|  |  |
| --- | --- |
| 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**

**Вариант № 12**

**Название:** Spark

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

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

Москва, 2024

**Цель:** ознакомиться с работой Spark в языке программирования Java.

**Задание 1:** Сделать 10 выборок данных по выбранной предметной области

Был выбран Dataset с рейтингом лапши разных производителей в разных странах

Код класса Main:

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("Ramen Rating")

.master("local[\*]")

.getOrCreate();

spark.sparkContext().setLogLevel("ERROR");

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

.option("header", "true")

.load("ramen-ratings.csv");

dataset.createOrReplaceTempView("ramen\_rate");

System.out.println("1. Show top 10 highest rated ramens:");

spark.sql("SELECT Brand, Variety, Stars FROM ramen\_rate ORDER BY Stars DESC LIMIT 10").show();

System.out.println("2. Count of ramen's varieties by brand:");

spark.sql("SELECT Brand, COUNT(Variety) as VarietyCount FROM ramen\_rate GROUP BY Brand ORDER BY VarietyCount DESC").show();

System.out.println("3. All ramen from South Korea with a rating of 5:");

spark.sql("SELECT Brand, Variety, Style, Stars FROM ramen\_rate WHERE Country = 'South Korea' AND Stars = 5").show();

System.out.println("4. Average stars rate of ramen by country:");

spark.sql("SELECT Country, AVG(Stars) as AvgStars FROM ramen\_rate GROUP BY Country ORDER BY AvgStars DESC").show();

System.out.println("5. All Nissin ramen with rate above 4:");

spark.sql("select Variety, Style, Country, Stars from ramen\_rate where Brand = 'Nissin' and Stars >= 4 order by Stars asc").show();

System.out.println("6. List ramen from Japan that are not cups:");

spark.sql("SELECT Brand, Variety, Style FROM ramen\_rate WHERE Brand = 'Japan' AND Style <> 'Cup'").show();

System.out.println("7. Distribution of ramen styles in Japan:");

spark.sql("SELECT Style, COUNT(\*) as StyleCount FROM ramen\_rate WHERE Country = 'Japan' GROUP BY Style ORDER BY StyleCount DESC").show();

System.out.println("8. List distinct flavors of ramen:");

spark.sql("SELECT DISTINCT Variety FROM ramen\_rate").show();

System.out.println("9. Average rate of ramen by style:");

spark.sql("SELECT Style, AVG(Stars) as AvgStars FROM ramen\_rate GROUP BY Style ORDER BY AvgStars DESC").show();

System.out.println("10. Count of ramen brands per style:");

spark.sql("SELECT Style, COUNT(DISTINCT Brand) as BrandCount FROM ramen\_rate GROUP BY Style").show();

spark.stop();

}

}

Работа программы показана на рисунке 1.

Изображение выглядит как Мультимедийное программное обеспечение, текст, Графическое программное обеспечение, программное обеспечение

Автоматически созданное описание

Рисунок 1 – Работа программы

**Вывод:** был изучен Spark в языке программирования Java.