UNIT IV
BACKUP, ARCHIVE AND REPLICATION

Introduction to Business Continuity, Backup architecture, Backup targets and methods, Data deduplication, Cloud-based and mobile device backup, Data archive, Uses of replication and its characteristics, Compute based, storage-based, and network-based replication, Data migration, Disaster Recovery as a Service (DRaaS).

INTRODUCTION TO BUSINESS CONTINUITY :

- Continuous access to information is a must for the smooth functioning of business operations today, as the cost of business disruption could be catastrophic.
- There are many threats to information availability, such as natural disasters (e.g., flood, fire, earthquake), unplanned occurrences (e.g., cybercrime, human error, network and com- puter failure), and planned occurrences (e.g., upgrades, backup, restore) that result in the inaccessibility of information.
- It is critical for businesses to define appropriate plans that can help them overcome

these crises. *Business continuity is an important process to define and implement these plans.*

Business continuity (BC) **is an integrated and enterprise wide process that includes all activities (internal and external to IT) that a business must perform to mitigate the impact of planned and unplanned downtime**.

BC entails preparing for, responding to, and recovering from a system outage that adversely affects business operations. It involves proactive measures, such as business impact analysis and risk assessments, data protection, and security, and reactive countermeasures, such as disaster recovery and restart, to be invoked in the event of a failure. The goal of a business continuity solution is to ensure the "information availability" required to conduct vital business operations.

INFORMATION AVAILABILITY

➢ *Information availability (IA)* refers to the ability of the infrastructure to function according to business expectations during its specified time of operation. Information availability ensures that people (employees, customers, suppliers, and partners) can access information whenever they need it.

Information availability can be defined with the help of reliability, accessibility and timeliness.

Reliability: This reflects a component's ability to function without failure, under stated conditions, for a specified amount of time.

Accessibility: This is the state within which the required information is accessible at theright place, to the right user. The period of time during which the system is in an accessible state is termed *system uptime;* when it is not accessible it is termed *system downtime*.

Timeliness: Defines the exact moment or the time window (a particular time of the day, week, month, and/or year as specified) during which information must be accessible.

For example, if online access to an application is required between 8:00 AM and 10:00 pM each day, any disruptions to data availability outside of this time slot are notconsidered to affect timeliness.

Causes of Information Unavailability

- Various planned and unplanned incidents result in data unavailability.

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

✓ *Unplanned outages* include failure caused by database corruption, component failure, and human errors.

- Another type of incident that may cause data unavailability is natural or man-made disasters such as flood, fire, earthquake, and contamination.
- As illustrated in Figure 11-1, the majority of outages are planned. Planned outages are expected and scheduled, but still cause data to be unavailable.

Statistically, less than 1 percent is likely to be the result of an unforeseen disaster.

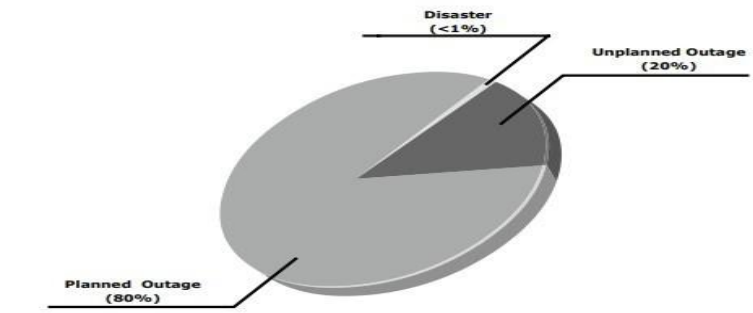


Figure 11-1: Disruptors of data availability

Measuring Information Availability

- Information availability relies on the availability of the hardware and software components of a data center. Failure of these components might disrupt information availability.
- A failure is the termination of a component's ability to perform a required function. The component's ability can be restored by performing an external corrective action, such as a manual reboot, a repair, or replacement of the failed component(s).
- Repair involves restoring a component to a condition that enables it to perform a required function within a specified time by using procedures and resources.
- Proactive risk analysis performed as part of the BC planning process considers the component failure rate and average repair time, which are measured by MTBF and MTTR:

Mean Time Between Failure (MTBF): It is the average time available for a system or component to perform its normal operations between failures.

Mean Time To Repair (MTTR): It is the average time required to repair a failed component. While calculating MTTR, it is assumed that the fault responsible for the failure is correctly identified and that the required spares and personnel are available.

MTTR includes the time required to do the following: detect the fault, mobilize the maintenance team, diagnose the fault, obtain the spare parts, repair, test, and resume normal operations.

- **IA is the fraction of a time period that a system is in a condition to perform**

its intended function upon demand. It can be expressed in terms of system uptime and downtime and measured as the amount or percentage of system uptime:

IA = system uptime / (system uptime + system downtime)

In terms of MTBF and MTTR, IA could also be expressed as

IA = MTBF / (MTBF + MTTR)

Table 11-1 lists the approximate amount of downtime allowed for a service to achievecertain levels of 9s availability.

For example, a service that is said to be "five 9s available" is available for 99.999percent of the scheduled time in a year (24 × 7 × 365).

Table 11-1: Availability Percentage and Allowable Downtime

UpTIme(%)	DownTIme (%)	Down TIme per Year	DownTIme per week
98	2	7.3 days	3 hr 22 minutes
99	1	3.65 days	1 hr 41 minutes
99.8	0.2	17 hr 31 minutes	10 minutes20 sec
99.9	0.1	8 hr 45 minutes	10 minutes 5 sec
99.99	0.01	52.5 minutes	1 minute
99.999	0.001	5.25 minutes	6 sec
99.9999	0.0001	31.5 sec	0.6 sec

Consequences of Downtime

- Data unavailability, or downtime, results in loss of productivity, loss of revenue, poorfinancial performance, and damages to reputation.
- *Loss of produc tivity* reduces the output per unit of labor, equipment, and capital.
- *Loss of revenue* includes direct loss, compensatory payments, future revenue losses,billing losses, and investment losses.
- *Poor financial performance* affects revenue recognition, cash flow, discounts, paymentguarantees, credit rating, and stock price.

An important metric, *average cost of downtime per hour*, provides a key estimatein determining the appropriate BC solutions.

It is calculated as follows:

Average cost of downtime per hour = average productivity loss per hour + average revenue loss per hour

Where:

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

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

Common terms of BC

- **Disaster recovery:** This is the coordinated process of restoring systems, data, and the infrastructure required to support key ongoing business operations in the event of a disaster.
- **Disaster restart:** This is the process of restarting business operations with mirrored consistent copies of data and applications.
- **Recovery-Point Objective (RPO):** This is the point in time to which systems and data must be recovered after an outage. It defines the amount of data loss that a business can endure.

For example, if the RPO is six hours, backups or replicas must be made at least once in 6 hours.

Figure 11-2 shows various RPOs and their corresponding ideal recovery strategies. For example:

RPO of 24 hours: This ensures that backups are created on an offsite tape drive every midnight.

RPO of 1 hour: This ships database logs to the remote site every hour.

RPO of zero: This mirrors mission-critical data synchronously to a remote site.

- **Recovery-Time Objective (RTO):** The time within which systems, applications, or functions must be recovered after an outage. It defines the amount of downtime that a business can endure and survive.

For example, if the RTO is two hours, then use a disk backup because it enablesa faster restore than a tape backup.

Some examples of RTOs and the recovery strategies to ensure data availability are

listed below

- **RTO of 72 hours:** Restore from backup tapes at a cold site.
- **RTO of 12 hours:** Restore from tapes at a hot site.
- **RTO of 4 hours:** Use a data vault to a hot site

BC PLANNING LIFECYCLE

- BC planning must follow a disciplined approach like any other planning process. Organizations today dedicate specialized resources to develop and main tain BC plans.
- The BC planning life cycle includes five stages (see Figure 11-3):

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

Figure 11-3: BC planning lifecycle

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

1. **Establishing objectives**

- Determine BC requirements.
- Estimate the scope and budget to achieve requirements.
- Select a BC team by considering subject matter experts from all areas of thebusiness, whether internal or external.
- Create BC policies.

2. **Analyzing**

- Collect information on data profiles, business processes, infra- structure 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.

3. **Designing and developing**

- Define the team structure and assign individual roles and responsibilities. For example, different teams are formed for activities such as emergency response, damage assessment, and infrastructure and application recovery.
- Design data protection strategies and develop infrastructure.
- Develop contingency scenarios.
- Develop emergency response procedures.
- Detail recovery and restart procedures.

4. **Implementing**

- Implement risk management and mitigation procedures that include backup, replication, and management of resources.
- Prepare the disaster recovery sites that can be utilized if a disaster affects theprimary data center.
- Implement redundancy for every resource in a data center to avoid single points of failure.

5. **Training, testing, assessing, and maintaining**

- Train the employees who are responsible for backup and replication of business-critical data on a regular basis or whenever there is a modi- fication 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 changeswithin the data center.

BACKUP PROCESS (BACKUP ARCHITECTURE)

- 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. The backup server depends on backup clients to gather the data to be backed up.
- The backup clients can be local to the server or they can reside on another server, presumably to back up the data visible to that server. The backup server receives backup metadata from the backup clients to perform its activities.
- 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). Typically, the storage node is integrated with the backup server and both are hosted on the same physical platform.
- 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.
- Storage nodes play an important role in backup planning because they can be used to consolidate backup servers.
- 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 (backups can also be initiated by a client). This request instructs the backup client to send its metadata to the backup server, and the data to be backed up to the appropriate storage node. On receiving this request, the backup client sends the metadata to the backup server.
- The backup server writes this metadata on its metadata catalog. The backup client also sends the data to the storage node, and the storage node writes the data to the storage device.
- 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.

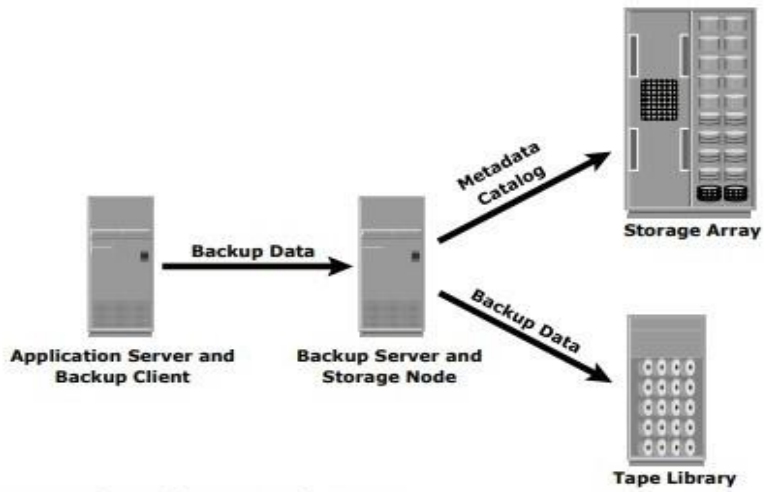


Figure 12-4: Backup architecture and process

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

BACKUP METHODS

- ***Hot backup and cold backup*** are the two methods deployed for backup. They are based onthe 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. 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. An open file is locked by the operating system and is not copied during the backup process until the user closes it.
- The backup application can back up open files by retrying the operation on files that were opened earlier in the backup process. During the backup process, it may be possible that files opened earlier will be closed and a retry will be successful.

- The maximum number of retries can be configured depending on the backup application. However, this method is not considered robust because in some environments certain files are always open.
- In such situations, the backup application provides open file agents. These agents interact directly with the operating system and enable the creation of consistent copies of open files. In some environments, the use of open file agents is not enough.
- For example, a database is composed of many files of varying sizes, occupying several file systems. To ensure a consistent database backup, all files need to be backed up in the same state. That does not necessarily mean that all files need to be backed up at the same time, but they all must be synchronized so that the database can be restored with consistency.
- Consistent backups of databases can also be done by using a cold backup. This requires the database to remain inactive during the backup. Of course, *the disadvantage of a cold backup is that the database is inaccessible to users during the backup process.*
- Hot backup is used in situations where it is not possible to shut down the database. This is facilitated by database backup agents that can perform a backup while the database is active. *The disadvantage associated with a hot backup is that the agents usually affect overall application performance.*
- A point-in-time (PIT) copy method is deployed in environments where the impact of downtime from a cold backup or the performance resulting from a hot backup is unacceptable.

- A pointer-based PIT copy consumes only a fraction of the storage space and can be created very quickly. A pointer-based PIT copy is implemented in a disk-based solution whereby a virtual LUN is created and holds pointers to the data stored on the production LUN or save location.
- In this method of backup, the database is stopped or frozen momentarily while the PIT copy is created. The PIT copy is then mounted on a secondary server and the backup occurs on the primary server.
- To ensure consistency, it is not enough to back up only production data for recovery. Certain attributes and properties attached to a file, such as permissions, owner, and other metadata, also need to be backed up.

- These attributes are as important as the data itself and must be backed up for consistency. Backup of boot sector and partition layout information is also critical for successful recovery.
- 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 BMRtechnologies can recover a server onto dissimilar hardware.

DATA DEDUPLICATION :

- Data deduplication emerged as a key technology to dramatically reduce the amount of space and the cost that are associated with storing large amounts of data. Data deduplication is the art of intelligently reducing storage needs in order of magnitude.
- This method is better than common data compression techniques.
- Data deduplication works through the elimination of redundant data so that only one instance of a data set is stored. IBM has the broadest portfolio of data deduplication solutions in the industry, which gives IBM the freedom to solve client issues with the most effective technology.
- Whether it is source or target, inline or post, hardware or software, disk or tape, *IBM has a solution with the technology that best solves the problem:*
- IBM ProtecTIER® Gateway and Appliance IBM System Storage N series DeduplicationIBM Tivoli Storage Manager
- Data deduplication is a technology that reduces the amount of space that is required to store data on disk. It achieves this space reduction by storing a single copy of data that is backed up repetitively.

- Data deduplication products read data while they look for duplicate data. Data deduplication products break up data into elements and create a signature or identifier for each data element.
- Then, they compare the data element signature to identify duplicate data. After

they identify duplicate data, they retain one copy of each element. They create pointers for the duplicate items,and discard the duplicate items.

- The effectiveness of data deduplication depends on many variables, including the rate of data change, the number of backups, and the data retention period.
- For example, if you back up the same incompressible data one time a week for six months, you save the first copy and you do not save the next 24. This method provides a 25:1 data deduplication ratio. If you back up an incompressible file on week one, back up the exact same file again on week two, and never back it up again, this method provides a 2:1 data deduplicationratio.
- A more likely scenario is that a portion of your data changes from backup to backup so that your data deduplication ratio changes over time.
- With data deduplication, you can minimize your storage requirements. Data deduplication can provide greater data reduction and storage space savings than other existing technologies.
- Figure 6-13 shows the concept of data deduplication.

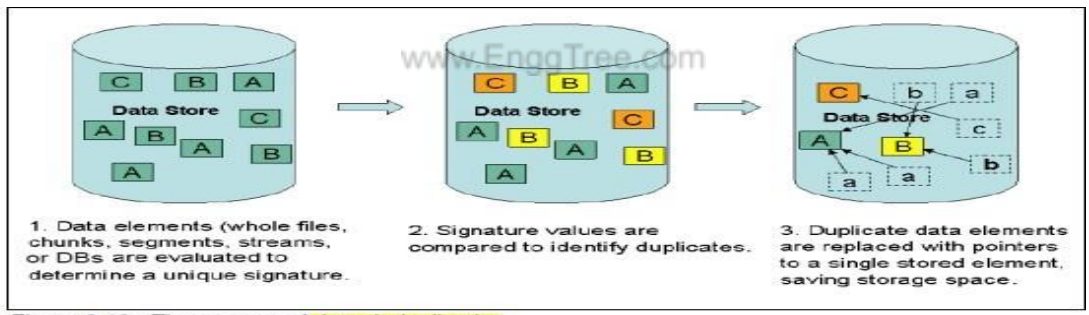


Figure 6-13 The concept of data deduplication

- Data deduplication can reduce your storage requirements but the benefit you derive is determined by your data and your backup policies. Workloads with a high database content have the highest data deduplication ratios.
- However, product functions, such as IBM Tivoli Storage Manager Progressive Incremental or Oracle Recovery Manager (RMAN), can reduce the data deduplication ratio.
- Compressed, encrypted, or otherwise scrambled workloads typically do not benefit from data deduplication.
- Good candidates for data deduplication are text files, log files, uncompressed and

non-encrypted database files, email files (PST, DBX, and IBM Domino®), and Snapshots (Filer Snaps, BCVs, and VMware images).

Types of data deduplication and IBM Hyper Factor

- Many vendors offer data deduplication products.
- Various methods are available to deduplicate data.
- The following *three methods* are used frequently for data deduplication:

✓ **Hash-based data deduplication** uses a hashing algorithm to identify chunks of data. Secure Hash Algorithm 1 (SHA-1) or Message-Digest Algorithm 5 (MDA-5) is commonly used. The details of each technique are beyond the intended scope of this publication.

✓ **Content-aware data deduplication** methods are aware of the structure of common patterns of data that is used by applications. The content-aware data deduplication method assumes that the best candidate to deduplicate against is an object with the same properties, such as a file name. When a file match is identified, a bit-by-bit comparison is performedto determine whether data changed and the changed data is saved.

✓ **IBM HyperFactor**® is a patented technology that is used in the IBM System Storage ProtecTIER Enterprise Edition and higher software. HyperFactor takes an approach that reduces the phenomenon of missed factoring opportunities, providing a more efficient process. HyperFactor data deduplication uses a 4 GB Memory Resident Index to track similarities for up to 1 petabyte (PB) of physical disk in a single repository.

HyperFactor technology uses a pattern algorithm that can reduce the amount of space that is required for storage by up to a factor of 25, based on evidence from existing implementations.

Figure 6-14 shows the HyperFactor technology.

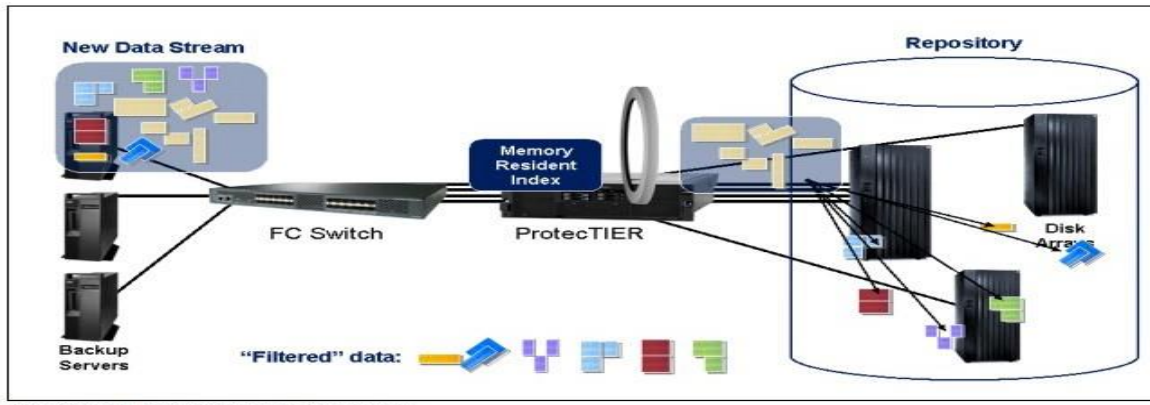


Figure 6-14 IBM HyperFactor technology

Data deduplication processing

- Data deduplication can either occur while the data is backed up to the storage media (real-time or inline) or after the data is written to the storage media (post-processing). Each method contains positive and negative aspects.
- These considerations must be evaluated by the engineer or technical specialist that is responsible for the concrete solution architecture and deployment. IBM decided to use inline data.
- deduplication processing because it offers larger target storage space without any need of atemporary disk cache pool for post-processed deduplication data.
- Bit comparison techniques, such as the technique that is used by Protec TIER, were designed to provide 100% data integrity by avoiding the risk of hash collisions.

CLOUD BASED AND MOBILE DEVICE BACKUPCLOUD BACKUP

- Cloud backup, also known as *online backup* or *remote backup*, is a strategy for sending a Copy of a physical or virtual file or database to a secondary, off-site location forpreservation in case of equipment failure, site catastrophe or human malfeasance.
- The backup server and data storage systems are usually hosted by a third-party

cloud or SaaS provider that charges the backup customer a recurring fee based on storage space or capacityused, data transmission bandwidth, number of users, number of servers or number of times data is retrieved.

- Implementing cloud data backup can help bolster an organization's data protection, business continuance and regulatory compliance strategies without increasing the workload of IT staff.

- There are a variety of approaches to cloud backup, with available services that can easily fit into an organization's existing data protection process. *Varieties of cloud backupinclude the following:*

Backing up directly to the public cloud.

- One way to store organizational workloads is by duplicating resources in the public cloud. This method entails writing data directly to cloud providers, such as AWS, Google Cloudor Microsoft Azure.
- The organization uses its own backup software to create the data copy to send to the cloud storage service. The cloud storage service then provides the destination and safekeeping for the data, but it doesn't specifically provide a backup application.
- In this scenario, it's important that the backup software is capable of interfacing with the cloud's storage service.
- Additionally, with public cloud options, IT professionals might need to look into supplemental data protection procedures, such as data encryption as well as identity and access management to secure backed up data.

Backing up to a service provider.

- In this scenario, an organization writes data to a cloud service or SaaS provider that offers backup services in a managed data center.
- The backup software that the company uses to send its data to the service might be provided as part of the service, or the service might support specific commercially available backup applications.

Choosing a cloud-to-cloud (C2C) backup.

- These services are among the newest offerings in the cloud backup arena. They specialize in backing up data that already lives in the cloud, either as data created using a SaaS application or as data stored in a cloud backup service.
- As its name suggests, a C2C backup service copies data from one cloud to another cloud. The cloud-to-cloud backup service typically hosts the software that handles this process.

Using online cloud backup systems.

- There are also hardware alternatives that facilitate backing up data to a cloud backup service. These appliances are all-in-one backup machines that include backup software anddisk capacity, along with the backup server.
- The appliances are about as close to plug-and-play as backup gets, and most of them also provide a seamless link to one or more cloud backup services or cloud providers.
- The list of vendors that offer backup appliances that include cloud interfaces is long, with Quantum, Unitrends, Arcserve, Rubrik, Cohesity, Dell EMC, StorageCraft and Asigra active in this arena.
- These appliances typically retain the most recent backup locally, in addition to copying it to the cloud backup provider, so any required recoveries can be made from the localbackup copy, saving time and transmission costs.

How data is restored

- Cloud backup services are typically built around a client software application that runs on aschedule determined by the purchased level of service and the customer's requirements.
- For example, if the customer has contracted for daily backups, the application collects, compresses, encrypts and transfers data to the cloud service provider's servers every 24 hours. To reduce the amount of bandwidth consumed and the time it takes to transfer files, the service provider might only provide incremental backups after the initial full backup.
- Cloud backup services often include the software and hardware necessary to protect an
- organization's data, including applications for Microsoft Exchange and SQL Server.

- Whether a customer uses its own backup application or the software the cloud backup service provides, the organization uses that same application to restore backed up data.

Cloud backup vs. cloud storage

- Although they share similarities, cloud backup and cloud storage aren't the same thing.
- Cloud storage is a service model in which data is stored on remote systems. Data in cloud storage is available to users over a network, typically the internet.
- Benefits of cloud storage include global availability, ease of use and off-site security. Potential drawbacks range from performance issues depending on network connection, to loss of complete control over the data, to escalating costs over time.
- Cloud backup is a service that sends an extra copy of an organization's data over a network
- to an off-site server, the typical user shouldn't need to access that data on a regular basis..

How does Cloud Backup Work?

- Cloud backup copies and stores data from a computer or other computing device to remote servers maintained by a cloud storage provider. For security reasons, the data is encrypted and delivered via the internet, guaranteeing that only authorized users may access thebackup data.

Here's how cloud backup actually works:

Installation

The first step is to install the cloud backup software on the device or devices that will be backed up. Typically, the steps will walk you through the setup process and assist you in configuring backup choices such as which data should be saved and how frequently should backups be performed.

Backup

Once the cloud backup software is installed, it will automatically copy and store your data on theremote servers.

This process is typically performed in the background, so you can continue to

use your devicewhile the backup is in progress.

Encryption

Data is encrypted before it is delivered over the internet to guarantee that it is safe from illegal access. The encryption method employs a one-of-a-kind key produced by the cloud backupprogram, and only the user has access to it.

Storage

After the data has been backed up, it is stored on remote servers operated by the cloud storage provider. The data is kept in a safe, off-site location, which adds an extra degree of security againstdata loss due to hardware failure, theft, or other sorts of calamities.

Recovery

To restore your data, just log into the cloud backup service and choose the files you want to recover. The data will subsequently, be sent from distant servers to your device. This technique is often quick and simple, and it does not require physical storage media or particular technological knowledge.

MOBILE DEVICE BACKUP

Back up or restore data on your Android device

- You can back up content, data, and settings from your phone to your Google Account. You can restore your backed up information to the original phone or to some other Android phones.
- You can't use back up when you set up a personal device with a work profile or for work only, or when you set up a company-owned device.
- Restoring data varies by phone and Android version. You can't restore a backup from a higher Android version onto a phone running a lower Android version.

Where your phone data is stored

- Backups are uploaded to Google servers and they're encrypted with your Google Account password. For some data, your phone's screen lock PIN, pattern, or password is also usedto encrypt your data so it can be backed up safely.

Your backup data (except what you back up to Google Photos) is erased if:

- You don't use your device for 57 days

- You turn off Android backup

Back up content

1. Back up photos and videos.
2. Back up files and folders.

Automatically back up your phone

- To help protect your backed-up data, use a PIN, pattern, or password screen lock, insteadof a swipe or Smart Lock.
- You can set up your device to automatically back up your files.

1. Open your device's Settings app.
2. Select **Google** › **Backup**.

Tip: If this is your first time, turn on **Backup by Google One** and follow the on-screeninstructions.

3. Tap **Back up now**.

Your Google One backup can take up to 24 hours. When your data is saved, "On" will be belowthe data types you selected.

Erase after backing up

- After you back up, you can reset your device by erasing everything on it.

Get your data onto a new phone

When you add your Google Account to a phone that's been set up, what you'd previously backed up for that Google Account gets put onto the phone.

- To restore a backed-up account to a reset phone, follow the on-screen steps. For morehelp, get help from your device manufacturer.
- Your photos and videos are already available in Google Photos. But you can restore the rest of the data you backed up while you set up your new phone for the first time or after a factory reset.
- At setup, to restore your data, follow the on-screen steps.The process can take up to 24 hours.

How Backup handles your data

- The data that backup collects is encrypted in transit.
- Backup sends your data to Google's backup servers and helps you transfer data between devices. Backup collects certain information to perform services on your device. Some of this functionality uses Google Play services.
- For example, backup collects:

Messages, contacts, app settings, and preferences are collected as part of your personalbackup.

DATA ARCHIVE :

An electronic data archive is a repository for data that has fewer access requirements.

Types of Archives :

It can be implemented as online, nearline, or offline based on the means of access:

■ **Online archive**: The storage device is directly connected to the host to make the dataimmediately available. This is best suited for active archives.
■ **Nearline archive**: The storage device is connected to the host and infor- mation is local, but thedevice must be mounted or loaded to access the information.
■ **Offline archive**: The storage device is not directly connected, mounted, or loaded. Manualintervention is required to provide this service before information can be accessed.

- An archive is often stored on a write once read many (WORM) device, such as a CD-ROM. These devices protect the original file from being overwritten. Some tape devices also provide this functionality by implementing file locking capabilities in the hardware or software.
- Although these devices are inexpensive, they involve operational, management, and maintenance overhead.
- Requirements to retain archives have caused corporate archives to grow at a rate of 50 percent or more per year. At the same time, organizations must reduce costs while maintaining required service-level agreements (SLAs). Therefore, it is essential to find a solution that minimizes the fixed costs of the archive's

operations and management.

- Archives implemented using tape devices and optical disks involve many hidden costs.
- The traditional archival process using optical disks and tapes is not optimized to recognize the content, so the same content could be archived several times.
- Additional costs are involved in offsite storage of media and media management. Tapes and optical media are also susceptible to wear and tear. Frequent changes in these device technologies lead to the overhead of converting the media into new formats to enable access and retrieval.
- Government agencies and industry regulators are establishing new laws and regulations to enforce the protection of archives from unauthorized destruction and modification.
- These regulations and standards affect all businesses and have established new requirements for preserving the integrity of information in the archives.
- These requirements have exposed the hidden costs and shortcomings of the traditional tape and optical media archive solutions

REPLICATION :

- Replication is the process of creating an exact copy of data. Creating one or more replicasof the production data is one of the ways to provide Business Continuity (BC).
- *Data replication*, where the same data is stored on multiple storage devices

Benefits of Data Replication

Data replication can be a cost-demanding process/operation in terms of computing power and storage requirements, but it provides an immense set of benefits that overshadow the cost aspect. *Some of the benefits of data replication are as follows:*

- **High Data Availability:** Data replication mechanisms ensures high availability and accessibility of the data by allowing users or applications to access the data from numerousnodes or sites even during an unforeseen failure or technical glitch. It stores data across multiple locations and thus enhances the reliability of systems.

- **Enhanced Data Retrieval:** With data replication in place, users can access data from a diverse set of regions/locations. With data available across different storage locations, data replication reduces latency and allows users to access data from a nearby data replica.
- **Enhanced Server Performance:** Data replication helps reduce the load on the primary server by distributing data across numerous storage regions/locations, thereby boosting the network performance.
- **Fault tolerance & Disaster Recovery:** With the rapid growth in the number of cyberattacks, data breaches, etc., most organizations face the issue of unexpected losses.

Uses of Data Replication

- One common use of data replication is for disaster recovery, to ensure that an accurate backup exists at all times in case of a catastrophe, hardware failure, or a system breachwhere data is compromised.
- Having a replica can also make data access faster, especially in organizations with a largenumber of locations.

COMPUTE BASED, STORAGE-BASED, AND NETWORK-BASED REPLICATION :

STORAGE-BASED REPLICATION (REPLICATION STORAGE)

- Replication Storage, also known as storage-based replication, is an approach to replicating data available over a network to numerous distinct storage locations/regions.
- It enhances the availability, accessibility, and retrieval speed of data by allowing users to access data in real-time from various storage locations when unexpected failures occur at the source storage location.

- Storage-based data replication makes use of software installed on the storage device to handle the replication.

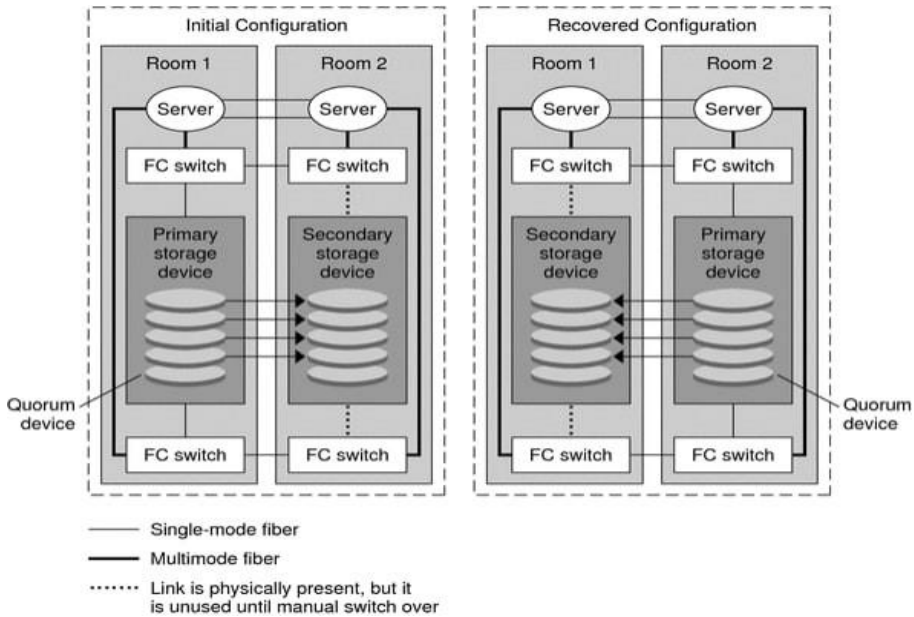


Image Source

- Storage system-based replication supports both local and remote replication.
- In storage system-based local replication, the data replication is carried out within the storage system.
- Local replication enables you to perform recovery operations in the event of data loss and also provides support for backup.
- Whereas in storage system-based remote replication, the replication is carried out between storage systems. In simple words, one of the storage systems is on the source site and the other storage system is on a remote site for data replication.
- Data can be transmitted between the two storage systems over a shared or dedicated network.

Advantages of Storage-Based Replication

- Storage-based replication follows a heterogenous storage mechanism and hence housessupport for numerous platforms.
- It operates independently of any server or storage-based device.

- It allows replicating data across multi-vendor products.

Disadvantages of Storage-Based Replication

- Setting up storage-based replication requires you to leverage proprietary hardware andhence, it has a high initial setup, operational, and management cost.
- It requires setting up and implementing a storage area network (SAN).

HOST-BASED DATA REPLICATION (COMPUTE BASED)

- Host-based data replication uses the servers to copy data from one site to another site. Host-based replication software usually includes options like compression, encryption and, throttling, as well as failover.

Advantages of Host-Based Replication

- Flexible: It can leverage existing IP networks
- Can be customized to your business' needs: You can choose what data to replicate
- Can create a schedule for sending data: allows you to throttle bandwidth
- Can use any combination of storage devices on each end

Disdvantages of Host-Based Replication

- Difficult to manage with a large group of servers if there is no centralized managementconsole
- Consumes host resources during replication
- Both storage devices on each end need to be active, which means you will need to purchasededicated hardware and OS
- Not all applications can support this type of data replication
- Can be affected by viruses or application failure

Use-Cases of Host-Based Replication

- Host-based replication finds application in various scenarios where organizations require flexible and granular control over data replication.

- Some common use cases include:

Application-specific replication: Host-based replication allows organizations to replicate specific applications or databases, ensuring data consistency and availability. For example, a company running a critical database application may utilize host-based replication to replicate the database to a secondary site for disaster recovery purposes.

Virtual machine replication: In virtualized environments, host-based replication is commonly used to replicate virtual machines (VMs) to remote hosts or data centers. This ensures VM availability in case of host failures or enables migration of VMs for load balancing purposes.

File and folder replication: Host-based replication enables the replication of specific files, directories, or folders based on predefined rules or policies. This is useful for organizations that need to replicate specific data sets, such as project files, user home directories, or shared folders, to remote locations for backup or collaboration purposes.

Cross-platform replication: Host-based replication offers the advantage of supporting heterogeneous environments, allowing replication between different operating systems and file systems. This is beneficial for organizations with mixed IT environments that require replication between Windows, Linux, or Unix-based hosts.

Data migration and consolidation: Host-based replication can be used for data migration or consolidation projects. It allows organizations to replicate data from multiple sources to a centralized target host or storage infrastructure. This is useful during system upgrades, data center migrations, or storage platform transitions.

Content distribution and caching: Host-based replication is employed in content delivery networks (CDNs) or caching solutions to distribute content closer to end-users. By remote sites or backup locations, organizations can ensure continuous operations and minimizedowntime in the event of hardware failures, natural disasters, or other disruptions.

NETWORK-BASED DATA REPLICATION

- Network-based replication is a data replication technique that operates at the network layer. It involves replicating data between source and target systems over a

networkinfrastructure. Unlike array-based or host-based replication, network-based replication is not tightly coupled to storage arrays or hosts but focuses on replicating data at the network level.

- In network-based replication, data is captured and replicated at the application or file system level. It intercepts the input/output (I/O) operations at the network layer, captures the changes made to data, and replicates them to the target system.
- This replication method allows for the replication of specific files, folders, or even individual application transactions.
- Network-based replication can be synchronous or asynchronous.
- In synchronous replication, the data changes are replicated to the target system immediately after they occur on the source system, ensuring a consistent copy of the data atall times.
- This method provides a higher level of data integrity but may introduce some latency due to the delay in acknowledging the write operation until the data is replicated.
- Asynchronous replication, on the other hand, introduces a slight delay between the data changes on the source and their replication to the target system. This delay allows for increased distance between the source and target systems, as well as a higher tolerance for network latency.
- Asynchronous replication is suitable for scenarios where minimal data loss is acceptable, and the focus is on optimizing performance and network utilization.
- Network-based data replication uses a device or appliance that sits on the network in the path of the data to manage replication.
- The data is then copied to a second device. These devices usually have proprietary replication technology but can be used with any host server and storage hardware.

Advantages of network-based data replication:

- Effective in large, heterogeneous storage and server environments
- Supports any host platform and works with any array
- Works separately from the servers and the storage devices
- Allows replication between multi-vendor products

Disadvantages of network-based data replication:

- Higher initial set-up cost because it requires proprietary hardware, as well as ongoingoperational and management costs
- Requires implementation of a storage area network (SAN)
- Some common use cases include:

Disaster Recovery: Network-based replication is widely used for disaster recovery purposes. Organizations replicate critical data and applications from their primary data center to a secondary or remote site. In the event of a disaster or site failure, the replicated data can be quickly activated, allowing for business continuity and minimal data loss.

Multi-site Deployments: Organizations with multiple geographically dispersed locations often utilize network-based replication to keep data synchronized across sites. This enables seamless collaboration, data sharing, and consistent access to up-to-date information. It particularly benefits distributed enterprises, branch offices, and global organizations.

Data Migration: When migrating data from one system or infrastructure to another, network- based replication simplifies the process. It allows for the smooth transfer of data, ensuring minimal downtime and disruption. Organizations can replicate data from the source system to the target system, validate its integrity, and seamlessly transition to the new environment.

High Availability and Load Balancing: Network-based replication is employed to achieve high availability and load balancing in environments where continuous data access and minimal downtime are critical. By replicating data across multiple systems, organizations can distribute the workload, handle increased traffic, and maintain service availability even in the event of hardware or system failures.

DevOps and Testing Environments: Network-based replication facilitates the creation of reliable and consistent testing environments. Development and testing teams can replicate production data to their test environments, ensuring realistic testing scenarios without impacting the production environment. This enables thorough testing, debugging, and validation of applications and infrastructure changes.

Data Archiving and Compliance: Network-based replication supports long-term data archiving and compliance requirements. Organizations can replicate data

to dedicated archival systems or cloud storage for regulatory compliance, data retention policies, or legal obligations. It ensures data integrity, security, and availability for archival purposes.

Cloud Data Replication: With the growing adoption of cloud services, network-based replication plays a crucial role in replicating data from on-premises environments to cloud-based infrastructure. Organizations can replicate data to the cloud for backup, disaster recovery, or as part of hybrid cloud strategies. It enables seamless data movement between on-premises and cloud environments.

DATA MIGRATION

- In general, data migration means moving digital information.
- Transferring that information to a different location, file format, environment, storage system, database, datacenter, or application all fit within the definition of data migration.
- Data migration is the process of selecting, preparing, extracting, and transforming data and permanently transferring it from one computer storage system to another.
- Data migration is a common IT activity. However, data assets may exist in many different states and locations, which makes some migration projects more complex and technically challenging than others.

- Examples of data assets include:

✓ Unorganized assortments of files stored across many different devices.
✓ Applications, operating systems, and environments.
✓ Relational databases like SQL Server, MySQL, PostgreSQL, and MariaDB.
✓ Unstructured databases such as MongoDB, Azure Cosmos DB, DocumentDB,Cassandra, Couchbase, HBase, Redis, and Neo4j.
✓ Data lakes, data blobs, and entire datacenters.

- Data migration projects require planning, implementation, and validation to ensure theirsuccess.

Importance of Data migration

- Data migration ensures that data is successfully and securely transferred to anotherapplication, storage system or cloud.
- Although moving data from one platform to another can be risky and costly, it alsoprovides an organization with numerous benefits.
- For example, in addition to upgrading applications and services, organizations can boosttheir productivity and reduce storage costs.

Types of data migrations and their challenges

Data migration is typically performed using one of the following methods:

- **Storage migration** transfers data from one storage device to another. This involves moving blocks of storage and files from storage systems, whether they're on disk, tape or in the cloud. During migration is also an optimal time for organizations to perform data validation and reduction by identifying obsolete or corrupt data.
- **Database migrat**ion moves database files to a new device. This is done when

organization changes database vendors, upgrades the database software or moves a database to the cloud. Databases must be backed up before migrating.

- **Application migration** moves an application or program from one environment to another.

Application migration typically occurs when an organization switches to another vendor, application or platform. This process is complex because applications interact with other applications, and each one has its own data model. Successful application migration may require using middleware products to bridge technology gaps.

Cloud migration moves data or applications from an on-premises location to the cloud or from one cloud service to another. Cloud migration is a common form of data migration. Cloud environments provide on-demand flexibility and scalability and reduce the capital expenditure (Capex) for on-premises infrastructures. Public cloud providers offer a variety of services for storage, database and application migrations.

- **Business process migration** moves business applications -- including

customer, product
and operational data -- and processes to a new environment.

During data migrations, teams must pay careful attention to the following challenges:

Source data. Not preparing the source data being moved might lead to data duplicates, gaps or errors when it's brought into the new system or application.

Wrong data formats. Data must be opened in a format that works with the system. Files might not have access controls on a new system if they aren't properly formatted before migration.

Mapping data. When stored in a new database, data should be mapped in a sensible wayto minimize confusion.

Sustainable governance. Having a data governance plan in place can help organizations track and report on data quality, which helps them understand the integrity of their data.

Security. Maintaining who can access, edit or remove data is a must for security.

