

Slovenská infraštruktúra pre vysokovýkonné počítanie

The Slovak Infrastructure for High Performance Computing

Tomáš LACKO – Peter HERMANN – Milan ŠUJANSKÝ

Abstract

Present circumstances in Slovakia into areas of High Performance Computing and comparison with foreign countries. Improvement position through develop project for Slovakia from structural fund "The Slovak Infrastructure for High Performance Computing". General purposes project, general architecture, super-computer technology and Gridcomputing. Software for High Performance Computing. Processing results of research needs of application software on university of Slovakia as supporting initiative corporations EUNIS Slovakia.

Keywords

High Performance Computing, Supercomputing, Grid, Gridcomputing

Úvod

Superpočítačové centrá sa začali budovať v 90-rokoch v ekonomicky rozvinutých krajinách, vrátane susedných krajín Poľsko, Rakúsko, Maďarsko a Česko, kde začiatky budovania superpočítačových centier siahajú do rokov 1995-1996. Na Slovensku sme evidovali niekoľko pokusov o vytvorenie takýchto centier (roky 2001 – 2007). Všetky pokusy narazili na nedostatok finančných prostriedkov a nedostatočné organizačné zabezpečenie.

V súčasnosti sa pripravuje projekt vo veľkej miere finančne zabezpečený zo štrukturálnych fondov EU o názve „Slovenská infraštruktúra pre vysokovýkonné počítanie“ (SIVVP), ktorý bude predstavovať významný inovačný prvok v oblasti aplikácie najvýkonnejších výpočtových prostriedkov vo vedeckom výskume a výučbe na Slovensku. Významnou podporou pre dostupnosť vysokovýkonných výpočtových systémov je už vybudovaná transparentná rýchla sieťová infraštruktúra SANET, ktorá navzájom prepája univerzitné a akademické pracoviská a tiež umožňuje prístup do celosvetovej siete INTERNET.

V ďalšom bude venovaná pozornosť súčasnému stavu na Slovensku a v zahraničí, budú podané informácie o pripravovanom projekte SIVVP a o situácii v oblasti predpokladaného aplikačného programového vybavenia.

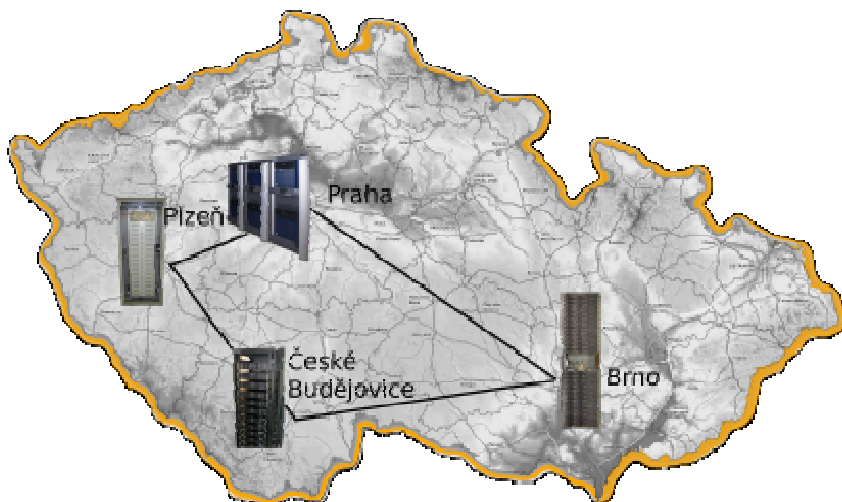
Súčasná situácia v oblasti vysokovýkonného počítania na Slovensku a v zahraničí

V súčasnosti prakticky neexistujú na vedeckovýskumných a univerzitných pracoviskách na Slovensku počítače, ktoré by spĺňali požiadavky kladené na superpočítačové systémy ako sú inštalované na špičkových výskumných a univerzitných pracoviskách nielen v Japonsku, v USA, v západnej Európe, ale aj v Poľsku, Maďarsku a Česku.

