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

**Вариант № 13**

**Название:** Spark

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент | ИУ6-23М |  |  | С.П.Пантелеев |
|  | (Группа) |  | (Подпись, дата) | (И.О. Фамилия) |
|  |  |  |  |  |
| Преподаватель |  |  |  | П.В. Степанов |
|  |  |  | (Подпись, дата) | (И.О. Фамилия) |

Москва, 2023

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

**Задание**: сделать 10 выборок данных по выбранной предметной области

Был выбран Dataset с аниме сериалами и фильмами, с показателем рейтингов, количеством зрителей, жанра и количеством серий.

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("RandomSplitExample")  
 .master("local[\*]")  
 .getOrCreate();  
  
 Dataset<Row> dataset = spark.read().format("csv")  
 .option("header", "true")  
 .load("C:/University/Java\_labs/Lab10/anime.csv");  
 dataset.createOrReplaceTempView("anime\_base");  
 spark.sql("select \* from anime\_base").show();  
 spark.sql("select \* from anime\_base where anime\_base.type='TV'").show();  
 spark.sql("select \* from anime\_base where anime\_base.episodes > 24").show();  
 spark.sql("select \* from anime\_base where anime\_base.rating > 9.5").show();  
 spark.sql("select \* from anime\_base where anime\_base.rating < 5 and anime\_base.members > 50000").show();  
 spark.sql("select \* from anime\_base where anime\_base.rating > 9 and anime\_base.members < 10000").show();  
 spark.sql("select \* from anime\_base order by 1").show();  
 spark.sql("select \* from anime\_base where anime\_base.genre='Sci-Fi'").show();  
 spark.sql("select \* from anime\_base where anime\_base.genre='Sci-Fi' and anime\_base.rating > 9").show();  
 spark.sql("select \* from anime\_base where anime\_base.genre='Sci-Fi' and anime\_base.rating > 8.7 and anime\_base.members > 50000 and anime\_base.type='Movie'").show();  
  
 spark.stop();  
 }  
}

Работа программы показана на рисунке 1.

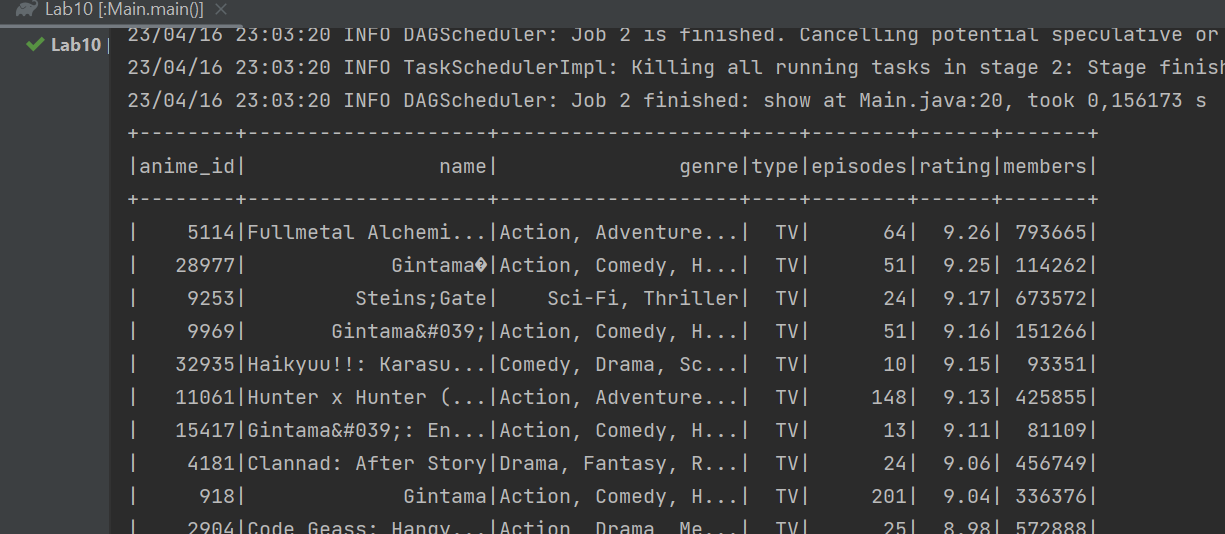


Рисунок 1 – Работа программы

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