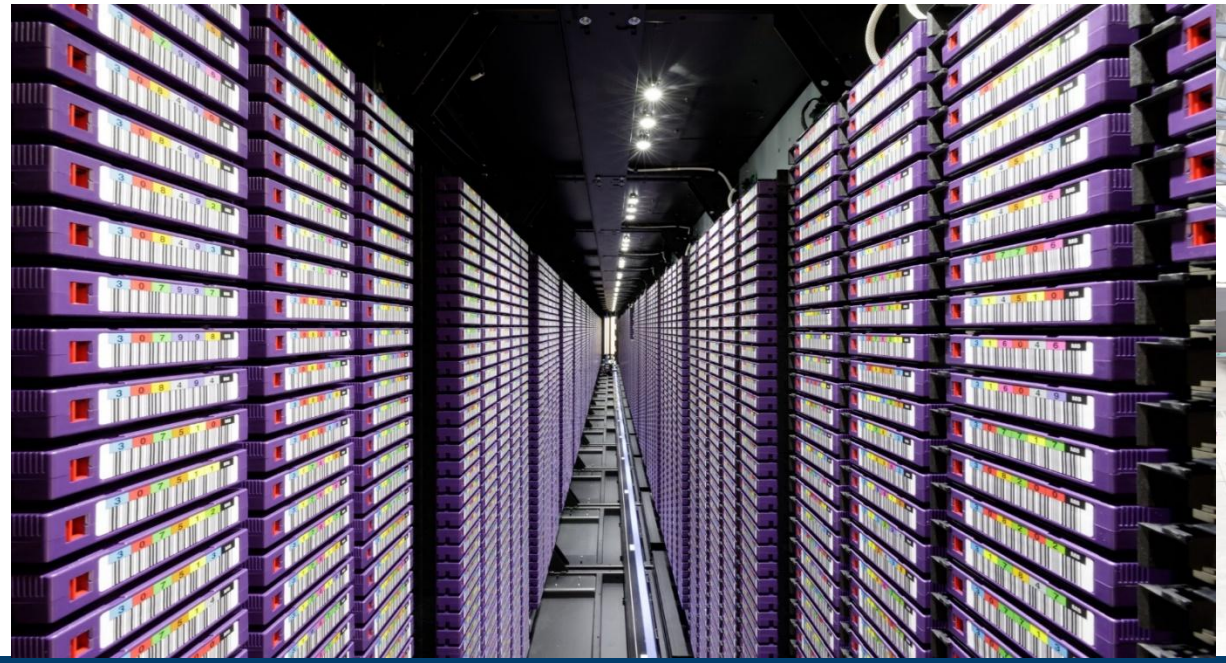
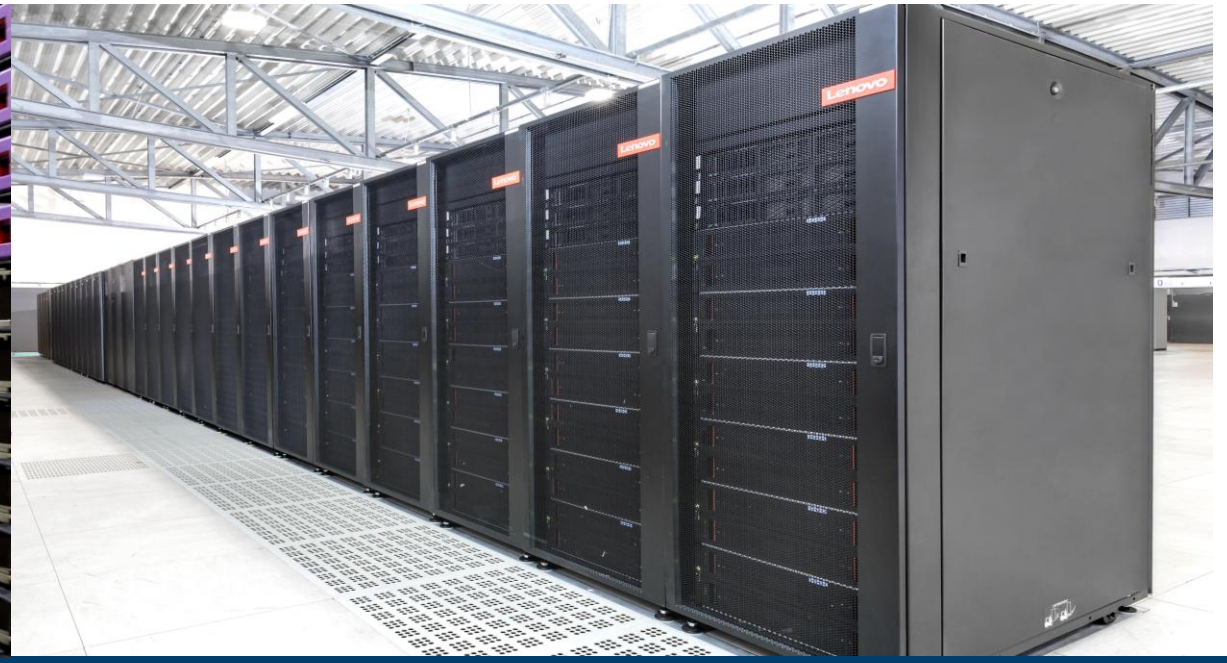




# MIGRATING THE STORAGE PROTECT INFRASTRUCTURE FROM POWER/AIX TO X86/LINUX INCLUDING THE SPACE MANAGEMENT CLIENT FOR STORAGE SCALE (GPFS)

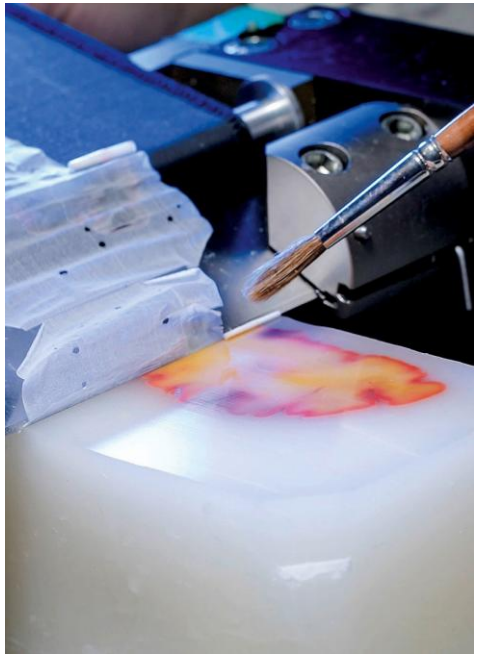
## DESIGN CHANGES IN THE 6TH GENERATION OF THE JÜLICH STORAGE CLUSTER JUST

24. SEPTEMBER 2024 I STEPHAN GRAF (JSC-HPCCDSS)

# FORSCHUNGSZENTRUM JÜLICH

Founded in 1956

Revenue:
€ 948 million in 2022



Research priorities:
information, energy, bioeconomy

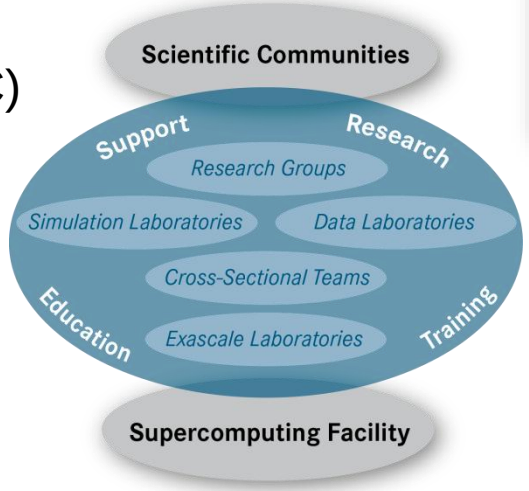
Research campus with 11 institutes and 18 branch offices in Germany and abroad

Shareholders:
Federal Republic of Germany (90%), federal state of North Rhine-Westphalia (10%)

Almost 7,250 employees from 111 countries

# JÜLICH SUPERCOMPUTING CENTRE

- **Supercomputer operation for**
  - Centre – FZJ
  - Region – RWTH Aachen University
  - Germany – Gauss Centre for Supercomputing (GCS)
    John von Neumann Institute for Computing (NIC)
  - Europe – PRACE, EU projects
- **Application support**
  - Unique support & research environment at JSC
  - Peer review support and coordination
- **R&D work**
  - Methods and algorithms, computational science, performance analysis and tools
  - Scientific Machine Learning and AI with HPC
  - Computer architectures, Co-Design, Exascale Labs together with IBM, Intel, NVIDIA
- **Education and training**


DEEP



JÜLICH Forschungszentrum

# JÜLICH SUPERCOMPUTING CENTRE

~300 Employees
(~280 FTE)

210 Scientists

22 PhD Students (+ externals)
27 Students (Bachelor/Master)

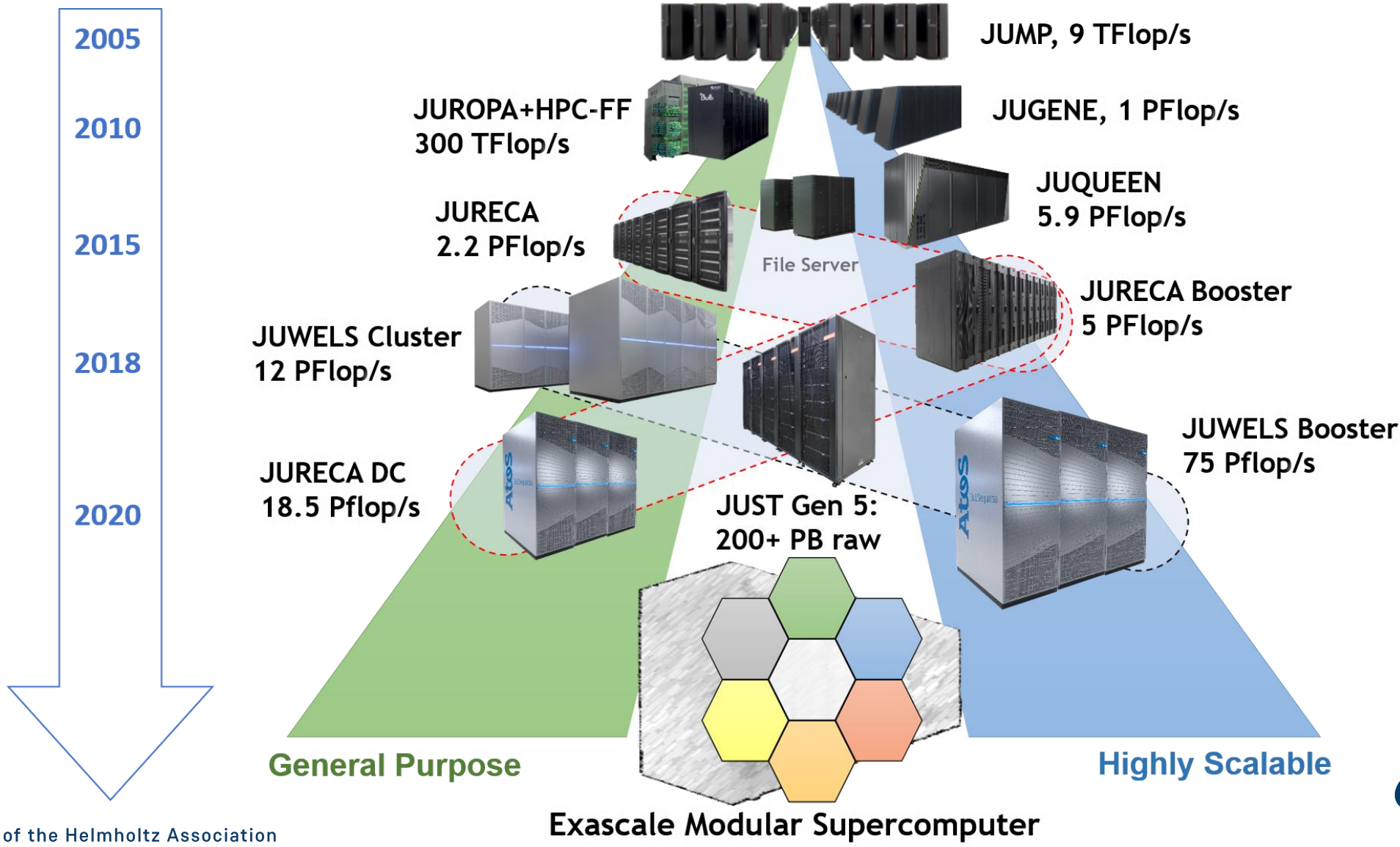
fz-juelich.de/jsc

also some JSC

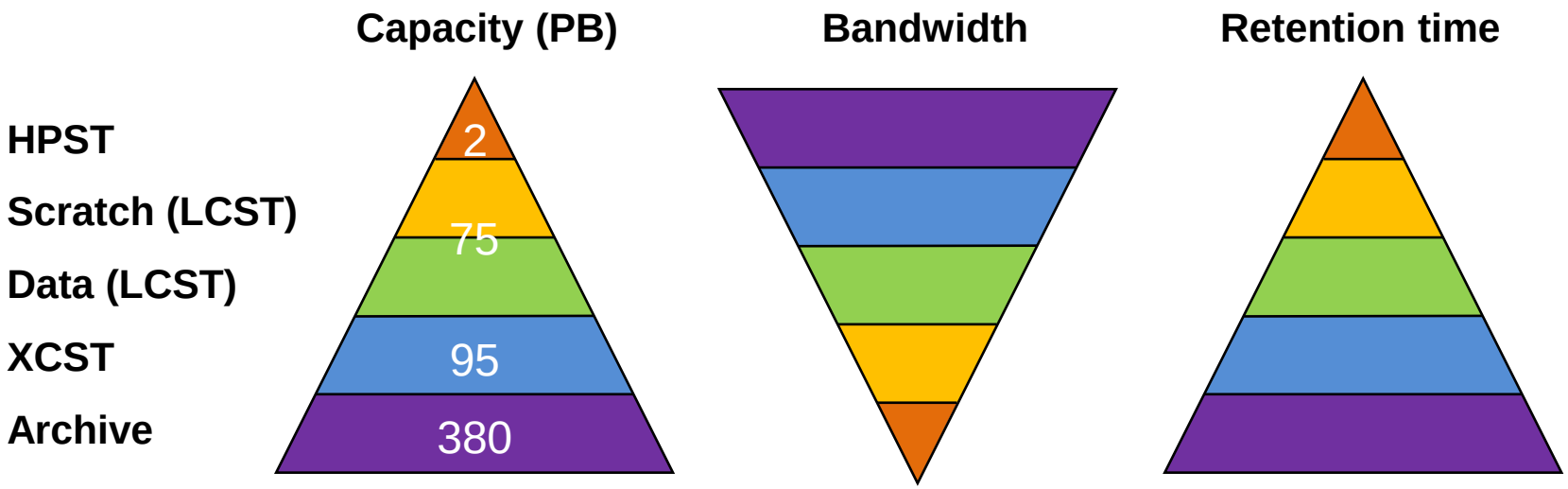
also some JSC

JSC

# (DUAL) HARDWARE STRATEGY AT JSC



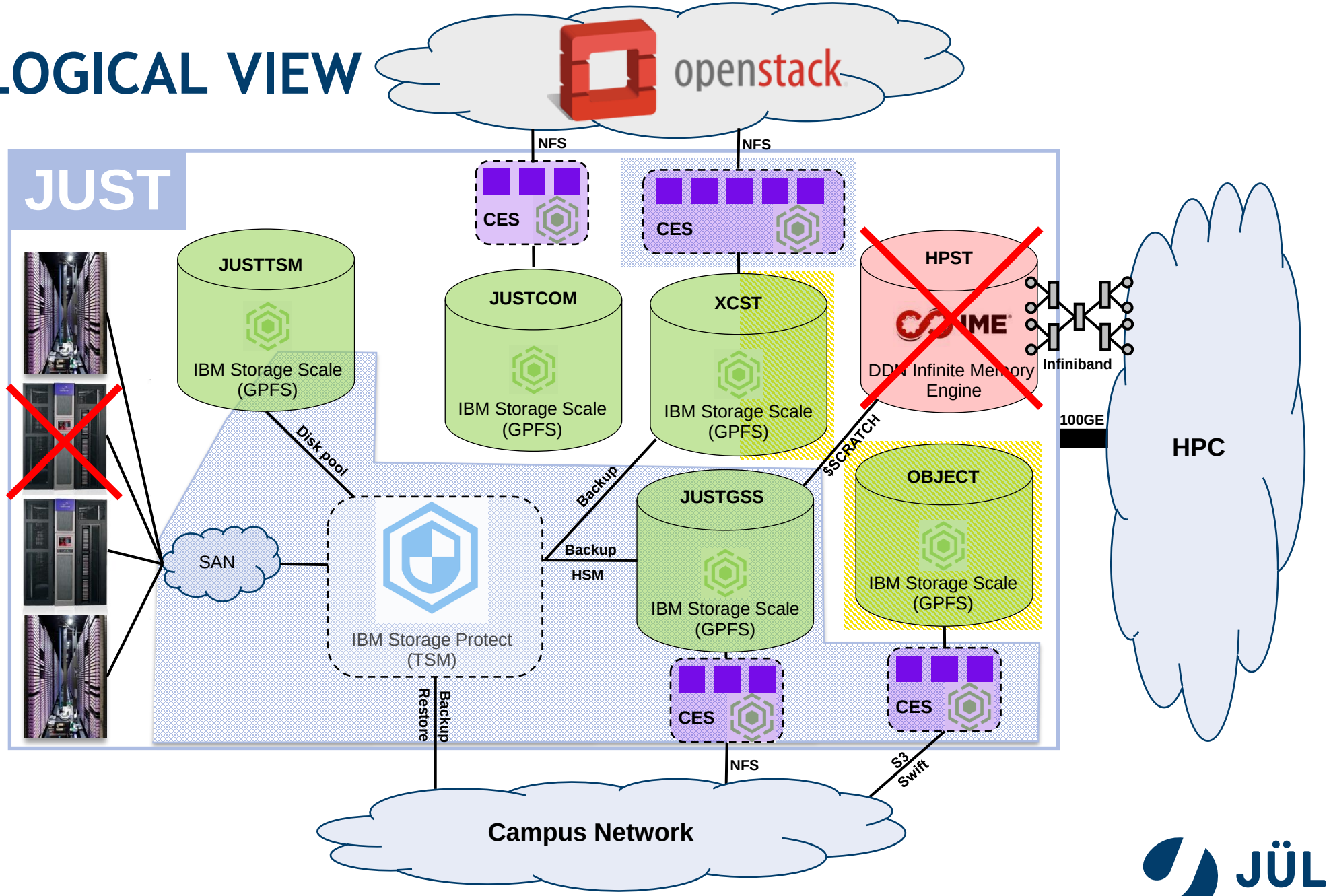
# JÜLICH STORAGE "JUST" – IN Q4/2023




- High Performance Storage Tier (HPST):  NVMe based Storage (low latency+high bandwidth)
- Large Capacity Storage Tier (LCST): Lenovo DSS Cluster (GNR, **5th** Gen. of JUST, bandwidth optimized)
- Extended Capacity Storage Tier (XCST): GPFS Building Blocks (target: capacity)
- Archive: Tape Storage Tier (Backup + GPFS & TSM-HSM)

# JUST - LOGICAL VIEW

## Q4/2023

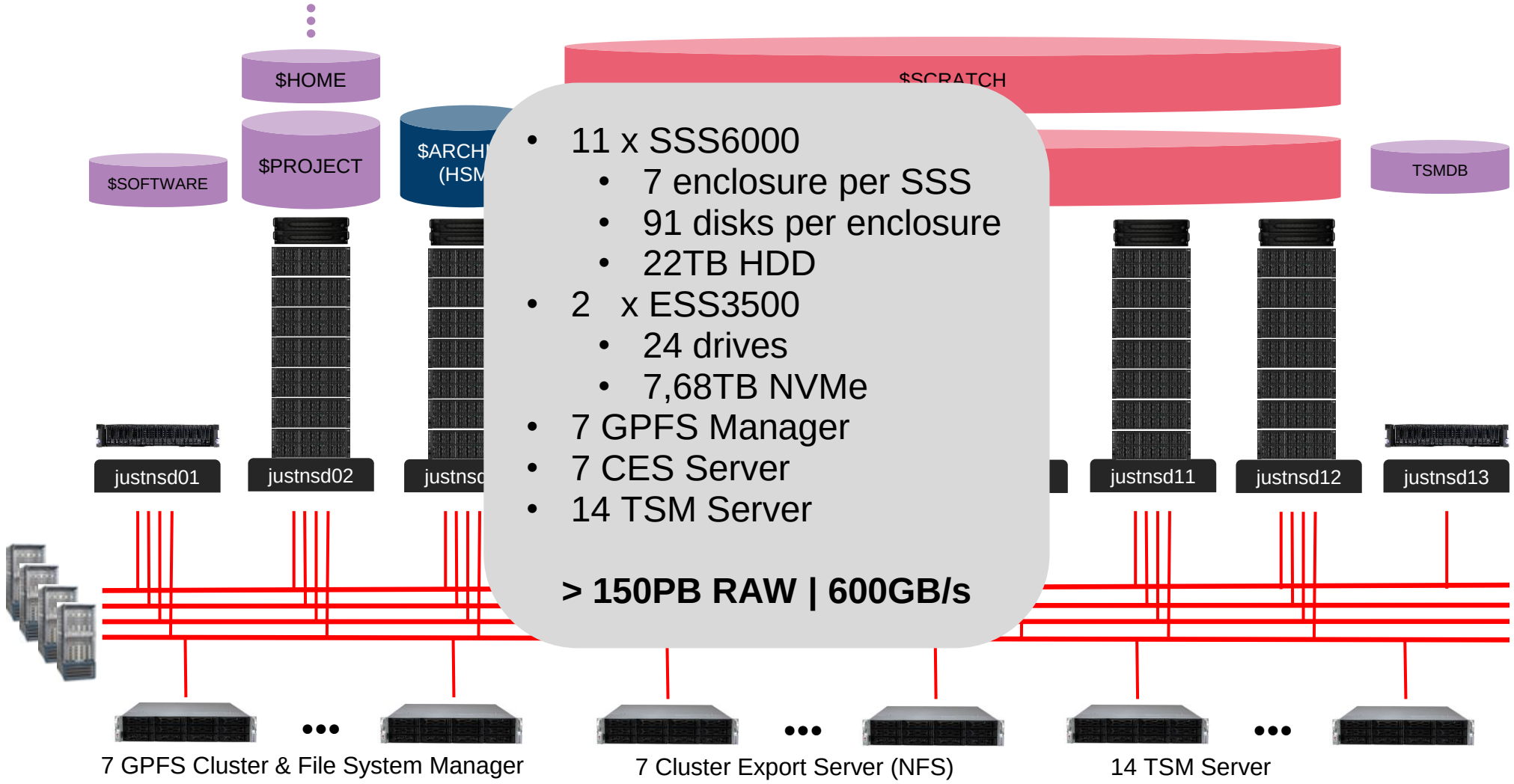


# ARCHITECTURAL CHANGE DECISIONS

## Procurement JUST (6th generation)

- Split core JUST cluster
  - Currently there is only one GPFS cluster
  - Overload situations lead to failing managers and thus global unavailability
  - Recovery process: In the past sequential, now parallel on three clusters
- **Storage Protect: Replace Power/AIX by x86/Linux**
  1. **Server – Server migration**
  2. **GPFS backup clients**
  3. **GPFS-HSM (client) migration**
- Cluster management
  - JUST5: xcat with 2 master servers and one monitoring server
  - JUST6: 3 master servers and ceph  with **ansible configuration management**
- SOFTWARE filesystem
  - Spinning disks with poor performance -> Dedicated NVMe Building Blocks

# JUST 6TH SERVERS AND JBODS




- 11 x SSS6000
  - 7 enclosure per SSS
  - 91 disks per enclosure
  - 22TB HDD
- 2 x ESS3500
  - 24 drives
  - 7,68TB NVMe
- 7 GPFS Manager
- 7 CES Server
- 14 TSM Server

**> 150PB RAW | 600GB/s**

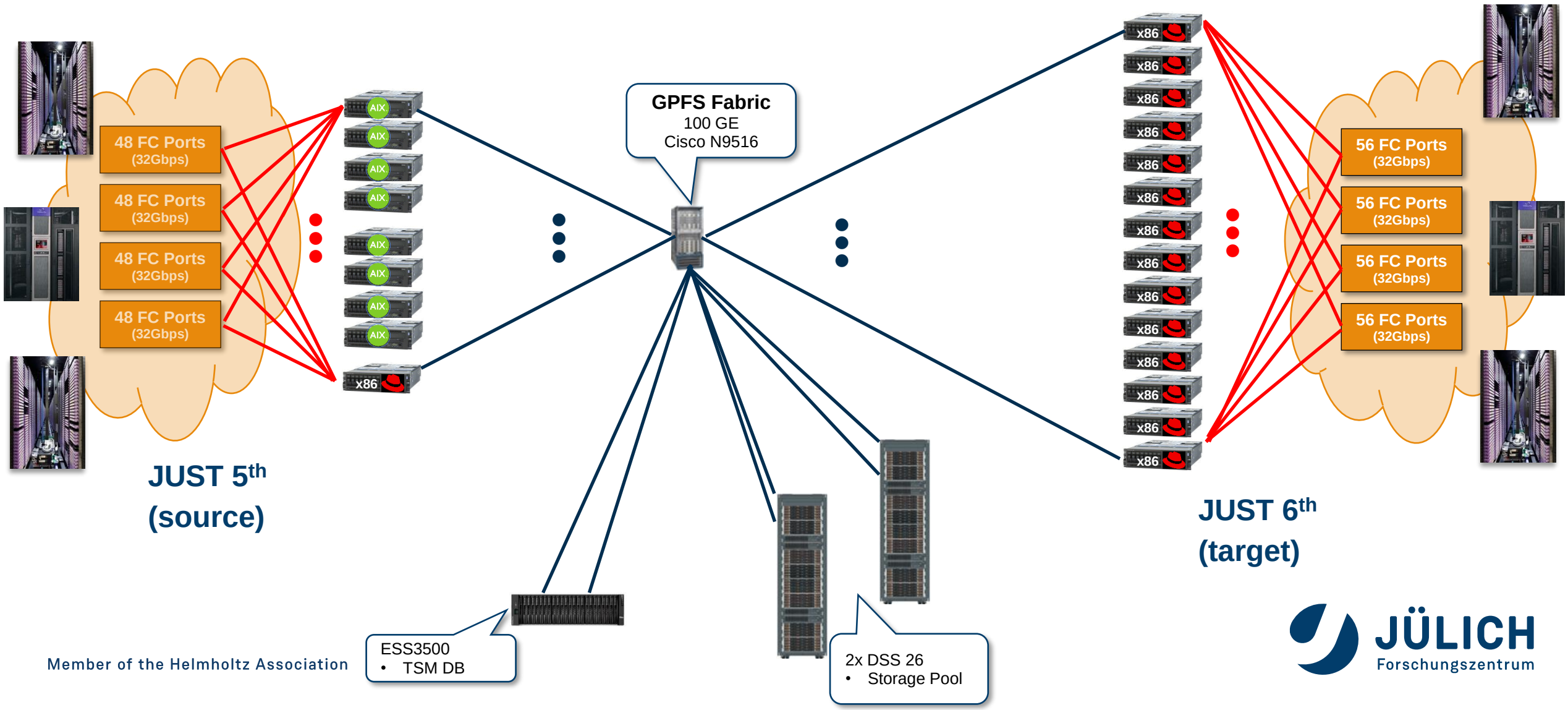
# HPC USER DATA MIGRATION

| Filesystem | Usage (GB) | Usage (Inodes) | (mm)Backup | HSM | Migration Method | old mount point | New mount point |
|---|---|---|---|---|---|---|---|
| **arch** | 27.262.976 (disk: 491.836) | 22.400.480 | ✓ | ✓ | mmrestripefs | /arch | /p/arch1 |
| **arch2** | 19.996.344 (disk: 454.473) | 8.879.869 | ✓ | ✓ | mmrestripefs | /arch2 | /p/arch2 |
| **largedata** | 15.578.638 | 110.496.477 | ✓ | | Copy/AFM | /p/largedata | /p/data1 |
| **fastdata** | 6.590.133 | 110.923.593 | ✓ | | Copy/AFM | /p/fastdata | /p/data1 |
| **project** | 3.625.717 | 702.994.066 | ✓ | | Copy/AFM | /p/project | /p/project1 |
| **scratch** | 9.501.539 | 419.819.397 | | | (selfservice) | /p/scratch | /p/scratch |
| **home** | 28.608 | 55.590.547 | ✓ | | Copy | /p/home | /p/home |
| **…** | **…** | … | … | … | Copy | … | … |
| ∑ | **~ 90.000.000** | **~ 1.600.000.000** | | | | | |

# SERVER – SERVER MIGRATION

# SPECTRUM PROTECT INFRASTRUCTURE

## Migration Setup



# SERVER – SERVER MIGRATION

## Preparation/Constrains

- **Create installation/migration plan (IBM & JSC)**
- New hardware installed (Rack,Server,Network,Power,BMC configuration, FC fabric, FC cabling,…)
- Prepare re-cabling of the tape drives to the new fabric
- FC Fabric configuration
- OS Deployment (Kickstart)
  - **No RHEL 9 support in SP 8.1.23, forced to install RHEL 8**
- Base OS customization (Ansible)
- Connectivity tests
  - GPFS network (done for IO acceptance tests)
  - Tape drive test (in OS visible)
- New GPFS file systems for the TSM instances
- Install IBM Storage Protect Server Extended Edition 8.1.23 (Linux x86_64)
  - **TSM ExtractDB/InsertDB from 8.1.17 -> 8.1.23 is supported (no update needed on source)**
- Configure all instances on target

# SERVER – SERVER MIGRATION

## Maintenance 2.7.2024 – 10.7.2024

### 1. Step: clean shutdown of running instances

- Disable sessions on all TSM instances
- Create DB Backup off all instances
- Halt all instances
- re-cabling of the tape drives to the new FC fabric (1 day)

### 2. Step: Start InsertDB on target

- Prereq: TSM instance is configured
- Start new instance as instance user: `dsmserv`
- Check server: `query session`
- Stop server: `halt`
- Remove existing DB: `dsmserv removdb TSMDB1`
- Delete all related DB/instance files

```
rm –fv $HOME/dsmkeydb.*
rm –fv $HOME/cert*
rm –Rfv /…/database/$USER
rm –Rfv /…/activelog/$USER/NODE0000
rm –rf /…/archivelog/$USER/USER
```

- Copy source instance cert files to target instance, adjust ownership :

```
cert.???, cert256.???, dsmkeydb.???
```

- Create new, empty DB by LOADFORMAT utility as instance user:

```
dsmserv loadformat dbdir=… activelogsize=…
activelogdir=… archlogdir=…
```

- Start insertion process as instance user

```
dsmserv insertdb sesswait=60
```

### 3. Step: Start ExtractDB on source

- Test connectivity to target node
- Start export as instance user:

```
dsmserv extractdb hladdress=tsm3cadm.zam.kfa-
juelich.de lladdress=1500
```

### 4. Step: finalize ExtractDB/InsertDB

- Check the output for any errors (extract/insert)
- Start instance as instance user on target in maintenance mode

```
dsmserv maintenance
```

- If all is fine, stop it: `halt`
- Adjust permission on disk storage pool
- Start instance again in maintenance mode and adjust some settings
- Run DB backup to check functionality

# SERVER – SERVER MIGRATION

## Migration overview

| TSM-Instanz | Description | Old DB Size (GB) | Old DB Dir | New DB Size (GB) | New DB Dir | TCPPort | TOTAL_MANAGED_TB | MAX_INGEST_IN_GB | AVERAGE_INGEST_IN_GB | AVERAGE_FILESIZE_IN_KB |
|---|---|---|---|---|---|---|---|---|---|---|
| tsmstk | Shared tape library manager | 12 | 4 | 2 | 4 | 1600 | 0,00 | 0,00 | 0,00 | 0,00 |
| tsmdc | dCache backend | 55 | 4 | 27 | 4 | 1560 | 22352.94 | 27.045,16 | 6.925,40 | 3.454.455.000.000,00 |
| tsmjsc | JSC Backup | 548 | 4 | 254 | | | | | 1.089,17 | 3.915,00 |
| jscwsg | JSC Backup | 966 | 4 | 585 | | | | | 284,69 | 207,00 |
| tsmarch | HPC Backup/HSM | 148 | 4 | 87 | | | | | 34.201,82 | 1.249.224.000.000,00 |
| tsmtest | Test | | | | | | | | - | - |
| justhome | HPC backup | 3566 | 4 | 2337 | | | | | 22.024,13 | 13.352,00 |
| tsmhome2 | HPC backup | | | 9 | | | | | 0,00 | 0,00 |
| tsm1a | FZJ backup | 120 | 4 | 58 | | | | | 2.397,44 | 987,00 |
| tsm1b | FZJ backup | 2158 | 4 | 557 | | | | | 2.706,46 | 686,00 |
| tsm1c | FZJ backup | 426 | 4 | 245 | | | | | 1.427,87 | 2.029,00 |
| tsm2a | FZJ backup | 752 | 4 | 415 | | | | | 3.771,40 | 1.537,00 |
| tsm2b | FZJ backup | 859 | 4 | 531 | | | | | 1.609,83 | 2.284,00 |
| tsm2c | FZJ backup | 767 | 4 | 354 | 12 | 1520 | 727.12 | 20.766,18 | 719,76 | 1.884,00 |
| tsm3a | FZJ backup | 429 | 4 | 218 | 8 | 1500 | 1135.97 | 51.426,77 | 5.690,25 | 4.828,00 |
| tsm3b | FZJ backup | 557 | 4 | 372 | 8 | 1510 | 733,00 | 27.840,29 | 4.707,30 | 1.582,00 |
| **tsm3c** | **FZJ backup** | **3959** | **4** | **2796** | **12** | **1520** | **2491.66** | **33.112,87** | **6.529,43** | **888,00** |
| tsm4a | FZJ backup | 551 | 4 | 339 | 8 | 1500 | 1497.96 | 43.994,50 | 7.279,73 | 3.573,00 |
| tsm4b | FZJ backup | 999 | 4 | 581 | 12 | 1510 | 1225.97 | 4.101,53 | 1.302,46 | 1.936,00 |
| adsmarc | FZJ Archive | 716 | 4 | 630 | 6 | 1550 | 3397.22 | 15.061,52 | 592,82 | 4.544,00 |
| tsmxcst | HPC backup | 162 | 4 | 109 | 6 | 1530 | 16909.65 | 160.245,82 | 118.259,83 | 156.568,00 |
| tsmxcst2 | HPC backup | 96 | 4 | 38 | 4 | 1540 | 7176.22 | 152.104,23 | 41.072,06 | 210.062,00 |
| tsmobj | Object backup | 5 | 4 | | 4 | 1535 | - | - | - | - |
| tsmdata1 | HPC backup | | | | 12 | 1545 | - | - | - | - |

**InsertDB and ExtractDB run very long**

For the biggest database with 3959 GB the InsertDB ran 27 hours.
The extract process typically finished typically much earlier and the insert process continues to run

# SERVER – SERVER MIGRATION

## Tape Access

- Install lin_tape + lin_taped packages for the IBM TS4500
- **Udev** Rules to create persistent device name links
  Oracle T10KD drives:

```
SUBSYSTEM=="scsi_generic", MODE="0666"
SUBSYSTEM=="scsi_generic", SUBSYSTEMS=="scsi", ATTRS{vendor}=="STK*", ATTRS{model}=="T10*", PROGRAM="/lib/udev/scsi_id --page=0x80 --
whitelisted -d $tempnode", RESULT=="*579001234567*", SYMLINK+="L2D01"
```

- IBM LTO drives:

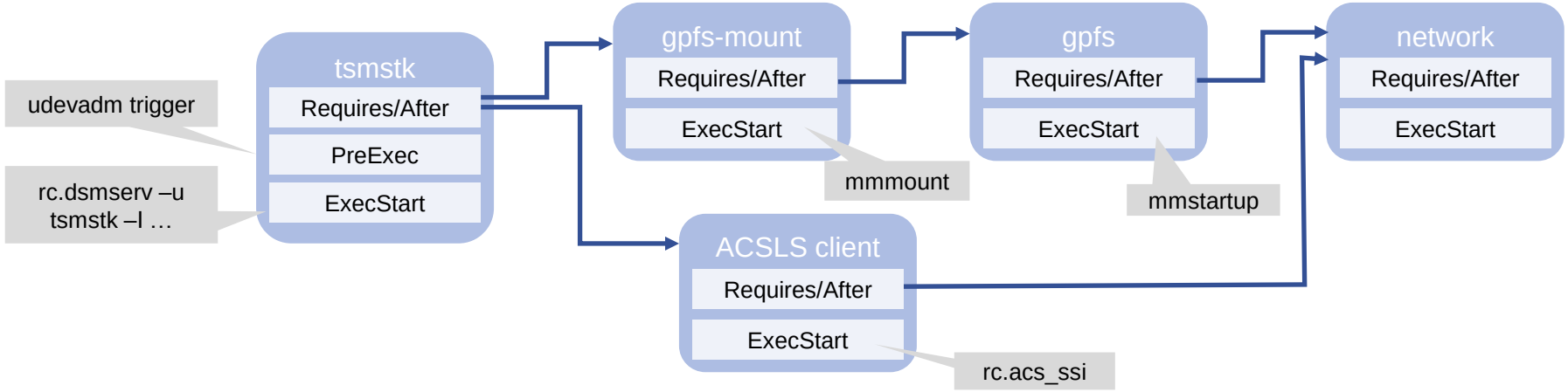
```
KERNEL=="TSMtape", MODE="0666"
KERNEL=="IBMchanger*", MODE="0666", ATTRS{serial_num}=="*78BB123*", ATTR{primary_path}=="Primary", SYMLINK+="TS3control"
KERNEL=="IBMtape*[0-9]", MODE="0666", ATTRS{serial_num}=="*10WT012345*", SYMLINK+="L3D01"
```

- Prepare systemctl service to start Oracle ACSLS service
  `/opt/Tivoli/tsm/devices/bin/rc.acs_ssi`
- Define path for all drives to all instances on the TSM shared tape library instance (tsmstk)

# SERVER – SERVER MIGRATION

## Production setup customization

- **Use systemd to start, stop and solve dependencies**




- **Prepare server/instance failover**
  - **One global dsm.sys/dsm.opt**
  - **one global (per group) DB2 port mapping in /etc/services**

```
# DB2:  20<srvnum><instnum><num>/tcp
# db2c: 25<srvnum><instnum><num>/tcp
```

  - `Db2nodes.cfg` **linked to** `/etc/tsm/...(per instance)`

```
DB2_tsmstk       20110/tcp
DB2_tsmstk_1     20111/tcp
DB2_tsmstk_2     20112/tcp
DB2_tsmstk_3     20113/tcp
DB2_tsmstk_4     20114/tcp
DB2_tsmstk_END   20115/tcp
db2c_tsmstk      25110/tcp
```

# HPC BACKUP CLIENT MIGRATION

# BACKUP CLIENT MIGRATION – CHALLENGES

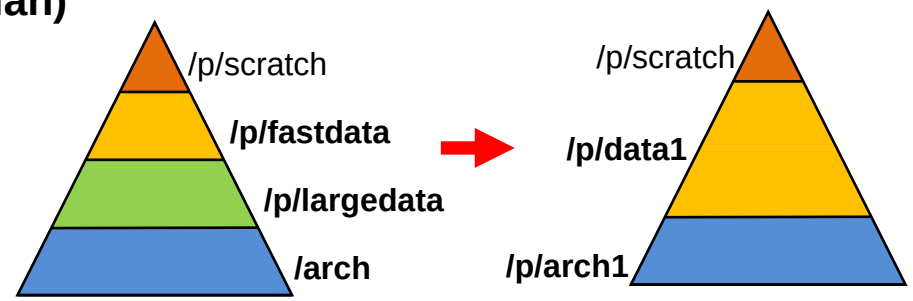
## Fact collection & Design changes

| Filesystem | Usage (GB) | Usage (Inodes) |
|---|---|---|
| **arch** | 27.262.976 | 22.400.480 |
| **arch2** | 19.996.344 | 8.879.869 |
| **largedata** | 15.578.638 | 110.496.477 |
| **fastdata** | 6.590.133 | 110.923.593 |
| **project** | 3.625.717 | 702.994.066 |
| **home** | 28.608 | 55.590.547 |

- GPFS file system backup (user data & statistic data)
- Backup tool: mmBackup
  - mmBackup is doing:

```
dsmc selective –filelist=…
dsmc incremental –filelist=…
dsmc expire –filelist=…
```

- **Backup client moves from AIX/Power (Big Endian) to Linux/x86 (Little Enian)**
- Backup client and target TSM instance are located on the same server
- **Some file systems will get new mount point**
  Motivation1: HPC admins decided to change the mountpoint in the past,
  but backup server still uses the old one
  Motivation2: rename of the directory to get a consistent name




- All online data (not the HSM managed file system) will be copied. **The goal is not to back up the data again** (if possible).
- Renaming the HSM managed file system **must not** trigger a new backup of the files.

# BACKUP CLIENT MIGRATION

## 1. Case: rename mount point

```
[root@x3850-x6-22 hsm]# dsmc incr F*
IBM Spectrum Protect
Command Line Backup-Archive Client Interface
  Client Version 8, Release 1, Level 20.0
  Client date/time: 03/22/2024 12:08:16
…
```

```
root@x3850-x6-22 hsm]# dsmc incr /d/gpfs/hsm/F*
IBM Spectrum Protect
Command Line Backup-Archive Client Interface
  Client Version 8, Release 1, Level 20.0
  Client date/time: 03/22/2024 12:19:32
…
Accessing as node: HSM-PROXY
…
Incremental backup of volume 'Fabian4'
Successful incremental backup of '/d/gpfs/hsm/Fabian4'
…
Total number of objects inspected:            5
Total number of objects backed up:            0
Total number of objects updated:              0
Total number of objects rebound:              0
Total number of objects deleted:              0
Total number of objects expired:              0
Total number of objects failed:               0
Total number of objects encrypted:            0
Total number of objects grew:                 0
Total number of retries:                      0
Total number of bytes inspected:           5.00 MB
Total number of bytes transferred:            0  B
Data transfer time:                        0.00 sec
Network data transfer rate:                0.00 KB/sec
Aggregate data transfer rate:              0.00 KB/sec
Objects compressed by:                        0%
Total data reduction ratio:                100.00%
Elapsed processing time:                   00:00:02
```

```
Protect: PROTECT1>q filespace HSM-PROXY

Node Name            Filespace            FSID     Platform     Filespac-     Is Filesp-        Capacity      Pct
                     Name                                       e Type        ace Unico-                      Util
                                                                                 de?
---------------      -----------      --------     --------     ---------     ----------     -----------     -----
HSM-PROXY            /p/gpfs/hsm             3                  GPFS              No              25 TB       1.1
```

```
[root@x3850-x6-22 /]# mmumount hsm -a
Fri Mar 22 12:09:11 CET 2024: mmumount: Unmounting file systems ...

[root@x3850-x6-22 /]# mmchfs hsm -T /d/gpfs/hsm
mmchfs: Propagating the cluster configuration data to all
  affected nodes.  This is an asynchronous process.

[root@x3850-x6-22 /]# mmmount hsm -a
Fri Mar 22 12:11:08 CET 2024: mmmount: Mounting file systems ...
```

```
Protect: PROTECT1>rename filespace HSM-PROXY /p/gpfs/hsm /d/gpfs/hsm

Do you wish to proceed? (Yes (Y)/No (N)) Y
ANR0822I RENAME FILESPACE: Filespace /p/gpfs/hsm (fsId=3) successfully renamed to /d/gpfs/hsm for node HSM-PROXY.


Protect: PROTECT1>q filespace HSM-PROXY

Node Name            Filespace            FSID     Platform     Filespac-     Is Filesp-        Capacity      Pct
                     Name                                       e Type        ace Unico-                      Util
                                                                                 de?
---------------      -----------      --------     --------     ---------     ----------     -----------     -----
HSM-PROXY            /d/gpfs/hsm             3                  GPFS              No              25 TB       1.1
```

# BACKUP CLIENT MIGRATION

## 2. Case: copy data to new device, but the same path (1/3)

Create initial data and backup

```
[root@x3850-x6-22 hsm]# mmlsfileset /dev/hsm
Filesets in file system 'hsm':
Name                     Status    Path
root                     Linked    /gpfs/hsm

[root@x3850-x6-22 hsm]# mmmount hsm
[root@x3850-x6-22 hsm]# cd /gpfs/hsm
[root@x3850-x6-22 hsm]# mkdir NEWBACKUP24.04.2024
[root@x3850-x6-22 hsm]# cd NEWBACKUP24.04.2024

# create files

[root@x3850-x6-22 NEWBACKUP24.04.2024]# ls -l
total 100352
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420240
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420241
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420242
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420243
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420244
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420245
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420246
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420247
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420248
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420249
```

```
[root@x3850-x6-22 NEWBACKUP24.04.2024]# dsmc incr /gpfs/hsm/NEWBACKUP24.04.2024/
IBM Spectrum Protect
Command Line Backup-Archive Cl
  Client Version 8, Release 1,
  Client date/time: 04/24/2024
(c) Copyright by IBM Corporati

Node Name: PROXYAGENT-22
Session established with serve
  Server Version 8, Release 1,
  Server date/time: 04/24/2024

Accessing as node: HSM-PROXY

Incremental backup of volume '
Directory-->                4,0
Normal File-->         10,485,7
…
Successful incremental backup

Total number of objects inspec
Total number of objects backed
Total number of objects update
Total number of objects rebound
Total number of objects delete
Total number of objects expire
Total number of objects failed
Total number of objects encryp
Total number of objects grew:
Total number of retries:
Total number of bytes inspecte
Total number of bytes transfer
Data transfer time:
Network data transfer rate:
Aggregate data transfer rate:
Objects compressed by:
Total data reduction ratio:
Elapsed processing time:                00:00:08
```

```
[root@x3850-x6-22 NEWBACKUP24.04.2024]# dsmc i /gpfs/hsm/NEWBACKUP24.04.2024/
IBM Spectrum Protect
Command Line Backup-Archive Client Interface
  Client Version 8, Release 1, Level 20.0
  Client date/time: 04/24/2024 15:57:08
(c) Copyright by IBM Corporation and other(s) 1990, 2023. All Rights Reserved.

Node Name: PROXYAGENT-22
Session established with server PROTECT1: Linux/x86_64
  Server Version 8, Release 1, Level 20.000
  Server date/time: 04/24/2024 15:57:08  Last access: 04/24/2024 15:56:44

Accessing as node: HSM-PROXY

Incremental backup of volume '/gpfs/hsm/NEWBACKUP24.04.2024/'
Successful incremental backup of '/gpfs/hsm/NEWBACKUP24.04.2024/*'


Total number of objects inspected:            11
Total number of objects backed up:             0
Total number of objects updated:               0
Total number of objects rebound:               0
Total number of objects deleted:               0
Total number of objects expired:               0
Total number of objects failed:                0
Total number of objects encrypted:             0
Total number of objects grew:                  0
Total number of retries:                       0
Total number of bytes inspected:          100.00 MB
Total number of bytes transferred:             0  B
Data transfer time:                         0.00 sec
Network data transfer rate:                 0.00 KB/sec
Aggregate data transfer rate:               0.00 KB/sec
Objects compressed by:                         0%
Total data reduction ratio:                100.00%
Elapsed processing time:                   00:00:01
```

# BACKUP CLIENT MIGRATION

## 2. Case: copy data to new device, but the same path (2/3)

### Move/copy data to new device

```
[root@x3850-x6-22 hsm]# pwd
/gpfs/hsm/NEWBACKUP24.04.2024
[root@x3850-x6-22 NEWBACKUP24.04.2024]# stat NEWBACKUP240420240
  File: NEWBACKUP240420240
  Size: 10485760        Blocks: 20480      IO Block: 1048576 regular file
Device: 38h/56d Inode: 510145       Links: 1
Access: (0644/-rw-r--r--)  Uid: (    0/    root)   Gid: (    0/    root)
Context: unconfined_u:object_r:unlabeled_t:s0
Access: 2024-04-24 15:55:42.589057000 +0200
Modify: 2024-04-24 15:55:42.607484005 +0200
Change: 2024-04-24 15:55:42.607484005 +0200
 Birth: -

[root@x3850-x6-22 hsm]# cd ..
[root@x3850-x6-22 hsm]# mv NEWBACKUP24.04.2024 NEWBACKUP24.04.2024_origin
[root@x3850-x6-22 hsm]# ls -l
total 1
drwxr-xr-x 2 root root 4096 Apr 24 15:54 NEWBACKUP24.04.2024_origin

[root@x3850-x6-22 hsm]# mmcrfileset /dev/hsm NEWBACKUP24.04.2024
Fileset NEWBACKUP24.04.2024 created with id 2 root inode 65795.

[root@x3850-x6-22 hsm]# mmlsfileset /dev/hsm
Filesets in file system 'hsm':
Name                    Status    Path
root                    Linked    /gpfs/hsm
NEWBACKUP24.04.2024     Unlinked  --

[root@x3850-x6-22 hsm]# mmlinkfileset /dev/hsm NEWBACKUP24.04.2024 -J /gpfs/hsm/NEWBACKUP24.04.2024
Fileset NEWBACKUP24.04.2024 linked at /gpfs/hsm/NEWBACKUP24.04.2024

[root@x3850-x6-22 hsm]# mmlsfileset /dev/hsm
Filesets in file system 'hsm':
Name                    Status    Path
root                    Linked    /gpfs/hsm
NEWBACKUP24.04.2024     Linked    /gpfs/hsm/NEWBACKUP24.04.2024
```

```
[root@x3850-x6-22 hsm]# mv NEWBACKUP24.04.2024_origin/* /gpfs/hsm/NEWBACKUP24.04.2024/

[root@x3850-x6-22 hsm]# ls -l /gpfs/hsm/NEWBACKUP24.04.2024/
total 77824
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420240
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420241
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420242
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420243
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420244
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420245
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420246
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420247
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420248
-rw-r--r--. 1 root root 10485760 Apr 24 15:55 NEWBACKUP240420249

[root@x3850-x6-22 hsm]# cd NEWBACKUP24.04.2024
[root@x3850-x6-22 NEWBACKUP24.04.2024]# stat NEWBACKUP240420240
  File: NEWBACKUP240420240
  Size: 10485760        Blocks: 20480      IO Block: 1048576 regular file
Device: 38h/56d Inode: 44294        Links: 1
Access: (0644/-rw-r--r--)  Uid: (    0/    root)   Gid: (    0/    root)
Context: unconfined_u:object_r:unlabeled_t:s0
Access: 2024-04-24 15:55:42.589057000 +0200
Modify: 2024-04-24 15:55:42.607484005 +0200
Change: 2024-04-24 16:00:08.931149977 +0200
 Birth: -
```

Goal: backup metadata only

# BACKUP CLIENT MIGRATION

## 2. Case: copy data to new device, but the same path (3/3)

Backup data using client option SKIPACLUP YES

```
[root@x3850-x6-22 NEWBACKUP24.04.2024]# dsmc incr /gpfs/hsm/NEWBACKUP24.04.2024/
```

```
[root@x3850-x6-22 NEWBACKUP24.04.2024]# dsmc i /gpfs/hsm/NEWBACKUP24.04.2024/
```

```
[root@x3850-x6-22 NEWBACKUP24.04.2024]# dsmc i /gpfs/hsm/NEWBACKUP24.04.2024/
IBM Spectrum Protect
Command Line Backup-Archive Client Interface
  Client Version 8, Release 1, Level 20.0
  Client date/time: 04/24/2024 16:04:09
…
Accessing as node: HSM-PROXY

Incremental backup of volume '/gpfs/hsm/NEWBACKUP24.04.2024/'
Successful incremental backup of '/gpfs/hsm/NEWBACKUP24.04.2024/*'


Total number of objects inspected:            11
Total number of objects backed up:             0
Total number of objects updated:               0
Total number of objects rebound:               0
Total number of objects deleted:               0
Total number of objects expired:               0
Total number of objects failed:                0
Total number of objects encrypted:             0
Total number of objects grew:                  0
Total number of retries:                       0
Total number of bytes inspected:          100.00 MB
Total number of bytes transferred:             0  B
Data transfer time:                         0.00 sec
Network data transfer rate:                 0.00 KB/sec
Aggregate data transfer rate:               0.00 KB/sec
Objects compressed by:                         0%
Total data reduction ratio:               100.00%
Elapsed processing time:                  00:00:01
```

5

```
[root@x3850-x6-22 NEWBACKUP24.04.2024]# cat /opt/tivoli/tsm/client/ba/bin/dsm.opt
Servername hsm
*SKIPACLUP YES

[root@x3850-x6-22 NEWBACKUP24.04.2024]# vi /opt/tivoli/tsm/client/ba/bin/dsm.opt

[root@x3850-x6-22 NEWBACKUP24.04.2024]# cat /opt/tivoli/tsm/client/ba/bin/dsm.opt
Servername hsm
SKIPACLUP YES
```

1

```
[root@x3850-x6-22 NEWBACKUP24.04.2024]# vi /opt/tivoli/tsm/client/ba/bin/dsm.opt

[root@x3850-x6-22 NEWBACKUP24.04.2024]# cat /opt/tivoli/tsm/client/ba/bin/dsm.opt
Servername hsm
* SKIPACLUP YES
```

4

**TS016056611: dsmc 'updatectime' does not work as documented**

- client user option for AIX® and Linux® clients on GPFS file systems only
- Default: 'no' (ignore ctime)
- **'yes'**: checks the change time (ctime attribute) during a backup operation. If the ctime attribute changed since the last backup operation, **the ctime attribute is updated** on the IBM® Storage Protect server. The object is not backed up unless it has either ACLs or extended attributes.

**But it does backup the whole file!!!**

**SKIPACLUPdatecheck yes** : If only ACLs are updated on a file, the next incremental backup will not recognize this ACL update, and the file is not backed up.

**But it triggered a Metadata update only on ctime changes**

# BACKUP CLIENT MIGRATION

## Summary

- Client migration from AIX/Power to x86 Linux to backup GPFS file system is working
- Renaming of mount points can be handled by keeping the old backup data
- Copying files to a new location (moving files to a new device) and keeping the path/mount point is also fine

Copying user data to a new storage cluster using a new mount point does not enforce a new backup

| Filesystem | Usage (GB) | Usage (Inodes) | Migration Method | old mount point | New mount point | Backup after migration |
|---|---|---|---|---|---|---|
| **arch** | 27.262.976 | 22.400.480 | mmrestripefs | /arch | /p/arch1 | **Keep old data** |
| **arch2** | 19.996.344 | 8.879.869 | mmrestripefs | /p/arch2 | /p/arch2 | **Keep old data** |
| **largedata** | 15.578.638 | 110.496.477 | Copy/AFM | /p/largedata | /p/data1 | **Keep old data** |
| **fastdata** | 6.590.133 | 110.923.593 | Copy/AFM | /p/fastdata | /p/data1 | **New backup** |
| **project** | 3.625.717 | 702.994.066 | Copy/AFM | /p/project | /p/project1 | **Keep old data** |
| **home** | 28.608 | 55.590.547 | Copy/AFM | /p/home | /p/home | **Keep old data** |

# GPFS-HSM: STUBFILE MIGRATION

# STUBFILE MIGRATION - BACKGROUND

## The Jülich GPFS-HSM file system for archiving user data

- Jülich is running two HSM file systems since decades:

```
[root@just6mgr04 hpo22]# mmlsfs arch -V
flag                value                    description
------------------- ------------------------ -----------------------------------
 -V                 25.00 (5.1.1.0)          Current file system version
                     9.03 (3.1.0.0)          Original file system version
```

| Filesystem | Usage (GB) | Usage (Inodes) |
|---|---|---|
| **arch** | 27.262.976 | 22.400.480 |
| **arch2** | 19.996.344 | 8.879.869 |

- User interface is POSIX file system, files are stored on disk (resident)
- The GPFS Information Lifecycle Management (ILM) will move the data to
  - File must be in backup
  - File must be older than 10 days
- Accessing the file (content) will trigger a Recall. File content will be stored

```
define(
    weight_expression,
    (CASE
        WHEN mb_allocated < 2  THEN 0
        WHEN access_age   < 10 THEN access_age
        WHEN is_premigrated    THEN KB_ALLOCATED * access_age
        ELSE                       mb_allocated * access_age
    END)
```

```
[zdv124@judac01:/p/arch1/zam/zdv124> ls -lisah
total 320K
  407977  128K drwx------   2 zdv124 zam   64K May 18 10:01 .
  407555  128K drwxr-xr-x 316 root   sys   64K May 24 15:00 ..
18062260   64K -rw-r--r--   1 zdv124 zam     5 Sep  2  2011 datum.txt
12920848     1 -rw-r--r--   1 zdv124 zam   12G Jun  3  2015 Vervet_s0050_tiff.tgz

[root@just6tsm11 zdv124]# dsmls  Vervet1818_60mu_70ms_s0050_tiff.tgz test.txt
IBM Storage Protect
Command Line Space Management Client Interface
…
      ActS         ResS         ResB   FSt    FName
 12188319595         1024            1   m      Vervet1818_60mu_70ms_s0050_tiff.tgz
          29           29            1   r      test.txt
```

```
' EXEC
cipts/gpfs_scripts/mmpolicyExec-hsm.sample' OPTS '-v'

1' MIGRATE FROM POOL 'system'
,35)
xpression)
```

# STUBFILE MIGRATION - CHALLENGES

## Fact collection & Design changes

- GPFS-HSM client has to move from AIX/Power (Big Endian) to Linux(x86 (Little Endian)
  - **Space Management clients on AIX and Linux use different data formats for the stub files**
- IBM provides **`StubConversionTool`** to convert the data format in two steps:
  1. attributes of all migrated files are converted into special attribute which are platform independent (serialize)
  2. these platform independent attributes are converted into attributes according to the target OS (deserialize)
- **IBM: The conversion does not trigger any recalls, re-migrations or new backups.**
- Requirements:
  - Update GPFS + GPFS-HSM Client (SP) to a tested version. (We used GPFS 5.1.9.2 and SP client 8.1.22.0)
  - Request the Stubfile Conversion Tool (TS015973491: EFIX for the stubbfilemigration tool in GPFSHSM)
  - Bring file system in a consistent state
    - **Ensure all data is in backup**
    - **Data is resident or migrated (premigrated is also fine)**
    - **Reconcile file system to synchronize the TSM database** (e.g. remove object from the TSM database for files which were deleted by the user)
  - **Stop all backups, migrations, recalls during the conversion**

