

Distribuované výpočty

Představujeme BOINC!

Víte o tom, že se díky svému počítači můžete účastnit skutečných vědeckých výzkumů? Stačí málo, nainstalujte si BOINC a zapojte se do některého z více než třiceti projektů.

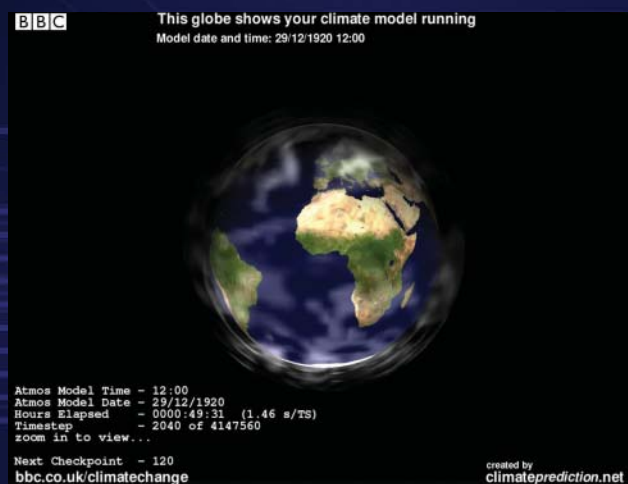
Text: Dušan Vykouřil, vykouril@centrum.cz

MNa úvod se zamyslete nad následující otázkou: Využíváte výkon svého počítače po celou dobu jeho provozu na 100 %? Pokud si s čistým svědomím odpovíte „Ano“, pak jste velká výjimka a dobrý uživatel. Většina počítačů na světě však svůj plný výpočetní potenciál využívá jen velice malou část své provozní doby. Když uvážíme, že existuje spousta různých způsobů, jak počítač využít na 100 % i při obyčejném psaní v textovém editoru, při práci ve skladovém programu nebo při popíjení šálku kávy či v době oběda, kdy se vám nechce na tu chvilku počítač vypínat, je obrovská škoda tohoto „lenušení“ počítače nějak efektivně nevyužít. Málokdo si uvědomuje, kolik takového nevyužitého výkonu na světě vlastně je a že je velmi snadné tento výkon věnovat například k vědeckým výzkumům.

Princip je zcela jednoduchý a jeho bezproblémová funkčnost léty ověřená. Stačí mít alespoň občas z konkrétního počítače přístup na internet, stáhnout si některého z klientů pro zpracovávání distribuovaných výpočtů (DC) a zaregistrovat se do některého z mnoha projektů. Chtěl bychom tímto článkem upozornit na ojedinělý systém, který spojuje několik takovýchto DC projektů a stará se o rozdělování poskytnutého výkonu mezi ně. Jedná se o systém BOINC (Berkeley Open Infrastructure for Network Computing).

Jak BOINC funguje?

O distribuci a základní správu výpočtů i jednotlivých projektů se stará program (klient) BOINC Manager, který si stáhnete z internetu a kterému je po nainstalování automaticky systémem přidělena nejnižší priorita. Jinými slovy, jakýkoliv běžný program má na vašem počítači přednost před BOINC. V praxi to funguje tak, že při práci na svém počítači vůbec nepoznáte, že je na něm BOINC nainstalován a že pracuje. Jediný rozdíl tedy bude v tom, že procesor bude vždy vytížen na 100 %. Klient využívá pouze zbytkový výpočetní prostor a maximálně pružně ustupuje programům, které používáte, takže vaše práce není vůbec bržděna. Pokud například spustíte nějakou hru či zpracování videa, které váš procesor využije na 100 %, klient BOINC se zastaví a bude automaticky pokračovat, až bude mít opět k dispozici alespoň nějaké procento výkonu CPU, které bude nevyužito. Existuje i několik nastavbových programů, díky kterým je možné obsluhovat více počítačů přes internet či v síti naráz a také jim přidělovat práci na různých projektech. Jedná se především o program BOINC Studio a BOINC View. Dokonce lze pro takovou správu použít i obsluhu čistě po internetu, takzvaný BOINC Account Manager.



PŘEDPOVĚĎ KLIMATU: Takto vypadá grafický výstup BOINC při zpracování projektu Climateprediction.net



BOJ PROTI RAKOVINĚ: Grafický výstup BOINC při zpracování projektu Rosetta

➔ Proč bych se měl do BOINC zapojit?

„A co z toho budu mít?“ – právě tato otázka patrně napadne leckterého jedince, který se o BOINC poprvé dozví. Odpověď je, jak by se dalo očekávat, velice snadná: „Téměř nic.“ Nečekejte, že vám za to někdo bude každý měsíc posílat šek nebo jiné peněžní dary anebo že dostanete lepší procesory, abyste toho mohli více spočítat. Na takovéto hmatatelné odměny můžete rovnou zapomenout.

Není ovšem pravda, že z toho nebudete mít zhola nic. Hlavním přínosem je pro většinu členů BOINC účast na skutečném vědeckém výzkumu a radost z toho, že pomáhají projektu, který podle jejich přesvědčení má opravdu smysl a který by bez lidí zapojených do BOINC prakticky neměl šanci existovat.

Míra práce každého jedince a jeho jednotlivých počítačů je také ohodnocena takzvaným kreditem, který je automaticky přidělován za každou zpracovanou jednotku či její část a jehož množství je odvozeno matematicky podle náročnosti výpočtu pro konkrétní jednotku i projekt. Podle účtu vašeho kreditu jsou vedeny i podrobné statistiky s pořadím nejen jednotlivců a projektů, ale v každém projektu i jednotlivých států, a dokonce i týmů. Existují i statistiky hromadné, které zahrnují všechny projekty současně a které veškeré výsledky v rámci BOINC projektů sčítají.

Co BOINC zpracovává?

Existuje spousta vědeckých projektů různých světových univerzit či skupin vědců, kteří nemají zrovna hromady peněz na nákup výpočetních center se superpočítači za miliony dolarů. A právě takovými projekty BOINC pomáhá.. Zpracování a dělení práce mezi jednotlivé počítače probíhá několika způsoby.

Tím nejčastějším je, že je rozdělena na mnoho dílků (pracovních jednotek) a tyto dílky jsou rozepisovány ke zpracování jednotlivým uživatelům, kteří si BOINC nainstalovali. Na vašem počítači je jedna takováto jednotka o minimální velikosti (několik kB) zpracována za pár desítek minut či několik hodin a poté automaticky odeslána zpět do centra projektu. Tam dochází k opětovnému spojení jednotlivých dílků a k průzkumu výsledků.

ČR v systému BOINC

I přes svou malou rozlohu a počet obyvatel na tom není naše republika ve statistikách BOINC tak špatně, jak by se mohlo zdát. Na rozdíl například od Slovenska, které se pohybuje až ve třetí desítky, ČR se nachází na devátém místě a zdárně se dotahuje na místo osmé. Zájem o DC projekty je z dlouhodobého hlediska u nás velký a nemálo počítačů se zasloužilo o rozvoj a propagaci BOINC u nás. Stále se však jedná o nepatrný zlomek nevyužitého počítačového výkonu, který leží ladem třeba zrovna u vás a který jen čeká na vhodné využití.

Dalším často používaným způsobem je, že zpracování probíhá za použití takzvaných modelů. Tyto modely jsou kompletní celky, které se zpracovávají, a rozdíl v nich je pouze ve specifickém nastavení vstupních dat.

Konkrétní projekty

Pod systémem BOINC již probíhá několik desítek projektů. Následující výčet není zcela kompletní, pokusíme se tedy představit alespoň ty nyní nejzajímavější.

Seti@Home

Na prvním místě nelze nezmínit projekt Seti@Home, jehož mnohaleté působení ve světě distribuovaných výpočtů se těší obrovské světové popularitě. Tento již více než 40 let trvající výzkum uvedl v roce 1999 mezi

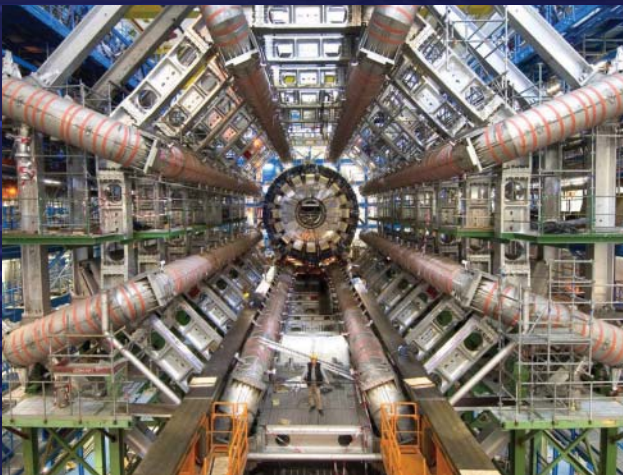
distribuované výpočty David Anderson, který v roce 2003 také stál u zrodu celého systému BOINC.

Jedná se o velice zajímavý a vskutku nadčasový projekt, jenž zajisté zaujme všechny nadšence vesmíru, kterým vrtá v hlavě pravděpodobně nejzásadnější otázka lidstva: „Jsme ve vesmíru sami?“ Díky největšímu radioteleskopu na světě, který se nachází v Arecibu (Puerto Rico) a který má centrum projektu k dispozici, zachytávají vědci vesmírný šum v jeho rádiové podobě, následně filtrují tento zvuk od běžných rušení a ve výsledném signálu hledají jakékoliv anomálie, které by mohlo mít na svědomí vysílání od mimozemských civilizací. Tento projekt měl za dobu své existence několik desítek kandidátů na takovýto signál, ale ten pravý důkaz prozatím stále nalezen nebyl.

Einstein@Home

Tento projekt se pomocí několika interferometrů obrovských rozměrů snaží zaznamenat a v další fázi zkoumat gravitační vlny ve vesmíru.

Existenci těchto gravitačních vln předpověděl již před 100 lety Albert Einstein a na konci dvacátého století se jí podařilo nepřímo ověřit. Stále však chybí přímý důkaz a také to hlavní, možnost jejich zkoumání, díky kterému bychom se mohli nepřímo podívat do historie celého vesmíru zpětně až k velkému třesku. Je to možná neuvěřitelné, ale v gravitačních vlnách by se opravdu dalo číst téměř jako ➔



SRAŽKY ČÁSTIC: Výstavba detektoru Atlas v novém urychlovači Large Hadron Collider



ŠUMY VE VESMÍRU: Radioteleskop, který využívá projekt Seti@Home k hledání mimozemských civilizací

→ v knize, protože mají tu ojedinělou vlastnost, že jejich síla časem polevuje jen velice mírně a je v nich zaznamenána každá velká událost, která se ve vesmíru stala.

LHC@Home

Společnost CERN staví obrovský částicový urychlovač, v němž chce velice přesně řídit oběh částic obrovskou rychlostí a vyvolávat jejich srážky. Cílem takového výzkumu je díky srážkám napodobujícím velký třesk vytvořit a zkoumat například částici Higgs Boson, která podle předpokladů při velkém třesku vznikla, ale v krátkém období po něm zanikla. Tato částice by měla zodpovídat i za hmotnost některých jiných částic a na svou velikost by měla mít obrovskou hmotnost.

V první fázi výpočtů jde převážně o simulace oběhů v urychlovači, aby se dalo dokončit ladění všech usměrňujících magnetů a přístroj mohl být v roce 2007 uveden do provozu. Přídělování práce na tomto projektu je nyní velice nahodilé, jelikož se vždy zpracovává určitý balík dat, která jsou v CERN potřeba při stavbě urychlovače.

Climateprediction.net (CPDN)

Jedná se o zpracovávání klimatických modelů s cílem předpovědět vývoj klimatu na Zemi v dalších desetiletích. Zpracování probíhá prostřednictvím klimatických modelů, kterých je více typů. Liší se od sebe jak počtem zpracovávaných let, tak i precizností výpočtů v jednotlivých letech. Modely fungují tak, že Země je po několika stupních zeměpisné šířky a délky rozdělena na několik set buněk (dle konkrétního typu modelu). Každá buňka má několik výškových úrovní (oceán + atmosféra). Pro každou buňku se počítá teplota, tlak, vlhkost, vektory větru, oblačnost a sníh/led.

Model se od nastavení počátečních podmínek a parametrů vyvíjí. Z toho také vyplývá větší náročnost těchto výpočtů oproti ostatním BOINC projektům. Klimatické modely se zpracovávají na běžném počítači několik set hodin.

