|  |  |
| --- | --- |
| 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 (ИУ-6)

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**

# Дисциплина: Языки программирования для работы с большими данными

**Название:** Scala. Spark

Студент гр. **ИУ6-23М \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А.П.Завьялова**

(Подпись, дата) (И.О. Фамилия)

Преподаватель  **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ П.В.Степанов**

(Подпись, дата) (И.О. Фамилия)

Москва, 2022

**Условие:**

1. Установить Hadoop+Spark
2. Выбрать любой датасет
3. Cделать 10 выборок данных на ваше усмотрение

**Фрагмент решения:**

# один из запросов на языке Python

df\_reader = spark.read.csv('hdfs://localhost:9000/library/reader\_file.csv', header=True)

df\_book = spark.read.csv('hdfs://localhost:9000/library/book\_file.csv', header=True)

df\_reader.createOrReplaceTempView("reader")

df\_book.createOrReplaceTempView("book")

result = spark.sql("""SELECT DISTINCT r.id, r.source\_fio, r.source\_registration\_date,

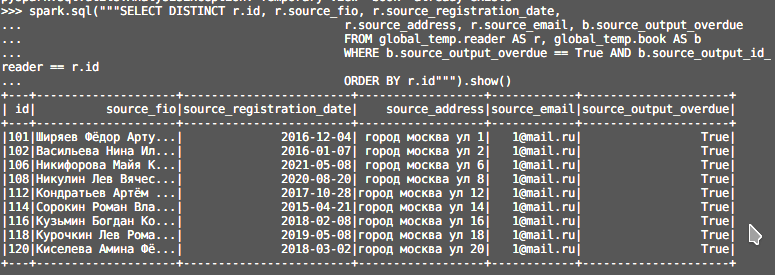
r.source\_address, r.source\_email, b.source\_output\_overdue

FROM reader AS r, book AS b

WHERE b.source\_output\_overdue == True AND b.source\_output\_id\_reader == r.id

ORDER BY r.id""")

result.show()



**Вывод:**

В ходе выполнения лабораторной работы, получены навыки работы с Scala и Spark, а именно: установлены все необходимые пакеты и переменные для работы Hadoop+Spark, а также написаны запросы на языке Python.

Репозиторий с кодом: <https://github.com/AnnaZav/Java_labs_bigdata.git>