ABHYASMITRA.IN

COMPREHENSIVE UNIVERSITY STUDY NOTES

CLOUD COMPUTING

Unit III — Cloud Platforms and Applications

Topics Covered in this Unit

Industrial Cloud Platforms: Amazon Web Services (AWS) Global Infrastructure, Regions & AZs
AWS Core Services: EC2 (Instances, AMIs, Pricing), S3 Storage Classes, EBS, Glacier & SimpleDB
AWS Database Ecosystem: Amazon RDS (Aurora, Multi-AZ) and DynamoDB Serverless NoSQL
Microsoft Azure Architecture: Core Concepts, ARM, SQL Azure & Azure App Services
Open-Source Private Cloud Platforms: OpenStack Modular Architecture, CloudStack & Eucalyptus
Comparative Evaluation: Commercial vs Open-Source Cloud Infrastructure Platforms
Data-Intensive & Emerging Cloud Applications: Smart Cities & Centralized IoT Dashboards
AI/ML in the Cloud: Deep-Dive Case Studies on Google Photos (Vision) & Amazon Alexa (NLU)
Cloud Computing in Healthcare & Biology: Gene Sequencing, Protein Folding & Wearable ECG Analysis
Geoscience & Earth Observation: Satellite Image Processing & Seismic Exploration on Cloud Clusters
High-Yield University Exam Question Bank with Detailed Model Answers & Unit Summary

Compiled in a structured, easy-to-understand, textbook style format
Specially curated for B.Tech / B.E. / BCA / MCA / B.Sc Computer Science & IT Students

COURSE SYLLABUS & MAPPING

[UNIT III SYLLABUS — PRESCRIBED UNIVERSITY CURRICULUM]

This section contains the official syllabus for Unit III. Verify with your university curriculum mapping.

Unit III Syllabus Breakdown:

Unit III Cloud Platforms and Applications: Industrial Cloud Platforms: Amazon Web Services (AWS)- AWS infrastructure, Components, Amazon Simple DB, Elastic Cloud Computing (EC2), Amazon Storage System, Amazon Database Services. Microsoft Azure: Azure core concepts, SQL Azure, and Application Services for managed runtimes. Open Source Platforms: Overview of OpenStack, CloudStack, and Eucalyptus for private cloud deployment. Cloud Applications: Data-Intensive & Emerging Applications: Smart Cities & IoT: Integrating sensor data from traffic, waste management, and power grids into a centralized cloud dashboard. AI/ML in the Cloud: Case study on Google Photos (image recognition) or Alexa (Natural Language Processing) using cloud-based TPU/GPU instances. Healthcare & Biology: Gene sequencing, protein folding and ECG analysis in the cloud. Geoscience: Satellite image processing and seismic data analysis using cloud clusters.

• Course Code: _____ • Semester / Year: _____

TABLE OF CONTENTS

3.1 Amazon Web Services (AWS) Ecosystem & Infrastructure

Definition: Amazon Web Services (AWS)

Amazon Web Services (AWS) is the world's most comprehensive and broadly adopted commercial cloud computing platform, offering over 200 fully featured services from data centers globally, spanning compute, storage, databases, analytics, networking, mobile, developer tools, and enterprise applications.

AWS Global Physical Infrastructure

- **1. AWS Regions:** A Region is a physical, separate geographic location in the world (e.g., us-east-1 N. Virginia, ap-south-1 Mumbai) that contains multiple isolated and physically separated data center clusters.
- **2. Availability Zones (AZs):** An AZ consists of one or more discrete data centers, each with independent redundant power, networking, and connectivity. AZs within a region are interconnected over high-bandwidth, ultra-low-latency private fiber networks.
- **3. Edge Locations / Points of Presence (PoP):** Hundreds of geographically distributed caching nodes used by Amazon CloudFront (CDN) and Route 53 (DNS) to deliver static and dynamic web content with minimum latency to global users.

