

Gridy jako klíčový fenomén informačních technologií nového tisíciletí

Jan Kmuníček, ÚVT MU

Již několik let se stále častěji a ve větší míře setkáváme s pojmem **Grid**. Proto je jisté na místě ujasnit si, je-li tento trend pouze marketingový „hype“ nebo skrývá-li se za ním něco skutečně přelomového. Následující text byl sepsán v naději, že vás přesvědčí o tom druhém.

1 Původ a stručná historie

Původ Gridů lze nalézt v počátcích distribuovaného clusterového počítání. Jako cluster přitom rozumíme spojení dvou a více počítačů (ve vztahu k softwaru i hardwaru) takovým způsobem, že se chovají jako jediný počítač. Clusterová řešení umožňují paralelní zpracování výpočetních úloh, vyrovnávání zátěže jednotlivých individuálních výpočetních zdrojů a toleranci chyb. Gridové počítání vychází z distribuovaného clusterového řešení, ale současně jej povyšuje na novou kvalitativně vyšší úroveň. V Gridu lze propojit cluster (a nejen je, výpočetním elementem mohou být i individuální superpočítače), které jsou geograficky vzdálené, a to takovým způsobem, aby poskytovaly široký fond výpočetní síly a úložného prostoru.

2 Co je to Grid?

Podat krátkou a přitom zcela vyčerpávající definici pojmu Grid není triviální úkol, neboť existuje velká řada různých Gridů. Nicméně obecně lze říct, že Grid je rozsáhlý distribuovaný systém organizačně samostatných elementů, kterými mohou být individuální počítače (paměti, pevné disky) a informační systémy vzájemně propojeny počítačovou sítí.

Všeobecně lze Grid vnímat podobně jako Internet. Internet spojuje miliony počítačů v celém světě a zpřístupňuje uživatelům informace dostupné na veřejných webových stránkách, umožňuje e-mailovou komunikaci, instant messaging, apod. Oproti tomu Grid (neexistuje jeden uniformní Grid, nýbrž celá řada cíleně orientovaných specifických Gridů) umožňuje uživatelům

pomocí specifických mechanismů (bezpečnost, plánování zdrojů, autentizace, autorizace, účtování, správa dat, monitorování, ...) využít výpočetní síly a diskové kapacity, které jsou nezbytné pro běh výpočetně náročných aplikací.

Pravděpodobně jedinou správnou možností, jak obecně popsat Grid resp. provést základní charakteristiku gridového prostředí, je uvést jednu z prvních definic pojmu Grid (výpočetní Grid), kterou použili I. Foster a C. Kesselman ve své knize [1], která se posléze stala neoficiální biblií gridového počítání:

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

Výraz výpočetní Grid (“computational grid”) byl víceméně použit jako analogie výrazu „power grid“, tj. elektrické rozvodné soustavy. Oba autoři srovnávali stav počítačů v roce 1999 se stavem elektriny kolem roku 1910. Došli k závěru, že obdobně jako v roce 1910, kdy každá budova měla vlastní generátor elektriny, její zavedení bylo drahé, využití neefektivní a skutečný rozvoj způsobilo až zavedení elektráren a rozvodné sítě, tak analogicky (v roce 1999) měla každá organizace vlastní výpočetní prostředky, ale neuměla je efektivně sdílet.

Tato definice byla následně zpřesněna o další klíčové vlastnosti, které musí prostředí, jež má být Gridem, vykazovat:

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

Gridy jsou často mylně interpretovány stejným způsobem jako standardní distribuované prostředí pro paralelní zpracování výpočetních úloh. Principiálně se však Grid/Gridy od distribuovaného prostředí liší především tím, že distribuované aplikace jsou vesměs chápány jako specializované systémy pro jediný cíl nebo skupinu uživatelů, zatímco Gridy jsou univerzální platforma. Ačkoliv Gridy staví na distribuovaném prostředí, tak jak je známe např. z homogenních

výpočetních clusterů, přináší ovšem kvalitativně nové vlastnosti. Gridy rozšiřují původní distribuované aplikace o nové vlastnosti:

1. různé druhy zdrojů
2. není vyžadován striktně stejný hardware, data a aplikace
3. různé druhy interakcí
4. různé uživatelské skupiny a aplikace interagují s Gridem různě
5. dynamická povaha (zdroje a uživatelé často přibývají/ubývají/mění se)

Nejjednodušeji se lze dívat na Grid jako na dynamický, virtuální, výpočetní, informační nebo znalostní systém, tj. soustavu především výkonných počítačů propojených vysokorychlostní sítí, určených pro řešení nejnáročnějších výpočetních a datových problémů.

3 Charakteristika gridového prostředí

Mezi základní vlastnosti charakterizující gridové prostředí patří zejména to, že Gridy (výpočetní) obsahují velmi početné výpočetní zdroje, které jsou

- heterogenní
- geograficky oddělené
- spojeny heterogenními sítěmi
- plně pod kontrolou jejich vlastníků (tj. vlastněny vzájemně nezávislými organizacemi a jedinci)

Pro přístup k výše uvedeným zdrojům jsou vyžadována různá bezpečnostní opatření a různá pravidla správy zdrojů. Současně ovšem může dojít i k zahrnutí potenciálně vadných zdrojů. Princip využívání Gridu je obdobný představě, že uživatelé si z Gridu budou brát prostředky, které zrovna potřebují, a to tehdy, když je potřebují a zrovna tam, kde je potřebují. Takovými prostředky mohou být procesory resp. jejich výkon, diskový prostor, přenosová kapacita sítí nebo speciální hardware (mikroskopy, senzory, ...). Grid naopak zajistí unifikovaný bezpečný přístup k těmto prostředkům, spolehlivou dodávku požadovaných služeb a následné vyúčtování za spotřebu. Proto, aby principy byly reálně použitelné, bylo třeba uvést v život koncept

„virtuálních organizací“, tj. dynamických asociací sdružujících dostupné výpočetní zdroje a jejich uživatele z různých administrativních domén s určitým specifickým společným cílem.

4 Typy Gridů

Principiálně lze rozlišit tři základní kategorie Gridů, a to:

- Výpočetní Grid
- Datový Grid
- Informační/Znalostní Grid

Výpočetní Grid: Grid poskytující výpočetní servis je nazýván výpočetní Grid. Takový Grid poskytuje zabezpečené služby pro spouštění aplikací na distribuovaných výpočetních zdrojích. Jistým způsobem se jedná o 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.

Datový Grid: Zpracování rozsáhlých datových sad pomocí služeb výpočetního Gridu se často nazývá datovým Gridem. Datový grid je charakterizován sdílením velkého množství dat, poskytováním zabezpečeného přístupu k těmto datům a umožněním jejich následné správy. To vše je řešeno formou replikovaných datových katalogů vytvářejících iluzi jednotného hromadného datového úložiště.

Informační/Znalostní Grid: Tento typ Gridů (často také nazýván jako kolaborativní či aplikační Grid) je charakterizován snahou o rozšíření možností datových Gridů o poskytování kategorizace dat, ontologií, sdílení znalostí a tvorby workflow. Nedílnou součástí tohoto typu Gridů jsou virtuální prostředí pro spolupráci resp. virtuální laboratoře umožňující vzdálenou kontrolu a správu vybavení, senzorů a zařízení.

5 Gridové projekty

To, že rozvoj gridových technologií je brán poměrně vážně, dokazuje podpora celého spektra různých gridových projektů ze strany Evropské Unie (EU). Počet podporovaných projektů, ať v 5.

nebo 6. rámcovém programu EU, je poměrně impozantní. Jejich přehled včetně příslušných odkazů je dostupný na webu¹. Obecně lze říct, že EU investuje nemalé finanční zdroje jak do infrastrukturních projektů (Damien, DataTAG, DEISA, SEE-GRID, EGEE) s cílem vytvořit jednotnou gridovou infrastrukturu; výzkumných projektů, jejichž cílem je rozvoj gridového middlewaru a gridových nástrojů (GRIP, Eurogrid, GridLab, DataGrid), tak i projektů aplikačně orientovaných (GRIA, EGSO, Crossgrid).

6 Gridové aplikace

Neméně zajímavou informací je přehled aplikačních oblastí, pro něž Gridy představují onen magický „grál“, který otevírá cestu k doposud ne-realizovatelným aplikačním úlohám (ať již z důvodu náročnosti na hardware či množství dat).

Obecně platí, že Gridy jsou vhodné pro paralelní (současné) zpracování úloh. Paralelní výpočetní úlohy jsou takové úlohy, jejichž zpracování na jediném procesoru trvá neúnosně dlouho nebo takové, jejichž celkové paměťové nároky jsou příliš velké. Gridy jsou aplikačně určeny jak pro velké množství malých, navzájem nezávislých úloh (např. parametrické studie) tak pro zpracování rozsáhlých datových souborů (např. digitální vyhlazení satelitních snímků).

Kromě řešení extrémně výpočetně náročných problémů slouží gridová prostředí jako řešení datových úložišť, umožňují přístup ke speciálnímu hardwaru a především zprostředkovávají efektivní využití dostupných zdrojů.

Pomineme-li nasazení gridových řešení v komerční sféře především k proprietárním řešením datově a výpočetně náročných operací (např. zpracování komplexních finančních a investičních modelů a jejich analýza, vládní aplikace apod.), existuje široké spektrum aplikačních oblastí v nichž může být právě gridově orientované řešení velmi vhodné. Z široké škály možností lze jako klíčové vybrat následující (poslední z nich spíše jako ukázkou možností):

1. Chemie a biologie

Využití Gridů pro analýzu a modelování chemických a biologických informací. Jedná se

především o realizaci výpočetních úloh z oblasti výpočetní chemie a molekulového modelování (např. generování 3D molekulárních struktur, molekulárně mechanické a kvantově chemické výpočty vlastností molekul a molekulárních deskriptorů) včetně zpracování bioinformatických dat.

