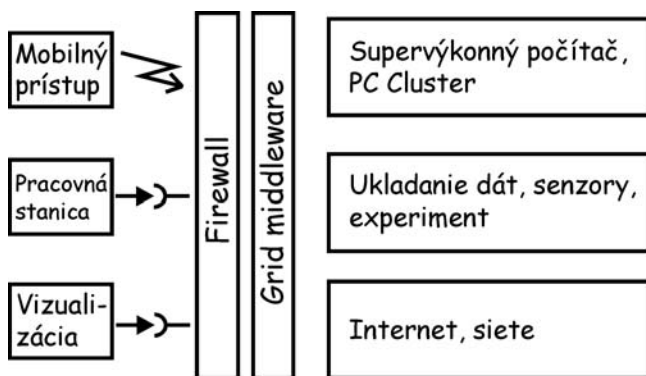


Využitie gridovej infraštruktúry EGEE vo farmácii a biomedicíne

Pojem grid sa prvýkrát objavil v septembri roku 1997, keď ho Ian Foster a Carl Kesselman použili na označenie novej infraštruktúry pre vedu 21. storočia. Grid je považovaný za ďalšiu evolučnú fázu vo vývoji distribuovaného počítania. Zatiaľ čo internetové a webové technológie umožňujú prístup k distribuovaným informáciám, gridové technológie umožňujú prístup k výpočtovým a pamäťovým prostriedkom priestorovo distribuovaným po celom svete.

Úvod

Na grid sa možno pozeráť ako na globálnu sieť počítačov reprezentujúcich jedno obrovské výpočtové prostredie. Tým, že poskytuje bezpečné, vysokovýkonné metódy na prístup k vzdialeným prostriedkom, umožňuje rôznym skupinám, inštitúciám a jednotlivcom, tzv. virtuálnym organizáciám, tieto prostriedky spoločne používať. Pritom pod pojmom prostriedok sa v tomto prípade chápu nielen samotné výpočtové prostriedky, ale napr. aj znalosti, vedecké expertízy, medzinárodné archívy a pod.



Obr.1 Základné prvky gridu

Elektronická veda je založená na spolupráci, ktorá je umožnená spoločným používaním prostriedkov cez internet, často je výpočtovo náročná a pracuje s obrovským množstvom dát (vytváranie nových dát alebo prístup k veľkým dátovým súborom) a prekračuje organizačné aj administratívne hranice. Gridová technológia umožňuje spoločné používanie distribuovaných prostriedkov [1] bez toho, aby sme sa vopred museli zaujímať o to, kde sa nachádzajú a kto ich vlastní. Na obr. 1 sú znázornené základné prvky gridu.

Na podporu jednotného a bezpečného prístupu k distribuovaným výpočtovým prostriedkom, dátam a nástrojom boli vyvinuté rôzne gridové služby [2]. Umožňujú vyhľadanie geograficky a organizačne distribuovaných výpočtových prostriedkov a dát, spravovanie úloh a bezpečnú komunikáciu medzi procesmi.

Výpočtové gridy, ktoré poskytujú výpočtové služby, možno považovať za virtuálne superpočítače, ktoré dynamicky spájajú veľké množstvo jednotlivých počítačov.

Dátové gridy (ale patria sem aj informačné a znalostné gridy) umožňujú spoločné využívanie dát a slúžia na podporu kategorizácie dát, vyhľadanie informácií, ontológie a spoločné využívanie znalostí.

Gridy na podporu spolupráce vytvárajú virtuálne prostredie, ktoré umožňuje geograficky vzdialeným jednotlivcom a skupinám spolupracovať, alebo na diaľku riadiť rôzne nástroje, senzory a pod.

Spolupráca v biomedicíne verus telemedicíne: napríklad o spolupráci v biomedicíne hovoríme, ak spolupracujú profesionáli, ktorí sa nachádzajú na rôznych miestach, pričom prítomnosť pacienta nie je rozhodujúca. Naopak telemedicina obvykle zahŕňa prenos dát priamo medzi pacientom a profesionálom.

