

(Disk) Space - the final frontier?

**Eine visuelle Reise zu den Sternen
mit OpenZFS**

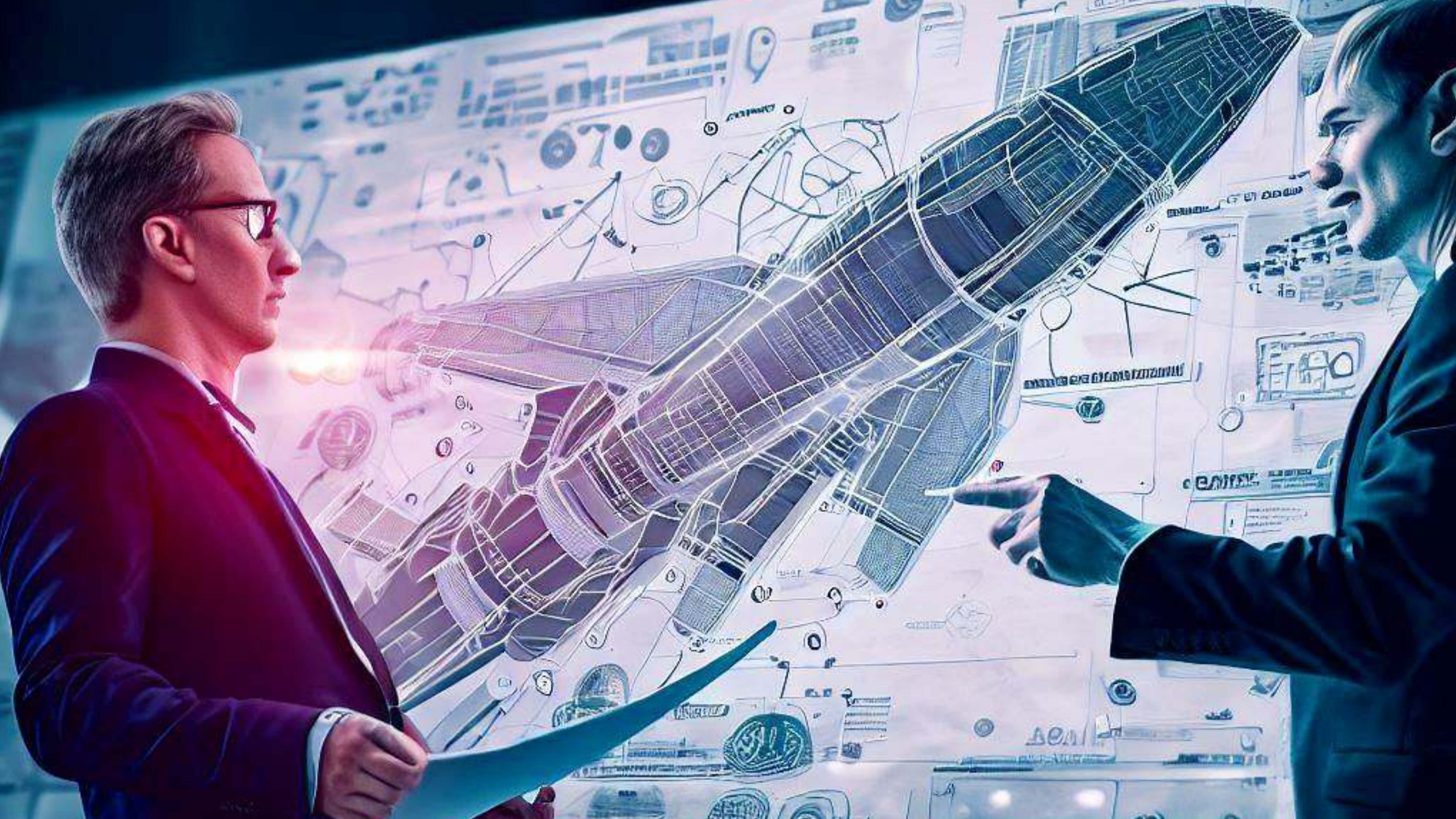
**Benedict Reuschling
benedict@reuschling.org
FSCK Backnang 2024**

Warum dieser Talk?

- Sonst eher trockene Tutorials
- Wenige Illustrationen, keine Story
- Diesmal: genau umgekehrt
- DALL-E Bildgenerator (Bing)
- Unterhaltsamer Ansatz, OpenZFS zu vermitteln
- Experiment

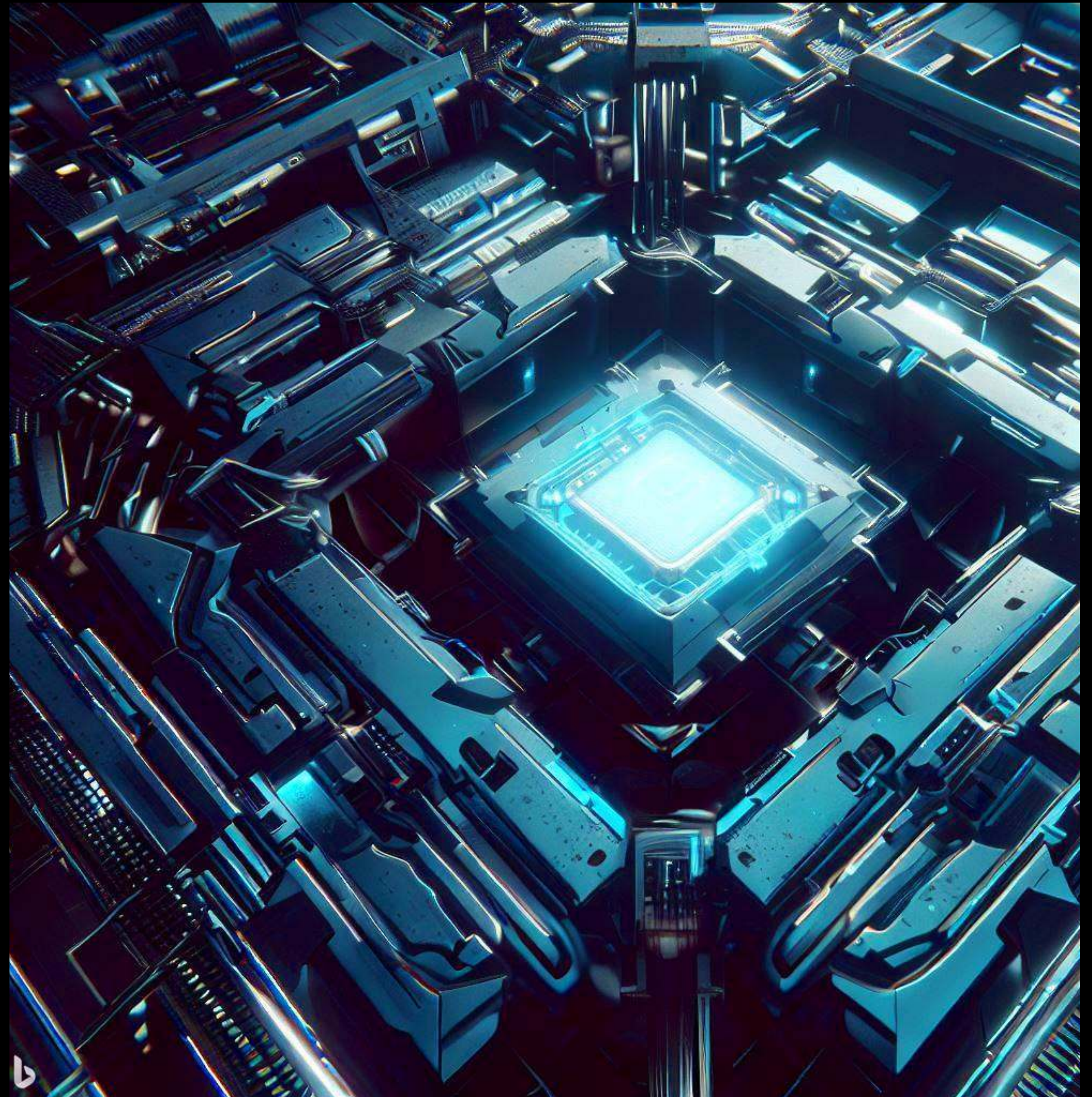


Der grosse Plan

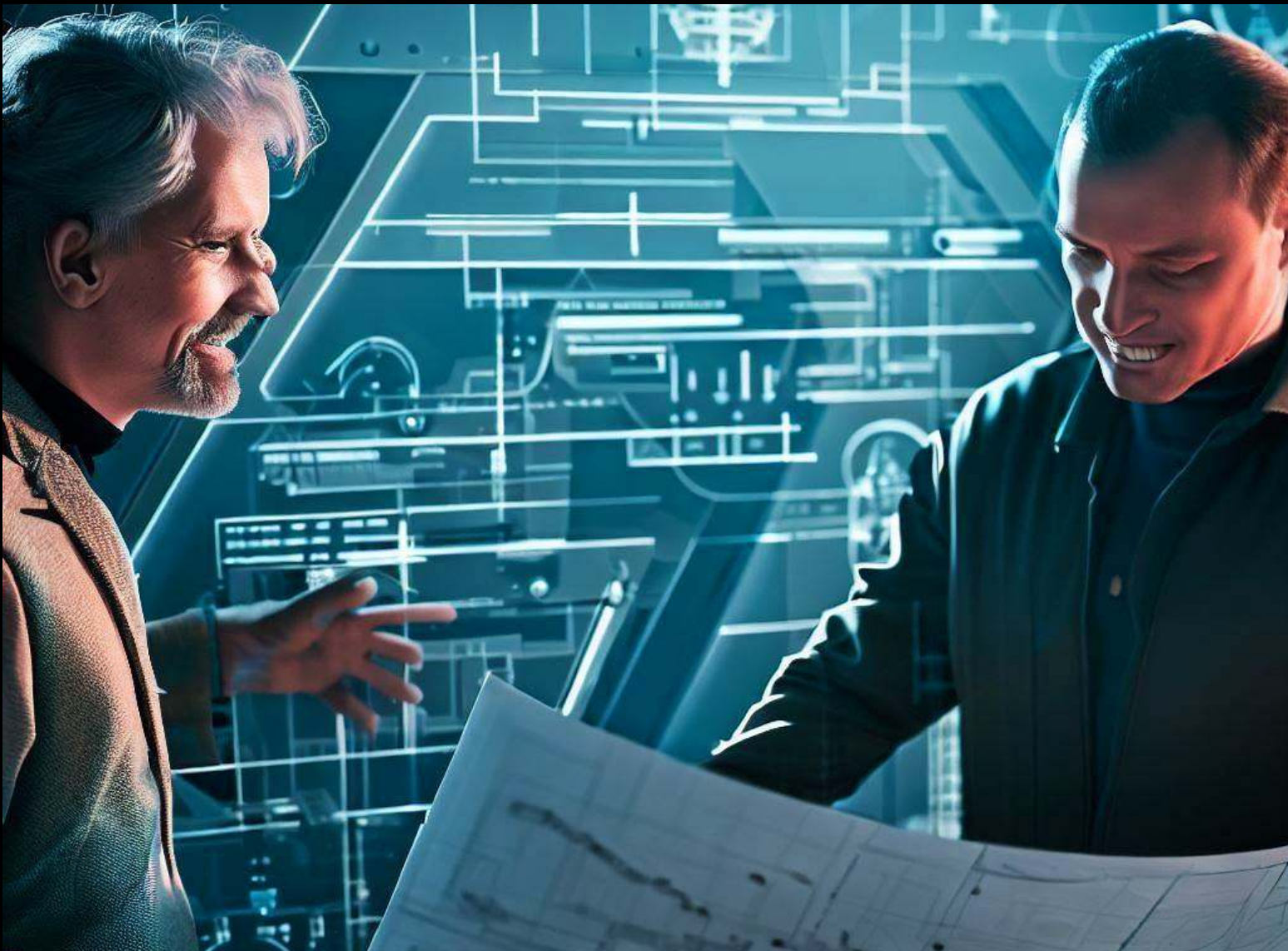


Reise zu den Sternen

- Ziel: Erforschung der Galaxie und Sammeln von Daten
- Schiffcomputer zentrales Element
- Ingenieure wählten OpenZFS Storagepool als Speicher
- Standard-Besatzung inkl. Chief Storage Engineer



**Und so machten sich die
Ingenieure ans Werk...**



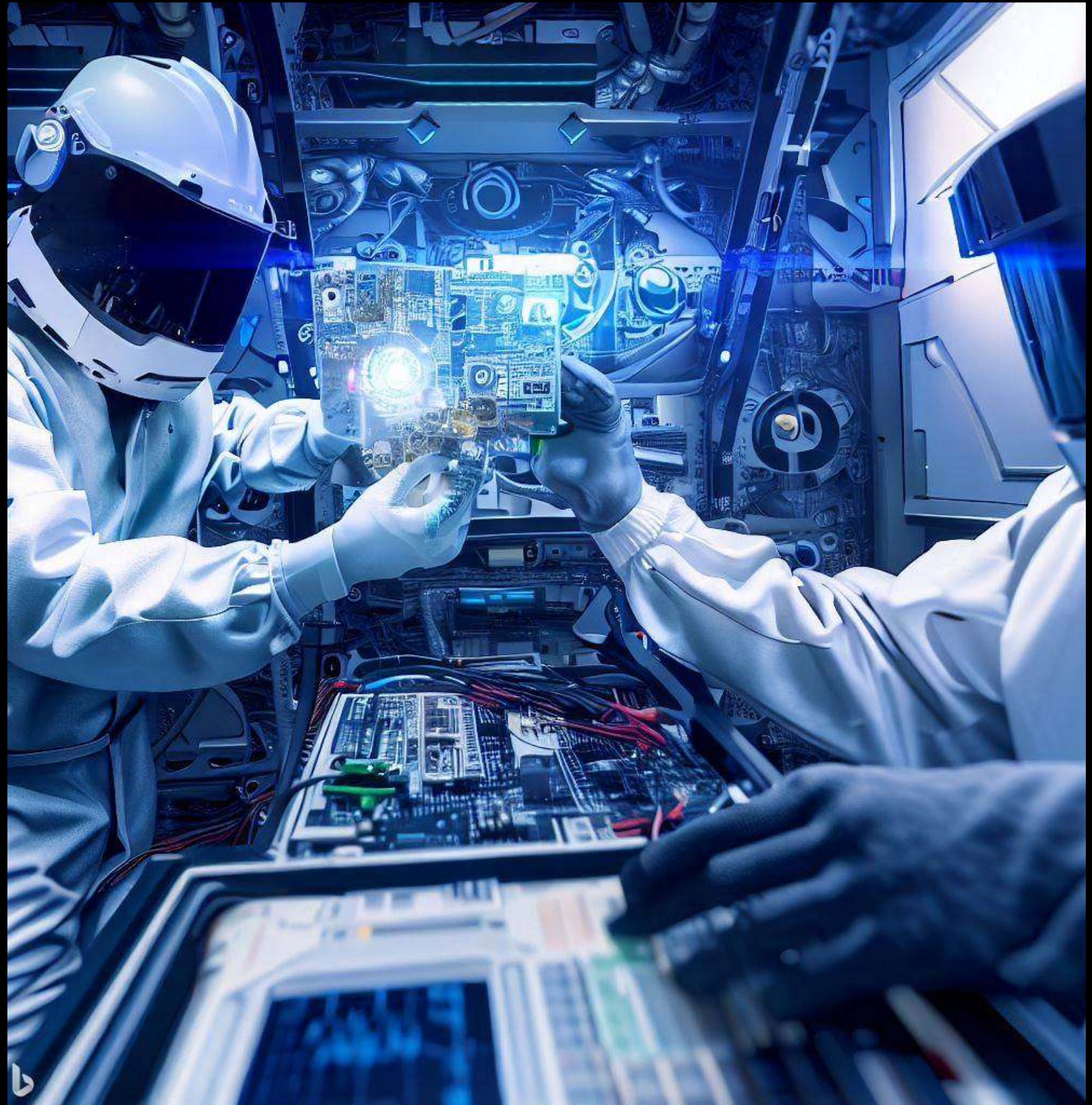
Raumschiff im Bau



Initialisiere Raumschiff- Speicher...

```
# zpool create FSCK24  
mirror medium1 medium2
```

Speicher sofort im System
verfügbar, gemountet, bereit
für I/O



Beginn der Reise

- Besatzung geht an Bord
- Abschied von der Erde
- Initialer Check der Systeme



Statuscheck des Storage-Pools:

zpool status und zpool list

zpool status

pool: FSCK24
state: ONLINE
scan:

config:

NAME	STATE	READ	WRITE	CKSUM
FSCK24	ONLINE	0	0	0
mirror	ONLINE	0	0	0
medium1	ONLINE	0	0	0
medium2	ONLINE	0	0	0

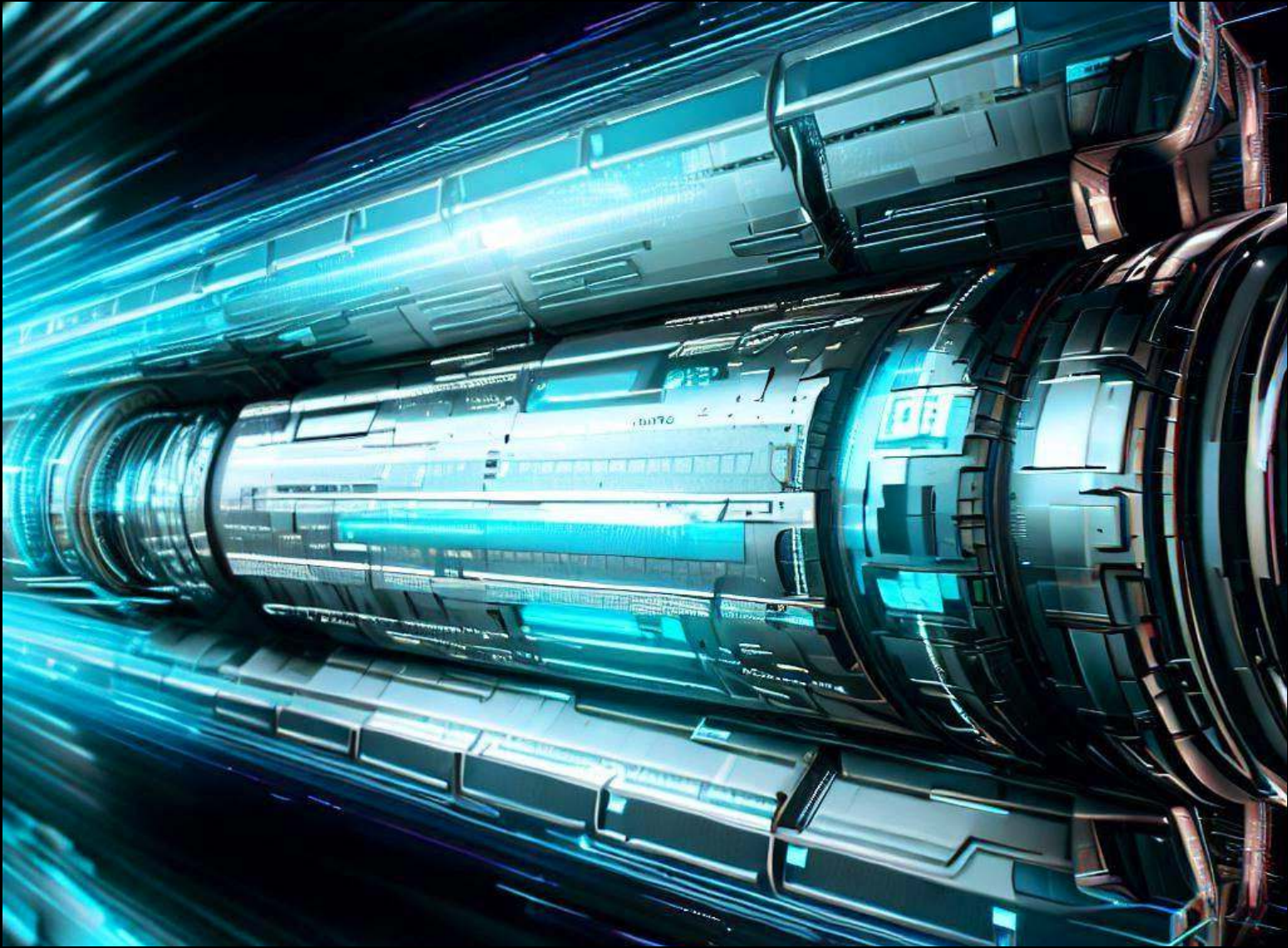
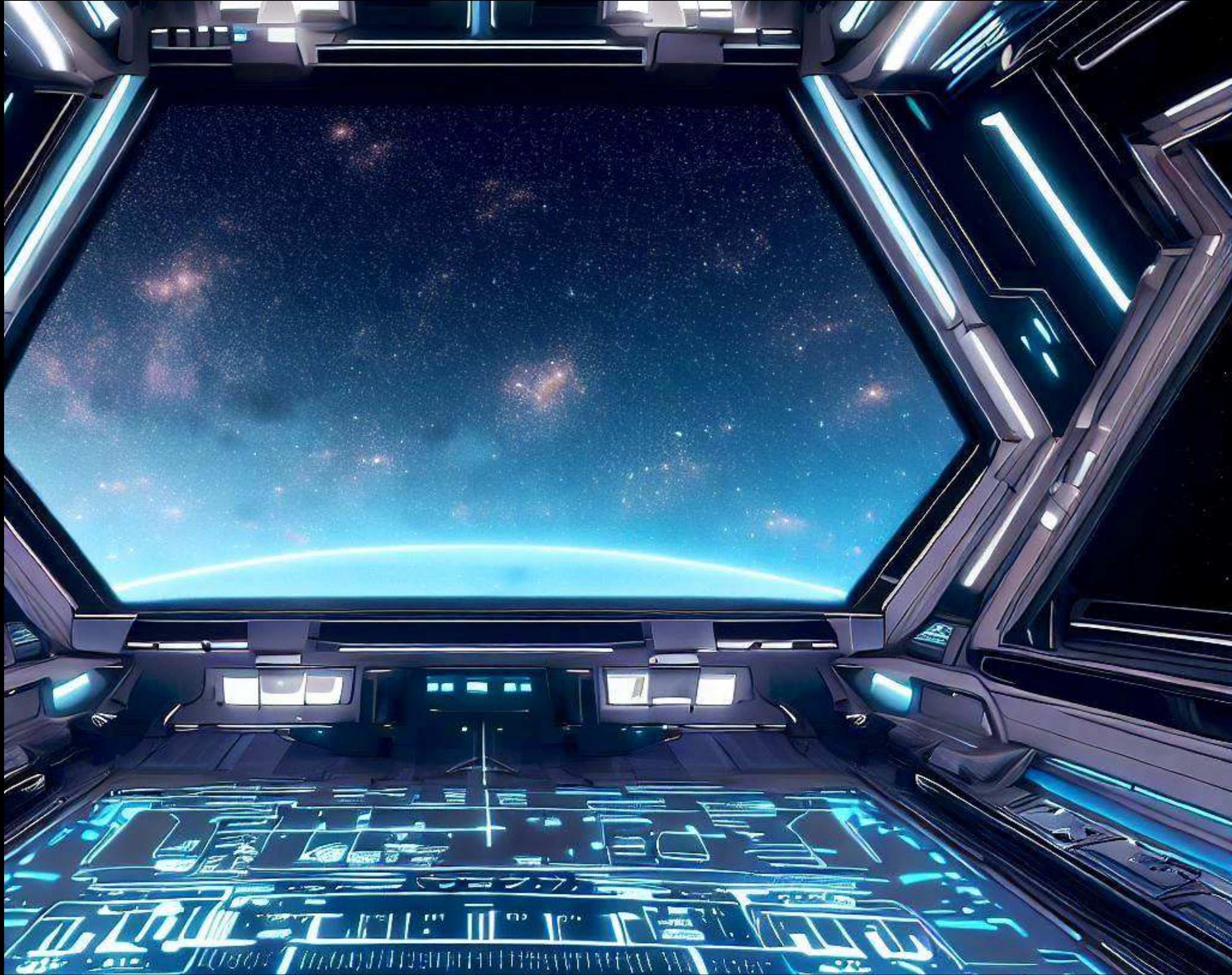
zpool list

NAME	SIZE	ALLOC	FREE	CKPOINT	EXPANDSZ	FRAG	CAP	DEDUP	HEALTH	ALTROOT
FSCK24	880T	15.3G	865T	-	-	0%	1%	1.00x	ONLINE	-



All systems go...





