Back it up NOW!

Data management for everyone

mel@melsinclair.com.au

With MEL SINCLAIR + SAM BUNN

This was originally a presentation

There are several parts to this presentation that, in printed form, may not make a lot of sense. This is because they were spoken and explained and a discussion was had.

It is really important that if you do not know what you are doing with backups and more complicated IT stuff, that you seek out professional assistance. Sometimes attempting stuff ourselves means we ruin the chance of a pro to fix it.

If you have any questions about this presentation or have found this through the QCG website, please contact me; mel@melsinclair.com.au

September 2023

As Photographers,
We are heavy data creators, distributors and hoarders.

A backup is our insurance policy, you hope you never
really need to use it,
but when you do, you'll be glad you have it.

Our intention tonight isn't to make you feel guilty if you
aren't backing up, but to make it known how easy it is to
start, with something, anything (which is better than
nothing)!

It's up to you to find that balance between what you can
afford to pay, to protect your important data, versus
what you are able to risk losing with minimal heartache.

WE ARE THE CONTENT WIZARDS!

- IBM estimates that 90% of the world's data has been created in the last 5 years alone. Think about that.

- This means that each day, the world is generating 2.5 quintillion+ bytes (18 zeros!). To put that in units we understand, it's 2.5 MILLON TERRABYTES. Per. Day.

About Mel + Sam

Professional nerds keen on making complicated IT stuffs, simpler.

Mel: -
Data creation wizard (aka Photographer)
Systems Administrator – Qld Government (15+ years)

Sam:-
PC fixing Wizard
Systems Administrator – Qld Government (14+ years)
Dell Certified Warranty Technician
Personal/Business PC Repair + Diagnostics 20+years

Have you already got a backup system?

If so:

- Do you count yourself as a beginner, intermediate, or expert?

- Do you know what it is?

  - Can you give us a brief statement of what your methods are?
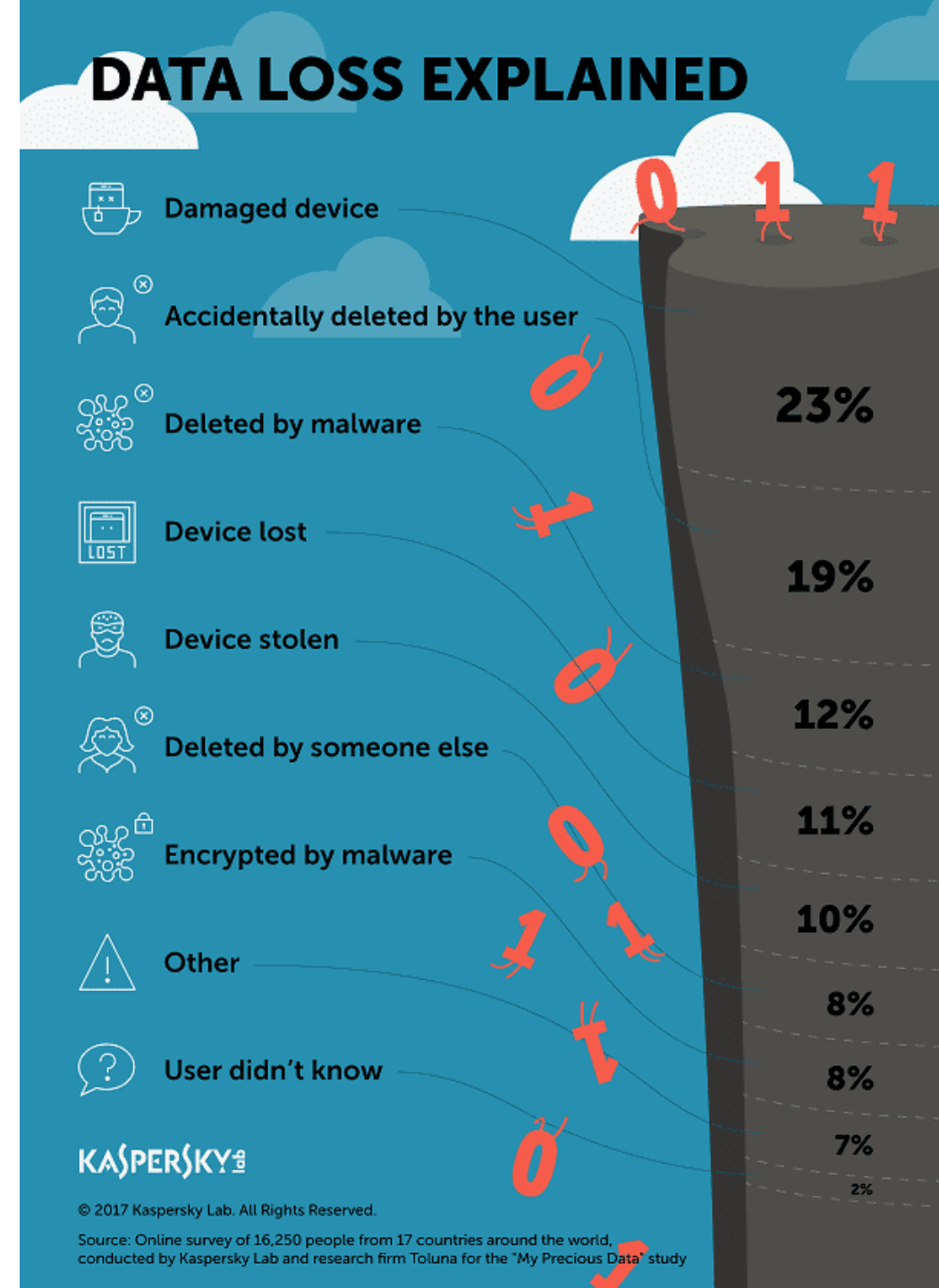   Backup locations, methods, how often?

THE TIMES YOU ARE GIVEN FRIENDLY ADVICE TO BACKUP YOUR COMPUTER

When your computer has just broken down

Some other random time
such as right now

Disaster can strike in many ways;

- Natural Disaster (Fire, Flood)
- Storm Damage (Electrical shock, Lightning)
- Overzealous cleanup of disks (whoopsies)
- Scammers erasing, locking or releasing a virus (everything you can see, they can see)
- Malware and Ransomware
- Theft / Loss (leaving your device on the train)
- Random drive death (happens-)
- All other instances of bad luck

DATA LOSS EXPLAINED

Damaged device		
Accidentally deleted by the user		
Deleted by malware	23%	
Device lost		
Device stolen	19%	
Deleted by someone else	12%	
Encrypted by malware	11%	
Other	10%	
User didn't know	8%	
	8%	
	7%	
	2%	

KASPERSKY

© 2017 Kaspersky Lab. All Rights Reserved.

Source: Online survey of 16,250 people from 17 countries around the world, conducted by Kaspersky Lab and research firm Toluna for the "My Precious Data" study

HARD DRIVES: HARD FACTS

- On average, each hard drive lives for only <u>3 years</u>. Anything longer than that is borrowed time or sheer luck.

- This rule applies whether you have a Mac or a Windows machine, Desktop or Laptop (or a mix)

- This applies irrespective of hard drive type, whether SATA/SAS, SSD, NVMe or whether you use server-grade storage drives.

- Best practice is to always have a replacement schedule. Plan to replace aging hardware and don't wing-it.

- Use label-makers to name the volume (disk name) and purchase date on the drive.

- Record this in a spreadsheet and save in safe places!

TRIPL3 REDUNDANCY

- If your data does not exist in at least 2 separate places (3 is ideal) then it does not exist.

- So what is ideal for triple redundancy?

- 1. **HOT** local copy – working or always required

- 2. **WARM** copy – NAS or DAS (on all the time)

- 3. **COLD -Physical** offline offsite storage, NAS or DAS (only on when needed)

- 4. **COLD online ARCHIVE** – cloud storage (icloud/ google drive / backblaze)

- Warning: Every method has its flaws, that's why we aim for 3, because the chance of all failing at once is slim.

Types of data for each location (serving suggestion)

- **Hot** storage: Working copy, currently using – recent work etc, photoshop files, LR Catalogues

- **Warm** storage: working with and referencing but not in constant need, may be network attached or only on when you need it.

- **Cold storage**: not accessed constantly but there when needed. Think older files that you don't always need but need to keep.

- **Cold archive**: only when needed. Off all other times. Anything not immediately needed, a day to a week to access won't hurt.

CLOUD

NAS

DAS

HOT COPY
(working)

EXTERNAL
HARD
DRIVE
(offline except
when needed)

OFFSITE HARD
DRIVE
(always off
unless recalled
for backup
retrieval)

One size does not fit all…

Whatever you choose should:
- Be easily usable
- Be easily understandable to you
- Be easily scalable (ie to upgrade)
- Fit within your budget
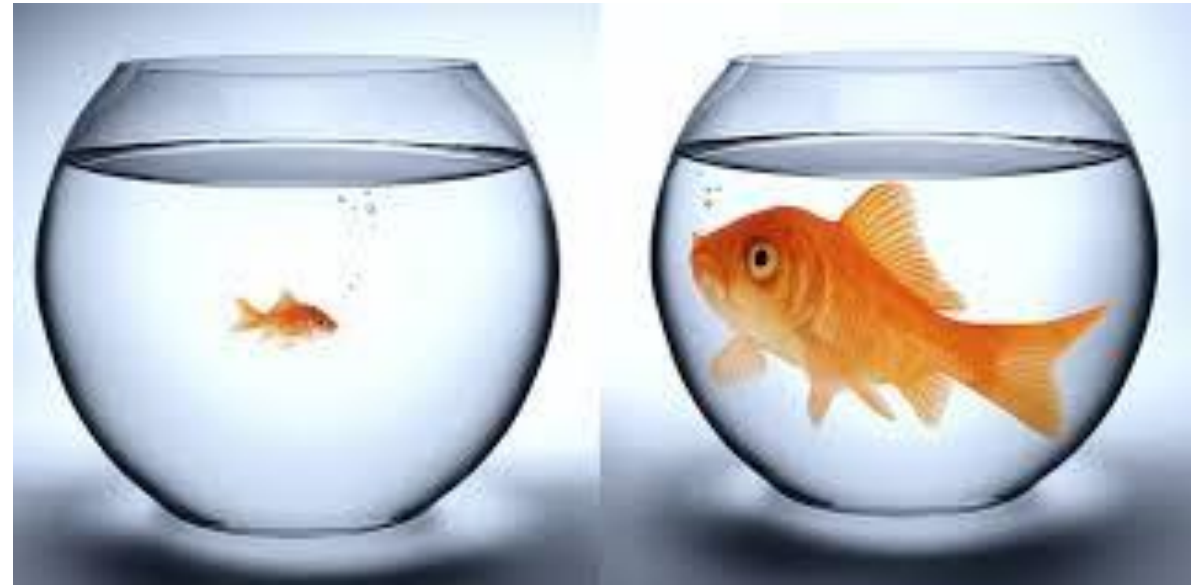
Type 1: JBOD – Just a bunch of disks…

- How it works (connection): USB 3.0+ or Thunderbolt
- Size: 1-5TB
- Why you'd use it: Simple to use, suitable for both PC and MAC, good for transporting large amounts of data. Readily available at JB HiFi and Officeworks.
- Cost: Usually Cheap but price will fluctuate with size
- Pros: Lightweight, takes up less space.
- Cons: The larger the disk, the longer the recovery time. High amount of organization required to keep updated. Can be cheap construction, can be proprietary connection inside.

Type 2: Cloud

- How it works (connection): Internet share
- Size: Limited by cost
- Why you'd use it: Offsite backup. Generally set and forget as long as you pay the subscription fee.
- Pros: Offsite backup, ease of use. Similar to iCloud, Google Drive, Dropbox etc similar interfaces used. For an extra cost, you can request them to send you a HDD to upload to, and post back, to save your internet connection.
- Cons: Can be slow to retrieve/download from. HDD option above goes the opposite way as well. Australian internet speeds suck unless you have FTTP NBN.

Type 3: DAS – Direct Attached Storage

- How it works (connection): Typically these are NAS boxes that connect straight to the PC/MAC. DAS is traditionally an Enterprise (Business) technology.

- Size: Small or large depending what you choose to put into it.

- Why you'd use it: If you had to constantly access a large amount of images Pros: Can be set and forget. Generally easy to setup. Typically low power use.

- Cons: Relying on a unique chassis/ brand specifics may lock you into a system that is not desirable long term. Monitoring of failed/failing drives may not be easy to do. May require additional licensing.

  - Cooling – Units may get hot in Australian Summer, some do not have much ventilation.

Type 3: NAS – Network Attached Storage

- How it works (connection): Connected to your home network, Ethernet or WiFi

- Size: Chassis is a base price, drives extra (size dep on cost)

- Why you'd use it: Fast to connect to, reliable, common connection protocol.

- Pros: Standardized, good for a Warm storage option.

- Cons: Some brands have extra cost licenses, proprietary software, extensive IT knowledge required to address issues.
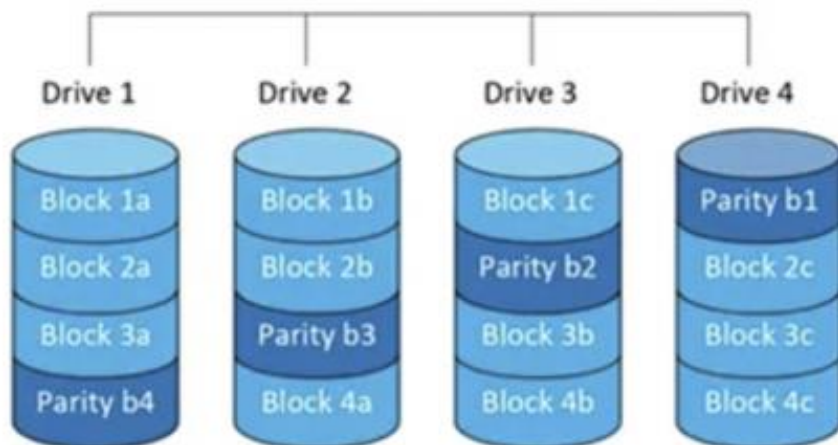
RAID 0
Striping

Drive 1 | Drive 2

Block 1	Block 2
Block 3	Block 4
Block 5	Block 6
Block 7	Block 8

RAID 1
Mirroring

Drive 1 | Drive 2

Block 1	Block 1
Block 2	=	Block 2
Block 3	Block 3
Block 4	Block 4
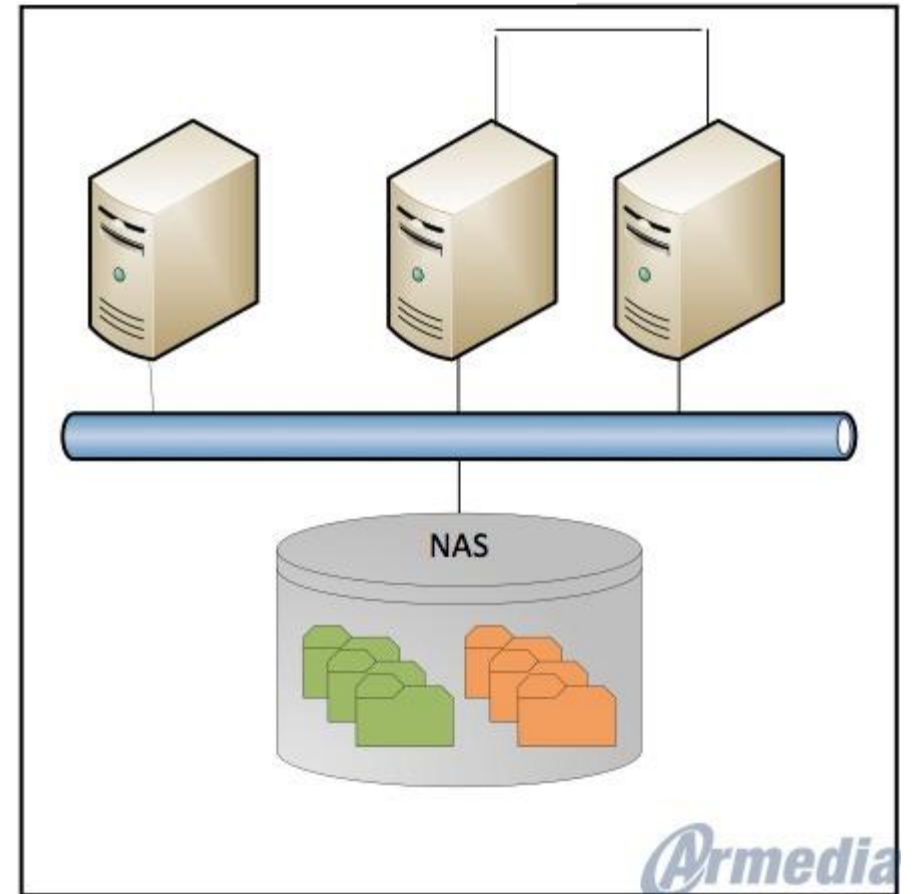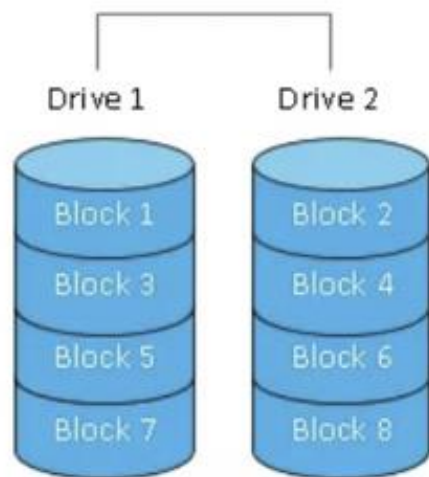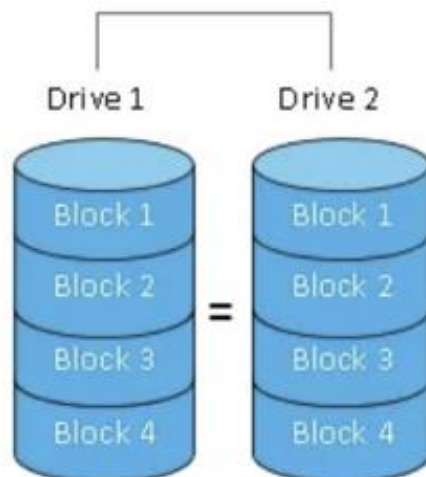
RAID 10
Mirroring + Striping