Grid umožňuje integráciu prostriedkov a znalostí a súčasne dekompozíciu funkcií a úloh. Kľúčom ku gridovej infraštruktúre je servisne orientovaná architektúra, volne prístupný softvér a otvorené štandardy.

Pre oblasť medicínskych snímok je všeobecne prijatým štandardom DICOM (Digital Imaging and Communication in Medicine) [4], ktorý stanovuje spôsob komunikácie rôznych zariadení, a teda umožňuje ich komunikáciu na diaľku cez sieť. Formát používa dohodnutý slovník (ontológiu) a možno ho rozšíriť pre rôzne oblasti medicíny. Formát DICOM v súčasnosti používajú kardiológovia, dentisti, gastro-enterológovia a iní.

Sprístupnenie gridu pre elektronickú vedu

Projekt **Sprístupnenie gridu pre elektronickú vedu** (Enabling Grids for E-sciencE) [5] spája vedcov a inžinierov z viac ako 90 inštitúcií v 32 krajinách sveta s cieľom poskytnúť spoločnú gridovú infraštruktúru pre e-vedu, prístupnú pre vedcov 24 hodín denne. Na začiatku bol koncipovaný ako 4-ročný projekt, druhá dvojročná fáza (EGEE-II) sa začala 1. apríla 2006. Projekt je financovaný Európskou komisiou. Projekt sa začal prácou dvoch vedeckých skupín – fyzika vysokých energií (High Energy Physics, HEP) a biomedicina, ale keď sa dostal do svojej druhej fázy, rozšíril sa o podporu takých rôznorodých výskumných oblastí, ako sú multimédiá, astrofyzika, archeológia či výpočtová chémia. Výskumníci vytvárajú virtuálne organizácie (VO), ktoré im umožňujú spolupracovať, spoločne využívať prostriedky a prístupovať k spoločným dátovým sadám prostredníctvom gridovej infraštruktúry EGEE. Projekt sa sústreďuje hlavne na tri základné oblasti:

- Prvou je budovanie konzistentnej, robustnej a bezpečnej gridovej siete, ktorá pritiahne ďalšie výpočtové prostriedky.
- Druhou oblasťou je neustále skvalitňovanie a udržiavanie middleware v snahe poskytnúť používateľom spoľahlivé služby.
- Tretou oblasťou je získavanie nových používateľov z priemyslu, ako aj z vedy a výskumu a poskytnúť im na vysokej úrovni potrebné školenia a podporu.

Projekt EGEE je pevne orientovaný na úzke spojenie s priemyslom. Ponúka viacero spôsobov, ako sa zapojiť, od spoločného využívania informácií až po plnú technickú spoluprácu, ako aj účasť v úvodných fázach plánovania trvalej gridovej infraštruktúry, ktorá bude nasledovať po EGEE. Integrovanou súčasťou EGEE je Priemyselné fórum (PF) EGEE, ktorého účelom je poskytnúť priemyslu príležitosť oboznámiť sa s EGEE a s najnovším vývojom v oblasti gridov a tiež poskytnúť priemyslu prostriedok na odovzdávanie požiadaviek na projekt s cieľom zaisťiť priemyselnú relevanciu projektu EGEE. Okrem stretnutí a účasti na takých akciách, ako sú EGEE konferencie, PF organizuje Priemyselné dni a na webovom portáli EGEE vedie samostatnú sekciu venovanú záujmom a potrebám priemyslu.

Biomedicina patrí medzi hlavné aplikačné oblasti projektu EGEE [6]. Z tejto oblasti už bolo prenesených na grid alebo je v štádiu prenosu 23 aplikácií, ktoré možno rozdeliť do troch podoblastí: spracovanie medicínskych snímok, biomedicina a vývoj liekov. V každej z týchto podoblastí už bolo na infraštruktúre EGEE realizovaných mnoho individuálnych aplikácií. Tieto aplikácie kladú na middleware (strednú vrstvu gridového výpočtového prostredia) špecifické požiadavky, najmä na bezpečnosť (citlivé dáta), spravovanie dát (zložité štruktúry dát a di-

stribuovanosť) a spúšťanie veľkého počtu malých, ale dátovo intenzívnych úloh. Biomedicínske aplikácie sú v súčasnosti pravidelnými používateľmi infraštruktúry (približne 15 000 úloh spúšťaných mesačne). Biomedicínske aplikácie sú v súčasnosti rozvinuté na infraštruktúre EGEE:

Oblasť medicínskeho spracovania obrazov sa týka počítačovej analýzy digitálnych medicínskych snímok. Obsahuje federáciu medicínskych dát, výpočtovo náročné medicínske procedúry, spracovanie veľkých dátových množín a štatistické štúdie nad veľkými populáciami dát.

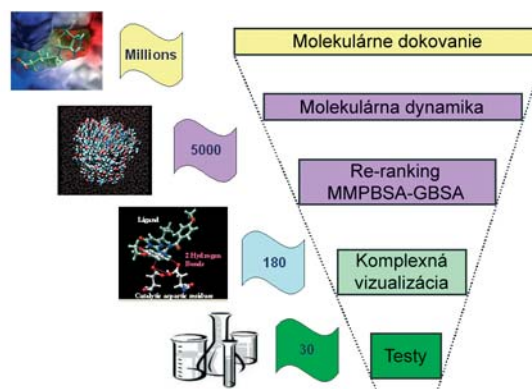
- **GATE** je simulátor založený na metóde Monte Carlo a určený na plánovanie rádioterapie na základe snímok pacienta. Gridovú infraštruktúru EGEE využíva na skrátenie času potrebného na simuláciu Monte Carlo na hodnotu prijateľnú na klinické použitie.
- Klinický systém na podporu rozhodovania (Clinical Decision Support System CDSS) využíva klasifikáciu obrazov, založenú na expertných znalostiach na podporu klinického rozhodovania. Grid sa využíva na zbieranie veľkých dátových súborov a na účinný tréning klasifikačného softvéru nad týmito rozsiahlymi množinami dát.
- Aplikácia **Pharmacokinetics** študuje difúziu kontrastnej látky v pečeni na základe postupnosti snímok magnetickej rezonancie, keďže artefakty spôsobené pohybom pacienta znemožňujú priame porovnanie snímok. Výpočet paralelizovanej ko-registrácie snímok, ktorý beží na gride, umožňuje analýzu takejto postupnosti v rozumnom čase.
- **SIMRI3D** je simulácia snímok magnetickej rezonancie s cieľom vytvoriť umelé, ale realistické 3D snímky magnetickej rezonancie (MR), ktoré umožnia analýzu snímok zo známych zdrojov, štúdium artefaktov a optimalizáciu pracovného postupu.
- Aplikácia **gPTM3D** umožňuje interaktívnu rekonštrukciu medicínskych 3D snímok, napr. na objemovú rekonštrukciu veľkých alebo zložitých orgánov. Aby bolo možné dosiahnuť kvalitu služby, ktorá je požadovaná pre interaktivitu, niektoré miesta na gride musia mať pre túto triedu úloh definovanú vysokú prioritu.
- **Bronze Standard** je aplikácia na vyhodnotenie algoritmov na registráciu medicínskych snímok. Množstvo manipulovaných dát a cena výpočtov presahujú možnosti štandardných počítačov, avšak na gride môže byť táto aplikácia distribuovaná prijateľne jednoduchým spôsobom.
- Softvérový balík **SPM** využíva neurologická výskumná komunita na včasnú diagnostiku Alzheimerovej choroby. Je založený na porovnaní vyšetreného prípadu s veľkou množinou normálnych prípadov. Gridové technológie umožňujú ľahký prístup k distribuovaným dátam, ako aj k distribuovaným výpočtovým prostriedkom.

Oblasť bioinformatiky sa týka analýzy génových sekvencií. Obsahuje genomiku, proteomiku a fylogénu.

- Gridová analýza proteínových sekvencií (Grid Protein Sequence Analysis GPS@) je webový portál, ktorý používateľom poskytuje priateľský prístup k bioinformatickým prostriedkom umiestneným na EGEE gride. Online prototyp portálu GPS@ sprístupňuje 13 programov, ktoré sú už k dispozícii na gride, z celkového počtu 46 programov na originálnom portáli.
- **xmipp_MLrefine** sa používa na 3D štruktúrnu analýzu veľkých makromolekulárnych komplexov. V rekonštrukčnom procese sa kombinuje mnoho snímok z elektrónového mikroskopu, ktoré korešpondujú s rôznymi pohľadmi na vzorku. Na nájdenie najpravdepodobnejšieho modelu, ktorý zodpovedá experimentálnym dátam, treba urobiť veľa iterácií, pretože snímky sú často veľmi zašumené.
- Snímky získané z elektrónového mikroskopu sú ovplyvnené mnohými aberáciami. Matematicky je rozdiel medzi teoretickou projekciou a aktuálnou experimentálnou projekciou modelovaný ako prenosová funkcia kontrastu (contrast transfer function CTF). Na určenie aktuálneho tvaru CTF, ktorý ovplyvňuje experimentálne snímky, je potrebná simulačná metóda **Xmipp_assign_multiple_CTFs**.
- **SPLATCHE** (SPatIAL And Temporal Coalescences in Heterogeneous Environment) je celulárny prostriedok na modelovanie rozvoja genómu. Umožňuje rekonštrukciu globálneho rozšírenia ľudí v minulosti v genograficky realistickej krajine a generovanie molekulárnej diverzity rôznych ľudských populácií.

Oblasť objavovania liekov má snahu urýchliť proces hľadania nových liekov prostredníctvom in silico simulácií proteínových štruktúr a dynamiky. Na začiatku in silico virtuálneho skríningu sú milióny komponentov, z ktorých treba vybrať vhodné komponenty na in vitro testovanie. Tento proces je výpočtovo veľmi náročný, ale potenciálne oveľa lacnejší a efektívnejší než klasické in vitro testovanie.

- Aplikácia **WISDOM** (obr. 2) zbieha rozsiahle výpočty na in silico objavovanie liekov proti novým a zanedbávaným chorobám. Výpočty molekulárneho dokovania určujú, ako sa skúmané lieky spájajú so špecifickými miestami cieľového vírusu – snahou je nájsť liek, ktorý bude s najväčšou pravdepodobnosťou aktívny proti vírusu. Tento postup bol úspešne použitý proti malárii a vtáčej chrípke a do budúcnosti sú plánované ďalšie podobné úlohy.
- **GridGRAMM** poskytuje jednoduchý prístup k molekulárnemu dokovaniu na webe. Výsledky obsahujú bodové hodnotenie kvality a rôzne metódy prístupu k 3D štruktúram komplexov. Molekulárne dokovanie možno použiť na štúdium molekulárnych interakcií, na analýzu interakcií enzým – substrát, na návrh liekov a na porozumenie morbidných mutácií.
- Cieľom **GROCK** (Grid Dock) je poskytnúť jednoduchý spôsob na uskutočnenie hromadného vyhľadávania molekulárnych interakcií prostredníctvom webu. Používateľ má možnosť porovnávať jednu molekulu s celou databázou známych štruktúr.



Obr.2 Tok práce v aplikácii WISDOM

Na rýchlom vyhľadaní zlúčenín schopných blokovat činnosť jedného z enzýmov na povrchu chrípkového vírusu vtáčej chrípky H5N1 sa nedávno podieľal aj Ústav informatiky SAV (ÚI SAV). Spolupracujúce laboratória v Ázii a Európe totiž analyzovali až 300 000 možných zložiek liekov proti tomuto vírusu použitím gridovej infraštruktúry EGEE. Cieľom bolo nájsť zlúčeniny schopné blokovat činnosť jedného z enzýmov nachádzajúcich sa na povrchu chrípkového vírusu, tzv. neuraminidázy subtypu N1. Použitie gridu na identifikáciu najslubnejších tipov pre reálne biologické testy môže urýchliť vývoj liekov proti vírusu vtáčej chrípky. Na obr. 2 je znázornený tok práce v aplikácii WISDOM (Wide In-Silico Docking on Malaria), ktorá je jedným z prvých úspešných používateľov gridu. Ukázalo sa, že túto metódu možno použiť takmer na všetky choroby a tým znížiť náklady na vývoj nových liekov a otvoriť nové perspektívne možnosti vo farmaceutickom priemysle.

Jedným zo základných cieľov pôsobenia liekov, ktoré sú dnes na trhu, je vírusová neuraminidáza – enzým, ktorý vírusu pomáha rozširovať sa a infikovať ďalšie bunky. Je známe, že pod vplyvom liekov tento proteín vytvára nové varianty, a preto sa v prípade chrípkovej pandémie potenciálnym problémom stáva jeho rezistentnosť voči liekom.

Výzvou pre in silico (t. j. počítačový) aplikáciu na vývoj liekov je nájsť molekuly, ktoré sú schopné zamedziť činnosti vírusu tým, že sa pripoja k jeho aktívnej strane. Na zistenie vplyvu malých mutácií na liekovú rezistenciu bola preskúmaná veľká množina zlúčenín voči tej istej neuraminidáze, avšak s rôznou, mierne odlišnou štruktúrou. Na základe takéhoto in silico virtuálneho testovania môžu výskumníci predpovedať, ktoré zlúčeniny a chemické látky najefektívnejšie blokuju aktívne neuraminidázy v prípade výskytu mutácií.



Použitie EGEE a spolupracujúcich výpočtových gridových infraštruktúr významne urýchľuje vývoj liekov. Na preskúmanie schopnosti pripojenia sa 300 000 zlúčenín k ôsmim rôznym cieľovým štruktúram neuraminidázy chrípky typu A bolo v apríli roku 2006 počas 4 týždňov použitých 2 000 počítačov – ekvivalent 100 rokov na 1 počítači. Vytvorených a uložených do relačnej databázy bolo viac ako 60 000 výstupných súborov s celkovým objemom 600 gigabytov. Potenciálne zložky liekov proti vtáčej chrípke sa následne identifikujú a hodnotia na základe väzbových energií dokovaných (pripojených) modelov. „Tieto výsledky demonštrujú, že grid je pre vedcov mocným a spoľahlivým nástrojom, otvárajúcim nové možnosti výskumu a zlepšujúcim existujúce metódy,“ povedala Viviane Redingová, európska komisárka pre informačnú spoločnosť a médiá. „Som veľmi rada, že európska vlajková loď gridovej infraštruktúry prispieva k riešeniu takých aktuálnych a sociálne závažných problémov, ako je vtáčia chrípka.“

Gridové výpočty na Slovensku

VOCE je oficiálna virtuálna organizácia (VO) projektu EGEE, ktorá bola vytvorená na podporu výskumníkov zo stredo európskeho regiónu. V rámci EGEE sem patria Česko, Poľsko, Rakúsko, Slovensko, Slovinsko, Maďarsko a Chorvátsko. Výskumníci z týchto krajín, ktorí nie sú členmi inej, profesijnej virtuálnej organizácie EGEE, môžu sa stať členmi VO VOCE a využívať prostriedky a služby produkčnej EGEE infraštruktúry pre svoju výskumnú činnosť.

Centrom poskytovania podpory v gridovej infraštruktúre pre používateľov, ale aj pre administrátorov sú regionálne operačné centrá (Regional Operations Centres ROC). Poskytujú odborné rady a technickú podporu v procese budovania a prevádzkovania infraštruktúry. Regionálne operačné centrum pre stredo európsku federáciu je Cyfronet [3].

Národné gridové infraštruktúry v strednej Európe spájajú národné a regionálne gridové infraštruktúry a vytvárajú silný a bezpečný grid. Ústav informatiky SAV poskytuje 64 CPU do niekoľkých gridových infraštruktúr. Tabuľka na obr. 3 ukazuje VO podporované site-mi v ROC CE. ÚI SAV prevádzkuje sajt (site) a poskytuje používateľskú podporu, ako aj činnosť Slovenskej gridovej certifikačnej autority a vydávanie digitálnych gridových certifikátov pre používateľov a hosťiteľské počítače zo Slovenska.

Obr.3 VO podporované site-mi v ROC CE

Informačný portál EGEE SK [8] (obr. 4) vytvorený v rámci projektu EGEE poskytuje informácie o gridoch (špeciálne o gridovej infraštruktúre EGEE) používateľom z rôznych oblastí (veda, výskum, vzdelávanie, priemysel, vývoj).

Elektronické zdravotníctvo (eHealth) je jednou z priorit akčného plánu e-Európa – programu EÚ, zameraného na internetizáciu európskych občanov a firiem. Moderné zdravotníctvo sa nezaobíde bez moderných technológií. Každá nová technológia, počítačové technológie nevyňímajúc, slúži dobre len tým, ktorí sú na jej používanie dobre pripravení. ÚI SAV, ktorý má viacročné skúsenosti v oblasti gridového počítania, usporiadal v spolupráci s medzinárodným konzorciom EGEE Informačný deň pre elektronické zdravotníctvo (Bratislava, 19. 9. 2007) [7]. Cieľom medzinárodného podujatia s názvom Využitie počítačového gridu vo farmácii a v biomedicíne (Grid for Pharmaceutical&Biomedical

Challenges) bolo informovať o možnostiach využitia moderných výpočtových gridových technológií v oblasti medicíny, farmaceutického priemyslu a ďalších príbuzných oblastí a sprostredkovať prístup reprezentantov slovenského zdravotníctva ku gridovej infraštruktúre EGEE.



Obr.4 EGEE SK portal

Počas Informačného dňa odborníci z rôznych európskych krajín (Francúzsko, Španielsko, Poľsko, Nemecko, UK, ČR, SR) predvádzali, aké sú výhody gridu pre firmy z oblasti farmácie a biotechnológií. Prezentované boli úspešne realizované aplikácie v oblasti farmaceutického priemyslu a zdravotníctva v krajinách EÚ, ako aj aktuálne informácie o súčasnom stave eHealth na Slovensku. Účastníci mali možnosť navštíviť stánok WISDOM a pozrieť si demonštráciu práce tejto aplikácie. (Obr. 2 je prevzatý z tejto prezentácie.)

Záver

ÚI SAV je zapojený do tvorby národných gridových iniciatív EÚ. Riaditeľ ÚI SAV, Ing. Ladislav Hluchý, CSc., má poverenie od ministra školstva SR koordinovať vznik Slovenskej gridovej iniciatívy za účelom jej integrovania s inými národnými gridovými iniciatívami, ktorých cieľom je budovanie elektronickej vedy v krajinách Európskej únie. Špecialisti z ÚI SAV poskytujú odborné konzultácie a pravidelne organizujú školenia pre záujemcov o gridové technológie.

Cieľom projektu EGEE je poskytnúť vedcom, výskumníkom, ale aj priemyslu prístup k rozsiahlym výpočtovým prostriedkom nezávisle od ich geografickej polohy. Ďalším nemenej dôležitým cieľom projektu je aj pritiahnutie nových používateľov gridových technológií. Táto práca je čiastočne podporovaná projektom VEGA 2/7101/27 a projektom EGEE-II pod číslom INFSo-RI-031688.

Literatúra

[1] Boyd, D.: CCLRC e-Science Centre <http://www.e-science.clrc.ac.uk/>

[2] Liu, Brent J. – Zhou, M. Z. – Documet, J.: Utilizing data grid architecture for the backup and recovery of clinical image data. In: Computerized Medical Imaging and Graphics 29 (2005), pp. 95 – 102. www.elsevier.com/locate/comprmedimag

[3] Cyfronet <http://egee.grid.cyfronet.pl/>

[4] DICOM štandard <http://www.leadtools.com/sdk/medical/DI-COM/dicomstdn.htm>

[5] Projekt EGEE <http://www.eu-egee.org/>

[6] Projekt EGEE – Biomedicínske aplikácie <http://www.eu-egee.org/sheets/uk/biomedical-applications.pdf>

[7] Projekt EGEE – Informačný deň pre elektronické zdravotníctvo (EGEE Industry Day in Bratislava) <http://www.ui.sav.sk/egee/BratislavaDay/>

[8] Webový informačný portál EGEE SK <http://www.ui.sav.sk/egee/>

Ing. Kurdel Peter
Ing. Sebestyénová Jolana, PhD.

Ústav informatiky SAV
Dúbravská cesta 9, Bratislava
e-mail: peter.kurdel@savba.sk,
sebestyenova@savba.sk