(Disk) Space - the final frontier?

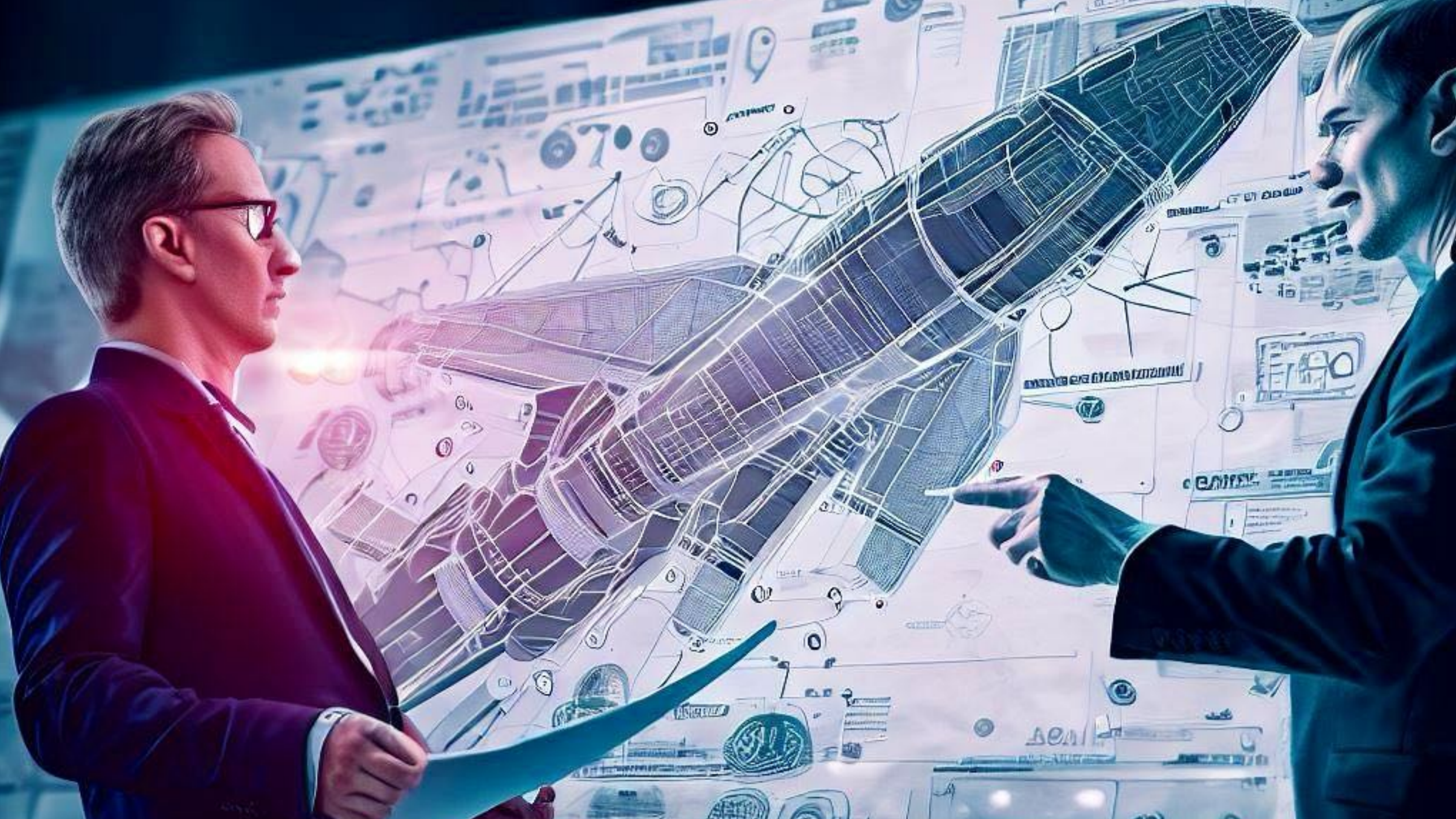
A Visual Journey to the Stars with OpenZFS

Benedict Reuschling
bcr@FreeBSD.org
BSD NL 2025

Why this talk?

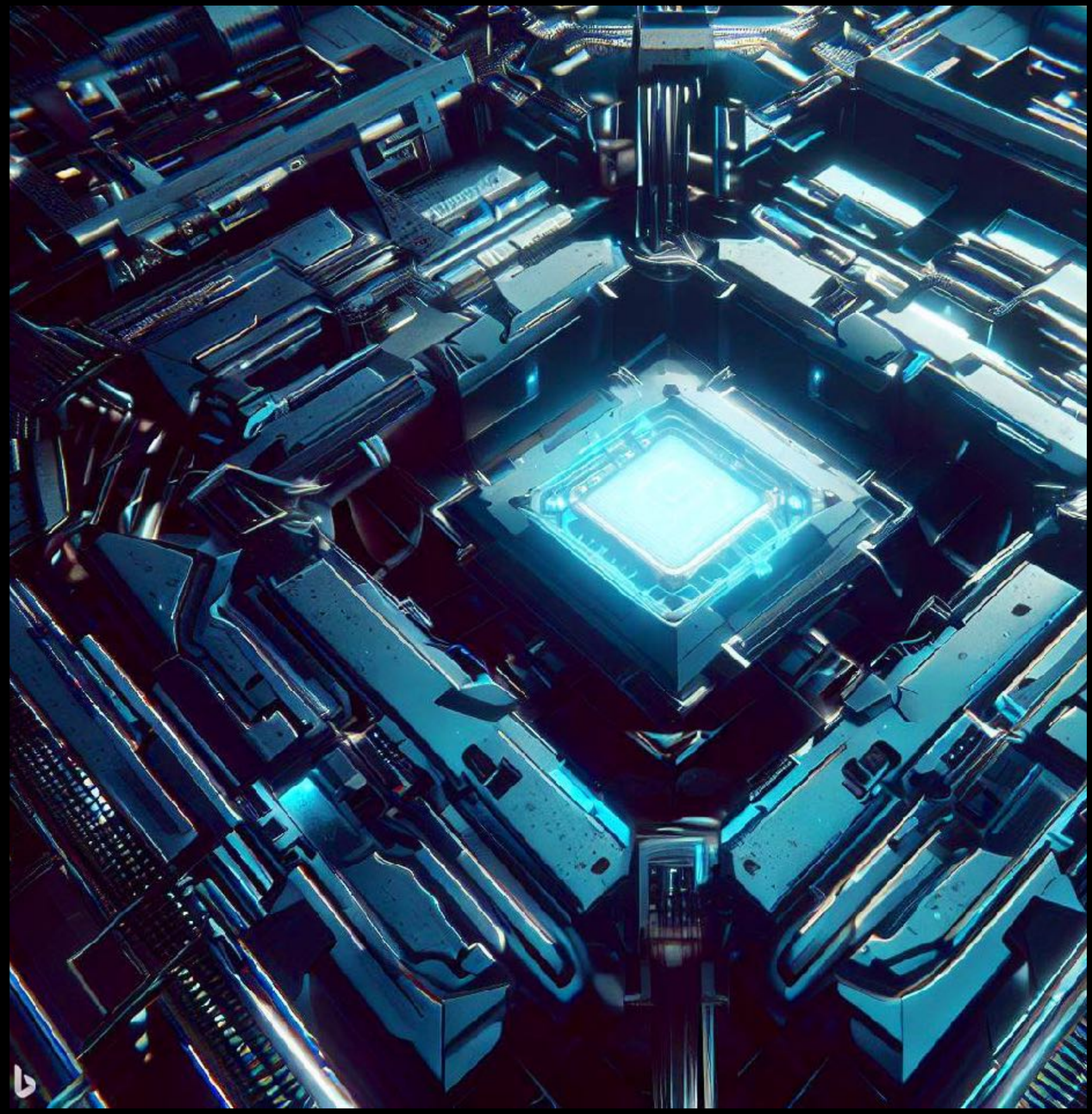
- I find my own tutorials boring (YMMV)
- Few illustrations, no story
- This time: exactly the opposite
- DALL-E Image Generator (Bing) helping my underdeveloped graphics talent
- Funny approach to teaching OpenZFS
- Experiment, let’s see how it goes

The Big Plan



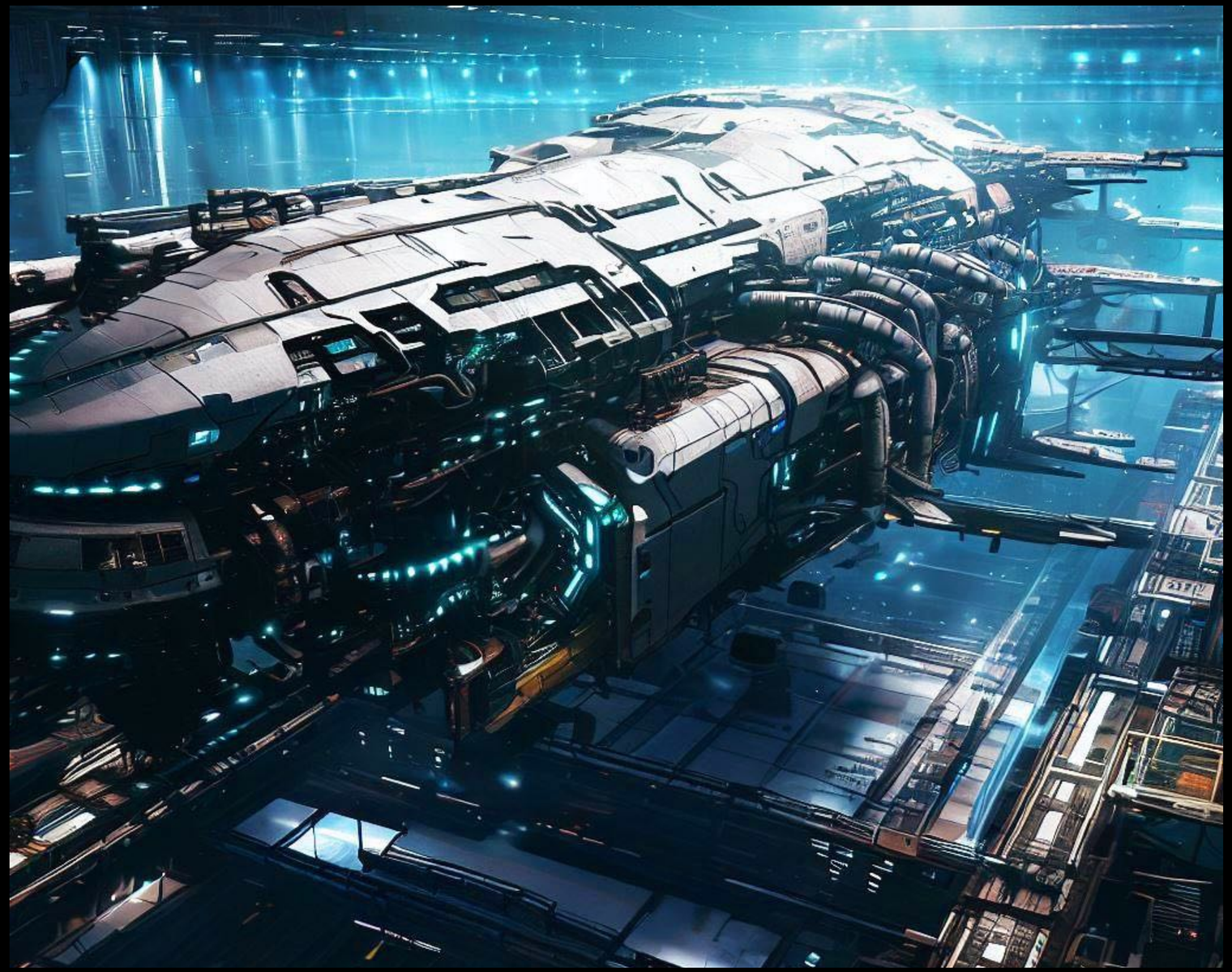
Journey to the Stars

- Goal: Explore the galaxy and collect research data
- Ship’s computer is the central element
- Engineers chose OpenZFS pool as primary on-board storage
- Standard crew including Chief Storage Engineer



And so the Engineers started to work…

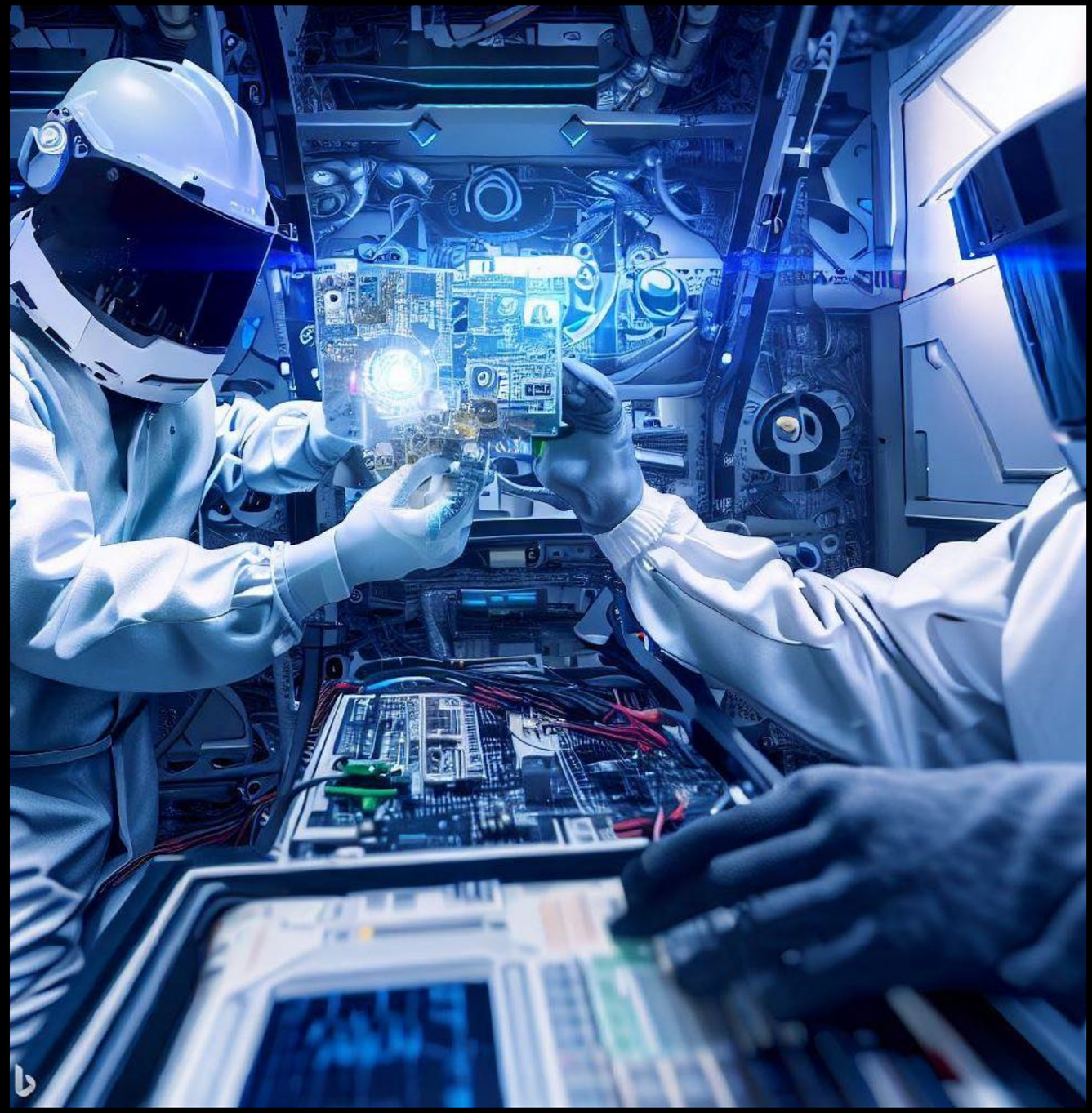
Spaceship under Construction



Initializing Spaceship Storage…

```
# zpool create BSDNL25
mirror media1 media2
```

Storage immediately available for the system, mounted and ready to do I/O



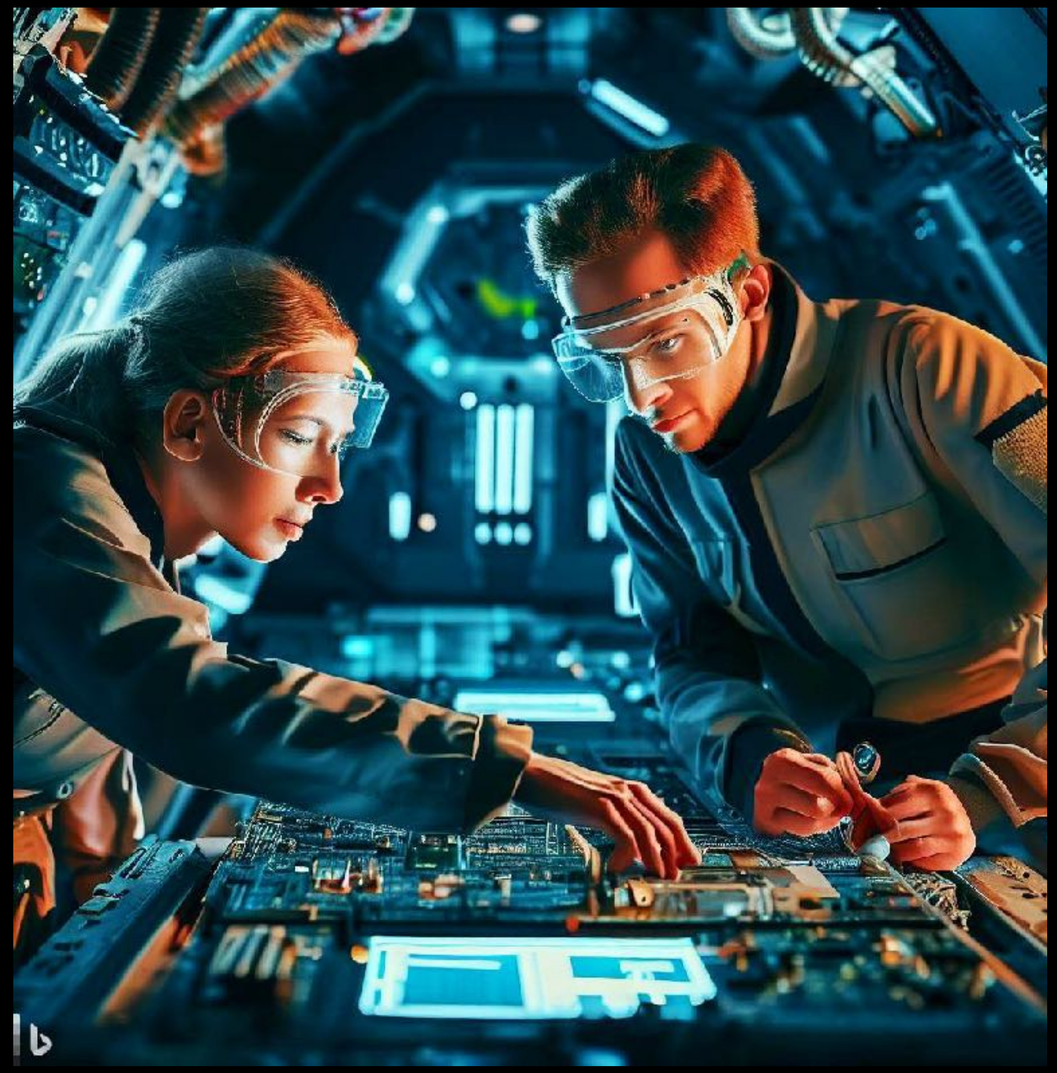
Start of the Journey

- Crew boards the ship
- Goodbye Earth
- Initial Systems Check

Status Check of the Storage Pool:

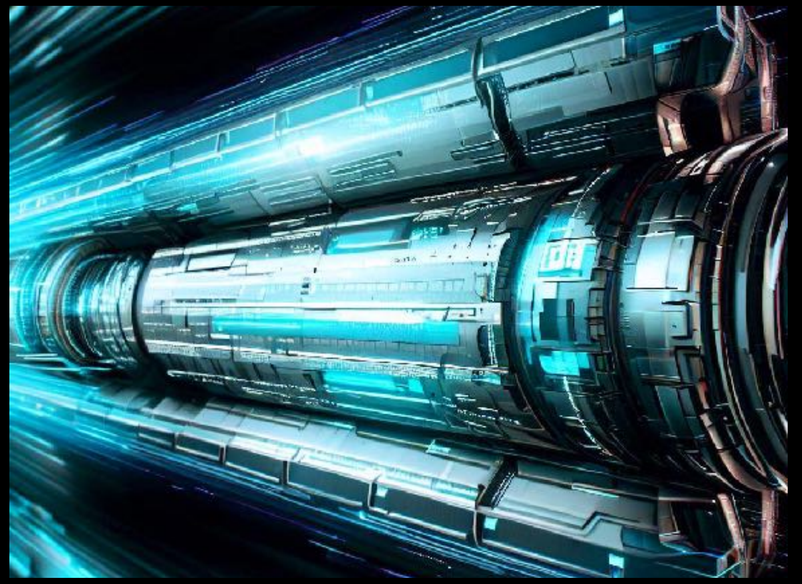
`zpool status` and `zpool list`

```
# zpool status

  pool: BSDNL25
 state: ONLINE
  scan:

config:

        NAME         STATE     READ WRITE CKSUM
        BSDNL25      ONLINE       0     0     0
          mirror     ONLINE       0     0     0
            media1   ONLINE       0     0     0
            media2   ONLINE       0     0     0
```



```
zpool list

NAME      SIZE  ALLOC   FREE  CKPOINT  EXPANDSZ   FRAG     CAP  DEDUP    HEALTH ALTROOT
BSDNL25   880T  15.3G   865T        -         -     0%      1%  1.00x    ONLINE -
```

