Unit-5
Cloud Security
Part-1

Part-1

- Cloud Security Fundamentals
- Cloud Risk
- Cloud Risk Division- Policy and Organizational Risks, Technical Risks, Legal Risks, Other Risks
- Cloud Computing Security Architecture
- VM Security Challenges

Cloud Computing

Cloud computing refers to the on demand delivery of computing services such as applications, computing resources, storage, database, networking resources etc. through internet and on a pay as per use basis. At the present time the demand for cloud computing services are increasing with respect to that demand for cloud computing skills is also increasing. It provides three main types of service models i.e.SaaS (Software as a Service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service). With this as starting from small to large organizations have started using cloud services so depending upon their requirement they go for the different types of cloud like Public cloud, Private cloud, Hybrid cloud, Community cloud.

Cloud Security Fundamentals

Cloud computing which is one of the most demanding technology of the current time, starting from small to large organizations have started using cloud computing services. Where there are different types of cloud deployment models are available and cloud services are provided as per requirement like that internally and externally security is maintained to keep the cloud system safe. Cloud computing security or cloud security is an important concern which refers to the act of protecting cloud environments, data, information and applications against unauthorized access, DDOS attacks, malwares, hackers and other similar attacks.

Community Cloud : These allow to a limited set of organizations or employees to access a shared cloud computing service environment.

- Cloud security is the first and foremost concern of every industry using cloud services. A cloud vendor must ensure that the customer does not face any difficulties such as loss of data or data theft. There is a possibility that a malicious user can go through the cloud by impersonating a legal user, thereby infecting the cloud services and hence affecting various customers sharing the malicious cloud services.

Cloud Risk

When infrastructure, applications, data and storage are hosted by cloud providers, there is a huge chance of risk in each type of service offering. This is known as cloud risk.

Organizations such as the Cloud Security Alliance (CSA) offer certification to cloud providers that meet their criteria. The CSA's Trusted Cloud Initiative program was created to help cloud service provider enable industry-recommended standards, secure access, compliance management, interoperable identity and follow best practices.

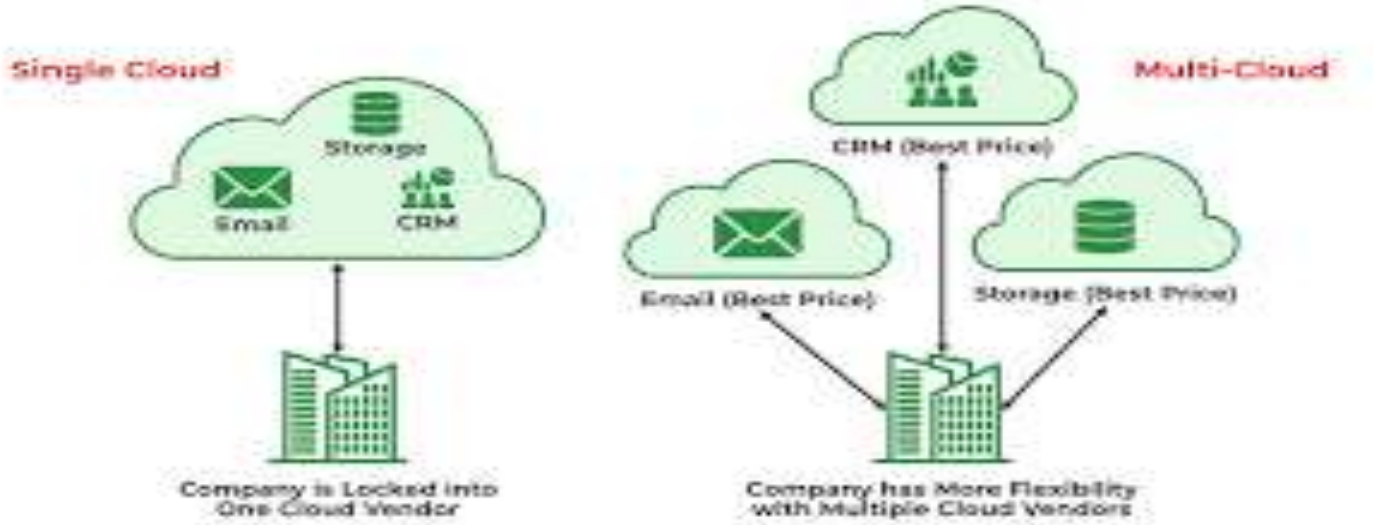
Cloud Risk Division

Cloud Risks can be divided into the following four major categories:

1. Privacy and organizational risks
2. Technical risks
3. Legal risks
4. Other risks

Privacy and organizational Risks

1. Lock-in: Cloud lock-in(also known as vendor lock-in or data lock-in) occurs when trasitioning data, products or services to another vendor’s platform is difficult and costly, making customers more dependent (locked-in) on a singlecloud storage solution.



2 Loss of governance: The loss of governance in cloud computing occurs when businesses migrate workloads from an exclusively on-premises IT infrastructure to the cloud without a suitable governance policy in place.

3 Compliance challenges: Cloud compliance is the art and science of complying with regulatory standards of cloud usage in accordance with industry guidelines and local, national, and international laws. (certification)

4 Cloud service termination or failure: There must be 24x7 support and high availability of all services, but in the competitive world of IT, an adequate business strategy, lack of financial support and other factors could lead some providers to go out of business or shut down their service portfolio offering. And it is possible that for a short or medium period of time some cloud computing services could be terminated.

5 Supply Chain Failure: Supply chain failure as a breakdown caused by either the internal operations or external suppliers that cause significant quality, delivery or cost impact to your company and/or customers. Many times, blame is place on outside suppliers or contract manufacturers without proper root cause analysis.

Technical Risks

1. Isolation failure: multi-tenancy and shared resources are defining characteristics of cloud computing. This risk category covers the failure of mechanisms separating storage, memory, routing and even reputation between different tenants.
2. Resource exhaustions : Resource exhaustion attacks generally exploit a software bug or design deficiency. In software with manual memory management (most commonly written in C or C++), memory leaks are a very common bug exploited for resource exhaustion.
3. Cloud provider malicious insider: malicious insider is **an insider who intends to cause damage to the organization for personal gain**. Because of their access and knowledge of the organization's most valuable assets, attacks involving malicious insiders are harder to identify and remediate than those that originate from outside the organization.

4 Intercepting data in transit: The data is vulnerable while it is being transmitted. Data can be intercepted and compromised as it travels across the network where it is out of a user's direct control. For this reason, data should be encrypted when in transit. Encryption makes the data unreadable if it falls into the hands of unauthorized users.

5 Insecure or ineffective deletion of data: When it comes to deleting or completely destroying old data from your computer, laptop, hard drive or other media devices, it is vital to keep safety and security the main priorities. Many people and even companies often use unsafe methods to destroy or erase confidential data. Simply deleting or reformatting your computer may not be secure or safe enough. Continuing to practice poor data destruction methods will inevitably lead to identity theft and data breaches.

6 Conflicts between customer hardening procedures and cloud environment:

Cloud providers follows different servers or different hardening mechanisms that are little different from traditional server hardening procedures.

7 Loss of encryption keys: This includes disclosure of secret keys(e.g file encryption, Customer private keys) or passwords to malicious parties, the loss or corruption of those keys.

8 Malicious probes or scams: Malicious probes or scams are indirect threats to the assets being considered. They can be used to collect information in the context of a hacking effort. A probable impact could be a loss of confidentiality, integrity and availability of service and data.

9 Compromise service engine: cloud provider rely on specific service engine that is placed on top of physical hardware. For IaaS, this can be hypervisor. For PaaS, it can be hosted application. Hacking the service engine may be useful to escape the isolation.

Legal Risks

1. Risk from changes of jurisdiction: Customer data may be kept in several jurisdictions, some of which may be high risk. If data centres are located in high-risks countries, sites could be attacked by local authorities and data or systems subject to enforced disclosure or seizure.
2. Licensing risks: Licensing conditions, such as per-seat agreements and online licensing checks unstable in a cloud environment.
3. Data protection risks: It can be tough for the cloud customer to efficiently check the data processing that the cloud provider brings out and hence be sure that data is handled in a lawful way.

