Unit-IV

Backup Purpose, Backup Considerations, Backup Granularity, Recovery Considerations, Backup Methods, Backup Process, Backup and restore Operations, Backup Topologies, Backup in NAS Environments, Backup .Technologies

Local Replication, Remote Replication: Source and Target, Uses of Local Replicas, Data Consistency, Local Replication Technologies, Restore and Restart Considerations, Creating Multiple Replicas, Management Interface, Modes of Remote Replication, Remote Replication Technologies, Network Infrastructure

Key Concepts:

- Operational Backup
- Archival
- Retention Period
- Bare-Metal Recovery
- Backup Architecture
- Backup Topologies
- Virtual Tape Library

- *A backup* is a copy of production data, created and retained for the sole purpose of recovering deleted or corrupted data.
- Evaluating backup technologies, recovery, and retention requirements for data and applications is an essential step to ensure successful implementation of the backup and recovery solution.

12.1 Backup Purpose

- Backups are performed to serve three purposes: **disaster recovery, operational backup, and archival.**

12.1.1 Disaster Recovery

- The backup copies are used for restoring data at an alternate site when the primary site is incapacitated due to a disaster.
- When a tape-based backup method is used as a disaster recovery strategy, the backup tape media is shipped and stored at an offsite location.
- These tapes can be recalled for restoration at the disaster recovery site.

12.1.2 Operational Backup

- *Operational backup* is a backup of data at a point in time and is used to restore data in the event of data loss or logical corruptions that may occur during routine processing.
- Operational backups are created for the active production information by using incremental or differential backup techniques.
- This ensures the availability of a clean copy of the production database if the batch update corrupts the production database.

12.1.3 Archival

- Backups are also performed to address archival requirements. Although CAS has emerged as the primary solution for archives, traditional backups are still used by small and medium enterprises for long-term preservation of transaction records, e-mail messages, and other business records required for regulatory compliance.
- backups serve as a protection against data loss due to physical damage of a storage device, software failures, or virus attacks and can also be used to protect against accidents such as a deletion or intentional data destruction.

12.2 Backup Considerations

- 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.

- Location, size, and number of files should also be considered, as they may affect the backup process.

- Backing up large-size files (example: ten 1 MB files) may use less system resources than backing up an equal amount of data comprising a large number of small-size files (example: ten thousand 1KB files). The backup and restore operation takes more time when a file system contains many small files.

- Backup performance also depends on the media used for the backup. The time-consuming operation of starting and stopping in a tape-based system affects backup performance, especially while backing up a large number of small files.

12.3 Backup 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.

- **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.

- ***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.

Recovery Considerations 12.4

- RPO and RTO are major considerations when planning a backup strategy.
- The retention period for a backup is also derived from an RPO specified for operational recovery.
- If short retention periods are specified for backups, it may not be possible to recover all the data needed for the requested recovery point, as some data may be older than the retention period.

Backup Methods 12.5

- **Hot backup** and **cold backup** are the two methods deployed for backup.
- 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.
- The backup of online *production data* becomes more challenging because data is actively being used and changed.
- The backup application can back up open files by retrying the operation on files that were opened earlier in the backup process.
- Consistent backups of databases can also be done by using a cold backup. This requires the database to remain inactive during the backup.
- Hot backup is used in situations where it is not possible to shut down the database.

Backup Process 12.6

- A backup system uses client/server architecture with a backup server and multiple backup clients. The backup server manages the backup operations and maintains the backup catalog, which contains information about the backup process and backup metadata.
- Figure 12-4 illustrates the backup process. The storage node is responsible for writing data to the backup device (in a backup environment, a storage node is a host that controls backup devices).
- A backup device is attached directly to the storage node's host platform. Some backup architecture refers to the storage node as the *media server* because it connects to the storage device.
- The backup process is based on the policies defined on the backup server, such as the time of day or completion of an event. The backup server then initiates the process by sending a request to a backup client.
- After all the data is backed up, the storage node closes the connection to the backup device. The backup server writes backup completion status to the metadata catalog.

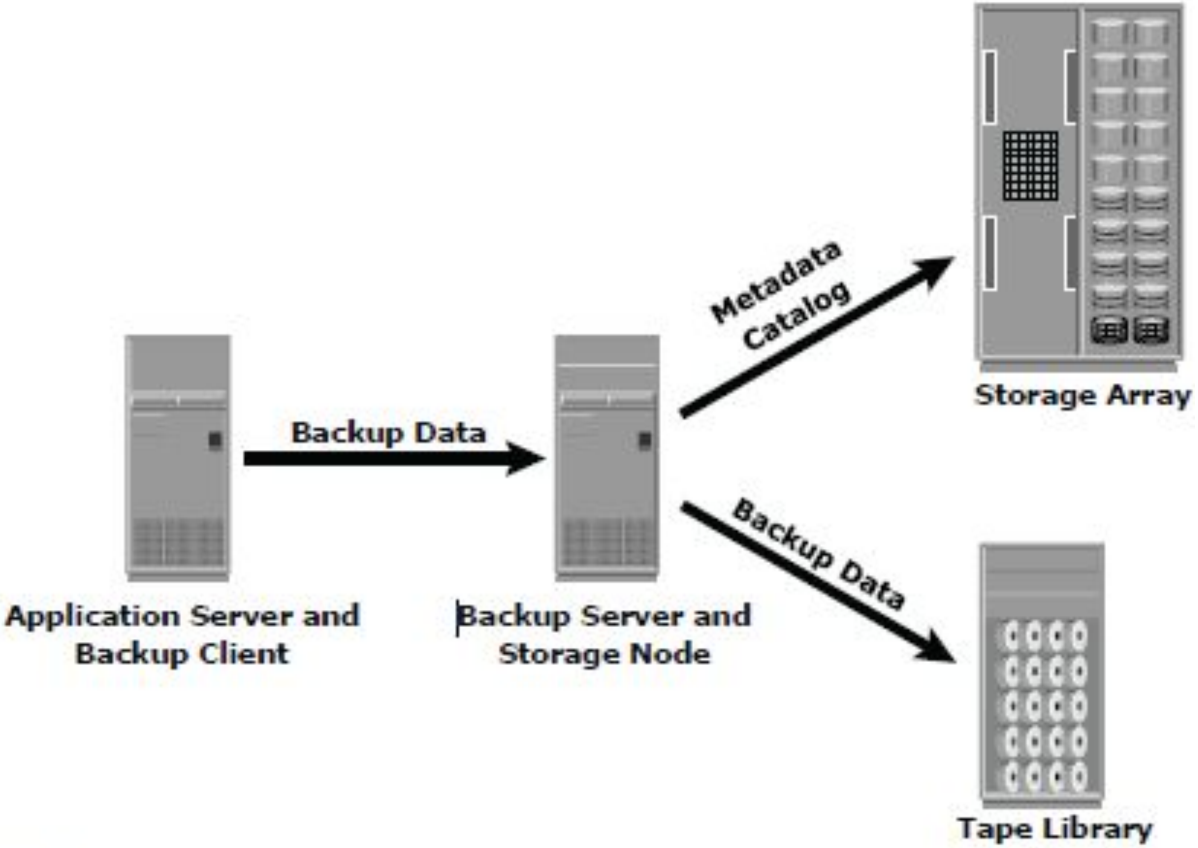

Metadata Catalog
Storage Array
Backup Data
Application Server and Backup Client
Backup Server and Storage Node
Backup Data
Tape Library

Figure 12-4: Backup architecture and process

Backup and Restore Operations 12.7

- When a backup process is initiated, significant network communication takes place between the different components of a backup infrastructure. The backup server initiates the backup process for different clients based on the backup schedule configured for them.
- The backup server coordinates the backup process with all the components in a backup configuration.
- The backup server maintains the information about backup clients to be contacted and storage nodes to be used in a backup operation.
- The storage node, in turn, sends metadata to the backup server to keep it updated about the media being used in the backup process. The backup server continuously updates the backup catalog with this information.

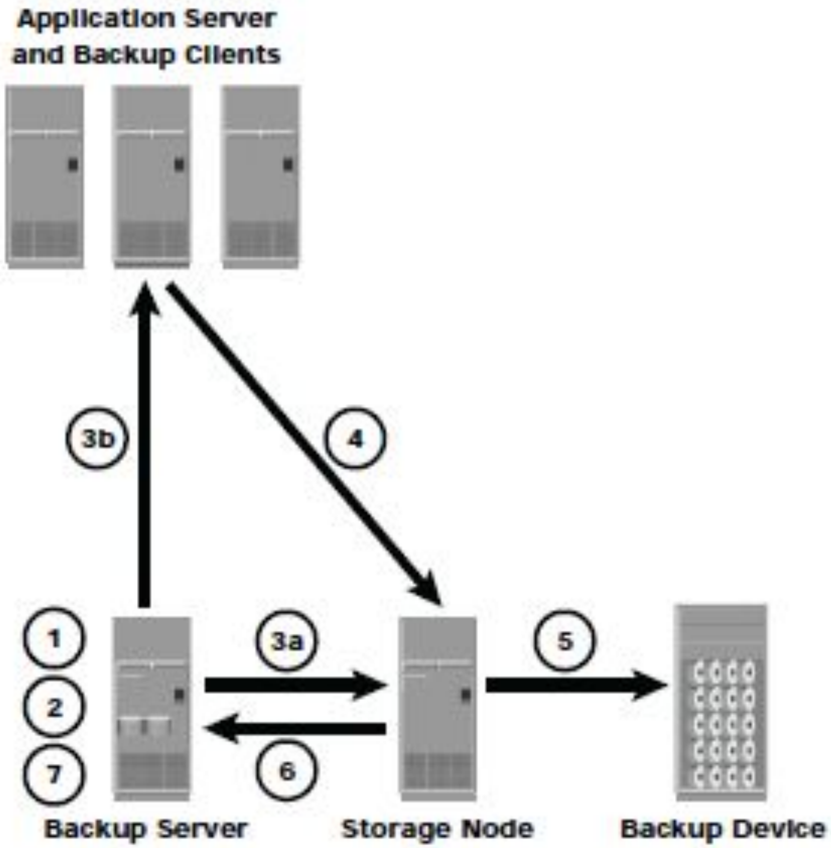


Figure 12-5: Backup operation

1 Start of scheduled backup process

2 Backup server retrieves backup related information from backup catalog

3a Backup server instructs storage node to load backup media in backup device

3b Backup server instructs backup clients to sends it's metadata to backup server and data to be backed up to storage node

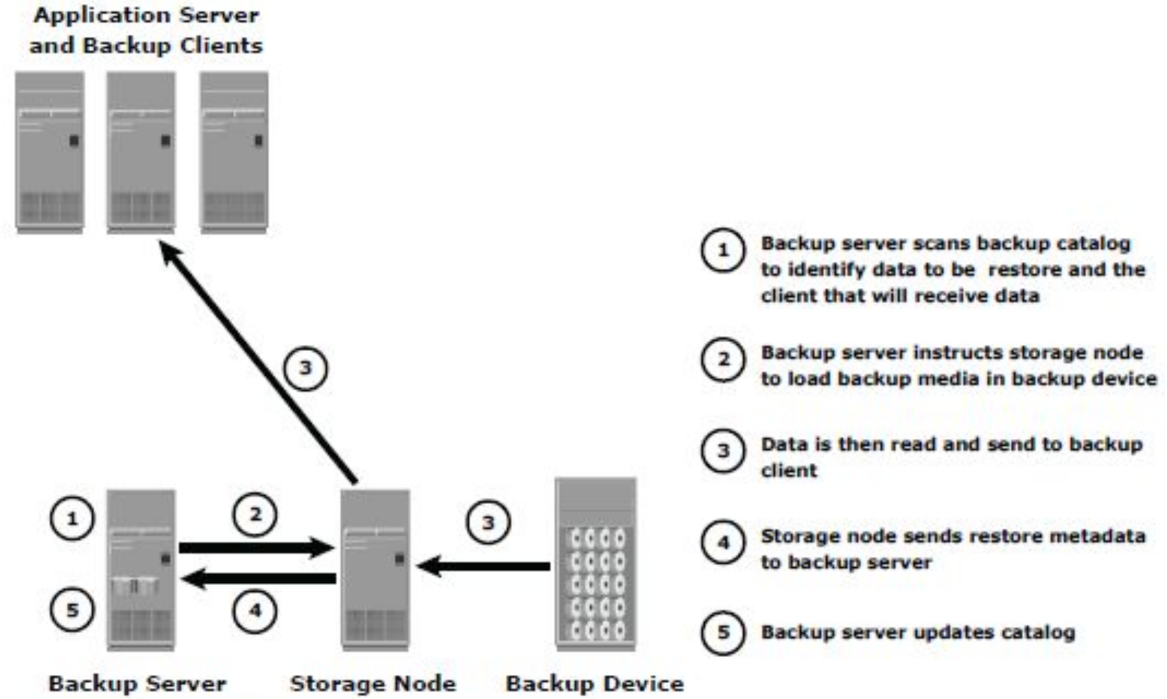
4 Backup clients send data to storage node

5 Storage node sends data to backup device

6 Storage node sends metadata and media information to Backup server

7 Backup server update catalog and records the status

After the data is backed up, it can be restored when required. A restore process
.must be manually initiated



1 Backup server scans backup catalog to identify data to be restore and the client that will receive data

2 Backup server instructs storage node to load backup media in backup device

3 Data is then read and send to backup client

4 Storage node sends restore metadata to backup server

5 Backup server updates catalog

Figure 12-6: Restore operation

Backup Topologies 12.8

- 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*, 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.

- 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. the client also acts as a storage node that writes data on the backup device.

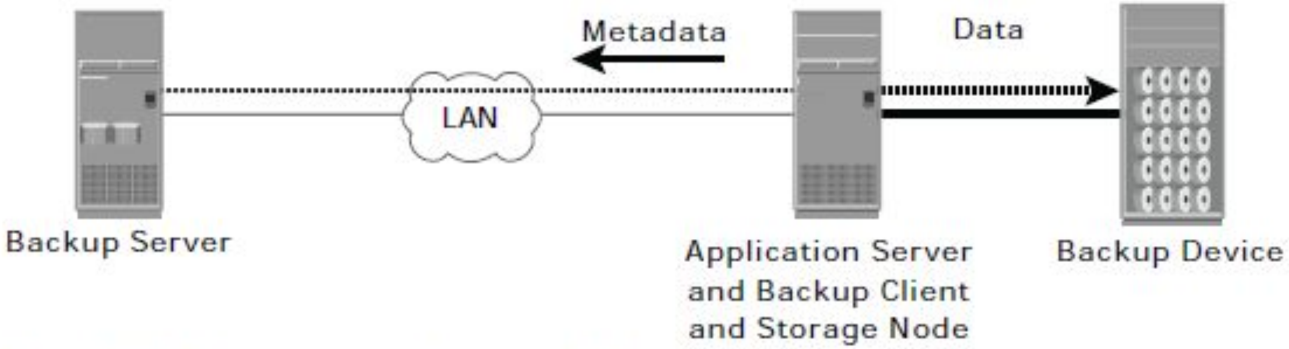


Figure 12-7: Direct-attached backup topology

- In *LAN-based backup*, all servers are connected to the LAN and all storage devices are directly attached to the storage node.
- 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. Streaming across the LAN also affects network performance of all systems connected to the same segment as the backup server.

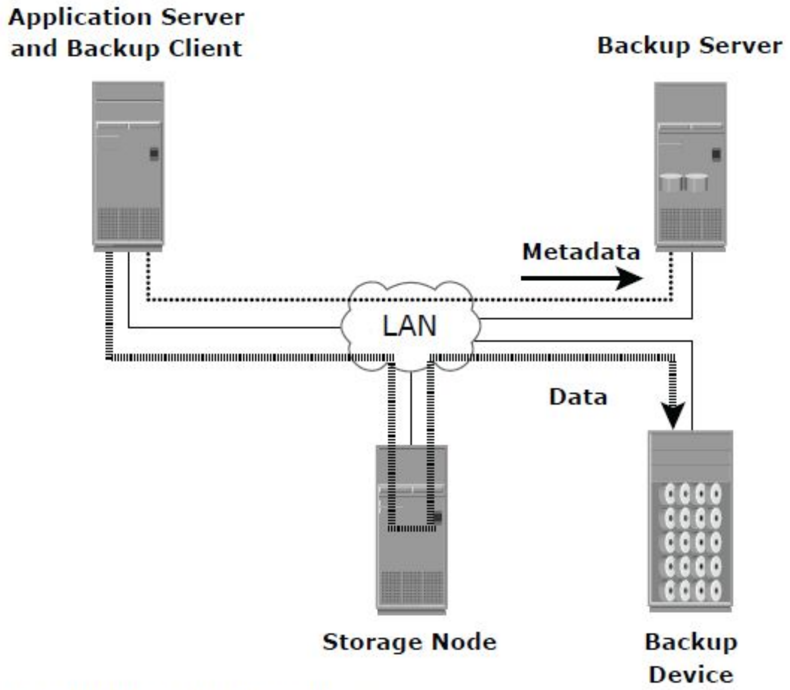


Figure 12-8: LAN-based backup topology

- The *SAN-based backup* is also known as the *LAN-free backup*. 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.

- clients read the data from the mail servers in the SAN and write to the SAN attached backup device. The backup data traffic is restricted to the SAN, and backup metadata is transported over the LAN.

- By removing the network bottleneck, the SAN improves backup to tape performance because it frees the LAN from backup traffic. At the same time, LANfree backups may affect the host and the application, as they consume host I/O bandwidth, memory, and CPU resources.

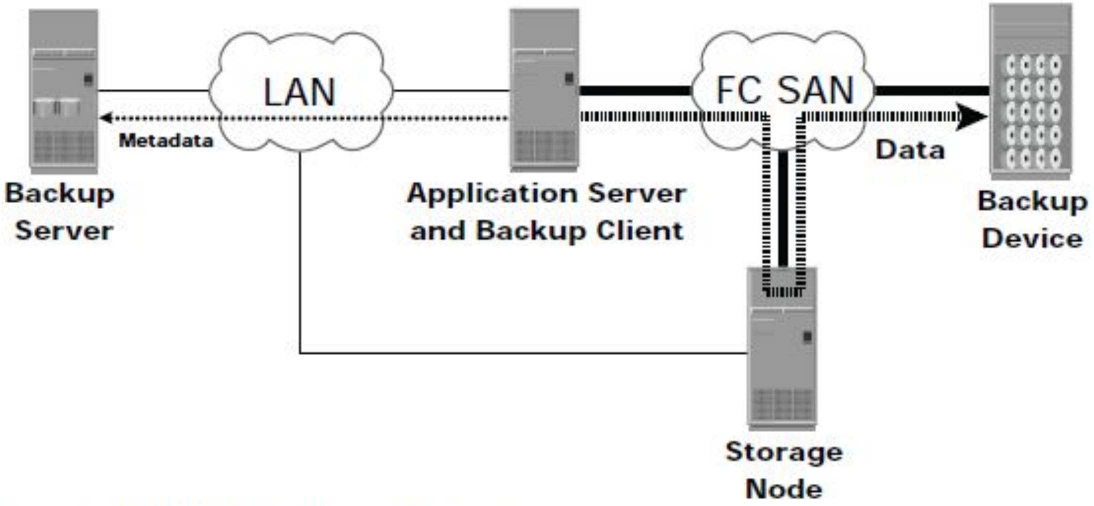


Figure 12-9: SAN-based backup topology

- The *mixed topology* uses both the LAN-based and SAN-based topologies, as shown in Figure 12-10. This topology might be implemented for several reasons, including cost, server location, reduction in administrative overhead, and performance considerations.

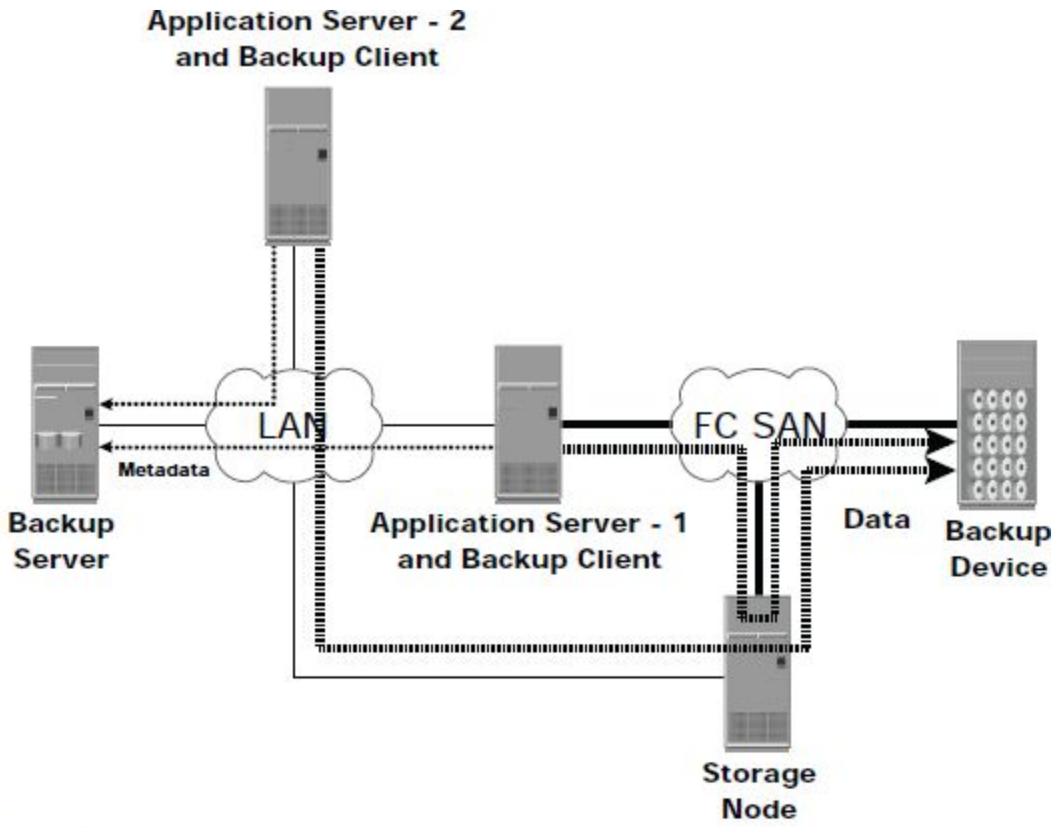


Figure 12-10: Mixed backup topology

Serverless Backup 12.8.1

- *Serverless backup* is a LAN-free backup methodology that does not involve a backup server to copy data. The copy may be created by a network-attached controller, utilizing a SCSI extended copy or an appliance within the SAN.
- Another widely used method for performing serverless backup is to leverage local and remote replication technologies. In this case, a consistent copy of the production data is replicated within the same array or the remote array, which can be moved to the backup device through the use of a storage node.

Backup in NAS Environments 12.9

- NAS heads use a proprietary operating system and file system structure supporting multiple file-sharing protocols.

- In the NAS environment, backups can be implemented in four different ways: **server based**, **serverless**, or using **Network Data Management Protocol** (NDMP) in either **NDMP 2-way** or **NDMP 3-way.**

- In *application* ***server-based*** *backup*, the NAS head retrieves data from storage over the network and transfers it to the backup client running on the application server.

- In *serverless backup*, the network share is mounted directly on the storage node. This avoids overloading the network during the backup process and eliminates the need to use resources on the production server.

- In **NDMP**, backup data is sent directly from the NAS head to the backup device, while metadata is sent to the backup server.

- In NDMP 2-way model, network traffic is minimized by isolating data movement from the NAS head to the locally attached tape library. Only metadata is transported on the network. This backup solution meets the strategic need to centrally manage and control distributed data while minimizing network traffic.

- In an *NDMP 3-way* file system, data is not transferred over the public network. A separate private backup network must be established between all NAS heads and the “backup” NAS head to prevent any data transfer on the public network in order to avoid any congestion or affect production operations. Metadata and NDMP control data is still transferred across the public network.

- NDMP 3-way is useful when you have limited backup devices in the environment. It enables the NAS head to control the backup device and share it with other NAS heads by receiving backup data through NDMP

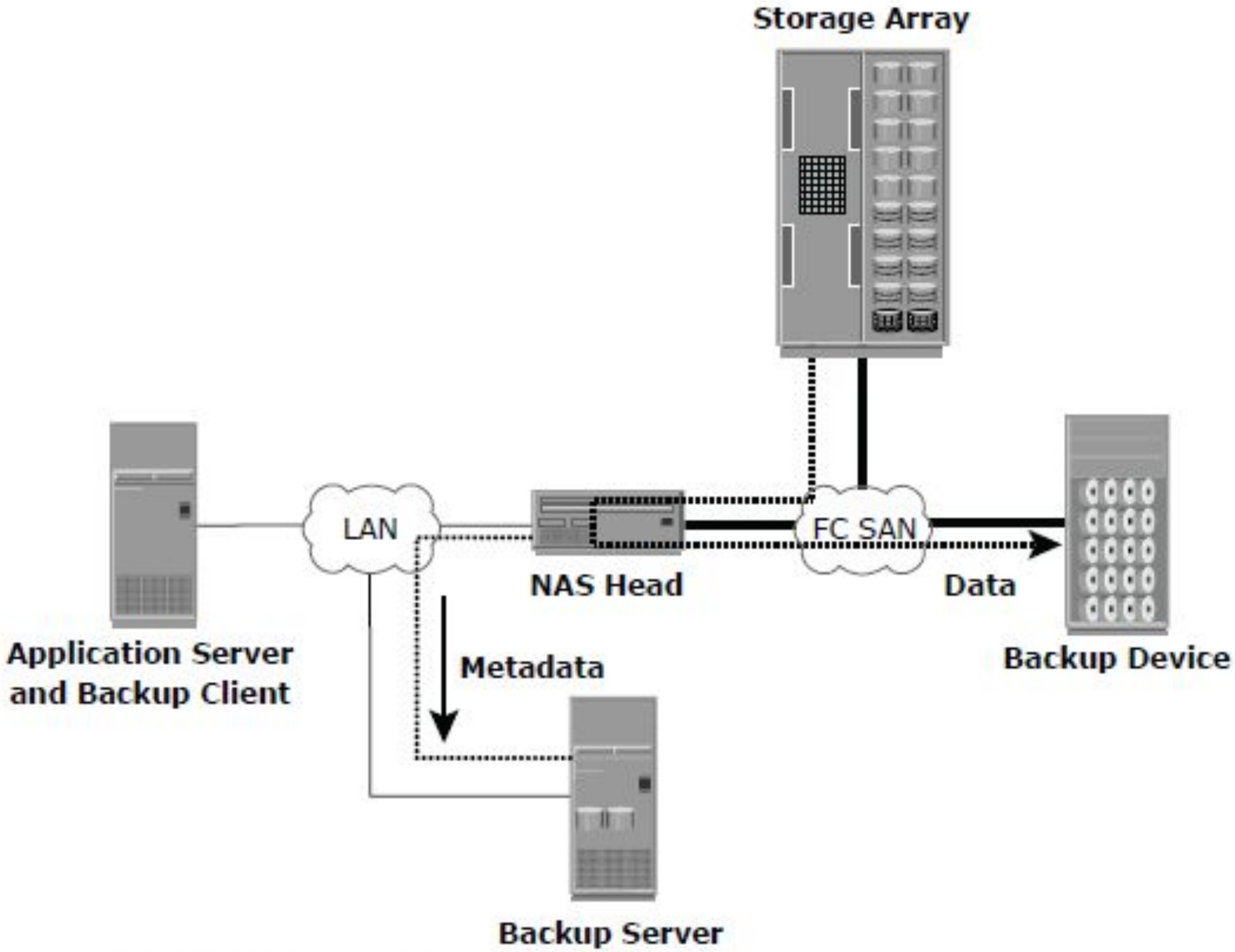


Figure 12-13: NDMP 2-way in NAS environment

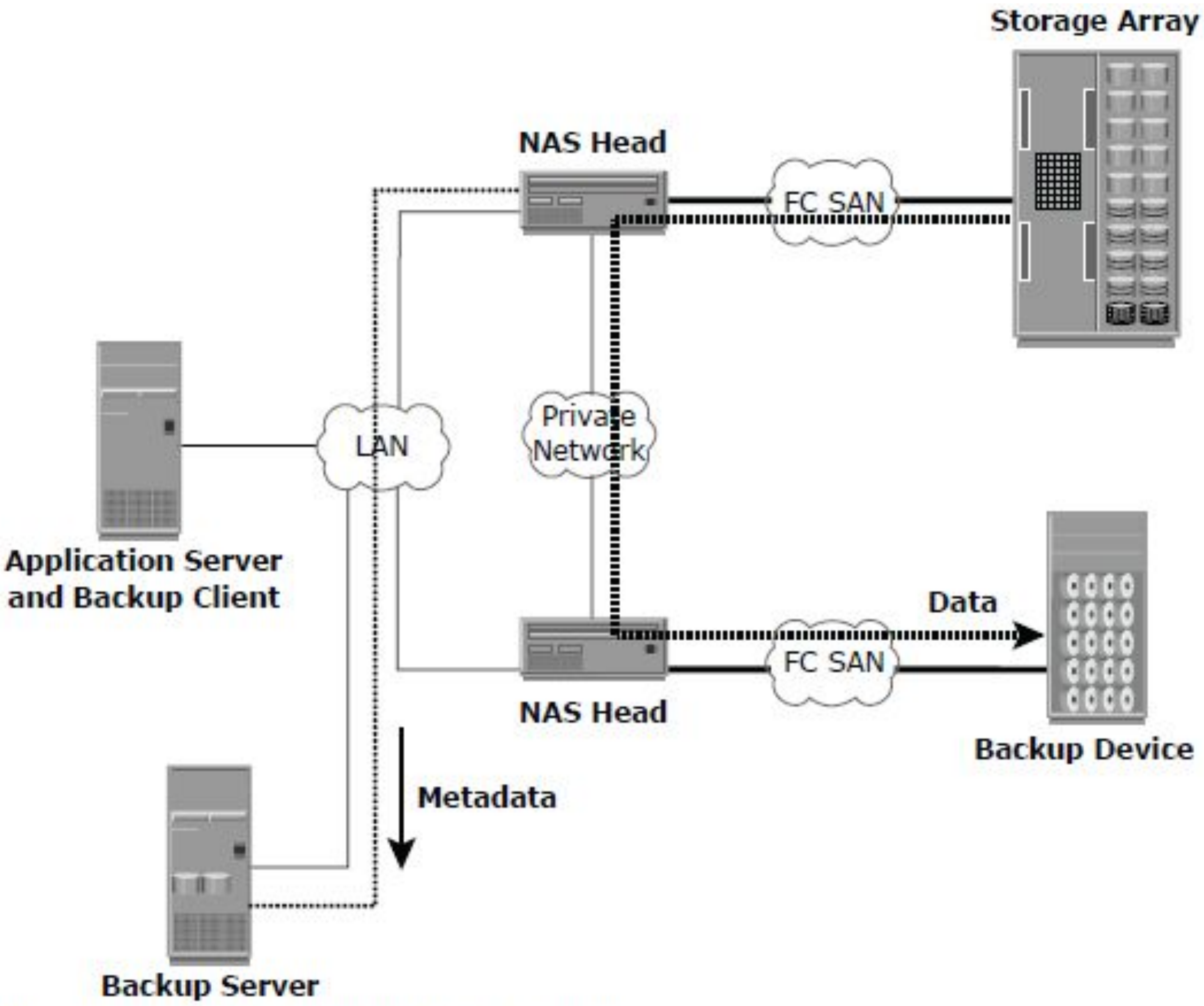


Figure 12-14: NDMP 3-way in NAS environment

Backup Technologies 12.10

- A wide range of technology solutions are currently available for backup. Tapes and disks are the two most commonly used backup media.
- Tapes are primarily used for long-term offsite storage because of their low cost.
- Tapes must be stored in locations with a controlled environment to ensure preservation of the media and prevent data corruption. Data access in a tape is sequential, which can slow backup and recovery operations.

Backup to Tape 12.10.1

- Tapes, a low-cost technology, are used extensively for backup. Tape drives are used to read/write data from/to a tape cartridge. Tape drives are referred to as sequential, or linear, access devices because the data is written or read sequentially.
- *Tape Mounting* is the process of inserting a tape cartridge into a tape drive. The tape drive has motorized controls to move the magnetic tape around, enabling the head to read or write data.
- A linear recording method was used in older tape drive technologies. This recording method consisted of data being written by multiple heads in parallel tracks, spanning the whole tape.

Physical Tape Library 12.10.2

- 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.

- *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. Another type of slot called a *mail* or *import/export slot* is used to add or remove tapes from the library without opening the access doors(Figure 12-15)

- When a backup process starts, the robotic arm is instructed to load a tape to a tape drive. This process adds to the delay to a degree depending on the type of hardware used, but it generally takes 5 to 10 seconds to mount a tape.

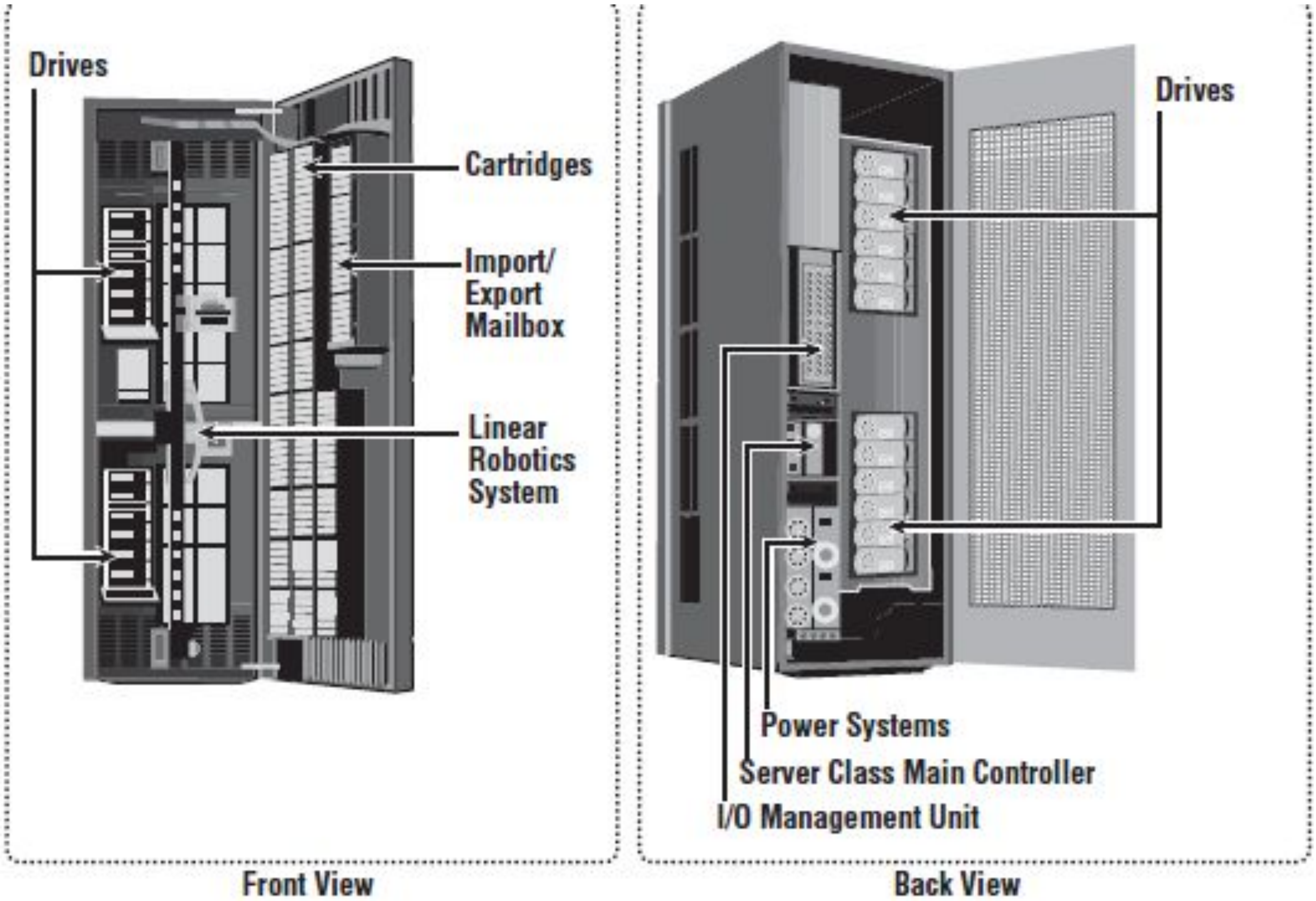


Figure 12-15: Physical tape library

Backup to Disk 12.10.3

- 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 cost, 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.
- Backup to disk does not offer any inherent offsite capability, and is dependent on other technologies such as local and remote replication.

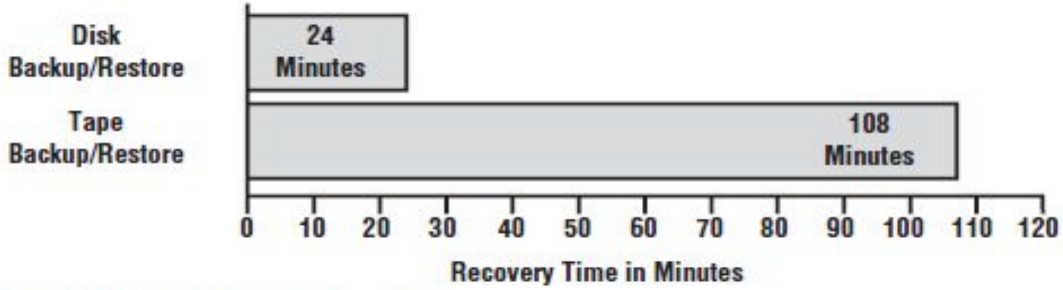


Figure 12-17: Tape versus disk restore

Virtual Tape Library 12.10.4

- 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.

- Virtual tape libraries use 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.

- Similar to a physical tape library, a robot mount is performed when a backup process starts in a virtual tape library. in a virtual tape library it is almost instantaneous. Even the *load to ready* time is much less than in a physical tape library.

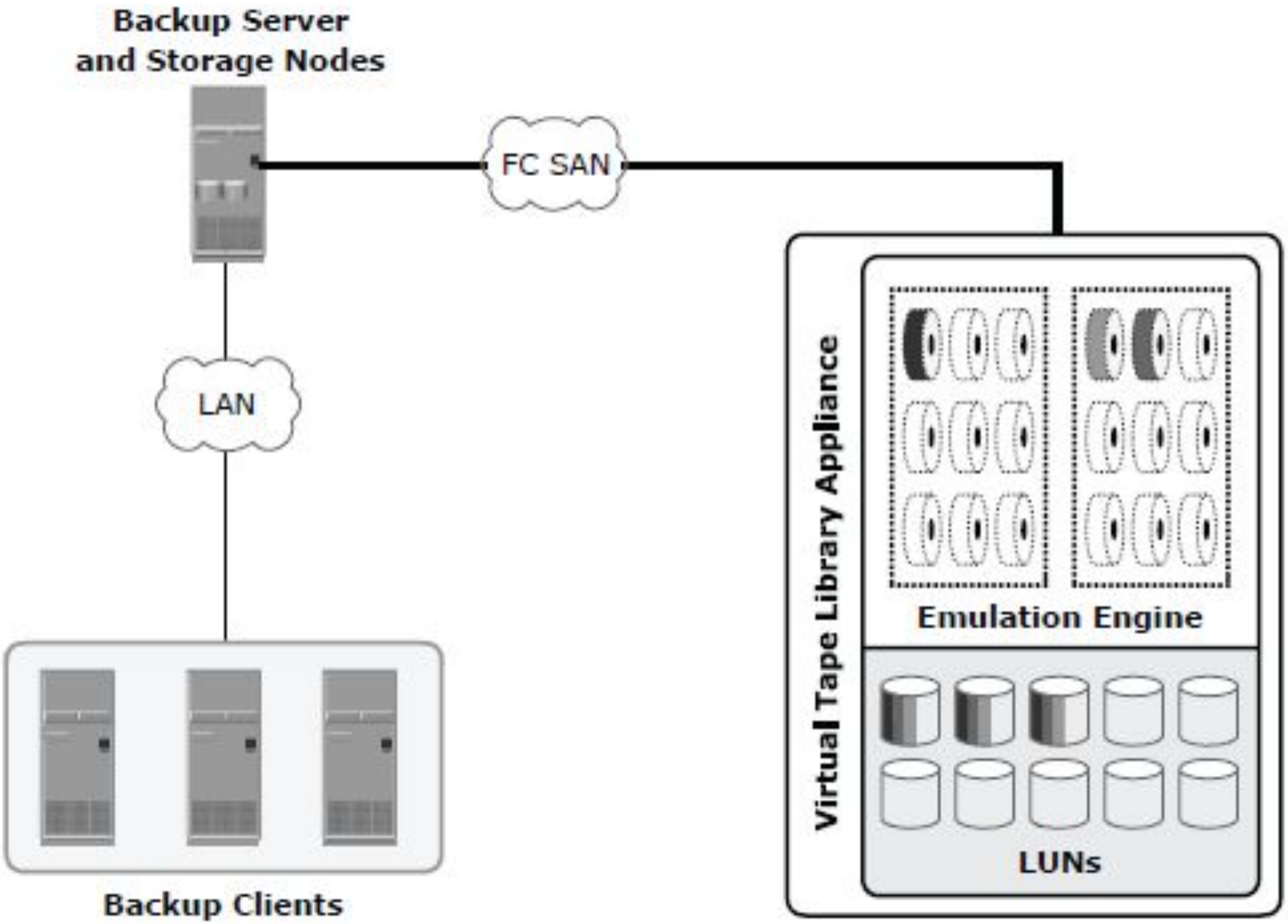

Backup Server and Storage Nodes
FC SAN
LAN
Backup Clients
Virtual Tape Library Appliance
Emulation Engine
LUNs

Figure 12-18: Virtual tape library

Concepts in Practice: EMC NetWorker 12.11

- EMC backup products provide a powerful and effective way to back up and recover data. This ensures higher information protection and enables compliance with regulations and corporate policies.
- NetWorker enables simultaneous-access operations to a volume, for both reads and writes, as opposed to a single operation with tapes. NetWorker works within the existing framework of the hardware, operating system, software, and network communication protocols to provide protection for critical business data by centralizing, automating, and accelerating backup and recovery operations across an enterprise.
- NetWorker provides cold and hot backups, and supports a wide range of applications for hot backups with granular-level recovery.
- NetWorker also provides centralized management of the backup environment through a GUI, customizable reporting, and wizard-driven configuration.
- NetWorker also supports Open Tape Format (OTF), a data format that enables multiplexed, heterogeneous data to reside on the same tape. Using OTF, a NetWorker storage node can be migrated to a host running a different operating system.

