



Posudek oponenta závěrečné práce

Student: Marek Burděl
Oponent práce: RNDr. Lukáš Hermann
Název práce: Design and implementation of data flow analysis in the StreamSets platform for Manta project
Obor: Webové a softwarové inženýrství

Datum vytvoření: 12. 6. 2020

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – následující škálou 1 až 4:</i>
1. Splnění zadání	<u>1=zadání splněno,</u> 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
<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.	
<i>Komentář:</i> Všechny body zadání byly splněny v popsaném rozsahu.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
2. Písemná část práce	85 (B)
<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.	
<i>Komentář:</i> Všechny části práce obsahují dostatečné množství informací pro pochopení zadání, návrhu, implementace i konkrétních výsledků. Práce neobsahuje faktické chyby, provedené rešerše obsahují informace platné k době vzniku práce. Práce je logicky rozdělená do pěti kapitol - první kapitola seznamuje se základními koncepty, druhá obsahuje analýzu nástroje StreamSets, seznam požadavků a použitých technologií, třetí návrh, čtvrtá implementaci datového modelu a analýzy datových toků a poslední testování prototypu. Jednotlivé kapitoly na sebe logicky navazují a postupně uvádí čtenáře do problematiky a jejího řešení. Práce obsahuje několik častěji se opakujících gramatických chyb v anglickém jazyce. Typografické zpracování je na vysoké úrovni. Práce obsahuje všechny náležitosti - abstrakty v obou jazycích, klíčová slova, obsah, seznamy obrázků, tabulek, použitých zkratk a literatury, závěrečné zhodnocení výsledků, přiložené CD se zdrojovými kódy a dokumentací. Všechny zdroje jsou řádně citovány a z textu je zřejmé, které části práce jsou implementovány studentem. Většina referencí je ve formě odkazů na webové stránky a je u nich uvedena doba, kdy byly informace z těchto zdrojů čerpány. Veškeré knihovny třetích stran byly využity v souladu s jejich licencí.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
3. Nepísemná část, přílohy	95 (A)
<i>Popis kritéria:</i> Dle 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	
<i>Komentář:</i> Práce obsahuje prototyp implementace analýzy datových toků v technologii StreamSets Data Collector. Pomocí prototypu lze extrahovat definice pipeline ze serveru StreamSets do souborů, načíst je do datového modelu a provést nad nimi analýzu datových toků, vracející graf datových toků. Všechny části implementace jsou funkčně otestované. Zdrojový kód je přehledný, až na výjimky okomentovaný a využívá dostupné knihovny. Prototypová implementace dostatečně pokrývá analyzované problémy a je dobře rozšiřitelná.	

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
4. Hodnocení výsledků, jejich využitelnost	95 (A)
<i>Popis kritéria:</i> Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.	
<i>Komentář:</i> Výsledky práce mohou být obratem použity v referovaném projektu pro produkční nasazení.	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – nehodnotí se</i>
5. Otázky k obhajobě	
<i>Popis kritéria:</i> Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odřázkami).	
<i>Otázky:</i> 1. V úvodu práce zmiňujete existenci topologií, které mohou obsahovat pipeline, jejichž analýzu dále popisujete. Můžete ve stručnosti popsat, jak existence topologií ovlivní vámi navržené algoritmy, především field's expansion. 2. V analýze uvádíte, že fieldy jednotlivých stage mohou být neatomické, tedy že mohou obsahovat další fieldy. Jak se při analýze datových toků vypořádáváte s přiřazením celých neatomických fieldů a přiřazením jejich vybraných částí? 3. Jako jeden z nefunkčních požadavků uvádíte výkon, konkrétně že analýza jedné pipeline by neměla při splnění minimálních hardwarových nároků trvat déle než 5 sekund. V práci ale není uvedeno, že byste performance testy prováděl. Můžete uvést, zda váš prototyp splnil zadaný požadavek na výkon?	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
6. Celkové hodnocení	90 (A)
<i>Popis kritéria:</i> 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.	
<i>Text hodnocení:</i> Předložená závěrečná práce obsahuje všechny podstatné části pro návrh a implementaci prototypu analýzy datových toků pro nástroj StreamSets Data Collector. Všechny části práce jsou zpracovány ve vysoké kvalitě z hlediska předložených informací, formálních náležitostí, použitých zdrojů i prototypové implementace, díky čemuž lze výsledky práce obratem využít v referovaném projektu, celkový dojem ovšem kazí gramatické chyby.	

Podpis oponenta práce: