



# Posudek oponenta závěrečné práce

<b>Oponent práce:</b>	Ing. Petr Košvanec
<b>Student:</b>	Illia Krauchenia
<b>Název práce:</b>	Návrh a prototypová implementace analýzy datových toků a procesů nástroje Matillion ETL do nástroje Manta
<b>Obor / specializace:</b>	Webové a softwarové inženýrství, zaměření Softwarové inženýrství
<b>Vytvořeno dne:</b>	6. února 2023

## Hodnotící kritéria

### 1. Splnění zadání

- ▶ [1] zadání splněno
- [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Práce bez výjimek splňuje zadání. Cíle jsou v úvodu vydefinovány v souladu se zadáním a splněny. Oceňuji, že student během práce aktivně konzultoval s vedoucím i oponentem.

### 2. Písemná část práce

80/100 (B)

Vypracovaný text je vhodného rozsahu a má výborně rozvržené kapitoly, ve kterých jsou pokryty všechny důležité aspekty softwarového inženýrství. Méně prostoru mohla dostat sekce vybraných technologií, neboť je většina z nich obecně známá a navíc jejich výběr byl dán platformou.

V práci je jen malé množství překlepů (detekovány 3), nevhodných slov nebo slovních spojení. Zkratky i pojmy jsou vysvětlovány dobře a až na drobné výjimky (SQL dialekt, Workflow) před jejich použitím nebo přímo po něm.

Nenašel jsem, co bych práci vytkl z hlediska typografie a i jazykově je práce výborná, což pomáhá k velmi dobré čtivosti velké většiny kapitol. Jedinou slabší kapitolou byl popis implementace, kde mohlo být místo většího množství kódu raději více diagramů.

Použité vizualizace a diagramy byly přínosné. Zklamala mě absence vývojového diagramu nebo pseudokódu při popisu algoritmu propagace a expanze. Jeho popis byl zmatený a nepřehledný. Formálních zápisů je v práci méně, než by bylo vhodné, např. Graf datových toků si jeden zasloužil.

Práce je obsahově velmi kvalitní a obsahuje málo faktických chyb nebo nepřesností. Uvedu například, že není uvedena závislost modulu dataflow-generator na connector-resolver, jenž je v kódu patrná a také je chybně vysvětlená motivace hlavního algoritmu expanze sloupců. Sekce funkčních a nefunkčních požadavků se mi celkově moc nelíbila, F2 popisuje spíše vnitřní implementaci než funkčnost, F4 měl být rozpadnut, F7 rozšířen a požadavek na end-to-end testy plynoucí z N5 např. nebyl žádoucí.

Citace jsou v textu dobře viditelné a s příslušnou referencí. Forma citací je v souladu s citační normou. Bylo použito 38 online zdrojů a 1 diplomová práce, zajímavým přídavkem mohl být např. tištěný zdroj věnující se problematice data lineage nebo ETL, ale práce tím příliš netrpí.

### 3. Nepísemná část, přílohy

95 /100 (A)

Dodaný kód je až produkční kvality. Je velmi dobře členěný a respektuje OOP. Kód je extrémně dobře udržovaný, statická analýza nástrojem SonarLint neodhalila žádnou chybu nebo zranitelnost. V modulech je jen pár identifikovaných "code smells", z nichž ale asi jen 3 identifikují drobnosti, které by se daly zlepšit. V modulu dataflow-generator dokonce není detekována žádná duplicita.

Je patrná snaha komentovat kód s využitím Javadoc, ale na několika místech je dokumentace neúplná, např. v řadě tříd resolveru chybí komentáře k public metodám.

Kód je vhodně pokryt testy. Zdánlivě nízké pokrytí connector modulů je zjevně způsobeno formou testů - jsou spíše integrační povahy, kdy funkčnost connector-resolver je efektivně testována v jiném artefaktu, což je ale z hlediska neduplikace testových vzorků žádoucí. Pokrytí dataflow-generator je naopak se 74% nadprůměrné v porovnání s jinými moduly v platformě. Výtku mám k nešikovné práci s cestami k testovým vzorkům (měly být použity cesty ke zkopírovaným vzorkům mavenem), což může potenciálně způsobit jejich nechtěný přepis.

Implementace respektuje platformu a dostupné technologie. Je evidentní snaha o budoucí snadnou rozšiřitelnost, i když občas to trochu hraničí s porušením principu YAGNI.

Líbí se mi, že connector-model poskytuje immutable rozhraní, což není u ostatních scannerů vždy zaručeno, přestože je to žádoucí.

Dané přílohy nedaly práci už nic moc navíc, ale nebylo to potřeba.

### 4. Hodnocení výsledků, jejich využitelnost

90 /100 (A)

Prototyp je připravený pro experimentální nasazení u zákazníků. Čekám, že první zkušenosti nejspíš odhalí skutečné limitace zaměření na jednu z pěti podporovaných databázových platform, ale scanner vypadá dostatečně robustně, aby poskytl zákazníkovi i vývojáři vhodnou zpětnou vazbu.

Je zde evidentní potenciál budoucí rozšiřitelnosti, který tkví zejména v podpoře ostatních typů DB platform a pokrytí jednotlivých Matillion komponent.

Práce se tématu využitelnosti mohla v závěru trochu více sama věnovat vzhledem k tomu, že je v tomto ohledu skutečně nadprůměrná.

## Celkové hodnocení

88 /100 (B)

Jako celek se mi práce velmi líbila. Opravdu dobře se četl text i zdrojový kód. Zdrojový kód je velmi kvalitní, respektuje zvyklosti a je dostatečně robustní a rozšiřitelný. Výhrady jsem měl zejména u textové části, kde mě neoslily vydefinované funkční a nefunkční požadavky, ani popis algoritmu propagace a expanze sloupců. Po zbývajících stránkách byl ale text kvalitní s malým množstvím nedostatků. Využitelnost prototypu je velmi slušná a čeká se jeho brzké experimentální nasazení přímo u zákazníků.

## Otázky k obhajobě

1. Popište předpokládané rozdíly jednotlivých Matillion typů platforem. Na jaký typ Matillion komponent bude mít volba platformy největší vliv a proč?
2. Diagram závislostí modulů odhaluje závislost extraktoru na modelu, což nemusí být vhodné. Jak konkrétně by se dané závislosti dalo zbavit?

## **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.