

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 14

Hазвание: Scala Spark

Дисциплина: Языки программирования для работы с большими

данными

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

Цель работы: получение навыков работы с Scala Spark.

Задание:

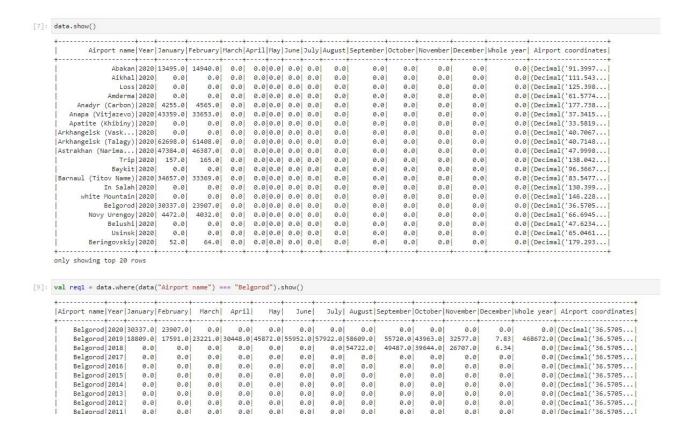
- 1. Выбрать любой датасет на kaggle.com
- 2. Сделать 10 выборок данных на ваше усмотрение

Выполнение.

```
[1]: import $ivy.`org.apache.spark::spark-sql:3.0.0`;
[1]: import Sivy.S
[2]: import org.apache.spark.sql._
                   val spark = SparkSession.
                                                   = SparkSession.
builder().
appName("scala-spark-notebook").
master("spark://spark-master:7077").
config("spark.executor.memory", "512m").
getOrCreate()
                22/05/26 15:58:26 INFO ResourceUtils: Resources for spark.driver:

22/05/26 15:58:26 INFO SparkContext: Submitted application: scala-spark-notebook
22/05/26 15:58:26 INFO SparkContext: Submitted application: scala-spark-notebook
22/05/26 15:58:26 INFO SecurityManager: Changing wiew acls to: root
22/05/26 15:58:26 INFO SecurityManager: Changing modify acls to: root
22/05/26 15:58:26 INFO SecurityManager: Annaging modify acls groups to:
22/05/26 15:58:26 INFO SecurityManager: Annaging modify acls groups to:
22/05/26 15:58:27 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(root); groups with wodify permissions: Set(root); groups with modify permissions: Set(root); groups with modify permissions: Set(root); groups with view permissions: Set(root); groups with modify permissions: Set(root); groups with view permissions: Set(root); groups with modify per
                22/05/26 15:58:29 INFO StandaloneSchedulerBackend: SchedulerBackend is ready for scheduling beginning after reached minRegisteredResourcesRatio: 0.0
                                       org.apache.spark.sql
                          ark: SparkSession = org.apache.spark.sql.SparkSession@2201bab
[3]: import org.apache.log4j.{Level, Logger};
Logger.getLogger("org").setLevel(Level.OFF);
[3]: import org.apache.log4i.{Level, Logger};
[4]: val data = spark.read.format("csv").option("sep", ",").option("header", "true").load("russian_passenger_air_service_2.csv")
 [4]: data: DataFrame = [Airport name: string, Year: string ... 14 more fields]
[5]: data.count
[5]: res4: Long = 3961L
[6]: data.printSchema
                     root

-- Airport name: string (nullable = true)
-- Year: string (nullable = true)
-- January: string (nullable = true)
-- January: string (nullable = true)
-- February: string (nullable = true)
-- March: string (nullable = true)
-- May: string (nullable = true)
-- June: string (nullable = true)
-- July: string (nullable = true)
-- August: string (nullable = true)
-- August: string (nullable = true)
-- September: string (nullable = true)
-- October: string (nullable = true)
-- November: string (nullable = true)
-- December: string (nullable = true)
-- Whole year: string (nullable = true)
-- Whole year: string (nullable = true)
-- Airport coordinates: string (nullable = true)
```



Ссылка на программное решение:

https://github.com/Time2HackJS/BigDataLanguages/tree/master/lr10

Вывод: в ходе лабораторной работы были получены навыки работы с Spark Scala.