All systems go…



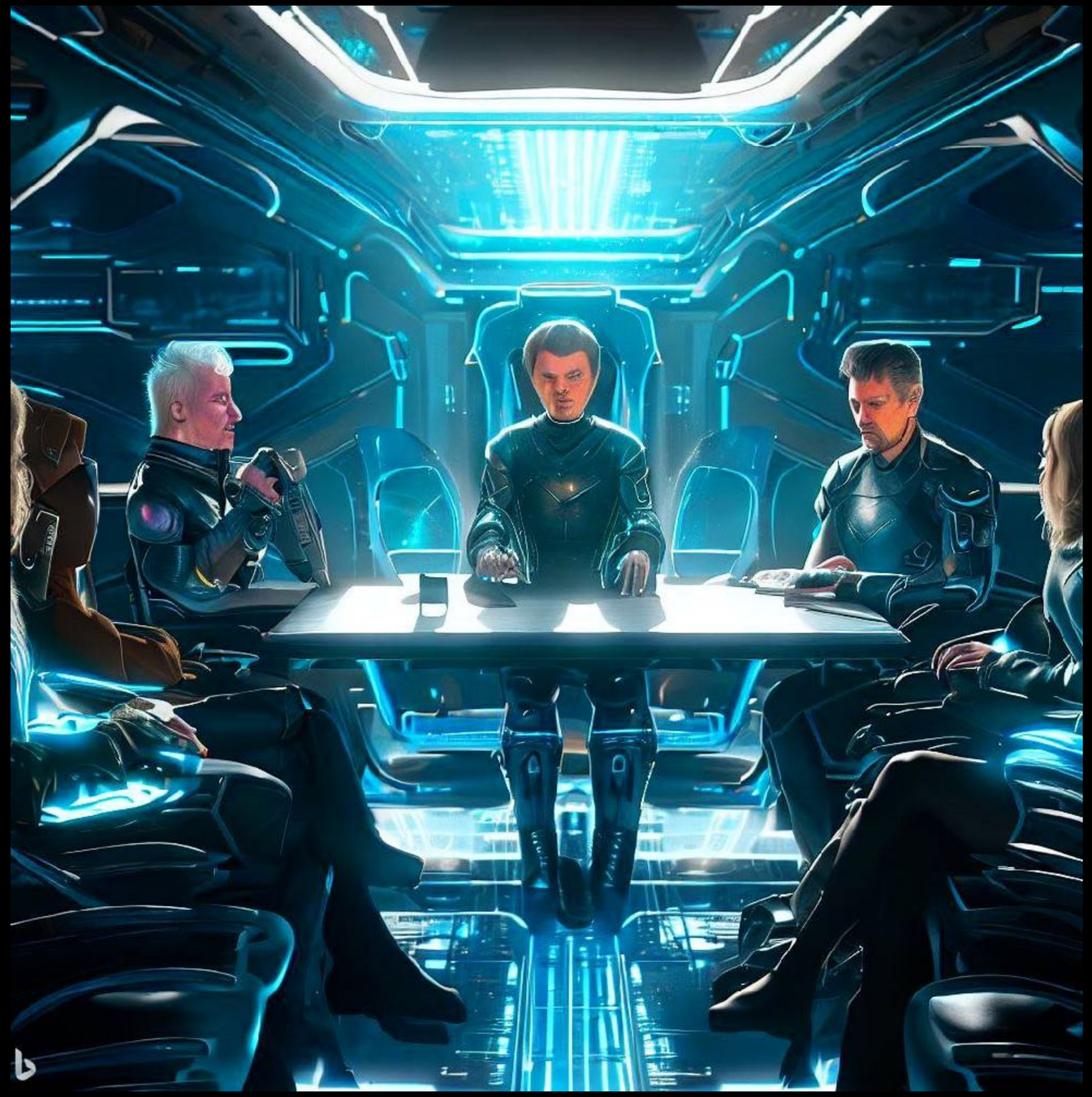
Meeting the Redundants

Day 23 since Start of Mission

Crew and `zpool status`: HEALTHY

Sensors detect a nebula some light years ahead

Captain decides to take a closer look



Sensors report Alien Ship in Nebula

BSDNL25 Captain: ***Hail the alien ship***

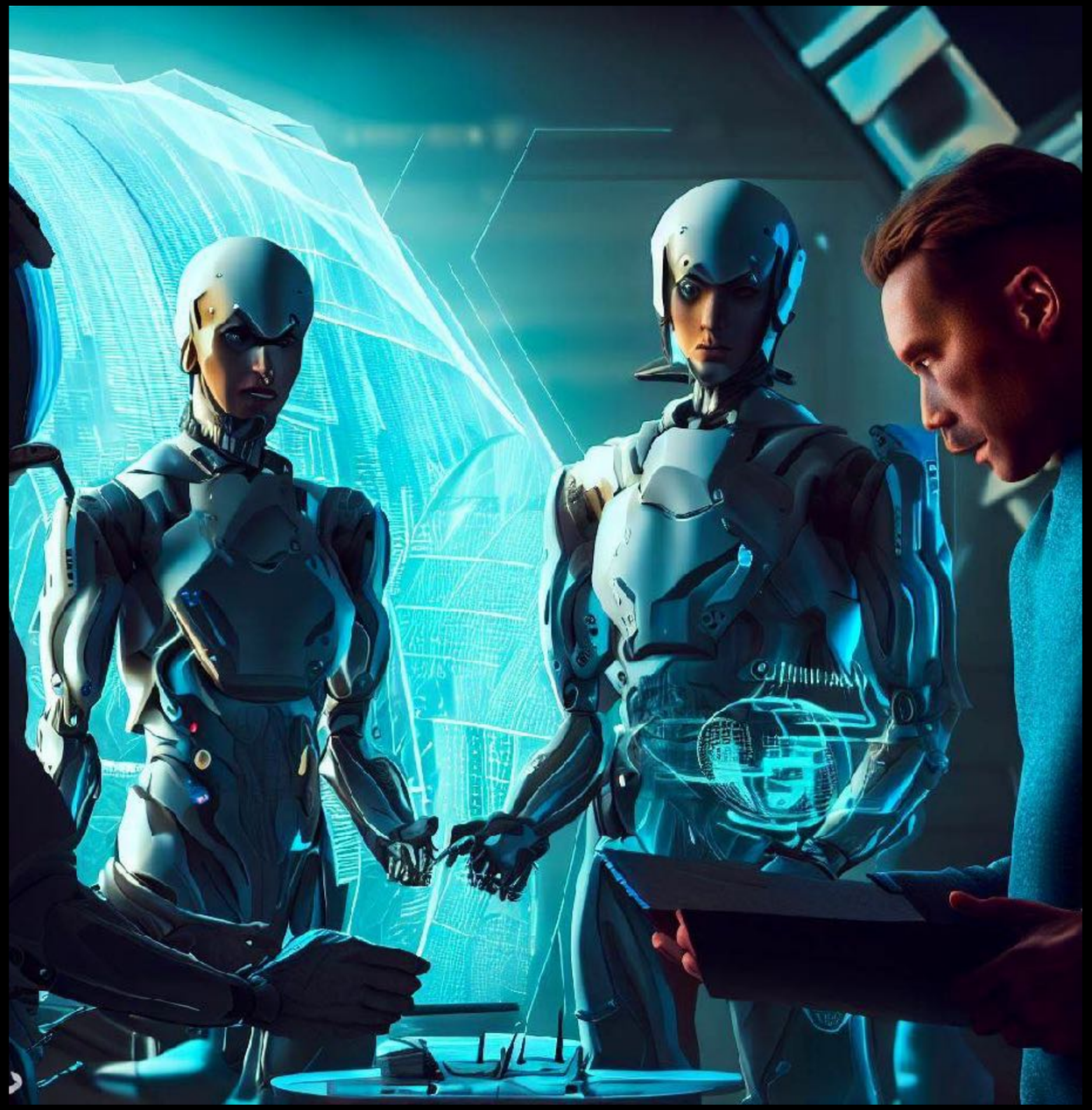
Answer *(2x)*: ***We call ourselves the Redandants, we're on a research mission***

BSDNL25 Captain: ***Send an invitation for dinner on the BSDNL25***

Answer: ***Agreed, agreed***

Touring the BSDNL

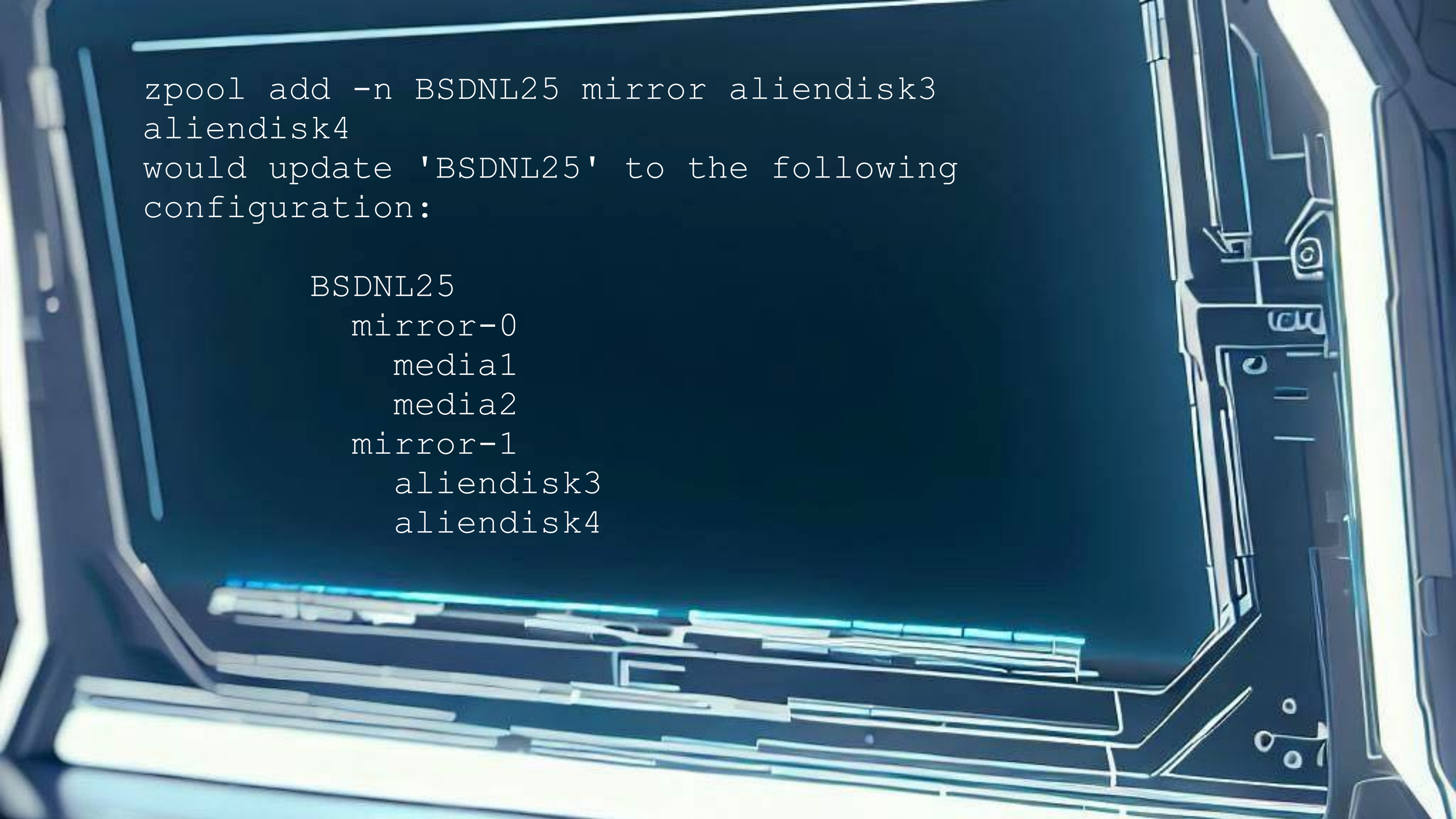
- Redandants are given a tour of the ship
- They suggest extending the pool to RAID10 to increase redundancy
- Offer storage media
- Chief Storage Engineer runs a simulation

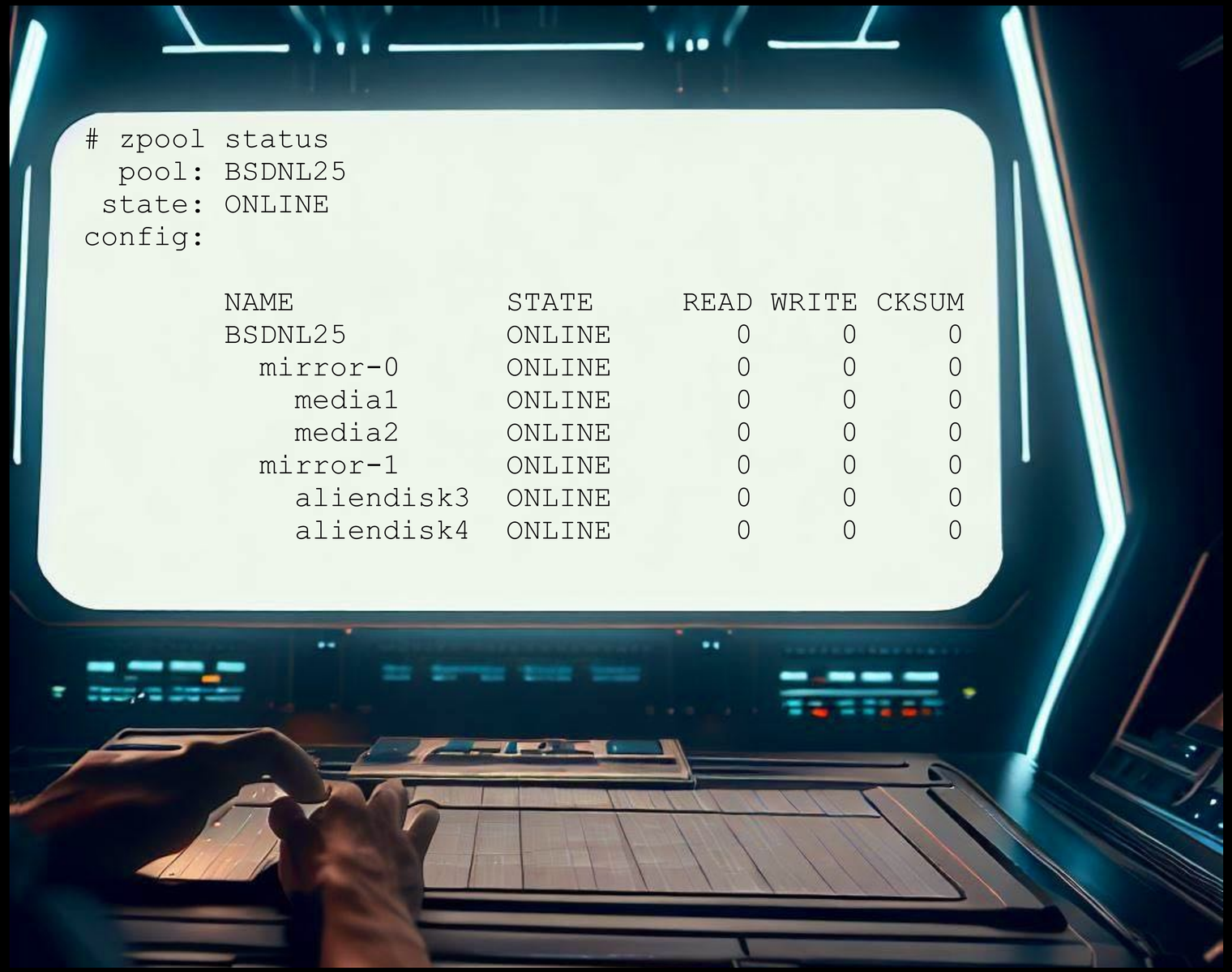


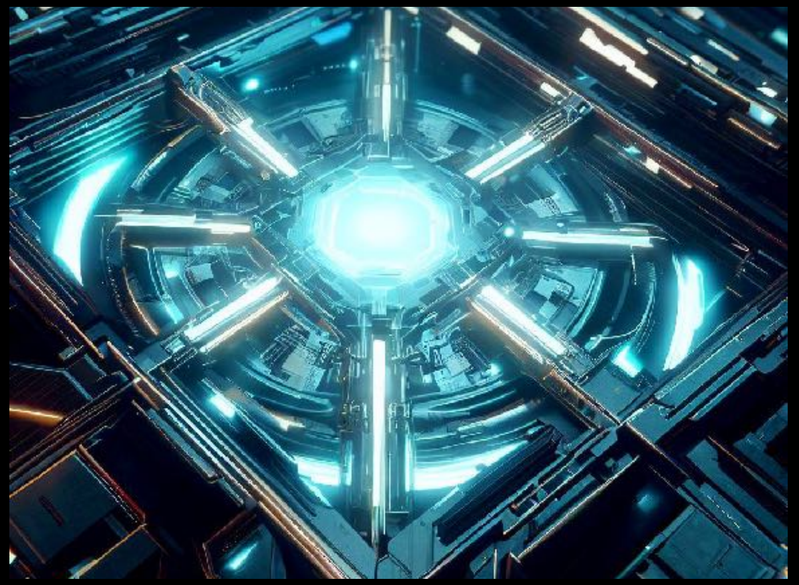
Transforming existing pool into RAID10 without Downtime

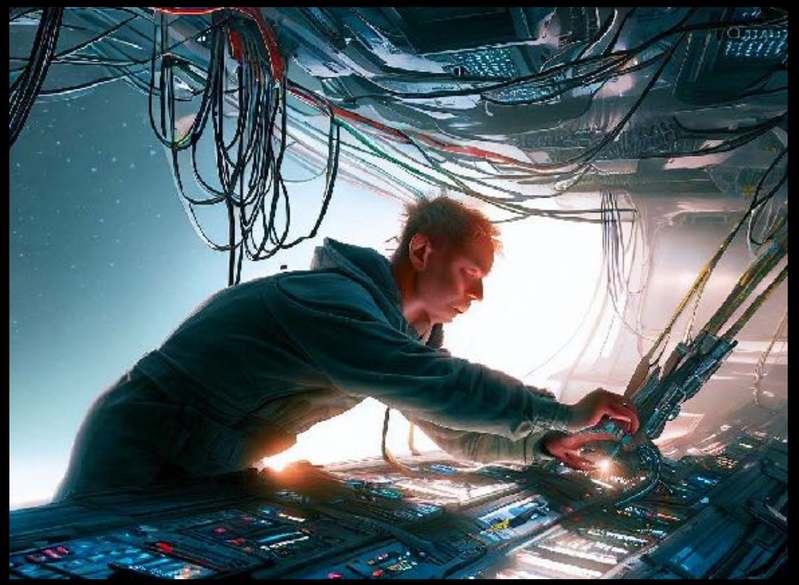
Dry-run simulation (-n):

```
zpool add -n BSDNL25
mirror aliendisk3 aliendisk4
```


zpool add -n BSDNL25 mirror aliendisk3
aliendisk4
would update 'BSDNL25' to the following
configuration:
BSDNL25
mirror-0
media1
media2
mirror-1
aliendisk3
aliendisk4


zpool status
pool: BSDNL25
state: ONLINE
config:
NAME STATE READ WRITE CKSUM
BSDNL25 ONLINE 0 0 0
mirror-0 ONLINE 0 0 0
media1 ONLINE 0 0 0
media2 ONLINE 0 0 0
mirror-1 ONLINE 0 0 0
aliendisk3 ONLINE 0 0 0
aliendisk4 ONLINE 0 0 0

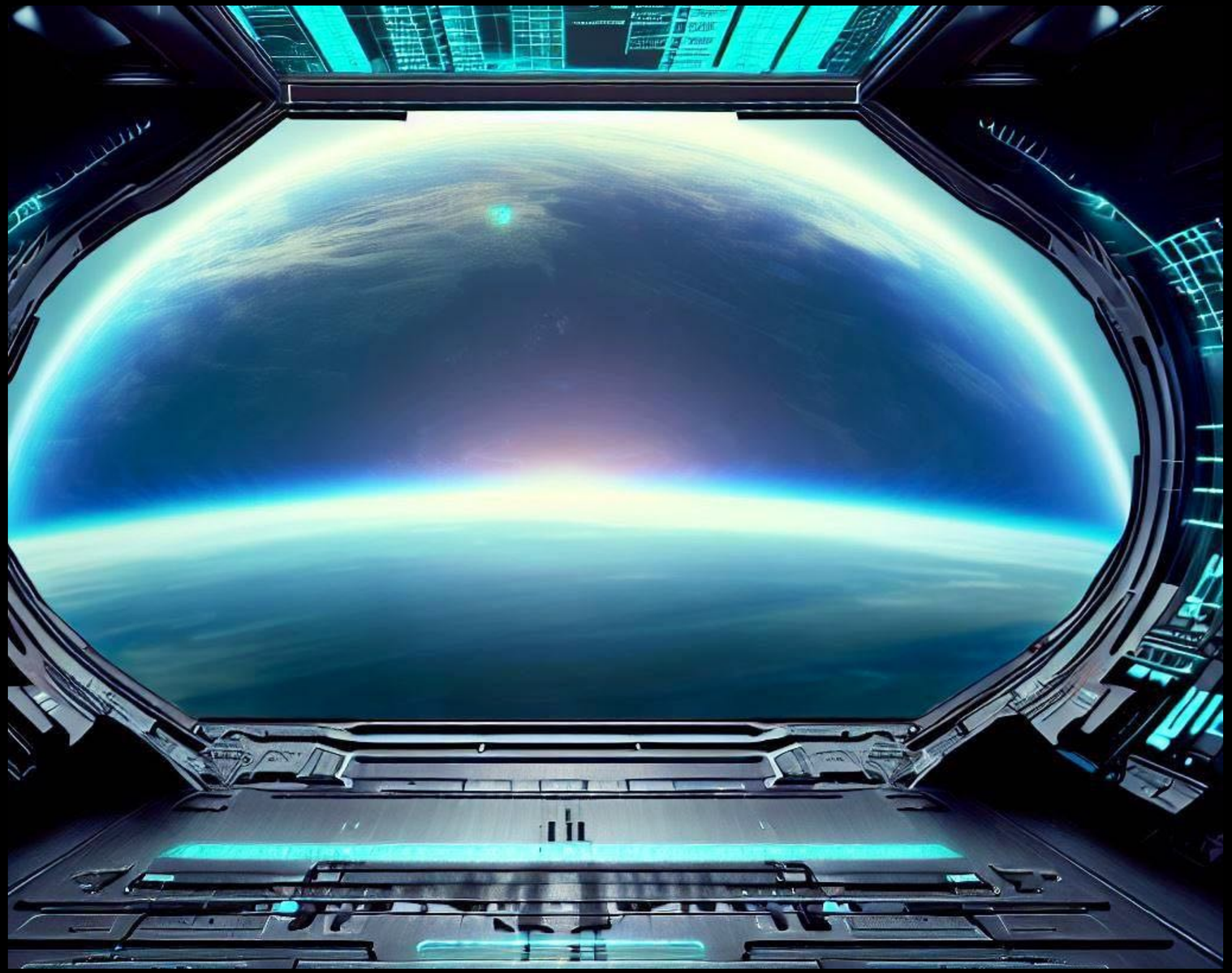




The Journey continues…

Day 42 since Mission did start

Setting course to nearby planet

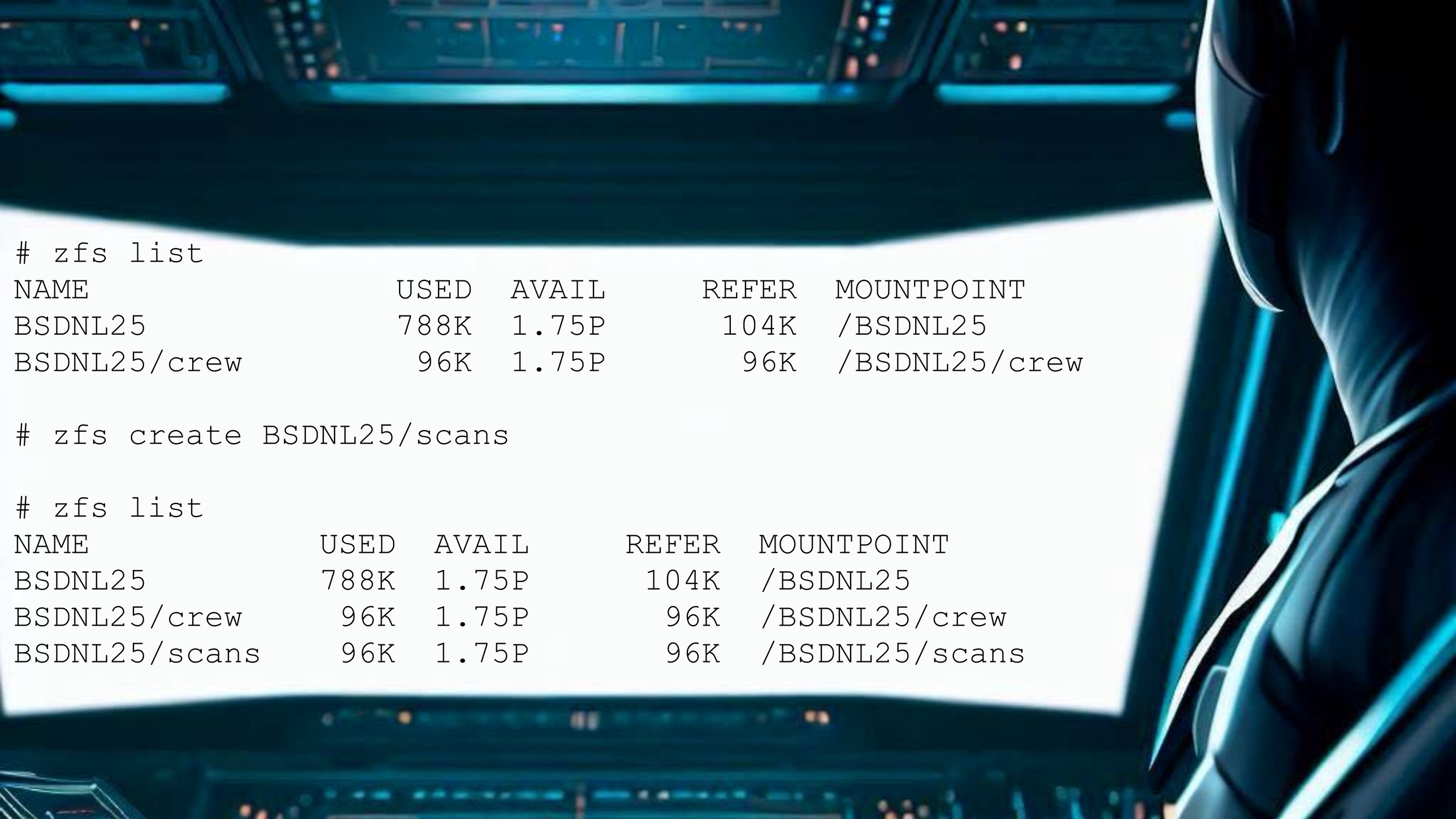


Meeting the Archivers

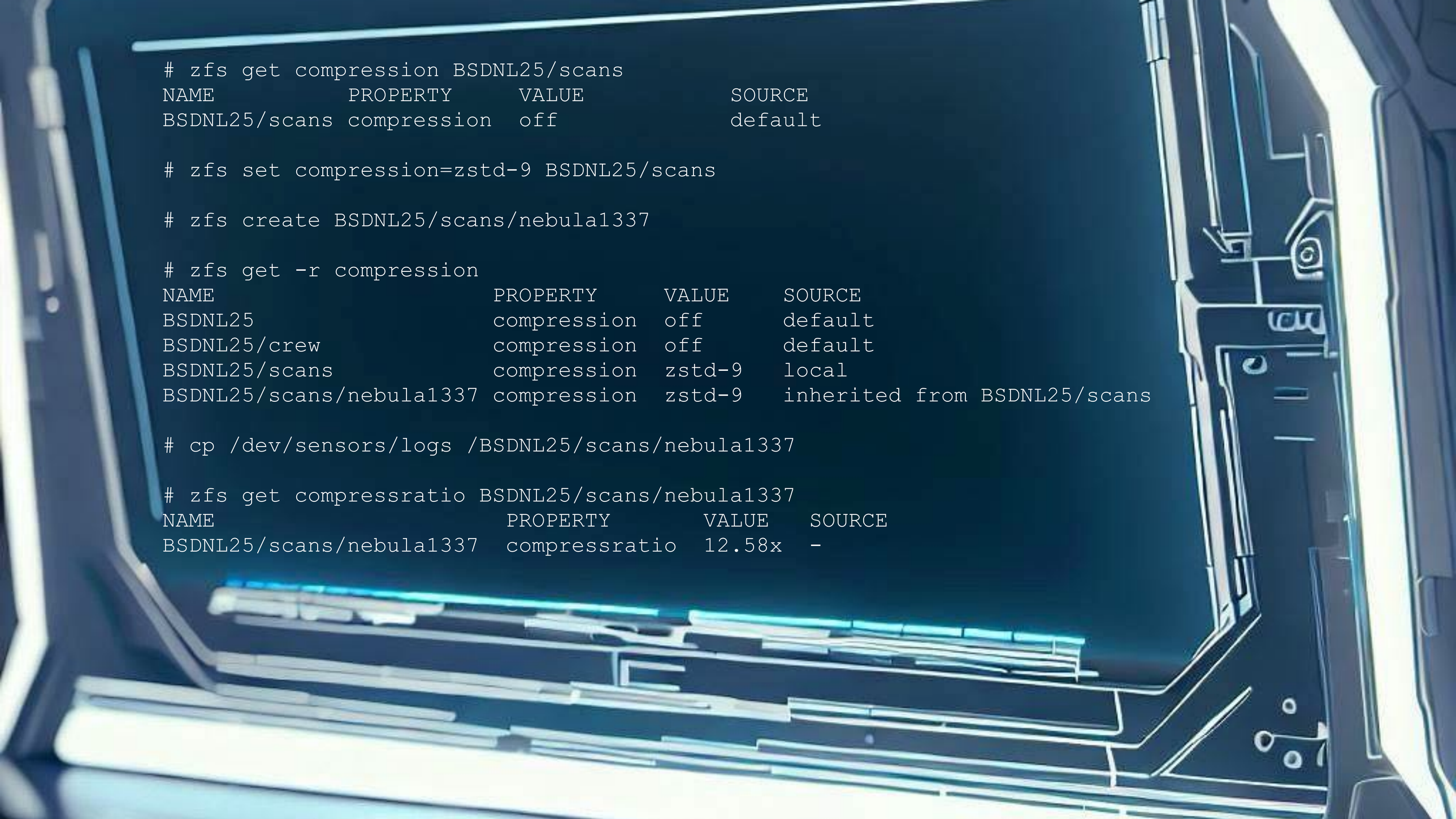
- Peaceful people
- Long and well documented history
- Record everything that happens
- Describe the following problem: storage space is running out
- Do not seem to know the concept of compression

Concept of Datasets

- Logical storage unit similar to partitions
- Use pool storage to store data
- Unlimited amount of datasets possible
- Diverse dataset properties available, i.e. compression
- Child datasets inherit properties of their parent dataset



```
# zfs list
NAME                 USED  AVAIL     REFER  MOUNTPOINT
BSDNL25              788K  1.75P      104K  /BSDNL25
BSDNL25/crew          96K  1.75P        96K  /BSDNL25/crew

# zfs create BSDNL25/scans

# zfs list
NAME                USED  AVAIL     REFER  MOUNTPOINT
BSDNL25             788K  1.75P      104K  /BSDNL25
BSDNL25/crew         96K  1.75P        96K  /BSDNL25/crew
BSDNL25/scans        96K  1.75P        96K  /BSDNL25/scans
```

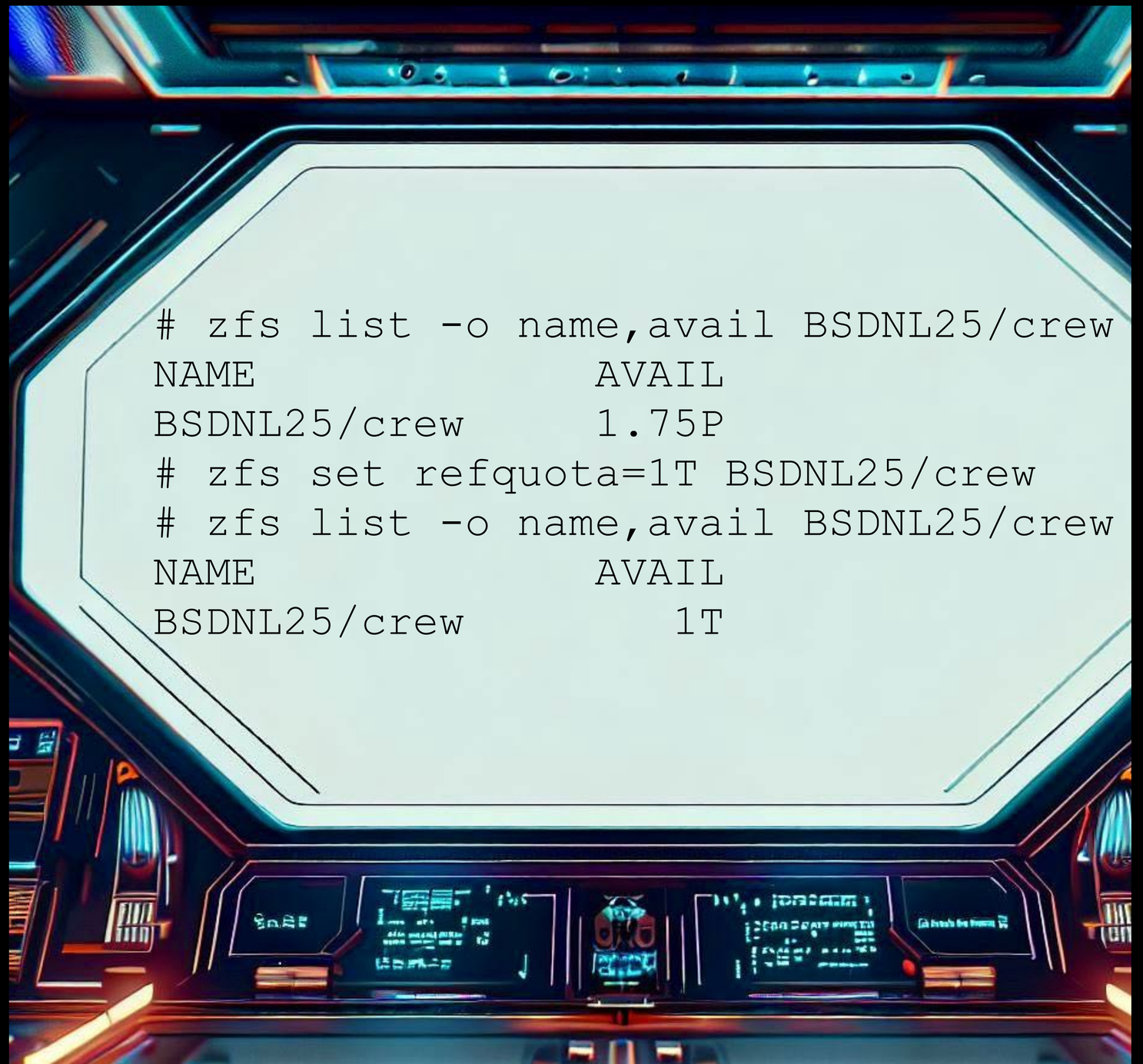

zfs get compression BSDNL25/scans
NAME PROPERTY VALUE SOURCE
BSDNL25/scans compression off default
zfs set compression=zstd-9 BSDNL25/scans
zfs create BSDNL25/scans/nebula1337
zfs get -r compression
NAME PROPERTY VALUE SOURCE
BSDNL25 compression off default
BSDNL25/crew compression off default
BSDNL25/scans compression zstd-9 local
BSDNL25/scans/nebula1337 compression zstd-9 inherited from BSDNL25/scans
cp /dev/sensors/logs /BSDNL25/scans/nebula1337
zfs get compressratio BSDNL25/scans/nebula1337
NAME PROPERTY VALUE SOURCE
BSDNL25/scans/nebula1337 compressratio 12.58x -

Quota and Reservation

- Archivers are impressed
- Chief Storage Engineer: *Would you like to know more?*
- Archivers: *How do you prevent datasets eating up all available pool space?*
- Concepts like quota and reservation are explained
- Mini trainings held on the BSDNL

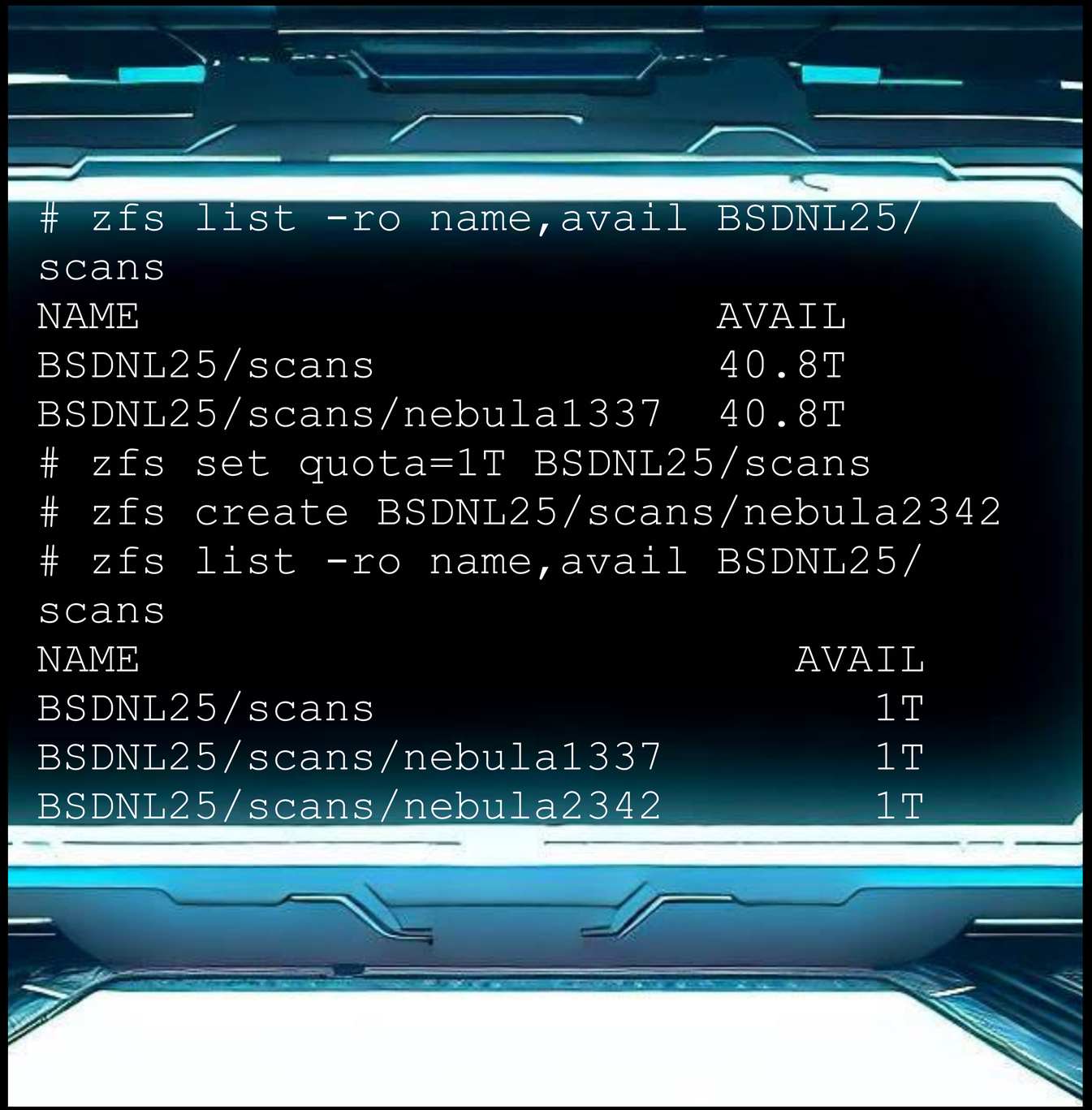
Quota

- Another property setting of datasets
- Limits how much data a dataset may store
- Pool prevents further allocations when quota is reached



Quotas for Child Datasets

- `zfs set refquota` refers only to this dataset
- `zfs set quota` refers to the specified dataset *and* all child datasets
- In the latter case all Datasets share the quota



Reservation

Disk Space Guarantees

- Reservations ensure a certain amount of free disk space to dataset
- Reserved space is immediately reduced from pool space
- Counterpart to quotas
- Same here: `reservation` refers to datasets including any children
- `refreservation` does only apply to this dataset, none below

```
# zfs list
NAME                USED  AVAIL     REFER  MOUNTPOINT
BSDNL25             788T  1.75P      104T  /BSDNL25
BSDNL25/crew         96T  1.75P        96T  /BSDNL25/crew
…

# zfs set reservation=1P BSDNL25/crew

# zfs list
NAME                USED  AVAIL     REFER  MOUNTPOINT
BSDNL25             788T  0.75P      104T  /BSDNL25
BSDNL25/crew         96T  0.75P        96T  /BSDNL25/crew
…
```

