

SPOJENOU SILOU

Na počátku 21. století se s počítačovou technikou setkáváme na každém kroku. Rychlost výpočtu na jednom počítači však mnohdy nesplňuje očekávání našeho uspěchaného světa, a proto není divu, že člověk se snaží využít více počítačů současně ve snaze získat tak rychleji výsledek.

Existuje mnoho počítačových úloh, jejichž řešení vyžaduje několik hodin, někdy i dní či dokonce měsíců. K těmto úlohám patří také úlohy z vizualizace dat či numerické výpočty, například sledování paprsku, výpočet diskrétní kosinové transformace, výpočet izoplochy pro implicitně zadanou funkci, numerické řešení soustavy diferenciálních rovnic apod. K řešení těchto úloh je na Západočeské univerzitě v Plzni vyvíjen systém pro distribuované zpracování GSD (General System for Distribution), umožňující uživatelům vytvářet jejich vlastní "distribuované aplikace" bez jakékoliv hlubší znalosti v oblasti distribuovaného zpracování.

Na systém GSD lze pohlížet jako na černou skříňku se vstupy a výstupy. Vstupem je uživatelská úloha, výstupem řešení uživatelského problému a uvnitř skříňky "magie", které je programátor používající systém ušetřen (na rozdíl od programátorů používajících profesionální systémy Message Passing Interface - MPI nebo Parallel Virtual Machine - PVM). Tento komfort však přináší řadu omezení. Nejvýznamnějším je použitelnost pouze pro úlohy paralelizovatelné ve smyslu master-slave (farmer-worker), tzn. nelze takto řešit úlohy, v nichž je nutná komunikace mezi pracovníky během jejich práce. Další omezení jsou tato: rozsah práce musí být znám před zahájením výpočtu, výpočet může být předčasně ukončen, systém nelze používat na strojích s odlišnou architekturou, platforma je pouze Windows NT 4.0 nebo 2000. Protože je systém psán obecně, nemusí být zcela efektivní pro konkrétní typ úlohy.

Systém GSD je určen pro uživatele, kteří nemají čas ani chuť pouštět se do zkoumání komunikace mezi distribuovanými komponentami, ale potřebují (nebo si chtějí vyzkoušet) výpočetní sílu distribuovaného zpracování. Pokud je požadavkem především efektivita, je vhodnější použít MPI nebo PVM. Plánovaná verze systému by měla být určena také pro všechny uživatele, kteří se rozhodnou nechat si svou úlohu spočítat "na internetu" (např. když je podniková síť příliš malá, aby výpočet mohl být dokončen v řádném termínu).

Algoritmus uživatelské úlohy může být napsán v libovolném programovacím jazyce, který produkuje knihovny DLL. Uživatel, který chce použít systém GSD, musí napsat dva moduly označované Master User (MU) a Slave User (SU) a tyto moduly umístit do jedné nebo dvou DLL knihoven. MU je modul pro načítání (resp. generování) vstupních dat, sběr výsledků a uložení na disk; modul SU obsahuje vlastní algoritmus. MU poběží na centrálním počítači v síti GSD, SU poběží na všech ostatních počítačích. Uživatel musí napsat funkce pro komunikaci s GSD (podle definice v modulech) a také tyto funkce označit jako DLL exportované. Činnost uživatelského programu je tedy řízená událostmi.

Systém GSD sestává ze tří částí: GSDMaster, GSDSlave a GSDLoader. GSDMaster je nainstalován na jediném počítači a v jeho adresním prostoru běží MU. Kromě toho komunikuje se všemi klony GSDSlave (GSDSlave je nainstalován na všech ostatních počítačích, které GSD využívá). V adresním prostoru GSDSlave běží SU. Tyto komponenty jsou napsány jako Win32 service, tj. jsou automaticky spuštěny při startování operačního systému (pokud není řečeno jinak) a k jejich činnosti není nutné přihlášení uživatele. S těmito komponentami přijde do styku administrátor systému, běžný uživatel přijde do styku pouze s GSDLoaderem.

GSDLoader zobrazí jednoduchý formulář, v němž uživatel vyplní informace o minimálním a maximálním počtu počítačů, na nichž bude modul SU distribuován - žádné ruční instalování na mnoha počítačích. Vyplňuje se rovněž umístění DLL knihovny s moduly MU a SU a umístění konfiguračního souboru. MU je inicializován na základě tohoto souboru, protože MU ani SU nesmí zobrazovat žádný dialog ani chybové hlášení - toto omezení je dáno vlastnostmi aplikace, která je spuštěna jako služba. Poslední nepovinná informace je stručný název úlohy a e-mailová adresa, na kterou se má po skončení výpočtu odeslat protokol o jeho průběhu.

Požadavek výpočtu úlohy se uloží do fronty, z ní je vybrán a úloha aktivována, jakmile je dostatek volných počítačů. Díky této vlastnosti je možné systém GSD "zaúkolovat" v pátek odpoledne, vypnout monitory, zhasnout světla, zamknout a po návratu v pondělí ráno budou výsledky k dispozici (pokud o víkendu nedošlo k odstávce elektřiny). Za předpokladu, že ve formuláři byla uvedena i e-mailová adresa,

na počítači je nainstalována podpora SMTP (MS Outlook Express ani MS Outlook nejsou schopny odeslat e-mail z programu, který běží jako service) a pokud se elektronická pošta posílá rovněž na mobilní telefon, může se uživatel již o víkendu radovat z informace, že se něco dopočetlo.

System se neustále vyvíjí, podle tvůrců se dočkáme integrované SMTP podpory, tzn. že nadále nebude nutný operační systém Windows 2000 s nainstalovaným IIS 5.0 ani Windows NT Server či MS Exchange Server. Pravděpodobně nejdůležitějším vylepšením by se mělo stát řízení systému GSD pomocí webového formuláře. FreeWare verzi GSD systému 1.2 BETA naleznete na přiloženém Chip CD. Při instalaci systému se řiďte pokyny uvedenými v souboru readme.txt.

Jan Doubek, Josef Kohout | cse@kiv.zcu.cz

Infotipy:

Bližší informace o systému, novinky, download > <http://herakles.zcu.cz/research.php>