# Wikiprint Book

Title: File Systems

Subject: DEEP - Public/User\_Guide/Filesystems

Version: 36

Date: 07.05.2024 09:47:24

# **Table of Contents**

File Systems	3
Available file systems	3
Stripe Pattern Details	4
Additional infos	4
Notes	4

## File Systems

## Available file systems

On the DEEP-EST system, three different groups of file systems are available:

- the <a href="2JSC GPFS file systems">2JUST</a> and mounted on all JSC systems;
- the DEEP-EST parallel BeeGFS file systems, available on all the nodes of the DEEP-EST system;
- · the file systems local to each node.

The users home folders are placed on the shared GPFS file systems. With the advent of the new user model at JSC (?JUMO), the shared file systems are structured as follows:

- \$HOME: each JSC user has a folder under /p/home/jusers/, in which different home folders are available, one per system he/she has access to.

  These home folders have a low space quota and are reserved for configuration files, ssh keys, etc.
- \$PROJECT: In JUMO, data and computational resources are assigned to projects: users can request access to a project and use the resources associated to it. As a consequence, each user can create folders within each of the projects he/she is part of (with either personal or permissions to share with other project members). For the DEEP project, the project folder is located under /p/project/cdeep/. Here is where the user should place data, and where the old files generated in the home folder before the JUMO transition can be found.

The DEEP-EST system doesn't mount the \$SCRATCH file systems from GPFS, as it is expected to provide similar functionalities with its own parallel and local file systems.

The following table summarizes the characteristics of the file systems available in the DEEP-EST and DEEP-ER (SDV) systems. Please beware that the \$project (all lowercase) variable used in the table only represents any JuDoor project the user might have access to, and that it is not really exported on the system environment. For a list of all projects a user belongs to, please refer to the user's <a href="YJuDoor page">YJuDoor page</a>. Alternatively, users can check the projects they are part of with the jutil application:

\$ jutil user projects -o columns

Mount Point	User can writelread toffrom	Cluster	Туре	Global / Local	SW Vention	Stripe Pattern Details	Maximum Measured Performance (see footnotes)	Description	Other
(phone	/phome/juses	SEW PLST	GPFS exported via NFS	Global				JUST GPFS Home directory; used only for configuration files.	
/p/project	/ріртірсі Бре	SOV. DEEP-EST	GPFS exponed via NFS	Global				JUST GPFS Project directory; GPFS main storage file system; not suitable for performance nelevant applications or benchmanking	
.larch	lands/Egroject	login rade only (deepv)	GPFS exported via NFS	Giobal				JUST GPFS Archive directory. Long-term storage solution for data used in a long time; Data migrated to tape - not insended for lots of small files. Recovery can take days.	If you plan to use the active, please get in contact to the system administration (e.g. via the support mailing lair). You call further information and some bints on using the archive these
Jafara.	iatum	DEEP-EST	DeeGFS	Global	BeeGFS 7.25			Fast work file system, no backup, hence not meant for permanent data storage	
/work_old	/work_old/\$pn	JAMESP-EST	BeeGFS	Global	BeeGFS 7.2.5			Work file system, no backup, hence not meant for permanent data storage. Deprecated	
/scratch	/scnach	DEEP-EST	afe local partition	Local				Node local soratch file system for temporary data. Will be cleaned up after job finishes. Size differs on the modules] "Recommend to use instead of Intop for storing temporary files	and
ltvme/scratch	/invine/scratch	DAM partition	local SSD (sfs)	Local				Screech file system for temporary data. Will be cleaned up after job finishes() *1.5 TB I treat Optano SSD Data Center (DC) P4800X (NVMe P450) 44, 2.57, 3D XPoint()	
/nvme/scratch	linvnelscratch	DAM partition	local SSID (molt)	Local				Screach file system for temporary data. Will be cleaned up after job finishes() *1.5 Total Optane SSD Data Center (DC) P4800X (NVMs PCliež M, 2.57, 3D XPoint()	
ilpmem/scrand	i įpmemisastd	DAM partition	DCPMM in appdirect mode	Local*			2.2 GBhs simple dd teat in dp-dam01		"3 TB is op-dam(01,02 2 TB is op-dam(01,02 2 TB is op-dam(03-16 issel Optane DC Persistent Memory (DCPMM) 256GB DMMs based on isselfs 2D XF0st non-volatile mon-volatile echnology

#### **Stripe Pattern Details**

It is possible to query this information from the deep login node, for instance:

```
manzano@deep $ fhgfs-ctl --getentryinfo /work/manzano
Path: /manzano
Mount: /work
EntryID: 1D-53BA4FF8-3BD3
Metadata node: deep-fs02 [ID: 15315]
Stripe pattern details:
+ Type: RAID0
+ Chunksize: 512K
+ Number of storage targets: desired: 4
manzano@deep $ beegfs-ctl --getentryinfo /sdv-work/manzano
Path: /manzano
Mount: /sdv-work
EntryID: 0-565C499C-1
Metadata node: deeper-fs01 [ID: 1]
Stripe pattern details:
+ Type: RAID0
+ Chunksize: 512K
+ Number of storage targets: desired: 4
```

#### Or like this:

```
manzano@deep $ stat -f /work/manzano
File: "/work/manzano"
         Namelen: 255
                              Type: fhgfs
Block size: 524288 Fundamental block size: 524288
Blocks: Total: 120178676 Free: 65045470 Available: 65045470
Inodes: Total: 0
                      Free: 0
manzano@deep $ stat -f /sdv-work/manzano
File: "/sdv-work/manzano"
  ID: 0 Namelen: 255
                             Type: fhgfs
Block size: 524288 Fundamental block size: 524288
Blocks: Total: 120154793 Free: 110378947 Available: 110378947
Inodes: Total: 0
                        Free: 0
```

See <a href="http://www.beegfs.com/wiki/Striping">http://www.beegfs.com/wiki/Striping</a> for more information.

#### Additional infos

Detailed information on the BeeGFS Configuration can be found ?here.

Detailed information on the BeeOND Configuration can be found ?here.

Detailed information on the Storage Configuration can be found ?here.

Detailed information on the Storage Performance can be found ?here.

#### Notes

• dd test @dp-dam01 of the DCPMM in appdirect mode:

```
[root@dp-dam01 scratch]# dd if=/dev/zero of=./delme bs=4M count=1024 conv=sync 1024+0 records in 1024+0 records out 4294967296 bytes (4.3 GB) copied, 1.94668 s, 2.2 GB/s
```

- The /work file system which is available in the DEEP-EST prototype, is as well reachable from the nodes in the SDV (including KNLs and ml-gpu nodes) but through a slower connection of 1 Gb/s. The file system is therefore not suitable for benchmarking or I/O task intensive jobs from those nodes
- Performance tests (IOR and mdtest) reports are available in the BSCW under DEEP-ER → Work Packages (WPs) → WP4 → T4.5 Performance measurement and evaluation of I/O software → Jülich DEEP Cluster → Benchmarking reports:
   ?https://bscw.zam.kfa-juelich.de/bscw/bscw.cgi/1382059
- Test results and parameters used are stored in JUBE:

```
user@deep $ cd /usr/local/deep-er/sdv-benchmarks/synthetic/ior
user@deep $ jube2 result benchmarks

user@deep $ cd /usr/local/deep-er/sdv-benchmarks/synthetic/mdtest
user@deep $ jube2 result benchmarks
```