

# Analýza a návrh abstraktní vícevrstvé architektury pro práci s grafovou databází realizující metadatové úložiště pro data lineage

Bc. Jakub Moravec

Vedoucí: Ing. Michal Valenta, Ph.D.

Oponent: Ing. Jiří Šebek

České vysoké učení technické v Praze

Fakulta Elektrotechnická

Otevřená Informatika, Softwarové Inženýrství

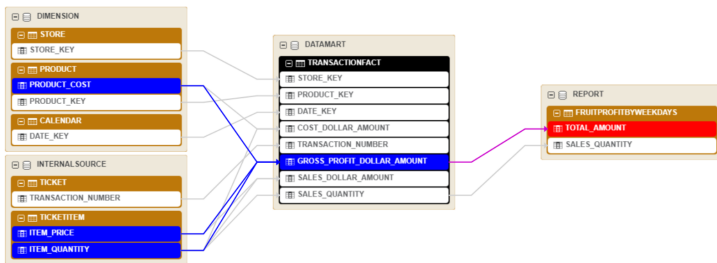
21.6.2018

# Obsah

- 1 Uvedení kontextu
- 2 Definice problému
- 3 Cíle práce
- 4 Dosažené výsledky
- 5 Základní charakteristiky navržené architektury
- 6 Přínosy práce

# Uvedení kontextu - Manta Flow

- Nástroj pro *data lineage* - analýza datových toků
- Parsuje zdrojové kódy informačních systémů (databáze, ETL nástroje) a sestavuje z nich datové toky
- Ze zdrojových kódů je sestaven syntakční strom, ten je ukládán v grafové databázi a dále analyzován



# Definice problému

- Existuje velké množství grafových databází
- Jednotlivé databáze se výrazně liší ve výkonostních parametrech
- Oblast je dynamická, vznikají nové nástroje, končí podpora pro některé stávající
- Neexistují obecné standardy pro jejich dotazování
- To vede k silné závislosti aplikací na používané grafové databázi a dotazovacím jazyce
  - Dochází k prolínání perzistentní a byznys logiky aplikace
  - Aplikace je obtížně spravovatelná a modifikovatelná
  - Při zvolení nevhodné granularity dotazů je degradována efektivita grafové databáze

# Cíle práce

- Seznámení se s grafovými databázemi a jejich API
- Analýza způsobu využívání grafové databáze v aplikaci *Manta Flow*
- Identifikace omezení stávající architektury aplikace vzhledem k práci s grafovou databází
- Rešerše existujících nástrojů pro abstrakci grafových databází
- Návrh vícevrstvé architektury abstrahující práci s grafovou databází
- Vytvoření prototypové implementace navržené architektury

# Dosažené výsledky

- 1 Rešerše grafových databází a možností jejich dotazování
- 2 Rešerše softwarových architektur vhodných pro návrh architektury
- 3 Analýza jednotlivých komponent aplikace *Manta Flow* a způsobu práce aplikace s grafovou databází
- 4 Identifikace omezení aplikace plynoucích ze stávající architektury aplikace
- 5 Specifikace konkrétních požadavků na navrhovanou architekturu na základě analýzy
- 6 Analýza existujících nástrojů pro abstrakci grafových databází

# Dosažené výsledky

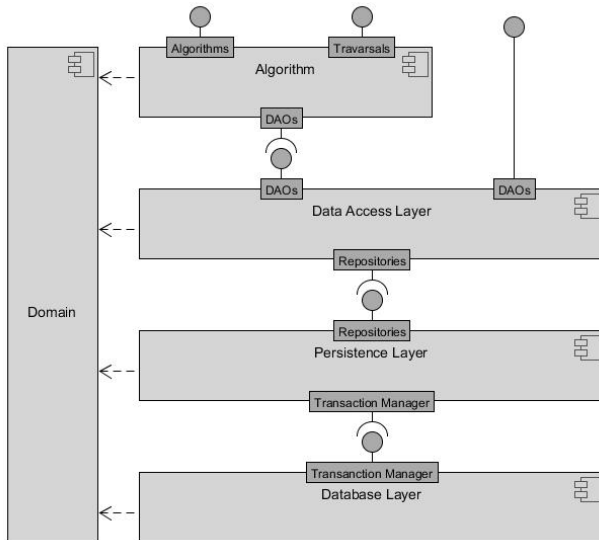
- 7 Návrh vícevrstvé architektury pro práci s grafovou databází a příslušných API
- 8 Prototypová implementace navržené architektury
- 9 Otestování jednotlivých vrstev prototypové implementace pomocí jednotkových testů
- 10 Implementace vybraných algoritmů tvořících byznys logiku aplikace pomocí prototypové implementace architektury
- 11 Otestování implementovaných algoritmů pomocí jednotkových a integračních testů
- 12 Návrh úpravy architektury dalších částí aplikace na základě navržené architektury pro práci s grafovou databází

# Základní charakteristiky navržené architektury

- Vícevrstvá architektura - funkcionality rozdělena do několika vrstev
- Definován specifický doménový model
- API perzistentní vrstvy vycházející z *Repository patternu*
- Oddělené metody využívající externí indexy
- Deklarativní transakční model



# Základní charakteristiky navržené architektury



# Přínosy práce

- Je navržena vícevrstvá architektura sloužící pro přístup ke grafové databázi
- Na základě prototypové implementace je ověřena vhodnost návrhu architektury
- Architektura odděluje perzistentní a byznys logiku aplikace a výrazně tak zlepšuje její modifikovatlnost
- Jsou navrženy dílčí kroky umožňující budoucí škálování aplikace

Děkuji za pozornost

Jakub Moravec  
jkb.moravec@gmail.com