

Všechno, co jste chtěli vědět o Gridech, ale báli jste se zeptat

Jan Kmuníček
Ústav výpočetní techniky MU & CESNET



Obsah



Charakteristika gridového prostředí

Typy Gridů

Vize výpočetního Gridu

Typy výpočetních problémů

Témata k diskuzi

Charakteristika gridového prostředí



- Definice Gridu

§ „A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities.“

§ **„Výpočetní grid je hardwarová a softwarová infrastruktura, která poskytuje spolehlivý, standardizovaný, všudypřítomný a levný přístup ke špičkovým výpočetním službám. “**

(I. Foster, C. Kesselman. -The Grid: Blueprint for a New Computing Infrastructure, 1998)

§ zpřesněno roku 2002:

- koordinuje zdroje nepodléhající centralizované správě
- používá standardní, otevřené, obecné protokoly a rozhraní
- poskytuje netriviální kvalitu i kvantitu služeb
(více než jednotlivé části každá zvlášť)

Charakteristika gridového prostředí



- Grid versus distribuované prostředí
 - § Distribuované aplikace jako specializované systémy pro jediný cíl nebo skupinu uživatelů.
 - § Gridy rozšiřují původní distribuované aplikace o nové vlastnosti:
 - různé druhy zdrojů
 - nevyžadován striktně stejný hardware, data a aplikace
 - různé druhy interakcí
 - **různé** uživatelské **skupiny a aplikace interagují** s Gridem **různě**
 - **dynamická povaha** (zdroje a uživatelé často přibývají / ubývají / mění se)

Charakteristika gridového prostředí

- Vlastnosti Gridu/Gridů

- § velmi početné výpočetní zdroje

- § výpočetní zdroje jsou

- heterogenní
- geograficky separované
- spojeny heterogenními sítěmi
- plně pod kontrolou jejich vlastníků

- § vlastnictví vzájemně nezávislými organizacemi & jedinci

- § vyžadovány různé bezpečnostní požadavky a pravidla

- § vyžadována různá pravidla správy zdrojů

- § potenciální zahrnutí vadných zdrojů



Charakteristika gridového prostředí



- Principy fungování
 - § uživatelé by si měli z Gridu vzít prostředky,
které zrovna potřebují, když je potřebují, tam kde je potřebují
 - procesory resp. jejich výkon
 - diskový prostor
 - přenosová kapacita sítí
 - speciální hardware (mikroskopy, senzory, ...)
 - § Grid zajistí
 - unifikovaný přístup
 - bezpečnost
 - spolehlivou dodávku
 - vyúčtování za spotřebu

Typy Gridů



- Základní klasifikace Gridů

- § výpočetní Gridy

- **virtuální superpočítač** pro řešení náročných aplikací, který dynamicky agreguje výpočetní kapacitu velkého počtu individuálních počítačů s cílem poskytnout platformu pro řešení náročných aplikací, které není možno řešit pomocí jediného systému

- § datové / informační / znalostní Gridy

- **sdílení velkého množství dat**
 - informační a znalostní Gridy rozšiřují možnosti datových Gridů o poskytování kategorizace dat, ontologií, sdílení znalostí a tvorby workflow

- § kolaborativní Gridy

- **virtuální prostředí pro spolupráci** / virtuální laboratoře
 - vzdálená kontrola a správa vybavení, senzorů a zařízení

Vize výpočetního Gridu



- Základní fakta
 - § výraz výpočetní Grid („**computational grid**“) jako analogie výrazu „power grid“, tj. elektrická rozvodná soustava
 - § pojem zavedli v roce 1998 Carl Kesselman a Ian Foster, kteří srovnávají současný stav počítačů se stavem elektřiny kolem roku 1910
 - každá budova měla vlastní generátor elektřiny, zavedení bylo drahé, využití neefektivní
 - rozvoj způsobilo až zavedení elektráren a rozvodné sítě
 - analogicky dnes má každá organizace vlastní výpočetní prostředky a neumí je efektivně sdílet

Typy výpočetních problémů



- Gridy jsou obecně vhodné pro **paralelní (současné) zpracování úloh**.
- Paralelní výpočetní úlohy takové úlohy, jejichž
 - § zpracování na jediném CPU trvá neúnosně dlouho
 - § celkové paměťové nároky jsou příliš velké
- Primární aplikační určení Gridů:
 - § velké množství malých, navzájem nezávislých úloh
 - příklad: parametrické studie
 - § zpracování rozsáhlých datových souborů „po částech“
 - příklad: digitální vyhlazení satelitních snímků

Typy výpočetních problémů



- **Granularita výpočetních problémů**
 - Výpočty s **velkou granularitou**
 - § každý podproblém nezávislý na ostatních
 - § přenos do gridového prostředí je relativně jednoduchý
 - příklad: Monte Carlo simulace, kde se obměňují parametry komplexního modelu reálného systému a výsledky se studují statistickými technikami – jedná se druh výpočetního experimentu
 - § vhodné řešit pomocí sítě **"volně vázaných" (loosely coupled)** počítačů, jelikož prodlevy při získání výsledků z jednotlivých procesorů neovlivní práci ostatních procesorů
-
- „high-throughput computing“**

Typy výpočetních problémů



- **Granularita výpočetních problémů**
- Výpočty s **malou granularitou**
 - § každý podproblém vysoce závislý na výsledku dalších podproblémů
 - § přenos do gridového prostředí je poměrně složitý
 - § nutno vymyslet algoritmy minimalizující množství a četnost dat, které si mezi sebou procesory vyměňují
 - příklad: výpočet počasí, který může být rozdělen do mnoha malých výpočtů počasí v malých objemech atmosféry
 - § vhodné řešit na velkých paralelních superpočítačích nebo alespoň velmi **„pevně vázaných“ (tightly coupled)** clusterech s procesory propojenými extrémně rychlou sítí
 - **„high-performance computing“**

Typy výpočetních problémů



- **Reálné výpočetní problémy**

- § kombinace obou extrémních typů

- Gridové prostředí je **nejvhodnější** pro řešení takto **komplexních úloh**

- § poskytuje vhodné výpočetní prostředky pro podúlohy s nízkou granularitou

- § jejich výsledky lze kombinovat formou podúloh s vysokou granularitou

- příklad: komplexní modelování podnebí Země
- hledání informace, jak výpočty závisí na jednotlivých parametrech použitých v modelech – spuštění mnoha podobných výpočtů
- každý z nich je paralelní výpočet s malou granularitou
- velké množství v podstatě nezávislých výpočtů může být současně distribuováno na mnoho různých clusterů na Gridu

—————▶ průběh simulace v kratším čase

Témata k diskuzi



- Rozšiřující náměty
 - § Gridy jako nadstavba distribuovaného počítání
 - § Celoevropský Grid založený na clusterových řešeních
 - § Webové / gridové služby
 - § Grid jako prostředí pro programátory / vývojáře
 - § Problémy pro řešení pomocí gridové infrastruktury