Cesta k petabajtům

autor: **Jan Kmuníček, Luděk Matyska**
rok vydání: **2004**
číslo vydání: **40**

ČR se podílí na budování evropského gridu

Virtuální výpočetní prostředí, doprovázené v poslední době i rozvojem distribuovaných datových skladů obé souhrnně označované jako grid má v České republice základnu budovanou již od roku 1996. Převážná většina aktivit souvisejících s gridy se tehdy soustředila kolem projektu

MetaCentrum, který se z původně samostatného projektu několika vysokých škol stal součástí výzkumného záměru sdružení Cesnet. MetaCentrum se současně účastní i mezinárodních projektů EU, především na jaře úspěšně dokončeného projektu DataGrid či stále běžícího projektu GridLab. V současnosti se tým MetaCentra soustřeďuje na řešení projektu EGEE (Enabling Grids for E-sciencE), který je realizován jako součást Šestého rámcového programu EU. Cílem projektu EGEE bude funkční celoevropská gridová infrastruktura, která umožní evropským vědcům přístup a využití výpočetních a datových zdrojů nemajících prozatím v evropském měřítku obdoby.

Počátky a plány
EGEE byl oficiálně zahájen 1. dubna 2004 jako dvouletý projekt s předpokládaným dalším pokračováním. S rozpočtem přes 35 milionů eur si klade za cíl propojit evropské národní, regionální i tematicky orientované gridy do jednotné evropské gridové infrastruktury. Ta bude již v průběhu řešení projektu k dispozici akademickým zájemcům o výpočetní nebo datové kapacity. Na řešení projektu se podílí 70 institucí ze všech zemí Evropy, z Ruska i USA, o spolupráci mají zájem i asijské země (např. Japonsko a Jižní Korea). Česká republika se projektu účastní opět prostřednictvím sdružení Cesnet.

Celá infrastruktura bude z velké části postavena na clusterových řešeních, čemuž bude odpovídat i portfolio aplikací cílených na tento typ gridu. Koncem roku 2004 má být vlastní gridová infrastruktura postavena na 25 uzlech (centrech) s celkovou kapacitou cca pět tisíc procesorů a 50 TB disků. Na konci dvouletého projektu by mělo být do EGEE gridu zapojeno na 100 center s 50 tisíci procesory a 1 PB (petabajtem, 1015) diskové kapacity.

Celý dvouletý projekt je rozčleněn do tří vzájemně souvisejících oblastí týkajících se vlastních servisních gridových služeb a provozu, dále vzdělávání a šíření informací formou podpůrných aktivit včetně identifikace vhodných kandidátských aplikací a také dalšího vývoje middlewaru, tj. programového vybavení, které odpovídá za propojení jednotlivých uzlů a pomáhá vytvořit uživatelskou iluzi jednotného výpočetního prostředí. Česká republika se podílí na všech těchto základních oblastech, přitom v oblasti vývoje middlewaru dokonce jako jediná ze zemí střední, ale i jižní či jihozápadní Evropy.

Výpočetní problémy
Jaký typ výpočetních problémů je vlastně řešitelný použitím nově vznikající gridové infrastruktury? Nejjednodušší situace nastává tam, kde uživatelé mají velké množství malých, navzájem nezávislých úloh typickým příkladem jsou třeba parametrické studie, kdy se mnohokrát opakuje stejný výpočet, vždy ale s lehce změněnými parametry. Výsledky všech výpočtů se kombinují a použijí pro určení vlivu změny parametru na vlastnosti celého systému.

Pro tyto typy výpočtů je grid optimálním prostředím. Uživatel připraví rozsáhlou dávku ta může být tvořena i milionem nezávislých úloh a gridové prostředí se postará o jejich naplánování, spuštění, sběr výsledků a případně i finální zpracování. Grid se rovněž postará o zhavarované úlohy, které lze v tomto případě automaticky znovu zadat.

Obdobný typ úloh představuje zpracování rozsáhlých datových souborů "po částech". Jednoduchým příkladem může být např. digitální vyhlazení satelitních snímků. Každý snímek může být zpracován samostatně, velké snímky je dokonce možno rozdělit na menší, jednotlivé části rovněž zpracovávat odděleně a výsledek opět spojit do jednoho snímku. Tímto paralelním zpracováním se dostáváme k nejzajímavější oblasti úloh, pro jejichž řešení byly gridy vyvinuty. Touto oblastí jsou obecně rozsáhlé úlohy, jejichž zpracování na jediném procesoru by trvalo neúnosně dlouho, případně jejich celkové paměťové nároky jsou příliš velké. Klasický přístup je řešit tyto úlohy na speciálních a velmi drahých paralelních počítačích. Úlohu v tomto případě současně zpracovává velký počet procesorů, které společně mají přístup k velmi rozsáhlé paměti (je-li k dispozici 1 000 procesorů, každý s 1 GB paměti, dohromady získáme úctyhodný 1 TB vnitřní paměti využitelný jedinou úlohou).

Paralelní počítače však v podstatě nejsou ničím jiným než množinou procesorů a jejich pamětí, propojenou vzájemně velmi rychlou sítí (často navíc se speciálními vlastnostmi). Na určité úrovni abstrakce je ovšem možné totéž realizovat počítačovým clusterem (skupina počítačů propojena "běžnou", byť stále velmi rychlou sítí a umístěná v malém prostoru) a na ještě vyšší úrovni právě gridem (vzdálenost mezi počítači je v tomto případě řádově vyšší než v clusteru)

Problém algoritmů
Mezi výše popsanými možnosti realizace výpočtu však existují významné rozdíly. Programování pro paralelní počítače je poměrně obtížné, protože úlohu je třeba rozdělit tak, aby každý z procesorů byl zaměstnán a procesory na sebe příliš nečekaly. Rychlost přenosu dat uvnitř paralelního počítače je však na druhé straně velmi vysoká, což umožňuje předávat poměrně často i velké objemy dat mezi procesory bez výrazné ztráty celkového výkonu.

Situace v gridovém prostředí je mnohem složitější. Vzdálenost mezi procesory a použitá síť vysokorychlostní internet vede k mnohem vyšší době přenosu dat. Algoritmy a postupy, které fungují na paralelních počítačích, nejsou v gridovém prostředí dostatečně efektivní a musejí být nahrazeny jinými, případně některé typy úloh nelze (alespoň zatím) efektivně na gridu řešit vůbec.

Otázky závislosti
Základním pojmem při paralelizaci úlohy je tzv. granularita problému. Pokud je každý podproblém vysoce závislý na výsledku dalších podproblémů, jedná se o systém s malou granularitou. Jako příklad může posloužit třeba výpočet počasí, který může být rozdělen do

mnoha malých výpočtů počasí v malých objemech atmosféry. Každý z těchto výpočtů je ale silně ovlivněn tím, co se děje v sousedních "sektorech". Ve skutečnosti i změny ve velmi vzdáleném objemu mohou mít velký vliv. Přenos tohoto typu úloh do gridového prostředí je velmi složitý; je přitom nutno vymyslet důmyslné algoritmy minimalizující množství a četnost dat, které si mezi sebou procesory vyměňují.

Na druhé straně spektra granularity stojí výpočty s velkou granularitou, u nichž je každý podproblém nezávislý na ostatních. Jako příklad mohou posloužit simulace označované jako Monte Carlo, při nichž se obměňují parametry komplexního modelu reálného systému a výsledky se studují statistickými technikami jedná se o druh výpočetního experimentu, který je často používán např. ve výpočetní chemii. V takovémto případě může být každý výpočet proveden nezávisle na ostatních. Gridové prostředí přirozeně preferuje úlohy s vysokou granularitou. Většina zajímavých vědeckých problémů však zpravidla pro řešení potřebuje kombinaci úloh obou typů, jen výjimečně jsou k dispozici homogenní algoritmy s vysokou granularitou. Právě pro tento typ úloh však může být gridové prostředí nejvhodnější, protože poskytuje vhodné výpočetní prostředky pro podúlohy s nízkou granularitou a jejich výsledky lze zpravidla kombinovat formou podúloh s vysokou granularitou.

Aplikace
Předpokládá se, že velký podíl na využití formující se infrastruktury budou mít v první fázi zejména aplikace z oblasti fyziky elementárních částic (HEP, High Energy Physics), kterou v ČR reprezentuje především Fyzikální ústav Akademie věd. Pro tuto uživatelskou komunitu je grid nezbytným a v podstatě také jediným řešením, jak ukládat a zpracovávat data v řádech PB, které jsou produkovány při experimentech částicové fyziky. Současně grid umožňuje spolupracovat ve skutečně celosvětovém měřítku ostatně světový web má rovněž svůj počátek ve švýcarském CERNu.

Jedním z klíčových cílů projektu EGEE je současně rozšíření uživatelské základny na co největší počet potenciálně vhodných aplikací. Portfolio problémů řešitelných pomocí výkonných gridových systémů je poměrně široké, nasazení v gridovém prostředí však většinou brání nedostatečná připravenost na obou stranách jak odborníků na gridy, tak i vlastních uživatelů. Vezmeme-li v úvahu výše popsané typy výpočetních problémů, charakteristické nasazení zahrnuje náročné výpočetně-chemické simulace biologicky důležitých systémů, zpracování bioinfomatických či lékařských dat a materiálově orientované výpočty a testy. Rostoucí zájem přichází také z oblasti monitorování a zpracování dat z dálkového průzkumu země, zpracování a distribuce dat z astronomických měření a experimentů částicové fyziky; v poslední době se přidává real-time zpracování videa a obrazu.

Podrobnější informace o projektu lze získat např. na adrese egee.cesnet.cz.

Rozdělení projektu EGEE
Veškeré aplikace využívající infrastrukturu EGEE by měly spadat do jedné ze 3 následujících kategorií.

Pilotní aplikace, sloužící pro prvotní testování implementace programového vybavení EGEE a ověření funkčnosti a výkonu gridového prostředí. V první fázi řešení projektu byl první pilotní aplikací výpočetní grid pro velký hadronový urychlovač částic (LCG, Large Hadron Collider

Computational Grid) sloužící jako model pro prostředí umožňující zpracování dat v řádech PB z experimentů částicové fyziky v CERNu. Druhou pilotní aplikací tvoří biomedicínské gridy; řada lékařských komunit je vystavena rostoucím výpočetním výzvám, jako je dolování dat z genomických databází nebo indexování lékařských databází v nemocnicích, jejichž datový obsah dosahuje mnoha TB dat pro jednu nemocnici ročně.

Interní aplikace, obsahující projekty, jejichž řešitelé se podílejí na projektu EGEE a současně spolupracují s institucemi mimo EGEE. V praxi se většinou jedná o národní projekty s nemalou zkušeností v oblasti gridových výpočtů. Externí aplikace, vyžadující explicitní podporu vybraných uživatelských skupin tak, aby jejich aplikace byly přeneseny do gridového prostředí, které EGEE buduje, a současně aby uživatelé získali dostatečné znalosti pro jejich efektivní využití. Součástí projektu je proto i rozsáhlá školicí aktivita.

Inženýrský přístup a zemské klima

Jako reálný příklad výpočtů realizovaných v gridů může posloužit komplexní modelování podnebí Země, při němž je snahou vědců zjistit, jak výpočty závisejí na jednotlivých parametrech použitých v jejich modelech. V podstatě tedy potřebují spustit mnoho podobných výpočtů. Každý z nich odpovídá paralelnímu výpočtu s malou granularitou, neboť předpověď podnebí je podobná způsobu předpovědi počasí ovšem na mnohem delší časové škále, což nakonec ústí do potřeby spustit každý výpočet na superpočítači nebo jednom klastru. Avšak mnoho v podstatě nezávislých výpočtů může být současně distribuováno na mnoho různých clusterů na gridu. Celá simulace tak proběhne v mnohem kratším celkovém čase. Vhodná dekompozice úlohy a následné přiřazení podúloh nejlépe odpovídajícím součástem gridu je oblastí, která se dnes pomalu posunuje od umění zvládaného jen malou skupinou zasvěcených přes "řemeslo" (které ovšem stále ještě vyžaduje příliš mnoho lidské práce) k inženýrské disciplíně, jejíž nástroje jsou dostupné pro stále rostoucí skupinu uživatelů.

Granularita a typ výpočtů

Míra vzájemné nezávislosti jednotlivých podproblémů určité úlohy se označuje jako granularita. Výpočty s malou granularitou je zpravidla vhodné řešit na velkých paralelních superpočítačích nebo alespoň velmi "pevně vázaných" (tightly coupled) clusterech s procesory propojenými extrémně rychlou sítí. Tento typ výpočtů se často popisuje jako "high-performance computing". Na druhé straně výpočty s vysokou granularitou je ideální řešit pomocí sítě "volně vázaných" (loosely coupled) počítačů, jelikož prodlevy při získání výsledků z jednotlivých procesorů neovlivní práci ostatních procesorů. Tento typ výpočtů se často popisuje jako "high-throughput computing".