Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Факультет інформатики та обчислювальної техніки

Кафедра автоматизованих систем обробки інформації та управління

Методи і засоби обробки великих даних

Лабораторна робота №5

«Робота з файлами в Spark Streaming»

Виконав:

студент групи ІС-12мп

Коноплянка Д. С.

Перевірила:

Тимофєєва Ю. С.

Київ 2021 р.

Варіант 6.

Файл Forest.csv.

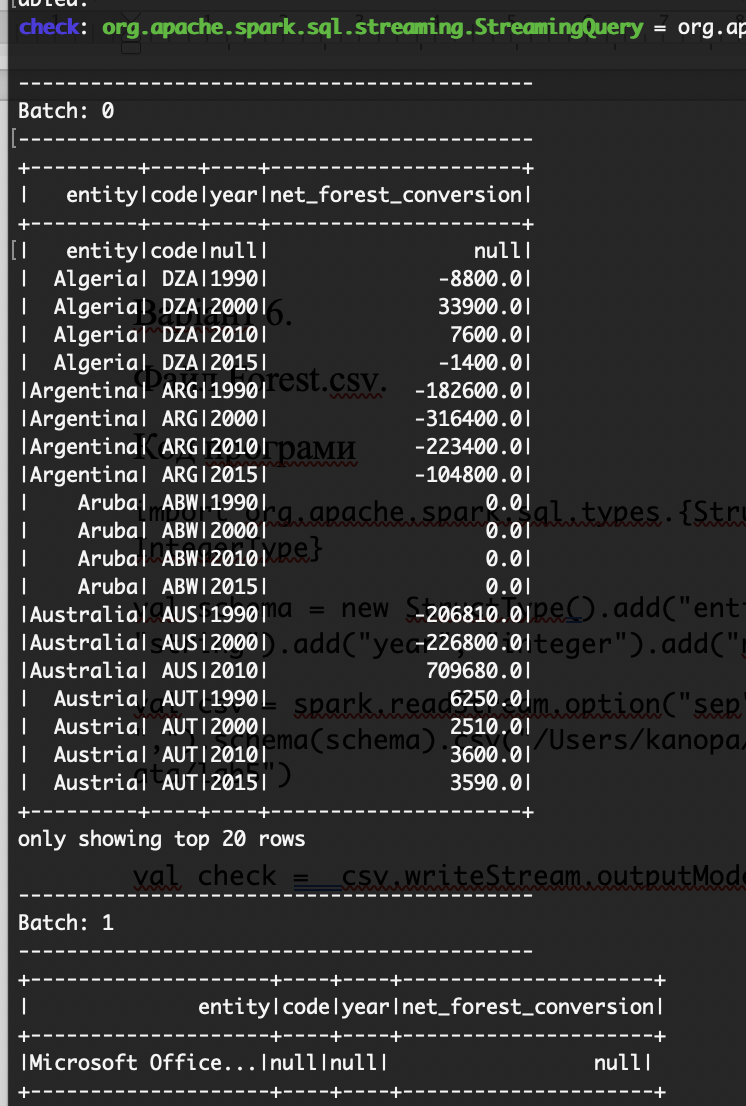
Код програми

import org.apache.spark.sql.types.{StructType, StructField, StringType, IntegerType}

val schema = new StructType().add("entity", "string").add("code", "string").add("year", "integer").add("net\_forest\_conversion", "float")

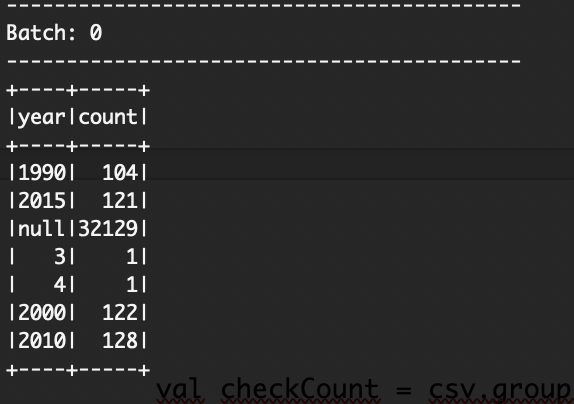
val csv = spark.readStream.option("sep", ",").schema(schema).csv("/Users/kanopa/Documents/education/kpi/5course/BigData/lab5")

val check = csv.writeStream.outputMode("update").format("console").start()



val checkCount = csv.groupBy("year").count()

checkCount.writeStream.outputMode("update").format("console").start()



val checkCount = csv.groupBy("entity").count()

checkCount.writeStream.outputMode("update").format("console").start()

