ICA0002: IT Infrastructure Services

Backups

Roman Kuchin
Juri Hudolejev
2020

What is a backup?

Backup

1. A copy of data that can be used to restore the original

2. A process of copying the data for restoration purposes

^^^ *these are fine if you want to explain your non-IT friend what a backups is*

Backup as an infrastructure service

Systematic established **process** of

- **Making** a copy of data that can be used to restore the original
- **Storing** that data for required period of time
- **Verifying** the usability of that data for restoration purposes
- **Documenting** the regular data restoration and disaster recovery plans
- **Verifying** the compliance of the documentation and the actual processes

Backup as an infrastructure service

In simple words, good backup

- Is created in time
- Is stored as long as needed
- Is available when needed
- Can be used to restore the needed service state

And every step of this process is documented and verified

Purpose of the backup

Ability to restore the **service** back to some certain state

- More than just restoring the data!

Important part of security policies

- Guarantees system availability and usability

Backup

1. Coverage
2. Frequency
3. Retention
4. Usability
5. Storage
6. Security
7. Documentation

Coverage: what to backup?

Ideally, everything

In real world, as much as your backup system can handle

Better to have multiple copies of the same data than no copies at all

Coverage: what to backup?

Data backups

- Must have in any case
- Usually universal format, can be restored to a different system
- Storage can be optimized (incremental / differential / deduplication)
- Creation and restore can be time consuming

Examples: SQL dump, JSON export, XML export, …

Coverage: what to backup?

File backups

- Why export data if we can backup the original file?
- Storage can still be optimized
- Creation and restore might require less time
- System being restored must support the file format

Example: cannot restore MySQL 5.6 files backup to MySQL 5.7 server

Coverage: what to backup?

Virtual machine and/or filesystem snapshots

- Creation and restore might require even less time
- May require machine or service downtime to create a backup
- Higher storage and bandwidth requirements

Example: your own machine (laptop, VirtualBox) snapshots

Coverage: what to backup?

Virtual machine images

- Why would you need them?
- The image can always be downloaded from the internet, right?

What if

- Public image is not available?
- You use custom modified images?

Frequency: how often to backup?

Examples: every hour, every night, every Sunday

More often:

- More backup storage is required
- More resources (CPU, memory) are needed to create and verify the backups
- Less service availability (downtimes or reduced HA during backup creation)

Less often:

- RPO: recovery point objective
- Time interval between two backups -- acceptable data loss

Retention: how long to store the backup?

Backup storage has its limits

Data value reduces over time

Retention: how long to store the backup?

Backup storage has its limits

Data value reduces over time

Backup retention must me longer than interval between backups!

-- Captain Obvious

How many versions of the same data are stored?

Ideally, as many versions as possible

In real world:

```
version_count =  frequency * retention
```

Usability: is the backup usable?

Ensure that you can restore the service from the backup

- Was the backup created **in time**?
- Is the stored backup **readable**?
- Is the stored backup **enough** to restore the service?
- Could the service be restored **to the needed state**?

Solution: test backup restores regularly

Usability is probably
the main feature of the backup

Storage: where to store backups?

3-2-1 rule:

At least 3 copies of the data -- 2 storage types onsite -- 1 copy offsite

Separate cloud or physical location -- preferably both!

Monitoring physical conditions: temperature, humidity etc.

Storage: where **not** to store backups?

"We have redundant hardware (RAID), so the backups are automated!"

"We are keeping backup on the same server to save costs!"

"We are keeping backup in the same datacenter, so the restore is really fast!"

"We are using cloud platforms, they do the backups themselves!"

What is wrong with these approaches? What are the risks?

Security: how to protect backups?

Good backup contains all the data needed to restore the system.

Backup should be protected as severely as the live system itself.

Store the backup in the safe place:

- Physical copies -- vault
- Online copies -- encrypt the transport, storage -- whatever possible

Documentation: how to describe backups?

Three main documents:

1. Backup SLA (service level agreement)
2. Backup restore plan
3. Technical documentation for backup operators

Related documents:

- DR (disaster recovery) plan

Documentation: how to describe backups?

Backup SLA:

- Coverage -- what is being backed up
- RPO: recovery point objective -- acceptable data loss
- Versioning -- how many versions of the data are stored
- Retention -- how long are the backups stored
- Usability -- how is the backup recovery verified
- Restoration criteria -- when to restore from the backup
- RTO: recovery time objective -- how long will it take to restore the service

RPO vs. RTO

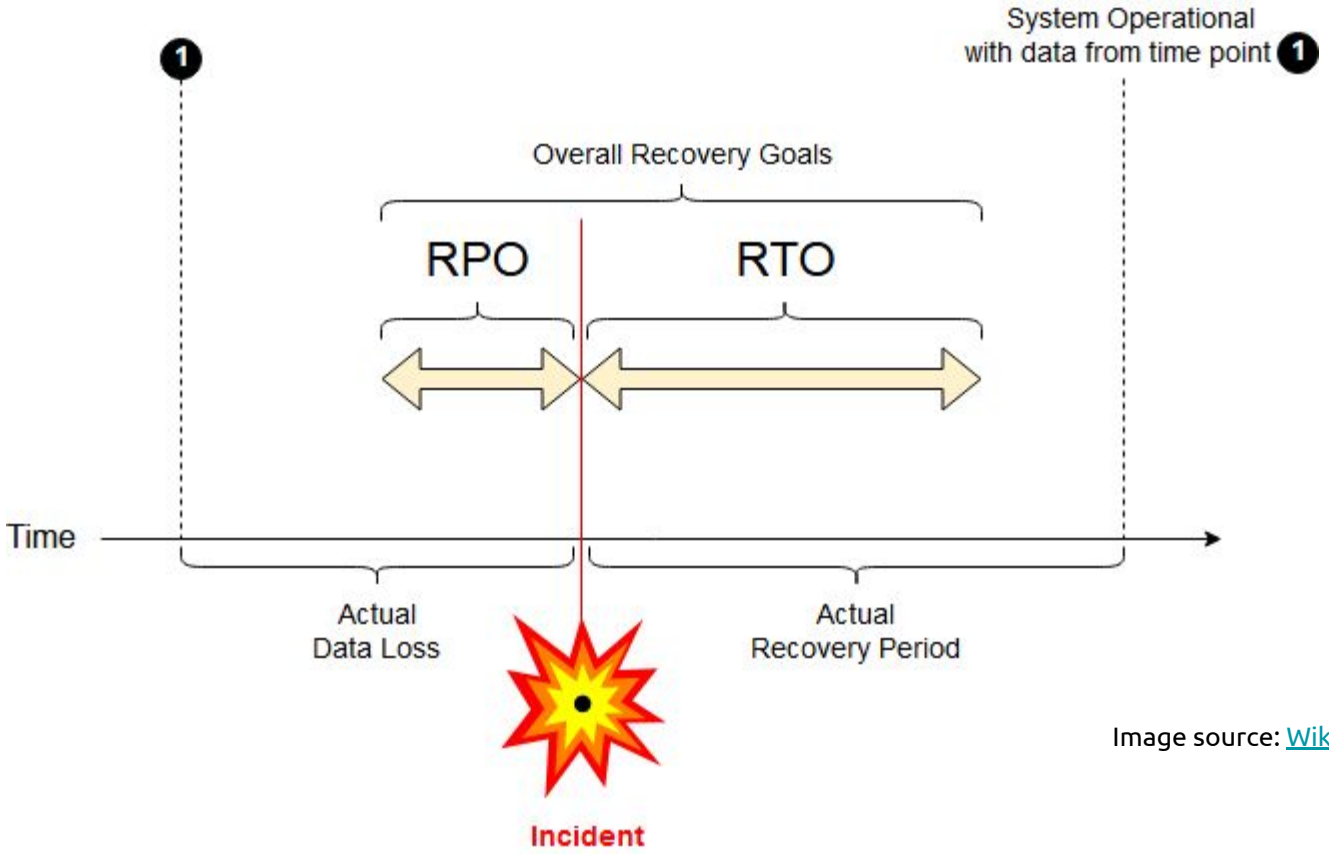

Image source: Wikipedia | CC BY 4.0

Documentation: how to describe backups?

Backup restore plan:

- Documented process of the backup restoration
- Explains in details who does what, how and then
- Ideally any employee with needed permissions could to restore the service

Documentation: how to describe backups?

Technical documentation for operators:

- How to check if the backup was created
- How to verify the backup
- How to install, configure and administer the backup system

Backup is your last resort!

Build your systems so that you would not need backups:

- High availability and redundancy
- Multiple copies of the data

But still do backups properly!

Never underestimate the importance of the backups

Questions?

What is **not** a backup?

Backup vs. archive

Backup:

- Main focus: fast service restore to a certain state
- Retention period: weeks, sometimes months

Archive:

- Main focus: data consistency + low cost of ownership
- Retention period: years, sometimes decades
- Ideally retyped to a new media from time to time

Backup vs. configuration management

Configuration management

- Helps to restore the service faster
- Does **not** replace the backup!

Goal: restore the **service**, not the data

Some components are easier to recreate than to recover

Better to have a backup unused than not have it all

Only wimps use tape backup; real men just upload their important stuff on ftp, and let the rest of the world mirror it ;)

-- Linus Torvalds

Motivational reading

The God of Backups: http://seer-grog.net/GoB.txt