



# Posudek oponenta závěrečné práce

**Student:** Peter Kolárovec  
**Oponent práce:** Ing. Magda Friedjungová  
**Název práce:** Využití metodiky Data vault pro ETL procesy  
**Obor:** Znalostní inženýrství

**Datum vytvoření:** 9. 6. 2019

<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení – následující škálou 1 až 4:</b>
<b>1. Splnění zadání</b>	<b>1=zadání splněno, 2=zadání splněno s menšími výhradami, 3=zadání splněno s většími výhradami, 4=zadání nesplněno</b>
<b>Popis kritéria:</b> 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.	
<b>Komentář:</b> Zadání je v plném rozsahu splněno.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</b>
<b>2. Písemná část práce</b>	<b>88 (B)</b>
<b>Popis kritéria:</b> 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.	
<b>Komentář:</b> Práce je logicky členěna. Teoretická část neobsahuje zbytečné informace, čtenář je dostatečně seznámen s kontextem potřebným k pochopení praktické části. Některé formulace, uvození sekcí, odkazování apod. nejsou nejšťastnější. Tabulky v sekci 5 mi přijdou informačně nezajímavé. Nicméně jsem nenarazila na žádný větší zádrhel. Seznam literatury je dostačující. Student úlohu dovedl do zdárného konce a svoji implementaci náležitě otestoval.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</b>
<b>3. Nepísemná část, přílohy</b>	<b>98 (A)</b>
<b>Popis kritéria:</b> 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	
<b>Komentář:</b> Student svojí práci chtěl navázat na již vyvinutý nástroj ETLTools, což je Python aplikace komunikující s Microsoft SQL Serverem. Přestože nakonec zvolil řešení nezávislé na stávajícím, jako programovací jazyk zachoval Python. Jako databáze byla zvolena MySQL. Oba výběry považuji za vhodné (více bych uvítala PostgreSQL kvůli jeho uživatelské základně). Kód by mohl být více komentovaný. Řešení je poněkud kompromisem (moduly jako obecné šablony) kvůli multiplatformnosti, což je velká přidaná hodnota. Jiné výhrady nemám, navržené řešení je funkční.	
<b>Hodnotící kritérium:</b>	<b>Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):</b>
<b>4. Hodnocení výsledků, jejich využitelnost</b>	<b>90 (A)</b>
<b>Popis kritéria:</b> 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.	

**Komentář:**

V této práci byl implementován ETL nástroj, který je dále využitelný. A to jak ve společnosti, pro kterou student nástroj navrhl, tak i ve společnostech/projektech jiných, po drobných úpravách. Na práci lze navázat, potenciál vidím minimálně v automatizaci plnění metadatových tabulek, způsobu verzování, rozšíření šablon apod.

*Hodnotící kritérium:*

*Způsob hodnocení – nehodnotí se*

## 5. Otázky k obhajobě

*Popis kritéria:*

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

*Otázky:*

- 1) Proč jste zvolili metodiku Data Vault a ne standardní relační modelování? (motivace, stručné zhodnocení výhod a nevýhod)
- 2) Jakým způsobem jsou metadata tabulky plněny?
- 3) Jak byste řešil situaci, kde klíč mappingu (ETL) je jiný než primární klíč? Případně, že 2 mappingy do stejné tabulky mají jiný klíč.

Příklady: Primární klíč historizované dimenze je surrogate key + valid\_to. Data do této tabulky ale nevkládáte podle tohoto klíče. Tabulka F\_SALE (F\_SALE\_SK,SALE\_ITEM\_SK,SALE\_DATE,PROFITABILITY\_SUM,REVENUE\_SUM) - PK je F\_SALE\_SK (IDENTITY (1, 1)). Mapping je však přírůstkový, delete insert pattern s klíčem SALE\_DATE (nebo SALE\_DATE a SALE\_ITEM\_SK). Delete insert pattern z toho důvodu, že se stává, že dodavatel dat převydá nějaký soubor a vy musíte přepočítat konkrétní prodejní den (nebo den a produkt).

*Hodnotící kritérium:*

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

## 6. Celkové hodnocení

92 (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 se věnuje metodice modelování Data Vault. Má praktický přínos - návrh a implementaci ETL nástroje, který má být k dispozici volně dalším uživatelům. Po textové i implementační stránce splňuje všechny parametry bakalářské práce a s výstupem jsem spokojená. Na práci lze navázat. Navrhuji práci hodnotit známkou A.

Podpis oponenta práce: