

Posudek oponenta na diplomovou práci  
Vojtěcha Kolomičenka

**Analysis and Experimental Comparison of Graph Databases**

Grafické databáze (GDB) jsou bezpochyby atraktivním tématem informatiky aktuálním zvláště v poslední době. Předložená práce mapuje stav dané oblasti (typy a funkce GBD s příklady některých dostupných produktů – viz kap. 2), ale pojednává i testování GBD obecně (kap. 3) a o stavu příslušných testovacích nástrojů (benchmarks) pro GBD (kap. 4). Jádro práce pak tvoří kapitoly 5 a 6 popisující návrh systému BlueBench a výsledky realizovaných testů. Výsledky jsou zajímavé, přínosné a poskytují materiál pro další diskusi. Bez ohledu na barevné grafy, mohly být ale popsány strukturovaněji. Pouhý verbální způsob snižuje čitelnost a informativnost textu. Na druhé straně je třeba ocenit autorův přístup obecně, protože pro hlubší proniknutí do problematiky byl nucen zpracovat velké množství materiálu a to nejen článků, ale i dokumentací řady softwarových nástrojů.

Zvláštní ohodnocení vyžaduje kapitola 1. Ta uvádí GBD do kontextu tzv. NoSQL databází. Jde o nejméně zdařilou část práce, která jde spíše mimo GBD a ve své stručnosti vede k mnoha sporým otázkám, jak v používaných pojmech, tak v pojetí apod. Pro práci navíc tato část není vůbec přínosná, neobsahuje ani žádné citace. Úsilí mohlo být naopak věnováno dalším částem práce (např. čl. 1.6.5 - o grafových jazycích, které jsou popsány více než nedostatečně). Také kap. 2 - popisuje GDB, ovšem stručně a s nejasným významem některých převzatých vět, např. že „link“ je kombinace „map“ a několika „bitmap“, které dovolují jistou konverzi. Co to asi znamená?

Detailní poznámky a připomínky:

1 – úvod do NoSQL – podle této specifikace NoSQL by sem bylo možné řadit i OO, OR apod., což ovšem není pravda.

8<sub>11</sub>- co je míněno „a map of keys and their associated values“. Co jsou ty klíče ve vztahu k atributům (vlastnostem)?

9 – řadit RDF mezi datové struktury není příliš šťastné v kontextu GDB.

13<sub>3</sub>- hodnota objektu – co to je?

14 – co je to element type a identifikátor elementu?

29 – 4.4 – název je zavádějící. Ty databáze se neporovnávají (také proč?). Porovnávají se zřejmě možnosti zpracování grafů v nich reprezentovaných. Ale to je nepřesné. Museli bychom totiž vědět, *jak* je graf v relačním SŘBD reprezentován. A těch způsobů je celá řada (např. i v objektově relačních strukturách, hrany, uzly + sousedi apod.)

34 – Pojmu CreateIndexes nerozumím. Proč „is begun“? Co se provede, když žádné elementy v DB nejsou?

46 – MySQLGraph. To vytvořil autor práce? Odkud se MySQLGraph vzal? Je implementován relačně?

47 – obr. 6.1 Znamená-li přerušovaná čára nepovinné členství, jak mohou existovat hrany bez uzlů?

- zmínka o spojeních a vnitřních SELECTech (co to je?) je sice relevantní, ale existuje i rekurzivní SELECT, kterým lze grafové úlohy řešit. A ten může mít jinou, efektivnější implementaci.

Drobnosti:

13 – zkratka *GDB* je uvedena, ale není důsledně používána.

32 – *testované DB*. Na předchozích stránkách se používaly „*systemy*“.

21 – *relationships* vs. *relations* (na předchozích stránkách). Preferoval bych *relationships* (relace je přetížený pojem).

Systematický nedostatek: Velmi nepořádné odkazy a bibliografické záznamy. Např. v kap. 1 – odkazy do (placené) databáze DL ACM – to není seriózní přístup vůči čtenáři. Odkaz má vést na bibliografický záznam o článku v Bibliografii (včetně autorů atd.).

**Závěr:** Autor se zhostil úkolu, pronikl do problematiky a realizoval software podporující navržený přístup. Zkušenosti z implementace mohou být inspirující zejména pro další experimenty s grafovými daty. Doporučuji práci přijmout za práci diplomovou.

V Praze dne 17. 5. 2013

Prof. RNDr. J. Pokorný, CSc.

KSI MFF UK