UNIT-5

Q1. What is Cloud Secutity? Explain the different types of Cloud Risks ?

Ans :- To ensure that the customers does not face any difficulties such as loss of data or data thief.

- Cloud security is the first and foremost concern of every industry and Organization.
- There is a possibility that a malicious user can go through the cloud by impersonating a legal user, Thereby after by infecting the cloud services.
- Most of the vendors are highly concern about the Data integrity , Privacy issues, authentication issue, Data loss, user-level security and vendor level security.

Cloud Risk

When infrastructure, Applications, Data and Storage are hosted by cloud providers, There is hugh chance of risk in each type of services offering. This is know as cloud Risk.

1. Organizations such as the cloud Security Alliance (CSA) offer certification to cloud providers that will meet their criteria.
2. The Cloud Security Alliance (CSA) is the world's leading organization dedicated to defining and raising awareness of best practices to help ensure a secure cloud computing environment.

3. **Core Cloud :-** Regardless if you already have a well established cloud security program or are starting your cloud migration for the first time, CSA can help you enhance your security strategy.

Q2. What are the different Cloud Risk Division Policy Organizational Risks ?

Ans :- **Policy and Organizational risks :-**

1. **Lock –in :-** When applications, data and services are dependent on only one cloud providers, it is know as a lock-in problem. There could be SaaS lock-in, PaaS lock-in and IaaS lock-in problems.
2. **Loss of governance** :- This issues comes when the cloud provider may sub-contract or outsource services to third parties (unknown providers) that may not compromise the same guarantees (Such as to provide the service in a lawful way).
3. **Compliance challenges** :- Cloud providers make huge investments for external certifications such as SAS(Statement on Auditing Standard No. 70 (**SAS 70**)) , PCI DSS (PCI DSS, is a formal process used by organizations to identify threats and vulnerabilities that could negatively impact the security of cardholder data)

   Compliance challenges continues…

   This certifications give them the reputation in the market that they are following the best security practices.

However, in some cases particular certification may also be a problem for accessing cloud services. For example, If a client using AWS cloud wants to use the EC2 (Amazon Elastic Compute Cloud (Amazon EC2)) and if the EC2 service does not have PCI compliance, then the EC2 service cannot be used for credit card-related transactions.

4. **Cloud service termination or failure :-** There must be 24x7 support and high availability of all services, but in the competitive world of IT**. Due to lack of financial support and other factors could lead some providers to go out of business or shut down their service 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 :-** There is a possibility that the cloud provider could outsource some services to other third parties. In that case, any interruption or corruption in the chain or a lack of coordination of responsibilities between all parties involved can lead to inaccessibility of services, Loss of data confidentially, availability and integrity and reputational losses, because of failure to meet customer demand such as cascading service failure and violating of SLA.

Q3. What is the different types of Technical Risks ? Explain briefly ?

Ans :-

1. **Isolation Failure —** This risk category covers the failure of mechanisms separating storage, memory, routing and even reputation between different tenants (e.g., so-called guest-hopping attacks).
2. **Resource exhaustions :-** Cloud service is fully on-demand pay-per-use service. There is a chance of risk in proper allocation of resources to cloud users.

3. **Cloud provider malicious insider :-** The malicious actions of an insider could possibly have an impact on the confidentiality, integrity and availability of all kind of data, IP, all kind of services. Taking care of this issue is extremely important in case of cloud computing because cloud architecture contain certain characteristics that are at very high risk.

4. **Intercepting data in transit :-** Cloud services are based on distributed architecture; therefore, transmission of data takes place across multiple physical machines, from one VM to another images distribution between cloud infrastructure and remote web clients, VPN environment and such. This risk is more vulnerable when data is being transferred from one-premises to cloud or cloud storage to on-premises. Spoofing, man-in the middle attacks and sniffing types of attacks could be possible during data transfer-related activities.
5. **Insecure or ineffective deletion of data :-** Whenever a provider is altered,

i. Resources are scaled down

ii. Physical hardware is moved

iii. Data may be available beyond the lifetime specified in the security policy

iv. This may be tough to carry out the procedures stated by the security policy because full data deletion is only imaginable by destroying a disk that also stores data from other clients. When a request to delete a cloud resource is made, this may not result in true wiping of the data. For this true data spreading is required and special procedures must be followed that may not be supported by the standard API.

6. Conflicts between customer hardening procedures and cloud environment :- Cloud providers follow different server or instances 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, SSL customer private keys) or passwords to malicious parties. The loss or corruption of those keys or their unauthorized use for authentication and digital signature.

8. Malicious probes or scams :- Malicious probes and scanning, as well as network mapping are indirect threats to the assets being considered. They can be used to collect information in the context of a hacking effort.

9. Compromise service engine :- All cloud providers rely either on an extremely specific platform, or the service engine that is placed just above the physical hardware and manages customers requests at different levels of abstraction.

Q4. What is the different types of Legal Risks and others Risk?

Ans :- **C. LEGAL RISKS :-**

1. **Risk from changes of jurisdiction :-** Customer data may be kept in several jurisdictions, some of which may be high risk. If datacenters are located in high-risk countries (e,g those who lack the rule of law) and enforcement, monocratic police states, states that do not respect international aggreements, sites could be attacked by local authorities.

2. **Licensing risks :-** Licensing conditions such as per-seat agreements and online licensing checks may become unusable in a cloud environment . (For example , software is charged as per instances basis so if our cloud-based instance increases, the cost of the software also increases exponentially.

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 the data is handled in a lawful way. **There may be data security breaches that are not intimated to the controller by the cloud provider. The cloud provider may misplace control of the data administered by the cloud provider.**

- This issue increases in the case of multiple transfers of data.

OTHERS 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 datacenters is possible. Generally cloud providers have large datacenters, therefore, physical control of a datacenter must be stronger because the impact 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 datacenters.

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.

Q5. Explain the Cloud Computing Security Architecture ?

Ans :- There are four layers in the generic architecture of cloud.

1. **Data Center 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.

2. **VM layer :-** This layer involves VM level security issues, VM monitoring, Hypervisor-related security issues and VM isolation management issues.

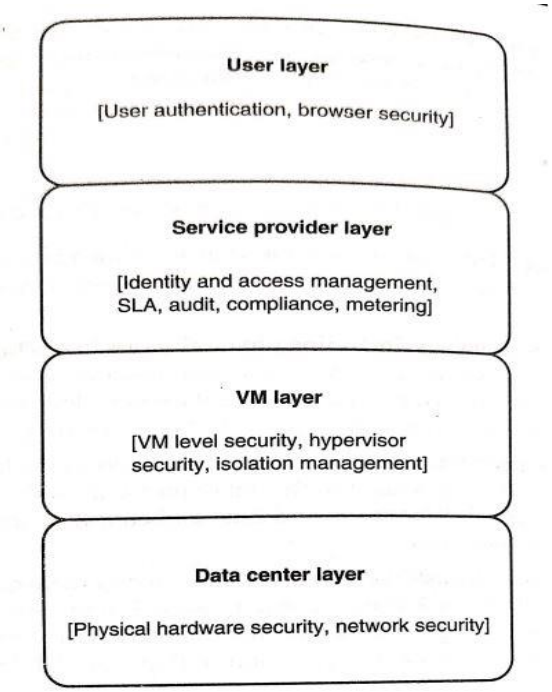


Figure 6.1 | Cloud security architecture.

3.. Service provider layer :- This layer is responsible for identity and access management , service level agreement (SLA), metering, compliance and audit-related issues.

4. User layers:- This is the first layer of user interaction. It is responsible for user authentication and authorization and all browser-related security issues.

Q6. What are the VM security Challenges ?

Ans:- **VM SECURITY CHALLENGES**

1. **Communication between VMs or between VMs and the Host :-** VMs serve some key requirements for any organization such as the following:-

i. Sharing one physical computer resource among multiple companies or organizations.

ii. Consolidation of different services into one physical computer.

iii. Providing a general hardware platform to host multiple operating system.

2. VM escape :- VMs allow us to share the resources of the host computer and provider isolation between VMs and their host.

i. In an ideal situation, any program that runs under the VM should not communicate to any other program inside that or any other VM, but because of some architecture limitations or some other bugs, software affect this isolation.

ii. It may so happen that a program running inside a VM can totally bypass the VM layer and acquire full access to the host system. Such a situation is known as VM escape.

