Module 3

1 What is business continuity? Explain the BC Terminology in detail.

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.

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, risk assessments, deployment of BC technology solutions (backup and replication), and reactive measures, such as disaster recovery and restart, to be invoked in the event of a failure. The goal of a BC solution is to ensure the "information availability" required to conduct vital business operations.

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, organizations plan for the minimum frequency with which a backup or replica must be made. For example, if the RPO is six hours, backups or replicas must be made at least once in 6 hours. Fig 3.3 (a) shows various RPOs and their corresponding ideal recovery strategies. An

organization can plan for an appropriate BC technology solution on the basis of the RPO it sets. For example:

→ RPO of 24 hours: This ensures that backups are created on an offsite tape drive every midnight. The corresponding recovery strategy is to restore data from the set of last backup tapes.

→ 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.

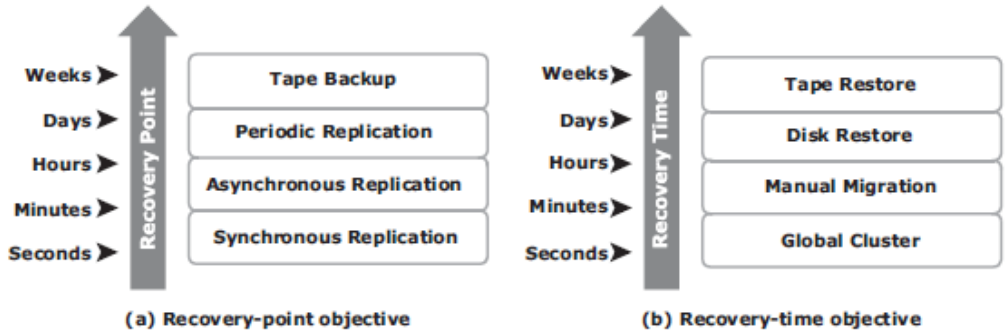

Fig 3.3: Strategies to meet RPO and RTO targets

➢ 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.

2 Explain Backup and Restore operations with neat diagram.

Refer page no. 234 from text-book

3 What is data deduplication? Explain the implementation of data deduplication.

➢ 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.

➢ Data deduplication helps to reduce the storage requirement for backup, shorten the backup window, and remove the network burden. It also helps to store more backups on the disk and retain the data on the disk for a longer time.

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.

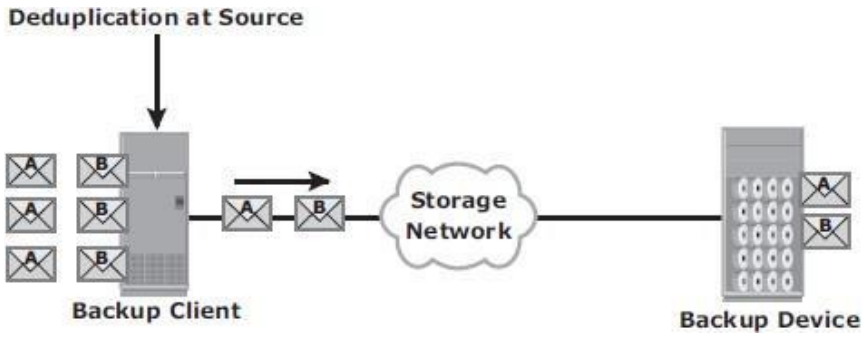

Fig 3.15: Source-based data deduplication

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.

➢ Inline deduplication introduces overhead in the form of the time required to identify and remove duplication in the data. So, this method is best suited for an environment with a large backup window.

➢ Post-process deduplication enables the backup data to be stored or written on the backup device first and then deduplicated later.

➢ This method is suitable for situations with tighter backup windows. However, post- process deduplication requires more storage capacity to store the backup images before they are deduplicated.

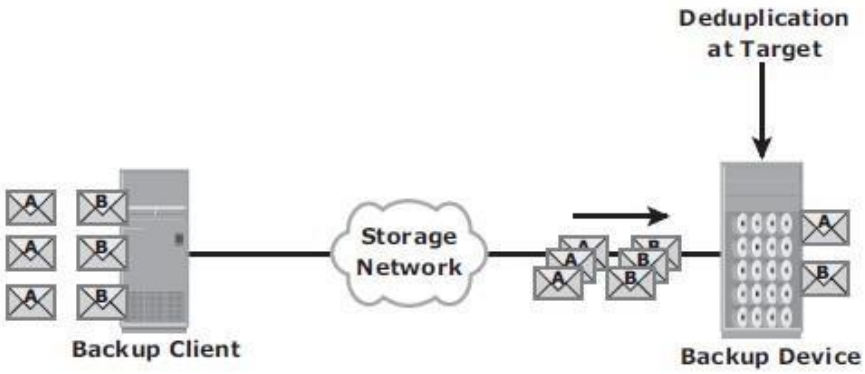

Fig 3.16: Target-based data deduplication

4 Explain Synchronous + Asynchronous and Synchronous + Disk Buffered methods of three-site replication with neat diagram.

