



Posudek oponenta závěrečné práce

Student: Bc. Marek Tornóci
Oponent práce: Mgr. Jiří Toušek
Název práce: Metadata extraction, parsing, and dataflow detection in Snowflake sql dialect
Obor: Webové a softwarové inženýrství

Datum vytvoření: 21. 8. 2020

Hodnotící kritérium:	Způsob hodnocení – následující škálou 1 až 4:
1. Splnění zadání	<u>1=zadání splněno,</u> 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
<p><i>Popis kritéria:</i> Posuďte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.</p> <p><i>Komentář:</i> Práce splnila zadání.</p> <p>Implementované prototypové řešení je dostatečně kvalitní a splňuje všechny technické požadavky softwarového systému Manta, je tak možné jej přímo využít v praxi.</p>	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
2. Písemná část práce	90 (A)
<p><i>Popis kritéria:</i> Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3. Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.</p> <p><i>Komentář:</i> Rozsah práce odpovídá řešenému problému. Práce je informačně bohatá, dobře strukturovaná, neobsahuje zbytečné části.</p> <p>Přes určité množství gramatických chyb a nevhodně zvolených formulací je práce dobře srozumitelná. Chválím volbu ukázkových vstupů (SQL skriptů) v práci, pěkně ilustrují popisovanou problematiku a místy doplňují detaily v textu přímo neuvedené.</p> <p>Závažné věcné chyby jsem v práci nenašel. Některé drobnější nepřesnosti a logické chyby:</p> <ul style="list-style-type: none">- první odstavec kapitoly 1.2 a diagram 1.6 uvádí jako jediný účel extrakce metadat získání datového slovníku, získání definic objektů (DDL) je přitom srovnatelně důležité;- kapitola 1.3 uvádí, že uzly v grafu datových toků reprezentují data - v implementovaném řešení reprezentují uzly kromě dat též operace s nimi;- v seznamu v kapitole 2.1.3 je položka "SCHEMATA view" uvedena dvakrát;- nesprávný popis klauzule USING v kapitole 2.3.9;- kapitola 2.3.10 pracuje s pojmem "namespace", který nebyl zaveden a nikde jinde v práci se nevyskytuje;- seznam zkratk uvádí u zkratky ETL "Extract, Load, Transform" - má být "Extract, Transform, Load" (uvedený přístup existuje, jde ale o odlišný přístup, zkracuje se ELT).	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
3. Nepísemná část, přílohy	95 (A)

Popis kritéria:

Die charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů

Komentář:

Prototypová implementace dodržuje základní zásady psaní kvalitního kódu, je dostatečně čitelná a okomentovaná. Implementace dodržuje standardy a strukturu projektu Manta. Implementace je řádně pokryta testy.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

4. Hodnocení výsledků, jejich využitelnost

95 (A)

Popis kritéria:

Die charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Komentář:

Výsledky práce jsou přímo uplatnitelné v praxi.

Vzhledem ke kvalitě prototypové implementace a dodržení technických požadavků systému Manta lze bez větších problémů implementovaný prototyp převzít jako základ pro praktické řešení.

Písemná část poskytuje dostatečně podrobný přehled o struktuře metadatového katalogu databáze Snowflake.

Hodnotící kritérium:

Způsob hodnocení – nehodnotí se

5. Otázky k obhajobě

Popis kritéria:

Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:

Lze extrahovat DDL definice materializovaných pohledů? Podle kapitoly 2.2.2.7 se materializované pohledy místo ze systémového pohledu VIEW (který má dle práce sloupec VIEW_DEFINITION) extrahují z TABLES, v kapitole 2.2.2.3 však není uveden žádný sloupec, který by DDL definici obsahoval.

Hodnotící kritérium:

Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):

6. Celkové hodnocení

95 (A)

Popis kritéria:

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.

Text hodnocení:

Práce řeší všechny podstatné aspekty poměrně rozsáhlé problematiky extrakce metadat systému Snowflake a analýzy datových toků.

Autor musel zvládnout širokou škálu problematik od extrakce metadat z relativně mladé databázové technologie (kde objevil několik problémů, které lze považovat za chyby implementace databáze) po tvorbu gramatiky pro parsování SQL skriptů a interpretaci výsledků parsování.

Implementované prototypové řešení je technicky na vysoké úrovni, dobře navržené a splňuje všechny na něj kladené požadavky. Bylo provedeno ověření korektnosti implementace.

Teoretické závěry práce i implementované řešení jsou přímo využitelné v praxi.

Podpis oponenta práce: