Working on TGCC

Table of Content
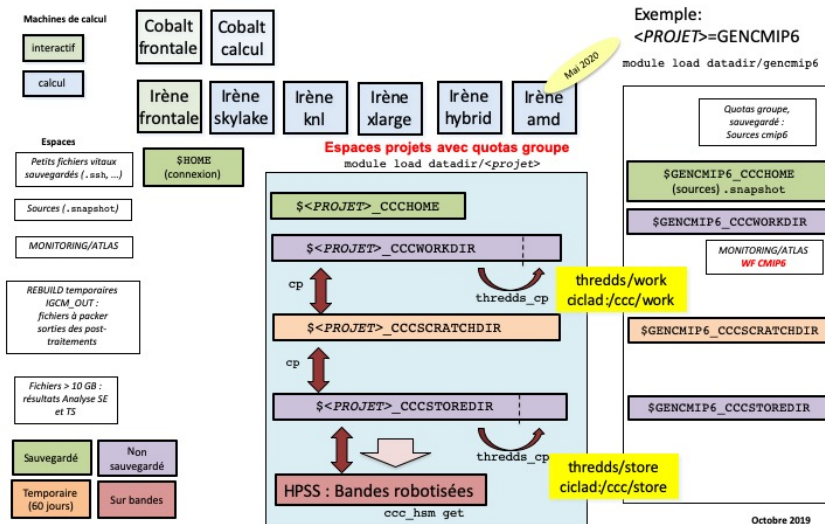
1. TGCC presentation

- ■http://www-hpc.cea.fr/en/complexe/tgcc.htm

2. TGCC's machines and file systems

Les machines et espaces disques du TGCC en une figure

3. How to install your environment on TGCC

- More information on the open-access website: ■http://www-hpc.cea.fr/en/complexe/tgcc.htm

  Online access to the machines' users manual : ■http://www-hpc.cea.fr/tgcc-public/en/html/tgcc-public.html
  - Online access to technical issues and news : ■https://www-tgcc.ccc.cea.fr/en/news/index.html
- TGCC's machine is :
  - Irene (Intel Skylake)
  - Irene-amd (AMD Rome)
- Note: the **$HOME/.snapshot** directory contains hourly, daily, and weekly backups of your $HOME files.

It is important to take the time to install a comfortable and efficient environment.

Your login will be linked to one or several GENCI projects. These projects will give you access to computing hours. Each GENCI project has its own spaces on the filesytem.
You need to work within these specifics spaces, and to have access to the specific environment variables of the project genXXXX you have to use the next command:

```
module switch dfldatadir dfldatadir/genXXXX
```

Add this command to your environment (see paragraph bellow).
You can find more informations on link between GENCI projects and filesytems here

3.1. Irene machine (Intel skylake and AMD Rome)

We suggest the user to use the igcmg environment (in bash) with a copy of the bashrc in his HOME.

```
ryyy999@irene: cp ~igcmg/MachineEnvironment/irene/bashrc  ~/.bashrc
```

Additionally, you need to copy and complete the example of bashrc_irene file to create your favorite environment (alias, module load ...). Don't forget to use it in .bashrc.

```
ryyy999@irene: cp ~igcmg/MachineEnvironment/irene/bashrc_irene ~/.bashrc_irene
ryyy999@irene: vi  ~/.bashrc  # to point your own .bashrc_irene
```

We strongly advice you to add the line `module switch dfldatadir dfldatadir/genXXXX` *in your own .bashrc_irene.*

WARNING : if you have a ~/.profile file, it's better to remove it to avoid any problem during the execution of a simulation with libIGCM

In this environment is specified:

the path to the compiler tool `fcm` and to the `rebuild` tool which recombines output files from a parallel model:

```
export PATH=$(ccc_home -u igcmg)/Tools/fcm/bin:$(ccc_home -u igcmg)/Tools/irene/bin:$PATH
```

- the load of modules giving access to computing or post processing libraries and tools needed on our platform (done in `ccc_home -u igcmg`/MachineEnvironment/irene/env_atlas_irene).

4. Repository IGCM with input files, also called R_IN

The shared repository with input files is stored at TGCC here:

```
R_IN=$CCCWORKDIR/../../igcmg/igcmg/IGCM
```

This folder is noted using the variable R_IN in the comp.card in libIGCM configurations. The folder R_IN is the same and regularly synchronized between the computing centers TGCC, IDRIS, ESPRI mesocenter(ciclad/climserv) and LSCE(obelix). Contact the plateforme groupe if you don't have read access to these files with your login at jean-zay.

5. Project and computing needs

- To find out the computing time used by the projects you are involved in (daily update):

```
ryyy999@irene: ccc_myproject
```

- When you will create a job you need to specify in the header the project from which you will use computing time:

```
#MSUB -A genxxx
```

6. About file systems

6.1. Quotas

To check the available and used storage capacities of `HOME`, `CCCSCRATCHDIR`, `CCCWORKDIR` and `CCCSTOREDIR`:

```
ryyy999@irene: ccc_quota
```

On the Irene machine this command will also return the space used by scratch (a specificity of the Irene machine).

This command has been improved and gives a lot of information : quotas and usage of shared space, type and duration of exception.

6.2. CCCSCRATCHDIR

The `$CCCSCRATCHDIR`directory is often cleaned and only files that are less than 40 days are stored.

6.3. CCCWORKDIR

The `$CCCWORKDIR` directory corresponds to the `$WORKDIR` directory on Irene. It is large but its content is not backed up. Don't forget to do a backup (tar) for important directories.

6.4. CCCSTOREDIR

To manipulate the files in /ccc/store a few commands are useful:

```
# Demigrate a list of files on CCCSTOREDIR, see also "ccc_hsm -h"
ccc_hsm get $CCCSTOREDIR/FILE1 $CCCSTOREDIR/FILE2 ...

# Demigrate recursively the files from a CCCSTOREDIR directory, see also "ccc_hsm -h"
ccc_hsm get -r $CCCSTOREDIR/DIRECTORY

# Find out the used space on CCCSTOREDIR
cd $CCCSTOREDIR ; find . -printf "%y %s %p \n"  | \
    awk '{ SUM+=$2 } END {print "SUM " SUM/1000000 " Mo " SUM/1000000000 " Go" }'

# or use --apparent-size with du :
du -sh --apparent-size
```

6.5. ccc_home command to know directory complete pathname

ccc_home could help you to find directory complete pathname for an other user or for you .

```
>ccc_home -h
ccc_home: Print the path of a user directory (default: home directory).
usage: ccc_home [ -H | -s | -t | -W | -x | -A | -a | -n] [-u user] [-d datadir]
               [-h, --help]

-H, --home          : (default) print the home directory path ($HOME)
-s, -t, --cccscratch : print the CCC scratch directory path   ($CCCSCRATCHDIR)
-X, --ccchome       : print the CCC nfs directory path ($CCCHOMEDIR)
-W, --cccwork       : print the CCC work directory path  ($CCCWORKDIR)
-A, --cccstore      : print the CCC store directory path ($CCCSTOREDIR)
-a, --all           : print all paths
-u user             : show paths for the specified user instead of the current user
-d datadir          : show paths for the specified datadir
-n, --no-env        : do not load user env to report paths
-h, --help          : display this help and exit

> ccc_home -A -u ryyy999
$CCCSTOREDIR/../../genXXX/ryyy999
```

6.6. Storage spaces available from ESGF/THREDDS

To store a file for the first time on esgf/thredds, you must ask for esgf/thredds write access by mail to the TGCC hotline access : `hotline.tgcc@cea.fr`. On Irene, files available on $CCCWORKDIR are candidates to be available from ESGF/THREDDS :

- use `thredds_cp` command (available here : ~igcmg/Tools/irene/thredds_cp)
- files will be hardlinked here : $CCCWORKDIR/../../thredds/login

From a server web, files are available here : ∎https://vesg.ipsl.upmc.fr/thredds/catalog/work_thredds/catalog.html

More information about output data available from ESGF/THREDDS here.

Final simulation outputs are stored in $CCCSTOREDIR/IGCM_OUT and on $CCCWORKDIR/IGCM_OUT regarding the ATLAS and MONITORING directories. These files are then available from ESGF/THREDDS access.

7. Specific directories for projects

You have a main home where you arrive when connecting to irene, called "home de connexion" by the TGCC. You also have a home, a storedir, a workdir, a scratchdir by project. For example if you are working with project gen2201 and gen2212 you will have all following directories:

```
/***/***/home/***/login                    # connexion home, where ***=your lab (lsce, ipsl, etc..)

/***/***/home/gen2201/login     # use it for sources, regular snapshot are in .snapshot
/***/***/home/gen2212/login

/***/store/***/gen2201/login
/***/store/***/gen2212/login

/***/work/***/gen2201/login
/***/work/***/gen2212/login

/***/scratch/***/gen2201/login
/***/scratch/***/gen2212/login
```

IMPORTANT : Check that you have read and write access to above directories (for your projects). Contact TGCC hotline if it is not the case.

On the SCRATCH space any files that stays 60 days without being read or modified will be purged(deleted), as well as any directory that remains empty for 30 days.

After connection to irene, load your project environment as default using the module dfldatadir. For example if you will work on the project gen2201, do following (**we strongly advice you to add the command into your `.bashrc_irene`**):

```
module switch dfldatadir dfldatadir/gen2201
```

By changing the dfldatadir, the variables $CCCHOME, $CCCWORKDIR, $CCCSTOREDIR and $CCCSCRATCHDIR point to the corresponding project directories. $HOME is always the main connection home.

