|  |  |
| --- | --- |
|  | **Министерство образования и науки 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

НАПРАВЛЕНИЕ ПОДГОТОВКИ: 09.04.01 Информатика и вычислительная техника

**О т ч е т**

**по лабораторной работе №10**

**Дисциплина: Языки программирование при работе с большими данными**

**Название:**

студент группы ИУ6-21М  **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Коваленко М.В.**

(Подпись, дата) (Фамилия И.О.)

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

(Подпись, дата) (Фамилия И.О.)

Москва, 2020

**Задание**

Разобраться с работой с большими данными, используя Scala, Spark и данные с Kaggle.

**Данные**

Для решения задачи были взяты данные о больничных койках в больницах США, их количестве и возможности заселения этих коек.

**Программа**

|  |
| --- |
| package LR10  import org.apache.spark.sql.SparkSession  import org.apache.spark.{SparkConf, SparkContext}  object LR10\_V1 {  def main(args: Array[String]): Unit = {  val conf = new SparkConf().setAppName("LR10\_V1").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/User/IdeaProjects/LR10/HRR\_Scorecard.csv")  //datafile.show()  datafile.createOrReplaceTempView("HRR\_Scorecard")  // Максимальное количество прогнозируемых зараженных COVID-19 в США старше 18 лет  spark.sql("SELECT MAX(`Projected Infected Individuals`) as max FROM HRR\_Scorecard group by `HRR`").show()  // Среднее количество пенсионеров в больницах США  spark.sql("SELECT `HRR`, AVG(`Population 65+`) as avg FROM HRR\_Scorecard group by `HRR`").show()  // Количество мест в больницах Hospital и ICU  spark.sql("select `Total ICU Beds`, `Total Hospital Beds`, `HRR` FROM HRR\_Scorecard WHERE `Total Hospital Beds` > 100").show()  spark.stop()  }  } |

**Вывод**

В результате выполнения данной лабораторной работы я научилась работать с большими данными, используя Scala и Spark.