

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Natálie Potočková

Název práce Analýza datových toků pro platformu Databricks

Rok odevzdání 2023

Studijní program Informatika **Studijní obor** Softwarové a datové inženýrství

Autor posudku Petr Škoda **Role** oponent

Pracoviště Katedra softwarového inženýrství

Text posudku:

Cílem práce je, zjednodušeně řešeno, implementovat podporu Databricks do produktu Manta Flow. Manta Flow je komerčním nástrojem pro automatizovanou extrakci data lineage. V důsledku zadání je tedy práce zasazena do prostředí firemního produktu.

Textová část práce je poměrně rozsáhlá. Její výrazná část se věnuje popisu integrace, technologií a jejich API. Další výraznou částí je kapitola design. Takto kapitola popisuje možná řešení, jejich vlastnosti a také zvolené řešení. Celkově textová část podrobně, a vcelku čitelně, uvádí čtenáře do problému a vysvětluje učiněná rozhodnutí. Některá popsaná řešení byla vytvořena mimo řešení samotné diplomové práce, případně třetí stranou. Tyto situace jsou v textu popsány, nicméně vzhledem k rozsahu práce je poměrně náročné je sledovat.

Kvalita textové části je na očekávané úrovni. V textu je možné najít poněkud neobvykle dělený text, například na stránkách 78 či 92. Případně nepříliš nápomocné zvýraznění klíčových slov na straně 94 v první sekci kódu.

Implementační část práce není samostatně funkční, neb je součástí produktu Manta Flow. Neb zdrojový kód pravděpodobně musí odpovídat firemním požadavkům, nepovažuji za přínosné ho detailněji komentovat. Celkově ho hodnotím jako dostatečně kvalitní.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 25.8.2023

Podpis: