



DATABÁZE PRO BIG DATA

CVIČENÍ VI.

Agregace a MongoDB

Lukáš Matějů

8.4.2024 | DPB



SAMOSTATNÁ ÚLOHA

- zaměřená na agregace v Mongo DB
 - agregační roura
- využívá stejnou databázi jako předchozí cvičení
 - restaurants

```
{
  "address": {
    "building": "1007",
    "coord": [ -73.856077, 40.848447 ],
    "street": "Morris Park Ave",
    "zipcode": "10462"
  },
  "borough": "Bronx",
  "cuisine": "Bakery",
  "grades": [
    { "date": { "$date": 1393804800000 }, "grade": "A", "score": 2 },
    { "date": { "$date": 1378857600000 }, "grade": "A", "score": 6 },
    { "date": { "$date": 1358985600000 }, "grade": "A", "score": 10 },
    { "date": { "$date": 1322006400000 }, "grade": "A", "score": 9 },
    { "date": { "$date": 1299715200000 }, "grade": "B", "score": 14 }
  ],
  "name": "Morris Park Bake Shop",
  "restaurant_id": "30075445"
}
```

- řešení možné odevzdat pomocí MongoDB shellu nebo PyMongo
 - na elearningu dostupný pomocný skript pro Python (s výpisy)

SAMOSTATNÁ ÚLOHA

1. zjistěte počet restaurací pro každé PSČ (zipcode)
 - a) seřadte podle zipcode vzestupně
 - b) seřadte podle počtu restaurací sestupně
 - výpis limitujte na 10 záznamů
 - úlohu vyřešte pomocí agregační roury

 2. restaurace obsahují pole grades, kde jsou jednotlivá hodnocení; vypište průměrné skóre pro každou hodnotu grades
 - v agregaci vynechte grade pro hodnotu „Not Yet Graded“ (může se vyskytovat místo A, B atd.)
 - výpis limitujte na 10 záznamů
 - úlohu vyřešte pomocí agregační roury
- A 12.5
B 10.4
C 28.1

BONUSOVÁ ÚLOHA

1. zjistěte 5 restaurací s nejlepším průměrným skóre pro známku A
 - restaurace s méně než třemi hodnoceními nebudou uvažovány
 - ve výsledku vypisujte kromě průměrného skóre i počet hodnocení
 - úlohu vyřešte pomocí agregační roury
2. nalezněte nejlepší restauraci pro každý typ kuchyně
 - rozšiřte předchozí úlohu
 - úlohu vyřešte pomocí agregační roury
3. nalezněte všechny restaurace s víceslovným názvem
 - restaurace musí mít alespoň 2 hodnocení vyšší než 10
 - úlohu vyřešte pomocí agregační roury