UNIT - 3

Business Continuity and Backup Recovery

SLO-1 : Business Continuity : Information Availability

Lesson Objective

After completing this Lesson, you will be able to:

- Define Business Continuity and Information Availability
- Detail impact of information unavailability
- Define BC measurement and terminologies
- Describe BC planning process
- Detail BC technology solutions

What is Business Continuity (BC)

- Business Continuity is preparing for, responding to, and recovering from an application outage that adversely affects business operations
- Business Continuity solutions address unavailability and degraded application performance
- Business Continuity is an integrated and enterprise wide process and set of activities to ensure "information availability"

Business Continuity

- **Business continuity (BC)**
  - *Integrated and enterprise-wide process* that includes all activities (internal and external to IT)
  - Business must perform to mitigate(defect) the impact of planned and unplanned downtime
  - **Proactive measures**
    - Business impact analysis
    - Risk assessments
    - BC technology solutions
    - deployment (backup and replication), and reactive measures, such as disaster
    - recovery and restart, to be invoked in the event of a failure

Business Continuity

- In a virtualized environment,
  - BC technology solutions need to protect both physical and virtualized resources
- Goal of a BC
  - ensure the **"information availability"** required to conduct vital business operations.

Information Availability

1. Causes of Information Unavailability
2. Consequences of Downtime
3. Measuring Information Availability

What is Information Availability (IA)

- IA refers to the ability of an infrastructure to function according to business expectations during its specified time of operation
- IA can be defined in terms of three parameters:
  - **Accessibility**
  - **Reliability**
  - **Timeliness**

Information Availability

- IA can be defined in terms of

Accessibility

- Information should be accessible at the right place, to the right user.

Reliability

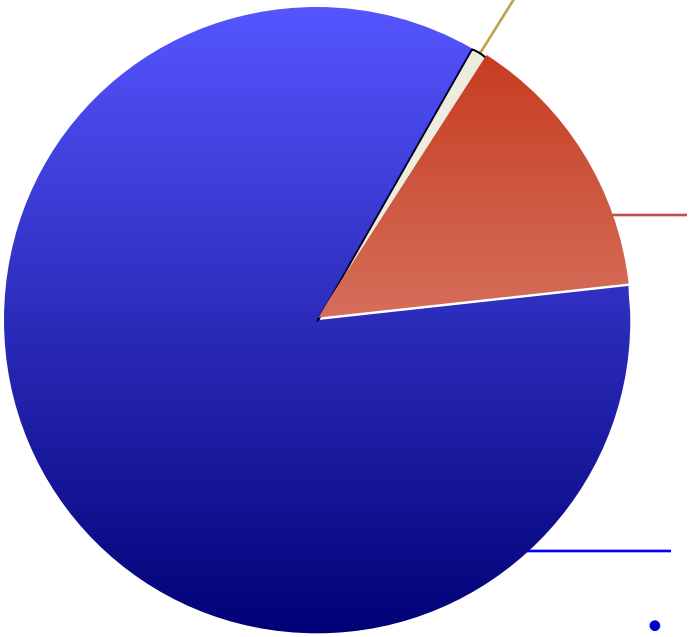
–The components delivering the information should be able to function without failure, under stated conditions, for a specified amount of time

- Information should be
  - Reliable
  - correct in all aspects.
  - "the same" as what was stored
  - no alteration or corruption to the information.

Timeliness of information

- Exact moment or the time window (a particular time of the day, week, month, and year as specified) during which information must be accessible.

1. Causes of Information Unavailability



Disaster (<1% of Occurrences)

Natural or man made

- Flood, fire, earthquake
- Contaminated building

Unplanned Outages (20%)

Failure

- Database corruption
- Component failure
- Human error

Planned Outages (80%)

Competing workloads

- installation/integration/maintenance of new hardware
- software upgrades or patches, taking backups,
- application and data restores,
- facility operations (renovation and construction), and refresh/migration of the testing to the production environment.

2. Consequences of Downtime / Impact of Downtime

Know the downtime costs (per hour, day, two days...)

Lost Productivity

- Number of employees impacted (x hours out * hourly rate)
- reduced output per unit of labor, equipment, and capital

Lost Revenue

- Direct loss
- Compensatory payments
- Lost future revenue
- Billing losses
- Investment losses

Damaged Reputation

- Customers
- Suppliers
- Financial markets
- Banks
- Business partners

Poor Financial Performance

- Revenue recognition
- Cash flow
- Lost discounts
- Payment guarantees
- Credit rating
- Stock price

Other Expenses

Temporary employees, equipment rental, overtime costs, extra shipping costs, travel expenses...

Impact of Downtime

Average cost of downtime per hour = average productivity loss per hour + average revenue loss per hour ₹ **31250** **(6250+25000)**

Where,

- Productivity loss per hour = (total salaries and benefits of all employees per week) / (average number of working hours per week)

  = 50000/8 6250

- Average revenue loss per hour = (total revenue of an organization per week) / (average number of hours per week that an organizations is open for business)

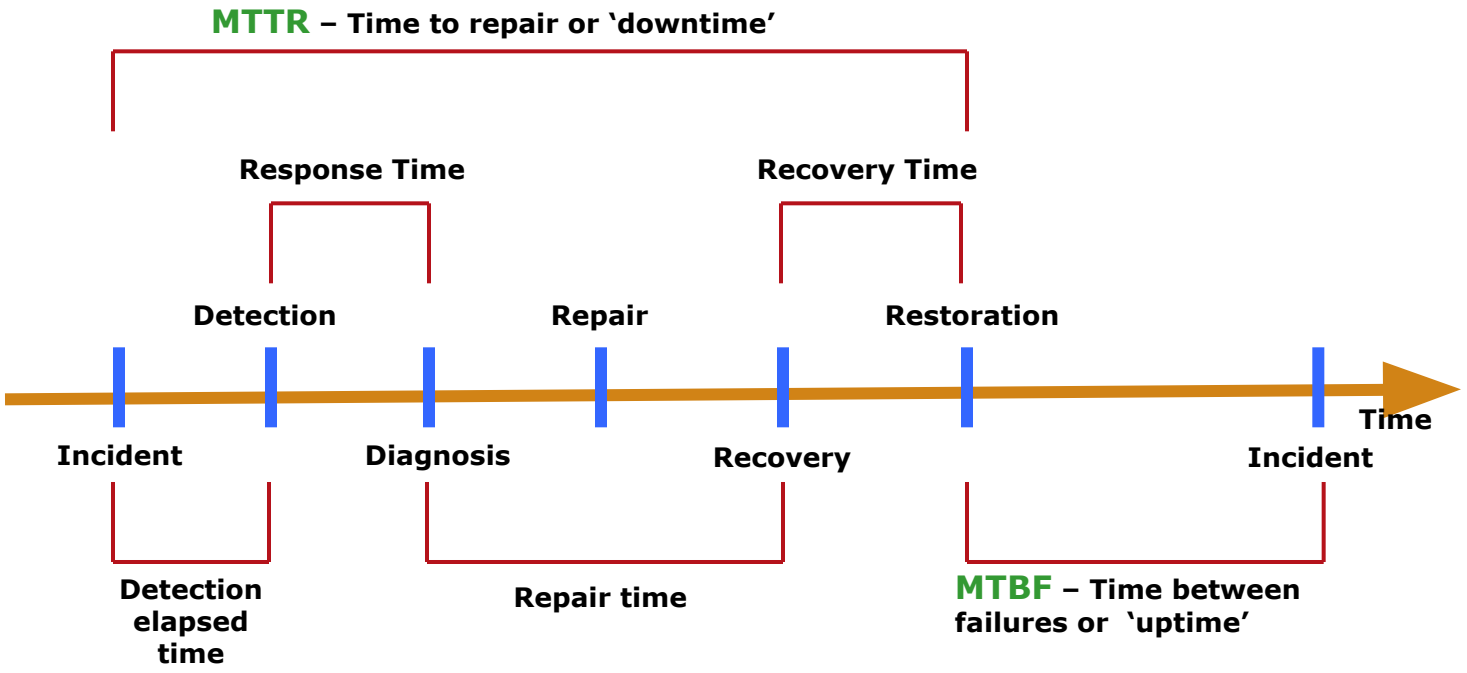
  = 200000/8 25,000

3. Measuring Information Availability

- IA relies on the availability of both physical and virtual components of a data center.
- Failure of these components might disrupt IA.
  - A failure is the termination of a component's capability to perform a required function.

- The component's capability can be restored by performing an external corrective action, such as a manual reboot, repair, or replacement of the failed component(s).
- Repair involves restoring a component to a condition that enables it to perform a required function.
- Proactive risk analysis, performed as part of the BC planning process,
  - considers the component failure rate and average repair time, which are measured by Mean Time Between Failure (MTBF) and Mean Time To Repair (MTTR):

Measuring Information Availability



- MTTR includes the total time required to do the following activities: Detect the fault, mobilize the maintenance team, diagnose the fault, obtain the spare parts, repair, test, and restore the data.

Measuring Information Availability (contd.)

- MTBF: Average time available for a system or component to perform its normal operations between failures

  = Total uptime/Number of failures

- MTTR: Average time required to repair a failed component

  = Total downtime/Number of failures

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

  IA = MTBF / (MTBF + MTTR)

  or

  IA = uptime / (uptime + downtime)

  - System uptime is the period of time during which the system is in an accessible state
  - System downtime is the period of time during which the system is not accessible state

Availability Measurement – Levels of '9s' Availability

- Uptime per year is based on the exact timeliness requirements of the service.
- This calculation leads to the number of "9s" representation for availability metrics.
- Table lists the approximate amount of downtime allowed for a service to achieve certain levels of 9s availability.
- For example, a service that is said to be "five 9s available" is available for percent of the scheduled time in a year (24 ×365)

% Uptime	% Downtime	Downtime per Year	Downtime per Week
98%	2%	7.3 days	3hrs 22 min
99%	1%	3.65 days	1 hr 41 min
99.8%	0.2%	17 hrs 31 min	20 min 10 sec
99.9%	0.1%	8 hrs 45 min	10 min 5 sec
99.99%	0.01%	52.5 min	1 min
99.999%	0.001%	5.25 min	6 sec
99.9999%	0.0001%	31.5 sec	0.6 sec

SLO-2 :

1. **BC Terminology,**
2. **BC Planning Life Cycle**

BC Terminology

BC Terminologies - Common terms related to BC operations

- Disaster recovery
  - Coordinated process of restoring systems, data, and infrastructure required to support ongoing business operations in the event of a disaster
  - Restoring previous copy of data and applying logs to that copy to bring it to a known point of consistency
  - Generally implies use of backup technology
- Disaster restart
  - Process of restarting from disaster using mirrored consistent copies of data and applications
  - Generally implies use of replication technologies

BC Terminologies (Cont.)

Recovery Point Objective (RPO)

- Point in time to which systems and data must be recovered after an outage
- Amount of data loss that a business can endure

Recovery Time Objective (RTO)

- Time within which systems, applications, or functions must be recovered after an outage
- Amount of downtime that a business can endure and survive

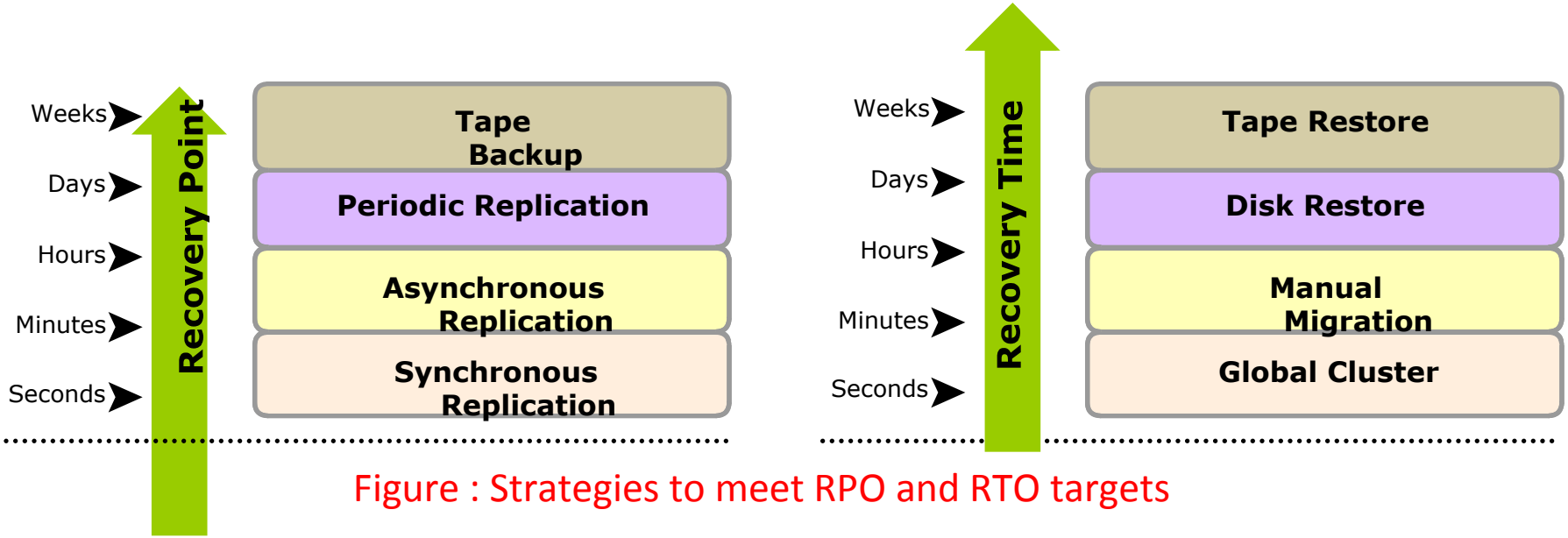


Figure : Strategies to meet RPO and RTO targets

Recovery-point objective

Recovery-time objective

BC Terminologies (Cont.)

- **RPO of 24 hours:** Backups are created at an offsite tape library 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 to the point of the last log shipment.
- **RPO in the order of minutes:** Mirroring data asynchronously to a remote site
- **Near zero RPO:** Mirroring data synchronously to a remote site

BC Terminologies (Cont.)

- **RTO of 72 hours:** Restore from tapes available at a cold site *(unused until a disaster occurs)*
- **RTO of 12 hours:** Restore from tapes available at a hot site *(defined as a backup site, which is up and running continuously)*
- **RTO of few hours:** Use of data vault at a hot site
- **RTO of a few seconds**: Cluster production servers with bidirectional mirroring, enabling the applications to run at both sites simultaneously

BC Terminologies (Cont.)

- **Data vault:** A repository at a remote site where data can be periodically or continuously copied (either to tape drives or disks) so that there is always a copy at another site
- **Hot site:** A site where an enterprise's operations can be moved in the event of disaster.
- It is a site with the required hardware, operating system, application, and network support to perform business operations, where the equipment is available and running at all times.
- **Cold site:** A site where an enterprise's operations can be moved in the event of disaster, with minimum IT infrastructure and environmental facilities in place, but not activated

BC Terminologies (Cont.)

- **Server Clustering:** A group of servers and other necessary resources coupled to operate as a single system.
- Clusters can ensure high availability and load balancing.
- Typically, in failover clusters, one server runs an application and updates the data, and another server is kept as standby to take over completely, as required.
- In more sophisticated clusters, multiple servers may access data, and typically one server is kept as standby.
- Server clustering provides load balancing by distributing the application load evenly among multiple servers within the cluster.

BC Planning Life Cycle

Business Continuity (BC) Planning Lifecycle

- BC planning must follow a disciplined approach like any other planning process.
- Organizations today dedicate specialized resources to develop and maintain BC plans.

The BC planning lifecycle includes five stages:

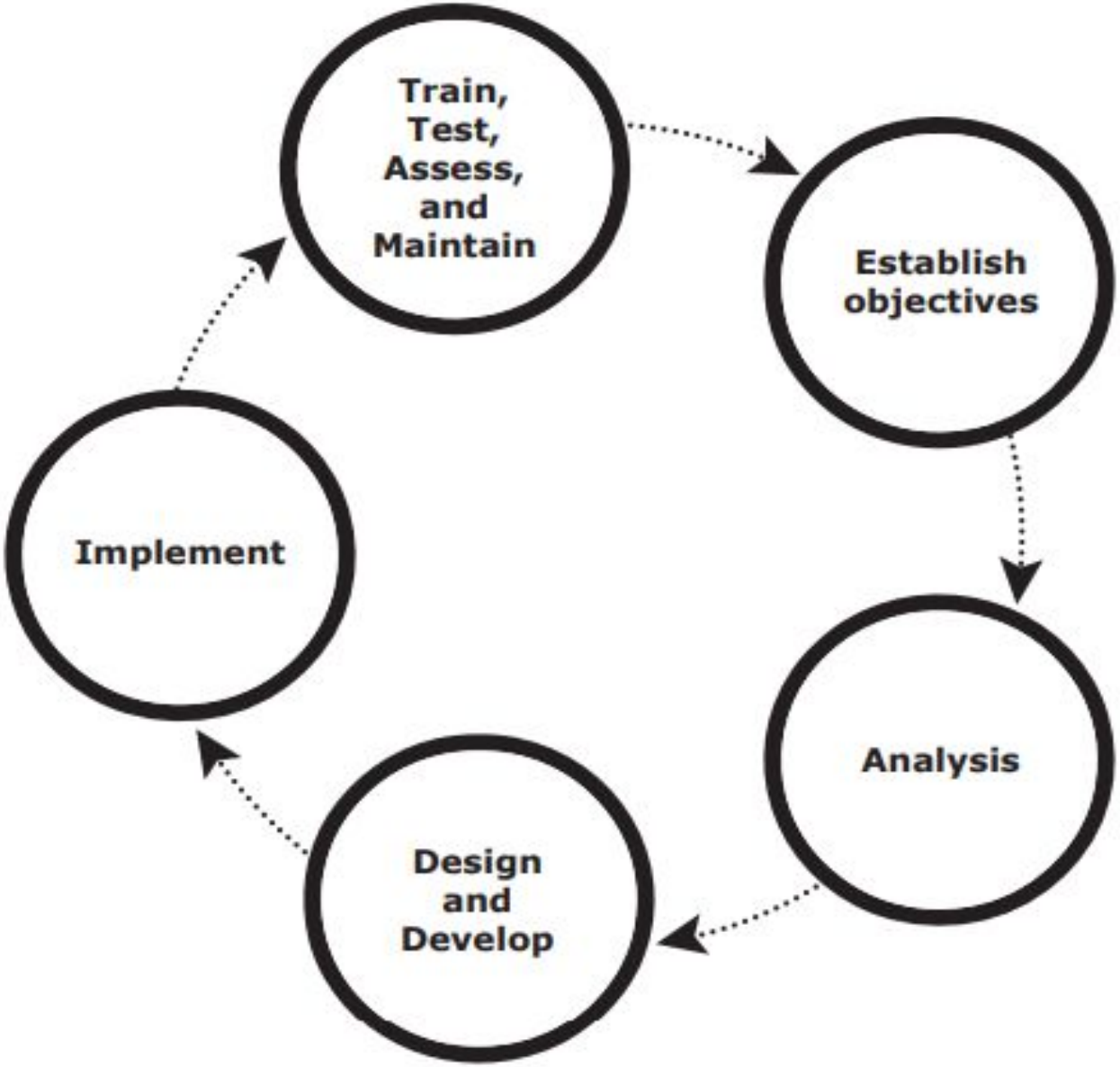
1. **Establishing objectives**
2. **Analyzing**
3. **Designing and developing**
4. **Implementing**
5. **Training, testing, assessing, and maintaining**

Establishing objectives

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

- 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.

Business Continuity (BC) Planning Lifecycle



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.
- Evaluate options.

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.

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.

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.

SLO-1 :

1. **Failure Analysis**
2. **Business Impact Analysis**

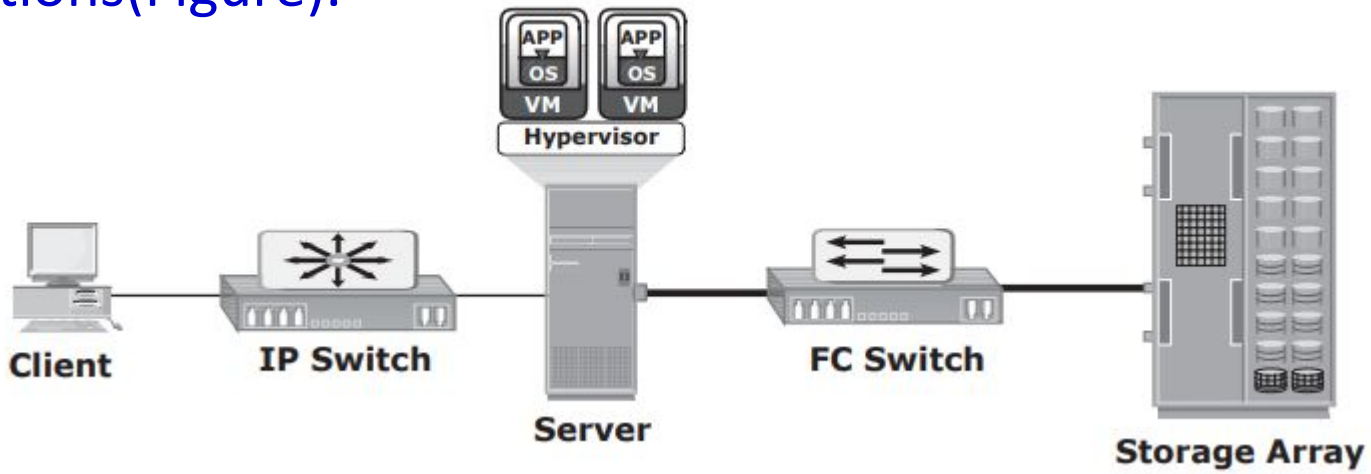
Failure Analysis

Failure Analysis

- Failure analysis involves analyzing both the physical and virtual infrastructure components
  - To identify systems that are susceptible to a single point of failure and implementing fault-tolerance mechanisms.
  - Single Point of Failure
  - Resolving Single Points of Failure
  - Multipathing Software

Failure Analysis: (1) Single Points of Failure

- A single point of failure refers to the failure of a component that can terminate the availability of the entire system or IT service.
- A system setup in which an application, running on a VM, provides an interface to the client and performs I/O operations(Figure).

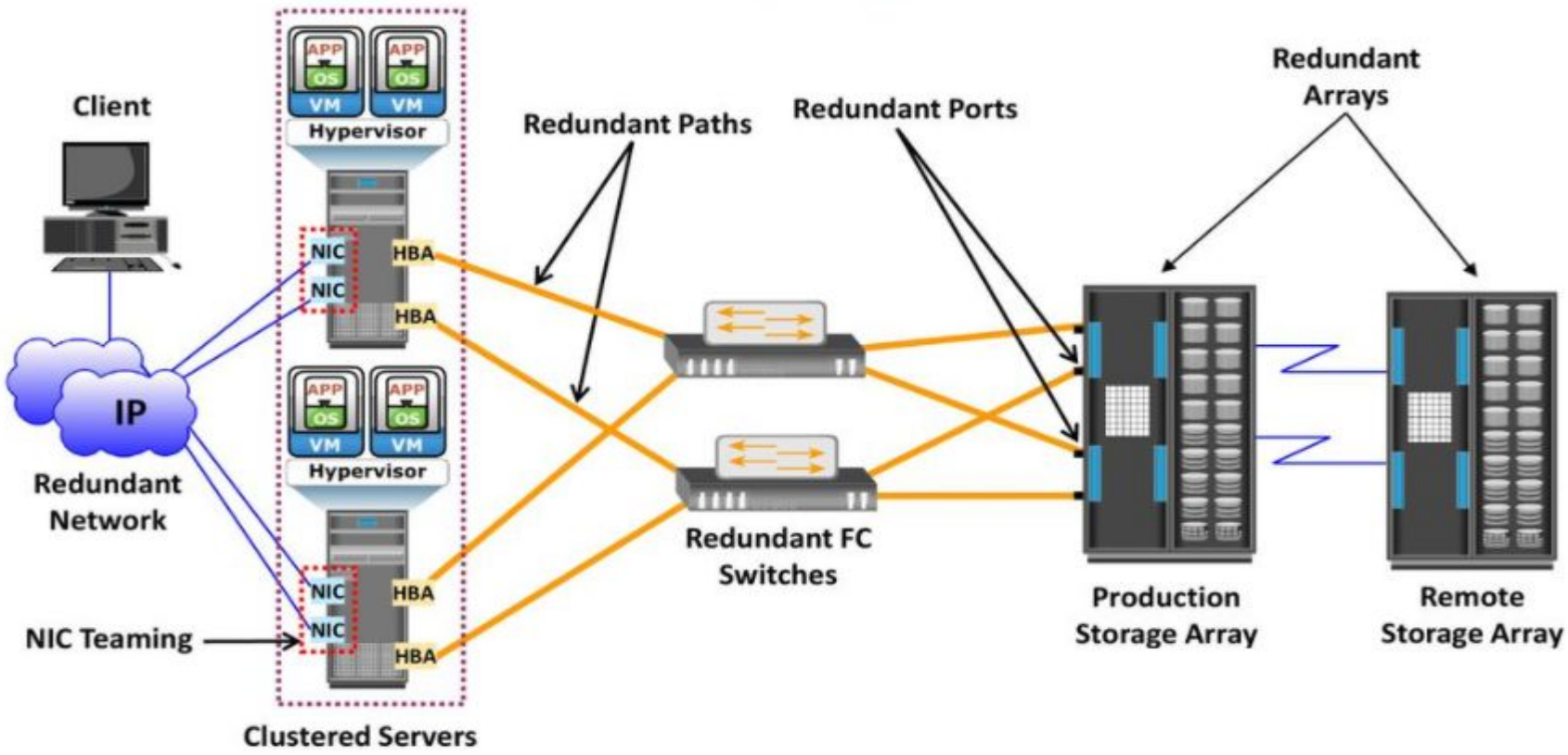


- The client is connected to the server through an IP network, and the server is connected to the storage array through an FC connection.

Failure Analysis: (1) Single Points of Failure

- In a setup in which each component must function as required to ensure data availability, the failure of a single physical or virtual component causes the unavailability of an application.
- This failure results in disruption of business operations.
- For example, failure of a hypervisor can affect all the running VMs and the virtual network, which are hosted on it.
- In the setup shown in Figure, several single points of failure can be identified.
- A VM, a hypervisor, an HBA/NIC on the server, the physical server, the IP network, the FC switch, the storage array ports, or even the storage array could be a potential single point of failure.

Failure Analysis: (2) Resolving Single Points of Failure / Fault tolerant



- To mitigate a single point of failure, systems are designed with redundancy, such that the system will fail only if all the components in the redundancy group fail.
- This ensures that the failure of a single component does not affect data availability.
- Careful analysis is performed to eliminate every single point of failure

Failure Analysis: (2) Resolving Single Points of Failure / Fault tolerant

- Based on the figure, implementation to resolve single points of failure includes:
  - Configuration of multiple HBAs to mitigate single HBA failure.
  - Configuration of multiple fabrics to account for a switch failure.
  - Configuration of multiple storage array ports to enhance the storage array's availability.
  - RAID configuration to ensure continuous operation in the event of disk failure.
  - Implementing a storage array at a remote site to mitigate local site failure.
  - Implementing server (host) clustering, a fault-tolerance mechanism whereby two or more servers in a cluster access the same set of volumes.
    - Clustered servers exchange heartbeats to inform each other about their health.
    - If one of the servers fails, the other server takes up the complete workload.

Multipathing Software

- Configuration of multiple paths increases data availability
- Even with multiple paths, if a path fails I/O will not reroute unless system recognizes that it has an alternate path
- Multi-pathing software helps to recognize and utilizes alternate I/O path to data
- Multi-pathing software also provide the load balancing
- Load balancing improves I/O performance and data path utilization

Business Impact Analysis

Business Impact Analysis(BIA)

- Identifies which business units, operations, and processes are essential to the survival of the business.
- Estimate the cost of failure for each business process.
- Calculate the maximum tolerable outage and defines Recovery Time Objective(RTO) for each business process.
- Businesses can prioritize and implement countermeasures to mitigate the likelihood of such disruptions.

BIA includes the following set of tasks:

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

SLO-2 :
BC Technology Solutions

BC Technology Solutions

- After analyzing the business impact of an outage, designing the appropriate solutions to recover from a failure is the next important activity.
- One or more copies of the data are maintained using any of the following strategies,
  - so that data can be recovered or business operations can be restarted using an alternative copy :
- **Backup:** Data backup is a predominant method of ensuring data availability. The frequency of backup is determined based on RPO, RTO, and the frequency of data changes.

BC Technology Solutions

- **Local replication:** Data can be replicated to a separate location within the same storage array. The replica is used independently for other business operations. Replicas can also be used for restoring operations if data corruption occurs.
- **Remote replication:** Data in a storage array can be replicated to another storage array located at a remote site. If the storage array is lost due to a disaster, business operations can be started from the remote storage array

SLO-1 :
Backup and Archive :
Backup Purpose

What is a Backup?

- Backup is an additional copy of data that can be used for restore and recovery purposes.
- The Backup copy is used when the primary copy is lost or corrupted.
- This Backup copy can be created as a:
  - Simple copy (there can be one or more copies)
  - Mirrored copy (the copy is always updated with whatever is written to the primary copy.)
    - Data archiving is the process of moving data that is no longer actively used, from primary storage to a low-cost secondary storage.

Backup Purpose

Backup Purpose

Backups are performed to serve three purposes:

1. Disaster recovery,
2. Operational recovery, and
3. Archival

It's All About Recovery

- Businesses back up their data to enable its recovery in case of potential loss
- Businesses also back up their data to comply with regulatory requirements

Backup purposes:

- Disaster Recovery
  - Restores production data to an operational state after disaster
- Operational recovery
  - Restore data in the event of data loss or logical corruptions that may occur during routine processing
- Archival
  - Preserve transaction records, email, and other business work products for regulatory compliance

Backup Considerations

Backup Considerations

- The amount of data loss and downtime that a business can endure in terms of RPO and RTO are the primary considerations in selecting and implementing a specific backup strategy
  - RPO refers to the point in time to which data must be recovered, and the point in time from which to restart business operations.
  - This specifies the time interval between two backups.
  - In another words, RPO determines backup frequency

Backup Considerations

- 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.
- The backup media type or backup target is another consideration, that is driven by RTO and impacts the data recovery time.
  - The time-consuming operation of starting and stopping in a tape-based system affects the backup performance, especially while backing up a large number of small fi les.

Backup Considerations

- The location, size, number of fi les, and data compression should also be considered because they might affect the backup process.
- Location is an important consideration for the data to be backed up.
  - Many organizations have dozens of heterogeneous platforms locally and remotely supporting their business.

SLO-1 & SLO-2:

1. **Backup Granularity**
2. **Recovery Considerations**

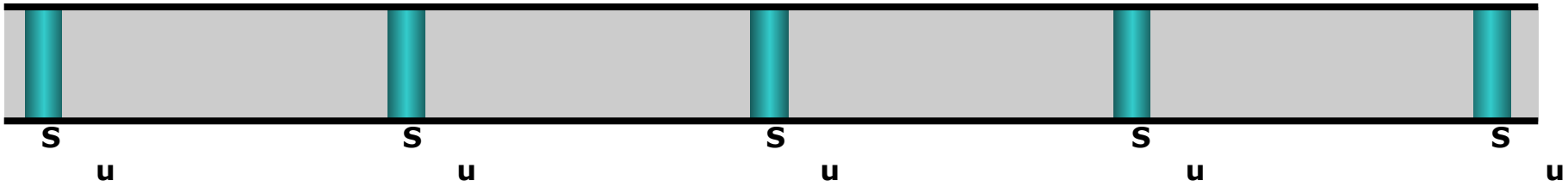
Backup Granularity

Backup Granularity

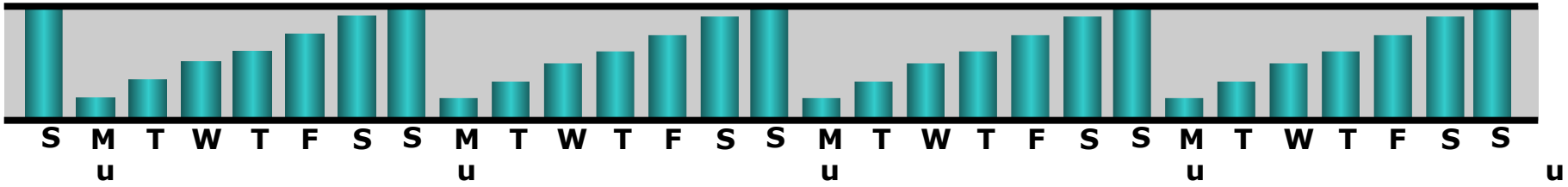
- Backup granularity depends on business needs and the required RTO/RPO.
- Based on the granularity, backups can be categorized as full, incremental and cumulative (differential).
- Most organizations use a combination of these three backup types to meet their backup and recovery requirements.

Backup Granularity

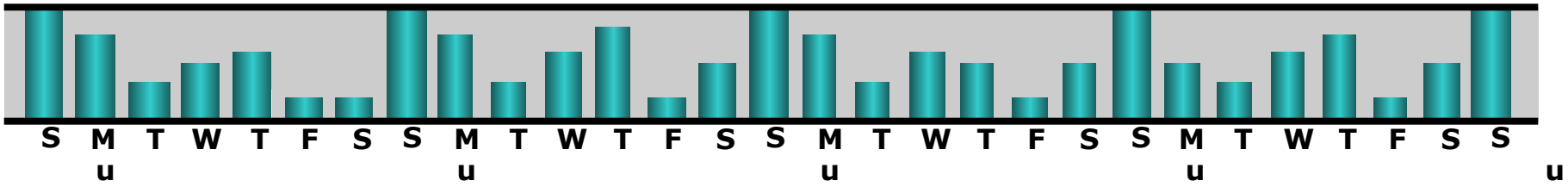
Full Backup



Cumulative (Differential) Backup



Incremental Backup



Different backup granularity

Full backup

- Backup of the complete data at a certain point in time
- Copy the data in the production volumes to a backup storage device
- Faster recovery but requires more storage space and also takes more time to back up

Incremental backup

- Copies data that has changed since the last full or incremental backup, whichever has occurred more recently
- Much faster because the volume of data backed up is restricted to the changed data only
- but takes longer to restore

Cumulative (differential) backup

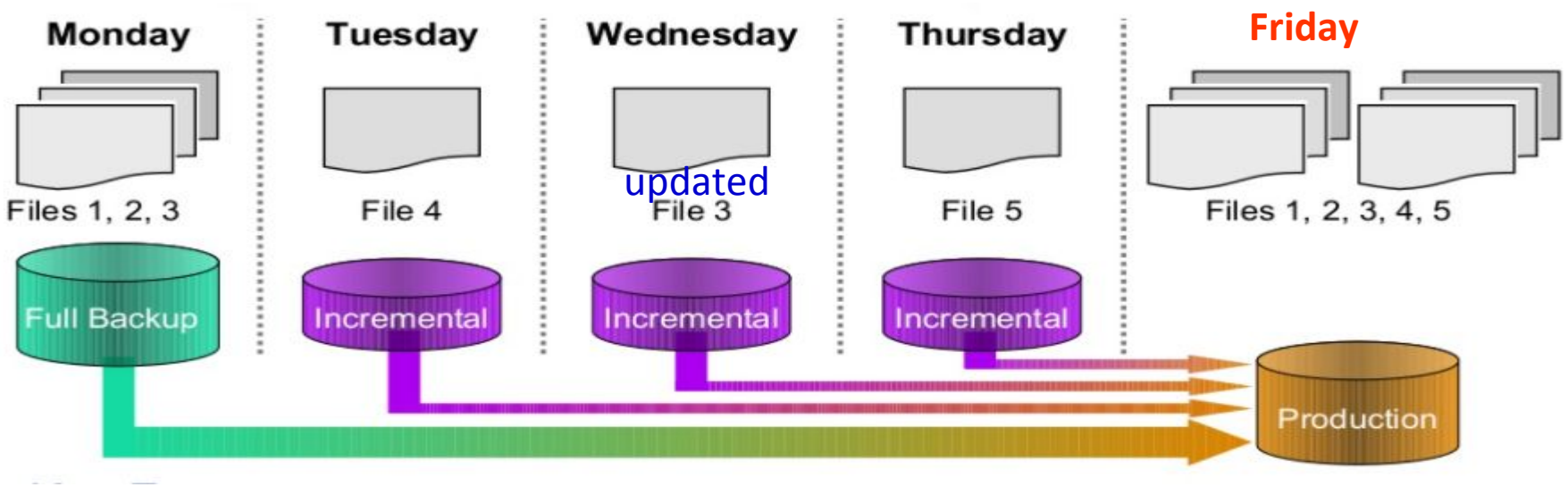
- Copies data that has changed since the last full backup
- Slower than incremental backup, but faster to restore

Synthetic (constructed) full backup

- A full backup generated from the latest full backup and all the incremental backups performed after that full backup
- Enables full backup copy to be done offline

- Restore operations vary with the granularity of the backup.
- A full backup provides a single repository from which the data can be easily restored.
- On Friday morning, data corruption occurs that requires data restoration using backup copies

Restoring from an incremental backup



- The process of restoration from an incremental backup requires the **last full backup and all the incremental** backups available **until the point of restoration.**

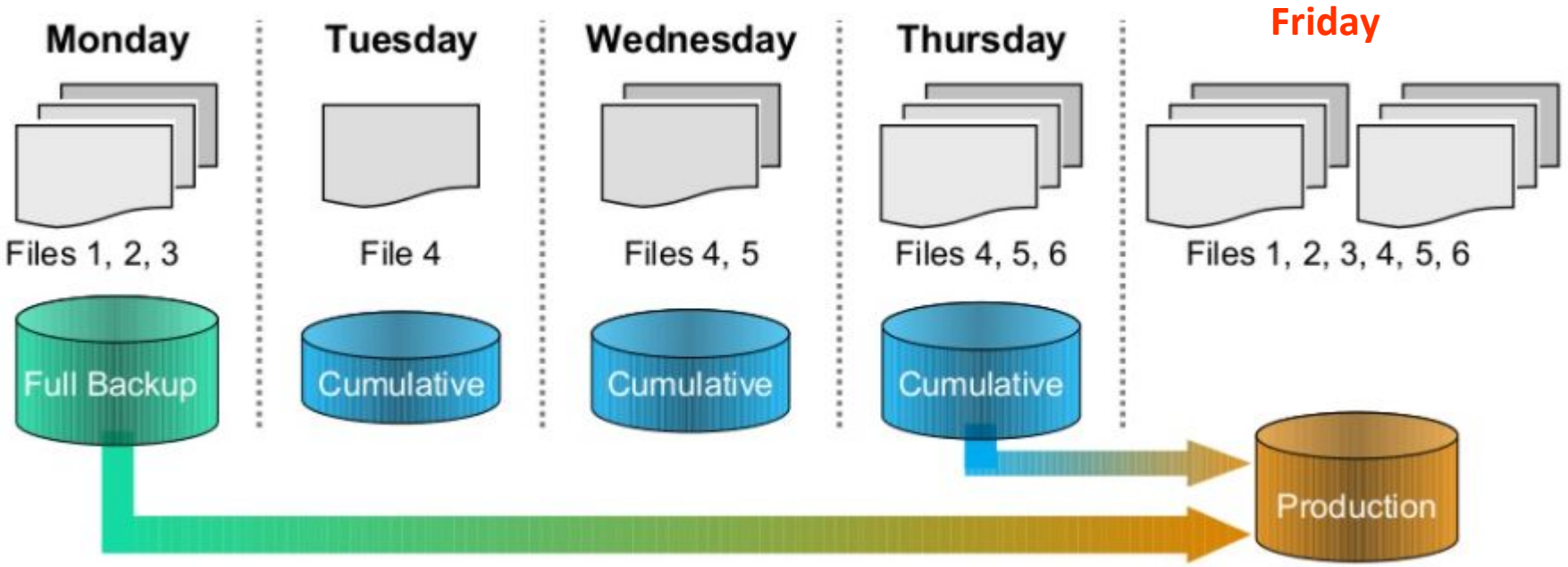
Key Features

– Files that have changed since the last full or incremental backup are backed up.

– Fewest amount of files to be backed up, therefore faster backup and less storage space.

– Longer restore because last full and all subsequent incremental backups must be applied.

Restoring a cumulative backup



- A restore from a cumulative backup requires the **last full backup and the most recent cumulative backup**.

Key Features

– More files to be backed up, therefore it takes more time to backup and uses more storage space.

– Much faster restore because only the last full and the last cumulative backup must be applied.

Recovery Considerations

Recovery Considerations

- The retention period is a key consideration for recovery. The retention period for a backup is derived from an RPO.
- For example, users of an application might request to restore the application data from its backup copy, which was created a month ago.
- This determines the retention period for the backup. Therefore, the minimum retention period of this application data is one month.
- However, the organization might choose to retain the backup for a longer period of time because of internal policies or external factors, such as regulatory directives.

Recovery Considerations

- If the recovery point is older than the retention period, it might not be possible to recover all the data required for the requested recovery point.
- Long retention periods can be defined for all backups, making it possible to meet any RPO within the defined retention periods.
- However, this requires a large storage space, which translates into higher cost.
- Therefore, while defining the retention period, analyze all the restore requests in the past and the allocated budget.

Recovery Considerations

- RTO relates to the time taken by the recovery process.
- To meet the defined RTO, the business may choose the appropriate backup granularity to minimize recovery time.
- A backup environment, RTO influences the type of backup media that should be used.
- For example, a restore from tapes takes longer to complete than a restore from disks.

SLO-1 :

1. **Backup Methods**
2. **Backup Architecture**

Backup Methods

Backup Methods

- Hot backup and cold backup are the two methods deployed for a backup.
- Both based on the state of the application when the backup is performed.
- In a hot backup, the application is up-and-running, with users accessing their data during the backup process.
  - it's is also referred to as an online backup.
- A cold backup requires the application to be shut down during the backup process.
  - Hence, this method is also referred to as an offline backup.

Hot backup

- The hot backup of online production data is challenging because data is actively used and changed.
- If a file is open,
  - it is normally not backed up during the backup process.
  - In such situations, an open file agent is required to back up the open file.
- Agents interact directly with the operating system or application and enable the creation of consistent copies of open files.
- The disadvantage associated with a hot backup is that the agents usually affect the overall application performance.

cold backup

- Consistent backups of databases can also be done by using a cold backup.
- This requires the database to remain inactive during the backup.
- The disadvantage of a cold backup is that the database is inaccessible to users during the backup process.
- A Point-In-Time (PIT) copy method is deployed in environments in which the impact of downtime from a cold backup or the performance impact resulting from a hot backup is unacceptable.
- The PIT copy is created from the production volume and used as the source for the backup.
- This reduces the impact on the production volume.

Bare-Metal Recovery (BMR)

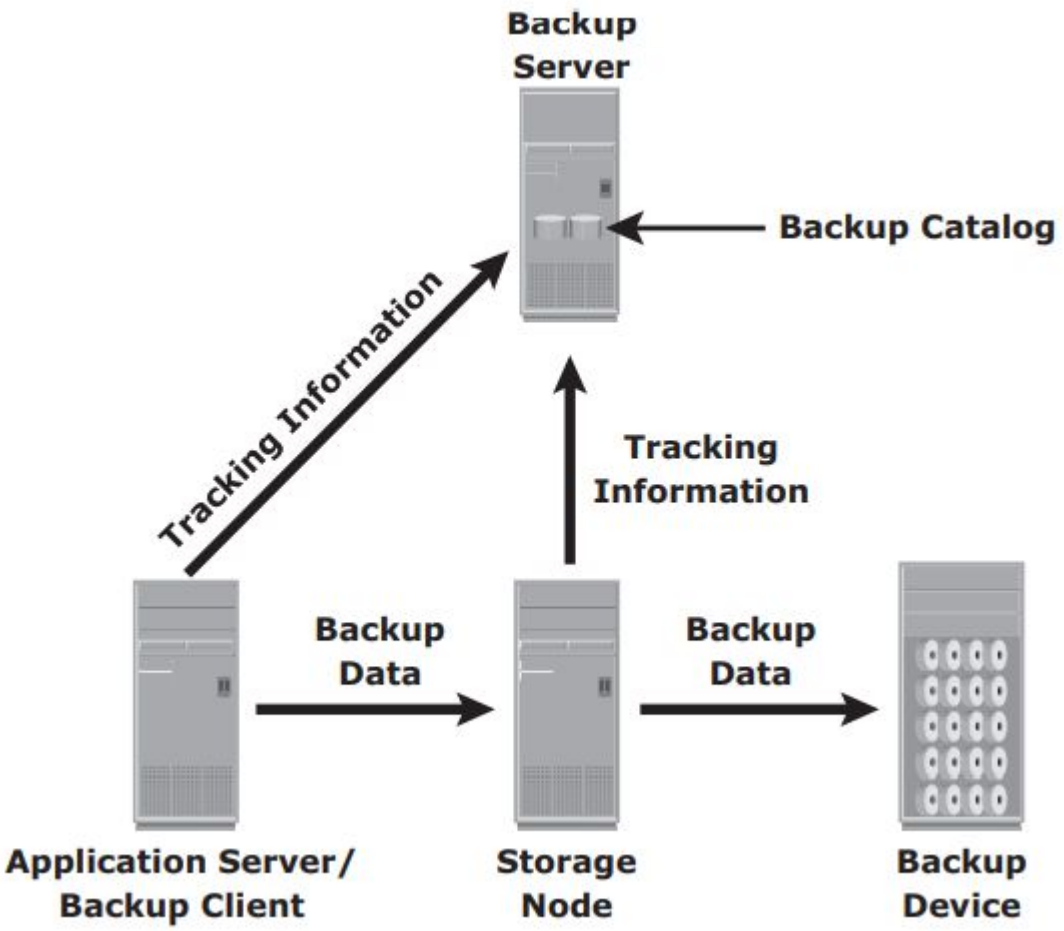
- In a disaster recovery environment, bare-metal recovery (BMR) refers,
  - to a backup in which all metadata, system information, and application configurations are appropriately backed up for a full system recovery.
- BMR builds the base system, which includes partitioning, the file system layout, the operating system, the applications, and all the relevant configurations.
- BMR recovers the base system first before starting the recovery of data files.
- Some BMR technologies
  - for example Server Configuration Backup (SCB)
    - can recover a server even onto dissimilar hardware.

Backup Architecture

Backup architecture

- A backup system commonly uses the client-server architecture with a backup server and multiple backup clients.
- The backup server manages,
  - backup operations and maintains the backup catalog, which contains information about the backup configuration,
    - contains information about when to run backups, which client data to be backed up, and so on
  - backup metadata
    - information about the backed up data

Backup architecture



Backup architecture

- The role of a backup client is to gather the data that is to be backed up and send it to the storage node.
- It also sends the tracking information to the backup server.
- Storage node is responsible for writing the data to the backup device

Backup architecture

- The storage node also sends tracking information to the backup server.
- In many cases, the storage node is integrated with the backup server, and both are hosted on the same physical platform.
- A backup device is attached directly or through a network to the storage node's host platform
- Some backup architecture refers to the storage node as the media server because it manages the storage device.

Backup architecture

- Backup software provides reporting capabilities based on the backup catalog and the log files.
- These reports include information, such as the amount of data backed up, the number of completed and incomplete backups, and the types of errors that might have occurred
- Reports can be customized depending on the specific backup software used.

SLO-2 :
Backup and Restore Operations

Backup Operation

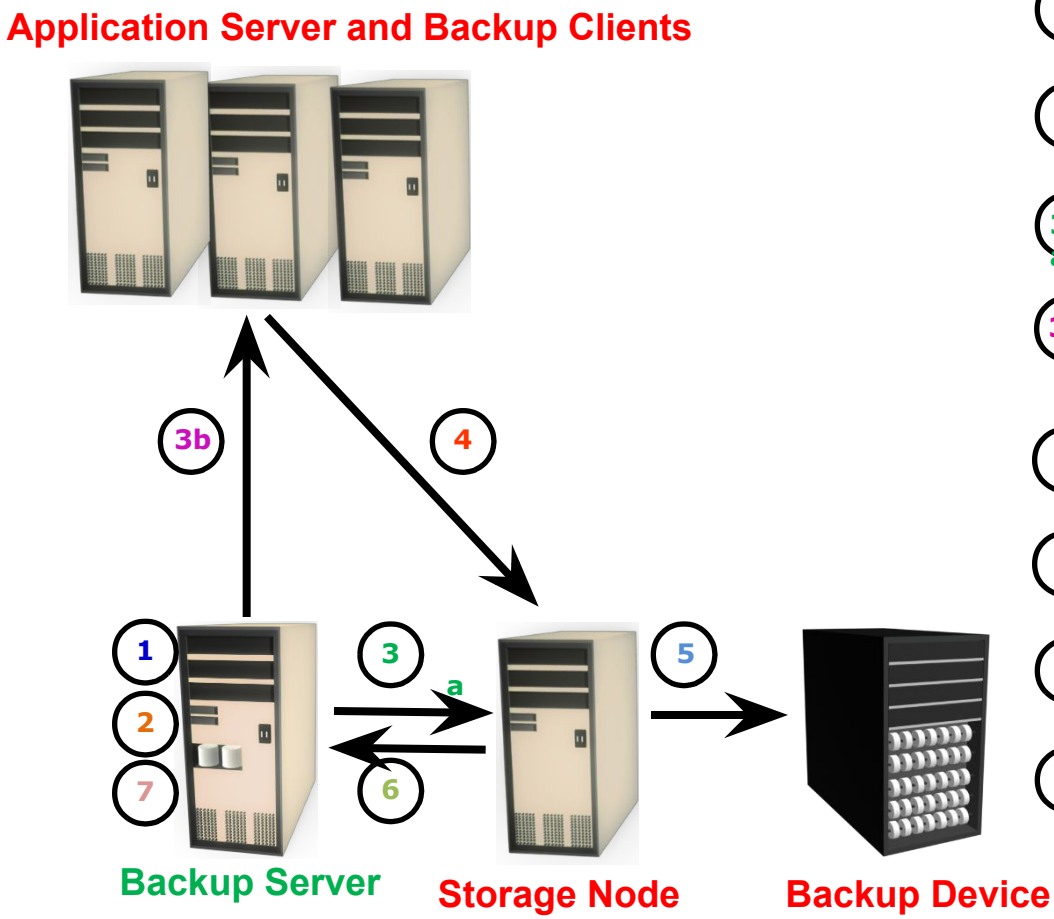
- When a backup operation is initiated,
  - network communication takes place between the different components of a backup infrastructure.
  - initiated by a server, but it can also be initiated by a client
- The backup server initiates,
  - backup process for different clients based on the backup schedule configured for them.
  - For example, the backup for a group of clients may be scheduled to start at 11:00 p.m. every day

Backup Operation

- The backup server maintains the information about,
  - backup clients to be backed up and storage nodes to be used in a backup operation.
- The backup server retrieves,
  - the backup-related information from the backup catalog and,
  - based on this information, instructs the storage node to load the appropriate backup media into the backup devices.
  - Simultaneously, it instructs the backup clients to gather the data to be backed up and send it over the network to the assigned storage node.

Backup Operation

- The backup server coordinates the backup process with all the components in a backup environment



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 send its metadata to the 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 media information to backup server
7. Backup server update catalog and records the status

Backup Operation

- The client sends some backup metadata (the number of files, name of the files, storage node details, and so on) to the backup server.
- The storage node receives the client data, organizes it, and sends it to the backup device.
- The storage node then sends additional backup metadata (location of the data on the backup device, time of backup, and so on) to the backup server.
- The backup server updates the backup catalog with this information

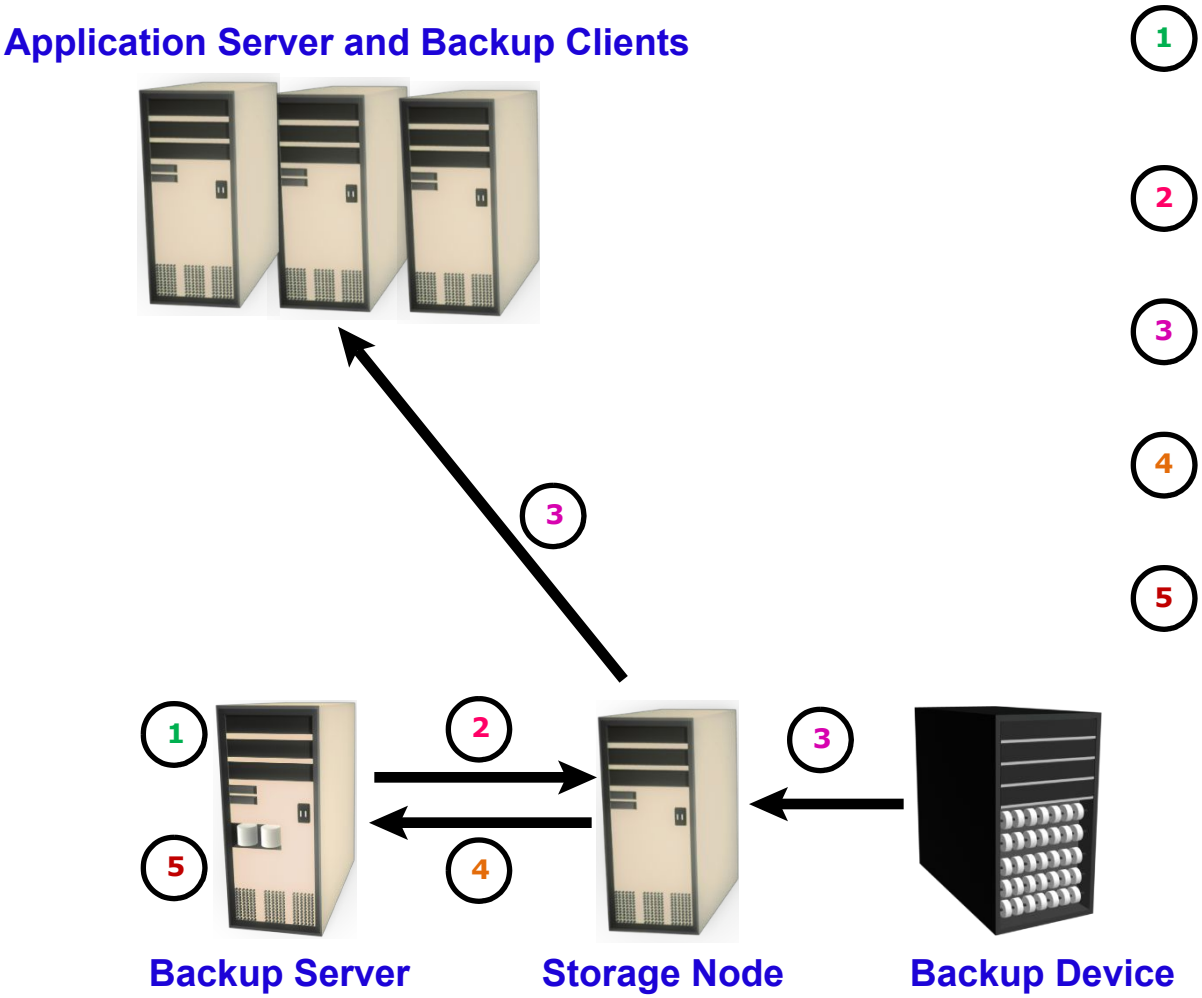
Restore Operation

- After the data is backed up, it can be restored when required.
- A restore process must be manually initiated from the client.
- Some backup software has a separate application for restore operations.
- These restore applications are usually accessible only to the administrators or backup operators.

Restore Operation

- Upon receiving a restore request,
  - an administrator opens the restore application to view the list of clients that have been backed up.
  - While selecting the client for which a restore request has been made,
  - the administrator also needs to identify the client that will receive the restored data.
- Data can be restored on the same client for whom the restore request has been made or on any other client.
- The administrator then selects the data to be restored and the specified point in time to which the data has to be restored based on the RPO.
- Because all this information comes from the backup catalog, the restore application needs to communicate with the backup server.

Restore Operation



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

SLO-1 :
Backup Topologies

Lesson: Backup/Recovery Topologies & Technologies

Upon completion of this lesson, you be able to:

- Describe backup topologies
  - Direct backup
  - LAN and LAN free backup
  - Mixed backup
- Detail backup in NAS environment
- Describe backup technologies
  - Backup to tape
  - Backup to disk
  - Backup to virtual tape

Backup Topologies

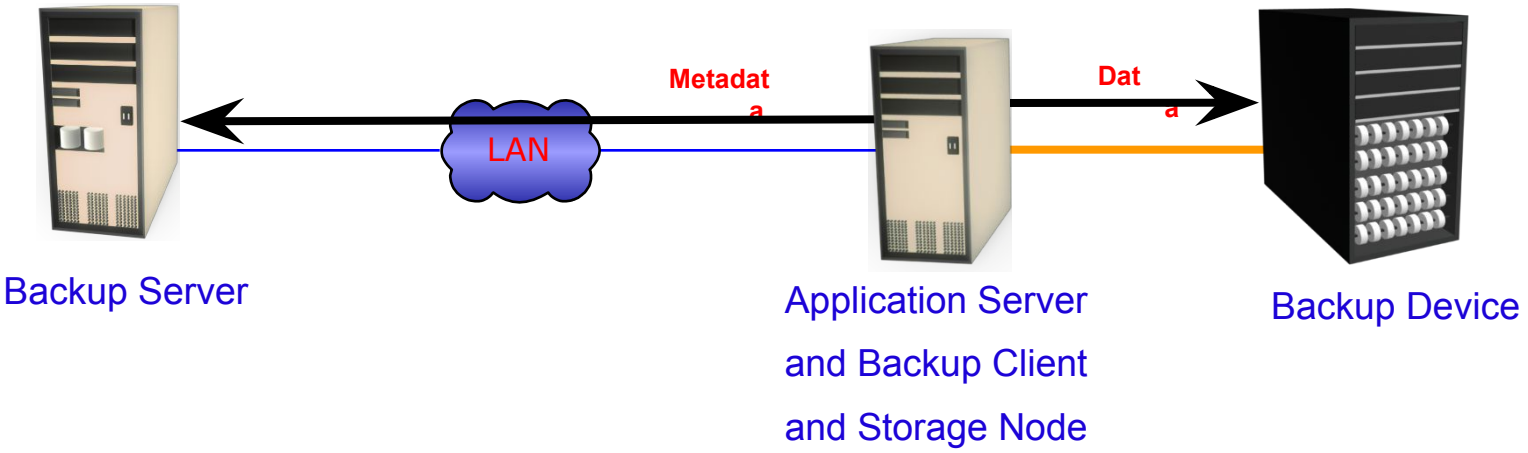
- There are 3 basic backup topologies:
  - Direct Attached Based Backup
  - LAN Based Backup
  - SAN Based Backup (LAN free Backup)

  A mixed topology is also used by combining LAN-based and SAN-based topologies.

Direct Attached Backups

- In a direct-attached backup,
  - the storage node is configured on a backup client,
  - the 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 there will be a need,
  - 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.

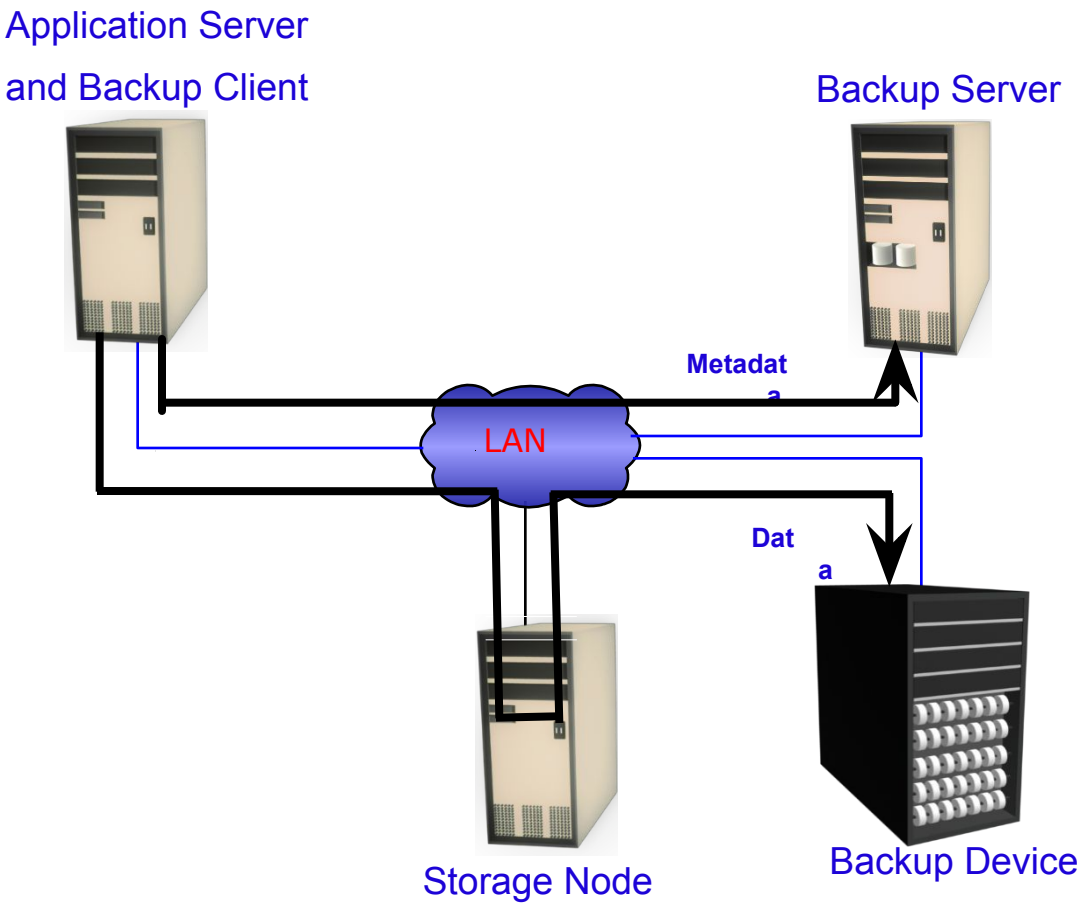
Direct Attached Backups



LAN Based Backups

- In a LAN-based backup,
  - the clients, backup server, storage node, and backup device are connected to the LAN.
- 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.
- This impact can be minimized by adopting a number of measures,
  - such as configuring separate networks for backup
  - installing dedicated storage nodes for some application servers.

LAN Based Backups



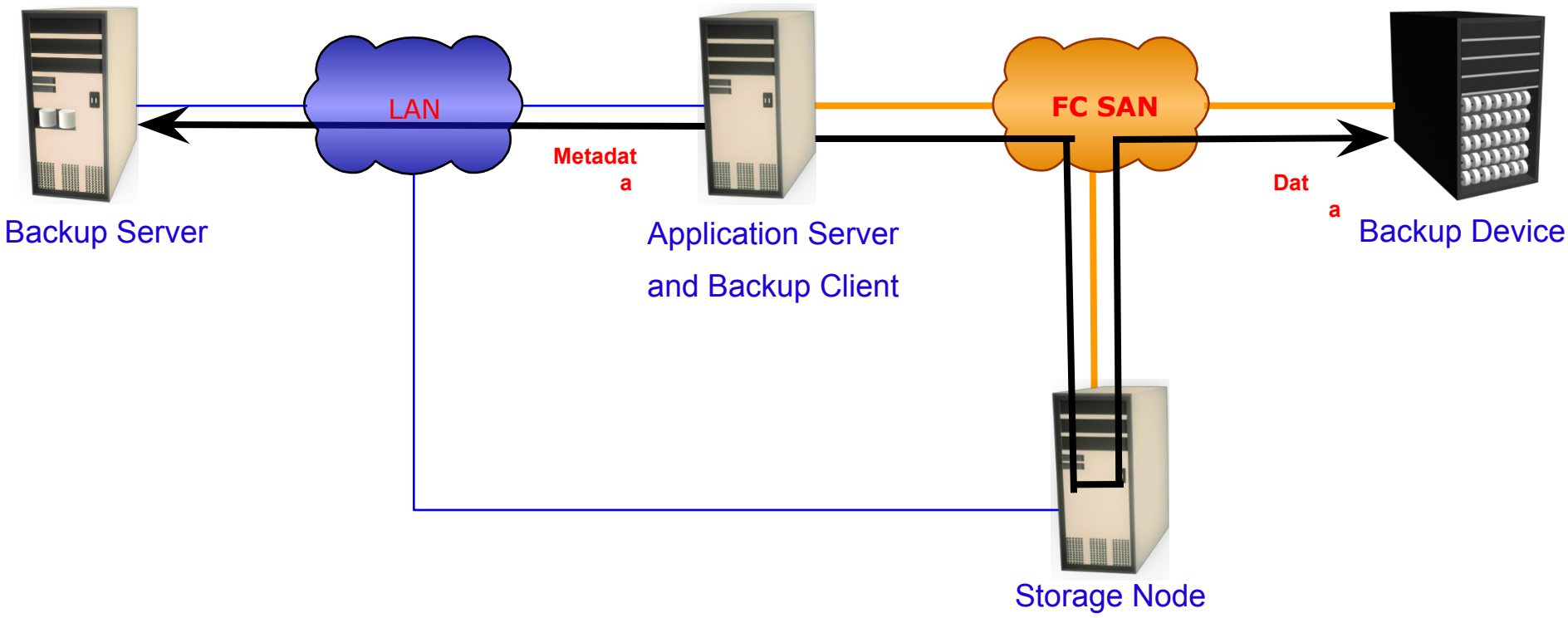
SAN Based Backups (LAN Free)

- The SAN-based backup topology,
  - most appropriate solution when a backup device needs to be shared among clients.
  - the backup device and clients are attached to the SAN.
- 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.

SAN Based Backups (LAN Free)

- 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.

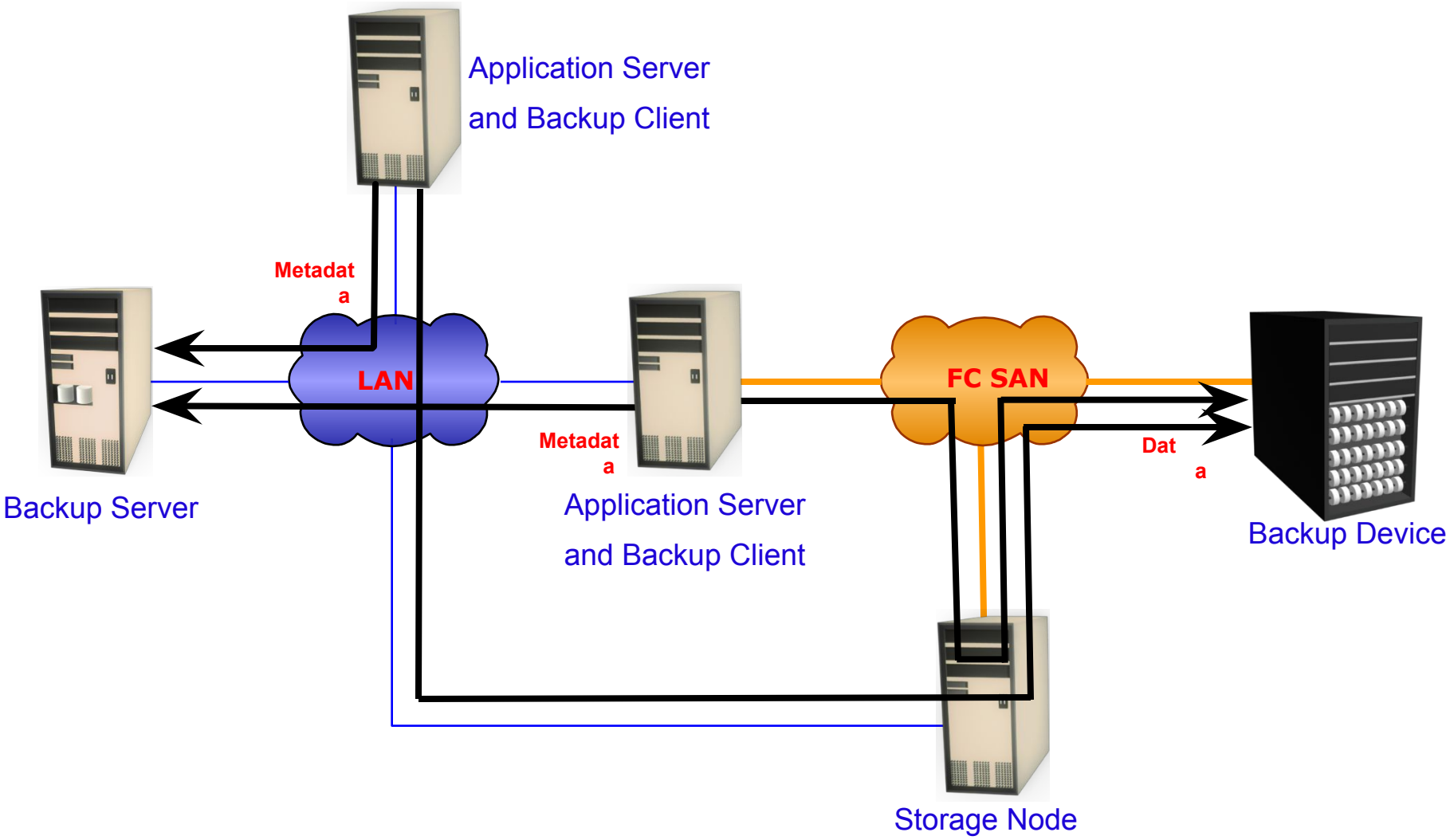
SAN Based Backups (LAN Free)



Mixed Backup

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

Mixed Backup



SLO-2 :
Backup in NAS Environments

Backup in NAS Environments

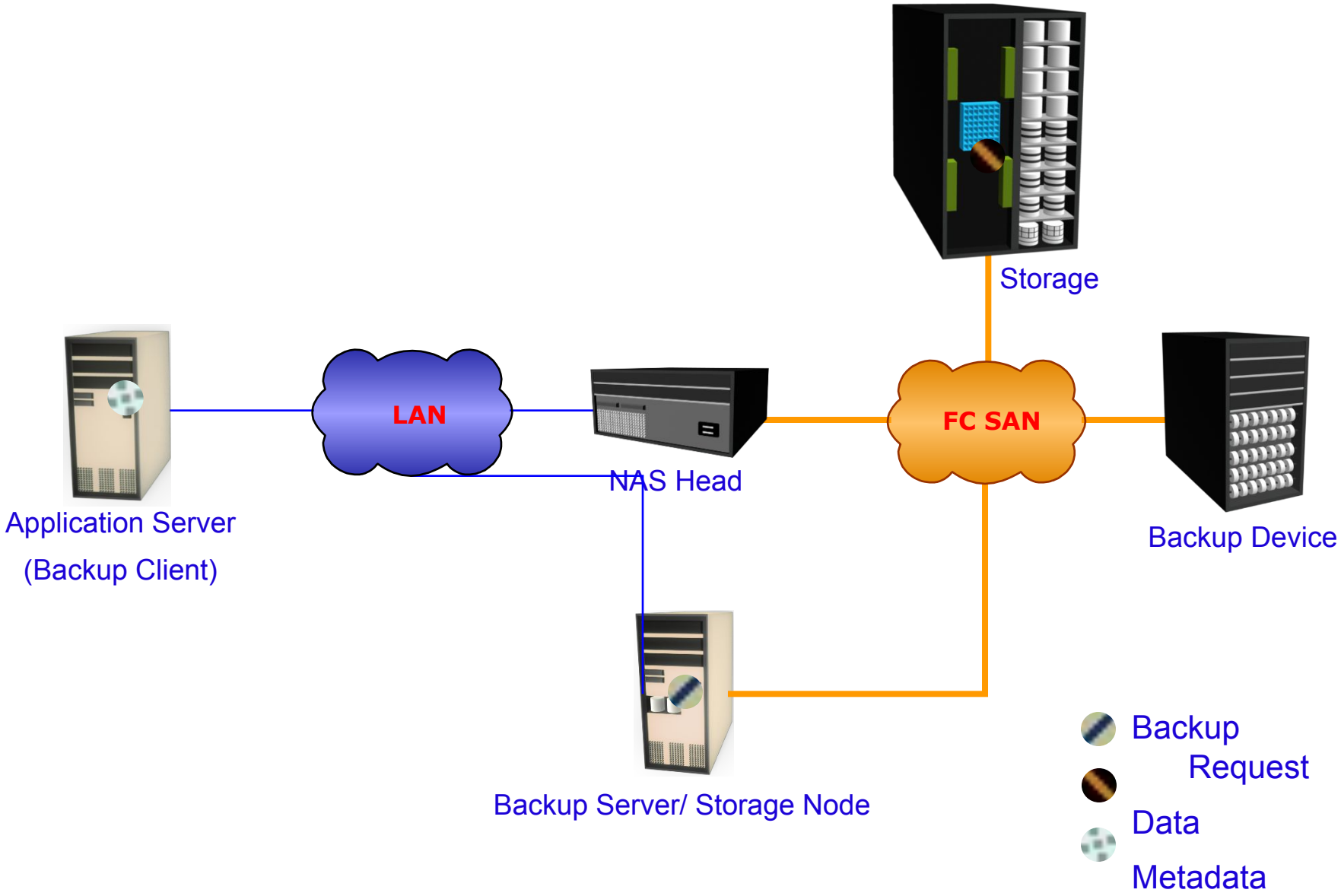
- The use of a NAS head imposes a new set of considerations on the backup and recovery strategy in NAS environments.
- NAS heads use a proprietary operating system and file system structure that supports multiple file-sharing protocols.
- In the NAS environment, backups can be implemented in different ways:
  1. Server based
  2. Serverless
  - or using Network Data Management Protocol (NDMP).
  - Common implementations are NDMP 2-way and NDMP 3-way.

Backup in NAS Environment – Server Based

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

  (storage array NAS head backup client)
- The backup client sends this data to the storage node, which in turn writes the data to the backup device.
  - backup client(data) storage node backup device
- This results in overloading the network with the backup data and using application server resources to move the backup data.

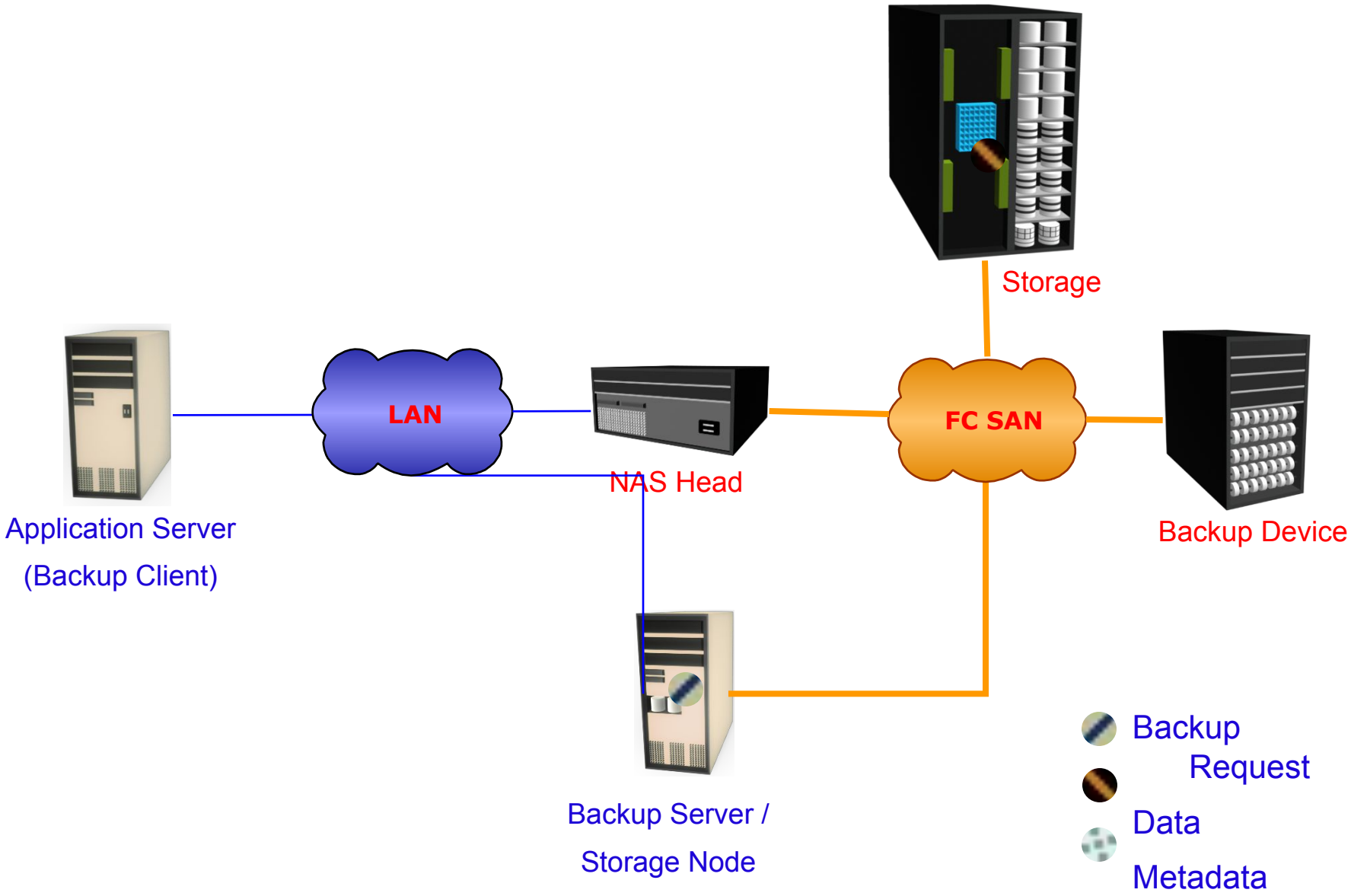
Backup in NAS Environment – Server Based



Backup in NAS Environment – Serverless

- In a 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 application server.
- In this scenario, the storage node, which is also a backup client,
  - reads the data from the NAS head and writes it to the backup device without involving the application server.
- Compared to the previous solution, this eliminates one network hop.