Drive 1 | Drive 2 | Drive 3 | Drive 4

Block 1	Block 1	Block 2	Block 2
Block 3	Block 3	Block 4	Block 4
Block 5	Block 5	Block 6	Block 6
Block 7	Block 7	Block 8	Block 8
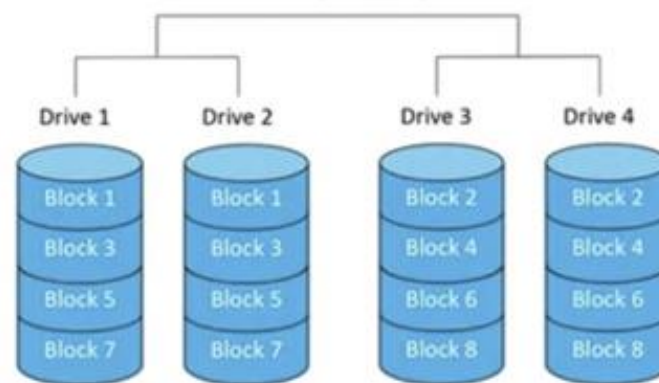
RAID 5
Striping with parity across drives

Drive 1 | Drive 2 | Drive 3 | Drive 4

Block 1a	Block 1b	Block 1c	Parity b1
Block 2a	Block 2b	Parity b2	Block 2c
Block 3a	Parity b3	Block 3b	Block 3c
Parity b4	Block 4a	Block 4b	Block 4c

RAID 6
Striping with dual parity across drives

Drive 1 | Drive 2 | Drive 3 | Drive 4

Block 1a	Block 1b	Parity b1	Parity b1*
Block 2a	Parity b2	Parity b2*	Block 2b
Parity b3	Parity b3*	Block 3a	Block 3b
Parity b4	Block 4a	Block 4b	Parity b4*
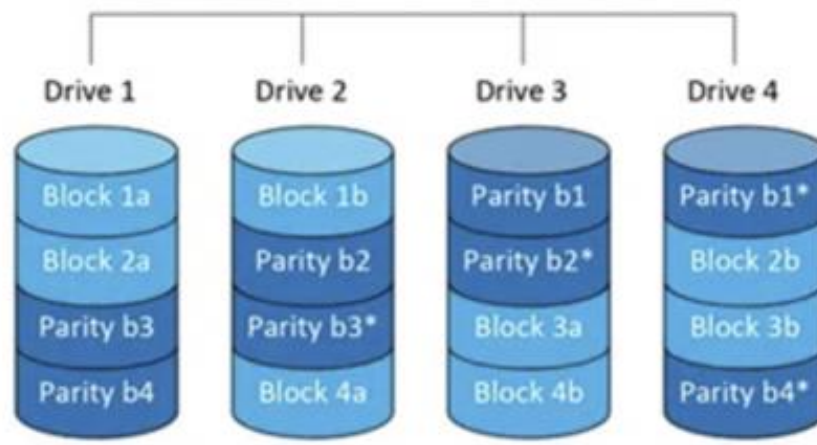
- Raid 0 = data striped across 2 or more disks, Lose one, lose them all. Extremely fast, with extreme risk. Do not use this as a backup method.
- Raid 1= Data mirrored across 2 disks. This means 2x read speed, 1x write speed. 1 drive failure tolerance, assuming the other disk is in good health also.
- Raid 5= Data split across 3 or more disks, 1 disk is a parity disks, this can be any of the disks as it is a 1-drive loss tolerance, however you are assuming that all the other disks are in good health also, when recovering. More disks = more risk in this scenario. Speed increases with more disks connected, but write is 1x speed.
- ADDRESSING HOT SPARES: people say they like to have a hot spare in their raid array, like it's a savior, but it is only a savour at a place in time. Only when a disk fails does it become active and is thrown into the array. For instance if R5, if a disk fails, it will automatically add the HS into the array and try to rebuild it to replace the disk that fauiled. However...depending on the health of the remaining disks, the array may fail if the disk fails during the rebuild process.
- All this applies to systems such as DROBO, SYNOLOGY, QNAP, + ENTERPRISE LEVEL STORAGE AND NAS.

- Raid 6 = 4 disks or more, like raid 5, but an extra disk for parity, More disks included, more speed for read, write is 1x speed.
- Raid 10 = A mixture of striped and mirrored disks, think Raid 0 but those two disks are mirrored. Has a 1.5 disk redundancy, if you lose 2 disks on one side of the mirror = dead array. Lose one disk on either side of the mirror, you can recover assuming the remaining disk are healthy.
- Raid 50 = Think Raid 10, but in a Raid 5 scenario, you need 6 disks or more as it is a Raid 5 but mirrored. Reduancy is ---
- RAID 60= Like raid 50 but with an extra redundant disk

Advanced: Understanding RAID

RAID is a technical term to explain a few different scenarios of data storage, using groups of disks to pool together.

RAID Level Comparison							
Features	**RAID 0**	**RAID 1**	**RAID 5**	**RAID 6**	**RAID 10**	**RAID 50**	**RAID 60**
Minimum number of drives	2	2	3	4	4	6	8
Fault tolerance	None	Single-drive failure	Single-drive failure	Two-drive failure	Up to one disk failure in each sub-array	Up to one disk failure in each sub-array	Up to two disk failure in each sub-array
Read performance	High	Medium	Medium	Medium	High	High	High
Write Performance	High	Medium	Low	Low	Medium	Medium	Medium
Capacity utilization	100%	50%	67% – 94%	50% – 88%	50%	67% - 94%	50% - 88%
Typical applications	High end workstations, data logging, real-time rendering, very transitory data	Operating systems, transaction databases	Data warehousing, web serving, archiving	Data archive, backup to disk, high availability solutions, servers with large capacity requirements	Fast databases, file servers, application servers	Large databases, file servers, application servers	Data archive, backup to disk, high availability solutions, servers with large capacity requirements

Buyer Beware: DROBO (Discontinued Jan 2023)

Drobo, while discontinued, was a bouquet of red flags. Do your research if you're going to buy a proprietary system

Examples of how you might build a solution… BUDGET / BEGINNER

JBOD ; Just a bunch of disks (Preferably 3 but understand if 1 or 2 is it)

- Hot: Internal disk in PC

- Warm/Cold: External Hard drives.
  Perhaps one on at a time, one off at a time.

- Pros: Cheap, readily available and convenient.

- Cons: While convenient, high overhead in data management and sync. Always a risk of drive or enclosure faults.

- TIP: If using these as a backup option, don't have them plugged in all the time!

Examples of how you might build a solution… INTERMEDIATE

A balanced Mix: Cloud and Ext HDD

- Hot: Internal disk in PC
- Warm: External Hard drives.
- Cold: Buy cloud storage to store critical files (Google Drive, Dropbox, iCloud, iDrive)

- Pros: A good mix of technologies, readily available.
- Cons: Will become a lot to manage
- TIP: You don't have to backup everything, just the stuff that'd hurt to lose.

Examples of how you might build a solution... ADVANCED

- **Hot:** Internal disk in PC
- **Warm**: Cloud storage to store files (Google Drive, Dropbox, iCloud, iDrive)
- **Cold:** NAS Box

- Pros: Cheap, readily available and convenient.
- Cons: While convenient, high overhead in data management and sync. Always a risk of drive or enclosure faults.
- TIP: You don't have to backup everything, just the stuff that'd hurt to lose.

Examples of how you might build a solution… EXPERT

- **Hot:** Internal disk in PC
- **Warm**: Home Server in Raid no lower than R6, monitoring enabled.
- **Cold:** Cloud storage to store files (Google Drive, Dropbox, iCloud, iDrive)
- **Cold Archive**: Offsite backup
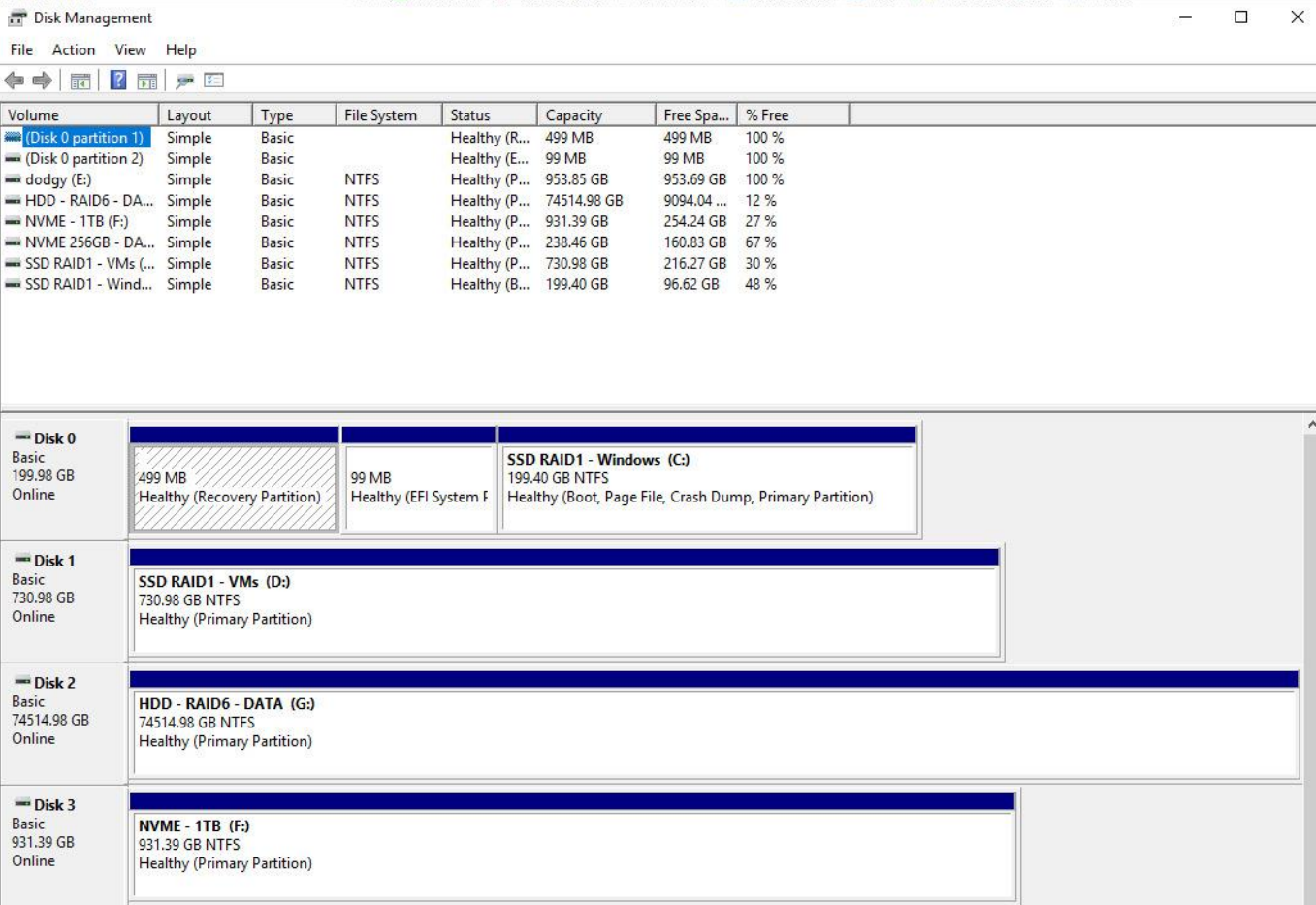
Pros: Security, multiple media types.

Cons: Wallet perpetually crying, IT knowledge is a must, or paying someone a nice wage to keep it running smoothly.

Our System (Mel + Sam)

- Home Server
- Large disk array dedicated to DFS / NAS system
- Mel backs up key business data to iDrive
- Not just backups, other home functions such as router, security
- You know how to manage such an environment and maintain it.

ON THE ROAD…

When you travel, which best describes your data habits?

- A: I back up my memory cards daily to two separate hard drives (or more)
- B: I back up my memory cards daily to my device (laptop, ipad, tablet etc)
- C: I backup as my memory card fills up
- D: I just hold onto my memory cards until I have time
- E: I don't do anything until I get home.

A and B are the ideal

Common Gotchas and Misconceptions

- Partitioning a drive (Splitting one physical into two virtual) is NOT a backup, if the hardware on the disk fails you will lose both drives.
- Disks will not last forever, and if they have, you are simply <u>lucky</u>.
- The storage realm is roughly set where we are now, disks will increase in space but the methods will not change much.
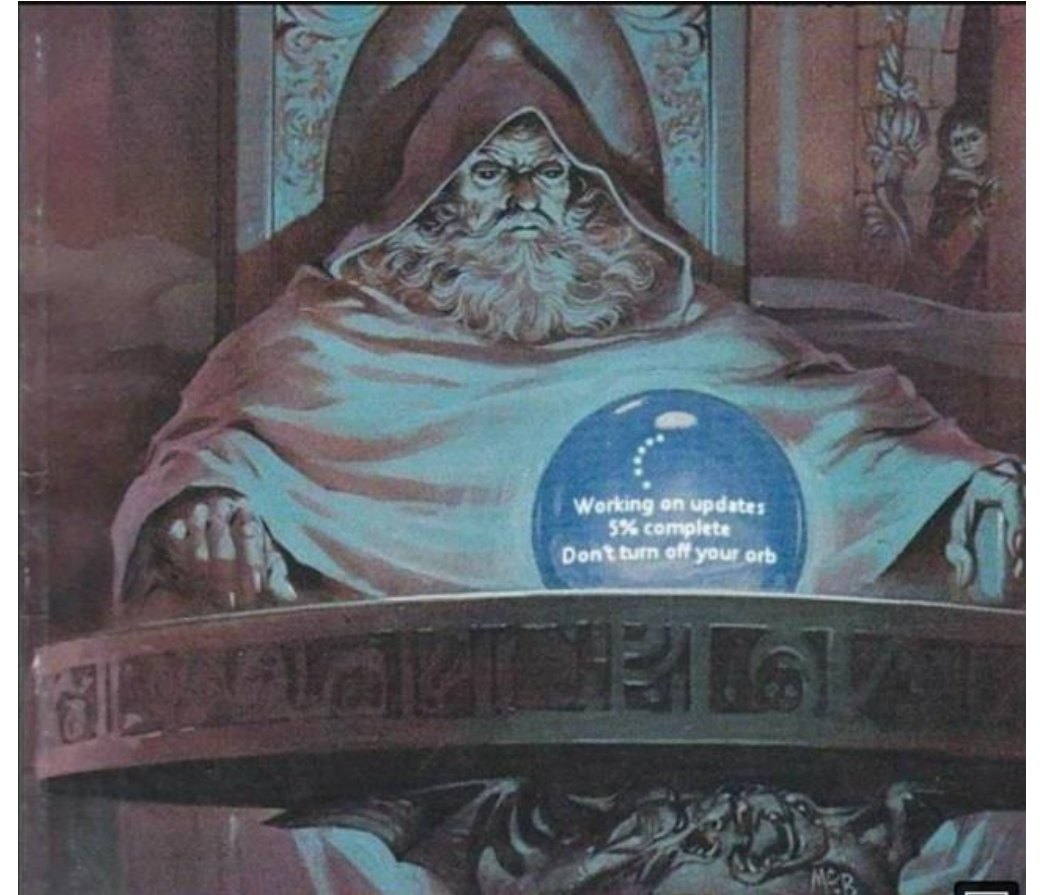
Handy programs

- **Beyond compare –** Beyond Compare is a software application to compare, merge, and synchronize data. It runs on Windows, macOS, and Linux. You can use it to compare your current drive to the backup, to see the differences and copy new files.

- **Treesize Free–** is a free disk space manager for Windows. The software shows you the sizes of folders including all subfolders. Handy for when you're trying to figure out what data is where.

Hard Disk Sentinel – is a multi-OS SSD and HDD monitoring and analysis software. Its goal is to find, test, diagnose and repair hard disk drive problems, report and display SSD and HDD health, performance degradations and failures.
Windows and Linux only

WHEN A DRIVE DIES… BY BEING BUMPED OR KNOCKED

- The hard drive may make a regular clicking noise when it is first powered on. This will last for approximately 20 seconds before the disk powers down.
- The hard drive may beep, or buzz when first powered on and will not spin up.
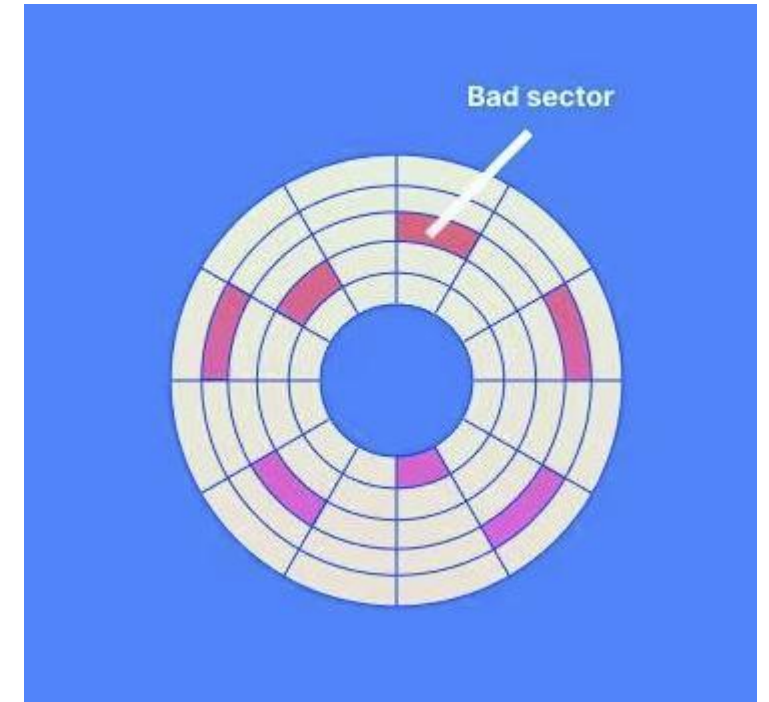- The drive will not be recognised by the computer.

Computer hard drive

spindle
disk
saved file
actuator
read/write head
arm
circuit board
power port
data cable port
drive configuration port

© 2010 Encyclopædia Britannica, Inc.
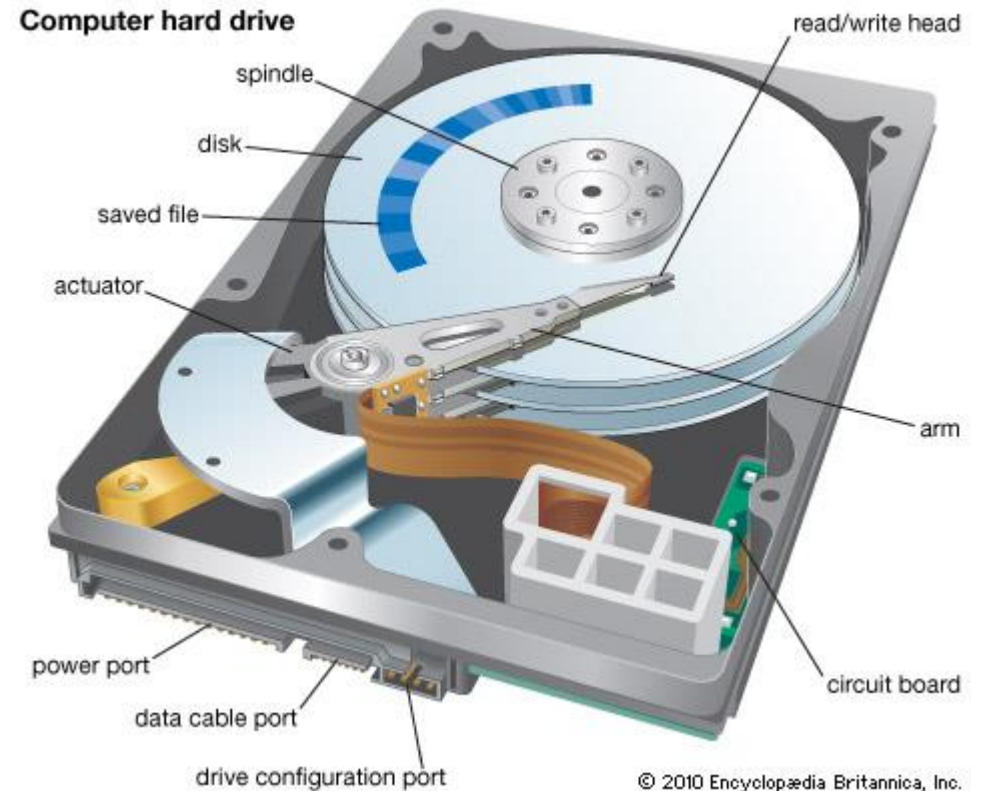
WHEN A DRIVE DIES…BECAUSE OF THE FILE SYSTEM

Failure Symptoms:

- Files and Folders will not open or have become inaccessible
- Files and Folders will not copy correctly – often CRC (Cyclical Redundancy Check) errors will be reported
- Files and or Folders may have been deleted or overwritten

- Areas of the hard drive may not be readable due to the development of bad sectors.
- Data structures on the hard drive may have been corrupted due to the disk not being ejected from the system correctly.

This happens most often on Memory Cards!

Bad sector

WHEN A DRIVE DIES…BECAUSE OF THE ELECTRONICS

Failure Symptoms:

- The drive will not spin when it is powered on
- Sometimes the drive will get hot / overheat
- Will not be recognized by the Computer
- Burn smell lingering from other components (power supply)

Causes:

- A power surge has caused damage to the drive
- The wrong power cable has been plugged in.

IF A HARD DRIVE EVER FAILS FOR YOU…

- If External: Immediately unplug the disk and do not turn it on again.
- If Internal: Turn computer off. If expert, isolate the disk or unmount (if not a C/System disk)
- Consult a data recovery expert and do not attempt to retrieve any files yourself – you may ruin the experts chance at getting your data back.
- Retrieval can cost anywhere from $300 to $1000 or MORE depending on the situation.

The (hard) truth

None of the methods detailed are 100% perfect, without flaws.

But the chances of all 3 failing at once are slim.

Insurance may cover the replacement of the disk, but the data is unable to be replaced.

RECOVERY COSTS AND CAVEATS

The Cost to recover is different, depending on how it failed, whether it's a…

- A single disk
- NAS or DAS chassis
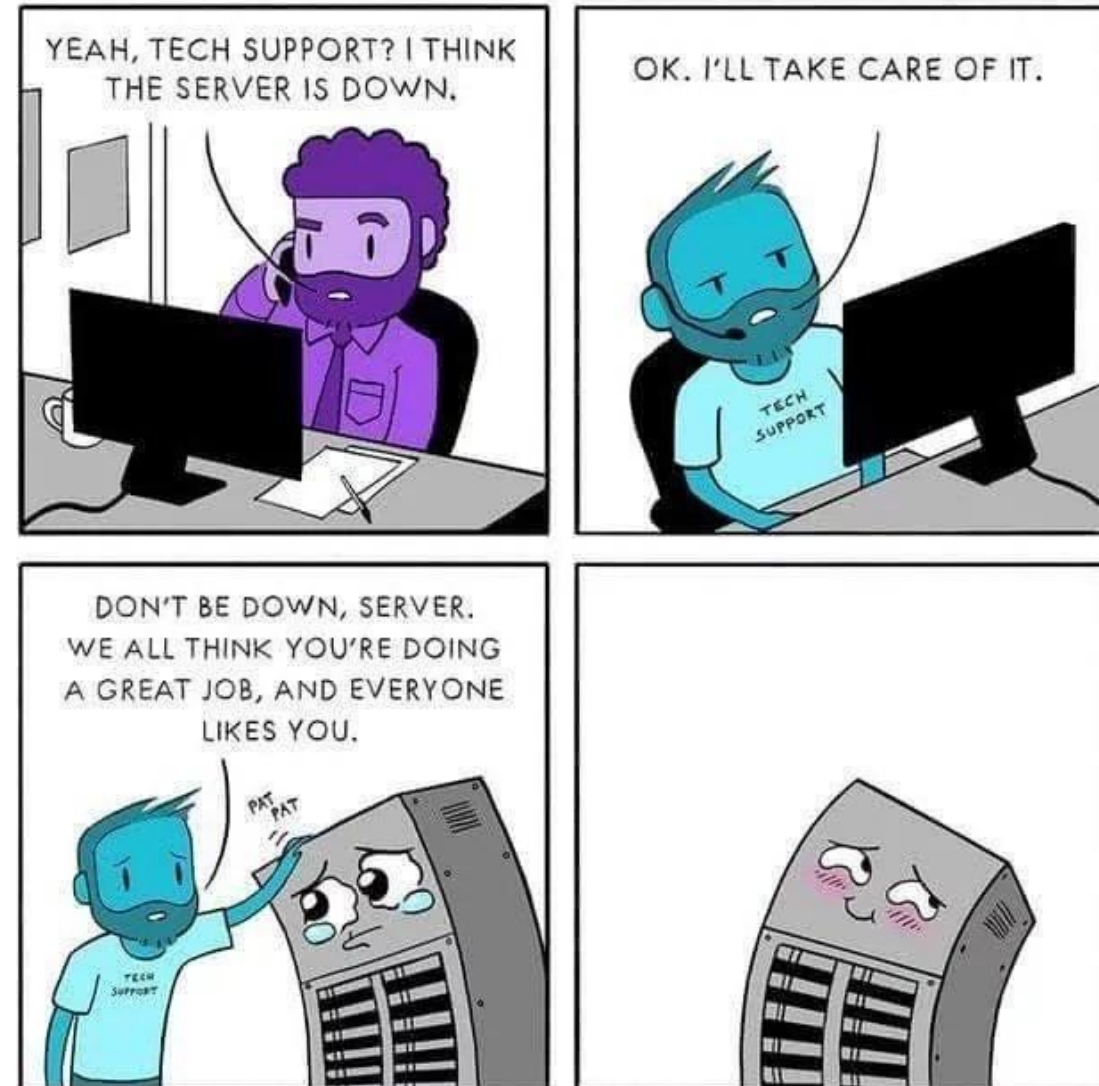- Disk arrays (RAID)

Caveats that make Data Recovery Expensive:

Type of disk

Size of disk

Type of failure

Where it is recovering to and how fast it can copy over

TAKEAWAYS: LESSONS YOU CAN LEARN

- Buy reputable hard drives: I prefer Western Digital Reds, Blacks and Blues OR for SSD: Samsung 850 EVO

- Invest in a NAS storage solution or cloud hosting plan, or both (Backblaze business is only $10 per month, per device)

- In the event of a drive failure, do not attempt to restore the data yourself. Isolate the disk or turn the computer off. Contact a trained professional.

- Disk / Data recovery is hellishly expensive and is a slow process.

- Have insurance for all your equipment. Data cannot be replaced if it is lost, stolen, damaged or destroyed.
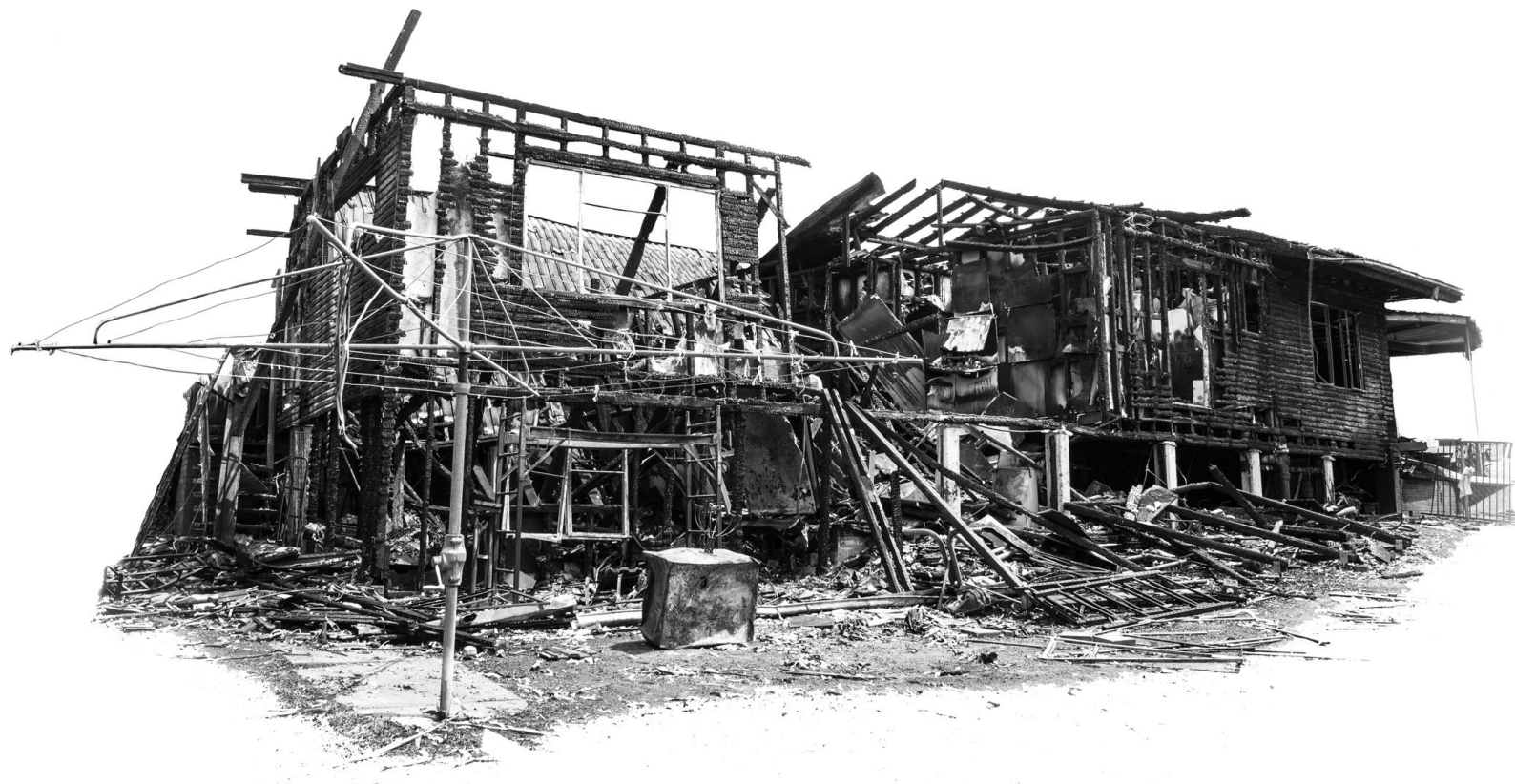
CASE STUDY: MEL (YES, THAT'S ME)

Warning: potentially distressing images ahead

In late 2017, two weeks before I was due to go on a trip of a lifetime to Antarctica.

my house burnt down.

Everything I owned, knew, loved and had created, was inside.

I WAS <u>EXTREMELY</u> LUCKY

- To ensure the data had the best chance of survival:

1. We immediately removed and discarded any external housing

2. We immediately dried in sun OR put in a bed of rice or silica and sealed container.

3. Waited until it was completely dry before turning on. Often you only get one chance.

4. Through patience and knowledge of what we were doing we retrieved all of my data. Bar one victim.

5. This was 12.71TB

DO IT RIGHT, PAY ONCE. DO IT WRONG, PAY TWICE

- REMEMBER: If disaster strikes and you do not have a backup, you will spend twice as much money.

- You will pay for the data recovery

- You will then pay for the new backup solution.

- So better just to go out and buy the backup solution now, before it's too late, eh?