

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 **e-infrastruktury CESNET**
 - datová úložiště, kolaborativní prostředí, ...

<http://metavo.metacentrum.cz>

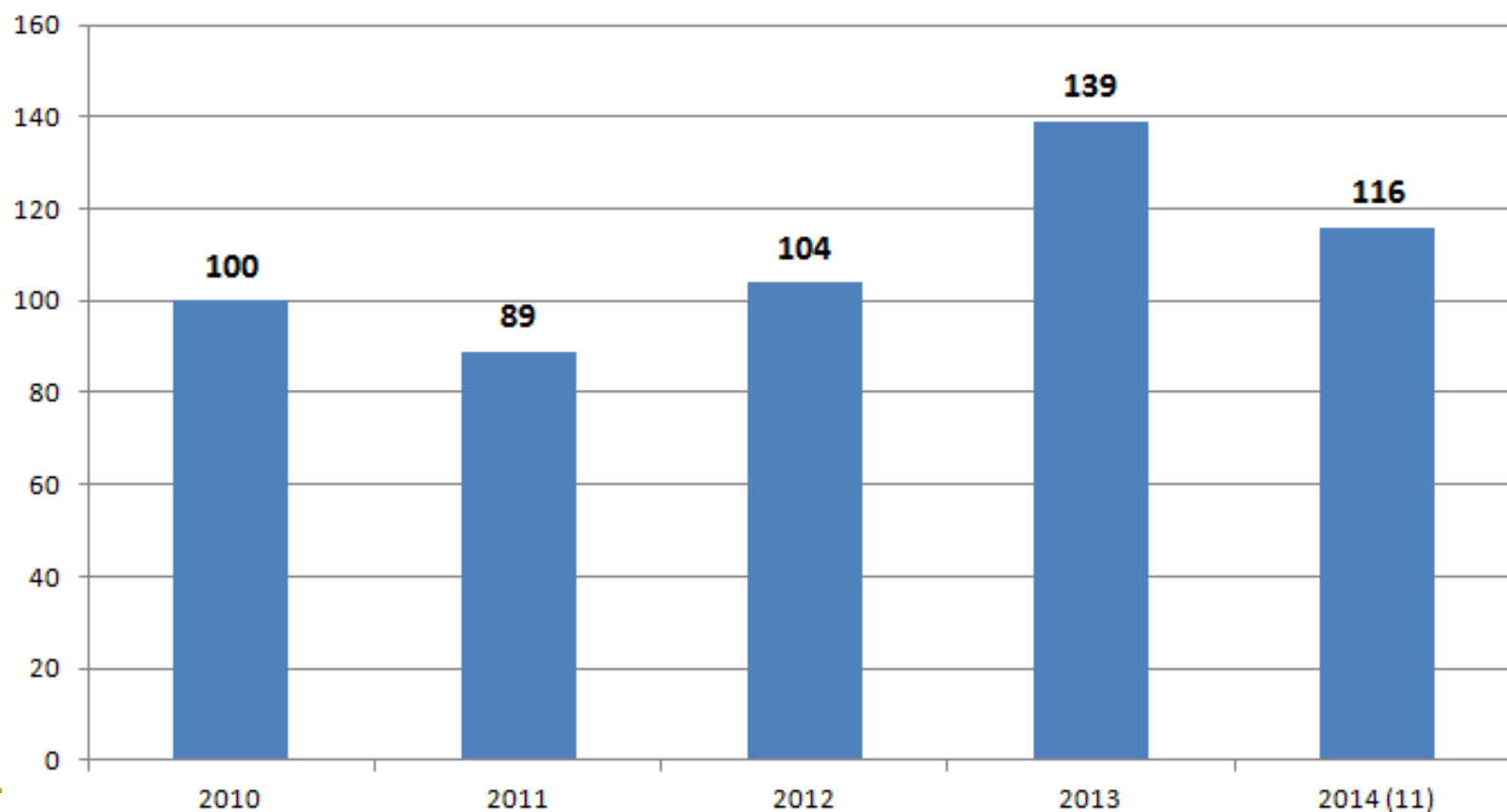


MetaVO – publikace s poděkováním I.

- **forma úplaty** za poskytnutí zdrojů/služeb
- doklad pro žádosti o **budoucí financování infrastruktury** z veřejných zdrojů (projekty)
- děkující uživatelé získávají **dočasnou prioritu (fairshare)**
 - publikace **rozděleny do kategorií** (různá váha)
 - navýšení priority **je dočasné**:
 - **klesá s propočítaným časem** (dle využitých zdrojů)
 - výpočet **využívá dominantní zdroj** (CPU, paměť, ...)
 - výpočet **nezávisí na stavu systému** (řízeno výhradně požadavky úlohy)
 - výpočet požadující **cennější zdroje platí víc**
 - **klesá se stářím publikace** (stáří max. 2 roky)

MetaVO – publikace s poděkováním II.

Publikace s poděkováním



MetaVO – publikace s poděkováním III.

- **Publikační úspěchy uživatelů (116 publikací za 2014)**
 - LL MU: Gora, A., Brezovsky, J., Damborsky, J.: Gates of Enzymes. CHEMICAL REVIEWS Volume: 113 Issue: 8, 2013. **IF 45,661**
 - CEITEC MU: VŠIANSKÁ, M. and M. ŠOB. The effect of segregated sp-impurities on grain-boundary and surface structure, magnetism and embrittlement in nickel. Progress in Materials Science, Elsevier, 56, 6, 2011. **IF 25,87**
 - LL MU: Sykora, J., Brezovsky, J., Koudelakova, T., Lahoda, M., ...: Dynamics and Hydration Explain Failed Functional Transformation in Dehalogenase Design. Nature Chemical Biology 10: 428-430, 2014. **IF 13,217**
 - UOCHB: L. Pospíšil, L. Bednářová, P. Štěpánek, P. Slavíček, ...: Intense Chiroptical Switching in a Dicationic Helicene-Like Derivative: Exploration of a Viologen-Type Redox Manifold of a Non-Racemic Helquat. Journal of the American Chemical Society, 2014 . **IF 11,444**



Meta VO – dostupné výpočetní zdroje

- výpočetní zdroje: **10024 jader (x86_64)**
- uzly s **nižším počtem výkonných jader**:
 - 2x4-8 jader
- uzly se **středním počtem jader (SMP stroje)**:
 - 32-80 jader
- paměť až **1 TB na uzel**
- uzly s **vysokým počtem jader: SGI UV 2000**
 - 288 jader (x86_64), 6 TB operační paměti
 - **384 jader (x86_64), 6 TB operační paměti**
- další „exotický“ hardware:
 - uzly s GPU kartami, SSD disky, ...



Meta VO – dostupné úložné zdroje

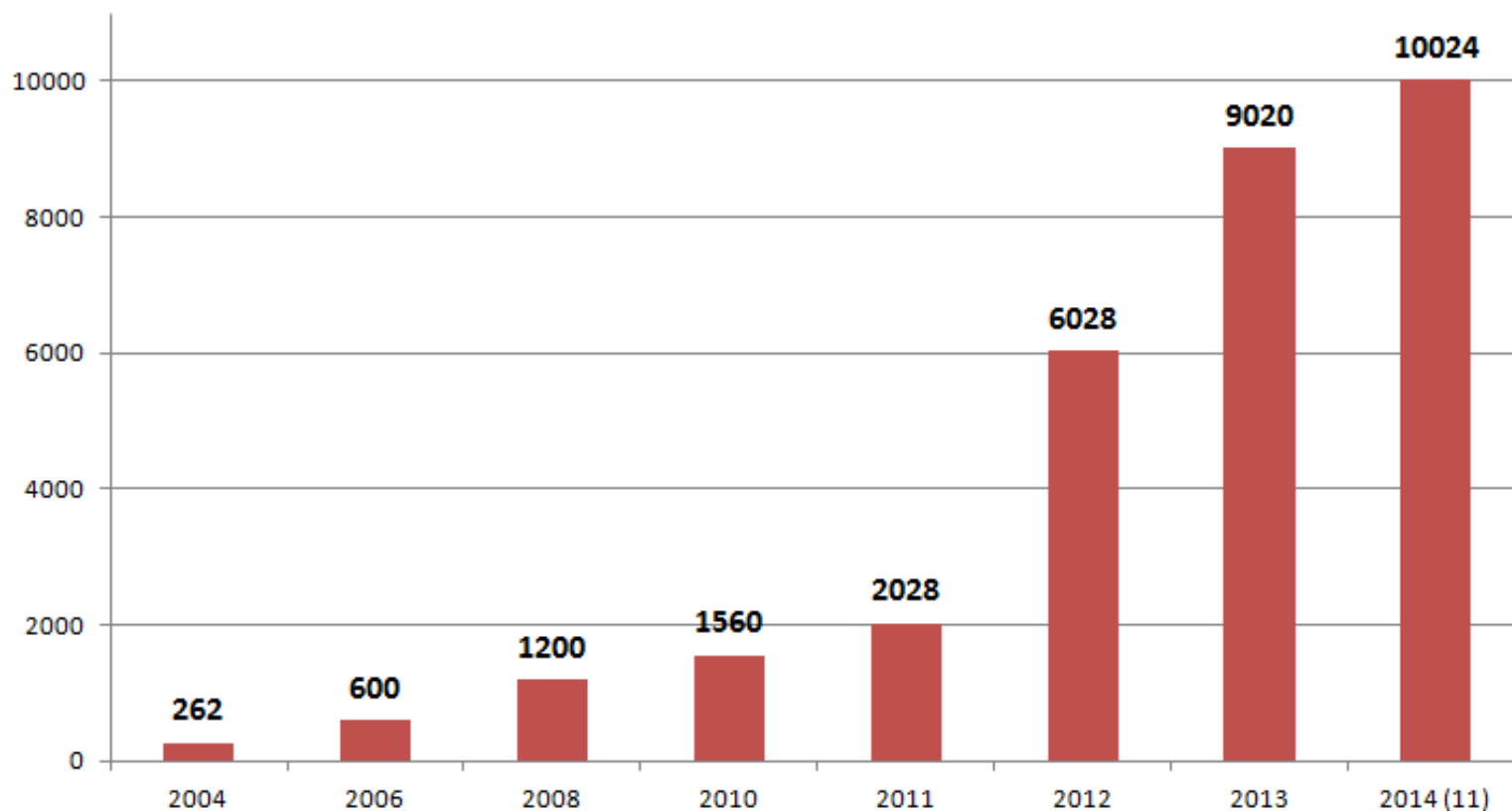
- **cca 1,1 PB (1136 TB)** pro pracovní data
 - úložiště 2x v Brně, 1x v Plzni, 1x v ČB, 1x v Praze, 1x v Jihlavě, 1x v Ostravě
 - uživatelská kvóta 1-3 TB na každém z úložišť
- **cca 13 PB** pro archivní data
 - (integrace DÚ Cesnet)
 - https://wiki.metacentrum.cz/wiki/Archivní_data_-_návod
 - **prosíme, využijte pro svá dlouhodobě nevyužívaná data**

Meta VO – dostupné aplikační programy

- ~ **300 různých aplikací**
 - viz <http://meta.cesnet.cz/wiki/Kategorie:Aplikace>
- průběžně udržované **vývojové prostředí**
 - GNU, Intel, PGI, ladící a optimalizační nástroje (TotalView, Allinea), ...
- generický **matematický software**
 - Matlab, Maple, Mathematica, gridMathematica, ...
- komerční i volný **software pro aplikační chemii**
 - Gaussian 09, Gaussian-Linda, Gamess, Gromacs, Molpro, Turbomole...
- **materiálové simulace**
 - Wien2k, ANSYS CFD (Fluent + CFX), Ansys Mechanical...
- **strukturní biologie, bioinformatika**
 - CLC Genomics Workbench, Geneious, Turbomole, Molpro, ...
 - řada volně dostupných balíčků
- **instalace aplikací uživateli** či **na žádost** (spolupráce na tvorbě dokumentace)

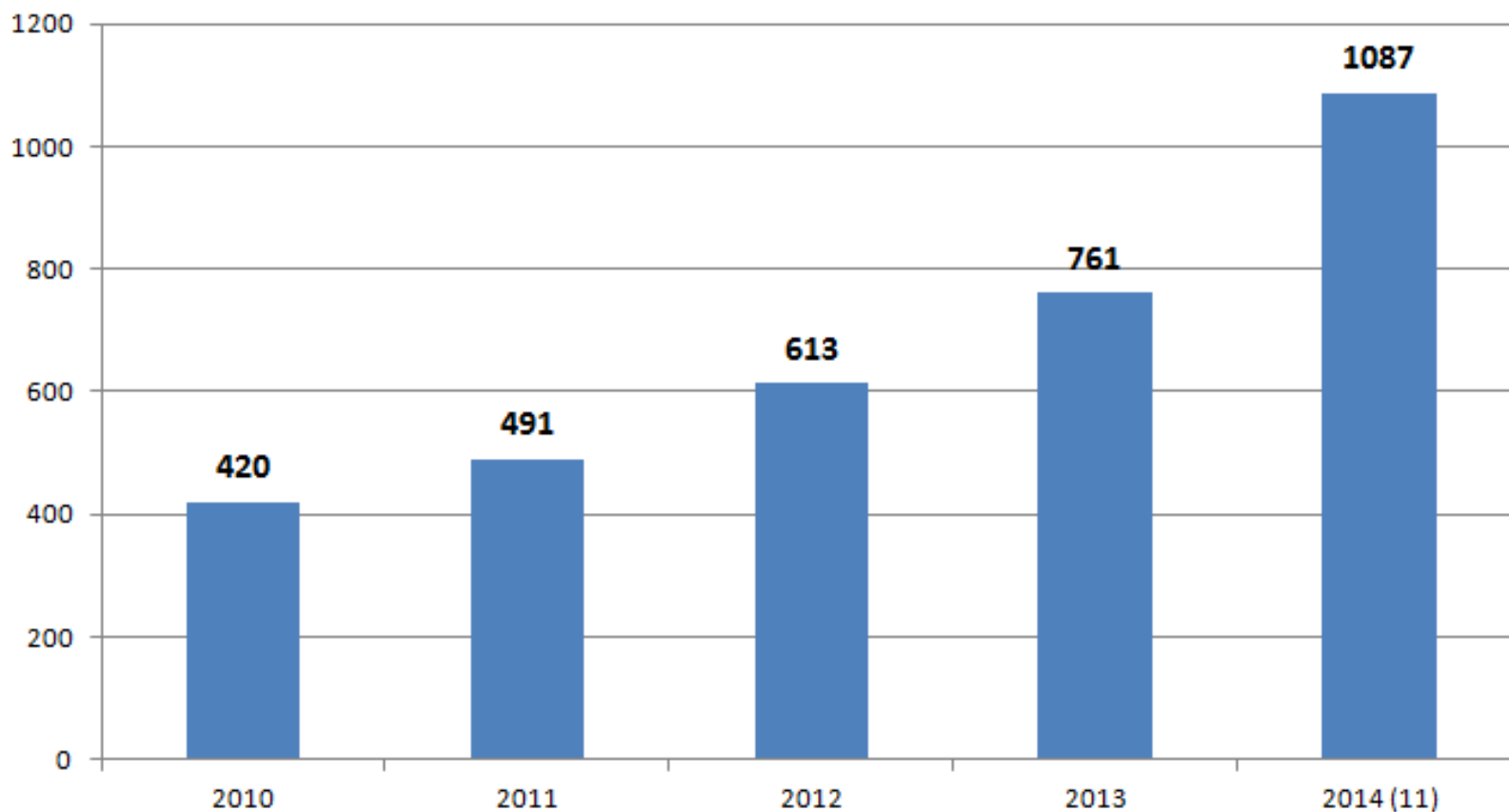
Meta VO v grafech... I.

Vývoj počtu CPU

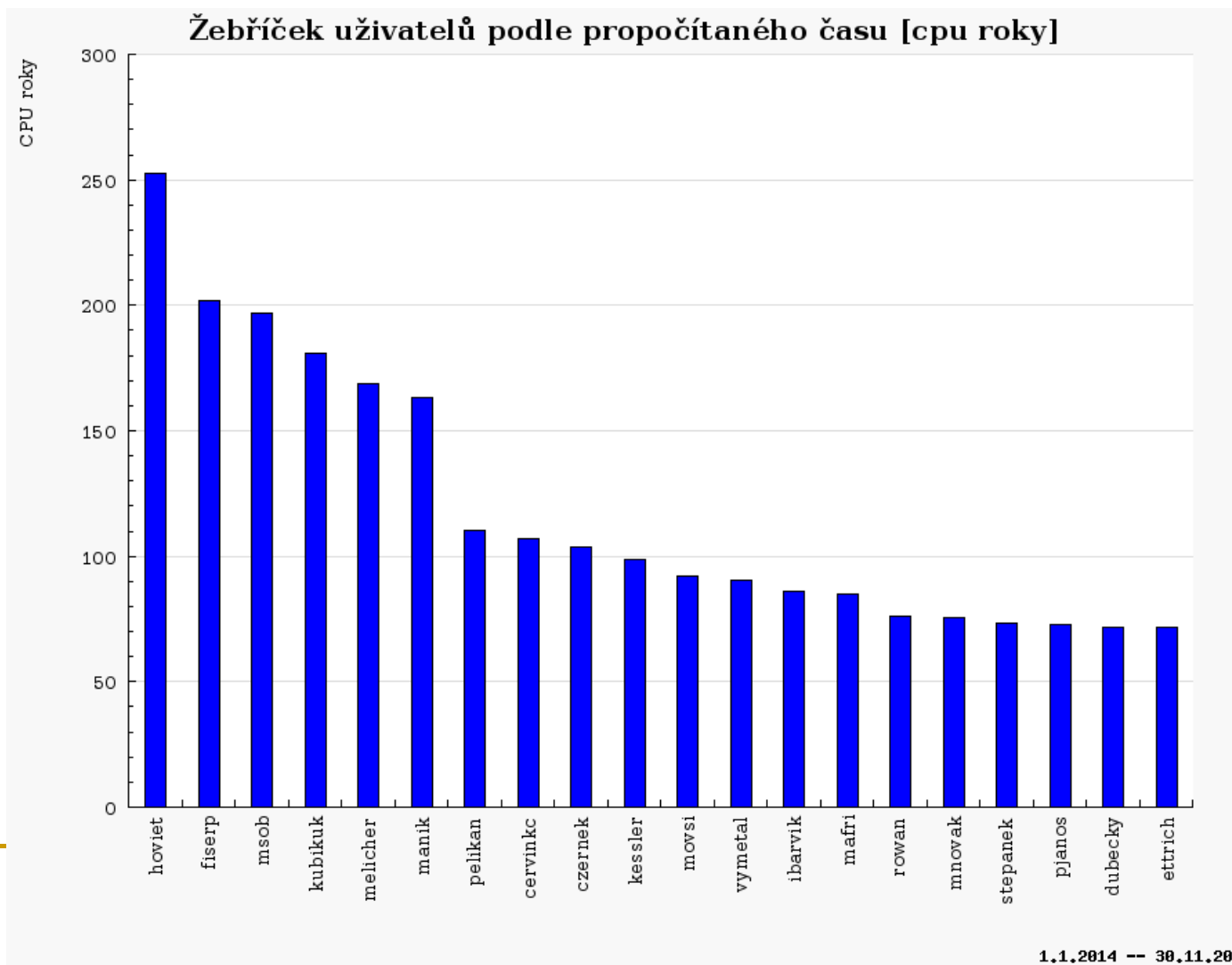


MetaVO v grafech... II.

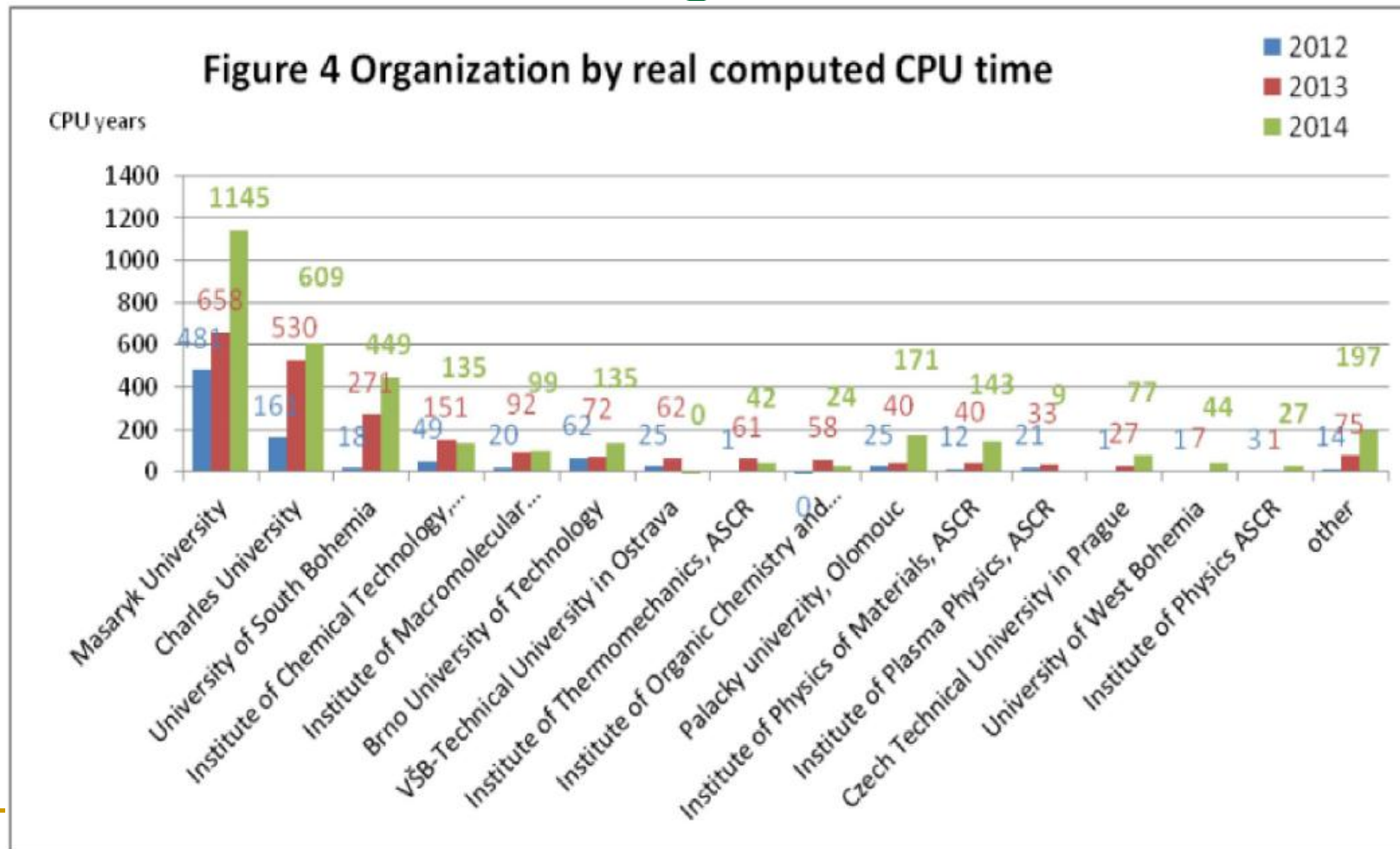
Vývoj počtu uživatelů



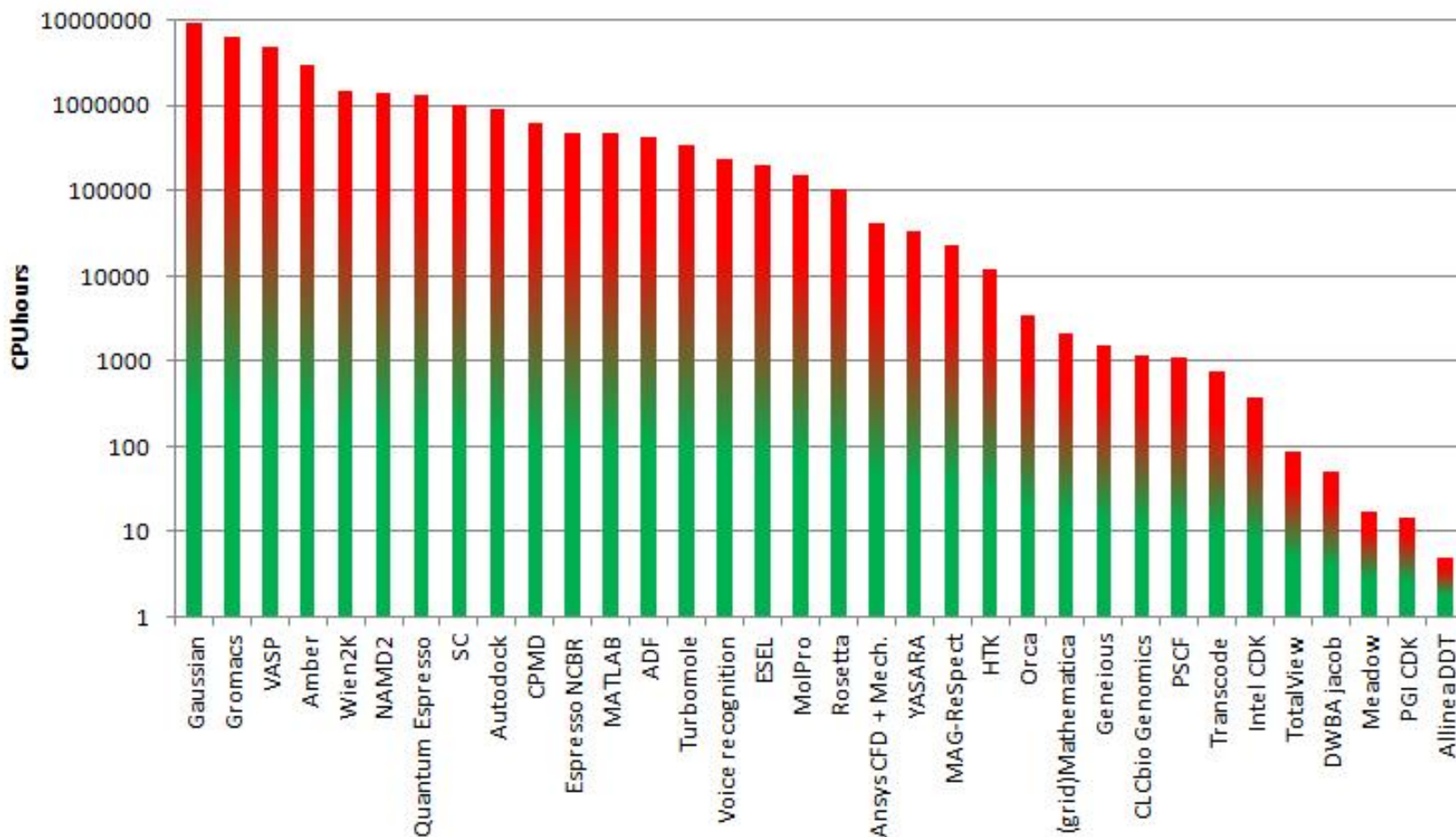
MetaVO v grafech... III.



Meta VO v grafech... IV.



Meta VO v grafech... V.

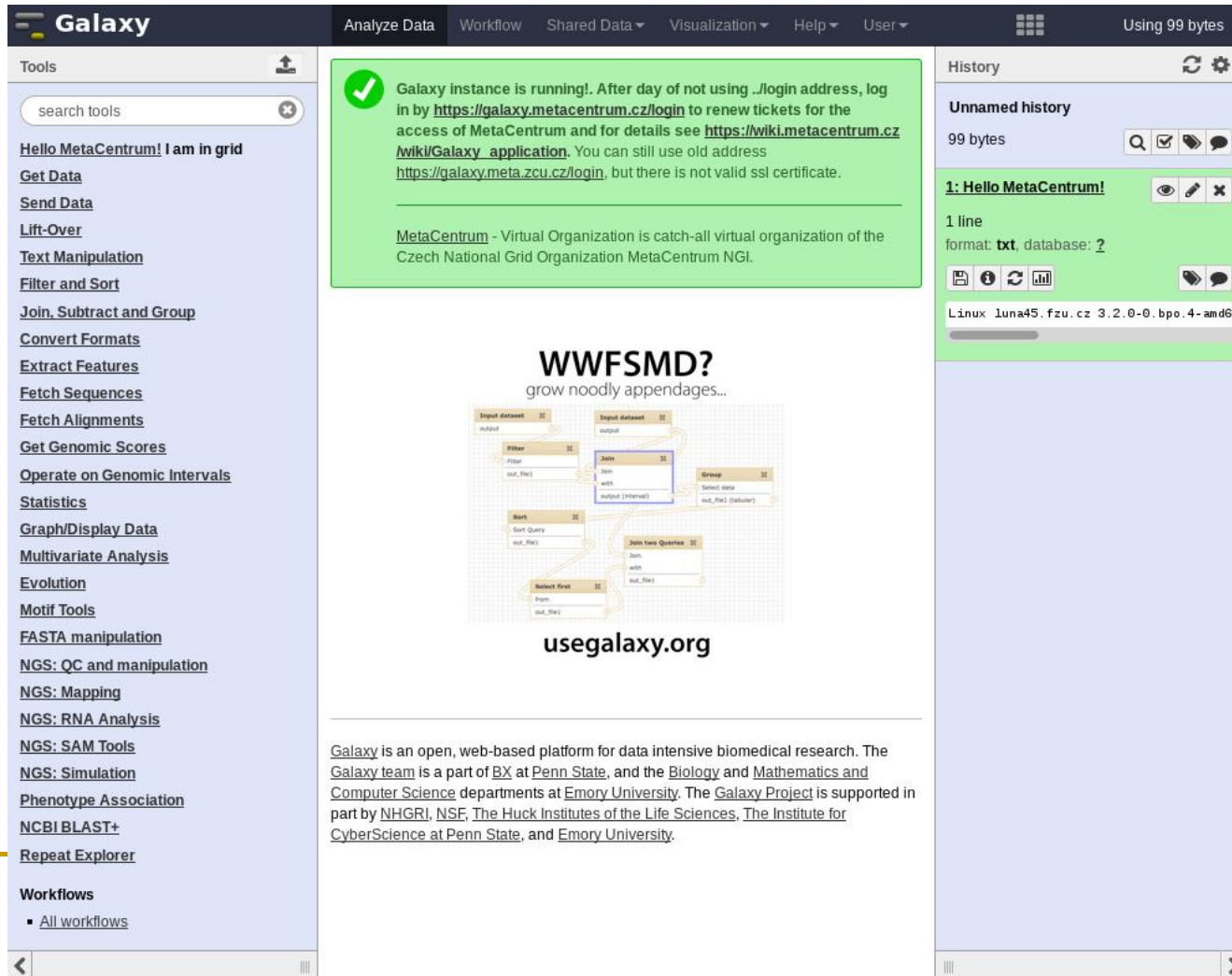


Co jsme od minula realizovali

Zprovoznění produkčního portálu GALAXY I.

- open-source **scientific workflow systém**
 - určený převážně pro bioinformatickou komunitu
- webové rozhraní <https://galaxy.metacentrum.cz>
 - autentizace standardně jméno&heslo
 - manipulace s daty a spouštění úloh **přímo z webového prohlížeče**
- návody a tutoriály
 - https://wiki.metacentrum.cz/wiki/Galaxy_application
- podpora více než **40 běžných bioinformatických aplikací**
 - např. Blast, Bfast, Bowtie, BWA, Cuff tools, RSEM, TopHat, ...

Zprovoznění produkčního portálu GALAXY II.