NetWorker Backup Operation 12.11.1

- In a NetWorker backup operation the NetWorker client pushes the backup data to the destination storage node.
- The client generates tracking information, including the file and directory names in the backup and the time of the backup, and sends it to the server to facilitate point-in-time recoveries.
- The storage node organizes the client's data and writes it to backup devices.
- NetWorker can initiate backup in two ways: client-initiated and server initiated.

NetWorker Recovery 12.11.2

- NetWorker is flexible in the way recovery operations are performed, and it maintains security to avoid recovery of data by unauthorized users.
- NetWorker detects, and can be configured to automatically resolve, naming conflicts.
- The three types of manual recoveries — brows able, save set, and directed — are all processes initiated from a NetWorker client.

EmailXtender 12.11.3

- EmailXtender is a comprehensive archive application that automatically collects, organizes, retains, and retrieves e-mail messages and attachments.
- It works with all major messaging environments.
- EmailXtender, companies can automatically enforce retention policies by periodically deleting messages from mail servers and archiving them to EmailXtender.
- The EmailXtract feature of EmailXtender enables removing messages from the mail server and replacing them with pointers or shortcuts to copies of the messages archived in EmailXtender.

DiskXtender 12.11.4

- DiskXtender is a robust storage management solution available only for the Microsoft Windows platform. DiskXtender extends the amount of space on the local NTFS (NT file system) volume.
- It does this by migrating files from the local drive to external media and purging files from primary storage. To a client retrieving files from the drive extended by DiskXtender.
- DiskXtender intelligently queues requests for files and accesses secondary media only when necessary.

Avamar 12.11.5

- Avamar is a comprehensive, client-server network backup and restore solution. Avamar differs from traditional backup and restore solutions by identifying and storing only unique sub-file data objects.

- Avamar uses standard IP network technology, so dedicated backup networks are not required. During backup, the Avamar client traverses each directory and examines the local cache to determine which files have not been previously backed up.

- Once an object is backed up on the server, it is never sent for backup again. This drastically reduces network traffic and enhances backup storage efficiency, guaranteeing the most effective de-duplication of the data.

Summary

- Data availability is a critical requirement for information-centric businesses. Backups protect businesses from data loss and also helps to meet regulatory and compliance requirements.
- This chapter detailed backup considerations, methods, technologies, and implementations in a storage networking environment. It also elaborated various backup topologies and architectures.
- Although the selection of a particular backup media is driven by the defined RTO and RPO, disk-based backup has a clear advantage over tape-based backup in terms of performance, availability, faster recovery, and ease of management. These advantages are further supplemented with the use of replication technologies to achieve the highest level of service and availability requirements. Replication technologies are covered in detail in the next two chapters.

Key Concepts:

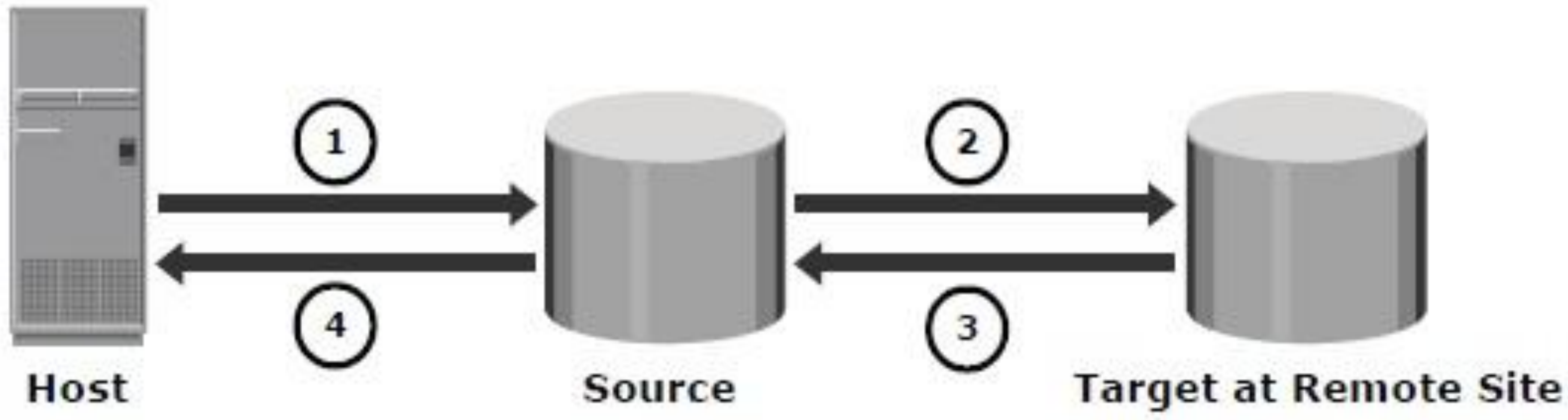
- Synchronous and Asynchronous Replication
- LVM-Based Replication
- Host-Based Log Shipping
- Disk-Buffered Replication
- Three-Site Replication
- Data Consistency

Remote replication is the process of creating replicas of information assets at remote sites (locations). Remote replicas help organizations mitigate the risks associated with regionally driven outages resulting from natural or human-made disasters. Similar to local replicas, they can also be used for other business operations.

This chapter discusses various remote replication technologies, along with the key steps to plan and design appropriate remote replication solutions. In addition, this chapter describes network requirements and management considerations in the remote replication process.

14.1 Modes of Remote Replication

The two basic modes of remote replication are synchronous and asynchronous. In synchronous remote replication, writes must be committed to the source and the target, prior to acknowledging “write complete” to the host (see Figure 14-1). 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. Further writes are transmitted to the remote site exactly in the order in which they are received at the source. Hence, write ordering is maintained. 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.



1 Host writes data to source

2 Data from source is replicated to target at remote site

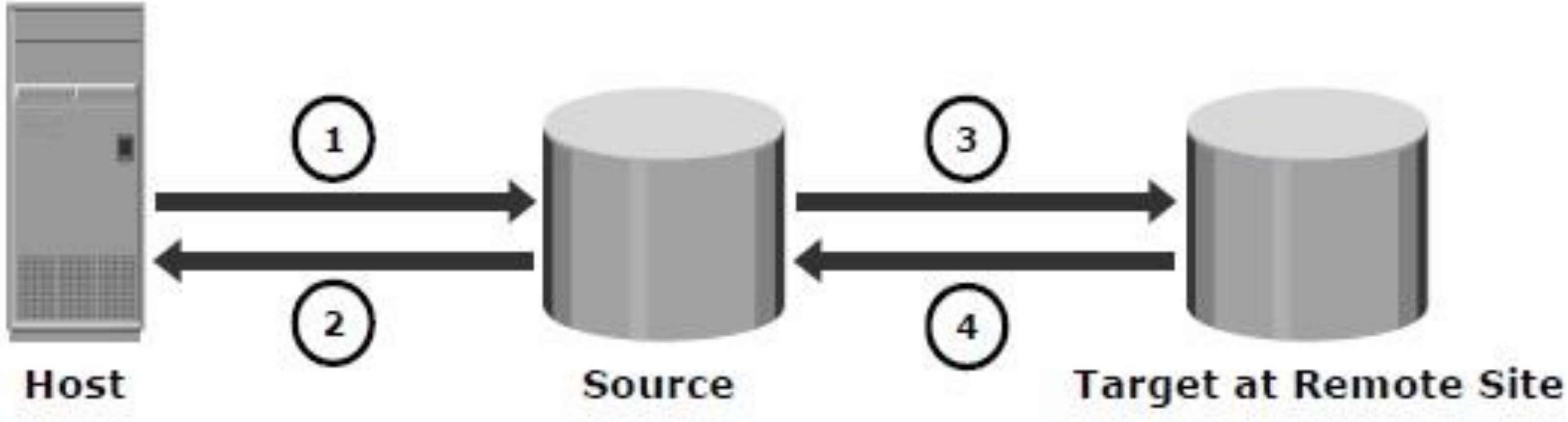
3 Target acknowledges back to source

4 Source acknowledges write complete to host

Figure 14-1: Synchronous replication

14.2 Remote Replication Technologies

Remote replication of data can be handled by the hosts or by the storage arrays. Other options include specialized appliances to replicate data over the LAN or the SAN, as well as replication between storage arrays over the SAN.



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

Figure 14-2: Asynchronous replication

Host-Based Remote Replication .14.2.1

Host-based remote replication uses one or more components of the host to perform and manage the replication operation. There are two basic approaches to host-based remote replication: LVM-based replication and database replication via log shipping.

- **LVM-Based Remote Replication**
- **Host-Based Log Shipping**

LVM-Based Remote Replication

LVM-based 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. Prior to the start of replication, identical volume groups, logical volumes, and file systems are created at the source and target sites. Initial synchronization of data between the source and the replica can be performed in a number of ways. One method is to backup the source data to tape and restore the data to the remote replica. Alternatively, it can be performed by replicating over the Until completion of initial synchronization, production work on the source volumes is typically halted. After initial synchronization, production work can be started on the source volumes and replication of data can be performed over an existing standard IP network (see Figure 14-3).

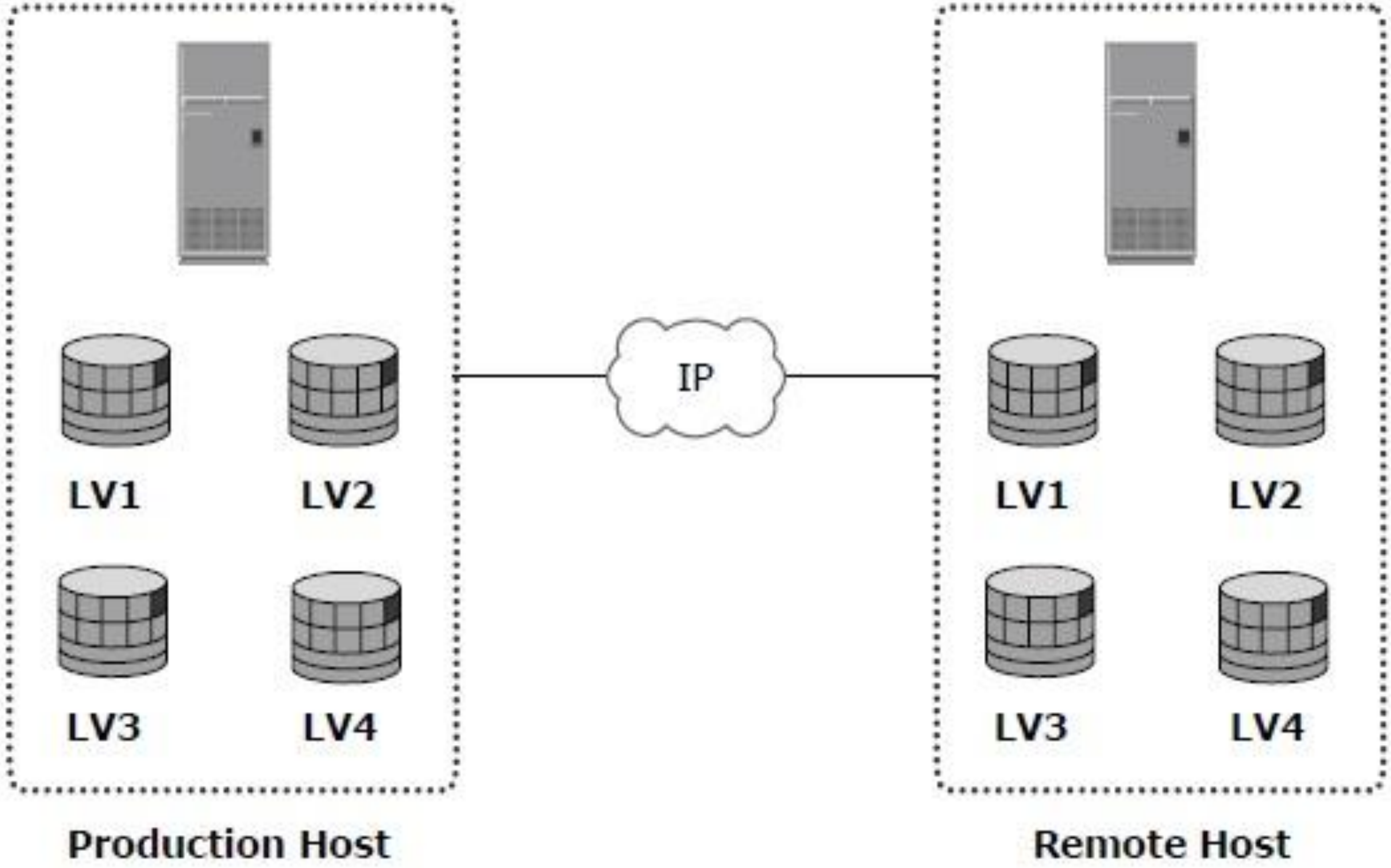


Figure 14-3: LVM-based remote replication

In asynchronous mode, writes are queued in a log file at the source and sent to the remote host in the order in which they were received. The size of the log file determines the RPO at the remote site. In the event of a network failure, writes continue to accumulate in the log file. If the log file fills up before the failure is resolved, then a full resynchronization is required upon network availability. In the event of a failure at the source site, applications can be restarted on the remote host, using the data on the remote replicas. LVM-based remote replication eliminates the need for a dedicated SAN infrastructure.

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 remote host (see Figure 14-4). The remote host receives the logs and applies them to the remote database.

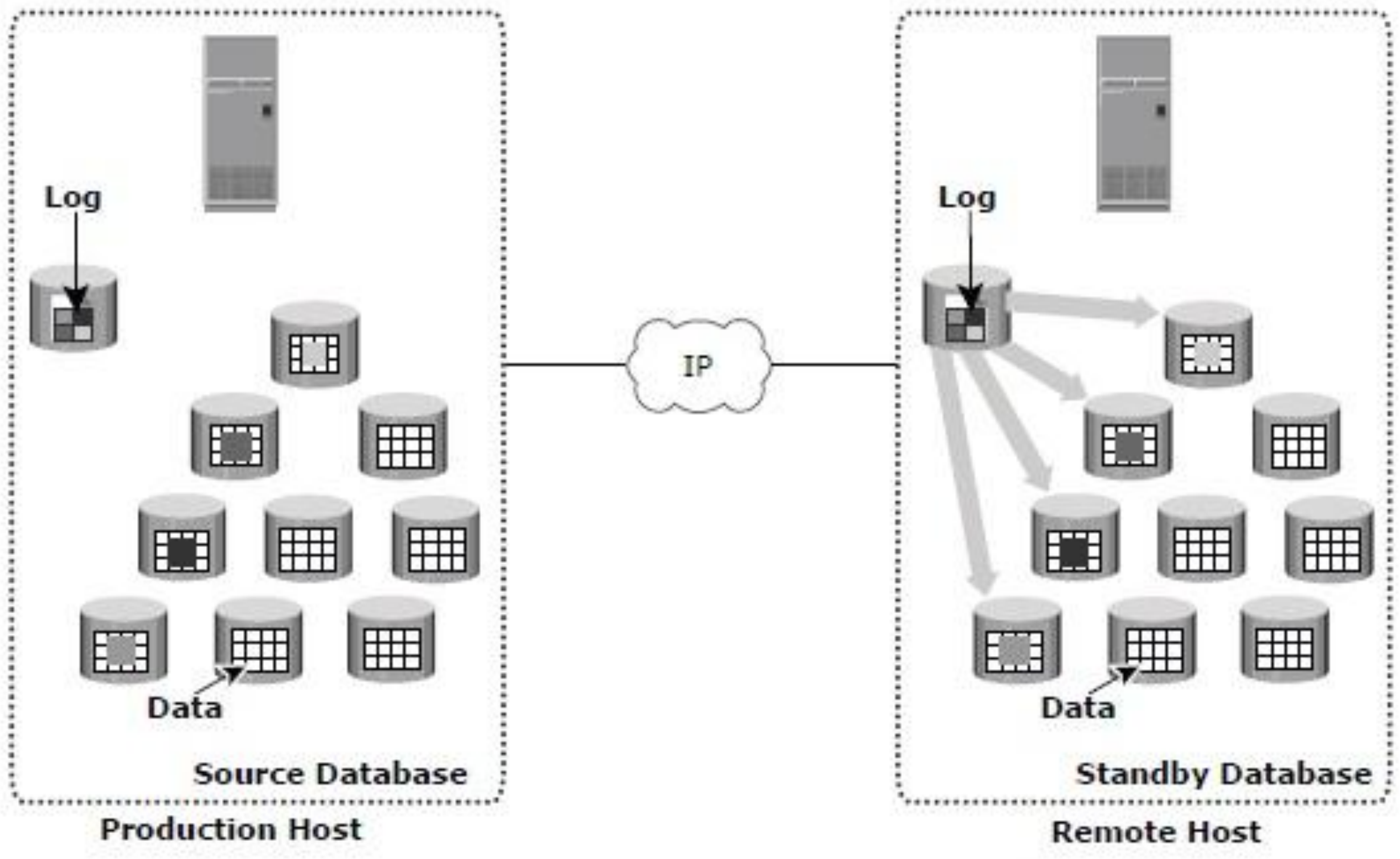


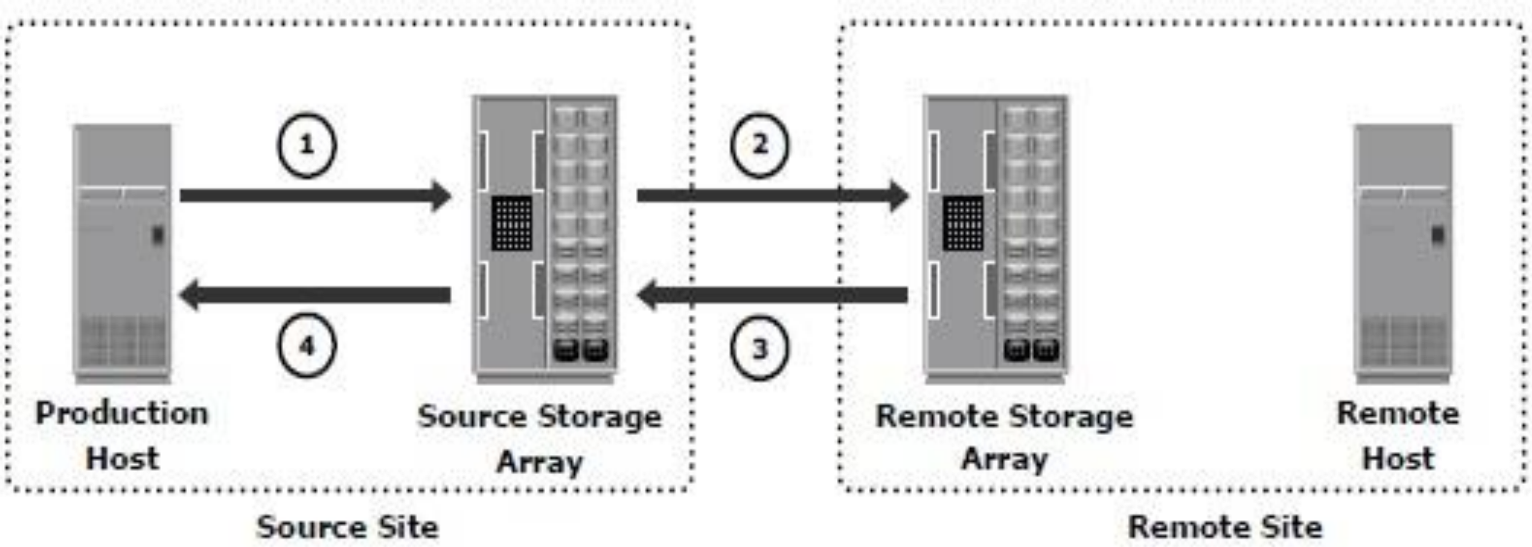
Figure 14-4: Host-based log shipping

Because the source host does not transmit every update and buffer them, this alleviates the burden on the source host CPU. Similar to LVM-based remote replication, the existing standard IP network can be used for replicating log files. Host-based log shipping does not scale well, particularly in the case of applications using federated databases.

Storage Array-Based Remote 14.2.2 Replication

In storage array-based remote replication, the array operating environment and resources perform and manage data replication. This relieves the burden on the host CPUs, which can be better utilized for running an application. A source and its replica device reside on different storage arrays. In other implementations, the storage controller is used for both the host and replication workload. Data can be transmitted from the source storage array to the target storage array over a shared or a dedicated network.

Synchronous Replication Mode



1 Write from the source host is recieved by the source storage array

2 Write is then transmitted to the remote storage array

3 Acknowledgment is sent to the source storage array by the remote storage array

4 Source storage array signals write-completion to the source host

Figure 14-5: Array-based synchronous remote replication

For synchronous remote replication, network bandwidth equal to or greater than the maximum write workload between the two sites should be provided at all times. Figure 14-6 illustrates the write workload (expressed in MB/s) over time. The “Max” line indicated in Figure 14-6 represents the required bandwidth that must be provisioned for synchronous replication. Bandwidths lower than the maximum write workload results in an unacceptable increase in application response time.

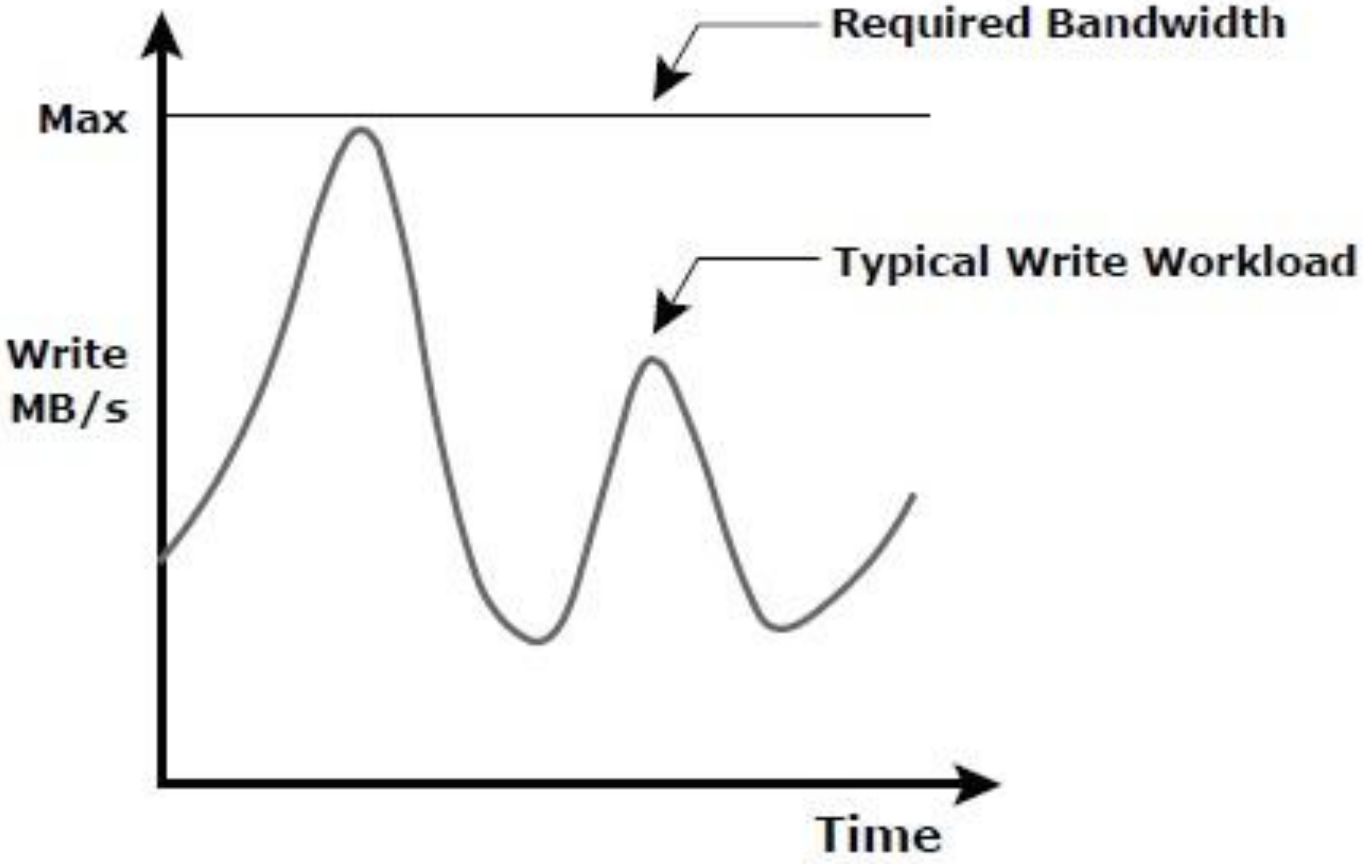
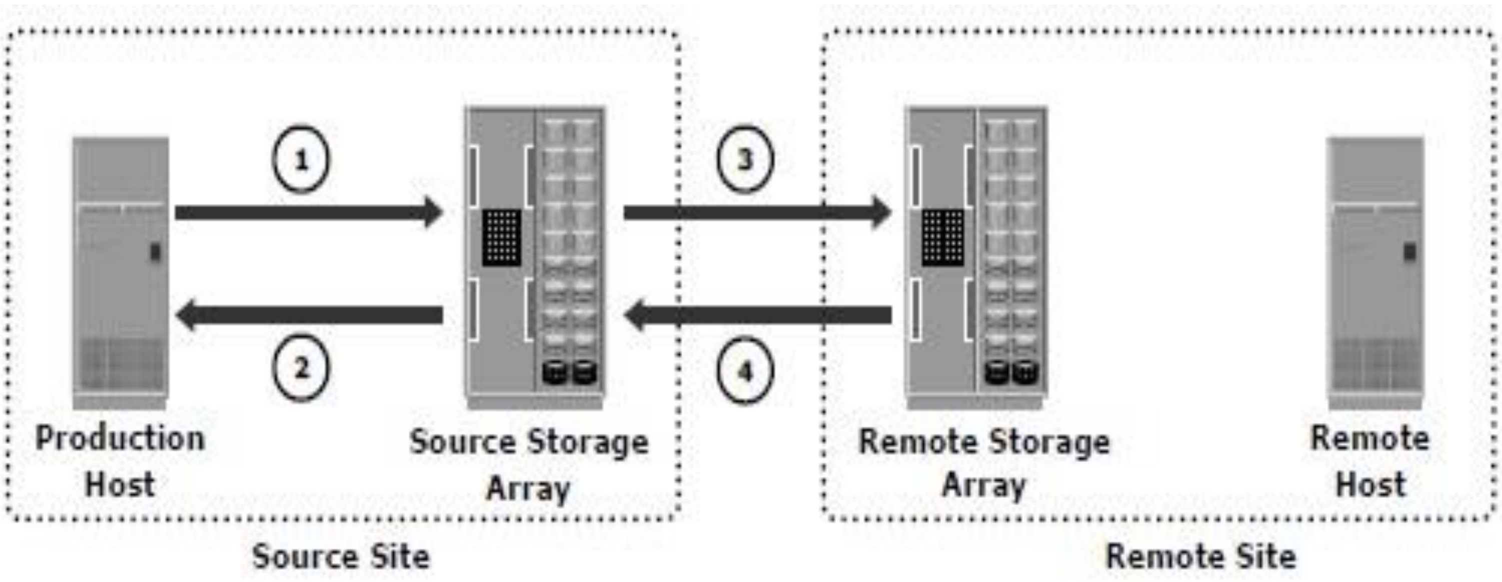


Figure 14-6: Network bandwidth requirement for synchronous replication

Asynchronous Replication Mode



1 Source host writes to the source storage array

2 Source array immediately acknowledges the source host

3 These writes are then transmitted to the target array

4 After the writes are recieved by the target array, it sends an acknowledge to source array

Figure 14-7: Array-based asynchronous remote replication

Some implementations of asynchronous remote replication maintain write ordering. A time stamp and sequence number are attached to each write when it is received by the source. Writes are then transmitted to the remote array, where they are committed to the remote replica in the exact order in which they were buffered at the source. This implicitly guarantees consistency of data on the remote replicas. Other implementations ensure consistency by leveraging the dependent write principle inherent to most DBMSs. The writes are buffered for a predefined period of time. At the end of this duration, the buffer is closed, and a new buffer is opened for subsequent writes. All writes in the closed buffer are transmitted together and committed to the remote replica.

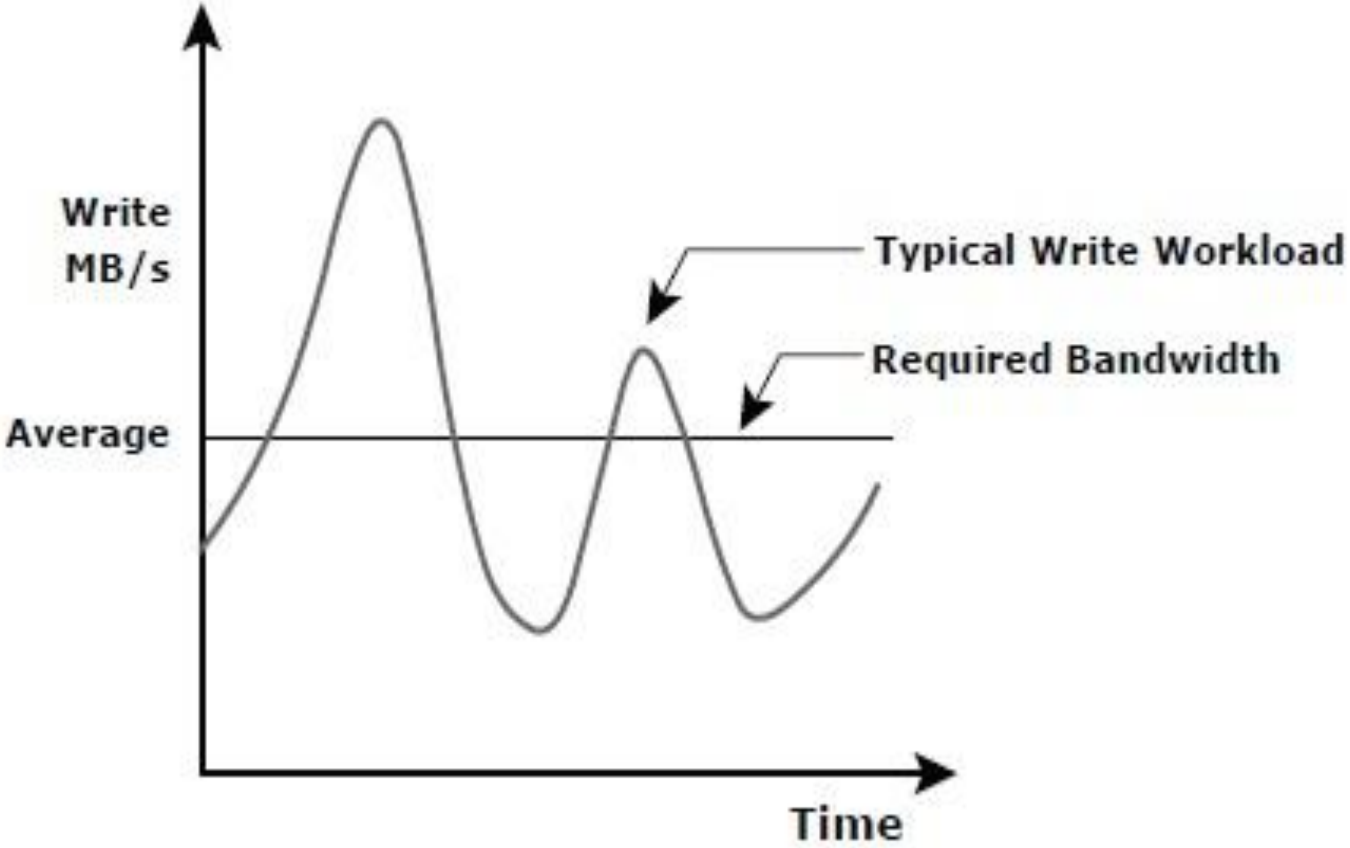
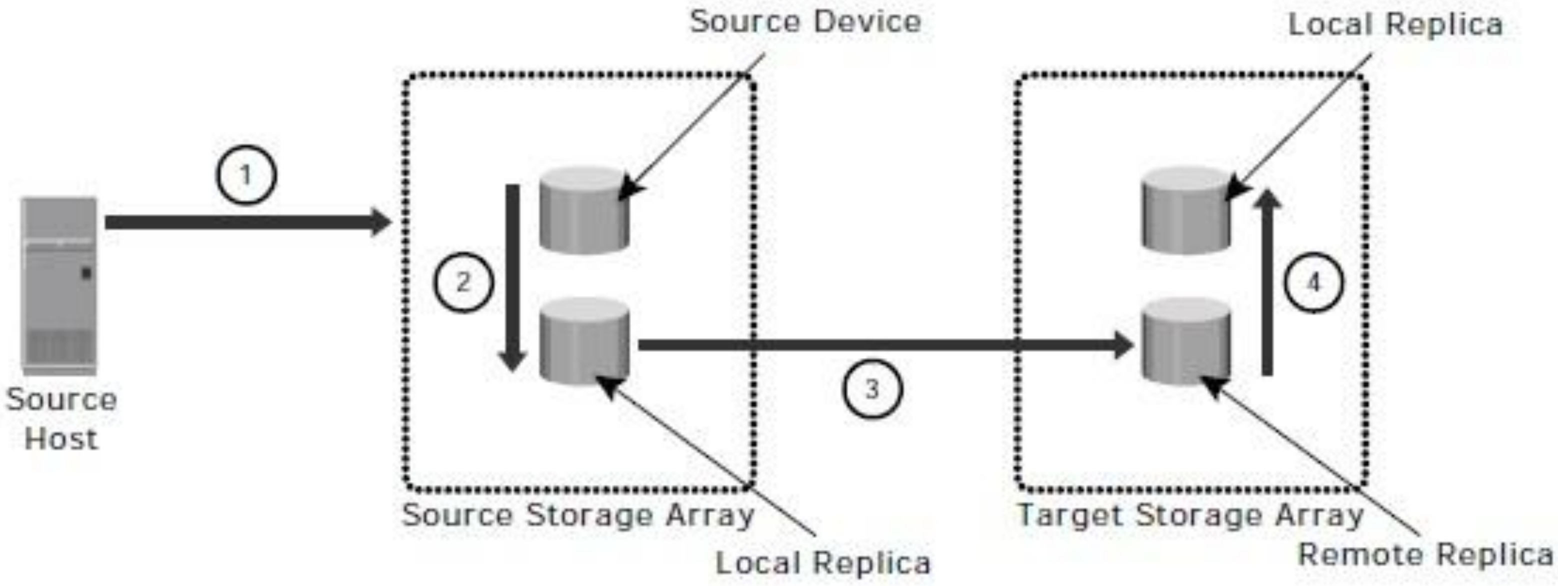


Figure 14-8: Network bandwidth requirement for asynchronous replication

Disk-Buffered Replication Mode



1 Source host writes data to source device

2 A consistent PIT local replica of the source device is created

3 Data from local replica in the source array is transmitted to its remote replica in the target array

4 A local PIT replica of the remote device on the target array is created

Figure 14-9: Disk-buffered remote replication

Three-Site Replication

- In synchronous and asynchronous replication, under normal conditions the workload is running at the source site.
- In synchronous replication, source and target sites are usually within 200 KM (125 miles) of each other.
- A regional disaster will not affect the target site in asynchronous replication, as the sites are typically several hundred or several thousand kilometers apart.
- Three-site replication is used to mitigate the risks identified in two-site replication. In a three-site replication, data from the source site is replicated to two remote data centers.

Three-Site Replication—Cascade/Multi-hop

In the **cascade/multi-hop** form of replication, data flows from the source to the intermediate storage array, known as a bunker, in the first hop and then from a bunker to a storage array at a remote site in the second hop. Replication between the source and the bunker occurs synchronously, but replication between the bunker and the remote site can be achieved in two ways: disk-buffered mode or asynchronous mode.

Synchronous + Asynchronous

This method employs a combination of **synchronous** and **asynchronous** remote replication technologies. Synchronous replication occurs between the source and the bunker.

Synchronous replication occurs between the source and the bunker. **Asynchronous** replication occurs between the bunker and the remote site.

The remote replica in the bunker acts as the source for the asynchronous replication to create a remote replica at the remote site. Figure 14-10(a) illustrates the **synchronous + Asynchronous** method.

Synchronous + Disk Buffered

This method employs a combination of local and remote replication technologies. Synchronous replication occurs between the source and the bunker: A consistent PIT local replica is created at the bunker. Data is transmitted from the local replica at the bunker to the remote replica at the remote site. Optionally, a local replica can be created at the remote site after data is received from the bunker. Figure 14-10(b) illustrates the **synchronous + disk buffered** method.

