|  |  |
| --- | --- |
| 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 | ИУ6-23М |  |  | В.Д. Рожнов |
|  | (Группа) |  | (Подпись, дата) | (И.О. Фамилия) |
|  |  |  |  |  |
| Преподаватель |  |  |  | П.В. Степанов |
|  |  |  | (Подпись, дата) | (И.О. Фамилия) |

Москва, 2023

**Задание:**

1. Выбрать любой набор данных на kaggle.com;
2. Cделать 10 выборок данных по выбранной предметной области.

Для выполнения данной лабораторной работы был выбран набор данных, включающий в себя информацию по змеям: место обитание, наличие яда, окрас и т. д. На основе данного набора данных было сделано 10 выборок по различным параметрам.

Код выполнения данных выборок представлен в листинге 1.

Листинг 1 – Программный код модуля Main.java для задания 1

package org.example;

import org.apache.spark.sql.Dataset;

import org.apache.spark.sql.Row;

import org.apache.spark.sql.SparkSession;

public class Main {

public static void main(String[] args) {

SparkSession spark = SparkSession.builder()

.appName("Snake Data Analysis")

.master("local[\*]")

.getOrCreate();

Dataset<Row> dataset = spark.read().format("csv")

.option("header", "true")

.load("/path/to/snake\_dataset.csv");

dataset.createOrReplaceTempView("snakes");

spark.sql("SELECT \* FROM snakes").show();

spark.sql("SELECT \* FROM snakes WHERE binomial = 'Agkistrodon contortrix'").show();

spark.sql("SELECT \* FROM snakes WHERE country = 'United States of America'").show();

spark.sql("SELECT \* FROM snakes WHERE poisonous = '1'").show();

spark.sql("SELECT \* FROM snakes WHERE family LIKE 'V%'").show();

spark.sql("SELECT \* FROM snakes WHERE age <= 5").show();

spark.sql("SELECT \* FROM snakes WHERE name LIKE '%cobra%'").show();

spark.sql("SELECT \* FROM snakes WHERE binomial LIKE 'B%'").show();

spark.sql("SELECT \* FROM snakes WHERE poisonous = '0'").show();

spark.stop();

}

}

**Вывод:** в ходе выполнения данной лабораторной работы была изучена библиотека Spark на языке Java и применены на практике методы, позволяющие произвести выборку по различным параметрам из заранее подготовленного набора данных.