Data Loss and Recovery

By Joshua Menezes
and Christian Tabbah

Table of contents

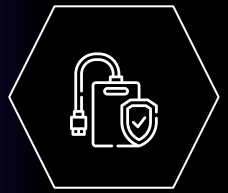
Introduction

Data loss is quite a common problem when it comes to using technology. In this presentation, we will focus on the different types of data loss as well as how you can recover from data loss if it were to happen.

Background

Background information about data loss recovery

Concepts



Physical Data storage

How is data stored physically?

- HDD’s
- SSD’s
- CDs and DVDs

Logical Data storage

How is data stored virtually?

- FAT32
- exFAT
- NTFS
- EXT4
- APFS
- ZFS

How is data stored physically?

Data is stored on Hard Disk Drives(hdd) or solid-state drives(ssd)

Hard Disk Drives – Key Components

Disk / Platter

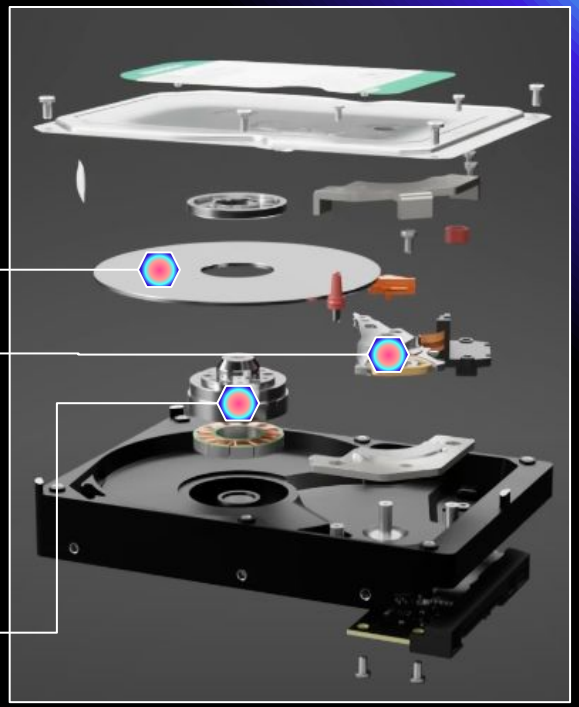
Stores all the data(can be multiple platters tall)

Head Stack Assembly

Used to read and write on different parts of the disk as it spins

Spindle

Spins the disk at a speed of 7200 RPM(using a motor)



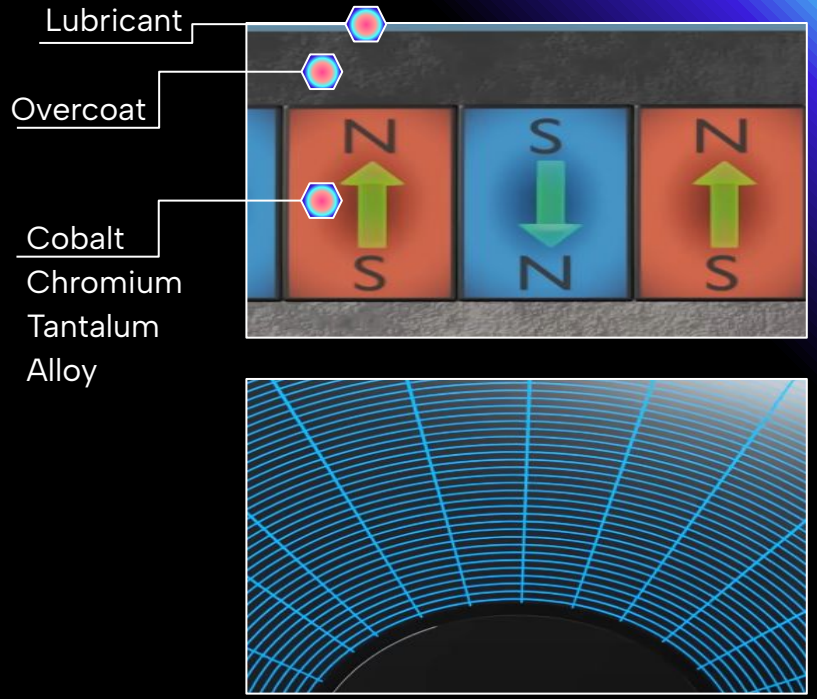
Hard Disk Drives – The Disk

Disk Material

A Cobalt Chromium Tantalum Alloy is chosen because it has small magnetic regions, who's direction can be manipulated using external magnetic fields

Tracks and Sectors

Each disk is divided into tracks and sectors. The head stack assembly will navigate these zones to find different parts of data on the disk



Hard Disk Drives – The Disk

Track Sectors

Preamble/ Synchronization Zone:

- Helps with the movement of the head stack assembly

Address:

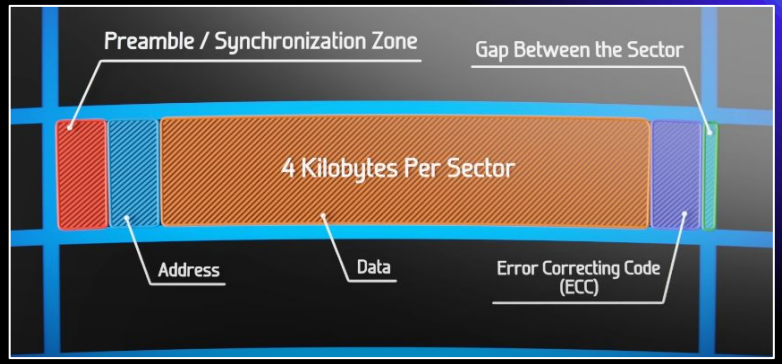
- Tells read-write head which track its on

Data:

- Actual stored data

Error Correcting Code(ECC):

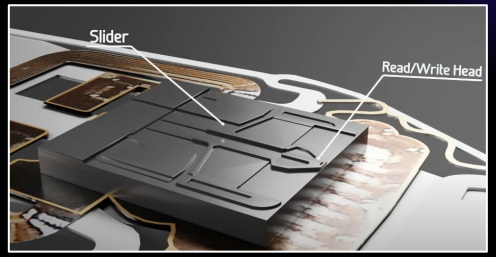
- Verifies that data is accurately written and properly read



Hard Disk Drives – Head Stack Assembly

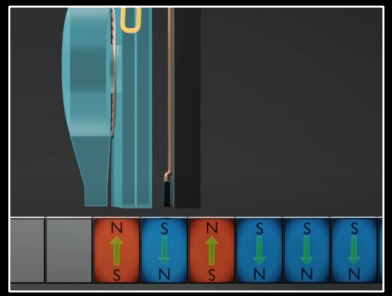
Writing to disk

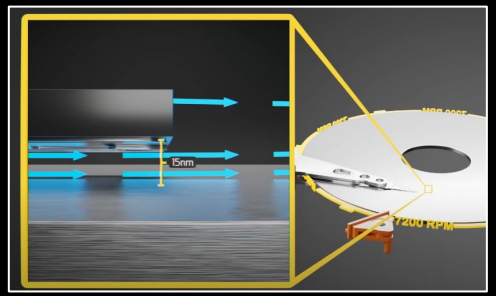
Writing to disk is done by manipulating the direction of magnetisation of each localized region in the disk. Each direction corresponds to either a 1 or a 0.



Reading from disk

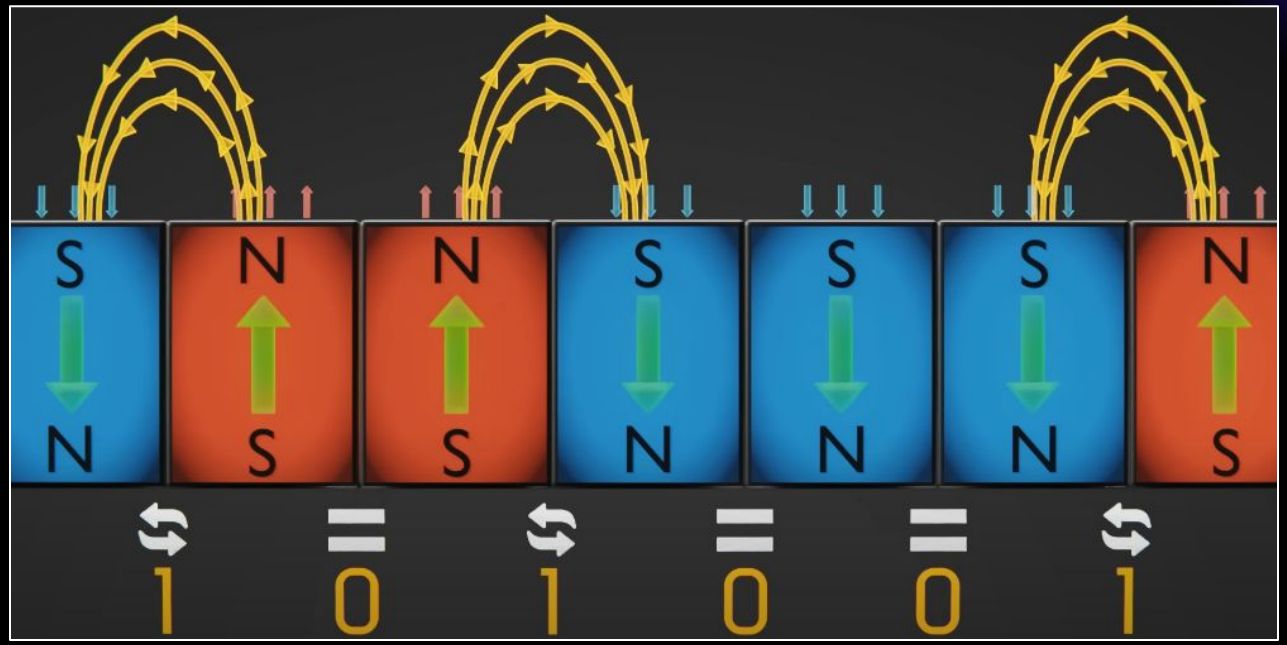
Reading from disk only requires the read head to sense the magnetic field direction of each localized region.





Hard Disk Drives – Clarification

What is a 1 and what is a 0?



Solid State Drives(SSDs) – Storage

SSD Storage

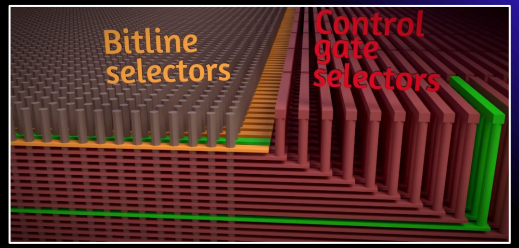
SSDs store data in billions of charge trap flash memory cells

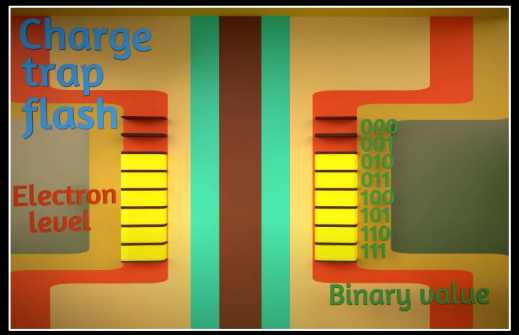
Charge Trap Flash

Each store 3 bits of information by trapping different levels of electrons.

Speeeeeeed

SSDs are much faster than HDDs due to their data access speed and the fact that they have no moving parts





Compact Discs(CDs) vs. Digital Versatile Discs(DVDs)

Compact Discs

- The data is stored on a single data track that spirals outwards
- The disk spins as the handle moves outwards to follow the spiral
- In the aluminum on the track, there are dents:
  - Each bump represents a 0
  - Each divot represents a 1
- Capacity of about 700MB, normally used for audio files

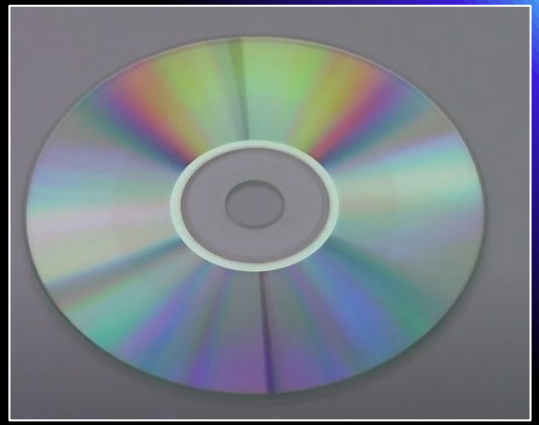
Digital Versatile Discs

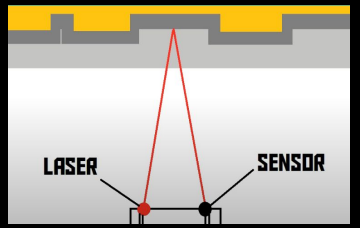
- The data is stored on a single data track that spirals outwards
- The disk spins as the handle moves outwards to follow the spiral
- In the aluminum on the track, there are dents:
  - Each bump represents a 0
  - Each divot represents a 1
- Capacity of about 4.7 GB on each side, normally stores anything

