

Computing cluster of the Institute of Botany, Czech Academy of Sciences

Overview, features, usage, discussion

Vojtěch Zeisek, Yann JK Bertrand

Institute of Botany, Czech Academy of Sciences, Průhonice
cluster@ibot.cas.cz, <https://sorbus.ibot.cas.cz/>

30. 10. 2019



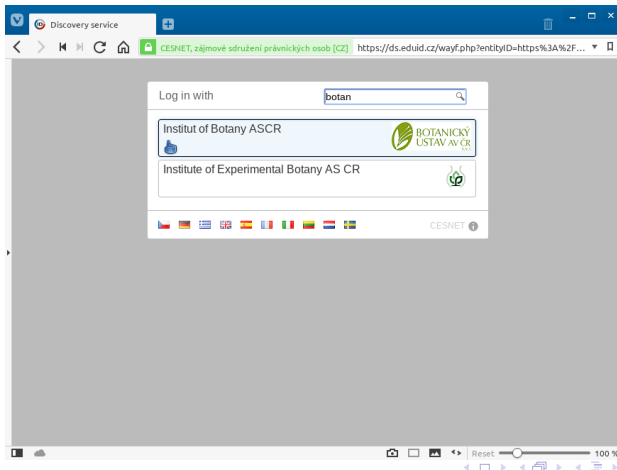
New research tool

- Purchased in 2019 from grant of the Czech Academy of Sciences (almost 3 500 000 CZK)
- Part of **MetaCentrum** national grid infrastructure connecting computing clusters
 - Maintained by MetaCentrum admins
 - All software and other resources available (as for any MetaCentrum computing node)
 - Can be used by any MetaCentrum user
- Connected together by 100 GB network cables
- Connected to the Internet (and rest of MetaCentrum) by optical fibre



Login to all MetaCentrum/CESNET services

- Login to all MetaCentrum/CESNET services, e.g. MetaCentrum computing grid, ownCloud or FileSender is done via EduID
- Logins are redirected to EduID login page

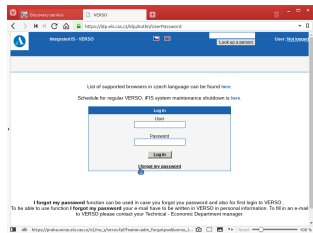


Login for employees of the Institute of Botany, CAS

- Our employees are from EduID login page redirected to custom login page (different for every institution)

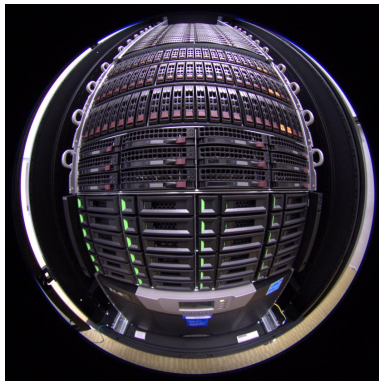


- User without password must first go to <https://praha.verso.eis.cas.cz/> and obtain new login name and password



Hardware overview

- Data storage and file server `tilia-nfs`
- HPC nodes for massive parallelization `draba1-3`
- ‘Standard’ computing nodes `carex1-6`
- Frontend `tilia`
- Database server `sorbus`



Disk system

- File server `tilia-nfs.ibot.cas.cz` (alias `storage-pruhonice1-ibot.metacentrum.cz`)
- Disk array with capacity 179 TB



HPC nodes for massive parallelization

- draba1, draba2 and draba3, each equipped with 4 CPU Intel Xeon Gold 6230 (4x 20 cores (4x 40 threads), 2.1 GHz, turbo 3.9 GHz), 1536 GB RAM and 1920 GB NVMe disk (RAID 0)
- Primarily intended for running computations requiring massive parallelization such as Apache Hadoop and Spark

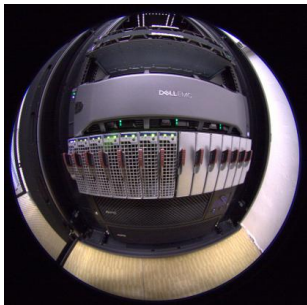


Standard computing nodes

- Each of the six nodes (carex1 to carex6) is equipped with 1 CPU AMD EPYC Naples 7261 (8 cores (16 threads), 2.5 GHz, turbo 2.9 GHz), 512 GB RAM and 1920 GB NVMe disk (RAID 0)
- Intended for any general purpose computations, running of any application



Frontend



- Standard MetaCentrum **frontend** — mainly advantageous to log there in order to process data stored on the Průhonice data array and to send requests to the Průhonice cluster
- 4 vCPU, 2.4 GHz, 8 GB RAM
- Home directories are at Průhonice storage
- Can be reached by SSH at `tilia.ibot.cas.cz` (alias `tilia.metacentrum.cz`)

Database server

- 8 vCPU, 2.4 GHz, 32 GB RAM and 60 GB SSD disk
- `sorbus.ibot.cas.cz` accessible only for dedicated users
- Mainly intended for running NoSQL databases such as Neo4j and MongoDB that are used during Apache Hadoop and Spark computations on the HPC nodes
- Can be used for any special supportive task



Apache Spark

- A computing framework for large-scale data processing in parallel
- The primary interface language is Scala but there are wrappers for java, Python and R
- Spark performs in memory computing which speeds up computation by at least x1000 compared to regular parallel implementations...it is the FUTURE
- See also <https://spark.apache.org/> and <https://hadoop.apache.org/>

When to use Spark?

What is big-data? Data larger than 0.5 GB

- Tabular data (genetic variants, species distribution, geographic location of features...)
- Genomic data (several genomes)
- Metagenomes comparisons
- Multiple sequence alignments (more than 1000) for gene expression, phylogenetics...

Large-scale spatial data with GeoSpark, SpatialSpark, Spark-GIS

- Co-Location Pattern Mining (e.g. in Africa do lions co-locate with zebras?)
- Multivariate Spatial Patterns (e.g. species density correlations, biomass correlations...)
- Classification and clustering (soil types, types of vegetation, response to perturbations)

Bioinformatics

- Read processing (quality cleaning, trimming) (QMSpy)
- Alignment (SparkSW, SparkBWA)
- Mapping (MetaSpark)
- Metagenomics (MetaSpark, SpaRC)
- Assembly (Spaler)
- Blasting (SparkBLAST)
- General sequence analysis (variants discovery, gene expression...) (GATK-Spark)
- Phylogenetics (CloudPhylo)
- Population genetics studies (SEQSpark)

Usage basics

- As any MetaCentrum nodes
- Knowledge of **Linux command line** is required...
- To copy data using graphical application like **FileZilla** connect with SFTP (and MetaCentrum credentials) on port 22 to `tilia-nfs.ibot.cas.cz` (alias `storage-pruhonice1-ibot.metacentrum.cz`)

```
1 # Connect to frontend
2 ssh USER@tilia.ibot.cas.cz # or
3 ssh USER@tilia.metacentrum.cz
4 # Copy data to/from data storage
5 scp -r local/data USER@tilia-nfs.ibot.cas.cz:~/
6 scp -r USER@tilia-nfs.ibot.cas.cz:~/remote/data ldir/
7 # To connect directly to any computing node, e.g.
8 ssh USER@carex1.ibot.cas.cz
9 ssh USER@carex4.metacentrum.cz
10 ssh USER@draba2.ibot.cas.cz
11 ssh USER@draba3.metacentrum.cz
```

SSH access

```

V(e) 16:02:09 veles.tbot očekává příkazy od vojta; ~
$ ssh gmetilla.tbot.cas.cz
gmetilla.tbot.cas.cz's password:
Linux tilla 4.9.0-11-amd64 #1 SMP Debian 4.9.189-3+deb9u1 (2019-09-20) x86_64
Last login: Wed Oct 30 16:02:09 2019 from na...
cesnet

  cesnet

Your Home Directories
-----
Size Avail S.Quota Used Directory
175T 122T 1G 200k /storage/brno11-elixir/home/g
88T 28T 102G 100k /storage/brno8/home/g
68T 32T 102G 60k /storage/budejovice1/home/g
30T 14T 10G 96k /storage/liberec3-tu/home/g
100T 35T 524G 13G /storage/praha1/home/g
142T 71T 1G 140k /storage/praha5-elixir/home/g
179T 179T 1T 72k /storage/pruhonice1-ibot/home/g

CPUs
-----
Free Used Total Centre
-253 5203 4950 CERIT
3727 7471 11198 META

Contact
-----
E-mail: meta@cesnet.cz
Web: http://www.metacentrum.cz/

V(e) 16:02:40 tilla očekává příkazy od gmetilla; ~
$
  