Begegnung mit den Redundanten

Tag 23 seit Missionsbeginn

Crew und zpool status:
HEALTHY

Sensoren erfassen einen
Nebel ein paar Lichtjahre
voraus

Captain beschliesst, diesen
zu untersuchen







Sensoren melden fremdes Schiff

FSCK Captain: *Grussbotschaft
senden*

Antwort: *Wir sind vom Volk der
Redundanten, auf Forschungsmission*

FSCK Captain: *Einladung zum
Abendessen auf der FSCK schicken*

Antwort: *Gerne, gerne*

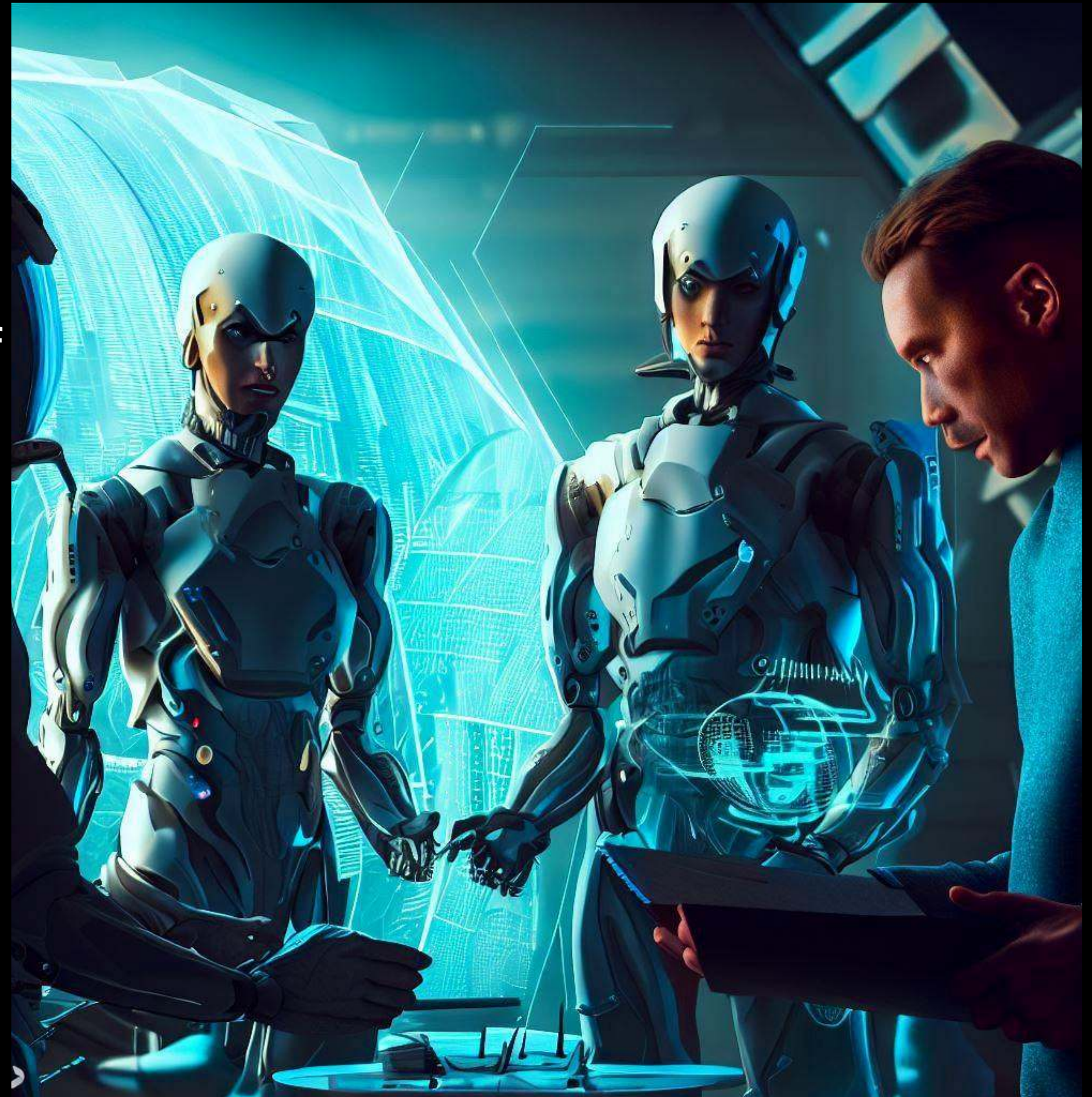




Schiffstour

Besichtigung der FSCK

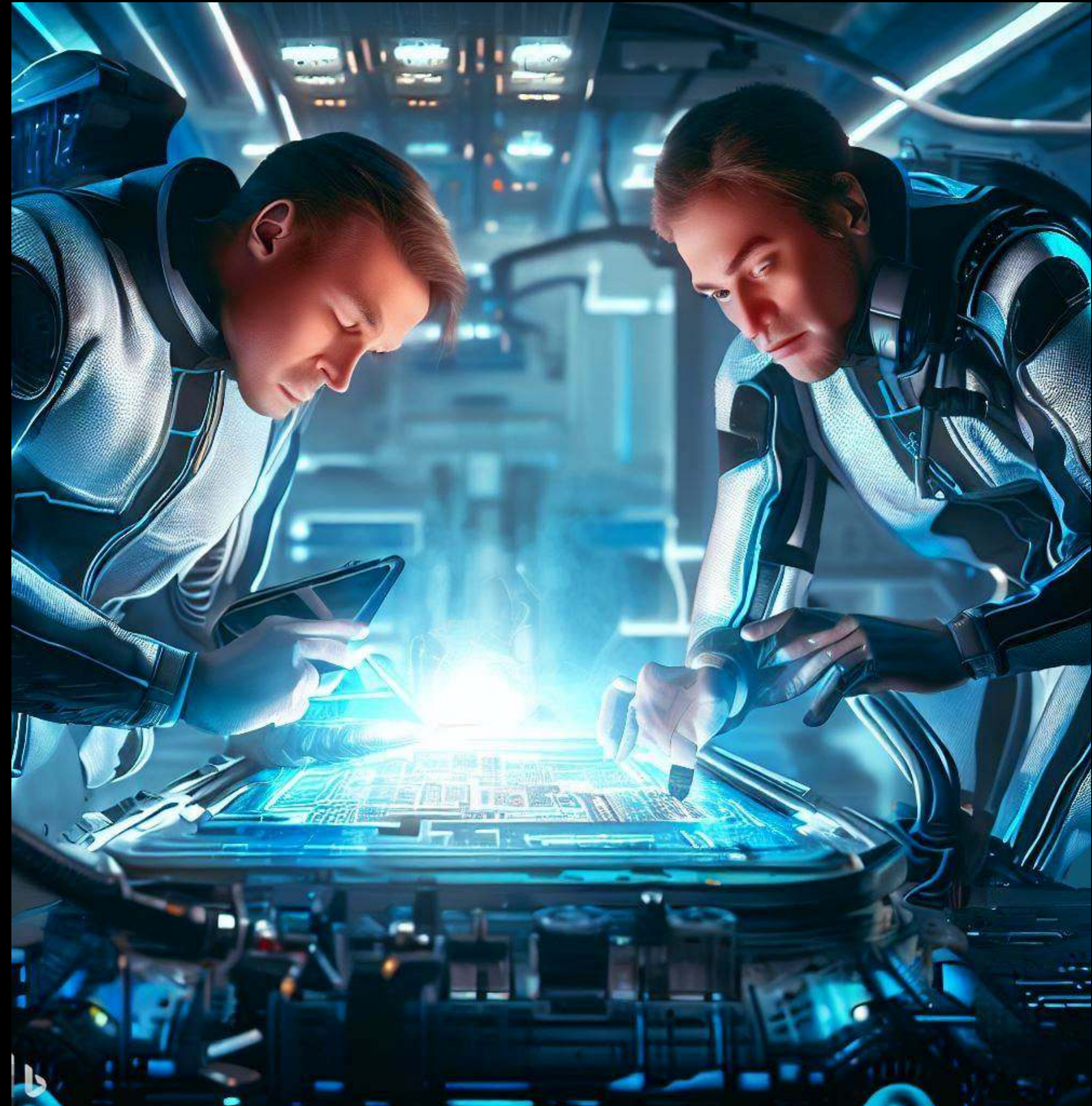
- Redundanten besichtigen das Schiff
- Schlagen vor, Pool in RAID10 zu erweitern zur Erhöhung der Ausfallsicherheit
- Bieten Speichermedium an
- Chief Storage Engineer schlägt Simulation vor



Erweiterung des Storage-Pools auf RAID10

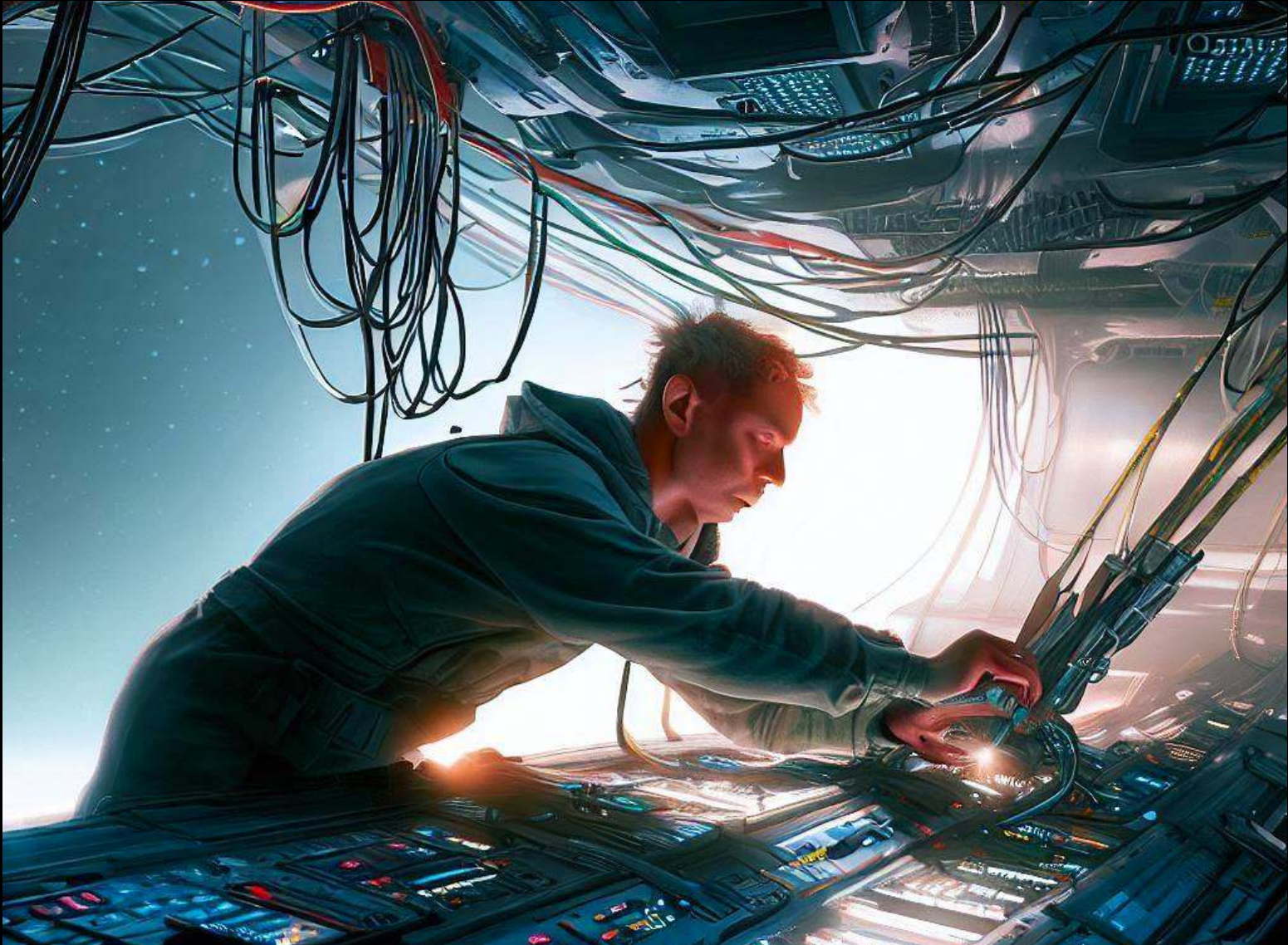
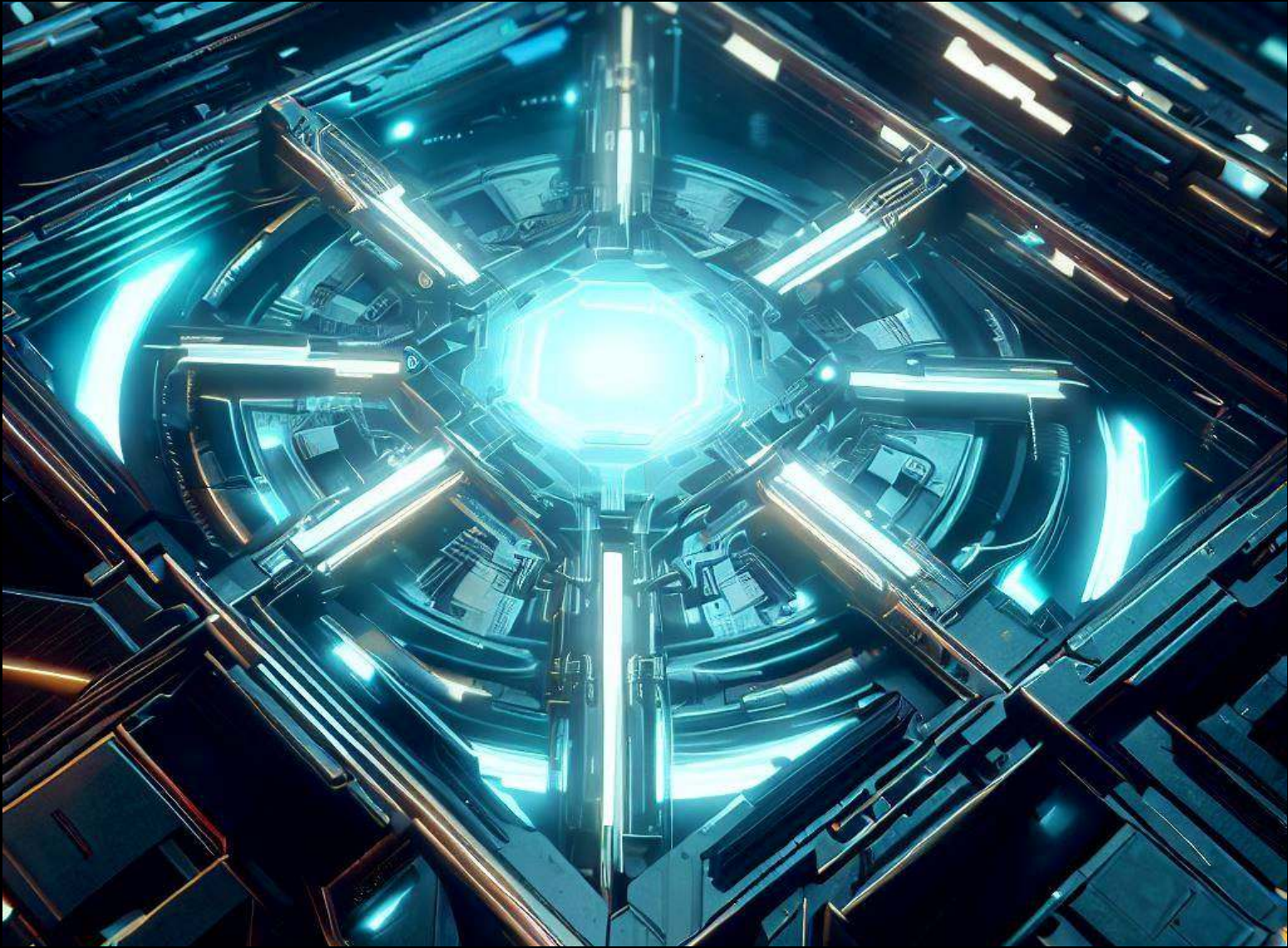
Simulation (-n):

```
zpool add -n FSCK24  
mirror aliendisk3 aliendisk4
```



zpool add -n FSCK24 mirror aliendisk3 aliendisk4
would update 'FSCK24' to the following
configuration:

```
FSCK24
  mirror-0
    medium1
    medium2
  mirror-1
    aliendisk3
    aliendisk4
```



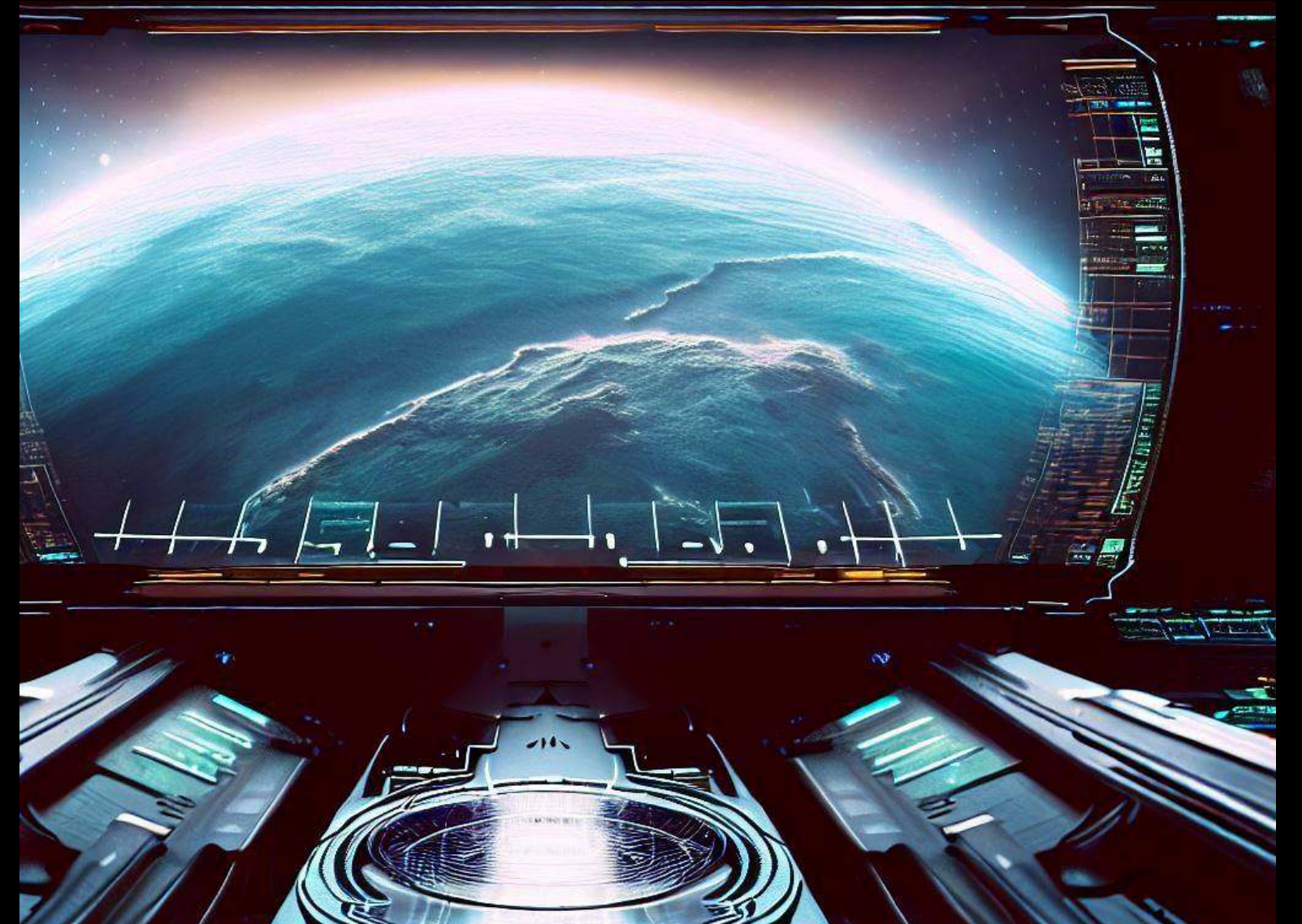
Die Reise geht weiter...

**Tag 42 seit
Missionsbeginn**

Kurs zu nahem Planeten







Volk der Archivaren

- Friedliches Volk
- Lange und gut dokumentierte Historie
- Zeichnen alles auf, was passiert
- Schildern folgendes Problem: Speicherplatz reicht nicht aus
- Kennen offenbar keine Kompression



Konzept des Datasets

- Logische Speichereinheit, ähnlich zu Partitionen
- Nutzen Speicher des Pools für Daten
- Unbegrenzt viele Datasets erstellbar
- Besitzen diverse Eigenschaften, z.B. Kompression
- Kind-Datasets erben Eigenschaften der Eltern




```
# zfs list
NAME                USED    AVAIL      REFER  MOUNTPOINT
FSCK24              788K    1.75P      104K    /FSCK24
FSCK24/crew         96K     1.75P      96K     /FSCK24/crew
```

```
# zfs create FSCK24/scans
```

```
# zfs list
NAME                USED    AVAIL      REFER  MOUNTPOINT
FSCK24              788K    1.75P      104K    /FSCK24
FSCK24/crew         96K     1.75P      96K     /FSCK24/crew
FSCK24/scans        96K     1.75P      96K     /FSCK24/scans
```



```
# zfs get compression FSCK24/scans
NAME          PROPERTY      VALUE          SOURCE
FSCK24/scans  compression  off            default
```

```
# zfs set compression=zstd-9 FSCK24/scans
```

```
# zfs create FSCK24/scans/nebula1337
```

```
# zfs get -r compression
```

NAME	PROPERTY	VALUE	SOURCE
FSCK24	compression	off	default
FSCK24/crew	compression	off	default
FSCK24/scans	compression	zstd-9	local
FSCK24/scans/nebula1337	compression	zstd-9	inherited from FSCK24/scans

```
# cp /dev/sensors/logs /FSCK24/scans/nebula1337
```

```
# zfs get compressratio FSCK24/scans/nebula1337
```

NAME	PROPERTY	VALUE	SOURCE
FSCK24/scans/nebula1337	compressratio	12.58x	-

Quota und Reservation

- Archivaren sind begeistert
- Chief Storage Engineer: *Möchten Sie mehr wissen?*
- Archivaren: *Wie verhindert ihr ein Überlaufen des Poolspeichers?*
- Konzepte wie Quota und Reservierung erläutert
- Mini-Schulungen auf der FSCK



Quota

- Lässt sich als Eigenschaft von Datasets festlegen
- Begrenzt wieviel ein Dataset speichern darf
- Pool verhindert weitere Allokation bei Erreichen der Quota

```
# zfs list -o name,avail FSCK24/crew
NAME                AVAIL
FSCK24/crew         1.75P

# zfs set refquota=1T FSCK24/crew
# zfs list -o name,avail FSCK24/crew
NAME                AVAIL
FSCK24/crew         1T
```


Quotas für Kind-Datasets

- `zfs set refquota` bezieht sich nur auf dieses Dataset
- `zfs set quota` bezieht sich auf das angegebene Dataset und alle Kind-Datasets
- Hier teilen sich also alle Datasets die Quota

```
# zfs list -ro name,avail FSCK24/scans
NAME                                AVAIL
FSCK24/scans                        40.8T
FSCK24/scans/nebula1337             40.8T
# zfs set quota=1T FSCK24/scans
# zfs create FSCK24/scans/nebula2342
# zfs list -ro name,avail FSCK24/scans
NAME                                AVAIL
FSCK24/scans                        1T
FSCK24/scans/nebula1337             1T
FSCK24/scans/nebula2342             1T
```


Reservierung

Platz garantieren

- Reservierungen garantieren eine Menge an Speicher
- Platz wird vom verfügbaren Poolspeicher abgezogen
- Gegenstück zu Quotas
- Auch hier: `reservation` bezieht sich auf Datasets inkl. Kinder
- `refreservation` hingegen nur auf dieses Dataset



zfs list

NAME	USED	AVAIL	REFER	MOUNTPOINT
FSCK24	788T	1.75P	104T	/FSCK24
FSCK24/crew	96T	1.75P	96T	/FSCK24/crew

...

zfs set reservation=1P FSCK24/crew

zfs list

NAME	USED	AVAIL	REFER	MOUNTPOINT
FSCK24	788T	0.75P	104T	/FSCK24
FSCK24/crew	96T	0.75P	96T	/FSCK24/crew

...