3. VM monitoring from the host:- It is not normally considered a limitation or a bug when one can start monitoring, changing or communication with a VM application from the host. In this case, the host itself starts controlling; therefore the host requires more strict security environments compared to each individual VM.

i. The host can affect VMs behavior in the following ways

ii. Start, stop, pause and restart VMs.

iii. Monitor and configure resources available to the VMs, including CPU, memory, disk and network usage of VMs.

iv. Adjust the number of CPUs, amount of memory, amount and number of victuals disks and number of virtual network interfaces to a VM

v. Monitor the applications running inside the VM.

vi. View, Copy and possibly modify the data stored on the VM's virtual disk.

4. **VM monitoring from another VM :-** Isolation is a basic characteristic of VM technology; it is usually referred as a security defect when one VM can easily monitor another without defined configuration and privilege to do so.

i. If the hypervisor memory is implemented properly then individual VM protection takes place automatically.

ii. It will not disturb others VMs memory address space. Because VMs do not have direct access to the Host system.

iii. If network traffic is more complicated then there could be an issue with isolation depending on how the network connections are set up with the VMs.

iv. There should be the case of a virtual hub also, if the Vm uses a virtual hub for connecting all VMs host machine, then guest VM my sniff the packets of host VM or other guest VMs using ARP.

5. **Denial of service :-** Because various computing resources like CPU, memory, network and hard disk are shared among multiple VMs and host machine. This may create a denial of service attack against another VM.

6. External modification of a VM: In a business application scenario, users VMs have the privilege of accessing employee databases through a secured application

i. Database security is more critical in a virtual environment.

ii. Database is placed inside a secured VM environment so that any external user is not allowed to access the database outside of the application.

iii. If a VM where database is installed becomes accessible from outside because of a malicious attack, then the database can be corrupted or modified and the system trust can be broken.

7. External modification of the hypervisor :- Because the hypervisor is mainly responsible for the enablement of virtualization while making the process of more self-protected and secure VM, it does not affect the working of any underlying hypervisor. Therefore the first thing is to protect the hypervisor from any external unauthorized access and changes

8. Mixed trust level VMs :- Enterprises must take care of mission critical-related information while leveraging the benefits of virtualization. After applying some self-protection system and some external security mechanism such as integrity checking, file monitoring, log assessment , firewall protection and antivirus detection, The VM can be more secure in mixed environment.

9. Resource contention :- Whenever some resource-consuming operations like malware or antivirus scanning, files and patch updates are executed on VMs, the result of these operations produce high loads on the systems and hamper server applications and VDI environment.

i. To avoid such situations, each VM requires additional significant memory footprint because just like traditional architecture, the antivirus must be installed on each operating system and the same kind of protection is required for each VM too.

ii. More virtualization-sensitive technology is needed for optimal resource utilization and increasing VM performance so that dedicated antivirus and file scanning should not affect the memory footprint on the virtual hosts.

Q7. What is Cloud Database? Types of Cloud Database ?

Ans :- Now a days cloud providers offer RDS (Cloud Relation Databases) now a days. Some of the popular and most adopted RDS across the globe are as follows.

1. Amazon relational database service:- Amazon RDS is very popular and widely adopted Web service. It looks like others AWS services and providers easy management consoles for operating RDS on cloud.

i. Amazon RDS is a highly cost-efficient and secured service.

ii. Currently it supports Oracle, SQL server, MySQL and PostgreSQL database.

iii. Amazon RDS specifically offers two types of RDS instances.

a. On-demand instances :- An on-demand instance offering is a pay-per-use instance with no long –term commitment.

b. Reserved DB instances:- Reserved DB instances give the flexibility of one-time payment for the DB instance if the database usage is predictable. There is also an offer of 30%-50% price cut over the on-demand price.

2. Google Cloud SQL:- Google cloud SQL is a MySQL database service that is managed by google and the entire management, data replication, encrpytion, security and backups are handled by google's cloud infrastructure.

i. Google claims maximum availability of its data because its data centers are located across every region of the world.

ii. Google cloud SQL is a very flexible, easy-to-use service.

iii. Which enables connecting and managing cloud SQL with an existing application, just as in done with MySQL.

3. Heroku Postgres:- Heroku Postgres is a relationship SQL database offered by Heroku. It is accessible through all programming lanaguages supported by Heroku.

i. Heroku postgres offers fully reliability of services, which means around 99.99% uptime and 99.99% durability of data.

ii. One of the advance features of Heroku Postgres is Dataclips, which enables users to send the results of the SQL query via the URL.

4. HP cloud relational database for MySQL:- HP cloud RDS automate application deployment, configuration management and patch-up task database.

i. It currently supports command line interface (CLI).

ii. But an easy-to-use Web-based console interface through API is excepted soon.

iii. It also provides database snapshot facility in multiple availability zones for providing more reliability.

5. Microsoft Azure SQL database:- Earlier it was known as SQL Azure. It is the most important component of the Microsoft Azure cloud service.

i. It can be operated as a standalone cloud database also.

ii. The database can be synched easily with other SQL server databases within the cloud infrastructure of the company or organization.

The performance of database can be predicted irrespective of whether the service chosen is basic, standard or premiu

6. Oracle database cloud service:- Oracle database cloud offers two options for users:

i. One is a single schema-based service and

ii. Another is fully configured Oracle database installed virtual machine

iii. It also provides flexibility in the management option: self-managed service or fully managed by Oracle.

7. Rackspace cloud database:- Rackspace cloud databases are based on open standards. These currently support MySQL, Percona and MariaDB databases.

i. Rackspace cloud provides high database performance using container-based virtualization.

ii. It provides automated configuration which reduces operational costs and team effort.

iii. Rackspace cloud is also connected to SAN storage, which built-in data replication for high data **replication.**

Q8. Explain briefly the Distributed File System Basics in Cloud Computing?

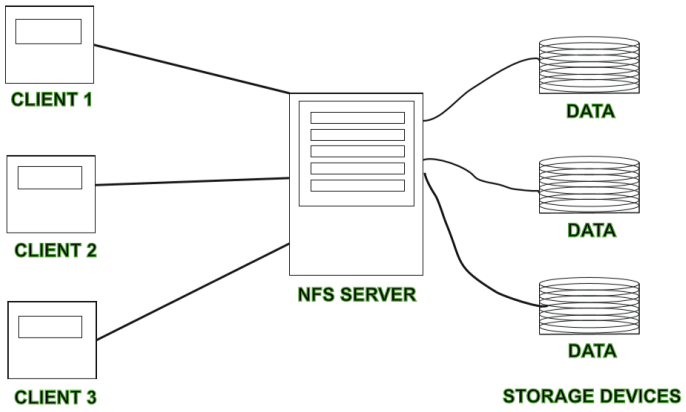
Ans :- A **Distributed File System (DFS)** as the name suggests, is a file system that is distributed on multiple file servers or multiple locations. It allows programs to access or store isolated files as they do with the local ones, allowing programmers to access files from any network or computer.

Features of DFS :

- **Transparency :**
  - **Structure transparency –**
    There is no need for the client to know about the number or locations of file servers and the storage devices. Multiple file servers should be provided for performance, adaptability, and dependability.
  - **Access transparency –**
    Both local and remote files should be accessible in the same manner. The file system should be automatically located on the accessed file and send it to the client's side.
  - **Naming transparency –**
    There should not be any hint in the name of the file to the location of the file. Once a name is given to the file, it should not be changed during transferring from one node to another.
  - **Replication transparency –**
    If a file is copied on multiple nodes, both the copies of the file and their locations should be hidden from one node to another.
- **User mobility :**
  It will automatically bring the user's home directory to the node where the user logs in.
- **Performance :**
  Performance is based on the average amount of time needed to convince the client requests. This time covers the CPU time + time taken to access secondary storage + network access time. It is advisable that the performance of the Distributed File System be similar to that of a centralized file system.
- **Simplicity and ease of use :**
  The user interface of a file system should be simple and the number of commands in the file should be small.
- **High availability :**
  A Distributed File System should be able to continue in case of any partial failures like a link failure, a node failure, or a storage drive crash.
  A high authentic and adaptable distributed file system should have different and independent file servers for controlling different and independent storage devices.
- **Scalability :**
  Since growing the network by adding new machines or joining two networks together is routine, the distributed system will inevitably grow over time. As a result, a good distributed file system should be built to scale quickly as the number of nodes and users in the system grows. Service should not be substantially disrupted as the number of nodes and users grows.
- **High reliability :**
  The likelihood of data loss should be minimized as much as feasible in a suitable distributed file system.

That is, because of the system's unreliability, users should not feel forced to make backup copies of their files. Rather, a file system should create backup copies of key files that can be used if the originals are lost. Many file systems employ stable storage as a high-reliability strategy.

- **Data integrity :**
  Multiple users frequently share a file system. The integrity of data saved in a shared file must be guaranteed by the file system. That is, concurrent access requests from many users who are competing for access to the same file must be correctly synchronized using a concurrency control method. Atomic transactions are a high-level concurrency management mechanism for data integrity that is frequently offered to users by a file system.
- **Security :**
  A distributed file system should be secure so that its users may trust that their data will be kept private. To safeguard the information contained in the file system from unwanted & unauthorized access, security mechanisms must be implemented.
- **Heterogeneity :**
  Heterogeneity in distributed systems is unavoidable as a result of huge scale. Users of heterogeneous distributed systems have the option of using multiple computer platforms for different purposes.
-



Advantages :

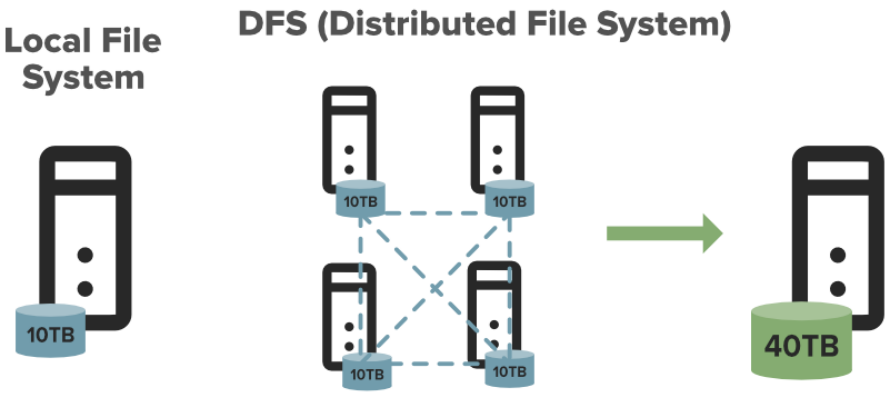
- DFS allows multiple user to access or store the data.
- It allows the data to be share remotely.
- It improved the availability of file, access time, and network efficiency.
- Improved the capacity to change the size of the data and also improves the ability to exchange the data.
- Distributed File System provides transparency of data even if server or disk fails.

Disadvantages :

- In Distributed File System nodes and connections needs to be secured therefore we can say that security is at stake.
- There is a possibility of lose of messages and data in the network while movement from one node to another.
- Database connection in case of Distributed File System is complicated.
- Also handling of the database is not easy in Distributed File System as compared to a single user system.
- There are chances that overloading will take place if all nodes tries to send data at once.

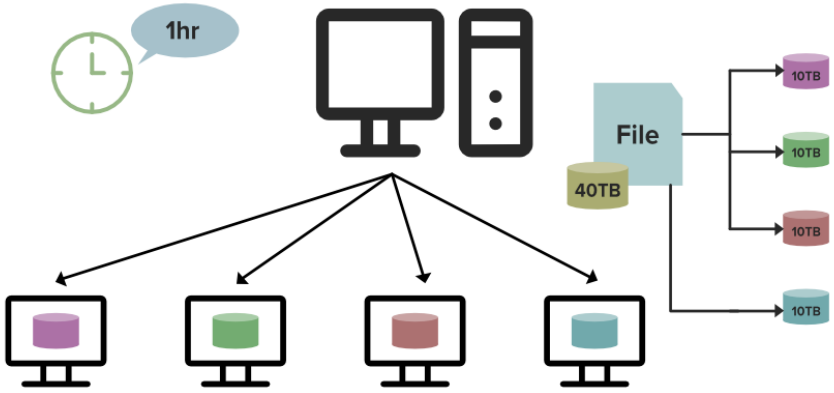
Q9. Explain the different types of DFS and HDFS with suitable diagram ?

Ans :- Before head over to learn about the HDFS(Hadoop Distributed File System), we should know what actually the file system is. The file system is a kind of Data structure or method which we use in an operating system to manage file on disk space. This means it allows the user to keep maintain and retrieve data from the local disk.



ou might be thinking that we can store a file of size 30TB in a single system then why we need this DFS. This is because the disk capacity of a system can only increase up to an extent. If somehow you manage the data on a single system then you'll face the processing problem, processing large datasets on a single machine is not efficient.

Let's understand this with an example. Suppose you have a file of size 40TB to process. On a single machine, it will take suppose 4hrs to process it completely but what if you use a DFS(Distributed File System). In that case, as you can see in the below image the File of size 40TB is distributed among the 4 nodes in a cluster each node stores the 10TB of file. As all these nodes are working simultaneously it will take the only 1 Hour to completely process it which is Fastest, that is why we need DFS.



HDFS is capable of handling larger size data with high volume velocity and variety makes Hadoop work more efficient and reliable with easy access to all its components. HDFS stores the data in the form of the block where the size of each data block is 128MB in size which is configurable means you can change it according to your requirement in *hdfs-site.xml* file in your Hadoop directory.

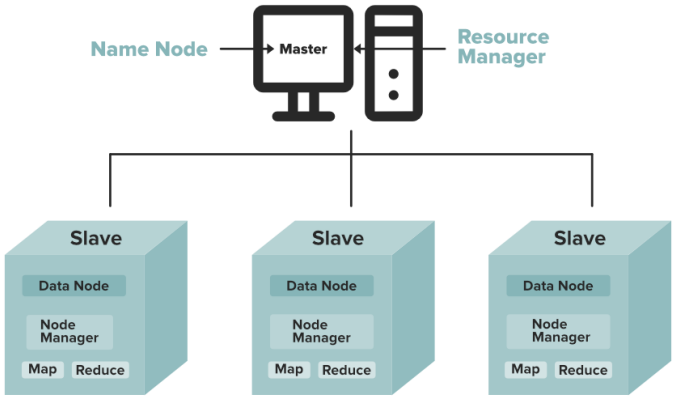
Some Important Features of HDFS(Hadoop Distributed File System)

- It's easy to access the files stored in HDFS.
- HDFS also provides high availability and fault tolerance.

- Provides scalability to scaleup or scaledown nodes as per our requirement.
- Data is stored in distributed manner i.e. various Datanodes are responsible for storing the data.
- HDFS provides Replication because of which no fear of Data Loss.
- HDFS Provides High Reliability as it can store data in a large range of *Petabytes*.
- HDFS has in-built servers in Name node and Data Node that helps them to easily retrieve the cluster information.
- Provides high throughput.

HDFS Storage Daemon's

As we all know Hadoop works on the MapReduce algorithm which is a master-slave architecture, HDFS has *NameNode* and *DataNode* that works in the similar pattern.
1. NameNode(Master)
2. DataNode(Slave)
1. NameNode: NameNode works as a *Master* in a Hadoop cluster that Guides the Datanode(Slaves). Namenode is mainly used for storing the Metadata i.e. nothing but the data about the data. Meta Data can be the transaction logs that keep track of the user's activity in a Hadoop cluster.



Objectives and Assumptions Of HDFS

1. System Failure: As a Hadoop cluster is consists of Lots of nodes with are commodity hardware so node failure is possible, so the fundamental goal of HDFS figure out this failure problem and recover it.
2. Maintaining Large Dataset: As HDFS Handle files of size ranging from GB to PB, so HDFS has to be cool enough to deal with these very large data sets on a single cluster.
3. Moving Data is Costlier then Moving the Computation: If the computational operation is performed near the location where the data is present then it is quite faster and the overall throughput of the system can be increased along with minimizing the network congestion which is a good assumption.
4. Portable Across Various Platform: HDFS Posses portability which allows it to switch across diverse Hardware and software platforms.
5. Simple Coherency Model: A Hadoop Distributed File System needs a model to write once read much access for Files. A file written then closed should not be changed, only data can be appended. This assumption helps us to minimize the data coherency issue. MapReduce fits perfectly with such kind of file model.