IT Infrastructure Administration

Day 9: Backup Concepts and Architecture

Ephrem Teshale(PhD)
Tadios Abebe

July 31, 2025

Backup Concepts and Architecture

Day 9
Training Outline

What is backup
Backup Strategy
RPO and RTO
Types of Backup
    Full Backup
    Incremental Backup
    Differential Backup
Backup devices and services
Proxmox Backup Server
Why PBS
Key concepts and terminology
Installation and Configuration
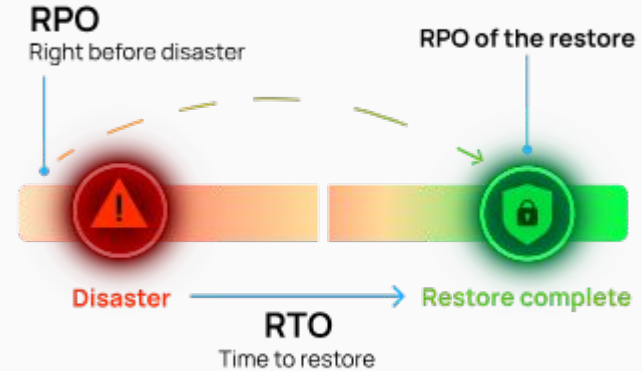Knowledge Check
Exercise

What is Backup

❏ Backup and restore refers to technologies and practices for **making periodic copies of data** and applications to a **separate, secondary device** and then using those copies to recover the data and applications—and the business operations on which they depend.

❏ Backup and restore is used if the original data and applications are lost or damaged due to a **power outage, cyberattack, human error, disaster** or some other unplanned event.

Backup Strategy

- How you backup your data and applications (in other words, how frequently you back them up and to what device or location) depends on the **cost of losing access** to the data and applications for any period of time and **the cost of replacing or re-creating** the data if it's lost for good.
- Typically, the **first step** in creating a backup strategy—especially an **enterprise backup strategy**—is to determine **recovery time objectives** and **recovery point objectives** for each data source and application.
- Recovery Point Objective (**RPO**) and Recovery Time Objective (**RTO**) are essential concepts for business-critical applications.
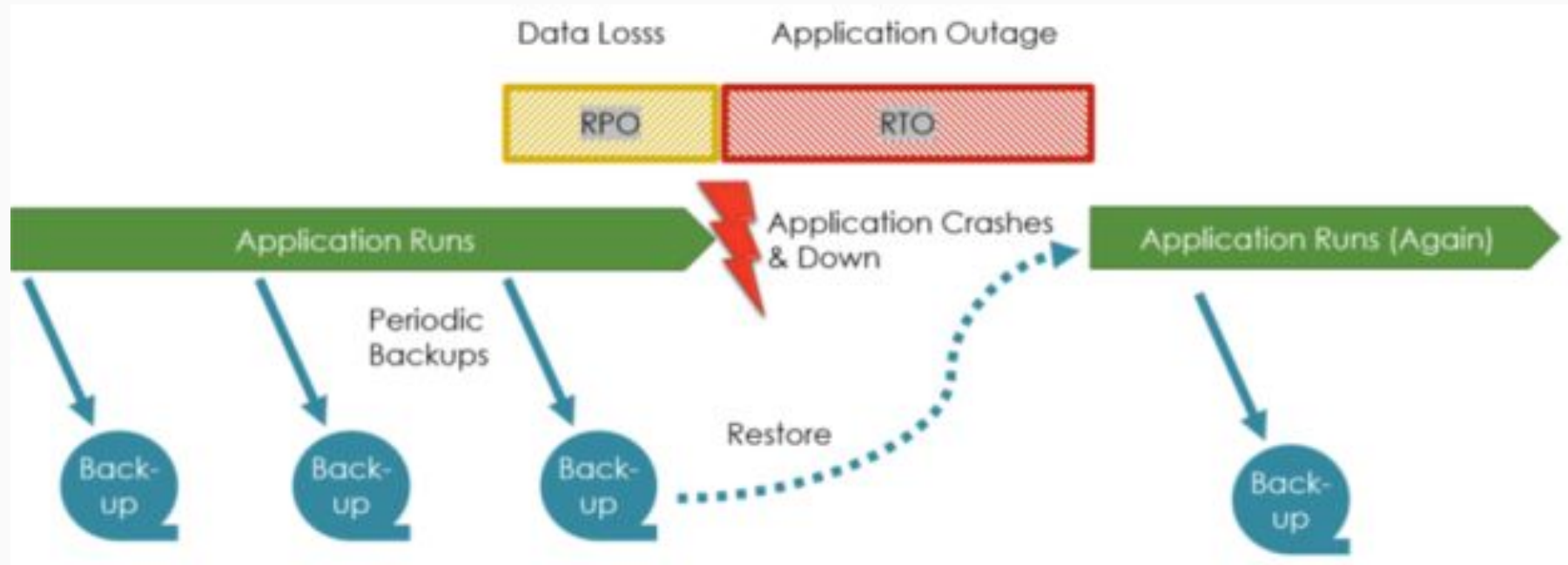

RPO
Right before disaster

RPO of the restore

Disaster

Restore complete

RTO
Time to restore

RPO

- ❏ Recovery point objective, or RPO, refers to the a**mount of data you can afford to lose** and effectively dictates how **frequently you need to back up** your data to avoid losing more.
- ❏ How much data might get lost max in such an event, this **defines the RPO**, the Recovery Point Objective. It depends on the **backup frequency**.
- ❏ One **factor** influencing the RTO is the time needed to restore the data from a backup. Network bandwidth and storage are significant factors, but so is the backup type. Full, incremental, and differential backups take different long.
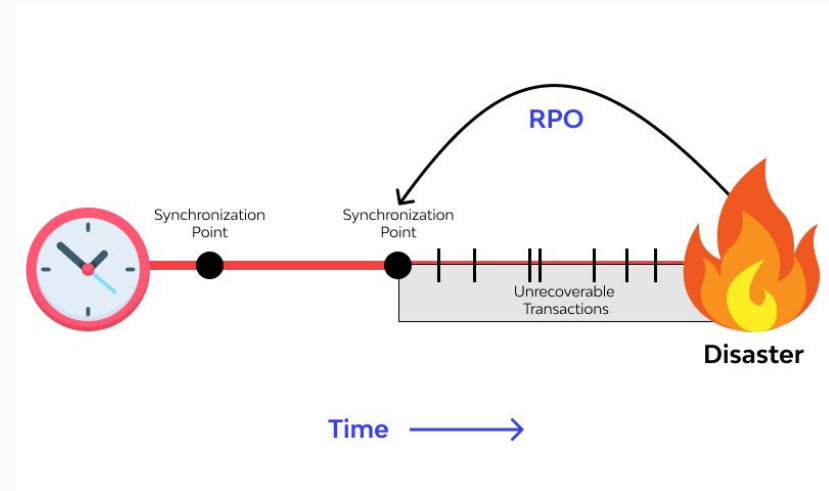
❏ Recovery time objective, or RTO, refers to the **maximum amount of time** the business can afford to be without access to the data or application—or, **how quickly** you need to recover the data and application.

❏ That is the time engineers (or automatic processes) have to start the application on other servers – potentially at a different location – and restore all necessary data from backups.

Type of Backups

Full Backup

❏ The starting point is always a full backup. The system writes all data to a secure backup storage. So, why would an architect not always go for a full backup? It is simple: Suppose an application is highly critical. Thus, the RPO is 5 min. Let us further assume that it is a massive database, but not much data changes within five minutes. Then, a full backup every five minutes seems too much. It might even be technically impossible, depending on network bandwidth and the amount of data. Incremental backups can be a better solution for such a scenario.
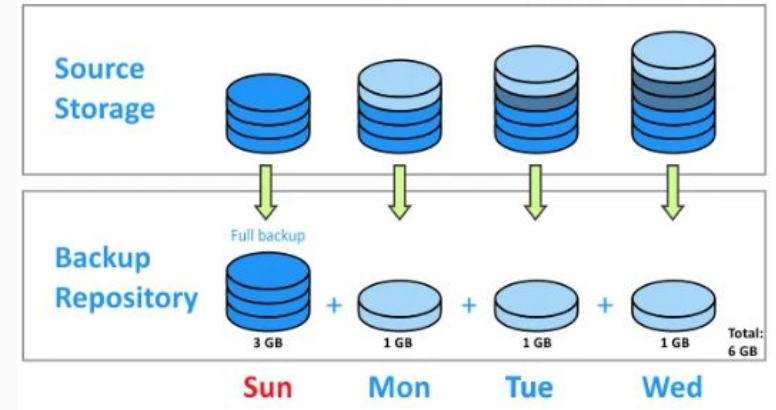

8

Incremental Backup

- ❏ An incremental backup copies only the most recent changes away. If the last full backup was at 4 am, an incremental backup at 4:05 am copies the changes from the last five minutes. The next backup at 4:10 am copies, as well, only the changes of the last five minutes, i.e., the changes between 4:05 am and 4:10 am.

- ❏ Incremental backups have on disadvantage compared to full backups: the restoration time might be longer. To restore the situation of 4:10 am, one has to fully restore the data from 4 am, then apply what changed till 4.05 am, and then the incremental changes till 4:10 am.

- ❏ Incremental backup starts with a full-image backup and then performs periodic backups of only the data that changed since the most-recent backup. Typically, after a set number of incremental backups, another full-image backup is performed and the cycle starts again.

Types of Backup

Differential Backup

- ❏ A middle way between incremental and full backups is the differential backup. A differential backup copies away all changes from the last full backup. At 4:05 am, a differential backup writes all changes of the last five minutes; at 4:10 am, all changes between 4:00 and 4:10, and at 4:15, all changes between 4:00 and 4:15. Thus, a restore of an incremental backup means: restore the full backup and (only) apply the changes of the last differential backup. However, these concepts get transparent with the cloud: the cloud providers implement standard configurations.

- ❏ Differential backup backs up all data that has changed since the last full-image backup. To restore data, you first apply the most-recent full-image backup and then the most-recent differential backup. Backup time increases with each successive differential backup, but restoring requires applying just two backup files—the latest full image backup and the differential backup).

Full vs Differential vs Incremental backup features

Feature	Full	Differential	Incremental
Storage Space	Large	Medium	Small
Backup time	Fast	Medium	Slow
Restore time	Fast	Medium	Slow
Data loss probability	High	Medium	Small
Backup frequency	Low	Medium	High
Granular recovery	No	Low	High

Backup devices and services

What type of device or service will you use to backup your data? Generally speaking, you have four choices.

Tape drive

- Tape is the oldest backup medium in use today. It offers low-cost, high-capacity data storage, but relatively slow read/write performance makes tape a poor choice for incremental backup, continuous data protection (CDP) or any other backup method that updates backups whenever data changes

- Tape is also more prone to physical wear and damage than other storage media so it needs to be closely managed and constantly tested to ensure that it works when it's time for recovery. For these reasons, tape is a better choice for nightly or weekly backups or for cost-effectively archiving data that your organization wants or needs to keep but doesn't need in order to quickly bring the business back online in the event of an outage or disaster.

Backup devices and services

**Hard disk drives (HDDs) or solid-state drives (SSDs)**

- ❏ Most data today is backed up to a hard disk drive (HDD) or solid-state drive (SDD), whether that drive is a stand-alone external drive or part of a backup server (see below). Both offer faster read/write performance than tape, making them a good choice for continually updated backups and short-RTO/RPO backup scenarios.

- ❏ SDDs are increasingly popular because they offer faster read/write times than HDDs, require less physical space to store the same amount of data, and consume less power (even if they are more expensive to purchase per gigabyte). If HDDs and SDDs have a drawback, it's that they aren't particularly scalable—if you need more backup capacity, you must purchase and install a new physical disk.

Backup devices and services

Backup server

- ❏ A backup server is a dedicated server built specifically for backing up files stored on multiple client computers on the same network. The server is outfitted with significant disk storage and specialized software for scheduling and managing backups.

- ❏ Backup server disks are often configured for redundancy to protect backup data and ensure that backups continue in the event of a disk failure. An onsite backup server can be a cost-effective backup solution for a small office but does not protect backup data against local outages or physical disasters.

Cloud backup

❏ Cloud backup backs up your data and applications via a corporate network or internet connection to a physical or (more likely) virtual backup server at a remote data center operated by your company, a hosting provider or a cloud services provider.

❏ Cloud backup is typically the most flexible type of backup. You can use it to back up files, application data, or entire physical or virtual servers. You can schedule backups as frequently or infrequently as you like. Because cloud backup servers are typically virtualized, you can scale easily and cost-effectively as needed.

❏ Cloud backup eliminates the need to physically move backup media to another location (and the significantly larger RTOs and RPOs that can result) for protection against local power outages or disasters.
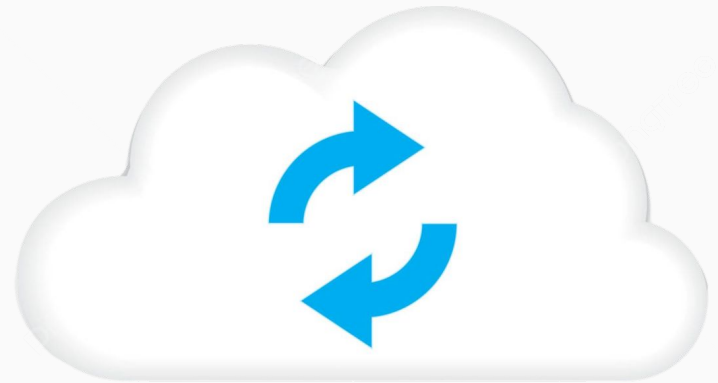
Proxmox Backup Server

An open-source, enterprise-grade backup solution from Proxmox.

Core Purpose: Efficiently store, manage, and restore backups of:

- Proxmox VE Virtual Machines (VMs)
- Proxmox VE Containers (LXC)
- Linux hosts (via proxmox-backup-client)

Key Design Principles:

- **Efficiency**: Deduplication, Compression, Incremental Backups.
- **Integrity**: Data verification, checksumming.
- **Security**: Encryption (client-side and at rest).
- **Integration**: Deeply integrated with Proxmox VE.

- **Cost-Effective:** Open-source, no licensing fees for core functionality.
- **Seamless Integration:** Works perfectly with our existing PVE infrastructure.
- **Optimized:** Built specifically for VM/CT backups.
- **Robustness:** Designed for large-scale, reliable backups.
- **Flexibility:** Adaptable to various storage backends.

PROXMOX

Key concepts and terminologies

- **Datastore**: The primary repository on PBS where backup data is stored. It's content-aware and handles deduplication.
- **Client**: The system sending backups to PBS (e.g., a PVE node, a Linux server).
- **Backup ID:** A unique identifier for each backup snapshot.
- **Chunk**: Small, variable-sized blocks of data. PBS breaks backups into chunks for deduplication.
- **Index**: Metadata that maps chunks back to the original files and VM/CT structure.
- **Snapshot**: A consistent, point-in-time copy of the VM/CT at the moment of backup.

- **Prune Job**: A scheduled task to automatically remove old backup snapshots based on defined retention policies.
- **Garbage Collection (GC):** The process that reclaims physical disk space after chunks are no longer referenced by any remaining backups.
- **Remotes**: A configuration on PBS to define another PBS instance for data replication.
- **Sync Jobs:** Tasks that replicate (synchronize) backup data between a local PBS datastore and a remote PBS datastore.

Installation and Configuration

PVE and PBS Setup Checklist

- remove enterprise repository
- add no sub repo
- add proxmox backup repo
- update and upgrade
- add pbs repo `deb http://download.proxmox.com/debian/pbs bookworm pbs-no-subscription`
- install tmux
- install proxmox backup server `apt install proxmox-backup-server`
- create zfs storage on pve
- edit the zfs storage for thin provisioning
- add the created zfs to the pbs
- create a namespace on the pbs for the main and sync site
- add the local pbs server as a backup storage
- remove backup content from local storage entry
- remove the default lvm thin storage from proxmox without clearing the storage
- resize the root lv `lvextend -l +100%FREE pve/root`
- resize the the root filesystem `resize2fs /dev/pve/root`

ETHIOPIAN IT PARK

CITCOT

ZERGAW CLOUD

Which of the following is a widely recommended backup strategy?

A) 1-1-1 Rule

B) 3-2-1 Rule

C) 5-4-3 Rule

D) 2-2-2 Rule

What does RTO refer to in backup planning?

A) Real-time Output

B) Recovery Task Objective

C) Recovery Time Objective

D) Remote Transfer Operation

If your RPO is 4 hours, what is the maximum acceptable data loss in a disaster?

A) 4 minutes

B) 4 hours

C) 4 days

D) No data loss

Which type of backup includes all data, regardless of whether it has changed?

A) Incremental

B) Full

C) Differential

D) Selective

What is the major benefit of incremental backups compared to full backups?

A) They require no storage

B) They are slower to perform

C) They use less storage and backup time

D) They include system settings only

Which backup type stores all data changed since the last full backup?

A) Incremental

B) Full

C) Differential

D) Snapshot

A full backup typically:

A) Requires the least storage

B) Is faster than incremental backup

C) Copies all selected data each time

D) Is only used with RAID systems

Which of the following is a drawback of incremental backups?

A) Takes the longest to perform

B) Requires the most storage

C) Slows down the restore process

D) Always includes unnecessary files

What is the main advantage of differential backups over incremental backups?

A) Faster backups than incremental

B) Smaller size than incremental

C) Quicker restoration as only 2 backups are needed

D) No need for full backups

ETHIOPIAN IT PARK

CITCOT

ZERGAW CLOUD

Which of the following is not a typical backup target?

A) Cloud storage

B) NAS

C) RAM

D) Tape drives

The 3-2-1 backup rule suggests:

A) 3 people managing 2 backups in 1 location

B) 3 types of backup on 2 formats in 1 week

C) 3 copies of data, 2 local, 1 offsite

D) 3 cloud accounts and 2 backups per day

Which storage format is used by PBS?

A) .vmdk

B) .bak

C) Chunk-based with Zstandard compression

D) ISO

Thank You

- +251977035511
- info@citcot.com
- citcot.com
  nunaethiopia.com