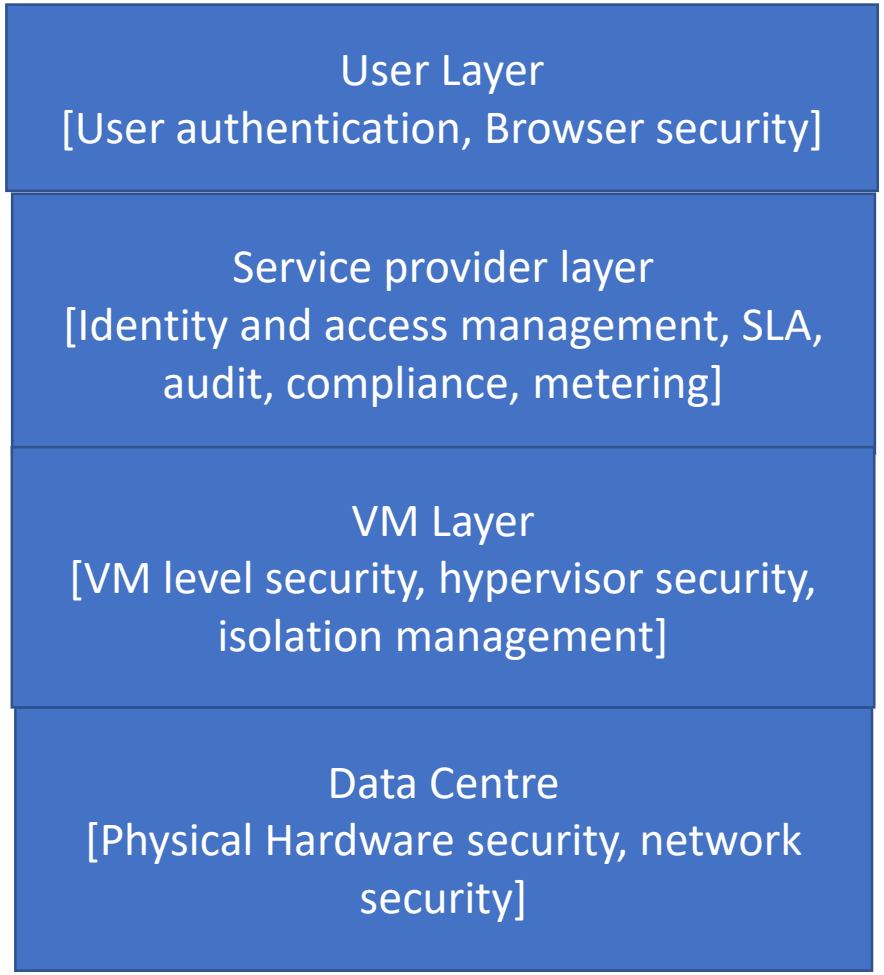
Other Risks

1. Backup lost or stolen: This risk is possible due to inadequate physical security procedures, AAA vulnerabilities, user provisioning vulnerabilities and user de-provisioning vulnerabilities.
2. Unauthorized access to premises: Because of inadequate physical security procedures, unauthorized access in datacentres is possible is possible. Generally, cloud providers have large datacentres; therefore, physical control of a datacentre must be stronger because the impact of a breach of this issue could be higher.
3. Theft of computer equipment: This risk is possible because of inadequate physical security procedures. This risk is mainly related to the datacentres, and dual authentication mechanism should be followed to accesses those machines.
4. Natural disasters: Natural disasters are possible any time so there must be a perfect disaster recovery plan. Although, the risk from natural disasters is quite less compared to traditional infrastructures because cloud providers offer redundancy and fault tolerance by default; for examples, AWS has various physical regions and multiple availability zone option within a region also.

Cloud Computing Security Architecture:

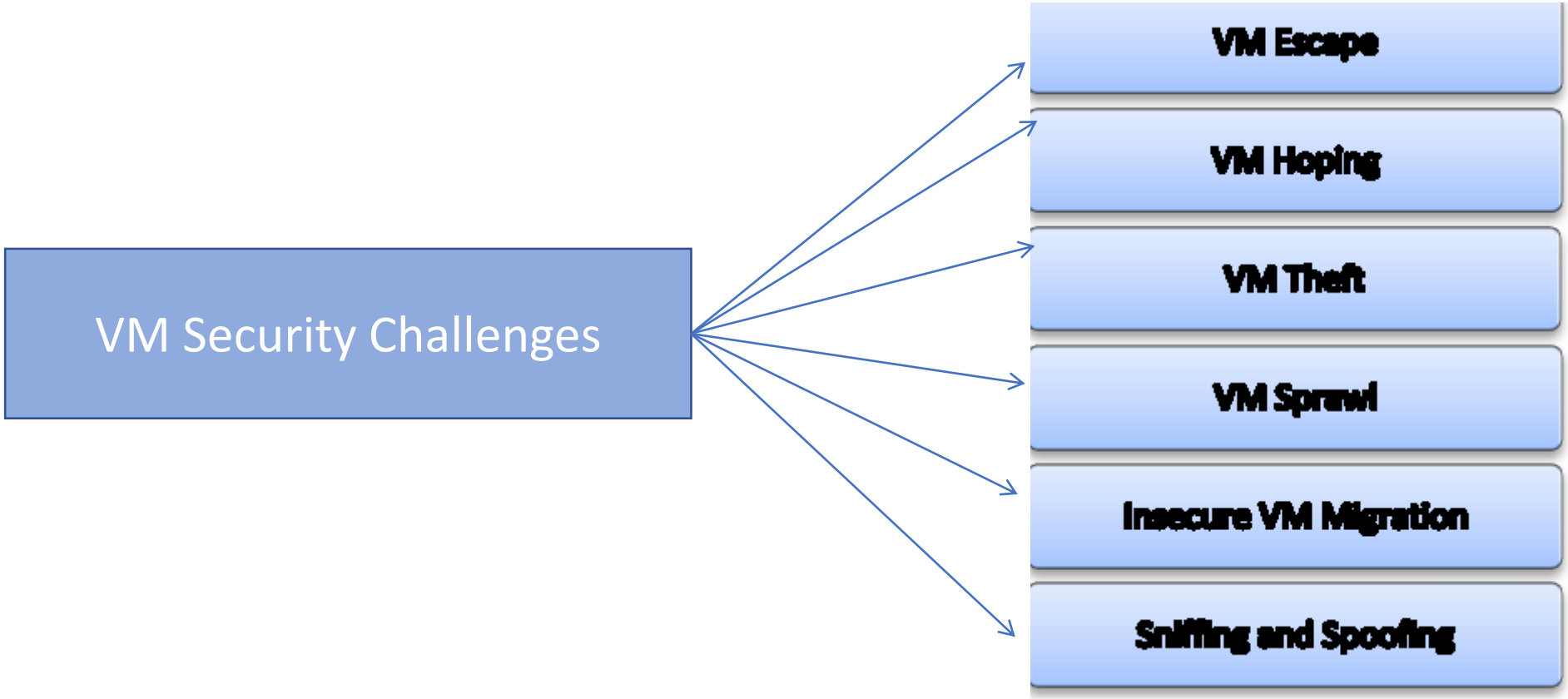
Architecture view of the security issues to be addressed in a cloud computing environment for providing security to the customer. This architecture defined four layers on the basis of cloud computing services categorization.

Four Layers of security Architecture:

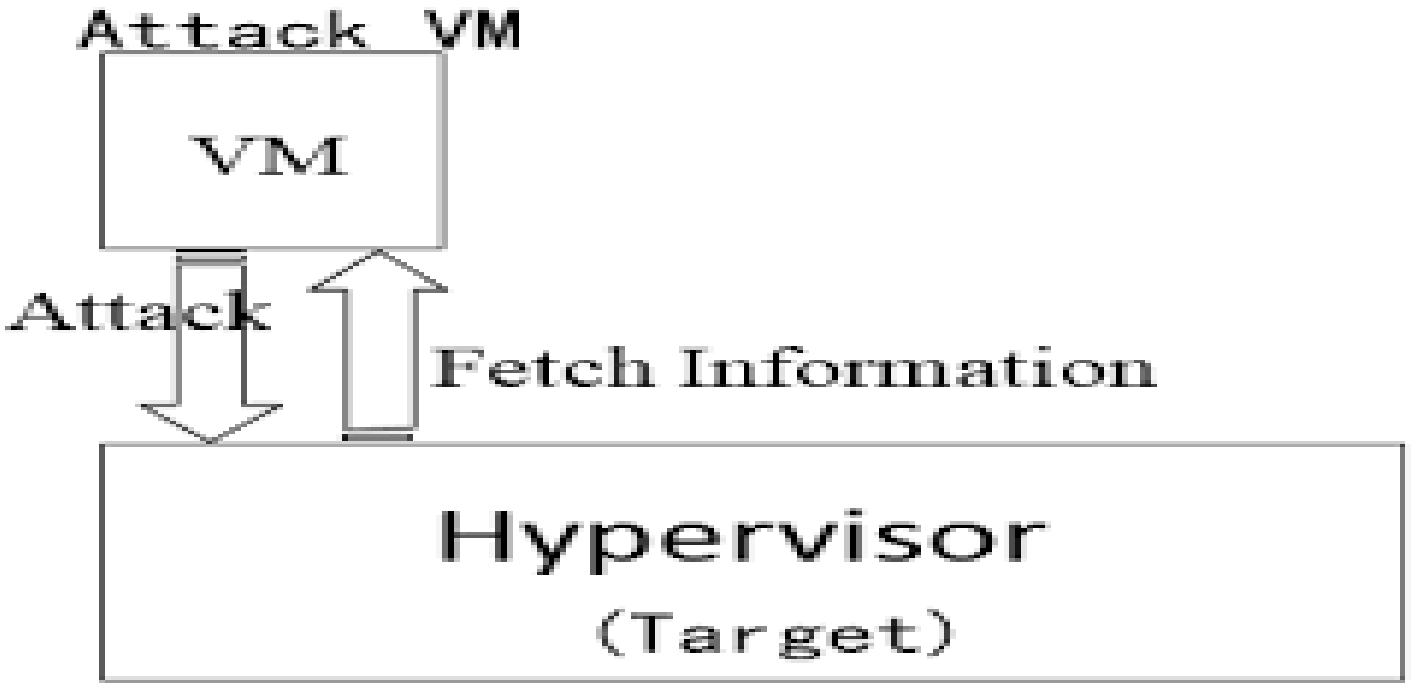


- Data Centre Layer: This layer is related to traditional infrastructure security concerns. It consists of physical hardware security, theft protection, network security and all physical assets security.
- VM Layer: This layer involves VM level security issues, VM monitoring, hypervisor-related security issues and VM isolation management issues.
- Service provider layer: This layer is responsible for identity and access management, service level agreement (SLA), metering, compliance and audit- related issues.
- User layer: This is the first layer of user interaction. It is responsible for user authentication and authorization and all browser- related security issues.

VM Security Challenges:



- **Virtual machine escape:-** It is an exploit in which the attacker runs code on a VM that allows an operating system running within it to break out and interact directly with the hypervisor. Such an exploit could give the attacker access to the host operating system.



VM Hoping: VM hopping is a common attack mode in virtualization security attacks. It means that an attacker attempts to gain access to other virtual devices on the same Hypervisor based on one virtual machine, and then attacks it.

Virtualization sprawl: It is a phenomenon that occurs when the number of virtual machines (VMs) on a network reaches a point where administrators can no longer manage them effectively. Virtualization sprawl is also referred to as virtual machine sprawl, VM sprawl or virtual server sprawl.

- **Insecure VM migration**: A workload cannot migrate to a destination server if it does not have the computing resources required to support it. Migration problems can occur when the destination server lacks adequate processor cores, memory space or NIC ports or has a storage shortage, and cannot reserve resources for the new workload.

- **Sniffing and Spoofing**: Spoofing is when an attacker creates TCP/IP using another person's IP address. A sniffer software is placed between two interactive endpoints in packet Sniffing, where the attacker pretends to be one end of the connection to the target and snoops on data sent between the two points.

Unit-5

Cloud Database

Part-2

Contents

- Cloud Database- Operational Model for Cloud Database, Types of Cloud Database

- Cloud File System- Distributed File System Basics, Concept of GFS and HDFS, Comparison of Features

Cloud Database

A cloud database is a database service built and accessed through a cloud platform. It serves many of the same functions as a traditional database with the added flexibility of cloud computing.

Key features:

- A database service built and accessed through a cloud platform
- Enables enterprise users to host databases without buying dedicated hardware.
- There are some SQL-based and some NoSQL- based databases offering.
- Accessed through a web interface or vendor-provided API.

Operation Model for Cloud Database

There are two primary methods to run a database on a cloud platform:

- **Virtual Machine image:** Cloud platforms allow users to purchase virtual-machine instances for a limited time, and one can run a database on such virtual machines. Users can either upload their own machine image with a database installed on it, or use ready-made machine images that already include an optimized installation of a database.

- **Database-as-a-service (DBaaS):** With a database as a service model, users pay fees to a cloud provider for services and computing resources, reducing the amount of money and effort needed to develop and manage databases.Users are given tools to create and manage database instances, and control users. Some cloud providers also offer tools to manage database structures and data. Many cloud providers offer both relational (Amazon RDS, SQL Server) and NoSQL (MongoDB, Amazon DynamoDB) databases. This is a type of software as a service (SaaS).

Architectural and common Characteristics

- **Fast Deployment**: Cloud databases are the perfect choice when you urgently need a database, as they can be up and running in minutes. Cloud databases eliminate the need to purchase and install hardware and set up a network.
- **Accessibility**: Users have quick access to cloud databases remotely through the provider's API or web interface.
- **Scalability**: You can expand cloud database storage capacity without disruptions and meet the requirements. Cloud database scalability is seamless due to DBaaS implementation, which is a major benefit for growing businesses with limited resources.
- **Disaster Recovery**: Data backups are regularly performed on cloud databases and kept on remote servers. These backups enable a business to stay online in cases of natural disasters, equipment failure, etc.

- **Lower Hardware Costs**: Cloud database service providers supply the infrastructure and perform database maintenance. Hence, companies invest less in hardware and have fewer IT engineers for database maintenance.
- **Value for Money**: Many DBaaS solutions are available in multiple configurations, allowing companies only to pay for what they use and turn off services when they don't need them. Cloud databases also save money by not requiring operational costs or expensive upgrades.
- **Latest Tech**: Cloud database providers upgrade infrastructure and keep it updated with new tech. This brings significant savings as companies don't have to allocate funds on new tech or staff training.
- **Security**: Most cloud database providers encrypt data and invest in the best cloud security solutions to keep the databases safe. Although there is no impenetrable security system, it is a safe way to protect data. Since cloud database providers use automation to enforce the best security practices, there is less room for human error compared to using on-premises databases.

Cloud Database Vendors

- Microsoft Azure
- Amazon Web Service (AWS)
- Oracle
- Google Cloud
- Rackspace

Types of Cloud database

- It is also important to differentiate between cloud databases that are relational as opposed to non-relational or NoSQL.
- The details of each type of cloud database are discussed in the following subsections:
- Cloud Relational Databases
- Cloud NoSQL Databases

Cloud Relational Database

- **Microsoft Azure:** Microsoft Azure cloud database is one of the most popular and globally widespread cloud platforms. In a nutshell, Azure is a cloud computing platform for VM creation, building and running web-based applications, smart client applications, and XML web services. It currently boasts the biggest and strongest global infrastructure, with 55 regions, more than any other cloud provider.
- **Weaknesses:** Iffy Customer Service, Not User Friendly

- **Oracle** Database provides companies with enterprise-scale database technology stored in the cloud. Despite its first offering being quite comprehensive, the Generation 2 offering has consistently higher performance with extensive governance and security controls.
- Data migration is also covered with a dedicated solution and tight customer support in case any technical issues or questions arise.
- **Weaknesses:** No Free Version, Pricey for Small Companies

- **Amazon Web Service (AWS):** AWS is one of the market leaders when it comes to DBaaS. Amazon offers various services for data management and integration. In Amazon RDS: Amazon Relational Database Service runs on either Oracle, SQL, or MySQL server instances. In Amazon SimpleDB. Designed for smaller workloads, SimpleDB is primarily a schema-less database.
- **Weaknesses**: Not Too Customizable, Downtimes as per Amazon's Schedule (The downside is that scaling and patching operations require downtime.)

- **Google Cloud:** Surprisingly, Google is still playing catch-up with the big players in the market. But its solutions are being adopted by more and more businesses of different sizes, thanks to its no-nonsense approach and comprehensive documentation which reduces stress on developers, IT professionals, and other stakeholders.
- The broad open-source compatibility also has its fair share of benefits, allowing you to scale while doing more with analytics and integrations.
- **Strengths:** Comprehensive Documentation, Good for Small and Big Businesses **Weaknesses:** Not Yet at the Level of the Big Three (AWS, Oracle, Azure) The downside is a lack of managed services and the high prices, including a costly support fee.

- **IBM Db2 on Cloud:** IBM Db2 on Cloud is a fully managed SQL database featuring a 99.99% uptime SLA, independent storage and compute scaling through UI and API, several disaster-recovery options, data encryption, and other features.
- IBM's relational database offers advanced data management and analytic capabilities for transactional and warehousing workloads. This database delivers high performance, boasts great insights, data availability, reliability, and broad operating system support.
- The downside of IBM Db2 is that it has fewer regional options, affecting performance in some cases.

- **Rackspace:** Rackspace offers scalable, fully managed, or hosted cloud databases, characterized by high performance and a storage area network (SAN) based on the OpenStack platform.
- Rackspace offers easy access to your cloud database via Cloud Control Panel, CLI (command Line Interface) or API, and features regular backups of all cloud databases.
- Redundant storage and synchronous data replication ensure data protection in case of disaster or hardware failure.
- The downside is a smaller number of data centers compared to the competition.

Cloud NoSQL Databases

- NoSQL database is “ not only SQL” database. The evolution of NoSQL database started in early 2009 and has been growing rapidly since because of some limitations with relational databases.

Limitations with Existing Database

- There are certain limitations with our traditional database system and they cannot fit into the current scenario of big data-related application because data is growing exponentially in every industry. Traditional databases are not capable of handling such data growth, which is now in terabytes(TB) and petabytes(PB).

Traditional databases are unable to:

- Store data in TB/PB; even a a good processor cannot process millions of rows.
- Process TB of data on a single machine.
- Be scalable after a certain limit.

Types of NoSQL Database

Here are the four main types of NoSQL databases:

- Document databases. (e.g MongoDB, CouchDB)
- Key-value stores.(AWS DynamoDB)
- Column-oriented databases.(Apache Hbase, Cassandra)
- Graph databases.(Neo4j)

Document-based database :The document-based database is a nonrelational database. Instead of storing the data in rows and columns (tables), it uses the documents to store the data in the database. A document database stores data in JSON(JavaScript Object Notation), BSON(Binary Javascript Object Notation), or XML(extensible markup language) documents.

- Documents can be stored and retrieved in a form that is much closer to the data objects used in applications which means less translation is required to use these data in the applications. In the Document database, the particular elements can be accessed by using the index value that is assigned for faster querying.

- **Key Values**: A key-value store is a nonrelational database. The simplest form of a NoSQL database is a key-value store. Every data element in the database is stored in key-value pairs. The data can be retrieved by using a unique key allotted to each element in the database. The values can be simple data types like strings and numbers or complex objects.
- **Column-oriented database:** It is a non-relational database that stores the data in columns instead of rows. That means when we want to run analytics on a small number of columns, you can read those columns directly without consuming memory with the unwanted data.
- Columnar databases are designed to read data more efficiently and retrieve the data with greater speed. A columnar database is used to store a large amount of data.

Graph-Based databases:

- Graph-based databases focus on the relationship between the elements. It stores the data in the form of nodes in the database. The connections between the nodes are called links or relationships.

NoSQL Databases:

- **MongoDB Atlas :** MongoDB Atlas is a cloud database created and managed by the same team that developed MongoDB. Mongo's cloud database is a fully managed NoSQL database that features flexibility, scaling, and database management automation. It allows most developers to go through various delivery models without requiring help from a database administrator.
- The downside is that MongoDB Atlas is **NoSQL only**, which means that SQL-reliant applications are not an option with this database.

- Amazon DynamoDB: Amazon DynamoDB is a fully managed, serverless, key-value NoSQL database designed to run high-performance applications at any scale. DynamoDB offers built-in security, continuous backups, automated multi-Region replication, in-memory caching, and data import and export tools.

- Apache Cassandra: Apache Cassandra, a distributed database management system, is built to **manage a large amount of data over several cloud data centers**

Cloud file system

A cloud file system is a hierarchical storage system in the cloud that provides shared access to file data. Users can create, delete, modify, read, and write files, as well as organize them logically in directory trees for intuitive access.

Distributed file system Basics

A distributed file system (DFS) is a file system that spans across multiple file servers or multiple locations, such as file servers that are situated in different physical places. Files are accessible just as if they were stored locally, from any device and from anywhere on the network. A DFS makes it convenient to share information and files among users on a network in a controlled and authorized way.

- An internet search engines is the most common example of DFS which is used for indexing millions of web pages.

  There are a numbers of DFS that solve this problem in different ways. Some popular file system are:
- Andrew file system(AWS)
- Network file system (NFS)
- Coda
- Microsoft distributed file system (MDFS)
- Apple filling protocol (AFP)
- Google file system (GFS)
- Hadoop distributed file system (HDFS)

Google file system (GFS)

- Google File System (GFS), a scalable distributed file system (DFS), to meet the company's growing data processing needs.
- GoogleFS is another name for GFS. It manages two types of data namely File metadata and File Data.
- The GFS node cluster consists of a single master and several chunk servers that various client systems regularly access. On local discs, chunk servers keep data in the form of Linux files. Large (64 MB) pieces of the stored data are split up and replicated at least three times around the network. Reduced network overhead results from the greater chunk size.

Google File System Architecture

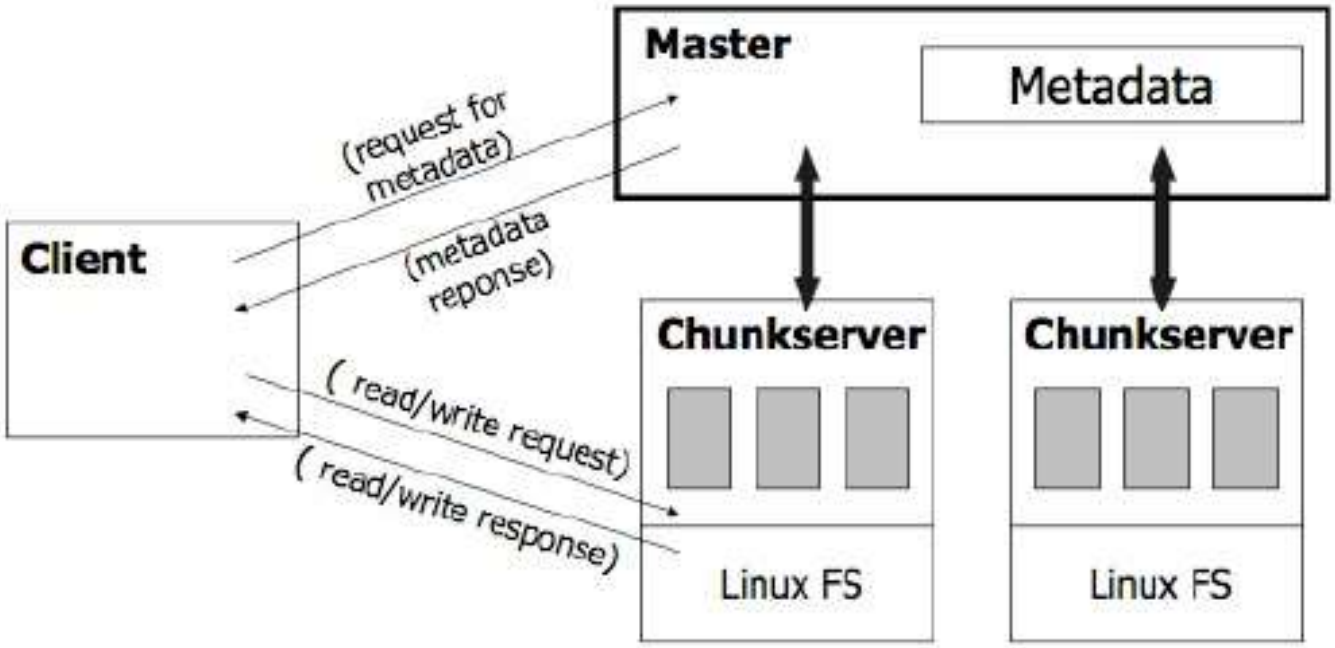


Figure 1

Components of GFS

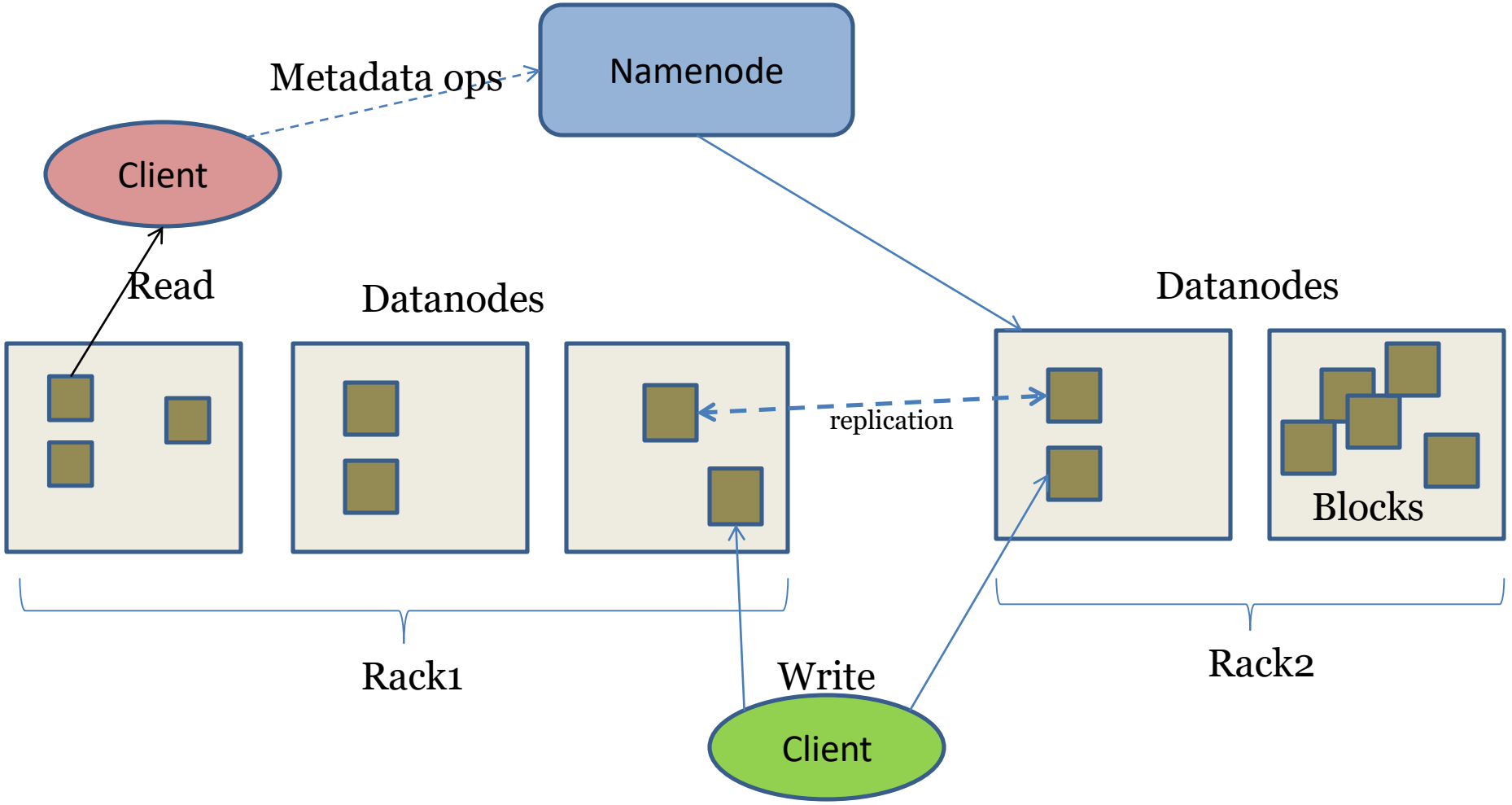
A group of computers makes up GFS. A cluster is just a group of connected computers. There could be hundreds or even thousands of computers in each cluster. There are three basic entities included in any GFS cluster as follows:

- **GFS Clients:** They can be computer programs or applications which may be used to request files. Requests may be made to access and modify already-existing files or add new files to the system.
- **GFS Master Server:** It serves as the cluster's coordinator. It preserves a record of the cluster's actions in an operation log. Additionally, it keeps track of the data that describes chunks, or metadata. The chunks' place in the overall file and which files they belong to are indicated by the metadata to the master server.
- **GFS Chunk Servers:** They are the GFS's workhorses. They keep 64 MB-sized file chunks. The master server does not receive any chunks from the chunk servers. Instead, they directly deliver the client the desired chunks. The GFS makes numerous copies of each chunk and stores them on various chunk servers in order to assure stability; the default is three copies.

Hadoop Distributed File System

- Hadoop is a DFS based on GFS that provides high throughput access to application data. Provides two main fetaures:
- Fault tolerance
- Portability.

HDFS Architecture



Hadoop Distributed File System Architecture

- Master/slave architecture
- HDFS cluster consists of a single **Namenode**, a master server that manages the file system namespace and regulates access to files by clients.
- There are a number of **DataNodes** usually one per node in a cluster.
- The DataNodes manage storage attached to the nodes that they run on.
- HDFS exposes a file system namespace and allows user data to be stored in files.
- A file is split into one or more blocks and set of blocks are stored in DataNodes.
- DataNodes: serves read, write requests, deletion, and replication upon instruction from Namenode.
- Note: chunk size is 128 MB

Hadoop Distributed File System HDFS	Google File System GFS
Cross Platform	Linux
Developed in Java environment	Developed in C,C++ environment
Initially it was developed by Yahoo and now its an **open source Framework**	It was developed & still owned by Google
It has Name node and Data Node	It has Master-node and Chunk server
128 MB will be the default block size	**64 MB** will be the default block size
Name node receive heartbeat from Data node	Master node receive heartbeat from Chunk server
Commodity hardware are used	Commodity hardware are used
"Write Once and Read Many" times model	Multiple writer , multiple reader model
Deleted files are renamed into particular folder and then it will removed via garbage	Deleted files are not reclaimed immediately and are renamed in hidden name space and it will deleted after three days if it's not in use
Edit Log is maintained	Operational Log is maintained
Only append is possible	Random file write possible