

METACENTRUM - COMPUTATION FOR THE ACADEMIC COMMUNITY

Roman Leontovč^{1,2}

¹ MetaCentrum, CESNET z.s.p.o.

² Faculty of Science, Charles University
leontovr@natur.cuni.cz

Outline

- **MetaCentrum - introduction**
- **MetaCentrum HW resources and software**
- **MetaCentrum infrastructure environment**
- **How to start**
- **How to compute**
- **Common issues and how to prevent them**
- **Other Cesnet services**
- **Other CESNET services**

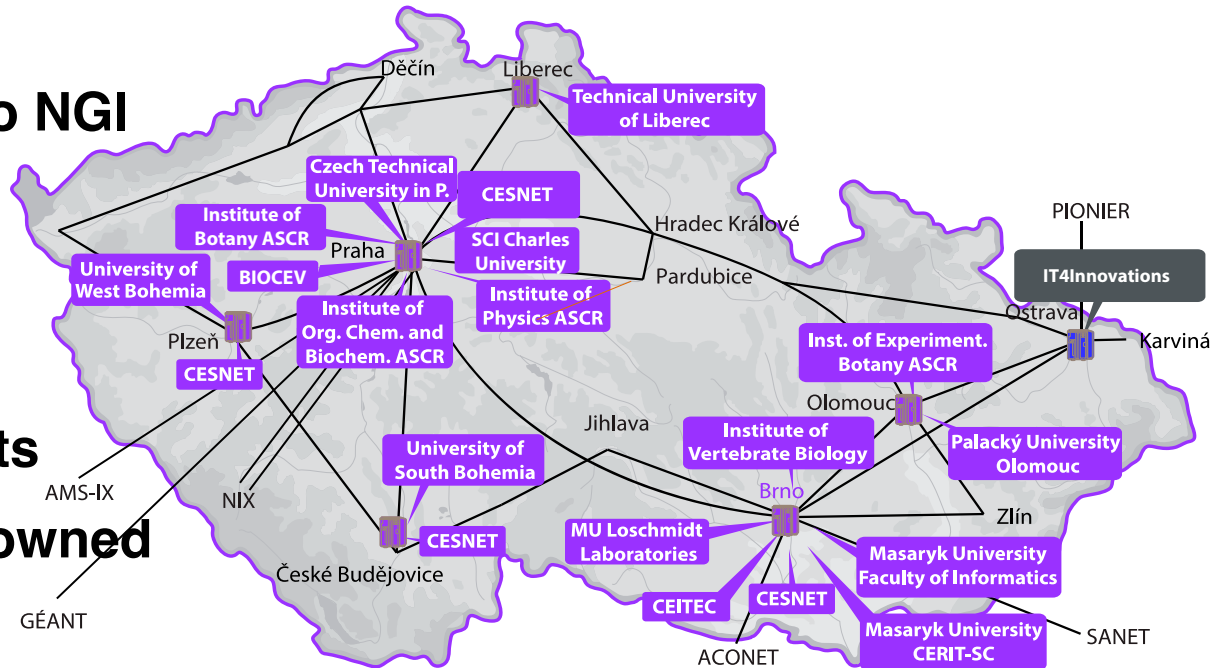
About MetaCentrum

■ MetaCentrum

- organization operating and managing distributed computing - National Grid Infrastructure (NGI)
- established 1996
- activity of the CESNET association (association of universities of the Czech Republic and the Czech Academy of Science)
- provides scientific computations, collaborative research and its support service
- integrated into European e-infrastructures EGI (grid), EOSC (European Open Scientific Cloud)
- completely free for students, academic staff of research institutions in Czech Republic (registration needed)
- 24x7 service

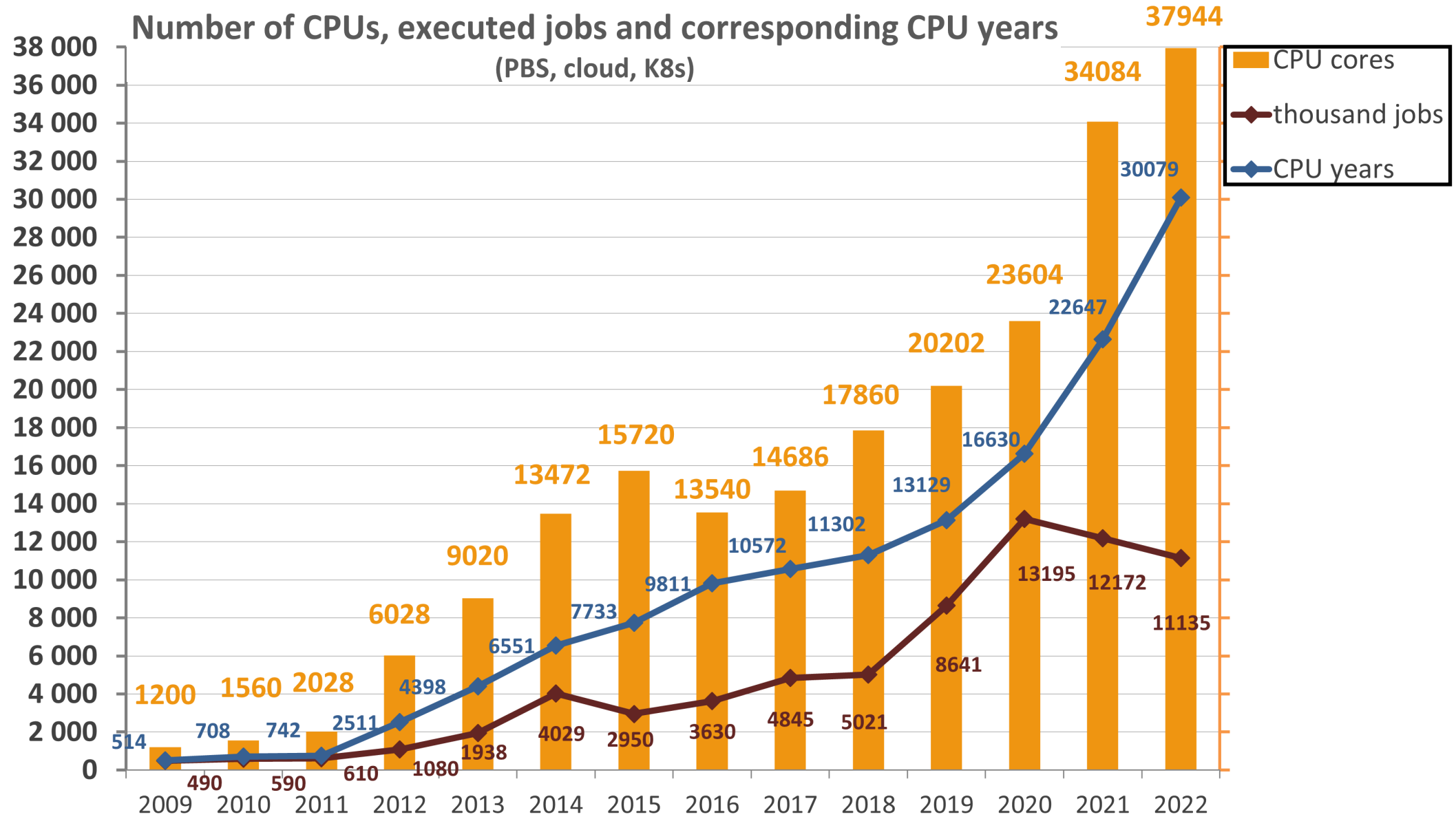
MetaCentrum NGI

- **Coordinator of national grid**
- **Assistance with:**
 - **purchase and integration computational resources into NGI**
 - **selection, installation and maintenance of the clusters**
 - **software maintenance**
 - **maintenance of user accounts**
 - **priority/exclusive access to owned clusters**



Number of CPUs, executed jobs and corresponding CPU years

(PBS, cloud, K8s)



MetaCentrum statistics of 2022

- 30,079 CPU years (upper paleolit)
- 2,710 users
- 11 135 000 executed jobs



Metacentrum services

- **Metacentrum services**
 - **individuals - computing, data processing**
 - **projects - sharing data in a group**
 - **institutions - integrating resources to NGI (maintenance of HW, SW)**
 - **research, development, education**
 - **immediately accessible resources (registration needed)**
 - **free of charge (acknowledgements in publications)**