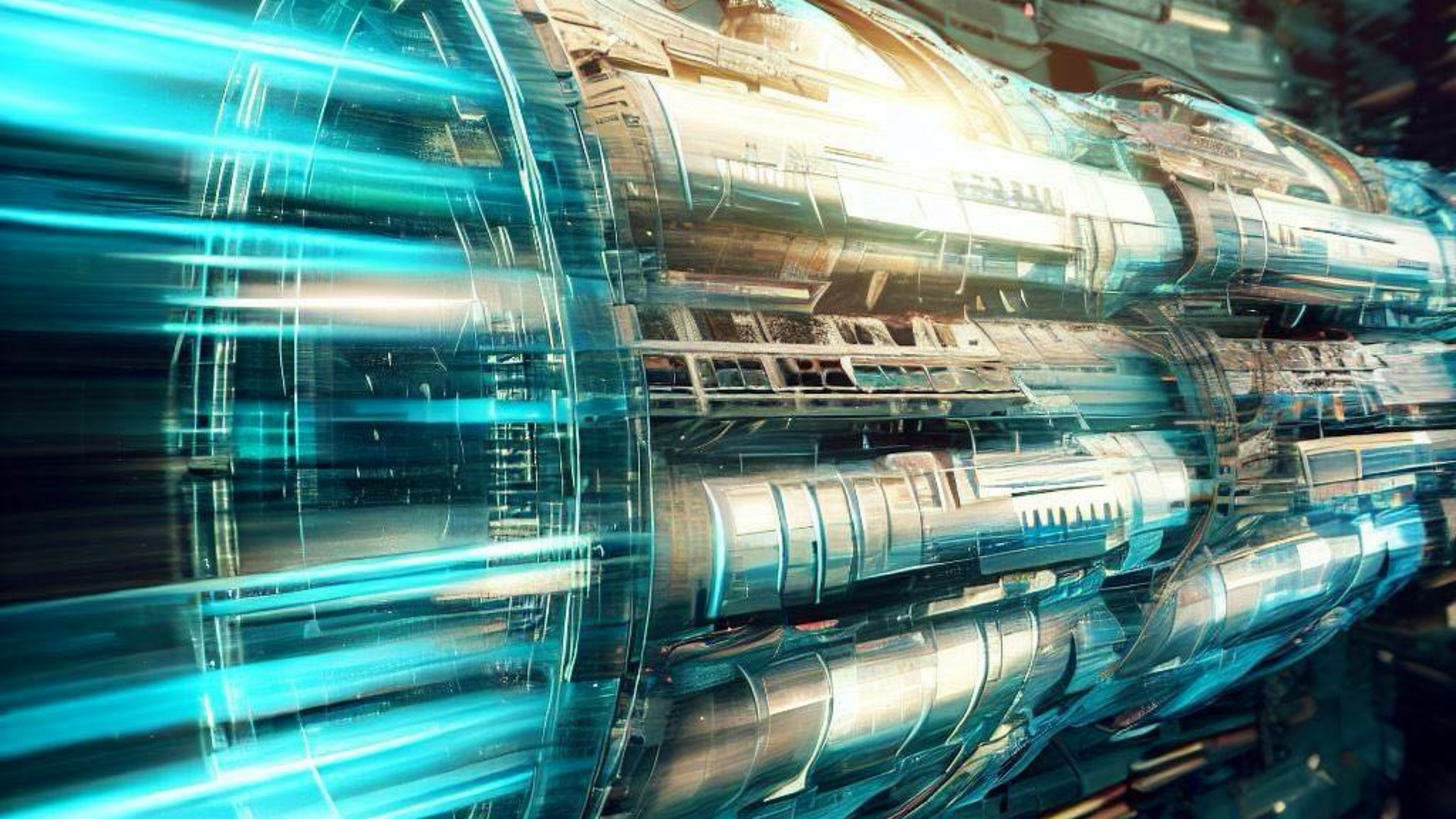
Technology Transfer

Necessary information given to Archivers

Compression reduces archives significantly
(copying to new dataset required)

Quota and reservation prevent running into shortages

BSDNL-Crew sets new course to other star systems



Time passes slowly…

… until one day…

Day 67 since Beginning of Mission

- INTERCOM: *Bridge to Chief Storage Engineer*
- Chief Storage Engineer: *Is there a problem?*
- Bridge: *You better see for yourself…*


CATS: ALL YOUR FILES ARE BELONG
TO US.


SFCYPAILTG HPOBOTTYIG
JOGE SBYE!,
SSTTPTARIIURT YOCR ESIMS.

Attack of the Crypto-Trojanides

- BSDNL Main Computer got encrypted!
- Without decryption key the ship is stranded in space
- Trojanides demand 23.000 Coins as ransom
- Give 24 hours time to think about it

Snapshots to the Rescue

- Snapshots are read only images of the state of a dataset at specific time
- Allows point-in-time recovery
- BSDNL25 pool creates such regular snapshots since leaving space dock
- Rollback to a snapshot from before the encryption happened is possible

```
# zfs list -t snapshot BSDNL25
NAME                                 USED  AVAIL     REFER  MOUNTPOINT
BSDNL@auto_2300-20-07-00:10           12T      -       12T  -
BSDNL@auto_2300-21-07-00:10            7G      -      112T  -
BSDNL@auto_2300-22-07-00:10          227G      -      124T  -
BSDNL@auto_2300-23-07-00:10           13M      -       36T  -

# zfs rollback BSDNL25@auto_2300-23-07-00:10

# ls /BSDNL25/scans/nebula1337
…

# zfs snapshot -r BSDNL25@systems_clean

#zfs list -t snapshot BSDNL25@systems_clean
NAME                        USED  AVAIL     REFER  MOUNTPOINT
BSDNL25@systems_clean        0B      -      112K  -
```

Ship Systems restored

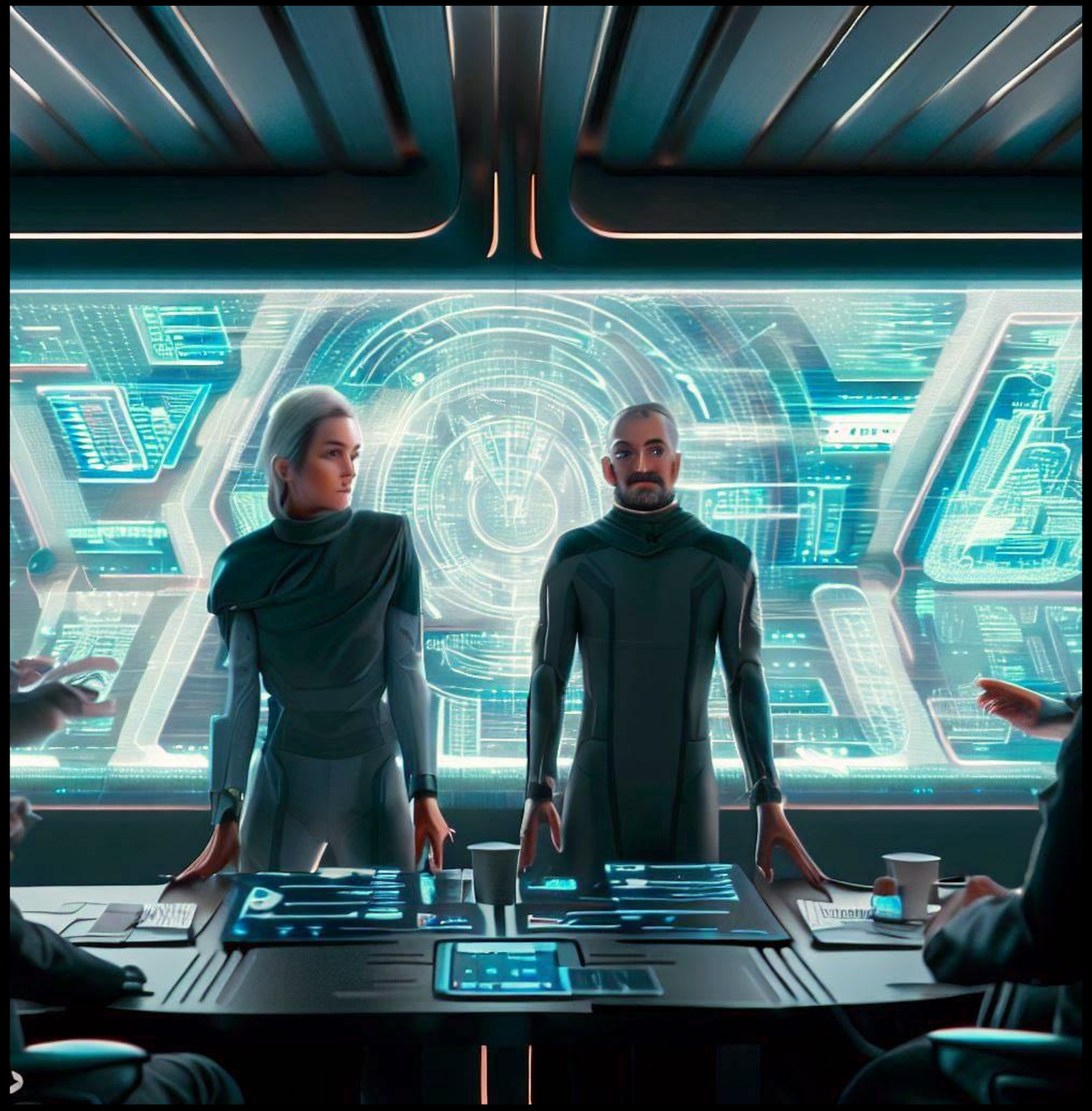
Returning to normal operation

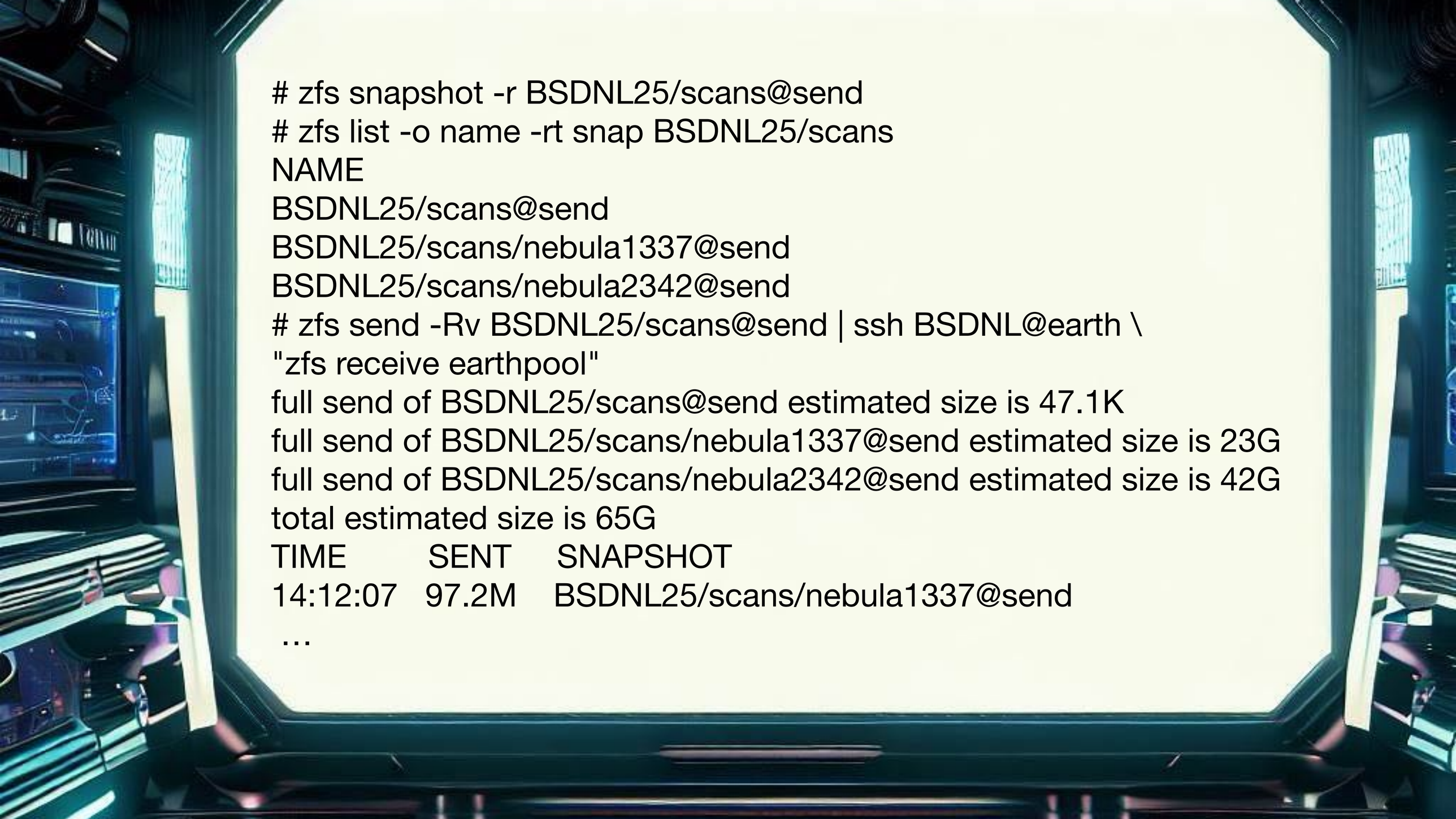
- Successful rollback to last snapshot before encryption
- BSDNL Captain: *Take us away from here, quickly…*
- Noone seems to follow
- Captain orders a full systems check

