

Hodnocení diplomové práce – vedoucí

Autor hodnocení: prof. Ing. Michal Krátký, 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. Zadání závěrečné práce.

Náročnost práce je průměrná. Student neimplementoval vlastní testovací prostředí (bod 2 zadání), ale použil existující (YCSB - Yahoo! Cloud Serving Benchmark). Vlastní implementace testovacího prostředí byla tedy nahrazena konfigurací testovacího prostředí a implementací skriptů pro automatické spouštění testů.

2. Aktivita studenta během řešení.

Konzultace začaly ve větší míře probíhat až před odevzdáním práce, což pravděpodobně souvisí se zdravotními problémy studenta. Student zpočátku pracoval bez konzultací, což se odráží především v rešeršní části textu, která je místy nesrozumitelná.

3. Aktivita při dokončování.

Práci jsem před odevzdáním několikrát četl, některé problémy textu se podařilo vyřešit, některé věty rešeršní části jsou ale stále nejasné. V dotazech se na některé problematické věty ptám, nicméně takových vět jsou v rešeršní části desítky. V textu také zůstaly nevhodné pojmy: databáze vs databázový systém, KDBS často dosahují horizontální škálovatelnosti, datová struktura hash, atd.

4. Hodnocení výsledků závěrečné práce.

Student pro testování zvolil prostředí Docker v rámci kterého pak testuje jeden ze 4 databázových systémů klíč-hodnota. Test se spouští skriptem v PowerShellu, v rámci kterého dojde ke spuštění kontejneru, operačního systému, databázového systému, rámce pro testování a spuštění vytížení pro daný databázový systém. Zvolené řešení umožňuje opakování testu v rámci stejného prostředí.

5. Hodnocení práce z hlediska přínosu nových poznatků.

Práce nepřináší nové poznatky, výsledky porovnání čtyř databázových systémů mohou být zajímavé pro příslušnou komunitu.

6. Charakteristika výběru a využití studijních pramenů.

Seznam literatury zahrnuje 153 studijních pramenů, což je na diplomovou práci vysoký počet. Zdá se, že všechny studijní prameny jsou v práci citovány.

7. Souhrnné hodnocení.

Rešeršní část je místy problematická, připomíná firemní reklamní dokumenty a čtenář musí často přemýšlet na skutečným významem vět. Prostorů kolem spuštění testů je povedené, stejně jako zpracování výsledků testování. Student místy neprezentuje všechny výsledky práce, například bych očekával rozsáhlejší popis skriptů pro spuštění testů. Práci hodnotím jako velmi dobrou.

8. Otázky k obhajobě.

1. Nejasnost mám s výpočtem času provedení operace (zde nazváno latence). 10tis. operací se provedlo za cca 10s, propustnost je 953.3 ops/s. Čas provedení operace by tedy měl být cca 1ms, nicméně aplikace reportuje pro čtení i aktualizaci cca 4ms, proč?
2. Na str. 58 je uvedeno: Na posledním místě se umístila KDBS Riak KV [28], nejspíše kvůli špatnému nastavení konfiguračních souborů. Můžete vysvětlit?
3. Výsledky testování jsou uvedeny pro 4 vlákna. Jaký smysl by mělo testování pro různý počet vláken resp. zasílání operací z více počítačů?

4. Na str. 19 je uvedeno: DynamoDB podporuje různé režimy konzistence dat (ACID) a umožňuje zápis na více replik v rámci své infrastruktury. Tzn. DynamoDB podporuje např. úroveň izolace serializable dle SQL? Jak je zajištěna vysoká dostupnost?
5. Co je to datová struktura hash (str. 21)?
6. Na str. 22 je uvedeno: Hybrid Memory architektura od Aerospike je implementována tak, že index je čistě paměťový (in-memory) a klíč-hodnota páry jsou uloženy perzistentně na SSD disku [91] a čtou se přímo z něj. Jak je uložení dat paměťové, když je uloženo na SSD?
7. Na str. 22 je uvedeno: Aerospike tvrdí, že poskytuje server-side clustering [94] a bezpečnost na transportní vrstvě. Databáze umožňuje customer deployment [95] s nepřetržitým během. Můžete vysvětlit?
8. Na str. 23 je uvedeno: Repliky umožňují čtecí škálovatelnost, rychlý fail-over [79], hot-standby [101] a další distribuované konfigurace, dodávající podnikové prostředky ve vestavěném balíčku. Vysvětlíte význam.
9. Databáze má hashovatelné schéma (str. 25). Co to znamená?

Celkové hodnocení: velmi dobře

Ostrava, 29.07.2024

prof. Ing. Michal Krátký, Ph.D.

V IS EDISON odevzdáno: 30.07.2024 00:27:24