up/down according to business demands.

4. Mobility: Cloud computing allows mobile access to corporate data via Smartphones and devices.

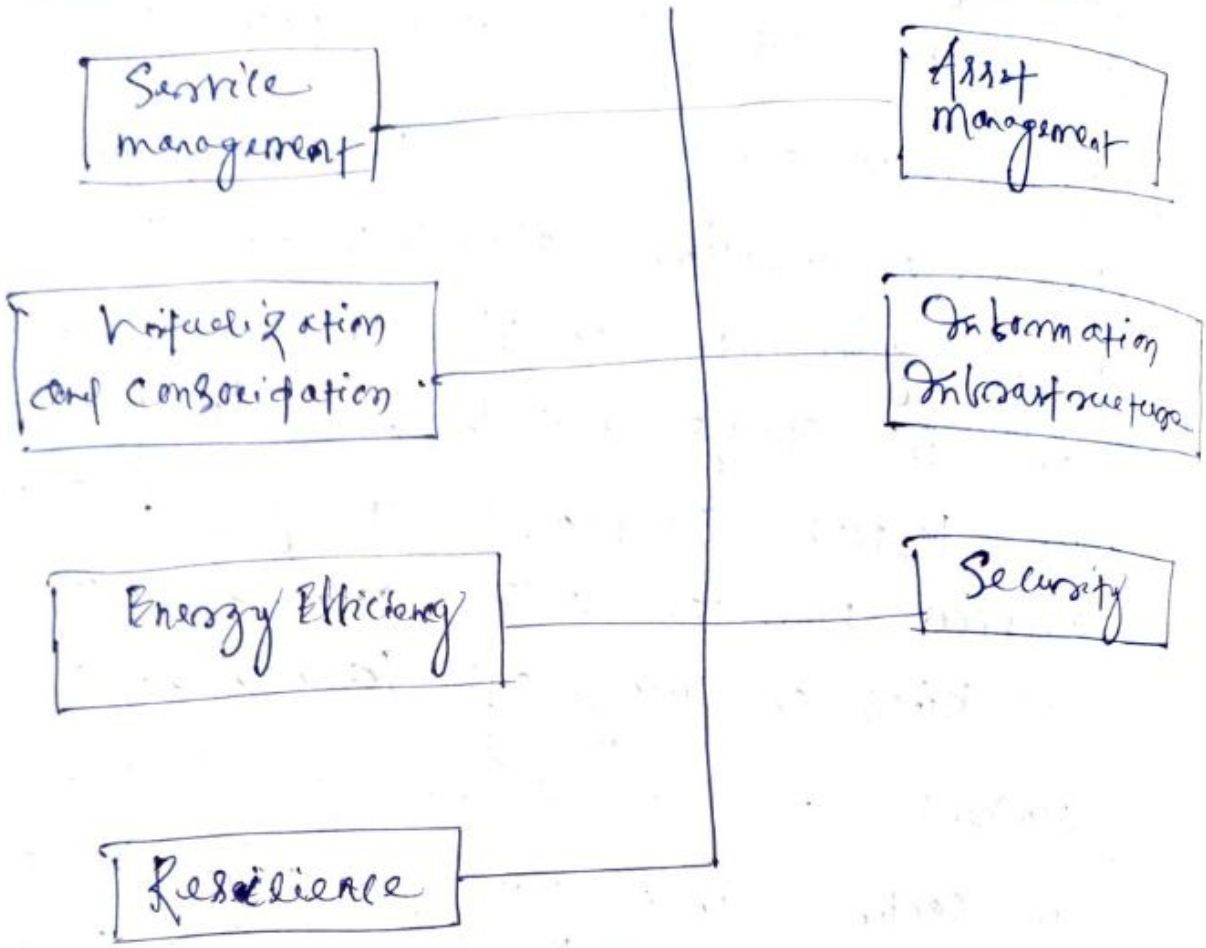
5. Disaster recovery: Data loss is major concern for all organizations along with data Security. Storing your data in the cloud guarantees that data is always available, even if your equipment like laptops or PCs, is damaged. Cloud-based services provide quick data recovery for all kinds of emergency scenarios.

6. Control: Cloud enables you complete visibility and control over your data. You can easily decide which users have what level of access to what data.

7. Market reach: Developing in the Cloud enables users to get their applications to market quickly.

8. Automatic Software updates: Cloud based applications automatically refresh and update themselves.

1.7 Cloud and dynamic Infrastructure



(Cloud and dynamic Infrastructure)

1. Service management: This type of Specific facility or functionality provided to Cloud IT Services by the cloud Service providers. This facility includes visibility, automation and Control to overseeing the first Cloud IT Services.

2. Asset management: In this assets or property which is involved in providing the cloud services are getting managed.

3. Virtualization and Consolidation: Consolidation is an effort to reduce the cost of a technology by improving its operating efficiency & effectiveness. it means migrating from large number of resources to fewer one, which is done by virtualization technology.

4. Information Infrastructure: it helps the business organizations to achieve the following information compliance, availability of resources retention and Security objectives.

5. Energy - Efficiency: Here the IT Infrastructure or organization sustainable. it means it is not likely to damage or effect any other thing

6. Security: This cloud infrastructure is responsible for risk management. Risk management refers to the risks involved in the Services which are being provided by the Cloud - Service providers.

7. Resilience: This infrastructure provides the feature of resilience means the services are resilient. it means the infrastructure is safe from all sides. The IT operations will not be easily get affected.

1.8 Cloud adoption

Cloud adoption is a strategy used by enterprise to improve the scalability of internet-based database capabilities while reducing cost and risk.

Who needs Cloud Adoption - and why?

A variety of industries benefit from cloud adoption, including healthcare, marketing and advertising, retail, finance, and education.

Benefits include:

1. Healthcare: Fueled by digital and social consumer behaviors and the need for secure and accessible electronic health records (EHRs), hospitals, clinics, and other

2. Marketing and Advertising: In an industry dependent on Social media, as well as the quick creation and publishing of Customer-relevant content, agencies are using hybrid cloud adoption strategies to deliver critical client messages to their local and around-the-world audiences.

3. Retail: A Successful e-Commerce Strategy requires a Sound Internet Strategy: with the help of Cloud adoption, internet based retail is able to effectively market to Customers and Save their Product data for less money.

4. Finance: Efficient expense management, human resources and Customer Communications are three of the most important business needs of today's finance organizations. For these reasons, financial Services institutions are now placing their e-mail platforms and marketing tools in the Cloud.

5. Education: Internet based education opportunities are now more popular than ever. The Cloud allows universities, private institutions, and K-12 public Schools to provide learning, homework, and grading systems online

1.9 Cloud Computing Applications

Cloud Computing is applied in almost all the fields like

- Business
- Entertainment
- data Storage
- Social networking
- education
- management

Some of the popular applications of cloud Computing are

(1) **Business Applications** : Cloud Computing Comprises more Collaborative and easy business with the help of different apps like MailChimp, Chatter, Google Apps for business and QuickBooks

* **MailChimp** – it provides one e-mail publishing platform.
  - it is a simple email marketing System
  - it provides a various options to design, Send and Save templates for emails.

* Chatter : This app helps to share important information about organization in real time.

* Google Apps for Business : google provides creating text documents, Spreadsheets, Presentations etc
— on Google Docs. it allows the business users to share them on a combined way.

* Quickbooks : it provides online accounting Solutions for business.
— it assists in monitoring cash flow, creating VAT returns and creating business report.

3 Data Storage and Backup Service Applications

* Box.com : it provides drag & drop Service for files. it is necessary for the users to drop the files into Box and access from anywhere

* mozy : it provides online backup service for files to prevent data loss

* Joukuu : it is a web base interface it helps to share a Single Unit of Contents for files stored in Google Docs, Box.net and Dropbox

(3) Management Applications :

* Toggl : it helps to track time spent working on a particular project

* Evernote : it is designed to create, organize and store different pieces of media. it keeps all stuff like text documents, photos, videos or even webpages in the Cloud.

* Outright : it is an accounting app that helps to track income, expenses, profit and losses in real time.

(4) Art Applications

* Moo : it provides art services like designing and printing business cards, post cards and mini cards.

(5) Entertainment Applications :

* Audio box.fm : it provides Streaming Service The music files are stored online and play from the cloud using own media player or Service.

(b) Social Applications :

* Facebook : it provides social networking services on facebook can share photos, videos, files, status and more.

* Twitter : it helps in interacting users from public directly. in this, users can follow any organization, celebrity or any person who is on twitter and can have latest updates regarding them.

Exercise

Important Questions (Short questions)

1. What is cloud computing
2. Define Grid Computing
3. Define utility computing
4. Define Iaas
5. Define Paas
6. Define Saas
7. What are the applications are used by cloud computing
8. Define cloud Adoption.

Important Questions (Long Questions)

1. Describe about the Historical development / evolution of cloud computing
2. Describe about the Characteristics of cloud computing.

(3) Describe about the Cloud Computing reference model.

(4) Describe about the Cloud Computing environment

(5) Describe about the Cloud Service requirements.

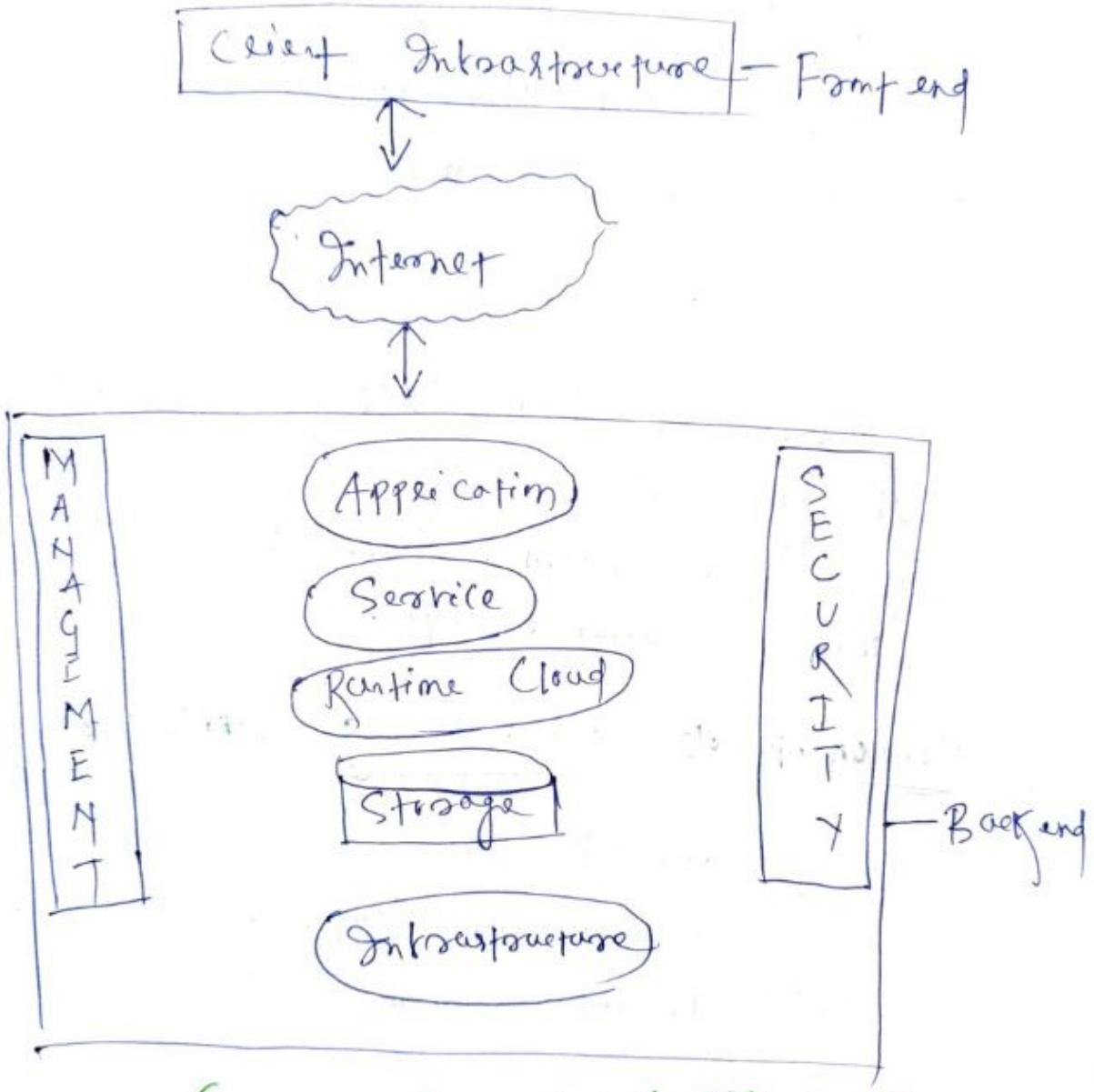
(6) Explain about the Cloud and Dynamic infrastructure.

(7) Describe about the Cloud applications.

X The End X

Chapter-2 (Cloud Computing Architecture)

2.1 Introduction



(Cloud Computing Architecture)

The cloud computing architecture comprises of many loosely coupled cloud components. Each of them the cloud architecture into two parts.

1. Front end
2. Back end

(1) Front end:

refers to the client part of cloud computing system. it consists of interfaces & applications that are required to access the cloud platform

Example: Web Browser

(3) Back end:

refers to cloud itself. it comprises of huge data storage, virtual machine, security mechanism, services, deployment models, servers etc.

Components of Cloud Computing Architecture

(1) Client Infrastructure:

it is a front end component (provides GUI to interact with cloud)

(2) Application: it may be any software or platform that a client wants to access.

(3) Services:

it manages that which type of services you access according to the client's requirement.

Cloud computing offers
SaaS, PaaS, IaaS.

(4) Runtime Cloud :
it provides "execution & runtime environment" to the virtual machines.

(5) Storage : one of the most important components it provides a huge amount of Storage Capacity in the Cloud to Store & manage data.

(6) Infrastructure : Cloud infrastructure includes hardware & Software components such as Servers, Storage, network devices & other resources needed for cloud computing model.

(7) Management : Manages components (like application, Service, infrastructure)

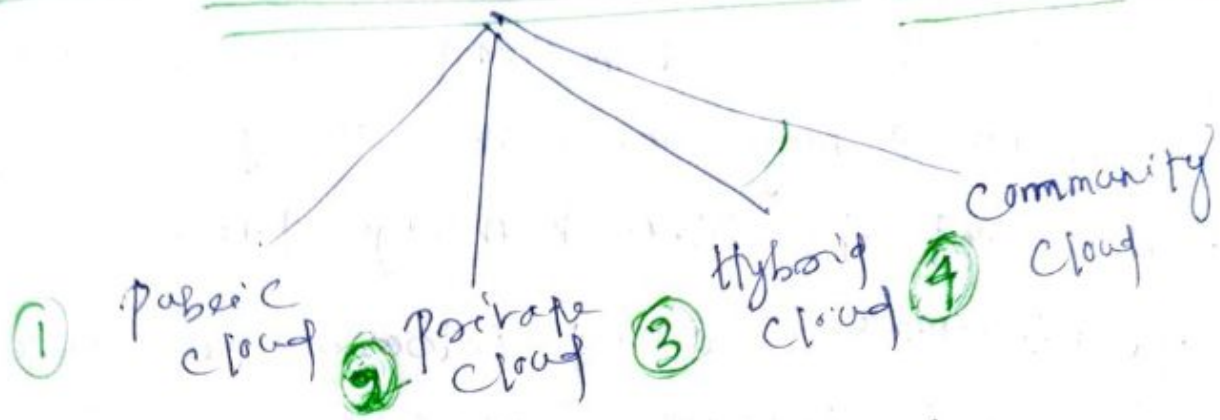
(8) Security : inbuilt backend component provides Security mechanism in the backend.

(9) Internet : Medium through which frontend & backend interacts.

2.2 Cloud Reference model

Same as 1.4

2.3 Types of clouds / Cloud Deployment model

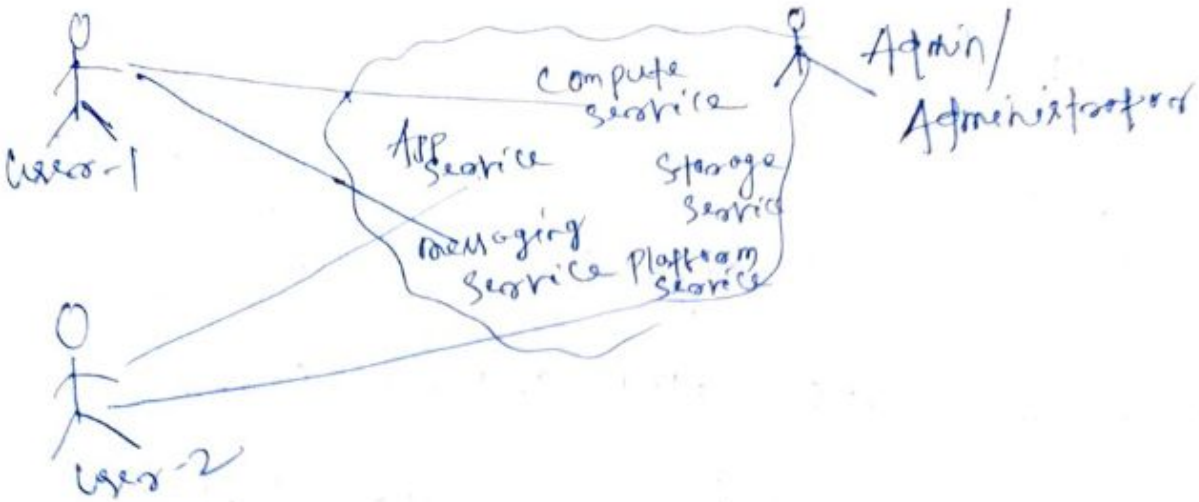


(1) Public cloud

- → open to all to store & access information via internet
- → pay as per use (for the services)
- → Managed by third parties (cloud service providers)
- → Fundamental characteristics of public cloud is MULTITENANCY

Example

EC2 (Amazon elastic Compute cloud), dropbox, Google drive



Advantages

→ it is maintained by Cloud Service Provider, So we need not maintain it.

→ Location independent because its Services are delivered through the internet.

→ high Scalability

example gmail offers 15GB. We can increase anytime & decrease also after changing

→ Cost effective and pay as per use.

Disadvantages

→ less Secure because resources are shared publically.

→ less Customizable as Compared to private Cloud.

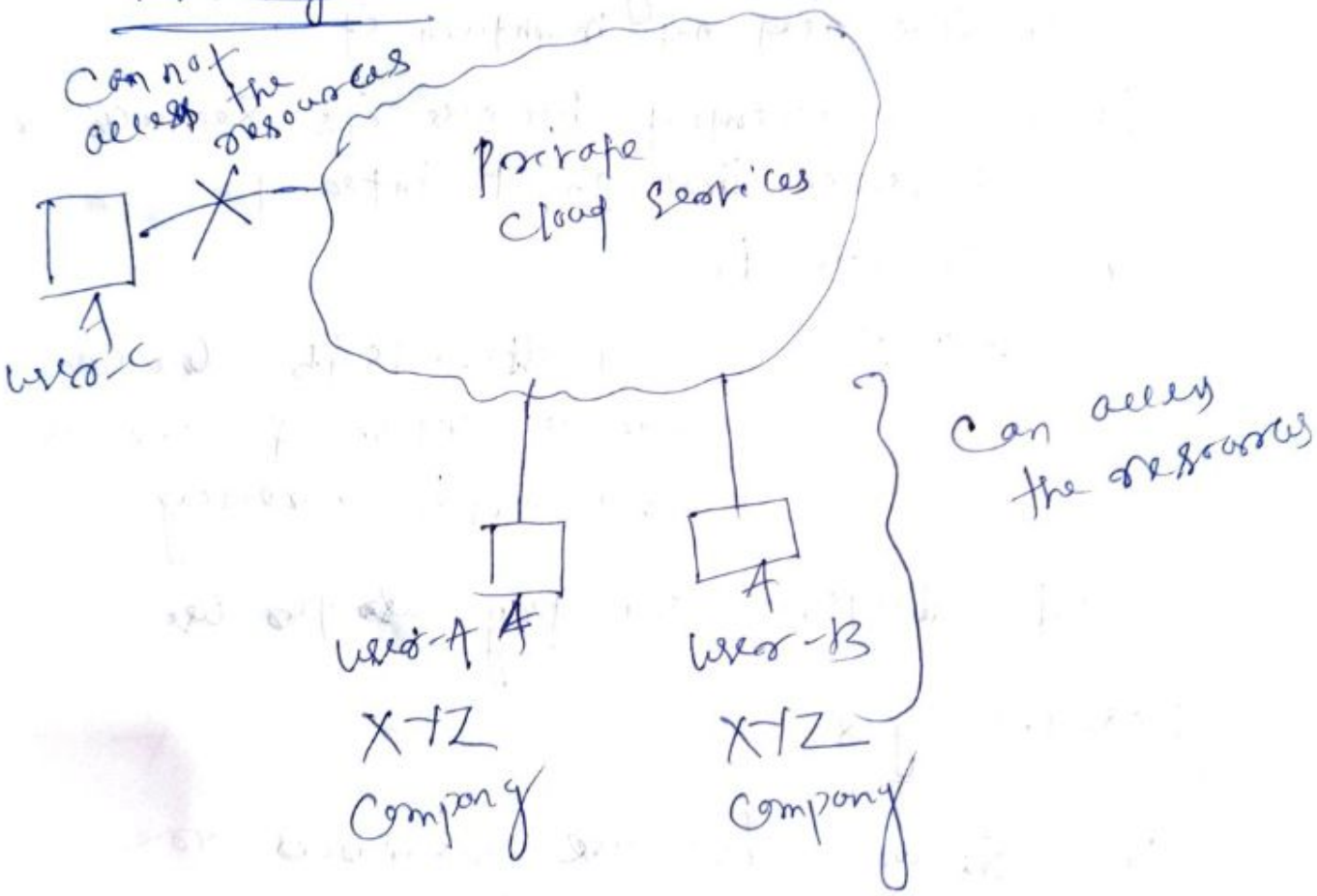
(2) Private Cloud

→ Services accessible within an organization i.e it belongs to a specific organization

→ Sometimes also called enterprise / Corporate cloud

→ Can be managed by organization, 3rd party etc.

~~Advantages~~



Advantages

(i) High Security : In private Cloud Security Concerns are high Since Customers data & other sensitive information flows through & stored in a private infrastructure.

(ii) data privacy : only authorized people can access the data

(iii) more Customizable : as Companies get to Customize their Solution as per requirement

(iv) improved reliability

Disadvantages

→ private Cloud is accessible within an organization, So, the area of operations is limited.

→ High Cost → We need to invest in Hardware & Software

→ Limited Scalability

(3) Hybrid cloud

→ Combination of public & private cloud
→ Critical activities performed by private cloud
→ non-critical activities performed by public cloud

Advantages

→ Scalability, Security, low cost (as compared to private cloud)

→ flexibility

Disadvantages

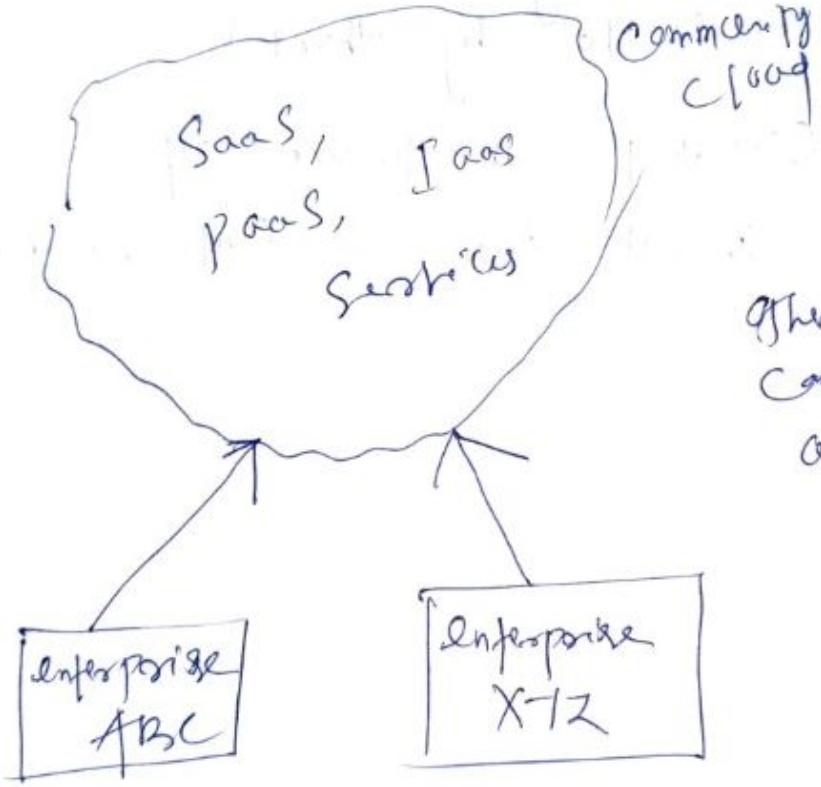
→ Managing is difficult/complex because there are more than one type of deployment model

→ dependency on infrastructure

(4) Community Cloud

→ allows services to be accessible by a group of several organizations to share the information between the organization & a specific community.

→ owned, managed & separated by one or more organizations in the community or 3rd party.



Advantages

(i) Cost reduction / cost effective

→ it is cheaper than private cloud (multiple companies share the bill, which lowers the cost.

(i) Sharing among companies (the resources)

(ii) More secure than public cloud but less than private cloud.

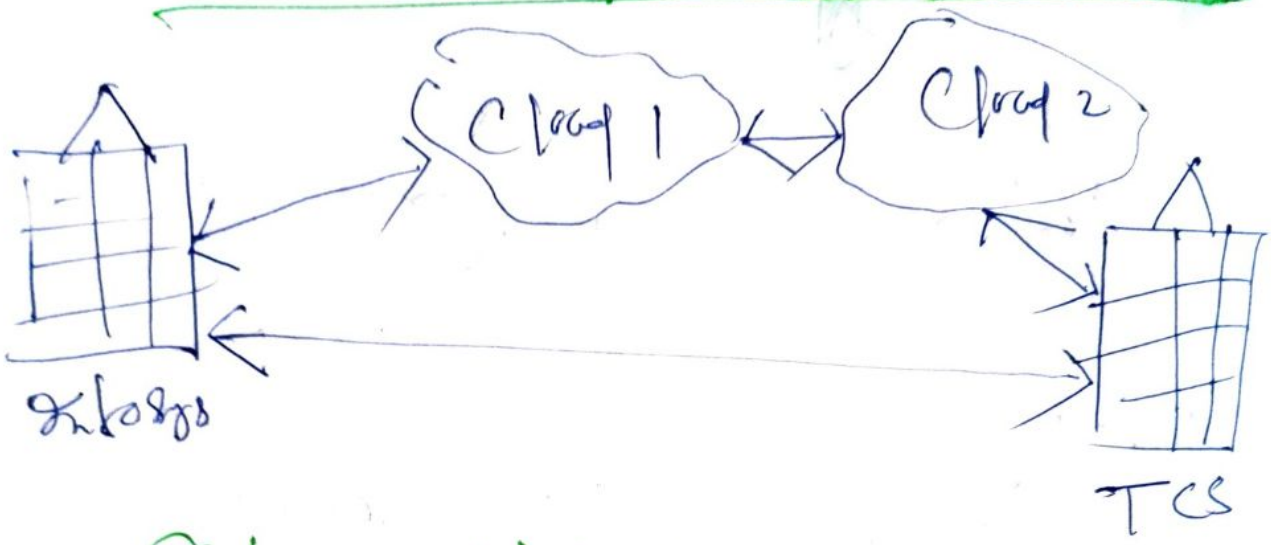
Disadvantages:

(i) data is accessible between organizations (because the data is stored at the same location any data stored there might be accessible by others)

(ii) Consistent maintenance cost

(iii) Overall increase cost than public cloud.

2.4 Cloud Interoperability and Standards



Interoperability

→ The ability of two or more systems, applications or components to exchange and use

information.

→ The ability of Systems to provide and receive Services from other Systems and to use the Services so interchanged to enable them to operate effectively together.

→ Interoperability is an enabler for interchange ability (Replacement of one element with another)

→ ~~Interability~~ Interoperability is the goal of Standards but Standards don't guarantee interoperability.

2.5 Cloud computing Interoperability use cases

→ Users of one cloud accessing Storage in another cloud (to provide elastic Storage)

→ Applications and Services running on (and communicating between) heterogeneous cloud platforms.

→ Application using resources (CPU, Storage) in another heterogeneous cloud platform (resource sharing)

→ Resource sharing across different time zones.

→ Demonstration of data portability (across Service Providers)

→ What is needed to transfer a running STATEFUL Service from Cloud provider A to B.

- Moving a file Sharing service between Cloud providers.
- moving a Streaming Service between Cloud providers

26 Role of Standards in Cloud computing environment

Different Standards are used in Cloud computing environment

1. Standards for Application Developers
2. Standards for messaging
3. Standards for Security

(1) Standards for Application Developers

- Browsers (Ajax)
- Data (XML, JSON)
- Solution Stacks (LAMP and LAPP)

Browsers (Ajax)

- Ajax is a technique, not programming language
- When we used ajax in website there is no need to refresh page.
- Small code

XML

- XML stands for Extensible Markup Language
- XML was designed to store & transport data
- XML was designed to be both human- and machine readable.

JSON

→ Javascript Object Notation is a lightweight data interchange format
→ it is easy for humans to read and write.
→ it is easy for machines to parse and generate.

Solution Stacks (LAMP and LAPP)

L - Linux
A - Apache
M - MySQL
P - Php or Python.

→ LAMP is a popular open source solution commonly used to run dynamic websites and servers.

(2) Standards for messaging

- Simple Message Transfer protocol (SMTP)
- Post office protocol (POP)
- Internet messaging Access Protocol (IMAP)
- Simple Object Access protocol (SOAP)

Simple Message Transfer protocol. (SMTP)

→ SMTP is the standard protocol for e-mail services on TCP/IP network.

→ SMTP provides the ability to send and receive email messages.

Post office protocol (POP)

→ POP is an application layer protocol in the OSI model that provides end users the ability to fetch and receive email

Internet Messaging Access protocol (IMAP)

→ IMAP is a standard protocol for accessing email on remote on a remote server from a local client

Simple Object Access Protocol (SOAP)

SOAP is a Protocol Specification for exchanging structured information in the implementation of Web Services in computer networks.

(3) Standards for Security

- Security Assertion Markup Language (SAML)
- Open Authentication (OAuth)
- SSL / TLS

Security Assertion Markup Language (SAML)

→ SAML is a language protocol for handling authentication and authorization in a network.

→ it is one of various XML based markup languages available to help with aspects of web development and use.

Open Authentication (OAuth)

OAuth is an open standard authorization protocol or framework that describes how unrelated servers & services can safely allow authenticated access

SSL/TLS

Secure Sockets Layer (SSL) and Transport Layer Security (TLS) both cryptographic protocols used to increase security by encrypting communication over computer networks.

Exercise

Important Questions (Short Questions)

1. Define Cloud computing Architecture
2. What are the Components are used in Cloud Computing Architecture.
3. Define public cloud
4. Define Private cloud.
5. Define Hybrid Cloud
6. Define Community cloud
7. What do you mean by Cloud Interoperability
8. Define SMTP, POP, IMAP, SOAP SAML, SSL/TLS.

Important Questions (Long Questions)

1. Describe about the Cloud reference model.
2. Explain about the types of Clouds.
3. Describe about the Cloud Interoperability and Standards & use cases.
4. Explain about the role of Standards in Cloud computing environment.

Chapter-3

3.0 (Scalability and Fault Tolerance)

3.1 Introduction / 3.2 Scalability and Fault Tolerance

Cloud Scalability is the ability to scale on demand the facilities and services as and when they are required by user.

Terms Related to Scalability

1. SCALE-UP – Increasing / Adding resources in existing server
2. SCALE-DOWN – Taking out resources that are added in existing server
3. SCALE OUT – Adding New / extra servers in the Cluster.
4. SCALE IN – Taking out Added servers from Cluster

Advantages of Scalability

1. More storage
2. More power
3. More scalability
4. Less time to create
5. Cost Savings.

Cloud Fault Tolerance is tolerating the faults by the cloud that are done by mistake by the user.

→ Fault tolerance refers to the ability of a system (computers, network, cloud clusters, etc) to continue operating without interruption when one or more of its components fail.

Metrics for fault Tolerance in Cloud computing

→ The existing fault tolerance technique in cloud computing considers various parameters

- Throughput
- response-time
- Scalability

- performance
- availability
- usability
- reliability
- Security and associated overhead

Types of Fault tolerance

(1) Reactive fault tolerance:

it techniques are used to reduce the impact of failures on a system when the failures have actually occured. Techniques based on this policy are checkpoint/restart and retry and so on.

- Check pointing/restart - The failed task is restarted from the recent checkpoint rather than from the beginning. it is an efficient technique for large application.

- **Replication** – In order to make the execution succeed, various replicas of task are run on different resources until the whole replicated task is not crashed; Hadoop and AmazonEC2 are used for implementing replication.
- **Job Migration**: On the occurrence of failure, the job is migrated to a new machine. HAProxy can be used for migrating job to other machines.
- **Retry**: This task level technique is simplest among all. The user resubmits the task on the same cloud resource.
- **Task Resubmission**: The failed task is submitting again either to the same machine on which it was operating or to some other machine.

(2) **Proactive Fault Tolerance**

Proactive fault tolerance predicts the faults proactively and replace the suspected components by other working components thus avoiding recovery from faults & errors.

- Software Rejuv - the System is planned for periodic reboots and every time the System starts with a new state.
- Self-healing - failure of an instance of an application running on multiple Virtual machines is controlled automatically.
- preemptive Migration - in this technique an application is constantly observed & analyzed. preemptive migration of a task depends upon feed-back-loop control mechanism.

3.4 Cloud Solutions

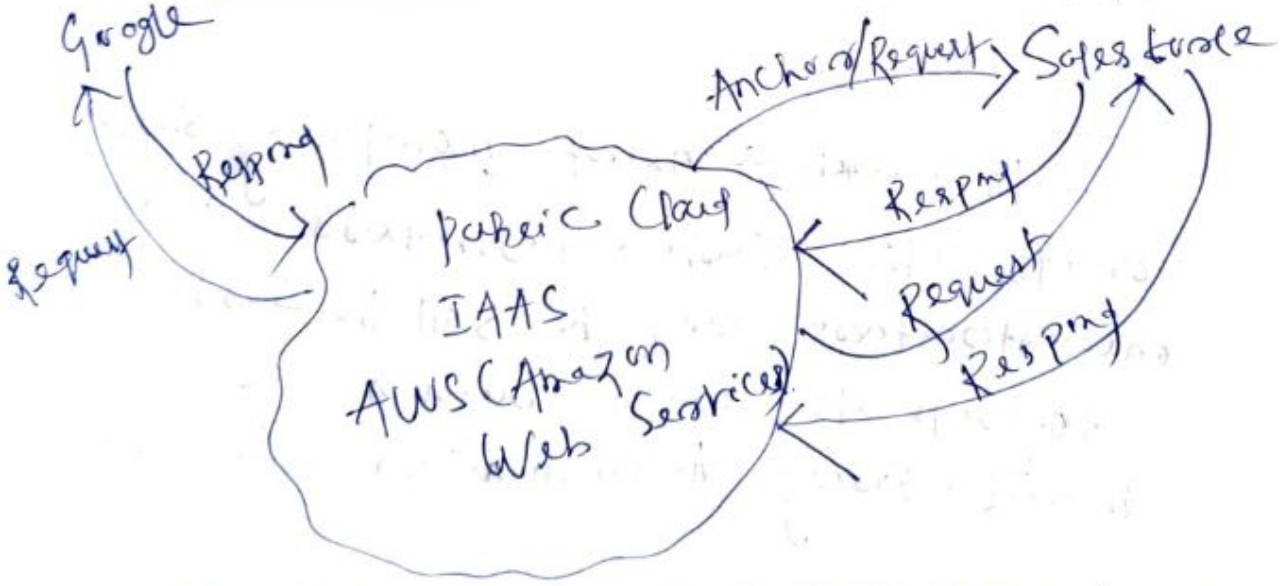
→ A cloud based solution refers to applications, storage, on-demand services, computer network or other resources that are accessed with an internet connection through another provider's shared cloud computing framework.

→ The simplest way to think of Cloud Computing is by comparing it to electricity. Your home and business have it. but you don't need a power plant on your property to use it. You just connect to the one that provides electricity to your area.

3.5 Cloud ecosystem

→ Cloud ecosystem is a term used to describe the complex system of interdependent components that work together to enable Cloud Services.

How Cloud ecosystem works?



→ The center of cloud ecosystem is public cloud service provider. it might be an IaaS provider such as Amazon Web Services (AWS) or a SaaS vendor such as Salesforce.

→ AWS is the center of its ecosystem, but it is also a part of the Salesforce ecosystem. Salesforce runs a number of its services on AWS's infrastructure.

→ AWS is also a part of the google ecosystem. google runs a number of its services on AWS's infrastructure.

Benefits of a cloud ecosystem

→ Companies can use a cloud ecosystem to build new business models. it becomes relatively easy for medical device Manufacturers

For Example

to launch a heart-monitoring service on its cloud service providers cloud infrastructure and then sell the service alongside its main business of manufacturing heart monitors for Hospitals

→ in cloud ecosystem, it is also easier to aggregate data and analyze how each part of the system affects the other parts

For example

It can ecosystem consists of patient records, Smart device logs and healthcare provider records, it becomes possible to analyze patterns across an entire patient populations.

3.5 Cloud Business Process Management

→ Business process Management (BPM) is a mature business discipline that has spawned a number of technologies to support it.

→ Today it is the agile who survive those organizations who are able to adapt to change, to innovate as well as continously improve, and to continously monitor & analyze the results of these adaptations

→ In the current web enabled business environment processes in many cases depend on the discovery and orchestration of components that exist as web services

→ The current trend is towards emphasis on mobility and collaboration as essential elements to support the agility and ~~current~~ currency of business processes.

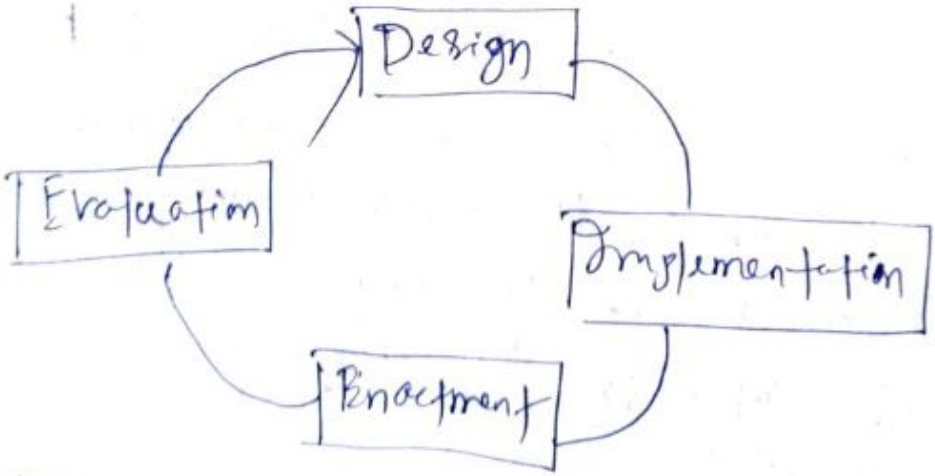
→ This means that BPM vendors are increasingly seeking to augment their BPM packages by incorporating Web 2.0 type functionality.

→ Cloud based BPM is one response to these new demands.

→ BPM governs organizations cross functional, customer focused end to end core business process.

BPM Lifecycle

The BPM lifecycle is an iterative process in which all of the BPM aspects are covered.



① Design:

The design phase consists of identifying existing process and capturing the business processes in processes models.

(2) Implementation

In the implementation phase, the designed process is implemented in an executable process language, which can be deployed in a BPMS.

(3) Enactment

The enactment phase is the runtime phase of the life cycle. The business process is deployed and monitored by a BPMS.

(4) Evaluation

In the evaluation phase the monitored information that is collected by the BPMS is used to review the business process. The conclusions drawn in the evaluation phase are input for the next iteration of the lifecycle.

3.7 Portability and Interoperability

Cloud Portability

→ Cloud portability is the ability to move applications and data from one cloud computing environment to another with minimal disruption

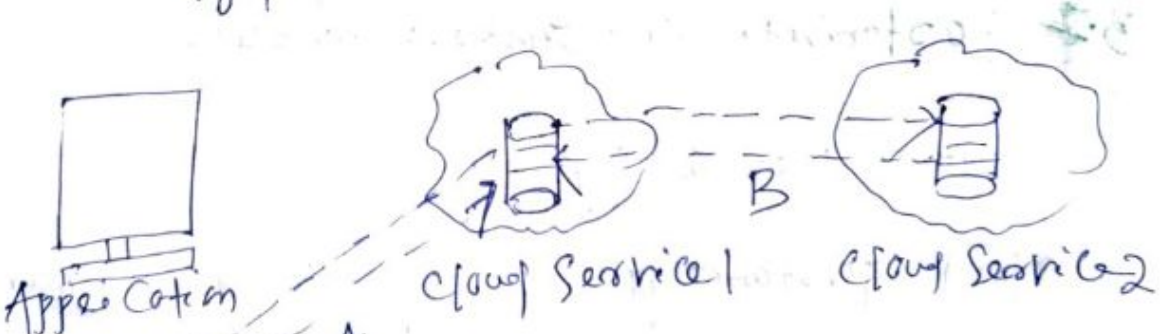
→ Cloud portability enables the migration of cloud services from one cloud provider to another or between public cloud and a private cloud.

→ Two types of cloud portability

(1) Data portability (2) Application portability

(1) Data Portability

Data portability means the ability to move data (Files, documents, database tables etc) from one cloud system to another and have that data usable in the other system.



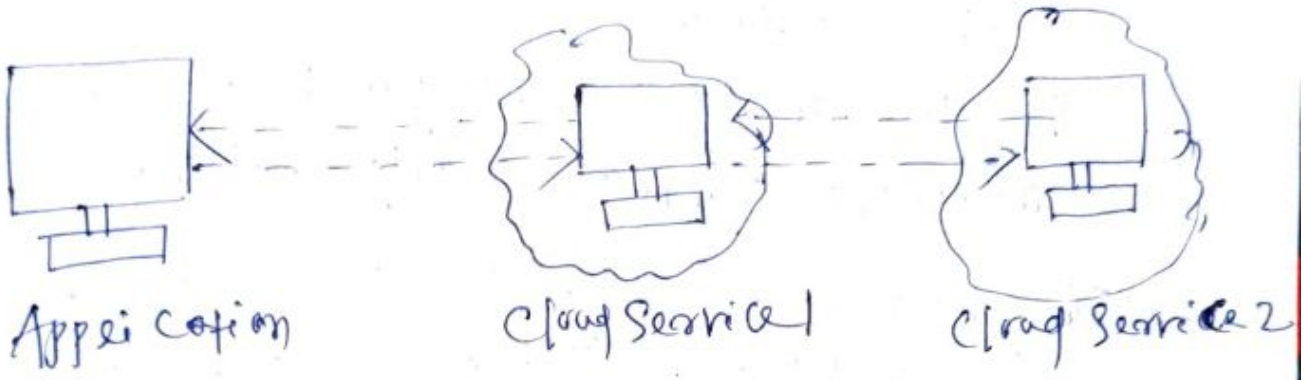
The arrows show examples of cloud data portability

A - Cloud Service System to / from cloud service

B - Cloud Service to / from cloud service

(2) Application portability

Application portability means the ability to move executable software from one cloud system to another, and be able to run it correctly in the destination system.



The arrows show examples of cloud application portability

A - Cloud service system to/from cloud service

B - Cloud service to/from cloud service.

Cloud Interoperability refers to the 2.4

3.5 Cloud Service Management

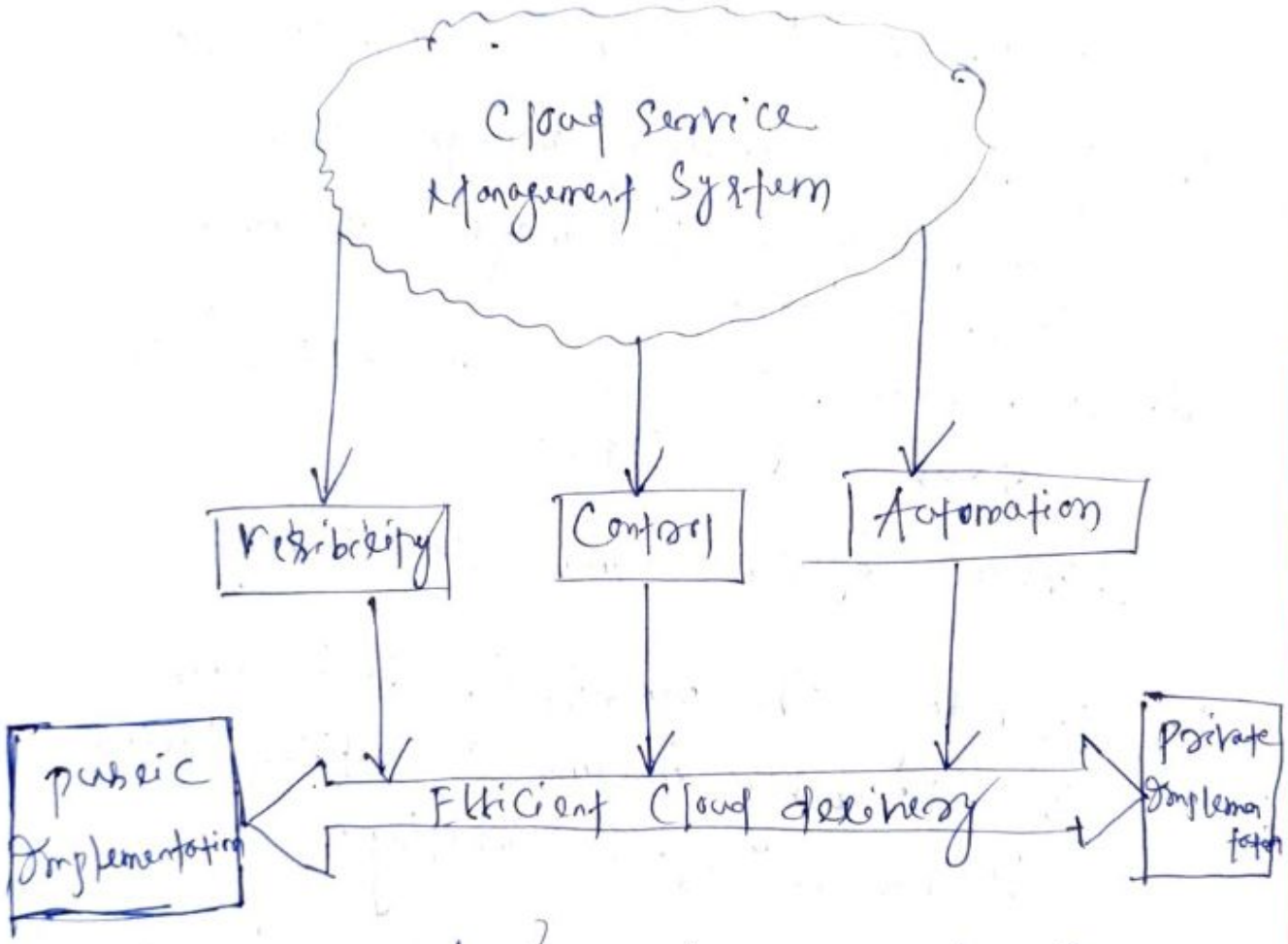
Service Management

1. A System integral of Supply Chain Management that Contents actual Company Sales and the Customers.
2. The goal of Service management is to maximize Service Supply Chains.
3. The purpose of Service management are to reduce high cost by Integrating products and Services.

Cloud Service management

1. Cloud monitoring and cloud Service management tools allow Cloud providers to ensure optimal performance, Continuity and efficiency in virtualized, on demand environments.
2. The designing of dynamic, Cloud-based infrastructure, platform and application Services doesn't occur in a Vacuum.

3. 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.



4. The above figure illustrates that Service management provides the visibility, control and automation needed for efficient cloud delivery in both public and private implementation.

Simpelity User interaction with it

1. The user friendly Self Service accelerates time to value.
2. Service Catalouge enables Standards which drives Consistent Service delivery.

Enable policies to lower Cost with provisioning

1. Automatic allocating and de-allocating of resources will make delivery of Services fast.
2. Provisioning policies allow release and reuse of assets.

Increase System admin productivity

1. Providing the benefits to the broker will probably become a critical Success factor in cloud Computing.
2. Due to the growth of Service brokerage business will increase the ability of cloud Consumers to use Services in a trustworthy manner.

3. These cloud mediators will help companies to choose the right platform, deploy the apps across multiple clouds.

3.8 Cloud offerings

patterns of this category cover different functionality found in clouds regarding the functionality they provide to customers and the behavior they display

(1) Cloud Environments

patterns of this category describe the hosting environments of cloud in detail and refer to other offerings composed to form these environments.

(2) Processing Offerings

Computation facility by the cloud.

(3) Storage Offerings

Storage facility by the cloud.

(4) Communication Offerings

data exchange facility between more than one user by the cloud.

(5) Security offerings

Copy of our data on various services etc I fails data is safe on the other.

3.9. Testing under control

→ Cloud Testing becomes ubiquitous wherein resources such as the software, hardware etc, are checked in a thorough testing effort.

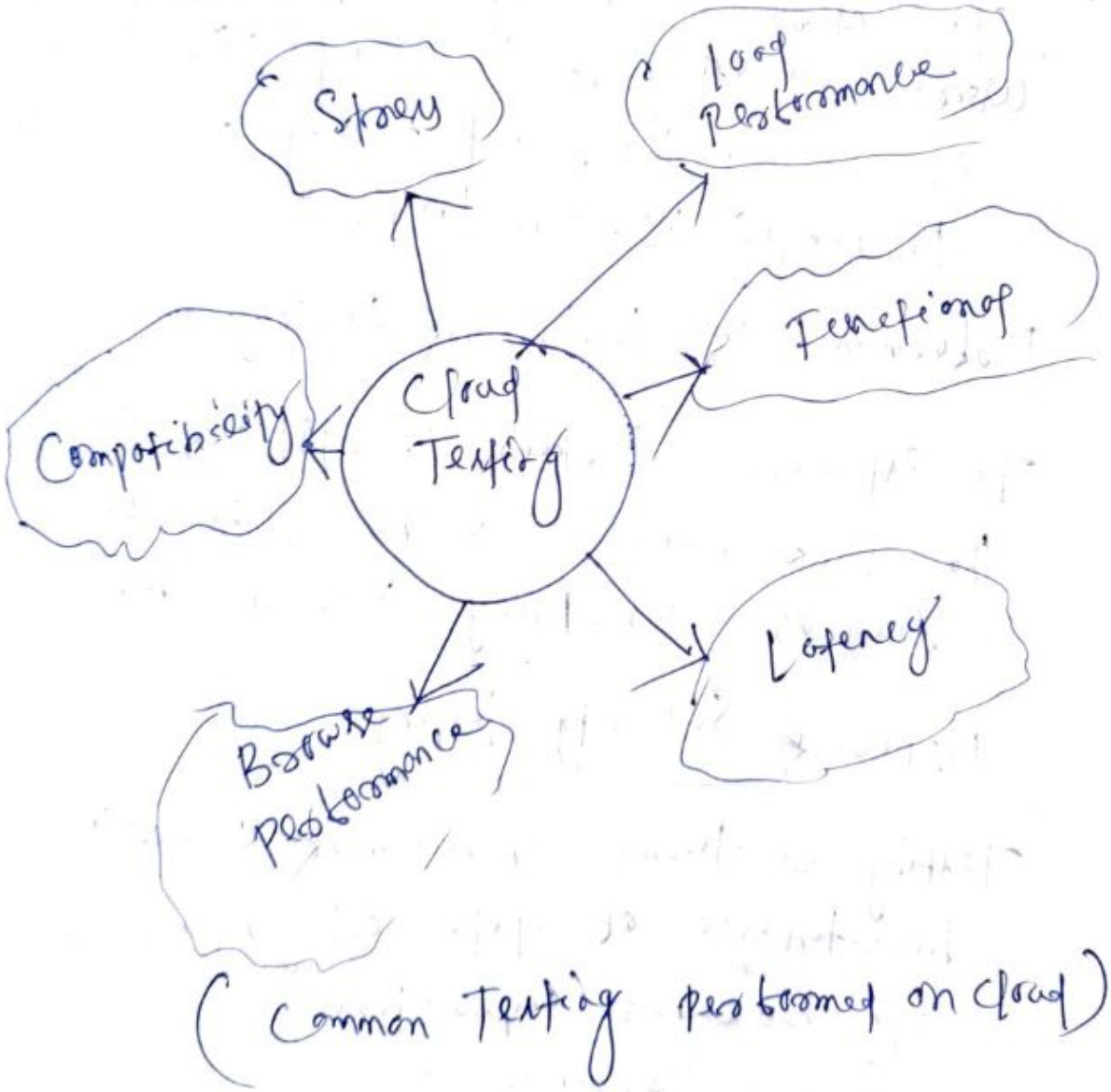
→ Due to certain challenges like:

1. High costs
2. Restrictive budget
3. Numerous test cases
4. Various users across the globe etc.

→ Cloud testing typically involves monitoring and reporting on real-world user traffic conditions as well as load balance and stress testing for a range of simulated usage conditions.

→ Main aim of the cloud testing is consumers can access the IT resources in the test environment.

→ Effective testing becomes essential wherein availability of flexible and scalable infrastructure, distributed test environment or unlimited storage helps in saving time as well as the cost.



(Common Testing performed on cloud)

Functional testing

To verify the basic functionalities with respect to valid input that should match the expected output such as user login, shutdown of system, etc.

Load testing

To ensure stability with a number of users accessing the cloud with scaling-in or scaling-out, load testing is conducted to handle variable load.

performance and Benchmark testing

To establish certain yardsticks considering the performance of the application such as consistency across devices.

Network Security testing

Testing in terms of network connectivity maintenance of data integrity, protocol etc becomes imperative to ensure a secure environment.

Interoperability and Compatibility Testing

→ To test seamless functionality across browsers and platforms.

→ This test is to check the system with various operating system environment. This can be done using cloud testing services.

Stress Testing

This kind of test is to ensure the degree of endurance of the system. That is, to what extent the system is able to perform under excessive pressure. Simulators are used to create peak load situations for conducting stress test.

~~Load Test~~

Browser performance Testing

Different browser platforms are used to check the compatibility of the application with the browsers.

Latency Testing

It is the measurement of latency (speed) between the action and the corresponding response after deployment of the system on the cloud.

3.10 Cloud Service Controls

→ Cloud Service Controls improves your ability to mitigate the risk of data exfiltration from cloud services such as Cloud Storage and BigQuery.

→ With cloud service controls, you can create perimeters that protect the resources and data of services that you explicitly specify.

3.11 Virtual Desktop Infrastructure (VDI)

→ VDI is a technology used to create a virtualized desktop environment on a remote server setup.

→ In simple terms by using VDI you can access your virtual desktops remotely.

→ What are the basic components of VDI

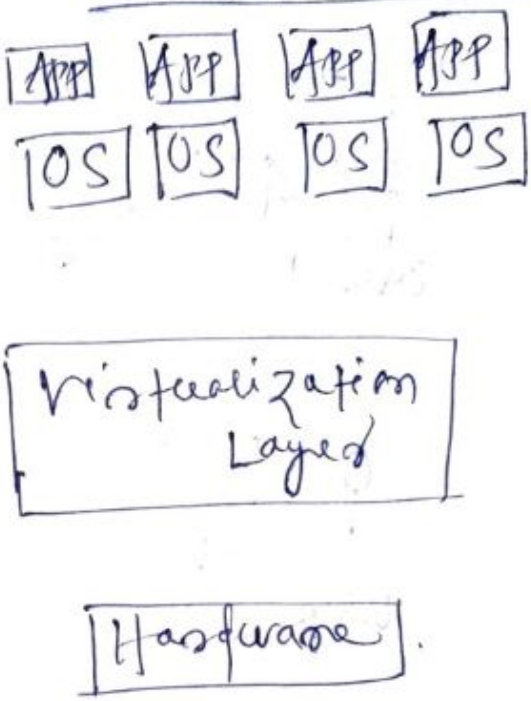
(1) Virtualization
(2) Hypervisor
(3) Connection Broker
(4) Desktop Pools
(5) Application Virtualization.

① Virtualization

Virtualization is the creation of a virtual version of a desktop, OS, Server, or storage

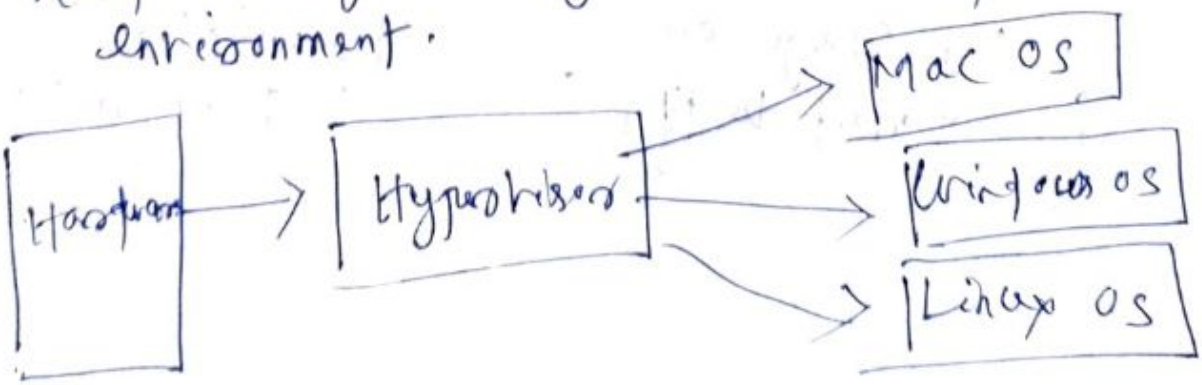
② Traditional Architecture

Virtual Architecture



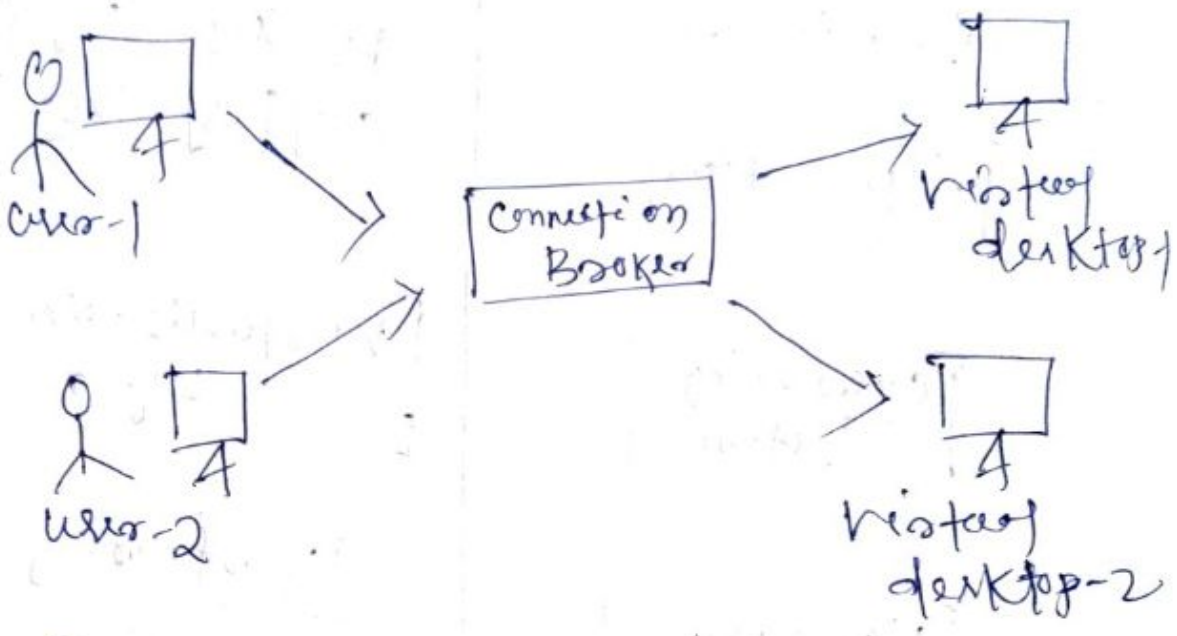
③ Hypervisor

Hypervisor is a software that separates the Operating System from the underlying hardware by creating a virtualized environment.



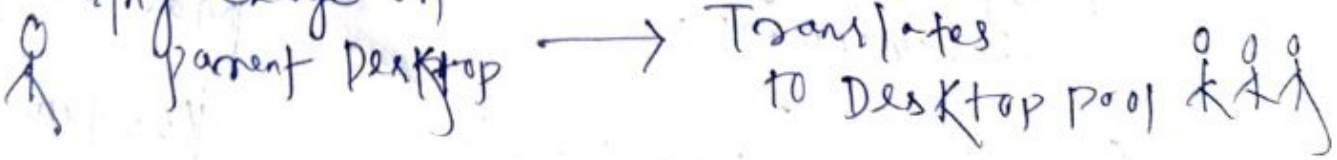
(3) Connection Broker

A Connection broker is a software program that allows the end-user to connect to remote virtual desktop.



(4) Desktop Pools

A Desktop pool is a group of virtual desktops with an identical configuration such as OS, storage, and applications.



(5) Application Virtualization

Application virtualization is the technology used to create a virtualized application image and replicate it to all the virtual desktops in a desktop pool.

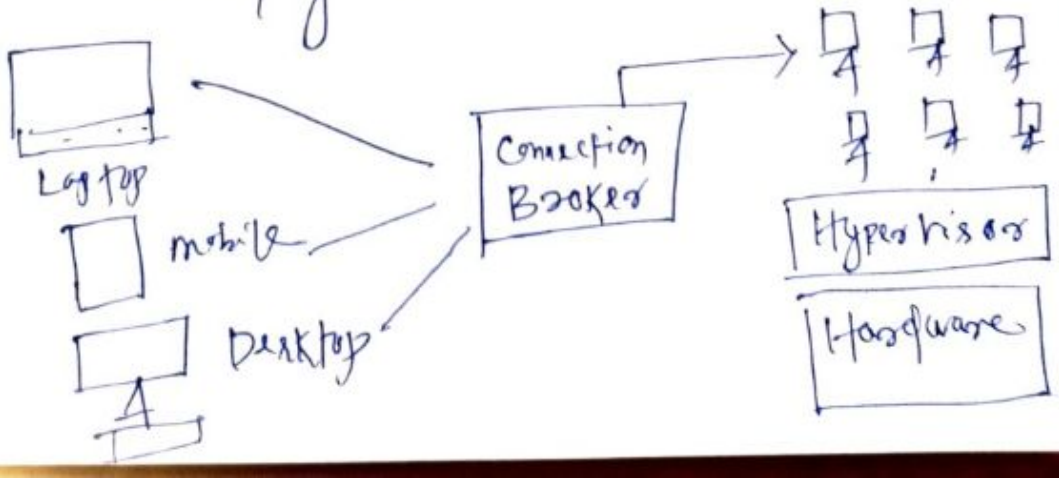
(word, excel, powerpoint)

1. packaging → 2. Installing applications → 3. Mirroring → 4. Application virtualization

How VDI works

After knowing all the components, lets see how VDI works.

1. User sends login request to their end point device.
2. Connection broker accepts the request.
3. Now the user can use desktop according to their process.



What are the benefits of VDI

(1) Access Anywhere
(2) Easy Backup
(3) Bring your own device
(4) High level Security
(5) Cost reduction.

Exercise

Important Questions (Short Questions)

1. What do you mean by cloud Scalability?
2. Define cloud fault tolerance.
3. What are the metrics used for fault tolerance in cloud Computing?
4. Define cloud Solutions.
5. Define cloud ecosystem.
6. What is cloud portability?
7. Define cloud Interoperability.
8. Define cloud Testing.
9. What is Stress Testing?
10. Define Functional Testing.

(11) Define Cloud Service Controls
(12) Define VDI.
(13) What is Hypervisor.
(14) Define Virtualization.
(15) Define Connection Broker.

Important Questions (Long Questions)

(1) Explain about cloud Fault Tolerance with its type.
(2) How cloud ecosystem works?
(3) Describe about the cloud Business process management
(4) Explain about the cloud Portability and its type.
(5) Describe about the cloud Service Management
(6) What are offering by the cloud?
(7) Explain about the cloud Testing and its type.
(8) Describe about the VDI and How VDI works.

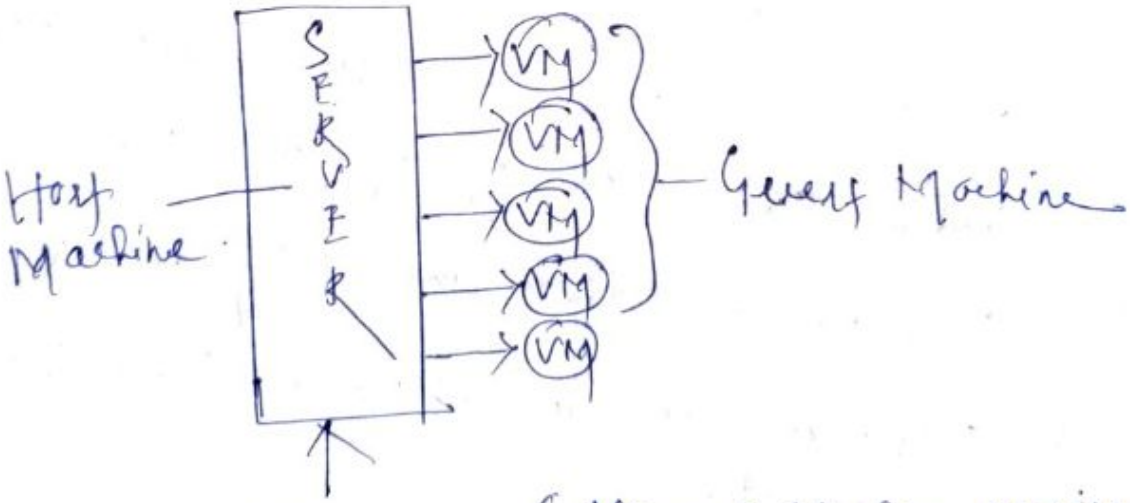
Chapter - 4

Cloud Management and Virtualisation Technology

4.1 Create Virtualised Architecture

→ Virtualization is a technique, which allows to share a single physical instance (server) of a resource or an application among multiple customers and organizations.

→ it does by assigning a logical name to a physical storage and providing a pointer to that physical resources when demanded.



Creation of a virtual machine (VM) over existing operating system and hardware known as hardware virtualization

→ A virtual Machine provides an environment that is logically separated from the underlying hardware.

→ The machine on which the virtual machine is going to create is known as Host Machine and that virtual machine is referred as a Guest Machine.

Types of virtualization

1. Hardware virtualization
2. operating System virtualization
3. Server virtualization
4. Storage virtualization.

1. Hardware virtualization

→ Hypervisor (virtual Machine Manager) is directly installed on the hardware System is known as hardware virtualization.

→ The main Job of hypervisor is to control and monitoring the processor, memory, and other hardware resources.

→ After virtualization of hardware System we can install different operating System on it and run different applications on those OS

Usage

Hardware Virtualization is mainly done for the Server platforms, because Controlling Virtual Machines is much easier than Controlling a Physical Server.

2. Operating System Virtualization

→ When the Virtual Machine Manager (VMM) is installed on the host operating System instead of directly on the hardware System is known as operating System Virtualization.

Usage

operating System Virtualization is mainly used for testing the applications on different platforms of OS.

3. Server Virtualization

→ When the Virtual Machine Manager is directly installed on the Server System is known as Server Virtualization

Usage

Server Virtualization is done because a Single Physical Server Can be divided into Multiple Servers on the demand basis and for balancing the load.

4. Storage Virtualization

Storage virtualization is the process of grouping the physical storage from multiple network storage devices so that it looks like a single storage device.

Usage

Storage virtualization is mainly done for back-up and recovery purposes.

4.2 Data Centers

→ A Data Centers is a facility on some location where multiple servers/machines are engaged in collection, storing, processing and distribution of massive amount of data.

→ Data Centers have a network's most critical systems and vital to the continuity of daily operations

→ The security and reliability of data Centers and their information is a top priority for organizations.

Why are need of Data Centers

→ for storing massive amount of data

→ for providing the 24X7 services to the customers.

→ for data Safety and Security
→ for Conducting day to day business operations.
→ Google and facebook are investing $700 millions.

Components of Data Center

A data Center Consists of

→ A bunch of Servers Connected through network to run Complex applications.
→ A Cooling System to manage the heat released by machines.
→ proper ventilation Systems to ensure optimal air-flow.
→ Sophisticated Security Systems to prevent unauthorized access to data across Centers.
→ power distribution & Backup units (gensets, batteries, etc) for Smooth execution using power Supply Units.
→ ~~Redundant~~ Redundant Units / Backup Systems to ensure maximum Uptime.

Types of Data Centers

- Internet Data Centers (IDCs)
- Cloud Data Centers (CDCs)
- Dark Data Centers (DPCs)

4.3 Resiliency

→ Resiliency is the ability to handle failures gracefully and recover the whole System. This is a huge challenge for services and applications where the components compete for resources and depend on other internal or external components/ services that fail, or may rely on defective Software

→ Cloud resiliency is the capacity to rapidly adapt and respond to risks, as well as opportunities; in simple words resiliency refers to improve our business for handle risks.

→ This also maintains the continous business operations that support growth.

Resiliency Capabilities : The Strategy combines multiple parts to mitigate risks and improve business resilience.

1. From a facilities perspective, one may want to implement power protection.
2. From a Security perspective, to protect our data and applications one may want to implement remote backup, identity management, email filtering or email archiving.
3. From a process perspective, one may implement identification and documentation of most critical business processes.
4. From a organizational perspective, one may want to implement a virtual workstation environment.
5. From Strategy & vision perspective, one may want to look at the key of crisis management process.

Agility

→ in a Cloud Computing Context, agility often refers to the ability rapidly develop, test and launch applications that drive business growth in a constantly Changing IT environment

→ Cloud technology offers business a key means of promoting agility, and is a key tool in the enterprise push toward better adaptability.

Advantages of agility

(1) Greater Business Continuity and flexibility:

Cloud services can be rolled up or down as per business requirements without increasing the pool of IT equipment that company must purchase and manage.

(2) Infrastructure Agility:

Cloud allows companies to significantly decrease the time it takes to provision and de-provision

(3) Automated allocation of resources

Cloud computing relies on distributing workloads and sharing of resources to achieve coherence and economic of scale.

(4) up-to-date technology upgrades

The refresh cycles try on upgrade can be long as there are plenty dependencies that need to be planned out infrastructure, operations, and storage.

4.5 Cisco Data Center Network Architecture

A comprehensive Architecture that enables IT executive to

→ Consolidate and Virtualize Computing Storage and network resources.

→ Deliver Secure and optimize employee, partner and Customer access to information and applications.

→ protect and rapidly recover IT resources and applications.

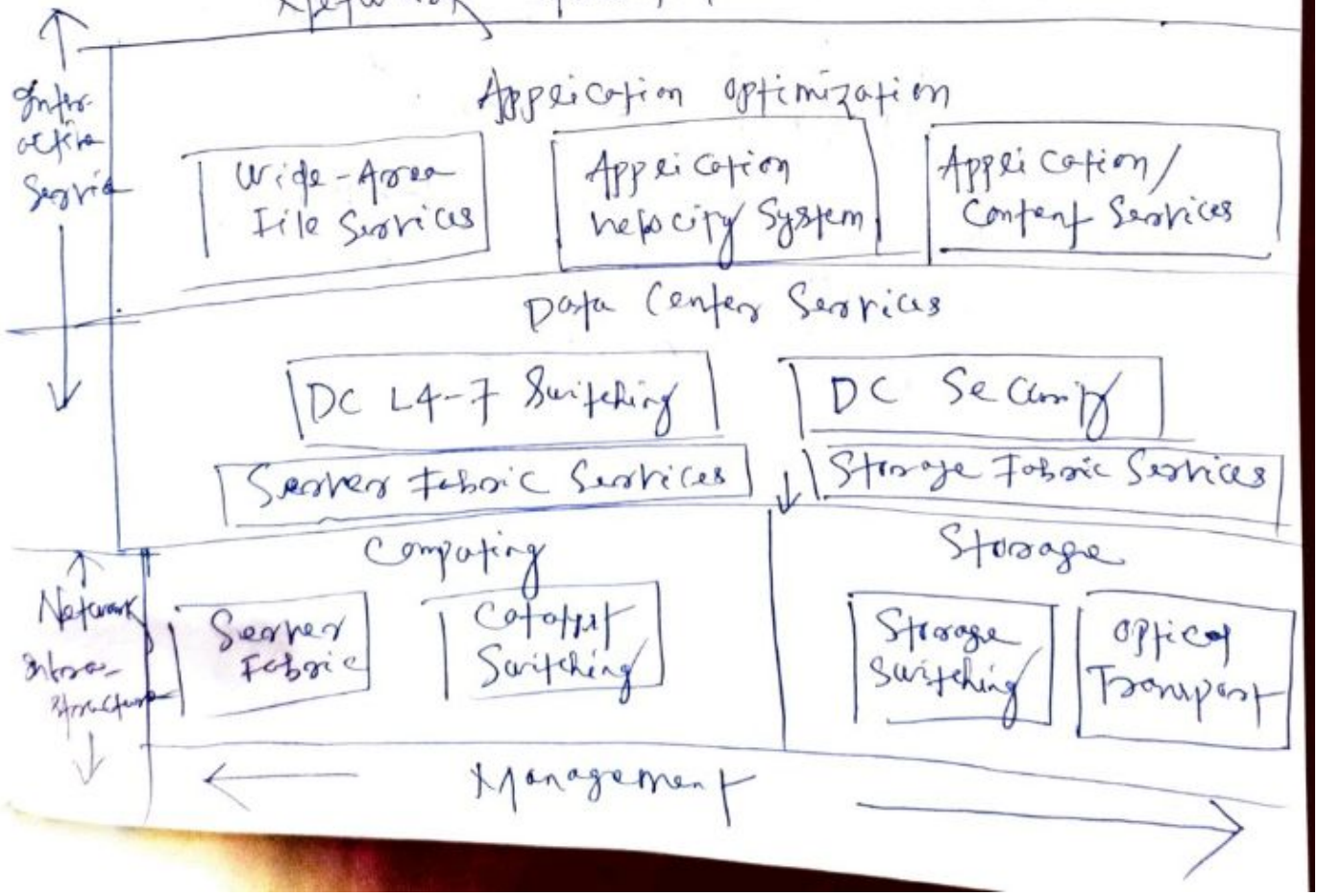
Built with

Network Infrastructure : Gigabit/10 Gigabit, InfiniBand and Storage Switching and optical transport.

Interactive Services : Storage Fabric Services, Computer Services, Security Services and application optimization services

Management : Fabric Manager (element and network management and Cisco Vframe (Server & Service provisioning).

→ CISCO Data Center Network Architecture in Support of SOA (Service-oriented Network Architecture



Benefits

→ Low priced Server and Storage infrastructure.

→ Increased business agility and adaptability

→ Ability to meet regulatory Compliance Standards with integrated network Security and Support for business Continuance.

→ Tested and verified design and extensive Service offerings for lower Implementation Costs & reduced risk.

→ Investment Protection for Core data Center platforms offering multiyear deployment lifecycles.

→ Rapid application development and time to market of business critical Services.

4.6 Cloud Storage

→ It is a service model in which data is transmitted and stored on remote storage system where it is maintained, managed, back up and made available to the users over internet.

→ Cloud storage is based on virtualized infrastructure with accessible interfaces.

→ With the help of RESTful APIs user can retrieve and access data from storage.

Advantages

→ pay for what is used.
→ Utility billing
→ Global availability
→ Ease of use
→ Recovery, Security & accessibility

Disadvantages

→ Back-ups may be slower depends upon the internet speed
→ Higher internet utilization
→ Privacy Concerns.

4.1 Cloud provisioning

→ "The cloud provisioning is the allocation of 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

(1) advance provisioning
(2) dynamic provisioning
(3) User self provisioning

The term provisioning simply means "to provide".

→ Cloud provisioning primarily defines how, what and when an organization will provision cloud services. These services can be private/internal, public or hybrid cloud products & solutions.

→ Cloud providers deliver cloud solutions through on-demand, pay-as-you-go systems as a service to customers and end users. Cloud provider customers access cloud resources through.

→ Internet and programmatic access and are only billed for resources and services used according to a subscribed billing method.

→ Depending on the business model, a cloud provider may provide various solutions, such as.

Infrastructure as a Service (IaaS): may include virtual servers, virtual storage and virtual desktops / computers.

Software as a Service (SaaS): Delivery of simple to complex software through the internet

Platform as a Service (PaaS): A combination of IaaS and SaaS delivered as a unified service.

Types of provisioning

(1) Advance provisioning
(2) Dynamic provisioning
(3) User self provisioning

(1) **Advanced / post sales provisioning:**

The Customer is provided with the resource upon Contract / Service Sign up.

(2) **Dynamic / on-Demand provisioning:**

The user / Customer adds a Cloud Service or device themselves.

(3) **User-self provisioning:**

The user / customer adds a Cloud Service or device themselves.

(2) **Dynamic / on-Demand provisioning**

The Customer or requesting application is provided with resources on run time.

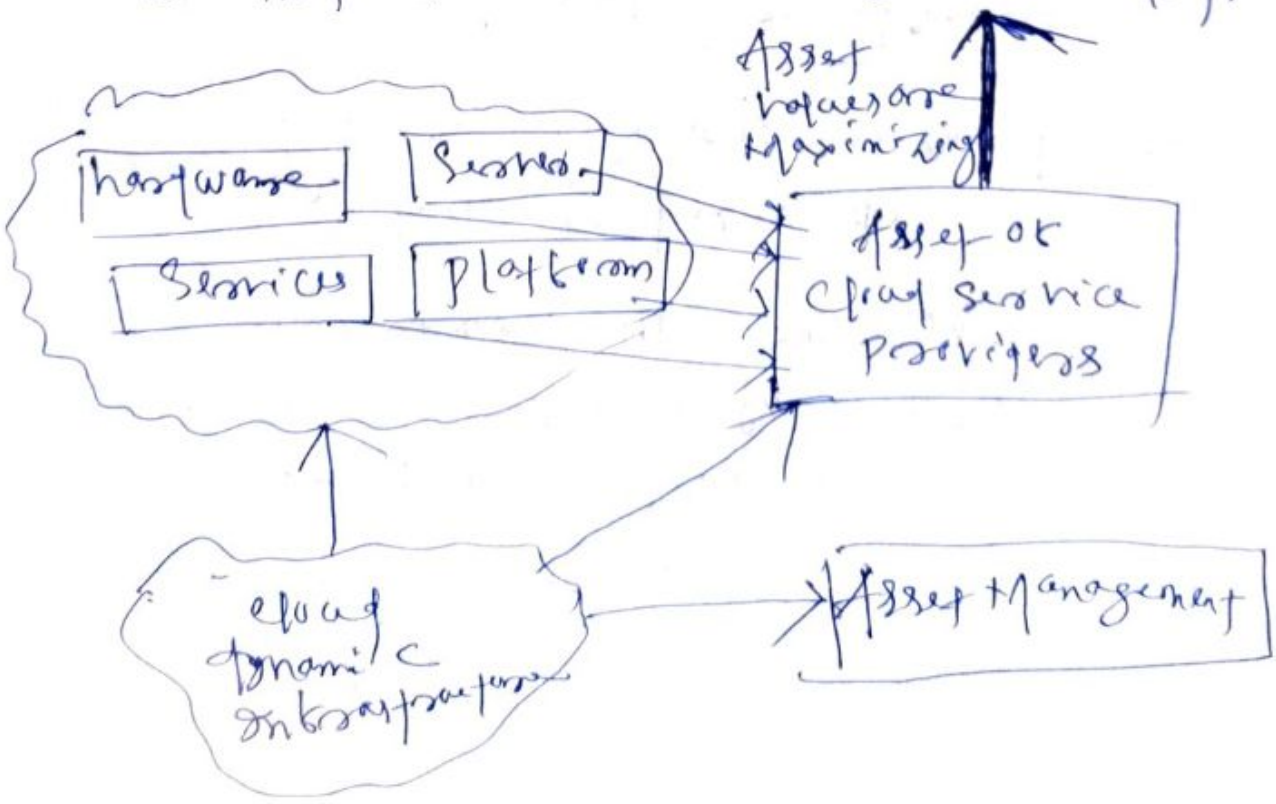
4.8 Cloud Asset Management (CAM)

→ in this especially the assets are the property which is involved in providing the cloud services are getting managed.

here property means

(1) Software
(2) Hardware
(3) Customer
(4) Policies
(5) migration
(6) Cloud adoption
(7) Availability.

→ They are getting managed in such a way so that their value will get maximized.



→ CAM is primarily about managing the challenges of cloud applications, platforms and infrastructure (SaaS, PaaS, IaaS). For instance;

1. Inability to track and manage the growing use of SaaS applications and providers.
2. Lack of a centralized view of cloud resources and consumption.
3. Limited access to SaaS subscription data
4. Limited access to actual SaaS, IaaS, and PaaS usage data.

Benefits of Cloud Asset Management (CAM)

1. Accurate tracking of key applications deployed in cloud.
2. Overcome the limitations of cloud portals by providing access to a single centralized view.
3. Expanded access to data and improved analysis and reporting

4. Accurate, complete view of investments and their usage across the whole IT estate enables better cost control.

4.9 MapReduce

→ Mapreduce is a software framework and programming model used for processing huge amounts of data

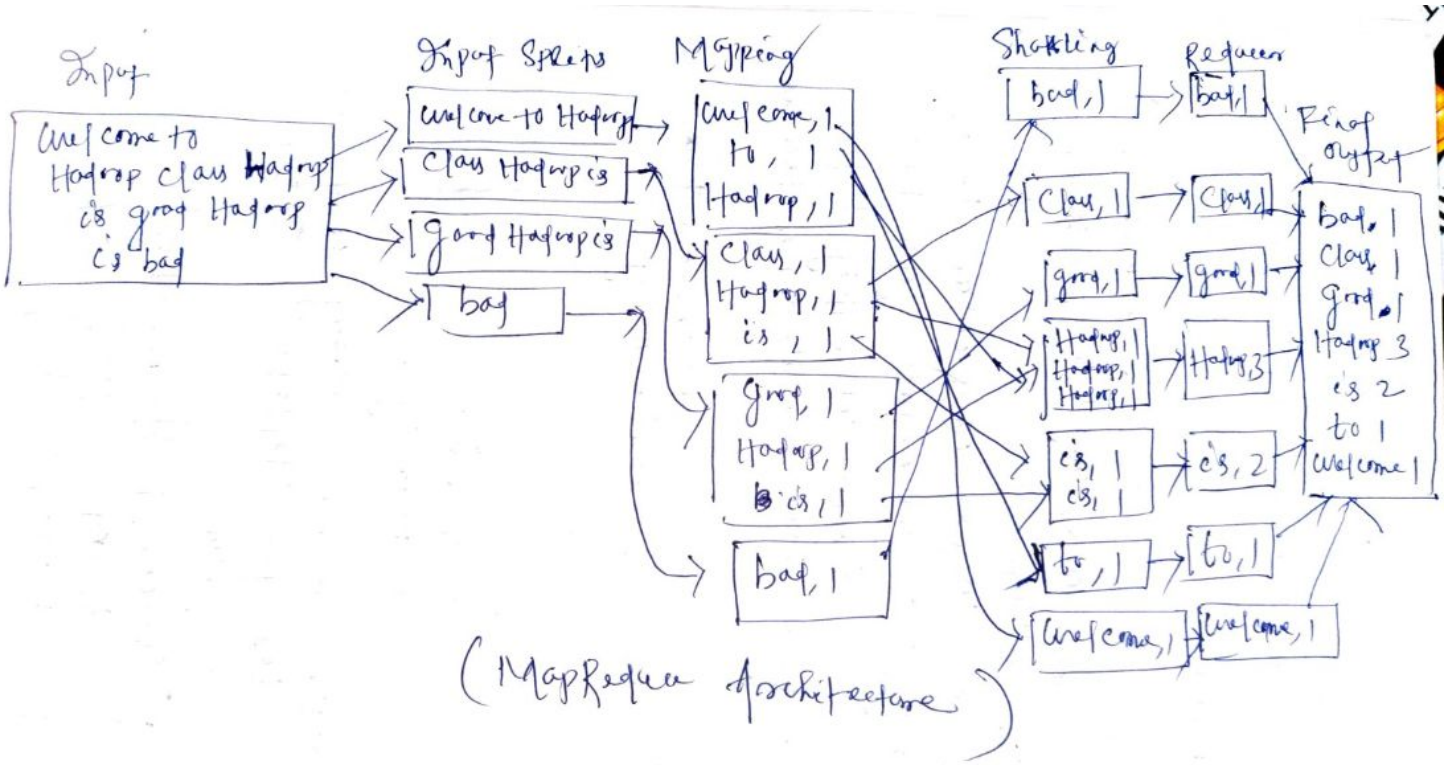
→ Mapreduce program work in two phases, namely Map and Reduce. Map tasks deal with splitting and mapping of data while Reduce tasks shuffle and reduce the data

→ The whole process goes through four phase of execution namely

(1) Splitting
(2) Mapping
(3) Shuffling
(4) Reducing.

Consider you have following input data for your MapReduce in Big data program

```
Welcome to Hadoop class
Hadoop is good
Hadoop is bad.
```



(MapReduce Architecture)

Input Splits

An input to MapReduce in Big Data Job is divided into fixed-size pieces called input Splits. Input Split is a chunk of the input that is consumed by a single map.

Mapping

This is the very First Phase in the execution of Map-reduce program. In this phase data in each Split is passed to a mapping fr produce output values.

Shuffling

The phase consumes the output of mapping phase. Its task is to Consolidate the relevant records ~~from~~ from mapping phase output. In our example, the same words are clubed together along with their respective frequency.

Reducing

In this Phase, output values from the Shuffling phase are aggregated. This Phase combines values from Shuffling phase and returns a single output value. In short, this phase Summarizes the Complete dataset.

In our example, this phase aggregates the values from Shuffling Phase. i.e Calculates total occurrences of each word

1.10 Cloud Governance

→ Cloud governance is a general term for applying specific policies or principles to use of Cloud Computing services.

→ In other terms we can say that Cloud governance refers to the decision making processes, criteria and policies involved in planning, architecture, acquisition, deployment, operation and management of a Cloud Computing capability.

→ The goal of Cloud governance is to secure applications and data when they are located remotely.

There are three reasons of Cloud governance

(1) Enable "business of "Cloud Speed" and refabricate a Cloud ~~Service~~ Centric IT operating model based on the Speed, agility and cost of Cloud Computing.

(2) Enable appropriate Cloud decision making without friction.

(3) Integrated with existing enterprise IT governance processes, policies, boards and tools.

(4) Balanced appropriate coverage for key decisions, investments and risks while achieving the benifits of clouds.

(5) proactive to anticipate and prevent shadow clouds and unauthorized cloud ~~services~~ activities that expose organizational risks.

The governance is applied in cloud for

(1) Setting Company policies in Cloud Computing

(2) Risk based decision which cloud provider, if any, to engage.

(3) Assigning responsibilities for enforcing and monitoring of the policy compliance.

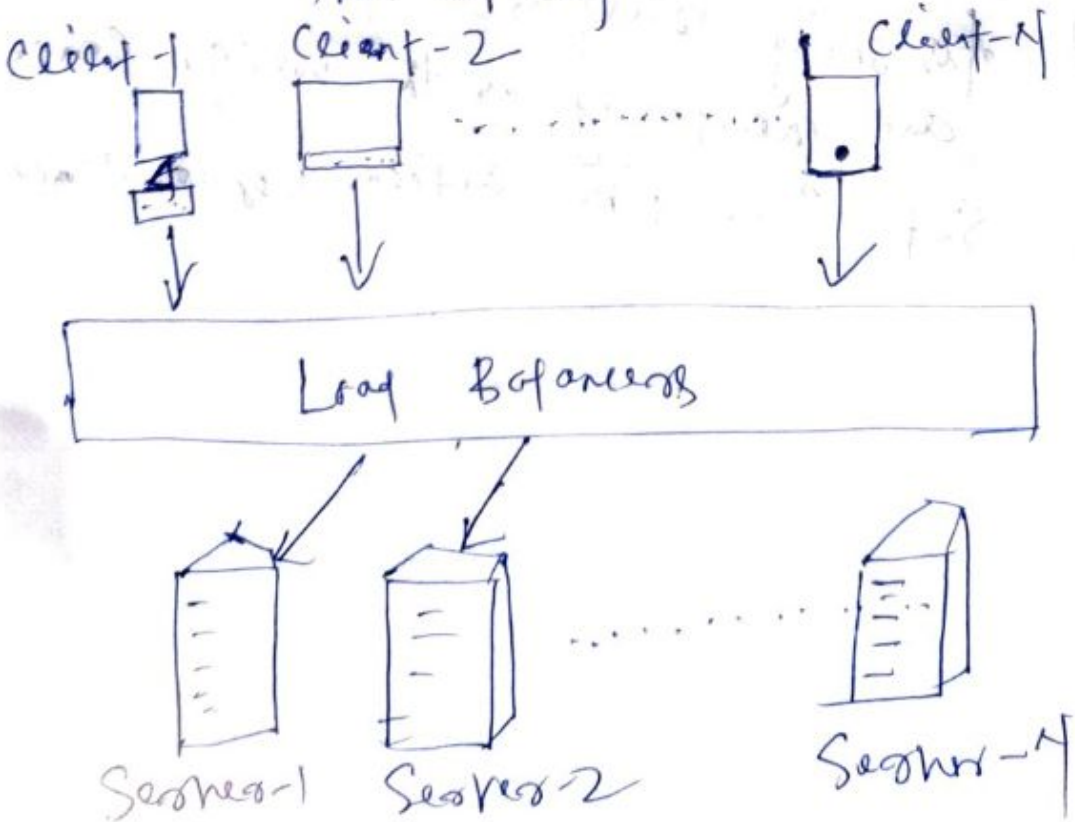
(4) Set Corrective action for non compliance.

4.11 Load balancing

→ Cloud load balancing is defined as the method of splitting workloads and computing properties in a cloud computing.

→ It enables enterprise to manage workload demands or application demands by distributing resources among numerous computers, networks or servers.

→ Cloud load balancing includes holding the circulation of workload traffic and demands that exist over the internet.



Load Balancers

→ Load Balancers allocates the workload and balances it between two or more Cloud Servers.

→ We can so outline our infrastructure to permit it to meet activity spikes, optimize the allocation of resources and guarantee a minimal response time.

→ Using a load balancer is recommended in all cases, whether we require one or more of the following

- Guaranteed Service Continuity
- Handle high traffic
- Be prepared for sudden request spikes.

Load Balancers objective

→ Maintain System fitness

→ Improve System performance

→ protect against System failures.

4.12 High Availability

1. In simple words we can say that high availability refers to the availability of resources in a computer system.
2. In terms of cloud computing it refers to the availability of cloud services.
3. High availability is the heart of the cloud.
4. It provides the idea of anywhere, any time access ~~to~~ to service of cloud environment.
5. Availability is also related to reliability.
6. Availability is a technology issue as well as business issue.
7. High Availability can be simply defined by the simple equation

$$HA = \frac{MTBF}{MTBF * MTTR}$$

Where MTBF – Mean Time Between Failures
MTTR – Mean Time To Repair
HA – High Availability.

8. There is two way improve the availability
   (a) Increase MTBF to very large values.
   (b) Reduce MTTR to very low values.

To maintain High Availability using Four things

1. make ready for Server failure
2. make ready for Zone failure
3. make ready for cloud failure
4. Automate and Test everything.

4.13 Disaster Recovery

1. A disaster recovery is the process by which an organization can recover and access their Software, data and hardware.
2. It is necessary for faster disasters recovery to have an infrastructure supporting high availability.
3. The failure of disaster recovery plan mainly due to lack of high availability preparation, planning, and Maintenance to occurance of the disaster.

Chapter-5

Virtualisation

5.1 Virtualisation

refers to the 4.1

5.2 Network Virtualisation

→ Network virtualisation is the process of Combining hardware and Software network resources and network functionality into a Single.

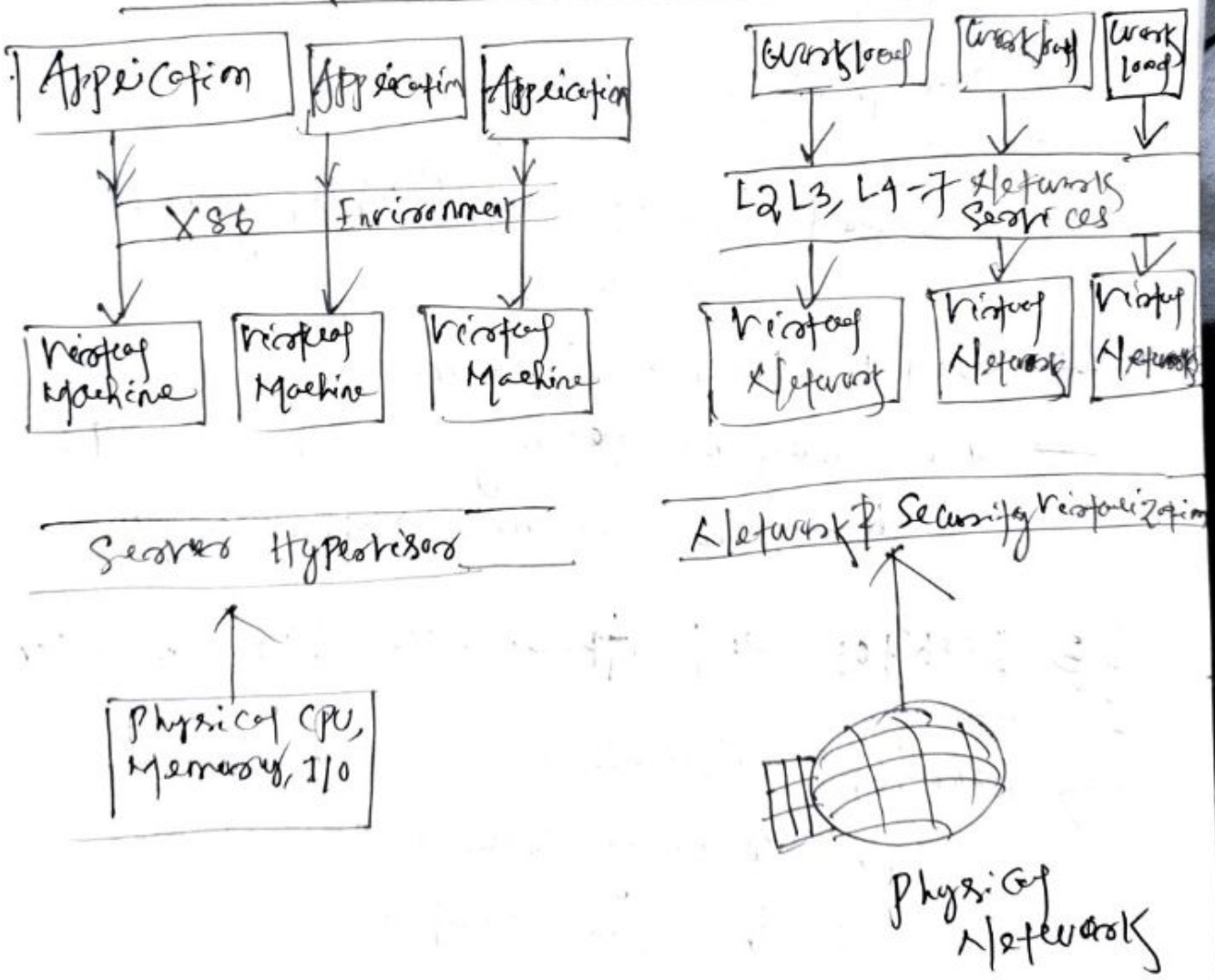
→ Network virtualisation is a process of logically grouping Physical networks and making them operate as Single or multiple independent networks Called Virtual Networks.

→ In Case Virtual network each application sees its own logical network, independent of Physical network.

→ Network virtualization is an Abstraction of the physical network.

- Support for multiple logical networks running on a common shared physical substrate
- A container of network services.

Network virtualization Architecture



Advantages

→ more productive IT environments (i.e efficient Scaling)
→ improved Security and recovery times.
→ Faster in application delivery.
→ more efficient networks
→ Reduced overall Costs.

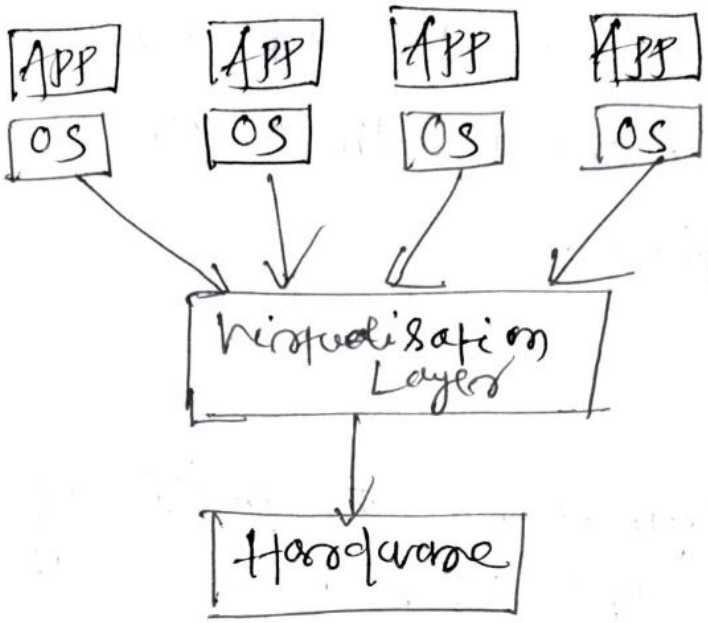
Disadvantages

→ increased Upfront Costs (investing in Virtualization Software).
→ Need to License Software
→ There may be learning Curve if IT Managers are not Experienced..

3 Desktop and Application Virtualisation

→ Desktop Virtualisation allows the User's OS to be remotely stored on a Server in the data Center
→ it allows the user to access their desktop virtually, from any location by different machine.

They can launch applications, open files, resize windows, edit documents and more.



5.4 Desktop as a Service (DaaS)

→ Desktop-as-a-Service or simply DaaS, securely delivers virtual apps and desktops from the cloud to any device or location.

→ The DaaS virtualization solution provisions secure SaaS and legacy applications as well as full-Windows-based virtual desktops and delivers them to your workforce.

→ DaaS offers a simple and predictable pay-as-you-go subscription model, making it easy to scale up or down on-demand.

→ Desktop as a service is also known as a virtual desktop or hosted desktop services.

→ DaaS is a Cloud Computing offering where a Service provider delivers virtual desktops to end users over the internet, licensed with a per-user subscription.

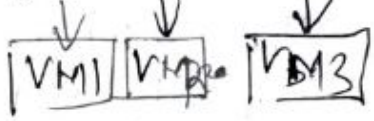
Advantages

→ Faster deployment and provisioning of active end users.
→ Reduced downtime for IT support
→ Cost Savings
→ Increased device flexibility
→ ~~Enchance~~ Enhanced Security
→ Reliability
→ Uninterrupted Connectivity
→ Disaster recovery.

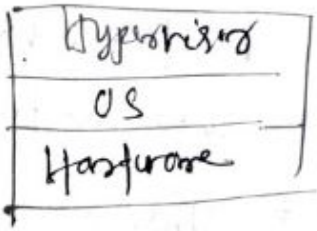
5.5 Local Desktop Virtualisation

→ Local Desktop Virtualisation means the operating System runs on a client device using hardware virtualisation, and all processing and workloads occur on local hardware.

→ This type of desktop virtualisation works well when users do not need a continuous network connection and can meet application computing requirements with local System resources.



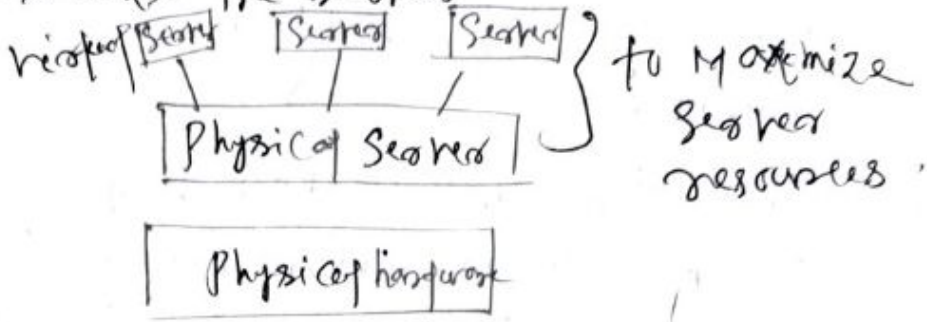
5.6 Virtualisation Benefits



→ Economical
→ Flexible operations
→ Security
→ Eliminates the risk of System failure
→ Flexible transfer data
→ Better resource utilisation
→ Remote access
→ pay per use of the IT infrastructure on demand.
→ enables running multiple OS.
→ If any virtual machine is not working or having any problem, others will not be affected.

5.7 Server Virtualisation

→ It is the partitioning of Physical Server into several virtual Servers. It is used to maximize the Server resources.



Usage of Server Virtualisation

→ The Server virtualisation technology is mainly used in Webservers by using virtual Webservers, it provides low-cost web hosting services.

→ Instead of having Separate Computer for each Web Server, we can have number of virtual Servers on the same Computer.

Server Virtualisation is used:

→ to make more efficient use of Server resources.

→ to improve the Server availability

→ to help in disaster recovery

→ development & testing, and

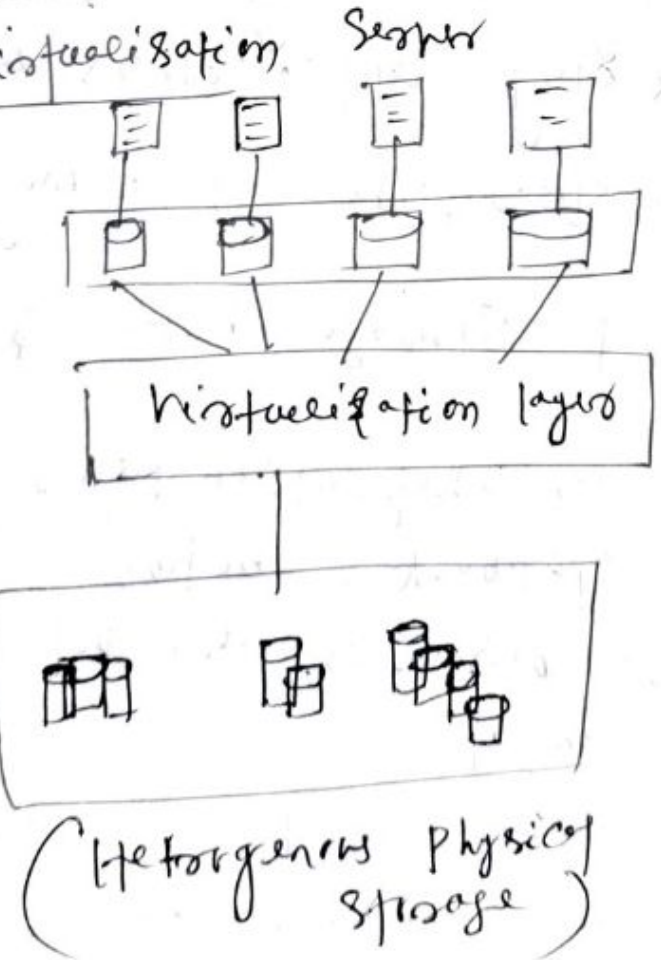
→ to centralized the Server administration.

Advantages of Server Virtualisation

1. Each virtual Server can be independently rebooted.
2. Server virtualisation reduces the costs because less hardware is required.

5.8 Block and File level Storage Virtualisation

Storage virtualisation



→ process of presenting a logical view of physical storage resources to hosts

→ Logical storage appears and behaves as physical storage directly connected to host.

Benefits of Storage Virtualization

→ Increased storage utilization

→ Adding or deleting storage without affecting application's availability.

→ Non-disruptive data migration.

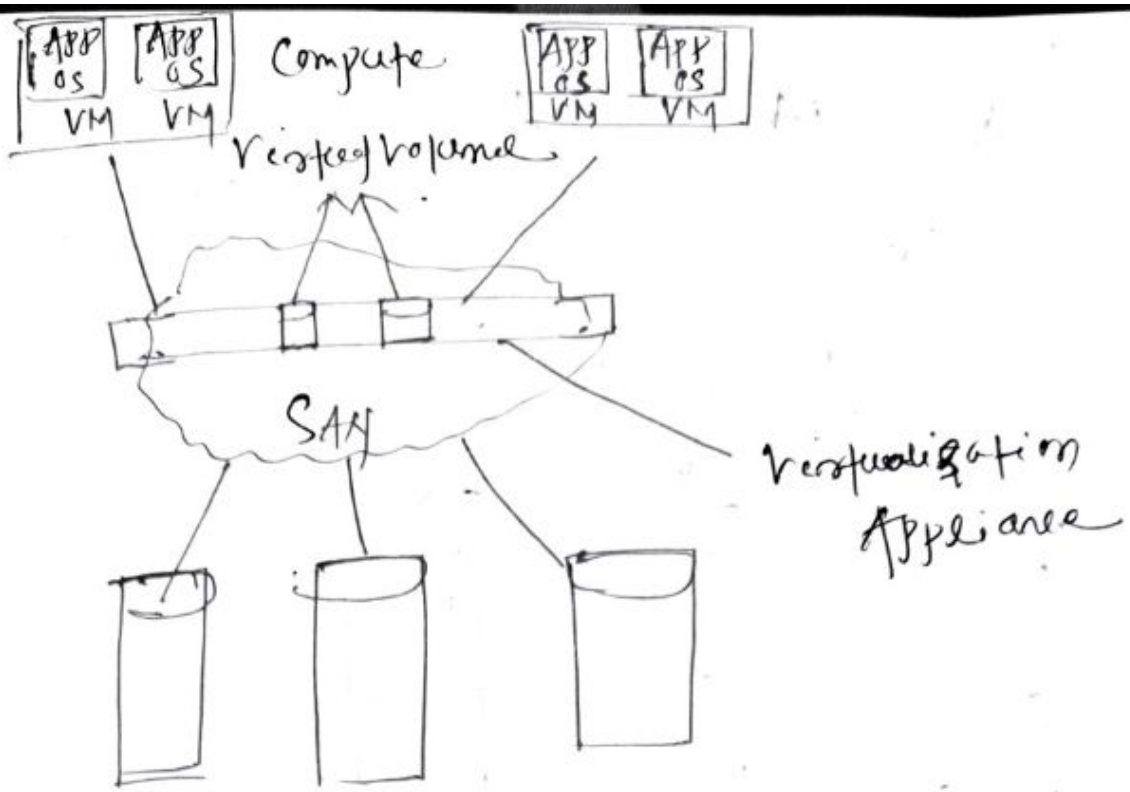
Block-level Storage Virtualization

→ Creates an abstraction layer at Storage Area Network, between Physical storage resources and volumes presented to compute.

→ Uses virtualization appliances to perform Mapping operation

→ makes underlying storage infrastructure transparent to compute

→ Enables significant cost and resource optimization



File Level Storage Virtualisation

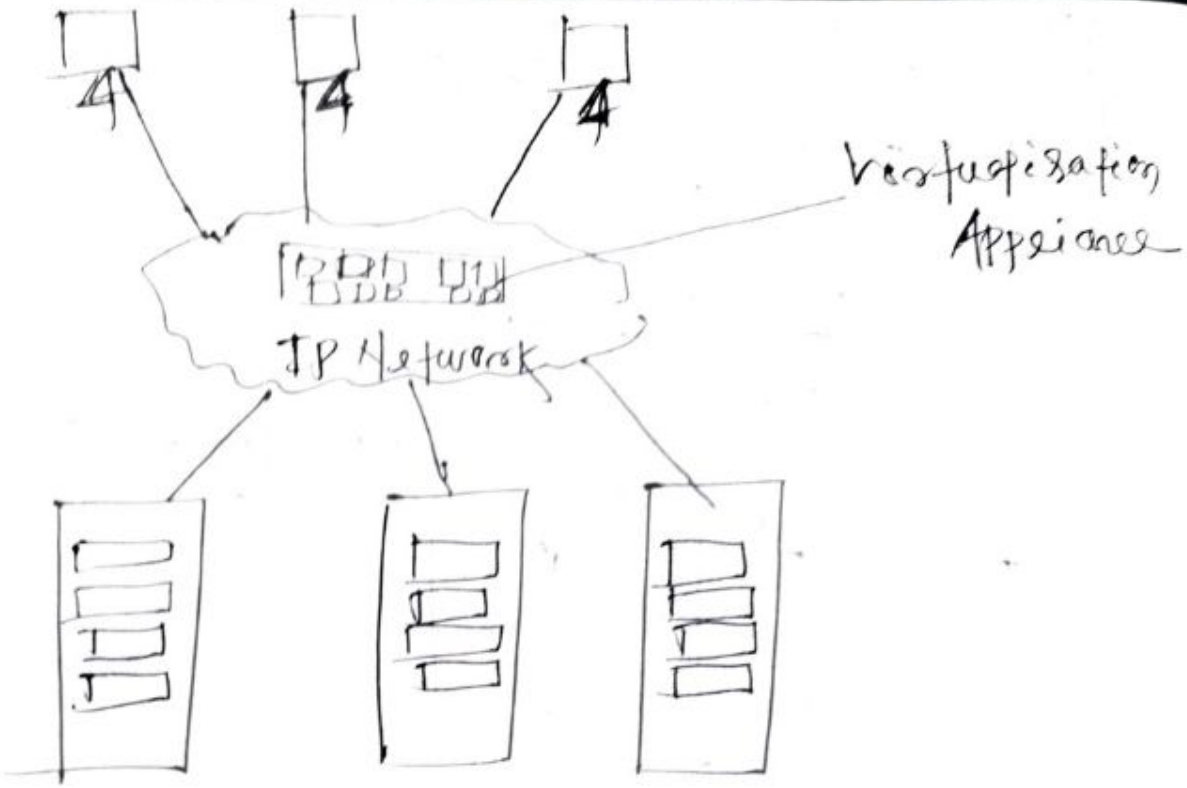
→ provides an abstraction in the NAS / File Servers environment

- Eliminates dependencies between the file and its location

→ Enables movement of files between NAS Systems without impacting client access

→ provides opportunities to optimize storage utilization

→ implemented using global namespace



Multi-headed NAS Systems.

(NAS - Network - Attached Storage)

5.9 Virtual Machine Monitor

Same as Hypervisor

5.10 Infrastructure requirements

refers to the 1.6

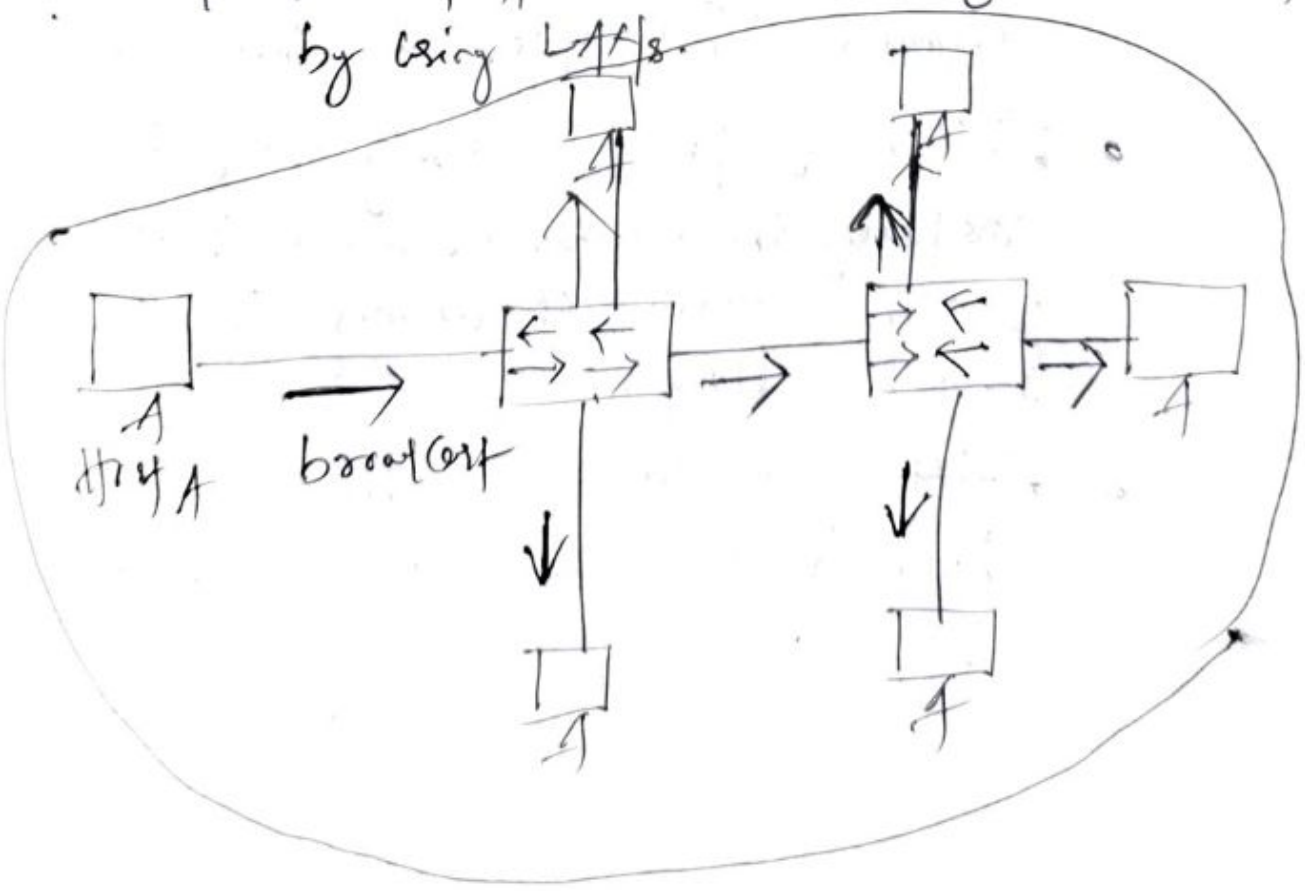
5.4 VLAN and VSAN

VLAN

→ VLAN stands for virtual Local Area Network

→ A VLAN allows Several networks to work virtually as one LAN one of the most beneficial elements of a VLAN is that it removes latency in the network, which saves network resources and increase network efficiency.

→ In addition, VLANs are created to provide Segmentation and assist in issues like Security, network management and Scalability Traffic patterns can also easily be controlled by using LANs.



→ VLANs can spread across multiple switches, with each VLAN being treated as its own subnet or broadcast domain. This means that frame broadcasted onto the network will be switched only between the ports within the same VLAN.

→ A VLAN acts like a physical LAN, but it allows hosts to be grouped together in the same broadcast domain even if they are not connected to the same switch.

→ Here the main reasons why VLANs are used (Advantages)

- VLANs increase the number of broadcast domains while decreasing their size.
- VLANs reduce security risk by reducing the number of hosts that receive copies of frames that the switches send.
- VLAN you can keep hosts that hold sensitive data on a separate VLAN to improve security.

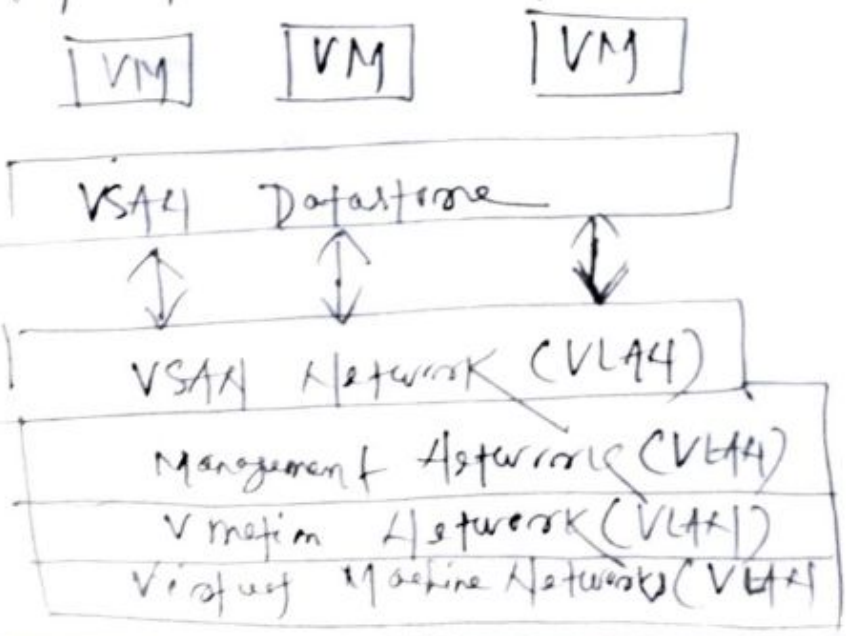
- You can create more flexible network designs that group users by department instead of physical location.
- network changes are achieved with ease by just configuring a port into the appropriate VLAN.

VSAN

→ VSAN stands for Virtual Storage Area Network

→ VSAN is a logical Partitioning created with a physical storage area network.

→ This implementation model of a Storage virtualization technique divides and allocates some or an entire Storage area network into one or more logical SANs to be used by internal or external Services and Solutions.



The benefits of VSAN include

- performance, since the local server can access data at fast speed and low latency
- Low infrastructure cost, since there are no networked storage appliances.
- High ~~scale~~ scalability - simply put, add more servers and get more storage
- IO boot storms since os/app stack images are stored locally.

Chapter - 6
Cloud Security

6.1 Cloud Security Fundamental

Cloud Security Consists of

→ Set of policies
→ Controls
→ procedures
→ Technologies.

that work together to protect cloud based Systems, data, & infrastructure.

→ Service delivery depends upon the integrity of cloud service provider or the cloud security solutions.

6.2 Cloud Security Services

→ Identity and Access management Provides controls for assured identities and access management.

→ Data Loss Prevention monitoring, protecting and verifying the security of data at rest, in motion and use in the Cloud and on-premises.

→ Security information and event management Systems accept log and event information

→ Business continuity and disaster recovery are the measures designed and implemented to ensure operational resiliency in the event of any service interruptions.

→ Network Security provide deal with Security control within in a cloud environment is generally provided through virtual devices.

→ in Encryption, there are, distinctive algorithms that computationally difficult or almost impracticable to break.

6-3 Cloud Security Design Principles

=

1. Governance framework

When procuring a cloud service, ensure that the Supplier has a suitable Security governance framework in place. A governance framework will ensure that procedure, personnel, physical and technical controls remain effective through the lifetime of services.

2. operational Security

The Service provider should have processes and procedures in place to ensure the operational Security of the Service. The Service will need to be operated and managed Securely in order to detect or protect attacks against it.

3. Data in transit protection

Consumer data transiting network should be adequately protected against integrity and confidentiality. This should be achieved via network protection and encryption.

4. Asset protection and resilience

Consumer data, and the assets storing or processing it, should be protected against physical tempering, loss, damage or seizure.

5. Separation between Consumers

Separation between different consumers or service providers and malicious or compromised consumer from affecting the service or data of another.

6. Identity and authentication

Consumer and service provider access to all service interface should be constrained to authenticated and authorized individuals.

→ All Cloud Services will have some requirement to identify and authenticate users wishing to access Service interfaces.

7. Secure deployment

Services should be designed and developed to identify and mitigate threats to their security.

8. Audit information provision to consumers

Consumers should be provided with the audit records they need to monitor access to their Service and the data held within it.

9. Disaster Recovery

A Disaster Recovery is the process by which an organization can recover and access their Software, data, and hardware.

6.4 Secure Cloud Software requirements

Dynamic Risk Assessment

→ enterprise framework that support Machine to machine data collection for continous monitoring

→ Comprehensive assessment for Vulnerability behavior, Configuration and impact.

→ Real-time discovery Capability for assets, applications and data.

Threat-Based Defense

→ Defend the key attack vectors and priority targets based on intelligence.

→ Automated assessments with counter-measure awareness.

→ No impact to availability or performance of Critical Systems.

Monitoring across several domains

→ Integration of IT risk data elements with cyber physical data for impact decisions & high-level decision support systems.

→ Handling Big Security data.

6.5. Cloud Security Policies

→ Security policy is an overall general statement produced by senior management, a selected policy board or committee of an organization that dictates what role security plays within that organization.

There are different types of Security policies

(1) Regulatory

- Regulatory policy ensures that the organization is following Standards set by specific industry regulations. These policies that an organization must implement due to compliance, regulation, or other legal requirements.
- These companies can be financial institutions, public utilities, or some other type of organization that operates in the public interest.

(2) Advisory

- Advisory policy strongly advises employees on the behaviours and activities which should and should not take place within the organization. These policies are not mandatory but are strongly suggested, perhaps with serious consequences defined.
- Failure to follow them will result in consequences such as termination or a job action warning. A company with such policies wants most employees to consider these policies mandatory.

(3) Informative

→ The purpose of this policy is to communicate information to a specific audience that audience is generally any individual who has the opportunity or cause to read the policy.

→ Informative policies typically carry less importance than regulatory or advisory policies, they can carry strong messages about specific situations to the audience.

6.6 Cloud Computing Security Challenges

Cloud Computing Security Challenges. fall into three broad categories

① Data Protection

→ Securing your data both at rest and in transit.

② User Authentication

→ Limiting access to data and monitoring who access the data.

(3) Disaster and Data Breach

Contingency planning

Security issues in Cloud Computing

(1) Data breaches

→ it is an incident in which sensitive, protected or confidential data has potentially viewed ~~or~~ stolen or used by an individual unauthorized.

(2) Data Loss

→ valuable data disappear into the either without a trace.

(3) Account or Service traffic Hijacking

→ An attacker gains access to your account he or she can eavesdrop on your activities and redirect your clients to illegitimate sites.

(4) ~~Insecure~~ Insecure Interfaces and APIs

→ Cloud Computing providers expose a set of software interfaces or APIs that customers use to manage and interact with Cloud services.

(5) Denial of Service

→ DoS outages can cost service providers customers and prone. pricey to customers who are billed based on disk space consumed.

(6) Malicious insiders

→ It can be a current or former employee, a contractor or a business partner who ~~hacker~~ gains access to a network, system, or data for malicious purposes.

(7) Shared technology vulnerabilities

→ Cloud service providers share infrastructure, platforms, and applications to deliver their services in a scalable way.

Chapter 7

Cloud Computing Security Architecture

7.1 Architectural Considerations:

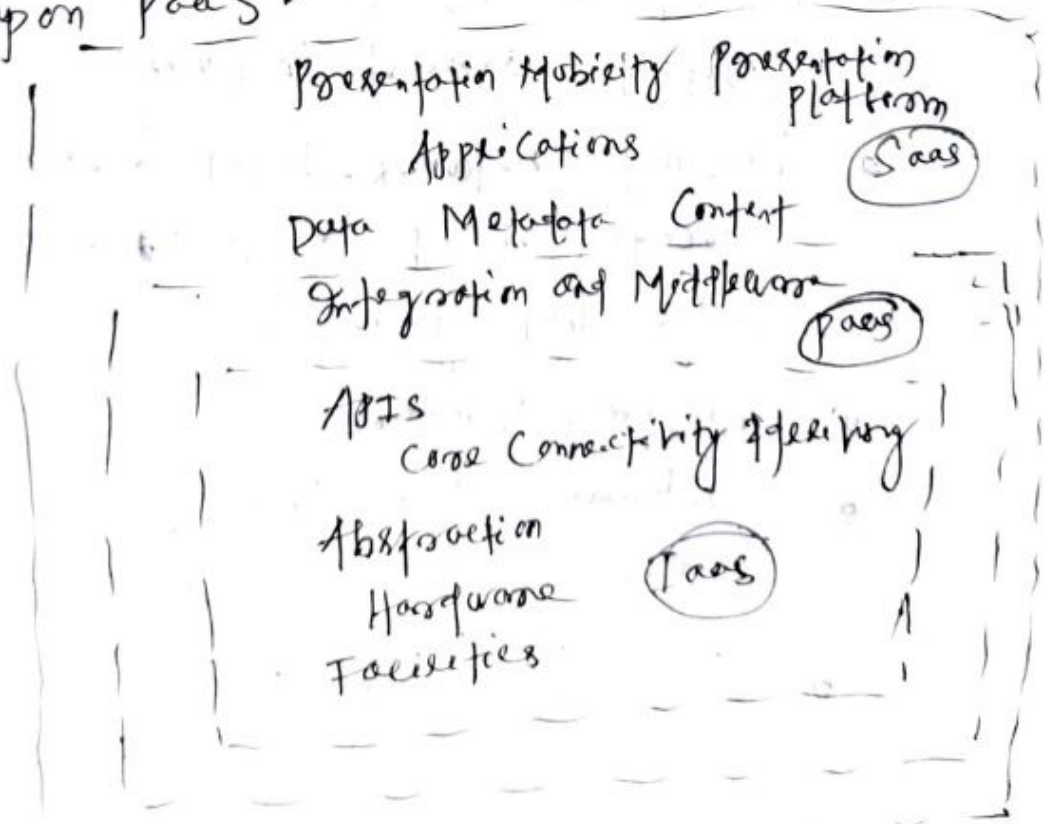
While all Cloud architecture models require performance management tools and strategy, the security architecture varies based on the type of cloud model

(1) Software-as-a-Service (SaaS)

(2) Infrastructure-as-a-Service (IaaS)

(3) Platform-as-a-Service (PaaS)

It is important to distinguish the different services models; as The Cloud Security Alliance notes: "IaaS is the foundation of all Cloud Services, with PaaS building upon IaaS, and SaaS in turn building upon PaaS."



IaaS Cloud Computing Security Architecture

→ This infrastructure provides the storage and networking components to cloud networking. It relies heavily on application programming interfaces (APIs) to allow enterprises to manage and interact with the cloud. However, Cloud APIs tend to be insecure as they are open and readily accessible on the network.

→ The Cloud Service provider (CSP) handles the security of the infrastructure and the abstraction layers.

IaaS cloud Computing Service models require these additional security features

- Virtual Web application firewalls placed in front of a Web site to protect against malware.
- Virtual network-based firewalls located at the cloud network's edge that guards the perimeter.
- Virtual routers.
- Intrusion Detection Systems and Intrusion prevention Systems (IDS/IPS)
- Network Segmentation.

SaaS Cloud Computing Security Architecture

→ SaaS Centrally hosts software and data that are accessible via a browser.

The enterprise normally negotiates with the cloud service provider the terms of security ownership in a legal contract.

→ Cloud Access Security Brokers (CASB) play a central role in dis Cressing security issues within a SaaS cloud service model as it logs, audits, provides access control, and oftentime includes encryption capabilities

Other security features for the SaaS cloud environment include

- Logging
- IP restrictions
- API gateways.

PaaS Cloud Computing Security Architecture

→ Cloud Security Architecture defines PaaS as the "deployment of applications without the cost and complexity of buying and managing the underlying hardware and software provisioning hosting capabilities"

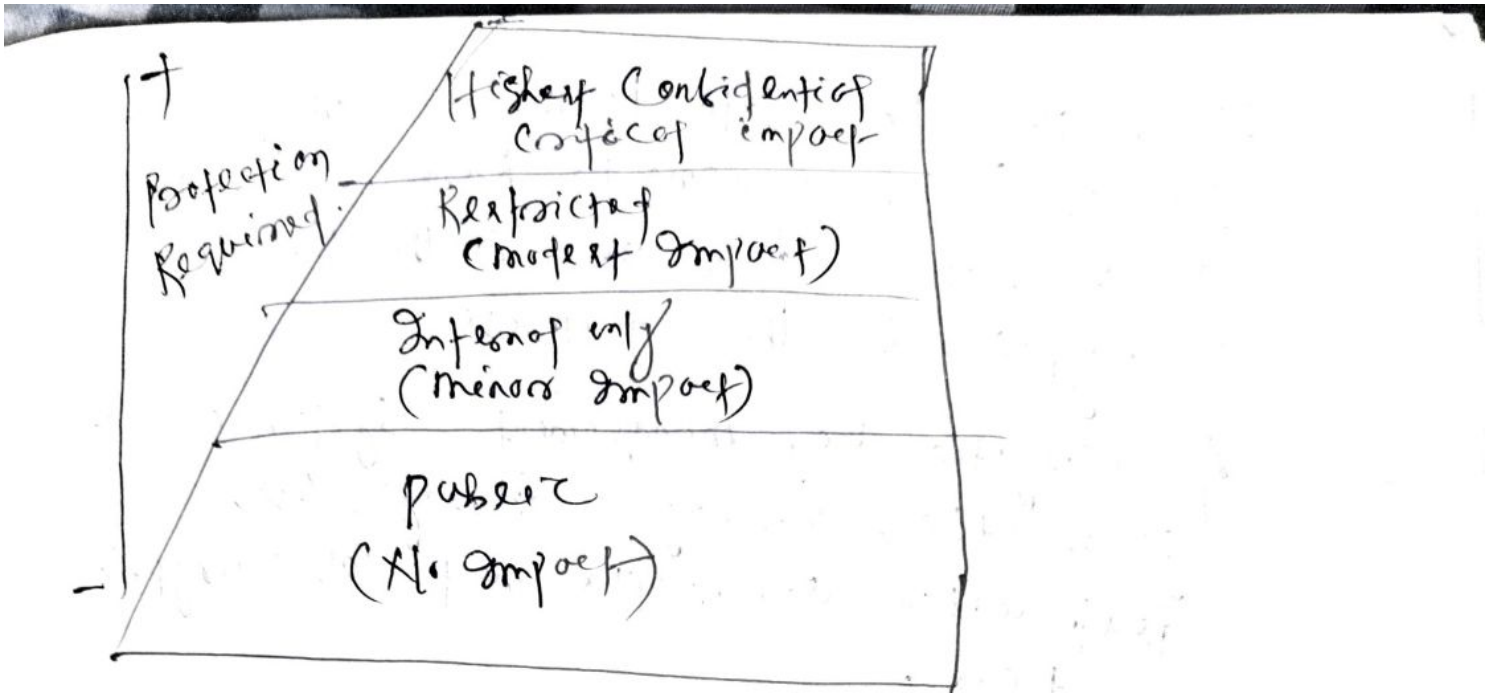
→ The Cloud Service Provider Secures a majority of a PaaS Cloud Service Model. However the Security of applications only with enterprise. The essential Components to Secure the PaaS Cloud include

- Logging
- IP restrictions.
- API gateways
- Cloud ~~Security~~ Access Security Brokers.

7.2 Information Classification

→ Information Classification is fundamental to asset management, risk assessment and the optimal use of Security Controls within the IT Infrastructure of any organization.

→ The aim of our information classification program is to help organizations improve the effectiveness and efficiency of Controls applied in protecting the Confidentiality, Integrity and Availability of the information



7.3 Virtual Private Networks

refers to the CNS Chapter-7 (7.4)

7.4 Public Key and Encryption Key Management

→ Encryption is a process that uses algorithms to encode data as ciphertext. The ciphertext can only be made meaningful again, if the person or application accessing the data has the data encryption keys necessary to decode the ciphertext. So, if the data is stolen or accidentally shared, it is indecipherable, thanks to data encryption.

→ Encryption key management is the administration of tasks involved with protecting, storing, backing-up, and organizing encryption keys.

→ High public data losses and regulatory compliance requirements have caused a dramatic increase in the use of encryption in the enterprise.

→ A encryption key management system includes generation, exchange, storage, use, destruction and replacement of encryption keys.

→ There are several encryption key management standards efforts underway. The best known is the Key Management Interoperability protocol (KMIP). The goal of KMIP is to allow users to attach any encryption device to a key management system.

7.5. Digital Certificates

~~Refer to~~ Refers to the CRLs Chapter-4(4.1)

7.6 Key Management

→ The main aim of Key management is to generate a Secret Key between two parties and store it to prove the authenticity between communicating users.

→ Key management is the techniques which support Key generation, Storage and maintenance of the Key between authorized users.

→ Key management plays an important role in cryptography as the basis for securing cryptographic goals Confidentiality, authentication, data integrity, and digital Signatures.

→ Basic purpose of Key management is Key generation, Key distribution, Controlling the use of Keys, updating, destruction of Keys and Key backup/recovery.

Following point to be executed in Key management

(1) User Registration
(2) User Initialization
(3) Key generation
(4) Key Installation

5. Key registration
6. Normal use
7. Key backup
8. Key update
9. Key de-registration and revocation.

7.7 Memory Costs

Home assignment.

7.8 Implementing Identity Management

→ Identity and Access Management encompasses the components and policies necessary to control and track user identities and access privileges. For IT resources, environment and systems.

→ • Specifically, identity and Access Management mechanism exist as system comprise of four main components.

1. Authentication
2. Authorization
3. User management
4. Credential Management.

(1) Authentication

Username and password combinations remains the most common forms of user authentication managed by Identity and Access Management. which can also support digital signatures, digital certificate, biometric hardware (fingerprint), specialized software (voice analysis)

(2) Authorization

The authorized component defined the correct granularity for access controls and oversees the relationship between identities, access control rights and IT resource availability.

(3) User Management

Related to the administrative capability of the system the user management program is responsible for new user identities and access groups, resetting the passwords, defining password policies and managing the privileges.

(4) Credential Management

This system establishes identities, access control rules for defined user accounts, which mitigates the threat of insufficient authorization.

→ The Identity Access Management mechanism is primarily used to counter the insufficient authorization, denial of service, and overlapping trust boundaries threats.

7.9 Controls and autonomic System

Autonomic computing

→ it is initiated by IBM

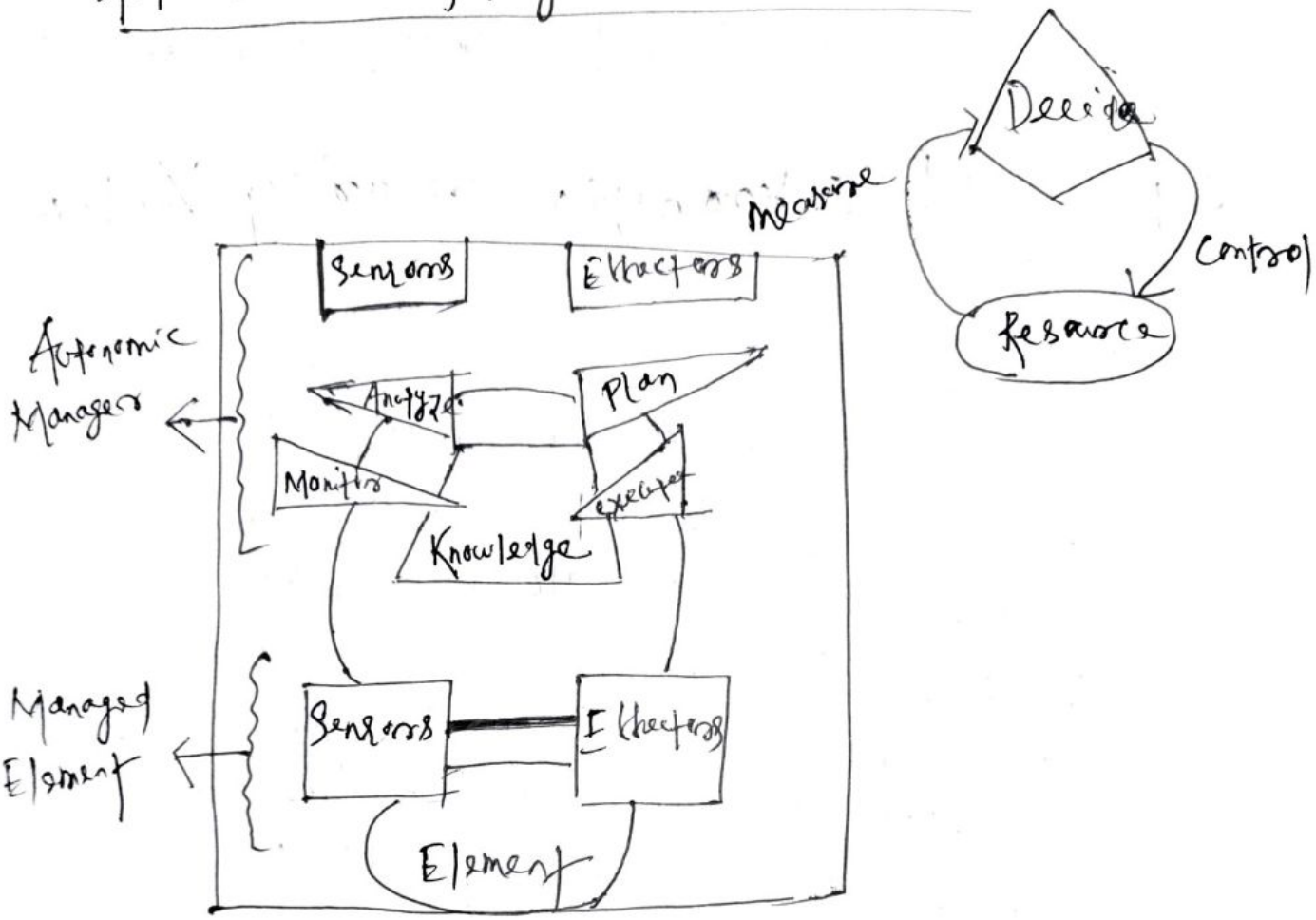
→ it is a System that supports computing to perform and work without any outer control

→ Aim is to make the computer carry out critical functions without any human intervention.

→ Key elements of autonomic Computing

- Knows Configure itself
- optimizes itself
- Heal itself
- protect itself
- Adapt itself.

Autonomic Computing Architecture



What is the Memory Card?

A memory card is a type of storage device that is used to *store videos, photos, or other data files*. It offers a volatile and non-volatile medium to store data from the inserted device. It is also referred to as a **flash memory**. Commonly, it is used in devices like phones, digital cameras, laptops, digital camcorders, game consoles, MP3 players, printers, and more.

History of Memory Card

The flash memory is the basis for memory card technology, which was invented by **Fujio Masuoka** at Toshiba in **1980**. Later in **1987**, it was commercialized by Toshiba.

Types of Memory Cards

There are several types of memory cards in the market, most commonly used types of memory cards are given below:

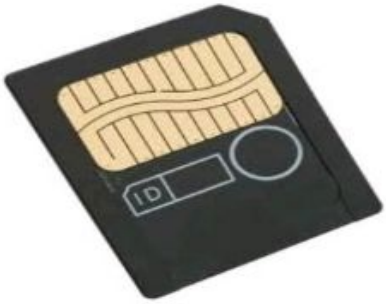
- SD Card
- MicroSD
- SmartMedia Card
- Sony Memory Stick
- CF (CompactFlash)
- xD-Picture Card
- SDHC Card
- MMC

SD Card: It is one of the most common types of memory cards, stands for **Secure Digital card** that is designed to provide high-capacity memory in a small size. Mainly, it is used in numerous small portable devices such as handheld computers, digital video camcorders, digital cameras, mobile phones, etc. Approximately, more than 8000 different models and over 400 brands of electronic equipment use SD technology. It measures 32 x 24 x 2.1 mm and weighs approximately 2 grams and is considered a standard for the industry due to widespread use.

MicroSD: It is a type of removable flash memory card that is also known as T-Flash or TransFlash used for storing information. SanDisk developed the first microSD card and approved as a standard on 13 July 2005. It is often used with mobile phones and other mobile devices that are available in sizes from 128 MB to 4 GB.

Some of the laptops include a feature of the MicroSD slot that allows users to insert a MicroSD to download data or files on the laptop. If you have a desktop computer or your laptop has no MicroSD slot, you can use a media card reader that also allows you to view data on the MicroSD card, and transfer that data to the computer.

SmartMedia card: It is a type of memory that comprises of a Flash-Memory chip that stores data. Toshiba developed the first SmartMedia card and had a smaller storage capacity from 2 MB to 128 MB. It has a single NAND flash chip that is embedded in a thin plastic card. It is the smallest memory card, only 0.76mm thick, and easy to maintain a favorable cost than others.



Sony Memory Stick: It is a type of flash memory that was introduced by Sony in October 1998. It is used with Sony's digital cameras and other types of electronics for storing data. Almost all of Sony's products that use flash media use a memory stick as it is a proprietary Sony product. Sony released different kinds of memory stick as well as Memory Stick Micro, Memory Stick PRO, Memory Stick Duo, Memory Stick PRO Duo, and Memory Stick PRO-HG. The memory stick is designed with storage from 4 MB to 256 GB and a maximum capacity of 2 TB.

CompactFlash: It is a very small removable mass storage device that is commonly found in PDAs, digital cameras, and other portable devices. SanDisk Corporation invented the CompactFlash memory card in 1994. It is a 50-pin connection storage device that supports 3.3V and 5V operation and relies on flash memory technology. It does not need a battery to retain data indefinitely. The storage capacity of the CF card is large, that is from 2 MB to 128 GB.

xD-Picture Card: It is a flash memory card designed for use in many models of digital cameras. In 2002, it was developed by Olympus and Fuji film. The size of xD (Extreme Digital) Picture Card is 20mm x 25mm x 1.7mm, and its capacity for the original version is up to 512 MB, and for the type H and M/M+ versions up to 2 GB.

SDHC Card: It stands for **Secure Digital High Capacity**, based on the SDA 2.00 specification. It is an extended version of the standard SD card having storage capacity up to 32 GB. The SDHC works differently as compared to the standard SD card as it uses new technology. Furthermore, it provides different data transfer speed for consumers by using below three-speed class system:

- Class 2 - minimum sustained DTS of 2MB/sec
- Class 4 - minimum sustained DTS of 4MB/sec
- Class 6 - minimum sustained DTS of 6MB/sec

MMC (MultiMediaCard): It is a tiny memory card as flash memory, developed by SanDisk and Siemens AG/Infineon Technologies AG. It is used to make storage portable among several devices, such as car radios, cell phones, digital cameras, car navigation systems, PDAs, printers, music players, cellular phones, video camcorders, and personal computers. It is much like to the SD card, and smaller as compared to older memory card formats, such as CompactFlash and SmartMedia card. The MMC provided storage capacities up to 128 MB until October 2002.

Advantages of Memory Card

1. Increased Storage
2. Cost-Effective
3. Reduce Phone Memory Consumption
4. Removable & Portable
5. Non-volatile Memory
6. Easily Accessed on PC

Disadvantages of memory card

1. Break Easily
2. Low-class Card Can reduce Phone Performance
3. Slower than Primary Memory
4. Apps Disappear after Removing It

Chapter - 8

Market Based management of Clouds

8.1 Cloud Information Security Vendors

(1) **Datadog**

→ Datadog Security monitoring detects Cloud Security threats in real-time across your applications, network, and infrastructure.

→ It investigates Security threats and provides detailed data through metrics, traces, logs, etc.

(2) **HackerOne**

HackerOne is the no.1 hacker-powered security platform, helping organizations find and fix critical vulnerabilities before they can be exploited.

(3) **Intruder**

Intruder helps Organizations to reduce their attack exposure by providing an all-in-one Cybersecurity Solution.

(4) Cipher

Cipher can protect your internet-connected services and devices.

(5) Sophos

Sophos is a hardware and Software Security company that provides Co-ordinated Security between firewalls and the end points with real-time aptitude.

(6) Hytrust

→ Hytrust is a Cloud Security Automation company that has the automated the Security controls related to networking, computing etc

→ Hytrust offers various services like cloud and virtualization Security, cloud encryption, encryption Key management, automated compliance etc.

(7) Cipher Cloud

CipherCloud is a privately held leading Cloud Security Company that protects your data flawlessly and more effectively by incorporating data monitoring & protection, risk analysis and cloud detection.

(8) Portnoy

Portnoy is a Network Security and Compliance Company that offers enterprise and Corporate level cloud based encryption Solutions.

(9) NetSkope

NetSkope is a Cloud Security Company that uses some patented technology to provide Security over various networks like remote, Corporate, mobile, etc.

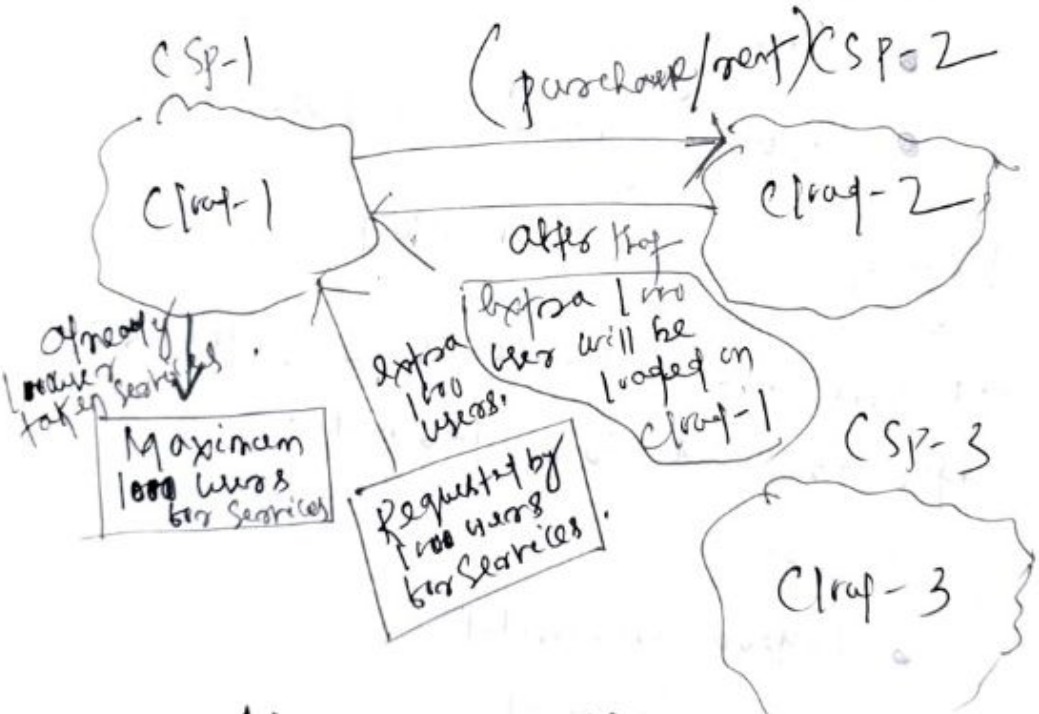
(10) Twistlock

Twistlock is a privately held information technology and Services Company that provides uninterrupted and end-to-end Security for Containerized applications.

S:3 Cloud Federation, Characterization

Cloud Federation is the practice of interconnecting the Cloud Computing environment of two or more Cloud Service providers for the purpose of load balancing and providing Services to the users.

Example



Advantages of Cloud Federation

1. Load balancing & Capacity management.
2. Scaling data to other Cloud Service providers
3. efficient use of resources.
4. prevention from failures.
5. prevention from inter-locking

8.3. Cloud Federation Stack

Cloud Federation Stack is also known as Level of Cloud Federation

(1) Conceptual Level

At this level various needs to join the Cloud Federation is defined.

Features

- motivations
- Advantages
- Trust agreement between Service Providers.

(2) Logical & Operational level

At this level policies & rules for interoperation are defined.

Features

- Federations model
- CSP agreement
- Market & pricing model.
- Service level agreement.

(3) ~~This level~~ Intrastructural Level

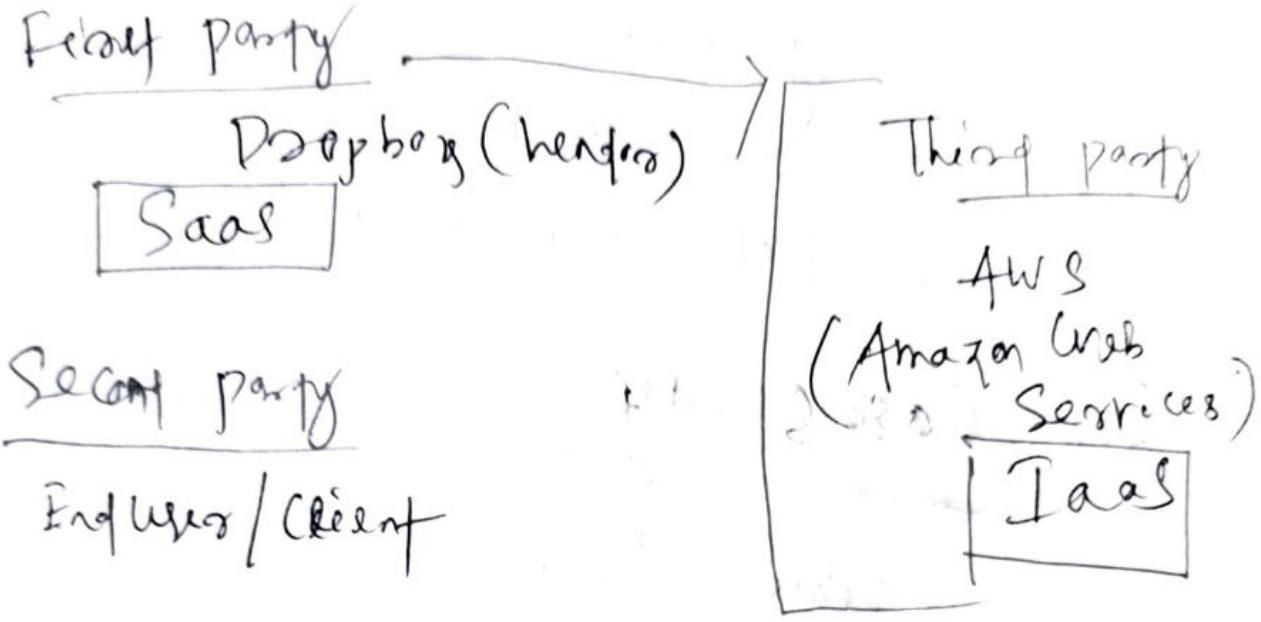
This level addresses the technical challenges involved in enabling heterogeneous cloud.

Deterrent issue

- protocol, interfaces & standards
- programming Interoperations.
- Federation Platforms.
  (Intercloud, RESERVOIR)

8.4 Third Party Cloud Services

Example



The third party Cloud Services is the Services in which user wants to acquire when he/she is not getting that Service with acquired or hired Cloud Service provider.

Some of the third party Cloud Services are

1. Amazon Web Services.
2. Microsoft Azure.

Advantages

1. Maintenance & Support.
2. Skilled Company with all the resources
3. Security Benefits.
4. Cost advantages.

Disadvantages

1. Security worries.
2. Lack of Control
3. potential Cost drawback

8.5 Case Study

(1) Microsoft Azure

→ previously Cridgous Azure

→ Supports IaaS and PaaS

→ Supports extensive Set of Services to quick create, deploy and manage applications.

→ many programming languages and frameworks are supported.
→ Available across a worldwide Microsoft-managed datacenters.

Azure Services

→ Compute
→ Mobile Services
→ Storage Services
→ Data management
→ messaging
→ Media Services
→ Content Delivery Network (CDN)
→ Developer
→ Management
→ Machine learning.

(2) Amazon Web Services (AWS)

→ AWS is a comprehensive, evolving cloud computing platform provided by Amazon.
→ Supports IaaS, PaaS, SaaS.
→ AWS services can offer an organization tools such as compute power, database storage and content delivery services.

→ AWS launched in 2006 from the infrastructure that Amazon.com built to handle its online retail operations.

→ AWS was one of the first companies to introduce a pay-as-you-go cloud computing model that scales to provide users with compute, storage or throughput as needed.

AWS Services

→ Compute
→ Storage databases
→ Data Management
→ Migration
→ Hybrid cloud
→ Networking
→ Development tools
→ Management
→ Monitoring
→ Security
→ Governance
→ Big data Management
→ Artificial intelligence
→ Mobile development
→ Message & Notification

VPN

Virtual private Network

1. What is VPN ?
2. Why you should use a VPN ?
3. How Does a VPN works ?
4. Types of VPN
5. Advantages.
6. Disadvantages.

(1) What is VPN?

→ VPN stands for Virtual Private Network

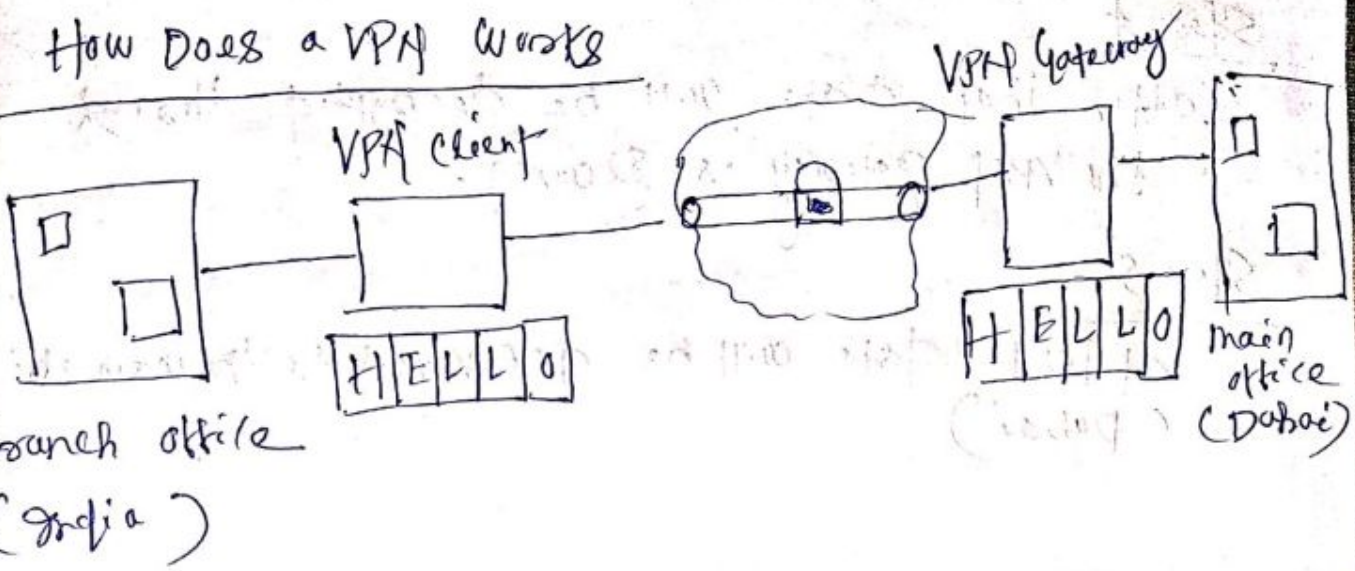
→ VPN is an encrypted connection over Internet from a device to network.

→ The encrypted connection helps ensure that sensitive data is safely transmitted.

(2) Why you should use a VPN?

1. Browse the web securely on public wifi
2. Save money shopping online
3. Automatically Encrypt everything
4. Improve online Gaming Speed
5. Enjoy private and Secure voice chat
6. Complete Sensitive Research without Interference
7. Completely private Collaboration

(3) How Does a VPN Works



Step-1

Branch office in India transfer the data through the help of VPN Client.

Step-2

VPN Client uses the encryption technique

Step-3

Data will be passed through the tunnel which is secured. and that tunneling uses the two protocol

(1) PPTP (Point to Point Tunneling Protocol)
(2) IPSec (Internet Protocol Security)

Step-4

after that data will be decrypting through the VPN gateway is secured.

Step-5

at last data will be delivered to the main office (pune)

④ Types of VPN

Virtual Private Network is basically of 2 types

(1) Remote access VPN
(2) Site to Site VPN

(1) Remote access VPN

→ Remote Access VPN permits a user to Connect to a private network and access its services and Resources remotely.

→ The connection between user and private network occurs through the Internet and Connection is Secure and private.

(2) Site to Site VPN

→ A Site-to-Site VPN is also called as Router to Router VPN and is commonly used in the large Companies or organizations.

→ use Site-to-Site VPN to Connect the network of one office location to the network of another office location and it also two types.

(a) Intranet based VPN

when several offices of the same Company are Connected using Site to Site VPN type it is called Intranet based VPN

(4) Extranet based VPN

When Companies use site to site VPN type to connect to the office of another Company it is called Extranet based VPN.

(5) Advantages

(1) Enhance Security
(2) Remote Control
(3) Share Files
(4) Unblock websites and bypass filters
(5) Better Performance
(6) Reduce Costs.

(6) Disadvantages

(1) it might be difficult to set up for Business users.
(2) it might add more cost to your network connection
(3) it can slow down your internet speed.

Chapter - 9

Hadoop

9.1 Introduction : (Hadoop is developed by Apache)

Hadoop is an

open Source

Software framework

which provides huge Data Storage

" Hadoop is an open Source Software framework which huge data Storage facility ".

→ Hadoop is a framework for large-scale processing.

→ Hadoop written in Java and Shell Scripts.

Advantages

→ Computing power
→ Storage
→ Fault Tolerance
→ Flexibility
→ Low Cost
→ Scalability

Disadvantages

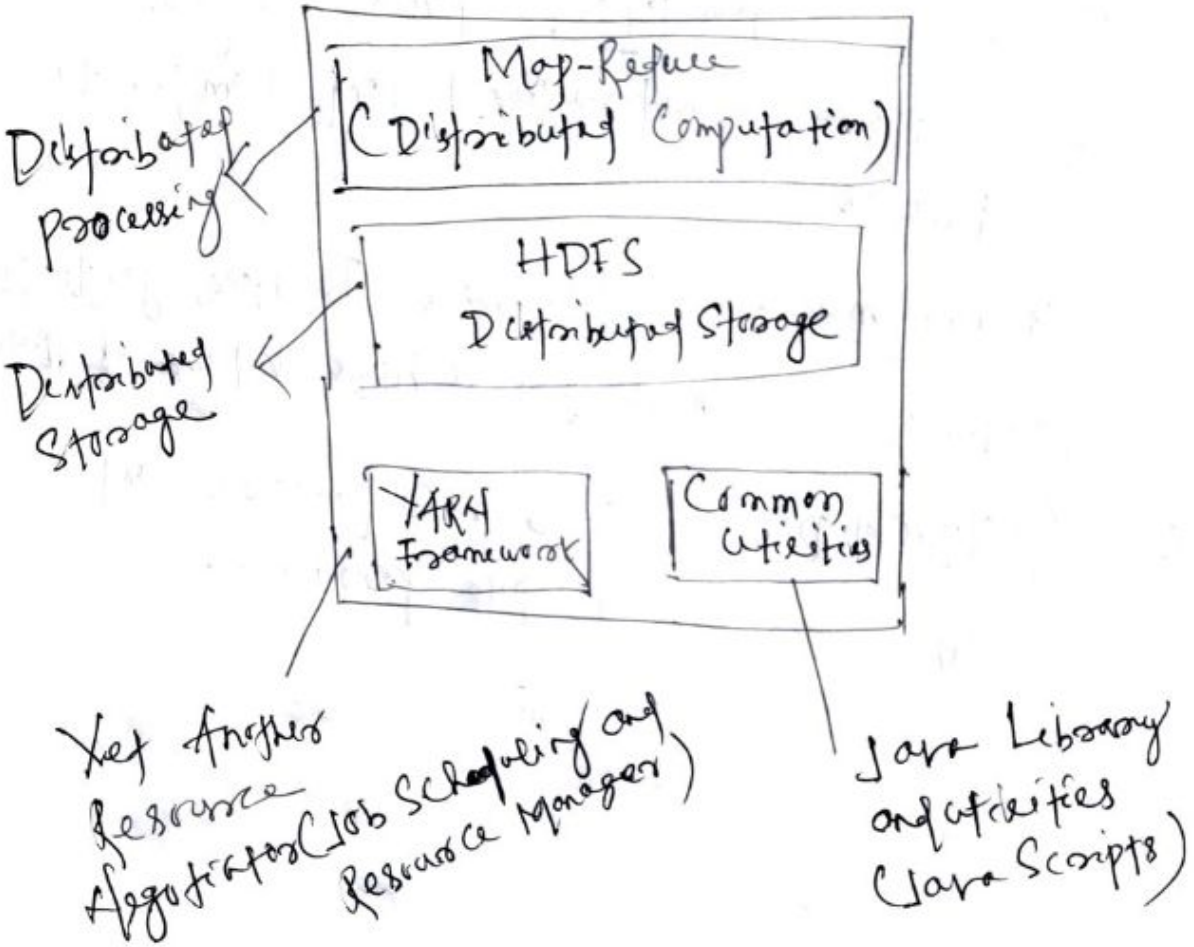
→ Not fit for small data
→ Security Concerns.
→ Programming model is restrictive
→ Joins of multiple datasets is complex & slow.

Hadoop Architecture

Hadoop has two major layers namely

(1) Processing / Computation layer (Map-Reduce)

(2) Storage layer (Hadoop Distributed File System)



(1) MapReduce

Map Reduce is a parallel programming model for writing distributed applications devised at google for efficient processing of large amounts of data (Multi-terabyte data-sets) on large clusters (thousands of nodes), fault-tolerant manner.

(3) HDFS (Hadoop Distributed File System)

→ The HDFS is based on the Google File System (GFS) and provides a distributed file system that is designed to run on commodity hardware.

→ It is highly fault-tolerant and is designed to be deployed on low cost hardware

→ It provides high throughput access to application data and is suitable for applications having large datasets.

Hadoop framework also includes the following two modules.

- Hadoop Common / Common utilities

These are Java libraries and utilities required by other hadoop modules.

- Hadoop YARN

This is a framework for job scheduling and cluster resource management.

How Does Hadoop Work

Steps

(1) → Data is initially divided into directories and files. Files are divided into uniform sized blocks of 128M and 64M (preferably 128M)

(2) → These files are then distributed across various cluster nodes for further processing.

(3) → HDFS, being on top of the local file system, supervises the processing.

(4) → Blocks are replicated for handling hardware failure.

(5) → Checking that the code was executed successfully.

(6) → Performing the sort that takes place between the map and reduce stages.

(7) → Sending the sorted data to a certain computer.

(8) → Writing the debugging logs for each job.

Big Data

→ Big Data is defined as data that is huge in size. Big data is a term used to describe a collection of data that is huge in size and yet growing exponentially with time.

→ Big Data analytics examples includes stock exchanges, Social media sites, Jet engine, etc.

9.2 Data Sources

→ Due to the capability of processing variety of data and volume of data, data sources for hadoop has increased and along with that the complexity has increased enormously.

→ Let's look at some of the data sources which can produce enormous volume of data or consistent data continously.

(1) Data Sensors: These are the thousands of sensors, producing data continously.

(2) Machine Data: produce data which should be processed in near real time for analysing huge loss.

(3) Telco Data: CDR (Call Detail Record) and other telecom data generates high volume of data.

(4) Healthcare System data: Genes, images, ECR records are unstructured and complex to process.

(5) Social media: Facebook, Twitter, google plus, YouTube and others get a huge volume of data.

(6) Geological Data: Semi conductors and other geological data produce.

9.3 Data Storage and analysis

→ The problem is simple: while the storage capacities of hard drives have increased massively over the years, access speeds the rate at which data can be read from drives. have not kept.

→ one typical drive from 1990 could store 1370MB of data and data transfer speed of 4.4MB/s, so you could read all the data from a full drive in around five minutes.

→ over 20 years later, one terabyte drives are the norm but the transfer speed is around 100MB/s, so it takes more than two and a half hours to read all the data off the disk.

→ This is a long time to read all data on a single drive and writing is even slower. The obvious way to reduce the time is to read from multiple disks at once. Imagine if we had 100 drives, each holding one hundredth of the data. Working in parallel we could read the data in under two minutes.

→ This parallel nature is working in mapreduce of hadoop concept. and also working HDFS (Hadoop Distributed File System).

→ The storage is provided by HDFS and analysis by MapReduce in hadoop concept.

9.4 Comparision with other System

Difference between Relational Database Management System (RDBMS) and Hadoop.

RDBMS	Hadoop
1. Traditional row-column based databases, basically used for data storage, manipulation and retrieval	1. An open-source software used for storing data and running applications or processes concurrently.

RDBMS

2. in this structured data is mostly processed
3. it is best suited for OLTP (online Transaction processing) environment
4. it is less Scalable than Hadoop
5. Data normalization is required in RDBMS
6. it Stores transformed and aggregated data
7. it has no latency in response
8. The data Schema of RDBMS is Static type.
9. High Data integrity available
10. Cost is applicable for licensed software

Hadoop

2. in this both Structured and unstructured data is processed.
3. it is best Suited for BIG data.
4. it is highly Scalable
5. Data normalization is not required in Hadoop.
6. it Stores huge volume of data.
7. it has some latency in response.
8. The data Schema of Hadoop is dynamic type
9. Low data integrity available than RDBMS
10. Free of cost, as it is an open Source Software