

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Josef Kumštýř
Název práce Precise and Efficient Incremental Update of Data Lineage Graph
Rok odevzdání 2022
Studijní program Informatika **Studijní obor** Softwarové systémy

Autor posudku RNDr. Filip Zavoral, PhD. **Role** Oponent
Pracoviště KSI

Text posudku:

Hlavním cílem diplomové práce bylo výrazné zefektivnění stávajícího systému pro data lineage. Ten při aktualizaci vždy analyzoval všechna data, což pro velké objemy trvalo i přes týden strojového času. Základní myšlenkou zefektivnění je identifikace jak změněných souborů tak jejich okolí, tj. množiny dalších souborů, jejichž analýza může být změnami ovlivněna.

Stěžejní částí práce je kapitola 4, ve které autor popisuje navržené řešení ve formě algoritmů, jejich diskuse a vylepšení. Navržené metody a algoritmy autor pilotně implementoval v scanneru Oracle procedur. Vyhodnocení výkonnosti ukazuje, že zvolené metody jsou vysoce efektivní, analýza v případě změny pouze části souborů je mnohonásobně rychlejší (v závislosti na poměru změněných souborů) než kompletní analýza všech dat. Bylo by zajímavé analyzovat i vliv stupně paralelizace na celkový čas analýzy, to však text práce neobsahuje. Netriviálním aspektem implementace byla nutnost integrace navržených algoritmů do již rozsáhlého software a zajištění bezchybné funkčnosti s ostatními komponentami systému. To se autorovi podařilo, ukázal, že jeho řešení je plně použitelné a že je možné navrženými metodami efektivně zpracovávat výrazně rozsáhlejší data.

Text práce i prototypová implementace jsou na vysoké úrovni, k práci nemám vážnější připomínky a doporučuji ji k obhajobě.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

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

Datum 25. května 2022

Podpis