Cluster computing

Old school

- **Computing cluster**
 - group of interconnected “classical” computers



Cluster computing

Now



MetaCentrum Resources

■ CPU

- ~ 45,000 CPU (x86_64)
- Intel, AMD
- Linux - Debian 11, CentOS

heterogenous environment

- nodes with:
 - low number of cores (2x4-32 cores)
 - average number of cores - SMP machines (32-128 cores)
 - high number of cores
 - 384 cores 6 TB RAM - SGI UV 2000
 - 504 cores 10 TB RAM - HPE Superdome Flex

<https://metavo.metacentrum.cz/pbsmon2/hardware>

■ GPU

- 17 clusters, over 400 GPU cards (NVIDIA A10, A40, A100, RTX A4000, Tesla, GeForce)
- 8.5 PB disk space (working data), ca 3 TB/user
- 21 PB archive storage, “unlimited”/user (Cesnet Storage Department)

https://wiki.metacentrum.cz/wiki/GPU_clusters



How to start

- **Employees and students of research or academic organizations in the Czech Republic**
- **Registration** (<https://metavo.metacentrum.cz/en/application/index.html>)
- **Rules of use** (<https://www.metacentrum.cz/en/about/rules/index.html>)
 - **science, research, education, development, (commercial)**
 - **free, acknowledgement in publications**
- **Read our documentation, FAQ and tutorial for beginners**
https://wiki.metacentrum.cz/wiki/Main_Page https://wiki.metacentrum.cz/wiki/Beginners_guide
https://wiki.metacentrum.cz/wiki/FAQ/Grid_computing <https://wiki.metacentrum.cz/wiki/Troubleshooting>

Acknowledgement for 2023

MetaCentrum official acknowledgement formula

Computational resources were provided by the e-INFRA CZ project (ID:90140), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

Acknowledgement formula for ELIXIR CZ resources

Computational resources were provided by the ELIXIR-CZ project (ID:902553), part of the international ELIXIR infrastructure

MetaCentrum environment

Grid computing https://wiki.metacentrum.cz/wiki/Beginners_guide

- Batch jobs, long-time running (days, weeks, months)
- Interactive jobs (direct control)
- Primarily for command line control
- Graphical environment (X-windows, Remote desktop, Open OnDemand)
 - Open OnDemand - web base GUI, no scripting skills needed)
<https://wiki.metacentrum.cz/wiki/Singularity>
- Containers (Singularity)

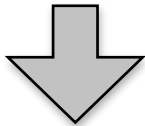
Cloud computing (MetaCentrum cloud)

- Virtual machines
- Personalization of environment (OS, SW versions etc.)

Metacentrum environment

Frontends

- mainly virtual machines
- direct log in without reservation (ssh, kerberos, putty)
- for file/directory manipulation, basic unix operation, preparation of the batch jobs, submitting the jobs
- not for demanding activities (large data manipulation, computing, software compilation etc.)



submit the batch or interactive job

Computing nodes

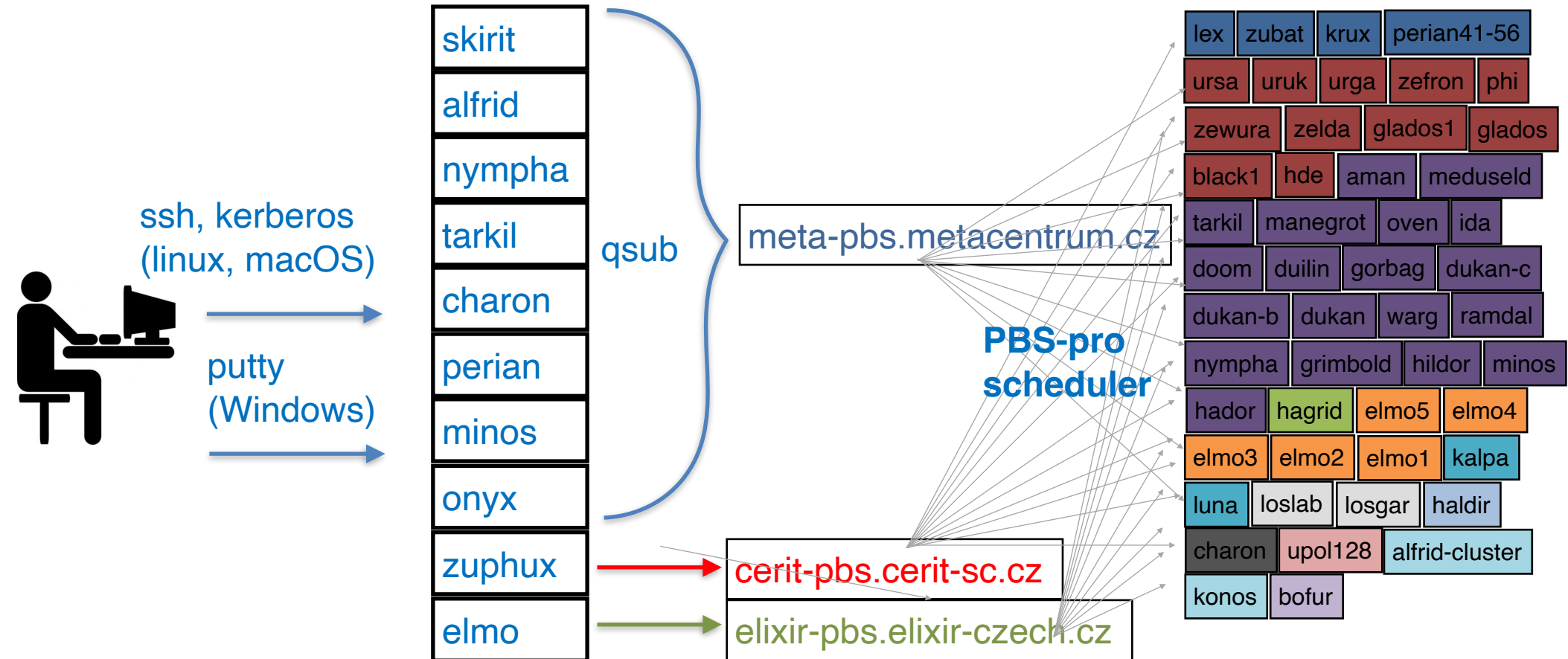
- physically process the submitted jobs
- mainly for non-interactive work, jobs are assigned by PBS (scheduling system)
- exceptional direct login (manual deleting/transfer data after job failure)
- heterogenous environment (different number of cores per node, GPU cards...)

MetaCentrum environment

Frontends

PBS servers

Computing nodes



Metacentrum Software

- **modular subsystem** (https://wiki.metacentrum.cz/wiki/Application_modules#Usage)
 - frontends/working nodes are “application-free environment”
 - applications/software are provided as modules, user specify which applications and versions will be used
- **~3000 modules** (computational chemistry, molecular modeling, bioinformatics, technical and material simulations, mathematical and statistical modeling, image video and sound processing, development tools and environments etc...)

Metacentrum Software

■ Commercial software

- paid and maintained by MetaCentrum (regular updates), for all users, license agreement may be required
- paid by users, maintained by MetaCentrum, limited to users/groups

■ Open source software

- installation/updates by MetaCentrum on request, for all users
- installation/updates by users in their environment.

Metacentrum Software

- **New version of modular system will be released soon**

- **Activation of new modular system via command:**

```
source /cvmfs/software.metacentrum.cz/modulefiles/5.1.0/loadmodules
```

- **Basic module commands**

- **module ava - all available modules**
- **module ava blast/ - all available versions of blast**
- **module add - add module to the environment (module add blast/2.10.0)**
- **module list - list of loaded modules**
- **module rm - unload module (module rm blast/2.10.0)**
- **module purge - unload all modules**

Conda modules

- **Conda**
 - **open-source package management system**
 - **quickly installs, runs and updates packages and their dependencies**
 - `module add conda/modules` (load conda module)
 - `conda env list` (list of environments/software available)
 - `conda activate busco_v5.4.3_py3.8` (activate the environment)
 - `conda deactivate` (deactivate the environment)

Python modules

- **Python packages installed and managed by pip package installer**
- **Dependent on the python version and compiler**
 - module add python36-modules (load modules for python 3.6)
 - pip3.6 list (list available packages, which can be used)

Perl modules

- **Packages dependent on perl**
 - `module avail bioperl/` (available bioperl modules)
 - `module add bioperl/1.7.8-gcc` (load specific bioperl module)
 - `perl_installed_modules.pl` (list of available perl packages)

Data storages in MetaCentrum

https://wiki.metacentrum.cz/wiki/Types_of_data_storages

■ Scratch storages

- fast storages with minimum data capacity
- working with data during computations
 - **scratch_local**
 - on every node, HDD, default
 - **scratch_ssd**
 - fast SSD, typically smaller in volume, not everywhere
 - **scratch_shared**
 - network volume, which is shared between all clusters in a given location, not everywhere
 - **scratch_shm**
 - scratch held in RAM, very fast, on every node
- automatically cleaned

■ Disk arrays

- data storing between computation
- several storages geographically distributed (Praha, Brno, Plzeň, Jihlava etc)

■ Hierarchical storages

- data archiving

Batch jobs

- **high demands on time and computing power**
- **on the frontend**
 - prepare the input data
 - prepare the batch job script in text editor (vim, pico, nano...) or graphical qsub assembler (<https://metavo.metacentrum.cz/pbsmon2/person>)
 - specifications of requested resources (number of nodes, CPUs, walltime, memory, scratch space etc.)
 - copy input data to working node scratch directory
 - load software modules,
 - execute the computation,
 - copy the results
 - clean the scratch directory
 - **submit the batch job (qsub)**
 - https://wiki.metacentrum.cz/wiki/Beginners_guide#Run_batch_jobs

```
#!/bin/bash
#PBS -l walltime=00:10:00
#PBS -l select=1:ncpus=2:mem=4gb:scratch_local=10gb
#PBS -N my_job
#PBS -m abe
#PBS -M leontovyc.roman@seznam.cz

# set a DATADIR variable
DATADIR=/storage/plzen1/home/leontovyc_roman/quality_control_batch

# copy input file "ERR10177551.fastq" to the scratch directory
cp /storage/plzen1/home/leontovyc_roman/quality_control_data/ERR10177551.fastq $SCRATCHDIR

# move into the scratch directory
cd $SCRATCHDIR

# load a module for your application
module add fastQC-0.11.5

# make the output directory
mkdir fastqc_result

# run the calculation
# do not forgot to use reserved CPUs by '-num_threads' flag
fastqc -t 2 -o fastqc_result ERR10177551.fastq

#copy results
cp -r fastqc_result $DATADIR

# clean the scratch directory
clean_scratch
```

Batch jobs

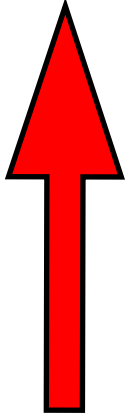
■ Monitoring

- in terminal - qstat
- graphical view on <https://metavo.metacentrum.cz/pbsmon2/jobs/detail>
- complete info about the job (walltime, memory used, effectivity of CPU usage, working node etc.)
- Walltime can be increased by user https://metavo.metacentrum.cz/en/news/novinka_2021_0013.html or on request
- submitted/running job can be terminated by user anytime - qdel

Interactive jobs

- For interactive work
- Testing calculations
- requesting resources same way as for batch job with -l (upper case “i”) option
 - https://wiki.metacentrum.cz/wiki/Beginners_guide#Run_interactive_job

```
qsub -I -l walltime=1:0:0 -l select=1:ncpus=2:mem=4gb:scratch_local=10gb
```



Fairshare

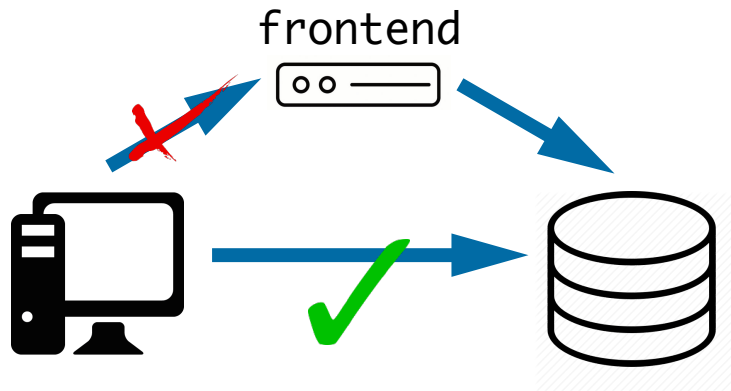
- **mechanism which allows historical resource utilization information to be incorporated into job feasibility and priority decisions**
- **more you compute longer you wait**
- **continuously increasing in time to initial value**
- **separated on different schedulers (meta-pbs.metacentrum.cz, cerit-pbs.cerit-sc.cz , elixir-pbs.elixir-czech.cz)**
- **significantly increased by acknowledgement of MetaCentrum in publications**

cesnet
metacentrum
.....

e-INFRA
CZ

Common issues and how to prevent them

- Do not use frontends, copy data directly on storage, use compressed files (.tar, .zip, .gz, etc.)
- SFTP client for Windows users (WinSCP, FileZilla, CyberDuck)



```
scp my_data.gz  
leontovyc_roman@skirit.metacentrum.cz:\  
/storage/brno2/home/leontovyc_roman
```

```
scp my_data.gz leontovyc_roman@storage-  
brno2.metacentrum.cz:
```


- Optimise your calculations (hardware usage)
- Reservation of too many resources decreases your fairshare score and reduces the priority for your future jobs

<https://wiki.metacentrum.cz/wiki/Fairshare>


- You can increase your fairshare score by acknowledgement to MetaCentrum in your publications

https://wiki.metacentrum.cz/wiki/Usage_rules/Acknowledgement

- Effectivity can be checked on the computation node by standard Linux tools (top, htop) or on metavo.metacentrum.cz web portal

- Is not appropriate to run long and demanding calculations directly on frontends and/or on clusters outside of PBS
- Ask for an **Interactive job...**

```
qsub -I -l select=1:ncpus=2:mem=4gb:scratch_local=10gb -l walltime=1:00:00 \  
-m abe
```



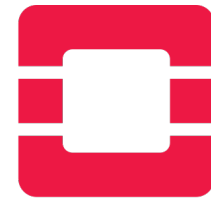
cesnet
metacentrum
.....

 **e-INFRA**
CZ

Other Cesnet services

Cloud computing

- **Run in Openstack** (<https://cloud2.metacentrum.cz>)
- **virtual machines - fully controlled by user, different OS (Linux, Windows etc.)**
 - predefined images
 - user defined
- **primarily for testing/computing not for web hosting**



openstack.®

Data storages

cesnet

datacare
■■■■■■

- <https://du.cesnet.cz/>
- data storages for scientific research
- more than 21 PB storage
- data backups, data archiving, data sharing
- fully integrated with MetaCentrum
- hierarchical - different speed and capacity, MAID and tape storages

FileSender



FILESENDER



- <https://www.cesnet.cz/sluzby/filesender/>
- web “depository-like” service for file share/transfer
- files up to 500 GB
- at least one participant with eduID identity
- free of charge

OwnCloud

- <https://www.cesnet.cz/sluzby/owncloud/https://www.cesnet.cz/services/owncloud/?lang=en>
- **synchronisation between various computers/mobile devices**
- **100 GB/user (may be increased)**
- **for Windows, Linux, Os X, iOS, Android**
- **access via web interface**
- **data sharing with other users**
- **free of charge**



Advantages

- **100% free of charge**
- **strong computational power**
- **large data storage capacity (backups, archives)**
- **heterogenous environment**
- **full service administration (hardware, software, user accounts etc.)**
- **user support**

Disadvantages

- **sharing resources**
- **interactive work might be complicated, more suitable for batch jobs**
- **waiting for jobs to run**
- **might be difficult to understand the structure/environment**

THANK YOU FOR ATTENTION

