cesnet

metacentrum

METACENTRUM - COMPUTATION FOR THE ACADEMIC COMMUNITY

Roman Leontovyč[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
- 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 sevice

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**

Metacentrum NGI

- Metacentrum services
  - individuals - computing, data processing
  - 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

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

Old school

Cluster computing

Now

MetaCentrum Resources

- 53 machines, 30,734 cores
- Linux - Debian, 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
    - GPU cards (NVIDIA T4, 1080 Ti, 2080 Ti, A100)
    - Xeon Phi etc.
- 8.5 PB disk space (working data), ca 3 TB/user
- 21 PB archive storage, “unlimited”/user (Cesnet Storage Department)

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

MetaCentrum statistics 2021

- 2,606 users
- 13 100 000 executed jobs
- 22,647 CPU years (upper paleolit)

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**

Acknowledgement:

Computational resources were supplied by the Ministry of Education, Youth and Sports of the Czech Republic under the Projects CESNET (Project No. LM2015042) and CERIT-Scientific Cloud (Project No. LM2015085) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

MetaCentrum environment

Grid computing, HTC, HPC clusters

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

- **Batch jobs, long-time running (months)**
- **Interactive jobs**
- **Containers (Singularity, Docker)**
- **Graphical environment (X-windows, Remote desktop, Open OnDemand)**
  - **Open OnDemand - web base GUI, no scripting skills needed)**

Cloud computing (MetaCentrum cloud)

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

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
  - automatically cleaned
- **Disk arrays**
  - data storing between computation
  - several storages geographically distributed (Praha, Brno, Plzeň, Jihlava etc)
- **Hierarchical storages**
  - data archiving

Metacentrum Software

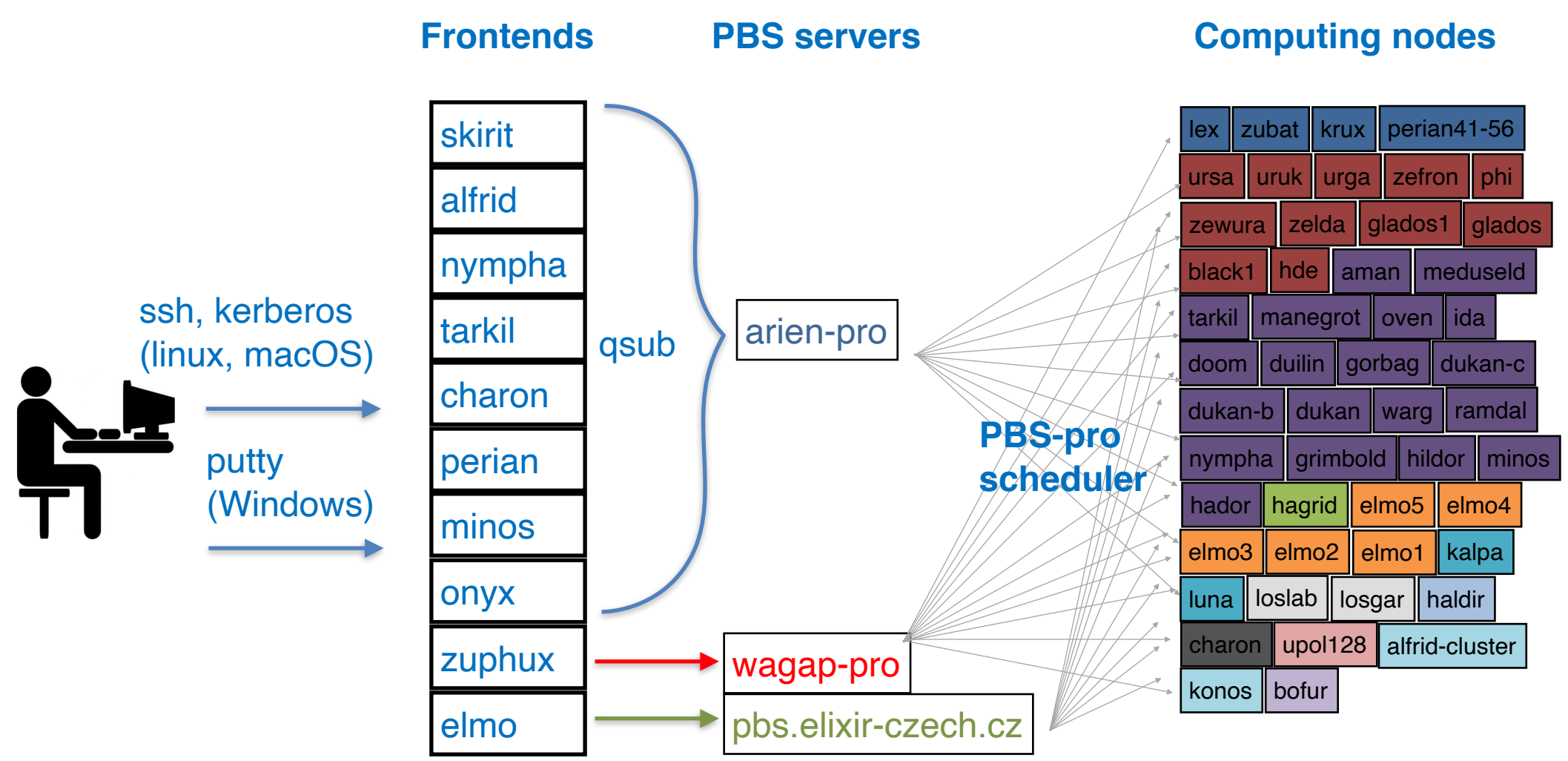
- **modular subsystem** (https://wiki.metacentrum.cz/wiki/Application_modules#Usage)
  - frontends/working nodes are “application-free environment”
  - applications/software are provided as modules, user specify which applications and versions will be used
- **hundreds of software modules** (computational 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
  - 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.

