



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)»
(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

О Т Ч Е Т

по лабораторной работе № 1 0

Название: Spark. Вариант 9

Дисциплина: Языки программирования для работы с большими
данными

Студент ИУ6-22М
(Группа)

К.Ю. Каташинский
(Подпись, дата) (И.О. Фамилия)

Преподаватель

П.В. Степанов
(Подпись, дата) (И.О. Фамилия)

Москва, 2024

Цель работы

Ознакомиться с работой Spark в языке программирования Java.

Ход работы

Задание:

- 1) Выбрать любой датасет на kaggle.com.
- 2) Сделать 10 выборок данных по выбранной предметной области.

Для лабораторной работы был выбран датасет с демографическими данными по России за 1990–2017 года. Код программы представлен в листинге 1, результат – на рисунке 1.

Листинг 1 – Код программы задания 1

```
package Lab10;

import org.apache.spark.sql.Session;
public class Variant {
    public static void main(String[] args) {
        var spark =
SparkSession.builder().appName("Lab10").getOrCreate();
        var statistics =
spark.read().format("csv").option("header",
"true").load("src/Lab10/dataset.csv");
        statistics.createOrReplaceTempView("statistics");
        spark.sql("SELECT * FROM statistics");
        spark.sql("SELECT * FROM statistics WHERE region =
'Amur Oblast' AND year > 2000 AND year < 2005").show();
        spark.sql("SELECT * FROM statistics WHERE region =
'Altai Krai' AND year > 2005 AND year < 2015").show();
        spark.sql("SELECT * FROM statistics WHERE region =
'Kaliningrad Oblast'").show();
        spark.sql("SELECT AVG(birth_rate) FROM statistics
WHERE year < 2000").show();
        spark.sql("SELECT AVG(death_rate) FROM statistics
WHERE year < 2000").show();
```

```

        spark.sql("SELECT AVG(urbanization) FROM statistics
WHERE year < 2000").show();

        spark.sql("SELECT AVG(birth_rate) FROM statistics
WHERE year >= 2000").show();

        spark.sql("SELECT AVG(death_rate) FROM statistics
WHERE year >= 2000").show();

        spark.sql("SELECT AVG(urbanization) FROM statistics
WHERE year >= 2000").show();

        spark.stop();
    }
}

```

```

+-----+
|  avg(death_rate) |
+-----+
|14.364929859719474|
+-----+

+-----+
|avg(urbanization)|
+-----+
|           69.3149|
+-----+

Process finished with exit code 0

```

Рисунок 1 – Результат задания 1

Вывод

В ходе выполнения лабораторной работы было произведено знакомство с работой Spark в языке программирования Java.