```

- Check e.g. PDF <https://github.com/V-Z/course-linux-command-line-bash-scripting-metacentrum/releases/tag/v2019>

Priority access to storage and computing resources

- Resources are available to all MetaCentrum users
 - Any user may send task of the length up to 24 hours (-l walltime=24:0:0)
 - Default user quota for data storage is 10 GB
- Members of group **ibot** can submit tasks of any length requesting any resources (no limits) — tasks are running immediately as the resources are available
- Members of group **ibot** have default quota on the storage 2 TB (can be anything higher — mail cluster@ibot.cas.cz)
- Group **ibot** is available for anyone having affiliation with the Institute of Botany, or for collaborators
- To become member of the **ibot** group mail cluster@ibot.cas.cz
- It is possible to establish space for data shared by group of users on the storage — such requests must be sent to cluster@ibot.cas.cz

Graphical connection to the storage

Use any favourite SFTP/SCP client...

- MetaCentrum USER connecting on port 22 to `tilia-nfs.ibot.cas.cz` via SFTP/SCP/SSHFS/rsync/...

The screenshot shows a file manager window titled "sftp://tilia-nfs.ibot.cas.cz - FileZilla". The address bar shows the connection path. The status bar at the bottom indicates the total size of the files and folders.

Left Panel (Local File System):

- home/vyuka/dokumenty/
 - x-format
 - zisteni
 - CLC_Data
 - CLC_references
 - k
 - lei
 - calbre
 - dokumenty

Right Panel (Remote File System):

- /
 - home

File List (Left Panel):

Název souboru	Velikost	Typ souboru	Poslední změna
botanik		Složka	25.10.2019 14...
chs		Složka	25.10.2019 13...
clady		Složka	25.10.2019 13...
man		Složka	25.10.2019 13...
metacentrum		Složka	25.10.2019 13...
ekst		Složka	2.9.2019 09:47...
veda		Složka	21.5.2019 14:5...
vyuka		Složka	10.1.2019 09:4...
avatac_42.jpg	25 391	png-soubor	20.3.2015 16:1...
vajtech_zenok.jpg	4 767	png-soubor	28.4.2016 14:0...
zavazek_avajtech_zenok_a_jpic...	525 123	png-soubor	10.7.2017 16:3...
zavazek_eat	26 721	odk-soubor	25.11.2009 19...
zavazek_gacj_paf_2017.odt	17 201	odk-soubor	4.1.2017 10:59...

File List (Right Panel):

Název souboru	Velikost	Typ souboru	Poslední změna	Opisování	Vlastník / Skupina
cache		Složka	10.10.2019 15:17...	drwx	meta
config		Složka	10.10.2019 15:17...	drwx	meta
local		Složka	10.10.2019 15:17...	drwx	meta
parallel		Složka	11.10.2019 11:57...	drwx	meta
sib		Složka	11.10.2019 14:32...	drwx	meta
Authority	51	Soubor	23.10.2019 15:01...	-w	meta
alias	608	Soubor	11.10.2019 15:02...	-w	meta
bash_history	5 067	Soubor	23.10.2019 15:02...	-w	meta
bashrc	2 480	Soubor	11.10.2019 15:01...	-w	meta
fehervpl	8 963	pl-soubor	23.10.2019 14:08...	-w	meta
xlogin	28	Soubor	8.10.2019 13:20:25...	-w	meta
profile	414	Soubor	11.10.2019 15:02...	-w	meta

Status Bar:

5 souborů a 8 složek. Celková velikost: 599 203 bajtů

7 souborů a 5 složek. Celková velikost: 17 011 bajtů

Server nebo místní soubor:

Server nebo místní soubor	Směr	Vzdálený soubor	Velikost	Priorita	Stav
---------------------------	------	-----------------	----------	----------	------

Formy ve formě: Neúspěšné přenosy, Úspěšné přenosy

Access to data from any MetaCentrum node and from scripts

- Home of fontend `tilia` is on the storage `tilia-nfs`
- Storage is available from any MetaCentrum node in `/storage/pruhonice1-ibot/home/USER/`
- As any other MetaCentrum storage, just note particular path...

```
1 # In scripts use something like
2 cp -a /storage/pruhonice1-ibot/home/USER/data \
3     "$SCRATCHDIR"/
4 # to copy data to computing node, and
5 cp -a "$SCRATCHDIR"/results \
6     /storage/pruhonice1-ibot/home/USER/
7 # to copy results back to the storage
```

Sending tasks to Průhonice cluster

- Submit task script as usually using qsub command, just add parameter -q ibot
- Try more options in https://metavo.metacentrum.cz/pbsmon2/qsub_pbspro

```
1 # Submit task by something like
2 qsub -l walltime=1:0:0 -q ibot \
3     -l select=1:ncpus=4:mem=4gb:scratch_local=1gb \
4     -m abe script.sh
5 # Or specifically select carex node
6 qsub -l walltime=1:0:0 -q ibot -l cluster=carex \
7     -l select=1:ncpus=4:mem=4gb:scratch_local=1gb \
8     -m abe script.sh
9 # Or specifically select draba node
10 qsub -l walltime=1:0:0 -q ibot -l cluster=draba \
11     -l select=1:ncpus=4:mem=4gb:scratch_local=1gb \
12     -m abe script.sh
```

Requesting whole node and using hyperthreading

- Queuing system reserves only physical CPU cores and does not work well with hyperthreading
- To use hyperthreading (advantageous for some applications, like Java and some parallelization), user must request whole node using `-l place=exclhost`
- MetaCentrum admins are working on better reservation of whole nodes and handling of hyperthreading

```

1 # To reserve whole carex node use something like
2 qsub -l walltime=1:0:0 -q ibot \
3     -l select=1:ncpus=8:mem=500gb:scratch_local= \
4     1600gb:hyperthreading=True:cluster=carex \
5     -l place=exclhost -m abe script.sh
6 # To reserve whole draba HPC node use something like
7 qsub -l walltime=1:0:0 -q ibot \
8     -l select=1:ncpus=80:mem=1500gb:scratch_local= \
9     1600gb:hyperthreading=True:cluster=draba \
10    -l place=exclhost -m abe script.sh

```

Default limits

- Members of group **ibot**
 - No limits for computing resources
 - Default quota for data storage is 2 TB — requests to change to be send to cluster@ibot.cas.cz
 - We can discuss and set different default limits for our users
- Any other MetaCentrum users
 - Any number of CPU, RAM, disk space, etc., but the task must end up within 24 hours (no longer walltime than 24:0:0)
 - Default quota for data storage is 10 GB
 - We can discuss and set different default limits for other users
- Content of shared directories is counted to quotas of respective users — their storage quotas must be set accordingly

Support and maintenance

- The cluster is managed by administrators of the **MetaCentrum** — regarding general problems related to computation, application etc. please use the **MetaCentrum help support**
- Specific questions related to the Průhonice cluster, e.g. issues connected with the usage of the database server, should be addressed to cluster administrators Vojtěch Zeisek and Yann JK Bertrand
- If you desire to join the **ibot** group, create or modify shared folder and change your allowed disk space on the disk storage, please contact Vojtěch Zeisek
- Questions related to Apache Hadoop, Spark, NoSQL databases (Neo4j and MongoDB) and other issues specific computing on the HPC nodes should be directed to Yann JK Bertrand
- **Contact Vojtěch Zeisek and/or Yann JK Bertrand on cluster@ibot.cas.cz**

Thank you for your attention...

Let's discuss...

- Which default limits do we wish to set?
- How do we wish to use shared data folders? Any use case? Any requests?
- Any requests to higher quota on the storage?
- Any requests to run something special on the database server?
- **Any question from anyone? :-)**