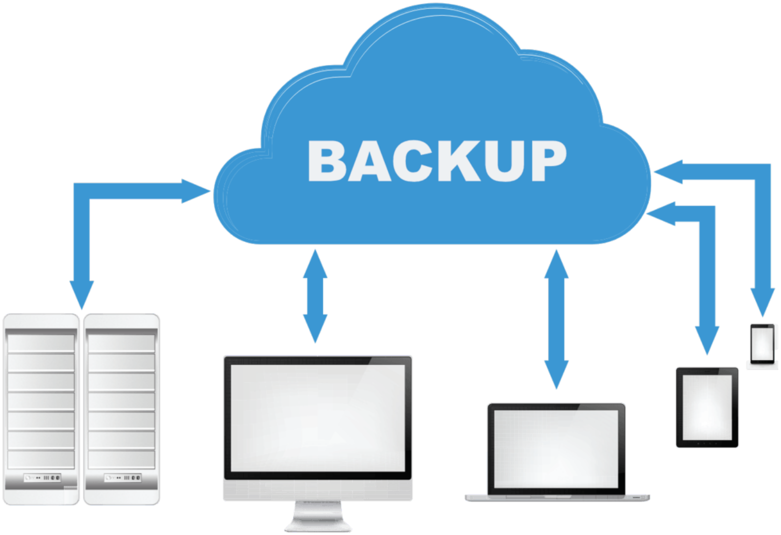
System Administration & Security

COMP 175 | Fall 2021 | University of the Pacific | Jeff Shafer



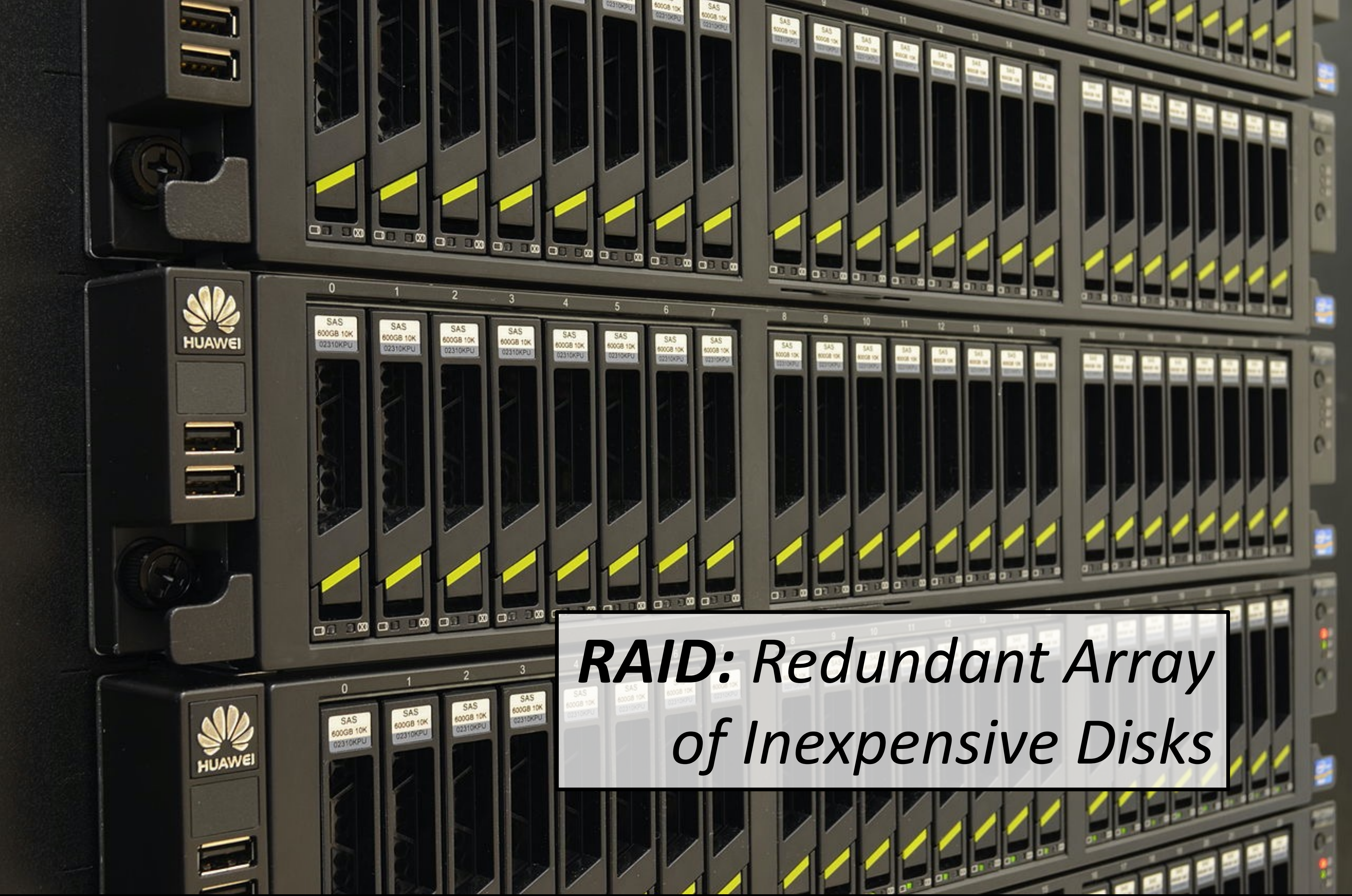
Backups

How Toy Story 2 Almost Got Deleted

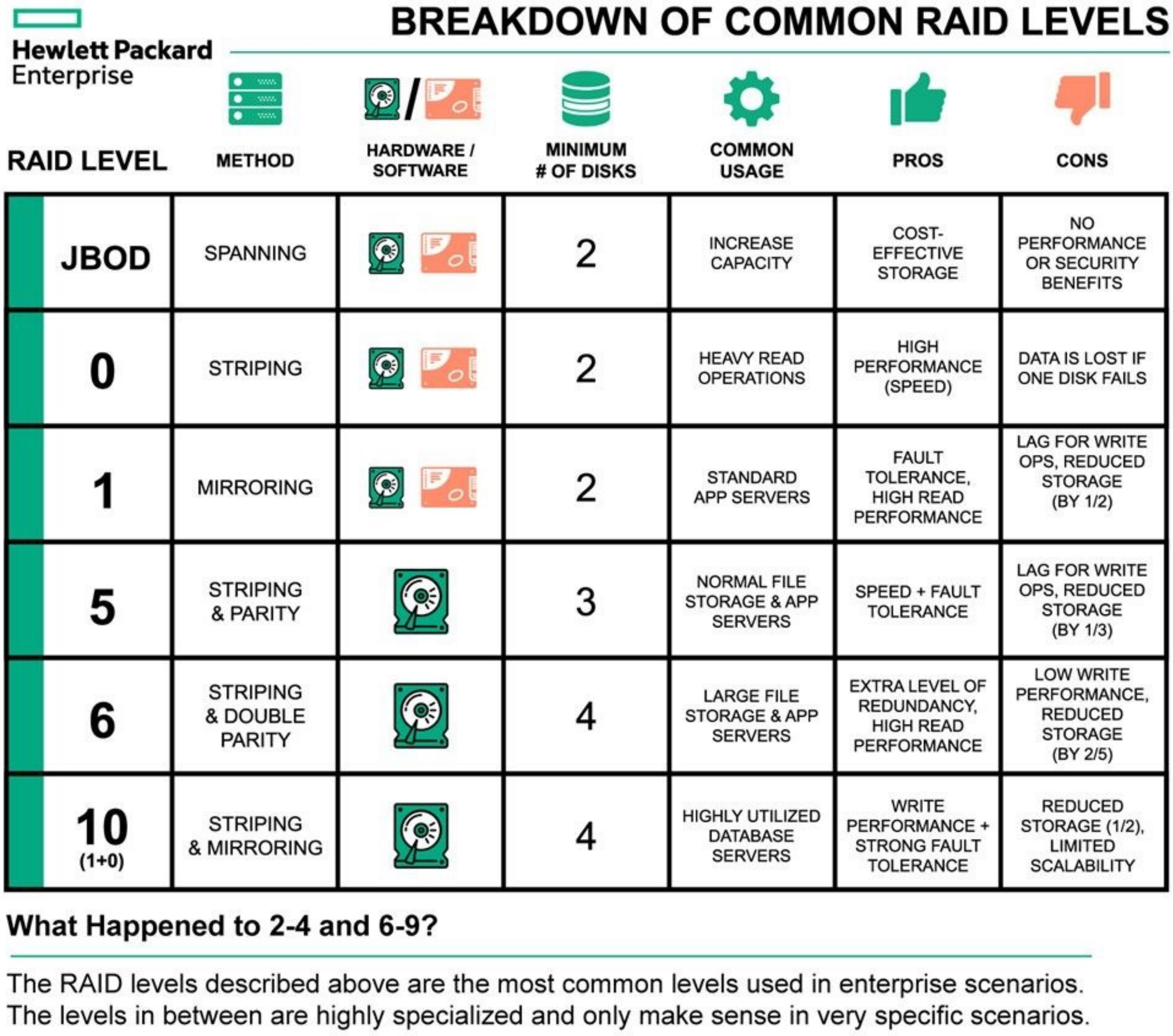
https://www.youtube.com/watch?v=8dhp_20j0Ys

Data Loss Scenarios

- Employee deletes the wrong file
- A single hard drive / SSD fails
- Corporate laptop is stolen
- A power supply fails and shorts out multiple drives in an enclosure simultaneously
- The datacenter catches on fire
- Malware infects and encrypts 1000 systems across the enterprise (roughly) simultaneously



RAID: *Redundant Array of Inexpensive Disks*



Hewlett Packard Enterprise

BREAKDOWN OF COMMON RAID LEVELS

RAID LEVEL	METHOD	HARDWARE / SOFTWARE	MINIMUM # OF DISKS	COMMON USAGE	PROS	CONS
JBOD	SPANNING	🖴 ▨	2	INCREASE CAPACITY	COST-EFFECTIVE STORAGE	NO PERFORMANCE OR SECURITY BENEFITS
0	STRIPING	🖴 ▨	2	HEAVY READ OPERATIONS	HIGH PERFORMANCE (SPEED)	DATA IS LOST IF ONE DISK FAILS
1	MIRRORING	🖴 ▨	2	STANDARD APP SERVERS	FAULT TOLERANCE, HIGH READ PERFORMANCE	LAG FOR WRITE OPS, REDUCED STORAGE (BY 1/2)
5	STRIPING & PARITY	🖴	3	NORMAL FILE STORAGE & APP SERVERS	SPEED + FAULT TOLERANCE	LAG FOR WRITE OPS, REDUCED STORAGE (BY 1/3)
6	STRIPING & DOUBLE PARITY	🖴	4	LARGE FILE STORAGE & APP SERVERS	EXTRA LEVEL OF REDUNDANCY, HIGH READ PERFORMANCE	LOW WRITE PERFORMANCE, REDUCED STORAGE (BY 2/5)
10 (1+0)	STRIPING & MIRRORING	🖴	4	HIGHLY UTILIZED DATABASE SERVERS	WRITE PERFORMANCE + STRONG FAULT TOLERANCE	REDUCED STORAGE (1/2), LIMITED SCALABILITY

What Happened to 2-4 and 6-9?

The RAID levels described above are the most common levels used in enterprise scenarios. The levels in between are highly specialized and only make sense in very specific scenarios.

https://community.hpe.com/t5/servers-systems-the-right/what-are-raid-levels-and-which-are-best-for-you/ba-p/7041151

RAID is NOT a Backup

RAID is NOT a Backup

- Operator makes a human error and erases `customer_data.dat`?
  - RAID won’t protect against that ☹
- Programmer screws up and business system corrupts running instance of `customer_data.dat`?
  - RAID won’t protect against that ☹
- Ransomware infects system and encrypts `customer_data.dat`?
  - RAID won’t protect against that ☹
- Firestorm burns down your Corporate HQ with scenic mountain views?
  - RAID won’t protect against that ☹

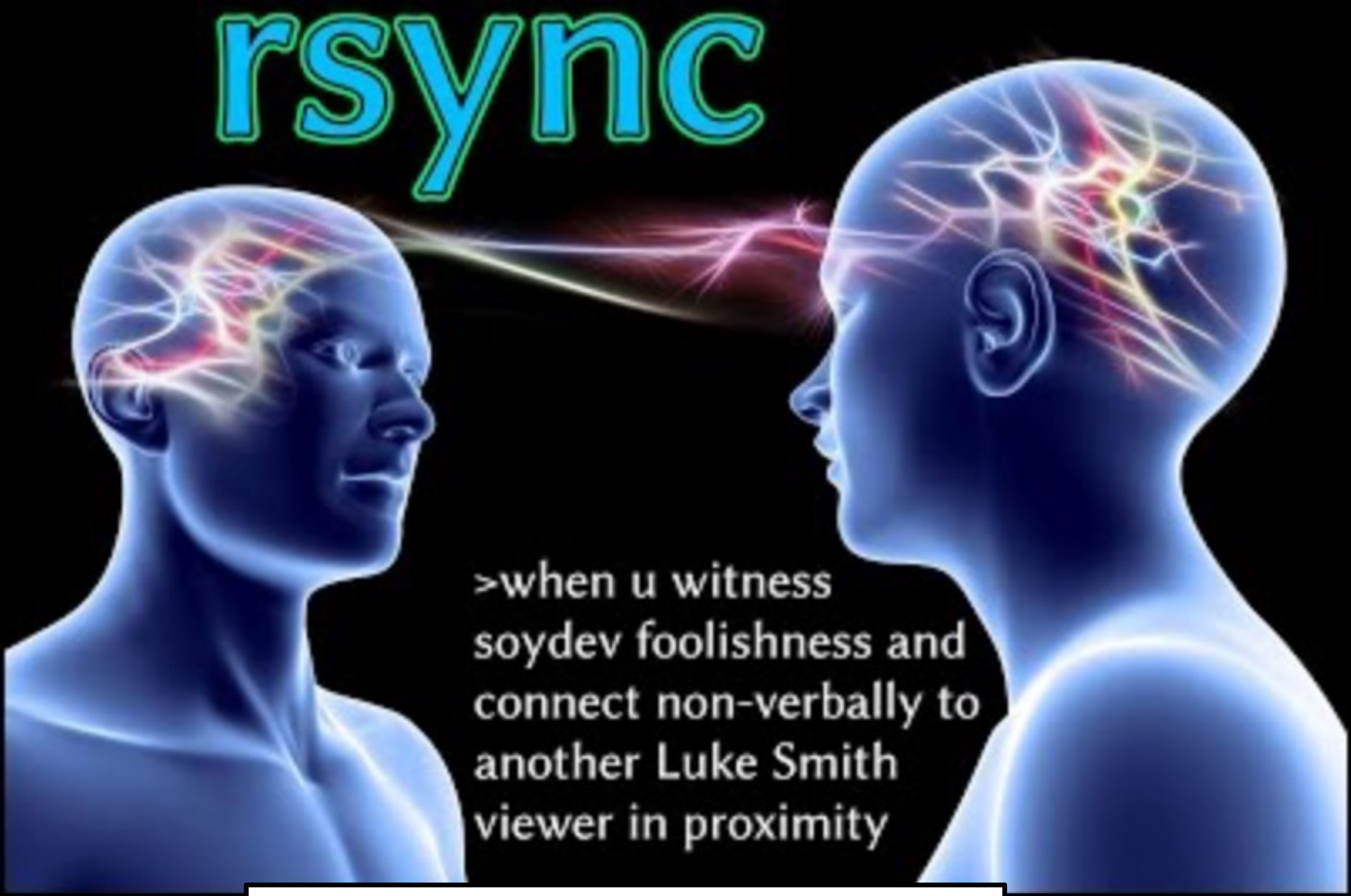
Backup Requirements

- **3-2-1 "Rule"**
  - **3 copies** of the data
  - **2 copies** in **different media** (disk vs tape vs "cloud")
  - **1 copy** being in a different location (offsite)
- **Periodic**
  - Hourly?
  - Daily?
  - Weekly?

Backup Requirements

- **Versioning**
  - Old copies of user data should be recoverable (Report v1, Report v2, …)
- **Write-once** backups (“**append-only**”)
  - A hacked (or corrupted) system shouldn’t be able to erase all its old backups too!
- **Encrypted** (data in flight and at rest should be secure)
- **Testable**

Rsync


rsync
>when u witness soydev foolishness and connect non-verbally to another Luke Smith viewer in proximity

https://www.youtube.com/watch?v=iTnWIKHtOnA

Additional Tools

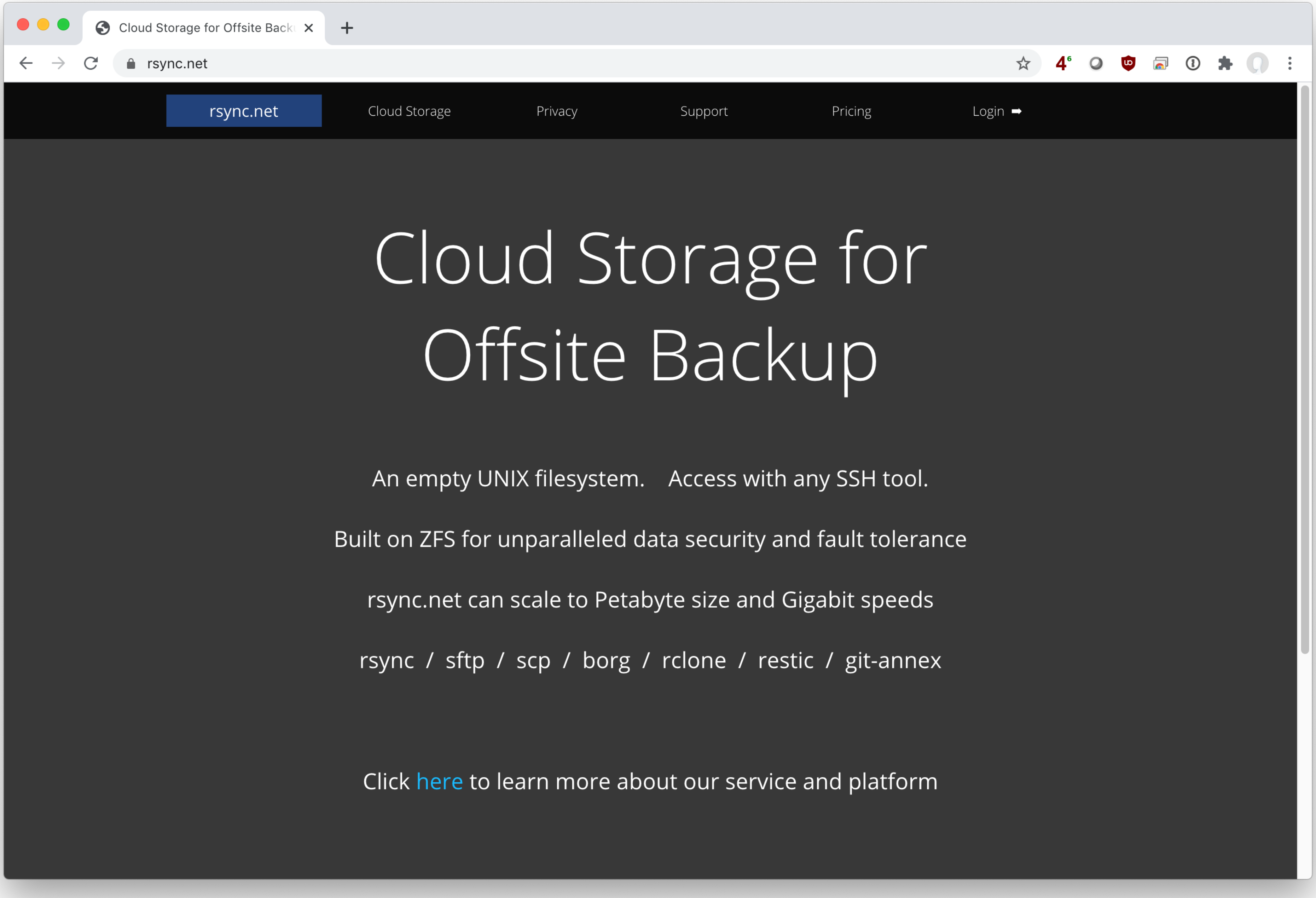
- **rsnapshot -** https://rsnapshot.org/
  - Rsync + Periodic snapshots
  - First run uses rsync to copy all files
  - Subsequent runs use rsync to copy only *modified files*
    - Unchanged files are “hard linked” to the original file to save disk space
  - Logrotate prunes the backups
    - `snapshot.0` is the newest, `snapshot.1` is one older, …

Additional Tools

- **Borg** - https://www.borgbackup.org/
  - Archive with **Deduplication**
  - Works with any backup server accessible over SSH
    - Performance bonus if Borg is also installed remotely
  - Encryption: 256-bit AES
  - Data integrity: SHA256
  - Compression: lz4 (fast) … lzma (space efficient)
  - Backups can be mounted as *userspace filesystems*
    - Restore files using regular file manager!
  - Supports *append-only* repository mode
    - A hacked client can not delete its older backups!

Deduplication

- Deduplication reduces the number of bytes stored
  - Each file is split into variable length chunks
  - Only chunks that have never been seen before are added to the repository
- How to identify uniqueness? Cryptographically strong **hash**
- All chunks in the same *repository* are considered
  - Typically 1 client = 1 repository
- Works great for
  - VM images (w/ sparse file support)
  - Physical disks
  - Renaming large directories
  - Historical duplication



Cloud Storage for Offsite Backup

An empty UNIX filesystem.   Access with any SSH tool.

Built on ZFS for unparalleled data security and fault tolerance

rsync.net can scale to Petabyte size and Gigabit speeds

rsync / sftp / scp / borg / rclone / restic / git-annex

Click here to learn more about our service and platform

https://www.rsync.net/

Lab / Project Idea for 2022? ☺

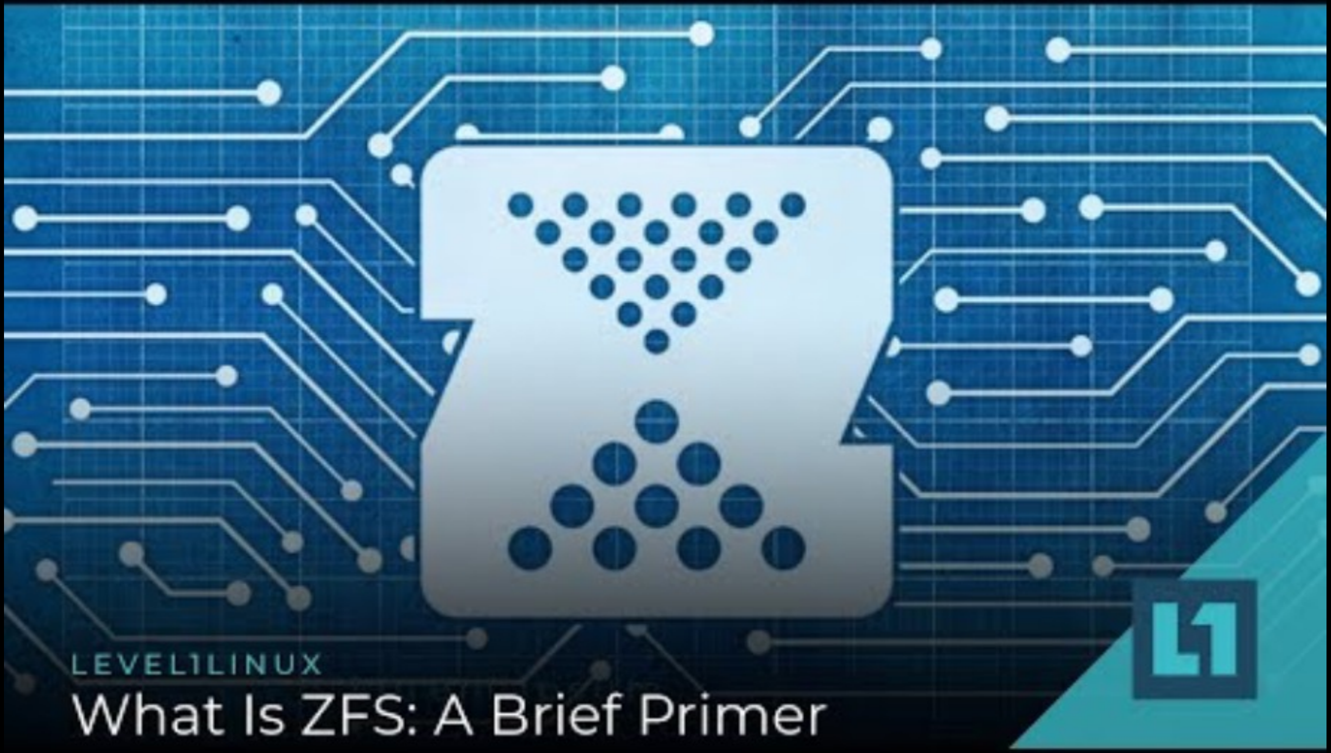
- AWS instance – “Backup Server”
  - Multiple EBS images (representing multiple disks)
  - ZFS file system tying all disks together into large storage pool w/checksums & redundancy, etc…
  - Separate repositories for each backup client (recommended security practice)
- Use Ansible to push Borg installation and configuration to all Tiger Enterprise systems
- Test backup **recovery** *(a critical step!)*

Wrap-Up

- Questions?
- Concerns?

- **Last Lab**
  - **Lab 14** – Ansible
- **Last Lecture**
  - System Monitoring
  - *Closing Thoughts*

What is ZFS?



https://www.youtube.com/watch?v=lsFDp-W1Ks0

“Don’t eat my data - 30+ years of storage war stories”

Steven Ellis (LCA 2020)

https://www.youtube.com/watch?v=PUuTUxnqX0g