The screenshot shows the Galaxy II web interface. At the top, there is a navigation bar with 'Galaxy' and 'Using 99 bytes'. Below this is a 'Tools' sidebar on the left with a search bar and various tool categories like 'Hello MetaCentrum!', 'Get Data', 'Send Data', etc. The main content area features a green notification box with a checkmark stating: 'Galaxy instance is running!. After day of not using ..login address, log in by <https://galaxy.metacentrum.cz/login> to renew tickets for the access of MetaCentrum and for details see https://wiki.metacentrum.cz/wiki/Galaxy_application. You can still use old address <https://galaxy.meta.zcu.cz/login>, but there is not valid ssl certificate.' Below the notification is a workflow diagram titled 'WWFSMD? grow noodly appendages...' with the URL 'usegalaxy.org'. The diagram shows a sequence of tools: 'Input dataset', 'Filter', 'Join', 'Sort', 'Join two Queries', and 'Select Row'. To the right of the workflow is a terminal window showing the command 'Linux luna45.fzu.cz 3.2.0-0.bpo.4-amd64' and the output '1 line format: txt, database: ?'. At the bottom of the terminal, there is a progress bar.

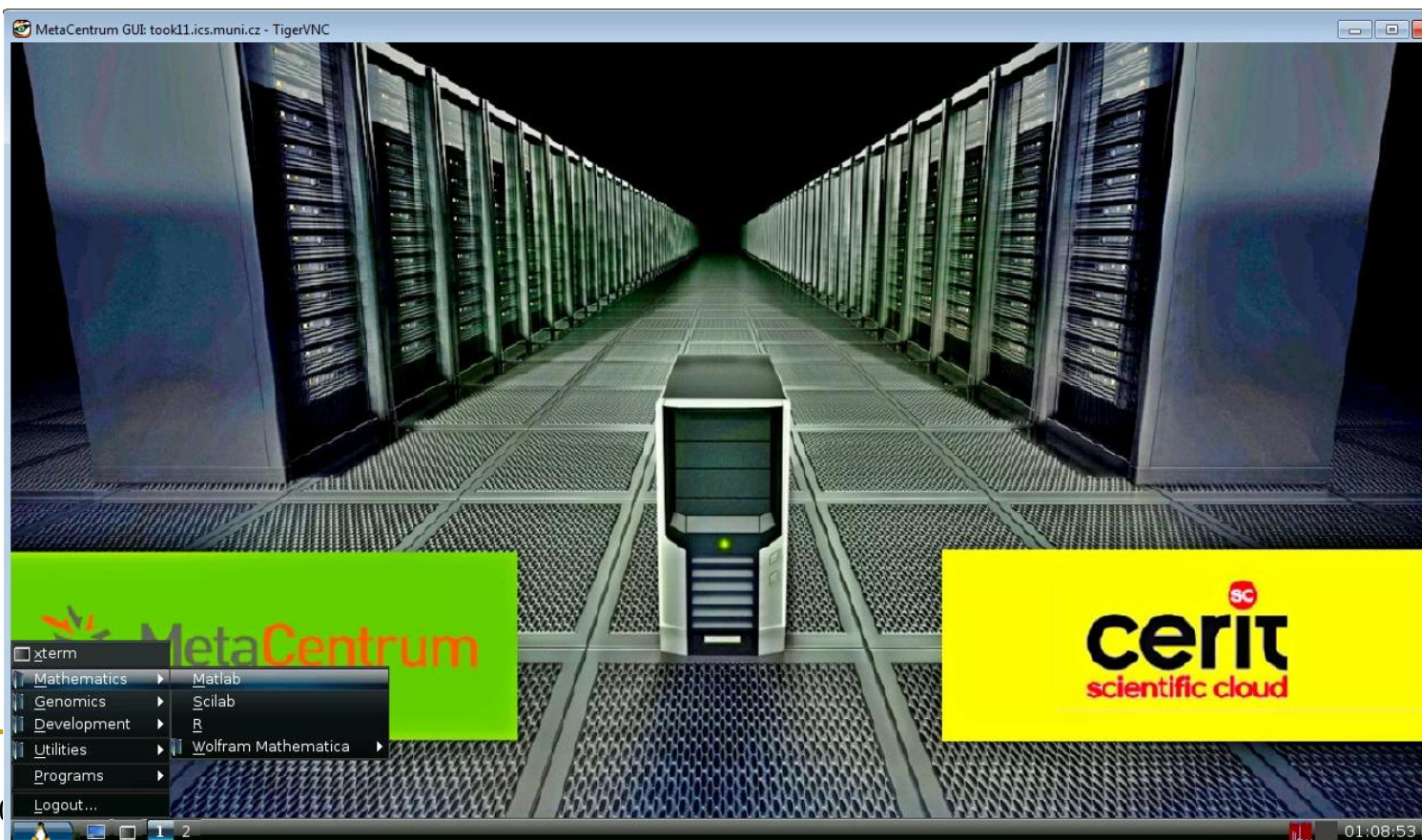
Interaktivní práce s grafickými aplikacemi I.

- narůstající množství výpočetně náročných aplikací vyžadujících grafické prostředí
- dosavadní způsob práce (export/tunelování displeje) nepříliš vhodný pro interaktivní práci
- **realizované řešení:**
 - **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**
 - více viz https://wiki.metacentrum.cz/wiki/Vzdálený_desktop

Interaktivní práce s grafickými aplikacemi II.

```
$ module add gui
```

```
$ gui start
```



Optimalizace plánovacích systémů I.

