QUESTION BANK

Name of the Department : M.E - COMPUTER SCIENCE AND ENGINEERING
Subject Code & Name : INFORMATION STORAGE MANAGEMENT
Year & Semester : II & III

UNIT 1

PART-A

1. Define Data with example.

Data is a collection of raw facts from which conclusions may be drawn. Handwritten letters, a printed book, a family photograph, a movie on video tape, printed and duly signed copies of mortgage papers, a bank's ledgers, and an account holder's passbooks are all examples of data.

2. What do you mean by information? List the characteristics of information

Information is organized or classified data, which has some meaningful values for the receiver. Information is the processed data on which decisions and actions are based.

For the decision to be meaningful, the processed data must qualify for the following characteristics

- Timely − Information should be available when required.
- Accuracy − Information should be accurate.
- Completeness − Information should be complete.

3. Differentiate between data and information. (NOV 2015)

	DATA	**INFORMATION**
Meaning	Data is raw, unorganized facts that need to be processed. Data can be something simple and seemingly random and useless until it is organized.	When data is processed, organized, structured or presented in a given context so as to make it useful, it is called information
Example	Each student's test score is one	The average score of a class or of the

	piece of data.	entire school is information that can be derived from the given data.
Etymology	Data" comes from a singular Latin word, datum, which originally meant "something given." Its early usage dates back to the 1600s. Over time "data" has become the plural of datum.	"Information" is an older word that dates back to the 1300s and has Old French and Middle English origins. It has always referred to "the act of informing, " usually in regard to education, instruction, or other knowledge communication.

4. List of some of the factors that have contributed to the growth of digital data.(NOV 2019)

The following is a list of some of the factors that have contributed to the growth of digital data:

- **Increase in data processing capabilities:** Modern-day computers provide a significant increase in processing and storage capabilities. This enables the conversion of various types of content and media from conventional forms to digital formats.
- **Lower cost of digital storage:** Technological advances and decrease in the cost of storage devices have provided low-cost solutions and encouraged the development of less expensive data storage devices. This cost benefit has increased the rate at which data is being generated and stored.
- **Affordable and faster communication technology:** The rate of sharing digital data is now much faster than traditional approaches. A handwritten letter may take a week to reach its destination, whereas it only takes a few seconds for an e-mail message to reach its recipient.

5. What is the use of data creation? (NOV 2018)

Data Creation links customer, sales and web data together to create a single view of a customer. By recording every visitor and every visit to a website, Data Creation builds a behavioral picture of visitors as anonymous data, and then links this data to known ‘contacts’ from web, email and CRM systems.

6. List the types of data based on how it is stored and managed

Data can be classified as structured or unstructured based on how it is stored and managed. Structured data is organized in rows and columns in a rigidly defined format so that applications can retrieve and process it efficiently. Structured data is typically stored using a database management system (DBMS).

Data is unstructured if its elements cannot be stored in rows and columns, and is therefore difficult to query and retrieve by business applications. For example, customer contacts may be stored in various forms such as sticky notes, e-mail messages, business cards, or even digital format files such as .doc, .txt, and .pdf. Due its unstructured nature, it is difficult to retrieve using a customer relationship management application. Unstructured data may not have the required components to identify itself uniquely for any type of processing or interpretation. Businesses are primarily concerned with managing unstructured data because over 80 percent of enterprise data is unstructured and requires significant storage space and effort to manage.

7. List the five core elements are essential for the basic functionality of a data center. (OR) What are the key challenge that formulate the data center? (NOV 2016)

Five core elements are essential for the basic functionality of a data center:

- Application
- Database:
- Server and operating system
- Network
- Storage array

8. What do you mean by storage device? Give example.

Data created by individuals or businesses must be stored so that it is easily accessible for further processing. In a computing environment, devices designed for storing data are termed *storage devices* or simply *storage*. The type of storage used varies based on the type of data and the rate at which it is created and used. Devices such as memory in a cell phone

or digital camera, DVDs, CD-ROMs, and hard disks in personal computers are examples of storage devices.

9. List the solutions available for data storage. (NOV 2015) (NOV 2018)

As the need for storing large amount of data grows, each of these can be combined and housed in central units

- A collection of tape drives and tapes-Tape Library
- A collection of optical disks and drives-Jukeboxes
- A collection of Hard disks-Disk arrays

Each solution addresses specific need for data storage and management

- Tape library-backup/restore; archival of data
- Jukeboxes-typically to store non changing content over long period of time
- Disk arrays-to store data that has to be immediately accessible andonline

10. List the storage technology evolved from non-intelligent internal storage to intelligent networked storage

- Redundant Array of Independent Disks (RAID)
- Direct-attached storage (DAS)
- Storage area network (SAN)
- Network-attached storage (NAS)
- Internet Protocol SAN (IP-SAN)

11. List the key characteristics of data center elements

The key characteristics of data center elements are

- Availability
- Security
- Scalability
- Performance
- Data integrity
- Capacity
- Manageability

12. List the key management activities involved in managing a (modern, complex) data center. (MAY 2017)

Managing a modern, complex data center involves many tasks. Key management activities include:

- Monitoring is the continuous collection of information and the review of the entire data center infrastructure. The aspects of a data center that are monitored include security, performance, accessibility, and capacity.
- Reporting is done periodically on resource performance, capacity, and utilization. Reporting tasks help to establish business justifications and chargeback of costs associated with data center operations.
- Provisioning is the process of providing the hardware, software, and other resources needed to run a data center.

13. List the two activities involved in Provisioning activities

Provisioning activities include capacity and resource planning.

- Capacity planning ensures that the user's and the application's future needs will be addressed in the most cost-effective and controlled manner.
- Resource planning is the process of evaluating and identifying required resources, such as personnel, the facility (site), and the technology. Resource planning ensures that adequate resources are available to meet user and application requirements.

14. What are the key challenges in managing information?(NOV 2016)

- **Exploding digital universe:** The rate of information growth is increasing exponentially. Duplication of data to ensure high availability and repurposing has also contributed to the multifold increase of information growth.
- **Increasing dependency on information:** The strategic use of information plays an important role in determining the success of a business and provides competitive advantages in the marketplace.

- **Changing value of information:** Information that is valuable today may become less important tomorrow. The value of information often changes over time.

15. Define Information lifecycle

The information lifecycle is the "change in the value of information" over time. When data is first created, it often has the highest value and is used frequently. As data ages, it is accessed less frequently and is of less value to the organization. Understanding the information lifecycle helps to deploy appropriate storage infrastructure, according to the changing value of information.

16. Define Information lifecycle management (ILM)

Information lifecycle management (ILM) is a proactive strategy that enables an IT organization to effectively manage the data throughout its lifecycle, based on predefined business policies. This allows an IT organization to optimize the storage infrastructure for maximum return on investment

17. List the characteristics ILM strategy

An ILM strategy should include the following characteristics:

- **Business-centric:** It should be integrated with key processes, applications, and initiatives of the business to meet both current and future growth in information.
- **Centrally managed:** All the information assets of a business should be under the purview of the ILM strategy.
- **Policy-based:** The implementation of ILM should not be restricted to a few departments. ILM should be implemented as a policy and encompass all business applications, processes, and resources.
- **Heterogeneous:** An ILM strategy should take into account all types of storage platforms and operating systems.
- **Optimized:** Because the value of information varies, an ILM strategy should consider the different storage requirements and allocate storage resources based on the information's value to the business.

18. List the four activities of (the process of) developing an ILM strategy

The process of developing an ILM strategy includes four activities—classifying, implementing, managing, and organizing:

- **Classifying**data and applications on the basis of business rules and policies to enable differentiated treatment of information
- **Implementing**policies by using information management tools, starting from the creation of data and ending with its disposal
- **Managing**the environment by using integrated tools to reduce operational complexity
- **Organizing**storage resources in tiers to align the resources with data classes, and storing information in the right type of infrastructure based on the information's current value

19. List the key benefits of implementing an ILM strategy

- **Improved utilization**by using tiered storage platforms and increased visibility of all enterprise information.
- **Simplified management**by integrating process steps and interfaces with individual tools and by increasing automation.
- **A wider range of options**for backup, and recovery to balance the need for business continuity.
- **Maintaining compliance**by knowing what data needs to be protected for what length of time.
- **Lower Total Cost of Ownership (TCO)** by aligning the infrastructure and management costs with information value.

20. What is bigdata?

Big data is a new and evolving concept, which refers to data sets whose sizes are beyond the capability of commonly used software tools to capture, store, manage, and process within acceptable time limits. It includes both structured and unstructured data generated by a variety of sources, including business application transactions, web pages,

videos, images, e-mails, social media, and so on. These data sets typically require real-time capture or updates for analysis, predictive modeling, and decision making.

21. List the components of the big data ecosystem

The big data ecosystem consists of the following:

- Devices that collect data from multiple locations and also generate new data about this data (metadata).
- Data collectors who gather data from devices and users.
- Data aggregators that compile the collected data to extract meaningful information.
- Data users and buyers who benefit from the information collected and aggregated by others in the data value chain.

PART-B

1. Illustrate briefly the evolution of storage technology from non intelligent internal storage to intelligent networked storage. (NOV 2016)
2. Explain how information is managed using information lifecycle. Brieflyexplain the characteristics and benefits of ILM. (NOV 2016)
3. Discuss in detail about challenges in data storage and data management. (NOV 2015)(DEC 2018)
4. Briefly discuss about core elements of a data center infrastructure. (NOV 2015) (DEC 2018)
5. Explain the key characteristics of data center and key activities involved in managing data center. (MAY 2017)
6. Explain different types of data with neat sketch.
7. Explain big data eco system with neat diagram.
8. What is VDC? Explain different types of data centers in detail.
9. A hospital uses an application that stores patient X-ray data in the form of large binary objects in an Oracle database. The application is hosted on a UNIX server, and the hospital staff accesses the X-ray records through a Gigabit Ethernet backbone. Storage array provides storage to the UNIX server, which has 6 terabytes of usable capacity. Explain the core elements of the data center. What are the typical challenges the storage management team

may face in meeting the service-level demands of the hospital staff? Describe how the value of this patient data might change over time.

10. An engineering design department of a large company maintains over 600,000 engineering drawings that its designer's access and reuse in their current projects, modifying or updating them as required. The design team wants instant access to the drawings for its current projects, but is currently constrained by an infrastructure that is not able to scale to meet the response time requirements. The team has classified the drawings as "most frequently accessed," "frequently accessed," "occasionally accessed," and "archive."

- Suggest a strategy for design department that optimizes the storage infrastructure by using ILM.
- Explain how you will use "tiered storage" based on access frequency.
- Detail the hardware and software components you will need to implement your strategy.
- Research products and solutions currently available to meet the solution you are proposing.

11. The marketing department at a mid-size firm is expanding. New hires are being added to the department and they are given network access to the department's files. IT has given marketing a networked drive on the LAN, but it keeps reaching capacity every third week. Current capacity is 500 gigabytes (and growing), with hundreds of files. Users are complaining about LAN response times and capacity. As the IT manager, what could you recommend to improve the situation?

12. A large company is considering a storage infrastructure—one that is scalable and provides high availability. More importantly, the company also needs performance for its mission-critical applications. Which storage topology would you recommend and why?

UNIT 2

PART-A

1. Define storage system environment.What are the three main components of storage system environment?

The data flows from an application to storage through various components collectively referred as a storage systemenvironment.

The three main components in this environment are the

- Host,
- Connectivity, and
- Storage

These entities, along with their physicaland logical components, facilitate data access.

2. Define Host.

Users store and retrieve data through applications. The computers on which these applications run are referred to as *hosts*. Hosts can range from simple laptops to complex clusters of servers. A host consists of physical components (hardware devices) that communicate with one another using logical components (software and protocols). Access to data and the overall performance of the storage system environment depend on both the physical and logical components of a host.

3. List the Physical and logical Components of Host.

A host has three key physical components:

- Central processing unit (CPU)
- Storage, such as internal memory and disk devices
- Input/output (I/O) devices

Following are the logical components of a host:

- Operating system
- Device drivers
- Volume manager
- File system
- Application

4. List the components of CPU

The CPU consists of four main components:

- **Arithmetic Logic Unit (ALU):** This is the fundamental building block of the CPU. It performs arithmetical and logical operations such as addition, subtraction, and Boolean functions (AND, OR, and NOT).
- **Control Unit:** A digital circuit that controls CPU operations and coordinates the functionality of the CPU.
- **Register:** A collection of high-speed storage locations. The registers store intermediate data that is required by the CPU to execute an instruction and provide fast access because of their proximity to the ALU. CPUs typically have a small number of registers.
- **Level 1 (L1) cache:** Found on modern day CPUs, it holds data and program instructions that are likely to be needed by the CPU in the near future. The L1 cache is slower than registers, but provides more storage space.

5. List the difference between RAM and ROM

- **Random Access Memory (RAM):** This allows direct access to any memory location and can have data written into it or read from it. RAM is volatile; this type of memory requires a constant supply of power to maintain memory cell content. Data is erased when the system's power is turned off or interrupted.
- **Read-Only Memory (ROM):** Non-volatile and only allows data to be read from it. ROM holds data for execution of internal routines, such as system startup.

6. List some examples of storage devices

Storage devices are less expensive than semiconductor memory. Examples of storage devices are as follows:

- Hard disk (magnetic)
- CD-ROM or DVD-ROM (optical)
- Floppy disk (magnetic)
- Tape drive (magnetic)

7. List the various types of communication of I/O Devices

I/O devices enable sending and receiving data to and from a host. This communication may be one of the following types:

- **User to host communications:** Handled by basic I/O devices, such as the keyboard, mouse, and monitor. These devices enable users to enter data and view the results of operations.
- **Host to host communications:** Enabled using devices such as a Network Interface Card (NIC) or modem.
- **Host to storage device communications:** Handled by a Host Bus Adaptor (HBA). HBA is an application-specific integrated circuit (ASIC) boardthat performs I/O interface functions between the host and the storage,relieving the CPU from additional I/O processing workload. HBAs alsoprovide connectivity outlets known as ports to connect the host to thestorage device. A host may have multiple HBAs.

8. Define connectivity.What are the components of connectivity in storage system environment?

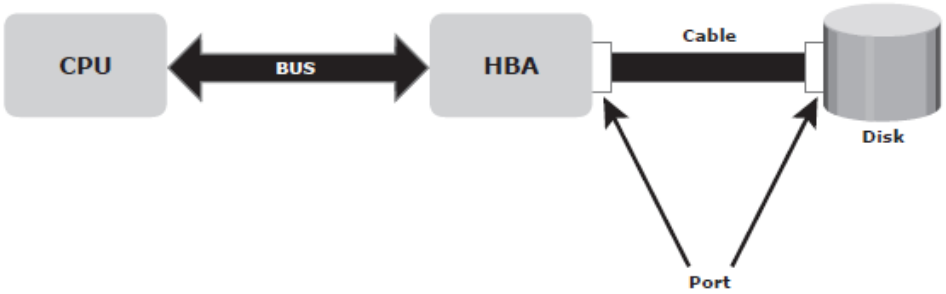
Connectivity refers to the interconnection between hosts or between a host and any other peripheral devices, such as printers or storage devices. The discussion here focuses on the connectivity between the host and the storage device.

The components of connectivity in a storage system environment can be classified as physical and logical.

- The physical components are the hardware elements that connect the host to storage
- The logical components of connectivity are the protocols used for communication between the host and storage.

9. List the Physical Components and logical components of Connectivity

The three physical components of connectivity between the host and storage are Bus, Port, and Cable



- The **bus**is the collection of paths that facilitates data transmission from one part of a computer to another, such as from the CPU to the memory.
- The **port** is a specialized outlet that enables connectivity between the host and external devices.
- **Cables**connect hosts to internal or external devices using copper or fiber optic media

Logical Components of Connectivity

The popular interface protocol used for the local bus to connect to a peripheral device is peripheral component interconnect (PCI). The interface protocols that connect to disk systems are Integrated Device Electronics/Advanced TechnologyAttachment (IDE/ATA) and Small Computer System Interface (SCSI).

10. What are the mechanisms by which the physical components communicate across the bus? (NOV 2016)

Physical components communicate across a bus by sending bits (control, data, and address) of data between devices. These bits are transmitted through the bus in either of the following ways:

- **Serially:** Bits are transmitted sequentially along a single path. This transmission can be unidirectional or bidirectional.
- **In parallel:** Bits are transmitted along multiple paths simultaneously. Parallel can also be bidirectional.

11. List the two types of buses.

The physical components communicate with one another by using a communication pathway called a bus. A bus connects the CPU to other components, such as storage and I/O

devices. Buses, as conduits of data transfer on the computer system, can be classified as follows:

- **System bus:** The bus that carries data from the processor to memory.
- **Local or I/O bus:** A high-speed pathway that connects directly to the processor and carries data between the peripheral devices, such as storage devices and the processor.

12. Mention the various layers in the SCSI communication model. (Nov 2019)

There are three layers in the SCSI communication model:

- **SCSI application layer (SAL):** This layer contains both client and serverapplications that initiate and process SCSI I/O operations using a SCSI application protocol.
- **SCSI transport protocol layer (STPL):** This layer contains the services andprotocols that allow communication between an initiator and targets.
- **Interconnect layer:** This layer facilitates data transfer between the initiatorand targets. The interconnect layer is also known as the service deliverysubsystem and comprises the services, signaling mechanisms, and inter-connects for data transfer.

