



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 № 10

Название: Spark

Дисциплина: Языки программирования для работы с большими данными

Студент

ИУ6-22М

(Группа)

\_\_\_\_\_  
(Подпись, дата)

Т.И. Кадыров

(И.О. Фамилия)

Преподаватель

\_\_\_\_\_  
(Подпись, дата)

П.В. Степанов

(И.О. Фамилия)

Москва, 2024

**Цель работы:** освоить базовые принципы работы со Spark на языке Java.

**Вариант: 8.**

**Задание 1:** Сделать 10 выборки данных по выбранной предметной области .

Код решения приведен в листинге 1.

Листинг 1 — реализация решения

```
package org.example;

import org.apache.spark.sql.*;
import org.apache.spark.api.java.JavaSparkContext;

public class Main {
    public static void main(String[] args) {
        SparkSession spark = SparkSession.builder()
            .appName(" Main")
            .master(" local[*]")
            .getOrCreate();
        spark.sparkContext().setLogLevel("ERROR");

        JavaSparkContext jsc = new JavaSparkContext(
            spark.sparkContext());

        Dataset<Row> dataset = spark.read().format(" csv")
            .option(" header", " true")
            .load(System.getProperty(" user.dir")+ " /src/main/resources/who_suicide_statistics.csv");

        dataset.createOrReplaceTempView("suicide_tab");

        System.out.println(" \n\n\n1. all");
        spark.sql("select * from suicide_tab").show();

        System.out.println(" \n\n\n2. Total number of suicides per country for 2007 year");
        spark.sql("select country, sum(suicides_no) as total_suicides from suicide_tab where year = 2007 group by country").show();

        System.out.println(" \n\n\n3. Total number of suicides per year");
        spark.sql("select year, sum(suicides_no) as total_suicides from suicide_tab group by year").show();

        System.out.println(" \n\n\n4. Total number of suicides per gender");
        spark.sql("select sex, sum(suicides_no) as total_suicides from suicide_tab group by sex").show();

        System.out.println(" \n\n\n5. Average population for each country");
        spark.sql("select country, avg(population) as avg_population from suicide_tab group by country").show();

        System.out.println(" \n\n\n6. Total number of suicides per age group");
        spark.sql("select age, sum(suicides_no) as total_suicides from suicide_tab group by age").show();

        System.out.println(" \n\n\n7. Countries with the highest total number of suicides");
```

```

    spark.sql("select country, sum(suicides_no) as total_suicides from suicide_tab group by country order by total_suicides desc nulls last").show();

    System.out.println("\n\n8. Years with the highest total number of suicides");
    spark.sql("select year, sum(suicides_no) as total_suicides from suicide_tab group by year order by total_suicides desc nulls last").show();

    System.out.println("\n\n9. Suicides in Russia by year");
    spark.sql("select year, sum(suicides_no) as total_suicides from suicide_tab where country = 'Russian Federation' group by year order by year desc").show();

    System.out.println("\n\n10. Suicides in Russia by sex and age group");
    spark.sql("select sex, age, sum(suicides_no) as total_suicides from suicide_tab where country = 'Russian Federation' group by sex, age order by sex, age").show();
    // Stop the SparkSession and JavaSparkContext
    spark.stop();
    jsc.stop();
}
}

```

**Вывод:** в ходе лабораторной работы были освоены базовые принципы работы со Spark на языке Java.