CERIT-SC

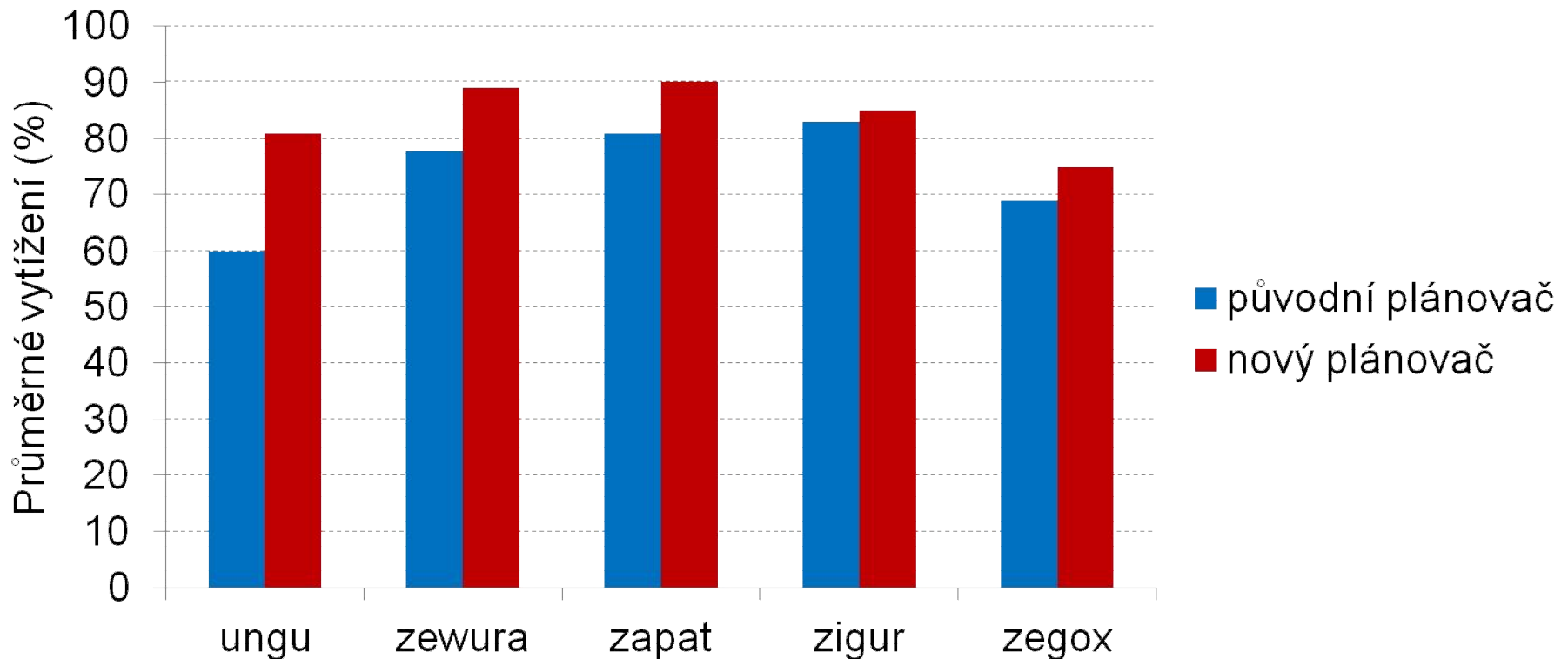
- nasazení experimentálního rozvrhového plánovače
- → **efektivnější plánování** díky implementaci backfillingu
- → vytváření rozvrhu => schopnost **odhadnout předpokládaný start úlohy**
- https://wiki.metacentrum.cz/wiki/Manual_for_the_TORQUE_Resource_Manager_with_a_Plan-Based_Scheduler

MetaCentrum

- **optimalizace algoritmu střádání zdrojů** pro čekající úlohy za účelem zvýšení dostupnosti systému
- doplněna **predikce startu úlohy**

Optimalizace plánovacích systémů II.

CERIT-SC – ilustrace zlepšení vytížení clusterů



Scratch adresáře

- **různé typy scratchů**
 - klasické + SSD
 - sdílené (clustery *hildor*, *mandos*, *ramdal* a *doom*)
 - podpora množstevních jednotek (KB, MB, GB, ...)
 - `-l scratch=SIZE:TYPE` (např. `-l scratch=50gb:shared`)
- **uzavření dat jednotlivých úloh do vlastního adresáře**
 - usnadnění identifikace dat historických úloh (odmazávání po 14 dnech)
 - na přidělený prostor odkazuje **proměnná \$SCRATCHDIR**
- **uzavření základního scratch adresáře pro zápis**
 - opět z důvodu **snadnější identifikace historických dat** (odmazávání)
 - pokus o promazání `"rm -rf $SCRATCHDIR"` může vést na hlášku
`rm: cannot remove ...: permission denied`

Doporučená kostra skriptů

Vytvoření **utility pro promazávání scratchů (clean_scratch)** =>
nová doporučená kostra skriptu:

- kromě bezpečného promazání **informuje o opomenutých datech** na uzlu

```
#!/bin/bash

# set a handler to clean the SCRATCHDIR once finished
trap 'clean_scratch' TERM EXIT
# if temporal results are important/useful
# trap 'cp -r $SCRATCHDIR/data $DATADIR && clean_scratch' TERM
...

# copy out the output data
# if copying fails, let data in SCRATCHDIR and inform user
cp $SCRATCHDIR/output.txt $DATADIR || export CLEAN_SCRATCH=false
```

Projektové adresáře & sdílení dat

- semipermanentní adresáře určené pro **sdílení objemnějších dat v rámci skupiny nebo projektu**
 - uloženo v `/storage/projects/XXX`
 - (*fyzická pozice: HSM v Jihlavě*)
 - mj. určeno i pro **sdílené databáze SW**
 - např. pravidelně aktualizované **zrcadlo všech BLAST databází** v `/storage/projects/BlastDB/`

Sdílení dat v rámci skupin:

- podpora sdílení libovolného NFS adresáře (i v dom. adresářích)
- unixová skupina + **webové rozhraní pro správu členů**
 - rozšíření plánovače pro **možnost specifikace uživatelské skupiny**
 - `qsub ... -W group_list=SKUPINA`

Nový systém modulů

- transparentně nasazena **nová verze modulů** (*modulefiles*)
- **nejvýznamnější uživatelské výhody:**
 - *bash-completion* („*tab-completion*“)
 - snadnější vyhledávání: `module avail matl`
 - ...

II. etapa nasazení (plán 2015):

- transparentní změna **struktury modulů** a **struktury zápisu**
- `matlab-8.4` → `matlab/8.4` (kompatibilitu zachováme)

III. etapa nasazení (plán 2015):

- implementace **doplňkových informačních funkcí**
- např. nutnost odsouhlasit licenci, použití na nelicencovaných zdrojích, atp.

