

Posudek oponenta závěrečné práce

České vysoké učení technické v Praze

Fakulta informačních technologií

Student: Bc. Robert Kotlář
Oponent práce: Ing. Magda Friedjungová
Název práce: Datový sklad ČVUT - způsoby datové integrace
Obor: Webové a softwarové inženýrství

Datum vytvoření: 5. 6. 2017

Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 5:
1. Náročnost a další komentář k zadání	1=mimořádně náročné zadání, 2=náročnější zadání, 3=průměrně náročné zadání, 4=lehčí, ale ještě dostatečně náročné zadání, 5=nedostatečně náročné zadání
Popis kritéria: Podrobněji charakterizujte diplomovou (bakalářskou) práci a její případné návaznosti na předchozí nebo běžící projekty. Dále posuďte, čím je zadání této ZP náročné. (U obtížnější ZP lze dále tolerovat některé nedostatky, které by u ZP standardní obtížnosti tolerovány nebyly; a naopak u jednoduché ZP mohou být zjištěné nedostatky hodnoceny přísněji.)	
Komentář: Student se musí v práci vypořádat hned s několika požadavky, které vyžadují určitou technickou zdatnost. Obzvláště pak návrh a implementace staging area a ETL procesů pro nahrávání dat ze zdrojových systémů považují za časově náročné úkoly, k jejichž úspěšné realizaci je nezbytná i dávka předchozích zkušeností. Návrh vlastní historizační procedury také skýtá mnohá úskalí. Vzhledem k tomu, že mají být výsledky práce nasazeny v ostrém provozu datového skladu ČVUT, považují zadání za náročnější.	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
2. 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 splňuje zadání. V komentáři uveďte body zadání, které nebyly zcela splněny, případně rozšíření ZP oproti původnímu zadání. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.	
Komentář: Zadání práce bylo v celém rozsahu splněno a výstupy práce jsou plně funkční.	
Hodnotící kritérium:	Způsob hodnocení - následující škálou 1 až 4:
3. Rozsah písemné zprávy	1=splňuje požadavky, 2=splňuje požadavky s menšími výhradami, 3=splňuje požadavky s většími výhradami, 4=nesplňuje požadavky
Popis kritéria: Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.	
Komentář: Písemná zpráva splňuje doporučený rozsah.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
4. Věcná a logická úroveň práce	88 (B)
Popis kritéria: 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.	
Komentář: Práce je logicky členěna. Student poměrně velkou část práce věnuje teoretické části, která by si zasloužila menší pozornost, a to především z důvodu, že jsou v ní definovány opravdu základní pojmy týkající se datových skladů obecně. Student v rešerši nezmiňuje konkrétní v praxi použitá řešení, což může být způsobeno i tím, že firmy svá řešení běžně nezveřejňují. V této sekci se občas vyskytují tvrzení jako "nejpoužívanější" apod., která by si zasloužila referenci. K praktické části nemám žádné výhrady. Student pečlivě a věcně popisuje návrh a implementaci jednotlivých bodů zadání. Navržená řešení v kapitole 4 testuje a potvrzuje jejich funkčnost.	
Hodnotící kritérium:	Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):
5. Formální úroveň práce	87 (B)
Popis kritéria: Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 14/2015, článek 3.	

Komentář:

Práce je psána srozumitelně, nicméně po stylistické stránce by se na textu dalo ještě zapracovat. Po gramatické stránce se v práci vyskytuje minimum chyb a překlepů (bohužel např. v obrázku 1.2). Některé tabulky (např. 1.1, 1.2 aj.) a obrázky (např. 2.1, 2.2 aj.) přetékají za okraje stránky. Obrázky 1.2, 1.3 a 1.4 nejsou seřazeny podle číslování. Některá použití tučného písma nejsou na místě (např. str. 27, zkratky V3S a SSP). Text v obrázcích 2.4, 2.5, 2.6 je hůře čitelný kvůli jeho velikosti.

Hodnotící kritérium:

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

6. Práce se zdroji

90 (A)

Popis kritéria:

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení ZP. Charakterizujte výběr studijních pramenů. Posuďte, zda student využil všechny relevantní zdroje nebo zda se pokoušel řešit již vyřešené problémy. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a uváž, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Komentář:

Pro řešení problému student využil relevantní zdroje včetně těch, které jsou pro danou problematiku stěžejní. Využívá jak zdrojů anglických, tak českých. Dále cituje i jiné závěrečné práce, na které svojí prací navazuje. Bohužel pro danou problematiku není mnoho dostupných akademických zdrojů, a tak reference nejsou tak bohaté, jako je tomu u jiných závěrečných prací.

Student občas volí nesprávné položky v šabloně pro citace, a tak jsou některé citace nesourodé.

Hodnotící kritérium:

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

7. Hodnocení výsledků, publikační výstupy a ocenění

100 (A)

Popis kritéria:

Vyjádřete se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvořil sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. Popište případnou publikační činnost a získaná ocenění související s řešením této ZP.

Komentář:

Práce je velmi cenná svojí funkčností, která byla ověřena nejen v kapitole 4, ale skutečným nasazením v projektu Datový sklad ČVUT. V rámci práce se podařilo zrychlit nahrávání dat z původních 8-12 hodin na 1 hodinu, což je zásadní rozdíl.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

8. Komentář o využitelnosti výsledků

Popis kritéria:

Uveďte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uveďte možnosti využití výsledků ZP v praxi.

Komentář:

Student splnil všechny požadavky zadání a své řešení úspěšně nasadil v praxi. Navrhl šablonu ETL procesů pro plnění datového skladu ČVUT, která bude v budoucnu dále využívána. Stejně tak bude využívána i navržená historizační procedura, která výrazně přispěje k naplnění cílů datového skladu ČVUT. Díky ní bude možné uchovávat historická data a dále s nimi operovat. Automatizace celého procesu nahrávání dat do datového skladu ČVUT včetně monitoringu ETL procesů je beze sporu velkou přidanou hodnotou této závěrečné práce.

Hodnotící kritérium:

Způsob hodnocení - nehodnotí se

9. Otázky k obhajobě

Popis kritéria:

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

Otázky:

- K rozpoznávání změn v záznamech využíváte MD5 hash. Z jakého důvodu jste zvolil právě MD5 hash? Mají na vaši práci dopad bezpečnostní rizika této hashe?
- Pro hledání změn v záznamech jste použil jazyk SQL a pro nahrání pre-stage clean vrstvy využíváte vlastní PL/pgSQL skripty. Napříč prací používáte open-source produkty společnosti Pentaho, ve kterých jste si mohl jednotlivé moduly dle vlastních potřeb upravit - proč jste nevyužil této možnosti?

Hodnotící kritérium:

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

10. Celkové hodnocení

90 (A)

Popis kritéria:

Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nesmí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

Text hodnocení:

Práce je velmi kvalitní, její implementace pomohla hned v několika bodech rozvojového projektu Datový sklad ČVUT a má tak praktické využití. Datový sklad tak může poskytovat kvalitní data, na základě kterých bude možné zpracovávat relevantní analýzy a reporty.

Podpis oponenta práce: