

Hodnocení diplomové práce – oponent

Autor hodnocení: doc. Ing. Radim Bača, Ph.D.
Vedoucí diplomové práce: prof. Ing. Michal Krátký, Ph.D.
Oponenti: doc. Ing. Radim Bača, Ph.D.
Téma: Key-value databázové systémy
Verze ZP: 1
Student: Bc. Jan Jedlička

1. Splnění požadavků zadání.

Práci považuji za spíše lehčí. Práce odpovídá zadání, nicméně většina bodů je realizována spíše minimalisticky. Zejména druhý bod, kde měl student navrhnout a implementovat testovací prostředí, je v jeho podání realizován nastavením a zprovoznění existujícího benchmarku. To v zásadě není špatně, nicméně jak popisuji dále, chybí mi tam nějaký větší vklad ze strany studenta.

2. Hodnocení formální stránky závěrečné práce.

Text práce je hlavním výstupem celé práce, jelikož student nevytvářel vlastní kód a v příloze je tedy prakticky jen readme soubor. Z tohoto pohledu jsem měl ohledně textu vyšší očekávání než obvykle, která bohužel nebyla naplněna.

Za jeden problém považuji skutečnost, že student popisuje možnosti nastavení konzistence v jednotlivých NoSQL databázích nepřesně nebo neúplně. Konzistence je přitom ústřední koncept NoSQL databází. U Aerospike DBS se mluví o tom, že poskytuje silnou konzistenci pod čímž si představuji ACID. Nicméně v tabulce 2.1 je u Aerospike v položce týkající se konceptu transakcí nasáno pouze "atomické". Tak jak to je? Aerospike navíc nabízí nastavení módu konzistence (AP vs ACID), což text vůbec nezmiňuje. Na straně 36 je zmínka, že u RIAK KV musela být změněna konzistence, nicméně popis v kapitole 2.5.6 vůbec neuvádí, že RIAK KV něco takového umožňuje a tabulka 2.1 jen nalezneme údaj, že RIAK KV nepodporuje transakce. To je zcela nedostačující popis.

Za ještě důležitější nedostatek textu ale považuji to, že není jasně popsána použitá konfigurace DBS v testech. Kupříkladu již zmíněné nastavení konzistence. Je Aerospike nastaven na AP nebo ACID? Jistě to může mít vliv na výkon databáze a bylo by vhodné, aby databáze byly alespoň do jisté míry nastaveny podobně.

Co se týče jazykové stránky a vlastní prezentace výsledků tak v textu se vyskytují drobné nedostatky. Drobné překlepy jako použít slova letence, místo latence a také mi vadilo jiné pořadí DBS v tabulkách a grafech.

3. Hodnocení výsledků závěrečné práce.

Student připravil nasazený YSBC benchmark na virtuálním stroji, kde jsem jej také mohl vyzkoušet.

Výsledkům provedených testů se věnuje kapitola 4. Bohužel musím konstatovat, že výsledky jsou popsány velmi špatně. Text se omezuje na konstatování naměřených výsledků a v podstatě pouze slovně shrnuje hodnoty, které můžeme jednodušeji vidět v grafech. Vyhodnocení zcela minimálně interpretuje naměřené výsledky. V textu se stále opakuje, že Riak KV má výrazně horší propustnost čtení a zápisu, ale nějaká hlubší analýza proč to tak je, zcela chybí. Pouze na konci student spekuluje, že by se mohlo jednat o špatné nastavení konfiguračních souborů. V tom případě bych očekával test, kde se zkusí jiné nastavení, aby se tato hypotéza zkusila potvrdit nebo vyvrátit.

Také mi vadí, že zpracování vstupů aplikace je založeno na přepisování do Excelovské tabulky. Očekával bych nějaký krátký program, třeba v Pythonu, který výstup přečte, interpretuje, uloží a zobrazí grafy. Toto mi chybí i vzhledem k tomu, že student jinak nic neprogramoval.

4. *Hodnocení práce z hlediska přínosu nových poznatků.*

Práce měla potenciál provést zajímavé srovnání několika key-value DBS. Nicméně vzhledem k chybějícím informacím o nastavení DBS jsou výsledky spíše jen orientační.

5. *Charakteristika výběru a využití studijních pramenů.*

Práce cituje ohromné množství zdrojů. Některé se mi jevily jako spíše okrajové a do práce bych je neuváděl. Například reference 34 uvedená za pojmem o vysoké úrovni konzistence tento pojem stejně neupřesňuje.

6. *Otázky k obhajobě.*

Jaké nastavení konzistence jste využil u jednotlivých DBS při provádění experimentů? Byla nějaká distribuce? Propagovaly se změny do více replik?

Co se stane když u Riak KV zkusíte změnit konfiguraci? Má to vliv na výsledné časy?

7. *Souhrnné hodnocení.*

Práce má mnoho nedostatků, kde za nejvýraznější považuji nejasnou konfiguraci testovaných DBS. U studenta navazujícího studia očekávám hlubší vhled do problematiky. Práce je tedy z mého pohledu na hraně obhajitelnosti.

Celkové hodnocení: nevyhověl

Ostrava, 31.07.2024

doc. Ing. Radim Bača, Ph.D.

V IS EDISON odevzdáno: 31.07.2024 11:15:47