

METACENTRUM

zpracování (velkých) dat

Miroslav Ruda

září 2026

Osnova přednášky

- Představení MetaCentra pro nové uživatele
- Vývoj, novinky, kam směřujeme

- Představení MetaCentra pro nové uživatele
- Vývoj, novinky, kam směřujeme
 - co je to e-INFRA CZ a EOSC CZ, proč v MetaCentru?
 - co se děje na evropské úrovni
 - co nového s LLM, AI, co poskytujeme, co plánujeme

- Představení MetaCentra pro nové uživatele
- Vývoj, novinky, kam směřujeme
 - co je to e-INFRA CZ a EOSC CZ, proč v MetaCentru?
 - co se děje na evropské úrovni
 - co nového s LLM, AI, co poskytujeme, co plánujeme
- Technické podrobnosti zítra, na semináři MetaCentra

Národní distribuované výpočetní prostředí, koordinace CESNET

- zdroje CESNET i CERIT-SC (55.000 CPU jader, 500 GPU)
- výpočetní zdroje umístěné na CESNETu, univerzitách, AV ČR
 - původní motivace sdílení zdrojů (HW) stále platí
 - poskytování dočasně volných zdrojů pro vzdálené uživatele
 - využití vzdálených zdrojů v případě naléhavého/velkého zatížení
 - použití jiných zdrojů během výpadku
 - zdroje pro spuštění projektu, kontrola vhodnosti HW
 - myšlenka funguje i pro cloud, drahé komerční licence
- komunitní přístup, centrální správa a AAI
- grid, IaaS cloud a PaaS cloud, LLM služby
- virtualizační platforma pro vysoce dostupné služby

- distribuovaný model užitečný pro různé případy použití
 - integrace zdrojů vlastněných jinými VI (ELIXIR)
 - zpracování velkých dat - není potřeba vzdáleného přenosu
 - architektura kompatibilní s datovými úložišti v EOSC CZ
- NGI v evropské e-infrastruktuře EGI, mandated org. v EOSC-A
- cílená podpora velkých projektů (VI, ESFRI)
- vývoj nových služeb/nástrojů podle potřeb uživatelů/projektů
 - OnDemand, Jupyter, Galaxy, Kubernetes
 - podpora zpracování citlivých dat
- služby volně dostupné pro akademické využití a spolupracující partnery ze zahraničí
 - "placení" publikacemi s poděkováním, určují prioritu uživatelů

Spolupráce s vědeckými skupinami I

Potřeby uživatelů jako dlouhodobá motivace vývoje služeb

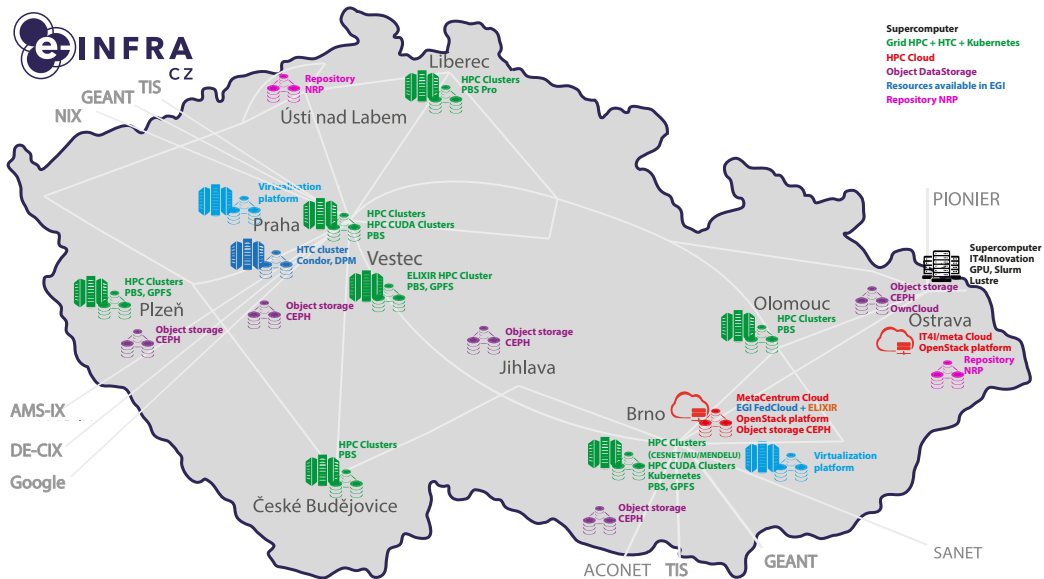
- HPC začátky, MPI úlohy - výpočetní chemie, matematika, simulace
- HEP , astrofyzika - LHC, Auger, CTA, Belle
 - původní motivace gridu, nadále aktivní v EGI
- Life Science - zejména ELIXIR
 - spolupracující VI, zdroje začleněné do MetaCentra
 - spolupráce při provozu služeb ELIXIRu, Galaxy, LS AAI
 - life-science je největší konzument zdrojů
- kratší interaktivní výpočty - ekonomie, ENVRI projekty
 - Open OnDemand GUI - Matlab, R Studio
 - Jupyter Notebooky
 - Kubernetes a Rancher GUI

Spolupráce s vědeckými skupinami II

- dostupnost software, reprodukovatelnost
 - modules pro HPC prostředí
 - možnosti využití virtuálních strojů a containerů
- ESA – CollGS, Data Relay Hub
 - řada skupin a projektů využívajících Copernicus data
- zpracování dat jako další motivace růstu
 - velká data - možnost využití S3 úložišť CESNETu/ OneData
 - paralelní zpracování - sdílené scratch filesystemy
 - citlivá data (BBMRI, medicína, sociální vědy) - SensitiveCloud
- EOSC CZ repozitáře
 - integrace s HPC cmdline, Galaxy, Notebooky
 - mezinárodní projekty (LS, ENVRI), národní EOSC uzel

Spolupráce s vědeckými skupinami III

- podpora AI, LLM modely
 - desítky GPU dovolujících provoz vlastních AI modelů
 - HPC grid, Kubernetes
 - centrálně spravované LLM modely, dostupné přes chat a OpenAI API
- centrální uživatelská podpora
 - zpětná vazba přes roční zprávy a dotazníky na spokojenost
 - tento rok zahájený program ambasadorů



e-INFRA CZ, EOSC CZ a mezinárodní spolupráce



- MetaCentrum je součást infrastruktury CESNETu od roku 1998
 - původně projekt MU, ZČU, UK, VUT
- od roku 2010 součást evropské gridové infrastruktury EGI
- oddělené VI CESNET a VI CERIT-SC od 2012
 - MetaCentrum stále integruje zdroje CERIT-SC
 - společný vývoj prostředí, včetně cloudových služeb
 - integrace zdrojů dalších organizací, včetně VI ELIXIR CZ
- od roku 2020 jediná společná e-infrastruktura e-INFRA CZ
 - CESNET, CERIT-SC (MU), IT4Innovations (VŠB)
 - sbližování prostředí MetaCentra a IT4Innovations
 - další služby MetaCentra ve společném vývoji CESNET a CERIT-SC



e-INFRA CZ, EOSC CZ a mezinárodní spolupráce



- navrženo pokračování VI e-INFRA CZ, 2027-2032
 - z pohledu MetaCentra zejména posílení infereční AI vrstvy
- e-INFRA CZ koordinuje i vývoj EOSC infrastruktury v ČR
 - ekvivalent webu pro sdílení vědeckých dat
 - základní služby EOSCu (IP EOSC CZ 2023)
 - Národní repozitářová platforma (NRP 2024, OS I.)
 - vědecké tématické repozitáře, FAIR data
- od roku 2023 se podílíme na budování EOSC federace v EU
 - CESNET jako koordinátor národního uzlu v evropské federaci
 - EOSC EU Node, CESNET zodpovědný na Jupyter Notebooky
 - národní uzel je součástí druhé vlny přijatých kandidátských uzlů (2025)
 - v roce 2026 aspirujeme na produkční uzel (10 uzlů v EU)



Výpočetní prostředí MetaCentra I

Grid, centrálně spravované HTC a HPC clustery

- dávkové, dlouhé (dny/týdny) úlohy, paralelní výpočty
- včetně interaktivních úloh, GUI
 - Galaxy, Jupyter, OnDemand
- trvalé úložiště (GPFS+NFS) a lokální/sdílené úložiště scratch
- výpočty také v kontejnerech (Singularity)
 - HPC přístup, podpora pro non-root obrazy Dockerů
 - Software NVIDIA GPU Cloud (PyTorch, TensorFlow)
- distribuované clustery členů a partnerů e-INFRA CZ
 - vývoj a výzkum v oblasti plánování
- podmnožina zdrojů dostupných také v infrastruktuře EGI

Výpočetní prostředí MetaCentra II

MetaCentrum cloud – virtuální stroj místo úloh

- obrazy poskytované MetaCentrem, EGI, projekty, uživateli
- cloudové výpočty a služby spojené s výpočty (OpenStack)
 - ale také školení, výuka, bezpečnostní polygon KYPO
- Terraform nebo EGI IM pro virtuální clustery/K8s
- centrální instalace v Brně, společné úsilí CESNET a CERIT-SC
 - od roku 2022 vývoj nové distribuce OpenStack - Bescar Cloud
 - od roku 2023 druhá instalace v IT4I
- přístupné i prostřednictvím EGI FedCloud, EOSC Marketplace
- v tomto prostředí provozujeme i služby EGI Jupyter a Binder
 - včetně EOSC EU uzlu a projektů ENVRI-Hub a EOSC Beyond

Containerový cloud, PaaS řešení

- containery místo úloh/virtuálních strojů
 - metoda balení softwarových balíčků, zapouzdření služeb
 - reprodukovatelnost, provenance
 - jak v HPC (Singularity) tak v cloudu (Docker, Kubernetes)
- Kubernetes pro mikroslužby, správa virtuálních prostředí
- spravovaná služba, non-root kontejnery, GUI Rancher
- silná podpora interaktivních požadavků a požadavků workflows
 - SaaS přístup pro Matlab, RStudio, NextFlow
- vývoj v oblasti converged computing
 - integrace dávkového systému a Kubernetes

Výpočetní prostředí MetaCentra IV

- Sensitive cloud – práce na případech použití souvisejících se zpracováním citlivých dat
 - koordinuje CERIT-SC
 - v současné době Kubernetes, později také OpenStack
 - od roku 2024 s certifikací ISO 27.000
- služba integrace zdrojů dalších organizací do MetaCentra
 - centrální správa, prioritní přístup vlastníků
 - dlouhodobě pro HPC clustery
 - od roku 2024 i pro OpenStack cloud
- LLM jako služba
 - koordinuje CERIT-SC
 - přístup k lokálně instalovaným LLM přes WebUI chat nebo OpenAI API
 - <https://docs.cerit.io/en/docs/ai-as-a-service/introduction>

MetaCentrum v číslech, duben 2026

55.000 CPU jader, 20 PB, 460 GPU, 4.500 uživatelů, 350+ WoS

- HD uzly - 32-128 CPU jader (x86_64), 256-1024 GB RAM
- SMP servery - 2-3 TB RAM
 - specializované servery s 6/10 TB RAM
- NVIDIA GP-GPU - 160 uzlů, 500 karet
 - NVIDIA DGX s H100/B200/B300
 - clustery s NVIDIA A100, A40, H100, H200
- CPU jádra: grid - 42.500+6.000, IaaS 10.400, K8s 3.000, SC 1.000
- CPU roky: grid - 26.000+13.000, IaaS 10.000, Kubernetes 1.100
- clustery vlastněné rostoucím počtem partnerů
 - CESNET 25.000 a CERIT-SC 9.800
 - VI ELIXIR 8.300
 - FZU 3.200 a 6.000 pro LHC/EGI
 - ZČU, MU, UK, ČVUT, TUL, AV ČR

Využití MetaCentra za 10 let

	2012	2016	2018	2020	2021	2022	2023	2024	2025
Počet uživatelů MC	613	1611	2020	2225	2606	2710	3055	3490	4621
Noví uživatelé (Meta)	312	742	713	774	792	767	850	1232	1960
Počet úloh [miliony] Meta/EGI	1,1/ n/a	3,6/ 6	5/ 6,7	13,1/ 10	12,1/ 9,3	11,1/ 14,2	11,7/ 5,4	15/4,9	14,4/ 11
CPU čas [CPU let] Meta/EGI	2500/ n/a	9475/ 5963	11357 /4074	16630/ 9160	22647/ 9581	27547/ 9218	31858/ 14770 HEPScore	37552/ 217886 HEPScore	39609/ 5901
Počet CPU jader vč. EGI	6028	17234	21344	29874	34084	44088	47748	53004	53778/ 6960

MetaCentrum leden – prosinec 2025	Celkem MetaCentrum	PBS		MetaCentrum Cloud	Kubernetes	Sensitive Cloud	EGI		
		PBS celkem	z toho ELIXIR PBS				Celkem EGI	EGI HPC grid	EGI FedCloud
Počet úloh / spuštěných VM strojů / kontejnerů	14,4 mil.	7,6 mil.	135 211	80 684	6,42 mil	37 833	11 mil.	11 mil.	36
Propočítaný CPU čas (walltime) [CPU let]	39609	25 392	1 214	10 586	1760	657	5901	5860	41
Propočítaný GPU čas (walltime) [GPU let]	420	283	0	59	43	35	n/a	n/a	n/a
Počet uživatelů		4621 s aktivním účtem v MetaCentru						6 Vos	12 projektů
		2357 (alespoň 1 úloha)	111	2175 projektů	772	55 projektů			
Počet CPU jader konec 2025	53778	39988	5984	9670	3008	1112	6960 + část z FedCloud	6960 bez HT 6000 FZÚ 960 CESNET	část kapacity MetaCentrum Cloud
Počet evidovaných GPU karet ke konci roku 2025	503	332	0	84	47	40		0	část kapacity MetaCentrum Cloud
Publikace s poděkováním e-INFRA CZ v roce 2025		502 (stav v lednu, ještě přibývá, WoS e-INFRA CZ)					n/a		

<https://metavo.metacentrum.cz/cs/state/stats/2025/>

Děkuji za pozornost

`support@metacentrum.cz`

`https://www.metacentrum.cz`

`https://docs.metacentrum.cz/`