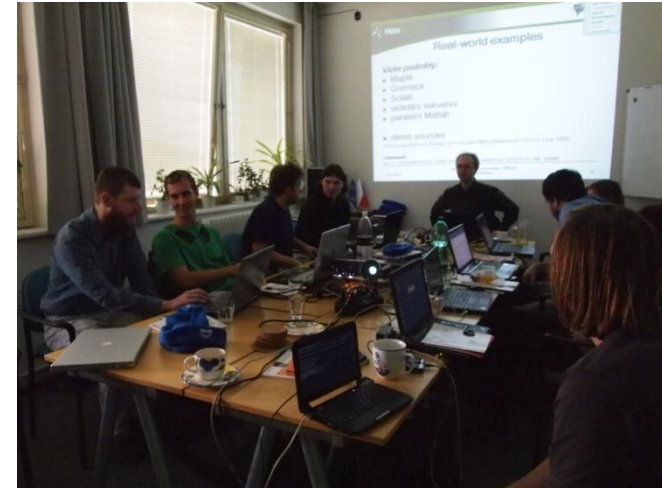
Integrace vlastních zdrojů

Možnosti integrace vlastních zdrojů:

- ***plná integrace*** – výkonný cluster integrovaný do infrastruktury MetaCentra/MetaVO
 - možnost prioritizace vlastníka
 - správa v režii MetaCentra
- ***privátní čelní uzel / výpočetní cluster*** – podpora skupin s HW (uzel, cluster) nevhodným pro plnou integraci
 - poskytnutí obrazu/VM 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,...)

Informovanost uživatelů

- **výjezdní hands-on semináře**
 - prakticky orientovaná školení (4-5 hodin)
 - v úzkém okruhu zájemců (do 10 osob)
 - orientace na školenou uživatelskou skupinu (praktické příklady)
 - Praha (3x), HK, Pardubice, Brno (3x), Olomouc (2x), Plzeň, Zlín – **100+ uživatelů**
- **odborné semináře**
 - *Architektura SGI UV2000, Využití portálu GALAXY*
 - úzce specializovaná školení
- **úpravy PBSmon portálu** (např. GPU využití, úložiště) a **dokumentace, rozesílání průběžných novinek, atp.**



Provozní změny – různé

- specifikace **odhadu času úlohy** skrze parametr `walltime`
 - a zaslání bez specifikace fronty
- **inteligentní vlastnost infiniband** (`nodes=...:infiniband`)
- **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)
 - **vytvoření utility** `check-local-quota` pro nalezení nepořádku
- dokončen **plný upgrade** na systém **Debian 7**
- průběžná implementace **nového designu stránek** MetaCentra
- původní `zewura9-zewura20` **přejmenovány na cluster zebra**
- ...

Co dalšího plánujeme?
(o zavedení budete informováni aktualitou)

Chystané novinky I.

Stránka „Moje problémy“ na portále:

- na jednom místě poskytne informace o pro Vás relevantních problémech (a akcích, kterými je řešit)
- například:
 - opomenutá data na scratchích (po ukončených úlohách)
 - vyčerpané kořínkové/domovské kvóty
 - nespustitelné úlohy
 - neefektivní úlohy
 - zabitě úlohy
 - staré soubory vhodné k archivaci
 - atp.
- upozornění na problémy **motd/email notifikacemi**

Chystané novinky II.

Experimentální nasazení spolupráce plánovačů MetaCentra & CERIT-SC:

- testovací příprava pro budoucí plné nasazení (fronta **p2ptest**)
 - jakákoliv úloha zařazená do této fronty a zaslána na libovolný plánovač **může být předána** spolupracujícímu plánovači (pokud vyhovují požadavky)
- *aktuální fáze*: ladění posledních detailů

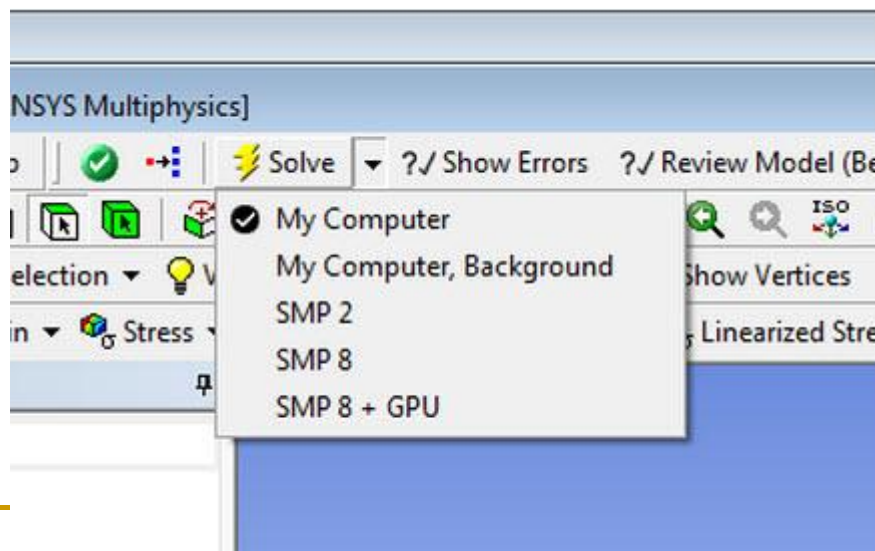
Sandbox pro úlohy:

- **tvrdé omezení CPU** (nepřekročitelné) → úloha nebude zabita
- **paměť zřejmě omezovat nebudeme** (probíhající diskuze)
 - *problémy*: při horním omezení paměti a dosažení maximální hodnoty výpočet stejně spadne, mnohdy s nicneříkající chybou
 - zabití & email lépe informují o důvodech problémů

Chystané novinky III.

Ansys RSM:

- **Ansyes Remote Solve Manager** upravován pro spolupráci s gridovou infrastrukturou
- zatím ve stádiu funkčního prototypu, nutný přechod na produkční systém



Chystané novinky IV.

Propagační publikace:

- **akceptace propagačních publikací/výstupů** pro navýšení fairshare
 - propagační informace o Vašem výzkumu realizovaném na naší infrastruktuře
 - (odborný/veřejný) web, veřejnoprávní média, propagační materiály, atp.
- **po oznámení dočasné zvýšení fairshare**
 - avšak nižší bonus než časopisecké/konferenční články
- **v budoucnu i bonus za výpomocnou činnost MetaVO**
 - lokální podpora konkrétní skupiny, instalace aplikací, úprava dokumentace, atp.

Závěrem ...

- infrastruktura se neustále vyvíjí
 - způsoby (efektivního) využití se průběžně mění
 - snaha adaptovat infrastrukturu Vašim potřebám
- snaha o udržení informovanosti uživatelů
 - školicí semináře
 - pravidelné oznamy
 - stránky dokumentace
- o Vaši zpětnou vazbu Vás prosíme formou dotazníku
 - *Co se Vám u nás líbí/nelíbí? Máte nápady na vylepšení? atd. atd.*
 - za jeho vyplnění budete **odměněni malým dárkem** 😊

Děkuji Vám za pozornost!

`rebok@ics.muni.cz`



metavo.metacentrum.cz

www.cerit-sc.cz