|  |  |
| --- | --- |
| 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 Информатика и вычислительная техника**

**Отчет**

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

**Название:** Обработка больших данных

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент | ИУ6И-21М |  |  | Ши Жуньфан |
|  | (Группа) |  | (Подпись, дата) | (И.О. Фамилия) |
|  |  |  |  |  |
| Преподаватель |  |  |  | П. В. Степанов |
|  |  |  | (Подпись, дата) | (И.О. Фамилия) |

Москва, 2020

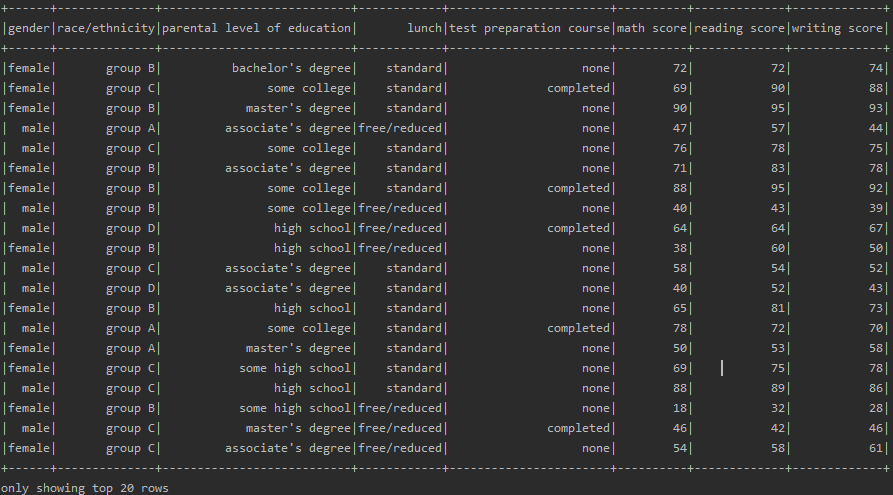
**Вариант 1**

Скачать датасет с Kernel и выполнить несколько запросов.

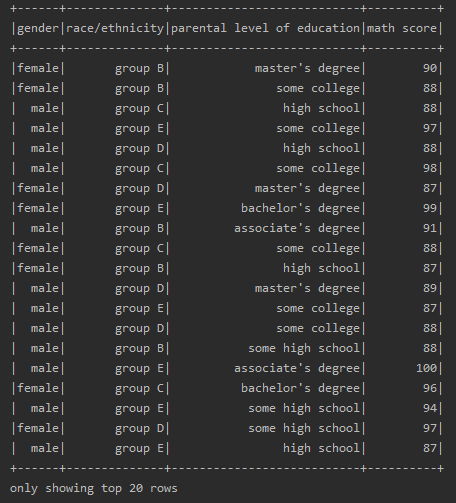
|  |
| --- |
| import org.apache.spark.sql.SparkSession import org.apache.spark.{SparkConf, SparkContext} object Lab10 {  def main(args: Array[String]): Unit = {  val conf = new SparkConf().setAppName("CounterDemo").setMaster("local[\*]")  val sc = new SparkContext(conf);  val spark = SparkSession.*builder*.appName("Test app").getOrCreate()  val datafile = spark.read  .format("com.databricks.spark.csv")  .option("header",true)  .load("C:\\Users\\Stone\\Desktop\\StudentsPerformance.csv")  // datafile.show()  datafile.createOrReplaceTempView("students")  // spark.sql("SELECT COUNT(\*) FROM students").show()  spark.sql("select \* from students").show()  //spark.sql("select `race/ethnicity`, count(`race/ethnicity`) from students group by `race/ethnicity`").show()  spark.sql("select gender,`race/ethnicity`, `parental level of education`, `math score` from students WHERE `math score` > 85").show()  spark.sql("select gender,`race/ethnicity`, `parental level of education`, `math score` ,`writing score`,`reading score`from students WHERE `math score` > 75 and`reading score`>80 ").show()  spark.stop()    }  } |

**Результаты**

select \* from students：



select gender,`race/ethnicity`, `parental level of education`, `math score` from students WHERE `math score` > 85：



select gender,`race/ethnicity`, `parental level of education`, `math score` ,`writing score`,`reading score`from students WHERE `math score` > 75 and`reading score`>80：