Refer page no. 300 text-book

5 Explain with a neat diagram BC planning lifecycle.

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):

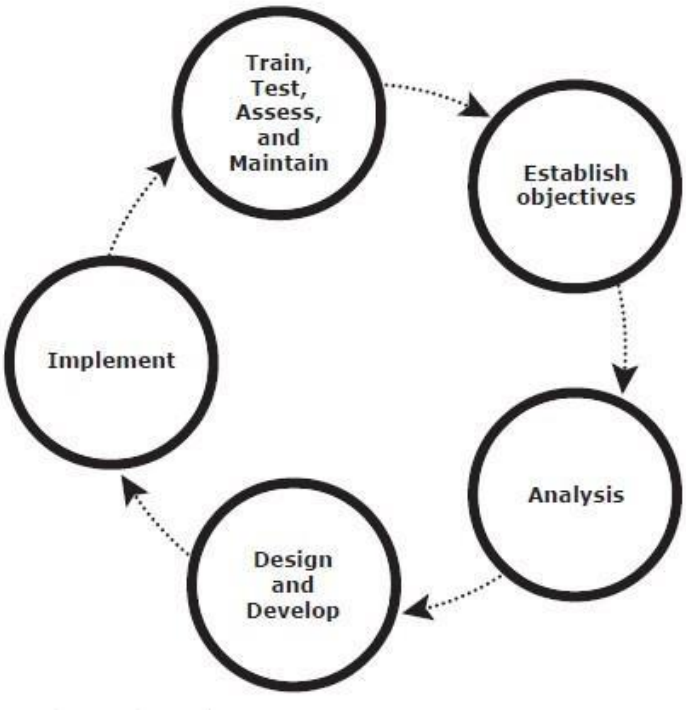

Fig 3.4: BC Planning Lifecycle

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.

→ Update the BC plans and recovery/restart procedures to reflect regular changes within the data center.

6 Mention backup topologies. List various backup forget solution and explain anyone with a neat diagram.

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.

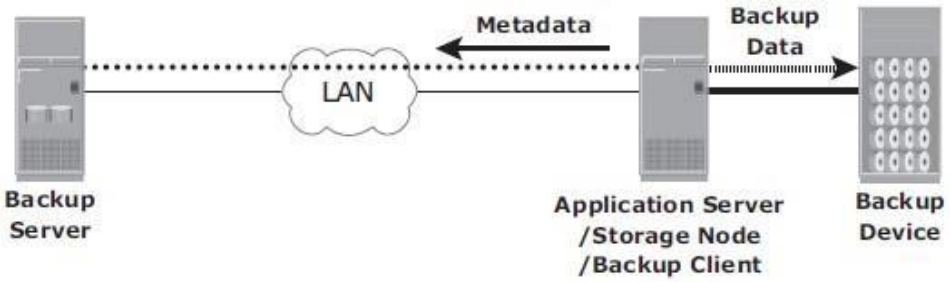

Fig 3.7: Direct-attached backup topology

➢ 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.

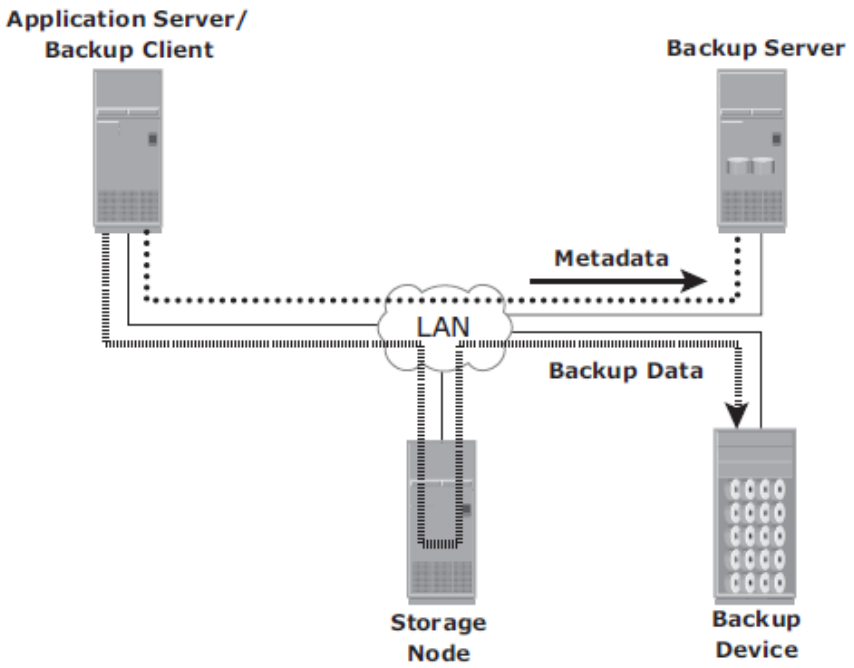

Fig 3.8: LAN-based backup topology

➢ 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.

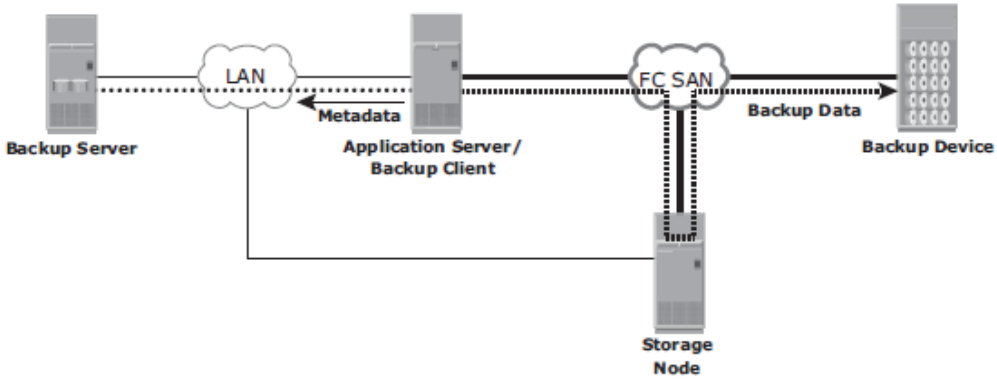

Fig 3.9: SAN-based backup topology

➢ The emergence of low-cost disks as a backup medium has enabled disk arrays to be attached to the SAN and used as backup devices. A tape backup of these data backups on the disks can be created and shipped offsite for disaster recovery and long-term retention.

➢ The mixed topology uses both the LAN-based and SAN-based topologies, as shown in Fig 3.10. This topology might be implemented for several reasons, including cost, server location, reduction in administrative overhead, and performance considerations.

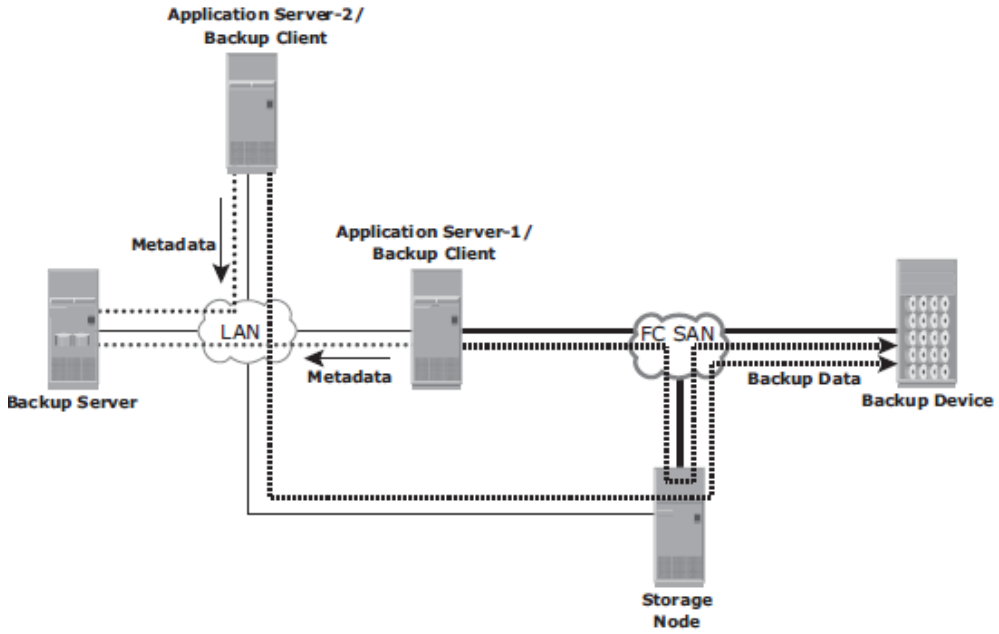

Fig 3.10: Mixed backup topology

7 List various uses of local replication. Explain storage array based local replication with a neat diagram.

Page 272 – text-book

8 Differentiate between Synchronous and Asynchronous based remote replication model.

Refer page 295 from textbook

9 Explain local Replication technology using Host based methods.

Refer page: 269 from text-book

10 Write a short notes on the following ; i) Three site Replications ii) Network based Remote Replication.

Refer page: 298 from text-book