

## **Databázové systémy**

### **Analytický report**

Tomáš Krupa

Cvičiaci: Ing. Eduard Kuric, PhD.

Študijný odbor: Informatika-4

Ročník: 4. Bc

Akademický rok: 2016/2017

## Zadanie

Stručný opis spôsobu, akým boli vygenerované dáta

Počet riadkov

Pre každý numerický atribút sumarizáciu min, max, avg, dolný kvartil, medián, horný kvartil (pre istotu: aj foreign key je numerický atribút, primary key netreba riešiť, ak je riešený ako sekvencia)

Pre každý nominálny atribút početnosti jednotlivých hodnôt

Minimálne pre tú tabuľku s rádovo miliónmi záznamov aj s uvedením postupu (SQL), akým sa študent k daným číslam dopracoval

## Opis scenárov

Doménou mojej aplikácie je realitný portál. Používateľ aplikácie je záujemca o kúpu alebo predaj nehnuteľnosti alebo predavač. Vie vyhľadať inzerát podľa vlastných kritérií a získať tak kontakt na predajcu nehnuteľnosti. Keď sa mu inzerát páči, má možnosť ho "likovať". Takisto ho môže kedykoľvek odlikovať. Vo svojom profile vidí svoje osobné údaje. Takisto vidí svoje inzeráty, ktoré môže odstrániť s inzercie a takisto svoje obľúbené inzeráty, ktoré môže odoberať z tohto zoznamu. Používateľ prihlásený ako predavač má možnosť vytvoriť a pridať inzerát do inzercie takisto ako ho aj vymazať. Aplikácia poskytuje full-textové vyhľadávanie v inzercii, ako aj vyhľadávanie podľa ponúknutých kritérií, ako napr. typ nehnuteľnosti, jej výbava, cenové rozpätie alebo vzdialenosť od zvoleného mesta. Po zobrazení výsledkov vyhľadávania môže používateľ zobraziť detaily konkrétneho inzerátu. Používateľ môže takisto napísať správu druhému používateľovi.

Diagram fyzického modelu dát sa nachádza v súbore `~/data/diagram.png`.

## Generovanie dát

Dáta boli generované skriptom napísaným v jazyku Python 3.5. Prístup do databázy PostgreSQL bol zabezpečený databázovým adaptérom Psycopg2. Celý skript je sa nachádza v súbore `~/data/db_generator.py`. Obsah tabuľky `cities` z mestami a ich koordinátami bol vložený ručne príkazom, ktorý sa nachádza v súbore `~/data/cities.sql`. Pri práci na generovaní skriptu bolo za potreby čistiť obsahy tabuliek. Na to poslúžil PL/pgSQL skript, ktorý sa nachádza v súbore `~/data/truncator.sql`. Zvyšné DDL skripty sa nachádzajú v súbore `~/schema/ddl_skripts.sql`.

## Počty riadkov tabuliek

users  
    buyers - 500 000  
    sellers - 10 000  
ads - 300 000  
likes - 10 000 000  
cities - 33  
messages - 2 550 000  
categories - 1 000 438  
features - 1 894 231  
photos - 3 743 923  
searches  
    seller\_searches - 40 753 112  
    buyer\_searches - 10 246 888

## Numerické atribúty

### **Cena** (*ads.price*)

min: 0  
max: 9 999 000  
avg: 657 677.26  
dolný kvartil: 57000  
medián: 137 000  
horný kvartil: 274000

Postup sa nachádza v súbore *~/data/aggregate.sql*.

## Početnosti nominálnych atribútov

### **Mená** (*users.name*)

594 mien

"Abdul";842  
"Abora";842  
"Adar";862  
"Adimír";796  
"Aglar";791  
"Ahn";839  
"Akil";821  
"Ako";820  
"Akol";819

"Alar";834

"Amas";817

.

.

.

"Vlano";789

"Vlartin";808

"Vlasa";857

"Vlater";855

"Vlath";828

"Vlatres";802

"Vlavan";815

"Vlavana";804

"Vlazuz";850