Amazon Elastic Compute Cloud (Amazon EC2)

- **• Architectural Overview:** Amazon EC2 provides resizable virtual compute capacity in the cloud. It reduces the time required to obtain and boot new server instances to minutes, allowing users to quickly scale capacity up and down as computing requirements change.
- **• Amazon Machine Image (AMI):** A pre-configured template that contains an operating system, application server, and software packages required to launch a virtual instance.
- **• Instance Purchasing Models:**
  - **1. On-Demand Instances:** Pay for compute capacity by the second with no long-term commitments or upfront payments. Best for unpredictable workloads.
  - **2. Reserved Instances (RI) / Savings Plans:** Significant discount (up to 72%) in exchange for a 1-year or 3-year term commitment. Best for steady-state workloads.
  - **3. Spot Instances:** Bid on unused EC2 spare capacity at steep discounts (up to 90%). Subject to 2-minute interruption notice if AWS reclaims capacity. Ideal for fault-tolerant batch processing.
  - **4. Dedicated Hosts:** Physical servers dedicated entirely to a single customer, satisfying strict regulatory licensing requirements (BYOL).

AWS Storage Systems

- • **Amazon Simple Storage Service (Amazon S3):** An object storage service offering 99.999999999% (11 9's) of data durability. Data is stored as objects within containers called 'buckets'. Offers storage tiers: S3 Standard (frequently accessed), S3 Intelligent-Tiering, S3 Standard-IA (Infrequent Access), S3 Glacier Instant Retrieval, and S3 Glacier Deep Archive (cost-effective long-term cold storage).
- • **Amazon Elastic Block Store (Amazon EBS):** Provides high-performance raw block-level storage volumes for persistent attachment to EC2 instances. Formats include General Purpose SSD (gp2/gp3) and Provisioned IOPS SSD (io1/io2) for transactional databases.
- • **Amazon Elastic File System (Amazon EFS):** A scalable, elastic, fully managed POSIX-compliant NFS file system that can be concurrently mounted by thousands of EC2 instances.

AWS Database Services

- • **Amazon SimpleDB:** One of AWS's earliest cloud NoSQL data stores (launched in 2007). Provided simple real-time querying of non-relational structured data with automated indexing, serving as the conceptual precursor to DynamoDB.
- • **Amazon RDS (Relational Database Service):** A managed service that automates database provisioning, OS patching, automated daily backups, multi-AZ synchronous replication for high availability, and read replicas. Supports engines: Amazon Aurora (MySQL/PostgreSQL compatible with 5x throughput), PostgreSQL, MySQL, MariaDB, Oracle, and Microsoft SQL Server.
- • **Amazon DynamoDB:** A fully managed, serverless, key-value and document NoSQL database designed for internet-scale applications requiring single-digit millisecond latency at any scale. Features built-in in-memory caching (DAX) and global multi-region active-active replication.

3.2 Microsoft Azure Platform & Core Services

Definition: Microsoft Azure

Microsoft Azure is a comprehensive public cloud computing platform developed by Microsoft that provides IaaS, PaaS, and SaaS solutions across an extensive global network of Microsoft-managed datacenters, with deep native integration for enterprise Windows Server, Active Directory, and .NET ecosystems.

Azure Core Concepts & Architecture

- • **Azure Resource Manager (ARM):** The centralized management and deployment layer that enables administrators to provision, manage, tag, and lock Azure resources using declarative JSON/Bicep templates.

- • **Resource Groups:** Logical containers that hold related Azure resources (VMs, storage accounts, virtual networks) that share the same application lifecycle, billing, and access control policies.
- • **Azure Subscriptions & Management Groups:** Hierarchical organizational boundaries used for billing allocation, spending quotas, and enterprise governance (Azure Policy).

Azure App Services & Managed Runtimes

- • **Azure App Service:** A fully managed HTTP-based Platform as a Service (PaaS) for hosting web applications, RESTful APIs, and mobile backends in diverse programming languages (.NET, Java, Node.js, Python, PHP) without managing underlying infrastructure.
- • **Key Capabilities:** Automated OS patching, integrated SSL certificates, custom domain binding, auto-scaling up to 100 instances, deployment slots (Blue/Green staging environments), and native CI/CD integration with GitHub Actions.
- • **Azure Functions:** An event-driven serverless compute platform that executes snippets of code in response to triggers (HTTP requests, database changes, queue messages) with zero infrastructure management.

SQL Azure & Azure Storage

- • **Azure SQL Database (SQL Azure):** A fully managed Database as a Service (DBaaS) running the latest stable Microsoft SQL Server engine with 99.995% availability SLA, automated AI-driven query tuning, threat detection, and active geo-replication.
- • **Azure Blob Storage:** Massively scalable object storage for unstructured data (audio, video, log files, backups) categorized into Hot, Cool, Cold, and Archive access tiers.
- • **Azure Cosmos DB:** A globally distributed, multi-model NoSQL database service offering guaranteed single-digit millisecond latency SLAs and supporting APIs for MongoDB, Cassandra, Gremlin, and SQL.