13. List the limitations of tape storage medium.

- Data is stored on the tape linearly along the length of the tape. Search and retrieval of data is done sequentially, invariably taking several seconds to access the data. As a result, random data access is slow and time consuming.This limits tapes as a viable option for applications that require real-time, rapid access to data.
- In a shared computing environment, data stored on tape cannot be accessed by multiple applications simultaneously, restricting its use to one application at a time.
- On a tape drive, the read/write head touches the tape surface, so the tape degrades or wears out after repeated use.
- The storage and retrieval requirements of data from tape and the overhead associated with managing tape media are significant.

14. What is disk drive? List the key components of disk drive. (NOV 2018)

Disk drives are the most popular storage medium used in modern computers for storing and accessing data for performance-intensive, online applications. Disks support rapid

access to random data locations. This means that data can be written or retrieved quickly for a large number of simultaneous users or applications. In addition, disks have a large capacity. Disk storage arrays are configured with multiple disks to provide increased capacity and enhanced performance.

Key components of a disk drive are

- Platter,
- Spindle,
- Read/write head,
- Actuator arm assembly, and
- Controller

15. Which components constitute the disk service time? Which component contributes the largest percentage of the disk service time in a random I/O operation?

The components constitute the disk service time are

- Seek time,
- Rotational latency and
- Data transfer rate.

Seek time is the component contributes the largest percentage of the disk service time in a random I/O operation.

16. Define seek time, rotational latency and data transfer rate.

- The **seek time**(also called access time) describes the time taken to position the R/W heads across the platter with a radial movement (moving along the radius of the platter).
- To access data, the actuator arm moves the R/W head over the platter to a particular track while the platter spins to position the requested sector under the R/W head**.** The time taken by the platter to rotate and position the data under the R/W head is called **rotational latency**. This latency depends on the rotation speed of the spindle and is measured in milliseconds. The average rotational latency is one-half of the time taken for a full rotation.

- The **data transfer rate**(also called transfer rate) refers to the average amount of data per unit time that the drive can deliver to the HBA.

17. Why do formatted disks have less capacity than unformatted disks?

In order to make the storage system functional, it need to be formatted. Common types are disk formats are fat 32, NTFS, EXT2. In each of the formatting schemes, the portion of the storage space is allocated to configured file system to enable cataloging the data on the disk drive.

18. What is RAID mechanism? (NOV 2015)

A redundant array of independent (inexpensive) disks or RAID uses multiple "smaller" disks that function as one large drive, and provide for data recovery if a single drive files in most cases. When given the task of maintaining large amounts of high availability storage there are different approaches for attaining this goal, use RAID technology or use single large expensive disks (SLEDs).

RAID technology has been developed to address three areas of disk storage:

- Large capacities
- Increase input/output performance
- Reliability through redundancy

19. What is software RAID?

Software RAID uses host-based software to provide RAID functions. It is implemented at the operating-system level and does not use a dedicated hardware controller to manage the RAID array.Software RAID implementations offer cost and simplicity benefits when compared with hardware RAID. However, they have the following limitations:

- **Performance:** Software RAID affects overall system performance. This is due to the additional CPU cycles required to perform RAID calculations.
- **Supported features:** Software RAID does not support all RAID levels.
- **Operating system compatibility:** Software RAID is tied to the host operating system hence upgrades to software RAID or to the operating system should be validated for compatibility.

20. What is hardware RAID?

In hardware RAID implementations, a specialized hardware controller is implemented either on the host or on the array. These implementations vary in the way the storage array interacts with the host.

Key functions of RAID controllers are:

- Management and control of disk aggregations
- Translation of I/O requests between logical disks and physical disks
- Data regeneration in the event of disk failures

21. List the various levels of RAID with their name.

- RAID 0 Striped array with no fault tolerance
- RAID 1 Disk mirroring
- RAID 3 Parallel access array with dedicated parity disk
- RAID 4 Striped array with independent disks and a dedicated parity disk
- RAID 5 Striped array with independent disks and distributed parity
- RAID 6 Striped array with independent disks and dual distributed parity
- RAID 1 + RAID 0 Nested Combinations of RAID levels.

22. Why is RAID 1 not a substitute for a backup?

- RAID 1 provides protection against disk failure and not a solution for data recovery due to disaster.
- RAID 1 isn't a substitute for backup because there are a lot of risks that it can't protect against.
- If you accidentally delete a file, it will instantly be removed from **both** mirrored copies.
- If your disk is corrupted by a software bug or virus, the corruption will be done to **both** mirrored copies simultaneously.

23. Why is RAID 0 not an option for data protection and high availability? (NOV 2016)

RAID 0 is data striping. This means that data will appear to a user as one logical disk but all the data is distributed among two or more physical disks. If one disk fails operation cannot continue and data is lost. So RAID 0 not an option for data protection and high availability

24. How data recovery is performed using Hot spares. (MAY 2017)

One of the following methods of data recovery is performed depending on the RAID implementation:

- If parity RAID is used, then the data is rebuilt onto thehot spare from the parity and the data on the surviving HDDs in the RAID set.
- If mirroring is used, then the data from the surviving mirror is used tocopy the data.

25. What happens if the failed HDD is replaced with a new HDD in hot spares?

When the failed HDD is replaced with a new HDD, one of the followingtakes place:

- The hot spare replaces the new HDD permanently. This means that it isno longer a hot spare, and a new hot spare must be configured on thearray.
- When a new HDD is added to the system, data from the hot spare iscopied to it. The hot spare returns to its idle state, ready to replace thenext failed drive.

26.What is meant by intelligent storage system? (NOV 2015)

With advancements in technology, a new breed of storage solutions known as an intelligent storage system has evolved. The intelligent storage systems are the feature-rich RAID arrays that provide highly optimized I/O processing capabilities. These arrays have an operating environment that controls the management, allocation, and utilization of storage resources. These storage systems are configured with large amounts of memory called cache and use sophisticated algorithms to meet the I/O requirements of performance sensitive applications.

27. How the cache utilization level influences the mode of flushing to be used? (MAY 2017)

The cache utilization level drives the mode of flushing to be used:

- **Idle flushing:** Occurs continuously, at a modest rate, when the cache utilization level is between the high and low watermark.
- **High watermark flushing:** Activated when cache utilization hits the high watermark. The storage system dedicates some additional resources to flushing. This type of flushing has minimal impact on host I/O processing.
- **Forced flushing:** Occurs in the event of a large I/O burst when cache reaches 100 percent of its capacity, which significantly affects the I/O response time. In forced flushing, dirty pages are forcibly flushed to disk.

28. Represent the major categorization of intelligent storage systems. (NOV 2019)

Intelligent storage systems generally fall into one of the following two categories:

- High-end storage systems
- Midrange storage systems

Traditionally, high-end storage systems have been implemented with *active-active arrays*, whereas midrange *storage systems* used typically in small- and mediumsizedenterprises have been implemented with *active-passive arrays*.

PART-B

1. Explain the physical components of Host in detail.
2. Explainin detail about the various logical components of the Host in detail with suitable sketch. (NOV 2019)
3. Explain the physical components and logical components of connectivity in detail.
4. Explain the key (physical) components of disk drive with neat sketch. (NOV 2015) (NOV 2019)
5. Explain Zoned bit recording and logical block addressing with neat sketch. (NOV 2019)
6. Explain the components that contribute to service time on a disk drive in detail. (MAY 2017)
7. Consider a disk I/O system in which an I/O request arrives at the rate of 80 IOPS. The disk service time is 6 ms.

a. Compute the following:

Utilization of I/O controller

Total response time

Average queue size

Total time spent by a request in a queue

b. Compute the preceding parameter if the service time is halved. (NOV 2016)

8. Explain the fundamentals laws governing disk performance with suitable data.

9. An application specifies a requirement of 200 GB to host a database and other files. It also specifies that the storage environment should support 5,000 IOPS during its peak processing cycle. The disks available for configuration provide 66 GB of usable capacity, and the manufacturer specifies that they can support a maximum of 140 IOPS. The application is response time sensitive and disk utilization beyond 60 percent will not meet the response time requirements of the application. Compute and explain the theoretical basis for the minimum number of disks that should be configured to meet the requirements of the application.

10. The average I/O size of an application is 64 KB. The following specifications are available from the disk manufacturer: average seek time = 5 ms, 7,200 RPM, transfer rate = 40 MB/s. Determine the maximum IOPS that could be performed with this disk for this application. Taking this case asan example, explain the relationship between disk utilization and IOPS.

11. Explain striping, mirroring and parity concepts with neat sketch.

12. Explain various RAID levels with neat sketch.

13. Explain the process of data recovery in case of a drive failure in RAID 5. (MAY 2017)

14. What are the benefits of using RAID 3 in a backup application? (MAY 2017)

15. Discuss the impact of random and sequential I/O in different RAIDconfigurations. (NOV 2016)

16. An application has 1,000 heavy users at a peak of 2 IOPS each and 2,000typical users at a peak of 1 IOPS each, with a read/write ratio of 2:1. It isestimated that the application also experiences an overhead of 20 percentfor other workloads. Calculate the IOPS requirement for RAID 1, RAID 3,RAID 5, and RAID 6.compute the number of drives required to support theapplication in different RAID environments if 10K RPM drives with a ratingof 130 IOPS per drive were used.(NOV 2016)

17. Compare and contrast integrated and modular storage systems. (NOV 2015)

18. Explain the components of intelligent storage system with neat sketch. (NOV 2018)

19. Explain hardware, software, key protocols and components of storage system architecture in detail. (Nov 2018)

UNIT 3

PART-A

1. Define Business Continuity.

Business continuity (BC) is an integrated and enterprisewide process that includes all activities (internal and external to IT) that a business must perform to mitigate the impact of planned and unplanned downtime.

