8.12.9 Backup Strategy Facts

File backups can be performed using a number of Linux commands. However, an overall backup strategy will make the backup process more efficient, and will ensure all files are backed up in a timely manner.

This lesson covers the following topics:

- Backup types
- Off-site and off-system storage plans
- Integrity checks

Backup Types

There are different types of backup that you should be familiar with. Each type has their own pros and cons, so consider which scenario works best for you.

Backup Type	Description
Full backup	A full backup is a copy of all the files that exists at the point in time when the backup is taking place. If a backup took place last week, when a full backup runs again this week, every single file will be backed up again, even if nothing has been updated. This option can be expensive, given that each time a new backup runs, you will need, at the very least, double the amount of backup storage as you did before.
Incremental backup	An incremental backup stores files that has been changed or added, since the last backup has been made. This type of backup can save a considerable amount of storage capacity, since only the updated files are backed up. In this scenario, every time a backup runs, only the files updated since the last backup are updated. Newly added files will also be backed up. This is a good option when running backups regularly.
Differential backup	A differential backup copies all of the files that have been updated, but only against the last full backup. An incremental backup would backup files changed since the last full or incremental backup. A differential backup will update files that have changed since the last full backup.
Snapshot clones backup	Snapshot clones is a technology that allow you take point-in-time snapshots of the files on a system without causing the downtime inherent in traditional backups. A snapshot is not an independent backup of a set of data files. Rather, a snapshot is set pointers to blocks of data that make up a set of files at a point-in-time. When a change is made, the original blocks are kept and changed blocks are added. When a new snapshot is taken, another set of pointers is created that point to the blocks that make up the file at that point-in-time. The term clone is used when restoring. The blocks for a selected snapshot are written to a new storage location as a traditional set of files. This set of files is considered a clone of the original snapshot. The benefits of snapshots and clones is the rapid accessibility of point-in-time data which also allows for a quick roll back when data is corrupted. The disadvantage of snapshot technology is that it requires more storage space and it may impact production performance during the clone process since primary data is being accessed.
Image backup	A disk image is a single computer file or set of files that contain the contents of a hard disk. It is usually created by coping the disk sector-by-sector, instead of file-by-file. A disk image is often called a system image, especially when it is an image that contains a computer operating system. Image backups are an important part of a backup strategy, especially after software or system updates are installed.

Off-Site and Off-System Storage Plans

You should consider having an off-site, or off-system, storage plan. This is especially important in cases where a natural disaster destroys the hardware and data stored locally. There are several tools that can assist you when backup up using off-site or off-system storage.

Backup Tool	Description
rsync	The rsync tool can copy local files from/to a remote host using a remote shell - SSH - or a remote rsync daemon. It is a file copying tool with the capability of reducing the amount of data transferred, making replicas, copies and backups. To make it even more difficult for others to access your data, there's the option to harden SSH, improving security.
SFTP	SFTP (SSH File Transfer Protocol or Secure File Transfer Protocol) is packaged with SSH, and works over a secure connection. It has the ability to leverage a secure connection to transfer files between the local and remote systems. SFTP is usually preferred over FTP because of its security features and ability to piggy-back on an SSH connection.

SCP	The SCP tool securely copies files and directories between remote hosts without starting an FTP session or even logging into the remote systems. SCP uses SSH to transfer data, so it requires authentication, but it does encrypt both the file and any passwords exchanged

Integrity Checks

You can check the integrity of a backup to ensure that the data has been backed up or transferred without error. One way to do this is to use a hash algorithm that produces a "fingerprint" of the downloaded file. A hash algorithm inputs the backup data and outputs a unique character string. If the source data and the backed up data produce the same fingerprint, you can be confident that the data is identical. There are two hash algorithms that are commonly used to check integrity.

Algorithm	Description
MD5	The MD5 (Message Digest Algorithm 5) algorithm produces a 128-bit hash value. It was originally used as a cryptographic hash function but was found to suffer from extensive vulnerabilities.
SHA	A cryptographic hash function which takes an input and produces a 160-bit (20-byte) hash value known as a message digest – typically rendered as a hexadecimal number, 40 digits long. The SHA (Secure Hash Algorithm) algorithm is slowly replacing the MD5 algorithm.