

# Lexikální analyzátor dialektu dotazovacího jazyka databáze MySQL pro zachytávání změn v databázi

Bc. Roman Kuchár

Vedúci: Ing. Jiří Pechanec

Oponent: Ing. Tomáš Černý, MSc., Ph.D.

České vysoké učení technické v Praze

Fakulta elektrotechnická

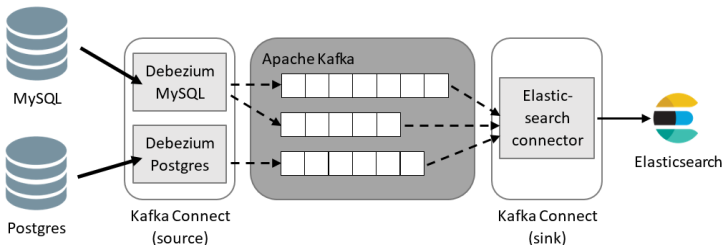
Otevřená informatika, Softwarové inženýrství

21.6.2018



# Debezium

- Open-source projekt
- Zachytávanie zmenených dát (CDC)
  - Stream událostí o zmenách v databázy
- Sledovanie transakčných logov
  - Analýza štruktúry tabuliek (DDL dotazy)



# Motivácia

- Parsovanie DDL dotazov
  - Špecifické typy DDL dotazov pre každé DBMS
  - Neexistuje použiteľná knižnica
- Aktuálne riešenie DDL parseru
  - Ručne implementované
  - Nekompletná implementácia
  - Náročné na správu

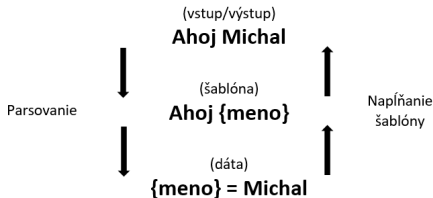
# Motivácia

- Parsovanie DDL dotazov
  - Špecifické typy DDL dotazov pre každé DBMS
  - Neexistuje použiteľná knižnica
- Aktuálne riešenie DDL parseru
  - Ručne implementované
  - **Nekompletná implementácia**
  - **Náročné na správu**

- ① Analýza možností nahradenia existujúceho riešenia
- ② Návrhnúť nové riešenie
  - Strojovo generovaný parser
  - Jenoduchý na spávu
- ③ Začlenenie nového parseru do projektu Debezium
  - Budúca podpora iných DBMS
- ④ Otestovanie parseru pomocou testovacej sady projektu Debezium

# Syntaktická analýza

- Teoreticky vyriešený problém
  - *Top-Down* parsovanie - LL gramatiky
  - *Bottom-Up* parsovanie - LR gramatiky
- V praxi sa stále vylepšuje
  - 2008 → **Bison** - IELR gramatika
  - 2014 → **ANTLR v4** - ALL(\*) gramatika

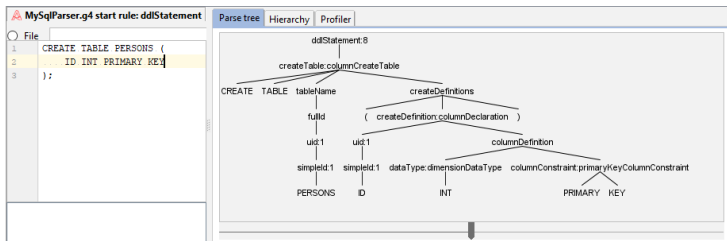


# ANTLR v4

- Veľká komunita užívateľov
- Mnoho podporných nástrojov
- Odstraňuje priamu ľavú rekurziu
- Dostupný repozitár vytvorených gramatík

# ANTLR v4

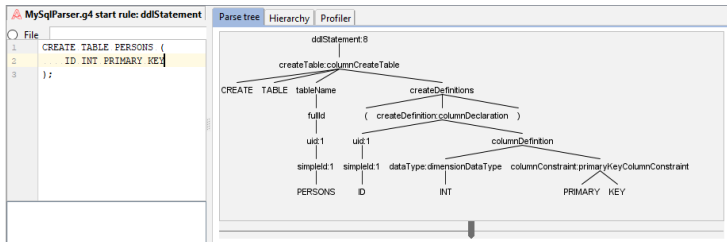
- Veľká komunita užívateľov
- Mnoho podporných nástrojov
- Odstraňuje priamu ľavú rekurziu
- Dostupný repozitár vytvorených gramatík





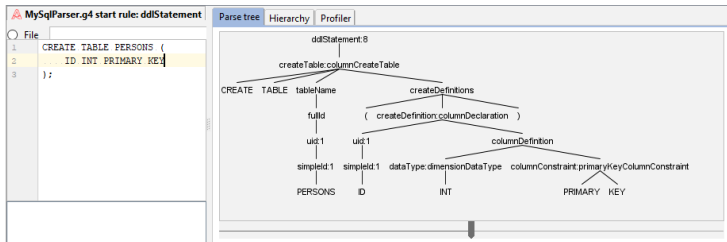
# ANTLR v4

- Veľká komunita užívateľov
- Mnoho podporných nástrojov
- Odstraňuje priamu ľavú rekurziu
- Dostupný repozitár vytvorených gramatík



# ANTLR v4

- Veľká komunita užívateľov
- Mnoho podporných nástrojov
- Odstraňuje priamu ľavú rekurziu
- Dostupný repozitár vytvorených gramatík



# Dosiahnuté výsledky

- Objavenie neregistrovaných chýb existujúceho parseru
  - Nesprávne chovanie parseru
  - Podpora syntakticky nevalidných dotazov
- Nový design pre parsovanie DDL
- Doplnenie chýbajúcej implementácie
- Oprava nevalidných testov
- Začlenenie parseru do projektu Debezium

# Dosiahnuté výsledky

- Objavenie neregistrovaných chýb existujúceho parseru
- Nový design pre parsovanie DDL
  - Generalizácia existujúcej implementácie
  - Generalizácia novej implementácie
  - Samostatný modul pre generovanie DDL parserov
- Doplnenie chýbajúcej implementácie
- Oprava nevalidných testov
- Začlenenie parseru do projektu Debezium

# Dosiahnuté výsledky

- Objavenie neregistrovaných chýb existujúceho parseru
- Nový design pre parsovanie DDL
- Doplnenie chýbajúcej implementácie
  - Parsovanie pohľadových dotazov (*VIEW*)
  - *TRUNCATE TABLE*
- Oprava nevalidných testov
- Začlenenie parseru do projektu Debezium

# Dosiahnuté výsledky

- Objavenie neregistrovaných chýb existujúceho parseru
- Nový design pre parsovanie DDL
- Doplnenie chýbajúcej implementácie
- Oprava nevalidných testov
  - Syntakticky nesprávne dotazy
  - Oprava použitej gramatiky
  - Implementácia nových testov
- Začlenenie parseru do projektu Debezium

# Dosiahnuté výsledky

- Objavenie neregistrovaných chýb existujúceho parseru
- Nový design pre parsovanie DDL
- Doplnenie chýbajúcej implementácie
- Oprava nevalidných testov
- Začlenenie parseru do projektu Debezium
  - Možnosť voľby parsovacej stratégie
  - Akceptovanie *coding rules* projektu

Ďakujem za pozornosť

Bc. Roman Kuchár  
kucharrom@gmail.com