BC entails preparing for, responding to, and recovering from a system outage that adversely affects business operations. It involves proactive measures, such as business impact analysis and risk assessments, data protection, and security, and reactive countermeasures, such as disaster recovery and restart, to be invoked in the event of a failure. The goal of a business continuity solution is to ensure the "information availability" required to conduct vital business operations.

2. Define Information availability (IA)

Information availability (IA)refers to the ability of the infrastructure to function according to business expectations during its specified time of operation. Information availability ensures that people (employees, customers, suppliers,and partners) can access information whenever they need it.

3. Define Information availability can be defined with the help of reliability, accessibility and timeliness.

Information availability can be defined with the help of reliability, accessibility and timeliness.

- **Reliability:** This reflects a component's ability to function without failure, under stated conditions, for a specified amount of time.
- **Accessibility:** This is the state within which the required information is accessible at the right place, to the right user. The period of time during which the system is in an accessible state is termed system uptime; when it is not accessible it is termed system downtime.

- **Timeliness:** Defines the exact moment or the time window (a particular time of the day, week, month, and/or year as specified) during which information must be accessible. For example, if online access to an applicationis required between 8:00 am and 10:00 pm each day, any disruptions to data availability outside of this time slot are not considered to affect timeliness.

4. Provide examples of planned and unplanned downtime in the context of data center operations.

Various planned and unplanned incidents result in data unavailability.

- Planned outages include installation/integration/maintenance of new hardware, software upgrades or patches, taking backups, application and data restores, facility operations (renovation and construction), and refresh/migration of the testing to the production environment.
- Unplanned outages include failure caused by database corruption, component failure, and human errors.
- Another type of incident that may cause data unavailability is natural or man-made disasters such as flood, fire, earthquake, and contamination.

5. Define MTBF and MTTR

- **Mean Time Between Failure (MTBF):** It is the average time available for a system or component to perform its normal operations between failures.
- **Mean Time To Repair (MTTR):** It is the average time required to repair a failed component. While calculating MTTR, it is assumed that the fault responsible for the failure is correctly identified and that the required spares and personnel are available. Note that a fault is a physical defectat the component level, which may result in data unavailability. MTTR includes the time required to do the following: detect the fault, mobilizethe maintenance team, diagnose the fault, obtain the spare parts, repair,test, and resume normal operations.

6. Write the formula to calculate IA

- It can be expressed in terms of system uptime and downtime and measured as the amount or percentage of system uptime:

  IA = system uptime / (system uptime + system downtime)

- In terms of MTBF and MTTR, IA could also be expressed as

  IA = MTBF / (MTBF + MTTR)

7. List the causes of information unavailability

- Data unavailability, or downtime, results in loss of productivity, loss of revenue, poor financial performance, and damages to reputation.
- Loss of productivity reduces the output per unit of labor, equipment, and capital. Loss of revenue includes direct loss, compensatory payments, future revenue losses, billing losses, and investment losses.
- Poor financial performance affects revenuerecognition, cash flow, discounts, payment guarantees, credit rating, and stock price.
- Damages to reputation may result in a loss of confidence or credibility with customers, suppliers, financial markets, banks, and business partners.
- Other possible consequences of downtime include the cost of additional equipment rental, overtime, and extra shipping.

8. How can you calculate,average cost of downtime per hour

An important metric, *average cost of downtime per hour*, provides a key estimate in determining the appropriate BC solutions. It iscalculated as follows:

- Average cost of downtime per hour = average productivity loss per hour + average revenue loss per hour

  Where:

- Productivity loss per hour = (total salaries and benefits of all employees per week) / (average number of working hours per week)
- Average revenue loss per hour = (total revenue of an organization per week) / (average number of hours per week that an organization is open for business)

9. Compare disaster recovery and disaster restart

- **Disaster recovery:** This is the coordinated process of restoring systems, data, and the infrastructure required to support key ongoing business operations in the event of a disaster. It is the process of restoring a previouscopy of the data and applying logs or other necessary processes to that copy to bring it to a known point of consistency. Once all recoveries are completed, the data is validated to ensure that it is correct.
- **Disaster restart:** This is the process of restarting business operations with mirrored consistent copies of data and applications.

10. List the difference between RPO and RTO

RTO and RPO are both business metrics that can help you calculate how often to perform data backups. However, there are some key differences:

- **Assessment basis**. RTO reflects your overall business needs. It's a measure of how long your business can survive with IT infrastructure and services disrupted. In contrast, RPO is just about data. It determines how often to back up data and does not reflect other IT needs.
- **Cost relevance**. The costs associated with maintaining a demanding RTO may be greater than those of a granular RPO. That's because RTO involves your entire business infrastructure, not just data.
- **Automation**. Meeting your RPO goals simply requires you to perform data backups at the right interval. Data backups can easily be automated, and an automatic RPO strategy is therefore easy to implement. RTO, on the other hand, is more complicated because it involves restoring all IT operations. It is virtually impossible to achieve RTO goals in a completely automated way (although you should automate as much of your recovery process as possible).
- **Ease of calculation**. In some ways, RPO is easier to implement because data usage is relatively consistent and there are fewer variables. Because restore times involve your entire operation, not just data, it is more complicated. Restore times can change based on factors such as the time of day or the day of the week at which a disaster occurs. The RTO must be aligned with what is possible by the IT organization. If the minimum restore time possible is 2 hours, then an RTO of 1 hour will never be met. It administrators must have a good understanding of the speeds with which different

types of restores can take place. Only then can an RTO be properly negotiated and met based on the needs of the business owners.

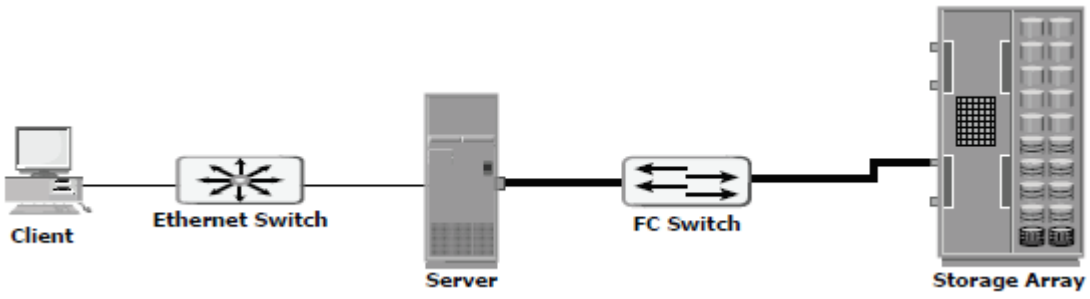
11. List the five stages of BC planning lifecycle. (DEC 2019)

The BC planning lifecycle includes five stages

- Establishing objectives
- Analyzing
- Designing and developing
- Implementing
- Training, testing, assessing, and maintaining

12 Define single point of failure (SPoF)

A single point of failurerefers to the failure of a component that can terminate the availability of the entire system or IT service. Figure illustrates the possibility of a single point of failure in a system with various components: server, network, switch, and storage array



13. Define business impact analysis (BIA)

A business impact analysis (BIA) identifies and evaluates financial, operational, and service impacts of a disruption to essential business processes. Selected functional areas are evaluated to determine resilience of the infrastructure to support information availability.

14. List the tasks of business impact analysis (BIA)

- Identify the key business processes critical to its operation.
- Determine the attributes of the business process in terms of applications,databases, and hardware and software requirements.

- Estimate the costs of failure for each business process.
- Calculate the maximum tolerable outage and define RTO and RPO for each business process.
- Establish the minimum resources required for the operation of business processes.
- Determine recovery strategies and the cost for implementing them.Optimize the backup and business recovery strategy based on business priorities.
- Analyze the current state of BC readiness and optimize future BC planning.

15. List the various BC technology solutions

- **Backup and recovery:** Backup to tape is the predominant method of ensuring data availability. These days, low-cost, high-capacity disks are used for backup, which considerably speeds up the backup and recovery process. The frequency of backup is determined based on RPO, RTO, and the frequency of data changes.
- **Storage array-based replication (local):** Data can be replicated to a separate location within the same storage array. The replica is used independently for BC operations. Replicas can also be used for restoring operations if data corruption occurs.
- **Storage array-based replication (remote):** Data in a storage array can be replicated to another storage array located at a remote site. If the storage array is lost due to a disaster, BC operations start from the remote storage array.
- **Host-based replication:** The application software or the LVM ensures that a copy of the data managed by them is maintained either locally or at a remote site for recovery purposes.

