

# Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования

## «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»

(национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)

#### ФАКУЛЬТЕТ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА КОМПЬЮТЕРНЫЕ СИСТЕМЫ И СЕТИ (ИУ6)

НАПРАВЛЕНИЕ ПОДГОТОВКИ **09.04.01 Информатика и вычислительная техника** МАГИСТЕРСКАЯ ПРОГРАММА **09.04.01/07 Интеллектуальные системы анализа, обработки и интерпретации больших** данных

## ОТЧЕТ

| 01121                              |               |             |                 |                              |  |  |  |  |  |  |  |
|------------------------------------|---------------|-------------|-----------------|------------------------------|--|--|--|--|--|--|--|
| по лабораторной работе № <u>10</u> |               |             |                 |                              |  |  |  |  |  |  |  |
| Название: Span                     | <u>·k</u>     |             |                 |                              |  |  |  |  |  |  |  |
| Дисциплина: <u>Я</u><br>данными    | зыки программ | ирования дл | я работы с бо   | <u>ЭЛЬШИМИ</u>               |  |  |  |  |  |  |  |
| Студент                            | ИУ6-22М       |             |                 | Д. Р. Григорян               |  |  |  |  |  |  |  |
| Преподаватель                      |               |             | (Подпись, дата) | П.В. Степанов (И.О. Фамилия) |  |  |  |  |  |  |  |

## Цель работы:

Ознакомиться с языком программирования Java, создать проект для написания 10 запросов из выбранного датасета, используя Spark.

## Выполнение:

### Задача 1:

Использовать Spark, Scala/Java. Выбрать любой датасет на kaggle.com и сделать 10 выборок данных по выбранной предметной области.

## Листинг 1 программы main:

## Результаты запросов показаны на рисунках 1-5:

```
Z3/04/14 17:29:25 INFO FileSourceStrategy: Pruning directories with:

23/04/14 17:29:25 INFO FileSourceStrategy: Pruning directories with:

23/04/14 17:29:25 INFO FileSourceStrategy: Post-Scan Filters:

23/04/14 17:29:25 INFO FileSourceStrategy: Output Data Schema: struct<INCIDENT_KEY: string, OCCUR_DATE: string, OCCUR_23/04/14 17:29:25 INFO FileSourceStrategy: Output Data Schema: struct<INCIDENT_KEY: string, OCCUR_DATE: string, OCCUR_23/04/14 17:29:25 INFO FileSourceScanExec: Pushed Filters:

23/04/14 17:29:25 INFO CodeGenerator: Code generated in 57.9883 ms

23/04/14 17:29:25 INFO MemoryStore: Block broadcast_17 stored as values in memory (estimated size 221.4 KB, free 2.5 C 23/04/14 17:29:25 INFO MemoryStore: Block broadcast_17_piece0 stored as bytes in memory (estimated size 20.6 KB, free
```

Рисунок 1 – Результат выполнения запроса

| +          | +-                                  |                  |                  | +-                         |                  | +                  | ₽ -                               |
|------------|-------------------------------------|------------------|------------------|----------------------------|------------------|--------------------|-----------------------------------|
| IN         | CIDENT_KEY OCCUR_DATE O             | CCUR_TIME        | BORO LO          | C_OF_OCCUR_DESC P          | RECINCT JURI     | SDICTION_CODE LOC_ |                                   |
|            |                                     |                  |                  |                            |                  |                    |                                   |
|            | 243566884 04/12/2022                | 22:08:00         | BRONX            | OUTSIDE                    | 49               | 0                  | STREET                            |
|            | 256484816 12/17/2022                | 04:08:00         | BRONX            | OUTSIDE                    | 52               | 0                  | STREET                            |
|            | 250216145 08/27/2022                | 00:21:00         | BRONX            | OUTSIDE                    | 44               | 0                  | VEHICLE                           |
|            | 239207164 01/15/2022                | 19:50:00         | QUEENS           | OUTSIDE                    | 113              | 2                  | HOUSING MULTI DW                  |
|            | 248013313 07/14/2022                | 01:19:00         | BROOKLYN         | INSIDE                     | 77               | 0                  | DWELLING MULTI DW                 |
|            | 247542571 07/04/2022                | 22:20:00         | BRONX            | OUTSIDE                    | 48               | 0                  | STREET                            |
|            | 251813843 09/29/2022                | 21:32:00         | BROOKLYN         | OUTSIDE                    | 81               | 0                  | STREET                            |
|            | 254000625 11/11/2022                | 15:24:00         | BROOKLYN         | OUTSIDE                    | 75               | 0                  | OTHER MULTI DW                    |
|            | 252647537 10/16/2022                | 20:05:00         | BROOKLYN         | OUTSIDE                    | 75               | 0                  | STREET                            |
|            | 249207672 08/08/2022                | 04:20:00         | IANHATTAN        | OUTSIDE                    | 20               | 0                  | STREET                            |
|            | 246104595 06/03/2022                | 15:41:00         | BRONX            | OUTSIDE                    | 41               | 0                  | STREET                            |
|            | 254062679 11/12/2022                | 22:18:00         | IANHATTAN        | OUTSIDE                    | 10               | 0                  | STREET                            |
|            | 247501139 07/02/2022                | 12:23:00         | ANHATTAN         | OUTSIDE                    | 7                | 0                  | STREET                            |
|            | 245766300 05/28/2022                | 00:41:00         | BROOKLYN         | OUTSIDE                    | 75               | 0                  | DWELLING MULTI DW                 |
|            | 247099746 06/23/2022                | 10:08:00         | BRONX            | OUTSIDE                    | 48               | 0                  | STREET                            |
|            | 245824483 05/29/2022                | 10:41:00         | BROOKLYN         | OUTSIDE                    | 73               | 0                  | HOUSING MULTI DW                  |
|            | 245491991 05/22/2022                | 14:48:00         | BRONX            | OUTSIDE                    | 46               | 0                  | STREET                            |
| <b>₽</b> A |                                     |                  |                  |                            |                  |                    |                                   |
| and Ma     | won library shared indoves // Alway | us download // F | )ourpload onco / | / Don't show again // Conf | iauro /2 minutos | 200)260            | DQ:1 IE LITE Q 4 chacos 12 main 0 |

Рисунок 2 – Результат выполнения запроса

Рисунок 3 – Результат выполнения запроса

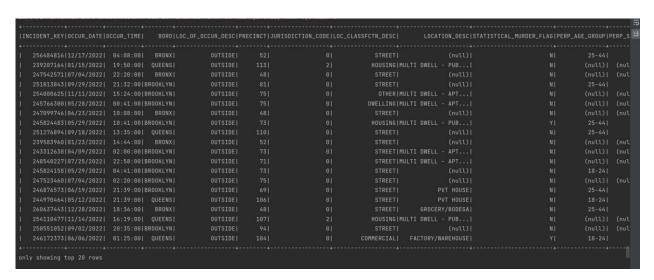


Рисунок 4 – Результат выполнения запроса

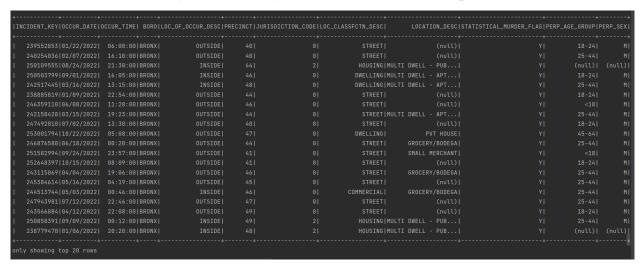


Рисунок 5 – Результат выполнения запроса

**Вывод:** в ходе выполнения лабораторной работы была написана программа, которая использует Spark Session для подключения к определенному датасету и с помощью Spark Sql выполняет из него 10 выборок.