3.3 Open Source Cloud Platforms: OpenStack, CloudStack & Eucalyptus

Open-source cloud software enables enterprises, governments, and research institutions to build private and hybrid cloud infrastructures on commodity hardware without proprietary vendor lock-in:

1. OpenStack

- • **Architecture & Philosophy:** OpenStack is a modular, open-source cloud operating system composed of interrelated microservices that control large pools of compute, storage, and networking resources throughout a datacenter:

- • **Nova (Compute):** Manages the lifecycle of virtual machines and bare-metal compute instances. Interacts with hypervisors like KVM, QEMU, and Xen.
- • **Swift (Object Storage):** Distributed, eventually consistent object storage system for storing unstructured media and backups.
- • **Cinder (Block Storage):** Provides persistent block-level storage volumes for Nova compute instances.
- • **Neutron (Networking):** Software-defined networking service managing IP addresses, VLANs, subnets, routers, and firewall rules.
- • **Keystone (Identity Service):** Centralized authentication and authorization engine for all OpenStack components via token-based access.
- • **Glance (Image Service):** Repository for discovering, registering, and retrieving VM disk images (e.g., ISO, QCOW2).
- • **Horizon (Dashboard):** Web-based graphical user interface for administrators and end-users.

OPENSTACK MODULAR ARCHITECTURE INTERACTION

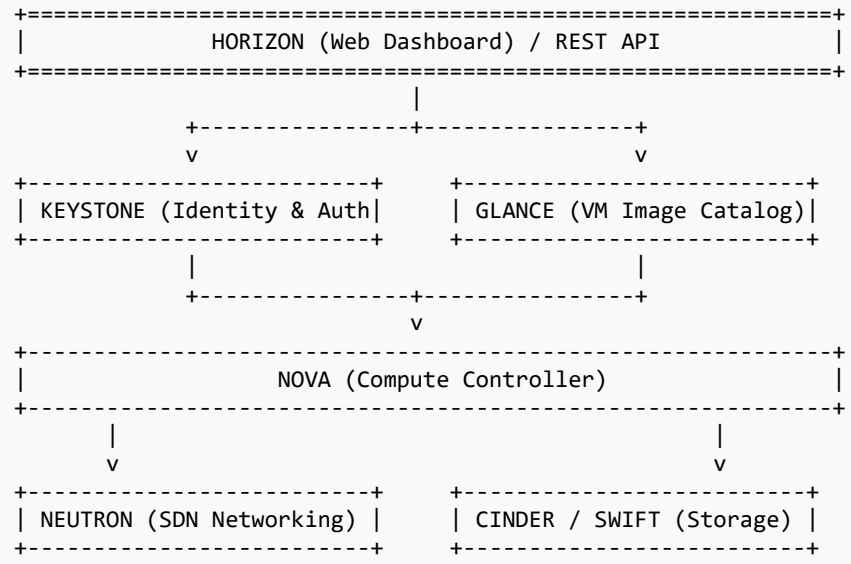

Figure: Inter-service API communication model across core OpenStack modules

2. Apache CloudStack

- • **Architectural Overview:** Apache CloudStack is a turnkey open-source Infrastructure as a Service (IaaS) platform designed for ease of installation, high scalability, and robust multi-tenancy.
- • **Hierarchy of Infrastructure:**
  - **1. Region:** The highest organizational level, consisting of multiple geographic Zones.
  - **2. Zone:** Corresponds to a single physical datacenter; contains one or more Pods and secondary storage.

- **3. Pod:** A single rack of hardware containing a layer-2 switch and one or more Clusters.
- **4. Cluster:** A collection of homogeneous physical hypervisor hosts sharing primary network storage.
- **5. Host:** A single physical server running a hypervisor (KVM, XenServer, VMware ESXi).

3. Eucalyptus

- • **Architectural Overview:** Eucalyptus (Elastic Utility Computing Architecture for Linking Your Programs To Useful Systems) is an open-source private cloud software platform uniquely designed from inception to be 100% API-compatible with Amazon Web Services (AWS).
- • **Core Components:** Cloud Controller (CLC - centralized gateway), Walrus (S3-compatible storage), Cluster Controller (CC), Storage Controller (SC - EBS-compatible), and Node Controller (NC - hypervisor host manager).

Platform	Architectural Modularity	AWS API Compatibility	Hypervisor Support	Deployment Complexity
OpenStack	Extremely modular (20+ independent Python microservices).	Partial (via ec2-api add-on).	KVM (primary), Xen, Hyper-V, VMware.	High; requires advanced networking and Linux orchestration skills.
Apache CloudStack	Monolithic Management Server with modular plugin agents.	High native AWS API support.	KVM, Citrix XenServer, VMware ESXi.	Moderate; turnkey installation with web UI out of the box.
Eucalyptus	Hierarchical five-tier controller architecture.	100% direct AWS EC2/S3 API compatibility.	KVM, VMware vSphere.	Moderate; purpose-built for AWS private cloud clones.

3.4 Data-Intensive & Emerging Cloud Applications

The elastic computational power and scalable storage of modern cloud infrastructure have catalyzed transformative innovations across diverse scientific, urban, and industrial domains:

1. Smart Cities & IoT: Centralized Sensor Ingestion Pipelines

- • **Architecture & Integration:** Millions of heterogeneous IoT sensors deployed across urban infrastructure (traffic cameras, smart water meters, GPS on public transit, air quality sensors, electrical power grids) generate continuous high-velocity telemetry data.
- • **Cloud Ingestion Pipeline:** Edge gateways collect sensor packets and transmit them over MQTT/HTTPS protocols into cloud stream ingestion buffers (AWS Kinesis, Apache Kafka on Azure Event Hubs).
- • **Real-Time Processing & Dashboarding:** Cloud stream processing engines (Apache Flink, Spark Streaming) aggregate and analyze metrics in real-time, feeding an integrated municipality dashboard for automated traffic light synchronization, dynamic power grid load shedding, and smart waste management dispatching.

2. AI/ML in the Cloud: Deep-Dive Domain Case Studies

- • **Case Study A: Google Photos (Image Recognition & Semantic Search):** Google Photos processes billions of user photos using cloud-hosted Tensor Processing Unit (TPU) clusters. Deep Convolutional Neural Networks (CNNs) perform facial recognition, landmark identification, object detection, and vector semantic clustering, allowing users to search natural terms like 'photos of sunset on the beach' without manual tagging.
- • **Case Study B: Amazon Alexa (Speech Processing Pipeline):** When a user speaks to an Echo device, the local device detects the wake word and streams compressed audio to AWS. Cloud-based Deep Learning pipelines execute:
  - **1. Automatic Speech Recognition (ASR):** Converts acoustic waveforms into raw text strings.
  - **2. Natural Language Understanding (NLU):** Extracts user intent and parameters (e.g., Intent: 'PlayMusic', Artist: 'Mozart').
  - **3. Text-to-Speech (TTS):** Synthesizes expressive audio responses in milliseconds using neural speech engines.

3. Healthcare & Biology: High-Throughput Genomic Computing

- • **Next-Generation Sequencing (NGS):** Sequencing a single human genome generates hundreds of gigabytes of raw FASTQ genomic data. Cloud compute clusters parallelize genome alignment (BWA-MEM) and variant calling algorithms (GATK), reducing processing time from weeks on local PCs to under an hour in the cloud.
- • **Protein Structure Prediction (AlphaFold):** Leverages cloud GPUs to model 3D protein folding configurations, accelerating drug discovery and vaccine synthesis.
- • **Real-Time Wearable ECG Anomaly Detection:** Wearable smartwatches stream ECG cardiac signals to cloud endpoints where recurrent neural networks (LSTM) detect arrhythmias, atrial fibrillation, and myocardial infarction risk in real time, triggering emergency medical alerts.

4. Geoscience & Earth Observation: Satellite & Seismic Cloud Processing

- • **Satellite Remote Sensing:** Petabytes of multi-spectral satellite imagery (NASA Landsat, ESA Sentinel) hosted in cloud data lakes (e.g., AWS Registry of Open Data, Google Earth Engine) are processed using distributed clusters to monitor deforestation, agricultural crop health, urban expansion, and climate change.
- • **Seismic Data Wave Simulation:** Oil and gas exploration involves processing terabytes of reflected acoustic wave data from geological seismic surveys. Massively parallel MPI clusters on cloud IaaS execute 3D Full Waveform Inversion (FWI) algorithms to map subterranean geological strata.

3.5 High-Yield University Exam Question Bank & Model Answers

Part A: Short Answer Questions (2–3 Marks)

- **Q1. What is an Availability Zone (AZ) in AWS? How does it differ from a Region?**

Answer: An AWS Region is a separate geographic area (e.g., ap-south-1 Mumbai). An Availability Zone (AZ) is one or more discrete, physically isolated data centers within that region, equipped with redundant power, cooling, and low-latency networking. Multiple AZs per region provide fault isolation and high availability.

- **Q2. Explain the purpose of Amazon S3 storage classes.**

Answer: S3 storage classes allow customers to optimize storage costs based on data access patterns. Frequently accessed data uses S3 Standard; infrequently accessed data with instant retrieval uses S3 Standard-IA; and cold, rarely accessed archival data uses S3 Glacier or Glacier Deep Archive at fraction of the cost.

- **Q3. What is OpenStack Nova and how does it interact with Keystone and Glance?**

Answer: Nova is OpenStack's core compute service managing VM lifecycles. It queries Keystone to authenticate user credentials and authorization tokens, and queries Glance to retrieve bootable VM disk images before launching instances on physical hypervisor hosts.

- **Q4. Describe the role of Azure Resource Manager (ARM).**

Answer: ARM is Microsoft Azure's centralized management layer that allows administrators to deploy, configure, secure, and manage resources (VMs, databases, networks) as a coordinated group using declarative JSON or Bicep templates.

Part B: Long Answer Questions (8–10 Marks)

- **Q5. Explain the modular architecture of OpenStack. Describe the function of Nova, Swift, Cinder, Neutron, Keystone, and Glance with an architectural diagram.**

Model Answer Outline:

- 1. Introduction to OpenStack as an open-source IaaS cloud operating system.
- 2. Architectural breakdown: Nova (Compute orchestration), Swift (Object storage), Cinder (Block storage volumes), Neutron (SDN networking, IPAM, firewalls), Keystone (Identity and token authentication), Glance (VM disk image catalog), Horizon (Web dashboard).
- 3. Interaction workflow: User authenticates via Keystone -> Selects image from Glance -> Requests Nova instance -> Nova attaches Cinder volume and Neutron virtual NIC.
- 4. Comprehensive architectural block diagram illustrating API gateway and message queue (RabbitMQ) communication.

- **Q6. Discuss the application of Cloud Computing in IoT-enabled Smart Cities and Healthcare with appropriate case studies and architectural pipelines.**

Model Answer Outline:

- 1. Smart City IoT Architecture: End-to-end data pipeline from physical urban sensors (traffic, waste, electricity) -> Edge Gateways -> Cloud Stream Ingestion (Kafka/Kinesis) -> Real-time Analytics (Spark/Flink) -> Central Municipal Dashboard.
- 2. Healthcare & Genomic Case Study: High-Throughput Computing (HTC) for DNA sequence alignment (BWA-MEM/GATK), AlphaFold protein folding on GPU instances, and cloud-assisted wearable ECG telemetry for real-time arrhythmia alerting.
- 3. Critical Benefits: Massive elastic scalability, fault tolerance, global accessibility, and cost reduction through serverless and automated data pipelines.

3.6 Unit III Summary & Quick Revision Sheet

Cloud Platform / Domain	Key Takeaways & University Exam Summary
AWS Cloud Ecosystem	EC2 (Compute, On-demand/Reserved/Spot), S3 (Object storage with Glacier tiers), EBS (Block storage), RDS (Relational DBs), DynamoDB (Serverless NoSQL).
Microsoft Azure	ARM (Template-based resource management), Azure App Services (Managed web runtime PaaS), SQL Azure (DBaaS), Cosmos DB (Multi-model NoSQL).
Open-Source Private Clouds	OpenStack (Modular Python microservices: Nova, Swift, Cinder, Neutron, Keystone, Glance), CloudStack (Turnkey Zones/Pods/Clusters), Eucalyptus (AWS API clone).
Smart Cities & IoT	Massive sensor data ingestion via MQTT/Kafka, real-time analytics for traffic, energy grid load shedding, and municipal waste monitoring.
AI/ML & Healthcare	Cloud GPU/TPU instances for Google Photos vision models, Alexa speech pipelines, Next-Gen DNA sequencing, and seismic exploration clusters.