

Czech National E-infrastructure Status and Plans: Storage



David Antoř

CESNET, Prague

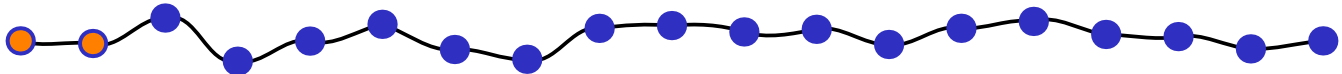
Czech Republic

`antos@ics.muni.cz`



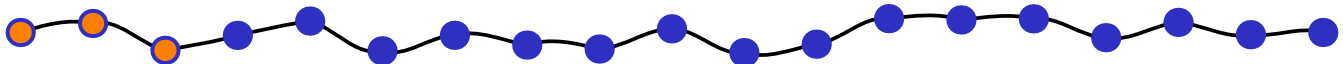
Overview

- MetaCentrum: the Czech NGI
 - resources and architecture
 - filesystem-based storage in MetaCentrum
 - concept of NGI storage development
- transforming into a national e-infrastructure
 - potential users
- e-infrastructure projects
 - planned storage infrastructure



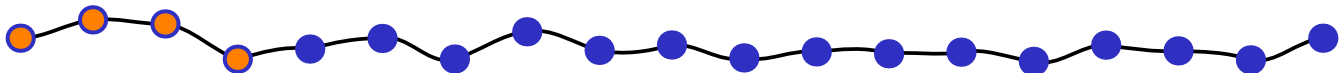
MetaCentrum—Czech NGI

- actors:
 - CESNET (Czech research and education network)
 - MetaCentrum: Czech National Grid Initiative, responsible for computing and storage grids
- about 3000 CPU cores
 - half of them dedicated to HEP and astrophysics communities
 - see <http://meta.cesnet.cz> for details

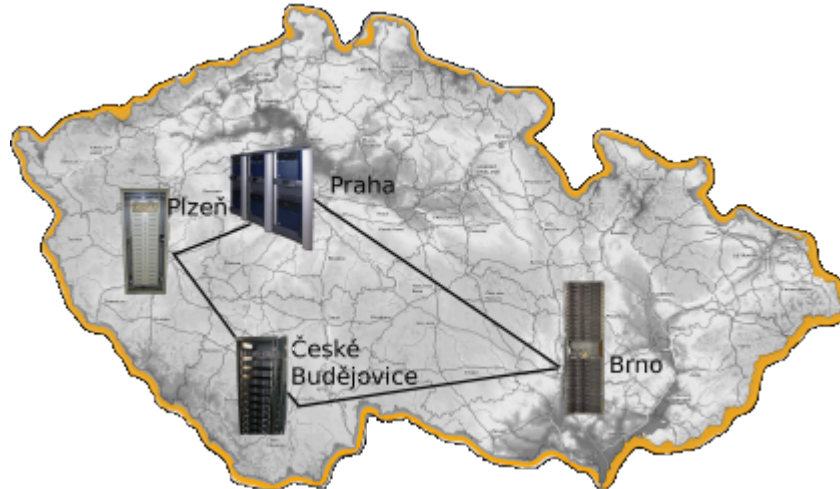


MetaCentrum Resources

- MetaCentrum resources are hosted and owned by
 - Institute of Physics, Academy of Sciences, Prague
 - Masaryk University, Brno
 - West Bohemian University, Pilsen
 - Charles University, Prague
 - University of South Bohemia, České Budějovice
 - University of Technology, Brno
 - Mendel University, Brno
 - CESNET, Prague
- heterogenous resources
 - SMP machines, clusters
- usage policies vary



Major MetaCentrum Sites

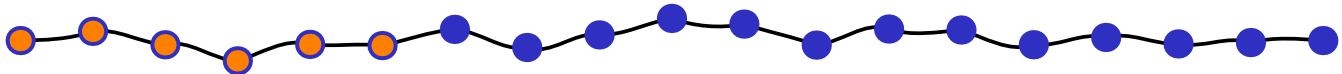


- up to 5 ms network latency over 10GE lines

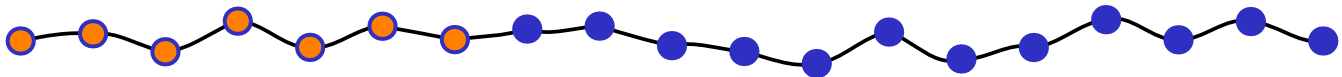


Current MetaCentrum Storage Experience

- local filesystems
 - /scratch
 - ★ internal disks of nodes
 - ★ Lustre on selected clusters, experimenting with PVFS
 - ▷ networked scratch, to achieve higher speeds and/or capacity
 - ▷ experimental
 - /home
 - ★ shared NFSv3 for a cluster
 - ★ eventually a site
 - ▷ physically collocated clusters (\approx single room)
 - ★ physically close to the cluster

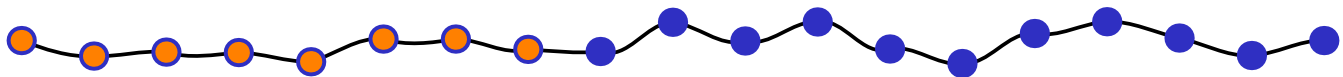
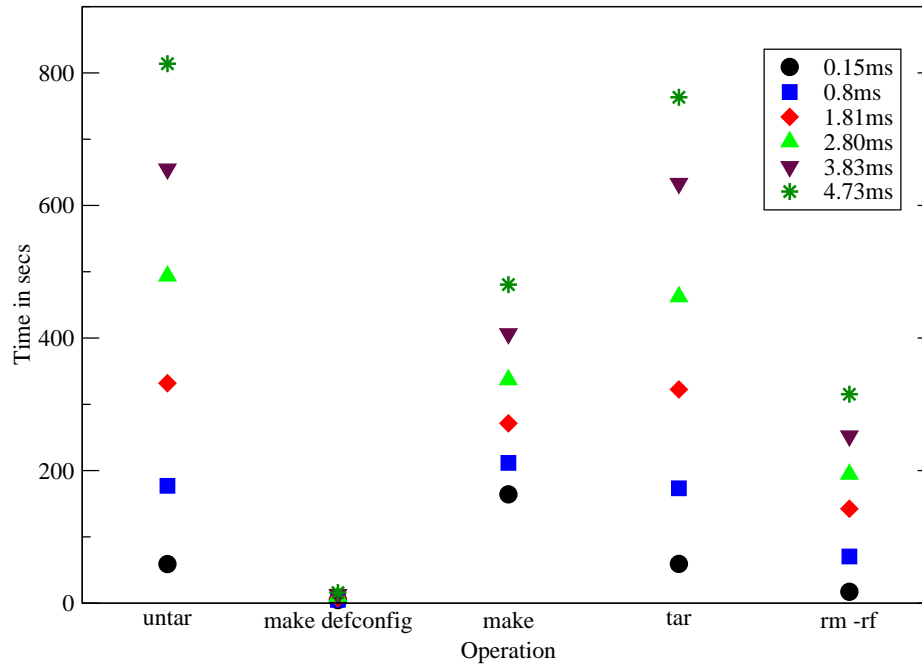


- global filesystems
 - shared by cluster nodes
 - can be also attached to user workstations
 - ★ AFS and NFSv4 with Kerberos
 - AFS for all nodes
 - ★ great concept
 - ▷ ACL setting, ideal for software installations/project data
 - ★ relatively slow
 - ★ implementations are not up-to-date, e.g., 2GB volume limits
 - NFSv4 for nearly all nodes
 - ★ good over local networks, reasonable over long distances
- overall capacity
 - about 200 TB in disk arrays (100 TB in Brno NFSv4)
 - 400 TB tape library for backups (Brno and Pilsen)



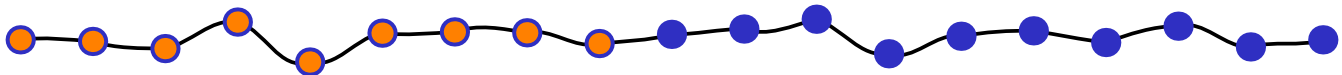
Throughput and Latency Relationship

NFSv4 and latency



Questions of the NGI Storage

- filesystem or storage depots (with simple put/get semantics)?
- local only or global filesystems?
 - local filesystems with automatic replication
 - ★ integrated with job scheduling
- more structured file access (e.g., databases)?
- access and use patterns versus technologies
 - performance vs. convenience
- metadata/catalogue implications for really large systems
 - relationship with filesystems



Transforming into an E-infrastructure

(aka Politics)

- partially caused by changes in financing research in the Czech Republic
 - consolidating enormous number of grant agencies
 - allowing to support big research projects and national level teams
 - ★ relationship with European Strategy Forum on Research Infrastructures et al.
- building complex coordinated e-infrastructure in the Czech Republic
 - coordinated by CESNET
 - but not limited just to CESNET activities



Potential Users (in Preparation or Evaluation)

- just examples
 - BIOCEV (Biotechnology and Biomedicine Centre)
 - CEITEC (Central European Institute of Technology)
 - CESSDA (Council of European Social Science Data Archives)
 - CLARIN (Common Language Resource and Technology Infrastructure)
 - INFRAFRONTIER (The European Infrastructure for Phenotyping and Archiving of Model Mammal Genomes)



Case Study: CEITEC—Life Sciences

- analysis of requirement for the project (first half of 2009)
- data is used from several months to 2 years and then archived
 - automation of archiving is preferred
- big data producers: instruments and computations
 - DNA sequencers—4 TB/day of primary data
 - confocal microscopes
 - amount of data from computations depends on available computation power
- availability of storage changes the way scientists work
 - time-scale comparisons, re-checking primary data



E-infrastructure Projects (in Evaluation)

- 


- coordinator of the national e-infrastructure
 - ★ NREN (since 1996)
 - ★ NGI
- storage (\approx 4M EUR, three sites)

- IT4Innovations

- Technical University of Ostrava
- standard self-contained supercomputing centre
- connected to the national e-infrastructure



E-infrastructure Projects (cont.)

- **CERIT**—Centre for Education, Research, and Innovation in IT
 - Masaryk University, Brno
 - CERIT-SC
 - ★ based on history of the Supercomputing Centre Brno (since 1994)
 - ★ high performance supercomputing centre
 - ★ and major grid centre in the Czech Republic
 - ▷ bridge between HPC and grid at national level
 - ★ storage facilities (\approx 4M EUR, two instances)



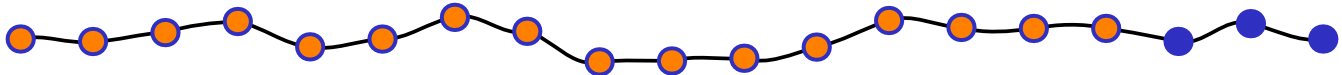
CESNET Storage: Initial Research

- we had long discussions with major storage vendors about
 - a global file system over the Czech Republic
 - but less sensitive to latency problems than NFSv4
- offered solutions were not flexible enough
 - typically HSM-based with proprietary software
- strategy
 - to buy just separate HSM systems
 - advanced services will be realised step-by-step
 - ★ either using third-party solutions
 - ★ or our own development



Planned CESNET Storage Infrastructure

- three sites with HSM systems
- built 2011–2013
- estimated capacity about 5 PB each
 - in current capacity/price ratio
- replications, geographic distribution, . . . managed by middleware
- with objectives strongly inspired by Norstore
- filesystem access, SRM, GridFTP, iRODS, databases, . . .



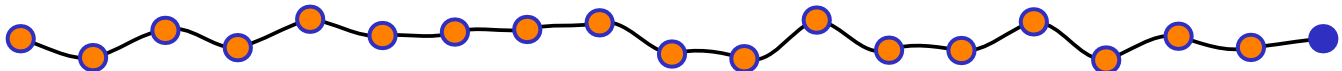
Planned CERIT Storage Infrastructure

- CERIT-SC: HPC and grid centre
 - located at single address/site
- storage resources closely coupled with processing facilities
 - large storage depots for distinct scientific communities
- need for storage hierarchy
 - top level directly connected to CESNET systems



Looking for a collaboration with...

- institutions and teams with similar problems
- running comparable infrastructures
- major topics include
 - less latency-sensitive filesystems
 - ★ MetaCentrum users (except HEP) insist on filesystems
 - ★ or: is it better to convince the users?
 - federated access to filesystems
 - long-distance replicas
 - archiving
 - long-term data preservation



Thanks for your attention

