

Posudok diplomovej práce: **Implementace paralelního zpracování dotazů v databázovém systému PostgreSQL**

Diplomant: **Daniel Vojtek**

Cieľom predkladanej diplomovej práce bolo implementovať paralelne niektoré z vykonávaných databázových operácií na PostgreSQL. V čase zadania diplomovej práce neboli v PostgreSQL žiadne z operácií vykonávané paralelne. Bolo potrebné si spraviť predstavu o možnostiach paralelizmu, čo všetko má zmysel paralelizovať, navrhnuť si architektúru paralelizmu a implementovať aj všetkú potrebnú infraštruktúru.

Kód databáze PostgreSQL je napísaný v jazyku C a jeho rozsiahlosť je v súčasnosti o niečo málo menej ako milión riadkov kódu (iba súbory \*.h a \*.c). Približne 600 tisíc riadkov kódu sa nachádza v databázovom backende, čo je časť, ktorá potenciálne môže súvisieť s touto prácou.

V čase zadania práce bola v Sun-e plná podpora snahe implementovať niektoré operácie v PostgreSQL paralelne. S kúpou MySQL Sun-om sa podpora PostgreSQL dostala do pozadia a následná akvizícia Sun-u Oracle-om spravila z tejto práce v pôvodnom Sun-e v podstate až prácu nežiadúcu. Tento vývoj sa výrazne prejavil na podpore, ktorú sme boli schopný riešiteľovi poskytnúť a musel preto pracovať samostatnejšie ako bol pôvodný zámer.

Riešiteľ počas práce vyskúšal implementovať paralelizmus najskôr pomocou vlákien. Rozsah potrebných zmien v už existujúcom kóde sa však ukázal príliš veľký a neistý, čo viedlo k prechodu na implementáciu pomocou procesov. Podobne sa ukázali nevhodné aj zvolené riešenia pri iných súvislostiach a veľakrát bolo potrebné kód prepísať úplne inak. Dopisovanie paralelnej funkcionality do pomerne rozsiahleho projektu výrazne zvyšovalo náročnosť práce.

Riešiteľovi sa podarilo úspešne navrhnuť infraštruktúru pre paralelné vykonávanie dotazu a implementovať v nej niektoré jednoduchšie operácie – triedenie a „union all“. Výsledné testy výkonnosti sú však príliš zjednodušené a na nie sú spravené na hardware vhodnom na testovanie paralelizmu databáz. Procesy, ktoré vykonávajú rôzne časti toho istého dotazu paralelne majú samostatne priradené číslo transakcie, čo môže viesť k nekonzistentnému chovaniu v transakcii. Odstránenie tohto problému by malo byť jednoduchšie v kóde novej verzie PostgreSQL. Pre konečného užívateľa databáze je implementácia zatiaľ príliš obmedzená a je potrebné ju ďalej rozširovať.

Práca overila možnosť implementácie paralelného spracovania dotazu v PostgreSQL, navrhla a implementovala potrebnú infraštruktúru. V práci je možné ďalej pokračovať a rozširovať implementáciu o ďalšie operácie s výrazne nižšou pracnosťou. Predkladanú prácu preto doporučujem uznať ako prácu diplomovú.

V Prahe, 24.1.2011



Mgr. Július Strottek