

Earth Simulator – největší superpočítač světa

Luděk Matyska, FI a ÚVT MU

V březnu 2002 byl u Yokohamy v Japonsku uveden do provozu největší současný superpočítač světa, systém zvaný Earth Simulator. Tento superpočítač byl vyvinut ve spolupráci celé řady národních agentur a výzkumných laboratoří Japonska, zejména pak National Space Development Agency, Japan Atomic Energy Research Institute a Japan Marine Science and Technology Center, s firmou NEC, která byla dodavatelem technické realizace společného řešení.

Projekt začal před více jak pěti lety v roce 1997 a zahrnoval nejen novou koncepci stavby superpočítače, ale i jeho umístění a vhodný způsob napájení. Pro Earth Simulator byla postavena vlastní budova v relativně seismicky klidné zóně Japonska, přitom stavba splňuje nejprísnejší bezpečnostní hlediska a měla by být schopna odolat zemětřesením všech stupňů kromě nejvyššího (kdy může dojít k puknutí kůry a propadne se celá budova). Earth Simulator (ve zkratce ES) je postaven uvnitř několika patrové konstrukce zavěšené na samostatných pylonech. Tímto způsob je dosaženo maximální stabilizace v případě otřesů půdy, kdy může dojít k případné změně vzájemné polohy nosných pylonů, avšak zavěšené „vana“ zůstane stabilní a pohyb zemské kůry by tak neměl ohrozit ani vlastní počítač ani diskové a páskové subsystemy, které jsou k němu připojeny.

ES je paralelní superpočítač s distribuovanou pamětí. Je tvořen celkem 640 procesorovými uzly (PN, Processor Node), vzájemně propojenými jednonaprúchodovým prepínačem 640×640. To znamená, že každý uzel může komunikovat s každým dalším, aniž by tím byla jakkoliv narušena možnost ostatních uzlů vzájemně komunikovat.

Každý procesorový uzel je tvořen 8 procesory vektorového typu, které sdílí celkem 16 GB hlavní paměti, vstup/výstupní jednotku a speciální procesor vzdáleného přístupu, který umožňuje realizovat přístupy do lokální paměti i vzdáleným procesorům bez účasti lokálního procesoru. Špičkový (teoretický) výkon každého procesoru je 8 Gflops (8 miliard operací v pohyblivé řádové

čárce za sekundu – pro srovnání, nejvýkonnější Intelovské procesory mají výkon pod 3 Gflops). Celkově je tak ES tvořen 5 120 procesory, 10 TB hlavní paměti a teoretickým špičkovým výkonem 40 Tflops (40 bilionů operací v pohyblivé řádové čárce za sekundu). Tyto parametry z něj dělají bezkonkurenčně nejrychlejší počítač na světě. Ve skutečnosti jeho odstup od ostatních systémů je tak veliký, že tři následně největší superpočítače dohromady stále nemají výkon jediného Earth Simulátoru.

Již zmíněný procesor vzdáleného přístupu je přímo napojen na prepínač a řídí meziuzlovou komunikaci. Každý procesor má propustnost 12,3 GB/s v obou směrech, což dává celkovou propustnost vnitřní propojovací sítě 8 TB/s (64 Tb/s, tj. 64 tisíckrát vyšší než je propustnost Gigabitového Ethernetu). Komunikační procesory podporují celou řadu způsobů přesunu dat (např. přesun částí 3D polí mezi výpočetními uzly je realizován přímo hardwarem ES), a to v rámci jediné operace. Tím je dosažena vynikající reálná rychlost přesunu dat mezi uzly, jejímž výsledkem je vynikající efektivita celého superpočítače v paralelních aplikacích.

ES je tvořen dvěma typy základních kabinetů: v jedněch jsou umístěny výpočetní uzly, zatímco druhé nesou propojovací síť (jedná se o Inter-Network Nodes, IN). Celkově základní sestava ES obsahuje 81 920 propojovacích komunikačních kabelů o celkové délce 2 890 km. Procesorové a propojovací nody, společně s diskovými kapacitami (řádu jednotek PB, tedy 10^{15} byte) a odpovídající páskovou knihovnou zabírají plochu $50 \times 65 = 3\,250$ m². Celá budova je pak 17 m vysoká.

Programování takto rozsáhlého paralelního systému není možné bez odpovídající podpory na straně překladačů a dalších podobných nástrojů. Autoři aplikací mají v podstatě k dispozici dva přístupy, které mohou kombinovat podle potřeby:

1. Pro úlohy, jimž stačí výkon jednoho uzlu, je možno použít Fortran 90 s tzv. OpenMP direktivami. Tento systém je schopen automatické paralelizace napsaných programů a využívá faktu, že všech 8 procesorů jednoho uzlu sdílí společnou paměť.

2. Pro větší úlohy, které vyžadují dva a více uzlů současně, je určen primárně High Performance Fortran (HPF), se speciálními rozšířeními, vyvinutými právě pro systém ES. HPF opět provádí automatickou či alespoň semi-automatickou paralelizaci programů, jako cílovou vrstvu používá systém MPI (Message Passing Interface), speciálně upravený pro maximálně efektivní využití hardware ES.

Programátoři mají samozřejmě k dispozici jak přímý přístup k využití MPI, tak i možnost používat tzv. hybridní programovací model, kdy je konkrétní úloha dekomponována tak, že větší celky vzájemně explicitně komunikují pomocí MPI, zatímco jejich součásti, jejichž velikost nepřesahuje jeden uzel, využívají jednoduššího paralelního modelu se sdílenou pamětí.

Jedním z velkých překvapení, které ES přichystal (snad s výjimkou autorů návrhu jeho architektury), je vysoká efektivita paralelních úloh. Superpočítače, tvořené tisíci procesory (např. ASCI Red s 8190 procesory) dosahují na reálných úlohách nejvýše několika desítek procent teoretického výkonu (30% teoretického výkonu je považována za vynikající úspěch, běžná efektivita se může pohybovat i pod 15%, tj. počítač s teoretickým výkonem 1 Tflops má reálných pouze kolem 150 Gflops). V případě Earth Simulátoru se však podařilo dosáhnout na rozsáhlých simulacích, používajících všechny výpočetní uzly tohoto systému, výkonu přes 26 Tflops, tj. efektivitu přes 65%. Podobná míra efektivit byla dosažena s různými výpočetními modely, takže se zdá, že se architektura ES je navržena mimořádně dobře a dovoluje vysoké využití teoretického výkonu pro různé typy úloh.

ES samozřejmě není určen pro řešení běžných úloh, kde stačí výkon slabších počítačů. Jak už naznačuje samotný název tohoto superpočítače, bude používán pro simulaci seismických dějů Země, zejména s ohledem na možné předpovídání zemětřesení v Japonsku. V současné době se možnosti ES v tomto směru ověřují matematickým modelováním dvou proběhlých zemětřesení - v Nankai v roce 1944 a v Kushiho v roce 1993. Realistický model vyžaduje na 3,2 miliard bodů 3D mřížky, na níž jsou zachyceny geologické poměry v rozsáhlém okolí vlastního epi-

centra zemětřesení. Výpočty vyžadují na 386 GB paměti na 1 024 procesorech a trvají cca 2 hodiny pro jeden model.

Druhou velkou oblastí je modelování zemského klimatu pomocí globálních atmosférických modelů. Výkon ES umožňuje modelovat stav atmosféry a její vývoj na celé Zemi a je schopen současně zpracovávat data o stavu oceánu - to je v současné době považováno za nezbytnou podmínku pro dosažení skutečně realistických předpovědí globálního klimatu.

A na závěr informace, která dává stavbu Earth Simulátoru do relace s možnostmi České republiky: vývoj a konstrukce stála cca 500 milionů dolarů, tj. kolem 15 miliard korun. To je více, než činí rozpočet vysokého školství ČR v běžném roce.

Další informace je možno získat na webových stránkách ES: <http://www.es.jamstec.go.jp/esc/eng/>. □