

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce: Martin Čorovčák

Název práce: Experimental Analysis of Query Languages in Modern Database Systems

Rok odevzdání: 2024

Studijní program: Informatika

Studijní obor: Informatika se specializací Databáze a web

Autor posudku: Ing. Pavel Koupil, Ph.D.

Role: Vedoucí

Pracoviště: Katedra softwarového inženýrství, MFF UK

Text posudku:

Cílem bakalářské práce bylo zaměřením se na různé databázové systémy (např. relační, dokumentové, sloupcové nebo grafové), vzájemně tyto systémy a jejich dotazovací jazyk(y) staticky porovnat a následně experimentálně porovnat výkon dotazování i s ohledem na škálovatelnost počtu uložených dat.

Konkrétně měl řešitel na základě statického porovnání vyjadřovací síly jednotlivých dotazovacích jazyků navrhnout scénáře dotazů a provést jejich dynamické (tj. experimentální) porovnání. Nakonec měl student navrhnout vhodná doporučení nebo diskutovat otevřené otázky a problémy plynoucí z výsledků experimentů.

K celé práci	lepší	obvyklé	horší	nevyhovuje
Obtížnost zadání	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Splnění zadání	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rozsah práce	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Práce se skládá ze 4 kapitol, úvodu a závěru. V úvodu je popsán kontext projektu a jsou vytyčeny hlavní cíle práce.

V kapitole 1 nazvané *Related work* řešitel diskutuje related work v oblastech benchmarků a měřený výkonu databázových systémů.

V kapitole 2 nazvané *Static Analysis* řešitel nejprve diskutuje cíle statické analýzy a způsob výběru NoSQL systémů. Následně velmi detailně porovnává vlastnosti vybraných databázových systémů a vyjadřovací sílu dotazovacích jazyků. Vyzdvihl bych především sekci 2.4 nazvanou Summary a tabulky 2.3 - 2.5, ve kterých je přehledně porovnána vyjadřovací síla vybraných dotazovacích jazyků (DDL a DML) a vlastnosti vybraných databázových systémů.

V kapitole 3 nazvané *ETL* řešitel diskutuje strukturu dat v jednotlivých databázových systémech s ohledem na vykonávané dotazy a ETL proces pro transformaci a import dat tak, aby výsledky měření byly porovnatelné. V této kapitole jsou také diskutovány dotazy, které se následně vykonávají. Dotazy jsou navíc vhodně klasifikovány do kategorií jako selekce, projekce, filtrování, agregace, spojování, třídění atd. Součástí kapitoly jsou také detailní statistiky ilustrující velikosti kolekcí v jednotlivých databázových systémech (měřeno v MB a počtu entit a typů) a tabulka s časy potřebnými pro import dat. To pouze svědčí o preciznosti řešitele a úplnosti řešení.

Hlavní přínos práce je pak v kapitole 4 nazvané *Dynamic Analysis*. Řešitel nejprve popisuje hardwarovou konfiguraci prostředí, ve kterém probíhaly experimenty. Následně detailně popisuje metodologii, ve které popisuje důvody pro opakované exekuce experimentů a vypořádání se s extrémní měření z důvodů externích a nekontrolovatelných vlivů. Řešitel se také zmiňuje o vypnutí vyrovnávací paměti na úrovni jednotlivých DBMS, aby výsledky měření byly porovnatelné a za stejných podmínek. Výsledky jsou pak vhodně ilustrovány pomocí grafů. V této kapitole bych vyzdvihl vizualizace 4.1 a 4.2 a především sekci 4.4 nazvanou Summary and Recommendations, ve které řešitel komentuje a porovnává výsledky měření a doporučuje vhodné systémy pro konkrétní scénáře dotazování.

V závěru řešitel shrnul výsledky práce a nastínil future work.

Práce je psána velmi dobrou angličtinou a sazba je profesionální. Řešitel cituje celkem

53 zdrojů. Celkový rozsah práce je také nadstandardní (96 stran).

Textová část práce	lepší	obvyklé	horší	nevyhovuje
Formální úprava	■	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Struktura textu	■	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Statická analýza	■	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dynamická analýza	■	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Součástí odevzdané práce jsou implementované ETL procesy, seznamy dotazů a skripty umožňující reprodukovatelnost naměřených časů.

Vzhledem k tomu, že se nejedná o klasickou implementační práci, hodnotím metodiku experimentů jako vhodnou pro daný účel. Dotazy v rámci jednotlivých experimentů byly vhodně navrženy s ohledem na měření výkonu konkrétního aspektu dotazování.

Implementační část práce	lepší	obvyklé	horší	nevyhovuje
Kvalita zpracování	■	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Metodika	<input type="checkbox"/>	■	<input type="checkbox"/>	<input type="checkbox"/>
Experimenty	■	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Reprodukovatelnost	<input type="checkbox"/>	■	<input type="checkbox"/>	<input type="checkbox"/>

Jedná se o velmi detailně a pečlivě provedené experimentální porovnání vybraných databázových systémů a dotazovacích jazyků. Zároveň má tato práce vysoký publikační potenciál.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

18. června 2024

Podpis: