



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 № 5

Название лабораторной работы: Spark

Дисциплина: Языки программирования для работы с большими данными

Студент гр. ИУ6-22М

(Подпись, дата)

Р.Г. Гаделия

(И.О. Фамилия)

Преподаватель

(Подпись, дата)

П.В. Степанов

(И.О. Фамилия)

Москва, 2023

Введение

Целью лабораторной работы является приобретение навыков работы со Spark на языке программирования Java.

Практическая часть

Задание 1

1) Выбрать любой датасет на kaggle.com

2) Сделать 10 выборок данных по выбранной предметной области Код

написанной программы представлен в листинге 1.

Листинг 1 – Программа для первого задания

```
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("Lab10")
            .master("local")
            .getOrCreate();

        Dataset<Row> flights = spark.read()
            .option("header", true)
            .option("inferSchema", true)
            .csv("/Users/ivangorshkov/Documents/BMSTU/BigDataJava/Java-IU6-12M/Lab10/russian_air_service_CARGO_AND_PARCELS.csv");
        flights.createOrReplaceTempView("flights");

        spark.sql("SELECT * FROM flights").show();
        spark.sql("select * from flights where Year = '2020'").show();
        spark.sql("select * from flights where Year = '2010'").show();
        spark.sql("select * from flights where May <
September").show();
        spark.sql("select * from flights where AirportName =
'Trip'").show();
        spark.sql("select AirportName, Year from flights order by
February desc").show();
        spark.sql("select AirportName, Year, March from flights where
January > 150.0 order by November desc").show();
        spark.sql("select AirportName, Year, April from flights order
by May desc").show();
        spark.sql("select AirportName, Year, May, June from flights
where Year = '2014' and AirportName = 'Trip' order by May
desc").show();
        spark.sql("select avg(January) from flights").show();
        spark.stop();
    }
}
```

Результат выполнения программы показан на рисунке 1.

AirportName	Year	January	February	March	April	May	June	July	August	September	October	November	December	Whole year	Airport coordinates
Abaken[2020]	43.58	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('91.3997...))
Aikhal[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('111.543...))
Loss[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('125.398...))
Anderna[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('61.5774...))
Anadny (Carbon)[2020]	122.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('177.738...))
Anapa (Vityazevo)[2020]	49.08	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('37.3415...))
Apatite (Khibiny)[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('33.5819...))
Arkhangelsk (Vask...)[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('40.7867...))
Arkhangelsk (Talag)[2020]	96.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('40.7148...))
Astrakhan (Narima...)[2020]	53.55	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('47.9998...))
Trip[2020]	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Baykit[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('96.3667...))
Barnaul (Titov Name)[2020]	163.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('83.5477...))
In Salah[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('130.399...))
white Mountain[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('146.228...))
Belgorod[2020]	5.24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('36.5785...))
Novy Urengoy[2020]	5.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('66.6945...))
Belushi[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('47.6234...))
Usinsk[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('65.0461...))
Beringovskiy[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('179.293...))

only showing top 20 rows

AirportName	Year	January	February	March	April	May	June	July	August	September	October	November	December	Whole year	Airport coordinates
Abaken[2018]	35.1	192.2	55.2	108.7	118.0	69.3	87.7	77.0	101.4	80.5	83.3	110.6	1207.0	0.0	0.0(Decimal('91.3997...))
Aikhal[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('111.543...))
Loss[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('125.398...))
Anderna[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('61.5774...))
Anadny (Carbon)[2018]	84.0	119.0	154.0	277.0	213.0	170.0	135.0	88.0	130.0	191.0	147.0	352.0	2852.0	0.0	0.0(Decimal('177.738...))
Anapa (Vityazevo)[2018]	14.62	21.43	25.1	25.25	34.66	59.19	89.76	91.2	62.42	47.53	31.6	29.72	532.48	0.0	0.0(Decimal('37.3415...))
Antypayuta[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Not found
Apatite (Khibiny)[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('33.5819...))
Arkhangelsk (Vask...)[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('40.7867...))
Arkhangelsk (Talag)[2018]	122.8	156.0	181.0	238.5	483.2	191.7	263.4	153.5	280.22	191.3	353.9	479.2	3815.52	0.0	0.0(Decimal('40.7148...))
Astrakhan (Narima...)[2018]	55.41	41.91	152.75	171.97	151.71	158.99	68.19	63.94	76.19	71.23	63.54	71.06	738.89	0.0	0.0(Decimal('47.9998...))
Achinsk[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('98.564...))
Trip[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Auctions[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('113.479...))
Baykit[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('96.3667...))
Barnaul (Titov Name)[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('83.5477...))
In Salah[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('130.399...))
white Mountain[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('146.228...))
Belgorod[2018]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('36.5785...))
Novy Urengoy[2018]	15.5	15.1	16.7	47.7	66.9	52.7	49.1	43.2	57.0	54.0	85.0	55.1	559.6	0.0	0.0(Decimal('66.6945...))

only showing top 20 rows

AirportName	Year	January	February	March	April	May	June	July	August	September	October	November	December	Whole year	Airport coordinates
Abaken[2020]	43.58	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('91.3997...))
Aikhal[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('111.543...))
Loss[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('125.398...))
Anderna[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('61.5774...))
Anadny (Carbon)[2020]	122.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('177.738...))
Anapa (Vityazevo)[2020]	49.08	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('37.3415...))
Apatite (Khibiny)[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('33.5819...))
Arkhangelsk (Vask...)[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('40.7867...))
Arkhangelsk (Talag)[2020]	96.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('40.7148...))
Astrakhan (Narima...)[2020]	53.55	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('47.9998...))
Trip[2020]	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Baykit[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('96.3667...))
Barnaul (Titov Name)[2020]	163.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('83.5477...))
In Salah[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('130.399...))
white Mountain[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('146.228...))
Belgorod[2020]	5.24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('36.5785...))
Novy Urengoy[2020]	5.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('66.6945...))
Belushi[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('47.6234...))
Usinsk[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('65.0461...))
Beringovskiy[2020]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('179.293...))

only showing top 20 rows

AirportName	Year	January	February	March	April	May	June	July	August	September	October	November	December	Whole year	Airport coordinates
Trip[2020]	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Trip[2019]	1.36	2.7	3.52	3.42	2.09	2.42	1.79	2.08	3.81	3.46	1.94	1.54	38.13	0.0	0.0(Decimal('138.842...))
Trip[2018]	0.0	0.0	0.0	0.0	0.0	0.0	1.82	2.1	5.54	2.32	4.89	5.2	0.0	0.0	0.0(Decimal('138.842...))
Trip[2017]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Trip[2016]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Trip[2015]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Trip[2014]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Trip[2013]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Trip[2012]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Trip[2011]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Trip[2010]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Trip[2009]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Trip[2008]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))
Trip[2007]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0(Decimal('138.842...))

AirportName	Year	January	February	March	April	May	June	July	August	September	October	November	December	Whole year	Airport coordinates
Anapa (Vityazevo)[2019]	45.92	53.15	54.0	54.72	52.0	67.45	172.31	72.57	70.0	63.0	69.0	82.1	856.22	0.0	0.0(Decimal('37.3415...))

```

+-----+-----+
|      AirportName|Year|
+-----+-----+
|Moscow (Sheremety...|2012|
|Moscow (Sheremety...|2019|
|Moscow (Sheremety...|2018|
|Moscow (Sheremety...|2017|
|Moscow (Sheremety...|2011|
|Moscow (Sheremety...|2016|
|Moscow (Sheremety...|2013|
|Moscow (Sheremety...|2014|
|Moscow (Sheremety...|2008|
| Moscow (Domodedovo)|2013|
|Moscow (Sheremety...|2010|
| Moscow (Domodedovo)|2011|
|Moscow (Sheremety...|2007|
| Moscow (Domodedovo)|2014|
| Moscow (Domodedovo)|2012|
|Moscow (Sheremety...|2015|
| Moscow (Domodedovo)|2018|
| Moscow (Domodedovo)|2008|
| Moscow (Domodedovo)|2015|
|Moscow (Sheremety...|2009|
+-----+-----+

```

only showing top 20 rows

```

+-----+-----+-----+
|      AirportName|Year|  March|
+-----+-----+-----+
|Moscow (Sheremety...|2019| 26703.6|
|Moscow (Sheremety...|2017| 25153.5|
|Moscow (Sheremety...|2018| 24478.7|
|Moscow (Sheremety...|2011| 22120.1|
|Moscow (Sheremety...|2010| 14919.0|
|Moscow (Sheremety...|2016| 17486.6|
|Moscow (Sheremety...|2013| 19005.2|
|Moscow (Sheremety...|2014| 21639.4|
|Moscow (Sheremety...|2012| 27617.9|
|Moscow (Sheremety...|2015| 14128.3|
|Moscow (Sheremety...|2009| 11742.7|
| Moscow (Domodedovo)|2012| 13593.3|
| Moscow (Domodedovo)|2013|14225.51|
| Moscow (Domodedovo)|2011|13339.73|
|Moscow (Sheremety...|2007| 13362.5|
|Moscow (Sheremety...|2008| 14409.4|
| Moscow (Domodedovo)|2014|14015.58|
| Moscow (Domodedovo)|2010|11290.43|
| Moscow (Domodedovo)|2007| 8754.34|
| Moscow (Domodedovo)|2009| 8011.42|
+-----+-----+-----+

```

only showing top 20 rows

```

+-----+-----+-----+
|      AirportName|Year|  April|
+-----+-----+-----+
|Moscow (Sheremety...|2012| 25020.2|
|Moscow (Sheremety...|2019| 28634.2|
|Moscow (Sheremety...|2018| 24709.4|
|Moscow (Sheremety...|2017| 23304.6|
|Moscow (Sheremety...|2011| 20935.6|
|Moscow (Sheremety...|2013| 18592.5|
|Moscow (Sheremety...|2016| 19383.1|
|Moscow (Sheremety...|2010| 15065.7|
|Moscow (Sheremety...|2014| 16924.1|
|Moscow (Sheremety...|2015| 15026.4|
| Moscow (Domodedovo)|2013|13922.46|
|Moscow (Sheremety...|2008| 14101.0|
| Moscow (Domodedovo)|2011|13999.56|
| Moscow (Domodedovo)|2012|14119.27|
| Moscow (Domodedovo)|2014|14121.12|
|Moscow (Sheremety...|2007| 12591.7|
| Moscow (Domodedovo)|2008|11837.18|
| Moscow (Domodedovo)|2018| 9397.39|
|Moscow (Sheremety...|2009| 11585.2|
| Moscow (Domodedovo)|2017|10918.44|
+-----+-----+-----+

```

only showing top 20 rows

```

+-----+-----+-----+
|AirportName|Year|May|June|
+-----+-----+-----+
|      Trip|2014|0.0| 0.0|
+-----+-----+-----+

```

```
+-----+
| avg(January) |
+-----+
| 146.47392072708885 |
+-----+
```

Рисунок 1 – Результат выполнения программы

Вывод: В результате выполнения лабораторной работы были приобретены навыки работы со Spark на языке программирования Java.