Module 1: Introduction to Information Storage

Lesson 1:

1. **List and describe the four types of digital data.**
    a. **Structured:** Organized in fixed fields (e.g., relational databases).
    b. **Semi-structured:** Has a pattern but no formal model (e.g., XML files).
    c. **Quasi-structured:** Erratic formats requiring processing (e.g., clickstream data).
    d. **Unstructured:** No inherent format (e.g., videos, PDFs).

Lesson 2:

2. **What are the three main components of a data center?**
    a. **Facility:** Building/floor space with power/cooling infrastructure.
    b. **IT equipment:** Compute, storage, and network systems.
    c. **Support infrastructure:** Power backups, HVAC, security systems.
3. **List five key characteristics of a data center.**
    a. Availability, security, capacity, scalability, performance, data integrity, manageability.
4. **Explain the data center management process of *provisioning*.**
    a. Allocating/configuring resources (e.g., storage, servers) to meet business needs for capacity, performance, and security.

Lesson 3:

5. **Compare the three computing platforms in terms of users and applications.**
    - **Platform 1 (Mainframes):** Millions of users, thousands of apps.
    - **Platform 2 (Client-Server):** Hundreds of millions of users, tens of thousands of apps.
    - **Platform 3 (Cloud/Big Data/Mobile/Social):** Billions of users, millions of apps.
6. **What were the limitations of mainframes (Platform 1)?**
        - High CAPEX/OPEX, significant space/energy requirements, limited scalability.
7. **Describe the *client-server model* and its challenges.**
- Distributed architecture where servers process client requests. Challenges: IT silos, maintenance overhead, scalability issues.
8. **What are the four pillars of the third platform?**
    - Cloud computing,
    - Big Data analytics,
    - mobile technologies,
    - social networking.

Module 2: Third Platform Technologies

Lesson 1:

1. **List and explain the five essential characteristics of cloud computing.**
    a. **On-demand self-service:** Consumers can provision resources automatically without human interaction.
    b. **Broad network access:** Services are accessible over the network via standard mechanisms (e.g., mobile, laptops).
    c. **Resource pooling:** Resources are pooled to serve multiple consumers using a multi-tenant model.
    d. **Rapid elasticity:** Resources can be scaled outward/inward dynamically to meet demand.
    e. **Measured service:** Resource usage is monitored, controlled, and reported for transparency.
2. **Compare IaaS, PaaS, and SaaS.**
    a. **IaaS:** Provides fundamental computing resources (e.g., virtual machines, storage). Consumers control OS/apps (e.g., AWS EC2).
    b. **PaaS:** Offers platforms for app development/deployment (e.g., Google App Engine). Consumers control apps but not infrastructure.
    c. **SaaS:** Delivers ready-to-use software (e.g., Gmail). Consumers have minimal control (limited configuration only).
3. **Describe the four cloud deployment models.**
    a. **Public cloud:** Open to the general public (e.g., AWS, Azure).
    b. **Private cloud:** Exclusive to a single organization (on-premise or hosted externally).
    c. **Community cloud:** Shared by organizations with common concerns (e.g., healthcare providers).
    d. **Hybrid cloud:** Combines two or more deployment models (e.g., private + public for cloud bursting).

Lesson 2:

4. **Define Big Data and explain its three key characteristics (3Vs).**
    a. *Answer:* Big Data refers to large, complex datasets requiring advanced processing. The 3Vs are:
        - **Volume:** Massive data quantities (e.g., petabytes).
        - **Velocity:** High-speed data generation/processing (e.g., real-time analytics).

- **Variety:** Diverse data types (structured, unstructured, semi-structured).

5. **Explain the roles of storage, MapReduce, and query in the SMAQ stack.**
    b. **Storage:** Distributed file systems (e.g., HDFS) for scalable data storage.
    c. **MapReduce:** Parallel processing framework for batch analytics (e.g., Hadoop).
    d. **Query:** NoSQL databases/user-friendly interfaces for data retrieval/analysis.
6. **How does MapReduce work? Describe the two phases.**
    e. **Map Phase:** Splits input data into chunks, processes them in parallel, and generates intermediate key-value pairs.
    f. **Reduce Phase:** Aggregates intermediate results to produce the final output (e.g., word count in documents).
7. **Compare data warehouses and data lakes.**
    a. **Data warehouse:** Structured, schema-on-write, optimized for SQL queries (e.g., historical sales data).
    b. **Data lake:** Stores raw data in native format, schema-on-read, flexible for diverse analytics (e.g., social media logs).
8. **Provide two use cases of Big Data analytics in healthcare and retail.**
    - **Healthcare:** Real-time patient monitoring, predictive diagnostics.
    - **Retail:** Customer behavior analysis, dynamic pricing optimization.
9. **Name challenges of each Big Data characteristic:**
    - **Volume:** Storage and analysis.
    - **Velocity:** Real-time processing demands.
    - **Variety:** Integration, and analysis.
    - **Variability:** Gathering and interpretation.
    - **Veracity:** transforming and trusting data.
10. **Name Big Data characteristic:**
    - **Volume:** Massive volumes of data.
    - **Velocity:** Rapidly changing data.
    - **Variety:** Diverse data from numerous sources.
    - **Variability:** Constantly changing meaning of data.
    - **Veracity:** Varying quality and reliability of data.
    - **Value:** Cost effectiveness and business value.

