

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, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА КОМПЬЮТЕРНЫЕ СИСТЕМЫ И СЕТИ (ИУ6)

НАПРАВЛЕНИЕ ПОДГОТОВКИ **09.04.01 Интеллектуальные системы анализа**, обработки и интерпретации больших данных

ОТЧЕТ

по лабораторной работе № 10

Вариант 13

Название: Spark

Дисциплина: Языки программирования для работы с большими данными

Студент	ИУ6-22М		В.А.Ловцов
	(Группа)	(Подпись, дата)	(И.О. Фамилия)
Преподаватель			П.В. Степанов
		(Подпись, дата)	(И.О. Фамилия)

Цель: ознакомиться с работой Spark в языке программирования Java.

Задание: сделать 10 выборок данных по выбранной предметной области

Был выбран набор данных об авиаперевозках.

```
Код класса Main:
```

```
package atomic.engineering
     import org.apache.spark.sql.SparkSession
     object SparkDemo {
       def main(args: Array[String]): Unit = {
         val
                                   spark
SparkSession.builder().master("local[1]").appName("AtomicEngin
eering").getOrCreate()
         println("name:" + spark.sparkContext.appName)
         val
                                   path
"src/russian_air_service CARGO AND PARCELS.csv"
         val df = spark.read.option("header", "true").csv(path)
     //
          df.show()
         df.createOrReplaceTempView("data")
         var sqlDF = spark.sql("SELECT * FROM data limit 10")
         sqlDF.show()
         sqlDF.printSchema
```

```
sqlDF = spark.sql("SELECT COUNT(*) AS row count FROM
data")
         sqlDF.show()
         sqlDF = spark.sql("SELECT DISTINCT Year FROM data")
         sqlDF.show()
         sqlDF = spark.sql("SELECT * FROM data WHERE `Airport
name` LIKE 'B%'")
         sqlDF.show()
         sqlDF = spark.sql("SELECT `Airport name`, Year FROM
data LIMIT 10")
         sqlDF.show()
         sqlDF = spark.sql("SELECT * FROM data WHERE February >
600")
         sqlDF.show()
         sqlDF = spark.sql("SELECT Year, SUM(September) AS total
FROM data GROUP BY Year")
         sqlDF.show()
         sqlDF = spark.sql("SELECT * FROM data ORDER BY June
DESC limit 10")
```

```
sqlDF.show()

sqlDF = spark.sql("SELECT AVG(October) AS average,
MAX(October) AS maximum FROM data")

sqlDF.show()

sqlDF = spark.sql("SELECT COUNT(DISTINCT `Airport
name`) AS airports FROM data")

sqlDF.show()

sqlDF = spark.sql("SELECT `Airport name`, SUM(December)
AS total_load FROM data GROUP BY `Airport name` ORDER BY
total_load DESC LIMIT 10")

sqlDF.show()

spark.stop()
}
```

Работа программы показана на рисунке 1.

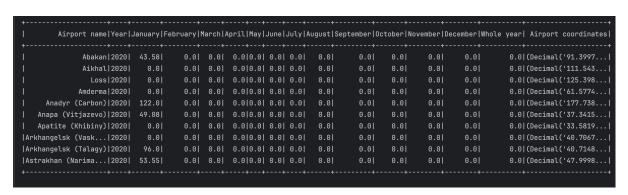


Рисунок 1 – Работа программы

Вывод: был изучен Spark, написана программа для осуществления запросов к данным