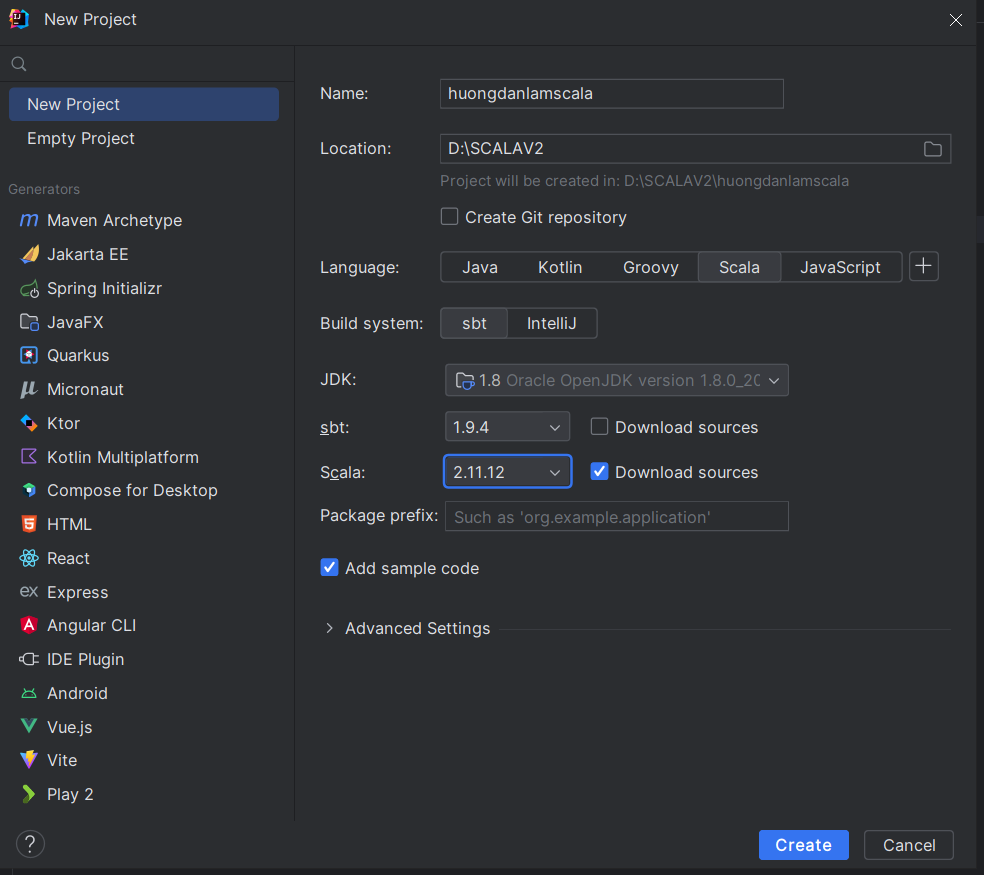
**Hướng dẫn cơ bản về Scala**

Chuẩn bị:

Cài biến môi trường Spark, JDK : [Apache Spark - Install Apache Spark On Windows 10 |Spark Tutorial | Part 1 - YouTube](https://www.youtube.com/watch?v=cYL42BBL3Fo)

Cài Inteliji IDEA: [Download IntelliJ IDEA – The Leading Java and Kotlin IDE (jetbrains.com)](https://www.jetbrains.com/idea/download/?section=windows)

1. tạo project

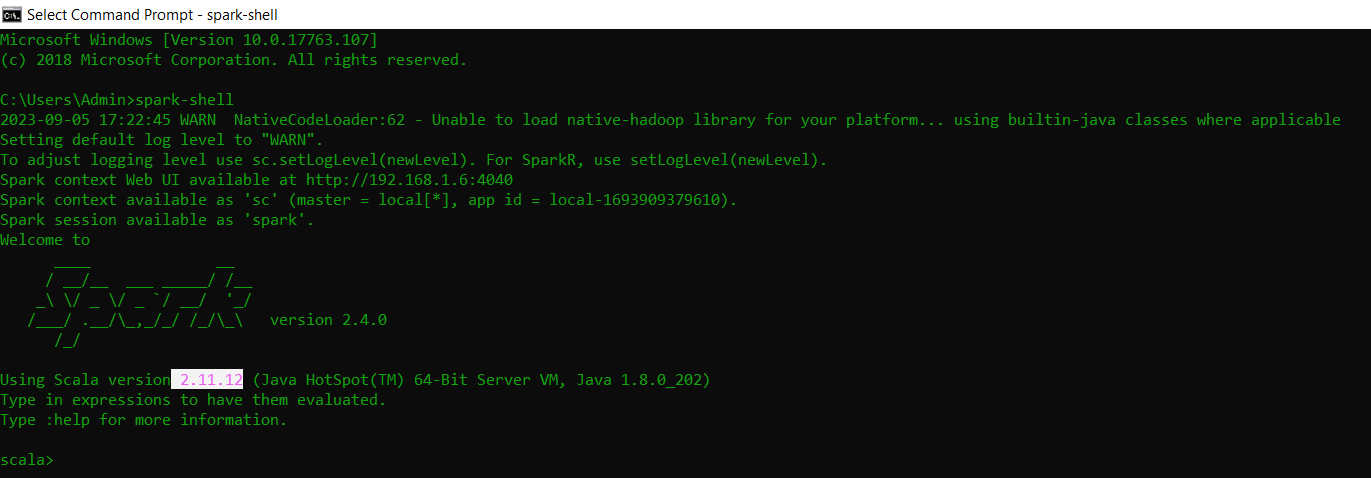


Hãy chọn như bước trên. Lưu ý :

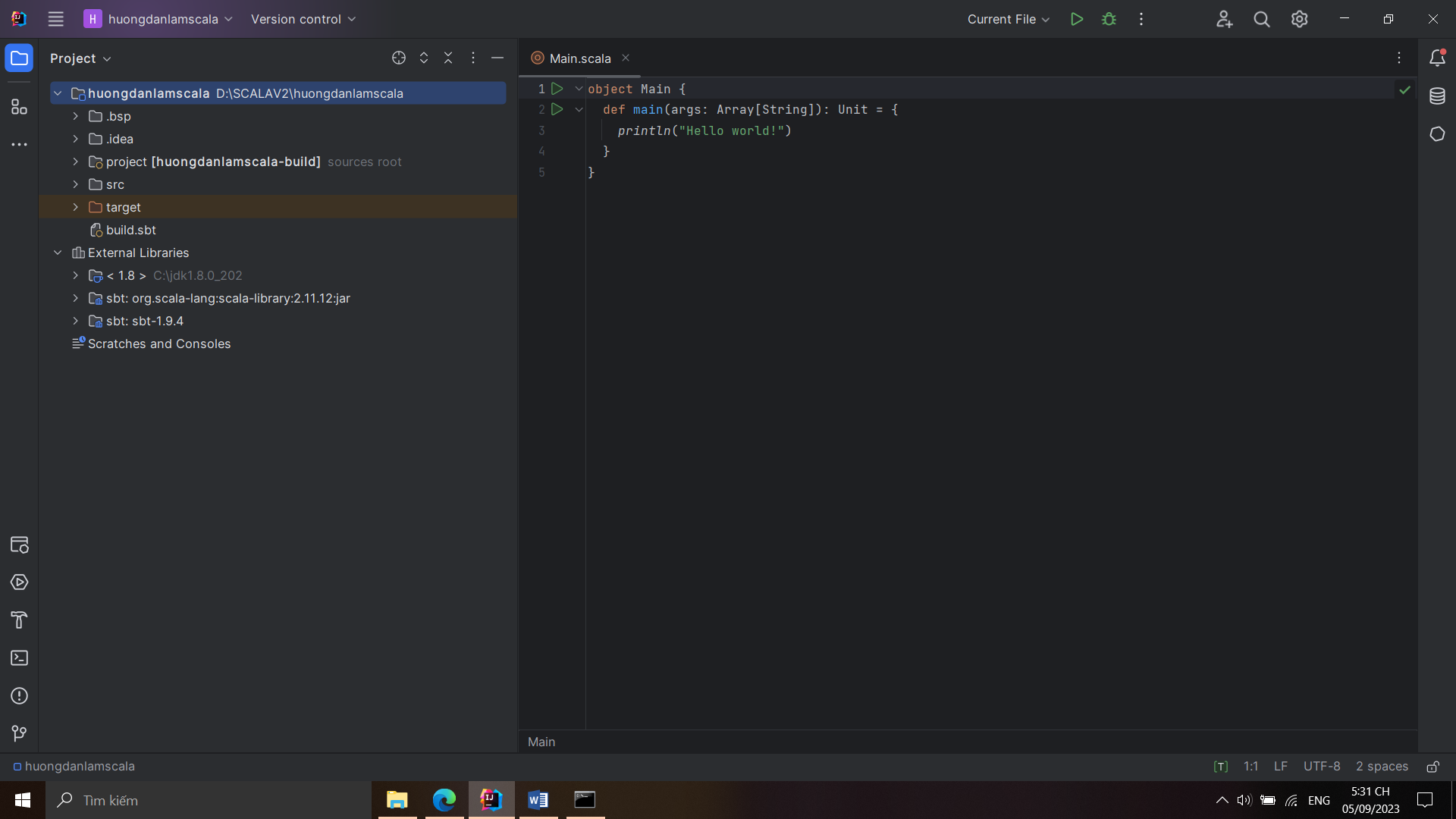
- location tùy chọn

- scala version : hãy vào cmd gõ lệnh spark-shell để check phiên bản tương ứng ( ở đây nó hiện là 2.11.12 nên sẽ chọn kia là 2.11.12 ) sbt để mới nhất