Module 3: Data Center Environment

Lesson 1:

1. **What are the five layers of a data center infrastructure?**
    a. Physical infrastructure
    b. Virtual infrastructure

c. Software-defined infrastructure
d. Orchestration
e. Services

2. **What are the three cross-layer functions in a data center?**
    a. Business continuity
    b. Security
    c. Management
3. **Compare best-of-breed infrastructure and converged infrastructure.**
    a. **Best-of-breed:** Integrates components from multiple vendors, prevents vendor lock-in, allows repurposing existing infrastructure, but requires more time and cost for integration.
    b. **Converged infrastructure:** Pre-packaged solution with optimized hardware/software, faster deployment, lower power/space requirements, but may limit vendor flexibility.
4. **What is the role of the orchestration layer in a data center?**
    a. Provides workflows for automated tasks, coordinates provisioning across layers, and interacts with software-defined infrastructure and cross-layer functions.
5. **How does the service layer function in a data center?**
    a. Delivers IT resources as services via a service catalog and self-service portal, allowing users to access resources without ownership liabilities.

Lesson 2:

6. **List the physical components of a compute system.**
    a. Processor, RAM, ROM, motherboard, chipset, secondary storage (HDD/SSD).
7. **What are the logical components of a compute system?**
    a. Operating system, virtual memory, logical volume manager (LVM), file system.
8. **Compare tower, rack-mounted, and blade compute systems.**
    - **Tower:** Standalone, bulky, high floor space.
    - **Rack-mounted:** Vertical stacking in racks, better space efficiency, but generates heat.
    - **Blade:** High-density, shared power/cooling, modular design.

Lesson 3:

9. **What is compute virtualization, and what are its benefits?**
    - Abstracts physical hardware to run multiple OS instances on a single compute system.

- Benefits: Server consolidation, improved resource utilization, flexible provisioning, reduced costs.

10. **Compare bare-metal and hosted hypervisors.**
    - **Bare-metal:** Installed directly on hardware, more efficient, used in enterprise environments.
    - **Hosted:** Runs on an OS, suitable for testing/development, adds overhead.
11. **What are the key files that make up a virtual machine (VM)?**
    - Configuration file,
    - virtual disk file,
    - memory state file,
    - snapshot file,
    - log file.
12. **Explain application virtualization and its techniques.**
    Decouples apps from OS/hardware. Techniques:
    - **Encapsulation:** Self-contained executable package.
    - **Presentation:** Remote execution with UI transmitted to client.
    - **Streaming:** App delivered in portions for local execution.
13. **Compare Remote Desktop Services (RDS) and Virtual Desktop Infrastructure (VDI).**
    - **RDS:** Shared OS/apps, high user density, cost-effective.
    - **VDI:** Dedicated VM per user, full isolation, persistent/non-persistent desktops.

Lesson 4:

14. **Compare server-centric and information-centric storage architectures.**
    - **Server-centric:** Storage dedicated to each server, limited scalability.
    - **Information-centric:** Shared storage pool via SAN, centralized management, better scalability.
15. **List and describe the types of storage devices.**
    - **Magnetic disk:** High capacity, random access (HDD).
    - **SSD:** Fast, low latency, high cost (flash-based).
    - **Tape:** Sequential access, low-cost archival.
    - **Optical disc:** WORM (Write Once Read Many), long-term storage.
16. **Explain compute-to-storage connectivity components.**
    - **Physical:** Host Bus Adapter (HBA), ports, cables.
    - **Protocols:** SCSI, FC, iSCSI, SATA, SAS.
17. **Compare IDE/ATA, SCSI, and Fibre Channel (FC) protocols.**
    - **IDE/ATA:** Low-cost, legacy (PATA), replaced by SATA.

- **SCSI:** High performance, supports multiple devices.
- **FC:** High-speed, used in SANs (16 Gb/s in FC 16).

