

Backup Done Right

QING Pei (@edwardtoday)

Permalink: <https://goo.gl/mTFFDt>

What is Backup?

A backup is a **second copy** of
all your important files

for example, your family photos, home videos, documents and emails

Why should I backup?

Having all of your data in only
one place is **dangerous**.

Losing your files is way **more**
common than you'd think.

Ever lost your phone?
And the contacts on it?

Ever lost your camera?
And the precious photos in it?

Ever lost your USB drive?
And the documents on it?

29% of people have never backed up.

- *Backblaze, 2013*

1 in 10 computers infected with
viruses each month.

- ICSA Labs/TruSecure, 2002

“All hard drives will eventually **fail**.”

– *Hard Drive Reliability Stats by Backblaze*

Data loss is almost inevitable.

- Accidental deletion
- Malfunctioning CPU, RAM, power supply, software, etc.
- Stolen laptop or USB drive
- Natural disaster (earthquake, hurricane, etc)

A Good Backup Strategy Will Help

Save your data on multiple

- Devices
- Locations
- Media

3 Kinds of Backups

- A Bootable Backup (or “Clone”)
- External Backup Drive
- Cloud Backup

Manual Backups

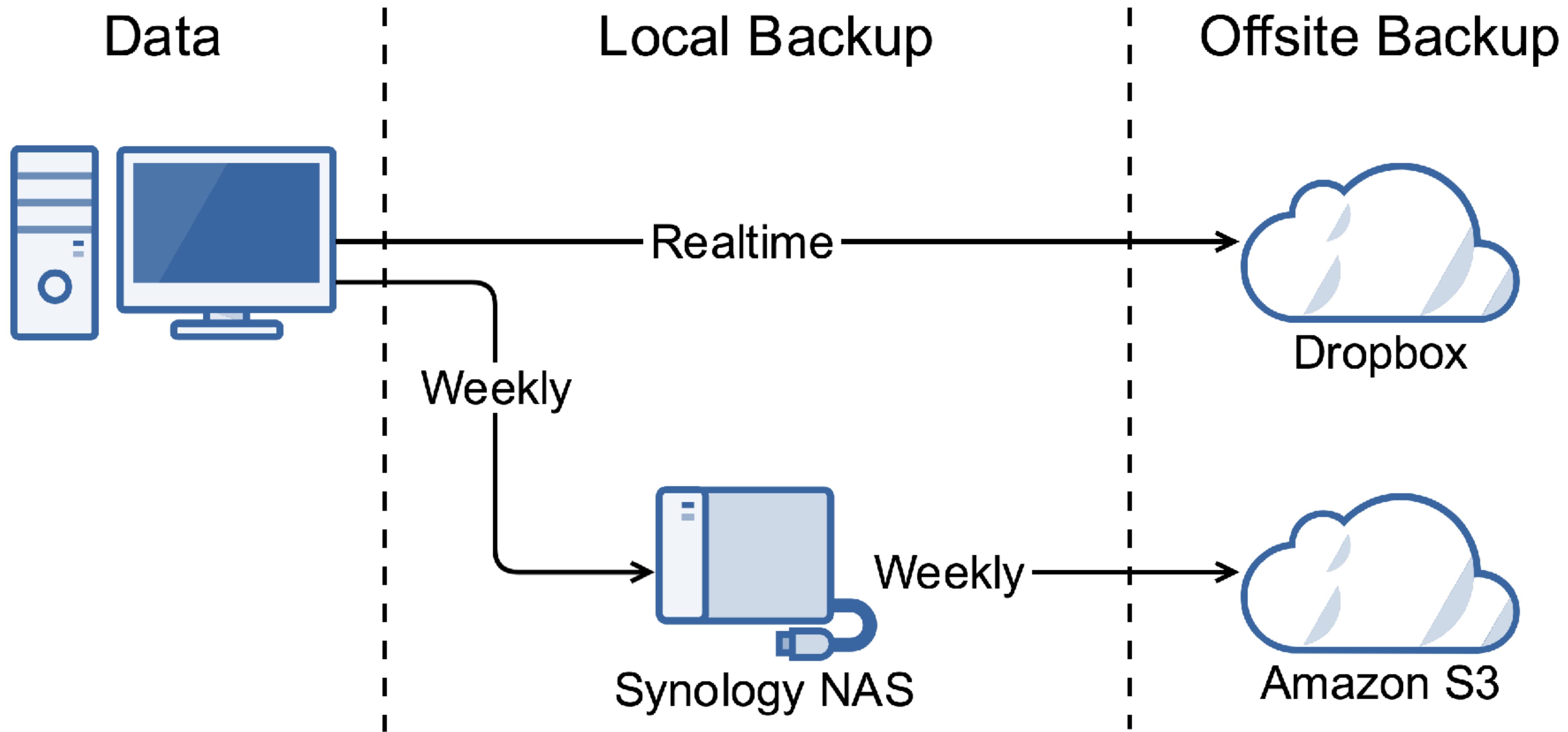
Manual or Automatic?

The 3-2-1 Backup Strategy

The 3-2-1 Backup Strategy

- (at least) **3 total copies** of your data
- **2** of which are **local** but on **different mediums** (read: devices)
- (at least) **1 copy offsite**

My Personal Strategy



Key factors to consider

when you develop a backup strategy

Role

- Who does the backup?
- Or who maintains the automated backup script?

Scope

- **What** to backup?

Time

- **When** to backup?
- **How** often?
 - Recovery Point Object (RPO) – maximum allowed time for lost data
- **How** long to keep the history for?

Cost

- How long does it take to backup?
- More importantly, how long does it take to recover?
 - Recovery Time Object (**RTO**) – amount of time you have to recover a system
 - Maximum Tolerable Period of Disruption (**MTPoD**) – how long can you go before the business REALLY begins to suffer
- How much space?
- Monthly or annual bills?

Effectiveness

- How do I recover from a dataloss?
- Is it still being backed up?
- Is the backup **valid**?

Security

- Encrypt or not?
- Where to keep the passwords and keys?

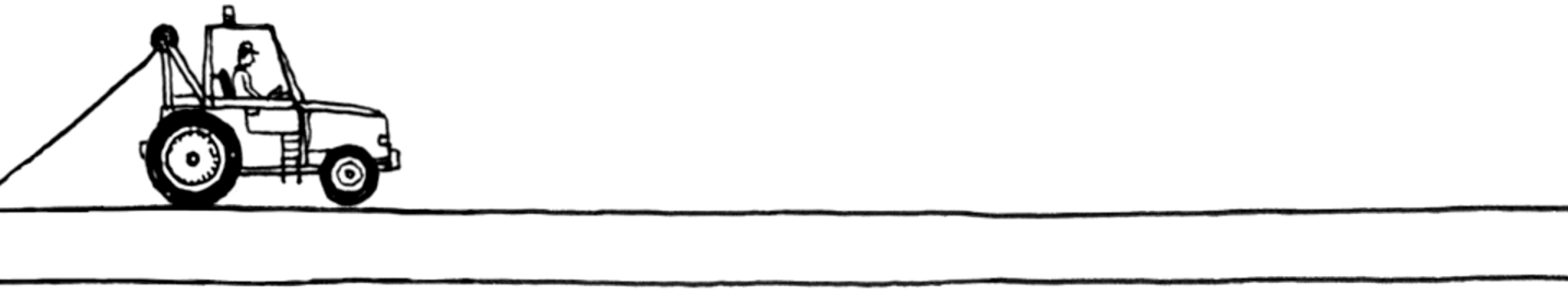
Common Pitfalls



Scheduling backups too infrequently.



Backing Up Manually



Not monitoring backup progress.



Not testing backups.



Not backing up ACLs.

Having a single point of failure



Backing up only data

Highlighted Tools

Backup Utilities

- rsync
- Robocopy
- Time Machine
- Windows Backup
- Duplicacy
- borg
- duplicity
- restic



A performance comparison of
Duplicacy, restic, Attic, and
duplicity

Cloud Backup Services

- Crashplan
- Backblaze
- rsync.net (sidenote: ZFS)
- And cloud storage providers such as Amazon S3