Backup in NAS Environment – Serverless



Network Data Management Protocol(NDMP)-Based Backup

- NDMP is an industry-standard TCP/IP-based protocol specifically designed for a backup in a NAS environment.
- It communicates with several elements in the backup environment (NAS head, backup devices, backup server, and so on) for data transfer.
  - enables vendors to use a common protocol for the backup architecture.
- Data can be backed up using NDMP regardless of the operating system or platform.
- Due to its flexibility, it is no longer necessary to transport data through the application server, which reduces the load on the application server and improves the backup speed.

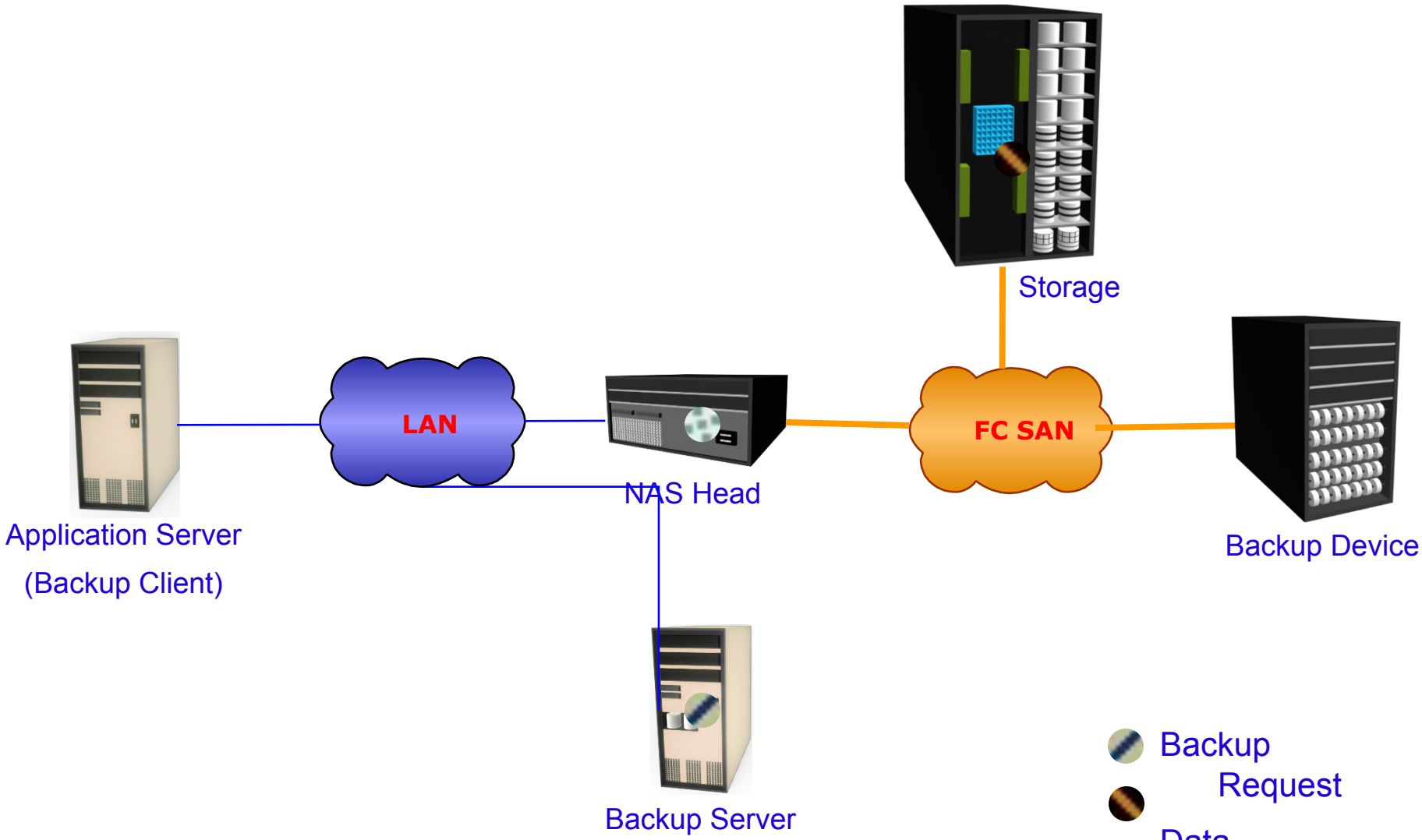
NDMP-Based Backup

- NDMP optimizes backup and restore by leveraging the high-speed connection between the backup devices and the NAS head.
- In NDMP, backup data is sent directly from the NAS head to the backup device, whereas metadata is sent to the backup server.

  data(NAS head → the backup device)

  metadata → backup server
- In this model, network traffic is minimized by isolating data movement from the NAS head to the locally attached backup device.
- Only metadata is transported on the network.

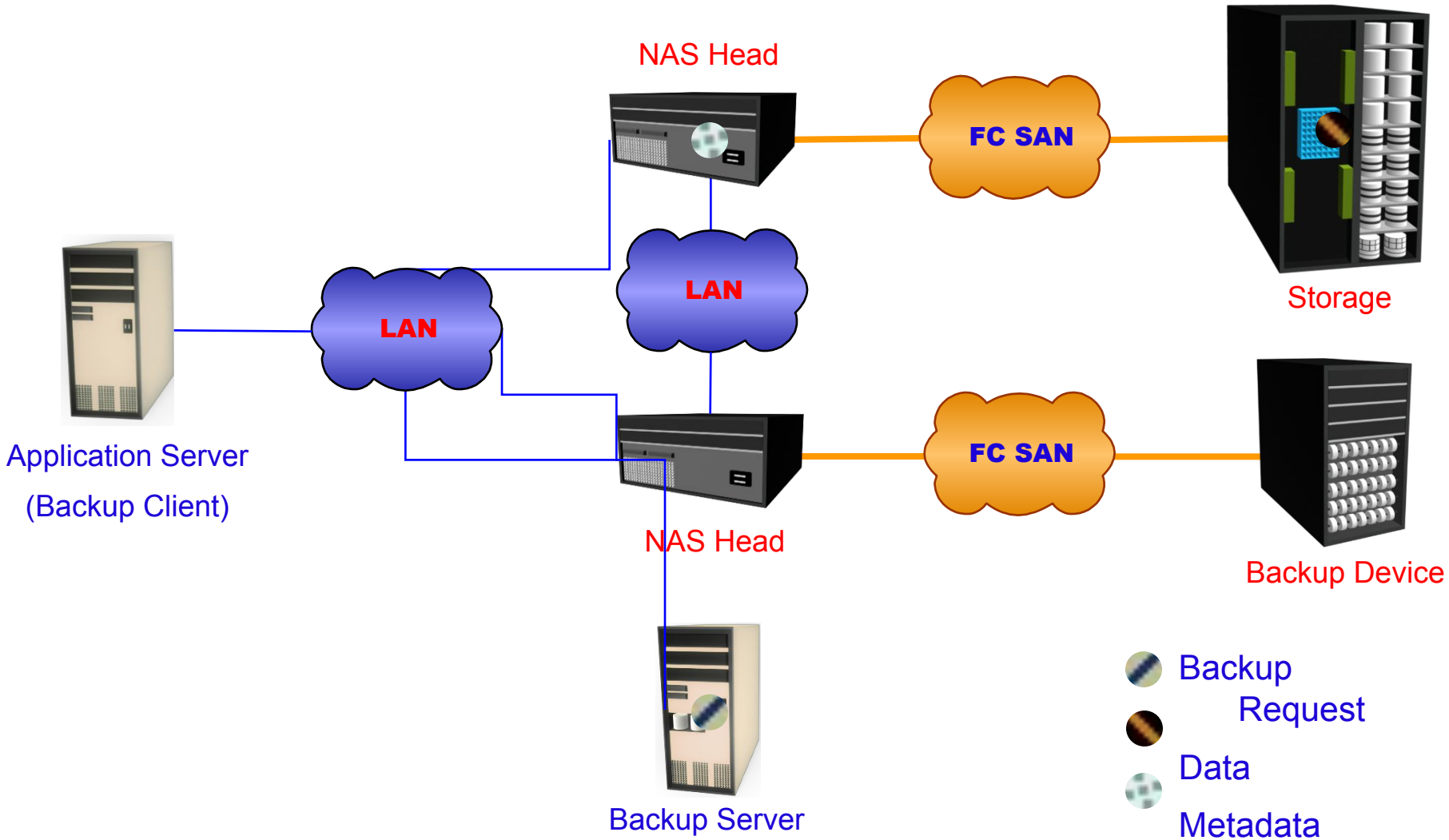
Backup in NAS Environment – NDMP 2-way



Backup in NAS Environment – NDMP 2-way

- In the NDMP 3-way method, a separate private backup network must be established between all NAS heads and the NAS head connected to the backup device.
- Metadata and NDMP control data are still transferred across the public network.
- An NDMP 3-way is useful when backup devices need to be shared among NAS heads.
- It enables the NAS head to control the backup device and share it with other NAS heads by receiving the backup data through the NDMP.

Backing up a NAS Device – NDMP 3-way



SLO-1 :

1. **Backup Targets**
2. **Data De-duplication for Backup**

Backup Targets

Backup Targets

- A wide range of technology solutions are currently available for backup targets.
- Two most commonly used backup targets
  - Tape and disk libraries
- In the past, tape technology was the predominant target for backup due to its low cost.
- A Virtual Tape Library (VTL) is one of the options that uses disks as a backup medium.
- VTL emulates tapes and provides enhanced backup and recovery capabilities.

Backup Technology options

1. Backup to Tape
   - Physical tape library
2. Backup to Disk
3. Backup to virtual tape
   - Virtual tape library

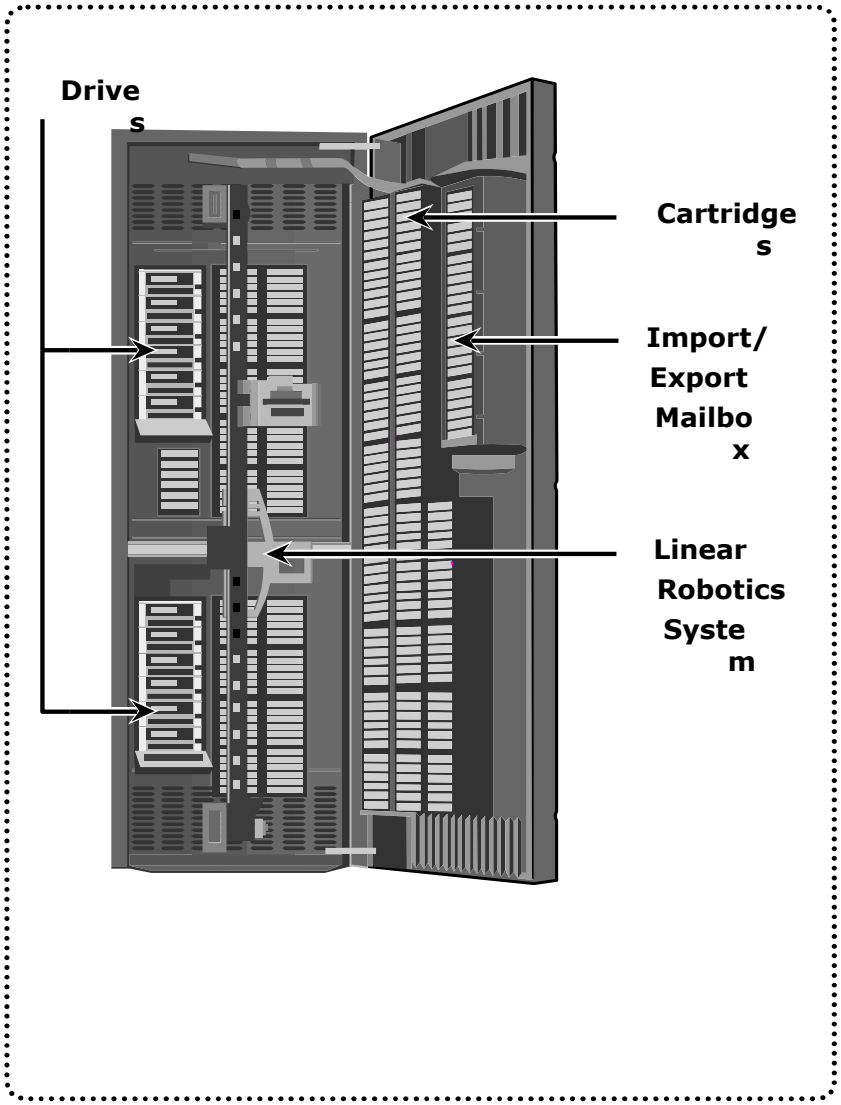
Backup to Tape

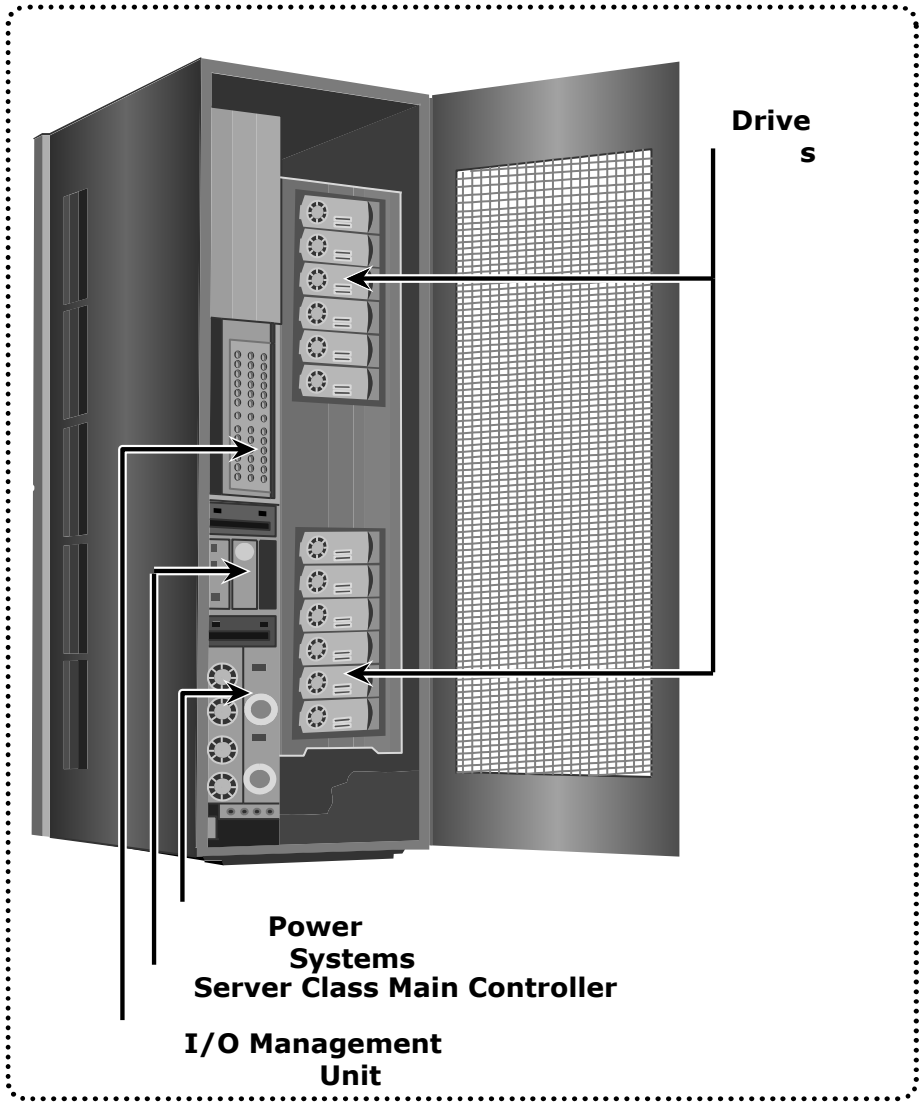
- Traditionally low cost solution
- Tape drives are used to read / write data from / to a tape cartridge (or cassette).
- Sequential / linear access
- Tape mounting is the process of inserting a tape cartridge into a tape drive.
- Several types of tape cartridges are available.
  - They vary in size, capacity, shape, density, tape length, tape thickness, tape tracks, and supported speed.

