



Posudek oponenta závěrečné práce

Student: Vladyslav Zavirskyy
Oponent práce: RNDr. Lukáš Hermann
Název práce: Design and implementation data flow analysis of jobs in IBM DataStage for Manta project
Obor: Webové a softwarové inženýrství

Datum vytvoření: 4. 6. 2019

Hodnotící kritérium:	Způsob hodnocení – následující škálou 1 až 4:
1. Splnění zadání	1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno
Popis kritéria: 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.	
Komentář: Všechny body zadání byly splněny v popsáném rozsahu.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
2. Písemná část práce	75 (C)
Popis kritéria: 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.	
Komentář: 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 požadavků, použitých technologií a nástroje IBM InfoSphere DataStage, třetí návrh datového modelu a analýzy datových toků, čtvrtá popisuje implementaci a poslední testování prototypu. Jednotlivé kapitoly na sebe logicky navazují a postupně uvádí čtenáře do problematiky a jejího řešení. Práce obsahuje velké množství gramatických chyb v anglickém jazyce, které v některých částech znesnadňují porozumění předávaných informací. Typografické zpracování je na vysoké úrovni, občas se vyskytuje přetečení řádky textu nebo obrázku. Práce obsahuje všechny náležitosti - abstrakty v obou jazycích, klíčová slova, obsah, seznamy obrázků, použitých zkratk a literatury, závěrečné zhodnocení výsledků, příložené CD se zdrojovými kódy a dokumentací. Všechny zdroje jsou řádně citovány a z textu je zřejmé, které části práce jsou implementovány studentem. Na mnoha místech je citace uvedena doslovně, i když by bylo lepší ji přeformulovat vlastními slovy a zkrátit pro potřeby textu. Skoro všechny reference jsou 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.	
Hodnotící kritérium:	Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):
3. Nepísemná část, přílohy	90 (A)
Popis kritéria: 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	
Komentář: Práce obsahuje prototyp implementace analýzy datových toků v technologii IBM InfoSphere DataStage. Pomocí prototypu lze úlohy v podobě XML souborů načíst 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 dosahuje produkční kvality.	

<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
4. Hodnocení výsledků, jejich využitelnost	90 (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 po přidání doplňujících funkcionalit 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 odrážkami).	
<i>Otázky:</i> 1. Ve vaší práci předpokládáte, že jsou vstupní soubory dodány ve formátu XML uživatelem. Analyzoval jste také možnost automatické extrakce těchto souborů přímo z nástroje IBM InfoSphere DataStage pomocí nějakého API? Pokud ano, můžete stručně popsat, jak byste toto API využil? 2. V kapitole věnované analýze představujete funkcionalitu nástroje DataStage, kterou v rámci prototypu nebudete pokrývat - Runtime column propagation. Můžete stručně uvést, jak byste analýzu této funkcionality implementoval za předpokladu, že znáte reálnou strukturu vstupních dat, která je odlišná od očekávané? 3. V kapitole věnované implementaci popisujete podrobněji analýzu stage Sort. Mohl byste stručně uvést implementaci komplexnější procesní stage, např. Join nebo Pivot?	
<i>Hodnotící kritérium:</i>	<i>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</i>
6. Celkové hodnocení	85 (B)
<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 IBM InfoSphere Datastage. 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 časté gramatické chyby a velké množství doslovných citací.	

Podpis oponenta práce: