Posudek oponenta bakalářské práce

Autor práce: **Petr Holický**

Název práce: **Uživatelské výpočty nad time-series databází**

Obsahově je text práce průměrný. Jako problematickou považuji hned druhou kapitolu textu, ve které jsou představeny hlavní reprezentanti tzv. *time-series* databází. U populárního systému *InfluxDB* je popisována *TICK stack* architektura, bohužel popis je doprovozen chybným obrázkem. Chybný popis systému *InfluxDB* je korunován obrázkem 2.2, na kterém je zachycena struktura dat, která správně přísluší databázovému systému *TimescaleDB*. Představení systému *TimescaleDB* je také problematické svým obrázkem. Ten rozhodně neukazuje relační datový model ale strukturu záznamu, kterou lze vložit do tabulky databáze. U všech představených systémů mi chybí výčet operací, které může programátor v databázi použít. Konstatovat, že do databáze lze data zapsat je nedostačující. Nelze z toho poznat, zda se jedná o vložení nových dat nebo jejich aktualizaci.

Čtvrtá kapitola popisuje strukturu ukládaných dat a vzorečků, které budou aplikovány na uložená data. Nejdůležitější kapitola teoretické části dokumentu bohužel obsahovala popis jen jediného vzorečku, i když odevzdaný archiv obsahuje 5 dalších vzorečků. Naštěstí to zachránila příloha A, která neobsahovala vzorečky, ale funkce, které mohou být v zápisu vzorečku použity. K zápisu vzorečku student zvolil prefixovou notaci bez vysvětlení proč.

Nejvíce ošizená je pátá kapitola, která se věnuje popisu klientské aplikace. Jednotlivé třídy jsou popsány pouze výčtem atributů a metod, kde každá metoda je stručně slovně popsána. Je škoda, že v celé kapitole není k dispozici jediný fragment zdrojového kódu vybraných metod.

První kapitola textu nezačíná na straně 1, ale stránkou 8. Protože poslední kapitola je zakončena na straně 42, nabízelo se ověření, jaký je rozsah dokumentu dle metodiky KIV – napočítal jsem 23 normostran. Na obranu studenta bych k rozsahu dokumentu připočetl celou přílohu A zahrnující bohatý výčet implementovaných pseudokonstant a funkcí.

V textu se nachází několik překlepů a zvláštní odkazování na obrázky a tabulky. Např. na str. 18: *„Při detailním výpisu času 3.1 je vidět“* není poznat, že se jedná o odkaz na Obr. 3.1. Podobných formulací je v textu více (strany 22, 23, 37, 38 a 39). Z internetu převzaté obrázky 2.4, 2.5, 2.7 až 2.10 mají černé pozadí a bílý text což výrazně snižuje jejich čitelnost na papíře. Vzhledem k *Apache* licenci těchto obrázků mohl student změnit jejich barvy pozadí a textu.

Programovým řešením bakalářské práce jsou dva programy napsané v jazyce *Java*. První program `SetupTestData.jar` je zmíněn jen v uživatelské příručce (Příloha B) a je použit pouze pro vytvoření potřebných tabulek v úložišti databáze *QuestDB* a jejich naplnění testovacími daty. Ty jsou součástí odevzdaného archivu. Druhým programem je klientská aplikace `questDB_Client.jar`, která nabízí dva způsoby využití: buď pro aplikaci vzorečků nad uloženými daty, anebo výpočet sumačních dat pro definované časové období. Klientská aplikace zahrnuje 4 třídy, zdrojové soubory obou programů jsou korektně komentovány a dostupné v odevzdaném archivu.

Pohledem na zdrojový kód klientské aplikace musím konstatovat, že není dobře dekomponován. Např. stěžejní metoda `selectData()` třídy `Client` obsahuje opakující se dva typy programových bloků, které se liší konstrukcí dotazu do databáze a zápisem odpovídajících dat do databáze. Samotná metoda obsahuje cca. 160 řádek programového kódu. Ještě více nepřehledná je metoda `parseFunction()`

třídy `EquationParser`, která obsahuje konstrukci `switch` na cca. 880 řádcích programového kódu.

Klientská aplikace vykazuje popisovanou funkcionalitu, jen časů uváděných v kapitole věnující se testování jsem nebyl schopen dosáhnout. Podle studenta použití klientské aplikace v režimu aplikování vzorečků nad testovacími daty (6 000 záznamů) trvá fáze výběru dat cca. 13 sekund. Ručním měřením času jsem při prvním puštění (vybírá 5 000 záznamů) ve fázi čtení naměřil cca. 90 sekund, při druhém (vybírá zbývajících 1 000 záznamů) cca. 6 minut a třetí spuštění (nevybere žádný záznam) překročilo 10 minut. Pokud jsem odpovídající dotaz spustil v konzoli databáze (nevybere žádný záznam), obdržel jsem výsledek také až po 10 minutách. Proto považuji studentem uvedené časové údaje v tabulce 6.1 za chybné.

Student použil pro napsání práce 14 elektronických zdrojů, které ve skutečnosti zahrnují 6 různých webů – 3x dokumentace představených *time-series* databází, 2x blog věnovaný těmto databázím a dokumentace *PostgreSQL* databáze. Všechny prameny jsou aktuálně dostupné a v dokumentu řádně citované.

Dotazy k práci

1. Aplikování vzorečků na dosud nezpracovaná data je realizováno jazykem Java nebo prostředky time-series databáze a proč?
2. Jak se klientská aplikace zachová, pokud bude z databáze načten vzoreček obsahující neznámou funkci či pseudokonstantu?

Zadání a zásady pro vypracování bakalářské práce student splnil i přes výše uvedené výhrady.

Navrhuji hodnocení známkou **dobře** a práci doporučuji k obhajobě.

V Plzni 26.5.2022 Ing. Martin Zíma, Ph.D.