2. Částicová fyzika

Distribuce dat z urychlovačů částic a jejich následná analýza.

3. Inženýrství a design

Výpočet a analýza modelů pomocí technik fluidní dynamiky a výpočetní pevné mechaniky včetně materiálových simulací.

4. Lékařství

Produkce interaktivních medicínských simulací, analýza a správa lékařských obrazů podpora virtuální spolupráce v e-nemocnicích.

5. Astronomie

Analýza terabajtů astronomických dat z teleskopů.

6. Životní prostředí

Příprava modelů a vyhodnocování monitorování znečištění prostředí, ukládání jaderného odpadu a předpovědi počasí.

7. Média

Produkce, vysílání a přehrávání interaktivního mediálního obsahu (audio, video, obraz) v reálném čase.

8. Spolupráce ve výzkumu

Budování virtuálních laboratoří umožňujících pokročilý, datově a výpočetně intenzivní výzkum.

9. Kolaborativní hry

Jako náhrada existujících serverů pro on-line hry za vysoce paralelní, masivně multiplayerové servery.

7 Gridy po česku

V České republice je gridová infrastruktura potenciálním uživatelům dostupná jednak pod záštitou projektu *META Centrum*² (jedna z klíčových aktivit instituce CESNET) a aktuálně také formou nově vybudované Virtuální organizace pro střední Evropu (VOCE).

¹<http://egee.cesnet.cz/cs/grid/links.htm>

²<http://meta.cesnet.cz>

Projekt *META Centrum* zastřešuje většinu aktivit souvisejících v České republice s Gridy, superpočítači, klastrovým nebo gridovým počítáním a/nebo výkonným počítáním obecně. Hlavním cílem projektu *META Centrum*, jehož počátky spadají do roku 1996 – tedy do doby, kdy pojem Grid nebyl prozatím ani formálně ustaven natož rutinně používán – je vytvoření virtuálního počítače, který umožní efektivní využití techniky instalované v rámci superpočítačového projektu, a současně umožní řešit výpočetní úlohy, které svými požadavky (na paměť, výkon centrálního procesoru, ...) přesahují možnosti jednotlivých dílčích superpočítačových center (uzlů). Cílem projektu *META Centrum* je umožnit uživatelům *META Centra* uniformní a jednotný přístup ke všem zdrojům bez nutnosti opakovaného přihlašování, a to i přes rozdíly v producentech hardwaru, operačních systémech i fyzickém umístění jednotlivých počítačů.

Zkušenosti z již téměř deseti let aktivního provozu *META Centra* jasně ukazují, že struktura *META Centra* je dostatečně flexibilní na to, aby v případě zájmu dalších výpočetních subjektů (uzlů) mohly tyto být jednoduše a účelně integrovány do stávajícího systému *META Centra* a byl tak zajištěn trvalý nárůst výpočetních kapacit.

Druhou dostupnou možností pro zájemce o využití Gridů v České republice je **prostředí VOCE** – Virtuální organizace pro Střední Evropu. Toto prostředí je budováno v rámci české účasti³ v projektu EGEE (Enabling Grids for E-Science) budujícího panevropskou gridovou infrastrukturu⁴ a aktuálně poskytuje kompletní plně funkční gridovou infrastrukturu pro region celé střední Evropy. Více technických informací společně s popisem vybraných unikátních služeb ve VOCE poskytovaných lze nalézt v článku „Do Gridu snadno a rychle – prostředí VOCE“ v příštím čísle Zpravodaje nebo přímo na stránkách VOCE⁵.

8 Gridy jako výzva

Pokud se mi vás výše uvedenými informacemi podařilo přesvědčit o tom, že budoucnost distri-

buovaného počítání, datových úložišť a souvisejících oblastí se realizuje již dnes a ÚVT MU je toho nejen svědkem, ale díky pracovníkům SCB také přímým aktivním účastníkem, přijměte prosím na závěr pozvání na posezení v GridCafé⁶.

Je naprosto zřejmé, že Gridy a technologie za nimi stojící jsou jedním z fenoménů IT nastupujícího tisíciletí. Zdá se, že po překonání dětských nemocí nastává nyní čas skutečného zúročení potenciálu, který v sobě Gridy skrývají. Jestli a v jaké míře se to po opravdu podaří, je již na vás – potenciálních uživatelích.

Literatura

- [1] I. Foster, C. Kesselman, eds. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann. San Francisco, California. 1999. □

³<http://egee.cesnet.cz>

⁴<http://www.eu-egee.org/>

⁵<http://egee.cesnet.cz/cs/voce/index.html>

⁶<http://gridcafe.web.cern.ch/gridcafe/>