

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Michal Jurčo
Název práce Data Lineage Analysis Service for Embedded Code
Rok odevzdání 2023
Studijní program Informatika **Studijní obor** Softwarové a datové inženýrství

Autor posudku Pavel Parížek **Role** Vedoucí
Pracoviště Katedra distribuovaných a spolehlivých systémů

Text posudku:

Cílem této práce bylo rozšíření platformy Manta Flow pro zjišťování datových toků o lepší podporu nástrojů datové analýzy, které používají "embedded" skripty napsané v univerzálních programovacích jazycích (Python, C#, Java). Mezi takové nástroje patří například AWS Glue, Snowflake a Databricks. Autor měl za úkol navrhnout obecnou službu pro analýzu takzvaného "embedded" kódu a integraci grafů datových toků, které jsou spočítány pro "embedded" skripty s těmi pro samotné nástroje datové analýzy. Součástí zadání bylo také vytvořit implementaci této obecné služby a její konkrétní instance pro nástroj AWS Glue se skriptami napsanými v jazyce Python.

Autor provedl důkladnou analýzu požadavků a diskuzi možných řešení, které jsou popsány v kapitolách 3-5 textu. Výstupem práce je také odladěná implementace, která již byla integrována do produktu Manta Flow. Největší výzvou při řešení této práce bylo navrhnout architekturu služby tak, aby podporovala všechnu požadovanou funkčnost ze strany existujících skenerů datových toků pro univerzální jazyky (například Python) a také ze strany skenerů pro nástroje datové analýzy. Proto návrh architektury probíhal v několika iteracích.

Kvalita zpracování textu je docela slušná, především vzhledem ke procesu jeho vzniku poměrně narychlo před termínem odevzdání. Nemám ale zásadní připomínky.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Pokud práci navrhujete na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).

Datum 21.8.2023

Podpis