Data Safety and Backups

Christian Lang (clang)

27. April 2020

A lot of private important or emotional relevant data is insufficiently protected against data loss.

My personal experience

- Not only business data is worth to protect:
  - private pictures,
  - password safe,
  - accounting and insurance data.
- The importance of this data is mostly discovered after loss.
- Rescuing data of a defective hard disc costs >1'000 sFr.

Scope

This presentation concentrates on data loss and not privacy.

Concepts

Data loss scenarios

- Hardware
  - Instantly defective hard disc.
  - Slow occurring problems: sporadic read errors.
  - Bit rot: sporadic bit-flips in physical memory segments.
  - Catastrophe: Fire or water damage, etc.
- User / Software
  - User deletes a file that he later needs again.
  - User changed file and later needs an earlier version of it.
  - Some software deletes/changes by itself (not intended by the user):
    - Errorneous software
    - Ransomware

Which one want you protect yourself from?
How?

RAID - Redundant Array of Inexpensive Disks

- Does only protect against “some” hardware problems.
  - A user that deletes his important data is not protected by RAID.
- Hardware RAID implementations require hardware replacements if the RAID controller dies.
- Different RAID configurations
- Software implementations use more flexible definitions.
  - E.g: RAID1 can also work with 3 discs. Of different sizes.

RAID is no backup!

Backups - Data copy elsewhere

- Can protect against all/various data loss scenarios.
- Big range of how to implement:
  - One simple USB stick with a copy of your data that is updated each evening by hand.
  - Multiple servers with multiple layers of backups and different used technologies.
- Location is important:
  - in my PC
  - in my separate NAS
  - not in the same house
  - in the cloud
- Different backup intervals protect against different problems:
  - monthly backup can recover old data that was thought of never needed anymore
  - hourly backup can recover files that I deleted by mistake

Backups are no version control!

How to design your Backup

Questions:

- How much time do I have to detect data loss?
  - Age of oldest kept backup. → Backup rotation scheme
- How long a data state needs to exist to go into backup?
  - Interval when backups are done.
- How much space can I spend for backups?
  - The more data versions the more space is consumed for backups.
- Where should I store my backup?
  - If I want to protect from Ransomware → at least not on the same host.
- Who has access to my backup?
  - Off-site backup in a public place (Cloud/Office) → encrypted.

Atomic Snapshots

- Traditional backup solutions use different kinds of backups:
  - full
  - incremental
  - differential
- Optimize trade-off between number of states and storage space.
- Creates binary images. → not directly accessible.
- This kind of "snapshots" can be simulated directly with hardlinks.
- More modern filesystems do provide atomic snapshot mechanism by itself.
  - E.g. by using internal tree data structure → similar to `git`
- Examples:
  - `btrfs`
  - `zfs`
  - `ntfs`
  - (`lvm`)

Monitoring

The best backup system is worthless if you do not detect failed hardware!

- Discs → `S.M.A.R.T.`
  - statistics: age, power on hours, temperature, failed sectors, etc.
  - tests
- Filesystem
  - fill state
  - read/write errors
- System log

→ Best case: automated email notifications

Migration

The migration to new hardware is a critical step!

- Should be planned
- Have extra copies ouf your data
- Keep old state of your data on old hardware until new setup is mature and runs at least $^1/_2$ year without data loss.

Example project

Requirements

In this example I want to focus on the most powerful but still simple solution.

Simple means:

- not too complex architecture
- not too much hardware
- done by a technology affine person

Possibilities

Hardware

- Specific NAS
  - Synology
  - Qnap
  - Western Digital
  - etc.
- Old PC
- Small single board computer
  - Raspberry Pi
  - Odroid HC1 or N2
  - etc.

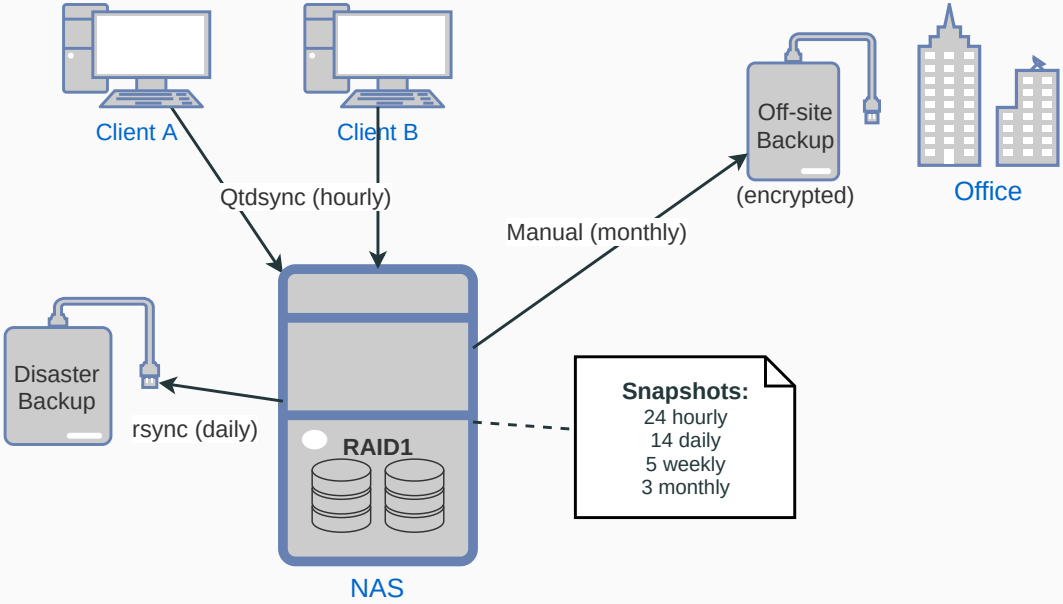
Software

- Proprietary
  - Synology DMS
  - etc.
- OpenSource linux distributions
  - FreeNAS $\rightarrow$ `zfs`
  - Rockstor $\rightarrow$ `btrfs`
  - OpenMediaVault
  - etc.
- Self-made $\rightarrow$ use some basic linux server distribution

Decisions

- Separate small linux server: NAS with `ubuntu-server`
- Flexible hardware configuration.
- Simple user structure.
- Filesystem with snapshots for backups: `btrfs`
- All daily data access via `samba` shares.
  - If the daily usage is too complicated no one uses the features.
  - Snapshots access via `shadow copy`.
- Separate disc for disaster backup:
  - `ext4` → "simpler" than `btrfs`
  - `rsync`
- Backups of client hosts flow to NAS: `qtdsync`
  - Therefore all client data can take advantage of the snapshot mechanism etc.

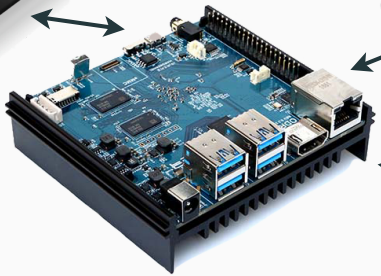
Structure example



Hardware example

USB3 - SATA

RAID1

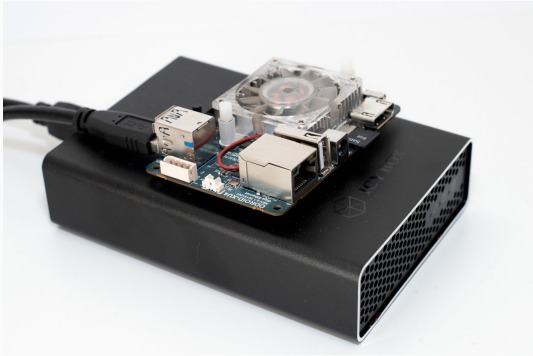


ARM SBC
Odroid N2
6 Cores 4GB
< 5W

Disaster
Backup

Off-site
Backup

Hardware example



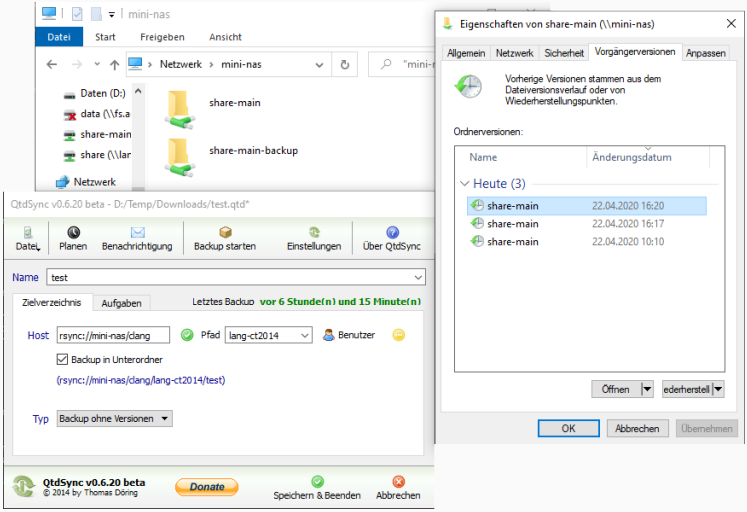
Odroid XU4 with 2x2.5 IcyBox

Odroid N2 with 2x3.5 IcyBox

Demo

github.com/langchr86/mini-nas

VM with `ubuntu-server` installed by `ansible` scripts.



Addendum

RAID configurations

Type	Min. disks	Usable Capacity (# of disks)	Protection
RAID0	2	2	Nothing. Higher risk for loss. Higher speed.
RAID1	2	1	All data is mirrored. Can withstand one failed disk.
RAID01	4	2	Combined RAID0 and RAID1. Similar to RAID1 but with higher speed and higher risk if one disk has failed.
RAID5	5	4	Distributed parity information. Can withstand one failed disk. Recovery is a very intense task.
RAID6	6	4	Same as RAID5 but can withstand 2 failed disks.

Backup methods

Traditional backup solutions use different kinds of backups:

- full
- incremental
  - depends on previous incremental
  - use N incremental + 1 full to recover
- differential
  - depends on last full
  - use 1 differential + 1 full to recover

Hardware-Costs

Usable storage: 2TB

Total costs: 467.-

- **2x Seagate Barracuda 2TB 2.5"** 2x 85.-
- **2x WD Elements 2TB 2.5"** 2x 71.-
- **1x IcyBox IB-RD2253-U31** 55.-
- **1x Odroid N2 4GB** 100.-

Can be reduced:

1. Off-site Backup: -1x WD Elements 71.-
2. RAID1: -1x Seagate 85.-