Module 4: Intelligent Storage Systems (ISS)

Lesson 1:

1. **What are the key components of an Intelligent Storage System (ISS)?**
    - **Answer:** The two key components are:
        - **Controller:** A compute system running a purpose-built OS, handling I/O processing, RAID protection, replication, provisioning, and cache management. Controllers can be block-based, file-based, object-based, or unified.
        - **Storage:** Comprises HDDs, SSDs, or a combination of both.
2. **Describe the components of a Hard Disk Drive (HDD).**
    - **Answer:** Key components include:
        - **Platter:** Rigid, magnetic-coated disks storing data.
        - **Spindle:** Rotates platters at speeds like 5,400–15,000 RPM.
        - **Read/Write Head:** Reads/writes data magnetically without touching the platter.
        - **Actuator Arm Assembly:** Positions R/W heads.
        - **Controller Board:** Manages data transfer, motor control, and communication.
3. **Explain Logical Block Addressing (LBA) and its advantage over CHS addressing.**
    - **Answer:** LBA uses linear addressing (block numbers) instead of Cylinder-Head-Sector (CHS) addresses. It simplifies OS interaction by abstracting disk geometry, allowing the controller to translate LBA to physical sectors.
4. **What factors determine HDD performance?**

    Answer:
        - **Seek Time:** Time to position R/W heads (average: 3–15 ms).
        - **Rotational Latency:** Time for platter to rotate data under the head (e.g., 2 ms for 15K RPM).
        - **Data Transfer Rate:** Speed of data movement (internal: platter-to-buffer; external: buffer-to-HBA).

Lesson 3: RAID

5. **What is RAID, and what are its primary purposes?**

a. **Answer:** RAID (Redundant Array of Independent Disks) combines disks into a logical unit to:
    - **Protect data** via redundancy (mirroring/parity).
    - **Improve performance** by parallel I/O processing.

6. **software RAID and hardware RAID.**
    b. **Answer:**
        - **Software RAID:** Uses OS-level tools; lower cost but impacts CPU performance and lacks advanced features.
        - **Hardware RAID:** Uses dedicated controllers; offloads processing, supports more RAID levels, and is scalable.
7. **Describe the three RAID techniques.**
    - **Answer:**
        - **Striping:** Spreads data across disks for parallel I/O (e.g., RAID 0).
        - **Mirroring:** Duplicates data on two disks (e.g., RAID 1).
        - **Parity:** Uses mathematical constructs for fault tolerance (e.g., RAID 5/6).
8. **How does RAID 6 differ from RAID 5?**
    - **Answer:** RAID 6 uses **dual parity** to tolerate two disk failures but has a higher write penalty (6 I/Os per write) and lower usable capacity.

Module 6: File-based Storage System (NAS)

Lesson 1:

1. **What is NAS, and how does it differ from general-purpose servers?**
    a. *Answer:* NAS (Network Attached Storage) is a dedicated, high-performance file-sharing device optimized for file I/O using protocols like NFS/CIFS. Unlike general-purpose servers, NAS has a specialized OS for file serving, enabling better performance and server consolidation.
2. **List the two main components of a NAS system and their functions.**
    a. **Controller/NAS Head:** Handles file I/O, runs a specialized OS, and manages RAID/LUNs.
    b. **Storage:** Persistently stores data using devices like SSDs, SAS, or SATA drives.
3. **Compare scale-up and scale-out NAS architectures.**
    a. **Scale-up:** Adds capacity/performance to a single NAS system but has a fixed ceiling.
    b. **Scale-out:** Pools multiple nodes into a cluster for unlimited scalability and parallel performance.
4. **Describe how a scale-out NAS handles a write operation.**

- *Answer:* The client sends a file to one node, which stripes the file across multiple nodes in the cluster for distributed storage.

5. **What are the three common file access methods for NAS?**
    - NFS (Network File System),
    - CIFS (Common Internet File System),
    - HDFS (Hadoop Distributed File System).

Module 9: Fibre Channel (FC) SAN

Lesson 1:

1. **List the key benefits of a SAN.**
    - Consolidation and sharing of storage resources across multiple compute systems, improving utilization.
    - Centralized storage management, reducing complexity and costs.
    - Connectivity across geographically dispersed locations for data access, replication, and remote backup.
    - Facilitates disaster recovery through long-distance data replication.
2. **What are the third platform requirements for an effective SAN infrastructure?**
    - High throughput for performance-intensive applications.
    - Interconnectivity among a large number of devices over wide locations.
    - Elastic and non-disruptive scaling for horizontally scaled applications.
    - Automated, policy-driven infrastructure configuration.
    - Simplified, flexible, and agile management operations.
3. **Name the technology solutions that meet third platform SAN requirements.**
    - Software-defined networking (SDN).
    - SAN implementations: Fibre Channel (FC) SAN, IP SAN, and Fibre Channel over Ethernet (FCoE) SAN.
    - Virtualization in SAN.
4. **What are the benefits of a software-defined SAN?**
    - Centralized control: Single point of management for the entire SAN.
    - Policy-based automation: Reduces manual, error-prone tasks.
    - Simplified, agile management: Abstracts operational complexity for easier configuration.