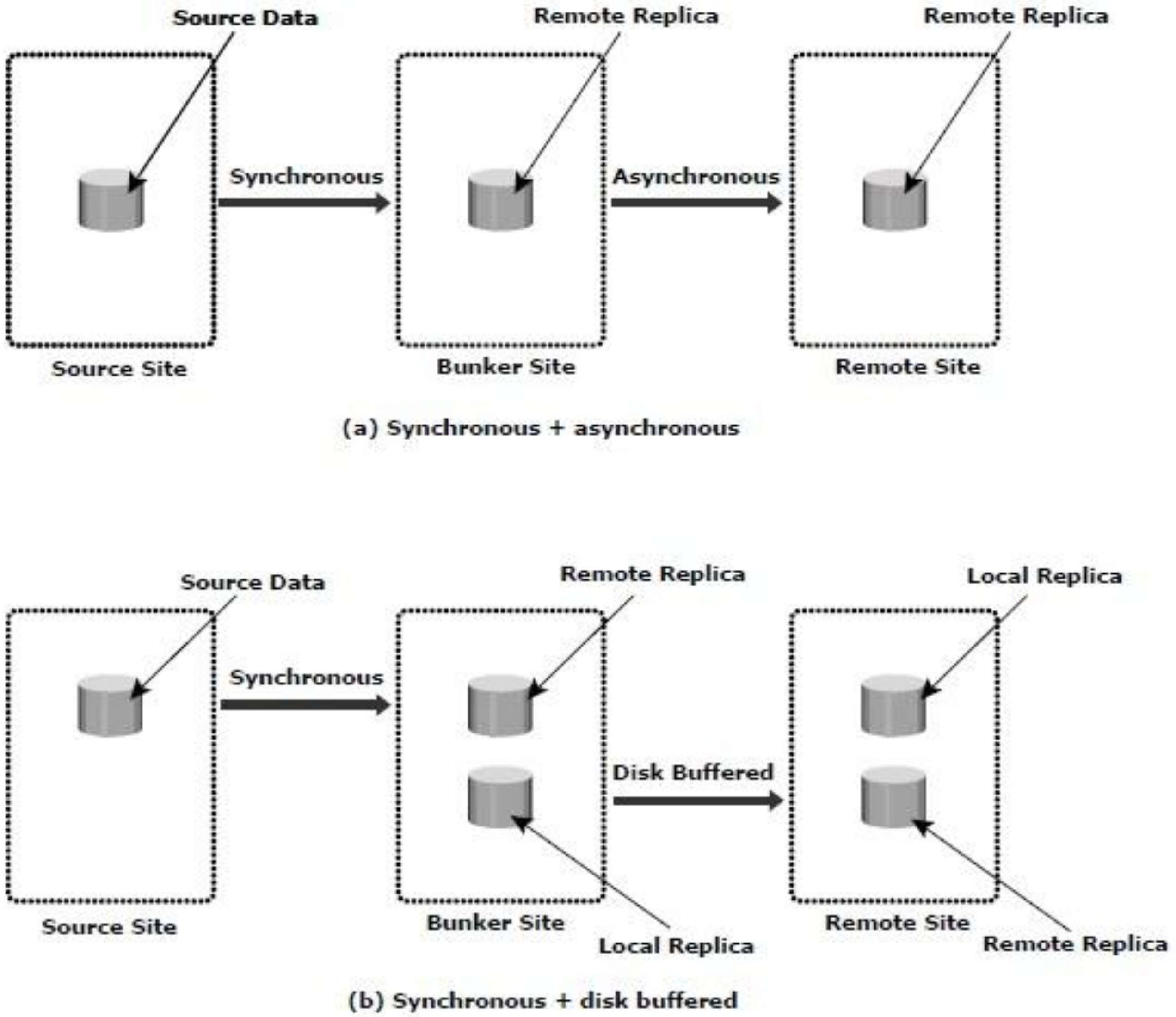


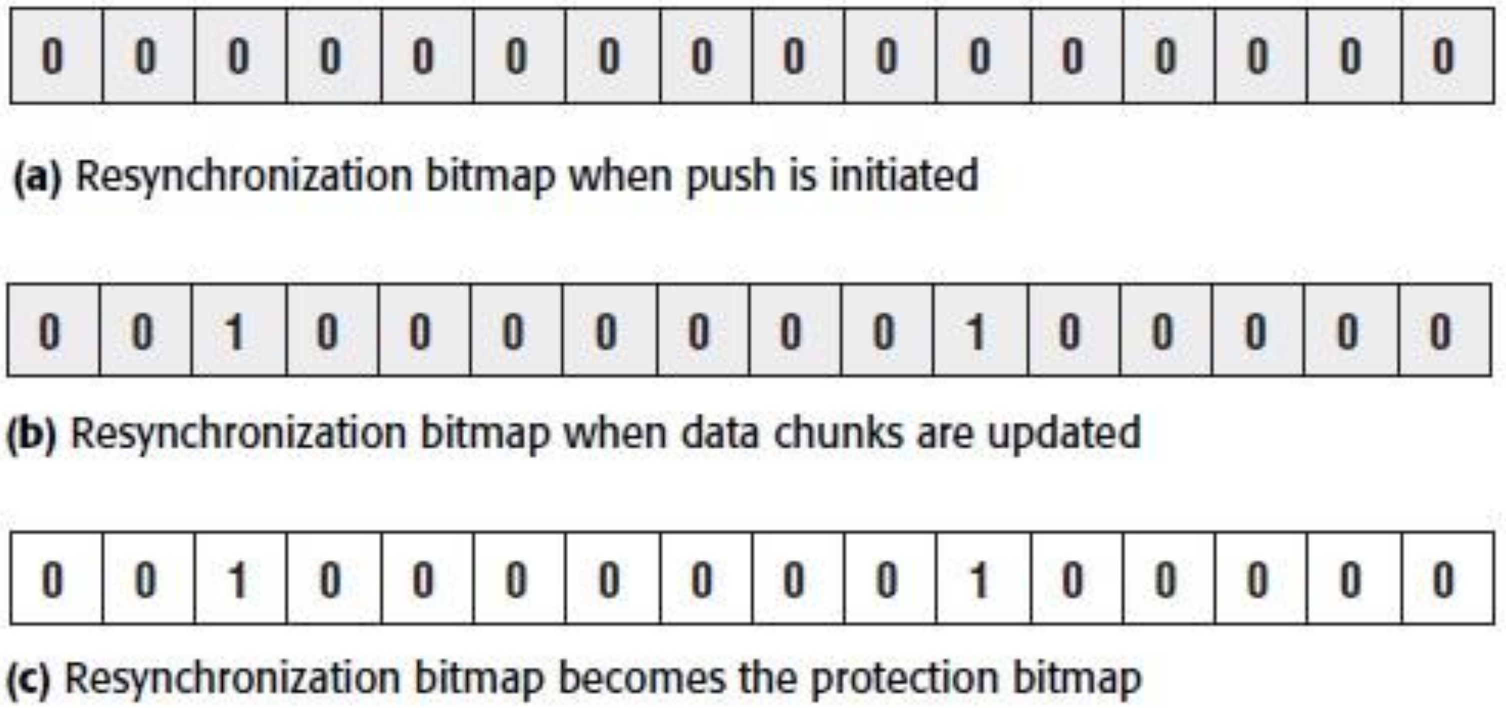
Figure 14-10: Three-site replication

Three-Site Replication—Triangle/Multi-target

In the three-site triangle/multi-target replication, data at the source storage array is concurrently replicated to two different arrays. The source-to bunker site (**target 1**) replication is synchronous, with a near-zero RPO. The source-to remote site (**target 2**) replication is asynchronous, with an RPO of minutes. The distance between the source and the remote site could be thousands of miles. This configuration does not depend on the bunker site for updating data on the remote site, because data is asynchronously copied to the remote site directly from the source.

SAN-Based Remote Replication 14.2.3

- **SAN-based** remote replication enables the replication of data between heterogeneous storage arrays. Data is moved from one array to the other over the SAN/ WAN.
- **SAN-based** remote replication is a point-in-time replication technology. Uses of SAN-based remote replication include data mobility, remote vaulting, and data migration.
- **SAN-based** replication uses two types of operations: push and pull. These terms are defined from the perspective of the control array.
- In **SAN-based** replication, the control array can keep track of changes made to the control devices after the replication session is activated. This is allowed in the incremental push operation only.



0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(a) Resynchronization bitmap when push is initiated

0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0

(b) Resynchronization bitmap when data chunks are updated

0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0

(c) Resynchronization bitmap becomes the protection bitmap

Figure 14-11: Bitmap status in SAN-based replication

Network Infrastructure 14.3

For remote replication over extended distances, optical network technologies such as dense wavelength division multiplexing (**DWDM**), coarse wavelength division multiplexing (**CWDM**), and synchronous optical network (**SONET**) are deployed.

DWDM 14.3.1

DWDM is an optical technology by which data from different channels are carried at different wavelengths over a fiber-optic link. It is a fiber-optic transmission technique that uses light waves to transmit data parallel by bit or serial by character. It integrates multiple light waves with different wavelengths in a group and directs them through a single optical fiber. Using DWDM, different data formats at different data rates can be transmitted together. Specifically, IP ESCON, FC, SONET and ATM data can all travel at the same time within the optical fiber (see Figure 14-12).

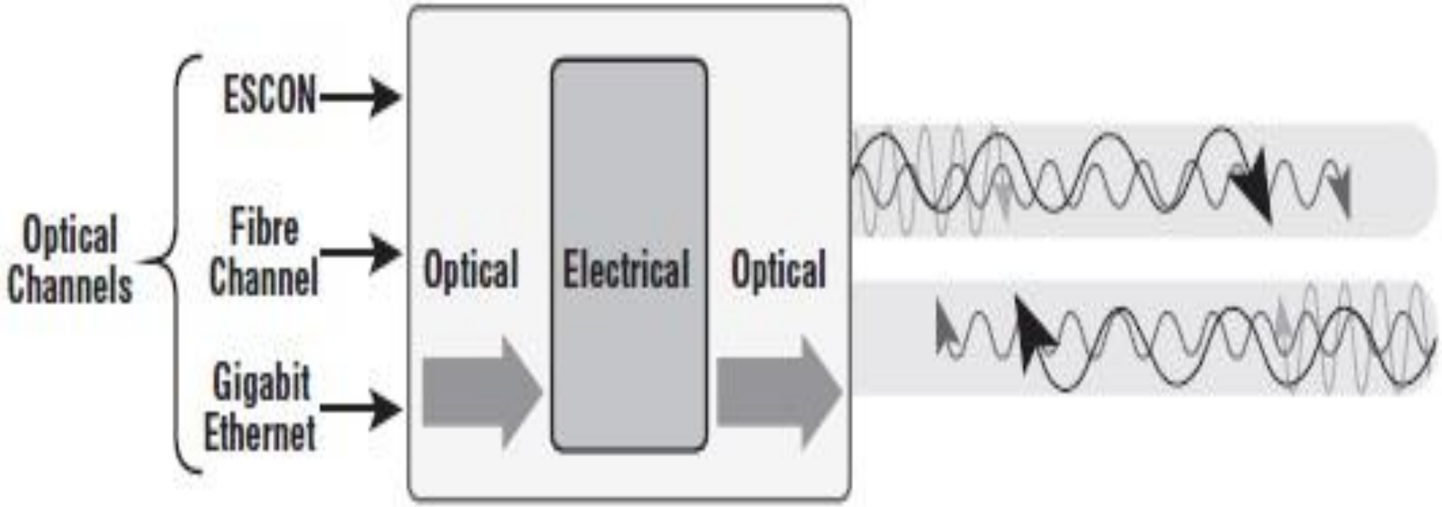


Figure 14-12: Dense wavelength division multiplexing (DWDM)

CWDM, like DWDM, uses multiplexing and demultiplexing on different channels by assigning varied wavelengths to each channel. Compared to DWDM, CWDM is used to consolidate environments containing a low number of channels at a reduced cost.

SONET 14.3.2

SONET (synchronous optical network) is a network technology that involves transferring a large payload through an optical fiber over long distances. SONET multiplexes data streams of different speeds into a frame and sends them across the network. The European variation of SONET is called synchronous digital hierarchy (SDH). Figure 14-13 shows the multiplexing of data streams of different speeds in SONET and SDH technologies. SONET/SDH uses generic framing procedure (GFP) and supports the transport of both packet-oriented (Ethernet, IP) and character-oriented (FC) data.

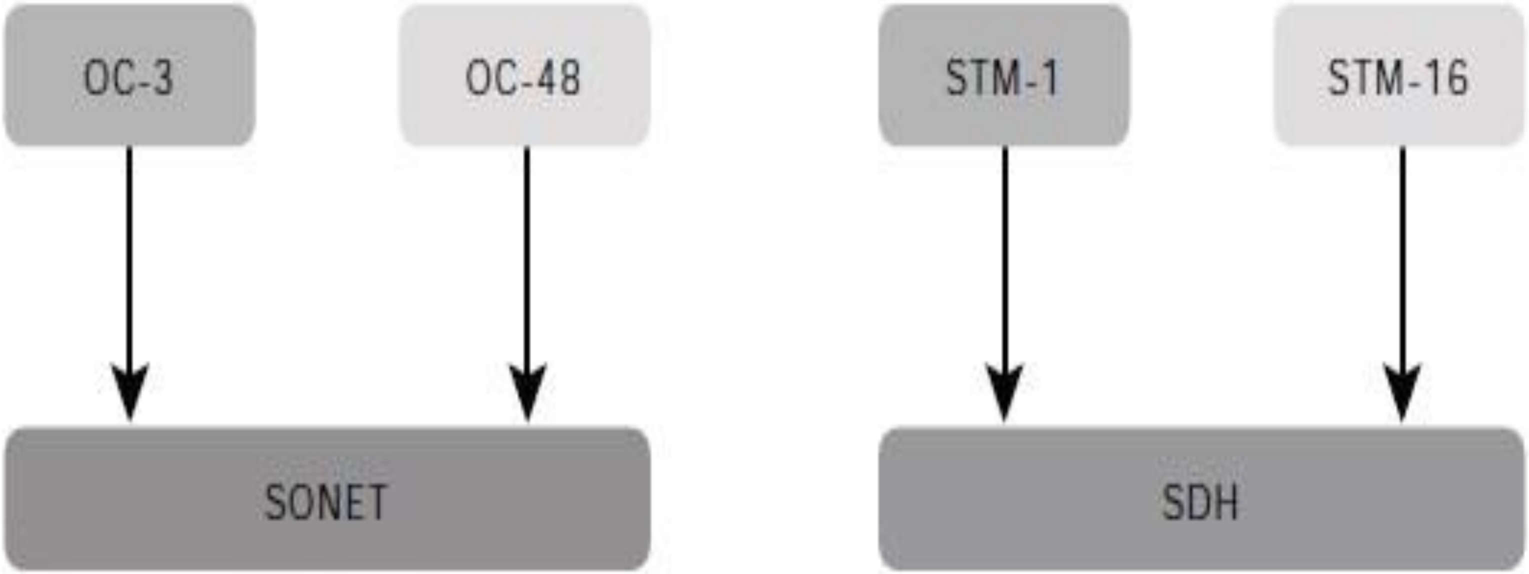


Figure 14-13: Data stream multiplexing in SONET and SDH

Concepts in Practice: EMC SRDF, EMC 14.4 SAN Copy, and EMC Mirror View

This section discusses three EMC products that use remote replication technology. EMC Symmetric Remote Data Facility (SRDF) and EMC Mirror View are storage array–based remote application software's supported by EMC Symmetric and Clarion respectively. EMC SAN Copy is SAN-based remote replication software deployed in an EMC Clarion storage array.

SRDF Family 14.4.1

SRDF offers a family of technology solutions to implement storage array-based remote replication technologies. The three Symmetric solutions are:

- **SRDF/Synchronous (SRDF/S)**
- **SRDF/Asynchronous (SRDF/A)**
- **SRDF/Automated Replication (SRDF/AR)**

Disaster Recovery with SRDF 14.4.2

The source arrays have SRDF R1 devices (source devices), and the target arrays have SRDF R2 devices (replica devices). Data written to R1 devices is replicated to R2 devices, either synchronously or asynchronously. SRDF R1 and R2 devices can have any local RAID protection, such as RAID 1 or RAID 5. SRDF R2 devices are in a read-only (R/O) state when remote replication is in effect. Hence, under normal operating conditions, changes cannot be made directly to the R2 devices. The R2 devices can only receive data from their corresponding R1 devices on the source storage array.

Failover

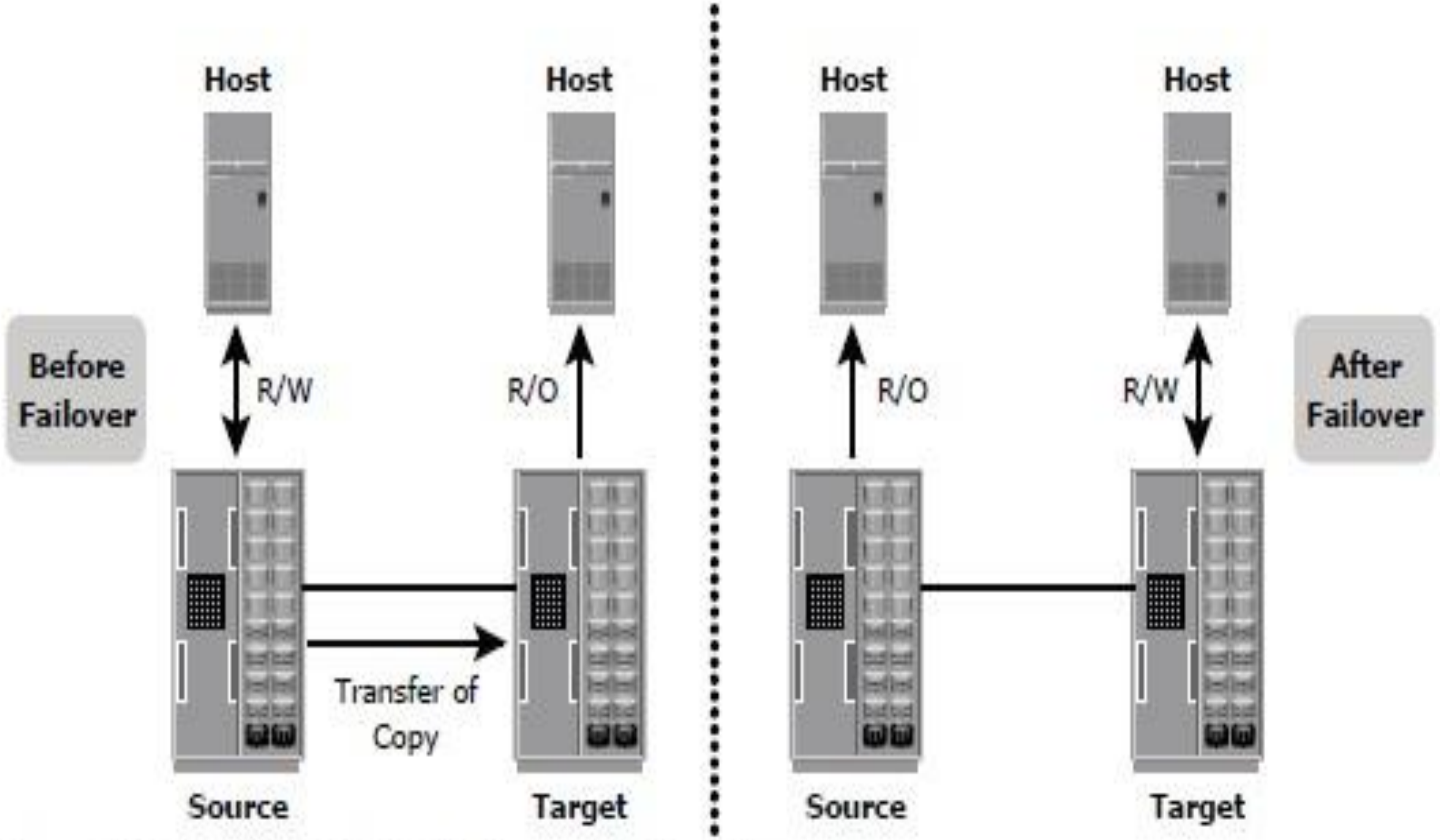


Figure 14-14: EMC SRDF - before and after failover

Failback

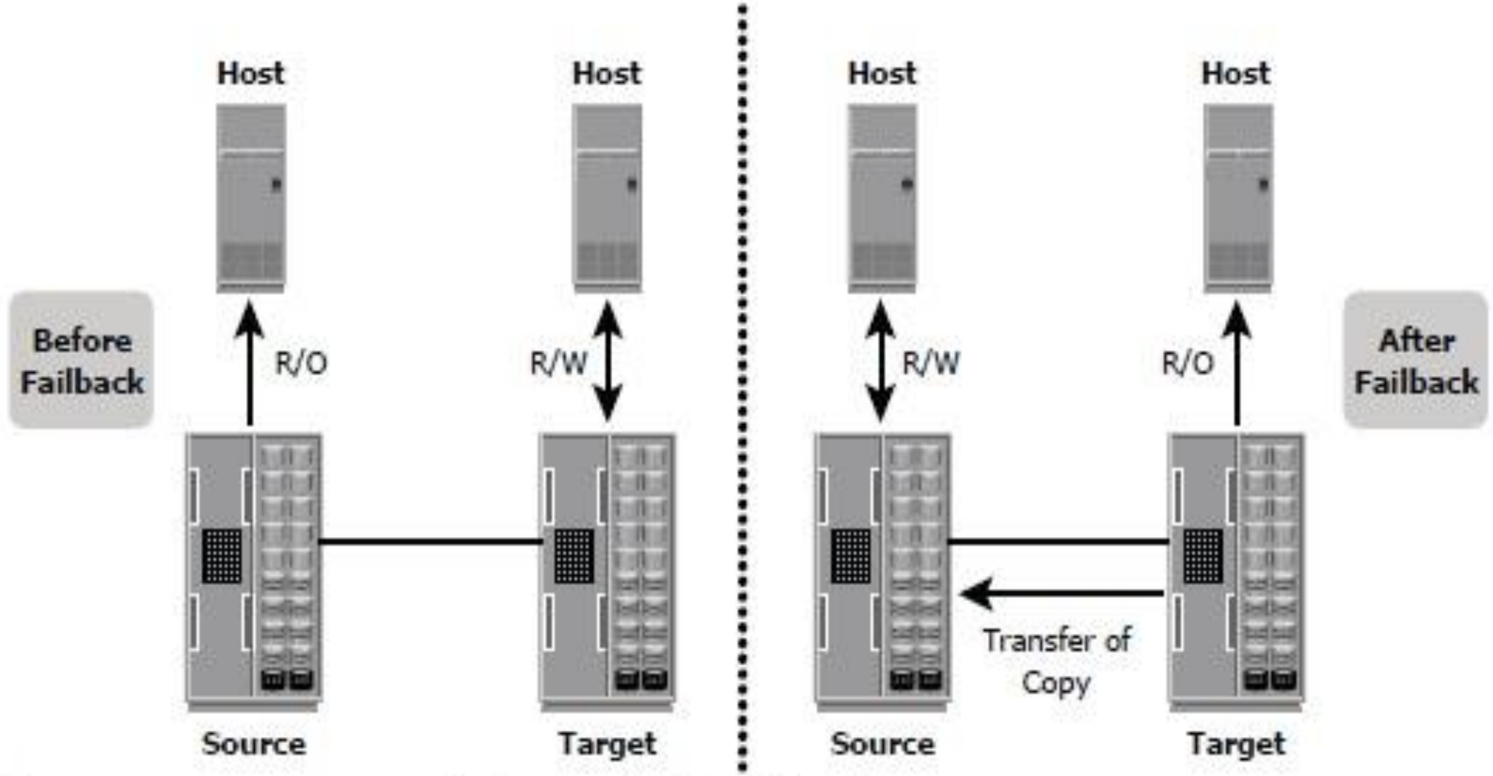


Figure 14-15: EMC SRDF - before and after failback

SRDF Operations for Concurrent 14.4.3 Access

SRDF provides split operations to enable concurrent access to both source and target devices. The establish and restore operations are used to return the source-target pairs to the normal SRDF state. In split operation, when R2 is split from R1, BC operations can be performed on R2. The split operation enables concurrent access to both the source and the target devices. In this operation, target devices are made R/W, and the SRDF replication between the source and the target is suspended, as shown in Figure 14-16.

