

Vědecké výpočty v MetaCentru

Miroslav Ruda

CESNET

CESNET Day, Liberec, 20.9. 2017



Národní distribuované výpočetní prostředí

- výpočetní zdroje umístěné na univerzitách, AV ČR
- gridové, cloudové a map-reduce výpočty
- NGI v evropské e-infrastruktuře EGI, EOSC

Původní motivace sdílení zdrojů (HW, SW) 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 u nás počítat, zpracovávat data
- projekty - chtějí počítat a potřebují spolupracovat
- 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

Projektové zajištění

- rozvoj infrastrukturních služeb - projekty H2020
 - INDIGO DataCloud, AARC2, GN4, EXCELERATE
- integrace s evropskou e-infrastrukturou EGI
 - EGI ENGAGE, EOSC



- aktuálně 14.000 CPU jader (x86_64)
 - HTC uzly s menším počtem jader (2x4-12)
 - 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
- největší poskytovatelé jsou CESNET, CERIT-SC, FZU
 - clustery CEITEC, ZČU, JČU, MU, AV ČR, ČVUT
- 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
 - jsou možné i interaktivní úlohy, grafické rozhraní
- cloudové výpočty - místo úloh celý virtuální stroj
 - obrazy MetaCentra, projektové, uživatelské
 - ne obecný web-hosting
- MapReduce - Hadoop/Spark, zpracování dat
- řada aplikačního software - komerční
 - mat. software (Matlab, Maple, Mathematica),
 - vývojové prostředí (Intel, PGI, Totalview, Allinea)
 - technické simulace (Ansys Fluent, CFX)
 - biochemie (Amber, Gaussian, Turbomole, Molpro)
- open-source (life-science, matematika, statistika)
 - nárůst balíků pro strukturální biologii, bioinformatiku, sekvenování a analýzu DNA

- primární je akademické užití - výzkumné, vývojové a výukové účely
 - podpora volného přístupu, 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 hardware a software
 - 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 na svoje zdroje, konkrétní forma podle dohody



- oblast life-science je dlouhodobě největší uživatel
- dedikace části zdrojů pro experimenty se změnou našeho prostředí, vývoj specifických řešení
- podpora specifických rozhraní (Mascot, Galaxy, Chipster)
- urychlení zapojení partnerů do mezinárodních projektů
- postupně se nové vlastnosti a balíky dostávají do celého MetaCentra
- CESNET je členem českého konsorcia ELIXIR, IT partnerem v připraveném projektu VI ELIXIR
 - zdroje dedikované pro projekt
 - IT podpora pro vybrané skupiny



Děkuji za pozornost

<http://www.metacentrum.cz>