

Jak ti vědci počítají

Zdeněk Šustr

5. 10. 2009

Superpočítače

- „*Supercomputer: any of a class of extremely powerful computers*“ (Encyclopedia Britannica)
- Zjevné řešení, zejména při:
 - Úlohách s extrémními nároky, zvláště pokud neumožňují masivní paralelizaci
 - Vysokém důrazu na zabezpečení
 - Např. vojenský výzkum
 - Schopnosti naplnit kapacity
 - Dostatku financí
- Přináší jisté nevýhody
 - Zastarávání
 - Omezená škálovatelnost
 - Komplikované spoluvlastnictví

TOP 500

TOP500.org je aktualizovaný přehled 500 nejvýkonnějších počítačů.

č.	Provozovatel	Popis	Výkon [PFLOPS]
1.	DOE/NNSA/LANL	Roadrunner – BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3,2 Ghz / Opteron DC 1,8 GHz, Voltaire Infiniband, IBM (129 600 jader)	1,105
2.	Oak Ridge National Laboratory	Jaguar – Cray XT5 QC 2,3 GHz, Cray Inc.	1,059
3.	Forschungszentrum Juelich (FZJ)	JUGENE – Blue Gene/P Solution, IBM	0,826
4.	NASA/Ames Research Center/NAS	Pleiades – SGI Altix ICE 8200EX, Xeon QC 3,0/2,66 GHz, SGI	0,487

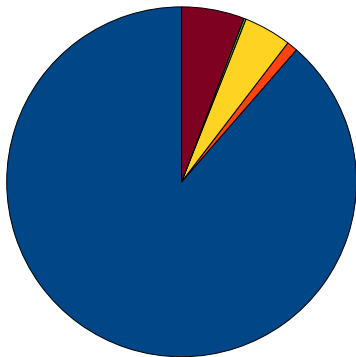
data: <http://www.top500.org>

TOP500

TOP 500 – Přehledové grafy

TOP500

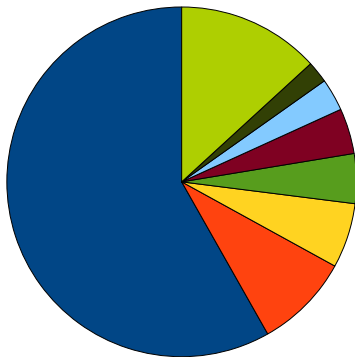
Zastoupení rodin operačních systémů



■ Linux ■ Windows ■ Unix
■ BSD ■ Mixed

TOP500

Zastoupení zemí



■ USA ■ UK ■ Německo ■ Francie
■ Čína ■ Japonsko ■ Švédsko ■ Ostatní

data: <http://www.top500.org>

Gridy – Jak?

- Distribuované prostředí složené ze „standardních“ PC
 - V množství je síla!
 - Slušné PC nebo katedrální cluster se najde takřka všude a ne vždy se plně využívá
 - V některých vědních oborech (částicová fyzika, výpočetní chemie) se výpočetní úkoly snadno paralelizují
- Zajímavé i pro provozovatele s vlastním HW
 - Možnost „balancingu“
 - *„Když nepočítám, dávám HW k dispozici. Když pak potřebuji, mohu krátkodobě využít i větší kapacitu než mám doma.“*
 - Dostupnost výpočetních kapacit i při výpadku vlastního zařízení
 - Vkladem je HW a práce správců, které tak jako tak mám

Gridy – Proč?

- Výhody
 - Dostupnost
 - Přidat se dá kdykoli
 - *Snadné zapojení*
 - Není třeba naplnit celou kapacitu
 - Škálovatelnost
 - Distribuované stárnutí/obměna
- „*Black box*“ z pohledu uživatele
 - Výpočetní úloha se „vypustí“ do gridu a z něj po čase „vypadne“ výsledek

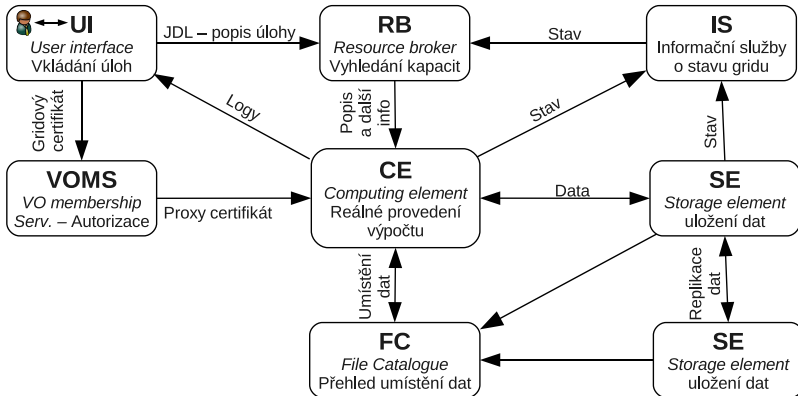


Gridový middleware

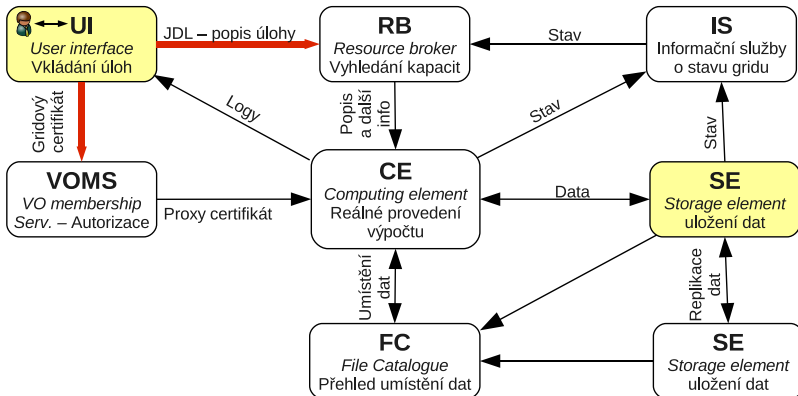
- Soubor služeb, které se starají o životní cyklus úlohy
 - Autorizační služby
 - Plánování
 - Výpočetní uzly
 - Uložení a katalogizace dat
- Základní služby (spuštění úlohy, přesuny I/O dat, dotaz na stav apod.) dostupné prostřednictvím standardního uživatelského rozhraní
- Rozsáhlé API pro implementaci specifických úkolů
- Existuje několik velkých middlewareových projektů po celém světě (EU, USA, Japonsko, JV Asie)
 - Pro ČR relevantní zejména gLite vyvíjený v rámci evropského projektu EGEE

Gridový middleware

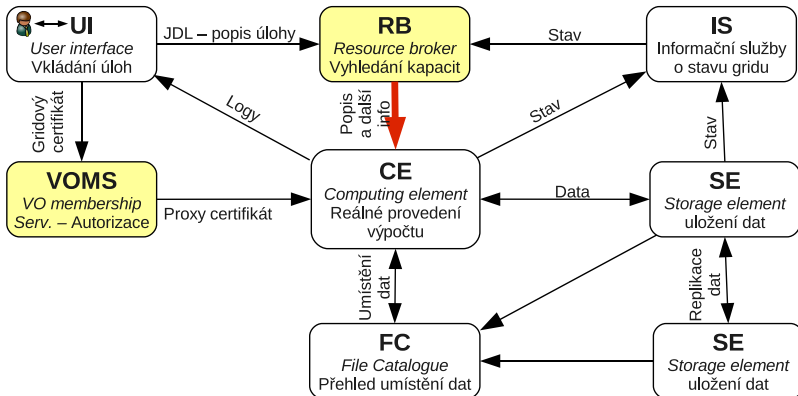
Jednoduchý přehled komponent



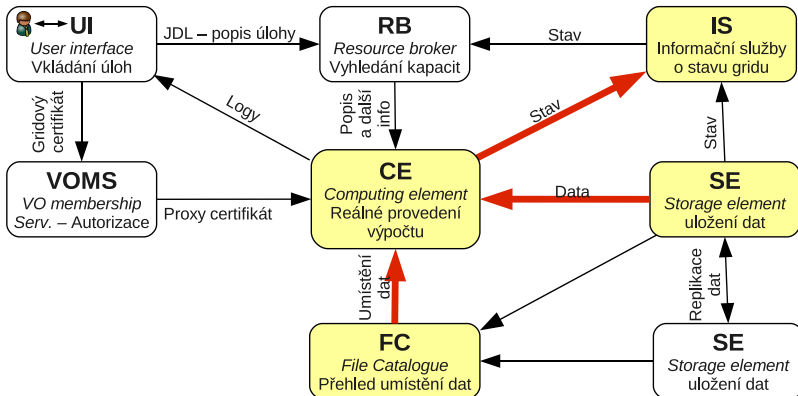
Spuštění úlohy



Plánování a fronty

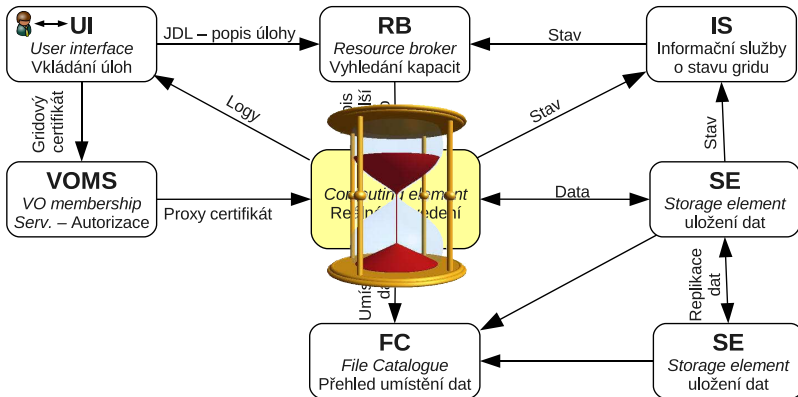


Spuštění výpočtu

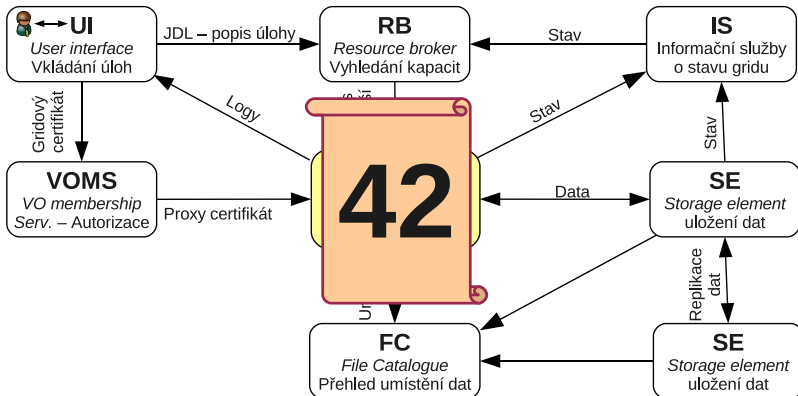


Gridový middleware

Výpočet

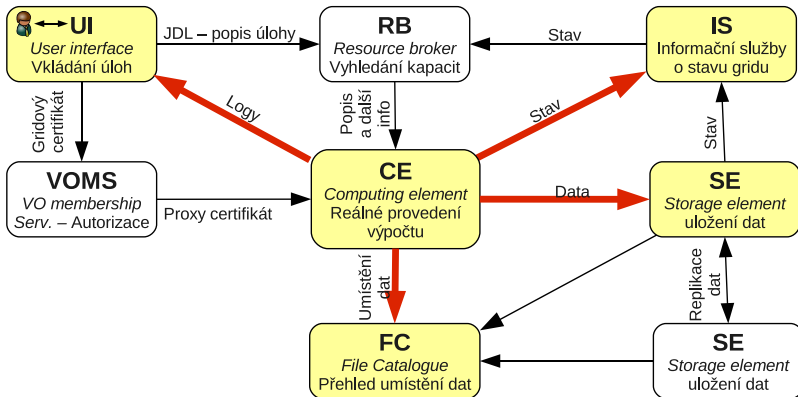


Výsledek



Gridový middleware

Uložení výsledků



Existující infrastruktura

- EGEE – největší produkční grid pro vědecké výpočty
 - gLite middleware
 - 92 000 procesorů (144 000 jader)
 - z toho přibližně 1 600 procesorů v ČR
 - cca. 25 PB diskového prostoru
 - benchmark?

Existující infrastruktura

- EGEE – největší produkční grid pro vědecké výpočty
 - gLite middleware
 - 92 000 procesorů (144 000 jader)
 - z toho přibližně 1 600 procesorů v ČR
 - cca. 25 PB diskového prostoru
 - benchmark?
 - Nelze spustit plošně
 - Výkon (velmi přibližně!) 0,75 PFLOPS?

Existující infrastruktura

- EGEE – největší produkční grid pro vědecké výpočty
 - gLite middleware
 - 92 000 procesorů (144 000 jader)
 - z toho přibližně 1 600 procesorů v ČR
 - cca. 25 PB diskového prostoru
 - benchmark?
 - Nelze spustit plošně
 - Výkon (velmi přibližně!) 0,75 PFLOPS?
- Open Science Grid (USA), NAREGI (Japonsko), NorduGrid (Skandinávie), EUAsiaGrid. . .
- MetaCentrum – národní grid pro ČR
 - dávkový systém PBS
 - 1 150 procesorů
 - clustery CESNETu a dalších akademických institucí v ČR

Cycle scavengers

- Speciální případ gridů – dobrovolná účast jednotlivců
 - Nízký počet zadavatelů
 - Max. jednotky algoritmů, ale podpora mnoha platforem
 - Vhodné pouze pro atraktivní obory
 - Genetika, astrofyzika, změny klimatu

Cycle scavengers

- Speciální případ gridů – dobrovolná účast jednotlivců
 - Nízký počet zadavatelů
 - Max. jednotky algoritmů, ale podpora mnoha platforem
 - Vhodné pouze pro atraktivní obory
 - Genetika, astrofyzika, změny klimatu
 - Koho láká ***Nuclear.Explosion@home?***

Cycle scavengers

- Speciální případ gridů – dobrovolná účast jednotlivců
 - Nízký počet zadavatelů
 - Max. jednotky algoritmů, ale podpora mnoha platforem
 - Vhodné pouze pro atraktivní obory
 - Genetika, astrofyzika, změny klimatu
 - Koho láká ***Nuclear.Explosion@home?***
 - Existují pilotní projekty (např. *Gomez*), které procesorový čas vykupují

Cycle scavengers

- Speciální případ gridů – dobrovolná účast jednotlivců
 - Nízký počet zadavatelů
 - Max. jednotky algoritmů, ale podpora mnoha platformem
 - Vhodné pouze pro atraktivní obory
 - Genetika, astrofyzika, změny klimatu
 - Koho láká ***Nuclear.Explosion@home?***
 - Existují pilotní projekty (např. *Gomez*), které procesorový čas vykupují

č.	Projekt	Výkon [PFLOPS]
1.	Folding@home	7,565
2.	BOINC (Seti@home, Einstein@home...)	2,411
3.	GIMPS (Great Internet Mersenne Prime Search)	0,041



Cloudy



- Virtuální výpočetní prostředí dimenzované na míru
 - Typicky jako komerční služba
- Homogenní
 - Výhoda i nevýhoda
 - Možnost dodat vlastní image zatím není samozřejmostí, ale není vyloučena
- Z pohledu uživatele obyčejný počítač přístupný např. přes SSH
 - Management úloh je třeba řešit dodatečně
- Předmět debat a nesmělých pokusů

Cloudy & gridy dohromady

- **Virtualizace** – Populární buzzword
 - V tomto kontextu „*Cloudifikace gridů*“ či „*gridifikace cloudů*“
 - tj. využití výhod obojího
- Gridové služby v cloudu?
 - Vznikne malý virtuální grid
- Cloudy v gridu?
 - Populární trend
 - Netušené možnosti?

Cloudy v gridech

- Virtuální stroj nebo celý cluster se spustí jako úloha v gridu
 - Souběžně s klasickými úlohami
 - Rozhodnutí zcela na uživateli
 - Vhodné pro úlohy se specifickými požadavky
 - Prodloužení životnosti zastarávajících systémů nebo softwarů
- Kvalitní podpora na linkové vrstvě
 - Společná VLAN i při geografickém odstupu uzlů
 - Zapojení do VLAN pracoviště

Typické použití

- **Vytvoření bezpečného prostředí** – VLAN odělená od okolního prostředí umožňuje provoz systémů se slabým zabezpečením nebo zastaralými knihovnamí

Typické použití

- **Vytvoření bezpečného prostředí** – VLAN odělená od okolního prostředí umožňuje provoz systémů se slabým zabezpečením nebo zastaralými knihovnamí
- **Provoz a škálování exotických systémů** – Není třeba přizpůsobovat úlohu gridovému prostředí. Stačí dodat vlastní image, naklonovat dle potřeby a počítat ve známém prostředí

Typické použití

- **Vytvoření bezpečného prostředí** – VLAN odělená od okolního prostředí umožňuje provoz systémů se slabým zabezpečením nebo zastaralými knihovnamí
- **Provoz a škálování exotických systémů** – Není třeba přizpůsobovat úlohu gridovému prostředí. Stačí dodat vlastní image, naklonovat dle potřeby a počítat ve známém prostředí
- **Usnadnění přechodu ke gridům** – Virtualizace existujícího prostředí a přechod na grid

Typické použití

- **Vytvoření bezpečného prostředí** – VLAN odělená od okolního prostředí umožňuje provoz systémů se slabým zabezpečením nebo zastaralými knihovnamí
- **Provoz a škálování exotických systémů** – Není třeba přizpůsobovat úlohu gridovému prostředí. Stačí dodat vlastní image, naklonovat dle potřeby a počítat ve známém prostředí
- **Usnadnění přechodu ke gridům** – Virtualizace existujícího prostředí a přechod na grid
- **Rozšíření nebo doplnění stávající infrastruktury** – Zapojení do lokální VLAN, posílení fyzických kapacit

Současný stav

- Probíhá horečný vývoj
- Řada pilotních instalací
 - Různý stupeň automatizace při dodávání, kopírování či spouštění obrazů virtuálních strojů
 - Různý stupeň propojení se síťovými službami
- V ČR má pilotní implementaci MetaCentrum

Shrnutí

- **Superpočítače**
 - Nepostradatelné, zejména pro některé typy úloh
 - Je třeba zdůvodnit vysoké pořizovací náklady a naplnit kapacitu kvalitním vědeckým programem
- **Gridy**
 - Vynikající pro řešení úloh, které umožňují masivní paralelizaci
 - Vhodné i pro ukládání velkého množství dat
 - Umožňují pozvolný vstup
- **Cloudy**
 - Nové, převážně komerční řešení
 - Prozatím celkově dražší než gridy
 - Výhodou je homogenní prostředí a podpora na komerční úrovni

Poděkování

Děkuji za pozornost 😊

Superpočítače
○○○

Gridy
○○○○○○○○○○○○○○

Cloudy
○

Cloudy & gridy
○○○○

Shrnutí
○○

Poděkování

To už je opravdu konec