# STUBFILE MIGRATION – GET CONSISTENT STATE

## 1. Ensure all data are in backup

- Start Maintenance: unmount file system on all clients
- Remove exports (GPFS deny access for all remote cluster)
- Create Backup of current state
  - run rebuild shadow database
  - run mmbackup to backup all data (include migrated)

```
mmbackup /<arch|arch2>  --backup-migrated  --incremental-backup-threads 4 --max-incremental-backup-count 100 --selective-backup-threads 10 \
                        --max-selective-backup-count 100 --expire-threads 4 --max-expire-count 10000  --noquote --tsm-servers tsmarch -v
```

| Filesystem | Rebuild Shadow DB | Backup include migrated (no changes) |
|---|---|---|
| **arch** | 17.06.2024 (52 m) | finished 19.06.2024 14:18:39 |
| **arch2** | 17.06.2024 (18 m) | finished 18.06.2024 09:19:18 |

## 2. Migrate all premigrated files

- Prepare GPFS Policy rule
- Run migration
  `mmapplypolicy <arch|arch2> -P <Policy file>` …

```
RULE EXTERNAL POOL 'hsm' EXEC '/var/mmfs/etc/mmpolicyExec-hsm.sample' OPTS '-v'


/* Exclude .SpaceMan  */
RULE 'exclude_spaceman'       EXCLUDE WHERE PATH_NAME LIKE '%/.SpaceMan/%'


RULE 'thresholdMigration' MIGRATE FROM POOL 'system'
 TO POOL 'hsm'
 WHERE (MISC_ATTRIBUTES LIKE '%M%' AND KB_ALLOCATED > 0
```

| Filesystem | Usage (GB) | Usage (Inodes) | moved data (GB) | Moved data (#) | duration |
|---|---|---|---|---|---|
| **arch** | 27.262.976 | 22.400.480 | 60.531 | 12.425 | 00:13:00 |
| **arch2** | 19.996.344 | 8.879.869 | 217.395 | 908 | 00:48:13 |

# STUBFILE MIGRATION – GET CONSISTENT STATE

## 3. Reconcile file systems

- Execute reconcile: **`dsmreconcileGPFS.pl … /arch`**

```
…
## create policy rule      : Wed Jun 19 14:41:27 2024 ####
## invoke policy engine    : Wed Jun 19 14:41:27 2024 ####
## prepare / sort filelist : Wed Jun 19 14:43:33 2024 ####
## number of files         : 8601867
## invoke dsmreconcile     : Wed Jun 19 14:48:50 2024 ####
…
Reconciling '/arch' file system:
  Querying the IBM Storage Protect server for a list of migrated files...
ANS5303I ***** 06/19/24   14:49:51 Processed      74383 IBM Storage Protect
…
ANS5303I ***** 06/19/24   16:39:51 Processed    8598729 IBM Storage Protect
         Received 8601340 entries
  Expiring migrated files on the server...
         6 files exceeded the 7 day expiration period; removed from server
         22 files newly marked for expiration
         0 files previously marked for expiration
  Updating migrated files on the server...
          2512 files updated on the IBM Storage Protect server
          8601312 valid IBM Storage Protect server objects found
          8601301 file list objects processed
          0 orphan files found
ANS9250I File system '/arch' reconciliation completed.

## dsmreconcile ends at     : Wed Jun 19 16:40:0   "###
```

1

```
…
## create policy rule      : Wed Jun 19 16:53:06 2024 ####
## invoke policy engine    : Wed Jun 19 16:53:06 2024 ####
## prepare / sort filelist : Wed Jun 19 16:56:07 2024 ####
## number of files         : 8601867
## invoke dsmreconcile     : Wed Jun 19 17:01:49 2024 ####
…
Reconciling '/arch' file system:
  Querying the IBM Storage Protect server for a list of migrated files...
ANS5303I ***** 06/19/24   17:02:50 Processed      79185 IBM Storage Protect server and      79185 file list
objects *****
…
ANS5303I ***** 06/19/24   18:49:50 Processed    8553154 IBM Storage Protect server and    8553121 file list
objects *****
         Received 8601334 entries
  Expiring migrated files on the server...
         0 files exceeded the 7 day expiration period; removed from server
         0 files newly marked for expiration
         22 files previously marked for expiration
  Updating migrated files on the server...
          8 files updated on the IBM Storage Protect server
          8601312 valid IBM Storage Protect server objects found
          8601301 file list objects processed
          0 orphan files found
ANS      File system '/arch' reconciliation completed.

#        econcile ends at     : Wed Jun 19 18:50:40 2024 ####
```

2

Missed to set **`migfileexpiration 0`** in dsm.sys

# STUBFILE MIGRATION

## Conversion process

- Get list of migrated files using the GPFS policy engine
  **`mmapplypolicy /arch -P migrated_policy.txt -I defer -f /tmp/arch.list`**

```
define(
    is_migrated,
    (MISC ATTRIBUTES LIKE '%M%' )
```

| Filesystem | Total files (Policy) | Migrated files (Policy) |
|---|---|---|
| **arch** | 21.645.952 | 8.601.878 |
| **arch2** | 8.867.516 | 3.696.586 |

**Failed for 567 files**

```
SCT:09.39.41:ERROR:dm_get_dmattr() failed
SCT:09.39.41:WARN :file /arch/…/3dt_decay_stat skipped from processing or processing failed
```

**Background:** Files are missing extended attribute "`HSMexOBjID`" which was introduced in the past by a SP-GPFS-HSM upgrade. The failed files were created before that update and were not accessed afterwards.

**Fix:** restore from backup

- Record the current HSM configuration of the file system
  **`dsmmigfs query <filesystem> -detail`**
- Delete or rename the **`.SpaceMan`** folder in the file system root
- On AIX (source) node run the StubConversionTool for serialization. We used a Python script to split file list in 10 chunks and call the tool in parallel.
  **`./StubConversionTool_AIX_8.1.21.0 -serialize -filelist=… -filesystem=/arch`**
- Afterwards execute on Linux (target) node the StubConverionsTool again to deserialize (also using python script to do it in parallel)
  **`./StubConversionTool_Linuxx86_8.1.21.0 -deserialize -filelist=... -filesystem=/arch`**

| Filesystem | Serialization | deserialization |
|---|---|---|
| **arch** | 01:26:50 | 00:53:18 |
| **arch2** | 00:37:15 | 00:16:12 |

# STUBFILE MIGRATION

## Setting up HSM on target node

- Install HSM client: `rpm -ivh ./8.1.22.0-20240402A/hsmgpfs/TIVsm-HSM.x86_64.rpm`
- Add file system to HSM

```
[root@just6tsm11 ~]# dsmmigfs add /arch2
IBM Storage Protect
Command Line Space Management Client Interface
  Client Version 8, Release 1, Level 22.0 - 216044 efix
…
Adding HSM support for /arch2 ...
ANS9087I Space management is successfully added to file system /arch2.

[root@just6tsm11 ~]# dsmdf
…
ManFs            FsSt    MigB    PmigB   MigF    PmigF   FreeI   FreeB
/arch2           a       0K      0K      0       0       31.13M  1.11P
```

- Put EFIX `dsmreconcileGPFS.pl` in place

```
[root@just6tsm11 bin]# pwd
/opt/tivoli/tsm/client/hsm/multiserver/bin
[root@just6tsm11 bin]# diff dsmreconcileGPFS.pl dsmreconcileGPFS.pl_patched
349c349
<     my $dsmreconcileCommand = "dsmreconcile -o -d -filelist=$prepOutput $fsMntPt 2>&1";
---
>     my $dsmreconcileCommand = "dsmreconcile -o -d -forceupdate -filelist=$prepOutput $fsMntPt 2>&1";
```

# STUBFILE MIGRATION

## Recreate HSM space management database (.SpaceMan folder)

- Run special **dsmreconcileGPFS.pl**

```
## dsmreconcile starts at  : Mon Jun 24 14:
…
## create policy rule      : Mon Jun 24 14:
## invoke policy engine    : Mon Jun 24 14:
## prepare / sort filelist : Mon Jun 24 14:
## number of files         : 3696566
## invoke dsmreconcile     : Mon Jun 24 14:
…
Reconciling '/arch2' file system:
  Querying the IBM Storage Protect server f
ANS5303I ***** 06/24/24   14:49:24 Processe
…
ANS5303I ***** 06/24/24   15:26:25 Processe
ANS9616E dsmreconcile: cannot get migration
```

```
[root@just6tsm11 bin]# diff
363c363,368
<         die "$_" if $_ =~ m/
---
>         if ( $_ =~ m/^(ANS.{
>         {
>             $returnCode = 10;
>         }
>
>         # die "$_" if $_ =~ m/^(ANS.{4}E).*/;
```

```
## dsmreconcile starts at  : Wed Jun 26 16:45:50 2024 ####
…
## create policy rule      : Wed Jun 26 16:45:50 2024 ####
## invoke policy engine    : Wed Jun 26 16:45:50 2024 ####
## prepare / sort filelist : Wed Jun 26 16:46:10 2024 ####
sort -T "/tmp" /tmp/dsmreconcile.gpfs.1719413149.list.reorg.out | tr -s '\n' '\0' > /tmp/dsmreconcile.gpfs.1719413149.list.prep.out
## number of files         : 3696566
## invoke dsmreconcile     : Wed Jun 26 16:46:48 2024 ####
dsmreconcile -o -d -forceupdate -filelist=/tmp/dsmreconcile.gpfs.1719413149.list.prep.out /arch2 2>&1
…
Reconciling '/arch2' file system:
  Querying the IBM Storage Protect server for a list of migrated files...
ANS5303I ***** 06/26/24   16:47:48 Processed      66844 IBM Storage Protect server and      66844 file list objects *****
…
ANS5303I ***** 06/26/24   17:26:49 Processed    2361345 IBM Storage Protect server and    2361345 file list objects *****
ANS9616E dsmreconcile: cannot get migration information for /hhh07/hhh073/CBL3D/Re100/5120x840x5120-70000-20130211/flow/flow.15500.1 on /arch2
…
ANS9616E dsmreconcile: cannot get migration information for /hhh07/hhh073/CBL3D/Re100/5120x840x5120-70000-20130211/flow/flow.21500.3 on /arch2
ANS5303I ***** 06/26/24   17:27:49 Processed    2407956 IBM Storage Protect server and    2407937 file list objects *****
…
ANS5303I ***** 06/26/24   17:50:49 Processed    3693384 IBM Storage Protect server and    3693365 file list objects *****
         Received 3696585 entries
  Expiring migrated files on the server...
         0 files exceeded the 7 day expiration period; removed from server
         0 files newly marked for expiration
         0 files previously marked for expiration
  Updating migrated files on the server...
          3696566 files updated on the IBM Storage Protect server
          3696585 valid IBM Storage Protect server objects found
          3696566 file list objects processed
          0 orphan files found
ANS9250I File system '/arch2' reconciliation completed.

## dsmreconcile ends at    : Wed Jun 26 17:50:54 2024 ####
```

