SATA

DAS

SSD

RAID

LTO

SAS

rsync

MD5

SAN

SCSI

HDD

JBOD

USB

NAS

SHA

LTFS

dd

Do You Have a DSS*?

*Digital Storage Strategy - An Introduction

Risk Factors and Causes of Digital Loss

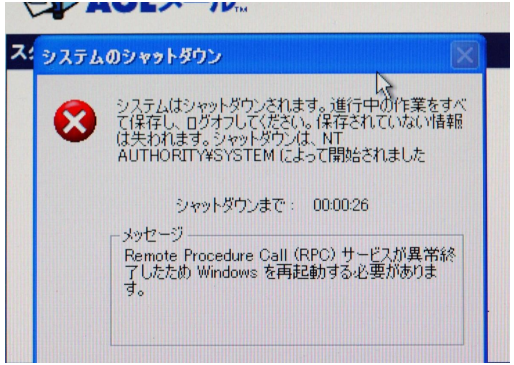
What are we scared of?

- Hardware failure
- Software failure/threats
- Degradation and bit rot
- Natural disasters
- Human error

Hardware Failure

- Mechanical failure
- Electrical failure
- Obsolescence (physical interfaces, peripherals, etc.)

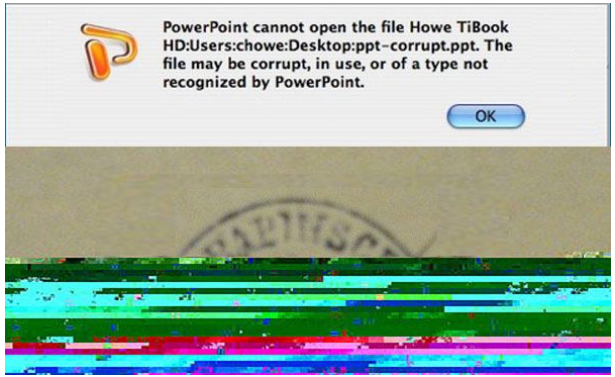
Software Failure/Threats



- Obsolescence ("software rot", changes in file formats, encoding)
- Incompatibility (between operating systems, file systems, partitioning)
- Lack of transparency
- Viruses/malware

Degradation and Bit Rot

- Physical deterioration of media (binder, dyes, etc.)
- Loss of magnetic orientation (hard disk, magnetic tape)
- “Bit rot” - gradual corruption of data due to accumulation of non-critical failures (e.g. random electric discharges altering data)



Natural Disasters

- Fire, earthquake, flooding
- Threat to both local and networked storage
- Stress need for geographic separation
- Affect manufacturing as well

Home > Data Center

NEWS ANALYSIS

Hurricane Sandy leaves wounded servers in its wake

As disaster recovery firms struggle to restore damaged data centers, experts warn of further storm-related breakdowns in the months ahead.

By Patrick Thibodeau | Follow
Senior Editor, Computerworld | NOV 19, 2012 6:00 AM PT

Data recovery experts have been busy in the wake of Hurricane Sandy, which left a slew of data centers underwater, damaging equipment and posing a significant threat to business-critical data.

Apparently disregarding weather forecasters' widespread warnings and underestimating the power of the storm that hit the East Coast late last month, many businesses didn't begin moving computer and IT communications equipment out of harm's way until it was too late, say officials at companies that specialize in data recovery.

Many data centers were casualties of the massive storm, and the damage threatened to shut down major New York-based businesses and interrupt Internet service across the country, according to experts.

For instance, the storm forced two so-called carrier hotels -- monolithic buildings that serve as major U.S. network hubs -- in lower Manhattan to operate on generator power for a significant period of time.

MORE LIKE THIS

- Sandy wounded servers, some grievously, say services firms
- Storm forces Internet hubs to run on generator power
- Drama in NYC as data center temp passes 100 degrees

VIDEO
Need a data center location? Iceland wants you

Computing

Thailand's devastating floods are hitting PC hard drive supplies, warn analysts

Country is centre of world hard drive manufacture but factories have been seriously affected by floods, which could affect supplies early in 2012

This article is 5 years old

Charles Arthur and agencies

@charlesarthur

Tuesday 25 October 2011 01.41 EDT

Thai residents make their way through a flooded street in Pathum Thani, Thailand. Photograph: Daniel Berehulak/Getty Images

Images: Eyebeam/Jonathan Minard

Human Error

- Poor handling of physical media
- Unintentional deletion/overwrite/modification
- Or this:

Summary of the Amazon S3 Service Disruption in the Northern Virginia (US-EAST-1) Region

We'd like to give you some additional information about the service disruption that occurred in the Northern Virginia (US-EAST-1) Region on the morning of February 28th. The Amazon Simple Storage Service (S3) team was debugging an issue causing the S3 billing system to progress more slowly than expected. At 9:37AM PST, an authorized S3 team member using an established playbook executed a command which was intended to remove a small number of servers for one of the S3 subsystems that is used by the S3 billing process. Unfortunately, one of the inputs to the command was entered incorrectly and a larger set of servers was removed than intended. The servers that were inadvertently removed supported two other S3 subsystems. One of these subsystems, the index subsystem, manages the metadata and location information of all S3 objects in the region. This subsystem is necessary to serve all GET,

