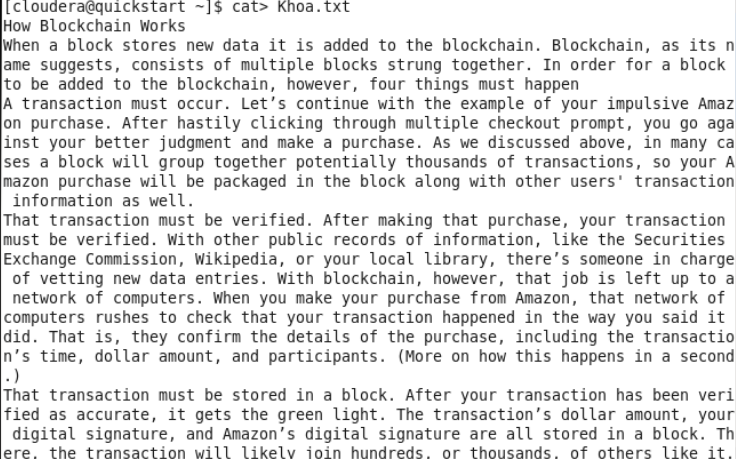
Tên: Lê Viết Khoa

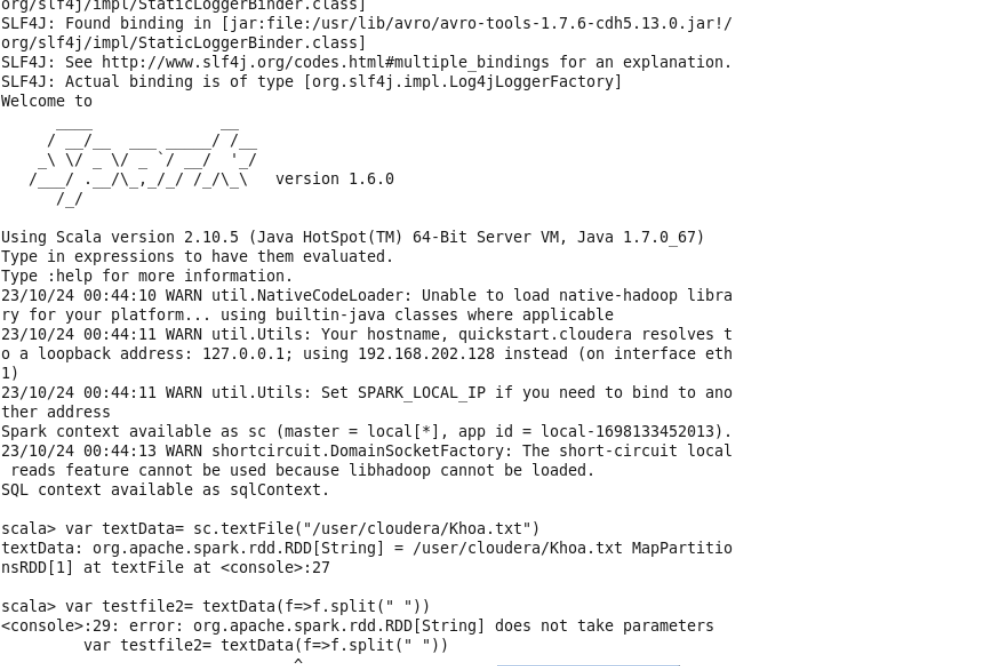
MSSV:1050080018

Bước 1: Khởi tạo file dữ liệu Khoa.txt



Bước 2: khởi động spark-shell trong terminal

Bước 3: var textData =sc.textFile("/user/cloudera/Khoa.txt") rồi nhấn enter



Bước 4: var testfile2 = textData.flatMap(f=>f.split(" ")) rồi enter

Bước 5: var testfile3= testfile2.map(m=>(m, 1)) rồi enter

Bước 6: var count = testfile3. reduceByKey(\_+\_) rồi enter

Bước 7: count.foreach(println) rồi enter

Sẽ hiện ra kết quả như hình ở dưới

