Introduction to file systems

Computer User Training Course 2014
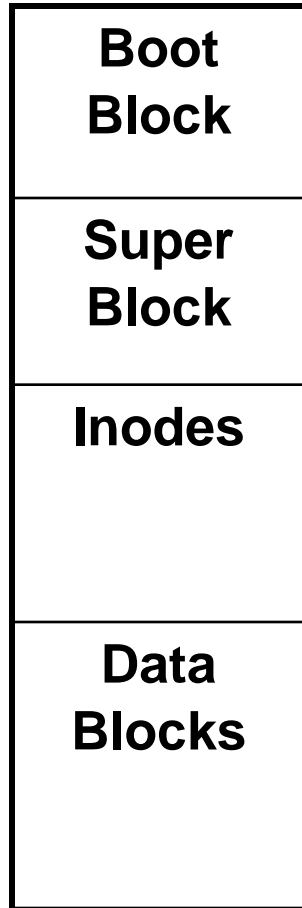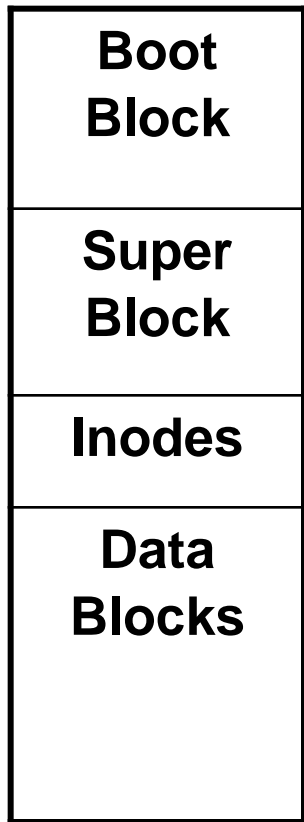
Carsten Maass

User Support

advisory@ecmwf.int

Overview

- **Workstation server ecgate & HPCF**

  - HOME

  - SCRATCH

  - SCRATCHDIR

  - PERM

- **File systems cross-mounted**

- **Summary / remarks**

- **Practical**

File system

| Boot Block |
| Super Block |
| Inodes |
| Data Blocks |

| Boot Block |
| Super Block |
| Inodes |
| Data Blocks |

State of FS
- Size
- No of files
- List of free inodes and blocks

- uid
- gid
- time stamp
- mode
- size
- link
- ...

ECMWF

File system types

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

- ## Permanent
    - $PERM relatively large permanent file system
    - no backups

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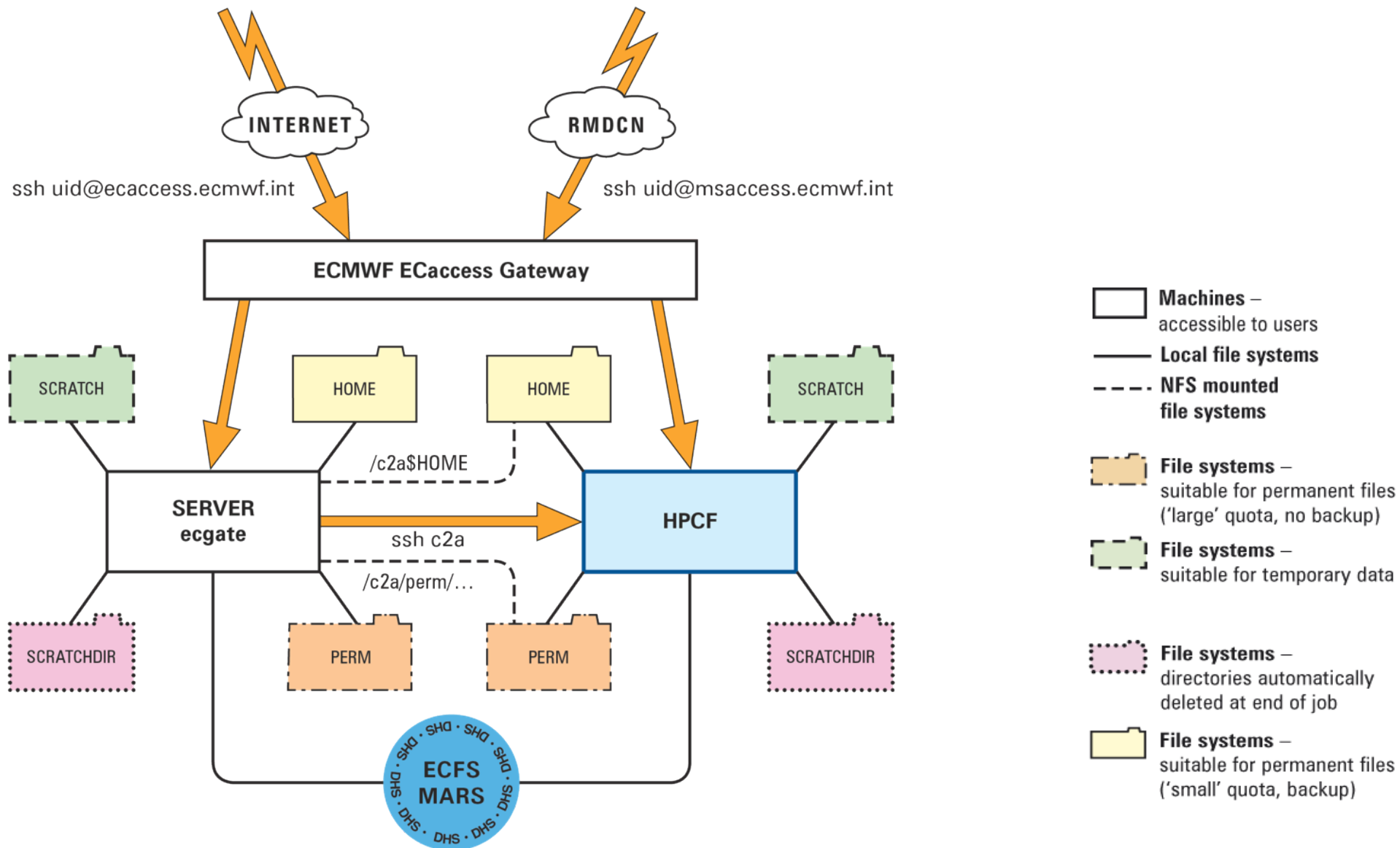   - no backups

- ## 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**

ECMWF

Overview of ECMWF file systems

INTERNET RMDCN

ssh uid@ecaccess.ecmwf.int
ssh uid@msaccess.ecmwf.int

ECMWF ECaccess Gateway

SCRATCH   HOME   HOME   SCRATCH

/c2a$HOME

SERVER
ecgate
ssh c2a
HPCF

/c2a/perm/…

SCRATCHDIR   PERM   PERM   SCRATCHDIR

ECFS
MARS

Machines – accessible to users

Local file systems

NFS mounted 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)

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`

- **backups / snapshots**

$HOME examples

```
/home/$GROUP/$USER

/home/ms/$GROUP/$USER

/home/ectrain/tra

/home/ms/it/cnv

/home/ms/spde/de01
```

ecgate: $HOME snapshots

- **snapshots taken during the day are available via**

  `cd .snapshot`

  in any sub-directory of the user's $HOME

- **daily snapshots can be found in** `/vol/nasa_snapshot/`

```
ecgb11{/vol/nasa_snapshot}:156 --> ls -la
total 180
drwxr-xr-x.  46 root root 4096 Feb 27 08:01 .
drwxr-xr-x.  53 root root    0 Feb 21 09:26 ..
drwxr-xr-x.  22 root root 4096 Sep  5 15:37 .nasa-2013-09-01
drwxr-xr-x.  22 root root 4096 Oct  4 21:27 .nasa-2013-10-01
drwxr-xr-x.  22 root root 4096 Nov  2 07:15 .nasa-2013-11-01
drwxr-xr-x.  22 root root 4096 Dec  2 05:06 .nasa-2013-12-01
drwxr-xr-x.  22 root root 4096 Jan  2 05:21 .nasa-2014-01-01
drwxr-xr-x.  22 root root 4096 Feb  2 11:33 .nasa-2014-02-01
…
```

ecgate: $SCRATCH

- **General Parallel File System (GPFS)**
- **(locally) mounted, currently 1 file system, ≈ 70 TB**
- **to be used for temporary data**
- **quota: 300 GB (soft limit 290 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 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
```

ECMWF

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
```

ECMWF

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
```

ECMWF

c2a: $HOME

GPFS, ≈ 4.3 TB (shared between all users)

- **permanent files: profile files + e.g. utilities, source, jobs**

- **daily backup**

- **block size 512 KB**

- **quota of 320MB and 21000 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
```

c2a: $SCRATCH

GPFS, ≈ 314 TB (shared between all users)

- **to be used for temporary data**
- **block size 4 MB**
- **quota of 26 TB and 1100000 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
```

ECMWF

c2a: $SCRATCH – select/delete

- **A select/delete runs on $SCRATCH every day**

- **Removes all files older than 21 days (expired)**

- **Removes additional files until usage < 83%**

  select/delete is currently under review

  $\longrightarrow$ **reduces retention period**

- **When usage is high, e.g. > 95%, a more aggressive cleaning is performed until usage < 70%**

- **Quota is shown as file system `s2ms_lb`**

ECMWF

c2a: $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:

```
/s2ms/TMP/JTMP/#/$HOST_$USER_PID

/s2ms/TMP/JTMP/79/c2a182_uid_p30081668
```

batch:

```
/s2ms/TMP/BATCH/#/$HOST.JobID.StepID_$USER
/s2ms/TMP/BATCH/22/c2a9u.802915.0_uid
```

ECMWF

c2a: $PERM

GPFS, ≈ 4.3 TB (shared between all users)

- **permanent but without backups**

- **enforced user quotas of 26 GB and 110000 files (as usual, usage can be checked with the quota command)**

- **accessible via $PERM environment variable**

- **mounted on ecgate as /c2a/perm/…**

- **512 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 c2a

(interactive node only)

ecgate	/ws$HOME
	/ws/$SCRATCH

on ecgate

c2a	/c2a$HOME
	/c2a/perm/ms/$GROUP/$USER

ECMWF

Usage hints

- **Transfer files between different platforms (e.g. ecgate – c2a) with**

```
scp / rsync
```

- **"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
```

ECMWF

Summary / Important remarks

Use _only_ the following file systems

ecgate / c2a	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

ECMWF

Summary / Important remarks

- **Only $HOME is backed up (snapshots are available)**

- **$HOME, $SCRATCH and $PERM on ecgate and HPCF are different**

- **Different select/delete policies may apply on temporary file systems**

- **Do not rely on select/delete**

- **Clear your space as soon as possible!**

- **Use the environment variables $HOME, $SCRATCH etc.**

- **Check your quota with**
    - `ecquota`   **# on ecgate**
    - `quota`   **# on HPCF**

- **ECFS archive (!) is accessible from both HPCF and ecgate**

Practical on Linux desktop or ecgate

1. **Display the full pathnames for the following file systems for your training ID : `$HOME, $TMPDIR, $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 c2a HOME directory!**

7. **How many file names starting with a dot (".") which are symbolic links do you find under your user-ID?**

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 .user_profile**