

*META***Centrum**

**Distribuované výpočetní a datové prostředí
pro vědeckou spolupráci**

Luděk Matyska, jménem celého týmu

Celkový přehled činnosti

- Virtualizace
 - Nasazení na infrastrukturu
 - Plánování – Magrathea
- IPv6
- Uživatelská podpora
- Úložné prostory
 - Analýza současné situace
 - Posílení kapacit
- Síť
- Nové investice

Virtualizace

- Nové instalace operačního systému už vždy virtualizovány
 - Plán na $>50\%$ uzlů do konce roku bude splněn
- V současné době virtualizace „statická“
 - Na každém uzlu dva trvale běžící virtuální počítače
 - Plánovač (Magrathea) zvolí vždy právě jeden
 - * V podstatě jde o přepínání virtuálních strojů
- Nástroj na vzdálené ovládání virtuálních strojů připraven (CLI)
 - Webové rozhraní plánujeme do konce roku
 - Statická konfigurace však jeho nasazení zatím nevyžaduje

Infiniband

- Opravy a zprovoznění low-level ovladače v jádře Linuxu
 - Včetně podpory Xenu
- Probíhají testy a měření skutečné propustnosti Infinibandu
 - Export PCI do DomU
 - Dosavadní výsledky slibné
 - * Propustnost >10 Gbps
 - * Latence 1/10 latence Ethernetu
- Podpora sdíleného IB zatím v zárodku

IPv6

- **Analýza skutečného stavu služeb**
 - Základní bezpečnostní služby schopny přechodu
 - * Kerberos, openssh
 - Apache OK
 - dhcpv6 OK (Dibbler)
 - Plánovač (PBSPro) ne, ale víme, jak obejít
 - Zásadní problémy s distribuovanými systémy souborů
 - * AFS prakticky beznadějná
 - * NFSv4 zatím v přípravě (vlastními vývojáři)
- **Nasazení na provozních službách před koncem roku**

Nasazení koprocesorů

- Reálně nedostatek zájmu uživatelů
- Plánujeme využít GPU ve strojích pořízených pro GLIF demo
 - První větší zkušenosti díky stáží studenta Iana Wesley-Smithe z LSU (primárně aktivita 615), komprese HD proudu s využitím GPU
 - Vypsány dvě práce na MU
- Nasazení FPGA v dohledné době neplánujeme
 - Pokus o spolupráci s ing. Lexou v rámci aktivity 602 nebyl úspěšný

Integrace clusterů

- Integrace nově pořízeného clusteru Loschmidtovy laboratoře (MU Brno)
- Uzavřeno výběrové řízení na další cluster NCBR (MU), který bude také integrován
- Do konce roku integrace clusteru v Českých Budějovicích

Uživatelská podpora

- Migrace portálu na nový hardware a upgrade software
- Reorganizace a doplňování statického obsahu portálu
- Vylepšení a personalizace stránek s aktuálním stavem výpočetních zdrojů (PBSMom)
- Navázání přihlášek a změny hesla na autentizační systémy spolupracujících organizací (zárodek federativního přístupu)
- Postupné vylepšování RT systému
- Prezentace

Příprava nové strategie

- Zahájili jsme sběr informací o potenciálních velkých uživateliích *METACentra*
 - Výzkumné záměry, výzkumná centra, velké výzkumné skupiny
 - V současné době vytipovány cca 4 skupiny
- Plánujeme je přímo oslovit
 - Nabídnout reálnou spolupráci a společné hledání, jak by jim mohlo *METACentrum* konkrétně pomoci
- Proběhl i sběr informací o instalovaných clusterech v akademickém prostředí ČR

Úložné kapacity

- Probíhá analýza možných řešení
 - Inventarizace kapacit a zejména jejich využití
 - Konzultace s potenciálními dodavateli
- Připravena koncepce interního testování
 - Vícenásobné headnody pro rozložení zátěže
 - Experimentální nasazení NFSv4
 - Ověření použitelnosti GPFS, VPFSv2 (paralelní scratch)
- Rozhodnutí o nákupu menších kapacit pro EGEE II
- Výběrové řízení na větší diskové kapacity

Vysokorychlostní síť

- Plány úzce propojeny s 615
- Experimentální 10 GE testbed
- Potřeba dokoupit značné množství 10 GE switchovaných portů
- Využito při demu na GLIFu
 - Postupný převod do provozu
 - Diskuse o vhodné kombinaci provozních a experimentálních požadavků

Nové investice

- Výpočetní zdroje
 - Specializované systémy
 - Náhrada skurutů v Praze/CESNET
 - Nový 16core stroj
- Diskové kapacity
 - Experimentální stroje
 - Kapacity pro EGEE
 - Nový diskový subsystém
- Síť
 - 10 GE Myrinet karty
 - 10 GE switche

Výpočetní kapacity

- Specializované systémy – GLIF demonstrace
 - 3 ks 1U serverů: dual-core dual Opteron 2,6 GHz, 4 GB RAM
 - Počítač na CineGRID demo: dual-core dual Opteron, 8800GTX
 - Dva řídicí počítače k diskovým polím
- Výpočetní výkon
 - 10 ks quad-core dual Intel, 16 GB RAM, 2*500 GB HDD
 - Společně s MediGRIDem nový 16core Opteron stroj

Diskové kapacity

- Experimentální systém
 - 3 ks 2U serverů pro experimenty na 10 GE páteři
- Dokoupeno 6 ks 500 GB disků do clusteru v Brně
- EGEE II diskové kapacity
 - Dvě disková pole 16 TB
- Nový diskový subsystém
 - Ve výběru, mid to high range systém značkového dodavatele s možností růstu
 - Propustnost alespoň 750 MB/s (reálně i více jak dvojnásobná)
 - Maximální kapacita až 120 TB

Síťová rozhraní

- Dva switche S2410 Force10
 - Až 24 10 GE portů, aktuálně osazeno 13 a 16 portů
- Šest 10 GE karet Myrinet

Další výsledky

- HA server
- Bezpečnost
 - Řada publikací
 - * Včetně vyžádaného příspěvku do odborného IEEE časopisu
 - Propagace a skutečné nasazení federativního přístupu
- Budování kontaktů na Taiwan (zajímavá 10 GE linka Amsterdam–Taipei)
- 1000+1 věc dle potřeby
 - Bereme jako součást naší práce

Mezinárodní aktivity

■ EGEE II

- Středně velký partner (příspěvek 250 tisíc Euro/rok)
- Účast na většině aktivit
- Zástup střední Evropy v Project Management Boardu

■ EGI_DS

- Člen konsorcia 7 partnerů
- Odpovědnost za grid use cases
- Definice funkcionalit

Nové projekty

- Příprava EGI_DS – přijat, před podpisem
- EGEE III
 - Intenzivní zapojení do přípravy
 - 23. října proběhly hearings
 - Nevíme rozpočet (návrh pro CESNET 300 tisíc Euro/rok)
- EUAsiaGrid
 - Práce na přípravě projektu
 - Návrh na vedení Aplikační work-package
 - Stav projektu zatím neznám

Shrnutí

■ Virtualizace

- Zcela nové prostředí
- Netriviální problémy se stabilizací
 - * Nové komplexní interakce
 - * Odhalování nedostatků Xenu
 - * Interakce s vysokorychlostními rozhraními nedokonalá
- Ministerské cíle budou splněny
- Některé interní bude třeba prodloužit do 2008

■ IPv6

- Podstatně horší situace než jsme očekávali
- To na druhé straně znamená, že děláme něco skutečně nového

Shrnutí II

- Lineární rozvoj provozu a uživatelské podpory
- Zatížení infrastruktury
 - Přes přidané kapacity roste doba čekání úloh
 - Narazili jsme již na limity multilicence MATLABu
- Úložné kapacity
 - Složitý úkol na delší období
- Adekvátní zapojení do mezinárodních aktivit/projektů
- Velmi úzká spolupráce s aktivitu 615

Plán na další období

- Pokračovat ve virtualizaci kapacit
 - Ověřit v provozu hypotézu schopnosti poskytovat různé prostředí na sdílené infrastruktuře
 - Vyřešit vysokorychlostní přístup k síti
 - * I za cenu režie v podobě dedikovaného jádra (quad-core)
 - Další rozvoj systému Magratea
- Uzavřít první etapu budování úložných kapacit
 - Koncepce, protokoly, nástroje, hardware
- Národní gridové prostředí
 - eInfrastruktura ČR
 - Vazba na Evropské strukturální fondy