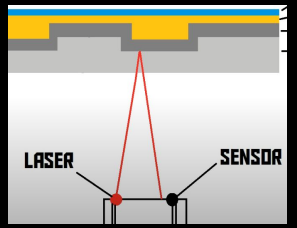
CDs and DVDs – Reading

Reading the bumps and divots

- On the read head of the player there is a laser and a sensor
- The laser sends a beam of light that is reflected off of the CD/DVD and onto the sensor.
- If the laser goes into the divot, the sensor will pick up the reflection and sense it
  - DVDs have much smaller grooves than CDs, which allow them to store much more data
- If the laser goes onto a bump, the reflection will miss the sensor





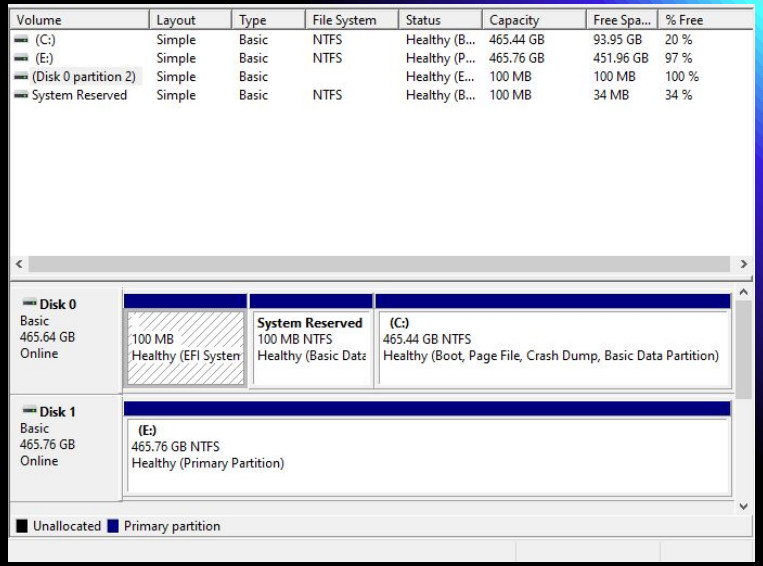


How is data stored logically?

File systems:
Windows: NTFS, FAT32, exFAT
Linux: ext4
Mac: APFS

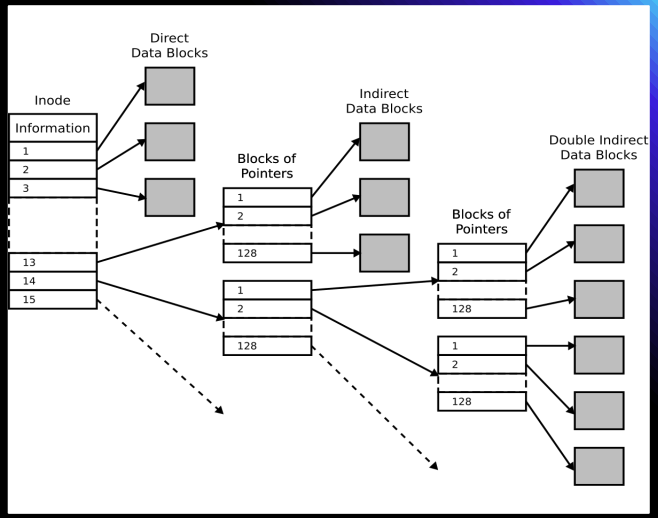
File systems – General Info

- Hard drives and SSDs do not come formatted, and need file systems to organize their data
  - They become formatted when installing an operating system, like windows.
- **Partitioning:**
  - Splitting one drive into multiple logical drives. Each logical drive is a partition.
  - ex) Your PC might have two, C: and D:
- Each partition in a SSD and HDD is formatted with a file system



File systems – General Info

- Remember what we learnt in CSC369!
  - Inodes will store the metadata to files and directories
  - Inodes will point to the Blocks which will store the data itself
  - These blocks are allocated and unallocated constantly



Important Note:
When a data partition is unallocated, it is not cleared.

Windows – FAT32

File Allocation Table 32

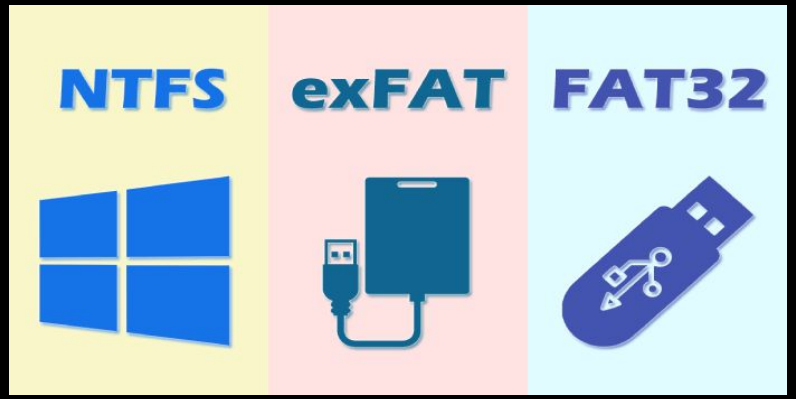
- Introduced with windows 95 (1996)
- Used in most removable drives at the time

- Advantages:
  - Most compatible file system, compatible with anything that has a usb port
- Disadvantages:
  - File size limit is 4GB
  - Max partition size of 8TB
    - Is you have a 16 TB drive, you will have to divide the storage space in half and make 2 partitions of 8TB each

Windows – exFAT

Extended File Allocation Table

- Latest microsoft file system(2006), and is an improvement of FAT32
- Designed to be used on flash drives and external storage devices