Physical Tape Library

- Provides housing and power for a large 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 between cartridge slots and tape drives.
- Mail or import/export slots are used to add or remove tapes from the library without opening the access doors (refer : Front View).

Physical Tape Library...





Tape drives are referred to as sequential, or linear, access devices because the data is written or read sequentially

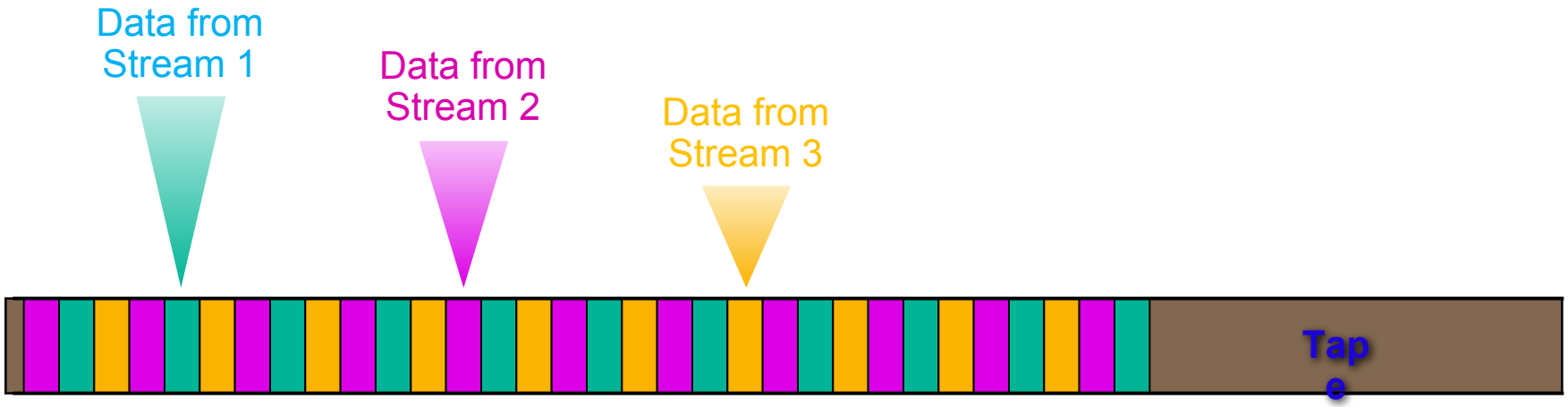
Physical Tape Library...

- When a backup process starts,
  - The robotic arm is instructed to load a tape to a tape drive. This process adds delay, but it generally takes 5 to 10 seconds to mount a tape.
  - After the tape is mounted, additional time is spent to position the heads and validate header information.
  - This total time is called load to ready time, and it can vary from several seconds to minutes.
- The tape drive receives backup data and stores the data in its internal buffer. This backup data is then written to the tape in blocks.

- The speed of the tape drives can also be adjusted to match data transfer rates.

Physical Tape Library… : Multiple streams on tape media

- Multiple streaming to improve media performance
  - Write data from multiple screams on a single tape



- But it has an associated disadvantage.
  - The backup data is interleaved because data from multiple streams is written on it.
  - Consequently, the data recovery time is increased because all the extra data from the other streams must be read and discarded while recovering a single stream.

Physical Tape Library…

- Many times, even the buffering and speed adjustment features of a tape drive fail to prevent the gaps, causing the "shoe shining effect" or "backhitching."
- Shoe shining is the repeated back and forth motion a tape drive makes when there is an interruption in the backup data stream.
  - For example, if a storage node sends data slower than the tape drive writes it to the tape, the drive periodically stops and waits for the data to catch up.
- After the drive determines that there is enough data to start writing again, it rewinds to the exact place where the last write took place and continues.
- This repeated back-and-forth motion not only causes a degradation of service, but also excessive wear and tear to tapes.

Physical Tape Library...

- When the tape operation finishes,
  - the tape rewinds to the starting position and it is unmounted.
  - The robotic arm is then instructed to move the unmounted tape back to the slot.
  - When a restore is initiated, the backup software identifies which tapes are required.
  - The robotic arm is instructed to move the tape from its slot to a tape drive.
- If the required tape is **not found** in the tape library, the backup software **displays a message**, instructing the operator to **manually insert** the required tape in the tape library.
- When a file or a group of files require restores, the tape must move to that file location sequentially before it can start reading. This process can take a significant amount of time, especially if the required files are recorded at the end of the tape.

Tape Limitations

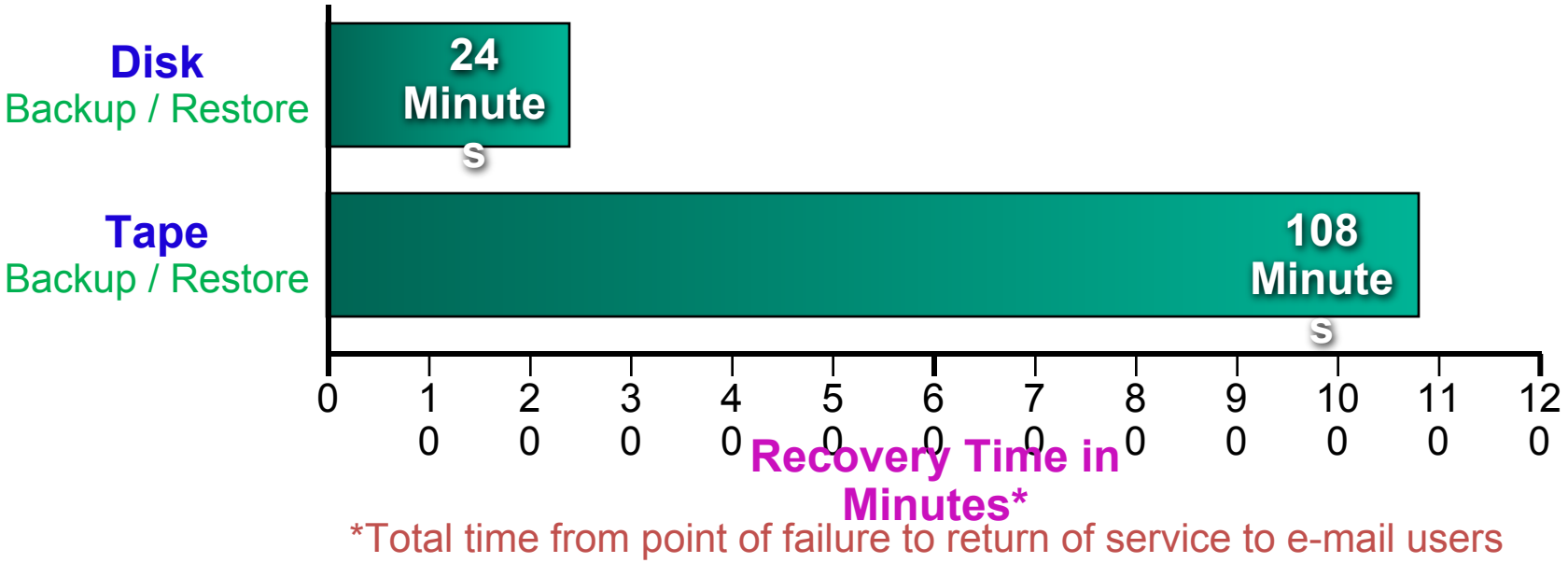
- Reliability
  - Restore performance
    - Mount, load to ready, rewind, dismount times
- Backup and recovery operation are slow due to sequential access
- Cannot be accessed by multiple hosts simultaneously
- Controlled environment for tape storage
- Wear and tear of tape
- Caused “show shining effect” or “backhitching”
- Shipping/handling challenges
- Tape management challenges(Physical transportation)

Backup to Disk

- Because of increased availability, low cost disks have now replaced tapes as the primary device for storing backup data because of their performance advantages.
- Backup-to-disk systems offer,
  - Ease of implementation
  - Reduced TCO (Total Cost of Ownership)
  - Improved quality of service.
  - Data transfer rates, disks also offer faster recovery when compared to tapes.
  - Clear advantages due to their inherent random access and raid-protection capabilities.
  - Enhances backup performance (Backup to disk is used as a staging area where the data is copied temporarily before transferring or staging it to tapes)
- Some backup products enables a much faster restore(ie. images to remain on the disk for a period of time even after they have been staged)

Tape versus Disk – Restore Comparison

Microsoft Exchange environment



*Total time from point of failure to return of service to e-mail users

Typical Scenario:

- 800 users, 75 MB mailbox
- 60 GB database

Source: EMC Engineering and EMC IT

Backup to Disk

- Recovering from a full backup copy stored on disk and kept onsite provides the fastest recovery solution.
- Using a disk enables the creation of full backups more frequently, which in turn improves RPO(Recovery-Point Objective) and RTO Recovery-Time Objective.

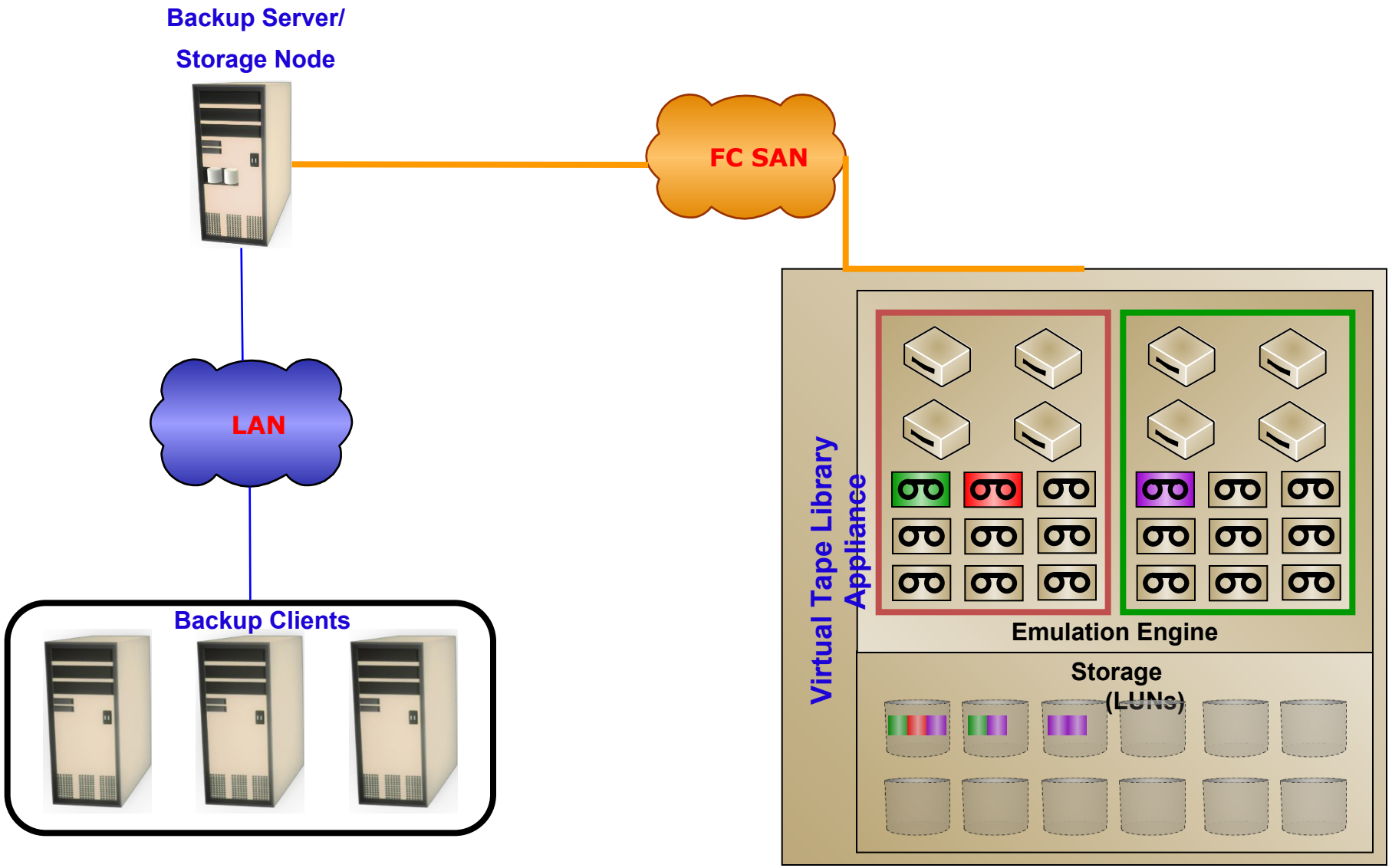
Backup to Virtual Tape

- Disks are emulated and presented as tapes to backup software
- Does not require any additional modules or changes in the legacy backup software
- Provides better single stream performance and reliability over physical tape
- Online and random disk access
  - Provides faster backup and recovery

Virtual Tape Library

- A virtual tape library (VTL) has the same components as that of a physical tape library, except that the majority of the components are presented as virtual resources.
- For the backup software, there is no difference between a physical tape library and a virtual tape library.
- 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 space on a LUN (Logical Unit Number)
- A virtual tape can span multiple LUNs if required.

Virtual Tape Library...



Virtual Tape Library…

- Similar to a physical tape library, a robot mount is virtually performed when a backup process starts in a virtual tape library.
- However, unlike a physical tape library, where this process involves some mechanical delays, in a virtual tape library it is almost instantaneous.
- Even the load to ready time is much less than in a physical tape library

Virtual Tape Library…

- After the virtual tape is mounted and the virtual tape drive is positioned, the virtual tape is ready to be used, and backup data can be written to it.
- Data is written to the virtual tape immediately.
- Unlike a physical tape library, the virtual tape library is not constrained by the sequential access and shoe shining effect.
- When the operation is complete, the backup software issues a rewind command.
- This rewind is also instantaneous.
- The virtual tape is then unmounted, and the virtual robotic arm is instructed to move it back to a virtual slot.

Virtual Tape Library…

- The steps to restore data are similar to those in a physical tape library, but the restore operation is nearly instantaneous.
- Even though virtual tapes are based on disks, which provide random access, they still emulate the tape behavior.

Virtual Tape Library…

- A virtual tape library appliance offers a number of features,
  - Some virtual tape libraries offer multiple emulation engines configured in an active cluster configuration.
- One engine can pick up the virtual resources from another engine in the event of any failure and enable the clients to continue using their assigned virtual resources transparently.

