WIPRO NGA Program – DC DWS Batch 7

Capstone Project Presentation – 4th and 5th Sept 2024

Project Title Here - STORAGE AND BACKUP

Presented by - MUKESH YADAV

STORAGE AND BACKUP

BY MUKESH YADAV

BATCH NAME : DATA CENTRE - BATCH 7

CLOUD USERNAME : 24NAG2180_U07

GIT USERNAME : b166user05

Contents

- Introduction to Storage and Backup
- Objectives
- Project Scope
- Prerequisites , Project Requirements
- Deliverables

What is storage ?

Storage refers to the systems and technologies used to save and manage data for everyday access and use. This includes physical hardware like hard drives and solid-state drives, as well as cloud storage solutions. Storage is designed to hold and provide quick access to data that is actively used or needed by users and applications.

What is the importance of understanding storage basics and topologies in IT?

Understanding storage basics involves knowing different types of storage devices (like hard drives and SSDs) and how they are used to store data. Storage topologies refer to the arrangement and connection of these storage devices within a network or system. This knowledge is crucial for designing efficient storage solutions, optimizing data access, and ensuring scalability and performance in IT environments.

What is backup ?

Backup is the process of creating copies of data to protect against loss, corruption, or disaster. Backups are stored separately from the primary storage systems and are used to restore data if the original data is lost or damaged. This can involve copying data to external drives, tapes, or cloud services on a regular basis to ensure data can be recovered in case of emergencies or failures.

Why is backup critical for data management and what should be included in a backup strategy?

Backup is critical because it protects against data loss due to hardware failures, accidental deletions, or disasters. A robust backup strategy includes creating regular copies of important data, using different types of backups (full, incremental, differential), and scheduling backups to ensure data is consistently protected. Effective backup solutions also involve configuring and managing backup software and understanding different backup types and their implementations to ensure reliable data recovery.

These questions and answers introduce the fundamental concepts of storage and backup, setting the stage for the detailed exploration of these topics in your project.

OBJECTIVE

Understand the Storage Basics, Topologies.

Understand the SAN fundamentals.

Explain the different RAID levels.

Difference between SAN and NAS.

Manage the storage spaces and backup software solutions.

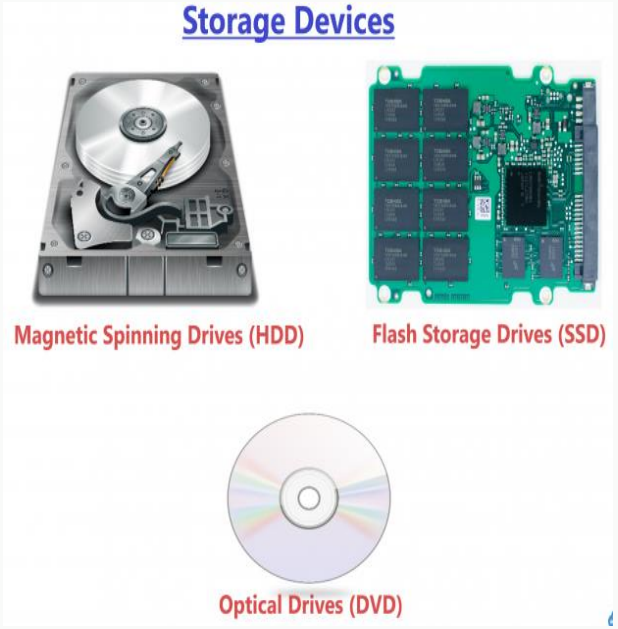
Understand the Storage Basics , Topologies

Storage Basics

Topologies

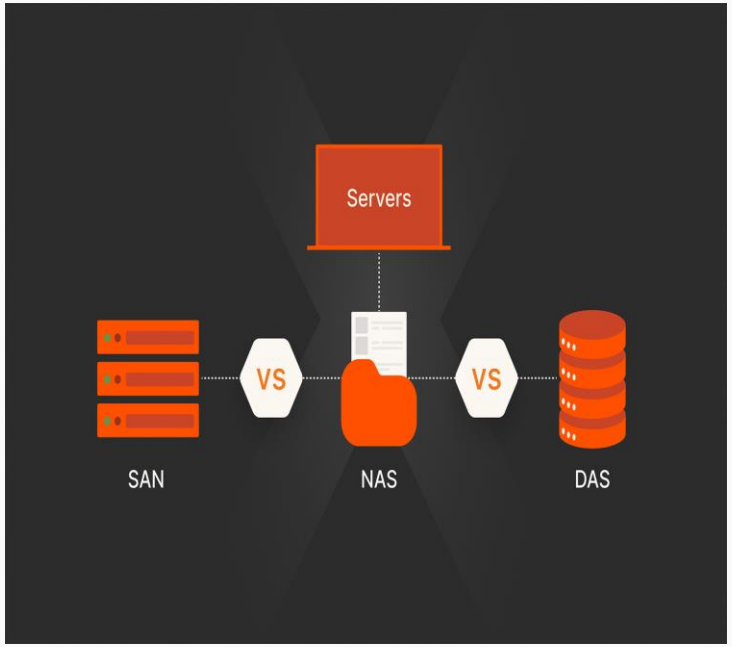
Storage Basics

- **Types of Storage Devices:** Different storage devices include hard drives (HDDs) for large capacity, solid-state drives (SSDs) for speed and durability, and optical drives like CDs/DVDs for data distribution and archival purposes.
- **Data Storage Mechanism:** Data on these devices is stored in bits, with HDDs using magnetic storage, SSDs using flash memory, and optical drives using laser technology to read/write data. Understanding these mechanisms is key to managing and optimizing storage performance.
- **Data Management:** Effective storage involves ensuring data is organized, accessed, and backed up efficiently. This includes understanding how data is written (sequential vs. random), read, and maintained to prevent loss and ensure quick retrieval when needed.



Topologies

- **Direct-Attached Storage (DAS):** Storage devices are directly connected to a single server. Ideal for straightforward, single-server setups.
- **Network-Attached Storage (NAS):** Storage is connected to a network, allowing multiple clients to access it. Suitable for shared access and file storage within a network.
- **Storage Area Network (SAN):** A high-speed network connecting multiple servers to shared storage. Best for environments requiring high performance and scalability.



Understand the SAN fundamentals

What is SAN and How It Works ?

Key Components and Benefits

What is SAN and How It Works ?

- **High-Speed Network:**

  - SAN is a specialized network that connects storage devices (e.g., disk arrays) to servers at high speeds, separate from the regular network used for other data traffic.

- **Block-Level Access:**

- - It operates at the block level, meaning servers interact directly with storage blocks rather than files, allowing for faster and more efficient data access.

- **Dedicated Network:**

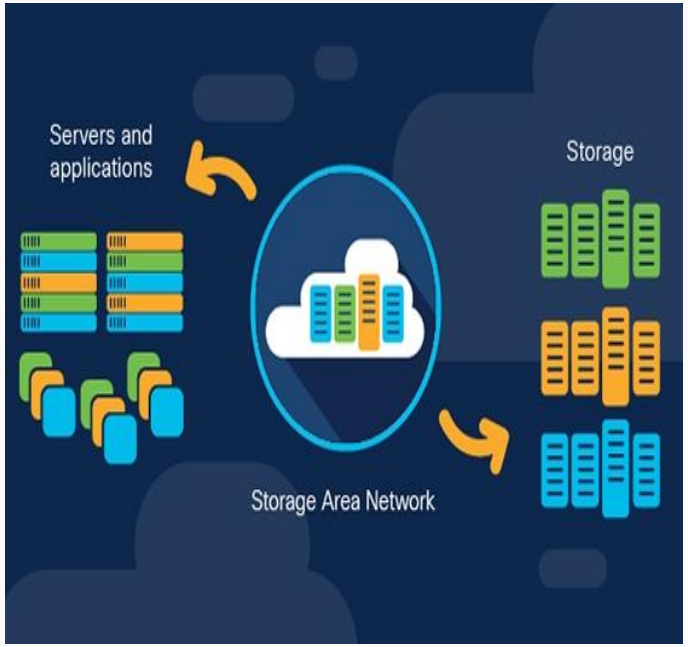
  - SAN uses its own network infrastructure, which ensures high performance and low latency, as it is not affected by general network traffic.

