

Internetové superpočítače

Využijte volné kapacity svého PC pro dobro všech

Vojtěch Bednář

Počítač každého z nás, který stojí doma nebo v kanceláři a je alespoň příležitostně připojený na internet, se může stát superpočítačem. Tedy přesněji řečeno jeho částí. Může být prvkem systému distribuovaného počítání. Co to je? Jak to funguje a jak je možné toho využít? Pokud vás zajímají otázky související s distribuovanými výpočetními systémy a pokud byste rádi viděli takový systém i u vás v pokoji, pak čtěte dál.

Distribuované počítání je poměrně starou technologií, přesněji řečeno technologickou koncepcí, jak řešit složité problémy, především matematického charakteru. Pokud máme například rovnici, jejíž řešení je enormně náročné, nebo sled postupů, jež je třeba extrémně mnohokrát opakovat, pak čelíme klasickému problému svěřovanému superpočítačům.

Byl jednou jeden superpočítač

Superpočítač v původním slova smyslu je strojem s enormním výpočetním výkonem a ohromnou pamětí, který je již od počátku konstruován pro řešení tohoto typu problémů. Tato konstrukce sice způsobuje, že takový stroj ztrácí většinu vlastností typických pro obyčejné, kancelářské či domácí počítače nebo dokonce servery, nicméně to je oběť, kterou je potřeba přinést zpracování velice složitých úloh. Typicky si takový počítač představujeme jako stroj zvící jednoho fotbalového hřiště.

Superpočítače v novějším pojetí již nejsou monolitním systémem, jako tomu bylo dříve, i když rozměr basketbalového hřiště ani zdaleka neztratily. Místo toho jsou tvořeny velkým množstvím desek, z nichž každá je technicky vzato počítačem sama o sobě. Obsahuje totiž základní sběrnici, procesor, část paměti a komunikační prvky, které ji propojují s ostatními částmi celého zařízení. Výsledkem je, že výpočetní kapacita takové desky je propojena s výpočetní kapacitou všech ostatních a vzniká zařízení, které se opět navenek chová, jako by šlo o jediný počítač, nicméně v něm dochází ke slučování výkonu všech komponent, které jej tvoří.

Superpočítač je, jak již doufejme vyplynulo z jeho popisu, nesmírně složitou a proto i nesmírně drahou záležitostí. Takové stroje se až na výjimky nestaví sériově. Každý je vlastně

originálem, který je konstruován některou ze specializovaných firem na zakázku pro každého jednotlivého zákazníka zvlášť. Zákazníky, kteří si mohou podobné stroje dovolit, jsou především vlády, armády, některé vysoké školy a instituce, které potřebují vysoký výkon k provozování své činnosti, například řízení leteckého provozu nebo meteorologická služba. Jinak řečeno, superpočítač je vynikající věc, ale ani omylem není pro každého. Kromě toho jsou tyto stroje také mimořádně náročné na provoz, investice do jejich stavby a oživení nejsou všechno. Superpočítače potřebují obrovské množství elektrické energie. Kromě toho musí být chlazeny, protože provozem velkého počtu výpočetních jednotek s vyšší výkonností než běžné PC vzniká velké množství zbytkového, odpadního tepla. Až donedávna musely být (a některé musejí být dodnes) umístovány do speciálního prostředí, kde se dbalo a dbá na velmi nízkou míru prašnosti, na čistotu ovzduší a na striktně udržovanou teplotu. Potřebují také specializovanou obsluhu – náročnost a tedy i cena jejich údržby je ve srovnání s klasickým stolním PC nesrovnatelná. To ovšem ještě ani zdaleka nejsme u otázky cenové náročnosti softwaru pro takové přístroje. Základnovým operačním systémem jsou různé specializované verze UNIXu, na nichž běží zvláštní aplikace určené pro ty úlohy, pro něž byl daný počítač sestaven. Jejich administrace se sice podobá řízení běžných unixových systémů, nicméně je ve skutečnosti mnohem náročnější, jak z hlediska nároků na erudici obsluhy, tak i z hlediska samotného řídicího a programovacího prostředí.

Rozdělená úloha

Již poměrně dávno přišel kdosi s myšlenkou, jak vyřešit úlohy určené pro superpočítače i bez jejich asistence. Byla to vlastně z nouze ctnost, ale znamenala počátek nápadu, který pak našel své naplnění až s masivním nasazením internetu. Tímto nápadem bylo suplovat funkci superpočítače v „moderním“ pojetí, tedy složeného z velkého množství relativně menších jader v síti vytvořené z obyčejných počítačů. Klíčovou podmínkou je, aby obrovskou úlohu, která stojí na začátku celého procesu, bylo možné rozdělit na velké množství relativně jednodušších úloh. Jednodušších v tom smyslu, že jejich výpočty je možné provést pomocí zařízení o běžně dostupné výpočetní síle v rozumném čase několika hodin, maximálně několika desítek hodin čistého početního výkonu, při zatížení daného stroje blížícímu se maximu jeho možností. Ne všechny úlohy určené pro superpočítače (jež se ve výsledku mohou chovat jako jediná logická jednotka) lze takto upravit, nicméně ty typické, postavené na velkém počtu opakování relativně konstantní úlohy, takto možné přepracovat je. Dalším krokem je vytvoření systému, jenž umí „porcovat“ výchozí úlohu na ony menší části a prostřednictvím nějakého komunikačního média, obvykle počítačové sítě (ale není to

podmínkou), ji distribuovat obyčejným počítačům. Ty pak úlohu prostřednictvím speciálně naprogramované aplikace, tzv. klienta distribuovaného počítání, v nastavené době, tedy tehdy, nejsou-li právě používány pro něco jiného, řeší. Jakmile je počítačem – uzlem sítě distribuovaného počítání – jednotlivá úloha dokončena, je její výsledek pomocí stejného média, jakým byla získána vstupní data, opět předán distribučnímu systému, který se postará o jeho správné zařazení do kontextu ostatních výsledků. Jakmile je tato operace provedena, zajistí odeslání dalšího zatím nezpracovaného balíku dat uvolněnému uzlu.

Protože v takovém prostředí záleží vlastně jen na výkonu distribučního systému a mimo něj mohou na jednotlivých úlohách zcela nezávisle na sobě pracovat stovky, tisíce a miliony samostatných a různě výkonných jednotek, je tak vlastně možné dosáhnout nejen početní síly blízké se superpočítačům, ale výkonnost mnohých z nich dokonce i několikanásobně překročit. Jak ale přinutit majitele, správce klasických počítačů, obyčejných PC, aby se tohoto účastnili? Snadno. Žádné obyčejné PC není využíváno po celou dobu, kdy je zapnuto.

V naprosté většině případů má systém definovaný spořič obrazovky, který je aktivován po určité době neaktivity, tedy pokud je zřejmé, že u stolního počítače už několik minut nesedí uživatel, nebo na tomto počítači nic nedělá. V takovém případě se spouští rybičky v akváriu, elegantní geometrické tvary nebo sled kreslených vtipů a výpočetní výkon stroje je bez užitku. Právě v tomto okamžiku ale může být aktivován systém distribuovaného počítání.

Již před lety byly provedeny pokusy spočívající ve vytvoření jakéhosi amatérského superpočítače tak, že bylo například v univerzitní tělocvičně sesíťováno několik stovek klasických PC a pomocí speciálního softwaru bylo dosaženo koordinace jejich činnosti. Systém, který vznikl, se sice superpočítačům podobal, ale pravá síť distribuovaného počítání to nebyla, protože o všechny tyto stroje se bylo nutné starat.

V souvislosti s rozvojem internetu a jeho proniknutí i do běžných domácností a na méně náročná a zatížená pracoviště se pro systémy distribuovaného počítání otevřelo nové pole působnosti, na němž mohou teprve předvést vše, co doopravdy umí. Projekt Seti@Home se dnes stal již legendou. Kromě něj ale existuje množství dalších, podobných systémů, které se zabývají jinými, rovněž velmi složitými úkoly z oblasti biologie, fyziky, medicíny, ale také například kryptografie. Díky systémům distribuovaného počítání se staly služby superpočítačů dostupnějšími než dříve, a mohou být tudíž nasazeny i pro takové úlohy, které by, ač jsou zajímavé, nikdo nikdy na klasickém superpočítači nepustil prostě proto, že jsou pro něj spekulativní a poněkud zbytečnou záležitostí – jejich okamžitý potenciální přínos nemusí být nutně roven vynaloženým nákladům.

*****S podtiskem*****

Podívejte se na seznam aktivních projektů

www.aspenleaf.com/distributed/distrib-projects.html

Distribuované počítání na internetu a jeho možnosti

V současné době existují díky internetu desítky sítí distribuovaného počítání. Některé z nich jsou určeny pro širokou veřejnost, jiné pracují pouze v daném prostředí, například ve školách, nemocnicích nebo společnostech, které tak efektivně využívají volného času a prostředků svých počítačů k pozitivním účelům. Některé z veřejně dostupných sítí jsou dobrovolnou záležitostí a každý, kdo se těchto projektů účastní, tak činí bez nároku za odměnu. Existovaly ale i takové, které slibovaly za investované hodiny procesorového času zaplatit určitou finanční částku. Sítě jsou realizovány jako spořič obrazovky (i v tomto ohledu bylo průkopníkem Seti@Home), nicméně mohou fungovat i nepřetržitě jako jeden z procesů hostitelského počítače. V minulosti se objevila přinejmenším jedna síť (Altnet), která fungovala na spywarovém základu. Jejím klientem se stal každý, kdo si nainstaloval výměnný program KaZaA Media Desktop (plnou, nikoliv pirátskou „Lite“ verzi) a byla určena k prodeji výpočetního času svých členů třetím stranám. Přestože oficiálně tato síť nebyla použita a její klienty dnes slouží k něčemu jinému (šíří prostřednictvím KaZaA chráněný obsah, soubory vybavené zlatou ikonkou), ukázala i ona možnosti distribuovaného počítání v kombinaci s jinými technologiemi, v tomto konkrétním případě s malwarem, tedy s nežádoucím softwarem bundlovaným sponzorskou formou s jinými aplikacemi.

Co tedy potřebuji, pokud chci pracovat v distribuční síti?

Pokud máte dojem, že váš počítač většinu svého času zahálí a pouze živí více či méně nudný spořič obrazovky, pak není důvod, proč byste se nemohli stát aktivním prvkem některé ze sítí distribuovaného počítání. Co k tomu potřebujete? Nic zvláštního, vlastně jen základní věci:

- dostatek výpočetního výkonu, procesor Intel PII a vyšší, Celerony, včetně kompatibilních,

- operační systém Windows konfigurovaný tak, že v době, kdy běží spořič obrazovky nefunguje na pozadí jiná služba, která by podstatným způsobem konzumovala výkon (například nepřetržitý antivirový scanner, nekonečná optimalizace disku či trvale využívaný server),

- minimální připojení k internetu, nemusí být trvalé (i když pro lepší výsledky se pochopitelně doporučuje) pro stahování zdrojových balíčků dat a odesílání výsledků,

- alespoň přibližnou představu, čemu vlastně byste chtěli výkon svého PC věnovat (na výběr je stále poměrně dost různých projektů),

- konfiguraci nastavenou tak, aby se systém aktivoval po rozumné době neaktivity počítače, přesněji neaktivity jeho hlavních aplikací.

Ještě je třeba říci, že existují placené systémy distribuovaného počítání. Nicméně není u nich příliš dobré počítat s nějakým horentním výdělkem, protože tak vysoký početní výkon, aby ho dosáhl, našinec stejně v naprosté většině případů nesežene, a pokud ano, pak je pravděpodobné, že by náklady v konečném důsledku bohužel byly vyšší než možný zisk.

*****Vložený velký článek*****

Některé zajímavé systémy distribuovaného počítání

Seti@Home

setiathome.berkeley.edu

*****obr.: s.tif*****

Seti@Home je nejstarším nepřetržitě běžícím globálním projektem distribuovaného počítání. Smyslem tohoto projektu je pomocí analýzy rádiových signálů získávaných prostřednictvím radioteleskopu v Arecibu v Novém Mexiku hledat možné stopy mimozemské inteligence, přesněji její snahy komunikovat s naší civilizací. Seti@Home bylo založeno univerzitou v Berkeley, nicméně na jeho provozu se podílí řada velkých společností, které tímto způsobem testují svůj hardware i software v podmínkách největšího systému distribuovaného počítání na světě. Je v provozu již čtyři roky a je pravděpodobné, že ještě v provozu nějakou dobu bude, nicméně žádné signály, které by k nám vysílal E.T., se prozatím nepodařilo objevit, byť existují vážní kandidáti z některých částí hvězdné oblohy.

Evolution@Home

www.evolutionary-research.org

*****obr.: evolut.tif*****

Tento projekt se podobá předchozímu pouze svým názvem smyslem je totiž něco zcela odlišného. Účelem je zjistit, jak funguje matka příroda, konkrétně evoluce. Pomocí speciálních simulátorů jsou zde s využitím počítačů dobrovolníků simulovány procesy, ke kterým docházelo v přírodě po miliony let. Díky tomu je možné lépe pochopit v mnoha ohledech vývoj života na Zemi a v důsledku toho i potenciálně na jiných místech této a jiných slunečních soustav. Protože evoluce je nesmírně složitým procesem, je stejně složité i její matematické modelování, které se opět děje pomocí speciálního klienta určeného pro několik různých systémových platforem.

Nedostatkem tohoto systému je menší míra automatizace než v případě Seti@Home, mnoho věcí je zde nutné dělat ručně, především se pak jedná o dodávání dalších pracovních jednotek. Také samotná aplikace není tak hezká a jak se můžete přesvědčit z obrázku, běží v konzolovém režimu. Přesto dělá velmi zajímavou práci – její výsledky na internetu vás mohou přesvědčit.

Climate Prediction

www.climateprediction.net

*****obr.: climate.tif*****

Také projekt Climate Prediction patří stejně jako předchozí do rodiny systémů enviromentálního počítání. Účelem v tomto případě není pochopit fungování evoluce, ale pokusit se předpovědět vývoj klimatu na příštích padesát let. Data jsou sbírána na základě současných globálních map světového klimatu a analyzována s ohledem na různé možné tendence a směry vývoje. Výsledkem tohoto procesu má být dosažení poměrně přehledné a dynamické mapy vývoje planetárního klimatu v perspektivě následujících padesáti let. Klient funguje jako spořič obrazovky, nabízí možnost výběru projektu (a nejen to), je poměrně obsáhle konfigurovatelný, má integrovanou službu zaslání zpráv uživatelům a několik dalších užitečných funkcí. Bohužel, čas od času na Windows XP/SP2 padá nebo má tendenci fungovat nekorektně, což je škoda.

Stavba urychlovače částic

stephenbrooks.org/muon1

*****obr.: urychlovac.tif*****

Z enviromentálních věd se přesuňme na chvíli do oblasti techniky. Nejmodernější fyzika vyžaduje nejnovější vybavení. Jeho konstrukce již dávno není jen otázkou inženýrů, ale i superpočítačů. V rámci projektu MUON1 se můžete účastnit vývoje urychlovače částic, který je konstruován opět pomocí dobrovolnických počítačů po celém světě. Jedná se o docela zajímavý a ve světle hledačů mimozemšťanů nebo evoluce mnohem praktičtější projekt, který však také vyniká poměrně velkými ambicemi konstrukce funkčního urychlovače. Bohužel, s klientem tohoto systému jsme měli trochu problémy. Pokud byste si jej chtěli vyzkoušet, doporučujeme určitou opatrnost, nicméně na Windows typu NT jej lze docela dobře zvládnout.

Hledání léků

www.find-a-drug.org

*****obr.: leky.tif*****

Hledání léků je projekt, který se pomocí služeb distribuovaného počítání snaží přispět k nalezení léků na běžné, avšak smrtelné a zničující choroby, jako je například malárie a další. Je provozován nezávislou neziskovou skupinou. Zajímavostí je, že Find-a-Drug se mimo jiné snaží najít i kvalitnější protilátky proti jedům a toxinům, které by mohly být potenciálně využity teroristy pro bioteroristické útoky. Projekt je rozdělen do několika dílčích částí, v rámci kterých se pracuje na jednotlivých úkolech. Smyslem je analýza velmi složitých molekul a jejich možných kombinací. Klient tohoto systému má standardní instalátor pro systém Windows. Jeho konfigurace je spojena s jednoduchou registrací. Ovládání klienta je realizováno přes ikonku v systémové oblasti Windows a přes okno, na němž je také zobrazována grafika, funkce systému je jinak řízena zcela automaticky. Nedostatkem byla v tomto případě kromě místy pochybné stability také skutečnost, že systém dokázal velmi rychle spotřebovat veškerý výpočetní výkon i poměrně silného počítače za jeho plného chodu, a tak bylo nutné jej ručně pomocí správce úloh „usměrnit“. Nakonec se to ovšem podařilo, a tak mohlo distribuované počítání hledání léků probíhat relativně bez problémů.

Výběr je větší

Toto bylo pět projektů distribuovaného počítání, nicméně jich existuje mnohem větší množství. Za pozornost stojí mimo jiné ještě například jeden, který se pokouší odhadovat sekvence lidského genomu z již známých částí a výrazně tak napomoci poznání genetické

stavby člověka a vlastně i dalších živých organismů (www.grid.org/projects/hpf). Pomoci hledání léků se snaží i jeden z projektů Standfordské univerzity (cmgm.stanford.edu/~cparnot/xgrid-stanford/index.html). Jiným zajímavým pokusem je snaha počítat proteinové sekvence pomocí projektu Predictor@Home (predictor.scripps.edu). Zajímavostí tohoto projektu je, že rozměrově se jedná o jednoznačně největšího klienta pro distribuované počítání, s nímž se můžete dnes na veřejném internetu setkat (jeho provoz zabere něco přes sto megabajtů na pevném disku). Zajímavé je i to, že tento systém používá standardizovanou platformu (tj. klient – server) pro distribuované počítání, s níž se je možné setkat i u několika dalších podobných projektů.

*****konec velkého vlož. článku*****

Tipy pro vás

Už jste si vybrali systém distribuovaného počítání, který vám vyhovuje? Nyní pro vás máme několik tipů, kterých je moudré se držet (ale v žádném případě to není povinné).

- Nechte svůj počítač pracovat nepřetržitě. Neustálý provoz PC nijak nevadí a rychleji dosáhnete zajímavých výsledků.
- Nechte systém distribuovaného počítání běžet kdykoliv, kdy je to možné, ale jen tak, aby vás jeho činnost nijak neomezovala při práci. Pamatujte, že to je sekundární úloha vašeho počítače a jako takový byl i navržen. Nemá smysl kvůli němu omezovat práci nebo hru a stavět vlastní počítač jen kvůli takovému systému mohou pouze ti největší nadšenci (ačkoliv i takoví zde jsou).
- Nepoužívejte současně více než jeden systém. Smysl by to mělo jedině v případě, že byste měli opravdu extrémně rychlý počítač, který by neměl co dělat. A to nepředpokládáme.
- Pravidelně sledujte novinky. Navštěvujte web provozovatele systému distribuovaného počítání, pročítejte si, na jakých úkolech se momentálně pracuje či momentálně pracujete, jak je celá věc daleko a co vše ještě zbývá vypočítat. Vyplatí se to.
- Pokud je to možné, nechte se zapsat na newsletter. Budete tak pravidelně informováni o novinkách. I to velmi oceníte.

Distribuované počítání představuje velmi zajímavou oblast využití internetu a elektronické komunikace v moderním světě vůbec. Pokud se pro ně rozhodnete, můžete se stát součástí

obrovského a potenciálně užitečného systému. Pouze je potřeba si opatrně a dobře vybírat a kvůli jedné aplikaci navíc rozhodně nepřestat dodržovat zásady bezpečného používání počítače a výměnných komunikačních systémů, jimiž aplikace pro distribuované počítání prakticky vzato jsou. Je třeba si také uvědomit, že jde o aplikace, které fungují nepřetržitě, využívají výkon počítače a mají přístup k jeho obsahu stejně jako ke komunikaci s internetem. Instalujte jen to, čemu důvěřujete.

*****Vloženo*****

Superpočítač versus distribuované počítání

Co je to superpočítač?

- jedno speciální zařízení pro extrémně náročné výpočty,
- může být složen z množství menších „jader“, tedy z malých počítačů,
- používá speciální verze obvykle UNIXových systémů,
- jeho součástí je speciální software,
- jedná se o náročnou a velmi drahou záležitost.

Co je to distribuované počítání?

- vyřešení úlohy pro superpočítač tím, že je tato úloha rozdělena na větší množství malých úkolů,
- tyto „malé úkoly“ mohou řešit i standardně výkonné stanice propojené sítí,
- na těchto stanicích je instalován speciální software,
- jednotlivé úlohy jsou přidělovány jednotlivým stanicím distribučním centrem,
- výsledky jsou pak předávány zpět do distribučního centra,
- jednotlivé stanice či prvky sítě distribuovaného počítání, která tak vzniká, mají různý výkon, různou kapacitu, jsou schopny zpracovávat výsledky svých částí výpočtu s různou frekvencí,
- protože tak ale činí současně, může síť o několika tisících uzlů vygenerovat během krátké doby několik tisíc hodin čistého výpočetního času, který by se na jednom stroji vytvářel velmi dlouho,

- výsledkem je tak vyřešení srovnatelně těžké úlohy jako pro superpočítač, a navíc v krátkém čase.

Obrázky popisky:

Distribinfo.tif: S výčtem aktivních projektů se nejlépe seznámíte na webové adrese distributedcomputing.info.

Grid.org.tif: Projekt, který se pokouší odhadovat sekvence lidského genomu z již známých částí se nachází na adrese www.grid.org/projects/hpf.

Protein.tif: Jiným zajímavým pokusem je snaha počítat proteinové sekvence pomocí projektu Predictor@Home (predictor.scripps.edu).

Setiklient.tif: Smyslem projektu Seti@Home je hledat možné stopy mimozemské inteligence, přesněji její snahy komunikovat s naší civilizací.