Virtual tapes : Advantages

- Using virtual tapes offers several advantages over both physical tapes and disks.
  - virtual tapes offer better single stream performance,
  - better reliability
  - random disk access characteristics.
  - always online and provide faster backup and recovery.
  - does not require the usual maintenance
  - easy installation and administration (preconfigured by the manufacturer)

Tape Versus Disk Versus Virtual Tape

FEATURES	Tape	Disk-Aware Backup-to-Disk	Virtual Tape
Offsite Capabilities	Yes	No	Yes
Reliability	No inherent protection methods	RAID, spare	RAID, spare
Performance	Subject to mechanical operations, load times	Faster single stream	Faster single stream
Use	Backup only	Multiple (backup/production)	Backup only

Data De-duplication for Backup

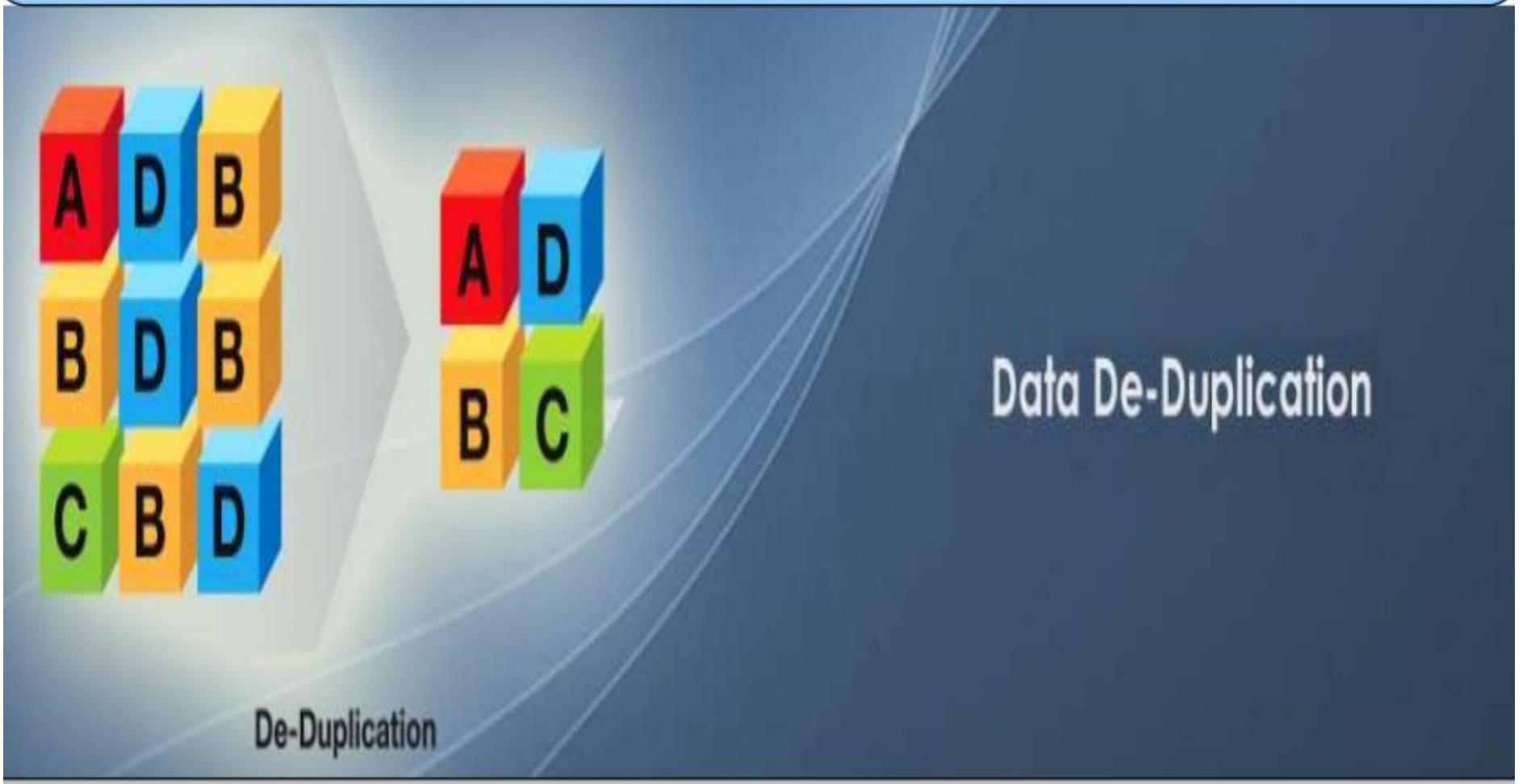
Introduction : Data De-duplication for Backup

- Traditional backup processes back up a lot of duplicate data, but do not provide any inherent capability to prevent duplicate data from being backed up.
- Backing up duplicate data significantly increases,
  - backup window size requirements
  - unnecessary consumption of resources
    - storage space and network bandwidth

Introduction : Data De-duplication for Backup…

- 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
  - shorten the backup window
  - remove the network burden.
  - to store more backups on the disk
  - retain the data on the disk for a longer time.

What Is Data Deduplication Technology?



Data deduplication is a technique for reducing the amount of storage space an organization needs to save its data.

Activate Windows
Go to Settings to activ

Data Deduplication Methods

There are two methods of deduplication:

1. **File level**
2. **Subfile level**

- Either method offers benefits; however, results can vary.

File-level deduplication (single-instance storage)

- Detects and removes redundant copies of identical files.
- It enables storing only one copy of the file;
  - the subsequent copies are replaced with a pointer that points to the original file.
- Simple and fast but does not address the problem of duplicate content inside the files
- For example, two 10-MB PowerPoint presentations with a difference in just the title page are not considered as duplicate files, and each file will be stored separately

Subfile deduplication

- breaks the file into smaller chunks and then uses a specialized algorithm to detect redundant data within and across the file.
- As a result, subfile deduplication eliminates duplicate data across files.
- There are two forms of subfile deduplication:
  - Fixed-length block
  - Variable-length segment

Fixed-length Block Deduplication

- Divides the files into fixed length blocks and uses a hash algorithm to find the duplicate data.
- Although simple in design, fixed-length blocks might miss many opportunities to discover redundant data because the block boundary of similar data might be different.
- Consider the addition of a person's name to a document's title page.
- This shifts the whole document, and all the blocks appear to have changed, causing the failure of the deduplication method to detect equivalencies.

Variable-length segment deduplication

- In variable-length segment deduplication, if there is a change in the segment, the boundary for only that segment is adjusted, leaving the remaining segments unchanged.
- This method vastly improves the ability to find duplicate data segments compared to fixed-block.

Data Deduplication Implementation

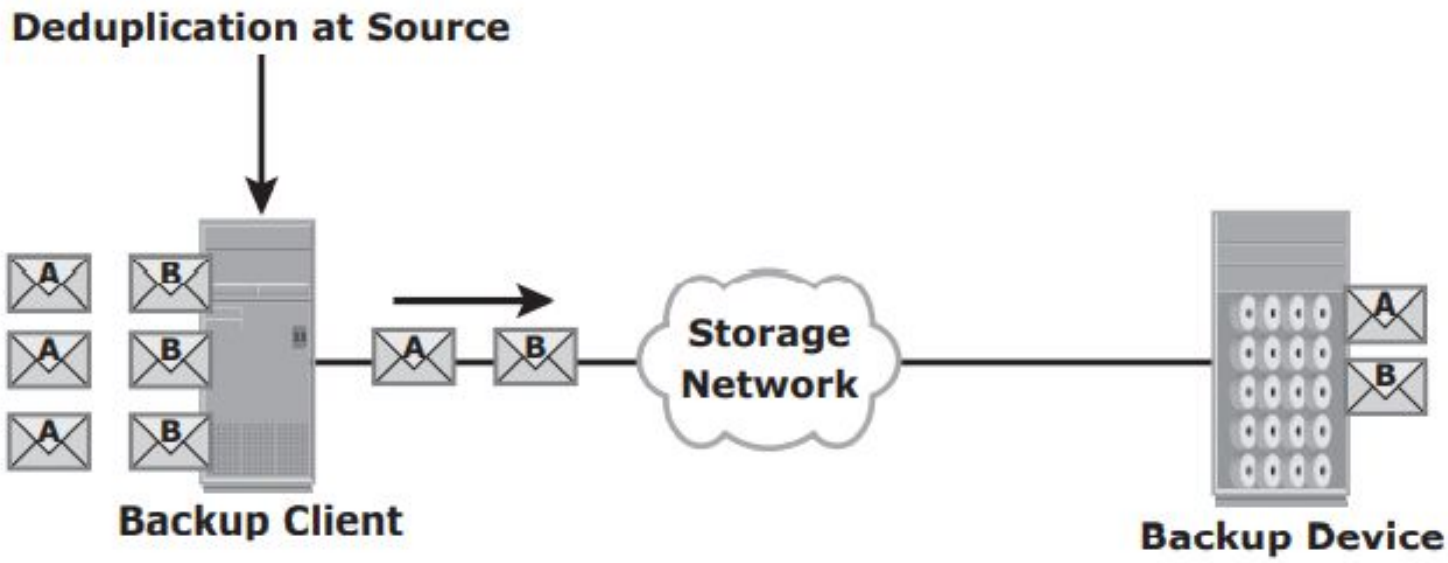
Deduplication for backup can happen at the data source or the backup target,

1. Source-Based Data Deduplication
2. Target-Based Data Deduplication

Source-Based Data Deduplication

- Eliminates redundant data at the source before it transmits to the backup device.
  - 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.

Source-Based Data Deduplication

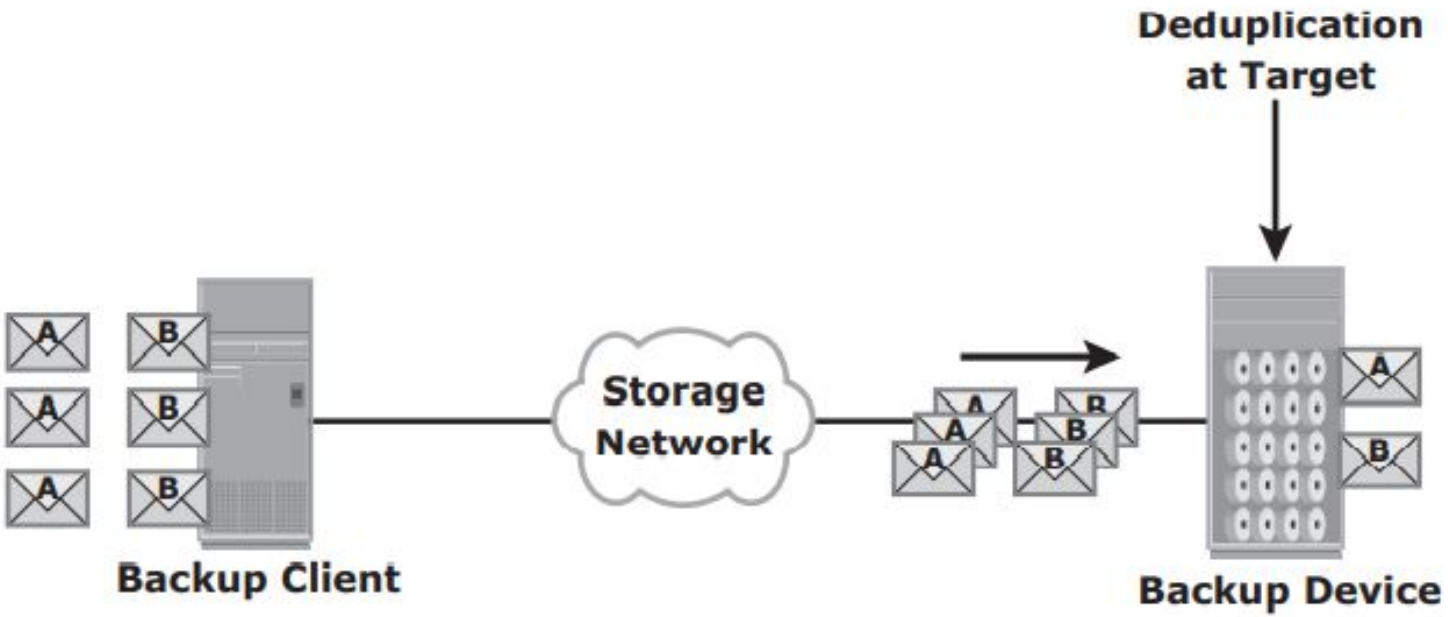


- Increases the overhead on the backup client, which impacts the performance of the backup and application running on the client.
- Source-based deduplication might also require a change of backup software if it is not supported by backup software.

Target-Based Data Deduplication

- 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.
- 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.

Target-Based Data Deduplication



Inline process deduplication

- Inline deduplication performs deduplication on the backup data before it is stored on the backup device.
  - reduces the storage capacity needed for the backup.
  - introduces overhead in the form of the time required to identify and remove duplication in the data.
  - 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.
- Suitable for situations with tighter backup windows
- Requires more storage capacity to store the backup images before they are deduplicated.

Benefits of Data Deduplication

- It reduces the amount of disk or tape that organizations need to buy.
- It can reduce storage requirements up to 95 percent.
- It can reduce the amount of network bandwidth required for backup processes.
- It can speed up the backup and recovery process.
- It can save money and time.

Activate Windo

SLO – 2 :
Backup in Virtualized Environments

Backup in Virtualized Environments

- In a virtualized environment, it is imperative to back up the virtual machine data (OS, application data, and configuration) to prevent its loss or corruption due to human or technical errors.
- Two approaches for performing a backup in a virtualized environment:
  1. **The traditional backup approach**
  2. **The image-based backup approach**

Traditional Backup Approaches

- Backup agent on VM
  - Requires installing a backup agent on each VM running on a hypervisor
  - Can only backup virtual disk data
  - Does not capture VM files such as VM swap file, configuration file
  - Challenge in VM restore
- Backup agent on Hypervisor
  - Requires installing backup agent only on hypervisor
  - Backs up all the VM files

Backup agent runs on each VM

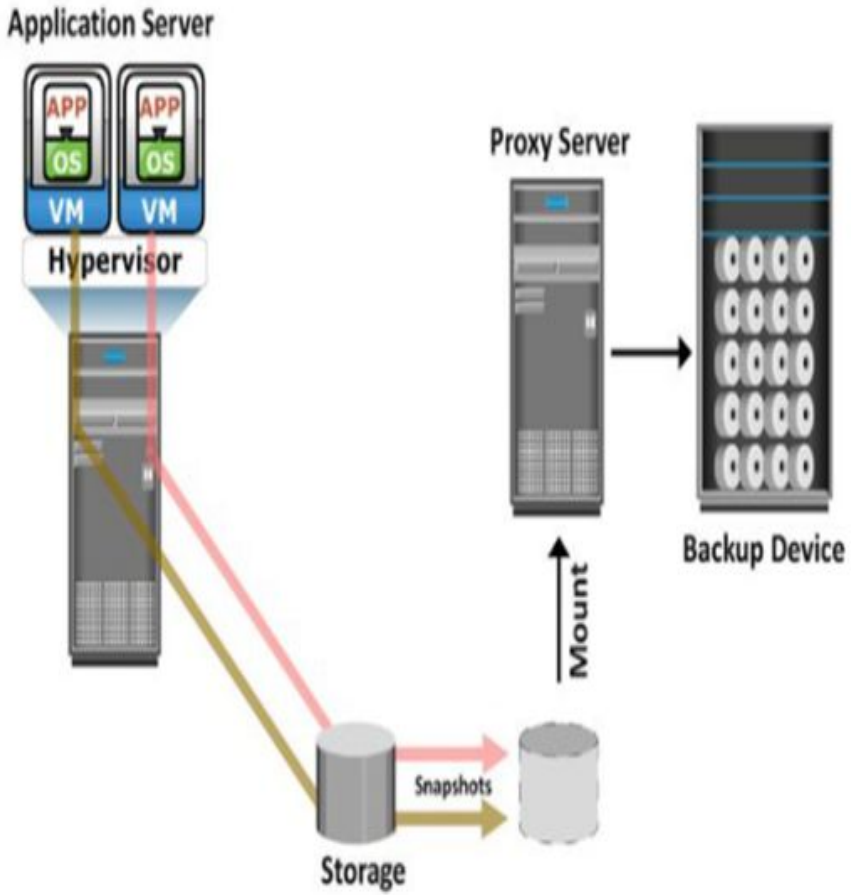
Backup agent runs on Hypervisor

= Backup Agent

Note: Hypervisor is a software program that manages multiple operating systems on a single computer system. It manages the system's processor, memory, and other resources to allocate what each operating system requires

Image-based Backup

- Creates a copy of the guest OS, its data, VM state, and configurations
  - The backup is saved as a single file – "image"
  - Mounts image on a proxy server



- **This effectively offloads the backup processing from the hypervisor and transfers the load on the proxy server, thereby reducing the impact to VMs running on the hypervisor.**
- **Image-based backup enables quick restoration of a VM.**