

Posudek vedoucího diplomové práce

Název DP: **Vícerozměrné indexování pro relační SŘBD**
Diplomant: **David Hoksza**

Obsah práce:

Předmětem diplomové práce (DP) bylo implementovat vícerozměrnou indexační metodu do stávajícího (komerčního nebo open-source) SŘBD. Předmětem implementace přitom nebylo použití vícerozměrné metody jako nástroje pro indexování speciálních dat (např. geometrií jako v případě GIS/CAD aplikací, což poskytují např. rozšíření Oracle, DB2), ale nativní vícerozměrné indexování složených klíčů klasických datových typů, kde indexovaný řádek tabulky je chápán jako bod ve vícerozměrném prostoru. Tato aplikace vícerozměrného indexování umožňuje efektivně (rychle) vyhodnocovat konjunktivní intervalové dotazy pomocí jediného indexu, a tudíž eliminovat buď sekvenční zpracování vynucené užitím indexu na zřetěžených attributech, anebo operace průniku výsledků z více indexů.

Cílem DP bylo prověřit, zda vícerozměrné indexování opravdu vykazuje významnou výkonostní výhodu oproti klasické technice zřetězení atributů. Přitom se nejedná o porovnání v laboratorních podmínkách na prototypch daných metod, ale přímo v rámci reálných SŘBD, které jsou léta vyvíjeny a optimalizovány. Takové srovnání je vypovídající nejen teoreticky pro účely základního výzkumu, ale svými výstupy může posloužit také odborné databázové veřejnosti. Je třeba podotknout, že cílem práce nebyla kompletní implementace vícerozměrného indexu tak, aby byl použitelný pro reálný provoz (tj. pro všechny datové typy a všechny operace s indexem), implementace byla z časových důvodů omezena pro potřeby experimentů.

Hodnocení:

Autor se zhostil nelehkého úkolu implementace vícerozměrné indexační metody (konkrétně R-stromu) do systému PostgreSQL (PG), která, jako open-source produkt, poskytuje širší možnosti customizace. Zde byl odveden časově nejnáročnější kus práce, neboť implementace “laboratorní” indexační metody do PG vyžadovala jednak důkladné seznámení se se softwarovou architekturou PG, jednak přizpůsobení metody různým omezením a rozhraním PG. Jako vedlejší produkt této snahy autor navrhl framework, který umožní případným dalším zájemcům o tuto problematiku zjednodušené přidávání dalších indexačních metod do PG.

Bylo rovněž potřeba prozkoumat a využít prostředků, kterými nabízejí srovnávané SŘBD statistiky o průběhu indexování a dotazování, a tyto “inter-platformní” statistiky využít pro korektní srovnání.

Hlavní (a pro mně nejcennější) částí práce pak byly experimenty, kde autor srovnává výkon jeho implementace vícerozměrného indexování v rámci PG s klasickým indexováním pomocí B-stromů v systémech PG, Oracle a MS SQL. Navíc se podařilo srovnat všechny systémy se SŘBD Transbase, což je v současné době jediný produkt nativně podporující vícerozměrné indexování složených klíčů pomocí UB-stromu, u jehož zrodu stáli Rudolf Bayer a Volker

Markl – autoři UB-stromu. Experimenty uvedené v DP přináší řadu zajímavých a dříve nepublikovaných poznatků, a to nejen o efektivitě zastoupených indexačních metod, ale také o vlivu optimalizací srovnávaných SRBD, které mohou výsledný výkon systému výrazně ovlivnit.

Text práce je zpracován přehledně, autor nejprve uvádí do problematiky indexování v SRBD a prezentuje nevýhody klasického a výhody vícerozměrného přístupu. Následuje popis základních struktur použitých v práci (B-strom, R-strom, UB-strom). Největší prostor v textu práce dostává popis architektury PG, možnosti poskytované pro implementaci uživatelských přístupových metod a jejich využití pro potřeby DP. Tento popis uvádí čtenáře do způsobu integrace externích přístupových metod a do jisté míry popisuje tuto integraci nejen v rámci PG, ale obecně v rámci libovolného SRBD. Podrobnější detaily jsou pak uvedeny ve dvou dodatcích. Rozsáhlá předposlední kapitola se věnuje experimentování a interpretaci výsledků, závěr práce nakonec shrnuje hlavní poznatky získané při implementaci a experimentování.

Formálně je práce na dobré úrovni, nicméně má dílčí nedostatky v gramatice a na některých místech ve stylistice.


Podrobnější připomínky, otázky, poznámky:

- 1) Autor by si měl zopakovat pravidla pro psaní 'í' a 'y'.
- 2) Některé uvedené reference nejsou v textu citovány.
- 3) Použití anglických výrazů tam, kde existují zavedené české ekvivalenty.

Závěr:

Práce splnila zadání a doporučuji ji k obhajobě.

V Praze dne 1. září 2006



RNDr. Tomáš Skopal, Ph.D.
vedoucí DP