

Introduction to file systems

Computer User Training Course 2016

Carsten Maass

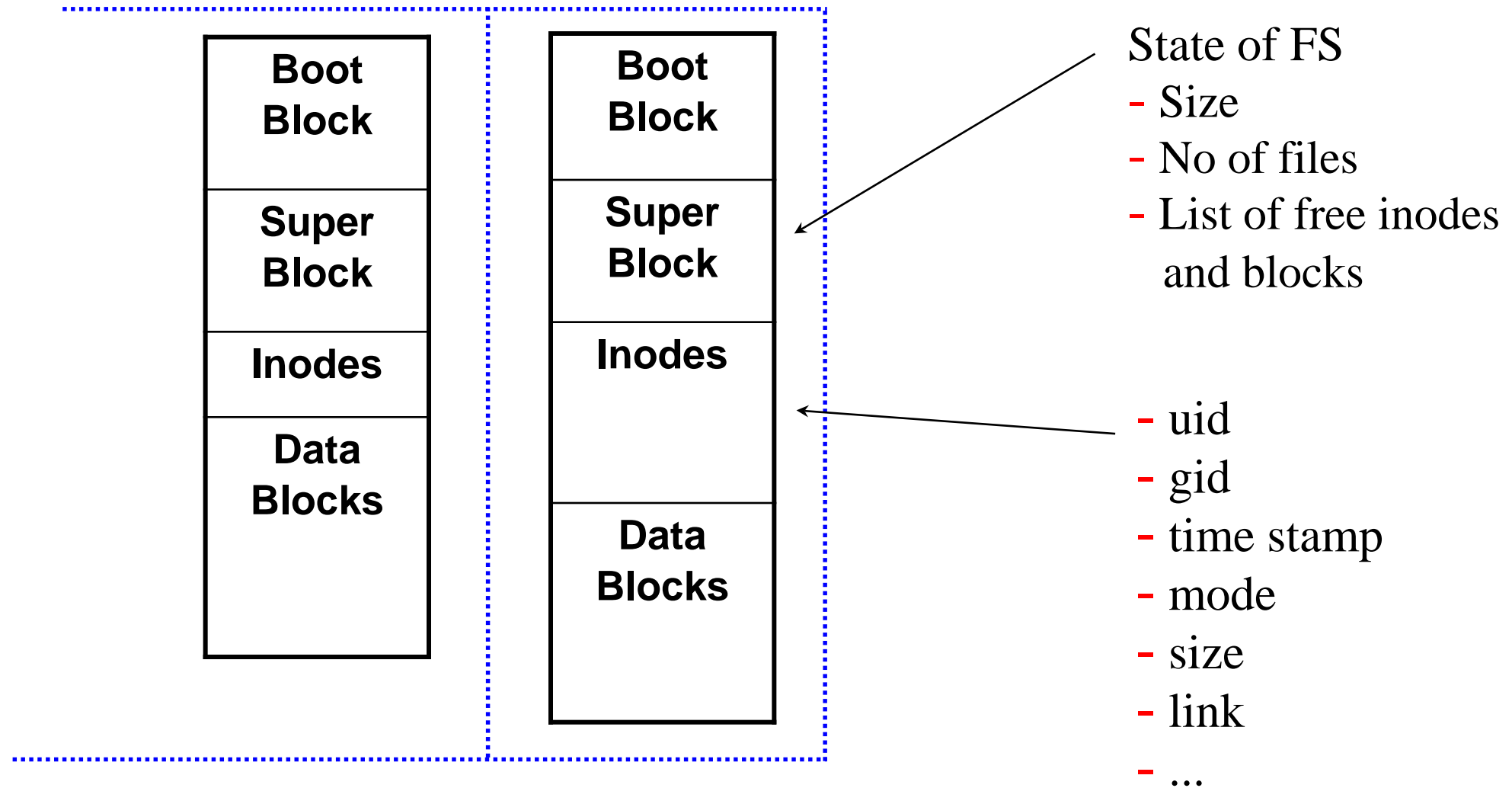
User Support

advisory@ecmwf.int

Overview

- Workstation server ecgate & HPCF
 - HOME
 - SCRATCH
 - SCRATCHDIR
 - PERM
- File systems cross-mounted
- Summary / remarks
- Practical

File system



File system types

- HOME

- \$HOME is a relatively small permanent file system
- snapshots / backups

- Permanent

- \$PERM relatively large permanent file system
- no backups

- Temporary

- files are kept as long as possible / deleted on a regular basis
- no backups


- Automatically deleted


- files are deleted at the end of your job or interactive session
- no backups


ECFS – User archive


- Tape archive – *Not* a file system!
 - long term archive
 - for excess data/files


Overview of ECMWF file systems

 **File systems** – suitable for permanent files ('large' quota, no backup)

 **File systems** – suitable for temporary data

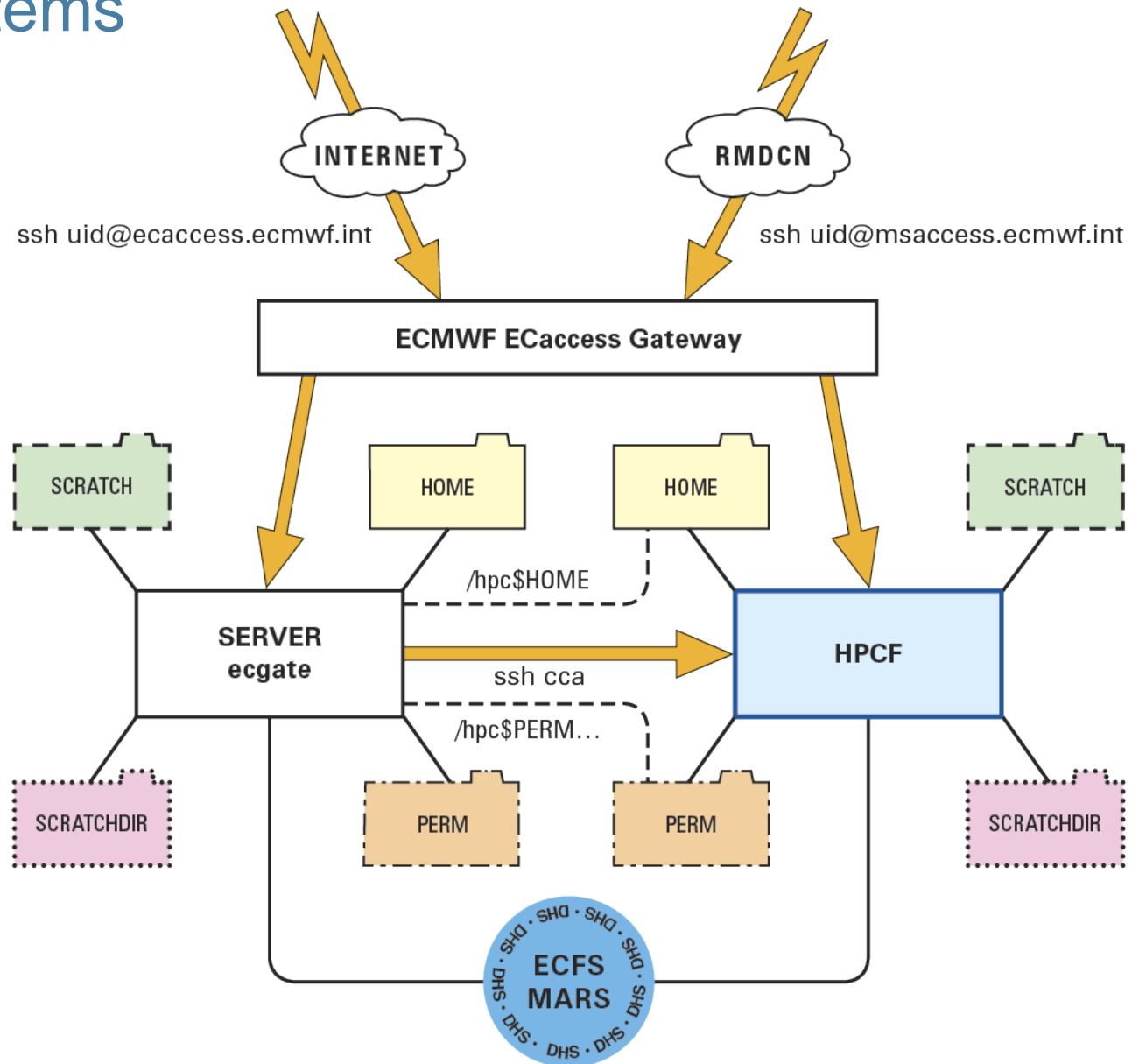
 **File systems** – directories automatically deleted at end of job

 **File systems** – suitable for permanent files ('small' quota, backup)

 **Machines** – accessible to users

 **Local file systems**

 **NFS mounted file systems**



ecgate: \$HOME

Filer appliance (NetApp), mirrored and redundant

- permanent files: profile files + e.g. utilities, source, jobs
- **quota** of 3.0 GB (soft limit 2.9 GB)
- check disk usage with command: **ecquota**
- snapshots
- backups

\$HOME examples

```
/home/$GROUP/$USER
```

```
/home/ms/$GROUP/$USER
```

```
/home/ectrain/tra
```

```
/home/ms/it/cnv
```

```
/home/ms/spde/de01
```

ecgate: \$HOME snapshots

- Most recent snapshots are in `.snapshot` in any sub-directory of the user's `$HOME`

```
cd .snapshot
```

- Additional snapshots can be found in e.g.

```
/vol/nasa_snapshot/.nasa-YYYY-MM-DD/vol_home_ms/...
```

```
ecgb11{/vol/nasa_snapshot/.nasa-2016-01-01/vol_home_ms}: --> ls -la
total 580
drwxr-xr-x. 145 root sys      4096 Dec  2 09:45 .
drwxr-xr-x.  22 root root     4096 Jan  2 04:00 ..
drwxr-xr-x.  81 root bin      4096 Apr 17  2015 at
drwxr-xr-x.  94 root bin      4096 Dec 31 10:49 be
drwxr-xr-x.  11 root root     4096 Sep 24 09:21 bg
drwxr-xr-x.  91 root bin      4096 Oct 27 13:25 ch
...
```


ecgate: \$SCRATCH

- General Parallel File System (GPFS)
- (locally) mounted, currently 1 file system, \approx 70 TB
- to be used for temporary data
- **quota**: 300 GB (soft limit 225 GB)
- select/delete is running:
 - On the 1. of every month files older than 1 year will be removed
 - Additional runs might be necessary at any time
 - Files newer than 32 days will (normally) not be removed
- **Please actively remove all files no longer needed!**

\$SCRATCH examples

```
/scratch/$GROUP/$USER  
/scratch/ms/$GROUP/$USER  
/scratch/ectrain/tra  
/scratch/ms/it/cn0  
/scratch/ms/spde/de01
```

ecgate: \$SCRATCHDIR

Part of \$SCRATCH (and its quota)

- automatically deleted at the end of job
- suitable for temporary data

\$SCRATCHDIR examples:

interactive:

```
/scratch/ms/$GROUP/$USER/scratchdir/$HOST.PID
```

```
/scratch/ms/uk/uk1/scratchdir/ecgb11.144600
```

batch:

```
/scratch/ms/$GROUP/$USER/scratchdir/$HOST.JobID.StepID
```

```
/scratch/ms/uk/uk1/scratchdir/ecgb03.ecmwf.int.31487.0
```

ecgate: \$PERM

- General Parallel File System (GPFS)
- Permanent but without backups
- **quota** of 30 GB (soft limit 27 GB)

\$PERM examples

```
/perm/$GROUP/$USER
```

```
/perm/ms/$GROUP/$USER
```

```
/perm/ectrain/tra
```

```
/perm/ms/it/cnv
```

```
/perm/ms/spde/de01
```

HPC: \$HOME

NAS (NFS), \approx 3 TB (shared between all users)

- permanent files: profile files + e.g. utilities, source, jobs
- mirrored
- snapshots (in `.snapshot` in any sub-directory of `$HOME`)
- block size 4 KB
- **quota** of 480 MB and 20000 files
- **quota** command will show disk usage

`$HOME` examples

```
/home/$GROUP/$USER  
/home/ms/$GROUP/$USER  
/home/ectrain/tra  
/home/ms/it/cn0  
/home/ms/spde/de01
```

HPC: \$SCRATCH

Lustre, ≈ 700 TB (shared between all users)

- to be used for temporary data
- default stripe size 1 MB / stripe count 1
- **quota** of 30 TB and 5000000 files
- select/delete is running
- **Please remove all files no longer needed!**

\$SCRATCH examples

```
/scratch/$GROUP/$USER  
/scratch/ms/$GROUP/$USER  
/scratch/ectrain/tra  
/scratch/ms/it/cn0  
/scratch/ms/spde/de01
```

HPC: Lustre striping

Lustre can stripe files over multiple Object Storage Targets (OSTs)

- stripe count: number of OSTs to use (default 1 MB)
- stripe size: size of stripes (default 1)
- Usage:

```
# lfs getstripe <PATH>
```

```
# lfs setstripe -c <COUNT> -s <SIZE> <PATH>
```

- Setting is inherited from parent directory
- For large files consider stripe count 2/4/8

HPC: \$SCRATCH – select/delete

- A select/delete runs regularly on \$SCRATCH
- Removes all files older than 30 days
- Retention period might change

HPC: \$SCRATCHDIR

Part of \$SCRATCH (and its quota)

- automatically deleted at the end of job or interactive login
- can be used for data

\$SCRATCHDIR examples:

interactive:

```
/lus/TMP/JTMP/#/$USER.PID.$HOST.DateTime
```

```
/lus/TMP/JTMP/94/uid.25695.cca-login2.20150227T095903
```

batch:

```
/lus/TMP/JTMP/#/$USER.JobID.$HOST.DateTime
```

```
/lus/TMP/JTMP/76/uid.7483867.ccapar.ccappn013.20150228T182811
```


HPC: \$TMPDIR

Used for e.g.

- Here documents
- Temporary files created by compiler or linker
- On login nodes cca/cca2
 - \$TMPDIR is backed by a node-local file system mounted under /tmpdir
 - Quota of 2GiB and 10000 files per user and login node
- On PPN batch nodes
 - **Currently:** \$TMPDIR=\$SCRATCHDIR
 - **Future plan:** A serial or fractional job can request a tmpfs of a suitable size be created on its execution PPN which the job's TMPDIR will then point to.

```
#PBS -l EC_job_tmpdir=<size>
```

where size is a number followed, optionally, by a unit specifier (no unit=bytes, K=kibibytes, M=mebibytes, G=gibibytes).

HPC: \$PERM

NAS (NFS), \approx 3 TB (shared between all users)

- permanent but without backups
- enforced user quotas of 26 GB and 200000 files (as usual, usage can be checked with the quota command)
- accessible via \$PERM environment variable
- mounted on ecgate as /hpc\$PERM
- 4 KB block size

\$PERM examples:

```
/perm/ms/$GROUP/$USER
```

```
/perm/ms/it/cn0
```

```
/perm/ectrain/tra
```

Cross mounted file systems

Purpose: facilitate commands like ls etc. on remote machines,

not suitable for data transfers and *not* to be used in batch

on cca

(interactive
node only)

ecgate	/ws\$HOME
	/ws\$SCRATCH

on ecgate

HPC	/hpc\$HOME
	/hpc\$PERM

Usage hints

- Transfer files between different platforms (e.g. ecgate – cca) with

```
scp / rsync
```

Much better performance if data is pulled from ecgate in cca batch jobs

- “Transfers” between \$SCRATCHDIR and \$SCRATCH on same platform

```
mv
```

- For important files on \$SCRATCH create a backup in ECFS and then use e.g.

```
#!/bin/ksh
cd $SCRATCH
if [[ -f large_file ]] then
    print “large_file exists already”
else
    ecp ec:large_file .
fi
```

Practical on Linux desktop or ecgate

1. Display the full pathnames for the following file systems for your training ID :
\$HOME, \$SCRATCH, \$SCRATCHDIR, \$PERM
2. Check your quotas!
3. Which is your largest directory/file (in kb)?
4. Find any file in your directories larger than 10 kb!
5. How much disk space (in kb) is available in the /home/ectrain file system?
6. List your cca HOME directory!
7. How many file names starting with a dot (“.”) which are symbolic links do you find in your \$HOME?
8. In your \$HOME create a public directory with permissions **rwxr-xr-x** and a private directory with permissions **rwx-----**
9. For your training ID find the available snapshots of your file \$HOME/DATES.txt and copy the latest and oldest version to your \$HOME!

Summary / Important remarks

Use *only* the following file systems

ecgate / cca	Suitable for
\$HOME	permanent files: sources, .profile, utilities, libs ...
\$SCRATCH	(large) temporary files
\$SCRATCHDIR	data to be automatically deleted at the end of a job
\$PERM	permanent, smaller input/output
cca:\$TMPDIR	compilation, here documents etc.

Summary / Important remarks

- **\$HOME, \$SCRATCH etc. on ecgate and HPCF are different**
- **Use the environment variables \$HOME, \$SCRATCH etc.**
- **Limit the number of files per directory**
- **Only \$HOME is backed up (snapshots are available)**
- **Different select/delete policies may apply on temporary file systems**
- **Do not rely on select/delete**
- **Clear your space as soon as possible!**
- **Check your quota with**
 - `ecquota` **# on ecgate**
 - `quota` **# on HPCF**
- **ECFS is an [archive](#) accessible from both HPCF and ecgate**