Technologietransfer

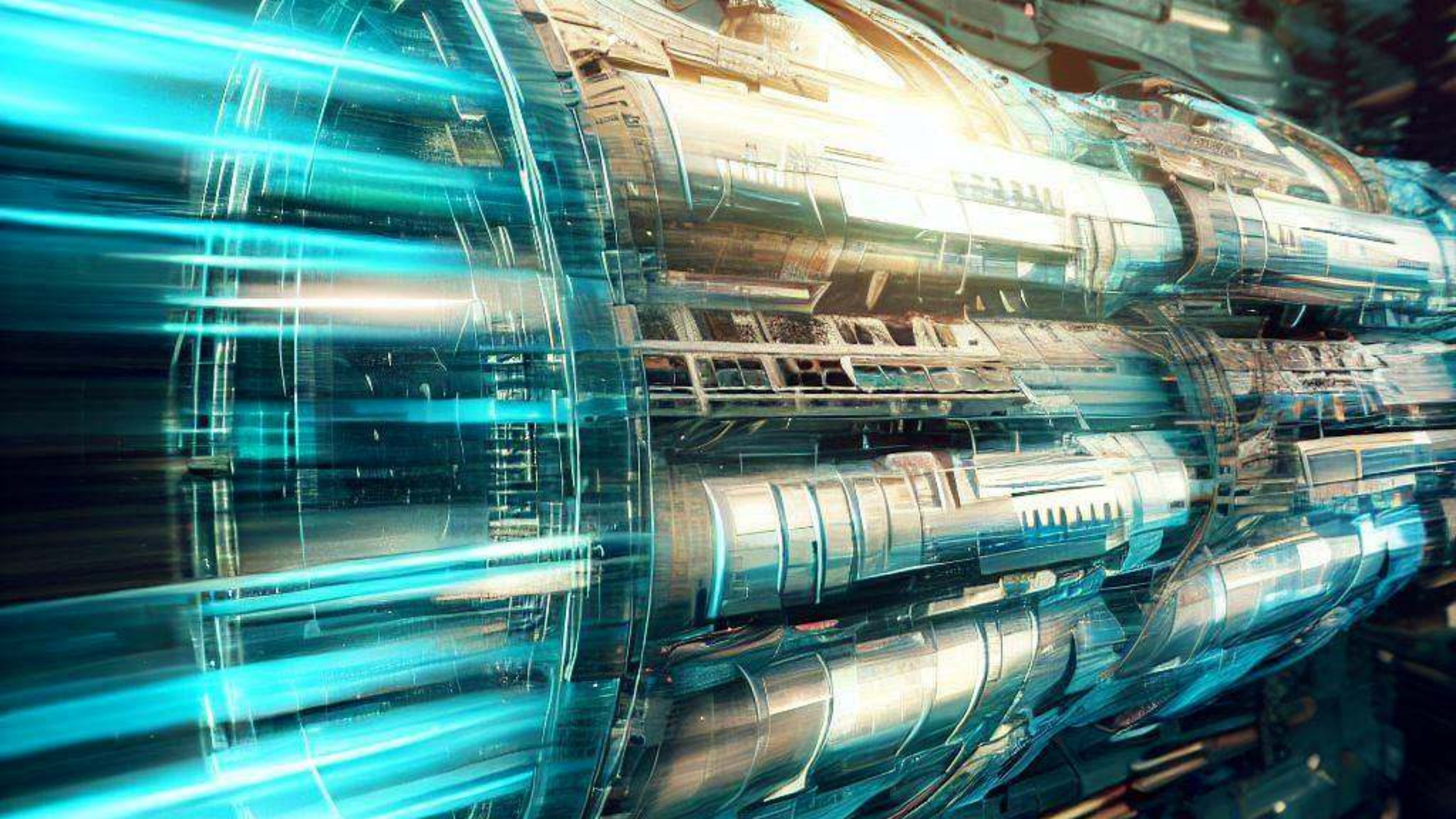
Notwendige Informationen an
Archivare gegeben

Kompression kann deren Archive
wesentlich verkleinern
(Kopieren auf neues Dataset nötig)

Quota und Reservierung
verhindern Überlauf

FSCK-Crew setzt wieder Kurs auf
neue Sternensysteme





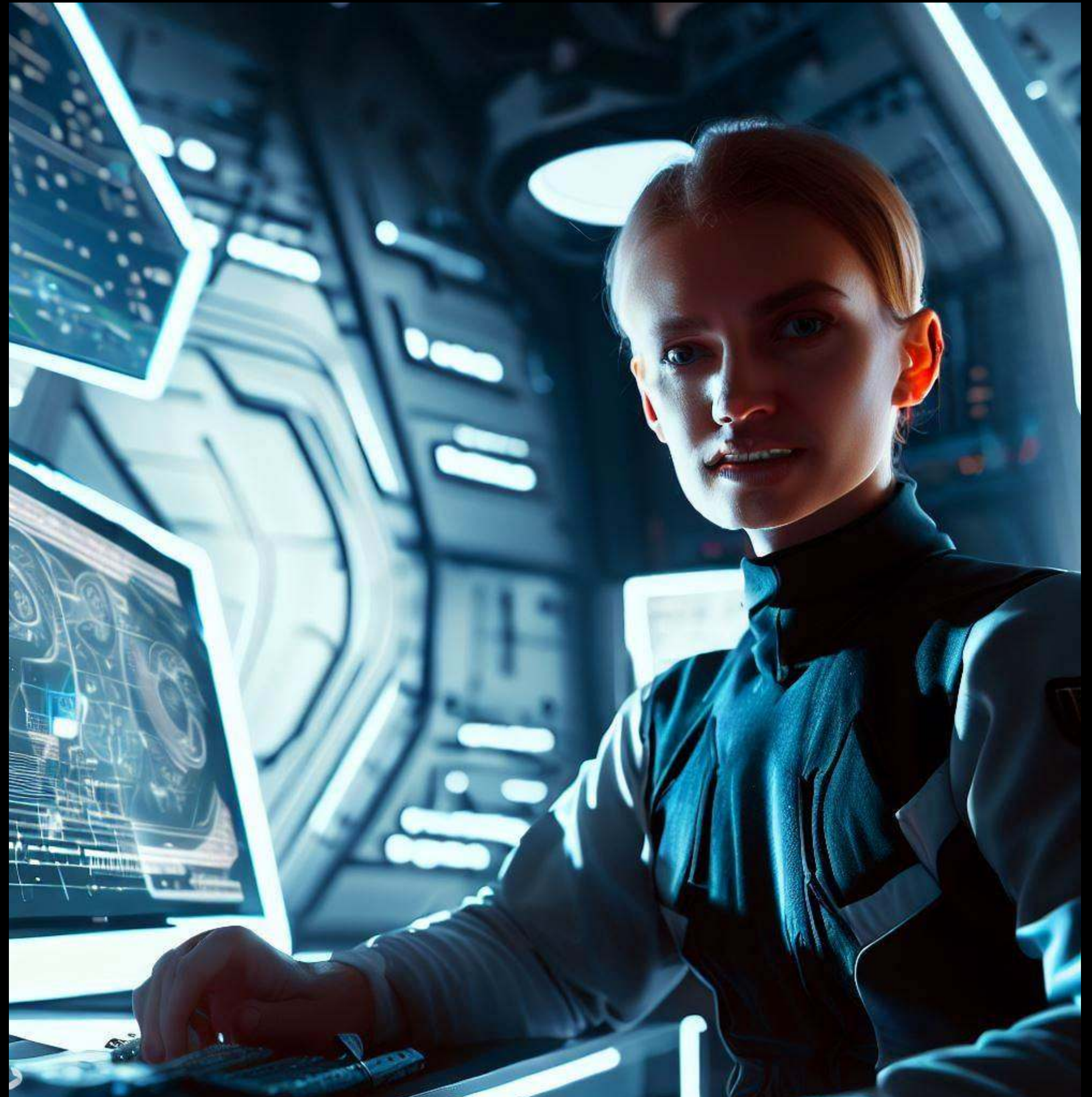
Die Zeit vergeht ereignislos...



... bis eines Tages...

Tag 67 seit Missionsbeginn

- INTERCOM: *Brücke an Chief Storage Engineer*
- Chief Storage Engineer: *Gibt es ein Problem?*
- Brücke: *Das sehen Sie sich am besten selbst an...*





CATS : ALL YOUR **FILES** ARE BELONG
TO US.





Angriff der Crypto-Trojaniden

- FSCK-Bordcomputer verschlüsselt
- Ohne Entschlüsselungskennwort kein Weiterflug möglich
- Trojaniden fordern 23.000 Coins als Lösegeld
- Geben 24 Stunden Bedenkzeit



Snapshots to the Rescue

- Snapshots sind nur-Lese Abbilder des Zustands der Datasets
- Erlaubt Point-in-time Recovery
- FSCK24-Pool erzeugt regelmässig Snapshots seit Missionsbeginn
- Zurückrollen zum Zeitpunkt vor der Verschlüsselung




```
# zfs list -t snapshot FSCK24
```

NAME	USED	AVAIL	REFER	MOUNTPOINT
FSCK@auto_2300-20-07-00:10	12T	-	12T	-
FSCK@auto_2300-21-07-00:10	7G	-	112T	-
FSCK@auto_2300-22-07-00:10	227G	-	124T	-
FSCK@auto_2300-23-07-00:10	13M	-	36T	-

```
# zfs rollback FSCK24@auto_2300-23-07-00:10
```

```
# ls /FSCK24/scans/nebula1337
```

```
...
```

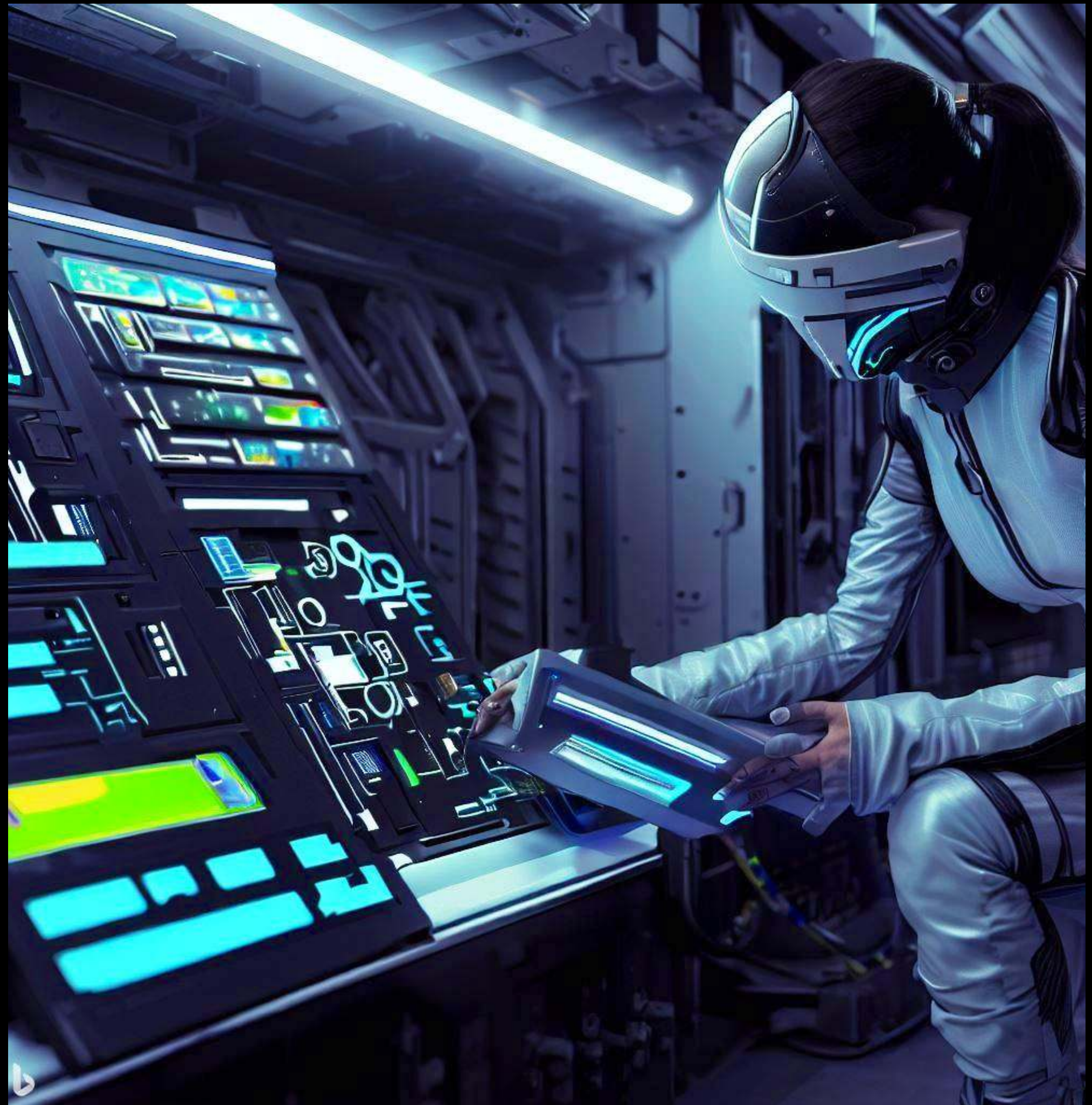
```
# zfs snapshot -r FSCK24@systems_clean
```

```
#zfs list -t snapshot FSCK24@systems_clean
```

NAME	USED	AVAIL	REFER	MOUNTPOINT
FSCK24@systems_clean	0B	-	112K	-

Schiffssysteme wiederhergestellt

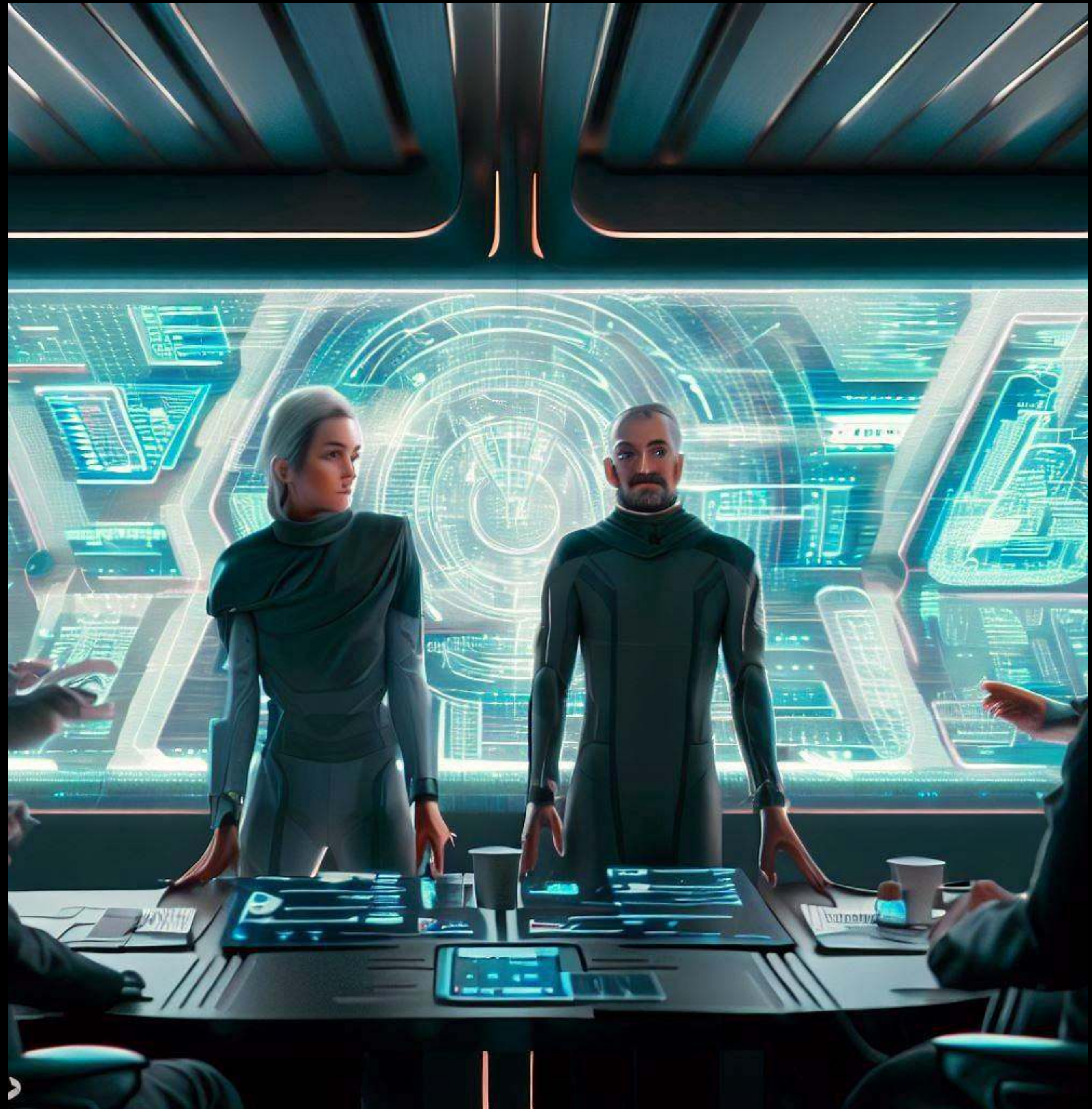
- Rückkehr auf letzten Snapshot vor der Infektion geglückt
- FSCK Captain: *Bringen Sie uns weg von hier...*
- Keine Verfolger sichtbar
- Captain befiehlt Überprüfung aller Systeme