Data migration strategies

Although implementation differs by migration type, there are still **two main strategies**
organizations use: ***big bang* and *trickle* migrations.**

- **Big bang migrations** transfer all associated data within a set time window. The advantages of creating a migration strategy around this method include lower cost, a quicker move andless complexity. The downside, however, is that big bang migrations require the system to be offline for the entire migration. There's also a risk of losing data if it isn't properly backed up to another location ahead of time.
- **Trickle migrations** complete a data migration within phases. During the migration, both old and new systems run at the same time, so there's no downtime, which means there's less risk of losing data. However, trickle migrations are more complicated and need more planning and time to implement properly.

How to create a data migration plan

- A data migration project can be challenging because administrators must maintain data integrity and time the project so there's minimal effect on the business and they can keep aneye on costs.
- Having a data migration plan helps to ensure there's minimal disruption and downtime to business processes.
- Factors to consider during a data migration project include how long the migration will take, the amount of downtime required, and the risk to the business due to technical compatibility issues, data corruption or application performance.

Phases of Data Migration

Discovery. This should include considerations such as data sources, destinations, security, costand which migration strategy to use.

Resource assessment. Identify who will be taking part in the migration.

Data inspection. Examine the data being migrated for data quality, anomalies or duplications.Data should also be backed up.

Design. Data is organized and mapped out for where it's being moved to.

Software tools. Any software that will help in the transition is purchased or created.

Migration. The migration process is initiated.

Cleanup. Old or legacy systems are shut down and decommissioned.

Examples of data migration tools

Microsoft SQL, AWS Data Migration Service, Varonis DatAdvantage and Varonis DataTransport Engine.

There are ***three broad categories*** of data movers: host-based, array-based and network appliances. **Host-based software** is best for application-specific migrations, such as platform upgrades,database replication and file copying. **Array-based-based software** is primarily used to migrate data between similar systems. **Network appliances** migrate volumes, files or blocks of data depending on theirconfiguration.

Data migration vs. data integration vs. data conversion

✓ Data migration is the process of transferring data between data storage systems or formats,

✓ Data integration is the process of combining data from multiple source systems -- creating a unified set of data for operational and analytical uses. The primary goal of data integration is to produce consolidated data sets that are clean and consistent. Integration isa core element of the data management process.

✓ Data conversion is the process of changing data from one format to another. If a legacy

system and a new system have identical fields, an organization could just do a data migration; however, the data from the legacy system is generally different and needs to be modified before migrating. Data conversion is often a step in the data migration process.

DISASTER RECOVERY AS A SERVICE (DRAAS)

- Disaster recovery as a service(DRaaS) is a cloud computing service model that allows an organization to back up its data and IT infrastructure in a third party cloud computing environment and provide all the DR orchestration, all through a SaaS solution, to regain access and functionality to IT infrastructure after a disaster.
- The as-a-service model means that the organization itself doesn't have to own all the

resources or handle all the management for disaster recovery, instead relying on the service provider.

- Disaster recovery planning is critical to business continuity.
- Many disasters that have the potential to wreak havoc on an IT organization have become more frequent in recent years:
- Natural disasters such as hurricanes, floods, wildfires and earthquakes
- Equipment failures and power outages
- Cyberattacks

Using DRaaS to prepare for a disaster

- True DRaaS mirrors a complete infrastructure in fail-safe mode on virtual servers, includingcompute, storage and networking functions.

An organization can continue to run applications—it just runs them from the service provider's cloud or hybrid cloud environment instead of from the disaster-affected physical servers. This means recovery time after a disaster can be much faster, or even instantaneous running from the cloud instead of from an on-site server, but the total business cost of downtime can be very high, so it's imperative that the business can get back up and running.

How does disaster recovery as a service work?

- DRaaS works by replicating and hosting servers in a third-party vendor's facilities versus in the physical location of the organization that owns the workload. The disaster recovery plan is executed on the third-party vendor's facilities in the event of a disaster that shuts down a customer's site.
- Organizations may purchase DRaaS plans through a traditional subscription model or a

pay-per-use model that allows them to pay only when disaster strikes.

- As-a-service solutions vary in scope and cost—organizations should evaluate potential DRaaS providers according to their own unique needs and budget.

DRaaS can save organizations money by eliminating the need for provisioning and maintaining an organization's own off-site disaster recovery environment. However, organizations should evaluate and understand service level agreements.
For instance, what happens to recovery times if both the provider and customer are affected by the same natural disaster, such as a large hurricane or earthquake. Different DRaaS providers have different policies on prioritizing which customers get help first in a large regional disaster or allowing customers to perform their own disaster recovery testing.

Advantages of Disaster recovery as a service

- Many businesses with lean IT teams simply can't afford to take the time needed toresearch, implement and fully test disaster recovery plans. DRaaS takes the burden of planning for a disaster off of the organization and puts it into the

hands of experts in disaster recovery.

- It can also be much more affordable than hosting your own disaster recovery infrastructure

in a remote location with an IT staff standing by if disaster strikes.

- If a disaster doesn't happen, that expensive second infrastructure and staff never get used. Many DRaaS providers charge you only if you need their services.
- For many organizations, DRaaS is a helpful solution to a nagging problem.

Is disaster recovery as a service right for you?

- Organizations may choose to hand over all or part of their disaster recovery planning to a DRaaS provider.
- There are many different disaster recovery as a service providers to choose from, with **three mainmodels:**
- **Managed DRaaS:** In a managed DRaaS model, a third party takes over all responsibility fordisaster recovery. Choosing this option requires an organization to stay in close contact with their DRaaS provider to ensure that it stays up to date on all infrastructure, application and services changes. If you lack the expertis

or time to manage your own disaster recovery,this may be the best option for you.

Assisted DRaaS: If you prefer to maintain responsibility for some aspects of your disaster recovery plan, or if you have unique or customized applications that might be challenging for a third party to take over, assisted DRaaS might be a better option. In this model, the service provider offers its expertise for optimizing disaster recovery procedures, but the customer is responsible for implementing some or all of the disaster recovery plan.

Self-service DRaaS: The least expensive option is self-service DRaaS, where the customer is responsible for the planning, testing and management of disaster recovery, and the customer hosts its own infrastructure backup on virtual machines in a remote location. Careful planning and testing are required to make sure that processing can fail over to the virtual servers instantly in the event of a disaster. This option is best for those who have experienced disaster recovery experts on staff.

✓ Whichever of these models suits you, VMware has a solution.

✓ If you would drive your own DRaaS solution to your own target DR site, you can consider solutions like Site Recovery Manager and VMware vSphere Replication.

✓ If you would like a service provider to assist you with DR, whether fully managed or self service, consider VMware Cloud Director Availability from one of our DRaaS Validate partners

DRaaS vs. BaaS :

- With **disaster recovery as a service**, the service provider moves an organization's computer processing to its cloud infrastructure in the event of a disaster. This way, the business can continue to operate, even if the original IT infrastructure is totally destroyedor held hostage.
- This differs from **backup as a service**, where only the data, but not the ability to process

the data, is duplicated by a third-party provider.

- Because BaaS is only protecting the data, and not the infrastructure, it is typically less expensive than DRaaS. BaaS can be a good solution for companies that need to archive data or records for legal reasons, but most organizations who use BaaS will want to combine it with another disaster recovery tool to ensure business continuity.
- Planning for disaster and getting the help you need is something every business needs to

consider. Whatever option you choose, a disaster recovery plan is essential for businesscontinuity, and organizations are increasingly turning to DRaaS.

UNIT V
SECURING STORAGE INFRASTRUCTURE

Information security goals, Storage security domains, Threats to a storage infrastructure, Security controls to protect a storage infrastructure, Governance, risk, and compliance, Storage infrastructure management functions, Storage infrastructure management processes.

INFORMATION SECURITY GOALS

- In Information security, it is a collection of practices intended to convey personal information secure from unapproved access and modification throughout of storing or broadcasting from one place to another place.
- Information security is designed and required to secure the print, digital, and some personal, sensitive, and private information from unapproved persons. It very well may be utilized to get information from being misused, affirmation, destruction, modification, and interruption.

- There are the major goals of information security which are as follows −

Confidentiality − The goals of confidentiality is that only the sender and the predetermined recipient should be adequate to approach the element of a message. Confidentiality have negotiated if an unauthorized person is capable to create the message.
For example, it can be a confidential email message sent by user A to user B, which is penetrated by user C without the authorization or knowledge of A and B. This kind of attack is known as interception.

Integrity − When the element of a message are transformed after the sender sends it, but since it reaches the intended recipient, and it can said that the principle of the message is lost.
For example, consider that user A sends message to user B and User C alter with a message basically sent by user A, which is absolutely intended for user B.
User C somehow handles to access it, modify its elements and send the changed message to user B. User B has no method of understanding that the element of the message changed after user A had sent it. User A also does not understand about this change. This kind of attack is known as modification.

Availability − The main goals of information security is availability. It is that resources must be available to authorized parties at all times.
For instance, because of the intentional actions of an unauthorized user C, an authorized user A cannot allow contact a server B. This can overthrow the principle of availability. Such an attack is known as interruption.

STORAGE SECURITY DOMAINS

- Storage devices that are not connected to a storage network are less vulnerable because they are not exposed to security threats via networks. However, with increasing use of networking in storage environments, storage devices are becoming highly exposed to security threats from a variety of sources.
- If each component within the storage network is considered a potential access point, one must analyze the attack surface that each of these access points provides and identify the associated vulnerability.
- In order to identify the threats that apply to a storage network, *access paths to data storage can be categorized into three security domains: application access,management access, and BURA (backup, recovery, and archive).*
- Figure 15-1 depicts the three security domains of a storage system environment.

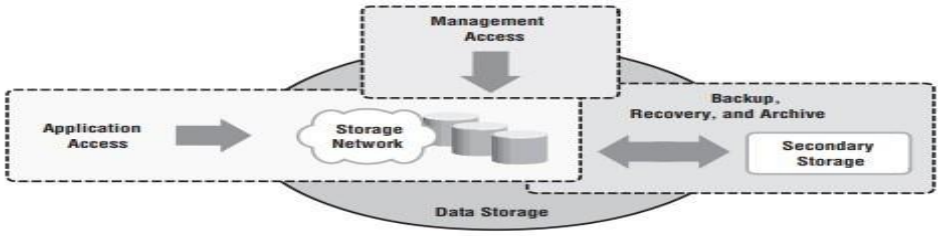


Figure 15-1: Three security domains of data storage

SECURING THE APPLICATION ACCESS DOMAIN

- The application access domain may include only those applications that access the data through the file system or a database interface.
- Figure 15-2 shows application access in a storage networking environment. Host A can access all V1 volumes; host B can access all V2 volumes.
- These volumes are classified according to access level, such as confidential, restricted, and public. Some of the possible threat in this scenario could be host A spoofing the identity or elevating the privileges of host B to gain access to host B's resources.
- Another threat could be an unauthorized host gain access to the network; the attacker on this host may try to spoof the identity of another host and tamper with data, snoop the network, or execute a DoS attack.
- Also any form of media theft could also compromise security.

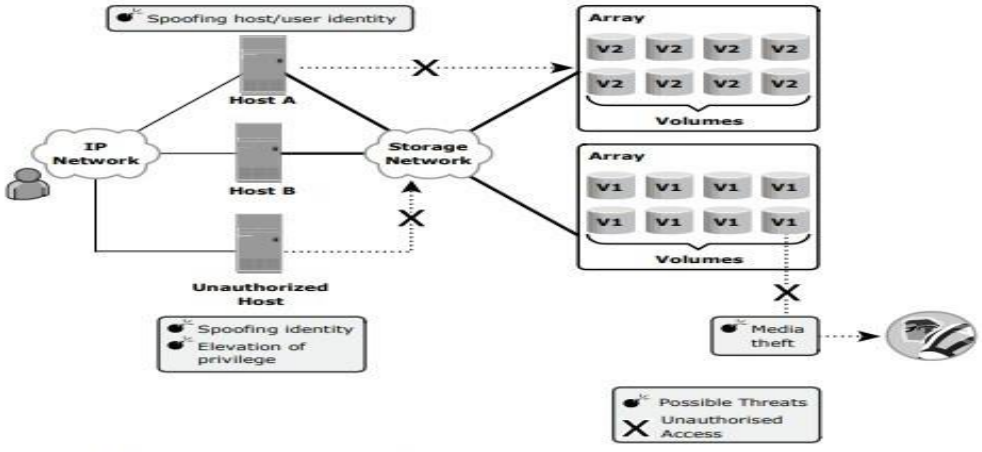


Figure 15-2: Security threats in application access domain

Controlling User Access to Data

- Access control services regulate user access to data. These services mitigate the threats of spoofing host identity and elevating host privileges. Both of these threats affect data integrity and confidentiality.
- Technical control in the form of user authentication and administrative control in the form of user authorization are the two access control mechanisms used in application access control

Protecting the Storage Infrastructure

- Securing the storage infrastructure from unauthorized access involves protecting all the elements of the infrastructure.
- Security controls for protecting the storage infrastructure address the threats of unauthorized tampering of data in transit that leads to a loss of data integrity, denial ofservice that compromises availability, and network snooping that may result in a loss of confidentiality.
- The security controls for protecting the network fall into two general categories: connectivity infrastructure integrity and storage network encryption

Data Encryption

- The most important aspect of securing data is protecting data held inside the storage arrays. Threats at this level include tampering with data, which violates data integrity, and media theft, which compromises data availability and confidentiality.
- To protect against these threats, encrypt the data held on the storage media or encrypt the data prior to being transferred to the disk.

SECURING THE MANAGEMENT ACCESS DOMAIN

- Management access, whether monitoring, provisioning, or managing storage resources, is associated with every device within the storage network. Most management software supports some form of CLI, system management console, or a web-based interface.
- Figure 15-3 depicts a storage networking environment in which production hosts are connected to a SAN fabric and are accessing storage Array A, which is connected to storage Array B for replication purposes.
- Further, this configuration has a storage management platform on Host B and a monitoring console on Host A.

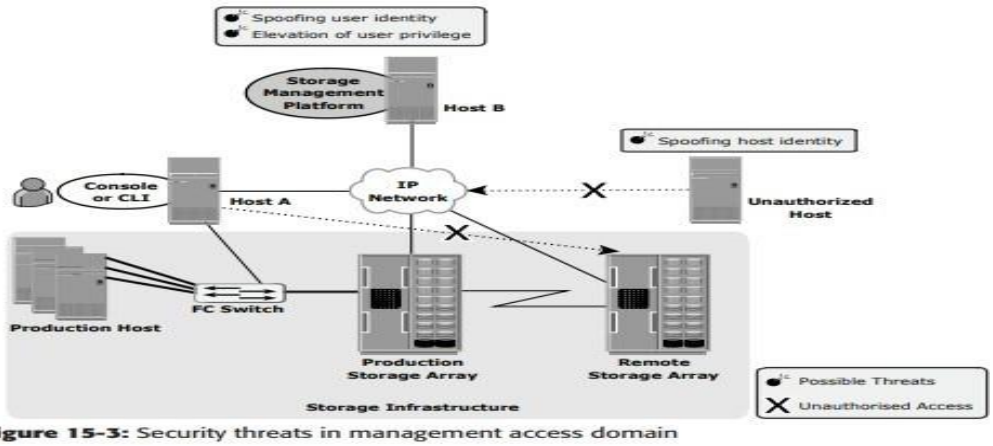


Figure 15-3: Security threats in management access domain

Controlling Administrative Access

- Controlling administrative access to storage aims to safeguard against the threats of an attacker spoofing an administrator's identity or elevating another user's identity and privileges to gain administrative access.
- Both of these threats affect the integrity of data and devices.
- To protect against these threats, administrative access regulation and various auditing techniques are used to enforce accountability.

Protecting the Management Infrastructure

- Protecting the management network infrastructure is also necessary. Controls to protect the management network infrastructure include encrypting management traffic, enforcing management access controls, and applying IP network security best practices.

SECURING BACKUP, RECOVERY, AND ARCHIVE (BURA)

- BURA is the third domain that needs to be secured against attack. A backup involves copying the data from a storage array to backup media, such as tapes or disks. Securing BURA is complex and is based on the BURA software accessing the storagearrays.
- It also depends on the configuration of the storage environments at the primary and secondary sites, especially with remote backup solutions performed directly on a remote tape device or using array-based remote replication.
- Protecting the BURA infrastructure requires addressing several threats, including spoofing the legitimate identity of a DR site, tampering with data, network snooping, DoS attacks, and media theft. Such threats represent potential violations of integrity, confidentiality, and availability.
- Figure 15-4 illustrates a generic remote backup design whereby data on a storage array is replicated over a disaster recovery (DR) network to a secondary storage at theDR site.

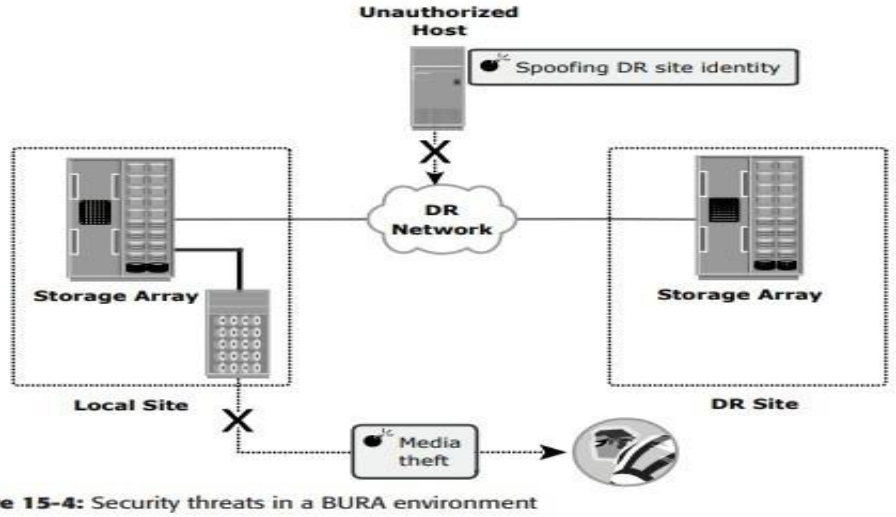


Figure 15-4: Security threats in a BURA environment

THREATS TO A STORAGE INFRASTRUCTURE

Risk Triad

- Risk triad defines the risk in terms of threats, assets, and vulnerabilities. Risk arises when a threat agent (an attacker) seeks to access assets by exploiting an existing vulnerability.
- To manage risks, organizations primarily focus on vulnerabilities because they cannot eliminate threat agents that may appear in various forms and sources to its assets. Organizations can install countermeasures to reduce the impact of an attack by a threat agent, thereby reducing vulnerability.
- Risk assessment is the first step in determining the extent of potential threats and risks in an IT infrastructure. To determine the probability of an adverse event occurring, threats to an IT system must be analyzed in conjunction with the potential vulnerabilities and the existing security controls.
- The severity of an adverse event is estimated by the impact that it may have on critical business activities. Based on this analysis, a relative value of criticality and sensitivity can be assigned to IT assets and resources.

Assets, threats, and vulnerability are considered from the perspective of risk identification and control analysis.

Assets

- Information is one of the most important assets for any organization. Other assets include hardware, software, and the network infrastructure required to access this information.
- To protect these assets, organizations must develop a set of parameters to ensure the availability of the resources to authorized users and trusted networks. These parameters apply to storage resources, the network infrastructure, and organizational policies.
- Several factors need to be considered when planning for asset security. Security methods have two objectives.
- *First objective is to ensure that the network is easily accessible to authorized users. It should also be reliable and stable under disparate environmental conditions and volumes of usage.*
- *Second objective is to make it very difficult for potential attackers to access and compromise the system. These methods should provide adequate protection against unauthorized access to resources, viruses, worms, Trojans and other malicious software programs.*

Threats

- Threats are the potential attacks that can be carried out on an IT infrastructure. These attacks can be classified as active or passive. Passive attacks are attempts to gain unauthorized access into the system.
- They pose threats to confidentiality of information. Active attacks include data modification, Denial of Service (DoS), and repudiation attacks. They pose threats to data integrity and availability. In a modification attack, the unauthorized user attempts to modify information for malicious purposes.

- A modification attack can target data at rest or data in transit. These attacks pose a threat to data integrity. Denial of Service (DoS) attacks denies the use of resources to legitimate users.
- These attacks generally do not involve access to or modification of information on the computer system. Instead, they pose a threat to data availability.
- The intentional flooding of a network or website to prevent legitimate access to authorized users is one example of a DoS attack. Repudiation is an attack against the accountability of the information.
- It attempts to provide false information by either impersonating someone or denying that an event or a transaction has taken place.
- Table 15-1 describes different forms of attacks and the security services used to manage them.

Table 15-1: Security Services for Various Types of Attacks

ATTACK	CONFIDENTIALITY	INTEGRITY	AVAILABILITY	ACCOUNTABILITY
Access	X			X
Modification	X	X		X
Denial of Service			X	
Repudiation		X		X

Vulnerability

- The paths that provide access to information are the most vulnerable to potential attacks. Each of these paths may contain various access points, each of which provides different levels of access to the storage resources.
- It is very important to implement adequate security controls at all the access points on an access path. Implementing security controls at each access point of every access path is termed as defense in depth.
- Attack surface, attack vector, and work factor are the three factors to consider when assessing the extent to which an environment is vulnerable to security threats. Attack surface refers to the various entry points that an attacker can use to launch an attack. Each component of a storage network is a source of potential vulnerability
- An attack vector is a step or a series of steps necessary to complete an attack. For example, an attacker might exploit a bug in the management interface to execute a snoop attack whereby the attacker can modify the configuration of the storage device to allow the traffic to be accessed from one more host.
- Work factor refers to the amount of time and effort required to exploit an attack vector.
- For example, if attackers attempt to retrieve sensitive information, they consider the time and effort that would be required for executing an attack on a database.
- The preventive control attempts to prevent an attack; the detective control detects whether an attack is in progress; and after an attack is discovered, the corrective controls are implemented.

- *Preventive controls avert the vulnerabilities from being exploited and prevent an attack or reduce its impact. Corrective controls reduce the effect of an attack, while detective controls discover attacks and trigger preventive or corrective controls.*

SECURITY CONTROLS TO PROTECT A STORAGE INFRASTRUCTURE

- Security controls for protecting the storage infrastructure address the threats of unauthorized tampering of data in transit that leads to a loss of data integrity, denial ofservice that compromises availability, and network snooping that may result in a loss of confidentiality.
- There are several types of security controls that can be implemented to protect hardware, software, networks, and data from actions and events that could cause loss or damage.
- For example:

Physical security controls include such things as data center perimeterfencing, locks, guards, access control cards, biometric access control systems, surveillance cameras, and intrusion detection sensors.

Digital security controls include such things as usernames and passwords, two-factor authentication, antivirus software, and firewalls.

Cybersecurity controls include anything specifically designed toprevent attacks on data, including DDoS mitigation, and intrusionprevention systems.

Cloud security controls include measures you take in cooperation with a cloud services provider to ensure the necessary protection for data and workloads. If your organization runs workloads on the cloud, you must meet their corporate or business policy security requirements *and* industry regulations.

GOVERNANCE, RISK, AND COMPLIANCE

- Governance, Risk, and Compliance (GRC) is a structured way to align IT with business goals while managing risks and meeting all industry and governmentregulations.
- It includes tools and processes to unify an organization's governance and risk management with its technological innovation and adoption.

Governance

Governance is the set of policies, rules, or frameworks that a company uses to achieve its business goals.

It defines the responsibilities of key stakeholders, such as the board of directors and senior management.

For example, good corporate governance supports your team in including the company's social responsibility policy in their plans.

Good governance includes the following:

- ✓ Ethics and accountability
- ✓ Transparent information sharing
- ✓ Conflict resolution policies
- ✓ Resource management

Risk management

Businesses face different types of risks, including financial, legal, strategic, and security risks. Proper risk management helps businesses identify these risks and find ways to remediate any that are found.

Companies use an enterprise risk management program to predict potential problems and minimize losses.

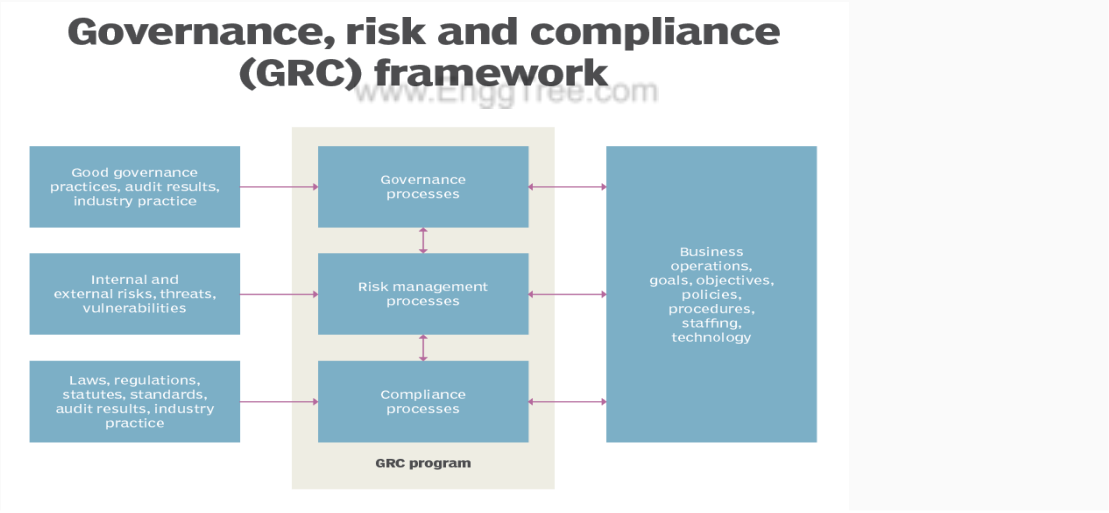
For example, you can use risk assessment to find security loopholes in your computer system and apply a fix.

Compliance

Compliance is the act of following rules, laws, and regulations. It applies to legal and regulatory requirements set by industrial bodies and also for internal corporate policies.

In GRC, compliance involves implementing procedures to ensure that business activities comply with the respective regulations.

For example, healthcare organizations must comply with laws like HIPAA that protect patients' privacy.



BENEFITS OF GRC :

- By implementing GRC programs, businesses can make better decisions in a risk-aware environment.
- An effective GRC program helps key stakeholders set policies from a shared perspective and comply with regulatory requirements.
- With GRC, the entire company comes together in its policies, decisions, and actions.

The following are some benefits of implementing a GRC strategy at your organization.

Data-driven decision-making

You can make data-driven decisions within a shorter time frame by monitoring your resources, setting up rules or frameworks, and using GRC software and tools.

Responsible operations

GRC streamlines operations around a common culture that promotes ethical values and creates a healthy environment for growth. It guides strong organizational culture development and ethical decision-making in the organization.

Improved cybersecurity

With an integrated GRC approach, businesses can employ data security measures to protect customer data and private information. Implementing a GRC strategy is essential for your organization due to increasing cyber risk that threatens users' data and privacy. It helps organizations comply with data privacy regulations like the General Data Protection Regulation (GDPR). With a GRC IT strategy, you build customer trust and protect your business from penalties.

IMPLEMENTATION OF GRC:

Companies of all sizes face challenges that can endanger revenue, reputation, andcustomer and stakeholder interest.

Some of these challenges include the following:

✓ Internet connectivity introducing cyber risks that might compromise data storage security

✓ Businesses needing to comply with new or updated regulatory requirements

✓ Companies needing data privacy and protection

✓ Companies facing more uncertainties in the modern business landscape

✓ Risk management costs increasing at an unprecedented rate

✓ Complex third-party business relationships increasing risk

WORKING OF GRC :

GRC in any organization works on the following principles:

Key stakeholders

GRC requires cross-functional collaboration across different departments thatpractices governance, risk management, and regulatory compliance.

Some examples include the following:

✓ Senior executives who assess risks when making strategic decisions

✓ Legal teams who help businesses mitigate legal exposures

✓ Finance managers who support compliance with regulatory requirements

✓ HR executives who deal with confidential recruitment information

✓ IT departments that protect data from cyber threats

GRC framework

A GRC framework is a model for managing governance and compliance risk in acompany

It involves identifying the key policies that can drive the company toward its goals. By adopting a GRC framework, you can take a proactive approach to mitigating risks, making well-informed decisions, and ensuring business continuity.
Companies implement GRC by adopting GRC frameworks that contain key policies that align with the organization's strategic objectives.
Key stakeholders base their work on a shared understanding from the GRC framework as they devise policies, structure workflows, and govern the company.
Companies might use software and tools to coordinate and monitor the success of the GRC framework.

GRC maturity

GRC maturity is the level of integration of governance, risk assessment, andcompliance within an organization.
You achieve a high level of GRC maturity when a well-planned GRC strategy resultsin cost efficiency, productivity, and effectiveness in risk mitigation.
Meanwhile, a low level of GRC maturity is unproductive and keeps business unitsworking in silos.

GRC CAPABILITY MODEL:

- The GRC Capability Model contains guidelines that help companies implement GRC and achieve principled performance.
- It ensures a common understanding of communication, policies, and training.
- You can take a cohesive and structured approach to incorporate GRC operations across your organization.

Learn

You learn about the context, values, and culture of your company so you can define strategies and actions that reliably achieve objectives.

Align

Ensure that your strategy, actions, and objectives are in alignment. You do so byconsidering opportunities, threats, values, and requirements when making decisions.

Perform

GRC encourages you to take actions that bring results, avoid those that hinder goals,and monitor your operations to detect sudden changes.

Review

You revisit your strategy and actions to ensure they align with the business goals. Forexample, regulatory changes could require a change of approach.

GRC TOOLS:

GRC tools are software applications that businesses can use to manage policies,assess risk, control user access, and streamline compliance.
There are some of the following GRC tools to integrate business processes, reducecosts, and improve efficiency.

GRC software

GRC software helps automate GRC frameworks by using computer systems. Businessesuse GRC software to perform these tasks:

- ✓ Oversee policies, manage risk, and ensure compliance
- ✓ Stay updated about various regulatory changes that affect the business
- ✓ Empower multiple business units to work together on a single platform
- ✓ Simplify and increase the accuracy of internal auditing

User management

You can give various stakeholders the right to access company resources with usermanagement software.

This software supports granular authorization, so you can precisely control who hasaccess to what information.

User management ensures that everyone can securely access the resources they needto get their work done.

Security information and event management

You can use security information and event management (SIEM) software to detect potential cybersecurity threats. IT teams use SIEM software like AWS CloudTrail to close security gaps and comply with privacy regulations.

Auditing

You can use auditing tools like AWS Audit Manager to evaluate the results ofintegrated GRC activities in your company.

By running internal audits, you can compare actual performance with GRC goals.

You can then decide if the GRC framework is effective and make necessary improvements.

CHALLENGES OF GRC IMPLEMENTATION

Businesses might face challenges when they integrate GRC components into organizational activities.

Change management

GRC reports provide insights that guide businesses to make accurate decisions, which helps in a fast-changing business environment. However, companies need to invest in a change management program to act quickly based on GRC insights.

Data management

Companies have long been operating by keeping departmental functions separated. Each department generates and stores its own data. GRC works by combining all the data within an organization. This results in duplicate data and introduces challenges in managing information.

Lack of a total GRC framework

A complete GRC framework integrates business activities with GRC components. It serves the changing business environment, particularly when you are dealing with new regulations. Without a seamless integration, your GRC implementation is likely to be fragmented and ineffective.

Ethical culture development

It takes great effort to get every employee to share an ethically compliant culture. Senior executives must set the tone of transformation and ensure that information is passed through all layers of the organization.

Clarity in communication

The success of GRC implementation depends on seamless communication. Information sharing must be transparent between GRC compliance teams, stakeholders, and employees. This makes activities like creating policies, planning, and decision-making easier.

How do organizations implement an effective GRC strategy?

You must bring different parts of your business into a unified framework to implement GRC. Building an effective GRC requires continuous evaluation and improvement. The following tips make GRC implementation easier.

Define clear goals

Start by determining what goals you want to accomplish with the GRC model. For example, you might want to address the risk of noncompliance to data privacy laws.

Assess existing procedures

Evaluate current processes and technologies in your company that you use to handle governance, risk, and compliance. You can then plan and choose the right GRC frameworks and tools.

Start from the top

Senior executives play a leading role in the GRC program. They must understand the benefits of implementing GRC for policies and how it helps them make decisions and build a risk-aware culture.

Use GRC solutions

You can use GRC solutions to manage and monitor an enterprise GRC program. These GRC solutions give you a holistic view of the underlying processes, resources, and records. Use the tools to monitor and meet regulatory compliance requirements.

For example, Netflix uses AWS Config to make sure its AWS resources meet security requirements. Symetra uses AWS Control Tower to quickly provision new accounts that fully adhere to their corporate policy.

Test the GRC framework

Test the GRC framework on one business unit or process, and then evaluate whetherthe chosen framework aligns with your goals. By conducting small-scale testing, you can make helpful changes to the GRC system before you implement it in the entire organization.

Set clear roles and responsibilities

GRC is a collective team effort. Although senior executives are responsible for setting key policies, legal, finance, and IT personnel are equally accountable for GRC success. Defining the roles and responsibilities of each employee promotes accountability. It allows employees to report and address GRC issues promptly.

Some examples of GRC products are the following:

- ✓ Diligent High Bond.
- ✓ IBM OpenPages.

- ✓ Logic Manager.
- ✓ Logic Gate Risk Cloud.
- ✓ MetricStream Enterprise GRC.
- ✓ Navex Global Lock path.
- ✓ ServiceNow Governance, Risk, and Compliance.

STORAGE INFRASTRUCTURE MANAGEMENT FUNCTIONS : Storage Management Activities

- All the management tasks in a storage infrastructure can be broadly categorized into availability management, capacity management, performance management, security management, and reporting.

Availability management

- The critical task in availability management is establishing a proper guideline for all configurations to ensure availability based on service levels.
- For example, when a server is deployed to support a critical business function, the highest availability standard is usually required.
- This is generally accomplished by deploying two or more HBAs, multipathing software with path failover capability, and server clustering. The server must be connected to the storage array using at least two independent fabrics and switches that have built-in redundancy. Storage devices with RAID protection are made available to the server using at least two front-end ports. In addition, these storage arrays should have built-in redundancy for various components, support backup, and local and remote replication. Virtualization technologies have significantly improved the availability management task. With virtualization in place resources can be dynamically added or removed to maintain the availability.

Capacity management

- The goal of capacity management is to ensure adequate availability of resources for all services based on their service level requirements.
- Capacity management provides capacity analysis, comparing allocated storage to forecasted storage on a regular basis.
- It also provides trend analysis of actual utilization of allocated storage and rate of consumption, which must be rationalized against storage acquisition and deployment timetables.
- Storage provisioning is an example of capacity management.
- It involves activities such as device configuration and LUN masking on the storage array and zoning configuration on the SAN and HBA components. Capacity management also takes into account the future needs of resources, and setting up monitors and analytics to gather such information.

Performance management

- Performance management ensures the optimal operational efficiency of all components.
- Performance analysis is an important activity that helps to identify the performance of storage infrastructure components.
- This analysis provides the information — whether a component is meeting expected performance levels. Several performance management activities are initiated for the deployment of an application or server in the existing storage infrastructure.
- Every component must be validated for adequate performance capabilities as defined by the service levels. For example, to optimize expected performance levels, activities on the server such as the volume configuration, designing the database, application layout configuration of multiple HBAs, and intelligent multipathing software must be fine-tuned. The performance management tasks on a SAN include designing sufficient ISLs in a multi-switch fabric with adequate bandwidth to support the required performance levels. The storage array configuration tasks include selecting the appropriate RAID type and LUN layout, front-end and back-end ports, and LUN accessibility (LUN masking) while considering the end-to-end performance.

Security Management

- Security management prevents unauthorized access and configuration of storage infrastructure components.
- For example, while deploying an application or a server, the security management tasks include managing user accounts and access policies, that authorizes users to perform role-based activities.
- The security management tasks in the SAN environment include configuration of zoning to restrict an HBA's unauthorized access to the specific storage array ports. LUN masking prevents data corruption on the storage array by restricting host access to a defined set of logical devices.

Reporting

- It is difficult for businesses to keep track of the resources they have in their data centers, for example, the number of storage arrays, the array vendors, how the storage arrays are being used, and by which applications.
- Reporting on a storage infrastructure involves keeping track and gathering information from various components/processes.
- This information is compiled to generate reports for trend analysis, capacity planning, chargeback, performance, and to illustrate the basic configuration of storage infrastructure components.
- Capacity planning reports also contain current and historic information about utilization of storage, file system, database tablespace, and ports.
- Configuration or asset management reports include details about device allocation, local or remote replicas, and fabric configuration; and list all equipment, with details such as their value, purchase date, lease status, and maintenance records.
- Chargeback reports contain information about the allocation or utilization of storage infrastructure components by various departments or user groups. Performance reports provide details about the performance of various storage infrastructure components.

STORAGE INFRASTRUCTURE MANAGEMENT PROCESSES :

- **Storage Management** is defined as it refers to the management of the data storage equipment's that are used to store the user/computer generated data.
- Hence it is a tool or set of processes used by an administrator to keep your data and storage equipment's safe.
- Storage management is a process for users to optimize the use of storage devices and to protect the integrity of data for any media on which it resides and the category of storage management generally contain the different type of subcategories covering aspects suchas security, virtualization and more, as well as different types of provisioning or automation, which is generally made up the entire storage management software market.

Storage management key attributes: Storage management has some key attribute which is generally used to manage the storage capacity of the system. These are given below:

1. Performance
2. Reliability
3. Recoverability
4. Capacity

Feature of Storage management: There is some feature of storage management which is provided for storage capacity. These are given below:

✓ Storage management is a process that is used to optimize the use of storage devices.

✓ Storage management must be allocated and managed as a resource in order to truly benefit a corporation.

✓ Storage management is generally a basic system component of information systems.

✓ It is used to improve the performance of their data storage resources.

Advantage of storage management:

There are some advantage of storage managementwhich are given below:

✓ It becomes very simple to manage a storage capacity.

✓ It generally reduces the time consumption.

✓ It improves the performance of system.

✓ In virtualization and automation technologies, it can help an organization improve its agility.

Limitations of storage management:

✓ Limited physical storage capacity: Operating systems can only manage the physical storage space that is available, and as such, there is a limit to how much data can be stored.

✓ Performance degradation with increased storage utilization: As more data is stored, the system's performance can decrease due to increased disk access time, fragmentation, and other factors.

✓ Complexity of storage management: Storage management can be complex, especially as the size of the storage environment grows.

✓ Cost: Storing large amounts of data can be expensive, and the cost of additional storage capacity can add up quickly.

✓ Security issues: Storing sensitive data can also present security risks, and the operating system must have robust security features in place to prevent unauthorized access to this data.

✓ Backup and Recovery: Backup and recovery of data can also be challenging, especially if the data is stored on multiple systems or devices.

Storage management consists of several different processes. Some storage management plans only use a few processes, while others might use them all. Below are the most common processes found in storage management:

PROVISIONING

- This method entails assigning storage capacity by analyzing current capabilities, such as storage on physical drives or the cloud, and deciding the proper information to store in each location.
- It's important to consider factors such as ease of access and security when determining where to store your data.
- Planning where to store data allows organizations to discover whether they have ample storage space available or whether they should reconfigure their system for better efficiency.

DATA COMPRESSION

- This is the act of reducing the size of data sets without compromising them. Compressing data allows users to save storage space, improve file transfer speeds and decrease the amount of money they spend on storage hardware and network bandwidth.
- Data compression works by either removing unnecessary bits of information or redundancies within data.
- For example, to compress an audio file, a data compression tool may remove parts of the file that contain no audible noise.

DATA MIGRATION

- This method entails moving data from one location to another. This can include the physical location, such as from one hard drive to another, or the application that uses the data.
- Data migration is often necessary when introducing new hardware or software components into an organization.
- For example, if a business purchases new computers for its office, it's important to transfer all data from the old systems to the new ones.
- Important factors to consider while implementing data migration include ensuring network bandwidth, effective transfer speeds, data integrity and ample storage space for the new location throughout the transfer.

DATA REPLICATION

- This process includes making one or more copies of a particular data set, as there are several reasons why a company may want to replicate its data.
- For example, you may wish to create a backup if there's a problem with an original data set. You may also want to replicate data so you can store it across different locations, improving the overall accessibility across your network.
- There are two types of data replication***: synchronous and asynchronous***. **Synchronous data** replication is when companies copy any changes to an originaldata set in the replicated data set. This type of replication ensures updated informationbut may also use require more resources than asynchronous replication. **Asynchronous replication** only occurs when a professional enters a command intothe database, so it's not an automatic process. With this type, your company has morecontrol over the resources used to replicate data but may not possess real-time databackups.

AUTOMATION

- Automation is the process of having tools automatically manage your data. Rather than updating your data manually, you can use software tools to accomplish this task for you.
- For example, you could use a tool to automatically update a shared database whenever you make a change on your local computer, rather than requiring manual updates. This would ensure that the database contains updated information for all users and prevents users from viewing outdated information if a user forgets to submit changes.

DISASTER RECOVERY

- Disaster recovery is a plan companies create for potential scenarios regarding data issues.
- For example, if the hard drive that stores your data breaks, it's important to have an effective plan that allows your business to return to normal operations. This planmight include switching to a backup hard drive, making a new copy of that backup and purchasing a new primary hard drive.
- Important elements in a disaster recovery plan include speed, data integrity and costs. Effective organizations often have plans that decrease technological downtime as much as possible

- In addition, it's important to prevent loss of essential data.
- Finally, organizations typically aim to reduce costs wherever possible, such as compressing data to save money on storage requirements.