|  |  |
| --- | --- |
| 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 **09.04.01/07 Интеллектуальные системы анализа, обработки и интерпретации больших данных**

**Отчет**

|  |  |
| --- | --- |
| **по лабораторной работе №** | 10 |

**Название:** Spark

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Студент | ИУ6-22М |  |  | Р.В. Баканов | |
|  | (Группа) |  | (Подпись, дата) | | (И.О. Фамилия) |
|  |  |  |  | |  |
| Преподаватель |  |  |  | | П.В. Степанов |
|  |  |  | (Подпись, дата) | | (И.О. Фамилия) |

Москва, 2024

**Цель работы:** освоить базовые принципы работы со Spark в Java.

**Задание:**

Выбрать любой датасет на kaggle.com и сделать 10 выборок данных по выбранной предметной области.

Код решения приведен в листинге 1.

Листинг 1 — Задание

import org.apache.spark.sql.\*;  
  
public class Main {  
 public static void main(String[] args) {  
 System.*out*.println("Инициализация spark");  
  
 SparkSession sparkSession = SparkSession.*builder*().appName("spark\_test.com").master("local[\*]").getOrCreate();  
 sparkSession.sparkContext().setLogLevel("ERROR");  
  
 String path = System.*getProperty*("user.dir") + "/pokemon\_data.csv";  
 Dataset<Row> dataset = sparkSession.read().format("csv").option("header", "true").load(path);  
 dataset.createOrReplaceTempView("pokemon");  
  
 System.*out*.println("\n\nРезультаты запросов:");  
  
 System.*out*.println("\n1. Пример записей");  
 sparkSession.sql("select \* from pokemon limit 10").show();  
  
 System.*out*.println("\n2. Процент покемонов со способностью летать");  
 sparkSession.sql("select round(sum(case when type1 = 'Flying' or type2 = 'Flying' then 1 end) / count(\*) \* 100, 1) as `Процент летающих покемонов` from pokemon").show();  
  
 System.*out*.println("\n3. Самые высокие покемоны");  
 sparkSession.sql("select name as `Покемон`, height as `Рост` from pokemon order by 2 desc limit 5").show();  
  
 System.*out*.println("\n4. Распределение по редкости");  
 sparkSession.sql("select special\_group as `Редкость`, count(\*) as `Кол-во` from pokemon group by special\_group order by 2").show();  
  
 System.*out*.println("\n5. Среднее кол-во опыта за победу для каждого класса");  
 sparkSession.sql("select type1 as `Класс`, ROUND(AVG(base\_exp), 1) as `Среднее кол-во опыта` from pokemon group by type1 order by 2 desc").show();  
  
 System.*out*.println("\n6. Самые сильные человекоподобные покемоны");  
 sparkSession.sql("select name as `Покемон`, total as `Общая сила` from pokemon where egg\_group1 = 'Human-Like' order by 2 desc limit 5").show();  
  
 System.*out*.println("\n7. Покемоны, содержащие 'duck' в имени");  
 sparkSession.sql("select name as `Покемон` from pokemon where name LIKE '%duck%'").show();  
  
 System.*out*.println("\n8. Кол-во покемонов со скрытой способностью");  
 sparkSession.sql("select count(\*) from pokemon where hidden\_ability is not null").show();  
  
 System.out.println("\n9. Покемоны со скоростью ниже среднего");  
 sparkSession.sql("select name as `Покемон`, speed as `Скорость` from pokemon where speed < (select AVG(speed) from pokemon) order by 2").show();  
  
 System.out.println("\n10. Пикачу");  
 sparkSession.sql("select \* from pokemon where name = 'Pikachu'").show();  
  
 sparkSession.stop();  
 }  
}

**Вывод:** в результате выполнения лабораторной работы были разработаны несколько программ в соответствие с требованиями задания.