USN | | | | | | | | | | |

Internal Assessment Test 2 – May 2025 – **Question Paper with solution**

Sub:	Cloud Computing					Sub Code:	BCS601	Branch:	CSE
Date:	27.05.2025	Duration:	90 mins	Max Marks:	50	Sem / Sec:	6 A,B,C		OBE

Answer any FIVE FULL Questions

		MARKS	CO	RBT
1 (a)	**With the help of a neat diagram, explain the stack of the extended cloud computing services** **SOLUTION:**	**6M**	**CO3**	**L2**

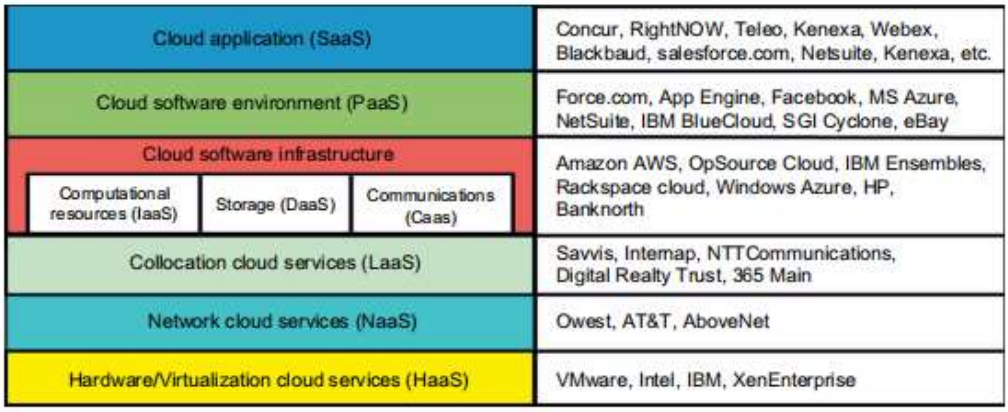
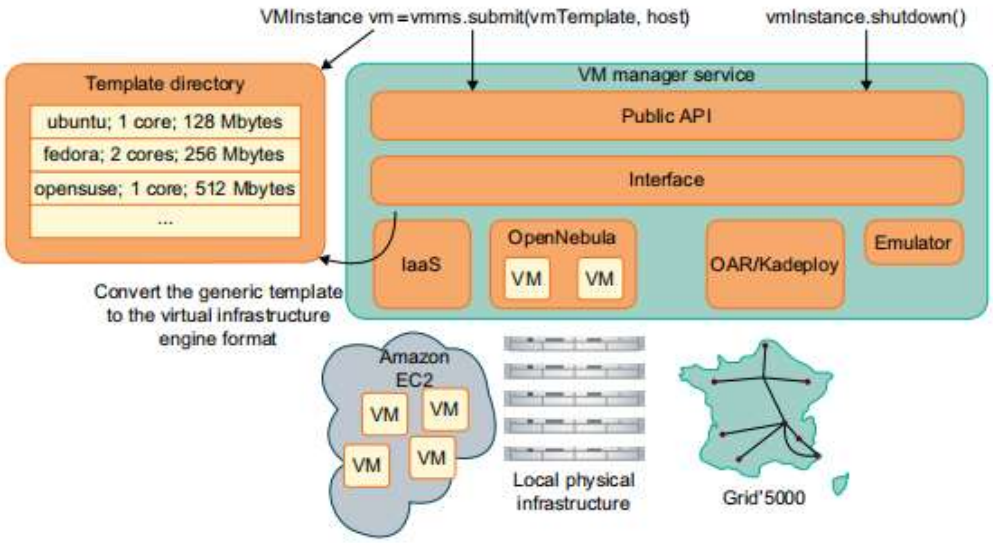


Figure 4.23 shows six layers of cloud services, ranging from hardware, network, and collocation to infrastructure, platform, and software applications. We already introduced the top three service layers as SaaS, PaaS, and IaaS, respectively. The cloud platform provides PaaS, which sits on top of the IaaS infrastructure. The top layer offers SaaS. These must be implemented on the cloud platforms provided. Although the three basic models are dissimilar in usage, as shown in Table 4.7, they are built one on top of another. The implication is that one cannot launch SaaS applications with a cloud platform. The cloud platform cannot be built if compute and storage infrastructures are not there.

The bottom three layers are more related to physical requirements. The bottommost layer provides *Hardware as a Service (HaaS)*. The next layer is for interconnecting all the hardware components, and is simply called *Network as a Service (NaaS)*. *Virtual LANs* fall within the scope of NaaS. The next layer up offers *Location as a Service (LaaS)*, which provides a collocation service to house, power, and secure all the physical hardware and network resources. Some authors say this layer provides *Security as a Service* ("SaaS"). The cloud infrastructure layer can be further subdivided as *Data as a Service (DaaS)* and *Communication as a Service (CaaS)* in addition to compute and storage in IaaS.

We will examine commercial trends in cloud services in subsequent sections. Here we will mainly cover the top three layers with some success stories of cloud computing. As shown in Table 4.7, cloud players are divided into three classes: (1) cloud service providers and IT administrators, (2) software developers or vendors, and (3) end users or business users. These cloud players vary in their roles under the IaaS, PaaS, and SaaS models. The table entries distinguish the three cloud models as viewed by different players. From the software vendors' perspective, application performance on a given cloud platform is most important. From the providers' perspective, cloud infrastructure

(b)	**Write a short note on VM Creation and Management** **SOLUTION:**	**4M**	**CO2, CO3**	**L2**



4.5.3.1 Independent Service Management

Independent services request facilities to execute many unrelated tasks. Commonly, the APIs provided are some web services that the developer can use conveniently. In Amazon cloud computing infrastructure, SQS is constructed for providing a reliable communication service between different providers. Even the endpoint does not run while another entity has posted a message in SQS. By using independent service providers, the cloud applications can run different services at the same time. Some other services are used for providing data other than the compute or storage services.

4.5.3.2 Running Third-Party Applications

Cloud platforms have to provide support for building applications that are constructed by third-party application providers or programmers. As current web applications are often provided by using Web 2.0 forms (interactive applications with Ajax), the programming interfaces are different from the traditional programming interfaces such as functions in runtime libraries. The APIs are often in the form of services. Web service application engines are often used by programmers for building applications. The web browsers are the user interface for end users.

In addition to gateway applications, the cloud computing platform provides the extra capabilities of accessing backend services or underlying data. As examples, GAE and Microsoft Azure apply their own cloud APIs to get special cloud services. The WebSphere application engine is deployed by IBM for Blue Cloud. It can be used to develop any kind of web application written in Java. In EC2, users can use any kind of application engine that can run in VM instances.

4.5.3.3 Virtual Machine Manager

The VM manager is the link between the gateway and resources. The gateway doesn't share physical resources directly, but relies on virtualization technology for abstracting them. Hence, the actual resources it uses are VMs. The manager manage VMs deployed on a set of physical resources. The VM manager implementation is generic so that it can connect with different VIEs. Typically, VIEs can create and stop VMs on a physical cluster. The Melbourne group has developed managers for OpenNebula, Amazon EC2, and French Grid'5000. The manager using the OpenNebula OS (www.opennebula.org) to deploy VMs on local clusters.

OpenNebula runs as a daemon service on a master node, so the VMM works as a remote user. Users submit VMs on physical machines using different kinds of hypervisors, such as Xen (www.xen.org), which enables the running of several operating systems on the same host concurrently. The VMM also manages VM deployment on grids and IaaS providers. The InterGrid supports Amazon EC2. The connector is a wrapper for the command-line tool Amazon provides. The VM manager for Grid'5000 is also a wrapper for its command-line tools. To deploy a VM, the manager needs to use its template.

2 (a)	**Explain in detail the following resource provisioning techniques along with use cases or examples:** **a. Demand-driven resource provisioning** **b. Event-driven resource provisioning** **c. Popularity-driven resource provisioning** **SOLUTION**	**6M**	**CO3**	**L2**

Provisioning of Compute Resources (VMs)

- Efficient VM provisioning depends on the cloud architecture and management of cloud infrastructures.
- Virtualized cluster of servers, demands efficient installation of VMs, live migration, and fast recovery from failures.
- Ex: EC2 and IBM Blue cloud uses Xen.
- Predefined VM templates are also provided.

- Demand Driven Provisioning
- Event-driven Provisioning
- Popularity-driven method

Demand-Driven Resource Provisioning

- Method adds and removes computing instances based on the current utilization level of allocated resources.
- Allocates a new processor if the already using processor crosses a threshold.
- Amazon implements auto-scale feature in its EC2 platform.

Event-Driven Resource Provisioning

- Adds and removes machine instances based on a specific time event.
- Scheme works for seasonal or predicted events.
- Scheme anticipates peak traffic before it happens.

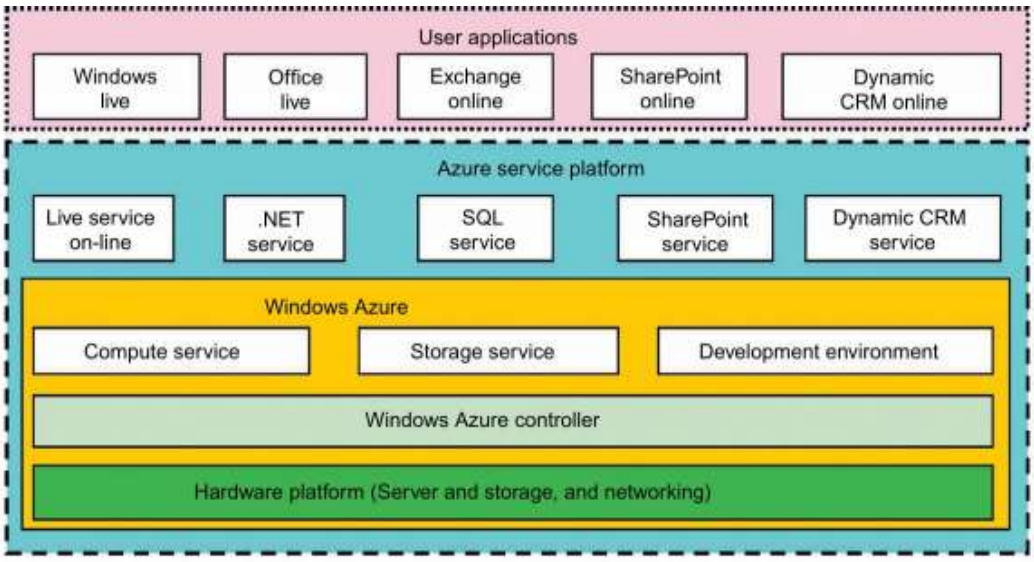
Popularity-Driven Resource Provisioning

- Internet searches for popularity of certain applications and creates the instance by popularity demand.
- Scheme anticipates increased traffic with popularity.
- Minimal loss of QoS, if the prediction is correct.

(b)	**Explain the architecture of Microsoft Azure with a neat diagram** **SOLUTION**	**4M**	**CO5**	**L2**

Microsoft Windows Azure

- Microsoft launched Microsoft Azure in 2008, to meet the challenges in cloud computing.
- Platform is built over Microsoft data centers.
- Platform is divided into three major component platforms.
- Applies standard web communication protocols SOAP and REST



Microsoft Windows Azure

Platform is divided into three major component platforms:

- Windows Azure: Offers a cloud platform built on Windows OS and based on Microsoft virtualization technology.
- Azure Service Platform: Built on top of the infrastructure are the services for cloud applications.
- User Applications: Cloud services in Azure can interact with traditional Microsoft software applications.

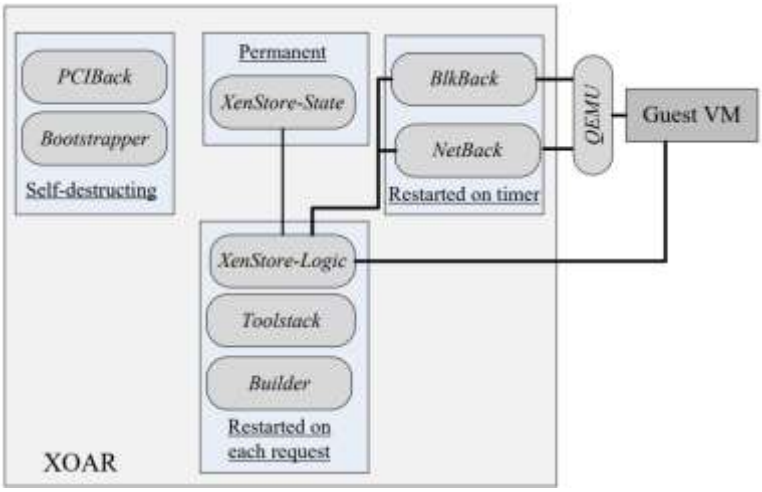
	Azure Services • **Live Service**: Used to apply the data involved across multiple machines concurrently. • **.NET Service**: Package support application development on local hosts and execution on cloud machines. • **SQL Azure**: Used for relational database associated with SQL server in the cloud. • **SharePoint Service**: Provides a scalable and manageable platform for users to develop their special business applications. • **Dynamic CRM Service**: Provides a platform for CRM applications .			
3 (a)	**What is Xoar? How does it break the monolithic design of the Trusted Computing Base?** ***Solution:*** *XOAR* XOAR is a modified version of the Xen hypervisor, designed to enhance security in cloud computing environments by breaking the monolithic design of the Trusted Computing Base (TCB). *Xoar - breaking the monolithic design of TCB* Xoar is a version of Xen designed to boost system security; based on micro-kernel design principles. The design goals are: 1. Maintain the functionality provided by Xen. 2. Ensure transparency with existing management and VM interfaces. 3. Tight control of privileges, each component should only have the privileges required by its function. 4. Minimize the interfaces of all components to reduce the possibility that a component can be used by an attacker. Eliminate sharing. 5. Make sharing explicit whenever it cannot be eliminated to allow meaningful logging and auditing. 6. Reduce the opportunity of an attack targeting a system component by limiting the time window when the component runs. The security model of Xoar assumes that threats come from: ● A guest VM attempting to violate data integrity or confidentiality of another guest VM on the same platform, or to exploit the code of the guest. ● Bugs in the initialization code of the management virtual machine. Xoar system components Permanent components XenStore-State maintains all information regarding the state of the system. Components used to boot the system; they self-destruct before any user VM is started. They discover the hardware configuration of the server including the PCI drivers and then boot the system: ● PCIBack - virtualizes access to PCI bus configuration. ● Bootstrapper - coordinates booting of the system.	**6M**	**CO3**	**L2**

Components restarted on each request:

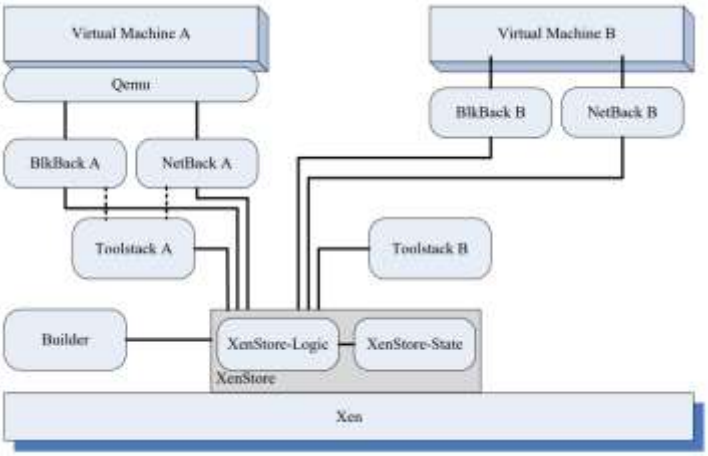
- XenStore-Logic.
- Toolstack - handles VM management requests, e.g., it requests the Builder to create a new guest VM in response to a user request.
- Builder - initiates user VMs.

Components restarted on a timer; the two components export physical storage device drivers and the physical network driver to a guest VM.

- Blk-Back - exports physical storage device drivers using udev rules.
- NetBack - exports the physical network driver.

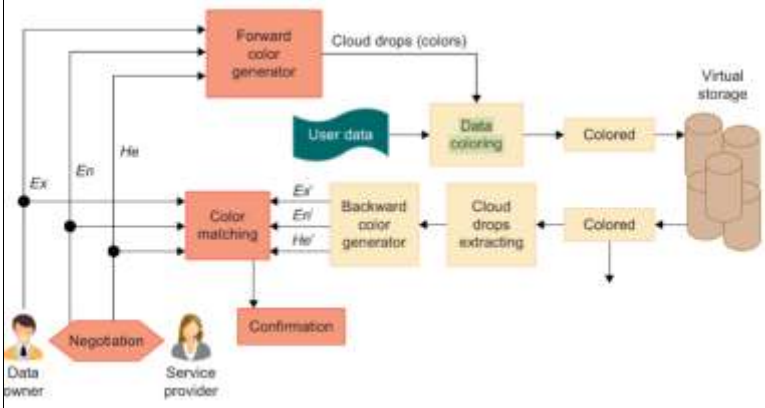


Xoar has nine classes of components of four types: permanent, self-destructing, restarted upon request, and restarted on timer. A guest VM is started using the Builder using the Toolstack; it is controlled by the XenStore-Logic. The devices used by the guest VM are emulated by the Qemu component. Qemu is responsible for device emulation



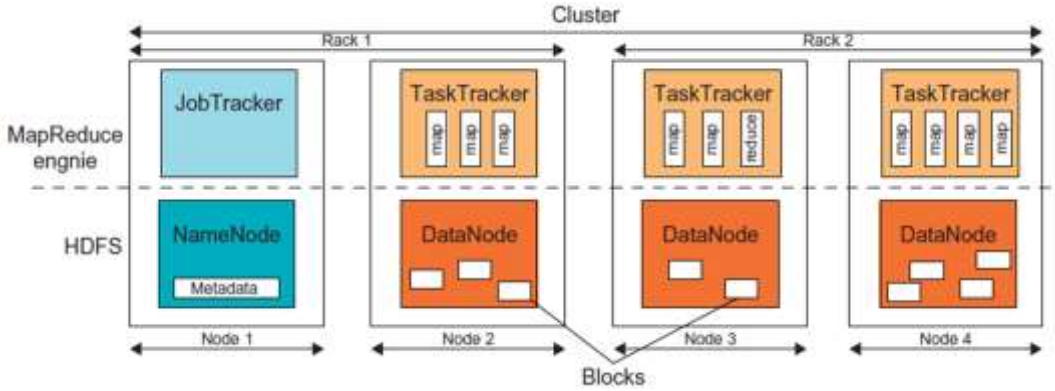
Component sharing between guest VMs in Xoar. Two VMs share only the XenStore components. Each one has a private version of the BlkBack, NetBack and Toolstack.

(b)	**Explain data coloring and watermarking for trust management at various levels in datacenter** ***Solution:*** 1. Data Coloring: • Data coloring is a technique where data is tagged or labeled with metadata indicating its sensitivity, origin, or security policies.	**4M**	**CO4**	**L2**

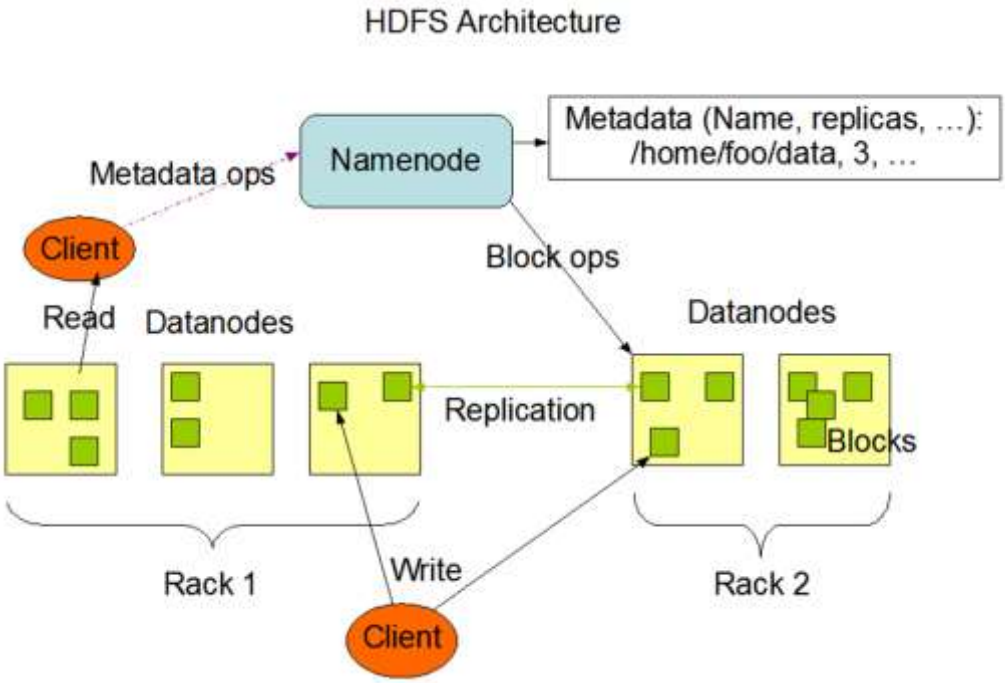
	• These labels help systems track, monitor, and control data flow in datacenters. • For example, data marked "confidential" may only be accessed by certain roles or VMs. 2. Watermarking: • Watermarking embeds invisible or visible markers into data to prove ownership or detect unauthorized use. • In cloud datacenters, watermarking is used to trace data leakage or tampering, ensuring data integrity and authenticity. 3. Role in Trust Management: • Both techniques enhance accountability and visibility in data handling. • They support fine-grained access control, audit trails, and compliance with trust policies across users, applications, and infrastructure. These methods are key components of reputation-based or policy-based trust management frameworks in cloud security. 			
4 (a)	**Describe the broad classes of cloud security risks as identified by recent studies.** **(a)Traditional Threats** **(b) Availability of Cloud Services** **(c) Third party control** ***Solution:*** Recent studies categorize cloud security risks into three broad classes: (a) Traditional Threats (b) Availability of Cloud Services (c) Third-Party Control (a) Traditional Threats: Traditional security threats in cloud computing include: • Phishing, SQL injection, cross-site scripting, and DDoS attacks. • These threats have a higher impact in cloud environments due to large user bases and shared infrastructure. • There is difficulty in assigning responsibility between cloud providers and users, and in identifying the source of attacks. • Authentication and authorization challenges are significant, especially with users having different privilege levels. • Multi-tenancy and hypervisor vulnerabilities introduce new attack channels.	**6M**	**CO4**	**L2**

	(b) Availability of Cloud Services ● Cloud service downtime due to failures, natural disasters, or attacks can severely impact users. ● Data lock-in restricts users from easily switching providers during outages. ● Users cannot always verify if the cloud application results are correct or if services are being delivered reliably. ● Complex system effects, like phase transitions, may cause unpredictable behaviors in cloud availability. (c) Third-Party Control ● Cloud users often have limited control or visibility over third parties used by the CSP. ● Subcontractors or hardware vendors may pose security risks if their reliability is not known. ● Risks include cloud provider espionage, data loss, or poor-quality infrastructure. ● Contractual agreements often place the entire responsibility on the user, not the CSP (e.g., AWS disclaimers). ● Auditing and compliance become difficult due to lack of transparency in CSP operations.			
(b)	**What are the main objectives of a Privacy Impact Assessment (PIA) tool?** ***Solution:*** A Privacy Impact Assessment (PIA) tool is designed to identify and assess privacy risks in information systems, particularly in cloud computing. The main objectives of a PIA tool include: 1. Risk Identification: To proactively identify privacy risks, including unauthorized data access, misuse, or disclosure in cloud environments. 2. Legal and Policy Compliance: To ensure that systems comply with privacy laws and policies such as the EU Data Protection Directive and national data protection regulations. 3. Stakeholder Awareness and Transparency: To inform stakeholders, including users and organizations, about how personal data is collected, stored, processed, and protected. 4. Support for Privacy-by-Design: To embed privacy protections into the design phase of cloud systems rather than making adjustments after deployment.	**4M**	**CO4**	**L2**

5	**Explain Hadoop architecture with a neat diagram**	**10M**	**CO5**	**L2**



Recent Hadoop Architecture :



- open source implementation of Map Reduce written in **Java**
- It uses HDFS
- **Hadoop core** is made up of the **Map Reduce Engine and HDFS**

HDFS follows the master slave architecture. the NameNode is the master and the DataNode is the slave in modern terminology.

The file is split into fixed sized blocks of 64 MB which currently is as big as 128 MB or 256 MB.

HDFS Features - HDFS fault tolerance provided by block replication, replica placement and heartbeat and blockreport messages.

High Throughput for large datasets primarily designed for batch processing using large block sizes to reduce overhead of metadata management. This allows for faster sequential reads.

The NameNode maps the blocks to the DataNodes. It also manages the file systems metadata and namespace.

DataNode : manages storage of the node responsible for storing and retrieving blocks.

To ensure High Availability, there are multiple namenodes, one that is active and others on standby.

HDFS Operation : control flow of HDFS read and write show distinct roles of NameNode and Data Nodes.

1. Reading a file :

- Client issues "open" request to namenode to get block locations
- NameNode returns replica locations of each block
- "read" function connects to closest / optimal DataNode
- The file is streamed to the client

2. Writing to a file

- client sends "create" request to NameNode.
- Data is streamed in blocks, replicated across dataNode using a pipeline.
- dataStreamer manages actual sending of data an handles replication.

In the Hadoop architecture the topmost layer is YARN-based resource management system. This orchestrates controla nd data flow of distributed applications.

In earlier versions, Job Tracker, TaskTrackers were replaced by YARN.

The resource scheduling in a cluster is done by the Resource Manager.

Each node runs NodeManger(NM) that manages the local resource and monitors execution containers.

For each MapReduce job, an Application Master (AM) is launched.

Hadoop 1.0 had slots for execution.

Hadoop 3.* has tasks that are executed within containers that are allocated flexible amounts of CPU and main memory.

Each map task processes one block of data.

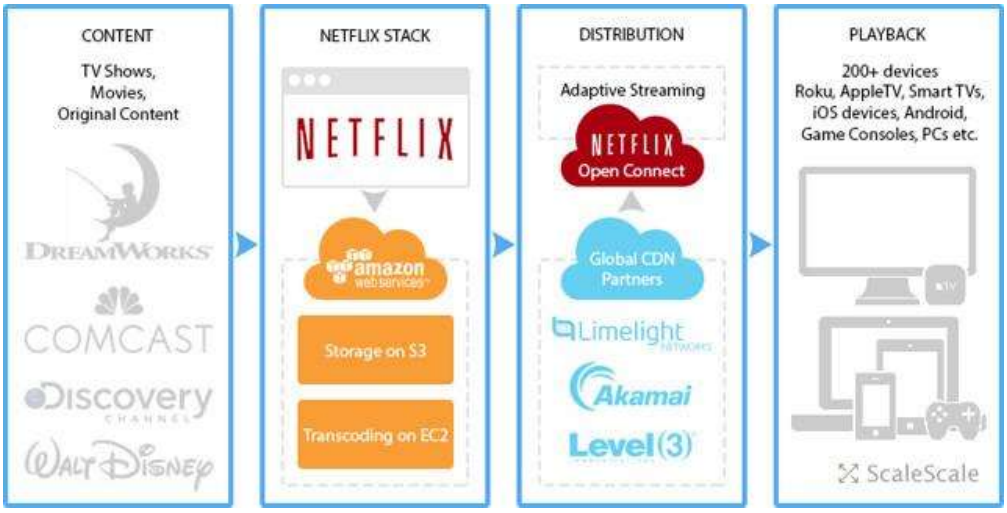
Running a Job in Hadoop :

- Job submission : submits a job via yarn.jar, hadoop.jar or by Hadoop API
- client contacts ResourceManager(RM) to request resources.
- RM launches Application Master (AM) for that job in a container on a NodeManager.
- Application Master(AM) for that job handles job lifecycle:
    - splits input files
    - requests containers from RM
    - assigns map & reduce tasks to containers on NodeManager(NM)
    - the job jar, input split metadata are submitted as part of the process

- Task Execution
    - tasks run in containers allocated by NodeManagers
    - Each container runs in separate JVM
    - Map tasks are placed near the data (data locality optimized)
    - Reduce tasks are not bound by data locality
- Monitoring & Heartbeat
    - NodeManagers(NM) send regular heartbeats to the ResourceManager(RM)
    - ApplicationMaster monitors progress, retrieves failed tasks, communicates with the client

- Completion
    - Once the tasks are done, signals completion to RM
    - containers are deallocated and the AM shuts down.

6	**Case Study : Netflix (A use case on AWS cloud)** **Netflix was started in 1997, with two founders. At first, Netflix was purely a movie rental service. In 2000, Netflix introduced a personalized movie recommendation system. Based on members' ratings to predict choices for all Netflix members. In 2007, streaming is introduced, allowing the members to watch films and series instantly. In 2008, Netflix was running relational databases in its own data centers.** **Because of growing customers, it is moved to AWS Cloud. They redesigned the architecture by choosing micro services. It took Netflix seven years to complete the migration to the cloud. In 2016, the last remaining data centers used by the streaming service were shut down.** **Bandwidth usage is more in 2018. Uses a separate network Netflix Open Connect. Company's proprietary Content Delivery Network (CDN). The Open Connect appliances store the video content and deliver it to client devices. After processing and storing, files are distributed across CDN locations. For processing requests EC2 instance is used. Membership surpassed 200 Million. Uses Amazon sage maker for accurate predictions. $1.67 billion revenue per month.**			



Service Type	Amazon web service
Hosting	EC2
Storage	S3
Network	Open connect
Database	DynamoDB
Events Driven Programming	AWS Lambda

	a. What are the advantages of Netflix moving to the cloud as compared to running their own streaming service? Scalability - ability to scale horizontally when workload increases or during peak times High Availability - through the Open Connect, and redundant fault tolerant storage, hardware failures are mitigated, if data in one zone is corrupted, it can be retrieved from another zone. Value Added Services - AWS Lamba is used to obtain real time insights for movie recommendations and other customer centric service Cost Benefits - compared to maintaining infrastructure and on premise, and spending for personnel, operational costs are reduced in cloud due to economies of scale	**3M**	**CO5**	**L2**
	b. How does Hadoop Distributed File System that Open Connect uses ensure fault tolerance of data using by Netflix? Block Replication - this is the core mechanism by HDFS to ensure fault tolerance and the default replication factor is usually set to 3. Replica Placement - the default strategy is one local, one same rack, one different rack. This optimizes communication costs. Currently this is improved by rack-aware placement also. Heartbeat and Block report messages : messages from DataNode to NameNode is given about the health of the blocks they store to regularly monitor the status.	**4M**	**CO5**	**L3**
	c. Eplain the nature of the data pertaining to Netflix that require DynamoDB for database and S3 for storage? Why are both required? Dynamo DB is a NoSQL database that may be used to store data about the users, watch history, session data, user preferences, information about the content that is hosted on netflix eg. movie - name, actor, title, producer, duration, etc. This data is semi-structured and needs to be scaled globally and it integrates the provisioning service given by AWS Lambda. S3 is used to store the transcoded video files, audio tracks, subtitles, logs for streaming sessions.	**3M**	**CO5**	**L3**

CI CCI HoD

M

CI CCI HoD

CO-PO and CO-PSO Mapping																			
Course Outcomes		**Blooms Level**	**Modules covered**	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**	**PSO1**	**PSO2**	**PSO3**	**PSO4**
CO1	Describe various cloud computing platforms and service providers.	**L2**	**1**	3	2	-	-	-	3	-	-	-	-	-	-	-	2	-	2
CO2	Illustrate the significance of various types of virtualization.	**L2**	**2**	3	2	-	-	2	-	-	-	-	-	-	-	-	2	-	2
CO3	Identify the architecture, delivery models and industrial platforms for cloud computing based applications.	**L2**	**3**	3	2	-	-	2	-	3	-	-	-	-	-	-	2	-	2
CO4	Analyze the role of security aspects in cloud computing.	**L2**	**4**	3	2	-	-	-	3	-	-	-	-	-	-	-	2	-	2
CO5	Demonstrate cloud applications in various fields using suitable cloud platforms..	**L2**	**5**	3	2	-	-	3	3	-	-	-	-	-	-	-	2	-	2

CO PO Mapping

COGNITIVE LEVEL	REVISED BLOOMS TAXONOMY KEYWORDS
L1	List, define, tell, describe, identify, show, label, collect, examine, tabulate, quote, name, who, when, where, etc.
L2	summarize, describe, interpret, contrast, predict, associate, distinguish, estimate, differentiate, discuss, extend
L3	Apply, demonstrate, calculate, complete, illustrate, show, solve, examine, modify, relate, change, classify, experiment, discover.
L4	Analyze, separate, order, explain, connect, classify, arrange, divide, compare, select, explain, infer.
L5	Assess, decide, rank, grade, test, measure, recommend, convince, select, judge, explain, discriminate, support, conclude, compare, summarize.

PROGRAM OUTCOMES (PO), PROGRAM SPECIFIC OUTCOMES (PSO)				CORRELATION LEVELS	
PO1	Engineering knowledge	PO7	Environment and sustainability	0	No Correlation
PO2	Problem analysis	PO8	Ethics	1	Slight/Low
PO3	Design/development of solutions	PO9	Individual and team work	2	Moderate/ Medium
PO4	Conduct investigations of complex problems	PO10	Communication	3	Substantial/ High
PO5	Modern tool usage	PO11	Project management and finance		
PO6	The Engineer and society	PO12	Life-long learning		
PSO1	Develop applications using different stacks of web and programming technologies				
PSO2	Design and develop secure, parallel, distributed, networked, and digital systems				
PSO3	Apply software engineering methods to design, develop, test and manage software systems.				
PSO4	Develop intelligent applications for business and industry				