



cesnet  
"...."

METACENTRUM

Miroslav Ruda

CESNET

---

Konference CESNET, Praha  
29.1. 2019



Dnes obecná přednáška

- představení, služby
- příklady spolupráce

Zítرا celodenní uživatelský seminář MetaCentra

- technické podrobnosti, hands-on
- kontext e-Infra CZ (CERIT-SC, IT4I), mezinárodní (EOSC)
- příklady využití našimi uživateli
- technická přednáška o NVIDIA akcelerátorech a cloudu

Národní distribuované výpočetní prostředí

- výpočetní zdroje na univerzitách, AV ČR
- gridové, cloudové a map-reduce výpočty
- NGI v evropské e-infrastruktuře EGI, EOSC
- zahrnuje zdroje tří VI (CESNET, CERIT, ELIXIR)

Původní motivace sdílení zdrojů (HW) stále platí

- přenesení nárazové zátěže na volnější zdroje
- a využití jiných zdrojů při výpadku
- poskytnutí vlastních dočasně volných zdrojů
- drahé komerční licence

Záběr postupně rozšířen o cloudové služby,  
zpracování rozsáhlých dat, cílenou podporu  
velkých projektů



## Služby MetaCentra (pro)

- jednotlivce - chtějí počítat, zpracovávat data
- projekty - chtějí počítat a potřebují spolupracovat, sdílet data
- instituce - chtějí poskytovat svoje výpočetní zdroje projektům, chtějí správu svých zdrojů

## Podpora uživatelských komunit/spolupráce

- LHC, Auger, CTA, Belle
- ELI, ELIXIR, BBMRI, CLARIN, ICOS
- CzechGlobe, CEITEC
- ESA – CollGS, Data Relay Hub

- **aktuálně 18.000 CPU jader (x86\_64)**
  - HTC uzly s menším počtem jader (2x4-16)
  - SMP servery - 32-80 jader, do 1 TB RAM
  - dva servery SGI UV2000 - 288/384 jader, 6 TB RAM
  - výpočetní uzly s GP-GPU kartami
- **VI CESNET (7.000) a VI CERIT-SC (6.000)**
  - FZU, VI ELIXIR, CEITEC, ZČU, JČU, MU
- **cca 5 PB diskových prostor na zpracování dat**
  - NFS, GPFS
  - HDFS pro Hadoop/Spark
  - objektová storage Ceph
  - DPM úložiště

- grid - centrální správa, primárně pro výpočty
  - dávkové, dlouhé, paralelní, master-slave, HTC
  - včetně interaktivních úloh, grafické rozhraní
- MetaCentrum cloud - místo úloh virtuální stroj
  - obrazy MetaCentra, projektové, uživatelské
  - ne obecný web-hosting, cloudové výpočty
- MapReduce - Hadoop/Spark, zpracování dat
- řada aplikačního software - komerční
  - mat. software (Matlab, Maple, Mathematica),
  - vývoj (Intel, PGI, Totalview, Allinea)
  - technické simulace (Ansys Fluent, CFX)
  - biochemie (Amber, Gaussian, Turbomole)
- open-source (life-science, matematika)
  - nárůst balíků pro strukturální biologii, bioinformatiku, sekvenování a analýzu DNA

- primární je akademické užití - výzkumné, vývojové a výukové účely
  - okamžitý přístup, bez podávání projektů
  - "placení" formou publikací s poděkováním
  - publikace využívány pro určení priority uživatele
- komerční využití je možné, ale při dodržení CESNET AUP, licenčních podmínek HW a SW
  - zpravidla na zdrojích vlastněných jiným subjektem
  - MetaCentrum zdroje za úplatu neposkytuje
- další zdroje poskytují vědecké skupiny
  - centrální správa zajištěná MetaCentrem
  - vlastníci mají dedikovaný nebo privilegovaný přístup, konkrétní forma podle dohody



## Projekty s vědeckými partnery

- ELIXIR - EXCELERATE, VI ELIXIR
- ESA - CollGS, DataHub Relay
- EOSC (Pilot, Hub) - cesta k dalším partnerům
- zkoušíme cestu k ELI – projekt PaNOSC

## Budování e-infrastruktury

- národní e-INFRA CZ (CESNET, CERIT-SC, IT4I)
  - VI ELIXIR, (OpenScreen, Biolmaging, CCP)
- EOSC projekty
- DEEP-HybridDataCloud
- ELIXIR AAI -> LifeScience AAI
- AARC2, GN4

- oblast life-science dlouhodobě výrazně aktivní
- dedikace části zdrojů pro experimenty
  - změna našeho prostředí
  - vývoj specifických řešení
- urychlení zapojení partnerů do projektů
- podpora specifických rozhraní
  - Mascot, Galaxy, Chipster
- postupně se nové vlastnosti a balíky dostávají do celého MetaCentra
- VI CESNET (a CERIT-SC) je členem českého konsorcia ELIXIR, IT partnerem v projektech
  - zdroje dedikované pro projekt
  - IT podpora pro vybrané skupiny
- provoz služeb ELIXIR CZ

## Spolupracující pozemní segment ESA Sentinel

- družicové snímky zemského povrchu
  - platformy S1 (radar), S2 (foto), S3 (multispektrální)
- „velká data“ blízko k výpočetním kapacitám
- plný archiv území ČR a nejbližšího okolí
- podpora uživatelů při přístupu k datům:
  - konzultace, „custom“ filtry, další regiony
  - účast v projektech jako technologický partner
    - zpřístupnění dat
    - automatizaci zpracování
    - poskytnutí zdrojů (při nekomerčním využití)
- součást sítě DataHub Relays
  - smluvní vztah s ESA: zrychlený přísun snímků celého světa.

The logo for cesnet, featuring the word "cesnet" in a white, lowercase, sans-serif font. Below the text is a graphic element consisting of a series of white dots of varying sizes, arranged in a pattern that suggests a network or data flow.

cesnet  
.....

Děkuji za pozornost

<http://www.metacentrum.cz>



- včera obecná přednáška o MetaCentru
  - oddělení CESNETu, součást VI CESNET
  - aktuální stav a plány
- dnes uživatelský seminář MetaCentra
  - novinky a plány z pohledu uživatelů
  - mezinárodní kontext (EOSC), projekty
  - širší pojetí MetaCentra v rámci nové VI e-INFRA CZ
    - novinky za CESNET, CERIT-SC, IT4I
    - tradičně i Datová úložiště CESNET
  - trochu více o MetaCentrum cloudu
  - využití systému Perun pro LifeScience AAI
  - dva příklady využití z řad uživatelů
  - dopoledne i představení NVIDIA akceleratorů a cloudové podpory

## Národní distribuované výpočetní prostředí

- výpočetní zdroje na univerzitách, AV ČR
- gridové, cloudové a map-reduce výpočty
- NGI v evropské e-infrastruktře EGI, EOSC
- zahrnuje zdroje tří VI (CESNET, CERIT, ELIXIR)

## Výpočetní zdroje

- aktuálně 18.000 CPU jader (x86\_64)
  - HTC uzly s menším počtem jader (2x4-16)
  - SMP servery - 32-80 jader, do 1 TB RAM
  - výpočetní uzly s GP-GPU kartami
- největší poskytovatelé - CESNET a CERIT-SC
- cca 5 PB diskových prostor na zpracování dat

## Rozvoj koordinovaný s dalšími VI

- e-INFRA CZ – CESNET, CERIT-SC, IT4I
- ELIXIR, OpenScreen, BioImaging, CCP

- obnova clusterů vlastněných CESNETem
  - nympa - 64 uzlů, 2xXeon 6130, 192 GB RAM
  - tento rok GP-GPU (cca 40 uzlů, NVIDIA T4)
- zapojen cluster TUL Liberec
- MetaCentrum Cloud na platformě OpenStack
  - společná aktivita VI CESNET a VI CERIT-SC
  - běží v testovacím režimu, tento rok produkční nasazení, odstavení staré OpenNebuly
  - podrobnosti v další přednášce
- rozvoj systému Perun pro potřeby ELIXIR/ LifeScience AAI
  - společně s CERIT-SC
- plány do dalšího VVV (2020-2022) - výrazný nárůst výpočetních kapacit

1. 1. - 31. 12. 2018	Celkem MetaVO	Torque + PBS Pro @arien	Torque + PBS Pro @wagap	Cloud (CESNET)	Cloud (CERIT-SC)	Hadoop	EGI HPC	EGI Cloud
Počet úloh / spuštěných WM strojů	5038 tis.	4012 tis.	1010 tis.	3 043	7 619	4265 (map/reduce tasks 17240)	6698 tis.	32 tis.
Propočítaný CPU čas (walltime) [CPU let]	11357	5192	3821	1420 (povolen overbooking)	912	12	4004	70
Počet uživatelů k 31. 12. 2018	2020 MetaVO			285 MetaCloud		109 alespoň jedna úloha, 154 registrováno	5 tis.	n/a
Počet evidovaných CPU jader ke konci období	17860 (včetně cloudu a Hadoop)	10580 CESNET 5824 ostatní 4756	6132 včetně cloud CERIT-SC	732	CPU jádra sdílená s jádry CERIT-SC zapojeními do PBS	416	3276 FZÚ 496 CESNET	208 CESNET
Storage	obsazeno 1,5 PB z celkových 2,9 PB		obsazeno 300 TB		obsazeno 360 TB z 1 PB		uloženo 3,7 PB	započítáno v národním cloudu
Publikace s poděkováním MC/NGI z 2018	276 (očekávaný nárůst s prodlužováním účtu)							



	2010	2012	2014	2016	2017	2018
Počet uživatelů	420	613	1112	1611	1908	<b>2020</b>
Noví uživatelé	139	312	605	742	732	<b>713</b>
Počet úloh	590 tis.	1,1 mil.	3,9 mil.	3,6 mil.	4,7 mil.	<b>5 mil.</b>
Počet CPU	1560	6028	10160	13540	14686	<b>17860</b>
Propočítaný CPU čas	730 CPU let	2 500 CPU let	6403 CPU let	9475 CPU let	10572 CPU let	<b>11357 CPU let</b>

- **life-science, dlouhodobě aktivní spolupráce**
  - CEITEC, Loschmidt Laboratories
  - stávající VI ELIXIR, nový projekt do VVV
    - dedikovaný hardware na výpočty VI
    - provoz služeb Galaxy, Chipster
    - provoz a vývoj služeb VI
  - ELIXIR EXCELERATE – cloudové prostředí pro life-science
  - další cluster bude zřejmě umístěn na Biocevu
    - podpora Bioimaging, OpenScreen, CCP
- **dálkový průzkum Země - ESA Sentinel**
  - data přímo dostupná pro další zpracování
  - hledáme partnery pro spolupráci
- **PaNOSC (Photon and Neutron ESFRI cluster)**
  - projekty se shodují na požadavcích na Jupyter Notebook a Hub

The logo for cesnet, featuring the word "cesnet" in a white, lowercase, sans-serif font. Below the text is a graphic element consisting of a series of white dots of varying sizes, arranged in a pattern that suggests a network or data flow.

cesnet  
.....

Děkuji za pozornost

<http://www.metacentrum.cz>

