

## *Posudek na diplomovou práci*

*Daniel Vojtek*

### **Implementation of parallel query processing in PostgreSQL**

Práce se zabývá implementací vybraných paralelních operací v databázovém systému PostgreSQL, konkrétně třídění a operátoru UNION ALL. První z nich je příkladem intraoperátorového a druhý interoperátorového paralelismu a dohromady tedy tvoří poměrně reprezentativní vzorek aplikací paralelismu v databázích, pomineme-li *partitioning*, který nelze implementovat bez podpory v kompilační části systému. Téma paralelismu v databázích by při pečlivém zpracování vydalo i na práci doktorskou a není tedy divu, že autorem předkládané výsledky jsou spíše na úrovni experimentu s vybranými operacemi než kompletním či reálně použitelným dílem.

Vzhledem k pozornosti, která je v posledním desetiletí věnována paralelismu nejen v databázích, je překvapivé, že práce obsahuje pouze 7 referencí na literaturu, přičemž paralelismu se týká jediná. To nelze hodnotit jinak než jako hrubé podcenění rešeršní části práce.

Ani struktura textu neodpovídá zvyklostem u prací tohoto typu – po velmi obecném úvodu, kde autor vysvětluje základní pojmy z paralelních databází, následuje náhle kapitola s technickými detaily systému PostgreSQL a popisem implementace – chybí tedy dostatečně přesný a přitom technickými detaily nezatížený popis funkce a algoritmů obou zkoumaných operátorů v seriové i paralelní verzi. Rovněž popis provedených měření, jejich výsledky a dokonce i závěr práce jsou nelogicky utopeny uvnitř kapitoly Implementace – tím vzniká podezření, že autor nikdy správně strukturovanou diplomovou práci neviděl (o vědeckém článku nemluvě).

Text práce je psán anglicky, bohužel často na úrovni doslovného překladu z češtiny.

Samotná implementace díla je provedena formou modifikací a dodatků k zdrojovým kódům systému PostgreSQL. Autor se zjevně pokusil úpravy provést tak, aby bylo jeho dílo nadále udržovatelné spolu s hlavní větví systému, nicméně ne zcela úspěšně: Při kompilaci upraveného systému je nutné zasahovat ručně do souborů generovaných utilitou *configure* a navíc po provedených úpravách přestává být systém kompilovatelný mimo platformu linux. Především by však měl autor vysvětlit, proč po jeho zásahu přestává fungovat cca 30% regresních testů (*gmake check*).

Autorem prezentované výsledky měření ukazují, že implementace je funkční a za příznivých okolností vykazuje určité zvýšení výkonu. To se ovšem týká pouze operátoru UNION ALL, pro paralelizované třídění žádné testy v textu popsány nejsou. Autor navíc neuvádí žádné parametry stroje, na kterém byly testy provedeny, takže z nich fakticky není možné zhodnotit, jak úspěšná je implementace.

Autor bezpochyby prokázal nadprůměrnou odvahu při volbě obtížného tématu a též dobrou schopnost orientace v komplikovaném komunitně vyvíjeném systému. Bohužel při své práci téměř úplně ignoroval akademickou stránku diplomových prací, tedy rešerši, využívání cizích výsledků a vymezení se vůči nim, i správné strukturování textu.

Vzhledem k náročnosti tématu doporučuji práci k obhajobě i přes výše uvedené významné nedostatky.

24.1.2011

David Bednárek