



METACENTRUM

Miroslav Ruda

CESNET

březen 2021



Výpočty a ukládání dat = MetaCentrum a Datová úložiště

- základní přehled pro nové příchozí, novinky pro naše uživatele
- konference CESNET 25 – trocha historie z pohledu MetaCentra
 - a zároveň osvětlení dnešní skladby prezentací
- služby a novinky MetaCentra
- služby a novinky Datových úložišť
- národní kontext - e-Infra-CZ – novinky z IT4Innovations
- mezinárodní kontext - European Open Science Cloud
- novinky z projektů dálkového průzkumu Země (Sentinel)

Odpoledne – hands-on/tutoriály Semináře gridového počítání



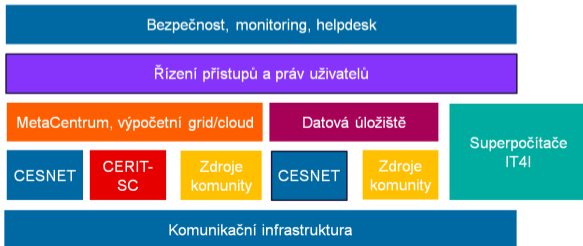
- 1994 - superpočítačová centra MU, VUT, UK, ČVUT, VŠB, ZČU
- 1996 - MetaCentrum jako grid - distribuovaný superpočítač
 - centrální správa AAI (Kerberos), HW, SW (modules)
- 1998 - MetaCentrum součást CESNETu (projekt TEN-155 CZ)
 - aplikace využívající národní síť, uživatel videokonferencí
 - motivace vývoje v AAI (certifikáty, federace)
 - správa identit a účtů Perun (vývoj od 1999, 2021 vlastní oddělení)
 - centrální úložiště uživatelských dat (NFS), software (AFS - 1996)
 - NFSv4 (2007), GPFS, Ceph, CVMFS (2020)
 - 2000 - první výpočetní clustery (Beowulf inspirace)
 - vysokorychlostní síť Myrinet, Infiniband (2007), OmniPath (2019)
 - RT ticketing system 2004, 654 lístků v roce 2006 (5000 v 2020)

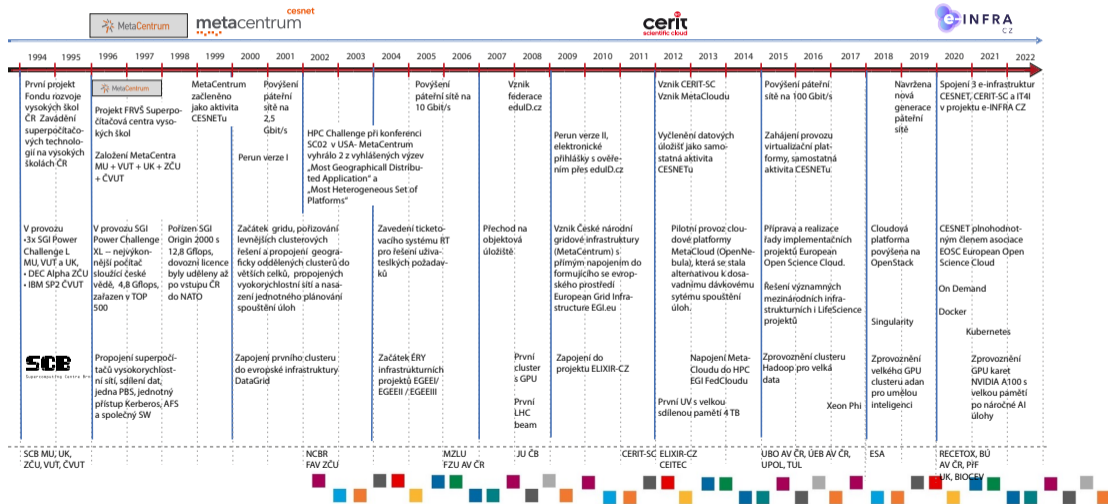


- evropské infrastrukturní projekty Datagrid, EGEE, EGI, EOSC
 - Egrid (1999), Datagrid (2000), EGI + EGEE (2003)
 - NGI v EGI (2009), EOSC mandated organization (2020)
- od počátku komunitní přístup (5 center na počátku)
 - NCBR a FAV ZČU (2002), VUTBR (2006), ČB (2008), ...
 - zdroje dalších výzkumných center a VI (NTIS ZČU, ELIXIR)
- 2007 virtualizace výpočetní infrastruktury -> výpočetní cloud
 - grid - Xen+Magrathea (2007), Singularity (2018)
 - cloud - OpenNebula (2012), EGI FedCloud (2013), OpenStack (2019), Kubernetes (2021)



- experimentální hardware, akcelerátory
 - Itanium II (2002), IBM Power (1994, 2004), Xeon Phi (2007)
 - GPU, první v roce 2008,
 - od roku 2013 NVIDIA CUDA – cluster doom s kartami K20
 - 2019 cluster adan s NVIDIA T4, 2020 cluster zia A100
- od roku 2020 jediná společná e-infrastruktura e-INFRA-CZ
 - CESNET, CERIT-SC (MU), IT4Innovations (VŠB)





Národní distribuované výpočetní prostředí

- výpočetní zdroje umístěné na CESNETu, univerzitách, AV ČR
- gridové, cloudové a map-reduce výpočty, velká data
- NGI v evropské e-infrastruktuře EGI, EOSC mandated org.
- komunitní přístup, centrální správa a AAI
- cílená podpora velkých projektů

Původní motivace sdílení zdrojů (HW) stále platí

- přenesení nárazové zátěže na volnější zdroje
- a využití jiných zdrojů při výpadku
- poskytnutí vlastních dočasně volných zdrojů
- zdroje pro nastartování projektu, ověření vhodnosti HW
- drahé komerční licence

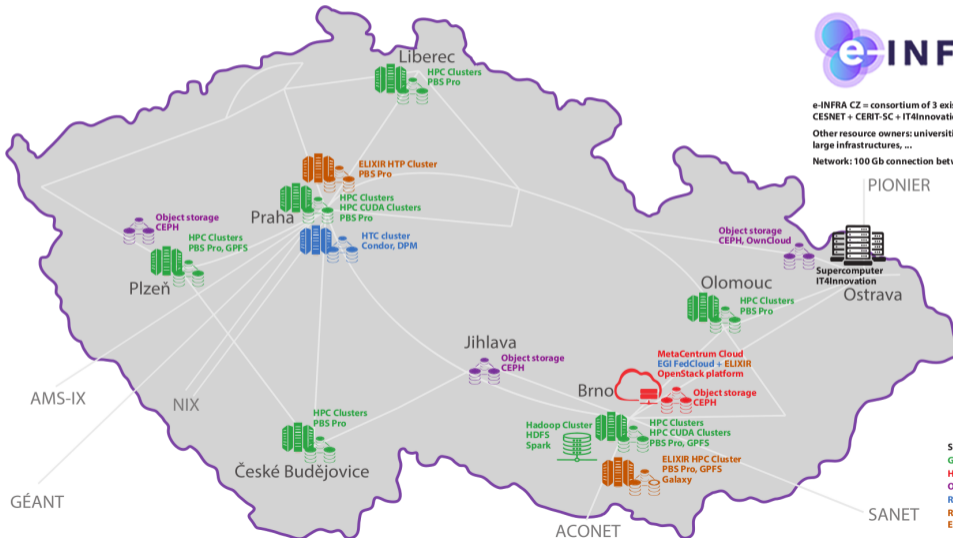




e-INFRA CZ = consortium of 3 existing e-infrastructures:
CESNET + CERIT-SC + IT4Innovation

Other resource owners: universities, research centers,
large infrastructures, ...

Network: 100 Gb connection between resource centers



Supercomputer
Grid HPC + HTC + Hadoop
HPC Cloud
Object DataStorage
Resources available in EGI
Resources dedicated to ELIXIR CZ

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

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



- grid, centrálně spravované HTC, HPC clustery
 - dávkové, dlouhé, paralelní výpočty (PBSPro)
 - včetně interaktivních úloh, grafické rozhraní
 - výpočty i v containerech (Docker, Singularity)
 - GPU karty, NVIDIA GPU Cloud software
 - distribuované clustery e-INFRA CZ a partnerů
- **MetaCentrum cloud - místo úloh virtuální stroj nebo container**
 - obrazy MetaCentra, projektové, uživatelské
 - cloudové výpočty a služby pro výpočty (OpenStack)
 - Kubernetes pro správu virtuálních prostředí
 - Terraform nebo Infrastructure Manager pro virtuální clustery
 - centrální instalace v Brně
- **MapReduce - Hadoop/Spark, zpracování dat**
 - dedikovaný cluster přejde do cloudu