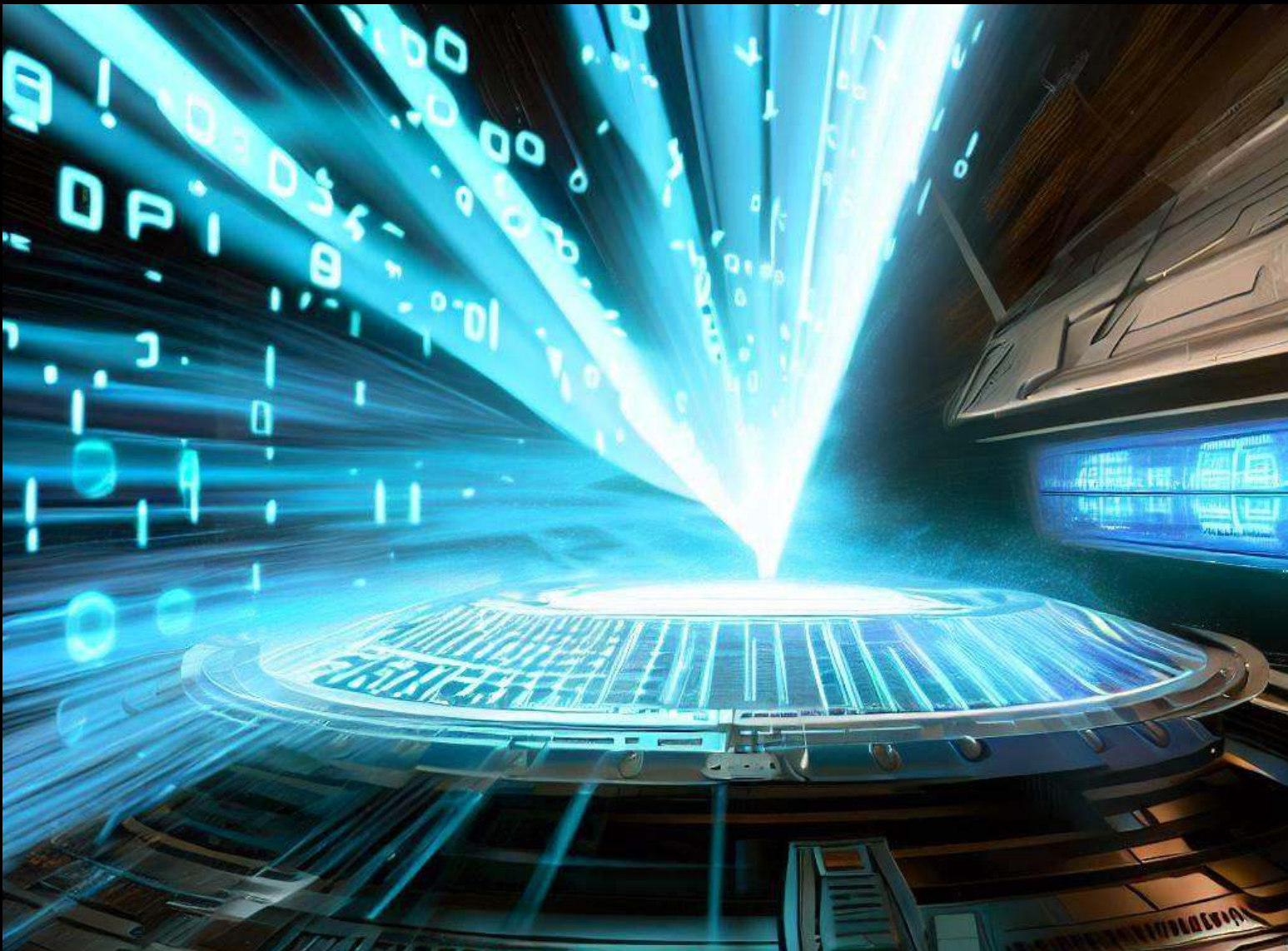
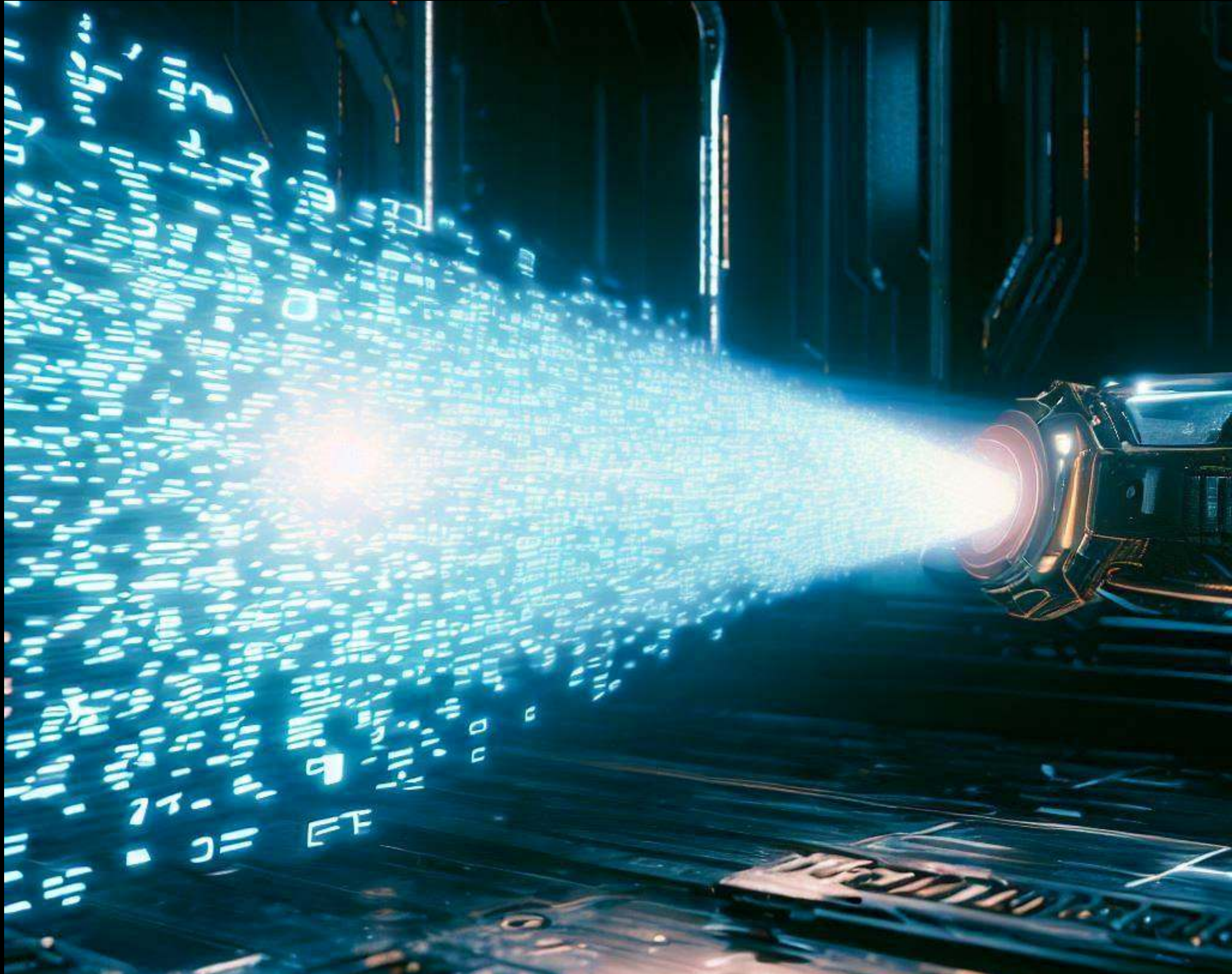
Senden von Daten an die Erde

Tag 100 seit Missionsbeginn

- Erde soll regelmässige Datensicherung erhalten
- Delta-Replikation von Snapshots
- Sicherer Übertragungskanal schützt vor Mithörern

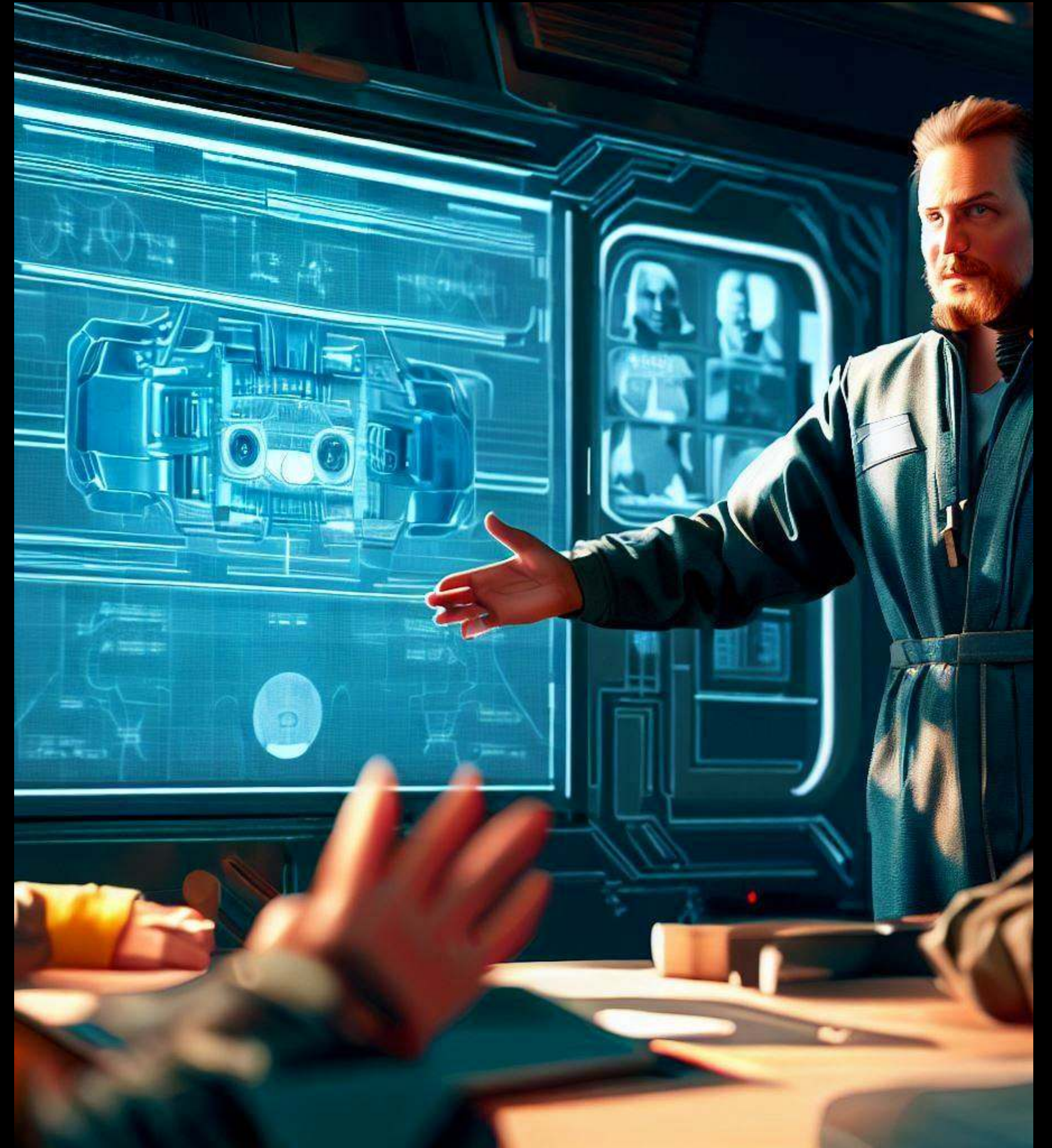



```
# zfs snapshot -r FSCK24/scans@send
# zfs list -o name -rt snap FSCK24/scans
NAME
FSCK24/scans@send
FSCK24/scans/nebula1337@send
FSCK24/scans/nebula2342@send
# zfs send -Rv FSCK24/scans@send | ssh fsck@earth \
"zfs receive earthpool"
full send of FSCK24/scans@send estimated size is 47.1K
full send of FSCK24/scans/nebula1337@send estimated size is 23G
full send of FSCK24/scans/nebula1337@send estimated size is 42G
total estimated size is 65G
TIME      SENT      SNAPSHOT
14:12:07  97.2M    FSCK24/scans/nebula1337@send
...
```

Die Reise geht weiter

- Es gibt noch viel zu entdecken
- Takeaway: OpenZFS hat viel zu bieten
- Technologie schon heute verfügbar
- Erprobt, stabil, zuverlässig
- www.openzfs.org



A woman with short dark hair, wearing a white flight suit with a dark blue collar, stands in a futuristic cockpit. She is holding a white rectangular sign in front of her chest. The cockpit is filled with blue light from various panels and screens. A large screen to her right displays colorful data. The background shows the complex structure of the cockpit with various instruments and lights.

Vielen Dank!