Tools and procedures for the Production Environment: Data Archiving, schedulers, Application software, ...

Since our HPC systems are shared among many users, long production jobs should be submitted using a scheduler. This guarantees that the access to our resources is as fair as possible.

Roughly, there are two different modes to use an HPC system:

- **interactive**, for data movement, archiving, code development, compilations, basic debugger usage: also for very short test runs and general interactive operations. A task in this class should not exceed 10 minutes CPU-time and is free of charge on HPC systems with the current billing policy.
- **batch**, for production runs. Users must prepare a shell script containing all the operations to be executed in batch mode, once the requested resources are available and assigned to the job. The job then starts and executes on compute nodes of the cluster. Remember to put all your data, programs and scripts in the $WORK or $CINECA_SCRATCH filesystem, that are the best storage areas accessible to execution nodes.

You must have valid active projects on the system in order to run batch jobs. Moreover, remember that on our systems there may be specific policies for the use of project budgets.

On all of our current HPC systems, the queuing system or scheduler is **SLURM** (that substitutes **PBSpro** starting from Jan 2018). The usage of slurm is described in separate section under this chapter, details are given for each single platform in the platform specific page.

The Software Catalog

CINECA offers a variety of third-party applications and community codes that are installed on its HPC systems. Most of third-party software is installed using software modules mechanism (see "The module command" later in this section).

The available packages and their detailed descriptions can be viewed in the full catalog or organised by discipline in our web site (http://www.hpc.cineca.it/content/resources/) by selecting "software" and "Application Software for Science".

If you do not see an application you are interested in, or if you have question about software that is currently available, please contact our specialists (superc@cineca.it).

The "module" command

All software programs installed on the CINECA machines (see *The Software Catalog* previous in this section) are available as modules.

A basic default modules environment is already set up by the system login configuration files.

In order to have a list of available modules and select a specific one you have to use the **module** command. The following table contains its basic options:
As you will see by typing "module avail", the software modules are collected in different profiles (base, advanced,...) and organized by functional category (compilers, libraries, tools, applications,...).

In order to detect all profiles, categories and modules available on both Marconi and Galileo the command "modmap" is available:

> modmap

ATTENTION: Remember to load the needed modules in batch scripts too, before using the related applications.

**Backup policies**

The $HOME filesystem is guaranteed by daily backups. In the following the standard backup policies are reported; in particular cases, a different agreement is possible: contact the HPC support (super@cineca.it) for further details.

The backup procedure runs daily and we preserve a maximum of three different copies of the same file. Older versions are kept for 1 month. The last version of deleted files is kept for 2 months, then definitely removed from the backup archive.

**Archive your data: $TAPE**

Please note: the previous data archive was managed via "cart" command, which will be available until December 31st, 2014.

A new Tape library, more large and powerful, has been acquired in October 2014. It will use a multi-level (LTFS) technology (data copied in a given filesystem will move to tape in automatic fashion following specific policies), and will be accessible via the environment variable $TAPE. The default user quota on tape, to be created upon request to super@cineca.it, is 1TB. Larger quota may be required in case of need. Unless differently specified or agreed, the data on tape follows the lifetime of the owner's username and will be preserved on tape for one year after the expiration of the last user's project.

**The project-oriented storage areas: $WORK and $DRES**

In 2014 the new storage areas $WORK and $DRES have been added for HPC users. From that time on, the $WORK area is available on all HPC systems.

The $WORK area is available for each active project and it is created by default.

The $DRES area (Data RESource) is available only on demand in order to host data files to be used by more than one project, and/or from different platforms and/or to archive data for medium/long time.

Both $WORK and $DRES are conceived as collaborative areas where all collaborators of a given project(s) can share data in secure way.

$WORK and $CINECA_SCRATCH are conceived as working directories for large files used and produced by batch jobs. Also the blocking features make these areas more suitable for large binary files.

The main differences are:
<table>
<thead>
<tr>
<th>$\text{CINECA_SCRATCH}</th>
<th>$\text{WORK}</th>
</tr>
</thead>
<tbody>
<tr>
<td>is defined on the username base, each username has its own area (and only one)</td>
<td>is defined on the project base. Each project has its own area where all collaborators can write. Each collaborator has as many $\text{WORK}$ areas as active projects.</td>
</tr>
<tr>
<td>a clean-up procedure could be active (files older than 40 days are cancelled daily)</td>
<td>data are preserved up to the end of the project</td>
</tr>
<tr>
<td>there is no quota, the only limitation is the physical size of the device</td>
<td>there is a default quota of 1 TB. Motivated requests for quota increase will be taken into account</td>
</tr>
<tr>
<td>by default files are public (read only). The user can change the permission (chmod) and make files private. It is not possible to restrict access to the group (all username share the same mail unix group)</td>
<td>by default files are private. The user can change the permission (chmod) and make files visible (R o R/W) with project collaborators</td>
</tr>
</tbody>
</table>

$\text{DRES}$ presents the same features than $\text{WORK}$, but it is conceived for sharing data across the different platforms. In fact, it is mounted ONLY on the login nodes of the HPC systems, so to favour the sharing among different systems.

