

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce: Bc. Ondřej Zálešák

Název práce: Experimental Analysis of Querying in Modern Database Systems

Rok odevzdání: 2024

Studijní program: Informatika

Specializace: Programování a vývoj software

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 type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rozsah práce	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Práce se skládá z úvodu, čtyř kapitol a závěru. V úvodu je popsán kontext a hlavní cíle práce (tj. statická a dynamická analýza ve vybraných databázových systémech).

Kapitola 1 je věnovaná statické analýze systémů PostgreSQL, Virtuoso, OrientDB, ScyllaDB, Couchbase a RavenDB. Řešitel vzájemně (ale stručněji až výčtovou formou) porovnává vlastnosti těchto systémů a dotazovací jazyky podle vhodně zadaných kritérií. Silnou stránkou kapitoly jsou srovnávací tabulky 1.1 až 1.3, nicméně jejich výklad je příliš stručný a pouze popisný.

Kapitola 2 je diskuse ETL procesu pro transformaci a import dat z kolekce IMDb, která byla použita pro vykonání experimentů (viz kapitola 3). Součástí kapitoly je také popis a dokumentace k frameworku, který zjednodušuje integraci dat pro účely (opakovaných nebo podobných) experimentů.

Kapitola 3 je věnována experimentům. Řešitel postupně popisuje vykonávané dotazy, experimentální sestavu (tj. HW a SW konfiguraci), import datových sad do vybraných systémů a výsledky experimentů. Součástí je např. porovnání velikostí importovaných datových sad, času potřebného pro jejich import a především naměřené výsledky experimentů (dotazování). Bohužel, přestože metodika experimentů je vhodná, diskuse nad výsledky experimentů je především stručnějším porovnáním naměřením časů bez hlubší analýzy a doporučení.

Kapitola 4 je věnována souvisejícím pracem (velmi stručně).

V závěru řešitel shrnul výsledky práce

Součástí textu je také příloha obsahující jednotlivé dotazy.

Práce je psána dobrou angličtinou s přijatelným počtem překlepů. Sazba je na dobré úrovni. Práce s literaturou je adekvátní, celkem je citováno 39 zdrojů – především technické dokumentace, ale to je u tohoto typu práce pochopitelné.

Textová část práce	lepší	obvyklé	horší	nevyhovuje
Formální úprava	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Struktura textu	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Statická analýza	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dynamická analýza	<input type="checkbox"/>	<input checked="" 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 naměření výkonu konkrétního aspektu dotazování.

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

Zadání práce považuji za splněné. Student se práci věnoval průběžně a samostatně. Po technické stránce se jedná o slušně provedené experimentální porovnání vybraných databázových systémů a dotazovacích jazyků s dostatečnou mírou detailů umožňující opakovatelnost experimentů. Nicméně diskuse a závěry nad výsledky (experimentálního) porovnání jsou stručné a bez hlubší interpretace.

Celkové hodnocení Velmi dobře

Práci navrhuji na zvláštní ocenění Ne

18. srpna 2024

Podpis: