Backups

COMP620

Protecting Ourselves

- It doesn't take a malicious hacker to make you lose your data
- You can do it all by yourself

When you use your computer, where do you usually store your data?

1. Hard drive
2. USB thumb drive
3. External hard drive
4. other

When was the last time you backed up the data on your computer?

1. Sooner than a week ago
2. One week ago
3. One month ago
4. Longer than a month
5. Never

When was the last time you backed up your phone?

1. Sooner than a week ago
2. One week ago
3. One month ago
4. Longer than a month
5. Never

Have you ever lost your data?

1. Never
2. Once
3. Several times
4. Many times

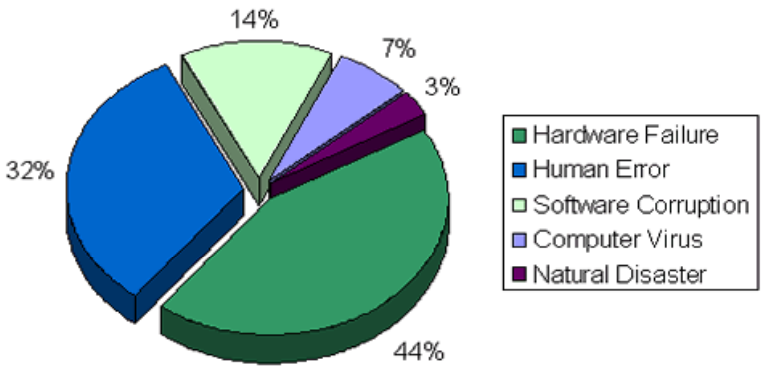
Your Disk Will Die!

It is a matter of **when**, not **if**.

Types of Data Loss

- User error (accidental deletion, lose CD, forget password, etc.)
- Disk failure
- Data corruption (file system failure)
- OS failure
- Power failure (cached data not saved)
- Natural disaster (fire, flood, plague of locust)
- Malware
- Theft of equipment

Causes of Data Loss



Information Assurance

It is less probable that somebody else will destroy your data than you will lose your data all by yourself.

Cost of Data

- How much can you afford to lose?
- Backup often enough so that you can afford to lose the data since the last backup.
- Many users can easily recover with daily backups. Some systems need immediate backups.

Risk Analysis

Consider

- The cost of the loss of data
- The cost of backing up your data
- The probability that a failure will occur over a period of time

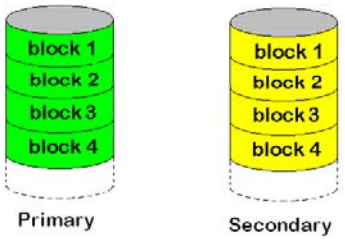
Concurrent Remote Backup

- Some systems are very concerned about losing data.
- Every update is transmitted to a remote backup facility as it is made
- Database audit trails record all changes made to the database

RAID

- RAID 1 – 6 will improve hardware reliability
- Most data loss is not the result of hardware failure
- If the user or software deletes a file, the hardware will dutifully delete it from all disks in a RAID bank.

RAID 1 (mirrored)



- Improved Reliability
- Slightly slower writes.
- Possibly faster reads
- Twice the disk space required

XOR Parity

- Consider the exclusive OR of several values

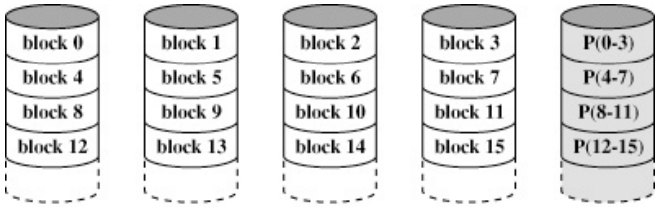
$$X = A \oplus B \oplus C$$

- If you XOR any of the three values, you will get the fourth.

$$B = X \oplus A \oplus C$$

- RAID 3, 4, 5 & 6 write the XOR of data to an additional disk to provide recovery in the event a disk fails.

RAID 4 (block-level parity)

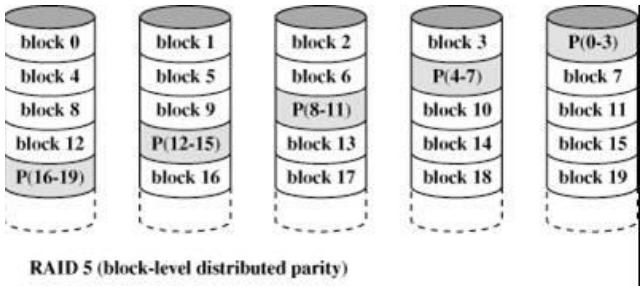


(e) RAID 4 (block-level parity)

Figure 11.9 RAID Levels (page 2 of 2)

Rarely used

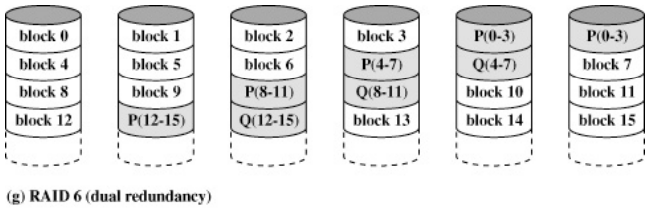
RAID 5 (distributed block parity)



RAID 5 (block-level distributed parity)

- Striping improves read performance
- Parity improves reliability
- N+1 disks are required

RAID 6 (dual redundancy)



(g) RAID 6 (dual redundancy)

Figure 11.9 RAID Levels (page 2 of 2)

- Like RAID 5 but with two parity blocks for each data block
- Slow writes
- N+2 disks required

Common Backup Scheme

- Backup all data on a quarterly basis *(possibly more or less frequently)*
- Do incremental backups of new and modified files daily.

Backup Media Considerations

- Cost
- Speed
  - Data transfer rate
  - Ability to restore an individual file
- Reliability
- Volume
- Life span

Backup Devices

- CD – read-only units of 600-750 MB
- DVD – currently rather expensive, price dropping
- USB memory sticks – expensive, limited size
- Remote server – initial expense, but effective
- External hard drive - initial expense, but effective

- Floppy disk – small and unreliable
- Tapes – slow and unreliable

Backup Data Storage

- Removable backup media should not be stored near the computer. A disaster destroying the computer will destroy the backups.
- It should not be too difficult to retrieve the backup media in case a file needs to be restored.

Sad Story

- A graduate student was almost finished with his thesis.
- He carefully backed up his work every day to floppy disks and stored the labeled floppy disks in a box next to the computer.
- A thief stole his computer and also took the box of floppies.

Saving the Backups

- We used to do an incremental backup of a mainframe system every day at 3:00am
- Friday nights we did a full backup
- The previous weeks tapes were stored in the vault of a local bank.

Security

- Your backup media might be sitting on a shelf while your computer is locked away.
- Your data can be read from your backup files.
- If someone can change your backups and then force a restore, they can change the data on your computer.
- Encryption can help protect backup data.

Backup Systems

- **Microsoft Home Server** – Once a day it will backup all new and changed files over a LAN to a PC with multiple disks.
  - It can RAID any collection of disks
  - Hibernating or standby computers will be waken to perform the backup and then returned to standby.

What to Backup

- It is difficult to do a full backup of a 300GB disk drive every night
- Some parts of your system probably do not change often.
- Identify the directories that hold your important and frequently changed information

Backup Tools

- Manually copying file to a device will work, but it is time consuming an error prone
- I use WinZip. A WinZip “job” compresses my new and changed files to a thumb drive with a filename based on the date
- Windows and Linux provide backup utilities

Shadow Copies

- Microsoft Windows Vista and 7 create backups of files that have been changed
- Creating a Restore Point creates backups of user and system files
- Restore Points are usually created once a day and before any system change
- You can recover a previous version of a file by right clicking and selecting restore