|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **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 **09.04.01/12 Интеллектуальный анализ больших**

**данных в системах поддержки принятия решений**

**Отчет**

**по лабораторной работе 10**

**Название:** Spark

**Дисциплина:** Языки программирования для работы с большими данными

Студент гр. ИУ6-23М **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_** И. Г. Калинин**\_\_**

(Подпись, дата) (И.О. Фамилия)

Преподаватель  **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_** П. В. Степанов **\_\_**

(Подпись, дата) (И.О. Фамилия)

Москва, 2023

**Цель лабораторной работы**

Познакомиться с базовыми понятиями языка Java, изучить работу Java+spark. Решить набор задач согласно индивидуальному варианту.

**Фрагмент программного кода**

В качестве демонстрации результата приведен один из листингов проекта. Код на языке Java представлен в листинге 1.

Листинг 1 – Код на языке Java

|  |
| --- |
| package org.example;  import org.apache.spark.SparkConf;  import org.apache.spark.api.java.JavaSparkContext;  import org.apache.spark.sql.Dataset;  import org.apache.spark.sql.Row;  import org.apache.spark.sql.SparkSession;  public class SampleData {  public static void main(String[] args) {  System.setProperty("hadoop.home.dir", "D:/BMSTU/NEW/java\_projs/lab10/spark-3.4.0-bin-hadoop3");  System.setProperty("java.util.logging.config.file", "D:/BMSTU/NEW/java\_projs/lab10/");  // Создаем объект SparkConf и SparkSession  SparkConf conf = new SparkConf()  .setAppName("YourAppName")  .setMaster("local");  JavaSparkContext sparkContext = new JavaSparkContext(conf);  SparkSession sparkSession = SparkSession.builder().sparkContext(sparkContext.sc()).getOrCreate();  // Загружаем данные из CSV файла с указанием кодировки  String path = "Rus\_schools\_final.csv";  Dataset<Row> data = sparkSession.read().option("header", "true").option("encoding", "CP1251").csv(path);  // Регистрируем временную таблицу для выполнения SQL-запросов  data.createOrReplaceTempView("schools");  // 1) Вывести все строки про "лицеи"  Dataset<Row> query1 = sparkSession.sql("SELECT \* FROM schools WHERE name LIKE '%лицеи%'");  query1.show(false);  // 2) Вывести все школы Татарстана  Dataset<Row> query2 = sparkSession.sql("SELECT \* FROM schools WHERE addr LIKE '%Татарстан%'");  query2.show(false);  // 3) Вывести все школы Казани  Dataset<Row> query3 = sparkSession.sql("SELECT \* FROM schools WHERE addr LIKE '%Казань%'");  query3.show(false);  // 4) Вывести все Муниципальные образовательные учреждения  Dataset<Row> query4 = sparkSession.sql("SELECT \* FROM schools WHERE struct LIKE '%Муниципальное образовательное учреждение%'");  query4.show(false);  // 5) Вывести все вечерние школы  Dataset<Row> query5 = sparkSession.sql("SELECT \* FROM schools WHERE name LIKE '%вечерние%'");  query5.show(false);  // 6) Вывести просто все записи  Dataset<Row> query6 = sparkSession.sql("SELECT \* FROM schools");  query6.show(false);  // 7) Все школы Новосибирска  Dataset<Row> query7 = sparkSession.sql("SELECT \* FROM schools WHERE addr LIKE '%Новосибирск%'");  query7.show(false);  // 8) Количество школ Новосибирска  Dataset<Row> query8 = sparkSession.sql("SELECT COUNT(\*) as count FROM schools WHERE addr LIKE '%Новосибирск%'");  query8.show(false);  // 9) Все школы с улицы "школьная"  Dataset<Row> query9 = sparkSession.sql("SELECT \* FROM schools WHERE addr LIKE '%школьная%'");  query9.show(false);  // 10) Все школы коррекции  Dataset<Row> query10 = sparkSession.sql("SELECT \* FROM schools WHERE name LIKE '%коррекции%'");  query10.show(false);  sparkContext.close();  }  } |

**Вывод**

В рамках данной лабораторной работы были реализованы все задачи, выданные преподавателем, а также изучены и выполнены все пункты целей данной лабораторной работы.