- data semi-permanentní = určená pro zpracování v MetaCentru
 - zálohy a archivy nutno přesunout do služeb DU
- sdílený /home pro všechny clustery v jednom městě
 - NFS, větší instalace s pomocí GPFS
 - na dalších clusterech viditelné ve /storage/
 - zpravidla disková pole, SSD cache, experimenty s all-SSD
- dočasný scratch prostor v každém uzlu clusteru
 - zpravidla lokální SSD disk, experimentálně sdílený filesystem
- ve struktuře /storage/ dostupné i archivy DU
- objektová storage Ceph pro cloudové prostředí OpenStack
 - přes protokol S3 bude dostupné i v gridu, spolupráce s DU
- speciální případy – HDFS pro Hadoop/Spark



Řada aplikačního software - komerční

- matematický software (Matlab, Maple, Mathematica),
- vývojové nástroje (Intel, AMD, NVIDIA, Totalview, Alinea)
- technické simulace (Ansys Fluent, CFX)
- biochemie (Amber, Gaussian, Turbomole)
- v rámci e-INFRA CZ připravujeme větší koordinaci s IT4Innovations

Open-source (life-science, matematika)

- nárůst balíků pro strukturální biologii, bioinformatiku, sekvenování a analýzu DNA
- za poslední rok více jak 100 nových/změněných programových balíků



Spolupráce s uživatelskými skupinami jako dlouhodobá motivace vývoje nových služeb

- LHC, Auger, CTA, Belle

- původní motivace gridu, nadále aktivní v EGI

- ELIXIR (OpenScreen, CCT)

- spolupracující VI, zdroje začleněné do MetaCentra
- spolupráce při provozu služeb ELIXIRu
- life-science je největší konzument zdrojů

- ELI, BBMRI, CLARIN, ICOS

- zejména na úrovni EGI

- výzkumná centra CzechGlobe, CEITEC, (Recetox)

- dlouhodobí uživatelé, vazba přes CERIT-SC

- ESA – CollGS, Data Relay Hub



- aktuálně 38.000 CPU jader (x86_64)
 - HD uzly s menším počtem jader (2x4-32)
 - SMP servery - 32-128 jader, do 3 TB RAM
 - specializované servery s 6/10 TB RAM
 - 100 výpočetních uzlů s GP-GPU kartami
 - NVIDIA T4, 1080 Ti, **2080 Ti, A100**
- CESNET (14916, z toho nárůst 5632 v roce 2020), CERIT-SC (6.176, z toho nárůst 1640)
 - VI ELIXIR (5824, z toho nárůst 2368)
 - FZU (5190 pro LHC/EGI)
 - CEITEC, ZČU, JČU, MU, UK
- cca 7 (z toho nárůst +2) PB diskových prostor
 - NFS, GPFS, HDFS pro Hadoop/Spark
 - objektová storage Ceph, DPM úložiště (EGI)



1. 1. - 31. 12. 2020	Celkem MetaVO	meta-pbs	cerit-pbs	elixir-pbs	MetaCentrum Cloud	Hadoop	Celkem EGI	EGI HPC grid	EGI FedCloud
Počet úloh / spuštěných VM strojů	13 194 686	7 645 666	4 552 910	982 832	13 278	4550 úloh, 118400 MapReduce tasks	10 197 909	10 141 117	56 792
Propočítaný CPU čas (walltime) [CPU let]	16630	8 248	4 369	1007	3006 (včetně 443 FedCloud)	0,35 CPU let	9196	8753 (elapsed time*no. of processors; HT)	443
Počet uživatelů k 31. 12. 2020	2225 MetaVO			72 ELIXIR-CZ		141		9 VO	22 projektů
Počet CPU jader ke konci prosince 2020	23604 (pbs+cloud+Hadoop)	12444 z toho CESNET 7756	5040	1492	3880 (MUNI, CESNET, CERIT-SC, ELIXIR-CZ,...) + 2560 export do PBS	416	6230 + část z MC cloud	6230 5190 FZÚ 1040 CESNET	část kapacity MetaCentrum Cloud
Alokovaná kapacita od ledna 2021	32036 (+8432 leden)	16284 CPU jader (+3840 leden)	5040	3136 CPU jader (+1472 leden)	7000 CPU jader (3120 nových jader leden), + 2560 export do PBS	416	6230 + část z MC cloud	6230 5190 FZÚ 1040 CESNET	část kapacity MetaCentrum Cloud
Počet evidovaných GPU karet ke konci období	242	213	29	0			-		
Storage	obsazeno 4,5 PB z celkových 10 PB			obsazeno 500 TB z 2 PB, z toho 1,7 PB pro citlivá data	obsazeno 200 TB z celkových 1,4 PB	obsazeno 270 TB z 1 PB		uloženo 4 PB	započítáno v národním cloudu
Publikace s poděkováním MC/NGI z 2020	236 (WoS)							n/a	

	2010	2012	2014	2016	2017	2018	2019	2020	1-2021
Počet uživatelů MC	420	613	1112	1611	1908	2020	2185	2225	
Noví uživatelé (Meta)	139	312	605	742	732	713	762	774	
Počet úloh [milion úloh] Meta/EGI	0,6/ n/a	1,1/ n/a	3,9/ n/a	3,6/ 6	4,7/ 7	5/ 6,7	8,6/ 6,8	13,1/ 10	
Počet CPU vč. EGI	1560	6028	14164	17234	18666	21344	26602	29874	38266
CPU čas [CPU let] Meta/EGI	730/ n/a	2500/ n/a	6403/ n/a	9475/ 5963	10572/ 4622	11357/ 4074	13129/ 4531	16630/ 9160	

První rok nového OP VV - nárůst investic, 21M Kč

- cluster kirke na ZČU, 60 uzlů (3800 jader)
 - HD, náhrada clusteru minos
 - 2xAMD 7532, 512 GB RAM, 4TB SDD
 - FDR Infiniband, propojení se zbytkem na ZČU
- diskové pole na ZČU, 1.2 PB HDD, 120 TB SSD
- doplnění EGI clusteru na FZU (8 uzlů, 512 CPU jader)
- doplnění cloudových kapacit v Brně
 - 10 výpočetních uzlů, 10 uzlů 2xT4 GPU
 - celkem 1280 CPU jader
- celkem 5632 CPU jader



- HD cluster gita, 28 uzlů, 1000 jader
 - 2x AMD 7F52, 512GB RAM
 - včetně 14x2 NVIDIA 2080Ti
 - prostor na další karty
 - diskless uzly
 - probíhá integrace do cloudu
- all SSD pole, 480 TB
- cluster zia - 5 uzlů, 640 jader CPU
 - 2x AMD EPYC 7662, 1TB
 - 5x4 NVIDIA A100



Poslední rok OP VV ELIXIR

- diskové pole na Biocev
 - 432 TB, +46 TB SSD disků
 - Biocev, data RE+analýza genomů
- 2 stroje s 3TB RAM pro analýzu genomu
- 20 strojů s 2TB RAM, služby, UOCHB
 - 2xAMD 7532, 2x7.6TB NVMe
 - v Praze migrace ELIXIR služeb na tyto stroje
- 6 uzlů s GPU 2xT4 do cloudu Brno
- 8 uzlů do cloudu Brno
- celkem 2368 CPU jader



- cluster doom s K20 bude odstaven
- cluster adan, hlavní zdroj pro HTC (122xT4)
- CERIT cluster glados s 20xNVIDIA 1080Ti v cloudu
- CERIT zprovoznil cluster gita s 28x 2080 Ti
- CERIT zprovoznil cluster zia
 - 5 uzlů, v každém 4xNVIDIA A100
- další T4 uzly do cloudu (EGI, ELIXIR)
 - 10 + 6 uzlů s 2x NVIDIA T4
- clustery NCBR MU(zubat) a KKY ZCU (konos)
 - 8x2 K20, 8x4 1080Ti
 - další plány v Recetoxu
 - server cha.natur.cuni.cz, 8x2080 Ti



- podpora grafických rozhraní pro interaktivní práci
 - JupyterHub, OnDemand (CERIT-SC)
- upgrade OpenStack cloudu, Ceph úložiště pro cloud
 - podpora Kubernetes (CERIT-SC), Bio/Ligth Portal
- Debian 10 na výpočetních uzlech, přechod na novou PBSPro
 - upgrade GPFS serverů, nasazení CVMFS filesystemu
- obrazy z NVIDIA GPU Cloud v MetaCentru - Singularity
- procesory AMD EPYC - architektury Intel i AMD v modulech
 - Spack pro kompilaci balíčků, plánujeme nové modules



Místo projektu EOSC-Hub začaly projekty EOSC-Future a EGI-ACE

- rozvoj základu/jádra EOSC a rozvoj EGI služeb zde zapojených

Nový projekt C-SCALE (oblast zpracování dat ESA)

- služby na zpracování Copernicus dat v EOSC

Project OCRE koordinovaný organizací GEANT

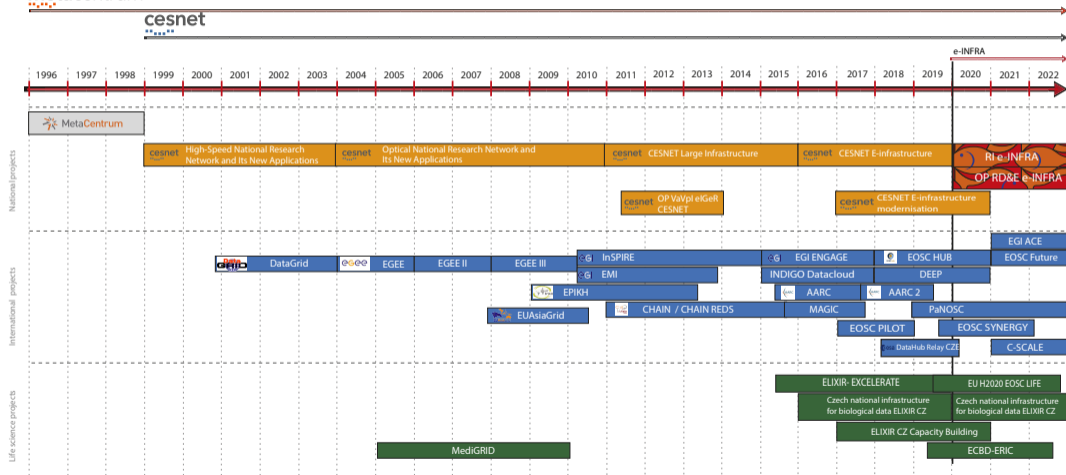
- rámcové smlouvy na cloudové služby IaaS, PaaS, SaaS
- pan-evropský tender podle EU direktivy č. 2014/24/EU
 - vysoutěženo 12 dodavatelů pro ČR
 - smlouvy a fakturace přes CESNET (underwriter)
 - predikovatelné čerpání, GDPR compliance, definované smluvní podmínky

- <https://www.cesnet.cz/sluzby/geant-iaas/>



cesnet
metacentrum

cesnet



- pokračujeme v průběžné obnově clusterů
 - životnost je 4 roky, zvýšený výkon podle Moorova zákona
 - CESNET - od roku 2020 zdvojnásobení investic
- hardware v 2021 - obnova clusterů v Brně
 - CESNET: SMP uzly, cca 20 strojů s 1–2TB RAM
 - uzly s GPU (90+ karet NVIDIA A40), 1 PB diskové pole
 - CERIT: HD cluster v Brně
- hardware v 2022 obnova HW na ZČU a JČU
 - CESNET: HD uzly, cca 5000 CPU jader, 1 PB diskové pole
 - CERIT: obnova velkého SMP stroje
- 2021: zapojen cluster Přírodovědecké fakulty UK
- první kroky ve zlepšení kompatibility MetaCentra a IT4I
 - společné účty, sdílení komerčních licencí, přístup k /software



- změny v oblasti uložení velkých dat
 - i s využitím GPFS limit růstu kapacity filesystemu pro /home
 - větší kapacity přes objektovou storage (vazba na cloud a DU)
 - nutnost začít podporovat S3, migrace dat mezi filesystemy a objekty, mezi městy, do DU
 - úvahy o sdíleném scratch (Lustre) v Brně
- další rozvoj služeb služeb v OpenStack cloudu
 - směr HPC cloudu a Sensitive cloudu
 - Kubernetes pro škálování aplikací v cloudu
 - data v objektové storage - S3
- škálující aplikace se musí začít připravovat na tyto změny
 - výzva ke spolupráci pro větší skupiny v ČR
 - včetně úvah o OCRE a komerčních cloudech



- Jak sejmout MetaCentrum v 10 krocích (Jiří Vorel, CESNET)
- Podporovaná GUI OnDemand, VNC (A. Melounová, CESNET)
- Podpora kontejnerových řešení (Honza Hoidekr, CESNET)
- Podpora AI, Nvidia cloud a využití s pomocí GPU (Honza Hoidekr, CESNET)
- Jupyter Notebooks (Aleš Křenek, CERIT-SC)
- Bio-portal: Zjednodušené rozhraní pro nezkušené uživatele (Zdeněk Šustr, CESNET)

