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**

О Т Ч Е Т

по лабораторной работе № 10

Вариант 12

Название: Scala Spark

Дисциплина: Языки программирования для работы с большими данными

Студент ИУ6-23М _____ П.А. Мартынюк
(Группа) (Подпись, дата) (И.О. Фамилия)

Преподаватель _____ П.В. Степанов
(Подпись, дата) (И.О. Фамилия)

Москва, 2022

Цель работы:

Получение навыков работы со Scala Spark.

Выполнение:

Задание:

1. Выбрать любой датасет (взят датасет из курсового проекта, тема «Поликлиника»)
2. Сделать 10 выборок данных на ваше усмотрение

Листинг выполнения одного из запросов (файл spark.scala)

```
import org.apache.spark.sql.SparkSession

object CounterDemo {
 def main(args: Array[String]): Unit = {
   val conf = new
SparkConf().setAppName("CounterDemo").setMaster("local[*]")
   val sc = new SparkContext(conf);
   val spark = SparkSession.builder.appName("Test app").getOrCreate()
   val path_1 = "/home/polina/Документы/scripts/visits.csv"
   val df_visits = spark.read.option("header", "true").csv(path_1)
   val path_2 = "/home/polina/Документы/scripts/patients.csv"
   val df_patients = spark.read.option("header", "true").csv(path_2)
   df_visits.createOrReplaceTempView("visits")
   df_patients.createOrReplaceTempView("patients")
   spark.sql("SELECT patient_age, count(patient_age) as counter FROM
patients, visits WHERE patients.patient_id = visits.patient_id AND
diagnosis = 'Грипп' GROUP BY patient_age").show()
   spark.stop()
 }
}
```

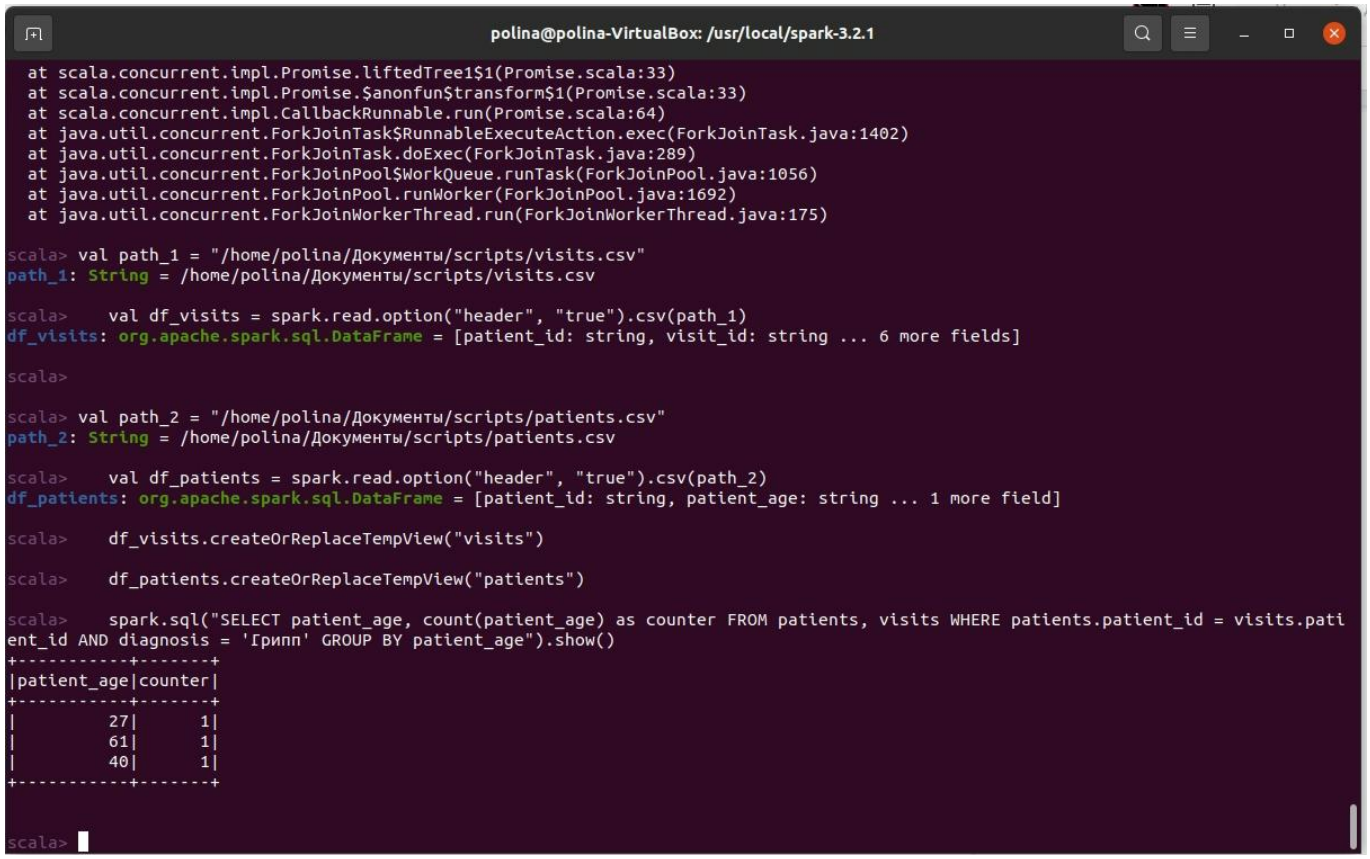

Рисунок 1 - Результат выполнения запроса

Ссылка на программное решение:

Программное решение представлено в репозитории распределённой системы управления версиями Git:

https://github.com/Owlfeather/JavaMagisterCourse/tree/main/Lab10/src

Вывод:

При выполнении лабораторной работы были получены навыки работы со Scala Spark.