Univerzita Hradec Králové
Fakulta informatiky a managementu

POSUDEK OPONENTA DIPLOMOVÉ PRÁCE

Jméno studenta: Bc. Lukáš Kučera

Název práce: Využití Time Series databází v energetice

Autor posudku: Ing. Monika Borkovcová, Ph.D

Cíl práce: Cílem diplomové práce je porovnat Time Series databázi s ostatními zvažovanými typy databázových systémů a ověřit přínos ukládání energetických dat do tohoto specializovaného databázového systému. V praktické části práce bude cílem navrhnout a implementovat benchmark testy s využitím InfluxDB databáze a porovnat s benchmarky ostatních typů zvažovaných databázových systémů

Povinná kritéria hodnocení práce	**Stupeň hodnocení** (známka)					
	A	**B**	**C**	**D**	**E**	**F**
Práce svým zaměřením odpovídá studovanému oboru	☒	☐	☐	☐	☐	☐
Vymezení cíle a jeho naplnění	☐	☒	☐	☐	☐	☐
Zpracování teoretických aspektů tématu	☐	☒	☐	☐	☐	☐
Zpracování praktických aspektů tématu	☐	☐	☒	☐	☐	☐
Adekvátnost použitých metod, způsob jejich použití	☐	☐	☒	☐	☐	☐
Hloubka a správnost provedené analýzy	☐	☐	☒	☐	☐	☐
Práce s literaturou	☒	☐	☐	☐	☐	☐
Logická stavba a členění práce	☒	☐	☐	☐	☐	☐
Jazyková a terminologická úroveň	☒	☐	☐	☐	☐	☐
Formální úprava a náležitosti práce	☒	☐	☐	☐	☐	☐
Vlastní přínos studenta	☐	☒	☐	☐	☐	☐
Využitelnost výsledků práce v teorii (v praxi)	☐	☐	☒	☐	☐	☐

Vyjádření k výsledku anti-plagiátorské kontroly

Antiplagiátorská kontrola eVSKP identifikovala celkovou podobnost: 4%, která vychází z citací definic či obecných pojmů vztažených k řešené problematice. Celkové hodnocení plagiátorství je tedy není plagiát.

Dílčí připomínky a náměty:

K práci mám následující připomínky:

- V kapitole 6 „Benchmarkové testování“ mi chybí přesnější informace o testovaných verzích a edicích a způsobu nasazení jednotlivých databázových systémů (je zmíněna pouze InfluxDB ve verzi 2.x).
- S touto připomínkou souvisí i popis architektury konkrétní verze jednotlivých DBS, které se významně liší.
- U relačních databází autor nevyužil žádné optimalizační techniky a neuvedl proč.

- U MongoDB je zase uvedeno, že lze využít možnost horizontálního škálování, byla tedy využita? Jelikož u NoSQL databází je horizontální škálování převážně nativní záležitostí.
- Dále například u Oracle není popsáno, jak bylo pracováno s multitentatní architekturou (PDB) a zda byly využity možnosti, které nabízí?
- Jak docházelo mezi jednotlivými testovacími iteracemi k promazání databáze? Jak bylo naloženo s pamětí knihoven nebo vyrovnávací pamětí datového slovníku, apod.?
- ….

Celkové posouzení práce a zdůvodnění výsledné známky:

Stanovený cíl práce byl splněn, diplomová práce splňuje požadavky kladené na tento typ odborné práce. Z pohledu řešené oblasti a úrovni závěrečné práce se jedná o průměrně náročné téma. Předložená práce popisuje problematiku ukládání dat, jejichž původce je vymezená část energetického průmyslu zabývající se analýzou a predikcí (spotřeba, plán výroby, cenová politika, modely pro sledování výkyvů během a mimo sezónu, apod.), do timeseries databází. Textová část práce je zpracována s odpovídající úrovni odborného textu, má akceptovatelnou jazykovou a terminologickou úroveň, splňuje formální náležitosti a zdroje jsou řádně zpracovány.

Autor se nejdříve zabýval časovými řadami a jejich použitím na energetickém trhu. Rozbor časových řad byl poměrně komplexně představen, a to z časového hlediska, periodicity, typu dat a v podstatě i podle stacionarity. Autor neopomněl pro oblast energetiky podstatnou cyklickou složku časových řad ale i trendovou, sezónní a iregulární. V kapitole tři autor rozebírá energetických trh, kde stručně popsal rozdělení energetického trhu, principy, účastníky a tržní subjekty a Imbalance Settlement Model (ISM) pro efektivní fungování energetických trhů a sítí a dále se zaměřuje na princip zúčtování odchylek v kapitole 3.1.5.2.

Technický základ předložené diplomové práce lze nalézt od kapitoly 4, kde jsou obecně a stručně popsány time-series databáze, přičemž kaptiola 5 se zaměřila na InfluxDB. Stěžejní částí celé práce je kapitola 6 „Benchmarkové testování“, jejíž výsledky jsou popsány adekvátně, ale mnoho vstupních podmínek nutných pro pochopení celého testování zde není uvedeno viz dílčí připomínky. Výsledky tak mohou být použitelné v praxi, ale vzhledem k dynamickému a rychlému vývoji v oblasti databázových systémů mohou být také považovány za zastaralé nebo mohou být nevhodně zpracované vzhledem k možnostem jednotlivých verzí, a to převážně při testování výkonu jednotlivých databázových systémů, např. není exaktně uvedeno s jakými verzemi a edicemi databázových systémů autor pracuje, zda například využívá nativní timeseries kolekce u MongoDB, apod. viz dílčí připomínky.

Celkové hodnocení práce tak vychází právě z nedostatečně specifikovaných informací o použitých databázových systémech při benchmarkovém testování, i když ostatní části práce jsou zpracovány na kvalitní úrovni. Po připojení na uvedený GitHub repositář (https://github.com/kucerlu1/tsstoreapp) jsem tyto informace také nezískala.

Otázky k obhajobě:

Prosím o vyjádření k dílčím připomínkám a námětům.

Práci doporučuji k obhajobě.

Navržená výsledná známka: C

V Hradci Králové**, dne 30. srpna 2024**

podpis