### File permissions

$\text{WORK}$ and $\text{DRES}$ are environmental variables, automatically set in the user environment.

$\text{WORK}$ variable points to a directory (fileset) specific for one of the user projects:

```
/gpfs/work/<account_name>
```

$\text{DRES}$ variable points to a space where all of the dres are defined:

```
/gss/gss_work/
```

In order to use a specific dres type the following path:

```
$\text{DRES}/\text{<dres}_\text{name>}
```

The owner of the root directory is the "Principal Investigator" (PI) of the project and the "owner" of the dres, the group corresponds to the name of the project and the name of the dres, the access permissions are:

- **own**: rwx
- **group**: rwx
- **other**: -

In this way all project collaborators, sharing the same project group, can read/write into the project/dres fileset, whereas other users can not.

Users are advised to create a personal subdirectory under $\text{WORK}$ and $\text{DRES}$. By default, files into the subdirectory are private, but the owner can easily share the files with the other collaborators by opening the subdirectory:

```
chmod 777 mydir
chmod 755 mydir
```

Since the $\text{WORK}$$\text{DRES}$ fileset is closed to non collaborators, the data sharing is active only for the collaborators of the project.

### Pointing $\text{WORK}$ to a different project: the chprj command

The user can modify the project pointed to by the variable (default project) using the command "change project" (chprj):

```
chprj -l
```

List all your accounts pointing out if active or completed and the default project.

```
chprj -d <account_no>
```

Set $\text{WORK}$ to point to the <account_no> project

```
chprj -h
chprj --help
```

Print the help page

### Monitoring the WORK and DRES occupancy
You should use the "cindata" command instead, listing all filesets containing any file owned by you.

```
> cindata

------------------------------------------ASyncronous Data report------------------------------------------
--------------------------------------------|----------USER USAGE--------|------TOTAL

<table>
<thead>
<tr>
<th>USER</th>
<th>AREADESCR</th>
<th>AREAID</th>
<th>FRESH</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPACE</td>
<td>QTA</td>
<td>QTA%</td>
<td>SPACE</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>erossi00</td>
<td>/gpfs/work/cin_tech</td>
<td></td>
<td>-3hou</td>
</tr>
<tr>
<td>1G</td>
<td>--</td>
<td>--%</td>
<td>2.0G</td>
</tr>
<tr>
<td>erossi00</td>
<td>/gpfs/work/cin_staff</td>
<td></td>
<td>-3hou</td>
</tr>
<tr>
<td>11G</td>
<td>--</td>
<td>--%</td>
<td>9.1T</td>
</tr>
<tr>
<td>erossi00</td>
<td>/gpfs/work/</td>
<td></td>
<td>-3hou</td>
</tr>
<tr>
<td>12G</td>
<td>--</td>
<td>--%</td>
<td>246T</td>
</tr>
<tr>
<td>erossi00</td>
<td>/gss/gss_work/DRES_test</td>
<td></td>
<td>3mon</td>
</tr>
<tr>
<td>erossi00</td>
<td>/gss/gss_work/DRES_test</td>
<td></td>
<td>3mon</td>
</tr>
<tr>
<td>1.4G</td>
<td>--</td>
<td>--%</td>
<td>109G</td>
</tr>
<tr>
<td>erossi00</td>
<td>/gss/gss_work/DRES_test</td>
<td></td>
<td>3mon</td>
</tr>
<tr>
<td>erossi00</td>
<td>/gss/gss_work/DRES_test</td>
<td></td>
<td>3mon</td>
</tr>
<tr>
<td>1.8T</td>
<td>--</td>
<td>--%</td>
<td>1.8T</td>
</tr>
</tbody>
</table>
```

The erossi00 user has files in two WORK areas ("cin_staff" and "cin Tech" projects) and in one DRES space ("DRES_test"). In "cin_staff", the total amount of data owned by erossi00 is 11GB, the total occupancy of the project area is 9.1TB, its quota is 10TB. The total dimension of the equipment hosting the WORK directory is 854TB (to be doubled in short time), with a total occupancy of 246TB (28.8%).

Note also that these statistics are 3 hours old.

For understanding correctly the occupancy of the DRES go to Data storage and filesystems chapter.

- UG2.5.1: Batch Scheduler SLURM