

Aktuality a plány virtuální organizace MetaVO

Tomáš Rebok

MetaCentrum, CESNET z.s.p.o.

(rebok@ics.muni.cz)

MetaCentrum VO (Meta VO)

- přístupná zaměstnancům a **studentům** VŠ/univerzit, AV ČR, výzkumným ústavům, atp.
 - komerční subjekty **pouze pro veřejný výzkum**
- nabízí:
 - výpočetní zdroje
 - úložné kapacity
 - aplikační programy
- po **registraci** k dispozici **zcela zdarma**
 - „placení“ formou publikací s poděkováním
- součást budované **e-infrastruktury CESNET**
 - datová úložiště, kolaborativní prostředí, ...

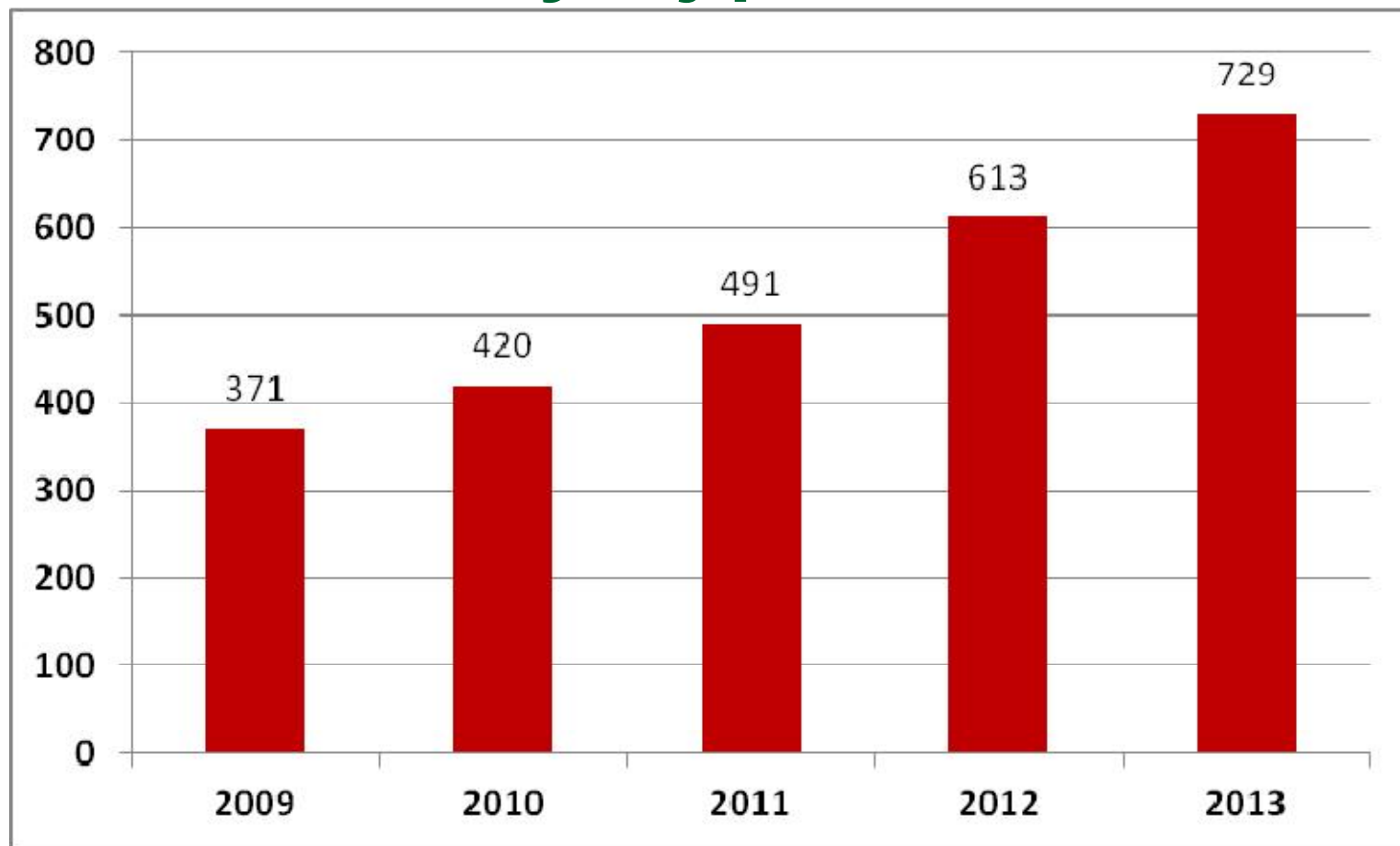
<http://metavo.metacentrum.cz>



MetaVO – základní charakteristika

- po registraci zdroje dostupné **bez administrativní zátěže**
 - → ~ okamžitě (dle aktuálního vytížení)
 - žádné žádosti o zdroje
- **každoroční** prodlužování uživatelských účtů
 - periodická informace o trvající akademické příslušnosti uživatelů
 - využití infrastruktury **eduID.cz** pro minimalizaci zátěže uživatele
 - oznamování publikací s poděkováním MetaCentru/CERIT-SC

MetaVO – vývoj počtu uživatelů



Meta VO – dostupné výpočetní zdroje I.

- výpočetní zdroje: **9028 jader** (x86_64)
 - zdroje CESNETu + zapojených organizací/institucí
 - klasické HD uzly (2x4-8 jader) i SMP stroje (32-80 jader, 288 jader)
 - paměť až 6 TB na uzel (+ 2x 1 TB na uzel)
 - Infiniband pro nízkolatenční komunikaci (MPI)
- příklady dostupného HW:
 - 1x **288 jader, 6 TB RAM** (SGI UV 2000, uzel *ungu*, CERIT-SC)
 - **20 x 80 jader**, 512 GB per node (cluster *zewura*, CERIT-SC)
 - 2 uzly s **1 TB RAM** - uzly *ramdal* (32 jader, CESNET) a *haldir* (64 jader, JČU)
 - až **2176 jader** (clustery *zewura+zegox*, CERIT-SC) přímo **propojených infinibandem**
 - 30 uzlů s **2x nVidia Tesla K20 5GB** (aka Kepler) per node (cluster *doom*, CESNET)
 - ...

Meta VO – dostupné výpočetní zdroje I.

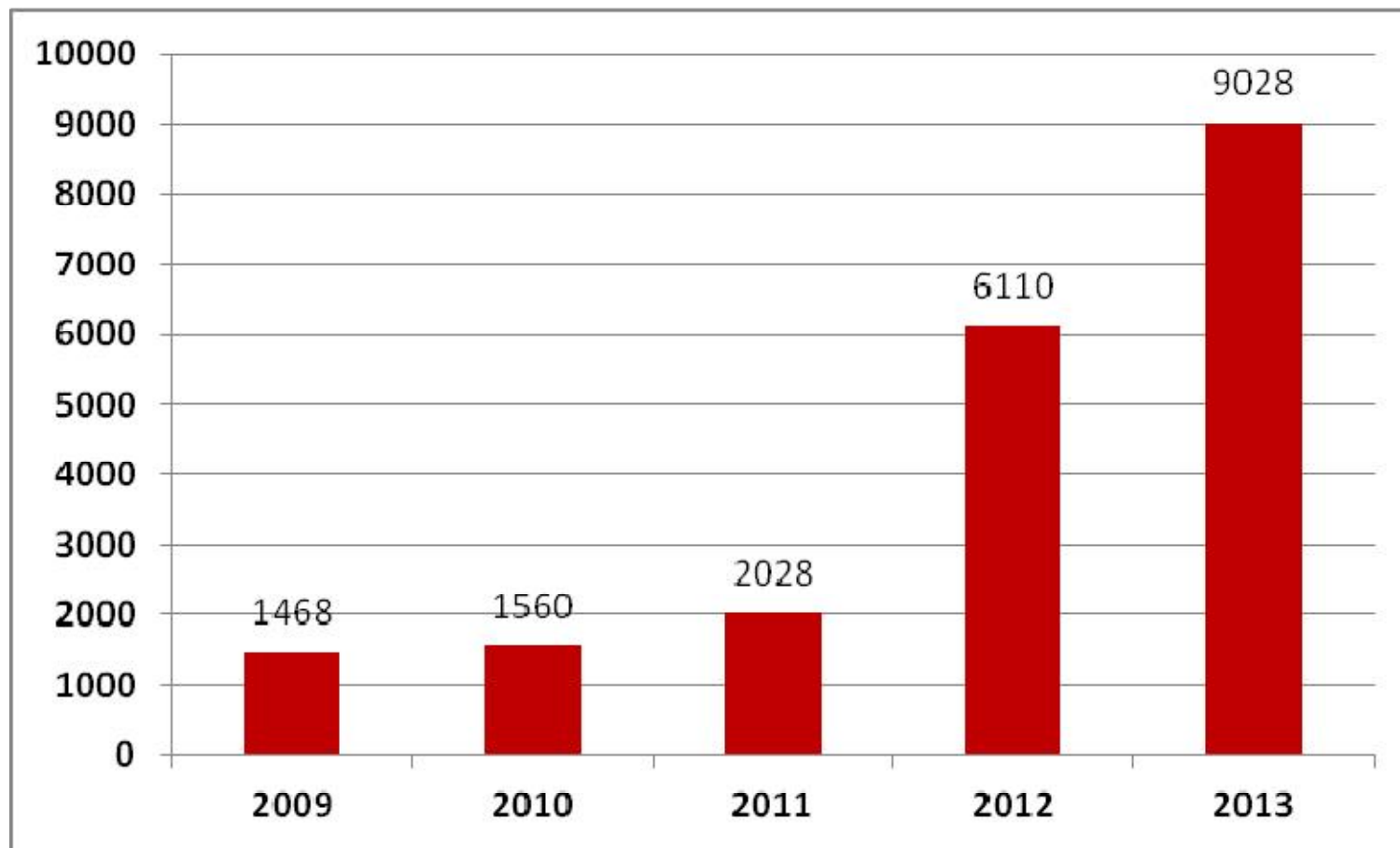
- výpočetní zdroje: **9028 jader** (x86_64)
 - zdroje CESNETu + zapojených organizací/institucí
 - klasické HD uzly (2x4-8 jader) i SMP stroje (32-80 jader, 288 jader)
 - paměť až 6 TB na uzel (+ 2x 1 TB na uzel)
 - Infiniband pro nízkolatenční komunikaci (MPI)

Nejvýznamnější novinky (posledních dnů)

- instalace uzlu **SGI UV 2000** (CERIT-SC, 288 Intel Xeon x86-64 jader, 6 TB sdílené RAM)
- instalace nového clusteru **doom** (30x 16 x86-64 jader, 64 GB RAM, CESNET, fyzické umístění Ostrava)
- ...

- ...

Meta VO – dostupné výpočetní zdroje II.



Meta VO – dostupné výpočetní zdroje III.

Možnosti integrace vlastních zdrojů:

- ***plná integrace*** – výkonný cluster integrovaný do infrastruktury MetaCentra/MetaVO
 - možnost prioritizace vlastníka
 - správa v režii MetaCentra
- ***částečná integrace*** – podpora skupin s vlastním HW (uzel, cluster) nevhodným pro plnou integraci
 - dodání obrazu s předpřipravenými službami a dostupnou SW výbavou
 - možnost využití jako **vlastní výpočetní uzel** a/nebo **vlastní čelní uzel**
 - správa **plně v režii vlastníka**
 - ze strany MetaCentra pouze základní podpora (konzultace, bezpečnostní updaty,...)

Meta VO – dostupné úložné zdroje

- **cca 1,1 PB (1151 TB)** pro pracovní data
 - úložiště 3x v Brně, 1x v Plzni, 1x v ČB, 1x v Praze, 1x v Jihlavě, 1x v Ostravě
 - uživatelská kvóta 1-3 TB na každém z úložišť
- **cca 3,8 PB (plán cca 16 PB)** pro archivní data
 - (integrace DÚ Cesnet)
 - uživatelská kvóta 5 TB
 - `/storage/plzen2-archive/home/$USER/VO_metacentrum-tape_tape`
 - **prosíme, využívejte pro svá dlouhodobě nevyužívaná data**

Meta VO – dostupné aplikační programy

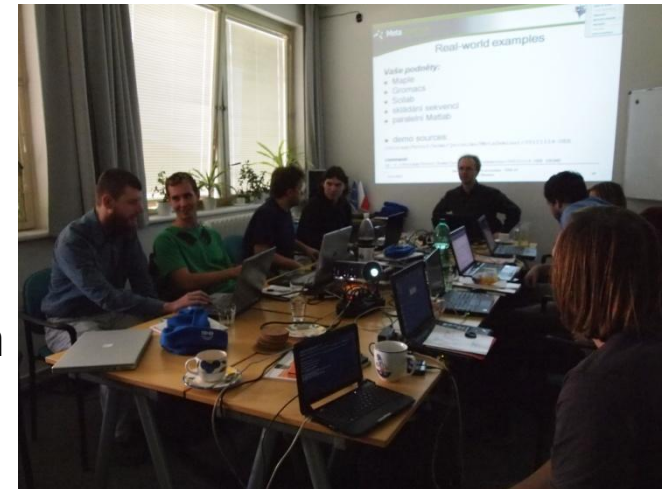
Nejvýznamnější novinky

- **Ansys CFD (CFX + Fluent)** – 25 spuštění, místně neomezená
- **Ansys Mechanical** – 5 spuštění, místně neomezená
- **Ansys HPC** – 60 licencí (dalších jader), místně neomezená
- **Matlab** - +100 licencí (450 celkem)
- **Wolfram Mathematica** – 10 licencí
- **Gaussian Linda**
- **Turbomole, Molpro**
- **CLC Genomics Workbench** (2 spuštění), **Geneious** (2 spuštění)
- **TotalView, Allinea DDT, PGI + Intel kompilátory**
- **Mascot Server**
- ...

Novinky – co jsme zavedli

Informovanost uživatelů I.

- **ročenka MetaCentra a CERIT-SC (2011-2012)**
 - aktuality infrastruktury
 - vlastní výzkum v oblasti služeb infrastruktury
 - výzkum realizovaný ve spolupráci s partnery
 - reporty významnějších uživatelských skupin (5+ publikací s poděkováním)
 - seznam všech publikací s poděkováním
- **výjezdní hands-on semináře**
 - prakticky orientované školení (3-4 hodiny)
 - v úzkém okruhu zájemců (do 10 osob)
 - zaškolení základních způsobů práce s gridem
 - orientace na školenou uživatelskou skupinu (praktické příklady)



Informovanost uživatelů II.

- **drobné provozní změny v sekci novinek**
 - nerozesílány emailem, možnost odběru RSS kanálem
- **pravidelné (měsíční) novinky (provozní + SW)**
 - (důležitější provozní novinky)
 - přehled instalovaného SW (komerční i volně dostupný)
- **přehlednější informace o využitém diskovém prostoru a kvótách**
 - úvodní přihlašovací obrazovka + MetaVO portál
- **postupné vylepšování dokumentace**

Výzkum a vývoj (CERIT-SC)

- **kolaborativní podpora výzkumu ve spolupráci s partnery**
 - = uživatelskými skupinami i jednotlivci
 - **cíl:** aplikace špičkové ICT za účelem překonání dosavadních limitů výzkumu
 - více viz přednáška prof. Matysky
- **vlastní plánovač úloh (`wagap.cerit-sc.cz`)**
 - pro ladění optimálního návrhu/rozložení front a nových plánovacích metod
 - úlohy zadávejte do předvolené fronty (nespecifikujte „-q“) a důsledně specifikujte odhad doby běhu (“-l walltime=HH:MM:SS”)
- **experimentální testbed s rozvrhovým plánovačem**
 - úprava plánovacího systému Torque na rozvrhový přístup (vlastní výzkum)
 - větší přehled o úlohách k naplánování
 - => **možnosti optimalizace plánu, možnost predikce času a lokace spuštění**
 - více viz
[https://wiki.metacentrum.cz/wiki/Uživatel:Vchlum/Instance_Torque_s_planovačem,_který_využívá_rozvrh](https://wiki.metacentrum.cz/wiki/Uzivatel:Vchlum/Instance_Torque_s_planovačem,_který_využívá_rozvrh)

Provozní změny – úložné prostory I.

- **různé typy scratchů**
 - klasické + SSD
 - sdílené (clustery *hildor*, *mandos*, *ramdal* a *doom*)
 - nově podpora množstevních jednotek (KB, MB, GB, ...)
 - `-l scratch=SIZE:TYPE` (např. `-l scratch=50gb:shared`)
- **uzavření dat jednotlivých úloh do vlastního adresáře**
 - příprava na zavedení automatizovaného odklizení scratchů
 - přesun dat na „hřbitůvek“
 - na přidělený prostor odkazuje **proměnná \$SCRATCHDIR**
- **disková kvóta pro kořenový adresář (1 GB / uživatel)**
 - ochrana strojů před pádem kvůli vyčerpání systémového svazku
 - může znamenat pád série úloh (chybová hláška zatím nepříliš srozumitelná)
 - pozor na `/tmp`, `stderr/stdout` úloh (`/var/spool/torque/spool`)

Provozní změny – úložné prostory II.

- **centralizace svazků pro pracovní data (/storage)**
 - => /storage/CITY/\$USER/home jako \$HOME adresář
 - cíl: urychlit informaci o aktuální lokaci uživatelům (pwd)
 - => v PBS skriptech **možno využívat \$PBS_O_WORKDIR**
 - jednoduchý přístup do adresáře, ze kterého byl skript zadán
- **podpora „mountování“ /storage svazků z uživatelských serverů/stanic**
 - jednoduchý a rychlý přístup ke svým datům
 - popis – viz dokumentace

Provozní změny – různé

- **ochrana úloh před nedostatkem rezervovaných zdrojů**
 - „zabíjení“ úloh/procesů využívajících větší než ohlášené množství zdrojů
 - zasílané oznamy o násilně ukončených úlohách
 - mechanismy řešení – viz přednáška M. Kuby
- **zavedení nového systému pro správu uživatelů (Perun)**
 - správa uživatelských účtů (přihlášky, prodlužování, ...)
 - podpora uživatelských skupin (usnadnění spolupráce mezi členy skupiny)
 - evidence publikací, přihlášky k licencím, atp.
 - (systémová pomoc s distribucí účtů, ...)
- **průběžné vylepšování uživatelského portálu**
- **zavádění nové verze systému Debian (Debian 7)**
- ...

Co dalšího plánujeme?
(o zavedení budete informováni aktualitou)

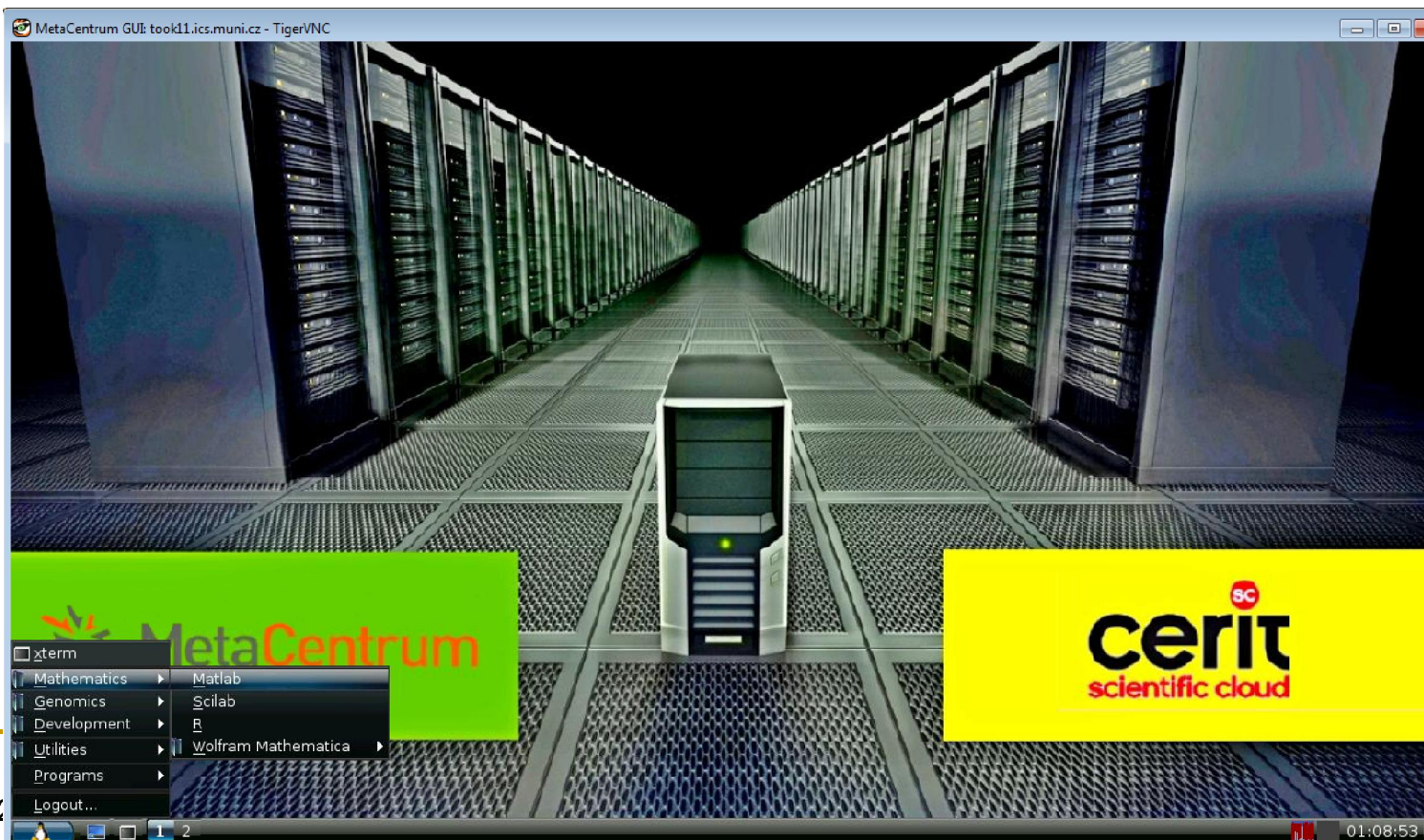
Interaktivní práce s grafickými aplikacemi I.

- narůstající množství aplikací s nezbytným grafickým prostředím
- dosavadní způsob práce (export/tunelování displeje) nevhodný pro interaktivní práci
- **připravované řešení** (dostupnost do konce roku 2013):
 - **grafické prostředí založené na VNC serverech**
 - nikoli náhrada desktopu, výhradně pro jednorázovou práci s grafickými aplikacemi
 - podpora **zabezpečení SSL** (jednodušší ustavení spojení) i **SSH tunelů**
 - podpora **připojení skrze webový prohlížeč**
 - **více viz** https://wiki.metacentrum.cz/wiki/Vzd%E1len%FD_desktop

Interaktivní práce s grafickými aplikacemi II.

```
$ module add gui
```

```
$ gui start [-w]
```



Paralelní/distribuované výpočty v Matlabu I.

- zadání **paralelního výpočtu v Matlabu** vyžaduje specifikaci počtu dostupných procesorů ve zdrojovém souboru
- zadání **distribuovaného výpočtu** výrazně složitější (nastavení integrace s PBS, ...)

- **připravujeme:** (cca leden 2014)

- podporu paralelních/distribuovaných výpočtů **skrze implementované nativní funkce Matlabu**

- ***paralelní:***

```
MetaParPool('open');
```

```
...
```

```
MetaParPool('close');
```

Paralelní/distribuované výpočty v Matlabu II.

- **připravujeme: (cont'd)**

- ***distribuovaný výpočet:***

```
jm = MetaGridPool('open');
```

```
...
```

```
MetaGridPool('close');
```

(spouštěno uvnitř úlohy, tj. v rámci aktuálního sandboxu)

Nová verze aplikačních modulů

- stávající verze sice funkční, ale zastaralá
- připravujeme **nasazení nové verze modulů** a současné **zpřehlednění struktury modulů**
 - snazší vyhledávání modulů
 - „*tab-completion*“
 - informační funkce v případě chyb (chybějící akceptace licenčních podmínek, nesprávná lokace spuštění, atp.)
 - atd.
- současně přejdeme na standardní formát zápisu jména/verze modulů („*modul/verze*“, např. „*gcc/3.0.2*“)
 - zachováme kompatibilitu s dosud používaným zápisem („*gcc-3.0.2*“)

Další chystané novinky

- „inteligentní“ vlastnost **infiniband**
 - přidělování uzlů **skutečně vzájemně propojených** Infinibandem (nutno brát v potaz jejich fyzickou lokaci)
- nová metrika pro spravedlivější účtování využitých zdrojů
 - nutno brát v úvahu nejen samotné CPU, ale **CPU + RAM + další zdroje** (např. GPU karty)
 - úprava Torque – zohlednění plného využití zdroje i při minimální CPU zátěži
- instalace nových clusterů, nových SWs, servisní služby, atp.
- ...

Závěrem ...

- infrastruktura se neustále vyvíjí
 - způsoby (efektivního) použití se vyvíjí též
 - dostupnost nejen technických služeb (podpora výzkumu)
- snaha o udržení informovanosti uživatelů
 - školicí semináře
 - pravidelné oznamy
 - stránky dokumentace
 - ročenka MetaCentra & CERIT-SC
- o Vaši zpětnou vazbu Vás prosíme formou dotazníku
 - za jeho vyplnění budete odměněni malým dárkem

Děkuji Vám za pozornost!

`rebok@ics.muni.cz`



metavo.metacentrum.cz

www.cerit-sc.cz