|  |  |
| --- | --- |
| 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 (ИУ6)**

НАПРАВЛЕНИЕ ПОДГОТОВКИ 09.04.01/07 **Интеллектуальные системы анализа, обработки и интерпретации больших данных**

**Отчет**

|  |
| --- |
| **по лабораторной работе № 10** |

**Название:**

Spark

**Дисциплина:** Языки программирования для работы с большими данными

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент | ИУ6-23М |  |  | Т.А. Малкина |
|  | (Группа) |  | (Подпись, дата) | (И.О. Фамилия) |
|  |  |  |  |  |
| Преподаватель |  |  |  | П.В. Степанов |
|  |  |  | (Подпись, дата) | (И.О. Фамилия) |

Москва, 2022

Вариант 1:

Cделать 10 выборок данных на ваше усмотрение

Решение:

|  |
| --- |
| from pyspark import SparkContext  from datetime import datetime  from pyspark.sql import SparkSession  import time  start\_time = time.time()  spark = SparkSession.builder.appName('abc').getOrCreate()  df=spark.read.csv('hdfs://localhost:9000/csv/order.csv',header=True)  my\_table = df.createOrReplaceTempView('order')  sql\_table1 = spark.sql('SELECT sum(order\_value) FROM order WHERE data\_order between "2021-01-01" and "2021-12-12"').show()  sql\_table2 = spark.sql('SELECT data\_order FROM orders') .show()  sql\_table3 = spark.sql('SELECT \* FROM order') .show()  sql\_table4 = spark.sql('SELECT sum(order\_value) FROM order WHERE data\_order between "2021-01-01" and "2021-12-12"') .show()  sql\_table5= spark.sql('SELECT \* FROM order WHERE data\_order between "2021-01-01" and "2021-12-12"') .show()  sql\_table6 = spark.sql('SELECT \* FROM order WHERE id\_order between 1 and 5') .show()#  sql\_table7 = spark.sql('SELECT \* FROM cutter') .show()  sql\_table8 = spark.sql('SELECT \* FROM cutter WHERE index between 1 and 5') .show()  sql\_table9 = spark.sql('SELECT feedback\_about\_cutter FROM cutter WHERE index between 1 and 5') .show()  sql\_table10 = spark.sql('SELECT cutter\_detail FROM cutter WHERE index=10') .show()  quit() |

Ссылка на репозиторий:

https://github.com/Carpediem2026/BigData