Sending Data back to Earth

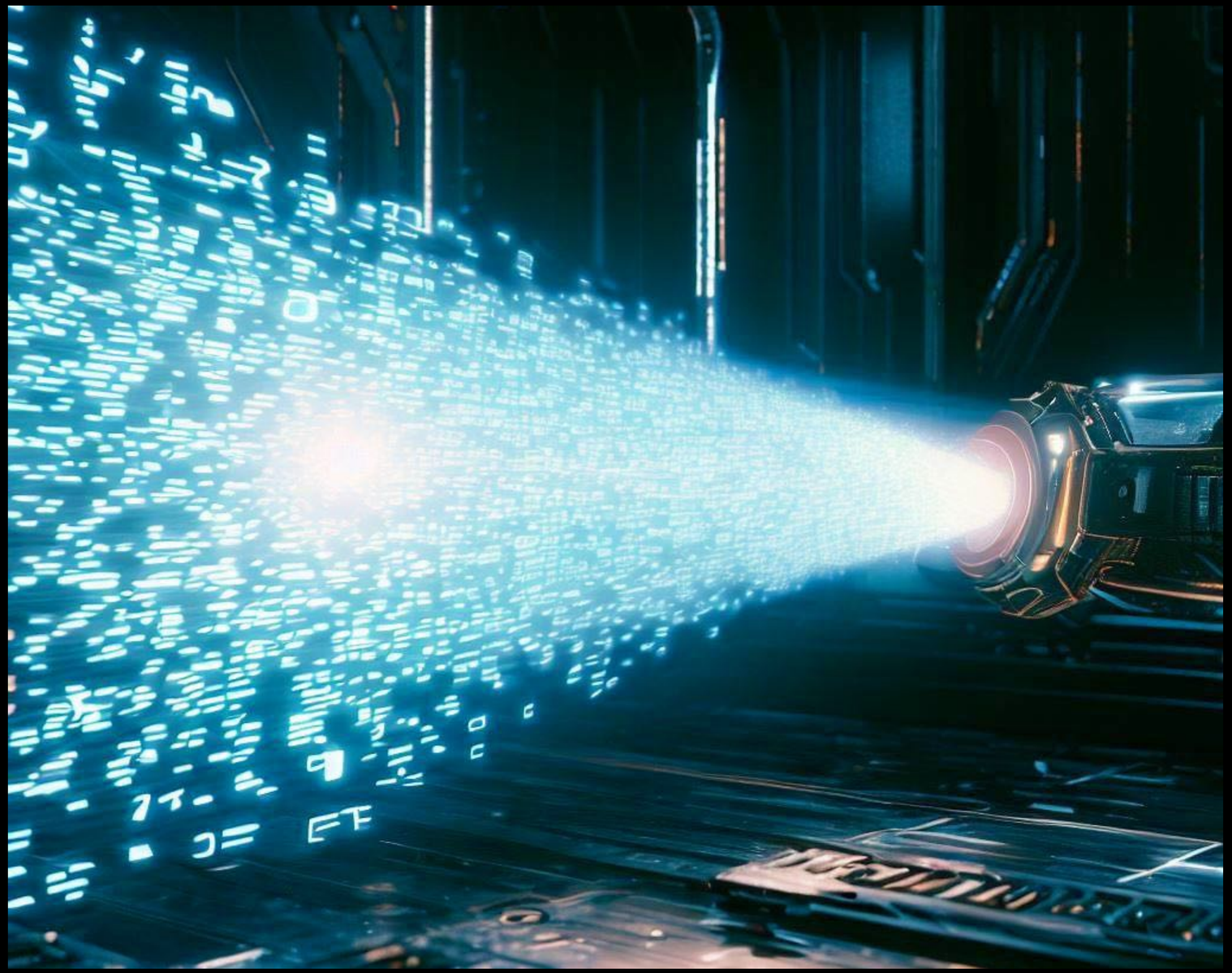
Day 100 since Mission Start

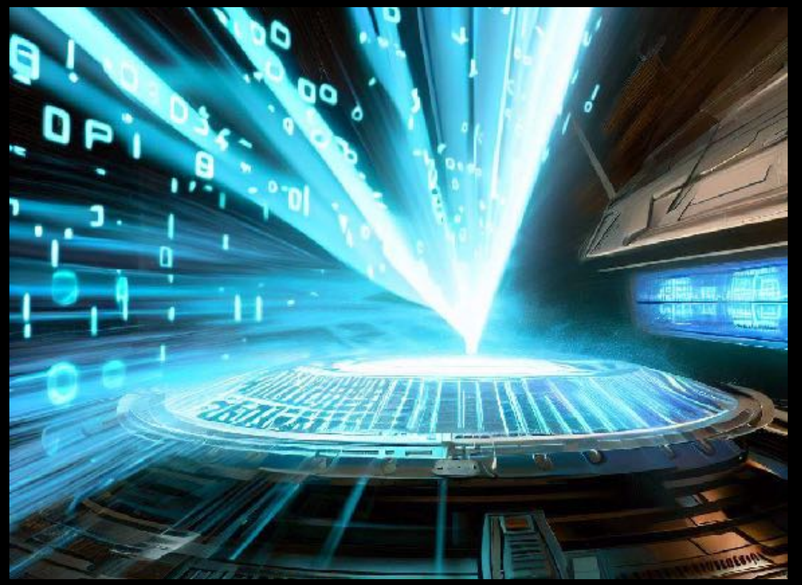
- Earth should receive regular backups of the research data
- ZFS supports delta replication of snapshots
- Piped through secure channel protects data from snoopers





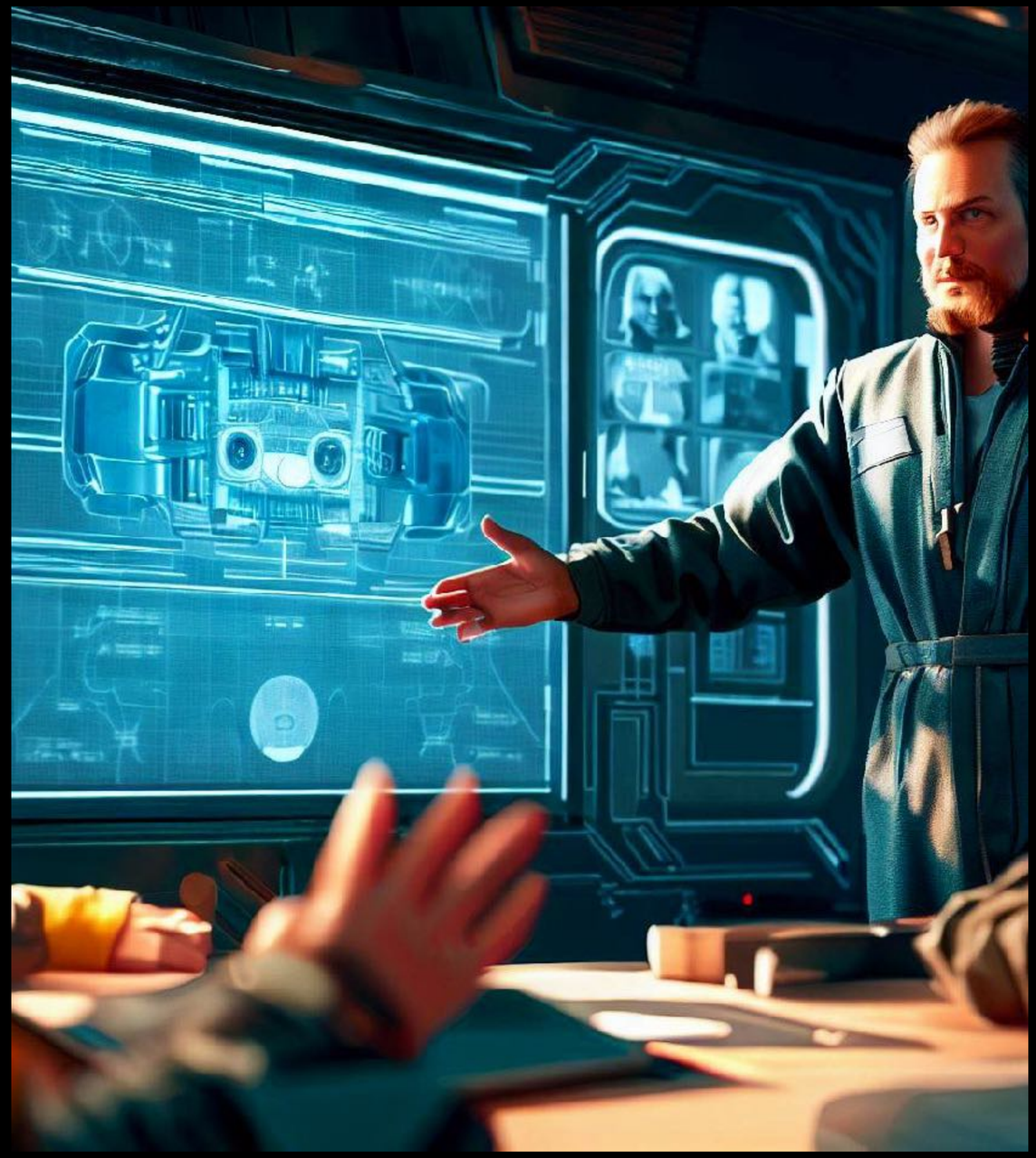
```
# zfs snapshot -r BSDNL25/scans@send
# zfs list -o name -rt snap BSDNL25/scans
NAME
BSDNL25/scans@send
BSDNL25/scans/nebula1337@send
BSDNL25/scans/nebula2342@send
# zfs send -Rv BSDNL25/scans@send | ssh BSDNL@earth \
"zfs receive earthpool"
full send of BSDNL25/scans@send estimated size is 47.1K
full send of BSDNL25/scans/nebula1337@send estimated size is 23G
full send of BSDNL25/scans/nebula2342@send estimated size is 42G
total estimated size is 65G
TIME         SENT     SNAPSHOT
14:12:07   97.2M    BSDNL25/scans/nebula1337@send
 ...
```





The Journey is still going on

- There is much more to discover
- Takeaway: OpenZFS has a lot to offer, for free
- No science fiction tech: it’s available today
- Tried, stable, reliable, trustworthy
- www.openzfs.org




Thank you!