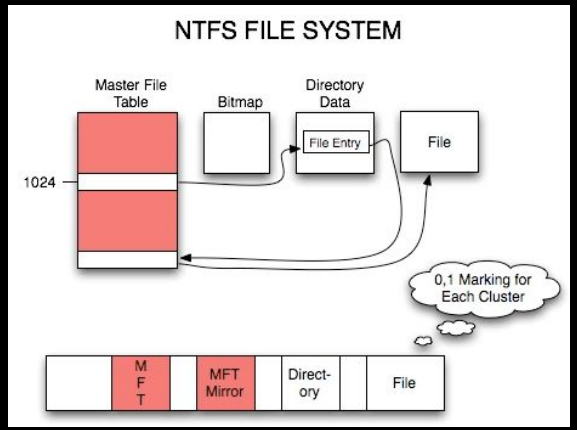
- Advantages:
  - Limitless when it comes to file and partition size limits
- Disadvantages:
  - Not as compatible as FAT32
  - Some older devices and versions of linux do not support it

Windows – NTFS

New Technology File System

- Default Windows file system, started being used in windows XP(2001)

- No file and partition size limits



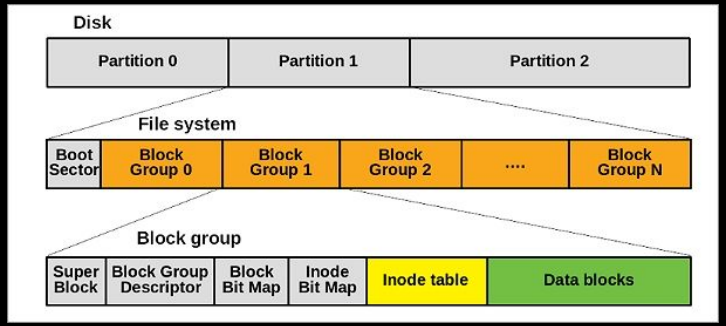
- Advantages:
  - Security permissions
  - Journaling
  - Hard and symbolic links(mutli access to the same file)
  - Encryption

- Disadvantages:
  - Lacks compatibility with some linux and mac devices
    - Everything is read-only

Linux – EXT4

Fourth Extended File System

- This is the format we saw in 369
- Found in Ubuntu, RHEL and other unix-like systems



- Advantages:
  - AES-256 encryption
  - Hard and symbolic links
  - Journaling
  - compression
- Disadvantages:
  - Has a limited max file and partition size(16TB)

Apple – APFS

Apple File System

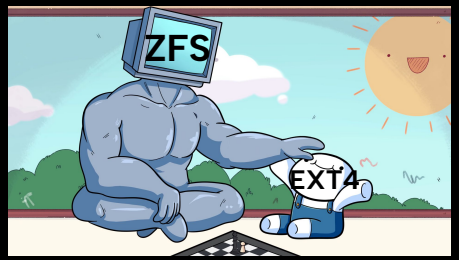
- Default file system for macs
- Limitless file sizes

- Advantages:
  - Strong encryption
  - Space sharing
  - Snapshots
  - Fast directory sizing
  - Supports traditional unix permissions
- Disadvantages:
  - Limited compatibility
    - For old version of macOS
    - For some third party apps

Sun Microsystems – ZFS

Zettabyte File Systems

- Found in Solaris, FreeBSD, Linux(third party implementations)
- It is a journaling file system and volume manager (2001)
- Is very extensive in its features compared to EXT4 and other file systems



- Advantages:
  - Designed for improved security, reliability and performance
  - Is a 128-bit file system, with virtually unlimited capacity
  - Is self-healing(self-corrects data) and has snapshot and cloning capabilities
- Disadvantages:
  - Memory intensive, degrades at higher capacities
  - Complex for less experienced users

Data Loss

How is data lost?

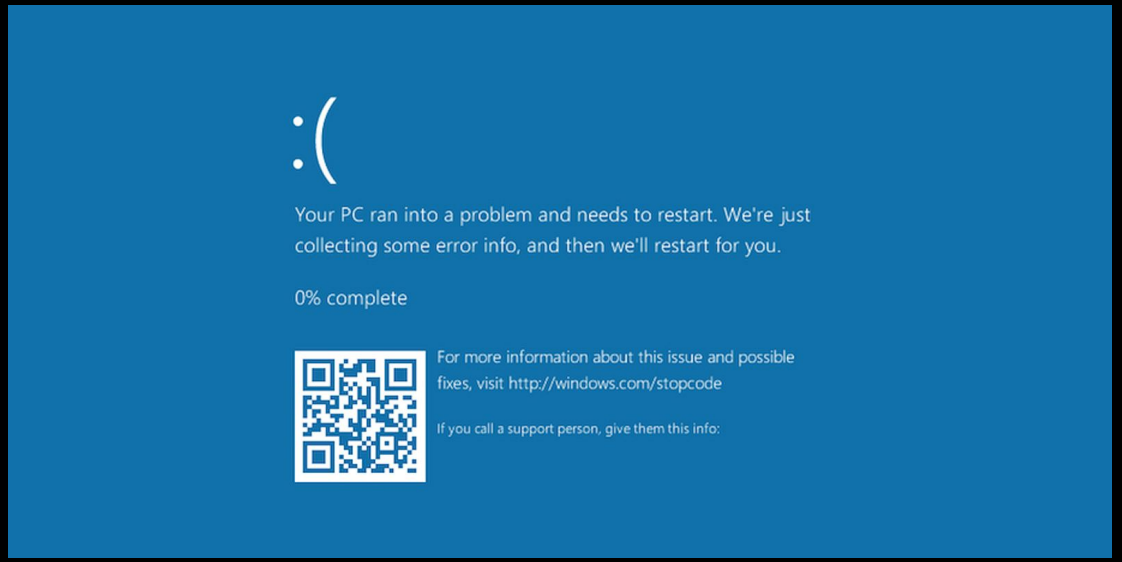
Concepts

Data Loss

What are the primary causes of data loss, and how does it happen?

Logical Data Loss

- Data loss that occurs via software



Logical Data Loss

- Human Error
  - Accidental or unintended file deletion (permanent file deletion vs soft delete Recycling Bins and Trash)
  - Incorrectly formatting or partitioning drives

Logical Data Loss

- File System Corruption
  - Could be caused by power failure, system crashes
- Lossy Compression
  - In an effort to save space on larger files, some compression algorithms attempt to erase irrelevant data on files...

Logical Data Loss

- Malware
  - Unwanted data encryption (i.e. ransomware)
  - Forced data deletion (wipers)

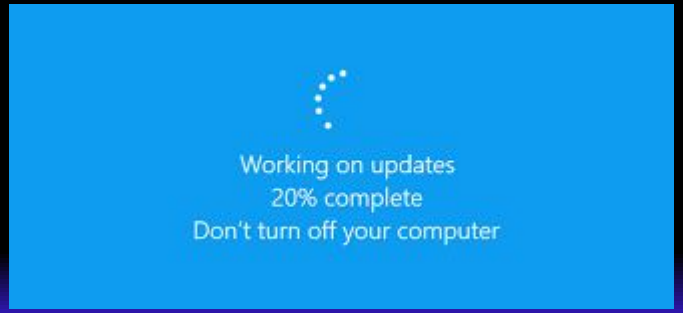
Notable wiper incidents

Narilam	Shamoon	Groove Monitor	Dark Seoul (Lazarus)	Sony (Lazarus)	BlackEnergy	Shamoon 2.0	StoneDrill
2008 - 2012	2012	2012	2013	2014	2015 - 2016	2016 - 2017	2016 - 2017
Middle East	Middle East	Middle East	South Korea	USA	Ukraine	Middle East	Middle East
Self spreading, affects database software through slow corruptions	Targets in the energy sector, oil, critical infrastructure	Time bomb (predefined dates), deletes files on all disks	Targets in financial and media sectors	Similarities with Shamoon. (EldoS raw disk driver)	Targets in the energy sector, media, transportation, government	Malware repackaged from first wave in 2012. Most victims in Saudi Arabia, with top targets including government, industry, transport, telecoms	Style similarities with Shamoon 2.0. Heavy use of evasion techniques to avoid detection by sandboxes

GREAT KASPERSKY lab

Logical Data Loss

- Software bugs
  - Operating system bugs can have a large impact as they deal directly with the file system
  - E.g. Windows 10 October 2018:
    - An update for windows 10 resulted in the deletion of user files who had the “Known Folder Redirection” feature enabled.
    - This deleted files in the Documents, Pictures, and Desktop folders.



Aside: Deliberate Data Loss

- In some instances, data loss can occur within files deliberately
- Compression algorithms are used to compress data into smaller segments, in order to save space
    - These algorithms can preserve all data during compression and decompression (lossless), or lose data in the process of compression (lossy)
- Lossy algorithms exist in order to improve data compression to substantially shrink files. This typically happens when processing images, in order to eliminate redundant pixels

Figure 1:

Figure 2:

Notice a difference between the two?

Physical Data Loss

Data loss resulting from storage devices

Physical Data Loss

- Hardware Failure
  - Depending on the data storage medium, hardware can physically fail which results in stored data being lost.
    - Unreliable power sources may cause data to be lost during writing periods
    - Improper cooling can lead to overheating
    - Other dependent hardware fails, causing the storage devices to fail
    - Interruption during firmware upgrades

DVDs

- A Digital Video Disc (DVD) is a form of digital storage that can be read from using an optical drive
  - Can be read only, read-write, single use, or multi use, depending on the layers on the DVD
- Data is stored on the device through a precise laser burning patterns
- Since data is read through an optical laser, smudges, fingerprints, scratches, or any external obstruction could result in playback issues and or difficulty accessing data burned on the disc.

SSDs

- A solid state drive (SSD) is a data storage device that uses flash based memory to store data.
  - Some varying differences between SSDs and other storage devices are that it has no moving parts, are significantly faster to read from, and can hold data without power

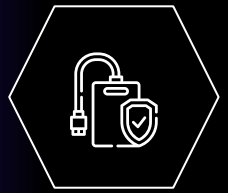
- SSDs typically have a shorter lifespan and become unreliable after time as the transistors used to store data wear out and lose their charging capacity.

- SSDs can further lose data by lack of use. NAND flash memory is typically used in modern SSDs that need a low charge to function consistently. If the device is starved of power for an extended period of time (> 2 years), it may lose data.

Data Recovery

How is data recovered?

Concepts



Physical Data Recovery

How is data recovered from physical data loss issues

- Hardware Repair
- Cool story

Logical Data Recovery

How is data stored virtual data loss issues

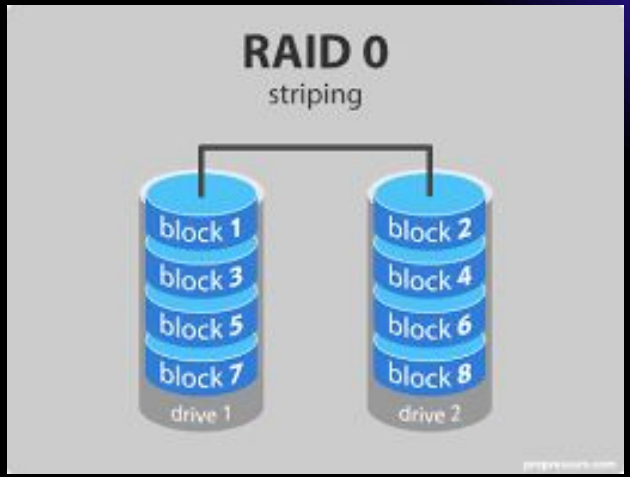
- Disk Drill

Physical Data Recovery

- Hardware repair
  - Replacing the physical components on a device to make it function properly (e.g. hard drives)

- Cleaning the device to allow other devices to read the data
  - Removing any dust / debris
  - For CDs, wiping the disc or polishing off any scratches may allow the optical reader to access any data

- Restoration via file systems
  - Redundant Array of Independent Disks (RAID)
    - Backup in case of drive failure



Physical Data Recovery – Cool story

- Damaged unreadable CD-ROM(read only memory) fixed by piece of tape

Broken

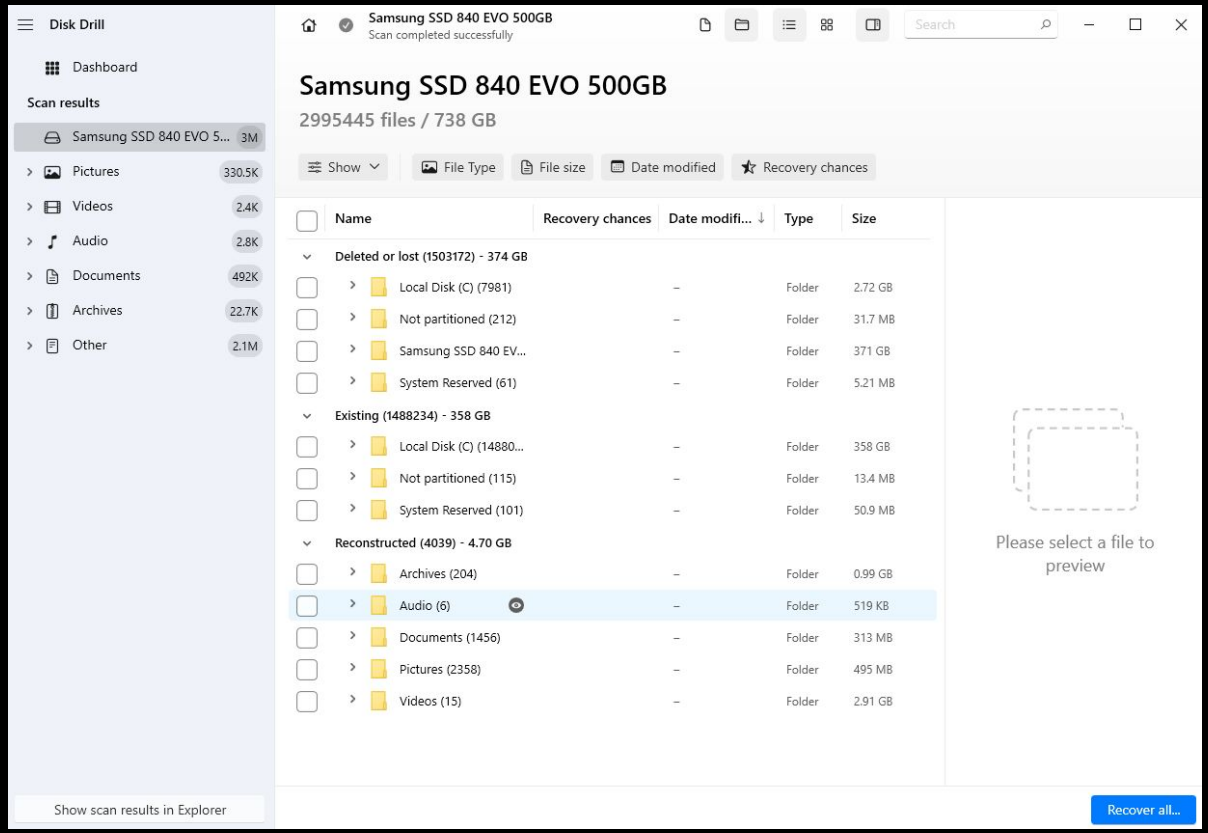
Fixed

- There are many recovery tools for CDs and DVDs, like minitool-power-data-recovery, which is similar to Disk Drill

Disk Drill

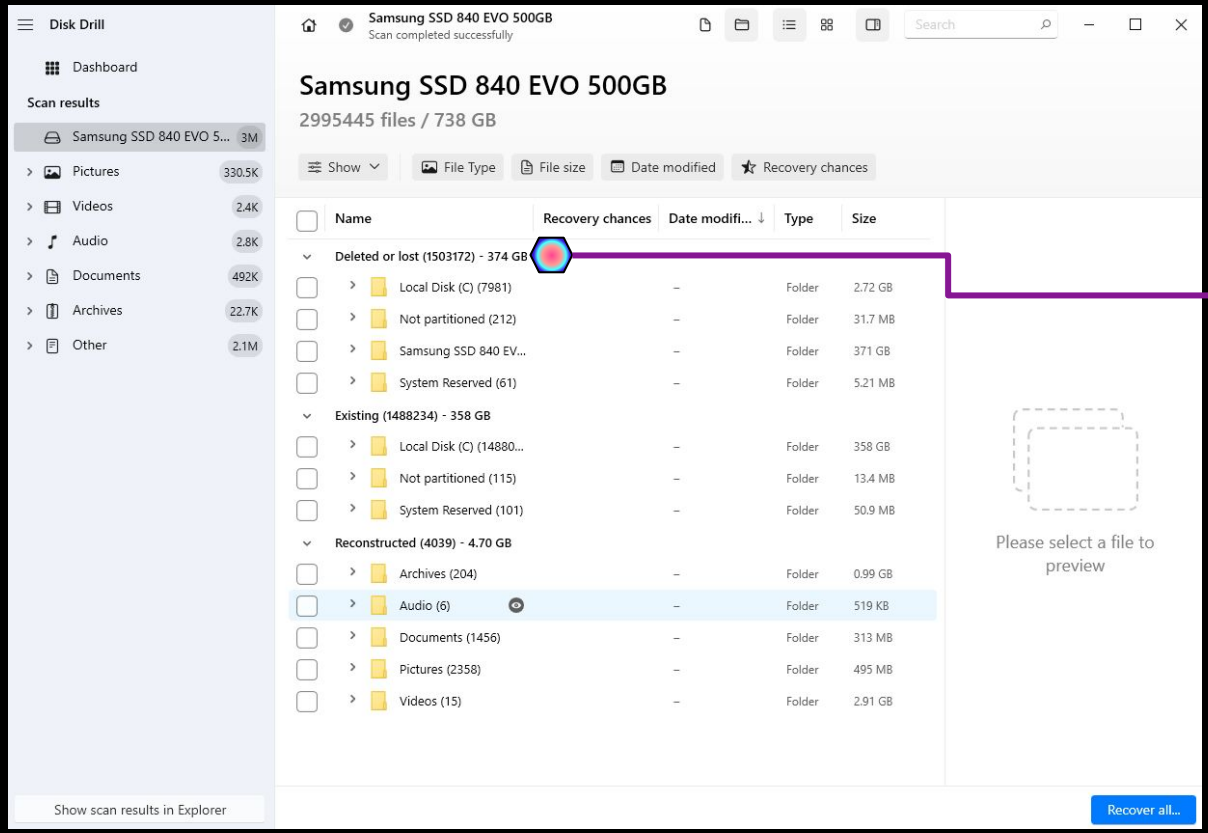
- Disk drill is a data recovery tool used to recover from data loss scenarios like:
  - Accidental deletion
  - Emptying recycling bin
  - Corrupted/damaged file systems
  - Partial file corruption
  - Ect.
- It can also reconstruct damaged file systems.
- It supports NTFS, FAT, exFAT, HFS+, APFS and more.

Disk Drill



Has a very user-friendly feel

Disk Drill

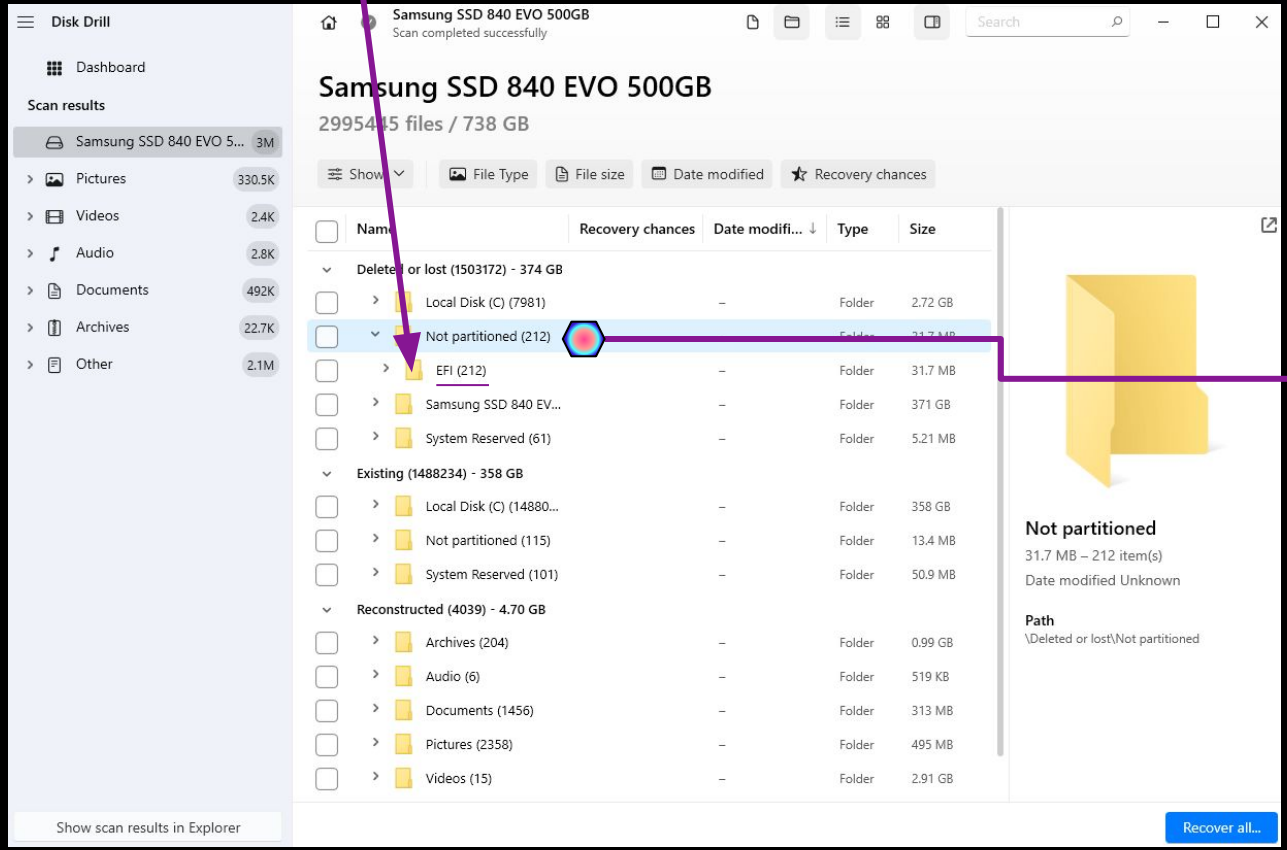


Files that have been deleted or lost.

Disk drill scans the storage device for file signatures metadata and other traces of deleted files

Notice this!

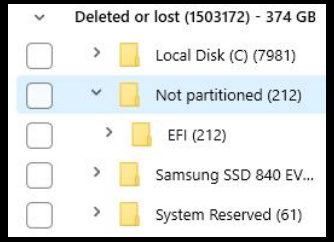
Disk Drill

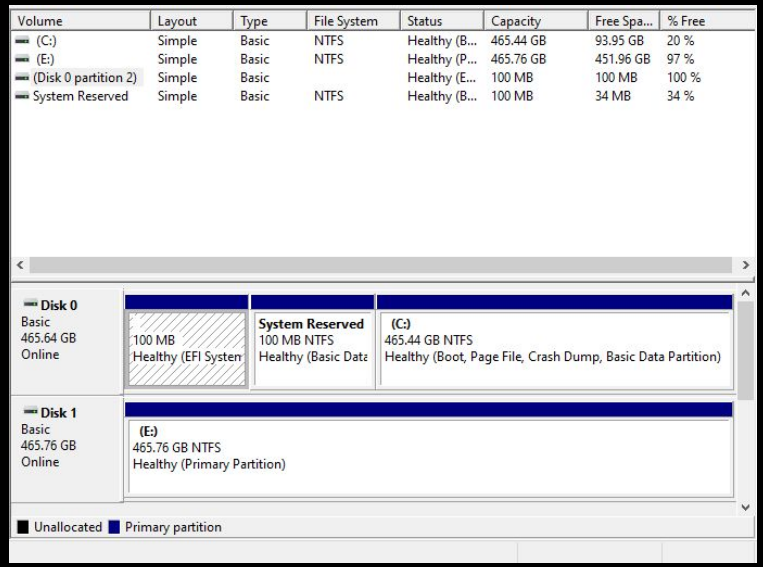


“Not partitioned” refers to raw, unused space on the drive.

This implies that Disk Drill checks every bit in the drive

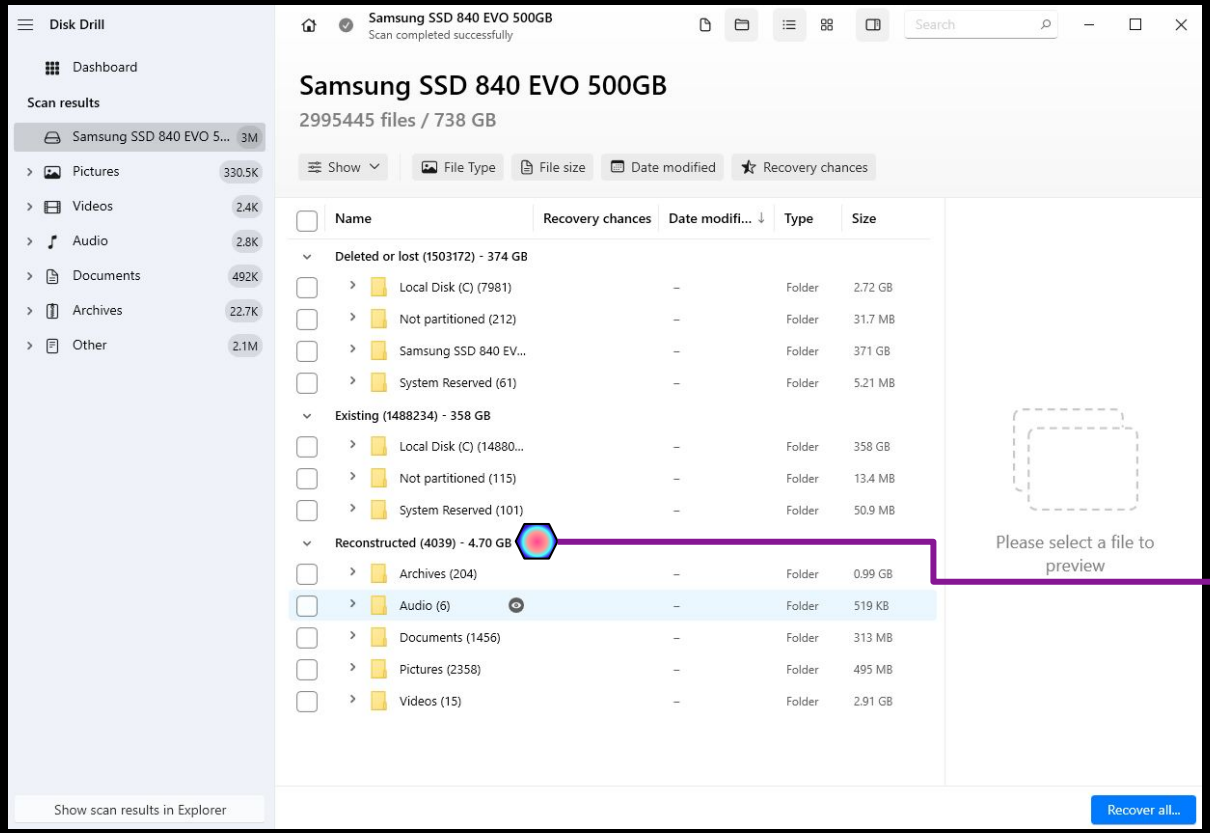
Disk Drill





- Notice that Disk Drill splits the data into its partitions
- In this case, the Not partitioned section of the Disk Drill is referring to the EFI System Partition(ESP)
- The ESP is used to store files necessary for system startup and boot management

Disk Drill



Files that disk drill has reconstructed from fragments and remnants of data found on the drives.

They might have been partially overwritten, damaged or had their metadata lost

Disk Drill – Reconstruction

How do they do it?!

1. Disk Drill scans the entire storage device bypassing the file system.(Deep scan)
   a. It will locate blocks specific to some files based on the file system's data allocation information

2. It identifies file signatures or headers. These are unique patterns of bytes at the beginning of specific file types.
   a. Different file types have different signatures:

89 50 4E 47 0D 0A 1A 0A	‰PNG CR LF SUB LF	0	png	Image encoded in the Portable Network Graphics format[21]
0E FE FF	so þÿ	0	txt *others*	SCSU byte order mark for text[26][25]
66 74 79 70 69 73 6F 6D	ftypisom	4	mp4	ISO Base Media file (MPEG-4)

Remember this

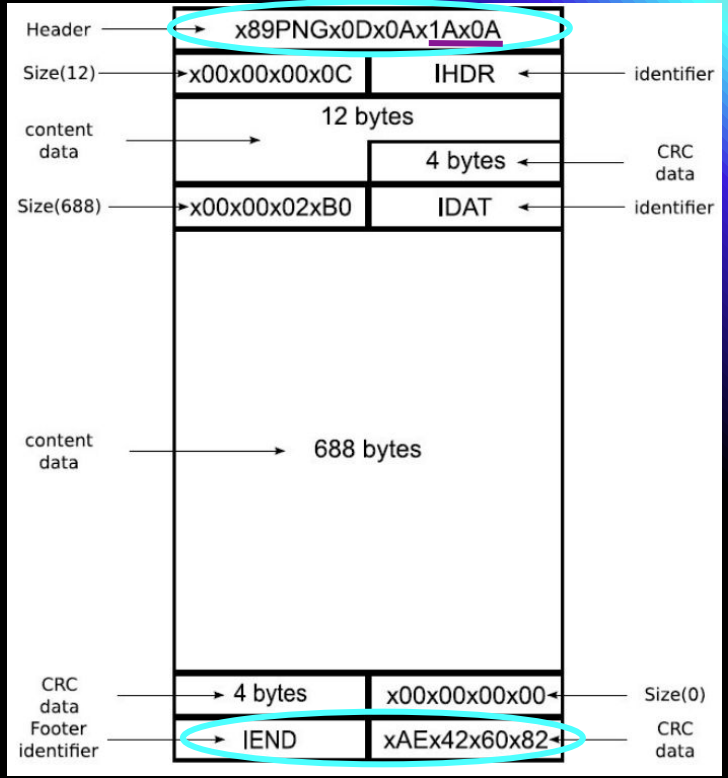
Disk Drill – Reconstruction

How do they do it?!

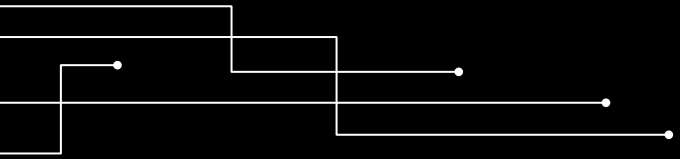
3. **Header and footer matching:**
   a. Disk Drill looks for both the beginning and end of files by matching signatures or patterns that indicate the file's structure

4. Fragmented file reconstruction:
   a. If a file is fragmented, Disk Drill tries to piece the segments together in the correct order(algorithm for this is a secret)

5. Data verification and integrity checks:
   a. Verifies checksums to ensure integrity of reconstructed data(other applications do this, not sure if disk drill does too.)



Demo



Notes

- It is cool to see that when you are in a virtual machine, Disk Drill is only able to access the chunk you allocated to this virtual machine, and is indeed completely isolated.

- We chose Disk Drill because it is the most reputable recovery tool, there are others like R-photo and DMDE, but we didn't want to risk our drives being compromised, so we stuck with Disk Drill.