- **Scalability and Performance:**

  - This setup provides scalability by easily adding more storage or servers and supports high-performance applications due to its dedicated and high-speed connections.

key components and benefits

- 1. **Components:** SAN includes storage arrays (for storing data), switches (to manage data traffic), and host bus adapters (HBAs) (to connect servers to the SAN).
- 2. **Centralized Management:** SAN provides centralized storage management and improves data access speeds by connecting multiple servers to a shared storage pool.
- 3. **Scalability and Advanced Features:** It offers easy scalability, along with advanced features such as data replication, high availability, and disaster recovery.



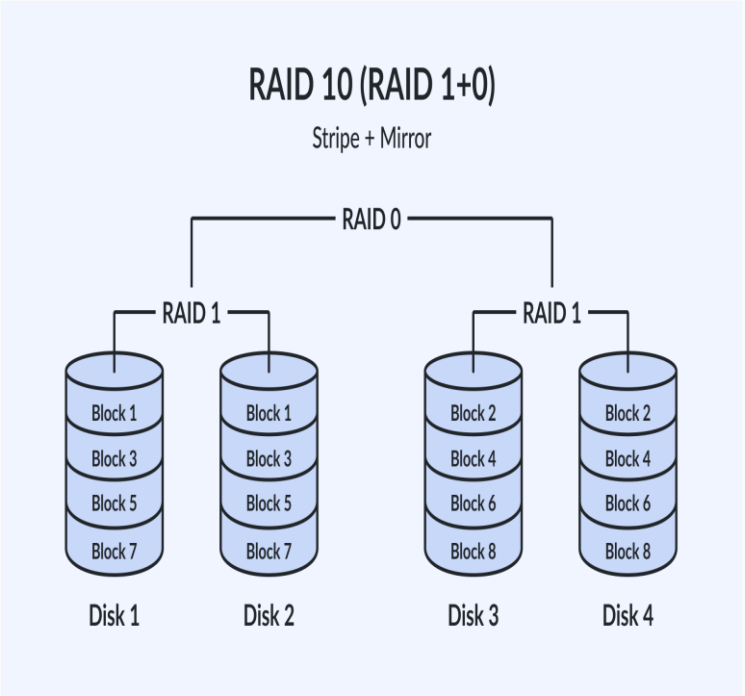
Explain the different RAID levels

RAID Levels Focused on Performance and Redundancy

RAID Levels Focused on Fault Tolerance and Efficiency

RAID Levels Focused on Performance and Redundancy

- **RAID 0 (Striping):** Splits data across multiple drives for faster performance but offers no redundancy.
- **RAID 1 (Mirroring):** Duplicates data across two or more drives, providing high redundancy and data protection.
- **RAID 10 (1+0):** Combines mirroring and striping to provide high performance and redundancy, handling multiple drive failures if no mirror loses all drives.



RAID Levels Focused on Fault Tolerance and Efficiency

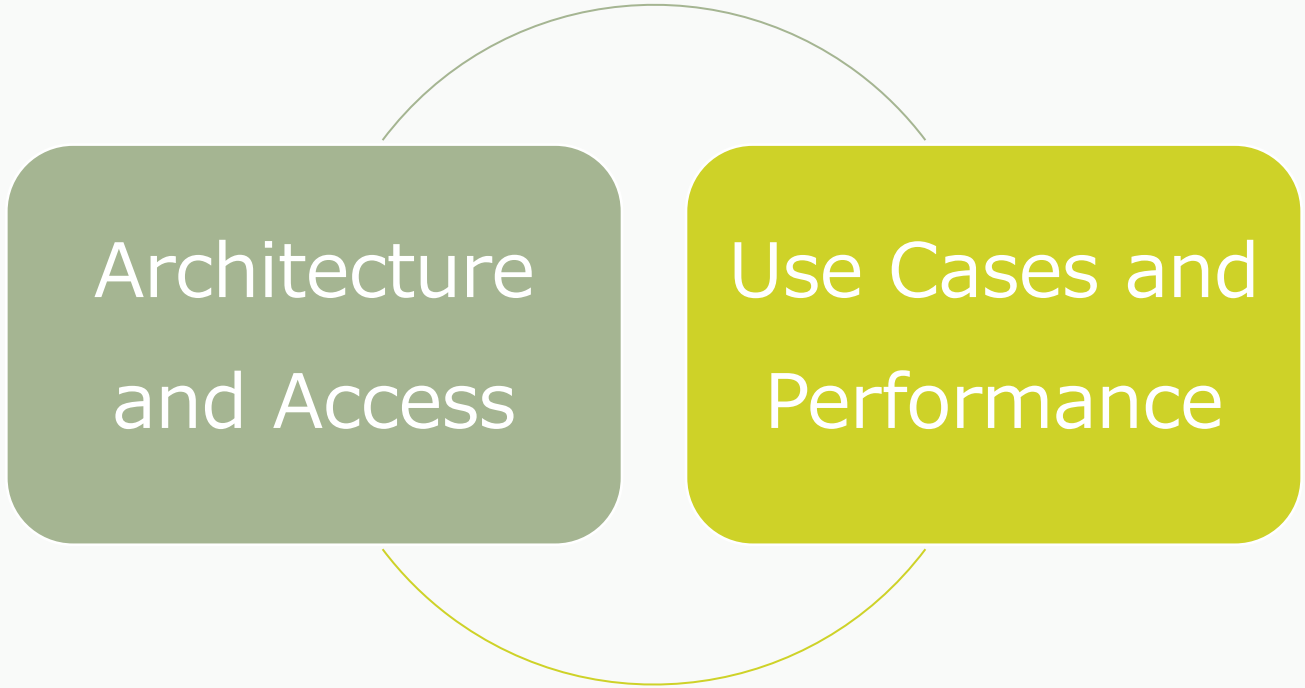
- **RAID 5 (Striping with Parity):** Distributes data and parity across three or more drives, balancing performance, redundancy, and efficiency with recovery from one drive failure.

- **RAID 6 (Striping with Double Parity):** Similar to RAID 5 but with additional parity for higher fault tolerance, protecting against two drive failures.

- **RAID 50 (5+0):** Combines RAID 5 and RAID 0 by striping across multiple RAID 5 arrays for enhanced performance and redundancy.

- **RAID 60 (6+0):** Combines RAID 6 and RAID 0 by striping across multiple RAID 6 arrays, offering high fault tolerance and improved performance.

Difference between SAN and NAS



Architecture and Access

- **SAN (Storage Area Network):**

  - Description: A high-speed network connecting storage devices (like disk arrays) directly to servers.

  - Access: Provides block-level access, meaning servers can interact directly with storage as if it were local, allowing for high performance and scalability.

- **NAS (Network Attached Storage):**

  - Description: A storage device connected to a standard network, providing file-level access to multiple users.

  - Access: Uses network protocols (like SMB or NFS) to allow file sharing and access over a network, suitable for general file storage and sharing.

Use Cases and Performance

- **SAN:**

  - Use Cases: Ideal for high-performance applications requiring fast, reliable access to large volumes of data (e.g., databases, virtualized environments).

  - Performance: Offers high speed and low latency due to dedicated storage network; more complex and expensive to set up.

- **NAS:**

- Use Cases: Suitable for file sharing, backup, and archiving needs across a network (e.g., office file storage, media sharing).

  - Performance: Provides adequate speed for general file access; easier and more cost-effective to set up but can be limited by network bandwidth.

Manage the storage spaces and backup software solutions

Assess and Implement Storage Needs: Evaluate data volume, access speed, and scalability requirements. Choose and configure appropriate storage solutions like HDDs, SSDs, or cloud storage, and set up RAID levels for performance and redundancy.

Monitor and Optimize Storage: Regularly check storage usage and performance metrics. Optimize storage through techniques like caching and tiering to maintain efficiency and prevent issues.

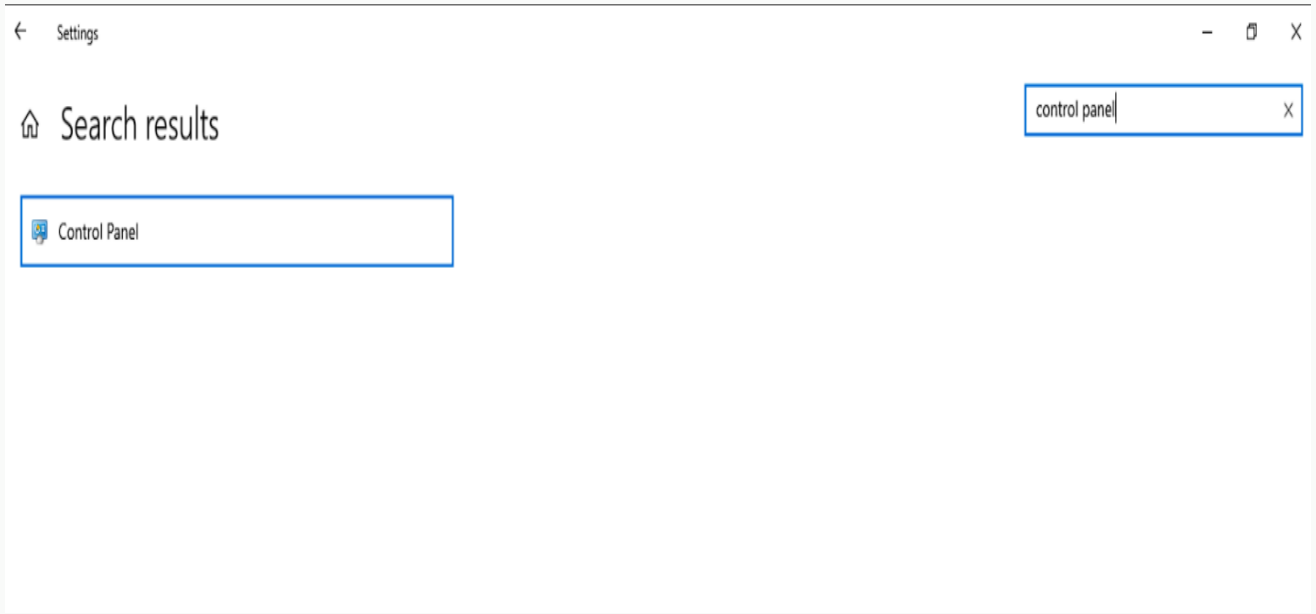
Select and Configure Backup Solutions: Choose backup software with essential features like automation and incremental backups. Schedule backups to minimize performance impact and ensure reliable data protection.

Ensure Security and Compliance: Encrypt backup data and implement access controls. Store backups in multiple locations and ensure compliance with data protection regulations and standards.

Backup

How to take backup in windows and Linux.

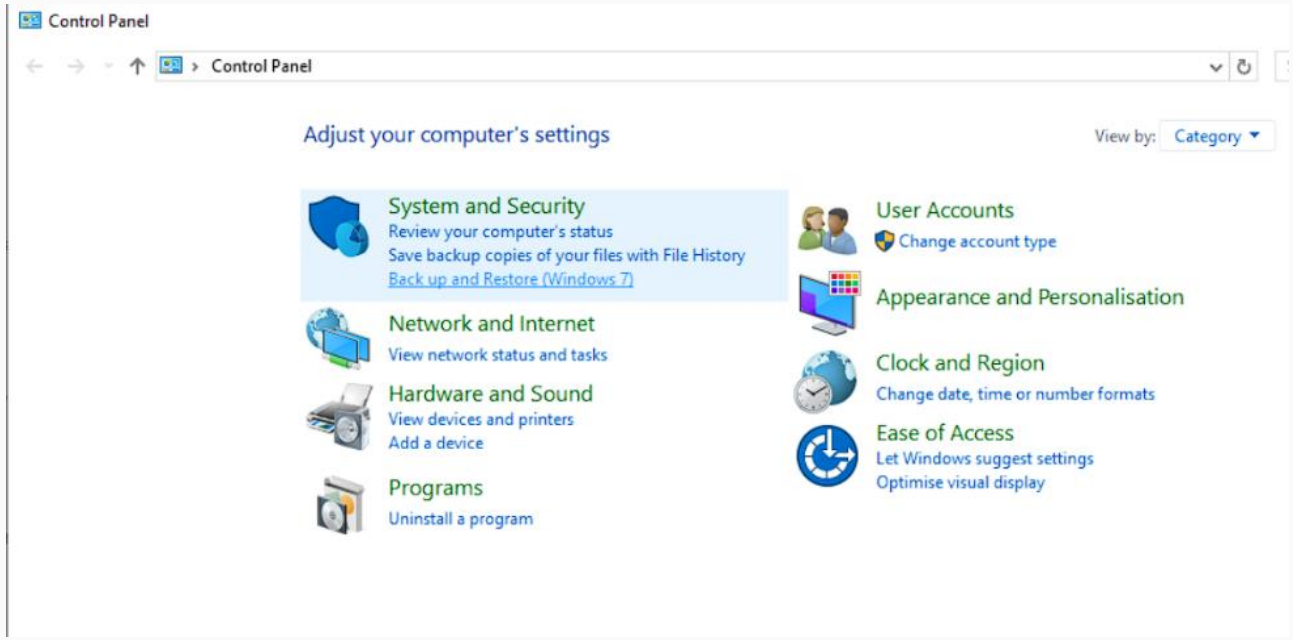
In Windows



1. Go to the Control Panel.

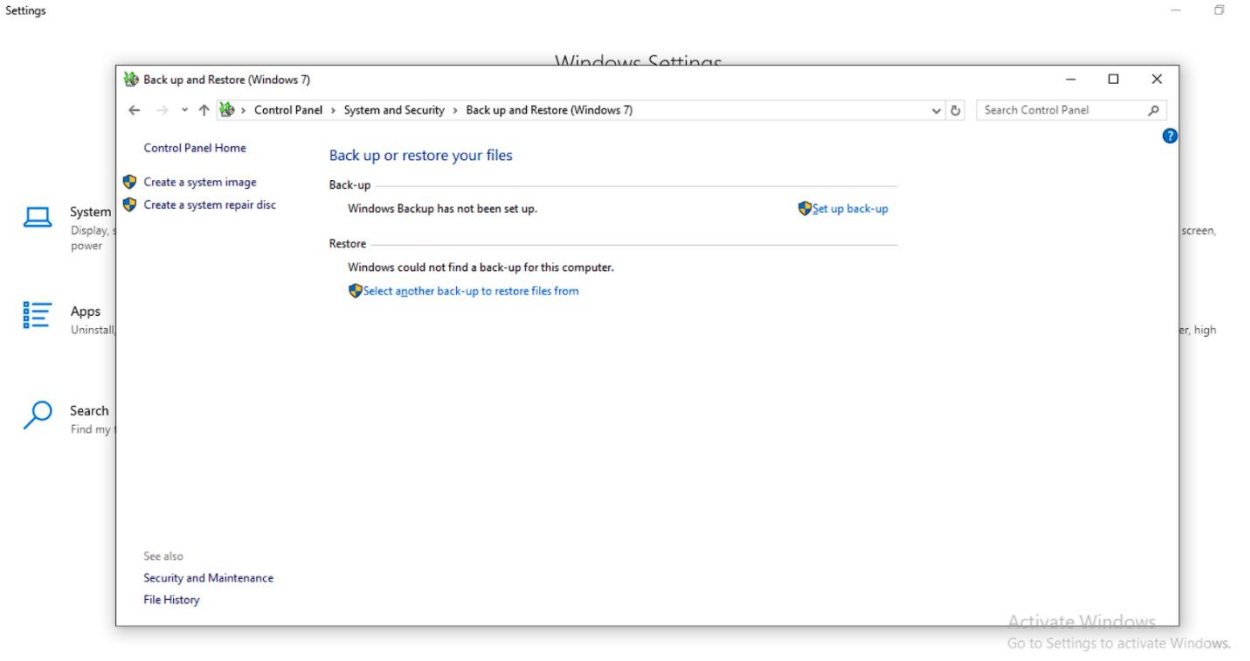
2. Open the Control Panel.

Step 1



- Navigate to the **System and Security** section.
- Click on the **Backup and Restore** option.

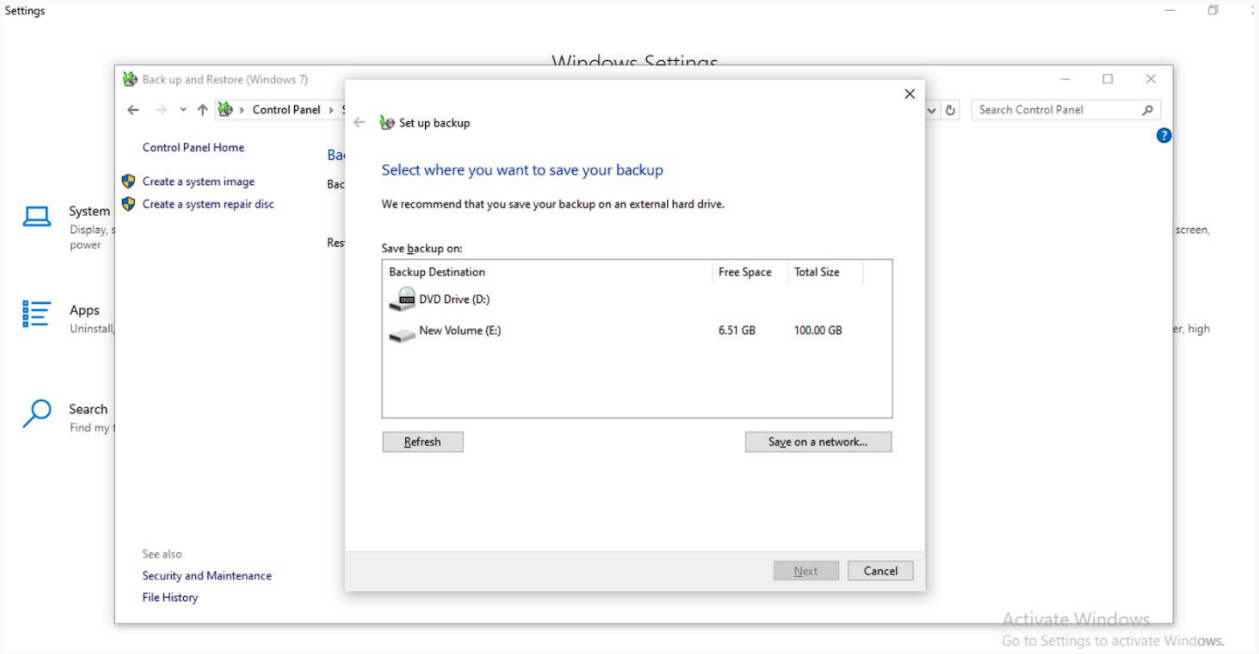
Step 2



1. Click on **“Set up backup”** to begin configuring the backup process.

2. Follow the prompts to complete the setup and wait for further instructions.

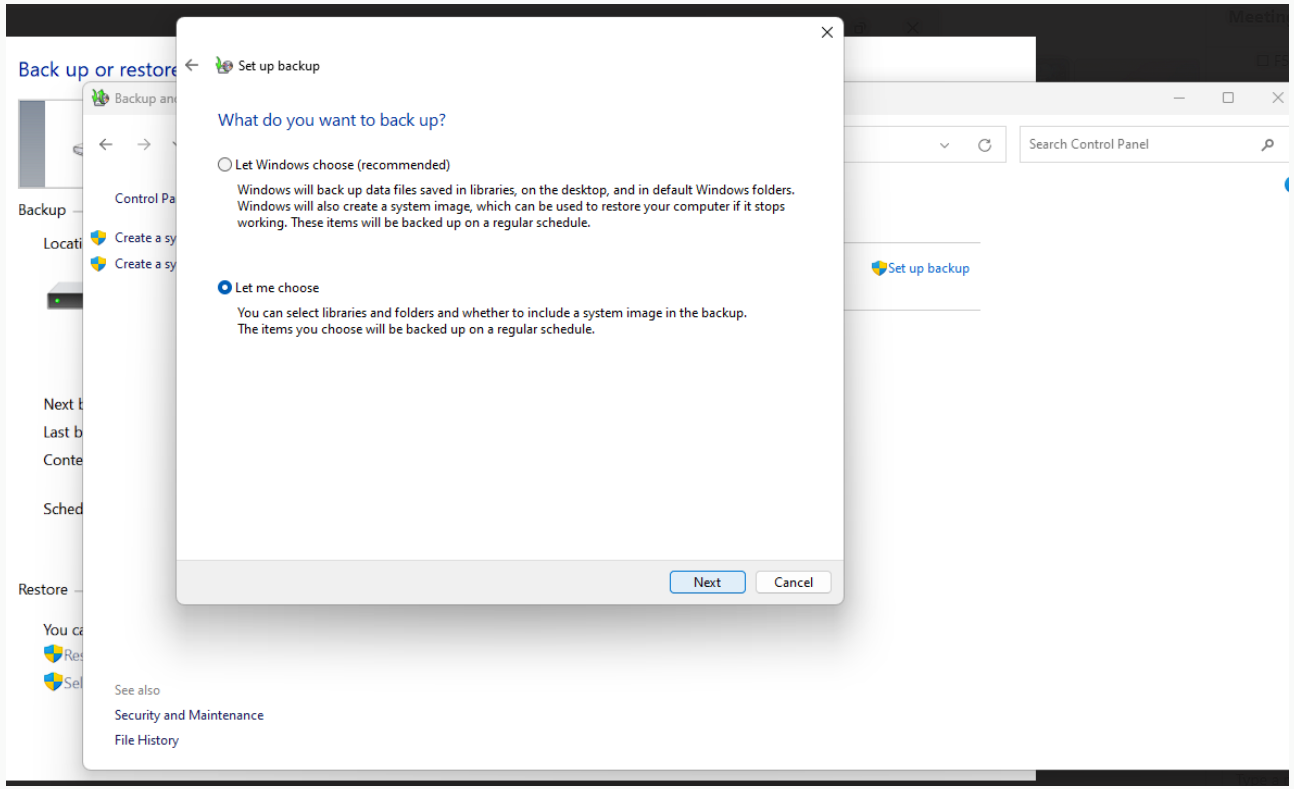
Step 3



1. **Select** the disk where you want to store the backup.

2. **Click on “Next”** to proceed with the setup.

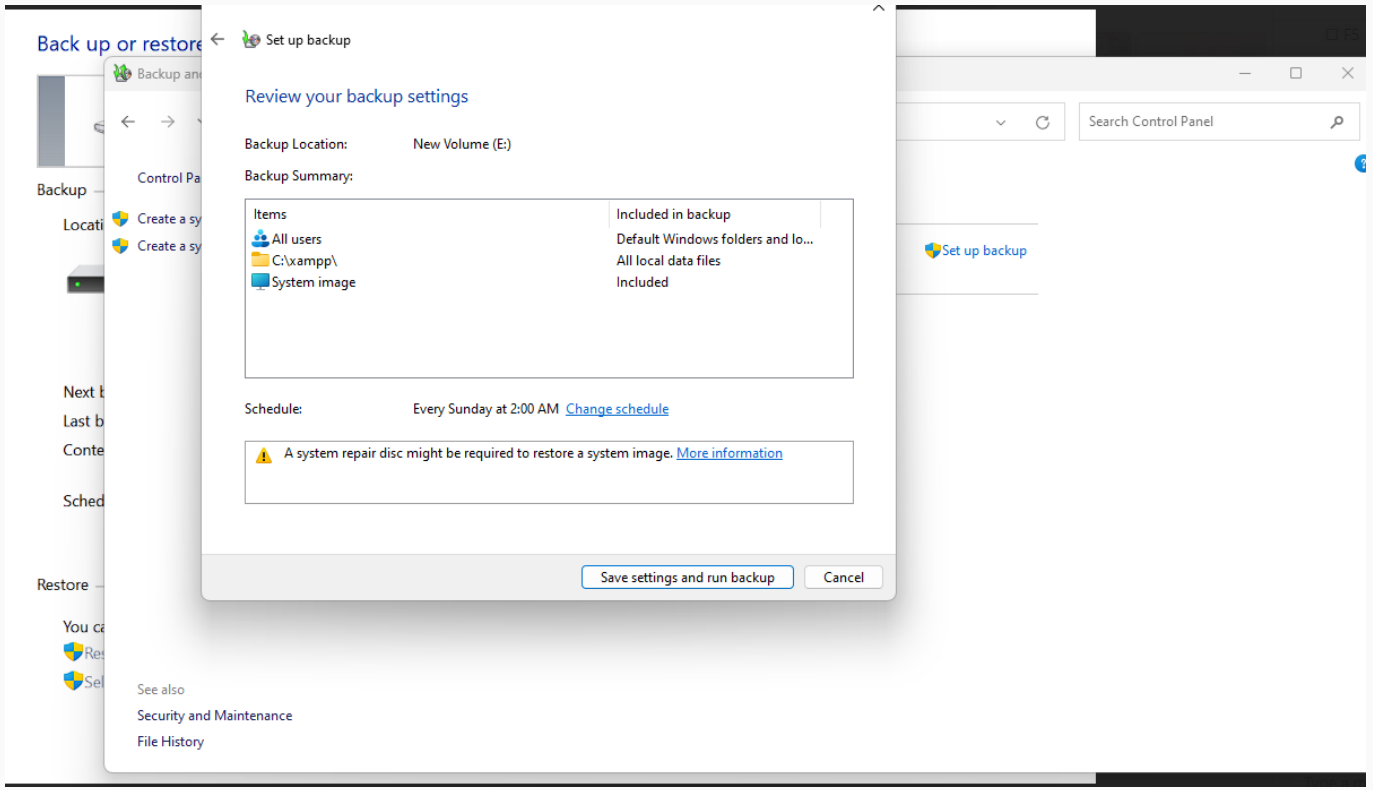
Step 4



1. **Click on** the "**Let me choose**" option. By default, "**Let Windows choose**" is selected, but selecting "**Let me choose**" allows you to customize the backup settings according to your preferences.