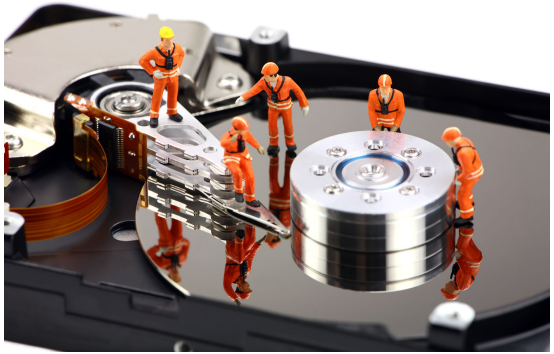
Storage Media and Characteristics

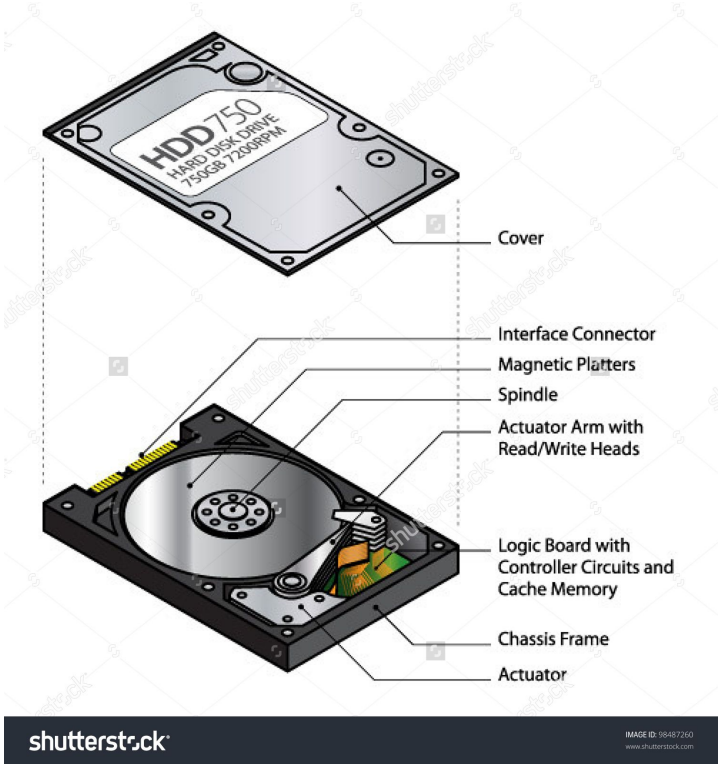
What are our options?

- Hard Disk Drive (HDD)
- Solid-State Drive (SSD)/Flash
- Optical Disc
- Magnetic/Data Tape

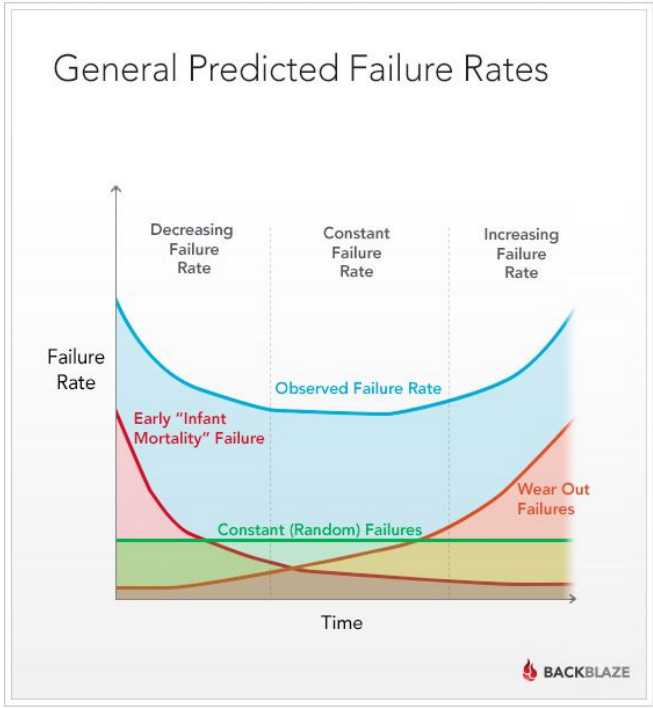
Hard Disk Drives (HDDs)

- Mechanical-magnetic (platter, spindle, head)
- Can be internal or external (interface varies)
- “Kryder’s Law” - exponential growth of areal density between 1990-2010; capacity has slowed but is still increasing





HDD Failure Rates



Hard Drive Failure Rates by Drive Size

For 2016, as of 12/31/2016

Drive Size	Drive Count	Drive Days	Drive Failures	Failure Rate
3 TB	6,605	2,416,353	93	1.40%
4 TB	54,189	18,504,977	1,042	2.06%
5 TB	45	16,425	1	2.22%
6 TB	2,335	850,992	41	1.76%
8 TB	8,765	1,093,435	48	1.60%
Totals	**71,939**	**22,882,182**	**1,225**	**1.95%**

BACKBLAZE

Hard Drive Failure Rates by Manufacturer

For 2016, as of 12/31/2016

MFG	Drive Count	Drive Days	Drive Failures	Failure Rate
HGST	24,545	8,028,068	132	0.60%
Seagate	45,531	14,194,485	1,029	2.65%
Toshiba	237	86,308	3	1.27%
WDC	1,626	573,321	61	3.88%
Totals	**71,939**	**22,882,182**	**1,225**	**1.95%**

BACKBLAZE

Hard Drive Annualized Failure Rates for 2016

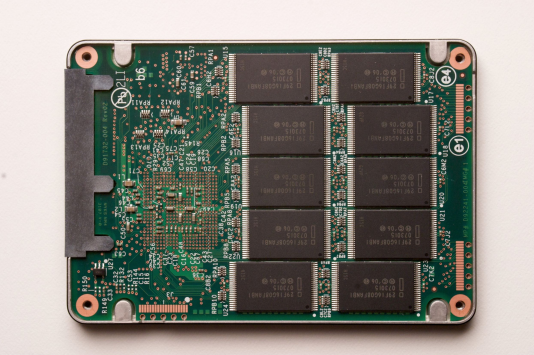
Reporting period 1/1/2016 - 12/31/2016 inclusive

MFG	Model	Drive Size	Drive Count	Avg Age (months)	Drive Days	Drive Failures	Failure Rate
HGST	HUH728080ALE600	8 TB	45	22.99	16,155	-	0.00%
Seagate	ST8000DM002	8 TB	8,660	4.72	1,075,720	48	1.63%
Seagate	ST8000NM0055	8 TB	60	1.44	1,560	-	0.00%
Seagate	ST6000DX000	6 TB	1,889	21.48	684,840	16	0.85%
WDC	WD60EFRX	6 TB	446	24.14	166,152	25	5.49%
Toshiba	MD04ABA500V	5 TB	45	22.15	16,425	1	2.22%
HGST	HDS5C4040ALE630	4 TB	2,625	45.35	987,011	14	0.52%
HGST	HMS5C4040ALE640	4 TB	7,014	29.48	2,579,698	28	0.40%
HGST	HMS5C4040BLE640	4 TB	9,407	15.51	2,436,130	34	0.51%
Seagate	ST4000DM000	4 TB	34,738	21.73	12,359,750	938	2.77%
Seagate	ST4000DX000	4 TB	184	38.54	72,615	27	13.57%
Toshiba	MD04ABA400V	4 TB	146	20.61	52,983	-	0.00%
WDC	WD40EFRX	4 TB	75	17.16	16,790	1	2.17%
HGST	HDS5C3030ALA	3 TB	4,476	55.87	1,647,137	34	0.75%
HGST	HDS723030ALA	3 TB	978	61.21	361,937	22	2.22%
Toshiba	DT01ACA300	3 TB	46	44.12	16,900	2	4.32%
WDC	WD30EFRX	3 TB	1,105	30.39	390,379	35	3.27%
	Totals		**71,939**		**22,882,182**	**1,225**	**1.95%**

BACKBLAZE

https://www.backblaze.com/blog/hard-drive-benchmark-stats-2016/

Solid-State Drives (SSDs)/Flash



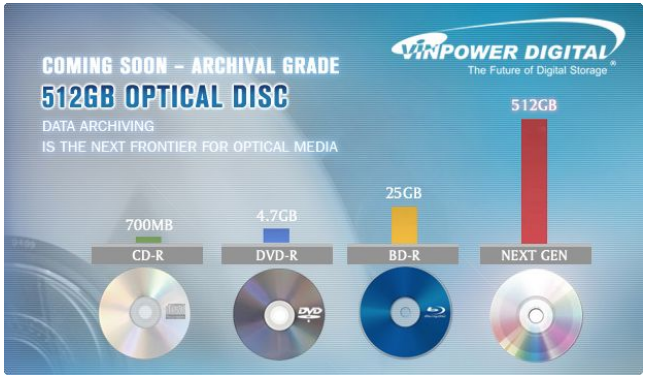
- Electrical (non-mechanical)
- Can be internal or external
- Can lose charge due to poor electrical insulation or frequent use over time (relatively poor “write-endurance”)
- Limited capacity/more expensive to manufacture

Hybrid (SSHD) Drives

- Physical combination of SSD (for application/OS data cache) and HDD (for user data) into one device
- **Self-optimized system** (drive itself determines what kind of data goes where) or **host-optimized** (OS, in conjunction with specialized software, determines what data goes where)
- Benchmarks somewhere between HDDs and SSDs for performance (duh)
- Mostly for laptops

Optical Discs

- Hardware becoming less common in consumer environments
- Prone to mishandling
- Lifespan depends on materials/dyes used, storage environment
- Limited capacity compared to other media - although Sony has soldiered on developing cartridge-based super optical formats

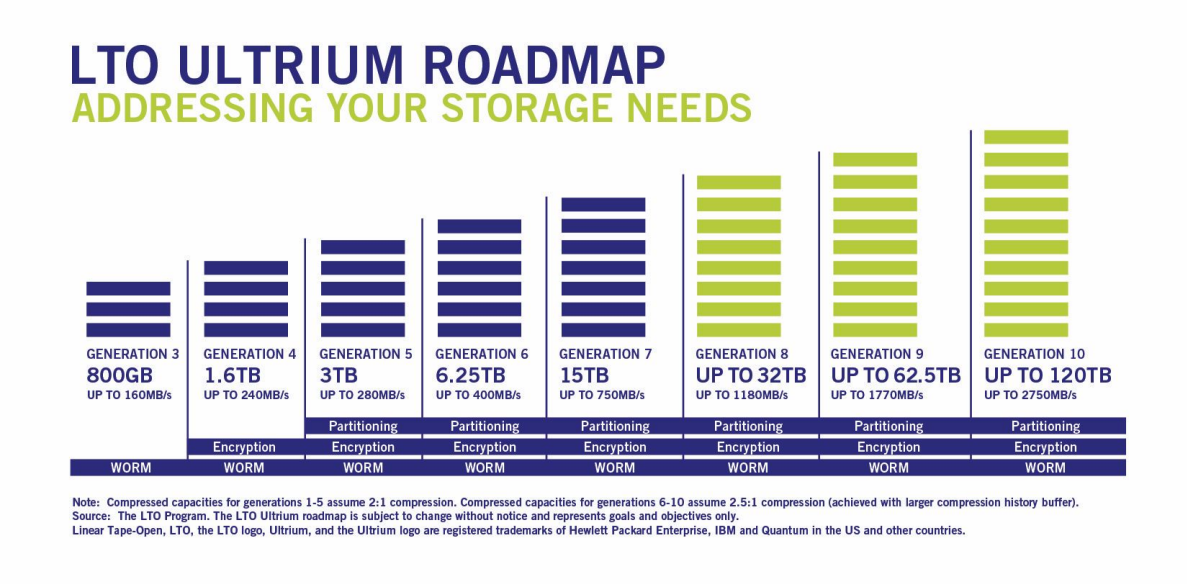


Magnetic/Data Tape

- Shares physical concerns of all magnetic tape formats
- Dominated by LTO (Linear Tape Open) standard
- Large capacity, quick write speed
- Sequential access = slow read speeds

LTO

- LTFS (Linear Tape File System) - XML-based, self-describing, makes LTO cartridges mountable
- Hardware is backwards-compatible for two generations



Physical Interfaces and Data Transfer Protocols

How do we access our storage?

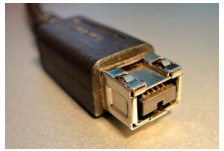
- USB
- FireWire/Thunderbolt
- eSATA
- SAS
- Ethernet and Fibre

USB (Universal Serial Bus)

- Wide adoption
- Backwards compatible
- Proliferation of physical interfaces/connectors
- Can power devices as well as transfer data

FireWire/Thunderbolt

- Proprietary Apple protocols
- Thunderbolt replaced FireWire, piggybacked on existing physical connectors
- Supports daisy-chaining many devices with no loss to transfer speed



SATA/eSATA

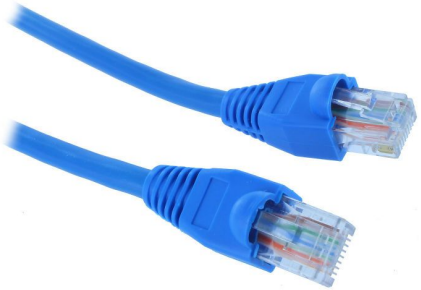
- Serial Advanced Technology Attachment common for internal HDD/SSD connections
- External version eSATA was faster than USB (until USB 3.1), non-proprietary
- eSATA not common on consumer desktops, require add-on expansion cards

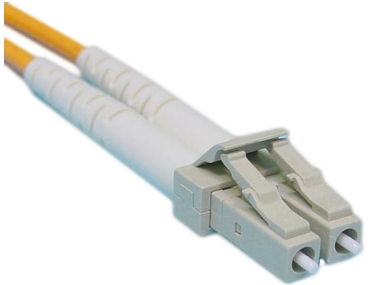
SAS (Serial Attached SCSI)

- Modern update of SCSI (Small Computer System Interface) protocol used for connecting peripheral devices
- Very rare with consumer or even many professional setups
- Enterprise-level connection for setting up servers and SANs

Ethernet and Fibre Channel

- Networked devices
- Fibre channel = typically seen with enterprise-level SANs, use optical cable





Transfer Speed Chart

USB 1.1	1.5 MB/s		
Firewire 400	49 MB/s		
USB 2.0	60 MB/s		
FireWire 800	98 MB/s		
SATA II	375 MB/s	3 Gb/s	
USB 3.0	625 MB/s	5 Gb/s	
SATA III/eSATA	750 MB/s	6 Gb/s	
USB 3.1	1250 MB/s	10 Gb/s	
Thunderbolt	1250 MB/s	10 Gb/s	x2 (dual channel)
Thunderbolt 2	2500 MB/s	20 Gb/s	x2 (dual channel)

Thunderbolt 3 | 5000 MB/s | 40 Gb/s | x2 (dual channel)
SATA III.2 | 1969 MB/s | 16 Gb/s

Fibre Channel | 8-128 Gb/s | cables can run for kilometers

Types of External Storage

What if I need more space? (you do)

- Direct Attached Storage (DAS)
- Network Attached Storage (NAS)
- Storage Area Network (SAN)
- Cloud storage

Direct Attached Storage (DAS)

- Disk subsystems (files on an external drive, disk or tape) are directly attached to workstation or server
- “Portable” (may or may not need dedicated power supply)
- Not networked
- May refer to a storage unit with more than one drive/disk
- USB, Thunderbolt, eSATA connections

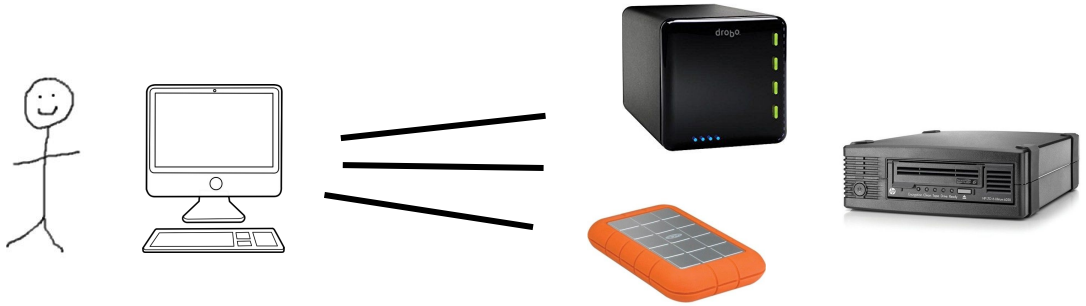
Network Attached Storage (NAS)

- Networked file-based storage (not a server)
- Flexible management software - CLI, mountable to local filesystem, browser GUI
- Accessible by multiple users/computers at the same time, remotely
- Usually built with multiple HDDs
- Transfer speeds limited to network capability
- Rack-mount/server or desktop units available
- NAS-enabled routers
- Usually Ethernet-connected (some sort of direct attachment for troubleshooting)

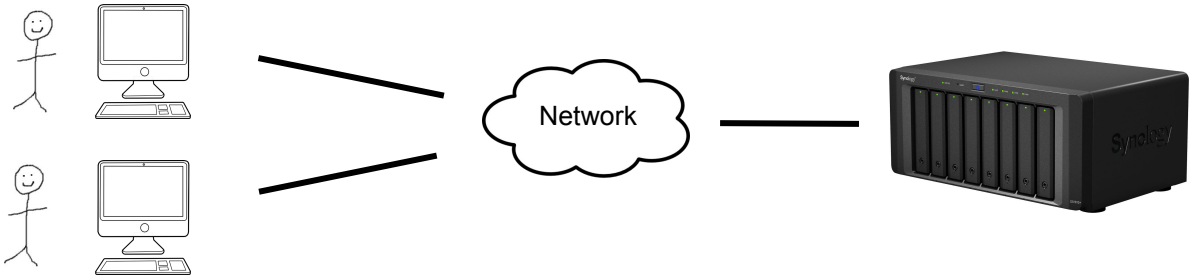
Storage Area Network (SAN)

- A group of storage devices made available to a server (and via the server to others, over network)
- High-speed for max performance
- Multiple users can access, remotely
- Block-based storage - requires IT management to install OS, distribute volumes and data
- Usually Fibre Channel and SAS connections

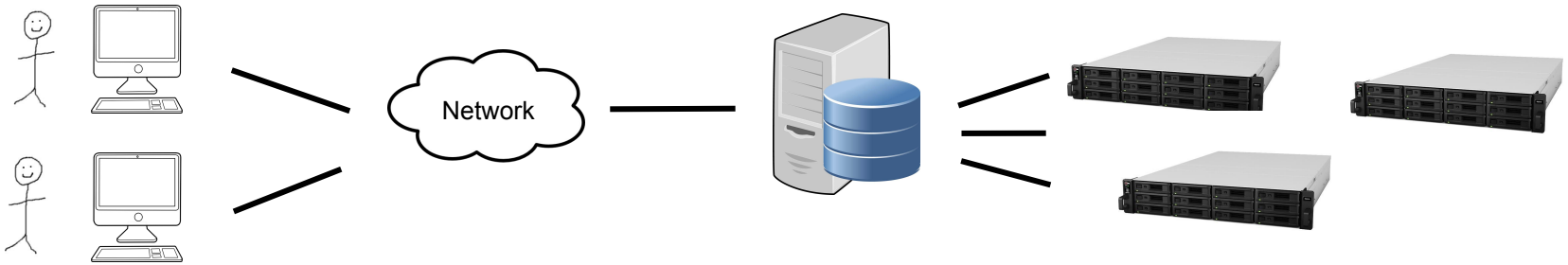
DAS



NAS



SAN



Cloud Storage

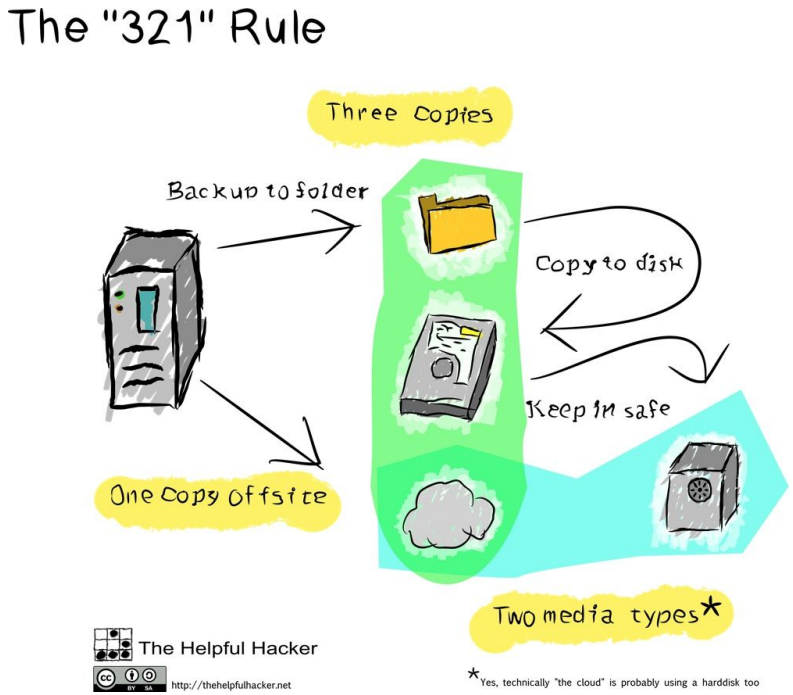
- Cheap
- Handled by third-party vendors
- Ability to access/manage/download files will vary depending on service
- No control over physical management

Administrative Strategy Beyond the Object

What else?

- Redundancy and Fault Tolerance
- File Fixity
- Migration
- Security and Access Control
- Tech Support/Administration

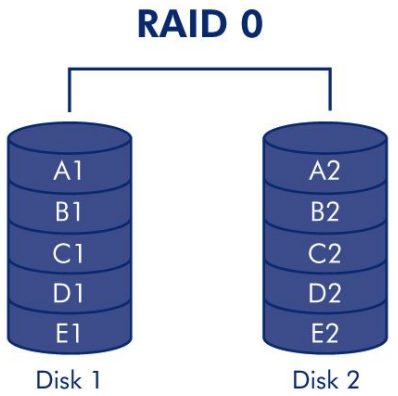
Redundancy and Fault Tolerance

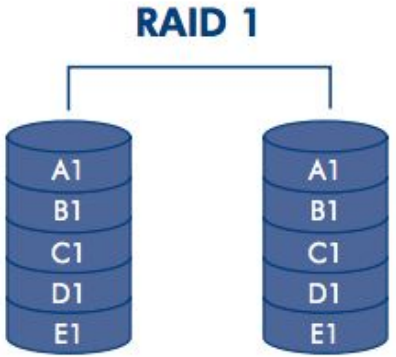


- LOCKSS
- 3-2-1 strategy:
  - 3 total copies
  - 2 local, on different devices
  - 1 offsite (preferably geographic separation)
- RAID (Redundant Array of Independent [Inexpensive] Disks) - make multiple disks into one logical volume

RAID Levels

- JBOD (Just a Bunch Of Disks)
  - [not really a RAID]
- RAID 0
  - Data striping
  - Improves read/write speed
  - Maintains total physical capacity
  - No redundancy - lose one disk, lose all data
- RAID 1
  - Data mirroring
  - Slower write speed
  - Half total physical capacity
  - Ensures data protection

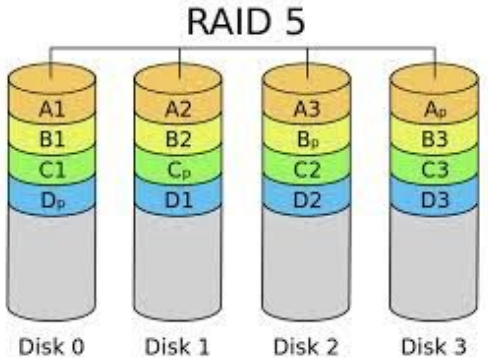


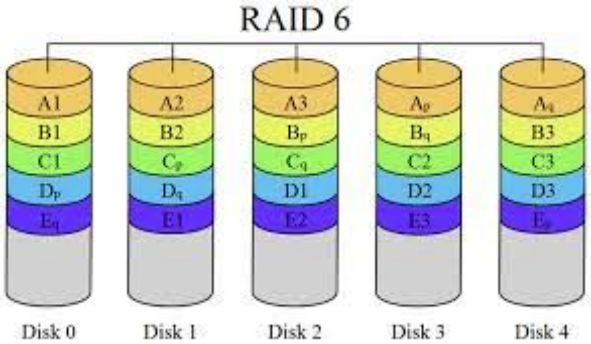


RAID Levels

- RAID 5
  - At least 3 drives
  - Data and parity distributed to all disks
  - If one drive fails, RAID can rebuild all data
  - Hot swappable

- RAID 6
  - At least 5 drives
  - Data and parity distributed to all disks
  - If two drives fail, RAID can rebuild all data
  - Hot swappable



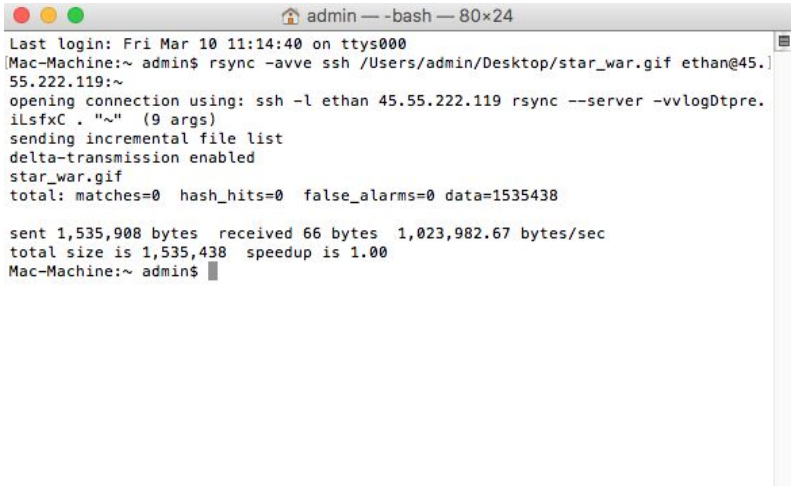


File Fixity

- Make sure files don’t change (corrupt) over time
- Accomplished via computing and validation of checksums
- Multiple checksum hash generators are common:
  - MD5
  - Security Hash Algorithm
    - SHA-1
    - SHA-2 (SHA-256, SHA-512)

Migration

- Regular, scheduled movement to new drives/media
- rsync = ❤□□❤□□❤□□
- Maintain fixity during transfer (Bag-It spec)



rsync : invoke command

-a : “archive” mode, preserve permissions, attributes, ownership, copy recursively

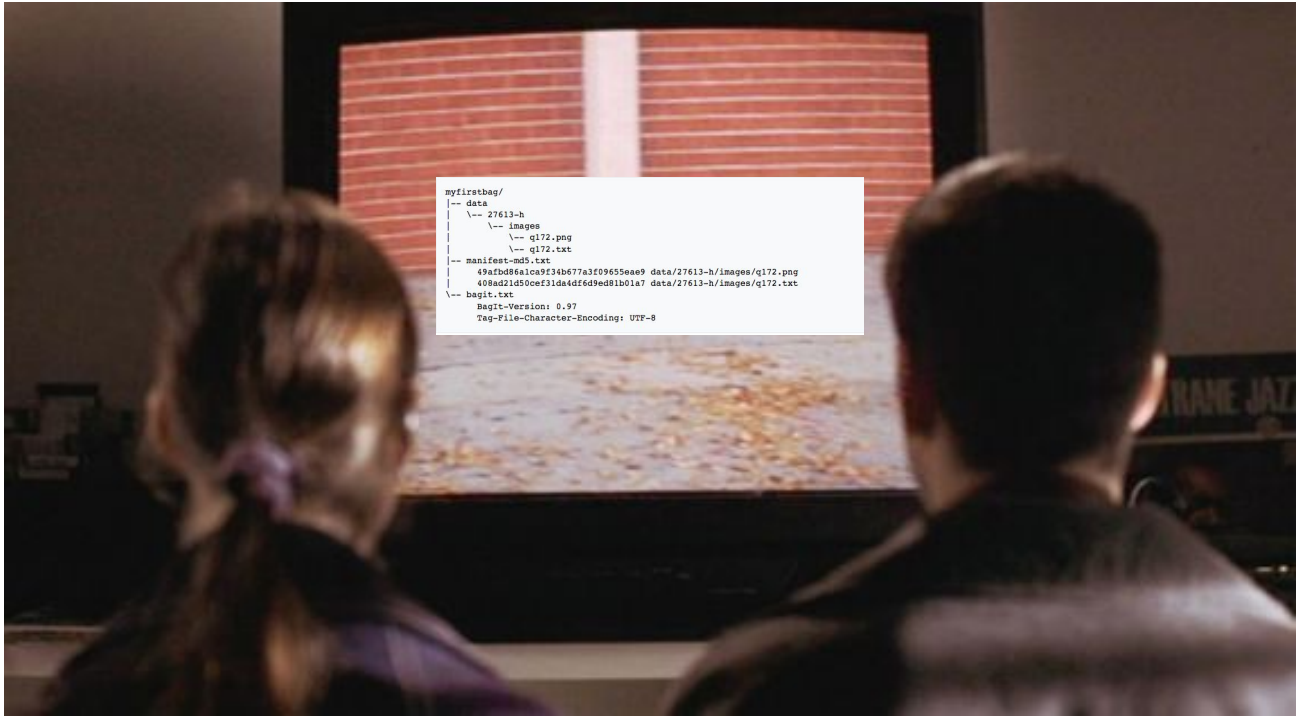
-v : “verbose”, displays information about transfer (double verbosity enabled here)

-e : edit configuration (here, specified SSH transfer protocol)

/input/file_path

user@server-IP:/output/file_path

BagIt


myfirstbag/
|-- data
| \-- 27613-h
| \-- images
| \-- q172.png
| \-- q172.txt
|-- manifest-md5.txt
| 49afbd86a1ca9f34b677a3f09655eae9 data/27613-h/images/q172.png
| 408ad21d50cef31da4df6d9ed81b01a7 data/27613-h/images/q172.txt
\-- bagit.txt
      BagIt-Version: 0.97
      Tag-File-Character-Encoding: UTF-8

Security and Access Control

- Permissions - who can view, modify, administrate
- Can vary at institutional level, volume level, collection level, file level
- Encryption

Tech Support/Administration

- Who’s managing all this?
- Archival vs. sysadmin concerns

Failure and Data Recovery

What if it's "too late"?

- RAID rebuilding
- ddrescue
- Hex code surgery

RAID Rebuilding

- Pre- and post-failure rebuilding options usually available (pre-failure requires hot spare drive in array)

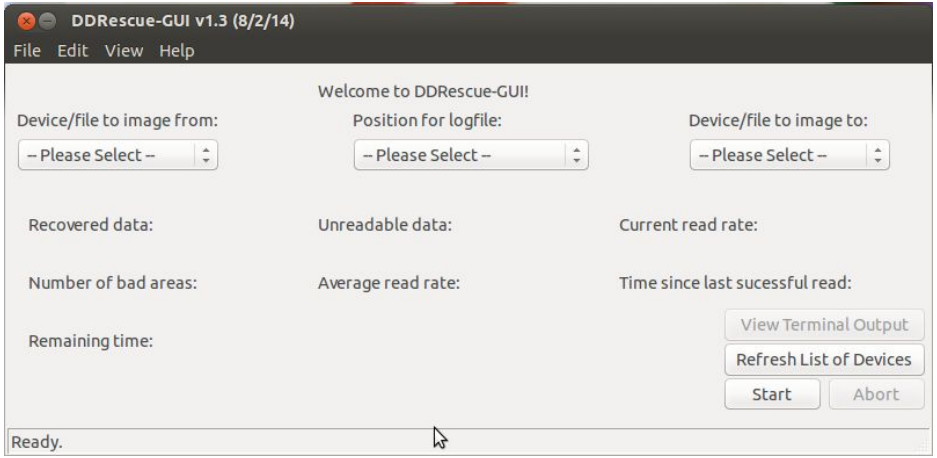
- Speed depends on:
  - Number of drives in array
  - Size of the drives
  - Priority given to rebuilding activities

ddrescue

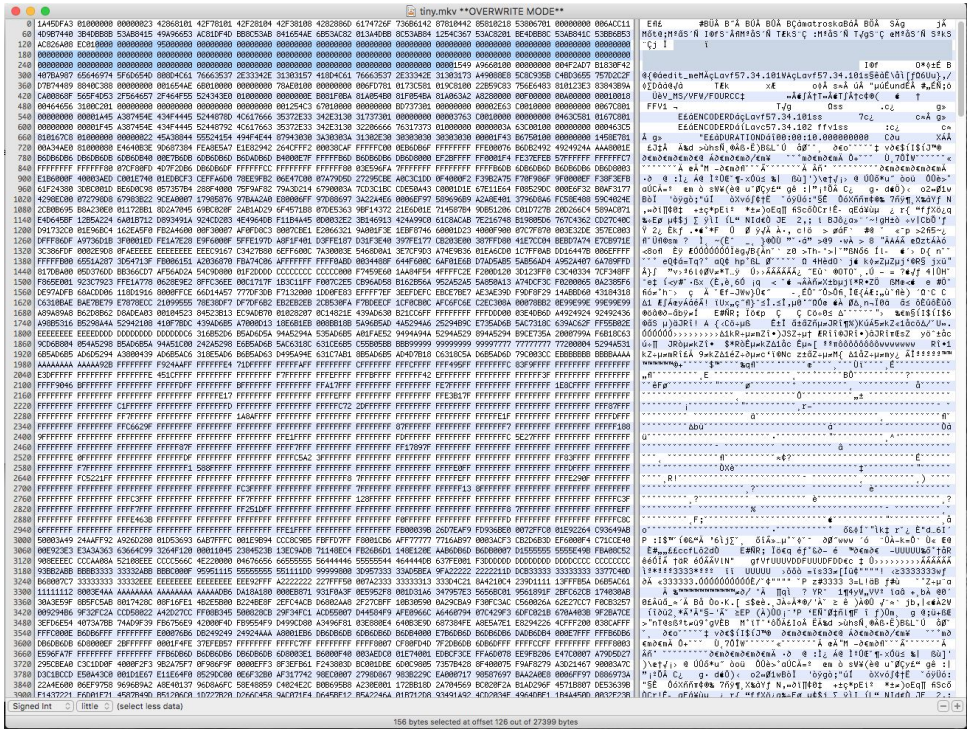
```
GNU ddrescue - Data recovery tool.
Copies data from one file or block device to another,
trying hard to rescue data in case of read errors.

Usage: ./ddrescue [options] infile outfile [logfile]
Options:
  -h, --help                    display this help and exit
  -V, --version                 output version information and exit
  -b, --block-size=<bytes>      hardware block size of input device [512]
  -B, --binary-prefixes         show binary multipliers in numbers [default SI]
  -c, --cluster-size=<blocks>   hardware blocks to copy at a time [128]
  -C, --complete-only           do not read new data beyond logfile limits
  -e, --max-errors=<n>          maximum number of error areas allowed
  -i, --input-position=<pos>    starting position in input file [0]
  -n, --no-split                do not try to split error areas
  -o, --output-position=<pos>   starting position in output file [ipos]
  -q, --quiet                   quiet operation
  -r, --max-retries=<n>         exit after given retries (-1=infinity) [0]
  -s, --max-size=<bytes>        maximum size of data to be copied
  -t, --truncate                truncate output file
  -v, --verbose                 verbose operation
Numbers may be followed by a multiplier: b = blocks, k = kB = 10^3 = 1000,
Ki = KiB = 2^10 = 1024, M = 10^6, Mi = 2^20, G = 10^9, Gi = 2^30, etc...

Report bugs to bug-ddrescue@gnu.org
```



Hex Code Example



Choosing Your Storage Strategy

How do I break it down?

- Total size of collections (GB/TB)
- Individual file sizes
- Growth rate
- Budget
- Users
- Network capacity
- IT support
- Vendor/developer support
- Community adoption/support

File Storage Calculation Tools

- https://www.digitalrebellion.com/webapps/videocalc
- File storage calculator
- AJA DataCalc app