Riešiť úlohy s využitím špeciálneho aplikačného programového vybavenia (často získaného v rámci spolupráce so zahraničnými univerzitami a výskumnými pracoviskami) a

tým sa účinne podieľať na rozvoji a aplikácii najnovších poznatkov vedy a techniky u nás nie je v súčasnosti možné, pretože naše špičkové pracoviská nie sú vybavené výkonnou výpočtovou technikou – superpočítačovým systémom. Zaostávanie v tejto oblasti predstavuje v súčasnej dobe minimálne 15 rokov za vyspelými krajinami.

Najjednoduchšie je porovnanie s Českou republikou. V rámci zvláštneho programu akcelerácie vedy a techniky MŠ ČR boli už v roku 1996 zriadené na vybraných univerzitách superpočítačové centrá za vyše 200mil. Kč, ktoré sú navzájom prepojené a organizačne sú začlenené do tzv. MetaCentra s jednotným organizačným vedením pod zastrešením združenia CESNET. Teritoriálne rozloženie výpočtových systémov je zrejmé z obr. 1. So súčasnými možnosťami vysokovýkonného počítania v rámci MetaCentra je možné sa oboznámiť na stránke [1].



Obr. 1

Superpočítanie v Nemecku. Podľa rebríčka TOP500 [2] je Nemecko so svojimi 46 inštaláciami superpočítačov po Spojených štátoch amerických (257) a Veľkej Británii (53) tretou krajinou v poradí čo sa týka počtu inštalácií. Nemecká vedecko-výskumná a univerzitná obec tvorí veľký počet spoločných domácich i medzinárodných projektov riešených v prostredí superpočítačových centier ako aj gridových zoskupení.

Podľa zverejnenej kategorizácie výpočtových systémov typu HPC v Nemecku je možné tieto systémy rozdeliť do troch vrstiev:

1. **vrstva** – cca 100 inštalácií – sú to univerzitné systémy alebo systémy patriace rôznym inštitúciám. Sem patria systémy, z ktorých mnohé síce svojim výpočtovým výkonom nezasahujú do rebríčka TOP500, ale sú pripojené do napr. do D-gridu, EGEEII a pod.
2. **vrstva** – cca 10 inštalácií. Sem patria superpočítače významných výkonov s regionálnym používateľským poslaním, umiestnené v Aachene, Berlíne, DKRZ, Drážďanoch, DWD, Karlsruhe, Hannoveri, MPG/RZG, Paderborne. Sú využívané lokálne alebo prostredníctvom gridov.
3. **vrstva** – tri inštalácie v Jülichu, Garchingu pri Mníchove a Stuttgarte. Tieto centrá sú považované za národné superpočítačové centrá a každé z nich je významné prevádzkovanou architektúrou superpočítača. Používateľ má tak možnosť použiť pre výpočet systém, ktorý je pre danú úlohu vhodnejší buď priamo alebo prostredníctvom siete DEISA.

Všetky tri superpočítačové centrá tvoria zoskupenie GCS (Gauss Centre for Supercomputing), ktoré je v partnerstve s PRACE kde zastupuje záujmy nemeckej superpočítačovej používateľskej komunity.

Ďalší prehľad o situácii v zahraničí je možné získať na internetových stránkach uvedených v nasledovnom prehľade:

DEISA: gridová infraštruktúra pre superpočítačové centrá:	http://www.deisa.eu
PRACE: Partnership for Advanced Computing in Europe	http://www.prace-project.eu/about-prace/partners
Europska Gridova Iniciativa	http://www.egi.org
Nemecko - špičkové stredisko	http://www.fz-juelich.de/jsc/jugene
Poľsko	http://plgrid.pl/partnerzy
Česká republika	http://meta.cesnet.cz/cs/
Maďarsko	http://www.lpds.sztaki.hu/index.php
Linky na niektoré centra v okolitých krajinách	http://www.ui.sav.sk/vvt/?lng=sk&charset=&doc=appl

Ciele projektu „Slovenská infraštruktúra pre vysokovýkonné počítanie“ [3]

Cieľom národného projektu „Slovenská infraštruktúra pre vysokovýkonné počítanie“ je vytvoriť vo vybraných univerzitách a pracoviskách SAV centrum vybavené vysokovýkonným systémom s výkonnosťou porovnateľnou s výpočtovými kapacitami obdobných centier vyspelých krajín. Tieto výpočtové systémy budú pripojené do akademickej siete SANET. Realizáciou národného projektu sa má odstrániť značné zaostávanie Slovenskej republiky v uvedenej oblasti nielen vzhľadom k vyspelým priemyselným krajinám, ale aj vzhľadom k susediacim krajinám našej republiky.

Centralizáciou umiestnenia počítačových systémov s niekoľkokrátovo vyššou kapacitou aká je v súčasnosti inštalovaná (alebo úplne chýbajúca) vo výpočtových strediskách na univerzitách a pracoviskách SAV, bude možné pripraviť pre vedecké a výskumné pracoviská v SR porovnateľné podmienky pre základný a aplikovaný výskum s vyspelými krajinami západnej Európy.

Nárast požiadaviek na riešenie úloh s vysokým stupňom výpočtovej zložitosti v mnohých odvetviach vedy a výskumu nemá už dlhšiu dobu podporu v adekvátnom vybavení vedecko-výskumných a univerzitných pracovísk zodpovedajúcou výkonnou výpočtovou technikou. Úlohy s takýmto stupňom výpočtovej zložitosti sú najmä z týchto vedných oblastí a konkrétnych procesov:

- chemické inžinierstvo,
- molekulárna chémia a fyzika,
- nukleárna chémia a fyzika,
- geológia, geotechnika, meteorológia a pozorovanie zeme,
- modelovanie a simulácia procesov v živých organizmoch (medicína, biotechnológia a pod.),
- technické disciplíny, napr. strojárstvo, elektrotechnika, architektúra, stavebníctvo, hutníctvo,
- ekonómia

- ďalšie vedné oblasti a procesy.

Pre zabezpečenie rýchleho, spoľahlivého a plnohodnotného prístupu všetkých vedeckých, univerzitných a výskumných pracovísk do vysokovýkonných centier je vybudovaná transparentná rýchla akademická sieťová infraštruktúra SANET. SANET je v prevádzke od roku 1992 a v dnešnej topológii s prenosovou rýchlosťou min. 10 Gbps transparentne spája všetky vedecké a akademické pracoviská v Slovenskej republike.

Superpočítačové systémy projektu „Slovenská infraštruktúra pre vysokovýkonné počítanie“

Z početných stretnutí používateľov zahraničných výpočtových zdrojov a potencionálnych používateľov domácej superpočítačovej výpočtovej techniky sa vyprofilovali tri kvalitatívne rozdielne prostredia, ktoré sa ukazujú ako najviac potrebné pre pokrytie ich požiadaviek. Sú to systémy :

- s vysokou hustotou výpočtových jadier a veľmi nízkou latenciou vedúcich k využívaniu výpočtových systémov typu masívneho paralelného procesingu – MPP s kapacitou RAM pamäte cca 2 až 4 GB na uzol,
- klastrové systémy s SMP uzlami vybavených kapacitou RAM pamäte veľkosti 16 až 128, príp. 256 GB, ktorých jednotlivé uzly – nody budú navzájom prepojené komunikačnou sieťou s nízkou latenciou (takými parametrami sa vyznačujú siete ako napr. Infiniband, Myrinet, Quadrics a pod.),
- a s globálne adresovateľnou pamäťou o kapacite RAM pamäte rádovo až do desiatok až stoviek TB.

Superpočítačové systémy budú napojené prostredníctvom SAN (Storage Area Network) na veľkokapacitné externé pamäťové médiá s diskovou kapacitou rádovo stoviek TB. Superpočítače určené pre gridové počítanie budú disponovať aj internou úložnou diskovou kapacitou cca 10 GB na jedno výpočtové jadro.

Uvedené architektúry budú rozmiestnené do navrhovaných lokalít tak, aby čo najviac pokrývali aj lokálne požiadavky na ich využívanie. Základným cieľom pri budovaní superpočítačových centier je koncentrovať maximálne možný výpočtový výkon danej architektúry do jednej lokality tak, aby boli splnené podmienky optimálnej konfigurácie na väčšinu úloh, ktoré budú na superpočítačoch zbiehané pri čo najväčšej využiteľnosti systému.. Všetky superpočítačové systémy sa navrhuje implementovať do sieťovej komunikačnej infraštruktúry SANET po jednotlivých etapách tak, aby bola dosiahnutá ich optimálna konfigurácia s odhadovaným výkonom 20 – 50 TFLOPs na systém.

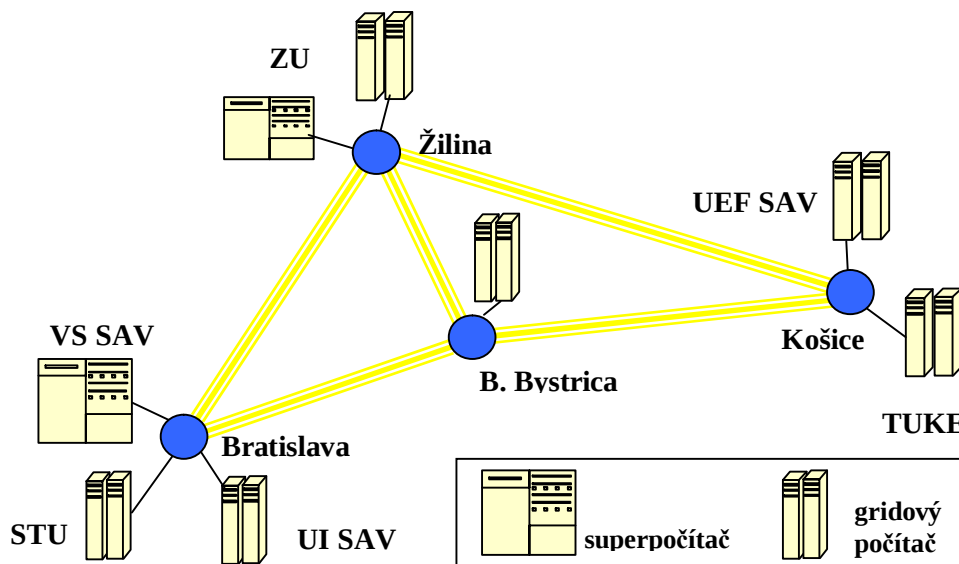
Superpočítače umiestnené v predpokladaných lokalitách – Bratislava a Žilina (Obr. 1/2) budú prepojené navzájom (Obr. 1/3), ako aj smerom k medzinárodnej superpočítačovej gridovej sieti DEISA a novokreovanej európskej superpočítačovej infraštruktúre PRACE.

Gridové systémy projektu „Slovenská infraštruktúra pre vysokovýkonné počítanie“

Gridové systémy budú tvorené vysokovýkonnými klastrovými systémami s vysokokapacitnými dátovými úložiskami a budú umiestnené v Bratislave, Žiline, Banskej Bystrici a v Košiciach. Budú tvoriť základ národnej gridovej infraštruktúry, do ktorej môžu byť zapojené aj už existujúce výkonné výpočtové prostriedky v ďalších organizáciách. Gridové centrá budú začlenené do súčasných európskych gridových infraštruktúr (EGEE) ako

aj do formujúcej sa Európskej Gridovej Infraštruktúry (EGI). Zapojenie do gridových infraštruktúr bude zabezpečené prostredníctvom gridových technológií (middleware) gLite a ARC.

Teritoriálne rozloženie jednotlivých počítačových systémov slovenskej infraštruktúry pre vysokovýkonné počítanie a zúčastnené organizácie (SAV a univerzity) sú zrejmé z obr. 2



Obr. 2

Softvérové zabezpečenie pre vysokovýkonné počítanie

Softvérové vybavenie bude pozostávať z optimálnej kombinácie voľne dostupných a komerčných softvérových balíkov implementovaných do prostredia operačného systému Linux. Požadované softvérové vybavenie sa má získavať postupne podľa vznikajúcich potrieb pri postupnom zavádzaní systémov do prevádzky.

Využitelnosť superpočítačovej a gridovskej techniky hlavne v univerzitnom prostredí bezprostredne závisí od zabezpečenia dostupnosti aplikačného softvéru hlavne komerčného charakteru. V oblasti prírodovedných disciplín a hlavne v oblasti technických disciplín, ktoré sú značne širokospektrálne, jedná sa o náročné programové balíky po odbornej a tiež po finančnej stránke.

V rámci združenia EUNIS Slovensko, ktoré zastrešuje univerzitné prostredie po stránke informačných technológií vznikla iniciatíva realizovať prieskum v oblasti používania aplikačného programového vybavenia. Cieľom bolo získať prehľad o používanom a hlavne z pohľadu budúcnosti o potrebnom programovom vybavení, včítane programového vybavenia použiteľného v oblasti SIVVP pre oblasť výučby a výskumu..

Celkový počet škôl, ktoré zareagovali na prieskum bol 9, vid' nasledovnú tabuľku. V rámci tohto bolo 174 požiadavok na využívanie v konkrétnych predmetoch a výskumných úlohách (len univerzitné pracoviská).

P.č.	Univerzita	Počet požiadavok
1	Akadémia Policajného zboru v Bratislave	3
2	Katolícka univerzita v Ružomberku	11

3	Slovenská poľnohospodárska univerzita v Nitre	12
4	Slovenská technická univerzita v Bratislave	45
5	Technická univerzita v Košiciach	38
6	Trenčianska univerzita A. Dubčeka v Trenčíne	11
7	Univerzita Komenského v Bratislave	25
8	Univerzita P. J. Šafárika v Košiciach	6
9	Žilinská univerzita v Žiline	23
	Spolu:	174

Z celkového počtu vysokých škôl 34 reagoval na prieskum pomerne malý počet. Aj niektoré odovzdané podklady boli mnohokrát nedostatočne vyplnené. Uvedené možno zdôvodniť aj tým, že prieskum sa vykonával na konci minulého akademického roka, nakoľko termín výzvy bol k 31.8.2009.

Okrem uvedených univerzít projektu sa zúčastňujú pracoviská Slovenskej akadémie vied a to Výpočtové stredisko, Ústav informatiky a Ústav experimentálnej fyziky. Po zohľadnení aj týchto pracovísk celkový počet druhov programového vybavenia vyplývajúceho z prieskumu je 136, z toho aplikačné programové vybavenie predstavuje 128 položiek a systémové programové vybavenie predstavuje 8 položiek. V uvedenom počte je zahrnutý okrem komerčného softvéru aj voľne dostupný a vlastný softvér.

Z uvedeného celkového počtu 136 položiek nie každé požadované programové vybavenie je vhodné na zaradenie do používania v oblasti superpočítačovej a gridovskej techniky. Do užšieho výberu sa dostalo po spracovaní požiadavok nasledovné programové vybavenie:

Systémový softvér. Operačný systém: Scientific Linux 5 (klon RedHat Enterprise Linux 5), x86_64; IMSL knižnica k FORTRAN 90; Kompilátor CAVE Comm.; Kompilátory Fortran (IV, 77, 90, 95) a C++ a systémové knižnice; MPICH;

Aplikačný softvér. 3D STUDIO MAX; ABAQUS; ACCELRYs (Discovery/Materials Studio); je v tom aj CHARMM; ADF; ADINA; AMBER; ANSYS; AVS/Express; Cambridge Structural Database; Cambridge Structural Database/CLIENT; CATIA V5; CFD-ACE+; COMSOL Multiphysics; Corina; CRYSTAL06; DEFORM; DYNAST; EXTEND-SIM7; FDTD Solutions; FlexX; Gaussian; ILOG CPLEX; JMatPro; MAGMASOFT; MATHEMATICA; MATLAB Distributed Computing Server; MATLAB Parallel Computing Toolbox; MOLPRO; Návrhový systém pre analýzu a dizajn komunikačných zariadení; Návrhový systém pre elektroniku a mikroelektroniku (napr. ORCAD, Mentor...); OPERA; Oracle Spracovanie informácií v prostredí HPC a GRID -- DBS, APC a RCA -kláštrová podpora; PamStamp; Pro/Engineer; ProCAST; QUICK FIELD; RT-LAB; Schrodinger; Sybyl; SYSWELD; TURBOMOLE; VASP; Vizualizačný SW pre HPC a GRID; WIEN2K;

Uvedené programové vybavenie je z oblasti prírodovedeckých a technických odborov. Je to dané skladbou reagujúcich organizácií na uvedený prieskum potrieb systémového a aplikačného softvéru.

Po predbežných obchodných jednaniach hodnota uvedeného programového vybavenia pri zohľadnení požadovaného počtu licencií a doby platnosti počas trvania projektu presahuje hodnotu 5 miliónov EUR. Treba zdôrazniť, že niektoré aplikačné programové vybavenie je značne finančne náročné. Jeho nákup ďalej komplikuje skutočnosť, že je rozdielna licenčná politika dodávateľov na jednotlivé druhy softvérov. Vzhľadom na to, že nie je možné pokryť z uvedeného projektu uvedenú finančnú sumu, bude potrebné vykonať ďalšiu redukciu v požiadavkách na aplikačný softvér.

V ďalšom postupe sa predpokladá vytvoriť zo záujemcov o používanie jednotlivých druhov softvéru odborné skupiny, ktoré budú garantovať odborné požiadavky pri vytváraní podkladov pre prípravu výberových konaní, pri realizácii vlastného nákupu a pri inštalácii a vlastnej prevádzke.

Záver

Snahou predkladateľa projektu, ako aj jeho tvorcov je nájsť cestu spoločného budovania slovenskej infraštruktúry pre vysokovýkonné počítanie so všetkými partnermi. Predkladaný spoločný projekt je vlastne popisom globálneho cieľa a metódou ako tento cieľ splniť s využitím štrukturálnych fondov EÚ v spolupráci troch organizácií SAV a štyroch univerzít. Sedem parciálnych cieľov je stanovených ako cesta k plneniu globálneho cieľa a predstavuje rozvoj počítačovej kultúry na slovenských univerzitách a v SAV až po európsku úroveň. Projekt dáva možnosť rozvoja nielen svojim partnerom, ale aj všetkým univerzitám a vedeckým ústavom SAV, tak ako i ďalším výskumným inštitúciám a výrobnéj praxi. Snahou tvorcov projektu bolo integrovať sily na pozdvihnutie technologickej a vedomostnej úrovne v oblasti vysokovýkonného počítania na Slovensku bez akéhokoľvek obmedzovania doteraz realizovaných aktivít a dosiahnutých výsledkov v tejto oblasti. Sme presvedčení, že realizácia projektu SIVVP bude významným krokom vpred pre slovenskú vedu a vzdelávanie.

Abstrakt

Súčasná situácia na Slovensku v oblasti vysokovýkonného počítania a porovnanie so zahraničím. Zlepšenie stavu prostredníctvom pripravovaného celoslovenského projektu zo štrukturálnych fondov "Slovenská infraštruktúra pre vysokovýkonné počítanie" (SIVVP). Hlavné ciele projektu, celková architektúra, superpočítačová technika a GRIDy. Softvérové zabezpečenie pre vysokovýkonné počítanie. Spracovanie výsledkov prebiehajúceho prieskumu potreby aplikačného softvéru na našich univerzitách ako podporná iniciatíva v rámci združenia EUNIS.

Kľúčové slová

Superpočítače, superpočítanie, vysokovýkonné počítače, grid, gridové počítanie.

Literatúra

- [1] MetaCentrum, <http://meta.cesnet.cz/cs>
- [2] TOP 500 supercomputer sites, <http://www.top500.org>
- [3] Opis projektu pre prioritnú os č. 1 – 4, Príloha č. 1 k žiadosti o nenávratný finančný príspevok

Kontakt

Ing. Tomáš Lacko, Slovenská akadémia vied Bratislava, Dúbravská cesta 9, 845 35
Bratislava, Slovenská republika, Tomas.Lacko@savba.sk

Ing. Peter Hermann, Slovenská technická univerzita v Bratislave, Vazovova 5,
812 43 Bratislava 1, Peter.Hermann@stuba.sk

doc. Ing. Milan Šujanský, CSc., Technická univerzita v Košiciach, Letná 9, 042 00 Košice,
Milan.Sujansky@tuke.sk

Recenzent: RNDr. Peter Škrovina, CIKT SPU v Nitre