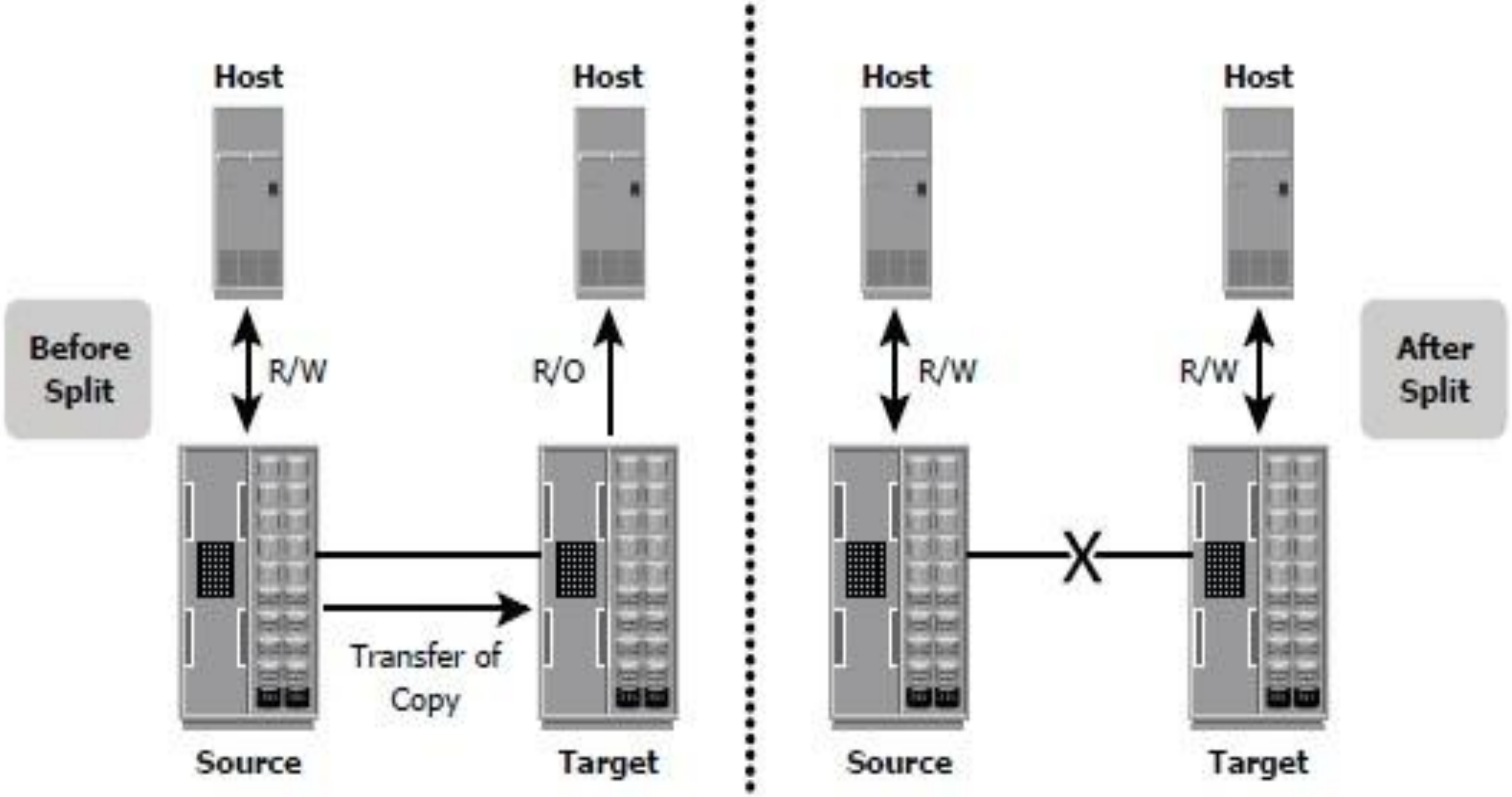


Figure 14-16: Concurrent access with EMC SRDF

During concurrent operations while in a SRDF split state, changes could occur on both the source and the target devices. Normal SRDF replication can be resumed by performing an establish or a restore operation. With either establish or restore, the status of the target device becomes R/O (see Figure 14-17).

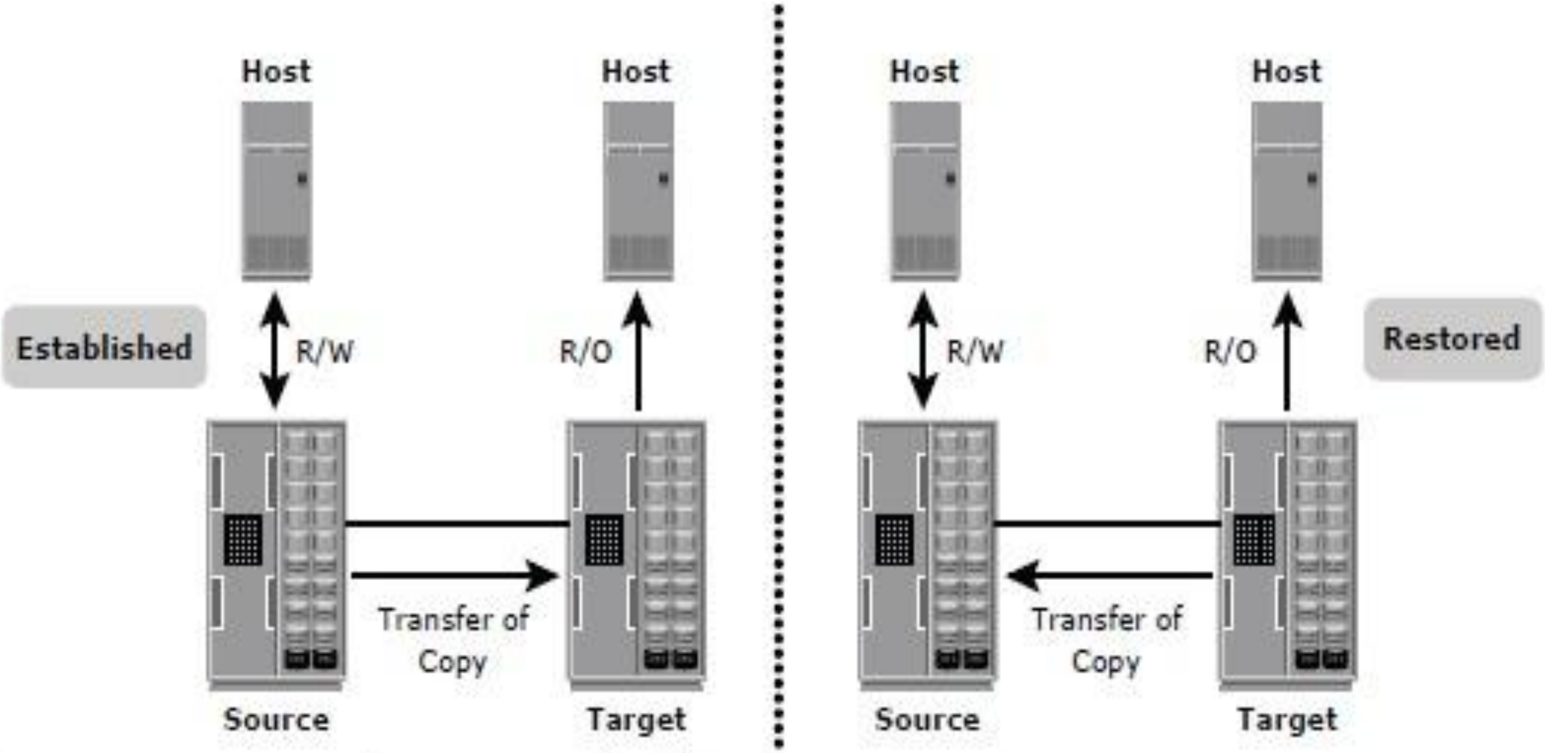


Figure 14-17: Restarting SRDF replication after concurrent access

EMC SAN Copy 14.4.4

SAN Copy is CLARiiON software that performs SAN-based remote replication between CLARiiON and Symmetric or other vendor storage arrays. It enables simultaneous creation of one or more copies of source devices to target devices through a SAN. Source and target devices could either be on a single array or on multiple arrays. SAN Copy software on the CLARiiON (designated as the control storage array) controls the entire replication process.

- **Automatic check pointing in the event of a link failure**
- **Transfer rate throttle**
- **Incremental SAN Copy**

EMC Mirror View 14.4.5

- **EMC MirrorView** is a CLARiiON-based software that enables storage array– based remote replication over FC SAN, IP extended SAN, and TCP/IP networks. MirrorView family consists of Mirror View/Synchronous (MirrorView/S), and MirrorView/Asynchronous (MirrorView/A). MirrorView software must be installed at both source and target CLARiiON in order to perform remote replication.
- **MirrorView** supports both synchronous and asynchronous replication of data on the same CLARiiON. It also supports consistency groups for maintaining data consistency across write-order dependent LUNs.

Mirror View Operations

- Initial Synchronization is a replication process that is used for new mirrors (target) to create an initial copy of the primary/primary image (LUN on source CLARiiON containing production data). During the initial synchronization process, the primary images remain online whereas the secondary/secondary image (LUN that contains a mirror of the primary image) is inaccessible.

- A fracture operation stops MirrorView replication. An administrator can initiate fracture to suspend the replication. MirrorView software can automatically fracture when it senses a connectivity failure between the primary and secondary LUNs.

- **MirrorView/S** invokes a fracture log when the secondary image is fractured. The fracture log is a bitmap held in the memory of the storage processor that owns the primary LUN.

- **MirrorView/A** does not use fracture and write intent logs, but it tracks locations (using Snap View technology) at the primary LUNs where updates occur. MirrorView/A utilizes the delta set mechanism to periodically transfer data to the secondary LUNs. MirrorView uses two bitmaps on the primary LUNs.

- A secondary image is promoted to the role of primary, when it is necessary to run production applications at the disaster recovery site.

Summary

This chapter detailed remote replication. As a primary utility, remote replication provides disaster recovery and disaster restart solutions. It enables business operations to be rapidly restarted at a remote site following an outage, with acceptable data loss. Remote replication enables BC operations from a target site. The replica of source data at the target can be used for backup and testing. This replica can also be used for data repurposing, such as report generation, data warehousing, and decision support. The segregation of business operations between the source and target protects the source from becoming a performance bottleneck, ensuring improved production performance at the source.

Remote replication may also be used for data center migrations, providing the least disturbance to production operations because the applications accessing the source data are not affected. This chapter also described different types of remote replication solutions.

The distance between the primary site and the remote site is a prime consideration when deciding which remote replication technology solution to deploy. Asynchronous replication may adequately meet the RPO and RTO needs, while permitting greater distances between the sites.

Storage management solutions provide the capability to not only automate business continuity solutions, but also enable centralized management of the overall storage infrastructure.

Organizations must ensure security of the information assets. The next chapter details storage security and management.

Key Concepts:

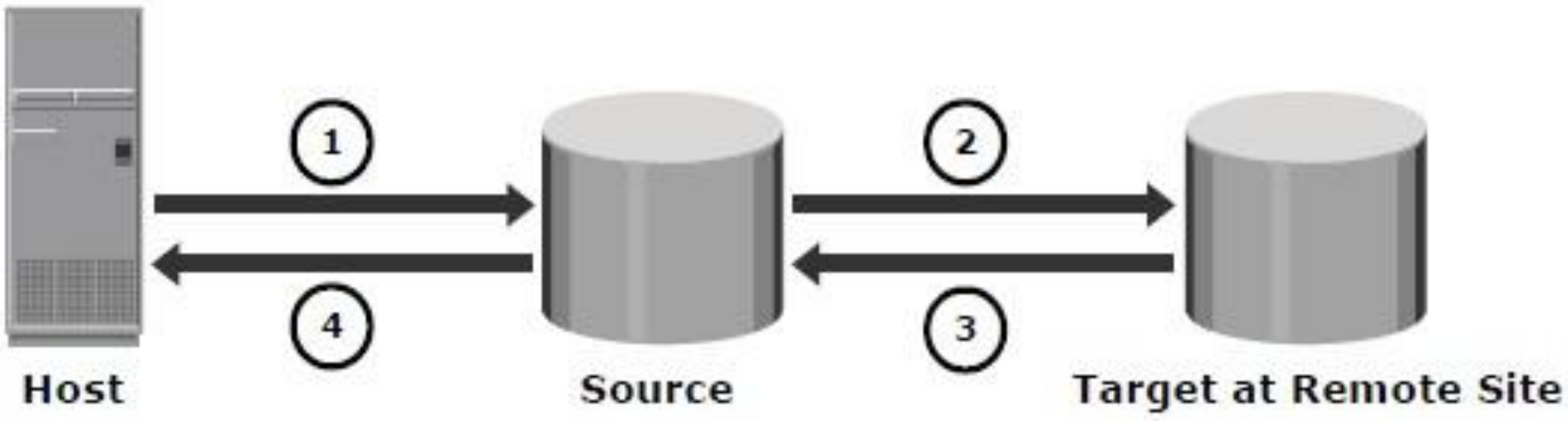
- Synchronous and Asynchronous Replication
- LVM-Based Replication
- Host-Based Log Shipping
- Disk-Buffered Replication
- Three-Site Replication
- Data Consistency

Remote replication is the process of creating replicas of information assets at remote sites (locations). Remote replicas help organizations mitigate the risks associated with regionally driven outages resulting from natural or human-made disasters. Similar to local replicas, they can also be used for other business operations.

This chapter discusses various remote replication technologies, along with the key steps to plan and design appropriate remote replication solutions. In addition, this chapter describes network requirements and management considerations in the remote replication process.

14.1 Modes of Remote Replication

The two basic modes of remote replication are synchronous and asynchronous. In synchronous remote replication, writes must be committed to the source and the target, prior to acknowledging “write complete” to the host (see Figure 14-1). 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. Further writes are transmitted to the remote site exactly in the order in which they are received at the source. Hence, write ordering is maintained. 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.

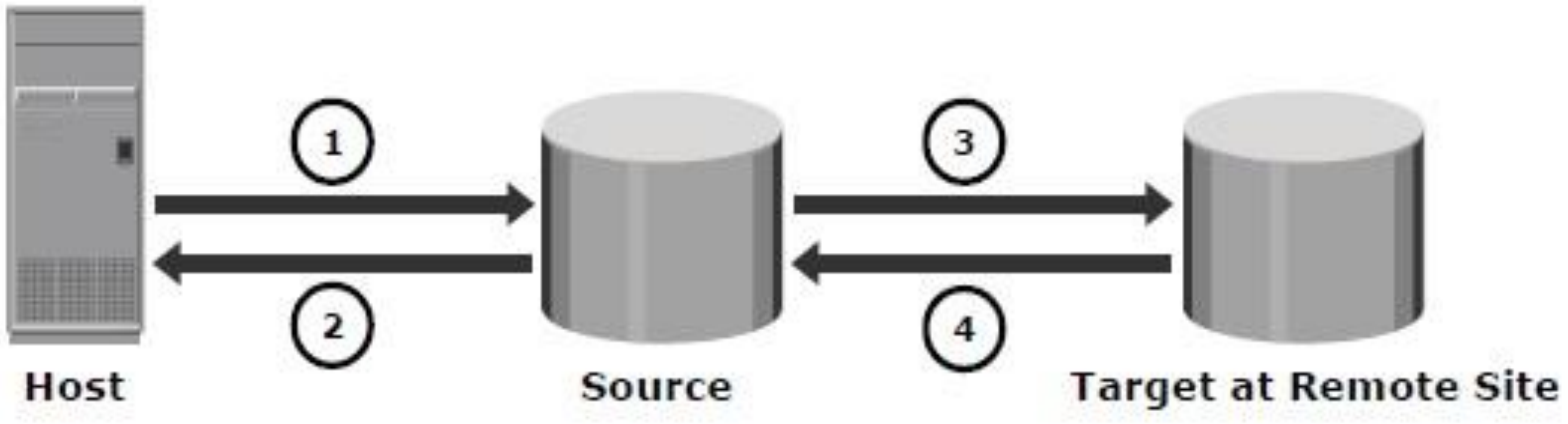


1. Host writes data to source
2. Data from source is replicated to target at remote site
3. Target acknowledges back to source
4. Source acknowledges write complete to host

Figure 14-1: Synchronous replication

14.2 Remote Replication Technologies

Remote replication of data can be handled by the hosts or by the storage arrays. Other options include specialized appliances to replicate data over the LAN or the SAN, as well as replication between storage arrays over the SAN.



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

Figure 14-2: Asynchronous replication

Host-Based Remote Replication .14.2.1

Host-based remote replication uses one or more components of the host to perform and manage the replication operation. There are two basic approaches to host-based remote replication: LVM-based replication and database replication via log shipping.

- **LVM-Based Remote Replication**
- **Host-Based Log Shipping**

LVM-Based Remote Replication

LVM-based 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. Prior to the start of replication, identical volume groups, logical volumes, and file systems are created at the source and target sites. Initial synchronization of data between the source and the replica can be performed in a number of ways. One method is to backup the source data to tape and restore the data to the remote replica. Alternatively, it can be performed by replicating over the Until completion of initial synchronization, production work on the source volumes is typically halted. After initial synchronization, production work can be started on the source volumes and replication of data can be performed over an existing standard IP network (see Figure 14-3).

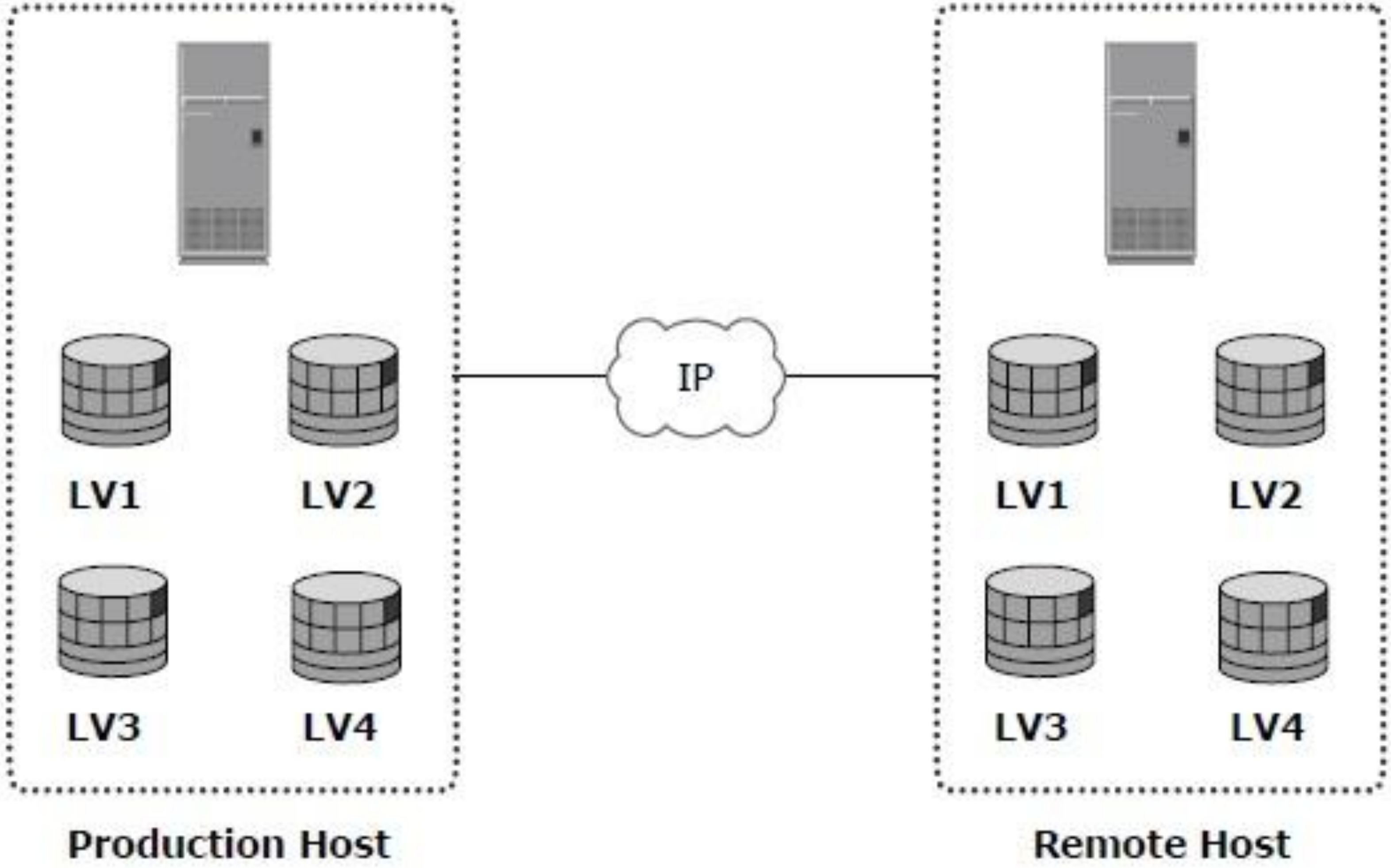


Figure 14-3: LVM-based remote replication

In asynchronous mode, writes are queued in a log file at the source and sent to the remote host in the order in which they were received. The size of the log file determines the RPO at the remote site. In the event of a network failure, writes continue to accumulate in the log file. If the log file fills up before the failure is resolved, then a full resynchronization is required upon network availability. In the event of a failure at the source site, applications can be restarted on the remote host, using the data on the remote replicas. LVM-based remote replication eliminates the need for a dedicated SAN infrastructure.

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 remote host (see Figure 14-4). The remote host receives the logs and applies them to the remote database.

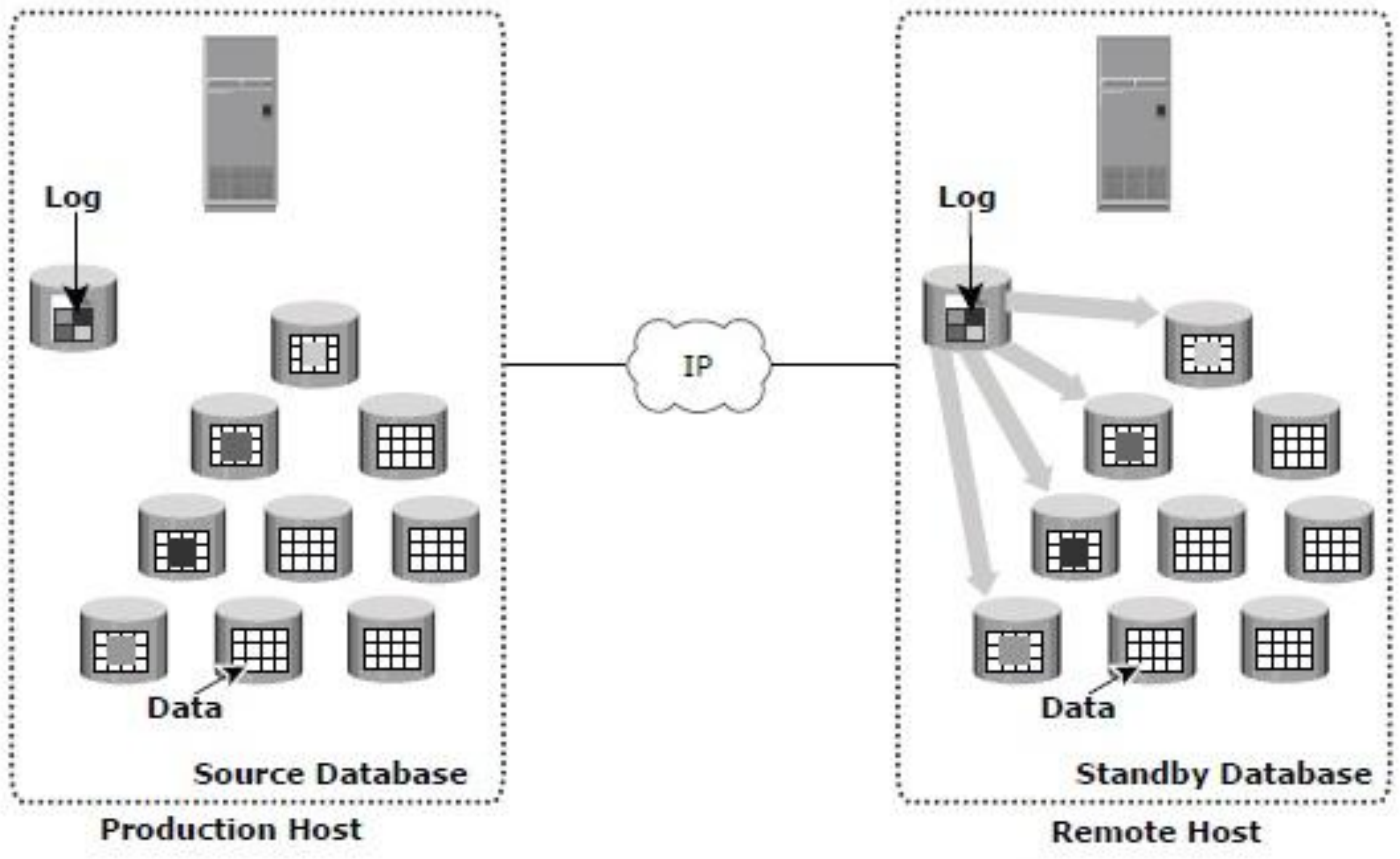


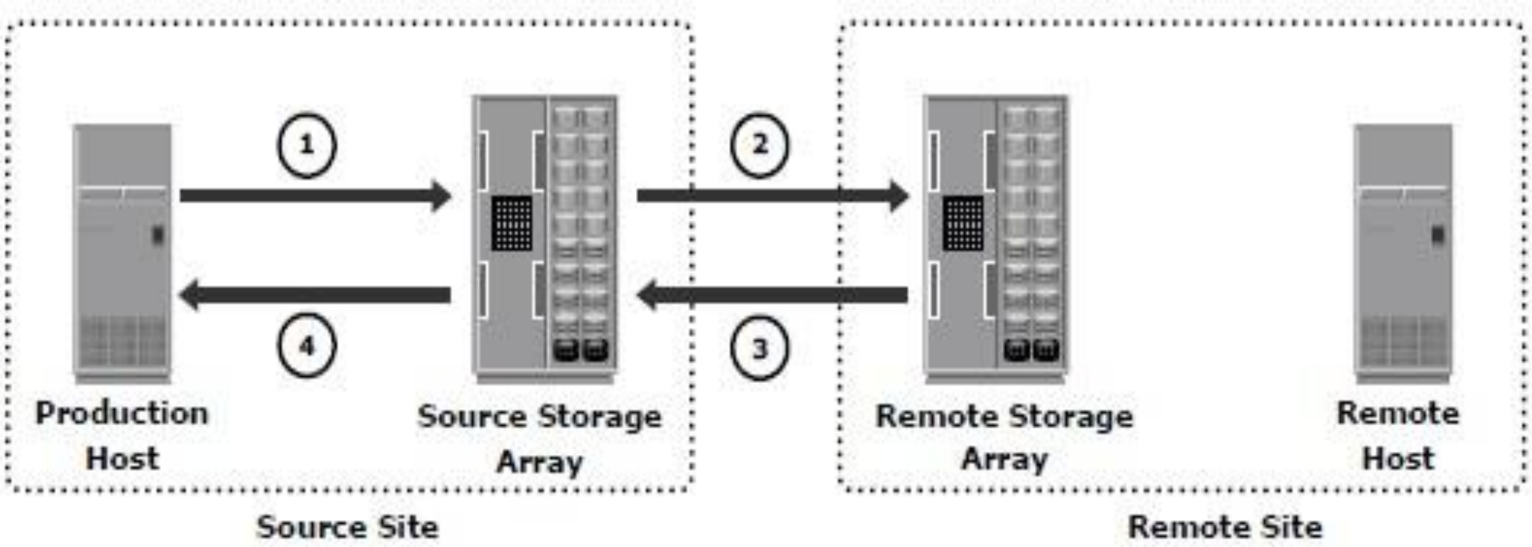
Figure 14-4: Host-based log shipping

Because the source host does not transmit every update and buffer them, this alleviates the burden on the source host CPU. Similar to LVM-based remote replication, the existing standard IP network can be used for replicating log files. Host-based log shipping does not scale well, particularly in the case of applications using federated databases.

Storage Array-Based Remote 14.2.2 Replication

In storage array-based remote replication, the array operating environment and resources perform and manage data replication. This relieves the burden on the host CPUs, which can be better utilized for running an application. A source and its replica device reside on different storage arrays. In other implementations, the storage controller is used for both the host and replication workload. Data can be transmitted from the source storage array to the target storage array over a shared or a dedicated network.

Synchronous Replication Mode



1 Write from the source host is recieved by the source storage array

2 Write is then transmitted to the remote storage array

3 Acknowledgment is sent to the source storage array by the remote storage array

4 Source storage array signals write-completion to the source host

Figure 14-5: Array-based synchronous remote replication

For synchronous remote replication, network bandwidth equal to or greater than the maximum write workload between the two sites should be provided at all times. Figure 14-6 illustrates the write workload (expressed in MB/s) over time. The “Max” line indicated in Figure 14-6 represents the required bandwidth that must be provisioned for synchronous replication. Bandwidths lower than the maximum write workload results in an unacceptable increase in application response time.

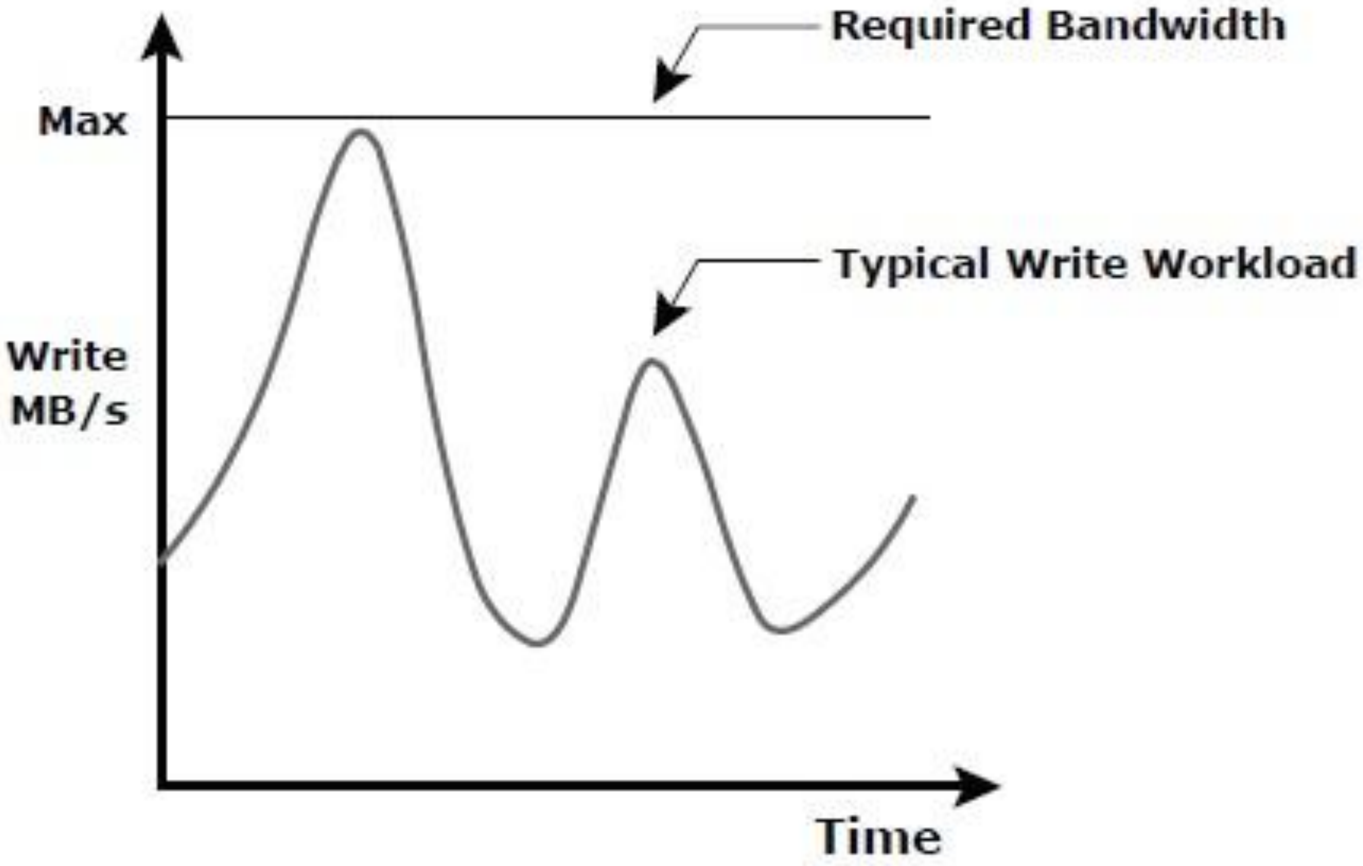
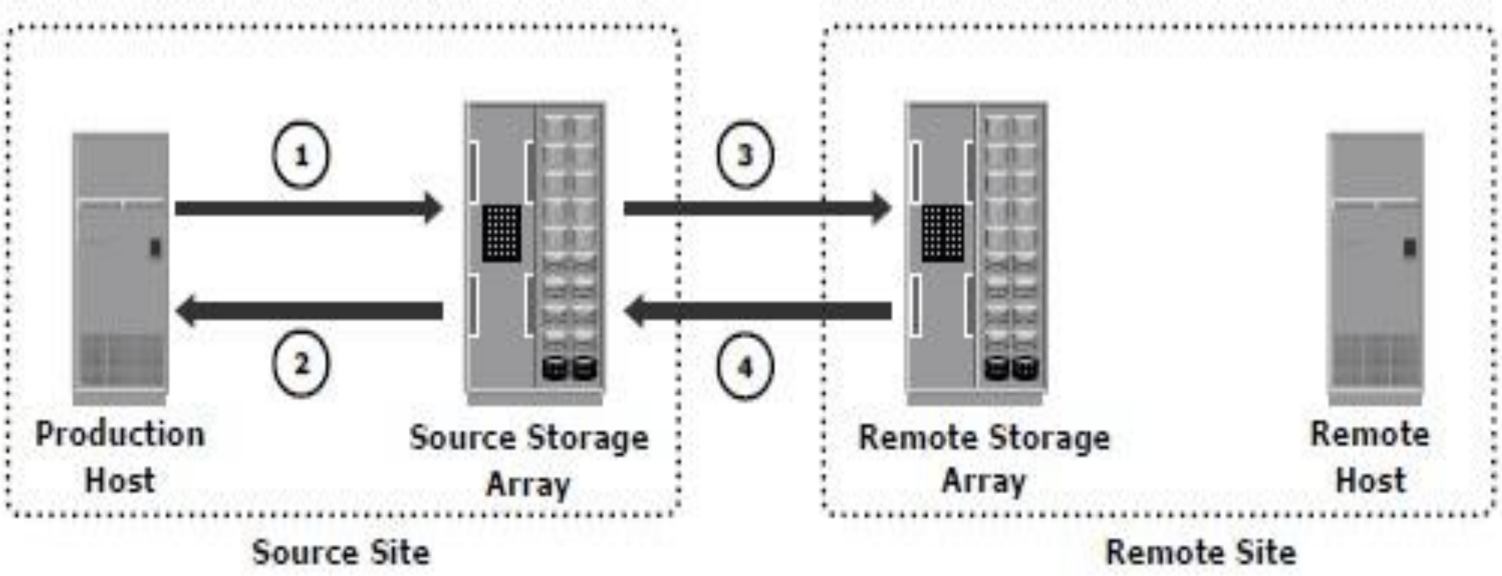


Figure 14-6: Network bandwidth requirement for synchronous replication

Asynchronous Replication Mode



1 Source host writes to the source storage array

2 Source array immediately acknowledges the source host

3 These writes are then transmitted to the target array

4 After the writes are recieved by the target array, it sends an acknowledge to source array

Figure 14-7: Array-based asynchronous remote replication

Some implementations of asynchronous remote replication maintain write ordering. A time stamp and sequence number are attached to each write when it is received by the source. Writes are then transmitted to the remote array, where they are committed to the remote replica in the exact order in which they were buffered at the source. This implicitly guarantees consistency of data on the remote replicas. Other implementations ensure consistency by leveraging the dependent write principle inherent to most DBMSs. The writes are buffered for a predefined period of time. At the end of this duration, the buffer is closed, and a new buffer is opened for subsequent writes. All writes in the closed buffer are transmitted together and committed to the remote replica.

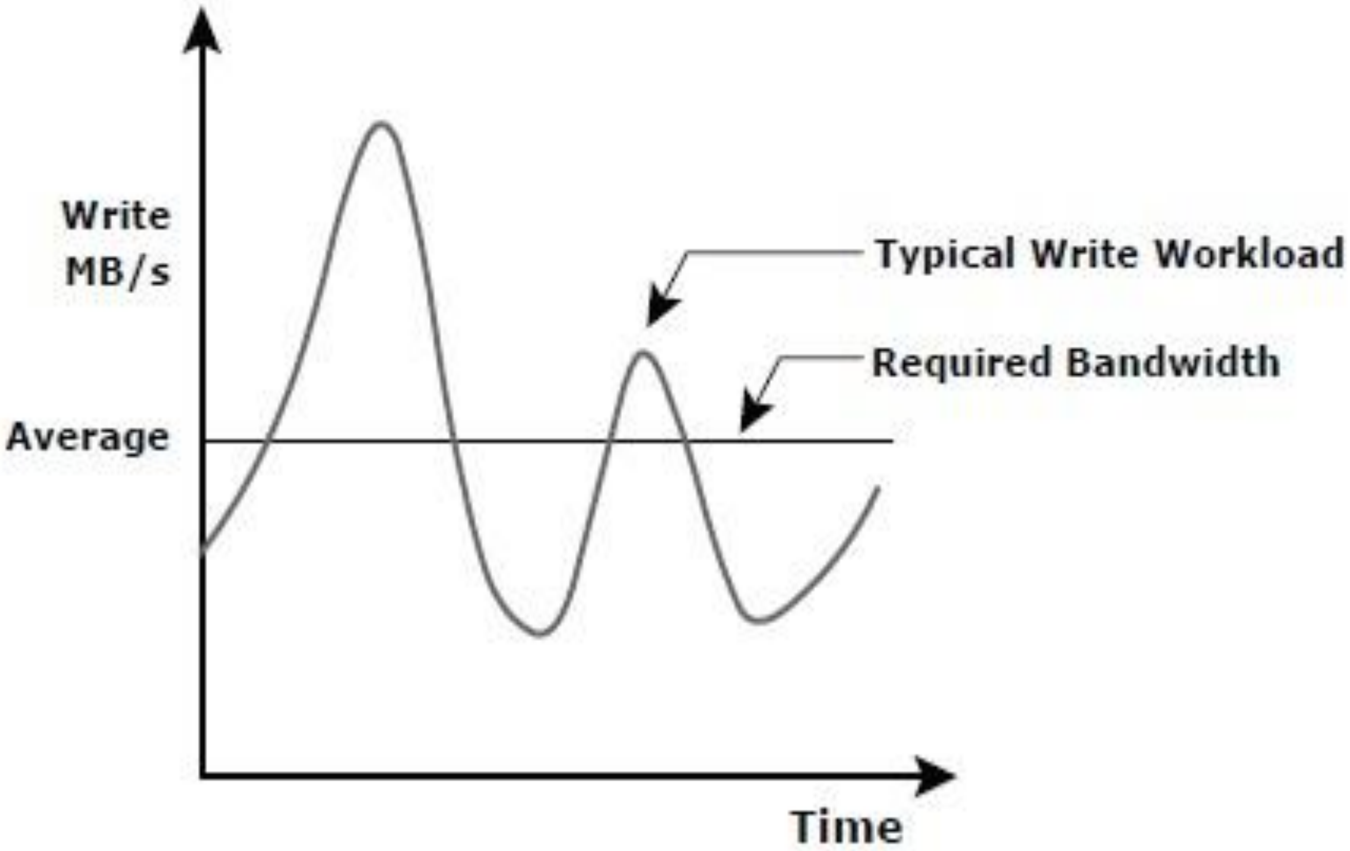
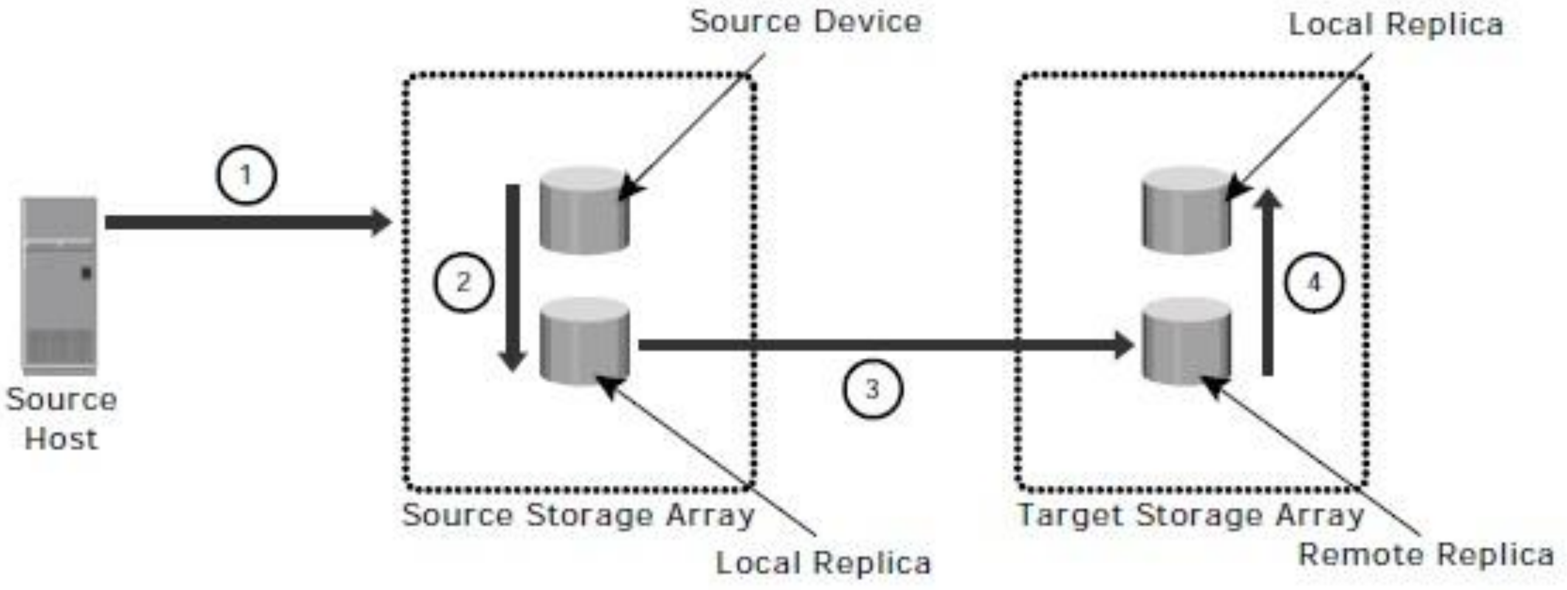


Figure 14-8: Network bandwidth requirement for asynchronous replication

Disk-Buffered Replication Mode



1. Source host writes data to source device
2. A consistent PIT local replica of the source device is created
3. Data from local replica in the source array is transmitted to its remote replica in the target array
4. A local PIT replica of the remote device on the target array is created

Figure 14-9: Disk-buffered remote replication

Three-Site Replication

- In synchronous and asynchronous replication, under normal conditions the workload is running at the source site.
- In synchronous replication, source and target sites are usually within 200 KM (125 miles) of each other.
- A regional disaster will not affect the target site in asynchronous replication, as the sites are typically several hundred or several thousand kilometers apart.
- Three-site replication is used to mitigate the risks identified in two-site replication. In a three-site replication, data from the source site is replicated to two remote data centers.

Three-Site Replication—Cascade/Multi-hop

In the **cascade/multi-hop** form of replication, data flows from the source to the intermediate storage array, known as a bunker, in the first hop and then from a bunker to a storage array at a remote site in the second hop. Replication between the source and the bunker occurs synchronously, but replication between the bunker and the remote site can be achieved in two ways: disk-buffered mode or asynchronous mode.

Synchronous + Asynchronous

This method employs a combination of **synchronous** and **asynchronous** remote replication technologies. Synchronous replication occurs between the source and the bunker.

Synchronous replication occurs between the source and the bunker. **Asynchronous** replication occurs between the bunker and the remote site.

The remote replica in the bunker acts as the source for the asynchronous replication to create a remote replica at the remote site. Figure 14-10(a) illustrates the **synchronous + Asynchronous** method.

Synchronous + Disk Buffered

This method employs a combination of local and remote replication technologies. Synchronous replication occurs between the source and the bunker: A consistent PIT local replica is created at the bunker. Data is transmitted from the local replica at the bunker to the remote replica at the remote site. Optionally, a local replica can be created at the remote site after data is received from the bunker. Figure 14-10(b) illustrates the **synchronous + disk buffered** method.

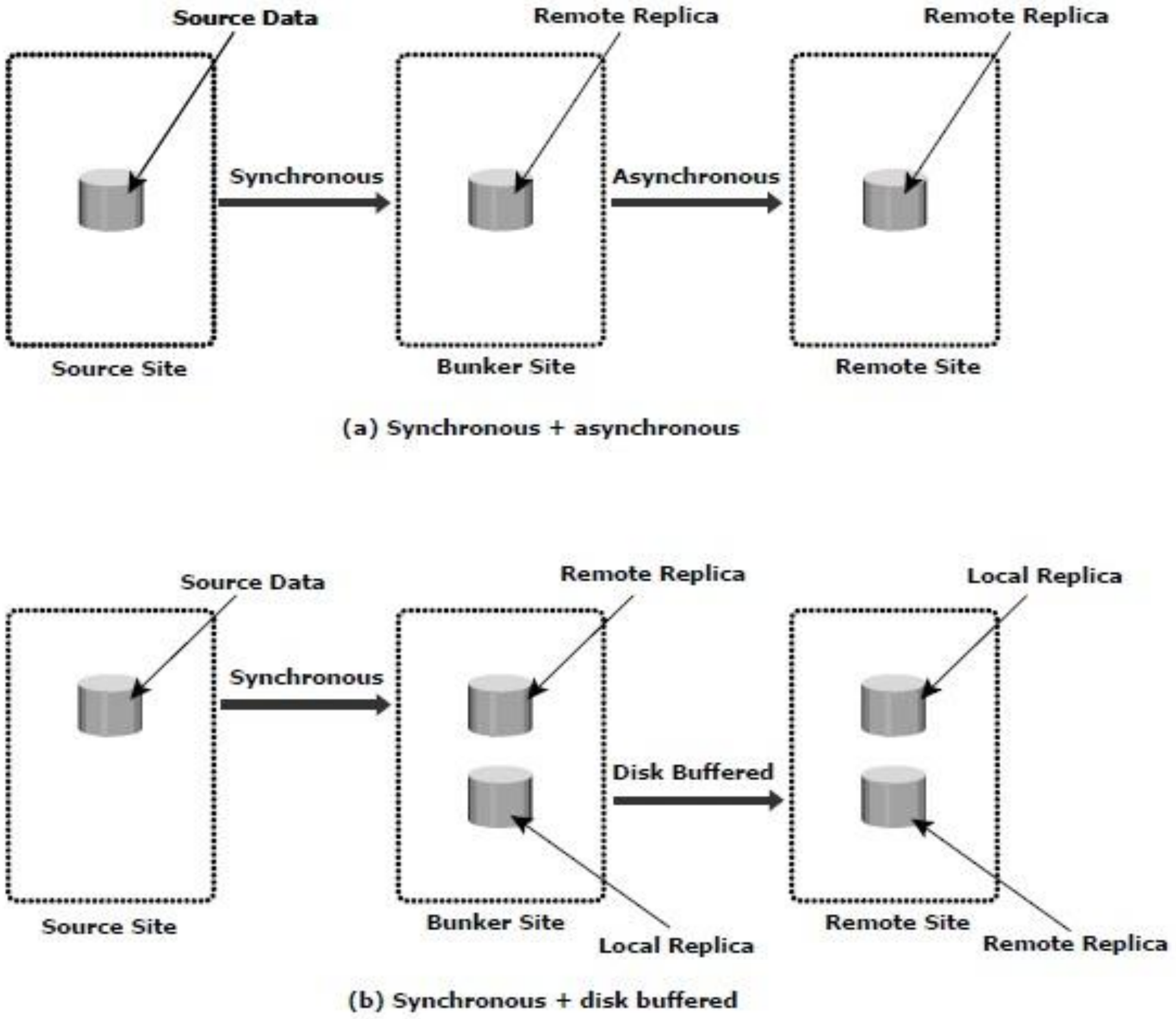


Figure 14-10: Three-site replication

Three-Site Replication—Triangle/Multi-target

In the three-site triangle/multi-target replication, data at the source storage array is concurrently replicated to two different arrays. The source-to bunker site (**target 1**) replication is synchronous, with a near-zero RPO. The source-to remote site (**target 2**) replication is asynchronous, with an RPO of minutes. The distance between the source and the remote site could be thousands of miles. This configuration does not depend on the bunker site for updating data on the remote site, because data is asynchronously copied to the remote site directly from the source.

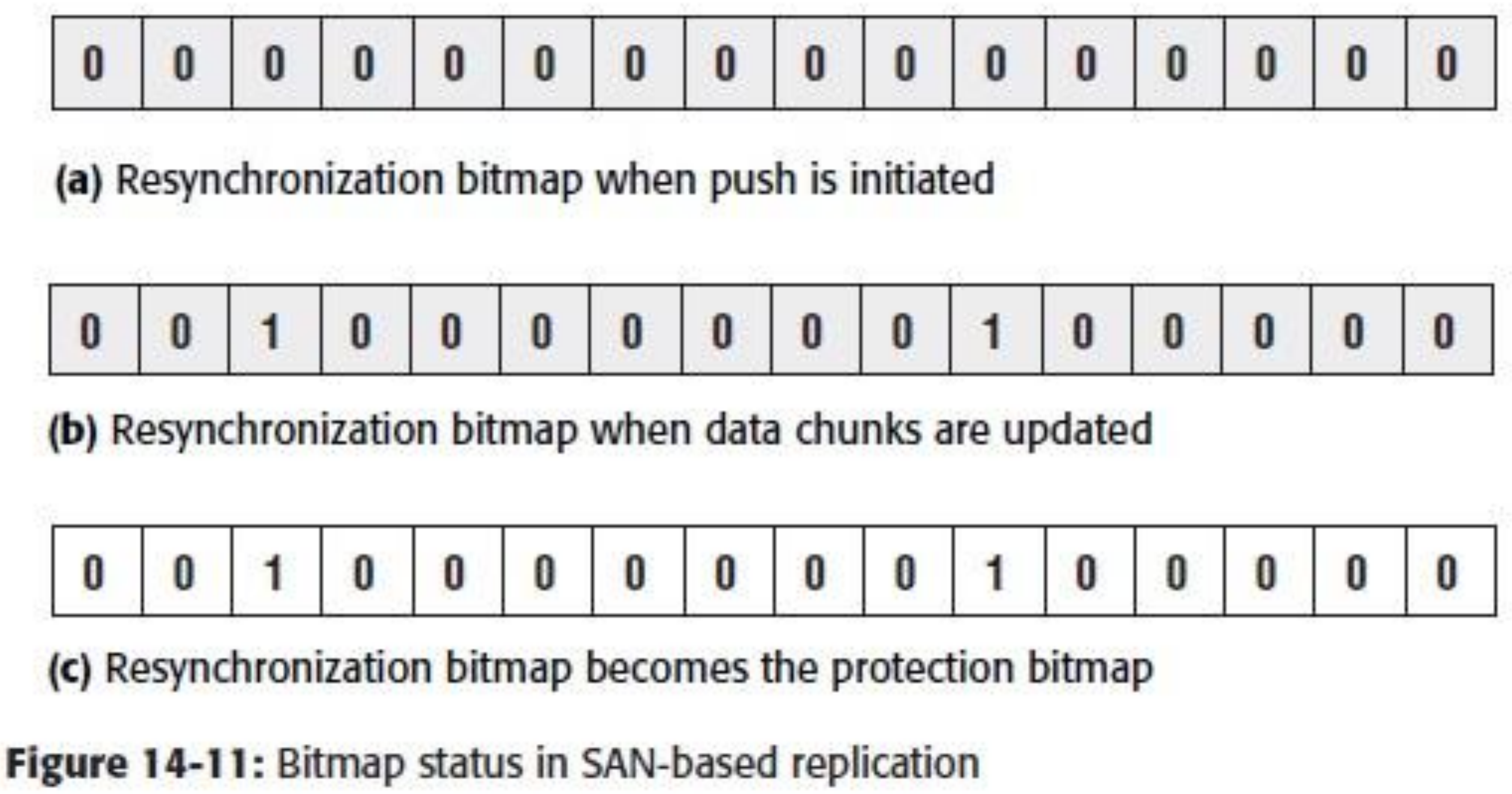
SAN-Based Remote Replication 14.2.3

- **SAN-based** remote replication enables the replication of data between heterogeneous storage arrays. Data is moved from one array to the other over the SAN/ WAN.

- **SAN-based** remote replication is a point-in-time replication technology. Uses of SAN-based remote replication include data mobility, remote vaulting, and data migration.

- **SAN-based** replication uses two types of operations: push and pull. These terms are defined from the perspective of the control array.

- In **SAN-based** replication, the control array can keep track of changes made to the control devices after the replication session is activated. This is allowed in the incremental push operation only.



0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(a) Resynchronization bitmap when push is initiated

0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0

(b) Resynchronization bitmap when data chunks are updated

0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0

(c) Resynchronization bitmap becomes the protection bitmap

Figure 14-11: Bitmap status in SAN-based replication

Network Infrastructure 14.3

For remote replication over extended distances, optical network technologies such as dense wavelength division multiplexing (**DWDM**), coarse wavelength division multiplexing (**CWDM**), and synchronous optical network (**SONET**) are deployed.

DWDM 14.3.1

DWDM is an optical technology by which data from different channels are carried at different wavelengths over a fiber-optic link. It is a fiber-optic transmission technique that uses light waves to transmit data parallel by bit or serial by character. It integrates multiple light waves with different wavelengths in a group and directs them through a single optical fiber. Using DWDM, different data formats at different data rates can be transmitted together. Specifically, IP ESCON, FC, SONET and ATM data can all travel at the same time within the optical fiber (see Figure 14-12).

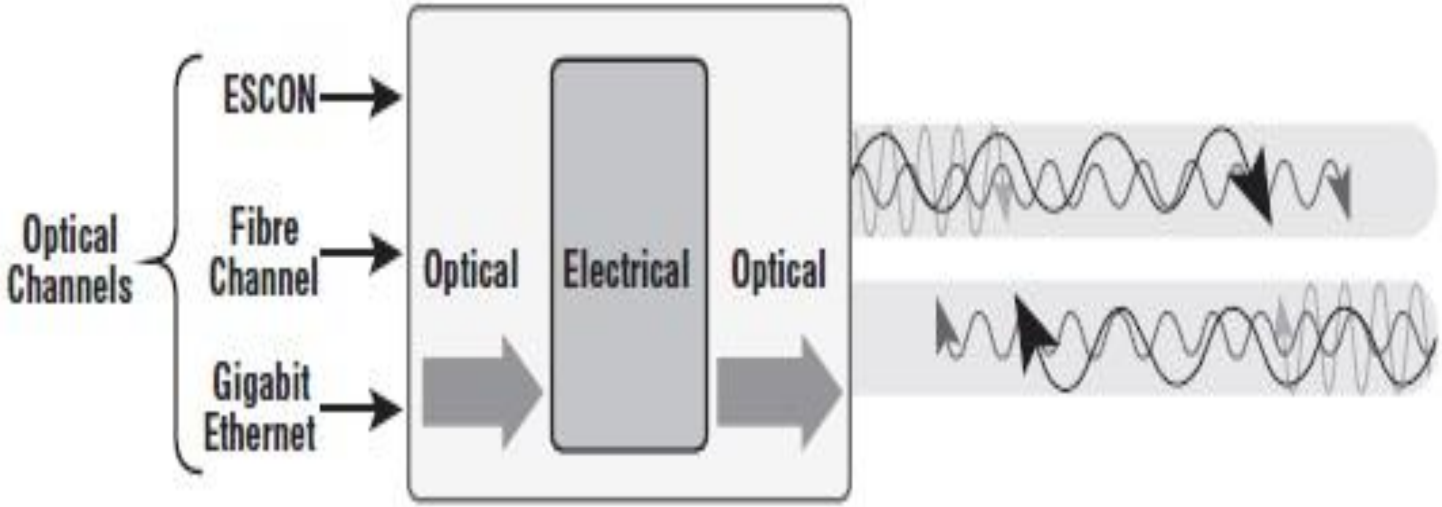


Figure 14-12: Dense wavelength division multiplexing (DWDM)

CWDM, like DWDM, uses multiplexing and demultiplexing on different channels by assigning varied wavelengths to each channel. Compared to DWDM, CWDM is used to consolidate environments containing a low number of channels at a reduced cost.

SONET 14.3.2

SONET (synchronous optical network) is a network technology that involves transferring a large payload through an optical fiber over long distances. SONET multiplexes data streams of different speeds into a frame and sends them across the network. The European variation of SONET is called synchronous digital hierarchy (SDH). Figure 14-13 shows the multiplexing of data streams of different speeds in SONET and SDH technologies. SONET/SDH uses generic framing procedure (GFP) and supports the transport of both packet-oriented (Ethernet, IP) and character-oriented (FC) data.

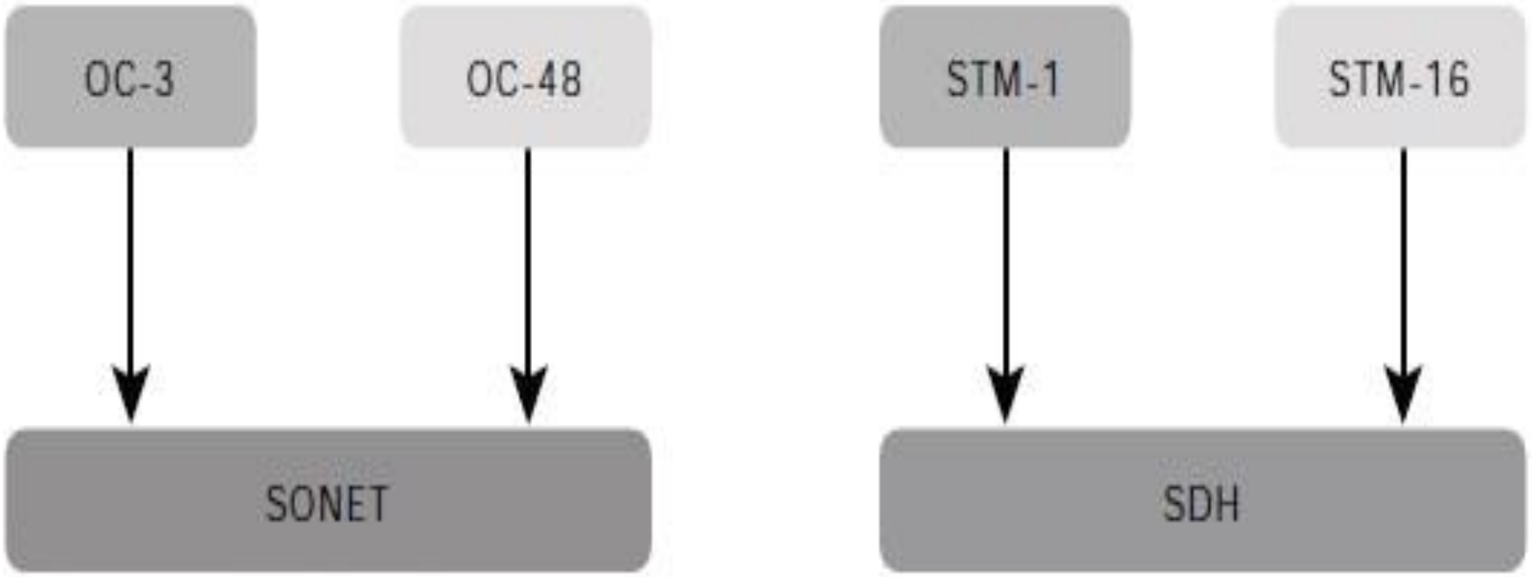


Figure 14-13: Data stream multiplexing in SONET and SDH

Summary

This chapter detailed remote replication. As a primary utility, remote replication provides disaster recovery and disaster restart solutions. It enables business operations to be rapidly restarted at a remote site following an outage, with acceptable data loss. Remote replication enables BC operations from a target site. The replica of source data at the target can be used for backup and testing. This replica can also be used for data repurposing, such as report generation, data warehousing, and decision support. The segregation of business operations between the source and target protects the source from becoming a performance bottleneck, ensuring improved production performance at the source.

Remote replication may also be used for data center migrations, providing the least disturbance to production operations because the applications accessing the source data are not affected. This chapter also described different types of remote replication solutions. The distance between the primary site and the remote site is a prime consideration when deciding which remote replication technology solution to deploy. Asynchronous replication may adequately meet the RPO and RTO needs, while permitting greater distances between the sites. Storage management solutions provide the capability to not only automate business continuity solutions, but also enable centralized management of the overall storage infrastructure. Organizations must ensure security of the information assets. The next chapter details storage security and management.