

System Infrastructure | Repository Backup Function for Repository Services

Introduction to Repository Backup Function:

Data protection is a complete workflow: backup, offsite storage, archiving, recovery.

From Wikipedia: “**Backup** in [computer engineering](#) refers to the copying of [data](#) for the purpose of having an additional copy of an original source. If the original data is damaged or [lost](#), the data may be copied back from that source, a process which is known as *Data recovery* or *Restore*. The "data" may be either data as such, or stored program code, both of which are treated the same by the backup software. Backups differ from an [archive](#) in which the data is necessarily duplicated, instead of simply moved.”

“Computer backups are useful primarily for two purposes, the first and most obvious is to restore a computer to an operational state following a disaster also called [disaster recovery](#). This includes loss of a hard disc or the file system becoming so badly corrupted it cannot be read. The second use, often overlooked but probably more common, is to facilitate the recovery of a single file or set of files when they are accidentally deleted or corrupted by the user or a program”.

“A backup is only as useful as its associated recovery strategy. Having a complete set of backup tapes is of no use if the only copy of the software required to read them is on one of the tapes. It is also possible for backup software to run successfully for several months, only to fail when it is needed most due to read errors on the backup media. Magnetic tapes in particular should be read-tested on a regular basis.”

Backup Systems

Backups are created on disk or tape (or a combination of both).

Disk vs Tape Cost Factors

The nature of computers is that there are always options and usually three points to consider when comparing performance and cost before buying equipment: [size](#) (measured in [bytes](#)), [rotational speed](#) (measured in [hertz](#) or other [cycles](#) such as [RPM](#)), [warranty](#) (measured in [years](#) or eventually [months](#)). Also consider the cost per gigabyte when comparing different solutions.

When measuring cost factors for data stored on disk also consider:

- Power consumption costs linked with keeping the spindles turning and keeping systems cool
- Wear-and-tear on all components of the array
- Data-center floor space
- General storage management costs linked to spinning disks

In contrast, archived tape offers:

- Greater portability
- Little power consumption
- Cheaper storage space
- Less management effort over the course of an extended archival life cycle

Disk Based Backup Advantages over Tape

- Improved reliability
- Higher performance
- Increased staff productivity
- Reduced data loss through **Continuous Data Protection (CDP)**

Disk Architectural Implementations:

- Backup to disk, which is sometimes referred to as D2D (*disk-to-disk*) or D2D2T (*disk-to-disk-to-tape*).
- Tape emulation, where software is used to present a virtual interface to disk, making it appear to backup applications as a tape device.
- Replication-based recovery, where data replication technologies are used to move data to a centralized location where backup can occur in a more controlled environment.
- Point-in-time copy, where an image of the data, such as a full mirror or pointer based snapshot of the data is stored on a RAID subsystem.
- A fifth category that is beginning to emerge is commonly referred to as continuous backup.

How is the Backup function implemented?

There are primarily three different types of backup - full, incremental, and differential. A [full backup](#) is simply the backing up of all the files on the system. An [incremental backup](#) will only backup files modified since the last backup. An incremental backup is also referred to as a cumulative incremental backup. A [differential backup](#) is a cumulative backup of changes made since the last full backup.

Metrics for Backups: There are six primary operational [metrics](#) relating to data backup i.e. pain points for recovery:

- **Recovery Point Objective (RPO)** is the point in time that the restarted infrastructure will reflect. Essentially, this is the roll-back that will be experienced as a result of the recovery. Reducing RPO requires increasing [synchronicity](#) or frequency of copying the data to be protected.
- **Backup Window** is the amount of time that is taken to copy a given data set to the backup device. Most traditional backup systems require a data set to be frozen for hours while the entire content of a [filesystem](#) is copied to [magnetic tape](#). Newer techniques use **incremental backup forever** as well as [mirror](#), [snapshot](#), effectively reducing the required backup window.

- **Restore Time** is the amount of time required to bring a desired data set back from the backup media.
- **Retention Time** is the amount of time in which a given set of data will remain available for restore. Some backup products rely on daily copies of data and measure retention in terms of days. Others retain a number of copies of data changes regardless of the amount of time.
- **Backup Validation**, also known as "Backup Success Validation", is the process by which Owners of data can get information regarding how their data was backed up. This same process is also used to prove compliance to regulatory bodies outside of the organization, for example, a biotech company might be required to show "proof" to the [FDA](#) that their test result data are backed up properly. Terrorism, data complexity, data value and increasing dependence upon ever-growing volumes of data all contribute to the anxiety around and dependence upon successful backups. For that reason, many organizations rely on third-party or "independent" solutions to test, validate, optimize and charge for their backup operations ([backup reporting](#)). See also [Backup validation](#). Some "[modern backup to disk software](#)" have built-in validation capabilities.
- **Open File Backup** is the ability to backup a file while it is in use by another application.

A backup should be planned carefully, and the following points should be considered:

- Cyclical backups improve [data recovery](#) reliability.
- Automated backup should be considered, as manual backups can be affected by human error.
- Making two copies of backup can potentially increase [security](#) for data recovery, to avoid accidents such as [fire](#) and [physics randomness](#).
- Using [standards](#) make them easier to recover, since that's the goal of a standard. Established standards are usually safer for recovery.
- New standards are generally faster and more powerful.
- [Data compression](#) might be important if there is more data than media space.
- Uncompressed data are mostly easier to recover if the backup media are damaged or corrupted, unless individual objects (files, folders, etc.) are compressed separately. Many backup programs allow an administrator to increase or decrease the scope of [solid compression](#) to [tradeoff](#) compression performance against durability.
- Backups might take a lot of time to accomplish, which can potentially become a problem in a work environment.
- Multiple media backup, for just one content, can be done with independent indexing to optimize individual data recovery.
- Backup depends both on [software](#) and [hardware](#) and so are exposed to [expiration](#) due to time issues.

In the case of a big accident or disaster, while it's important to keep calm it's predictable that most likely no one would be able to think clearly, and act accordingly. Prepare [checklists](#) and train staff in advance.

Appendix A

Current Specifications for Tape Backup:

Tape Formats Specifications:

Encoding :

 Helical Scan

 Linear serpentine

Recording Format

Track Density

Data Compression

Tape Specification

Archive Storage

Durability

Tape Media Type and Specifications

DLT

LTO

DAT/DDS

TRAVAN

Tape Drives for DLT Formats (Specifications)

DLT VS80

DLT VS160

DLT-V4

SDLT 320

SDLT 600

Tape Drives for LTO Formats (Specifications)

LTO-2

LTO-HH

LTO-3

Tape Drives for DAT/DDS Formats (Specifications)

DDS-4

DAT-72

Tape Drives for TRAVAN (Specifications)

TRAVAN 20

TRAVAN 40