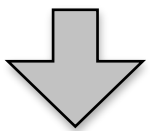
MetaCentrum environment



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, Xeon Phi)

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/sh
#PBS -N BWA_mapping_Echo2_NEW
#PBS -l walltime=24:00:00
#PBS -l select=1:ncpus=20:mem=50gb:scratch_local=50gb
#PBS -m abe
#PBS -M leontovyc.roman@seznam.cz

# set a handler to clean the SCRATCHDIR once finished
trap "clean_scratch"TERM EXIT
# set the location of input/output data
DATADIR="/storage/brno7-cerit/home/vorel/Echo/speedtest"
# prepare the input data
cp $DATADIR/speedtest-output-54-cores.fastq $SCRATCHDIR || exit 1
cp /storage/brno7-cerit/home/vorel/Echo/Drosophila/Drosophila_artificial_100/
reference_genome.fa $SCRATCHDIR || exit 1
# go to the working directory and perform the computation
cd $SCRATCHDIR
# load the appropriate module
module add bwa-0.7.5a
# run the computation
bwa index -a is reference_genome.fa
bwa mem  reference_genome.fa  speedtest-output-54-cores.fastq   > speedtest-output-54-
cores_NEW.sam
# copy out the output data (if it fails, let the data in SCRATCHDIR and inform the
user)
cp speedtest-output-54-cores_NEW.sam $DATADIR || export CLEAN_SCRATCH=false
```

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 on request
  - submitted/running job can be terminated by user anytime - qdel

Interactive jobs

- **for interactive work**
- **requesting resources same way as for batch job**
  - https://wiki.metacentrum.cz/wiki/Beginners_guide#Run_interactive_job
  - in terminal
  - graphical user interface (GUI) via VNC client e.g. TigerVNC

MetaCentrum environment

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 (arien-pro, wagap-pro, elixir-czech)
- significantly increased by acknowledgement of MetaCentrum in publications

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

- 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

FileSander

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/
- 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

cesnet
metacentrum

THANK YOU FOR ATTENTION