16. What is backup

A backup is a copy of production data, created and retained for the sole purpose of recovering deleted or corrupted data. With growing business and regulatory demands for data storage, retention, and availability, organizations are faced with the task of backing up an ever-increasing amount of data. This task becomes more challenging as demand for consistent backup and quick restore of data increases throughout the enterprise — which may be spread over multiple sites.

17. What are the purposes of backups?

Backups are performed to serve three purposes:

- disaster recovery,
- operational backup, and
- archival.

18. List the considerations for backups.

- The amount of data loss and downtime that a business can endure in terms of RTO and RPO are the primary considerations in selecting and implementing a specific backup strategy.
- Another consideration is the retention period, which defines the duration for which a business needs to retain the backup copies. Some data is retained for years and some only for a few days. For example, data backed up for archival is retained for a longer period than data backed up for operational recovery.
- It is also important to consider the backup media type, based on the retention period and data accessibility.
- Organizations must also consider the granularity of backup.
- The development of a backup strategymust include a decision about the most appropriate time for performing abackup in order to minimize any disruption to production operations.
- Similarly,the location and time of the restore operation must be considered, along with filecharacteristics and data compression that influences the backup process.
- Location, size, and number of files should also be considered, as they mayaffect the backup process. Location is an important consideration for the datato be backed up.
- Many organizations have dozens of heterogeneous platformssupporting complex solutions. Consider a data warehouse environment that usesbackup data from many sources. The backup process must address these sourcesin terms of transactional and content integrity. This process must be coordinatedwith all heterogeneous platforms on which the data resides.

- File size also influences the backup process. Backing up large-size files (example:ten 1 MB files) may use less system resources than backing up an equal amountof data comprising a large number of small-size files (example: ten thousand 1 KB files).
- The backup and restore operation takes more time when a file systemcontains many small files.
- Like file size, the number of files to be backed up also influences the backupprocess. For example, in incremental backup, a file system containing one million files with a 10 percent daily change rate will have to create 100,000 entries in the backup catalog, which contains the table of contents for the backed up data set

19. List the three categories of backups based on granularity

- Full backup is a backup of the complete data on the production volumes at a certain point in time. A full backup copy is created by copying the data on the production volumes to a secondary storage device.
- Incremental backup copies the data that has changed since the last full or incremental backup, whichever has occurred more recently. This is much faster (because the volume of data backed up is restricted to changed data), but it takes longer to restore.
- Cumulative (or differential) backup copies the data that has changed since the last full backup. This method takes longer than incremental backup but is faster to restore.

20. What is Synthetic (or constructed) full backup

Synthetic (or constructed) full backup is another type of backup that is used in implementations where the production volume resources cannot be exclusively reserved for a backup process for extended periods to perform a full backup. It is usually created from the most recent full backup and all the incremental backups performed after that full backup. A synthetic full backup enables a full backup copy to be created offline without disrupting the I/O operation on the production volume. This also frees up network resources from the backup process, making them available for other production uses.

21. Differentiate between hot backup and cold backup

Hot backup and cold backup are the two methods deployed for backup. They are based on the state of the application when the backup is performed. In a *hot backup*, the application is up and running, with users accessing their data during the backup process. In a *cold backup*, the application is not active during the backup process.

22. What is point-in-time (PIT)

A point-in-time (PIT) copy method is deployed in environments where the impact of downtime from a cold backup or the performance resulting from ahot backup is unacceptable. A pointer-based PIT copy consumes only a fractionof the storage space and can be created very quickly. A pointer-based PIT copy is implemented in a disk-based solution whereby a virtual LUN is created and holds pointers to the data stored on the production LUN or save location. In this method of backup, the database is stopped or frozen momentarily while the PIT copy is created. The PIT copy is then mounted on a secondary server and the backup occurs on the primary server.

23. List the Three basic topologies are used in a backup environment:

Three basic topologies are used in a backup environment:

- Direct attachedbackup,
- LAN based backup, and
- SAN based backup.

A mixed topology is alsoused by combining LAN based and SAN based topologies.

24. List the 4 different ways to implement backups in NAS environment.

In the NAS environment, backups can be implemented in four different ways:

- server based,
- serverless
- using Network Data Management Protocol (NDMP) in either NDMP 2-way or
- NDMP 3-way.

25. Name the various backup technologies and their pros and cons. (DEC 2019)

FEATURES	TAPE	DISK-AWARE BACKUP-TO-DISK	VIRTUAL TAPE
Offsite Capabilities	Yes	No	Yes
Reliability	No inherent protection methods	Yes	Yes
Performance	Subject to mechanical operations, load times	Faster single stream	Faster single stream
Use	Backup only	Multiple (backup/ production)	Backup only

26. What do you mean by replication?

Replication is the process of creating an exact copy of data. Creating one or more replicas of the production data is one of the ways to provide Business Continuity (BC). These replicas can be used for recovery and restart operations in the event of data loss. The primary purpose of replication is to enable users to have designated data at the right place, in a state appropriate to the recovery need.

27. List the uses of Local Replicas.

- Alternate source for backup
- Fast recovery
- Decision-support activities such as reporting
- Testing platform
- Data migration

UNIT 4

PART-A

1. Define Business Continuity.

Business continuity (BC) is an integrated and enterprisewide process that includes all activities (internal and external to IT) that a business must perform to mitigate the impact of planned and unplanned downtime.

BC entails preparing for, responding to, and recovering from a system outage that adversely affects business operations. It involves proactive measures, such as business impact

analysis and risk assessments, data protection, and security, and reactive countermeasures, such as disaster recovery and restart, to be invoked in the event of a failure. The goal of a business continuity solution is to ensure the "information availability" required to conduct vital business operations.

2. Define Information availability (IA)

Information availability (IA)refers to the ability of the infrastructure to function according to business expectations during its specified time of operation. Information availability ensures that people (employees, customers, suppliers,and partners) can access information whenever they need it.

3. Define Information availability can be defined with the help of reliability, accessibility and timeliness.

Information availability can be defined with the help of reliability, accessibility and timeliness.

- **Reliability:** This reflects a component's ability to function without failure, under stated conditions, for a specified amount of time.
- **Accessibility:** This is the state within which the required information is accessible at the right place, to the right user. The period of time during which the system is in an accessible state is termed system uptime; when it is not accessible it is termed system downtime.
- **Timeliness:** Defines the exact moment or the time window (a particular time of the day, week, month, and/or year as specified) during which information must be accessible. For example, if online access to an applicationis required between 8:00 am and 10:00 pm each day, any disruptions to data availability outside of this time slot are not considered to affect timeliness.

4. Provide examples of planned and unplanned downtime in the context of data center operations.

Various planned and unplanned incidents result in data unavailability.

- Planned outages include installation/integration/maintenance of new hardware, software upgrades or patches, taking backups, application and data restores, facility

operations (renovation and construction), and refresh/migration of the testing to the production environment.

- Unplanned outages include failure caused by database corruption, component failure, and human errors.
- Another type of incident that may cause data unavailability is natural or man-made disasters such as flood, fire, earthquake, and contamination.

5. Define MTBF and MTTR

- **Mean Time Between Failure (MTBF):** It is the average time available for a system or component to perform its normal operations between failures.
- **Mean Time To Repair (MTTR):** It is the average time required to repair a failed component. While calculating MTTR, it is assumed that the fault responsible for the failure is correctly identified and that the required spares and personnel are available. Note that a fault is a physical defectat the component level, which may result in data unavailability. MTTR includes the time required to do the following: detect the fault, mobilizethe maintenance team, diagnose the fault, obtain the spare parts, repair,test, and resume normal operations.

6. Write the formula to calculate IA

- It can be expressed in terms of system uptime and downtime and measured as the amount or percentage of system uptime:

  IA = system uptime / (system uptime + system downtime)
- In terms of MTBF and MTTR, IA could also be expressed as

  IA = MTBF / (MTBF + MTTR)

7. List the causes of information unavailability

- Data unavailability, or downtime, results in loss of productivity, loss of revenue, poor financial performance, and damages to reputation.
- Loss of productivity reduces the output per unit of labor, equipment, and capital. Loss of revenue includes direct loss, compensatory payments, future revenue losses, billing losses, and investment losses.

- Poor financial performance affects revenuerecognition, cash flow, discounts, payment guarantees, credit rating, and stock price.
- Damages to reputation may result in a loss of confidence or credibility with customers, suppliers, financial markets, banks, and business partners.
- Other possible consequences of downtime include the cost of additional equipment rental, overtime, and extra shipping.

8. How can you calculate,average cost of downtime per hour

An important metric, *average cost of downtime per hour*, provides a key estimate in determining the appropriate BC solutions. It iscalculated as follows:

- Average cost of downtime per hour = average productivity loss per hour + average revenue loss per hour

  Where:
- Productivity loss per hour = (total salaries and benefits of all employees per week) / (average number of working hours per week)
- Average revenue loss per hour = (total revenue of an organization per week) / (average number of hours per week that an organization is open for business)

9. Compare disaster recovery and disaster restart

- **Disaster recovery:** This is the coordinated process of restoring systems, data, and the infrastructure required to support key ongoing business operations in the event of a disaster. It is the process of restoring a previouscopy of the data and applying logs or other necessary processes to that copy to bring it to a known point of consistency. Once all recoveries are completed, the data is validated to ensure that it is correct.
- **Disaster restart:** This is the process of restarting business operations with mirrored consistent copies of data and applications.

10. List the difference between RPO and RTO

RTO and RPO are both business metrics that can help you calculate how often to perform data backups. However, there are some key differences:

- **Assessment basis**. RTO reflects your overall business needs. It's a measure of how long your business can survive with IT infrastructure and services disrupted. In contrast, RPO is just about data. It determines how often to back up data and does not reflect other IT needs.
- **Cost relevance**. The costs associated with maintaining a demanding RTO may be greater than those of a granular RPO. That's because RTO involves your entire business infrastructure, not just data.
- **Automation**. Meeting your RPO goals simply requires you to perform data backups at the right interval. Data backups can easily be automated, and an automatic RPO strategy is therefore easy to implement. RTO, on the other hand, is more complicated because it involves restoring all IT operations. It is virtually impossible to achieve RTO goals in a completely automated way (although you should automate as much of your recovery process as possible).
- **Ease of calculation**. In some ways, RPO is easier to implement because data usage is relatively consistent and there are fewer variables. Because restore times involve your entire operation, not just data, it is more complicated. Restore times can change based on factors such as the time of day or the day of the week at which a disaster occurs. The RTO must be aligned with what is possible by the IT organization. If the minimum restore time possible is 2 hours, then an RTO of 1 hour will never be met. It administrators must have a good understanding of the speeds with which different types of restores can take place. Only then can an RTO be properly negotiated and met based on the needs of the business owners.

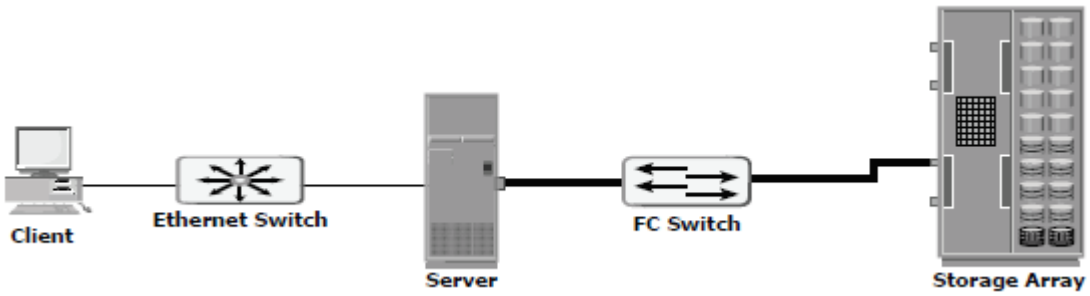
11. List the five stages of BC planning lifecycle. (DEC 2019)

The BC planning lifecycle includes five stages

- Establishing objectives
- Analyzing
- Designing and developing
- Implementing
- Training, testing, assessing, and maintaining

12 Define single point of failure (SPoF)

A single point of failurerefers to the failure of a component that can terminate the availability of the entire system or IT service. Figure illustrates the possibility of a single point of failure in a system with various components: server, network, switch, and storage array



13. Define business impact analysis (BIA)

A business impact analysis (BIA) identifies and evaluates financial, operational, and service impacts of a disruption to essential business processes. Selected functional areas are evaluated to determine resilience of the infrastructure to support information availability.

14. List the tasks of business impact analysis (BIA)

- Identify the key business processes critical to its operation.
- Determine the attributes of the business process in terms of applications,databases, and hardware and software requirements.
- Estimate the costs of failure for each business process.
- Calculate the maximum tolerable outage and define RTO and RPO for each business process.
- Establish the minimum resources required for the operation of business processes.
- Determine recovery strategies and the cost for implementing them.Optimize the backup and business recovery strategy based on business priorities.
- Analyze the current state of BC readiness and optimize future BC planning.

15. List the various BC technology solutions

- **Backup and recovery:** Backup to tape is the predominant method of ensuring data availability. These days, low-cost, high-capacity disks are used for backup, which

considerably speeds up the backup and recovery process. The frequency of backup is determined based on RPO, RTO, and the frequency of data changes.

- **Storage array-based replication (local):** Data can be replicated to a separate location within the same storage array. The replica is used independently for BC operations. Replicas can also be used for restoring operations if data corruption occurs.
- **Storage array-based replication (remote):** Data in a storage array can be replicated to another storage array located at a remote site. If the storage array is lost due to a disaster, BC operations start from the remote storage array.
- **Host-based replication:** The application software or the LVM ensures that a copy of the data managed by them is maintained either locally or at a remote site for recovery purposes.

16. What is backup

A backup is a copy of production data, created and retained for the sole purpose of recovering deleted or corrupted data. With growing business and regulatory demands for data storage, retention, and availability, organizations are faced with the task of backing up an ever-increasing amount of data. This task becomes more challenging as demand for consistent backup and quick restore of data increases throughout the enterprise — which may be spread over multiple sites.

17. What are the purposes of backups?

Backups are performed to serve three purposes:

- disaster recovery,
- operational backup, and
- archival.

18. List the considerations for backups.

- The amount of data loss and downtime that a business can endure in terms of RTO and RPO are the primary considerations in selecting and implementing a specific backup strategy.

- Another consideration is the retention period, which defines the duration for which a business needs to retain the backup copies. Some data is retained for years and some only for a few days. For example, data backed up for archival is retained for a longer period than data backed up for operational recovery.
- It is also important to consider the backup media type, based on the retention period and data accessibility.
- Organizations must also consider the granularity of backup.
- The development of a backup strategymust include a decision about the most appropriate time for performing abackup in order to minimize any disruption to production operations.
- Similarly,the location and time of the restore operation must be considered, along with filecharacteristics and data compression that influences the backup process.
- Location, size, and number of files should also be considered, as they mayaffect the backup process. Location is an important consideration for the datato be backed up.
- Many organizations have dozens of heterogeneous platformssupporting complex solutions. Consider a data warehouse environment that usesbackup data from many sources. The backup process must address these sourcesin terms of transactional and content integrity. This process must be coordinatedwith all heterogeneous platforms on which the data resides.
- File size also influences the backup process. Backing up large-size files (example:ten 1 MB files) may use less system resources than backing up an equal amountof data comprising a large number of small-size files (example: ten thousand 1 KB files).
- The backup and restore operation takes more time when a file systemcontains many small files.
- Like file size, the number of files to be backed up also influences the backupprocess. For example, in incremental backup, a file system containing one million files with a 10 percent daily change rate will have to create 100,000 entries in the backup catalog, which contains the table of contents for the backed up data set

19. List the three categories of backups based on granularity

- Full backup is a backup of the complete data on the production volumes at a certain point in time. A full backup copy is created by copying the data on the production volumes to a secondary storage device.
- Incremental backup copies the data that has changed since the last full or incremental backup, whichever has occurred more recently. This is much faster (because the volume of data backed up is restricted to changed data), but it takes longer to restore.
- Cumulative (or differential) backup copies the data that has changed since the last full backup. This method takes longer than incremental backup but is faster to restore.

20. What is Synthetic (or constructed) full backup

Synthetic (or constructed) full backup is another type of backup that is used in implementations where the production volume resources cannot be exclusively reserved for a backup process for extended periods to perform a full backup. It is usually created from the most recent full backup and all the incremental backups performed after that full backup. A synthetic full backup enables a full backup copy to be created offline without disrupting the I/O operation on the production volume. This also frees up network resources from the backup process, making them available for other production uses.

21. Differentiate between hot backup and cold backup

Hot backup and cold backup are the two methods deployed for backup. They are based on the state of the application when the backup is performed. In a *hot backup*, the application is up and running, with users accessing their data during the backup process. In a *cold backup*, the application is not active during the backup process.

22. What is point-in-time (PIT)

A point-in-time (PIT) copy method is deployed in environments where the impact of downtime from a cold backup or the performance resulting from ahot backup is unacceptable. A pointer-based PIT copy consumes only a fractionof the storage space and can be created very quickly. A pointer-based PIT copy is implemented in a disk-based solution whereby a virtual LUN is created and holds pointers to the data stored on the production LUN or save location. In this method of backup, the database is stopped or frozen momentarily while the

PIT copy is created. The PIT copy is then mounted on a secondary server and the backup occurs on the primary server.

23. List the Three basic topologies are used in a backup environment:

Three basic topologies are used in a backup environment:

- Direct attachedbackup,
- LAN based backup, and
- SAN based backup.

A mixed topology is alsoused by combining LAN based and SAN based topologies.

24. List the 4 different ways to implement backups in NAS environment.

In the NAS environment, backups can be implemented in four different ways:

- server based,
- serverless
- using Network Data Management Protocol (NDMP) in either NDMP 2-way or
- NDMP 3-way.

25. Name the various backup technologies and their pros and cons. (DEC 2019)

FEATURES	TAPE	DISK-AWARE BACKUP-TO-DISK	VIRTUAL TAPE
Offsite Capabilities	Yes	No	Yes
Reliability	No inherent protec-tion methods	Yes	Yes
Performance	Subject to mechan-ical operations, load times	Faster single stream	Faster sin-gle stream
Use	Backup only	Multiple (backup/ production)	Backup only

26. What do you mean by replication?

Replication is the process of creating an exact copy of data. Creating one or more replicas of the production data is one of the ways to provide Business Continuity (BC). These replicas can be used for recovery and restart operations in the event of data loss. The primary

purpose of replication is to enable users to have designated data at the right place, in a state appropriate to the recovery need.

27. List the uses of Local Replicas.

- Alternate source for backup
- Fast recovery
- Decision-support activities such as reporting
- Testing platform
- Data migration

UNIT 5

PART-A

1. List the four primary services of security

- **Accountability service:**Refers to accounting for all the events and operations that takes place in data center infrastructure. The accountability servicemaintains a log of events that can be audited or traced later for the purpose of security.
- **Confidentiality service:**Provides the required secrecy of information and ensures that only authorized users have access to data. This service authenticates users who need to access information and typically covers both data in transit (data transmitted over cables), or data at rest (data on a backup media or in the archives). Data in transit and at rest can be encrypted to maintain its confidentiality. In addition to restricting unauthorized users from accessing information, confidentiality services also implement traffic flow protection measures as part of the security protocol. These protection measures generally include hiding source and destination addresses, frequency of data being sent, and amount of data sent.
- **Integrity service:**Ensures that the information is unaltered. The objective of the service is to detect and protect against unauthorized alteration or deletion of information. Similar to confidentiality services, integrity services work in collaboration with accountability services to identify and authenticate the users. Integrity services stipulate measures for both in-transit data and at-rest data.

- **Availability service:**This ensures that authorized users have reliable and timely access to data. These services enable users to access the required computer systems, data, and applications residing on these systems. Availability services are also implemented on communication systems used to transmit information among computers that may reside at different locations. This ensures availability of information if a failure in one particular location occurs. These services must be implemented for both electronic data and physical data.

2. Define Risk Triad.

Risk triad defines the risk in terms of

- Threats,
- Assets, and
- Vulnerabilities.

Risk arises when a threat agent (an attacker) seeks to access assets by exploiting an existing vulnerability.

3. List the various assets of any organization

- **Information** is one of the most important *assets* for any organization.
- **Other assets** include hardware, software, and the network infrastructure required to access this information.
- To protect these assets, organizations must develop a set of parameters to ensure the availability of the resources to authorized users and trusted networks. These parameters apply to storage resources, the network infrastructure, and organizational policies.

4. List two objectives of security methods.

Security methods have two objectives.

- First objective is to ensure that the network is easily accessible to authorized users. It should also be reliable andstable under disparate environmental conditions and volumes of usage.

- Secondobjective is to make it very difficult for potential attackers to access and compromise the system. These methods should provide adequate protection against unauthorized access to resources, viruses, worms, Trojans and other malicious software programs.

5. How can you measure the effectiveness of a storage security methodology?

The effectiveness of a storage security methodology can be measured by two criteria.

- One, the cost of implementing the system should only be a small fraction of the value of the protected data.
- Two, it should cost a potential attacker more, in terms of money and time, to compromise the system than the protected data is worth.

6. Define eavesdropping and Snooping

- **Eavesdropping:** When someone overhears a conversation, the unauthorized access is called eavesdropping.
- **Snooping:** This refers to accessing another user's data in an unauthorized way. In general, snooping and eavesdropping are synonymous.

7. Differentiate active and passive attack

Passive attacks are attempts to gain unauthorized access into the system. They pose threats to confidentiality of information. Active attacks include data modification, Denial of Service (DoS), and repudiation attacks. They pose threats to data integrity and availability.

8. Define modification attack and DoS

- In a modification attack, the unauthorized user attempts to modify information for malicious purposes. A modification attack can target data at rest or data in transit. These attacks pose a threat to data integrity.
- Denial of Service (DoS)attacks denies the use of resources to legitimate users.These attacks generally do not involve access to or modification of information on the computer system. Instead, they pose a threat to data availability. The intentional

flooding of a network or website to prevent legitimate access to authorized users is one example of a DoS attack.

9. Define Repudiation

Repudiation is an attack against the accountability of the information. It attempts to provide false information by either impersonating someone or denying that an event or a transaction has taken place.

10. List the different forms of attacks and the security services used to manage them.

ATTACK	CONFIDENTIALITY	INTEGRITY	AVAILABILITY	ACCOUNTABILITY
Access	X			X
Modification	X	X		X
Denial of Service			X	
Repudiation		X		X

11. What is defense in depth?

Implementing security controls at each access point of every access path is termed as *defense in depth*.

12. List and explain the three factors to consider when assessing the extent to which an environment is vulnerable to security threats.

Attack surface, attack vector, and work factorare the three factors to consider when assessing the extent to which an environment is vulnerable to security threats.

- *Attack surface* refers to the various entry points that an attacker can use to launch an attack. Each component of a storage network is a source of potential vulnerability. All of the external interfaces supported by that component, such as the hardware interfaces, the supported protocols, and the management and administrative interfaces, can be used by an attacker to execute various attacks. These interfaces form the attack surface for the attacker. Even unused network services, if enabled, can become a part of the attack surface.

- An *attack vector* is a step or a series of steps necessary to complete an attack. For example, an attacker might exploit a bug in the management interface to execute a snoop attack whereby the attacker can modify the configuration of the storage device to allow the traffic to be accessed from one more host. This redirected traffic can be used to snoop the data in transit.
- *Work factor* refers to the amount of time and effort required to exploit an attack vector. For example, if attackers attempt to retrieve sensitive information, they consider the time and effort that would be required for executing an attack on a database. This may include determining privileged accounts, determining the database schema, and writing SQL queries. Instead, based on the work factor, they consider a less effort-intensive way to exploit the storage array by attaching to it directly and reading from the raw disk blocks.

13. Categorize the control, based on the role they play.

Based on the roles they play, controls can be categorized as preventive, detective, corrective, recovering, or compensating.

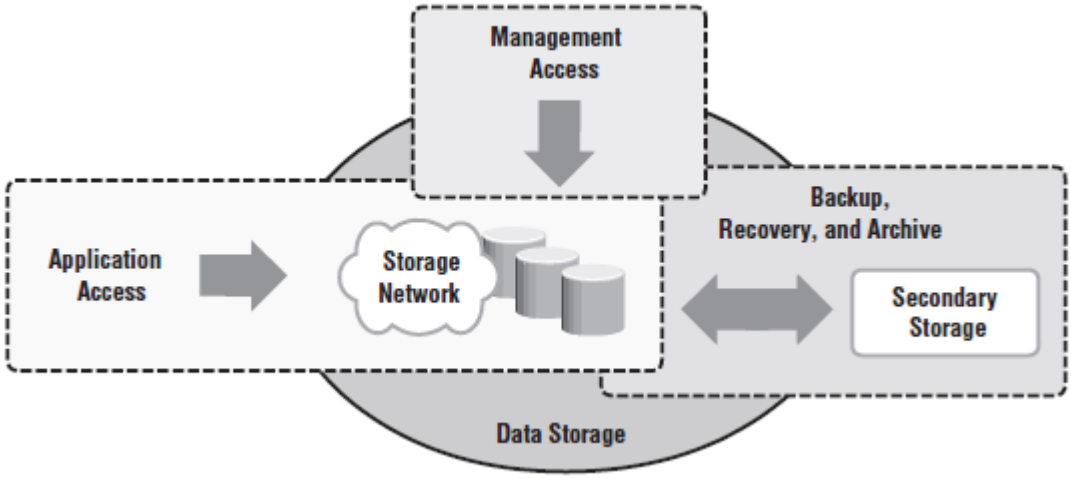
- The preventive control attempts to prevent an attack; the detective control detects whether an attack is in progress; and after an attack is discovered, the corrective controls are implemented. *Preventive* controls avert the vulnerabilities from being exploited and prevent an attack or reduce its impact.
- *Corrective* controls reduce the effect of an attack, while *detective* controls discover attacks and trigger preventive or corrective controls. For example, an Intrusion Detection/Intrusion Prevention System (IDS/IPS) is a detective control that determines whether an attack is underway and then attempts to stop it by terminating a network connection or invoking a firewall rule to block traffic.

14. Categorize theaccess paths to data storage (OR) State storage security domain. (DEC 2015)

Access paths to data storage can be categorized into three security domains:

- Application access,
- Management access, and

- BURA (backup, recovery, and archive).



15. List the various Security Zones and Protection Strategies

SECURITY ZONES	PROTECTION STRATEGIES
Zone A (Authentication at the Management Console)	(a) Restrict management LAN access to authorized users (lock down MAC addresses) (b) Implement VPN tunneling for secure remote access to the management LAN (c) Use two-factor authentication for network access
Zone B (Firewall)	Block inappropriate or dangerous traffic by: (a) Filtering out addresses that should not be allowed on your LAN (b) Screening for allowable protocols—block well-known ports that are not in use
Zone C (Access Control Switch)	Authenticate users/administrators of FC switches using RADIUS (Remote Authentication Dial In User Service), DH-CHAP (Diffie-Hellman Challenge Handshake Authentication Protocol), etc.
Zone D (ACL and Zoning)	Restrict FC access to legitimate hosts by: (a) Implementing ACLs: Known HBAs can connect on specific switch ports only (b) Implementing a secure zoning method such as port zoning (also known as hard zoning)
Zone E (Switch to Switch/ Switch to Router)	Protect traffic on your fabric by: (a) Using E_Port authentication (b) Encrypting the traffic in transit (c) Implementing FC switch controls and port controls
Zone F (Distance Extension)	Implement encryption for in-flight data: (a) FCsec for long-distance FC extension (b) IPSec for SAN extension via FCIP
Zone G (Switch-Storage)	Protect the storage arrays on your SAN via: (a) WWPN-based LUN masking (b) S_ID locking: Masking based on source FCID (Fibre Channel ID/Address)

16. List the most commonly used SAN security methods

- LUN masking and zoning,
- Switch-wide and fabric-wide access control,
- RBAC, and
- Logical partitioning of a fabric (Virtual SAN)

17. List two types of ACLs that windows support.

Windows supports two types of ACLs:

- Discretionary access control lists (DACLs) and
- System access control lists (SACLs).

The DACL, commonly referred to as theACL, is used to determine access control. The SACL determines what accessesneed to be audited if auditing is enabled.

18. What is Kerberos?

Kerberos is a network authentication protocol. It is designed to provide strong authentication for client/server applications by using secret-key cryptography. It uses cryptography so that a client and server can prove their identity to each other across an insecure network connection. After the client and server have proven their identity, they can choose to encrypt all of their communications to ensure privacy and data integrity.

In Kerberos, all authentications occur between clients and servers. The client gets a ticket for a service, and the server decrypts this ticket by using its secret key. Any entity, user, or host that gets a service ticket for a Kerberos service is called a *Kerberos client*. The term *Kerberos server* generally refers to the Key Distribution Center (KDC). The KDC implements the Authentication Service (AS) and the Ticket Granting Service (TGS). The KDC has a copy of every password associated with every principal, so it is absolutely vital that the KDC remain secure.

19. Why we need virtualization.

Virtualization provides flexibility while easing management of the existing infrastructure. Virtualization enables users to optimally utilize current processes, technologies, and systems. It allows for the addition, modification, or replacement of physical resources without affecting application availability. Virtualization technology offers high security and data integrity, which are mandatory for centralized computing environments. It also reduces performance degradation issues and unplanned downtime due to faults, and ensures increased availability of hardware resources.

20. List the different forms of virtualization. (DEC 2019)

Virtualization has existed in the IT industry for several years and in different forms, including

- Memory virtualization,
- Network virtualization,
- Server virtualization, and
- Storage virtualization.

21. List some of the features of VSAN

Some of the features of VSAN are:

- Fibre Channel ID (FC ID) of a host in a VSAN can be assigned to a host in another VSAN, thus improving scalability of SAN.
- Every instance of a VSAN runs all required protocols such as FSPF, domain manager, and zoning.
- Fabric-related configurations in one VSAN do not affect the traffic in another VSAN.
- Events causing traffic disruptions in one VSAN are contained within that VSAN and are not propagated to other VSANs.

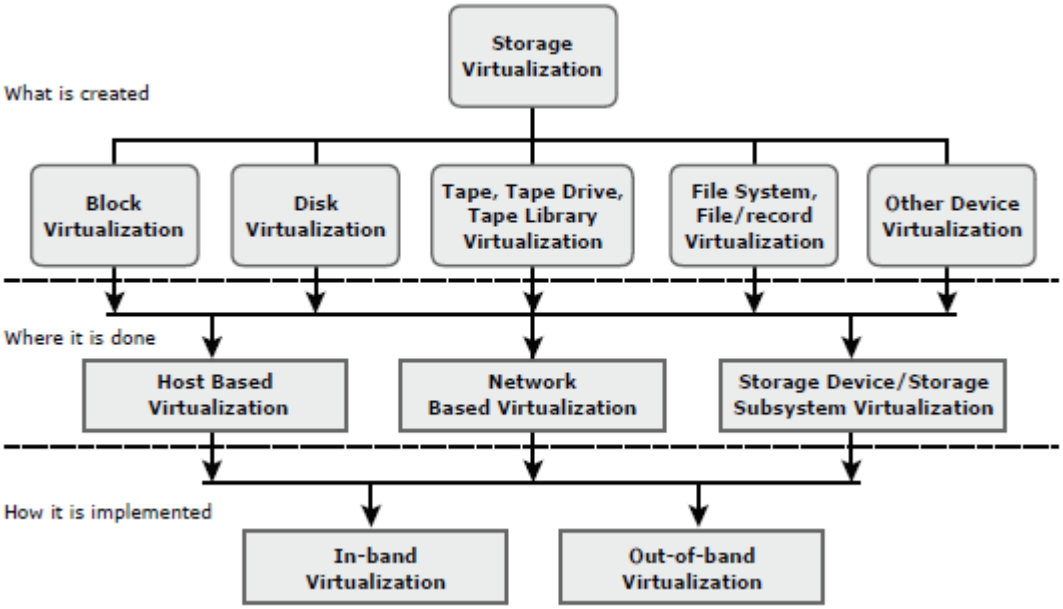
22. List the key benefits of storage virtualization

The key benefits of storage virtualization include

- increased storage utilization,
- adding or deleting storage without affecting an application's availability, and
- nondisruptive data migration (access to files and storage while migrationsare in progress).

23. Draw the SNIA (Storage Networking Industry Association) storage virtualization taxonomy. (DEC 2016)

The SNIA (Storage Networking Industry Association) storage virtualization taxonomy provides a systematic classification of storage virtualization, with three levels defining what, where, and how storage can be virtualized.



24. Represent the various issues that are addressed by storage virtualization solution. List the challenges of storage virtualization (DEC 2019)

Storage networking and feature-rich intelligent storage arrays have addressed and provided specific solutions to business problems. As an enabler, virtualization should add value to the existing solution, but introducing virtualization into an environment adds new challenges. The storage virtualization solution mustbe capable of addressing issues such as

- Scalability,
- Functionality,
- Manageability, and
- Support.

25. List the types of storage virtualization

Virtual storage is about providing logical storage to hosts and applications independent of physical resources. Virtualization can be implemented in both SAN and NAS storage environments. In a SAN, virtualization is applied at the block level, whereas in NAS, it is applied at the file level.

26. What is LUN masking? (DEC 2016)

LUN maskingdetermines which hosts can access which storage devices. (OR) A process that provides data access control so that the hostcan see only the LUNs it is intended to access.

PART-B

1. Explain in detail about storage security domains. (Dec 2019)
2. Explain in detail about the various types of storage virtualization with neat sketch (Dec 2019)(DEC 2015)
3. How can a block-level virtualization implementation be used as a data migration tool? Explain how data migration will be accomplished and discuss the advantages of using this method for storage. Compare this method to traditional migration methods. (DEC 2016)
4. Frequently, storage arrays in a data center are replaced with newer arrays to take advantage of technology advancements and cost benefits and to allow business growth. Migrating data from old arrays to a new array has now become a routinely performed activity in data centers. Do a survey of host-based, storage array–based, and virtualization appliance–based migration methods. Detail the advantages and disadvantages. Consider a migration scenario in which you are migrating from a DAS to a SAN environment.
5. Refer to question 4. Which method of migration will you use? Develop a short presentation explaining why you are recommending a particular method. Include a work breakdown structure for executing the migration with your recommended method.
6. A storage array dials a support center automatically whenever an error is detected. The vendor's representative at the support center can log on to the service processor of the storage array through the Internet to perform diagnostics and repair. Discuss the impact of this feature in a secure storage environment and provide security methods that can be implemented to mitigate any malicious attacks through this gateway.(DEC 2016)
7. Develop a checklist for auditing the security of a storage environment with SAN, NAS, and iSCSI implementations. Explain how you will perform the audit. Assume that you discover at least five security loopholes during the audit process. List them and provide control mechanisms that should be implemented to eliminate them. (DEC 2019)
8. Explain different forms of virtualization in detail.
9. Explain SNIA in detail.
10. Explain the security implementations in SAN.
11. Explain the security implementations in NAS.

12. Explain the security implementations in IP-SAN.