2. **Proceed with** the customized backup options as per your requirements.

Step 5

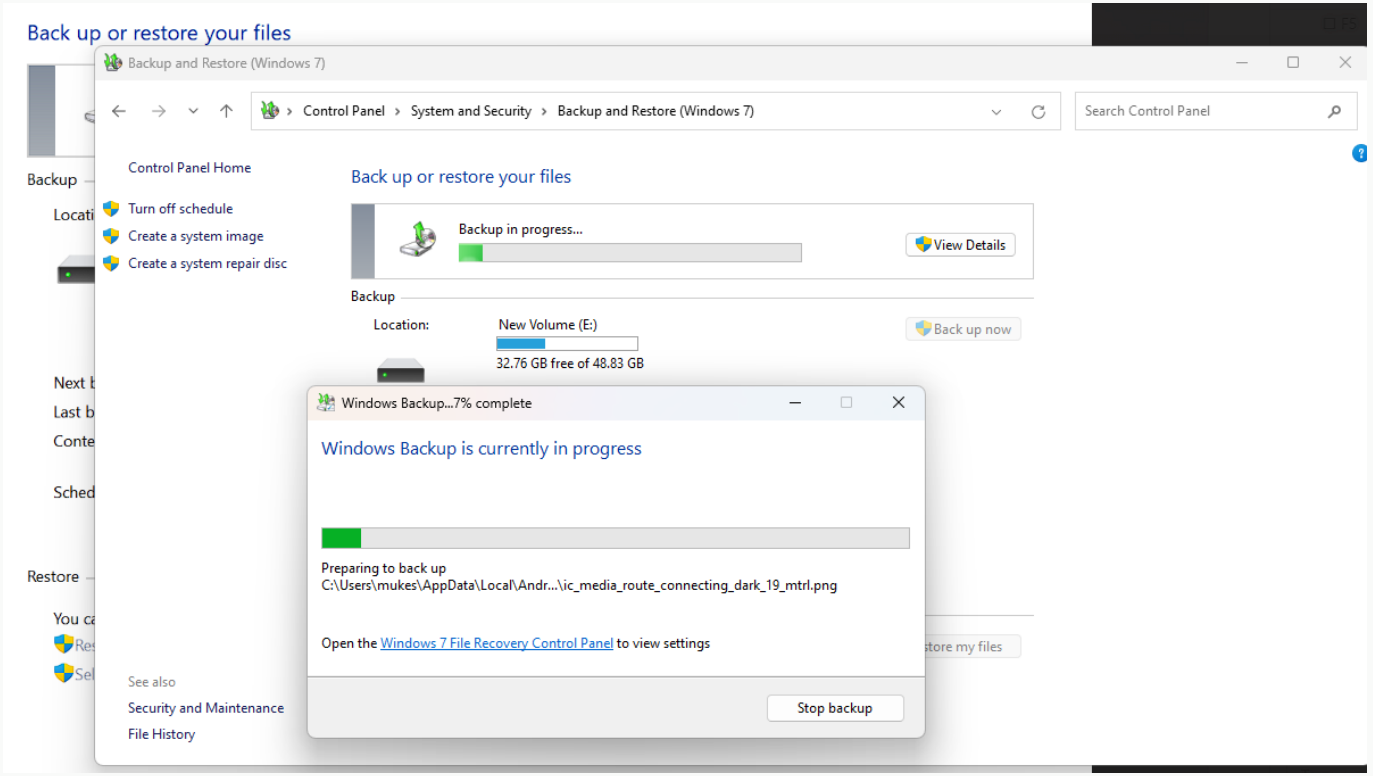


1. **Select** the files or folders you want to back up.

2. **Click on "Save settings and run backup"** to start the backup process.

3. **Optionally**, **schedule the backup task** according to your preference (e.g., minutes, hours, days, months, or years) by adjusting the schedule settings before proceeding.

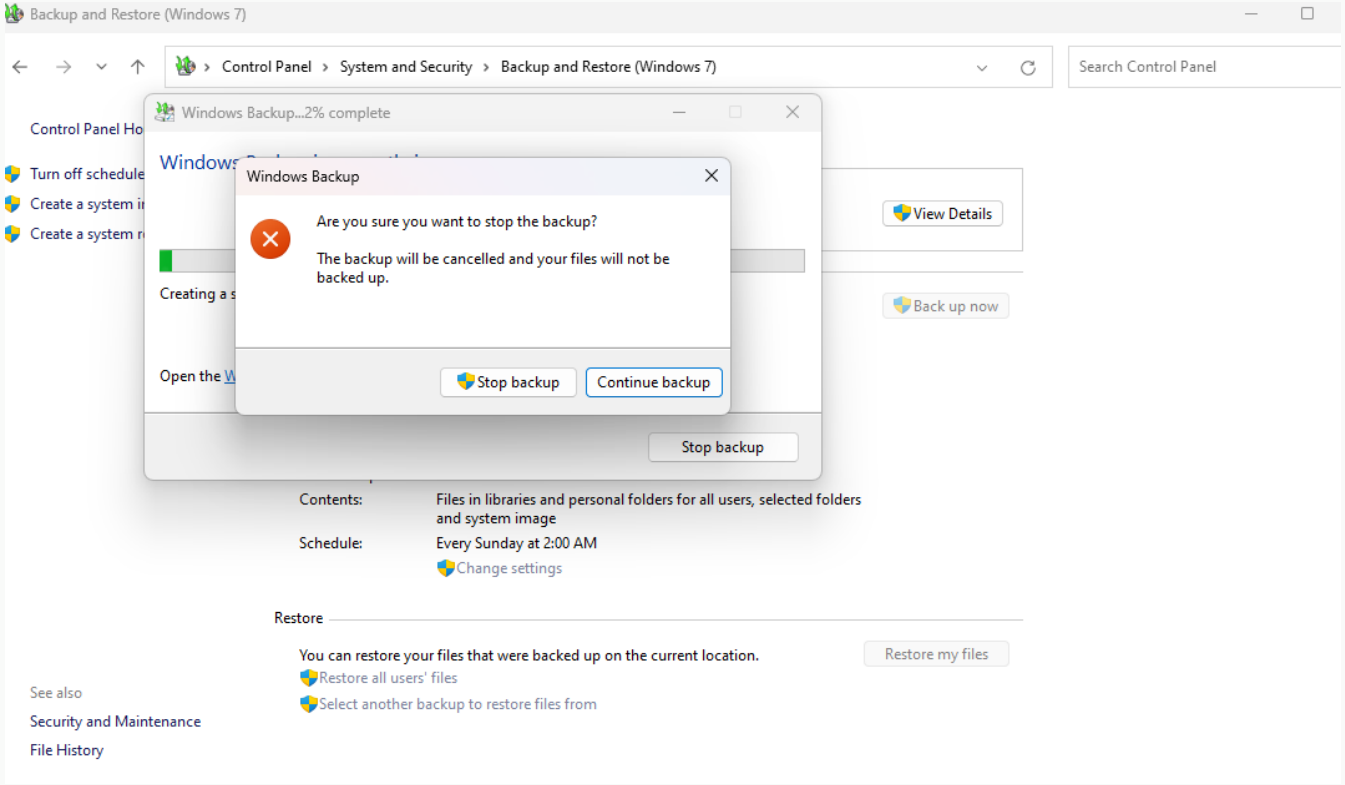
Step 6



1. **Backup Process**: Once you click "**Save settings and run backup**," the backup process will start.