Jak se daří našim týmům v systému BOINC?

Jednoznačně největším a nejuspěšnějším týmem u nás je Czech National Team.

Tento tým se postupně propracoval až na třetí místo v žebříčku několika desítek tisíc týmů na světě a jeho výkon stále pomalu stoupá. Do týmu se za více než dva a půl roku zapojilo již více než 4000 členů. Tým před několika měsíci po problémech ve vedení opustil dosavadní hostování týmových stránek a vznikly nové týmové stránky i týmové fórum na adrese www.CzechNationalTeam.cz, kde naleznete spoustu informací o projektech i návody, jak na instalaci BOINC či jak se zapojit do jednotlivých projektů. Na fóru se můžete setkat se členy týmu, kteří vám kdykoliv rádi pomohou s problémy.

Rosetta@Home

Projekt se zabývá vývojem léků proti rakovině, ale i jiným chorobám souvisejícím s DNA a funkcí proteinů. Vědci z desítek tisíců lidských proteinů totiž podrobně znají funkci pouze omezeného počtu z nich. Jelikož však již známe kompletní lidský genom (tedy posloupnost nukleotidů a genů), úkolem projektu je na jeho základě vypočítat tvar a z něho přímo vyplývající funkci těchto proteinů v našem organismu. Výsledky jsou zdarma odevzdávány do obecné vědecké databáze, ze které mohou čerpat všechny výzkumné týmy na světě.

ProteinPredictor

Projekt se zabývá zkoumáním struktury proteinů v různých jejich variantách a mutacích s jasným eventuálním výsledkem – pomocí zmapovat co nejvíce z nich. Díky těmto znalostem lze účinně bojovat nejen proti různým

nemocem (např. rakovině), ale také zkoumat různé mutace, které se v přírodě vůbec nevyskytují.

Malaria control.net

Tento projekt se za použití modelů, do kterých jsou vkládána skutečná vstupní data na základě terénních výzkumů, snaží vytvořit předpověď šíření malárie v dané oblasti a také efektivnost použitých konkrétních druhů vakcín. Pomáhá tak nejen v boji s touto velice zákeřnou chorobou, ale i v rozhodování, kde a jaké léky použít, aby jejich účinek byl co nejefektivnější.

Leiden Classical

Dynamický BOINC projekt, jehož úkolem je pokusit se o vysvětlení některých základů vědy, například kvantové mechaniky. Velice zajímavé je, že můžete v tomto projektu podrobit výpočtům svou vlastní práci. Projekt by měl sloužit jak studentům, tak nadšencům, kteří mají vědu jako svůj koníček.

Jádro programu je složeno z většího množství univerzálních knihoven, které je možné využít pro váš vlastní vědecký program.

Hlavní program je v jádru jednoduchý a snadno přístupný pro vysokoškoláky a má jim pomoci seznámit se se základy počítačové simulace a dynamiky. Poskytuje možnost rozsáhlého použití OpenGL, ale může být i pouze zkompileovaný pro zpracovávání v textové verzi.

Jaká je hardwarová náročnost?

Projekty nejsou nijak přehnaně náročné ani na operační paměť (doporučováno je 256 MB, minimem je 128 MB), ani na objem přenesených dat prostřednictvím internetu. Stačí i obyčejný modem s občasným připojením pro surfování, které systém využije k aktualizaci dat. S místem na pevném disku si také nemusíte lámat hlavu. Klient běžně zabírá i se staženou prací pro několik dní zhruba 30 MB. Výjimku v tomto ohledu tvoří pouze projekt CPDN, který bez větších problémů při výpočtu jedné pracovní jednotky nakype až na několik GB. Spotřeba počítače vám také nemusí vrtat hlavou, jelikož k nárůstu dojde pouze ve spotřebě CPU, ale tento rozdíl je u běžných PC v řádu jednotek W. Rovněž operační systém nerozhoduje, jelikož existuje klient jak pro Windows, Linux/x86 či Macintosh OS X, tak i pro Solaris/SPARC. ■ ■ ■



SDÍLENÝ RENDERING: V projektu BURP se nezpracovávají vědecká data, ale výpočetně náročná grafika a animace