Lesson 2:

5. **List the key components of an FC SAN.**
    a. **Network adapters:** FC HBAs (in compute systems) and front-end adapters (in storage systems).
    b. **Cables:** Copper (short distance) or optical fiber (multimode/single-mode for long distance).

c. **Interconnecting devices:** FC hubs, switches, and directors.

6. **Compare FC hubs, switches, and directors.**
    d. **FC Hub:** Connects nodes in a shared loop (FC-AL), limited scalability.
    e. **FC Switch:** Dedicated paths for nodes, fixed port count, some redundancy.
    f. **FC Director:** High-end modular switch with hot-swappable, redundant components and high port density.
7. **Describe the three FC interconnectivity options.**
    g. **Point-to-point:** Direct connection between two nodes (limited scalability).
    h. **FC-AL (Arbitrated Loop):** Shared loop where devices arbitrate for access (lower performance).
    i. **FC-SW (Switched Fabric):** Uses switches for dedicated paths, scalable, and minimally disruptive to changes.
8. **Explain the FC port types in a switched fabric:**
    a. **N_Port:** Node port (e.g., HBA or storage port connected to a switch).
    b. **E_Port:** Expansion port (connects two switches via ISL).
    c. **F_Port:** Fabric port (connects an N_Port to a switch).
    d. **G_Port:** Generic port that auto-negotiates as E_Port or F_Port.

Module 12: Introduction to Business Continuity

Lesson 1:

1. **Define Business Continuity (BC) and its primary goal.**
    a. *Answer:* Business Continuity is a process that prepares for, responds to, and recovers from system outages to ensure uninterrupted business operations. Its goal is to maintain information availability and meet service-level agreements (SLAs) through proactive and reactive measures.
2. **What are the three key aspects of Information Availability (IA)?**
    - **Accessibility:** Information must be available to authorized users when needed.
    - **Reliability:** Data must be accurate and uncorrupted.
    - **Timeliness:** Information must be accessible within specified time windows.
3. **List five causes of information unavailability.**
    - Application failure (e.g., logic errors).
    - Data loss (e.g., corruption or human error).
    - Infrastructure component failure (e.g., server/network downtime).
    - Data center/site outages (e.g., power failure, natural disasters).
    - IT infrastructure refreshes or maintenance.

4. **Explain the difference between planned and unplanned outages.**
    - **Planned outages:** Scheduled activities like maintenance, upgrades, or migrations.
    - **Unplanned outages:** Unexpected events like hardware failures, human errors, or disasters.
5. **Describe the financial and operational impacts of information unavailability.**
    - **Lost productivity:** Employee downtime × hourly wages.
    - **Lost revenue:** Direct sales loss, compensatory payments, future revenue decline.
    - **Reputation damage:** Loss of trust from customers, partners, or investors.
    - **Other costs:** Overtime, equipment rentals, or expedited shipping.
6. **How are MTBF and MTTR used to measure Information Availability (IA)?**
    - **MTBF (Mean Time Between Failures):** Average uptime between failures (reliability metric).
    - **MTTR (Mean Time To Repair):** Average downtime to restore a failed component.
7. **Define Disaster Recovery (DR) and its relationship to BC.**
    a. DR is a subset of BC focused on restoring IT systems after disasters. It involves secondary sites (DR sites) or cloud-based solutions (DRaaS) to resume operations quickly.
8. **What are the advantages of Disaster Recovery-as-a-Service (DRaaS)?**
    - Reduces costs (no need for physical DR infrastructure).
    - Scalable and flexible cloud-based solutions.
    - Faster recovery times and automated failover.
9. **List the stages of the BC planning lifecycle.**
    - **Risk Assessment:** Identify potential threats (e.g., disasters, cyberattacks).
    - **Business Impact Analysis (BIA):** Evaluate critical functions and downtime costs.
    - **Strategy Development:** Design BC/DR solutions (e.g., redundancy, backups).
    - **Implementation:** Deploy technologies (e.g., failover clusters, replication).
    - **Testing & Maintenance:** Regular drills and plan updates.
10. **What are the BC requirements for third-platform environments (cloud, mobile, big data)?**
    - High scalability to handle distributed workloads.
    - Automation for rapid recovery and policy-driven management.
    - Integration with cloud-based DR solutions (e.g., DRaaS).
    - Support for real-time data replication and analytics.