You will also have new environment variables to access working directories :

```
GEN2201_ALL_CCCSCRATCHDIR=/***/scratch/***/gen2201/gen2201
GEN2201_CCCWORKDIR=/***/work/***/gen2201/login
GEN2201_ALL_HOME=/***/***/home/gen2201/gen2201
GEN2201_CCCSTOREDIR=/***/store/***/gen2201/login
GEN2201_CCCSCRATCHDIR=/***/scratch/***/gen2201/login
GEN2201_ALL_CCCWORKDIR=/***/work/***/gen2201/gen2201
GEN2201_HOME=/***/***/home/gen2201/login
GEN2201_ALL_CCCSTOREDIR=/***/store/***/gen2201/gen2201
```

If you previously worked at curie and your directories were in **/*/dsm/login you will now find your data in a specific new project file system "dsmipsl". We recommend to move your data in your genci project file system. The TGCC hotline can help you if you want.**

8. Specific file systems for CMIP6

For gencmip6 project, and only for it, 3 more file systems and 4 more directories are available. Phase 1 have been installed in april 2016. Phase 2 and Phase 3 will come later in 2017 and 2018.

To use them, in interactive mode, you have to do : `module load datadir/gencmip6`.

Since libIGCM_v2.8.1, if you set your project to gencmip6/devcmip6, they are automatically used in place of usual HOME, CCCWORKDIR, CCCSTOREDIR and CCCSCRATCHDIR : `module switch dfldatadir dfldatadir/gencmip6` called from libIGCM.

8.1. GENCMIP6_HOME

- 50 TB
- gencmip6 group quota
- dedicated to sources and scripts
- strongly recommended for CMIP6 sources and simulations scripts

- regular snapshot are taken by the system. See $GENCMIP6_HOME/.snapshot Attention : you need an interactive connexion on a compute node :

```
> ccc_mprun -s -p standard -A devcmip6 -T 1800 -Q test
> cd
> . .bash_login
> cd .snapshot
> ls -l
total 44
drwxr-sr-x. 13 xxx gencmip6 4096 Dec 17 09:47 daily.2017-02-07_0010
drwxr-sr-x. 13 xxx gencmip6 4096 Dec 17 09:47 daily.2017-02-08_0010
...
```

8.2. GENCMIP6_CCCWORKDIR

- 2.5 PB in phase 1, 5 PB in phase 2
- gencmip6 group quota
- dedicated to small output files (ATLAS, MONITORING)
- available through ■https://esgf.extra.cea.fr following work_thredds
- no backup

8.3. GENCMIP6_CCCSTOREDIR

- 2.5 PB in phase 1, 5 PB in phase 2 and 14 PB on tape in phase 3
- gencmip6 group quota
- dedicated to large (more than 1GB) output files (Output, Analyse)
- available through ■https://esgf.extra.cea.fr following store_thredds
- linked with HSM (tapes)

8.4. GENCMIP6_SCRATCHDIR

- same file system as GENCMIP6_CCCWORKDIR
- used during batch execution (RUN_DIR) and erased at the end of the execution
- regular cleaning after 40 days

9. End-of-job messages

To receive the end-of-job messages sent by the job itself: end of simulation, error,... you must specify your address in the $HOME/.forward file.

News in June 2018 : On Irene you have to duplicate a .forward for each project HOME.

10. About password

ccc_password_expiration helps you to know expiration date of your password. Currently password have to be changed one time per year.

```
 > ccc_password_expiration
Password for xxxxx@USERS-CCRT.CCC.CEA.FR: PPPPPPPPPP
Your password will expire in 70 days on Fri Nov 22 08:42:59 2013
> ccc_password_expiration -h
Usage: ccc_password_expiration [username[@realm]]
```

11. Installing a missing Python package

To install a missing python package you may need, you must install it from its sources. In this example we try to install a package we call 'super_package' As IRENE has no http connexion to the internet, you must download it on you mesocentre account : On ciclad :

```
> wget <address of the source>.tar.gz
```

Then you must scp it to IRENE, on your WORKDIR. For this, log in to irene and :

```
> scp login@ciclad.address.fr:/path/to/your/archive $CCCWORKDIR/dossier_de_sources/super_package.tar.gz
```

From now on everything will be done on IRENE. Uncompress the archive :

```
> tar -xvzf  $CCCWORKDIR/dossier_de_sources/super_package.tar.gz
```

We want to install the package for a usage with python3.7, thus, we load the module and add our source folder to the PYTHONPATH :

```
> module load python3/3.7.5
> export PYTHONPATH="${PYTHONPATH}:$CCCWORKDIR/dossier_de_sources/lib/python3.7/site-packages"
# You may need to do some mkdir to create lib/python3.7/site-packages
```

Now we install the package :

```
> cd super_package
> python3 setup.py install --prefix=$CCCWORKDIR/dossier_de_sources
```

The package is now installed.

To use it on a next session, we will have to load the module and update the PYTHONPATH :

```
> module load python3/3.7.5
> export PYTHONPATH="${PYTHONPATH}:$CCCWORKDIR/dossier_de_sources/lib/python3.7/site-packages"
```

You can now import this package on your python scripts.

12. The TGCC's machines

12.1. **Irene**

See the documentation for Irene.

12.2. **Irene-amd**

See the documentation for Irene-amd.

12.3. **Porting On Redhat8**

See the documentation for Porting your models configurations On Redhat8