|  |  |
| --- | --- |
| 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 16** |

**Название:**

Работа со Scala и Spark

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

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

Москва, 2022

**Цель работы** –получение навыков работы с Scala и Spark.

**Ход работы**

**Задание:**

1. Выбрать любой датасет.
2. Сделать 10 выборок данных на ваше усмотрение

В качестве датасета был выбран соответствующий варианту датасет, а именно Russian\_passenger\_air\_service.

Листинг 1 – Код задания

package org.example

import org.apache.spark.sql.SparkSession

object SparkSessionTest extends App{

val spark = SparkSession.builder()

.master("local[1]")

.appName("SparkByExample")

.getOrCreate();

spark.sparkContext.setLogLevel("ERROR")

val datafile = spark.read

.format("com.databricks.spark.csv")

.option("header", true)

.load("C:\\Users\\fregat\\Downloads\\archive\\russian\_passenger\_air\_service\_2.csv")

datafile.show()

datafile.createOrReplaceTempView("passengers")

spark.sql("select \* from passengers").show()

spark.sql("SELECT COUNT(\*) FROM passengers").show()

spark.sql("SELECT \* from passengers WHERE passengers.Year = 2019").show()

spark.sql("SELECT \* from passengers WHERE passengers.January > 10000").show()

spark.sql("SELECT `Airport name` from passengers").show()

spark.sql("SELECT \* from passengers WHERE February BETWEEN 1000 AND 5000").show()

}

Пример выполнения запроса представлен на рисунке 1.

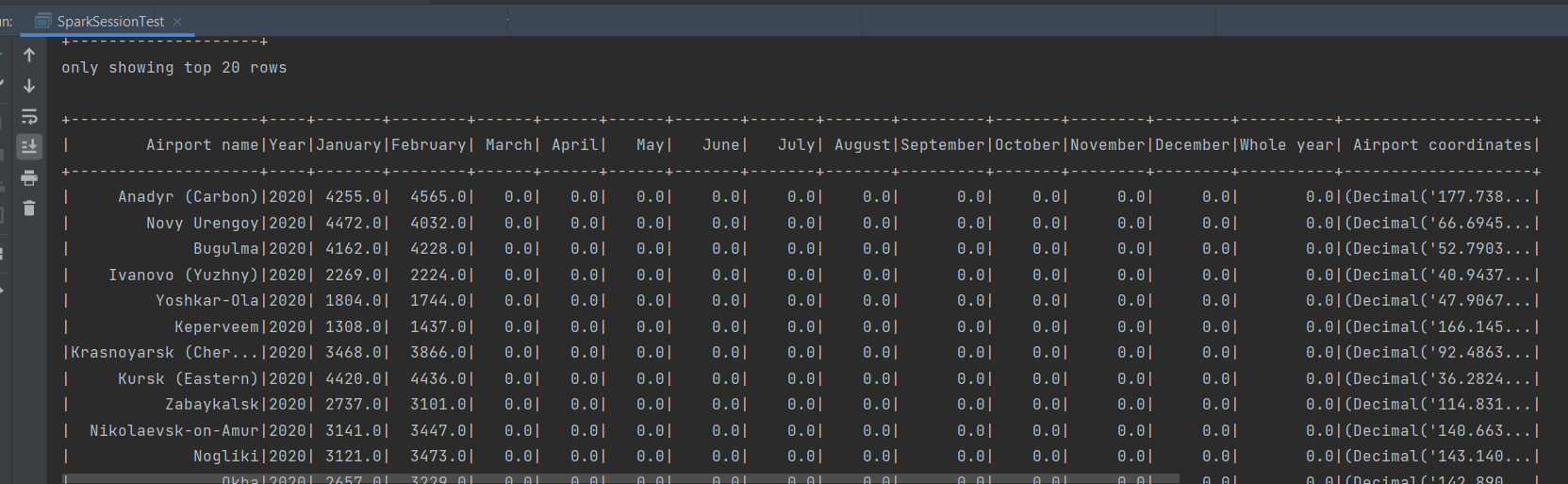


Рисунок 1 – Пример выполнения запроса

**Местоположение проектных файлов** –файлы проекта расположены в репозитории GitHub по следующим адресам:

<https://github.com/Fregatty/BigDataLanguages/tree/main/src/lab10>

**Вывод** – в результате выполнения лабораторной работы были получены навыки работы с Scala и Spark. Был выбран датасет и сделан ряд SQL запросов.