# STUBFILE MIGRATION

## Fixing the ANS9616E state

- https://www.ibm.com/support/pages/ans9616e-dsmreconcile-cannot-get-migration-information-fsfile1
- **Cause: Incorrect handling of HSM OS/hardware migration can cause the filesystem ID to change, which can then cause the DMAPI handle to change**
- **Run the "dsmmigquery -serverinfo /fs/file1" HSM client command. For example :**

```
Alias : file1
insert date: 12/03/09 02:44:47
extobjid : 0101040C000000000A010D2F06DB1A2CBC00AC405D2DB6937FCE5054
migr state : MIGRATED
times : c 10/22/09 02:48:18 m 10/22/09 00:27:00 a 12/21/09 15:52:58
mode OCT : 100640 uid: 904 gid: 206 size:4525124
inode : 1170182 ACL size: 0 checksum: 256
DMI handle : 8000003200000006-000000000011DB06-00000000BCB87AF3-0000010000000000
Alias : file1
insert date: 05/07/10 01:00:11
extobjid : 0101040C000000000A010D2F06E439C0240EFA2539CD21ABCE381884
migr state : MIGRATED
times : c 12/21/09 02:06:33 m 10/22/09 00:27:00 a 12/21/09 02:06:33
mode OCT : 100640 uid: 904 gid: 206 size:4525124
inode : 1170182 ACL size: 0 checksum: 256
DMI handle : 8000003500000006-000000000011DB06-00000000D390CF8F-0000010000000000
```

- **The above outputs reports 2 migrated copies of the same file, each with a different DMI handle, inserted (migrated) at different times. The DMI handle is unique for each filesystem. Un-proper filesystem change handling may cause the DMI handle change. The changed DMI handle can result in an orphaned migrated copy on server and the ANS9616E error during reconcile.**
- **Solution how to fix it provided**

# STUBFILE MIGRATION

## Verification

- **Reconcile:**

22 Objects marked as deleted

```
[root@just6tsm11 met_data]# dsmls cera20c.tar.gz
IBM Storage Protect
Command Line Space Management Client Interface
  Client Version 8, Release 1, Level 22.0 - 216044 efix
  Client date/time: 06/28/24   08:58:27
(c) Copyright by IBM Corporation and other(s) 1990, 2024. All Rights Reserved.

       ActS          ResS          ResB    FSt     FName
  4300160694          1024             1    m       cera20c.tar.gz
```

- **System information:**

```
[root@just6tsm11 met_data]# dsmls cera20c.tar.gz
IBM Storage Protect
Command Line Space Management Client Interface
  Client Version 8, Release 1, Level 22.0 - 216044 efix
  Client date/time: 06/28/24   09:01:27
(c) Copyright by IBM Corporation and other(s) 1990, 2024. All
Rights Reserved.

       ActS          ResS          ResB    FSt     FName
  4300160694    4300160694       4199424    p       cera20c.tar.gz

[root@just6tsm11 met_data]# md5sum cera20c.tar.gz
3ec97f0d3b393d7f2f9b87738ce9a251  cera20c.tar.gz
```

- **Recall test:**

```
[root@just6tsm11 met_data]# dsmrecall -d cera20c.tar.gz
IBM Storage Protect
Command Line Space Management Client Interface
  Client Version 8, Release 1, Level 22.0 - 216044 efix
  Client date/time: 06/28/24   08:58:58
(c) Copyright by IBM Corporation and other(s) 1990, 2024. All Rights Reserved.

Session established with server TSMARCH: AIX
  Server Version 8, Release 1, Level 17.000
  Server date/time: 06/28/24   08:58:58  Last access: 06/28/24   08:55:17

Recalling 4,300,160,694 /arch2/slmet/slmet111/met_data/cera20c.tar.gz [Done]

Recall processing finished.
```

### Finalizing/Cleanup

- Restore original dsmreconcileGPFS.pl
- Prepare systemd to run reconcile once per week
- Prepare systemd to run mmBackup daily
- rename mount point

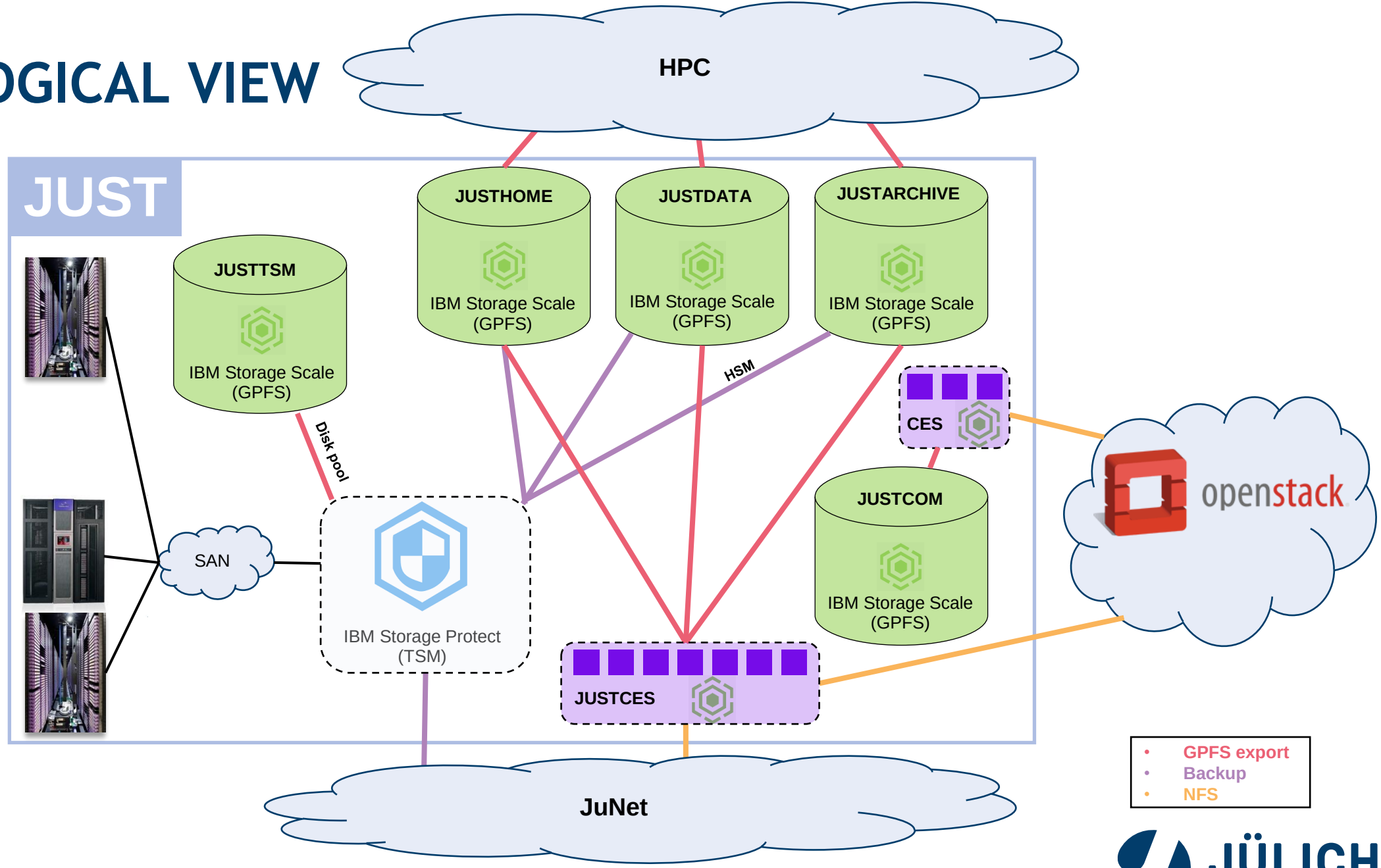
# SP MIGRATION FINISHED SUCCESSFULLY

## Summary

- We got rid of AIX/Power
  - ✓Server – Server migration
  - ✓GPFS backup client migration
  - ✓GPFS-HSM (client) migration
- We changed the GPFS mount points by keeping the backup data
- Detailed action plan prepared and tested by IBM (part of the contract)
- JSC & IBM together performed all actions (part of the contract)
- We learned a lot of new stuff for our Storage Protect environment
- Backup of your data is very important

# JUST-LOGICAL VIEW

## Today



# THANK YOU

Stephan Graf (JSC)

Martin Lischewski (JSC)

Fabián Kuhl (IBM)

Andre Gaschler (IBM)

# MIGRATION METHODS

- Mmrestripefs:
  - Well known procedure
  - Only possible for $ARCHIVE (will remain in old cluster JUSTGSS)
- Mpifileutils
  - Efficient way to copy/compare data using MPI
- GPFS AFM
  - IBM recommended way
  - Will be used for the large data sets

https://www.ibm.com/docs/en/storage-scale/5.1.9?topic=reference-active-file-management-afm

# STORAGE IN 2024

**JUST6**

- 154 PB raw, 116 PB net
- 600 GB/s write, 700 GB/s read
- 22 TB Spinning Disk
- IBM SSS6000
- IBM Storage Scale
- 100 GE fabric, like JUST5

**ExaSTORE**

- 308 PB raw, 229 PB net
- 1,1 TB/s write, 1,4 TB/s read,
- 22 TB Spinning Disk
- IBM SSS6000
- IBM Storage Scale
- InfiniBand and 100GE

**ExaFLASH**

- 29 PB raw, 21PB net
- 2,1 TB/s write, 3,1 TB/s read
- 30 TB NVME
- IBM SSS6000
- IBM Storage Scale
- InfiniBand