|  |  |
| --- | --- |
| 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 № 3**

**Название:** Spark

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

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

Москва, 2023

**Цель:** изучить и освоить принципы работы со Spark в Java.

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

В качестве предметной области был выбран датасет “Video Games Sales”.

Этот набор данных содержит список видеоигр, объем продаж которых превысил 100 000 копий. Описание полей датасета представлено на рисунке 1. Часть датасета представлена на рисунке 2.



Рисунок 1 – Описание полей датасета

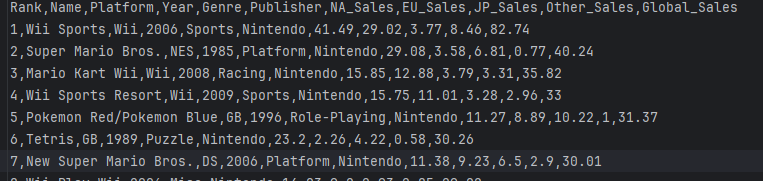


Рисунок 2 – Часть датасета

Код класса Main:

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("Video Game Sales")  
 .master("local[\*]")  
 .getOrCreate();  
  
 Dataset<Row> dataset = spark.read().format("csv")  
 .option("header", "true")  
 .load("vgsales.csv");  
 dataset.createOrReplaceTempView("game\_sales");  
  
 // 1 Выводим все  
 System.out.println("Весь датасет:");  
 spark.sql("select \* from game\_sales").show();  
  
 // 2 Выводим топ5 самых продаваемых игр за всё время  
 System.out.println("Топ5 самых продаваемых игр за всё время:");  
 spark.sql("SELECT \* FROM game\_sales ORDER BY Global\_Sales DESC LIMIT 5").show();  
  
 // 3 Выводим самые продаваемые игры в японии  
 System.out.println("5 самых продаваемых игр в японии");  
 spark.sql("SELECT \* FROM game\_sales ORDER BY JP\_Sales DESC LIMIT 5").show();  
  
 //4 5 самых продаваемых игр в японии жанра Action  
 System.out.println("3 самых продаваемых игр в японии жанра Action");  
 spark.sql("SELECT \* FROM game\_sales WHERE Genre = 'Action' ORDER BY JP\_Sales DESC LIMIT 3").show();  
  
 // 5 Вывести игры сони, сортируя по глобальным продажамм  
 System.out.println("Игры компании Sony Computer Entertainment");  
 spark.sql("SELECT \* FROM game\_sales WHERE Publisher = 'Sony Computer Entertainment' ORDER BY Global\_Sales DESC").show();  
  
 // 6 Самые продаваемые игры выпущенные в 2017  
 System.out.println("Самые популярные игры в 2017");  
 spark.sql("SELECT \* FROM game\_sales WHERE Year = '2017' ORDER BY Global\_Sales DESC LIMIT 5").show();  
  
 // 7 Самые продаваемые игры выпущенные на PS4  
 System.out.println("Самые продаваемые игры выпущенные на PS4");  
 spark.sql("SELECT \* FROM game\_sales WHERE Platform = 'PS4' ORDER BY Global\_Sales DESC LIMIT 5").show();  
  
 // 8 Самые продаваемые игры выпущенные на PS4  
 System.out.println("Самые непродаваемые игры выпущенные на PS4 в 2014 году");  
 spark.sql("SELECT \* FROM game\_sales WHERE Platform = 'PS4' and Year = '2014' ORDER BY Global\_Sales LIMIT 5").show();  
  
 // 9 Продажи игр серии Yakuza  
 System.out.println("Продажа игр серии Yakuza");  
 spark.sql("SELECT \* FROM game\_sales WHERE Name LIKE 'Yakuza%'").show();  
  
 // 10 Самые популярные игры Square Enix  
 System.out.println("5 покупаемых игр Square Enix");  
 spark.sql("SELECT \* FROM game\_sales WHERE Publisher = 'Square Enix' ORDER BY Global\_Sales DESC LIMIT 10").show();  
  
  
 spark.stop();  
 }  
}

Результат работы программы представлен на рисунке 3.

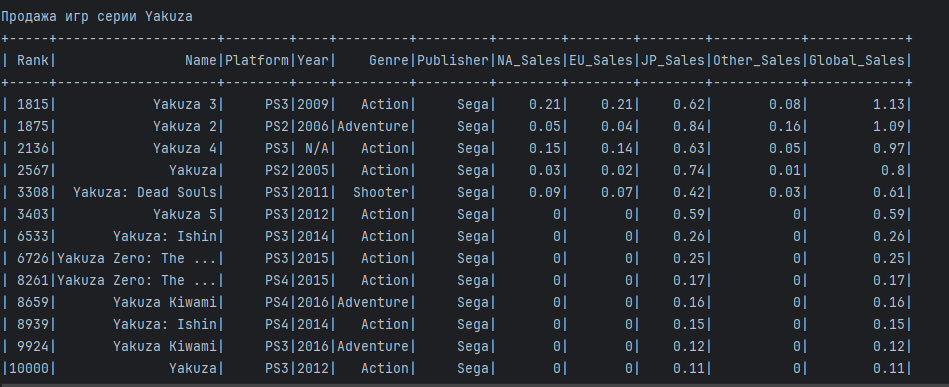


Рисунок 1 – Результат работы программы

**Вывод:** были освоены принципы работы со Spark в Java.