Performance comparisons of
cloud backup storages as
Duplicacy backends

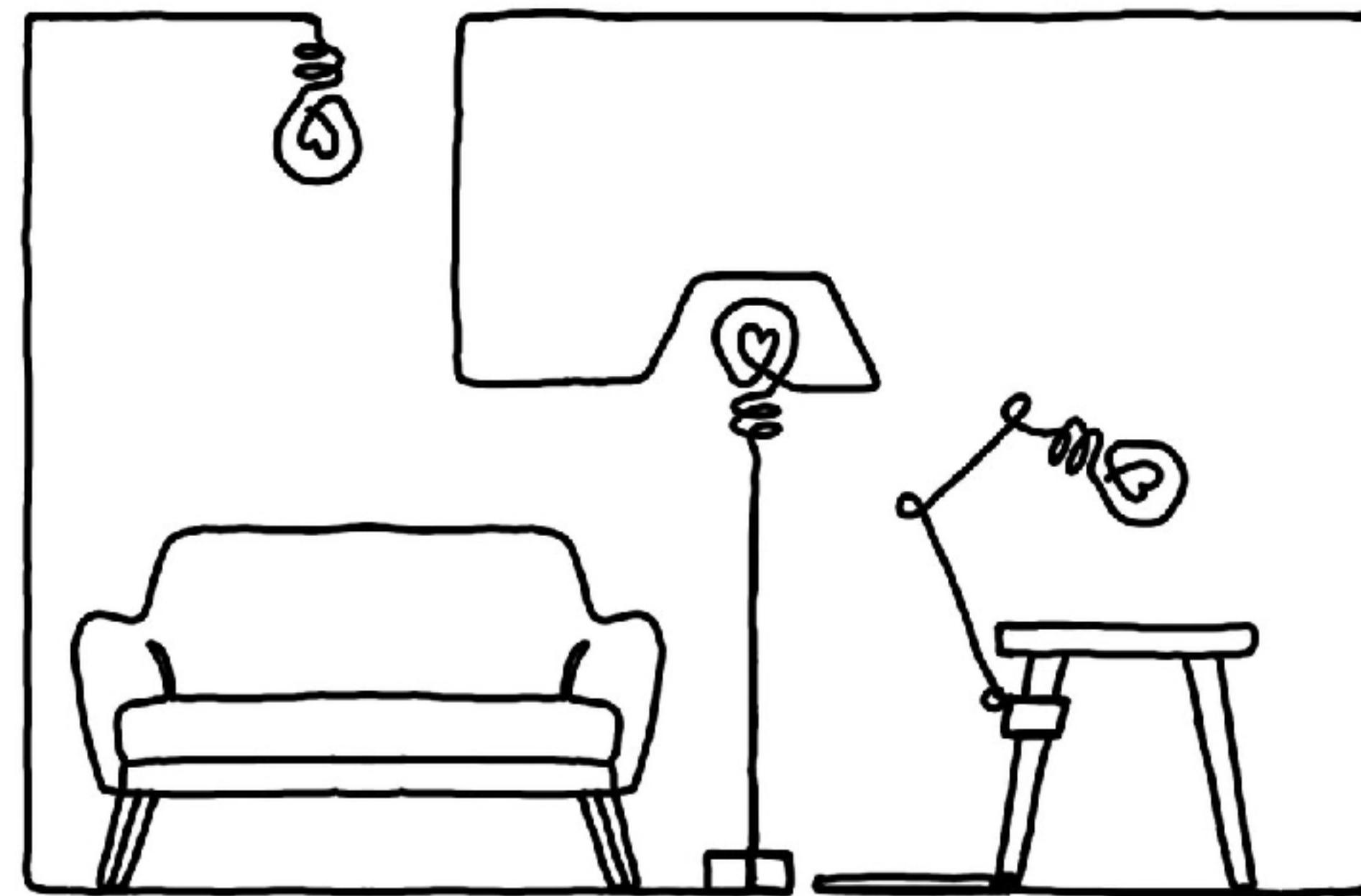
More on Restic

“... for Restic the focus is really on the one hand **security**, but on the other hand **speed** and **usability** ...”

– *Go Time #48: Restic and Backups (Done Right) with Alexander Neumann*

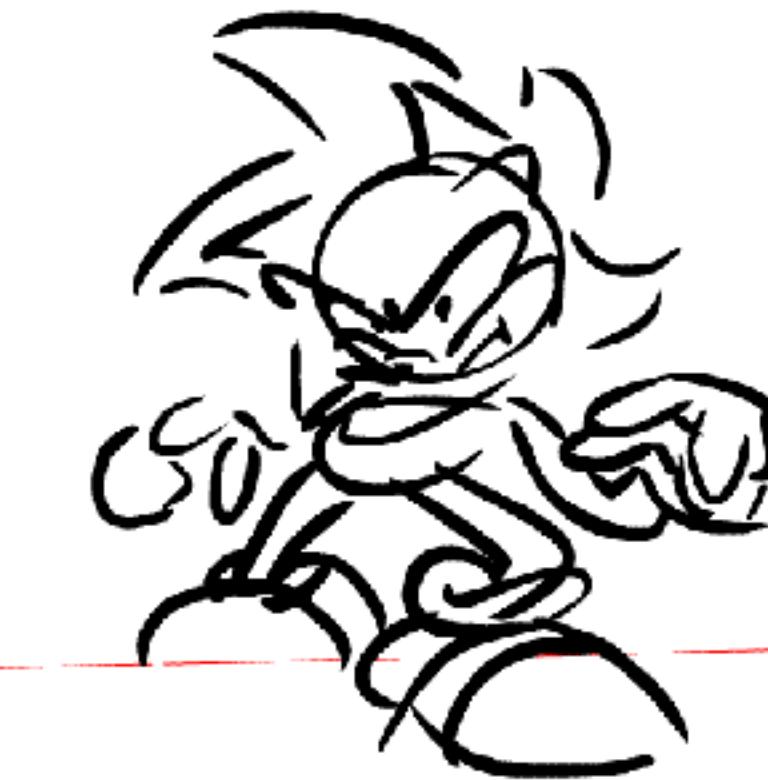
Restic is easy

Easy as 1..2..3



Restic is fast

R



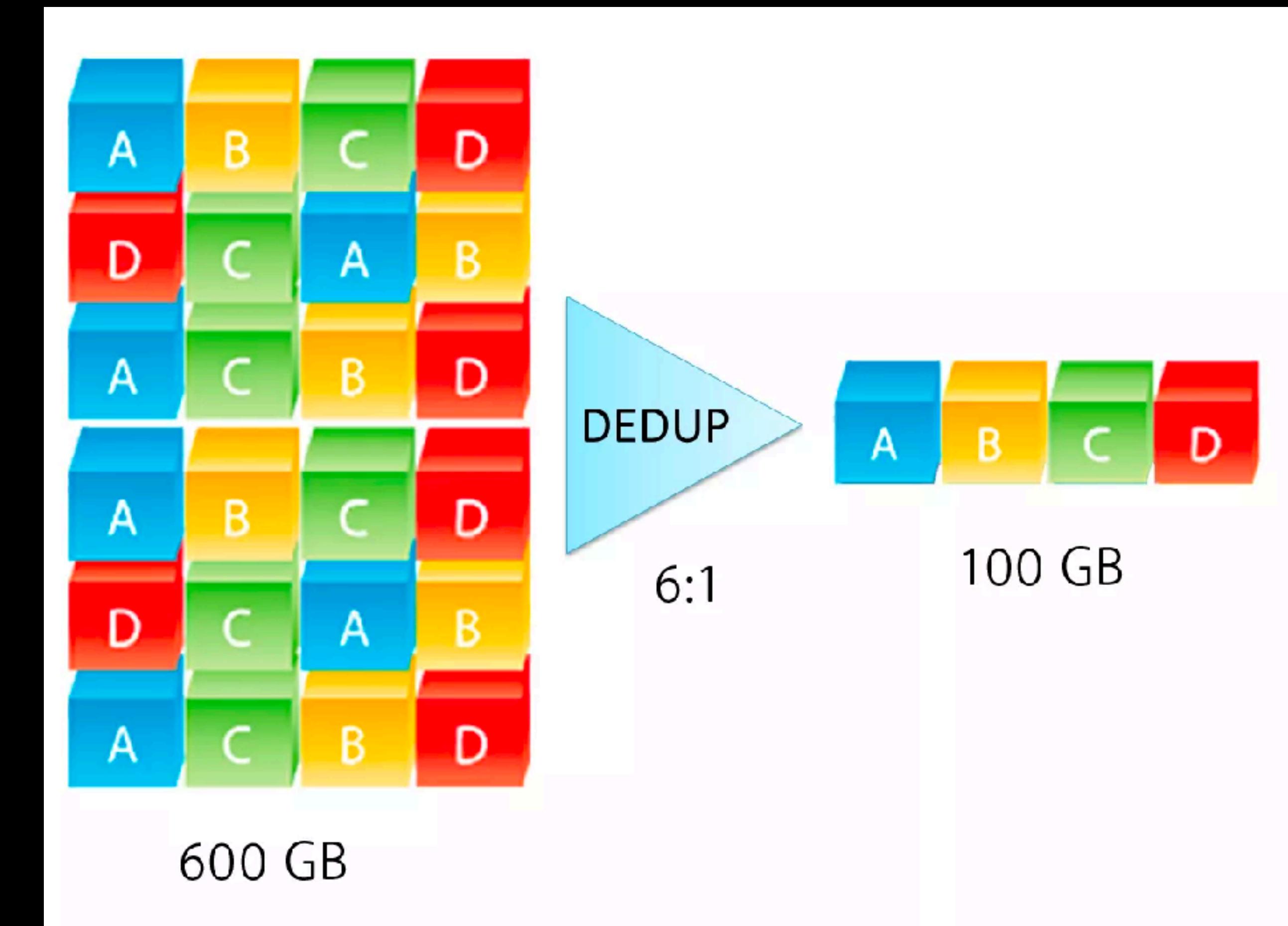
Restic is **verifiable**



Restic is **secure**



Restic is efficient



Restic is
transparent



Features

- Supported OS (Linux, OS X, *BSD, Windows)
- Deduplication
- Stable yet flexible repository format
- Very fast (although still improving)
- Multiple machine backups to one repository
- Pluggable back ends (local, SFTP, AWS S3, ...)
- Browse backups (FUSE support, embedded webserver)

Implementation

- Written in Go ($>=1.3$)
- Crypto: AES-GCM / Poly1305-AES
- Content Addressable Storage (like in Git/Camlistore)
- Metadata: JSON
- KDF: scrypt
- Content Defined Chunking (Rabin Fingerprinting)

Demo

Any questions?

Thank you.