Backup Strategies

CIS 2235 Linux System Administration

Reasons for Backup (duh)

- Disks fail
- Bugs in software can cause corruption
- Configuration mistakes by the administrator
  - Can you say "partitioning"?
- Accidental deletion or overwriting
  - e.g., with rm, mv, or cp
- Malicious virus attack
- Theft
- Fire or other disasters

Help sources

https://help.ubuntu.com/community/BackupYourSystem

Backup considerations

Why? What are you protecting yourself against?
What? What do you need to back up?
When? When will you do the backups?
Where? Where will you store your backups?
How? What type of media will you use?
To inform these decisions:

- Recovery time — how quickly do you need to recover?
- Recovery point — how much data can you afford to lose?

Backup Media

- Traditionally, backups have been made onto *tapes*
  - Can store lots of data on reasonably cheap tapes
- Copying to a different *hard disk*
  - There is a risk of losing the backup along with the original
  - Even better if on a remote computer
- CD/DVD *disk* writers can be used to store backups
  - Convenient for long-term storage
  - Handy to remove to remote locations
- More and more, *network* backup is becoming popular
  - AMANDA
  - a.k.a. "Cloud"

Types of Backup

Full backup

- includes everything of importance
- The current backup is contained in only one file – the whole thing
- The most complete, but takes the most time/space
- includes many files which hardly ever change

Differential backup

- only includes changes since the last full backup
- The current backup is contained in 2 files – the full and the last differential
- e.g., nightly backup only needs to include files changed since the last full backup (whenever that was, e.g., last Sat)
- Some files are backed up multiple times

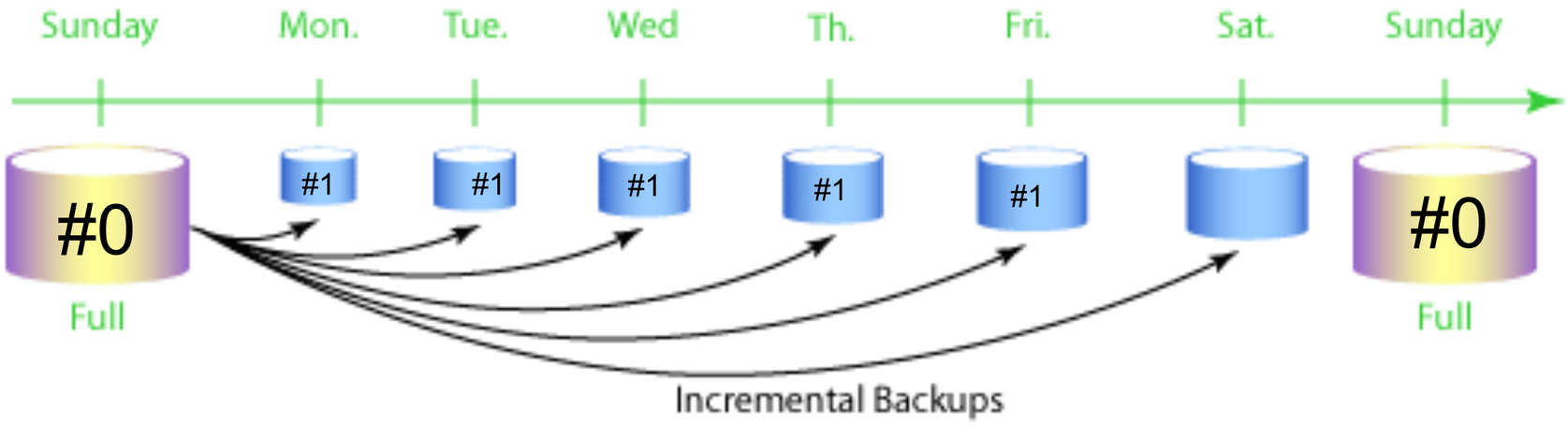
Types of Backup

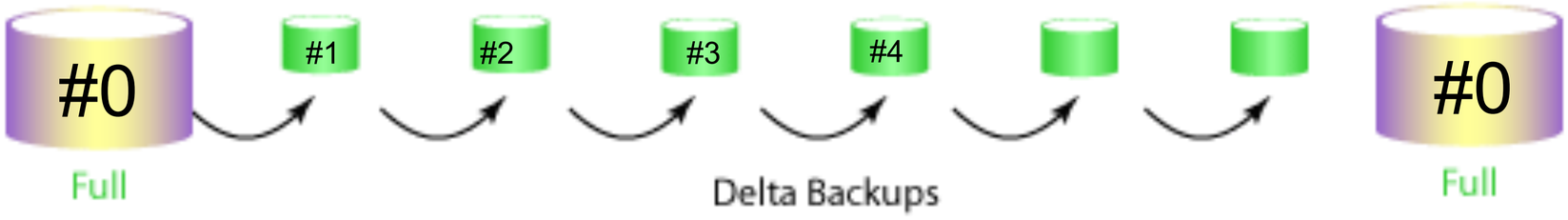
Incremental backup

- only includes changes since the last backup of any "type" (F or D)
- The current backup is contained in the full, plus every incremental since
- Most "efficient" – files are only backed up if they recently changed
- e.g., nightly backup only includes files changed in the last 24 hours

Can you identify the types of backups?

"Differential" – keep *replacing* the backup file from full





"Incremental" – keep *adding* more smaller files to the set

How do I get those files back?

Getting files from a backup is called restoring

Full backup

- Just restore everything from the one full backup
  - 1 step process

It takes longer to create a full backup

Easiest to restore

How do I get those files back?

Differential backup

- First, restore the full backup.
- Next, restore the most recent differential.
  - A 2-step process

Diff's save time on the backup but require a little more effort for the restore.

How do I get those files back?

Incremental backup

- First, restore the full backup, then any differential
- Then, restore all the incrementals since the last full or differential

Incrementals are the fastest to create as backups

Take the most work on the restore

Backup Strategy Considerations

1) Have a strategy
   Regular and verified
2) Document backup and restore procedures
   Backup medium, location, duration, frequency decisions
3) Minimize risk with multiple and various locations/time
   Balance backup throughput and resource costs with insurance and restore efforts if needed

Typical strategy:
- a full backup is done once a... week, month
- daily changes are differential or incremental each night
- Can mix and match differentials and incremental

Backup big picture

Property	FULL	DIFF	INC
Files included			
Run Frequency			
Backup file size			
Backup Speed			
Redundancy			
Ease of Recovery			

Backup Tool (Graphical User Interface)	
Areca Backup	File backup software developed in Java
BackupPC	High-performance, enterprise-grade system for backing up PCs
Bacula	Network backup, recovery and verification
fwbackups	Feature-rich backup sofware
Keep	Backup system for KDE
Simple Backup Solution	Set of backend backup daemon and Gnome GUI frontends

Backup Tool (Command-line)	
afbackup	Client-Server Backup System (GUI is also available)
AMANDA	Advanced Maryland Automatic Network Disk Archiver
Cedar Backup	Local and remote backups to CD or DVD media
Duplicity	Encrypted bandwidth-efficient backup
→ **Dump / restore**	Dump and restore utilities for ext2/ext3 filesystems
→ **tar**	Tar archiving utility

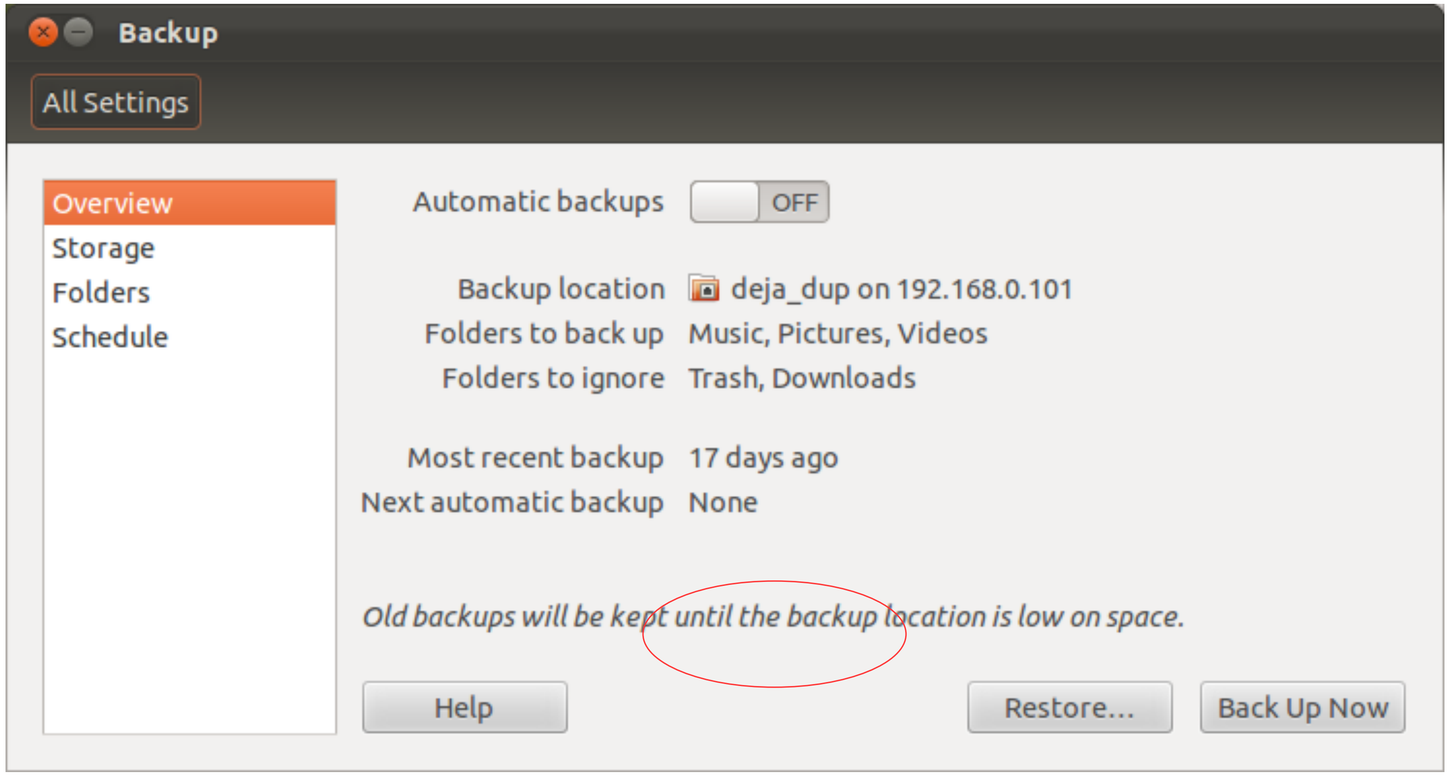
Snapshot backups	
FlyBack	Equivalent of OS X's Time Machine
Time Vault	Snapshotting daemon

Synchronisation	
rsnapshot	Local and remote filesystem snapshot utility
→ **rsync**	Fast remote file copy program

Disaster Recovery / Disk Cloning	
Clonezilla	Offers similar functionality to Symantec Ghost
Mondo Rescue	A powerful disaster recovery suite
PartImage	Backup partitions into a compressed image file
PING	Also offers similar functionality to Symantec Ghost

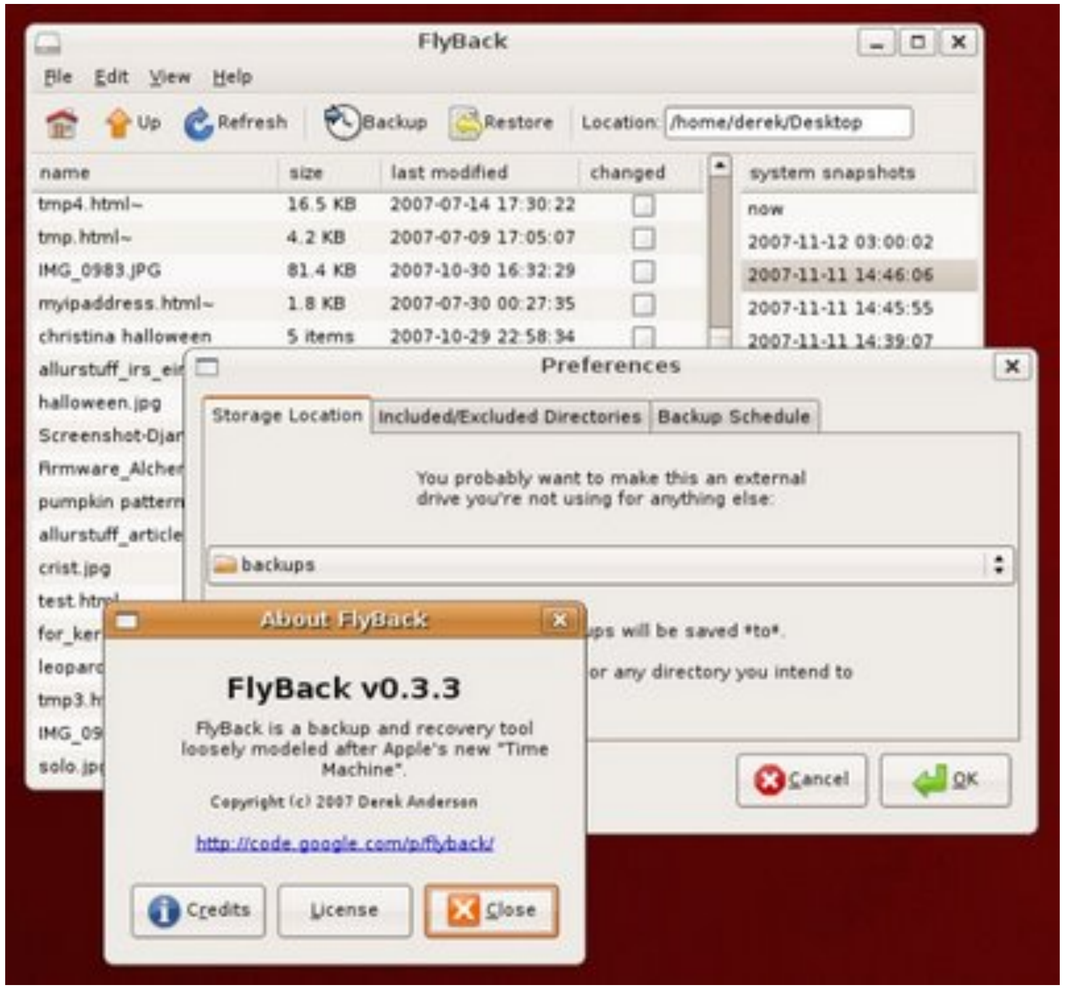
Specialist	
Zmanda	Perl-based utility to automate backup and recovery of MySQL databases

GUIs for linux

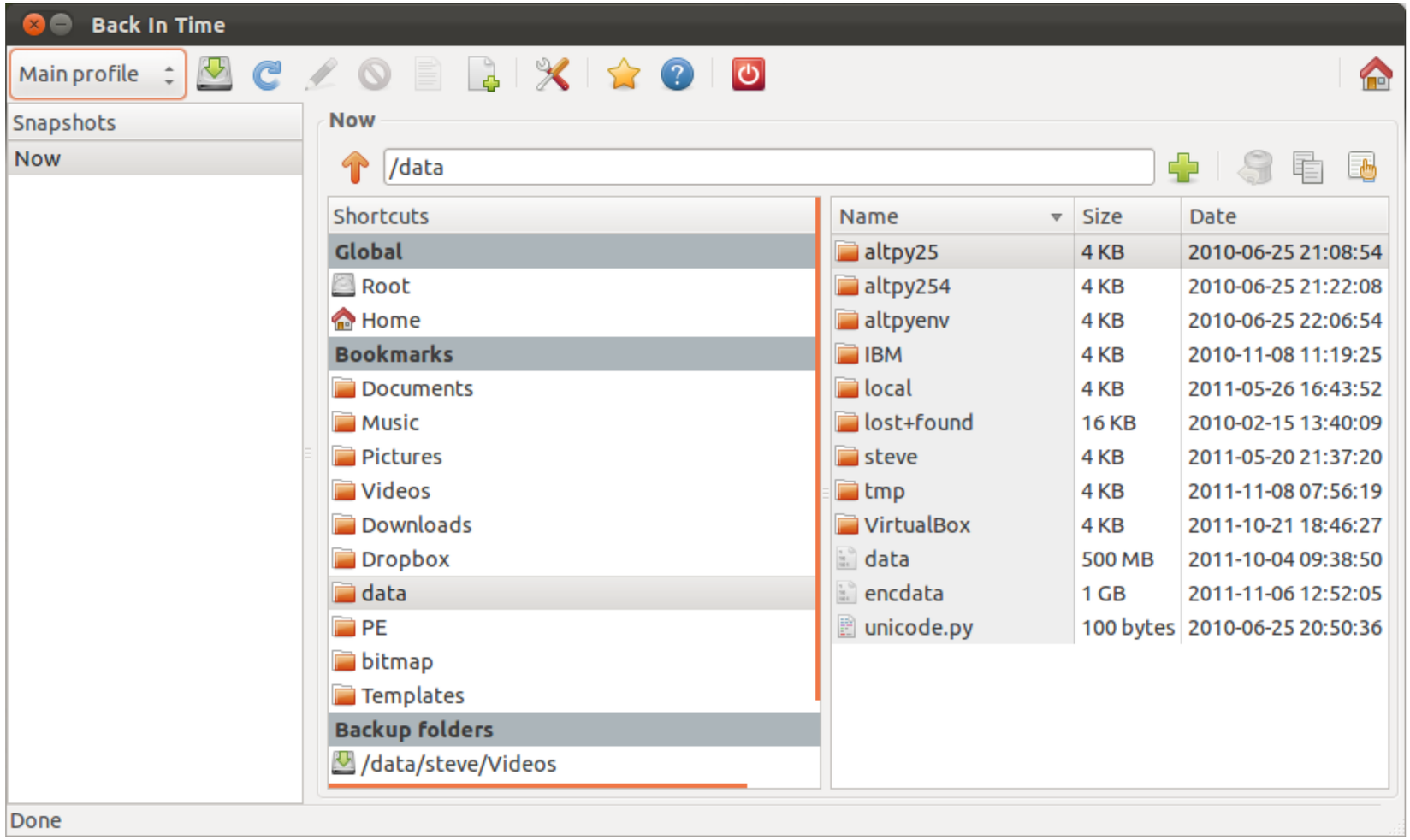


Default Ubuntu Backup = Deja Dup
Deja Dup is a front end to Duplicity, which uses rsync for backup

GUIs for linux



GUIs for linux



Unix dump/restore

"default" Unix backup strategy is `dump` & `restore`

dump creates a backup for entire partitions (not a directory)

fast - reads the inodes, not the directory tree

dump puts the whole partition into a single file

(some compression by combining partially-filled data blocks)

It also maintains permissions, owners, dates, etc. in the dump file

restore can do full or partial recoveries

Incorporates the concept of backup levels

Levels

Think of backups as "levels"

Level 0 is a full backup

Each higher level backs up files only modified since the most recent lower level.

- level 1 backs up files since the last level 0 (differential)
- level 2 backs up files since the last level 1 (incremental)
- level 3 backs up files since the last level 2 (smaller increment)

There is nothing "special" about each level, except level 0 is FULL, and each one "looks" to the previous level only.

Unix started with Levels 0-9 only.

You don't have to use the levels in any order. You can "skip around" to build your strategy!

Strategy 1: Daily Incremental

Keep increasing the level number
After L0, only backup new files since last backup ("Incremental")

Day 1	level 0	full	
Day 2	level 1	inc. since L0	1d of changes
Day 3	level 2	inc. since L1	1d of changes
Day 4	level 3	inc. since L2	1d of changes
Day 5	level 4		
Day 6	level 5		
Day 7	level 6		
Day 8	level 7		
Day 9	level 8		
Day 10	level 0	full	

Backup files are staying small and backup times are quick.

Strategy 1: Daily Incremental

To recover:

Restore L0, then *each and every* new level up the last day stored (n recovery steps)

Strategy 2: Daily Differential

Keep one level for all backups

Day 1 level 0 full

Day 2	level 1	diff. since L0	1d of changes
Day 3	level 1	diff. since L0	2d of changes
Day 4	level 1	diff. since L0	3d of changes
Day 5	level 1		
Day 6	level 1		

Backup files are getting bigger and backup times are getting longer.

Day 10 level 0 full

Strategy 2: Daily Differential

To recover:

Backup L0, then <u>only</u> the last day stored (just 2 recovery steps)

Strategy 3: mix of both Inc and Diff

Start with Full, use inc. most days, but add diff's as desired

Day 1	level 0 full		
Day 2	level 2	inc. since L0	1d
Day 3	level 3	inc. since L2	1d
Day 4	level 4	inc. since L3	1d
Day 5	level 1	diff. since L0	d 2-5
Day 6	level 2	inc. since L1	1d
Day 7	level 3	inc. since L2	1d
Day 8	level 4	inc. since L3	1d
Day 9	level 1	diff. since L0	d 6-9
Day 10	level 2	inc. since L1	1d
Day 11	level 3	inc. since L2	1d
Day 12	level 4	inc. since L3	1d
Day 13	level 5	inc. since L4	1d
Day 14	level 0	full	

Small, quick backup (Inc)

A little larger and longer, but not the whole thing (Diff)

Strategy 3: mix of both Inc and Diff

To recover day 12:

Restore L0,
Day 9 (L1),
Day 10 (L2) and
Day 11 (L3)

4 steps (but not 12)

A popular strategy 10-day based on Tower of Hanoi

"backup file"

Tape	Level	Backup *days*	Restore tapes
1	0	N.A.	1
2	3	1	1, 2
3	2	2	1, 3
4	5	1	1, 2, 4
5	4	2	1, 2, 5
6	7	1	1, 2, 5, 6
7	6	2	1, 2, 5, 7
8	9	1	1, 2, 5, 7, 8
9	8	2	1, 2, 5, 7, 9
10	9	1	1, 2, 5, 7, 9, 10

1's are inc from prev day
2's are diff from day before yesterday

Writing a backup plan

Sites should have a written backup strategy that answers at least the following:

Overall strategy:

- What data is backed up?
- What system or technology will perform the backups?
- Where will the backup data be stored?
- Will backups be encrypted? If so, where are the encryption keys stored?
- How much will it cost to store backups over time?

Writing a backup plan (2)

- Timelines
  - How often will backups be restored?
  - How often will backups be validated, restored, and tested?
  - How long will backups be retained?
- People
  - Who will have access to backup data?
  - Who will have access to encryption keys that protect backup data?
  - Who will be in charge of verifying the execution of backups?
  - Who will be in charge of validation and restore testing of backups?