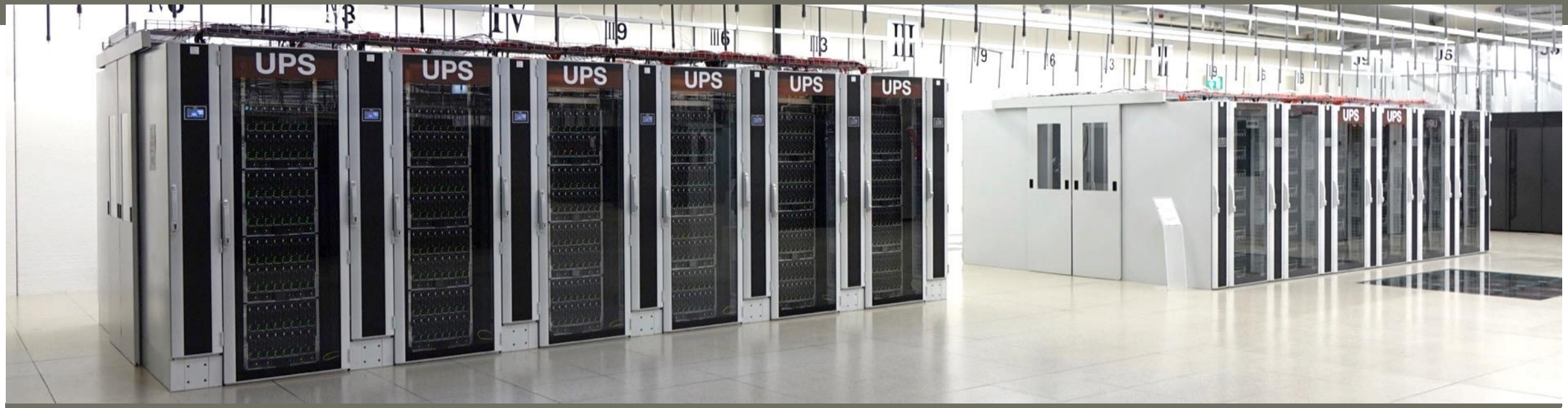

Getting started with Euler

Samuel Fux

High Performance Computing Group, Scientific IT Services, ETH Zurich

Outlook

- Introduction
- Accessing the cluster
- Data management
- Environment modules
- Applications
- Getting help

Outlook

- Introduction
- Accessing the cluster
- Data management
- Environment modules
- Applications
- Getting help

Intro > What is EULER?

- EULER stands for
  - Erweiterbarer, Umweltfreundlicher, Leistungsfähiger ETH Rechner
- It is the 5th central (shared) cluster of ETH
  - 1999–2007 Asgard ➔ *decommissioned*
  - 2004–2008 Hreidar ➔ *integrated into Brutus*
  - 2005–2008 Gonzales ➔ *integrated into Brutus*
  - 2007–2016 Brutus
  - 2014–2018+ Euler
- It benefits from the 15 years of experience gained with those previous large clusters

Intro > Shareholder model

- Like its predecessors, Euler has been financed (for the most part) by its users
  - In just 2 years, over 50 (!) research groups from almost all departments of ETH have invested in Euler
  - These so-called “shareholders” receive a share of the cluster’s resources (processors, memory, storage) proportional to their investment
- The small share of Euler financed by IT Services is open to all members of ETH
  - The only requirement is a valid NETHZ account
  - These “guest users” can use limited resources
  - If someone needs more computing power, he/she can invest in the cluster and become a shareholder at any time

Intro > Euler I (right) & II (left)

Outlook

- Introduction
- Accessing the cluster
- Data management
- Environment modules
- Applications
- Getting help

Access > Who can use Euler

- The only requirement to use Euler is a valid NETHZ account
- Members of ETH
  - Immediate access; no need to fill out an account request form
  - Use NETHZ credentials to login (via ssh) to *username*@euler.ethz.ch
  - New users must accept the cluster's usage rules upon first login
  - Euler uses NETHZ database to identify shareholders and guest users, and sets privileges and priorities automatically
- External collaborators
  - Members of other institutions who have a collaboration with a research group at ETH may use Euler for the purpose of said collaboration
  - Their counterpart ("sponsor") at ETH must create a NETHZ guest account for them, including e-mail address and VPN service
  - Once this is done, they can access Euler like members of ETH

Access > Legal compliance

- Euler is subject to ETH's acceptable use policy for IT resources (Benutzungs-ordnung für Telematik, BOT), in particular:
  - Euler accounts are strictly personal
  - Do not share your account (password, ssh keys) with anyone
  - Do not use someone else's account, even if they say it's OK
  - If you suspect that someone used your account, change your password and contact cluster-support@id.ethz.ch
- Consequences
  - In case of abuse, the offender's account may be blocked temporarily or closed
  - System administrators are obliged by law to investigate abusive or illegal activities and report them to the relevant authorities

Access > Security

- Firewall
  - Euler access is only possible via secure protocols (ssh, sftp, scp, rsync)
  - Euler is accessible only inside ETH ➔ use VPN to login from outside
  - Euler cannot access computers outside ETH ➔ use ETH's proxy service (http://proxy.ethz.ch:3128) to access external servers
  - Suspicious clients that repeatedly fail to login are automatically blacklisted
- Password
  - Your account will be temporarily blocked if you enter a wrong password too many times
  - You can change your NETHZ password at https://password.ethz.ch
  - Contact your local IT Support Group or the Service Desk if you need help with your NETHZ account (blocked account, forgotten password, etc.)

Access > SSH connection

- Unix, including Linux and macOS
  - Open a shell (Terminal in macOS) and use the standard `ssh` command:

    `ssh` *`username`*`@euler.ethz.ch`
- Windows
  - PuTTY (ssh client)
    - Download and install the free PuTTY program: http://www.chiark.greenend.org.uk/~sgtatham/putty/
    - Please refer to “Getting started” section of the manual: http://the.earth.li/~sgtatham/putty/0.60/htmldoc/Chapter2.html#gs
  - Cygwin (Linux environment under Windows)
    - Download: http://www.cygwin.com, http://cygwin.com/install.html
  - MobaXterm (ssh client + X-server)

Access > SSH connection

```
$ ssh leonhard@euler.ethz.ch
leonhard@euler.ethz.ch's password:
Last login: Mon Oct 16 12:48:55 2017 from somewhere.ethz.ch


    _____________________   ____
   /   _________   ___   /__/   /
  /   /_____/   /   /   /___    /
 /   _____/   /__/   /__/   /__/
/_______/   /__/   /__/   /__/
   Eidgenoessische Technische Hochschule Zuerich
   Swiss Federal Institute of Technology Zurich
   ------------------------------------------------------------------
                                E U L E R  C L U S T E R    CentOS 7


                https://scicomp.ethz.ch/wiki/Getting_started_with_clusters
                             NEW! -->  http://tinyurl.com/cluster-support
                                                cluster-support@id.ethz.ch


[leonhard@euler05 ~]$
```

Access > SSH keys

- SSH keys allow you to login without password
  - Especially useful for file transfers and automated tasks
  - When used properly, SSH keys are much safer than passwords
- SSH keys always come in pairs
  - A private key, stored on your local workstation (and nowhere else!)
  - A public key, stored on the computer(s) you want to connect to
- You can generate as many pairs as you like
  - For example, one for each computer you intend to connect to
- Keys can (should) be protected with a passphrase
  - You can use an SSH agent to unlock the key for the duration of a session, so that you do not need to enter the passphrase every time you want to use it

Access > SSH keys > Linux / macOS

- On your workstation, use `ssh-keygen` to generate a key pair
  - For extra security, enter a passphrase when prompted
  - By default the keys are stored in `$HOME/.ssh/id_rsa` and `$HOME/.ssh/id_rsa.pub`
  - Keep the private key secure; do NOT copy it or share it with anyone
- On Euler, create the directory `$HOME/.ssh`

```
mkdir -m 700 $HOME/.ssh
```

- On your workstation, use `ssh` to copy the public key to Euler

```
cat $HOME/.ssh/id_rsa.pub | \
ssh username@euler.ethz.ch "cat - >> .ssh/authorized_keys"
```

- With some Linux distributions, the two previous steps can be done at once using `ssh-copy-id`

```
ssh-copy-id username@euler.ethz.ch
```

Access > Graphical user interface

- Euler uses the X Window System (also called “X11” or simply “X”) to display a program’s graphical user interface (GUI) on your workstation
- You need to install an X11 server on your workstation to display X11 windows
- Linux
  - X11 (Xorg) is normally installed by default
- macOS
  - Since X11 is no longer included in macOS, you must install XQuartz
- Windows
  - X11 is not supported by Windows; you can install for example Cygwin/X, Xming or MobaXterm (all of these are freeware)

Access > SSH + X11

- The ports used by X11 are blocked by the cluster's firewall
  - To circumvent this, you must open an SSH "tunnel" and redirect all X11 communication through that tunnel, using the command:

    ```
    ssh -Y username@euler.ethz.ch
    ```

- Linux
  - Simply enter the command above in a shell
- macOS
  - Enter the command above in the Terminal; macOS should launch XQuartz automatically
  - If this does not work, launch XQuartz manually, select Terminal from the Applications menu, and enter the command above
- Windows
  - Open a terminal in Xming (Xming ➔ Xming) or Cygwin/x (Cygwin-X ➔ Xwin Server) or start and enter the command above

Outlook

- Introduction
- Accessing the cluster
- Data management
- Environment modules
- Applications
- Getting help

Data > Personal storage (every user)

- Home: `/cluster/`**`home`**`/`*`username`* (=`$HOME`)
  - Safe, long-term storage for critical data (program source, scripts, etc.)
  - Accessible only by the user (owner); other people cannot read its contents
  - Disk quota of 16 GB and a maximum of 100’000 files
  - Contents saved every hour (snapshot) and every night (tape backup)
- Scratch: `/cluster/`**`scratch`**`/`*`username`* (=`$SCRATCH`)
  - Fast, short-term storage for computations running on the cluster
  - Created automatically upon first access (`cd $SCRATCH`)
  - Visible (mounted) only when accessed
  - Strict usage rules; see `$SCRATCH/__USAGE_RULES__` for details
  - Disk quota of 2.5 TB and a maximum of 1’000’000 files
  - Absolutely NO backup and purged on a regular basis

Data > Group storage (shareholders only)

- Project: `/cluster/`**`project`**`/`*`groupname`*
  - Safe, long-term storage for critical data (like home)
  - Shareholders can buy as much space as they need
  - Access rights managed by the owner
  - Tape backup optional (for a fee)
- Work: `/cluster/`**`work`**`/`*`groupname`*
  - Fast, short- or medium-term storage for large computations (like scratch but without purge)
  - Shareholders can buy as much space as they need
  - Access rights managed by the owner
  - Visible (mounted) only when accessed
  - Absolutely NO backup

Data > Other storage options

- Local `/scratch` on each compute node (=`$TMPDIR`)
  - Intended for serial, I/O-intensive applications
  - Very short life span; data are deleted automatically when the job ends
  - Scratch space must be requested by the job (see “batch system” later on)
- Central NAS
  - Groups who have purchased storage on the central NAS of ETH can ask the Storage Group of IT Services to export it to Euler
  - Accessible from all nodes in the cluster
- Other NAS
  - Groups who are operating their own NAS can export a shared file system via NFS to Euler
  - Requirement: user and group IDs must be consistent with NETHZ
  - Mounted automatically on Euler under `/nfs/`*`servername`*`/`*`sharename`*

Data > File system comparison

File system	Life span	Backup	Max size	Small files	Large files
`/cluster/home`	Permanent	Yes	16 GB	+	o
`/cluster/scratch`	2 weeks	No	2.5 TB	o	++
`/cluster/project`	4 years	Optional	Flexible	+	+
`/cluster/work`	4 years	No	Flexible	o	++
local `/scratch`	Job	No	800 GB	++	o
central NAS	Flexible	Optional	Flexible	+	+

Data > Copying data from/to the cluster

- Secure copy (`scp`) is most commonly used to transfer files
- Syntax:

```
scp [options] source destination
```

- Example: copy a file from your workstation to Euler (home directory)

```
scp file username@euler.ethz.ch:
```

- Example: copy a file from Euler to your workstation (current directory)

```
scp username@euler.ethz.ch:file .
```

- Copy a whole directory

```
scp -r localdir username@euler.ethz.ch:remotedir
```

- Scp clients: WinSCP, PSCP, Filezilla
- Other commands like `sftp`, `rsync`, `svn`, `git`, `wget`

Outline

- Introduction
- Accessing the cluster
- Data management
- Environment modules
- Applications
- Getting help

Environment > Modules

- Euler provides environment modules to configure your environment for specific tools, e.g.
  - Development tools
  - Scientific libraries
  - Communication libraries
  - Third-party applications
- Advantages
  - Configuration is fully automatic and – hopefully – fool-proof
  - Different versions of the same software can co-exist and can be selected explicitly
  - You can easily try out different tools, switch between versions, to find out which one works best for you

Environment > Modules > Commands

`module`	get info about module sub-commands
`module avail`	list all modules available on Euler
`module avail` ***`name`***	list all modules that match *name*
`module key` *`keyword`*	list all modules whose description contains *keyword*
`module help` *`name`*	get information about module *name*
`module show` *`name`*	show what module *name* does (without loading it)
`module load` ***`name`***	load module *name*
`module list`	list all currently loaded modules
`module unload` *`name`*	unload module *name*
`module purge`	unload all modules at once

Environment > Modules > Life cycle

- All supported applications are readily accessible via the `module` command
- New, untested applications, and applications that are not fully supported yet, are placed in the “new” module category
  - This gives users a chance to test bleeding-edge software
- Applications that are no longer supported are moved into the “legacy” module category
  - Old modules are never deleted
- Table of all available applications and versions:
  - https://scicomp.ethz.ch/wiki/index.php/Euler_applications

Environment > Modules > Categories

Category	Support	Changes	How to enable	Purpose
New	Partial	Any time	`module load new`	Development & testing
Supported	Full	Quarterly	Enabled by default	Production
Legacy	Minimal	Never	`module load legacy`	Compatibility & continuity

Environment > Modules > Example

```
[leonhard@euler04 ~]$ module load new legacy
[leonhard@euler04 ~]$ module avail gcc
----------- /cluster/apps/modules/modulefiles -----------
gcc/4.4.7(4.4)     gcc/4.8.2(default) gcc/4.9.2
----------- /cluster/apps/modules/new -----------
gcc/4.8.4
----------- /cluster/apps/modules/legacy -----------
gcc/4.7.4
[leonhard@euler04 ~]$ module help gcc
----------- Module Specific Help for 'gcc/4.8.2' -----------

        GNU C/C++ and Fortran compilers version 4.8.2

[leonhard@euler04 ~]$ module load gcc/4.7.4
[leonhard@euler04 ~]$ module list
Currently Loaded Modulefiles:
  1) new         2) legacy      3) gcc/4.7.4
```

Outlook

- Introduction
- Accessing the cluster
- Data management
- Environment modules
- Batch system
- Applications
- Getting help

Applications > Centrally installed

- Applications that are needed by many users should be installed centrally, like compilers and libraries
  - Visible and accessible to all users via modules
  - Installed and maintained by cluster support
  - Commercial licenses provided by IDES
- Central applications are installed in /cluster/apps
- List of centrally installed applications on Euler:
  - https://scicomp.ethz.ch/wiki/index.php/Euler_applications
- Users can install additional applications in their home directory

Applications > Commercial / Open source

- Bioinformatics and life sciences
  - Bioconductor, BLAST, Bowtie, CLC Genomics Server, FSL, RAxML, TopHat
- Finite element methods
  - Ansys, Abaqus, FEniCS, MSC Marc, MSC Nastran
- Multi-physics phenomena
  - Ansoft Maxwell, COMSOL Multiphysics, Trilinos
- Quantum chemistry and molecular dynamics
  - ADF, CP2K, Gaussian, NWChem, Orca, Quantum Espresso, Turbomole
- Symbolic, numerical and statistical mathematics
  - Gurobi, Maple, Mathematica, MATLAB, R, Stata
- Visualization
  - Ffmpeg, ParaView, VisIT, VTK

Applications > Development

- Compiler
  - GCC, Intel, LLVM, PGI
- Scientific libraries
  - ACML, Boost, deal.II, Eigen, FFTW, GMP, GSL, HDF5, MKL, NetCDF, NumPy, OpenBLAS, PETSc, SciPy
- MPI libraries
  - Open MPI, MVAPICH2
- Build systems
  - GNU Autotools, Cmake, qmake, make
- Version Control
  - SVN, Git, Mercurial, CVS

Outlook

- Introduction
- Accessing the cluster
- Data management
- Environment modules
- Batch system
- Applications
- Getting help

Getting help

- Wiki
  - https://scicomp.ethz.ch
- Ticket system
  - http://tinyurl.com/cluster-support (NETHZ authentication)
  - Please describe your problem as accurately as possible
- E-mail
  - cluster-support@id.ethz.ch
  - Please do not send questions to individual members of the team
- Person-to-person
  - Contact us to set up an appointment at your place
  - Visit us at Weinbergstrasse 11, WEC, D floor (please call first)

Dos and don’ts

- Dos
  - Optimize your workflow to make it as efficient as possible
  - Understand what you are doing
  - Ask for help if you don’t understand what you are doing
  - Keep in mind that Euler is shared by many users
  - Choose the file system you want to use carefully
- Don’ts
  - Don’t waste CPU time or disk space
  - Don’t run applications on the login nodes
  - Don't use login nodes to transfer large amounts of data (do that in a batch job)
  - Don’t write large amounts of data to standard output
  - Don’t run hundreds of small jobs if the same work can be done in a single job

Questions?