



Posudek oponenta závěrečné práce

Oponent práce:	Ing. Pavel Karol
Student:	Adam Marhefka
Název práce:	Paralelizácia ETL procesov DW ČVUT s využitím nástroja Pentaho
Obor / specializace:	Webové a softwarové inženýrství, zaměření Softwarové inženýrství
Vytvořeno dne:	2. června 2023

Hodnotící kritéria

1. Splnění zadání

► [1] zadání splněno

[2] zadání splněno s menšími výhradami

[3] zadání splněno s většími výhradami

[4] zadání nesplněno

Student předkládá komplexní, vysoce specializovanou bakalářskou práci. BP obsahuje prvky modelování databází, business intelligence, softwarového inženýrství včetně následného programování navrhnuté aplikace.

Cíle této práce jsou částečně společné s konkurenční BP studentky Kristiny Zolochovské. Společně s druhou prací student identifikuje požadavky na paralelizaci ETL procesů Datového skladu ČVUT (dále jen DS) a společný je také výběr části zpracování, která se má optimalizovat do paralelního běhu k porovnání výkonnosti obou řešení.

Dále již tato práce má svoji část zadání, kdy student pracuje s aktuálně využívaným ETL nástrojem DS Pentaho Data Integration (dále jen PDI). Navrhuje a implementuje své řešení paralelizace ETL procesů a nakonec zhodnotí realizované řešení.

Zadání práce bylo tedy v plném rozsahu splněno.

2. Písemná část práce

95 /100 (A)

Práce je tvořena úvodem, vytyčenými cíli práce, 7 kapitolami a závěrem. Rozsahem odpovídá náročnému zadání. Má jazykově odpovídající úroveň a je po formální stránce v pořádku. Práce je psána ve slovenském jazyce, tudíž kvalitu textu nejsem schopen posoudit. Nicméně je logicky a přehledně strukturována s využitím dostatečného množství relevantních elektronických i knižních zdrojů.

V prvních třech kapitolách je obsažena teoretická část práce, kdy kapitola 1 je popis datových skladů obecně, kapitola 2 pak popisuje nástroj PDI. Kapitola 3 již popisuje aktuální nastavení datového skladu ČVUT a všechny jeho procesy. Tato kapitola je

perfektně zpracovaná a vhodná k využití v dokumentaci DS. S teoretickou částí práce jsem velmi spokojen.

Praktická část začíná 4. kapitolou a končí v kapitole 6. Kapitola 4 obsahuje detailní analýzu současného řešení a jeho možností. Na základě této analýzy student v 5. kapitole navrhuje řídicí aplikaci, kterou řeší zadaný problém paralelizace ETL procesů DS. Aplikace navíc počítá s obecným použitím tak, aby uměla spustit nejen soubory PDI, ale obecně jakýkoliv skript. V kapitole 6 následně tuto aplikaci implementuje a testuje.

Kratší kapitola 7 následně obsahuje zhodnocení řešení a jeho využití. Upozorňuje na signifikantní zrychlení zpracování, což byl jeden z primárních cílů zadání práce. Dále také ale upozorňuje na problém „black box“, tedy že dodaná aplikace do budoucna potřebuje programátorskou údržbu a rozvoj, což nemusí být aktuálně pro DS vhodné.

3. Nepísemná část, přílohy

95 /100 (A)

Přílohy bakalářské práce jsou rozsáhlé, vzhledem k řešení ale adekvátní. Student přikládá archiv se všemi zdrojovými kódy aplikace. Implementované řešení je plně funkční a otestované.

4. Hodnocení výsledků, jejich využitelnost

90 /100 (A)

Tato práce má dobrou využitelnost v praxi. Neoddiskutovatelný přínos má detailní analýza současného řešení. Vytvořená aplikace přináší novou funkcionalitu pro DS ČVUT z pohledu zpracování dat, velké zrychlení současného řešení a lepší možnost údržby. Zároveň se ale jedná o aplikaci, na jejíž rozvoj do budoucna v DS nemusí být kapacita.

Celkové hodnocení

95 /100 (A)

Práci navrhuji hodnotit klasifikačním stupněm A. Dle mého názoru přesahuje rámec bakalářské práce, je zpracována velmi kvalitně. Má velký přínos pro celou ČVUT. Umožňuje totiž do budoucna zrychlit zpracování DS a otevírá možnost jeho denní aktualizace dat, čímž zpřesní požadované výstupy.

Otázky k obhajobě

V textu jste zmínil, že jste použil aktuální verzi Pentaho DI. Na jaké jste narazil problémy v průběhu upgrade?

Existuje v aktuální verzi Pentaho nějaká nová možnost optimalizace zpracování?

Instrukce

Splnění zadání

Posudte, 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, posudte 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.

Písemná část práce

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, 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. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posudte, 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.

Nepísemná část, přílohy

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.

Hodnocení výsledků, jejich využitelnost

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.

Celkové hodnocení

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.