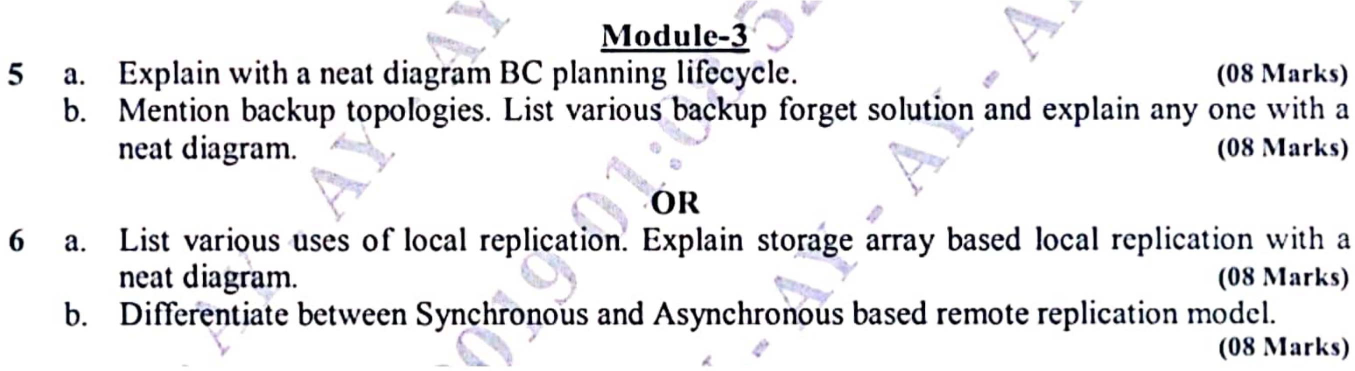


Module-3

5 a. Explain with a neat diagram BC planning lifecycle. (08 Marks)

b. Mention backup topologies. List various backup forget solution and explain any one with a neat diagram. (08 Marks)

OR

6 a. List various uses of local replication. Explain storage array based local replication with a neat diagram. (08 Marks)

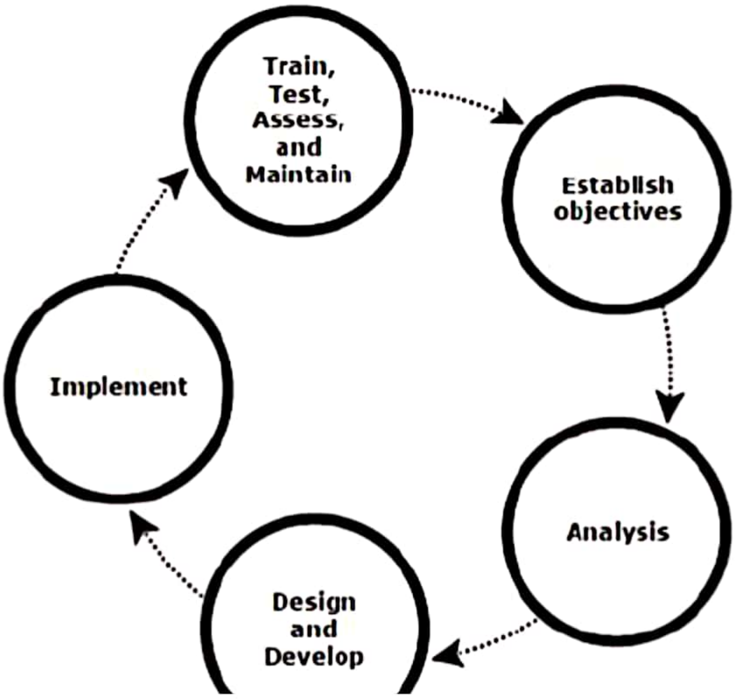
b. Differentiate between Synchronous and Asynchronous based remote replication model. (08 Marks)

5A)

3.1.3 BC Planning Life Cycle

BC planning must follow a disciplined approach like any other planning process. Organizations today dedicate specialized resources to develop and maintain BC plans. From the conceptualization to the realization of the BC plan, a life cycle of activities can be defined for the BC process.

The BC planning lifecycle includes five stages shown below (Fig 3.4):



Several activities are performed at each stage of the BC planning lifecycle, including the following key activities:

1. **Establishing objectives**

→ Determine BC requirements.

→ Estimate the scope and budget to achieve requirements.

→ Select a BC team by considering subject matter experts from all areas of the business, whether internal or external.

→ Create BC policies.

2. **Analyzing**

→ Collect information on data profiles, business processes, infrastructure support, dependencies, and frequency of using business infrastructure.

→ Identify critical business needs and assign recovery priorities.

→ Create a risk analysis for critical areas and mitigation strategies.

→ Conduct a Business Impact Analysis (BIA).

→ Create a cost and benefit analysis based on the consequences of data unavailability.

3. **Designing and developing**

→ Define the team structure and assign individual roles and responsibilities. For example, different teams are formed for activities such as emergency response, damage assessment, and infrastructure and application recovery.

→ Design data protection strategies and develop infrastructure.

→ Develop contingency scenarios.

→ Develop emergency response procedures.

→ Detail recovery and restart procedures.

4. **Implementing**

→ Implement risk management and mitigation procedures that include backup, replication, and management of resources.

→ Prepare the disaster recovery sites that can be utilized if a disaster affects the primary data center.

→ Implement redundancy for every resource in a data center to avoid single points of failure.

5. **Training, testing, assessing, and maintaining**

→ Train the employees who are responsible for backup and replication of business-critical data on a regular basis or whenever there is a modification in the BC plan

→ Train employees on emergency response procedures when disasters are declared.

→ Train the recovery team on recovery procedures based on contingency scenarios.

→ Perform damage assessment processes and review recovery plans.

→ Test the BC plan regularly to evaluate its performance and identify its limitations.

→ Assess the performance reports and identify limitations.

5B)

3.2.3 Backup Topologies

- Three basic topologies are used in a backup environment:

1. Direct attached backup
2. LAN based backup, and
3. SAN based backup.

- A **mixed topology** is also used by combining LAN based and SAN based topologies.
- In a **direct-attached backup**, a backup device is attached directly to the client. Only the metadata is sent to the backup server through the LAN. This configuration frees the LAN from backup traffic.
- The example shown in Fig 3.7 device is directly attached and dedicated to the backup client. As the environment grows, however, there will be a need for central management of all backup devices and to share the resources to optimize costs. An appropriate solution is to share the backup devices among multiple servers. Network-based topologies (LAN-based and SAN-based) provide the solution to optimize the utilization of backup devices.

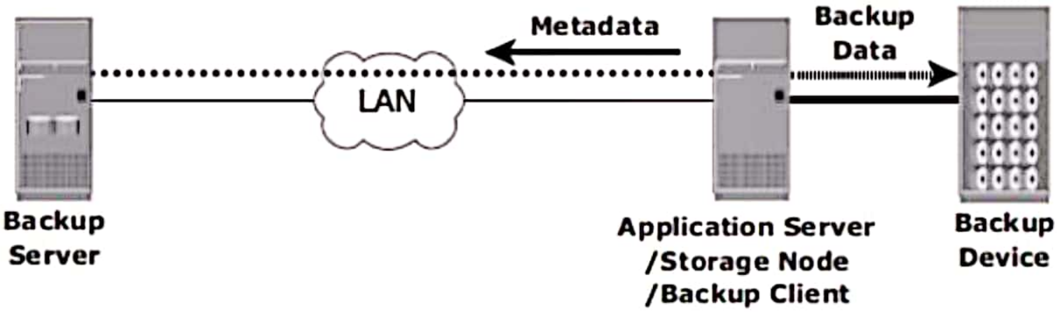


Fig 3.7: Direct-attached backup topology

- In **LAN-based backup**, the clients, backup server, storage node, and backup device are connected to the LAN (see Fig 3.8). The data to be backed up is transferred from the backup client (source), to the backup device (destination) over the LAN, which may affect network performance.
- This impact can be minimized by adopting a number of measures, such as configuring separate networks for backup and installing dedicated storage nodes for some application servers.

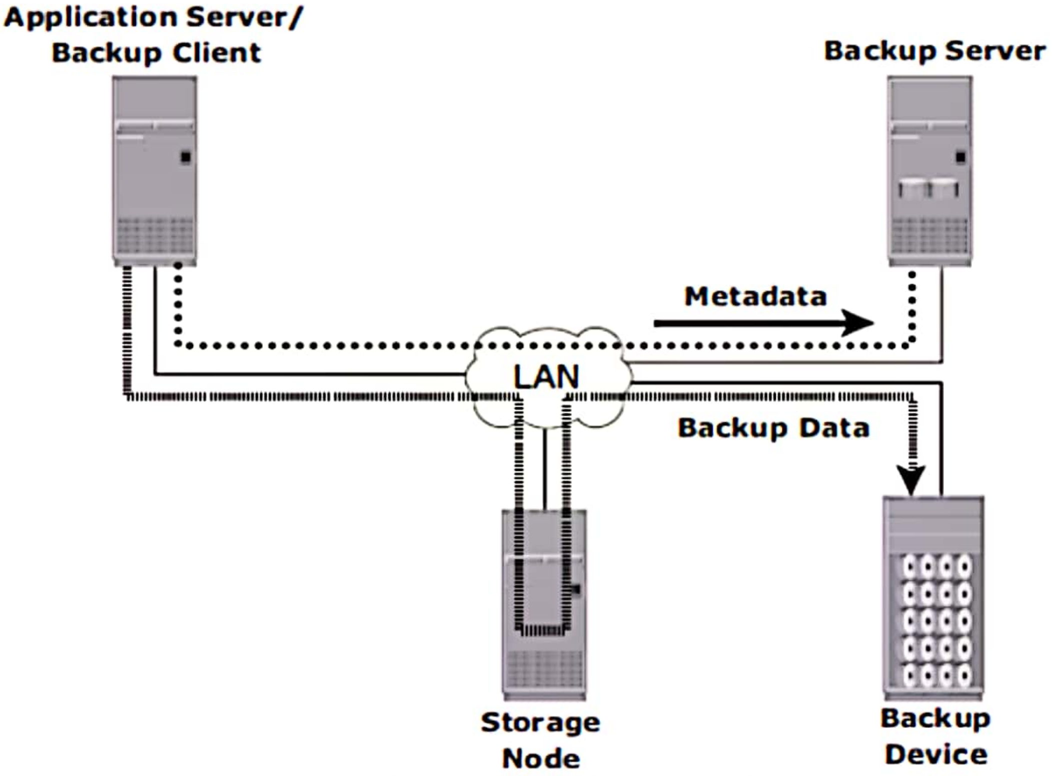


Fig 3.8: LAN-based backup topology

- The **SAN-based backup** is also known as the *LAN-free backup*. Fig 3.9 illustrates a SAN-based backup. The SAN-based backup topology is the most appropriate solution when a backup device needs to be shared among the clients. In this case the backup device and clients are attached to the SAN.
- In the example from Fig 3.9, a client sends the data to be backed up to the backup device over the SAN. Therefore, the backup data traffic is restricted to the SAN, and only the backup metadata is transported over the LAN. The volume of metadata is insignificant when compared to the production data; the LAN performance is not degraded in this configuration.

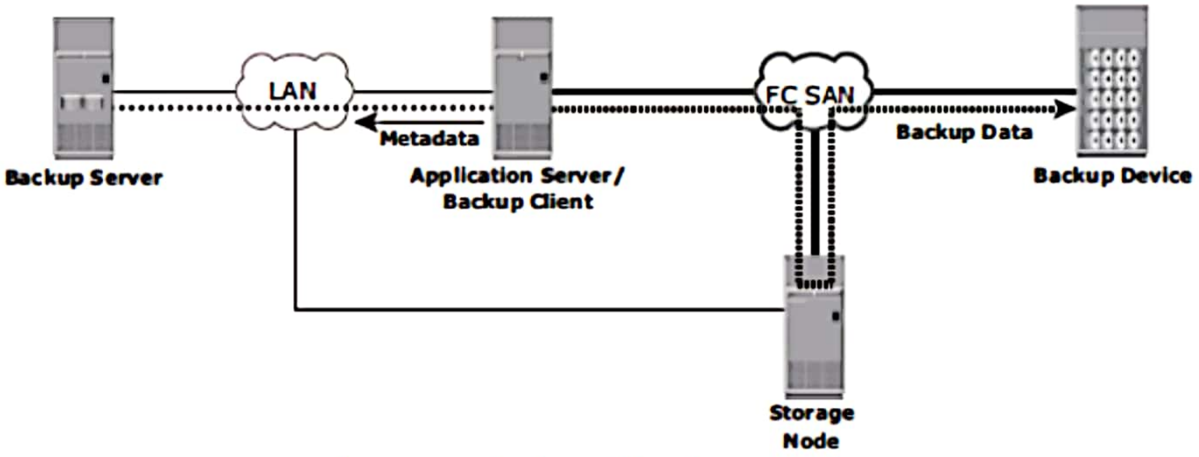
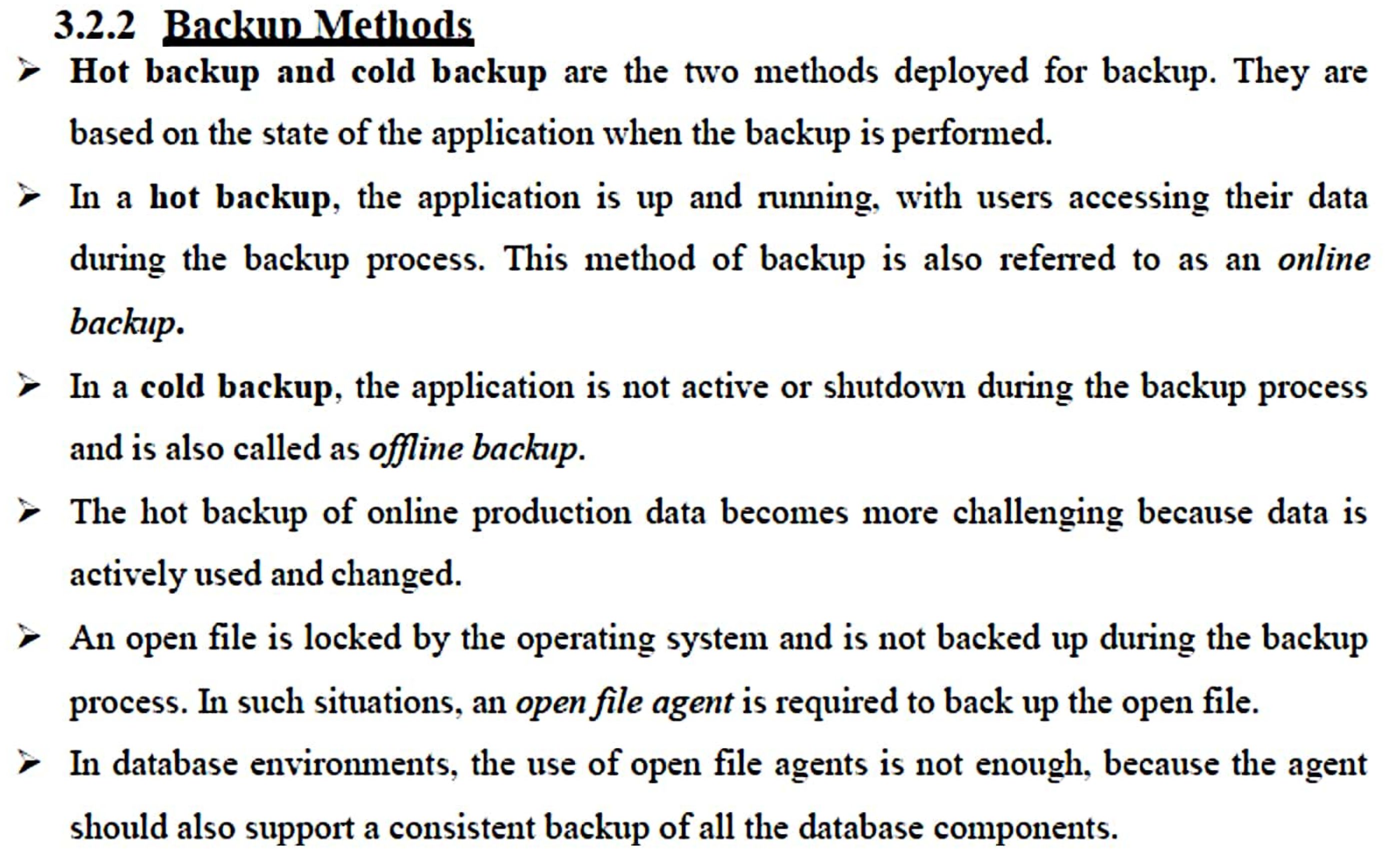
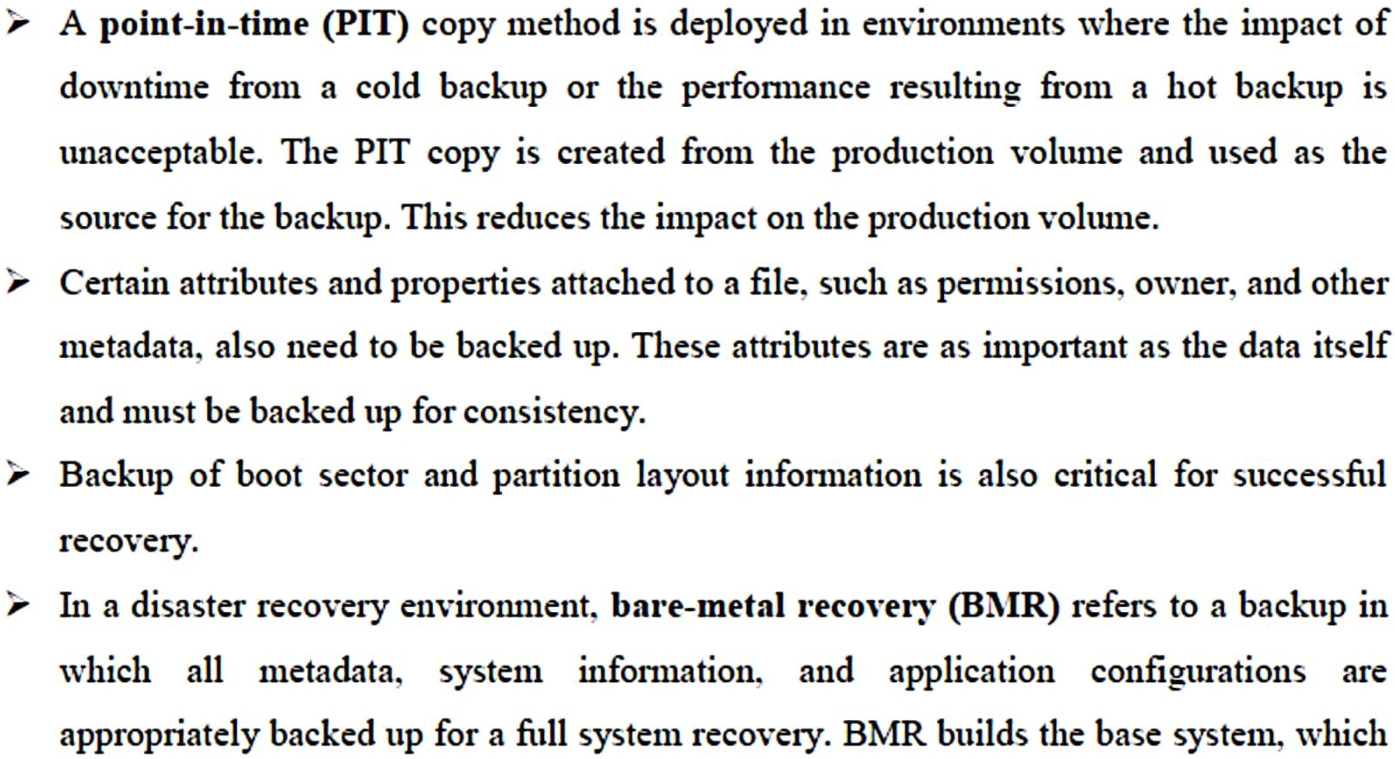


Fig 3.9: SAN-based backup topology



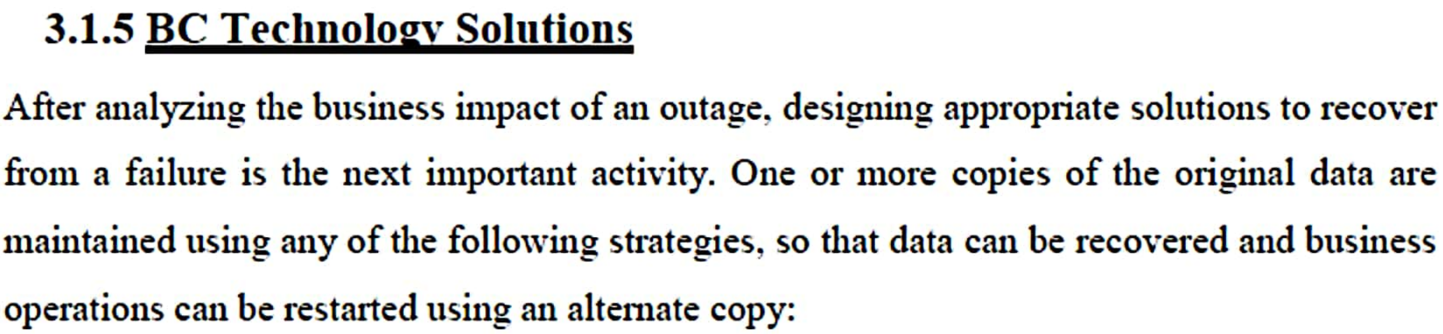
3.2.2 Backup Methods

- **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. This method of backup is also referred to as an ***online backup.***
- In a **cold backup**, the application is not active or shutdown during the backup process and is also called as *offline backup.*
- The hot backup of online production data becomes more challenging because data is actively used and changed.
- An open file is locked by the operating system and is not backed up during the backup process. In such situations, an *open file agent* is required to back up the open file.
- In database environments, the use of open file agents is not enough, because the agent should also support a consistent backup of all the database components.



- 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 a hot backup is unacceptable. The PIT copy is created from the production volume and used as the source for the backup. This reduces the impact on the production volume.
- Certain attributes and properties attached to a file, such as permissions, owner, and other metadata, also need to be backed up. These attributes are as important as the data itself and must be backed up for consistency.
- Backup of boot sector and partition layout information is also critical for successful recovery.
- In a disaster recovery environment, **bare-metal recovery (BMR)** refers to a backup in which all metadata, system information, and application configurations are appropriately backed up for a full system recovery. BMR builds the base system, which

6A)



3.1.5 BC Technology Solutions

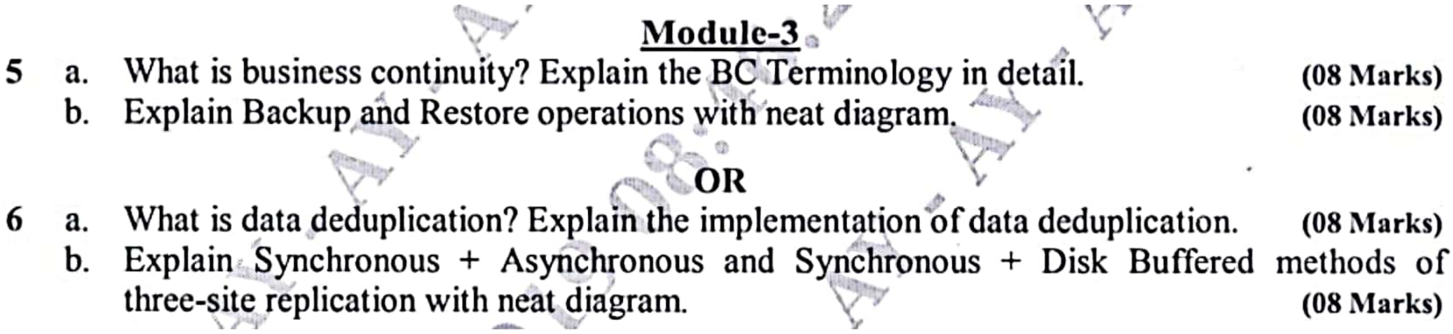
After analyzing the business impact of an outage, designing appropriate solutions to recover from a failure is the next important activity. One or more copies of the original data are maintained using any of the following strategies, so that data can be recovered and business operations can be restarted using an alternate copy:

1. **Backup:** Data backup is a predominant method of ensuring data availability. The frequency of backup is determined based on RPO, RTO, and the frequency of data changes.
2. **Storage array-based replication (local):** Data can be replicated to a separate location within the same storage array. The replica is used independently for other business operations. Replicas can also be used for restoring operations if data corruption occurs.
3. **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, business operations can be started from the remote storage array.

6B)

Synchronous vs. Asynchronous Replication: Main Differences

	Synchronous	**Asynchronous**
Distance	Works better when locations are in close proximity (performance drops in proportion to distance).	Works over longer distances (as long as network connection between datacenters is available).
Cost	More expensive	More cost-effective
Recovery Point Objective (RPO)	Zero	From 15 minutes to a few hours
Recovery Time Objective (RTO)	Short	Short
Network	Requires more bandwidth and is affected by latency; Can be affected by WAN interruptions	Requires less bandwidth and is not affected by latency; Is not affected by WAN interruptions
Data loss	Zero	Possible loss of most recent updates to data.
Performance	Low (waits for network acknowledgement from the secondary location).	High (does not wait for network acknowledgement from the secondary location).



Module-3

5 a. What is business continuity? Explain the BC Terminology in detail. (08 Marks)

b. Explain Backup and Restore operations with neat diagram. (08 Marks)

OR

6 a. What is data deduplication? Explain the implementation of data deduplication. (08 Marks)

b. Explain Synchronous + Asynchronous and Synchronous + Disk Buffered methods of three-site replication with neat diagram. (08 Marks)

5A)

Business Continuity (BC):

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

3.1.2 BC Terminology

This section defines common terms related to BC operations which are used in this module to explain advanced concepts:

- **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 previous copy 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.
- **Recovery-Point Objective (RPO):** This is the point in time to which systems and data must be recovered after an outage. It defines the amount of data loss that a business can endure. A large RPO signifies high tolerance to information loss in a business. Based on the RPO,

→ **RPO of 1 hour:** Shipping database logs to the remote site every hour. The corresponding recovery strategy is to recover the database at the point of the last log shipment.

→ **RPO in the order of minutes:** Mirroring data asynchronously to a remote site

→ **Near zero RPO:** This mirrors mission-critical data synchronously to a remote site.

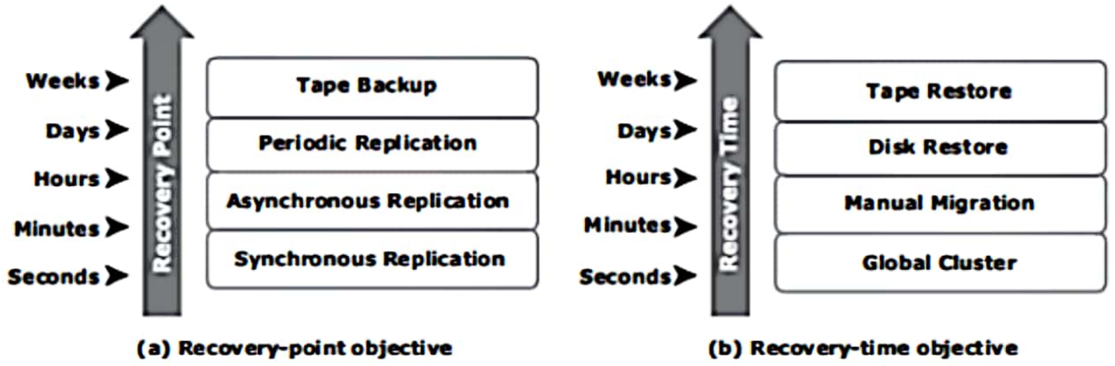


Fig 3.3: Strategies to meet RPO and RTO targets

- **Recovery-Time Objective (RTO):** The time within which systems and applications must be recovered after an outage. It defines the amount of downtime that a business can endure and survive. Businesses can optimize disaster recovery plans after defining the RTO for a given system. For example, if the RTO is two hours, then use a disk backup because it enables a faster restore than a tape backup. However, for an RTO of one week, tape backup will likely meet requirements. Some examples of RTOs and the recovery strategies to ensure data availability are listed below (refer to Fig 3.3 (b)):

  → **RTO of 72 hours:** Restore from backup tapes at a cold site.

  → **RTO of 12 hours:** Restore from tapes at a hot site.

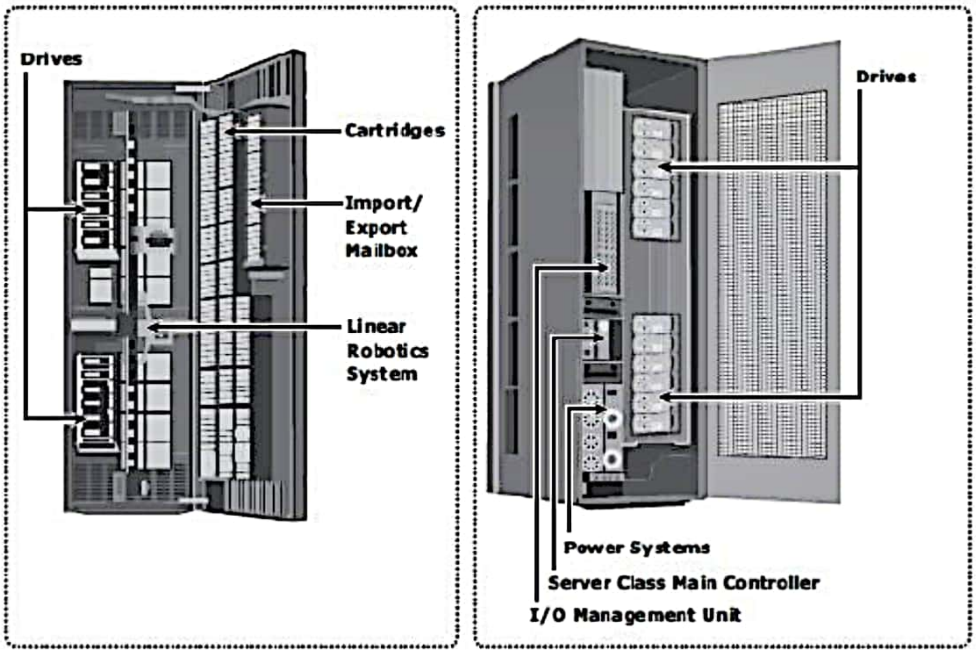
  → **RTO of few hours:** Use a data vault to a hot site.

  → **RTO of a few seconds:** Cluster production servers with bidirectional mirroring, enabling the applications to run at both sites simultaneously.

5B)BACKUP AND RESTORE OPERATIONS

Physical Tape Library

- The physical tape library provides housing and power for a number of tape drives and tape cartridges, along with a robotic arm or picker mechanism.
- The backup software has intelligence to manage the robotic arm and entire backup process. Fig 3-14 shows a physical tape library.
- *Tape drives* read and write data from and to a tape. Tape cartridges are placed in the slots when not in use by a tape drive. *Robotic arms* are used to move tapes around the library, such as moving a tape drive into a slot.



3.2.4.2 Backup to Disk

- Because of *increased availability*, low cost **disks** have now replaced tapes as the primary device for storing backup data because of their *performance advantages*. Backup-to-disk systems offer *ease of implementation*, *reduced TCO* (Total cost of ownership), and *improved quality of service*. Disks also offer *faster recovery* when compared to tapes.
- Backing up to disk storage systems offers clear advantages due to their inherent random access and RAID-protection capabilities.
- Fig 3.13 illustrates a recovery scenario comparing tape versus disk in a Microsoft Exchange environment that supports 800 users with a 75 MB mailbox size and a 60 GB database. As shown, a restore from disk took 24 minutes compared to the restore from a tape, which took 108 minutes for the same environment.
- Recovering from a full backup copy stored on disk and kept onsite provides the fastest recovery solution. Using a disk enables the creation of full backups more frequently, which in turn improves RPO and RTO.
- Backup to disk does not offer any inherent offsite capability, and is dependent on other technologies such as local and remote replication.

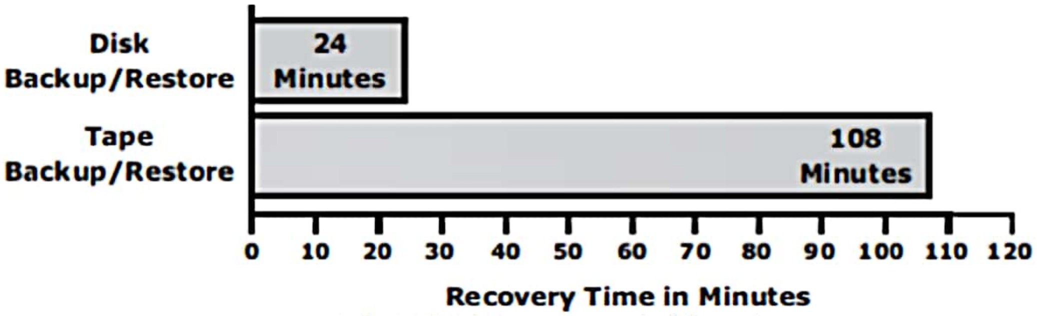


Fig 3.13: Tape versus Disk restore

Virtual Tape Library

- A virtual tape library (VTL) has the same components as that of a physical tape library except that the majority of the components are presented as virtual resources.
- For the backup software, there is no difference between a physical tape library and a virtual tape library.
- Fig 3.14 shows a virtual tape library that uses disks as backup media. Emulation software has a database with a list of virtual tapes, and each virtual tape is assigned a portion of a LUN on the disk. A virtual tape can span multiple LUNs if required.
- File system awareness is not required while backing up because virtual tape solutions use raw devices.
- Similar to a physical tape library, a robot mount is performed when a backup process starts in a virtual tape library. However, unlike a physical tape library, where this process involves some mechanical delays, in a virtual tape library it is almost instantaneous. Even the *load to ready* time is much less than in a physical tape library.

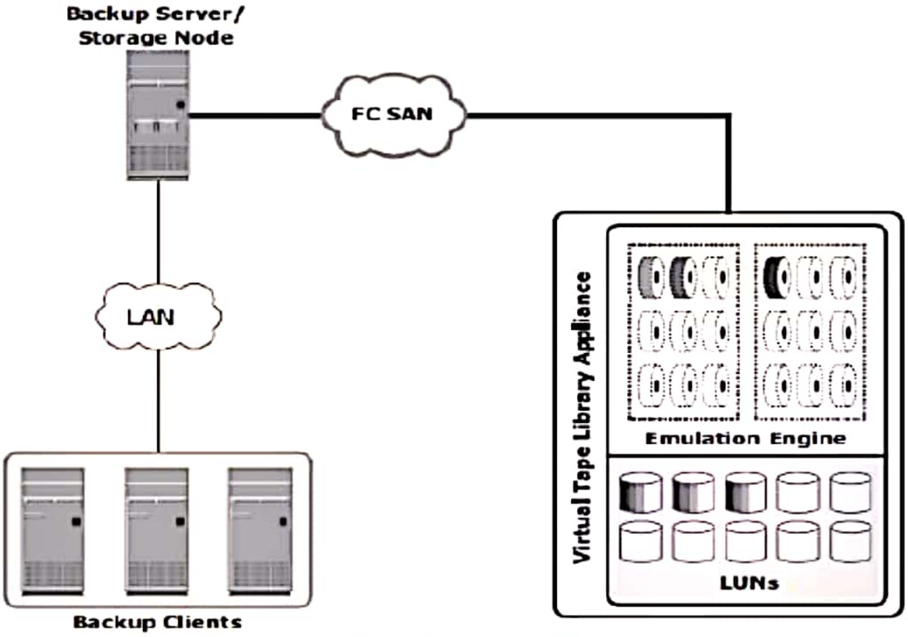


Fig 3.14: Virtual Tape Library

6A)

- **Data deduplication** is the process of identifying and eliminating redundant data. When duplicate data is detected during backup, the data is discarded and only the pointer is created to refer the copy of the data that is already backed up.

3.2.5.2 Data Deduplication Implementation

Deduplication for backup can happen at the data source or the backup target.

Source-Based Data Deduplication

- *Source-based data deduplication* eliminates redundant data at the source before it transmits to the backup device.
- Source-based data deduplication can dramatically reduce the amount of backup data sent over the network during backup processes. It provides the benefits of a shorter backup window and requires less network bandwidth. There is also a substantial reduction in the capacity required to store the backup images.
- Fig 3.15 shows source-based data deduplication.
- Source-based deduplication increases the overhead on the backup client, which impacts the performance of the backup and application running on the client.
- Source-based deduplication might also require a change of backup software if it is not supported by backup software.

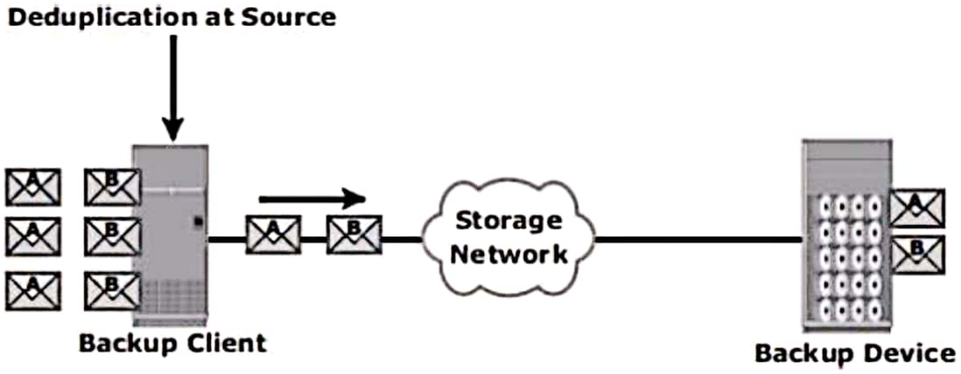


Fig 3.15: Source-based data deduplication

Target-Based Data Deduplication

- Target-based data deduplication is an alternative to source-based data deduplication.
- Target-based data deduplication occurs at the backup device, which offloads the backup client from the deduplication process.
- Fig 3.16 shows target-based data deduplication.
- In this case, the backup client sends the data to the backup device and the data is deduplicated at the backup device, either *immediately (inline)* or at a *scheduled time (post-process).*
- Because deduplication occurs at the target, all the backup data needs to be transferred over the network, which increases network bandwidth requirements. Target-based data deduplication does not require any changes in the existing backup software.
- *Inline deduplication* performs deduplication on the backup data before it is stored on the backup device. Hence, this method reduces the storage capacity needed for the backup.

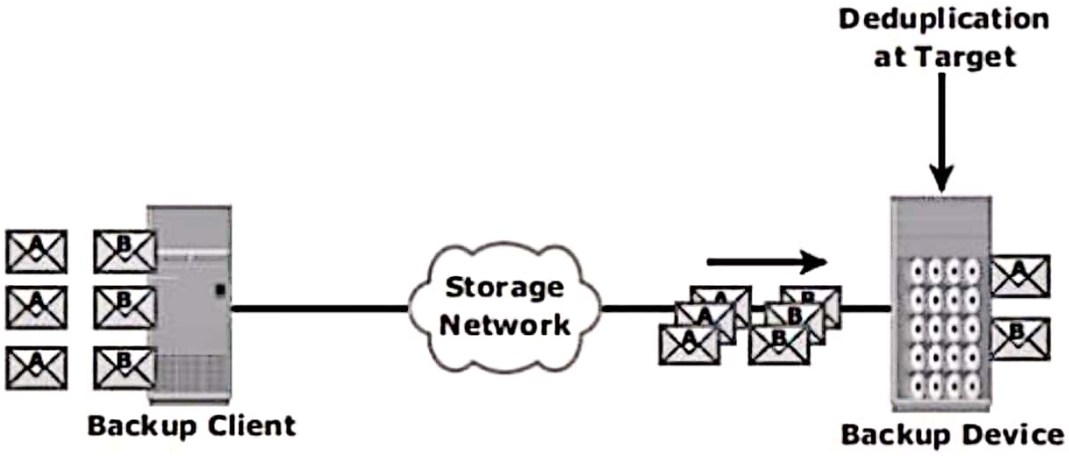
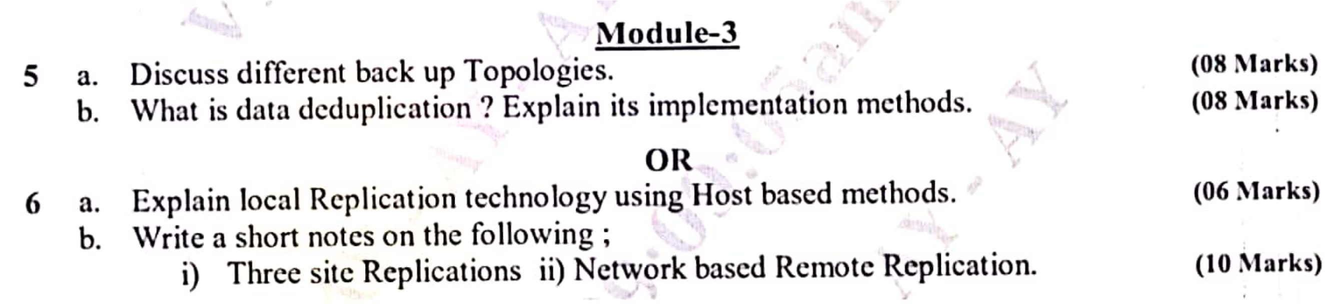


Fig 3.16: Target-based data deduplication

6B) UNAVAILABLE



Module-3

5 a. Discuss different back up Topologies. (08 Marks)

b. What is data deduplication ? Explain its implementation methods. (08 Marks)

OR

6 a. Explain local Replication technology using Host based methods. (06 Marks)

b. Write a short notes on the following ;

i) Three site Replications ii) Network based Remote Replication. (10 Marks)

5A)REPEATED (PAGE1) 5B) REPEATED

6A&6B NOT AVAILABLE