2. **Monitor and Manage**: You can monitor the progress of the backup. If needed, you have the option to **stop the backup** by clicking on the corresponding stop option. For now, keep the backup running to complete the process.

Step 7

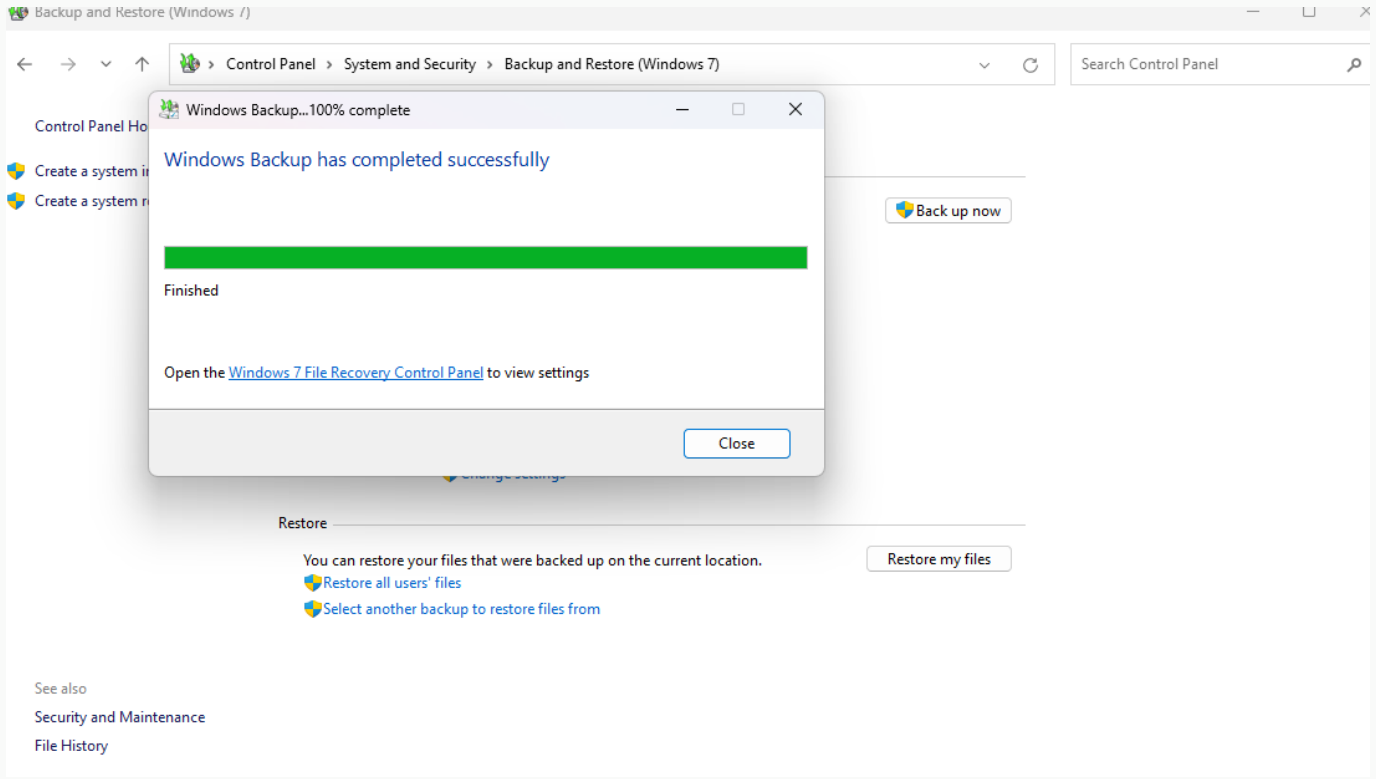


1. Click **“Stop Backup.”**

2. In the confirmation popup, click **“Continue”** to resume the backup (this is what I clicked).

3. Click **“Stop”** to halt the backup if needed.

Step 8



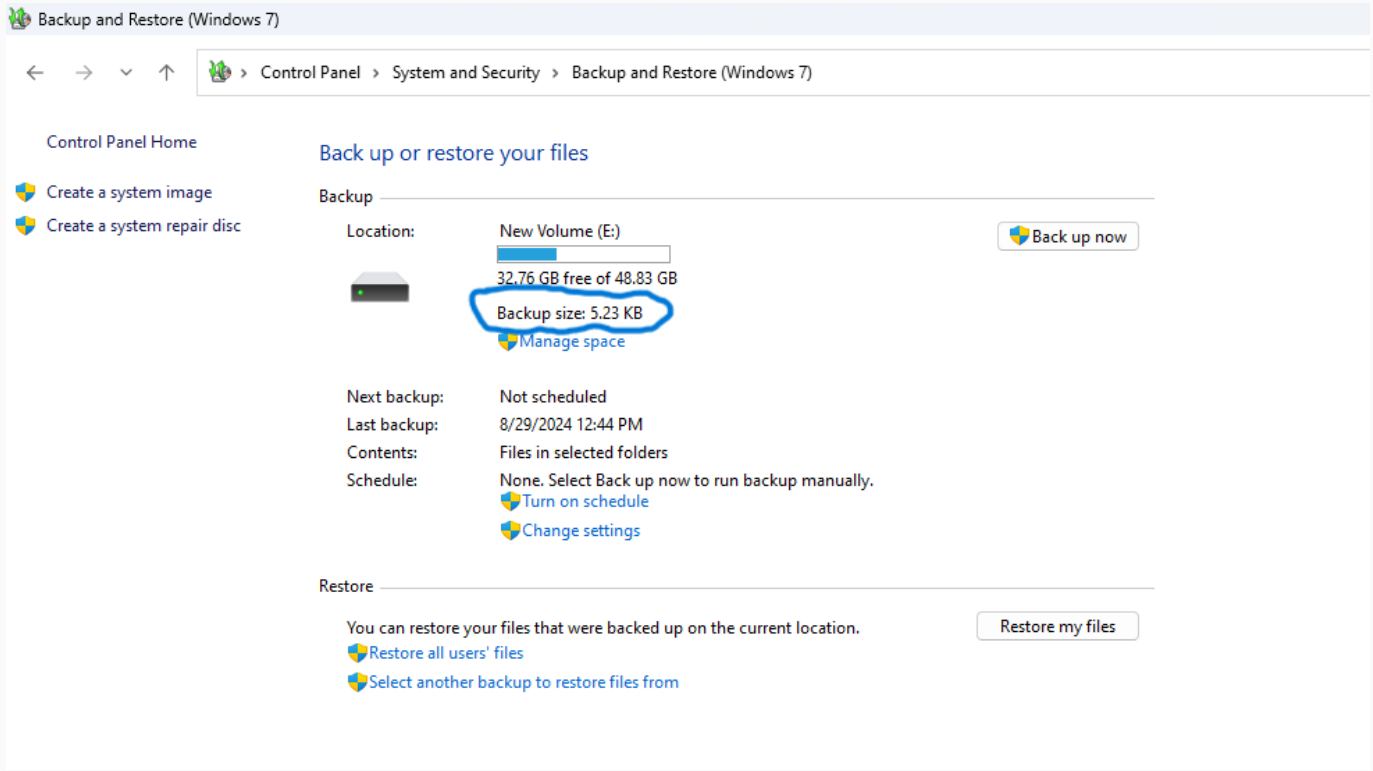
- Click **“Stop Backup.”**

In the confirmation popup, click “Continue” to resume the backup (this is what you clicked).

- After the backup is successfully completed, a popup will appear indicating that the backup is finished.
- Close the popup.

You can now access and view the backed-up file.

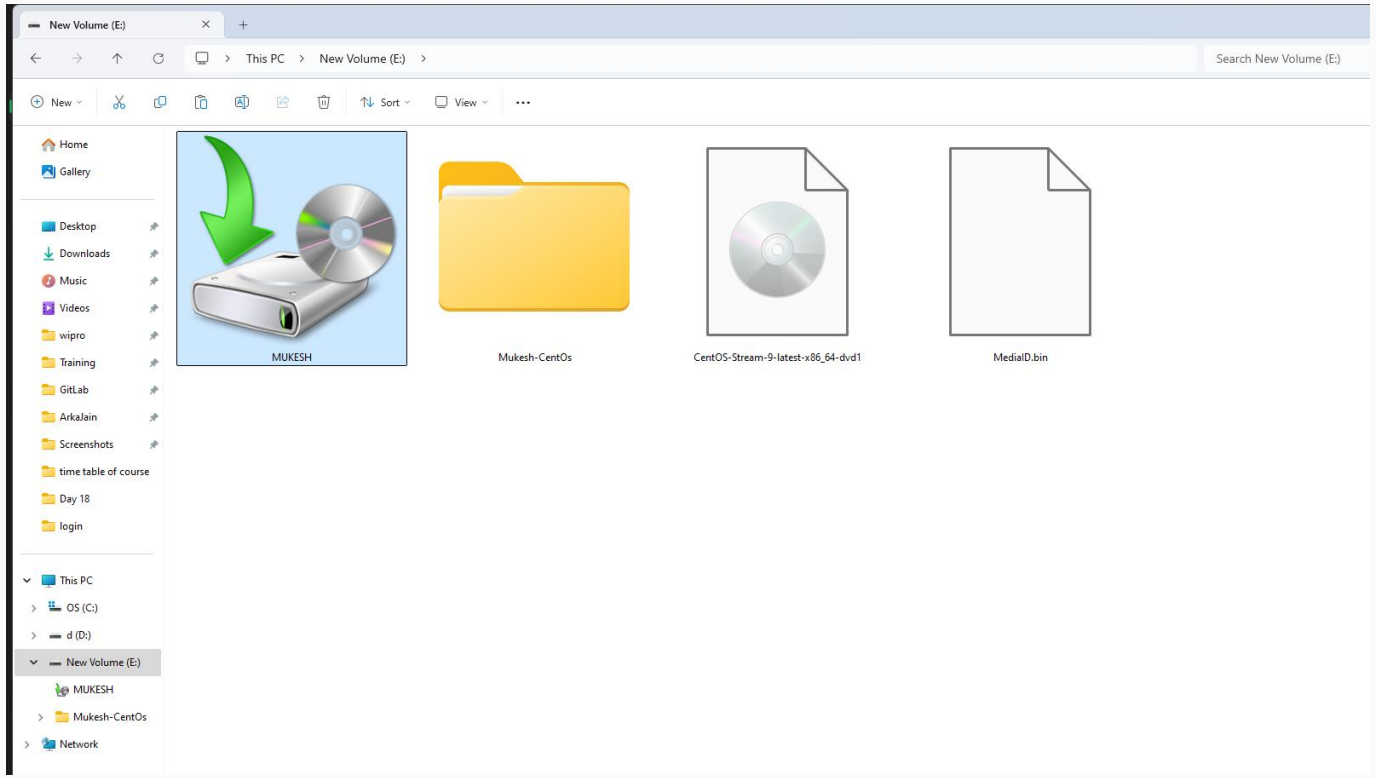
Step 9



1. After the backup is completed, view the “backup size.”

2. Check the size of the backed-up files.

Step 10



1. After the backup is completed, view the “backup size.”

2. Check the size of the backed-up files.

3. Locate the file named, Here I given “Mukesh” that contains the backup.

4. Double-click on the “Mukesh” file to restore the files.

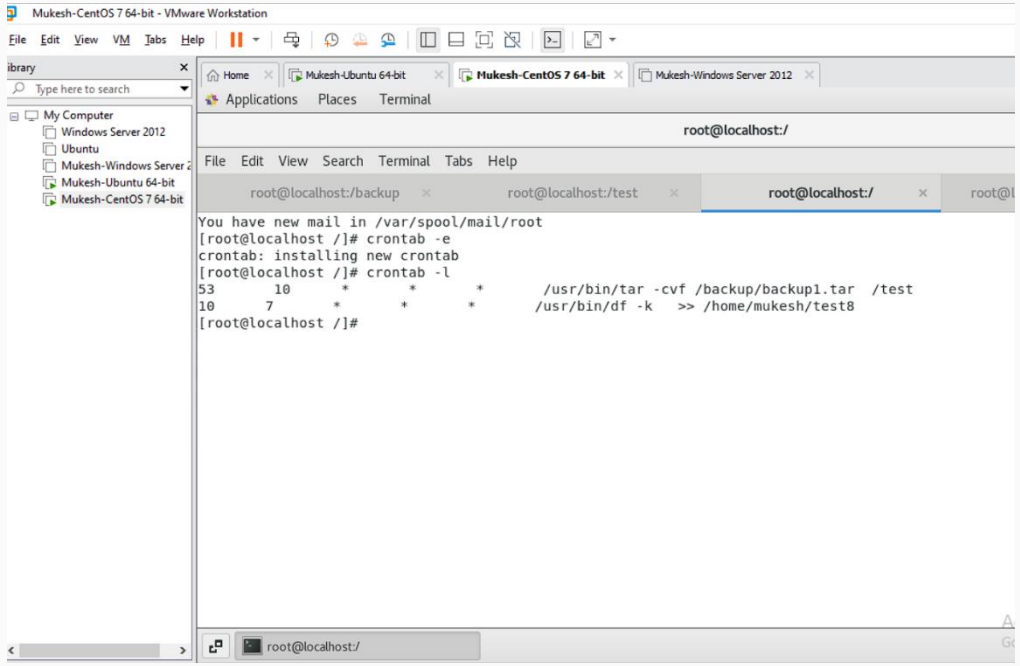
In Linux (CentOs)

```
[root@localhost tmp]# cd ..
[root@localhost /]# ls
bin  boot  dev  etc  home  lib  lib64  media  mnt  opt  proc  root  run  sbin  srv  sys  tmp  usr  var
[root@localhost /]# mkdir test
[root@localhost /]# cd test
[root@localhost test]# touch a.text
[root@localhost test]# touch b.text
[root@localhost test]# touch c.text
[root@localhost test]# ls
a.text  b.text  c.text
[root@localhost test]# pwd
/test
[root@localhost test]# ls
a.text  b.text  c.text
```

1. **Create Directory:** I used `**mkdir**` command to create a new directory.
2. **Change Directory:** Use `**cd**` to change to the directory named “test.”
3. Create Files: I used `**touch**` command to create four files within the directory.
4. **Set Source for Backup:** I used this directory as the source for your backup.

```
[root@localhost backup]# ls
backup1.tar  test
[root@localhost backup]# pwd
/backup
[root@localhost backup]# cd ..
[root@localhost /]# ls
backup  bin  boot  dev  etc  home  lib  lib64  media  mnt  opt  proc  root  run  sbin  srv  sys  test  tmp  usr  var
[root@localhost /]# cd backup
[root@localhost backup]# ls
backup1.tar  test
[root@localhost backup]# ls -la
total 12
drwxr-xr-x.  3 root root    37 Aug 29 10:56 .
dr-xr-xr-x. 19 root root   250 Aug 29 09:55 ..
-rw-r--r--.  1 root root 10240 Aug 29 10:53 backup1.tar
drwxr-xr-x.  2 root root    48 Aug 29 09:54 test
[root@localhost backup]# mv backup1.tar oldBackup.tar
[root@localhost backup]# ls -la
total 12
drwxr-xr-x.  3 root root    39 Aug 29 11:07 .
dr-xr-xr-x. 19 root root   250 Aug 29 09:55 ..
-rw-r--r--.  1 root root 10240 Aug 29 10:53 oldBackup.tar
drwxr-xr-x.  2 root root    48 Aug 29 09:54 test
[root@localhost backup]#
```

- **Backup Directory:** Store the backup in the backup directory.
- **List Files:** Use **ls** to check the list of files.
- **Present Working Directory:** Use **pwd** to display the current working directory.
- **Detailed File Listing:** Use **ls -la** to list all files with detailed permissions and attributes.



1. **Edit Crontab:** Run `**crontab -e**` to open the crontab editor.

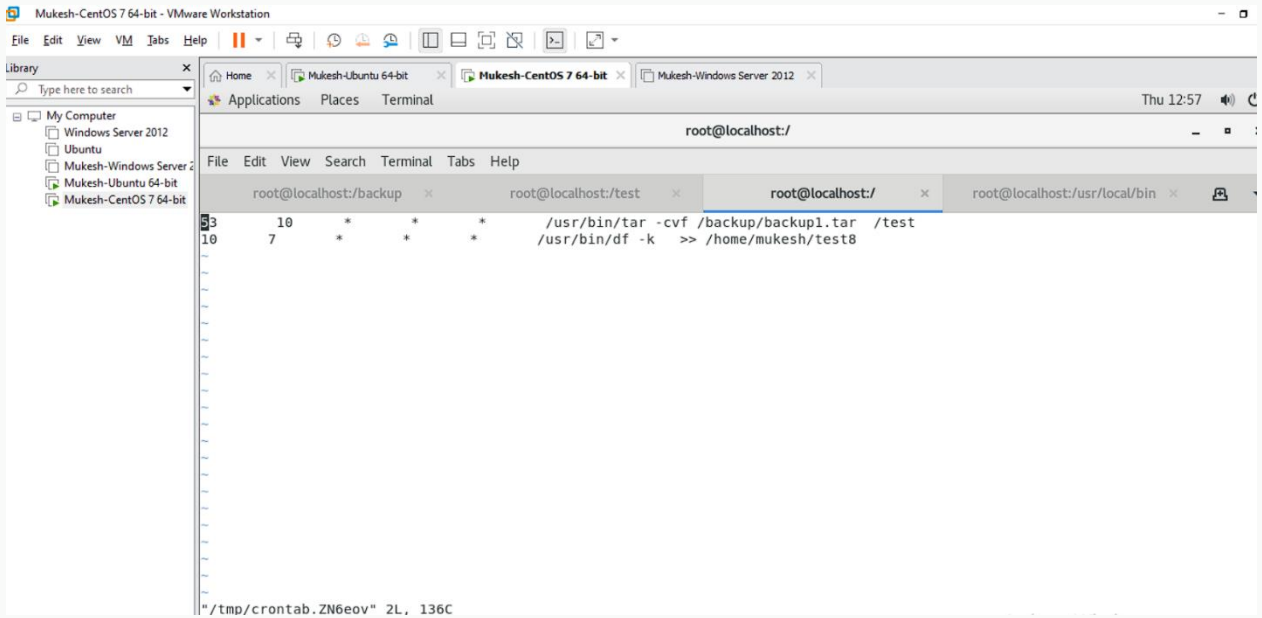
2. **Set Backup Schedule:** Add the following line to schedule the backup**: "53 10 * * * tar -cvf /backup/backup.tar /test"**

   This schedules the command to run at 10:53 AM every day. `**tar**`

3. **Create Backup:** The `**tar -cvf /backup/backup.tar /test**` command creates a backup of the `/test` directory and saves it as `backup.tar` in the `/backup` directory.

4. **Verify Execution:** After the scheduled time, the `**tar**` command will automatically back up the `**/test**` directory to the `**/backup**` directory as specified.

This setup ensures that the backup is performed daily at the scheduled time.



1. Edit Crontab: Use `**crontab -e**` to open the crontab editor and schedule tasks.

2. Here I scheduled Two Tasks.

3. List Scheduled Tasks: Use `crontab -l` to view the list of scheduled tasks to confirm they are set correctly.

This allows you to manage and verify your scheduled backup tasks.

```
[root@localhost backup]# ls
backup1.tar  test
[root@localhost backup]# pwd
/backup
[root@localhost backup]# cd ..
[root@localhost /]# ls
backup  bin  boot  dev  etc  home  lib  lib64  media  mnt  opt  proc  root  run  sbin  srv  sys  test  tmp  usr  var
[root@localhost /]# cd backup
[root@localhost backup]# ls
backup1.tar  test
[root@localhost backup]# ls -la
total 12
drwxr-xr-x.  3 root root    37 Aug 29 10:56 .
dr-xr-xr-x. 19 root root   250 Aug 29 09:55 ..
-rw-r--r--.  1 root root 10240 Aug 29 10:53 backup1.tar
drwxr-xr-x.  2 root root    48 Aug 29 09:54 test
[root@localhost backup]# mv backup1.tar oldBackup.tar
[root@localhost backup]# ls -la
total 12
drwxr-xr-x.  3 root root    39 Aug 29 11:07 .
dr-xr-xr-x. 19 root root   250 Aug 29 09:55 ..
-rw-r--r--.  1 root root 10240 Aug 29 10:53 oldBackup.tar
drwxr-xr-x.  2 root root    48 Aug 29 09:54 test
[root@localhost backup]#
```

1. Extract the Tarball: Use **`tar -xvf backup1.tar`** to extract the contents of the `backup1.tar` file.

2. Verify Extraction: Use `ls -la` to list the extracted files and check their permissions.

3. **Rename Backup File:** Use `mv` to rename the backup file for easy identification. For example:

“mv backup1.tar Oldbackup.tar”

4. **Monitor Files:** On the next day, we can distinguish between the previous and current backups by their names (`backup_yesterday.tar` and the new backup file).

PROJECT SCOPE

Explain the Storage Protocols.

Understand all types of Backup software used in IT companies.

Able to understand and implement the Backup types.

Configure SAN.

Explain the Storage Protocols

Storage protocols are essential for managing how data is transferred between storage devices and servers.

SCSI (Small Computer System Interface) is a standard for connecting and transferring data directly, commonly used in high-performance direct-attached storage setups.

iSCSI (Internet Small Computer System Interface) enables SCSI commands to be sent over IP networks, facilitating cost-effective and scalable network-based storage solutions.

Fibre Channel (FC) provides high-speed, low-latency connections over fiber optics for storage area networks (SANs), making it ideal for large-scale, mission-critical applications.

NFS (Network File System) allows file-based access to storage over a network, enabling multiple clients to access shared files, and is typically used in network-attached storage (NAS) environments. Each protocol serves distinct purposes, depending on the needs for data transfer, performance, and access.

Understand all types of Backup Software used in IT companies

Full Backup: Creates a complete copy of all data, providing comprehensive protection but requiring significant time and storage.

Incremental and Differential Backup: Incremental Backup saves only changes since the last backup, while Differential Backup saves changes since the last full backup, balancing efficiency and storage needs.

Snapshot and Cloud Backup: Snapshot Backup provides quick point-in-time copies for rapid recovery, and Cloud Backup stores data offsite for scalability and protection against local disasters.

Able to understand and implement the Backup types

Understand Backup Types: Learn the characteristics and benefits of various backup methods—Full Backup provides comprehensive coverage, Incremental Backup saves only new changes, Differential Backup captures changes since the last full backup, and Snapshot Backup offers quick recovery. Cloud Backup provides offsite storage for additional safety.

Implement Backup Strategies: Choose and configure backup solutions based on your needs. Use Full Backup for complete protection, Incremental or Differential Backup for efficient storage, and Snapshot or Cloud Backup for quick recovery and offsite protection. Schedule and automate backups to ensure regular and reliable data protection.

Configure SAN

Set Up SAN Hardware: Connect SAN components such as storage arrays, switches, and servers. Install and configure Host Bus Adapters (HBAs) in servers, connect them to SAN switches, and then link the switches to the storage arrays. Ensure proper cabling and hardware compatibility for seamless communication.

Configure SAN Software and Settings: Install and configure SAN management software to handle storage provisioning, zoning, and LUN (Logical Unit Number) mapping. Set up zoning to control access between servers and storage devices, and configure LUNs to allocate storage to specific servers. Verify connectivity, perform tests, and ensure that performance and redundancy settings meet your requirements.

PREREQUISITES & PROJECT REQUIREMENTS

Basic Knowledge Storage basics and storage devices.

Basic understanding how SAN is implemented.

Understanding of NAS and SAN.

Understanding of Storage disk arrays, volume manager, LVM.

DELIVERABLES

- Attach a storage disk array to a server using RAID software.
- Manage the storage file system space.
- Configure to backup file system on everyday and set it in scheduler.
- Identify and configure the RAID levels and backup the file system.
- Identify and troubleshoot the storage space issues.
- Implement the different types of backup’s and ensure its working.

Summary

I explored key storage concepts, including storage devices, topologies, SAN, RAID levels, and the differences between SAN and NAS.

I managed storage spaces, configured daily backups with scheduling, and troubleshooted storage space issues.

I gained hands-on experience by attaching a storage disk array to a server using RAID software.

I focused on backup software solutions and implementing different types of backups.

Conclusion

This project gave me a solid foundation in storage management, including SAN, RAID, and the differences between SAN and NAS

I developed practical skills in managing storage, configuring and scheduling backups, and troubleshooting storage issues.

I am now better equipped to handle real-world IT storage and backup challenges.

This project has prepared me for further learning in advanced storage and data protection techniques.

THANK YOU

Mukesh Yadav

Batch name : Data Centre - Batch 7
cloud username : 24NAG2180_U07

Git username : b166user05