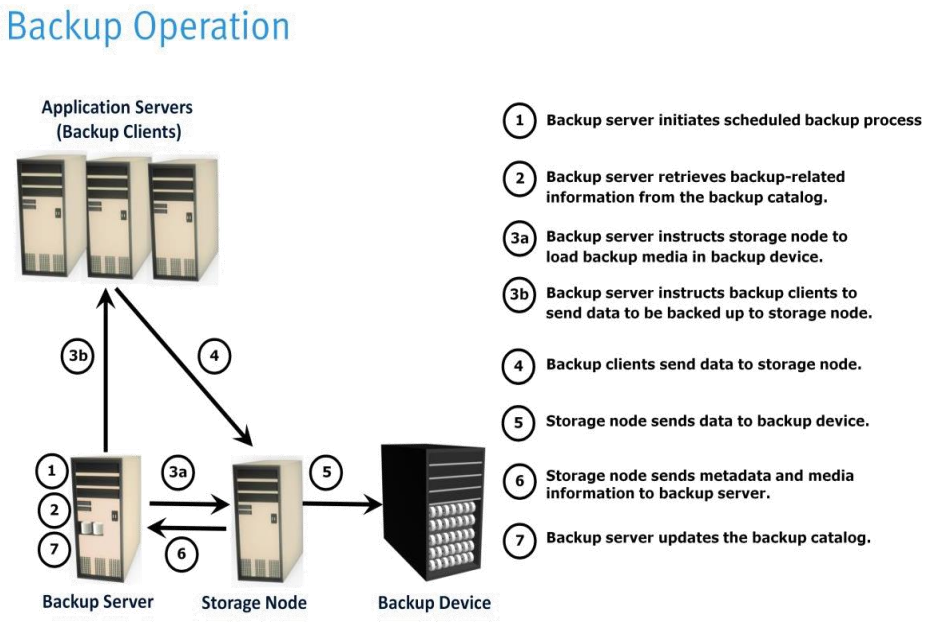
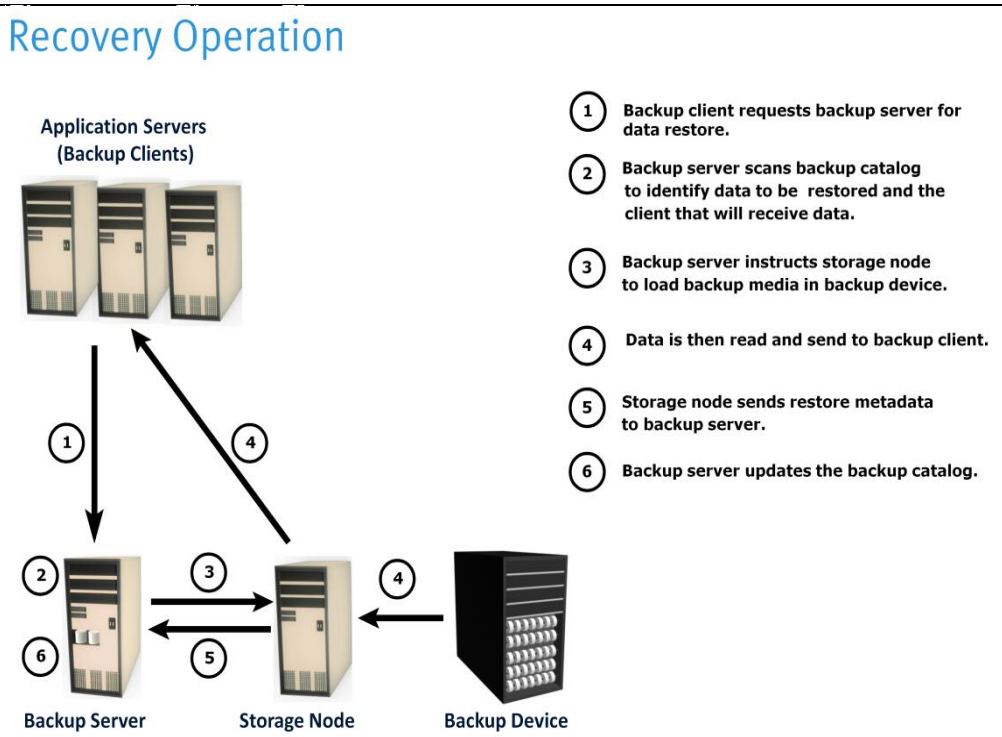
IAT 3 Scheme and Solution- Storage Area Network (18CS822)

USN | | | | | | | | | | |

Internal Assessment Test III – May 2024

Sub:	Storage Area Network					Sub Code:	18CS822	Branch:	ISE		
Date:	11-05-2024	Duration:	90 min's	Max Marks:	50	Sem/Sec:	VIII C			OBE	
Answer any FIVE FULL Questions									M	CO	RBT
1	a)Illustrates backup and recovery operations with a neat diagram								5+5	CO3	L3





1. b) Write short note on risk triad.

Solution:

Definition: 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.

Risk assessment is the first step in determining the extent of potential threats and risks in an IT infrastructure.

The process assesses risk and helps to identify appropriate controls to mitigate or eliminate risks.

Assets:

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. Several factors need to be considered when planning for asset security.

Security methods have two objectives.

1. It has to ensure that the network is easily accessible to authorized users. It should also be reliable and stable under disparate environmental conditions and volumes of usage.

2. It has 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.

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.

Threats:

Threats are the potential attacks that can be carried out on an IT infrastructure.

These attacks can be classified as active or passive.

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.

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.

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.

Vulnerabilities:

The paths that provide access to information are the most vulnerable to potential attacks. Each of these paths may contain various access points, each of which provides different levels of access to the storage resources.

It is very important to implement adequate security controls at all the access points on an access path. Implementing security controls at each access point of every access path is termed as defense in depth. Attack surface, attack vector, and work factor are 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.

An attack vector is a step or a series of steps necessary to complete an attack.

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. The controls can be categorized as preventive, detective, corrective, recovering, or compensating. 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.			
2	Analyze LVM based local replication, with a neat diagram. Discuss the advantages and Disadvantages. **Solution: 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.** **In LVM-based replication, logical volume manager is responsible for creating and controlling the host-level logical volume. An LVM has three components: physical volumes (physical disk), volume groups, and logical volumes. A volume group is created by grouping together one or more physical volumes. Logical volumes are created within a given volume group. A volume group can have multiple logical volumes.** **In LVM-based replication, each logical partition in a logical volume is mapped to two physical partitions on two different physical volumes, as shown in Figure. An application write to a logical partition is written to the two physical partitions by the LVM device driver. This is also known as LVM mirroring.** **Mirrors can be split and the data contained therein can be independently accessed. LVM mirrors can be added or removed dynamically.**	10	CO4	L3

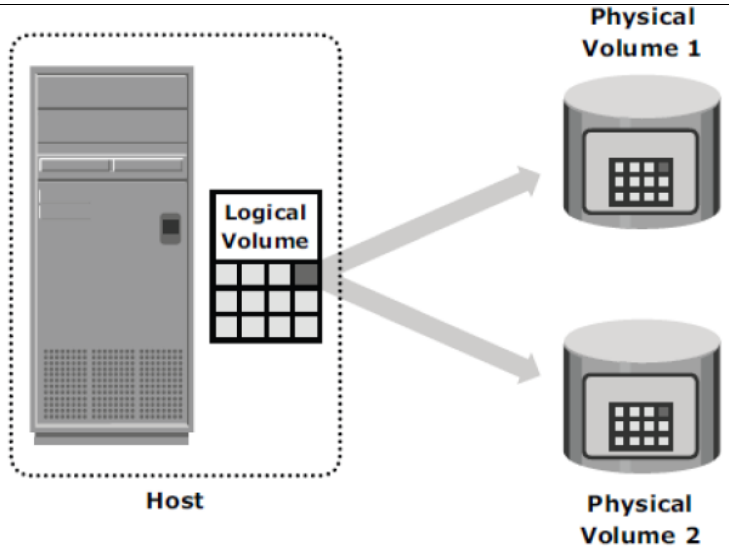


Fig: LVM Based Replication

Advantage:

The LVM-based replication technology is not dependent on a vendor-specific storage system. Typically,

LVM is part of the operating system and no additional license is required to deploy LVM mirroring.

Disadvantage:

As every write generated by an application translates into two writes on the disk, an additional burden is placed on the host CPU. This can degrade application performance.

Presenting an LVM-based local replica to a second host is usually not possible because the replica will still be part of the volume group, which is usually accessed by one host at any given time.

Tracking changes to the mirrors and performing incremental synchronization operations is also a challenge as all LVMs do not support incremental resynchronization. If the devices are already protected by some level of RAID on the array, then the additional protection provided by mirroring is unnecessary.

3	Describe synchronous and asynchronous modes of remote replication with a neat diagram **The two basic modes of remote replication are synchronous and asynchronous.** **Synchronous Replication:** **In synchronous remote replication, writes must be committed to the source and the target, prior to acknowledging "write complete" to the**	10	CO4	L2

host. Additional writes on the source cannot occur until each preceding write has been completed and acknowledged.

This ensures that data is identical on the source and the replica at all times.

In the event of a failure of the source site, synchronous remote replication provides zero or near-zero RPO, as well as the lowest RTO. However, application response time is increased with any synchronous remote replication.

The degree of the impact on the response time depends on the distance between sites, available bandwidth, and the network connectivity infrastructure.

The distances over which synchronous replication can be deployed depend on the application's ability to tolerate extension in response time. Typically, it is deployed for distances less than 200 KM between the two sites.

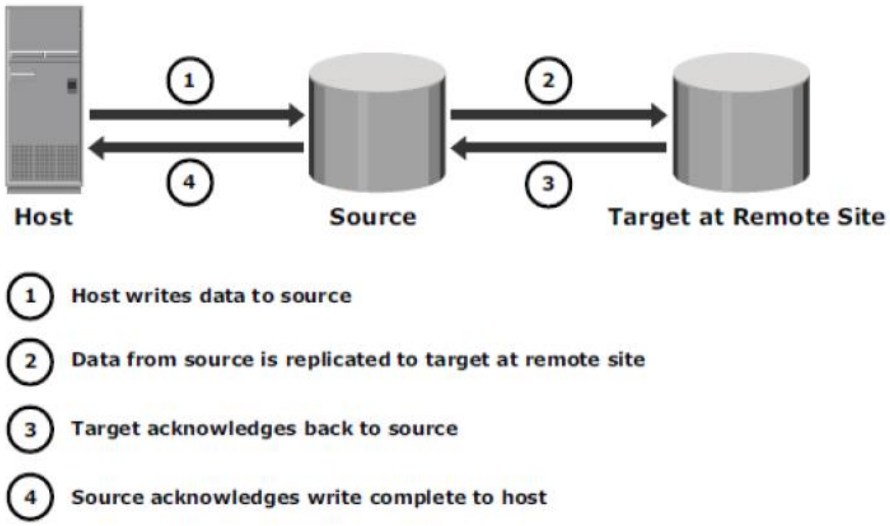


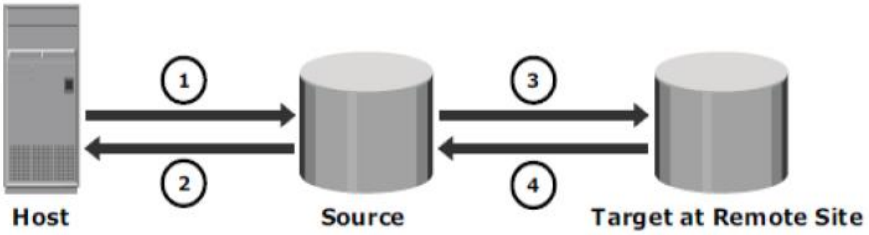
Fig: Synchronous Replication

Asynchronous replication:

In asynchronous remote replication, a write is committed to the source and immediately acknowledged to the host. Data is buffered at the source and transmitted to the remote site later.

This eliminates the impact to the application's response time.

Data at the remote site will be behind the source by at least the size of the buffer. Hence, asynchronous remote replication provides a finite (nonzero) RPO disaster recovery solution. RPO depends on the size of the buffer, available network bandwidth, and the write workload to the source.

	There is no impact on application response time, as the writes are acknowledged immediately to the source host. This enables deployment of asynchronous replication over extended distances. **Asynchronous remote replication can be deployed over distances ranging from several hundred to several thousand kilometers between two sites.**  1 Host writes data to source 2 Write is immediately acknowledged to host 3 Data is transmitted to the target at remote site later 4 Target acknowledges back to source **Fig: Asynchronous Replication**			
4	Elucidate the construction of different Backup Topologies **Solution:** Three basic topologies are used in a backup environment: direct-attached backup, LAN-based backup, and SAN-based backup. A mixed topology is also used by combining LAN-based and SAN-based topologies. In a direct-attached backup, the storage node is confi gured on a backup client, and the backup device is attached directly to the client. Only the metadata is sent to the backup server through the LAN. This confi guration frees the LAN from backup traffi c. The example in Figure 10-7 shows that the backup device is directly attached and dedicated to the backup client. As the environment grows, there will be a need for centralized management and sharing of backup devices to optimize costs. An appropriate solution is required 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.	10	CO3	L2

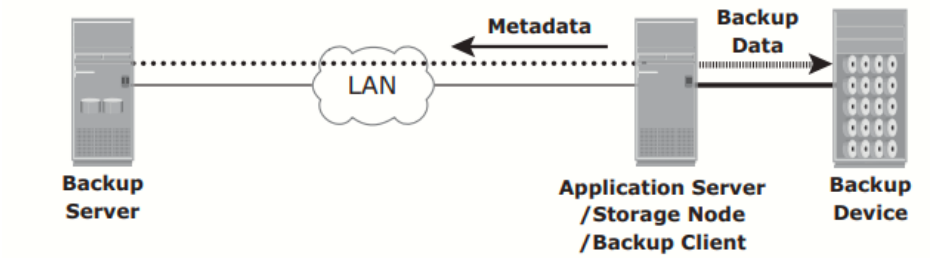
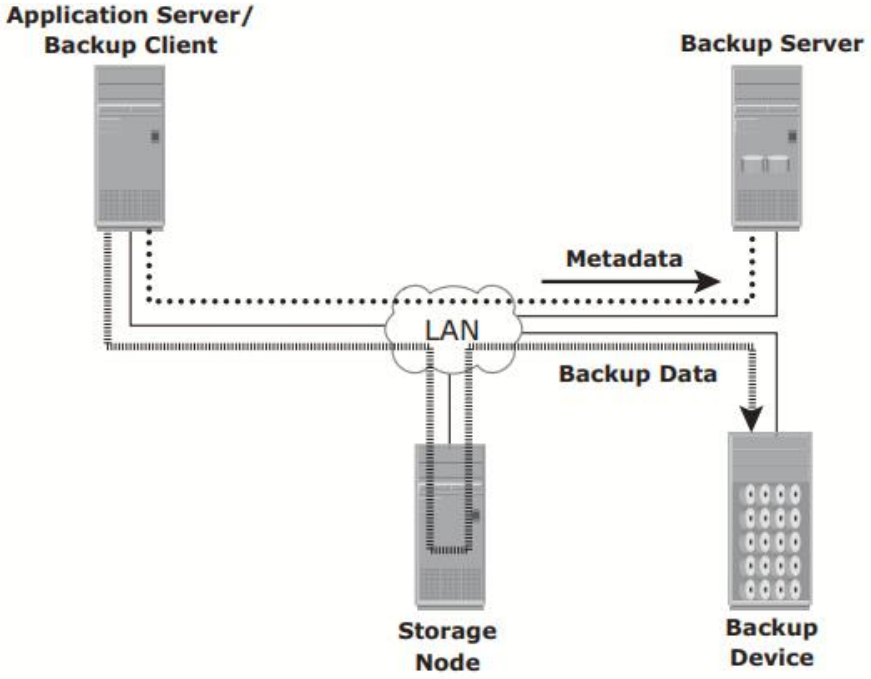


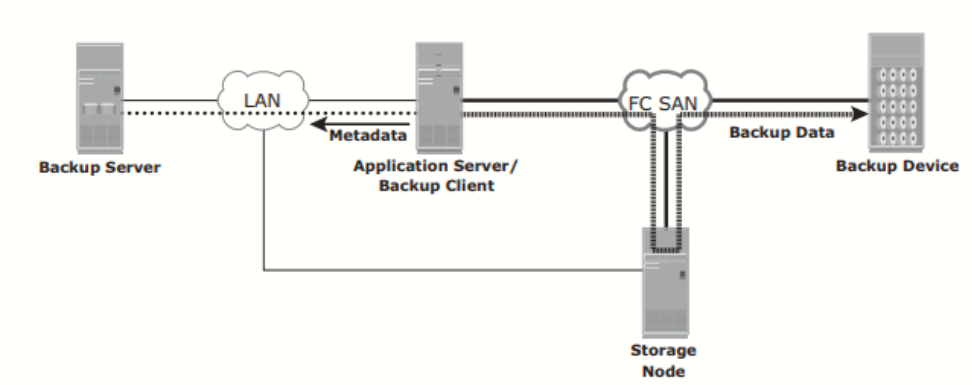
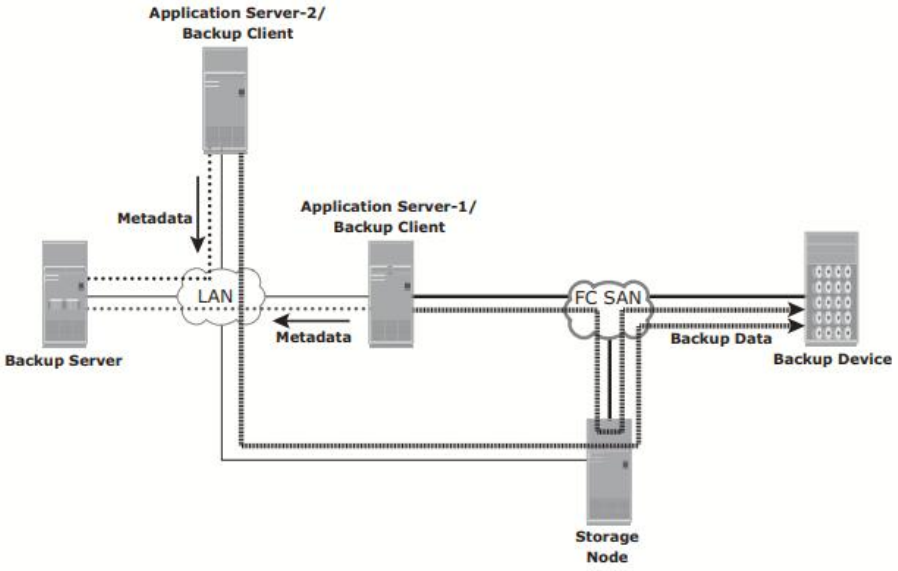
Figure 10-7: Direct-attached backup topology

In a LAN-based backup, the clients, backup server, storage node, and backup device are connected to the LAN. (see Figure 10-8). The data to be backed up is transferred from the backup client (source) to the backup device (destination) over the LAN, which might affect network performance.



igure 10-8: LAN-based backup topology

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. A SAN-based backup is also known as a LAN-free backup. The SAN-based backup topology is the most appropriate solution when a backup device needs to be shared among clients. In this case, the backup device and clients are attached to the SAN. Figure 10-9 illustrates a SAN-based backup. In this example, 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. 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.  **Figure 10-9:** SAN-based backup topology The mixed topology uses both the LAN-based and SAN-based topologies, as shown in Figure 10-10. This topology might be implemented for several reasons, including cost, server location, reduction in administrative overhead, and performance considerations.  **Figure 10-10:** Mixed backup topology			
5	a) List and explain the replica terminologies	7+3	CO4	L2

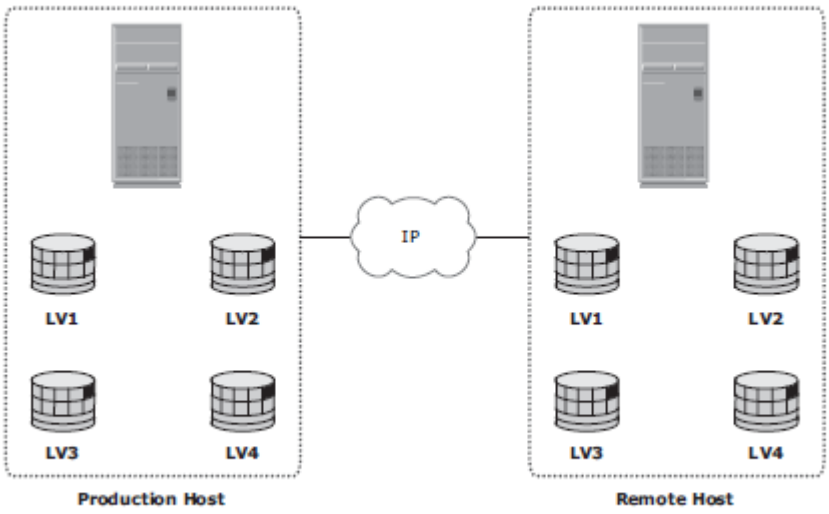
Source: A host accessing the production data from one or more LUNs on the storage array is called a production host, and these LUNs are known as source LUNs (devices/volumes), production LUNs, or simply the source.
Target: A LUN (or LUNs) on which the production data is replicated, is called the target LUN or simply the target or replica.
Point-in-Time (PIT) and continuous replica: Replicas can be either a PIT or a continuous copy. The PIT replica is an identical image of the source at some specifi c timestamp. For example, if a replica of a fi le system is created at 4:00 p.m. on Monday, this replica is the Monday 4:00 p.m. PIT copy. On the other hand, the continuous replica is in-sync with the production data at all times.
Recoverability and restartability: Recoverability enables restoration of data from the replicas to the source if data loss or corruption occurs.
Restartability enables restarting business operations using the replicas. The replica must be consistent with the source so that it is usable for both recovery and restart operations.

b) List and explain the uses of local Replica

Alternative source for backup: Under normal backup operations, data is read from the production volumes (LUNs) and written to the backup device. This places an additional burden on the production infrastructure because production LUNs are simultaneously involved in production operations and servicing data for backup operations. The local replica contains an exact point-in-time (PIT) copy of the source data, and therefore can be used as a source to perform backup operations. This alleviates the backup I/O workload on the production volumes. Another benefi t of using local replicas for backup is that it reduces the backup window to zero.

Fast recovery: If data loss or data corruption occurs on the source, a local replica might be used to recover the lost or corrupted data. If a complete failure of the source occurs, some replication solutions enable a replica to be used to restore data onto a different set of source devices, or production can be restarted on the replica. In either case, this method provides faster recovery and minimal RTO compared to traditional recovery from tape backups. In many instances, business operations can be started using the source device before the data is completely copied from the replica.

Decision-support activities, such as reporting or data warehousing: Running the reports using the data on the replicas greatly reduces the I/O burden placed on the production device. Local replicas are also used for data-warehousing applications. The data-warehouse application may be populated by the data on the replica and thus avoid the impact on the production environment.

	Testing platform: Local replicas are also used for testing new applications or upgrades. For example, an organization may use the replica to test the production application upgrade; if the test is successful, the upgrade may be implemented on the production environment. **Data migration:** Another use for a local replica is data migration. Data migrations are performed for various reasons, such as migrating from a smaller capacity LUN to one of a larger capacity for newer versions of the application.			
6	Explain the Remote Replication technologies **Host-Based Remote Replication** Host-based remote replication uses the host resources to perform and manage the replication operation. There are two basic approaches to host-based remote replication: Logical volume manager (LVM) based replication and database replication via log shipping. **LVM-Based Remote Replication** LVM-based remote replication is performed and managed at the volume group level. Writes to the source volumes are transmitted to the remote host by the LVM. The LVM on the remote host receives the writes and commits them to the remote volume group.  **Figure 12-5:** LVM-based remote replication **Host-Based Log Shipping** Database replication via log shipping is a host-based replication technology supported by most databases. Transactions to the source database are captured in logs, which are periodically transmitted by the source host to the	10	CO4	L2

remote host (see Figure 12-6). The remote host receives the logs and applies them to the remote database.

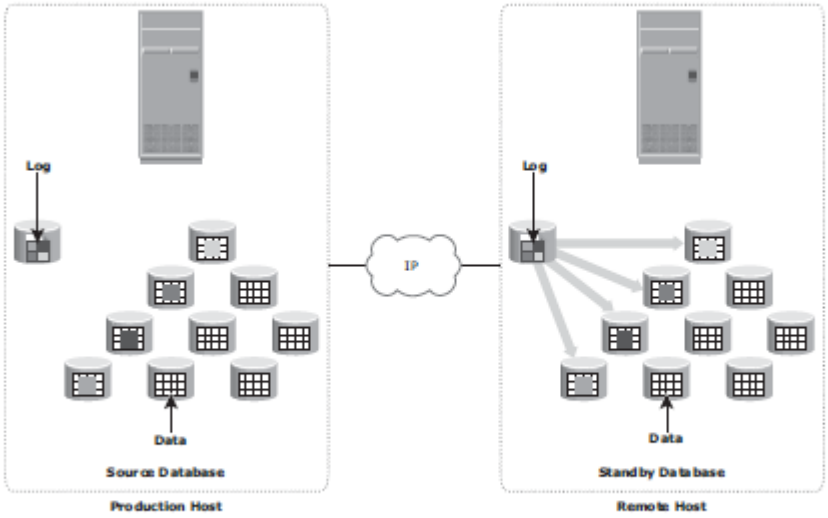


Figure 12-6: Host-based log shipping