Research e-Infrastructures in Czech Republic (e-INFRA CZ)

for scientific computations, collaborative research & research support

Tomáš Rebok

CERIT-SC, Institute of Computer Science MU

MetaCentrum, CESNET z.s.p.o.

(rebok@ics.muni.cz)

e-INFRA CZ

Covers three e-infrastructures – MetaCentrum (NGI), IT4I (SuperComputers) and CERIT-SC (NGI & Research Collaborations)

Available to all academic users from Czech universities, Academy of Science, research institutes, etc.

- commercial bodies mainly for public research

Offers:

- **computing resources**
- **storage resources**
- **application programs**

Computing services:

- **grid services**
- **supercomputing services**
- **cloud services**
- **storage services**
- **specialized environments**

https://www.e-infra.cz

Research e-Infrastructures in CR I.

MetaCentrum National Grid Infrastructure (NGI)

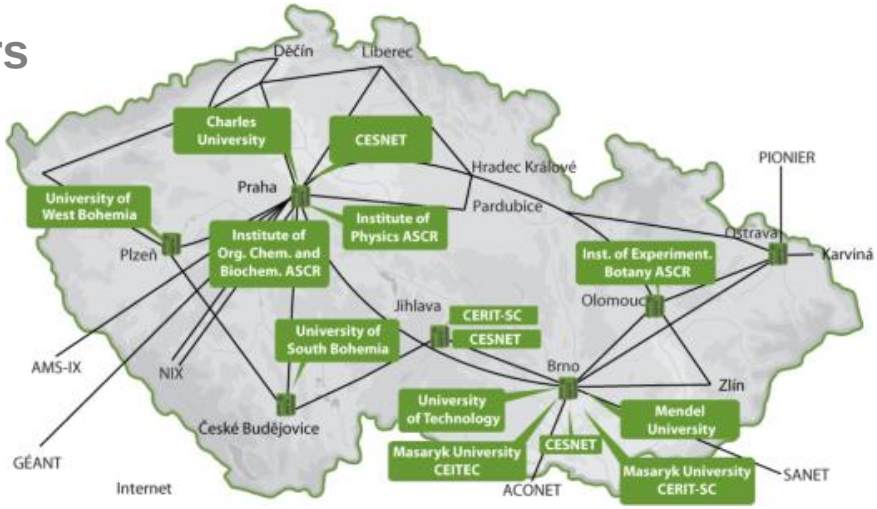
- operated by MetaCentrum NGI (CESNET) since 1996
- MetaCentrum responsible for management and coordination

https://www.metacentrum.cz

Distributed infrastructure

NGI integrates medium/large HW centers (clusters, powerful servers, storages) of several universities/institutions

- further integrated into the European Grid Infrastructure (EGI.eu)



BTW: **Computing clusters**

a set of interconnected („common“) computers

(oldschool)

BTW: **Computing clusters**

a set of interconnected („common“) computers

(nowadays)

Research e-Infrastructures in CR II.

- ***IT4innovations (Supercomputing Centre in Ostrava)***
  - **3344 computing cores** („small“ supercomputer/cluster)
  - **24192 computing cores** („big“ supercomputer/cluster)
  - attributes:
    - computing time assigned to **research projects**
    - **formal application** is necessary (evaluation of research and technical readiness + financial contributions)
    - **public competitions** 2x per year
    - if accepted, easier resource access (low number of competitive users)
  - purpose:
    - **large (proven) computations using homogeneous infrastructure**

https://www.it4i.cz

Research e-Infrastructures in CR III.

Centre CERIT-SC (Masaryk University in Brno)

- **long-term history (→ long-term experience in ICT science)**
  - CERIT-SC evolved from Supercomputing Center Brno (established in 1994), and
  - participates on the operation of National Grid Infrastructure

CERIT-SC mission:

- **production services for computational science**
  - high-performance computing clusters
  - large data storage, back-ups and data archives
  - web portals & projects' back-office
- **an application of top-level ICT in the science**
  - own research in e-infrastructures (know-how)
  - novel forms of infrastructure utilization (experimental usage support)
  - **research collaborations with other science areas**

https://www.cerit-sc.cz

e-INFRA CZ – Basic grid/cloud characteristics

- after registration, **grid and cloud resources** are available **without any administrative burden**
  - → ~ immediately (depending on actual usage)
  - **free of charge**
    - users "pay" via publications with acknowledgements
      - → results in user priorities in cases of high load
  - **no applications for resources**

- **user accounts extensions every year**
  - **validates users' relationship to an academic institution**
    - federated infrastructure eduID.cz used for minimalization of users' burden
  - **reports of user publications with acknowledgements to MetaCentrum/CERIT-SC**
    - used as a proof of infrastructure benefits for Czech research area
- **best-effort service**

e-INFRA CZ – Grid computing resources available

Grid computing resources: ca 14900 cores (x86_64)

- nodes **with lower number of computing cores**: 2x4-8 cores
- nodes **with medium number of comp. cores (SMP nodes)**: 32-80 cores
- memory (RAM) up to **1 TB per node**
- nodes with **high number of computing cores: SGI UV2000**
  - **384 cores (x86_64), 6 TB of RAM**
  - **288 cores (x86_64), 6 TB of RAM**
- other **„exotic“ hardware:**
  - **nodes with GPU cards, SSD discs, Xeon Phi, etc.**

http://metavo.metacentrum.cz/cs/state/hardware.html

e-INFRA CZ – Storage resources available

ca 3 PB for operational data

- centralized storage arrays distributed through various cities in the CR
- user quota 1-3 TB on each storage array

ca 22 PB for archival data

- (HSM – MAID, tapes)
- “unlimited” user quota

http://metavo.metacentrum.cz/cs/state/nodes

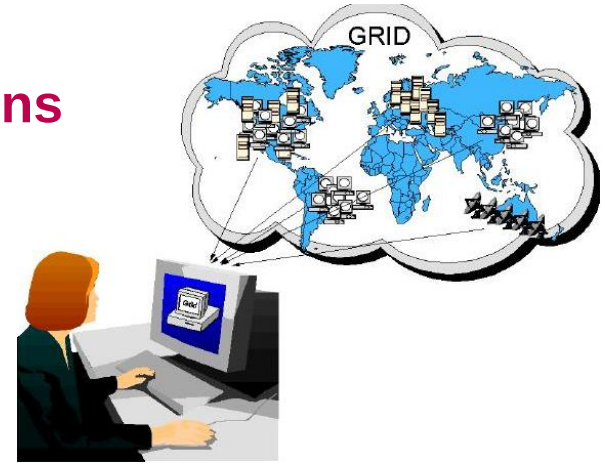
e-INFRA CZ – Software available

~ 300 different applications (commercial & free/open s.)

- see http://meta.cesnet.cz/wiki/Kategorie:Aplikace

- **development tools**
  - GNU, Intel, and PGI compilers, profiling and debugging tools (TotalView, Allinea), …
- **mathematical software**
  - Matlab, Maple, Mathematica, gridMathematica, …
- **application chemistry**
  - Gaussian 09, Gaussian-Linda, Gamess, Gromacs, …
- **material simulations**
  - Wien2k, ANSYS Fluent CFD, Ansys Mechanical, Ansys HPC…
- **structural biology, bioinformatics**
  - CLC Genomics Workbench, Geneious, Turbomole, Molpro, …

e-INFRA CZ – Grid computing services



Suitable for:

- **long-term and/or large-scale computations**
  - **(primarily batch processing)**
- **applications not requiring special OSs (features)**
  - **pre-installed or users' ones**

How to compute?

- ***batch jobs***
  - **the work described by job scripts and managed by scheduler(s)**
  - **sets of many jobs, long-running jobs, …**
- ***interactive jobs***
  - **text interface**
  - **graphical user interface (VNC)**

e-INFRA CZ – Cloud computing services

Suitable for:

- applications requiring special environment (OS) and/or features
  - one can run various operating systems (incl. Windows OS) and/or application equipment
  - administrator/root access is provided
  - still focused on research computations (not for webhosting)
    - Windows & Linux images provided, user-uploaded images also supported

How to compute?

- OpenNebula GUI for deployed nodes management
  - nodes simply created based on predefined templates
  - new version being built on OpenStack
- interact via:
  - cloud/VM console
  - internal OS services (SSH, VNC, Rdesktop, …)

e-INFRA CZ – Hadoop computing services

Hadoop:

- an open-source framework for distributed storage and distributed processing of large volumes of data
  - large data blocks splitted and distributed amongst nodes
  - a MapReduce-based algorithm (= data processing code) is distributed over the distributed blocks and processed in parallel

Suitable for:

- huge datasets to be processed
  - but NOT suitable for arbitrary data processing one can imagine
  - just for the processing meeting the MapReduce programming model
    - e.g., counting the number of times words occur in a corpus

e-INFRA CZ – Specialized frontends/environments

Suitable for:

- user communities with well-defined processing needs
  - workload & computing pipeline orchestrators available via GUI
  - usually adapted to user needs and/or because of interoperability with our infrastructure
    - in background, the proper computing method is used (grid, cloud, etc.)

How to compute?

- Galaxy, Chipster, etc.

e-INFRA CZ – How to become our user?

- ***register***
  - http://metavo.metacentrum.cz , section *„Application“*
  - EduID.cz => **proves your academic identity** using your home institution services (and credentials)
    - foreign users (researchers) have to be somehow related to a research group in CR
- ***make yourself familiar with basics of OS Linux***
  - http://metavo.metacentrum.cz , section *„Documentation“*
  - http://www.poznejlinux.cz/linuxbook/xhtml-chunks/ch07.html
- ***compute***

Conclusions

- **Research e-infrastructures in CR – e-INFRA CZ**
  - **comprises of MetaCentrum (CESNET), IT4I (Ostrava) and CERIT-SC (MUNI) infrastructures**
    - *ongoing tight integration*
  - *computing services* (grid, cloud, Hadoop/specialized and supercomputing services)
  - *data services* (archivals, backups, data sharing and transfers, …)
  - *services supporting collaborative research* (research collaboration with partners, project partners supporting projects‘ IT part, etc.)
  - further supporting services (remote collaborations support services, etc.…)

- *The message:* „If you cannot find a solution to your specific needs in the provided services, let us know – we will try to find the solution together with you…“

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

OP Research and Development for Innovation

The CERIT Scientific Cloud project (reg. no. CZ.1.05/3.2.00/08.0144) is supported by the *Operational Program Research and Development for Innovations*, priority axis 3, subarea 2.3 *Information Infrastructure for Research and Development.*

https://metavo.metacentrum.cz https://www.it4i.cz https://www.cerit-sc.cz