CS 200

System Management

Backups
Let’s discuss the facts of life...

You will die.

Your disk will die.

Your *disk* will die *first*.
Who does them?

The user, manually
   at day’s end, make copies on ...
   and, if you’re smart, before you make major changes

The user, by running a program, based on time-of-last-modification
   eg Retrospect, Norton Fastback, SmartBackup

Your PC, or a network server, by automatically running an application, based on time-of-last-modification
   eg Retrospect + Retrospect Client
Terminology (1)

A Full Backup
Back up everything, whether it’s changed or not

Incremental
Back up only files changed or created since the last backup

Duplicate sets
Make n copies of each backup \((n > 1)\)

Rotating sets
Rotate through n sets of backup tapes/CDs/DVDs/… \((n > 1)\)
Terminology (2)

Disk mirroring
  Replicating the contents (or changed contents, & deleting deleted files) of one disk to another

Incremental mirroring
  Copy changes to a duplicate disk

Offsite storage
  Far enough away as to eliminate the risk of one event destroying both your PC and the backup

Archival storage
  Periodically take a backup set out of service & “archive it”
## A Weekly-Cycle Incremental Backup Strategy

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>Incremental 1</td>
<td>Incremental 2</td>
<td>Incremental 3</td>
<td>Incremental 4</td>
</tr>
<tr>
<td>Full</td>
<td>Incremental 1</td>
<td>Incremental 2</td>
<td>Incremental 3</td>
<td>Incremental 4</td>
</tr>
<tr>
<td>Full</td>
<td>Incremental 1</td>
<td>Incremental 2</td>
<td>Incremental 3</td>
<td>Incremental 4</td>
</tr>
<tr>
<td>Full</td>
<td>Incremental 1</td>
<td>Incremental 2</td>
<td>Incremental 3</td>
<td>Incremental 4</td>
</tr>
<tr>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
</tr>
</tbody>
</table>

To recover an individual file

work backwards (4, 3, 2, 1, Full) until you find it

To restore the entire file system

work forwards (Full, 1, 2, 3, 4) restoring everything
More Terminology

If you keep a “catalog” of what’s in each backup

ie a table-of-contents (of file paths, size, dates of creation/modification, location in backup)

you could search that
to identify and restore the most recent version of each file
instead of copying them all & over-writing older versions

and you could likely keep the catalog on disk, making retrieval of an individual file much faster

A “session” is a particular backup (eg the Week 2, Weds incremental backup)

A “snapshot” is exactly what’s on your disk at the time of a backup

Files not changed since the last previous backup

are in the snapshot

but not in the session
A Duplicate Set (2) Incremental Backup Strategy

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full A</td>
<td>Incremental A1</td>
<td>Incremental A2</td>
<td>Incremental A3</td>
<td>Incremental A4</td>
</tr>
<tr>
<td>Full B</td>
<td>Incremental A1</td>
<td>Incremental A2</td>
<td>Incremental A3</td>
<td>Incremental A4</td>
</tr>
<tr>
<td>Full A</td>
<td>Incremental A1</td>
<td>Incremental A2</td>
<td>Incremental A3</td>
<td>Incremental A4</td>
</tr>
<tr>
<td>Full B</td>
<td>Incremental A1</td>
<td>Incremental A2</td>
<td>Incremental A3</td>
<td>Incremental A4</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

If one backup goes bad, you have another equally recent backup

If lightning fries your machine AND the backup you are making, you can restore from the other backup, and it’s equally recent

But … it takes twice as many tapes / disk(ette)s, and twice the time
A Rotating Set (2) Incremental Backup Strategy

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full A</td>
<td>Incremental A1</td>
<td>Incremental A2</td>
<td>Incremental A3</td>
<td>Incremental A4</td>
</tr>
<tr>
<td>Full B</td>
<td>Incremental B1</td>
<td>Incremental B2</td>
<td>Incremental B3</td>
<td>Incremental B4</td>
</tr>
<tr>
<td>Full A</td>
<td>Incremental A1</td>
<td>Incremental A2</td>
<td>Incremental A3</td>
<td>Incremental A4</td>
</tr>
<tr>
<td>Full B</td>
<td>Incremental B1</td>
<td>Incremental B2</td>
<td>Incremental B3</td>
<td>Incremental B4</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

If the most recent backup goes bad, you have another, though you will lose recently created files and recent changes.

If lightning fries your machine AND the backup you are making, you can restore from the other backup.

An advantage — if you’re infected by a virus

and some infected files were backed up

you can get older, uninfected versions from the older backup.

No more time / work is involved

but … it takes twice as many tapes / CDs / DVDs / disks / ...
### A Duplicate Set (2) + Rotating (2) Incremental Backup Strategy

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full A</td>
<td>Incremental A1</td>
<td>Incremental A2</td>
<td>Incremental A3</td>
<td>Incremental A4</td>
</tr>
<tr>
<td>Full B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full C</td>
<td>Incremental C1</td>
<td>Incremental C2</td>
<td>Incremental C3</td>
<td>Incremental C4</td>
</tr>
<tr>
<td>Full D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full A</td>
<td>Incremental A1</td>
<td>Incremental A2</td>
<td>Incremental A3</td>
<td>Incremental A4</td>
</tr>
<tr>
<td>Full B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full A</td>
<td>Incremental C1</td>
<td>Incremental C2</td>
<td>Incremental C3</td>
<td>Incremental C4</td>
</tr>
<tr>
<td>Full B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

... ... ... ... ...

**Combines the advantages of both**

**But takes more time, and more tapes / CDs / DVDs / ...**
Data Compression

x 2 (on average) is easy

many backup programs use data compression
must use the backup program to retrieve a file
whereas with a file-system-formatted backup
you just mount the disk & use standard file copying techniques

A related issue: O/S-format or proprietary format

disks are allocated in 512-byte “allocation blocks” (for efficiency)
some backup programs
create one gigantic file &
write your files inside it,
keeping track of file locations themselves
so as not to waste disk space
but then must use the backup program to retrieve the file;
if a piece of your disk / CD / DVD goes bad, or the vendor out of business, you’re probably up the creek
Security

Do you want a backup program (or file archiving program) that offers
(optionally) encrypted storage?
(optionally) encrypted data transmission (if done over a network)?
how good should the encryption be?
File > Save... Backups

Suppose your application **crashes while saving your work?**

When you select File > save, many apps will optionally change the name of the existing document file to something like “Backup of ...”, then write your current document to a *new* file on disk

MS Word, for example:
Ditto For Excel

Except that it’s a deeply buried File > Save As... option
And for TextWrangler

And so on...
Train yourself to do a File > Save

- automatically
- every time you pause to think between changes
Folder Archives

When you’re working with multiple files in a folder

Periodically make a compressed archive of the folder’s contents

eg in Mac OS X’s Finder, select the folder, then select File > Compress “• • •”
eg in Windows XP’s Explorer, File > Send To > Compressed (zipped) Folder
or the analogous items on the contextual (right-click) menus in each system
Numbering These “Checkpoints”

Numbered files (after a File > Save)

  someDocument_1.doc
  someDocument_2.doc
  someDocument_3.doc
  ...

Numbered archives

  someFolder_1.zip
  someFolder_2.zip
  someFolder_3.zip
  ...

Sometimes it’s useful to keep a list of what you’d just finished in each checkpoint
Time Machine (OS X 10.5 & Later)

Uses a separate disk (or partition)

Keeps
- hourly backups for the past 24 hours
- daily backups for the past month
- weekly backups until your backup disk is full

Each backup *looks* like an exact copy of your disks; *actually*, “hard links” are used so there’s only one copy of each version

Effectively a full backup + incrementals

Multiple (simultaneous) sets not possible

You can (manually) change backup disks ⇒ rotating backups and off-site archives *are* possible