-JDK cũng để tương ứng như scala version



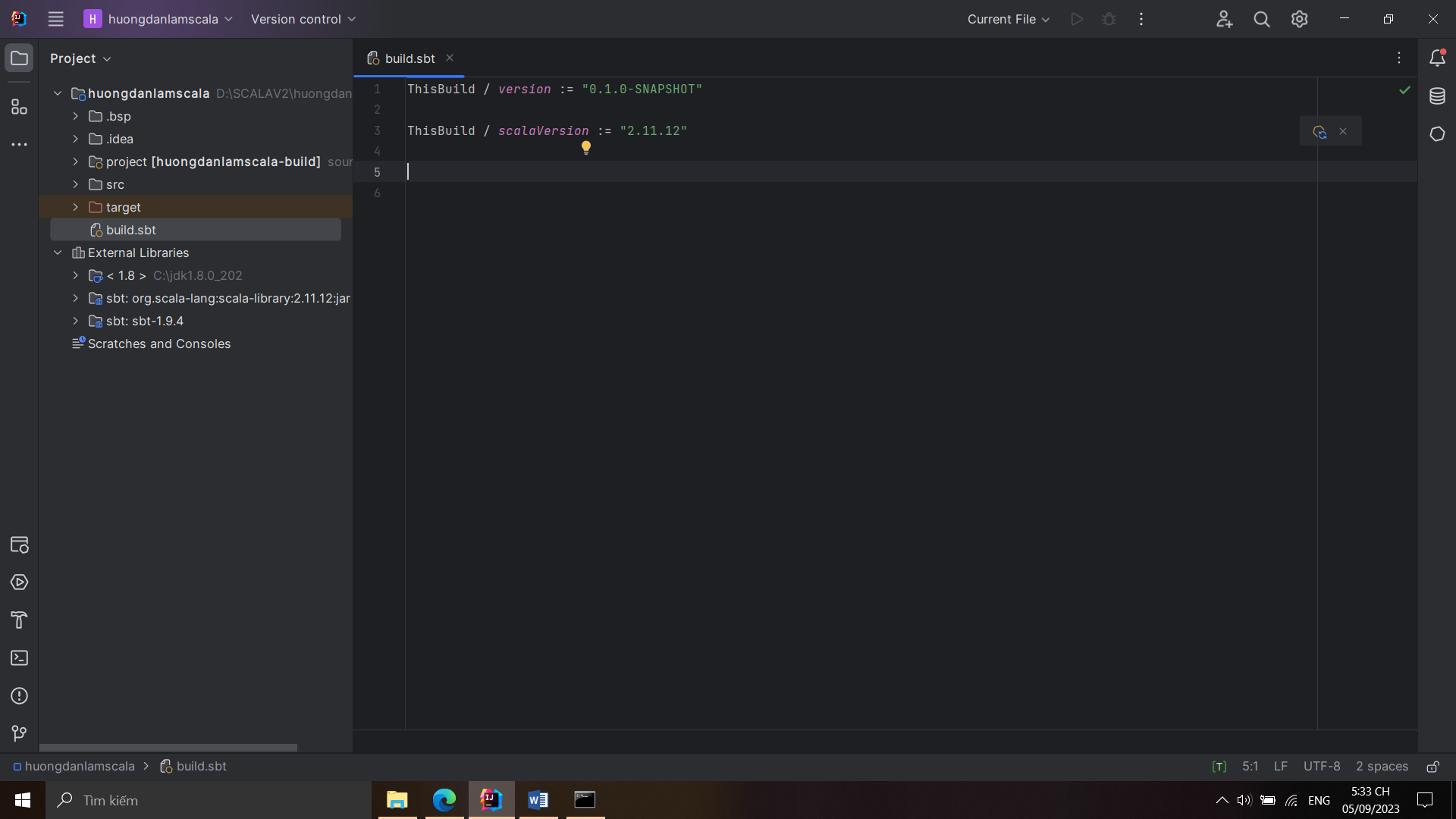
Sau khi xong kết quả sẽ hiện :



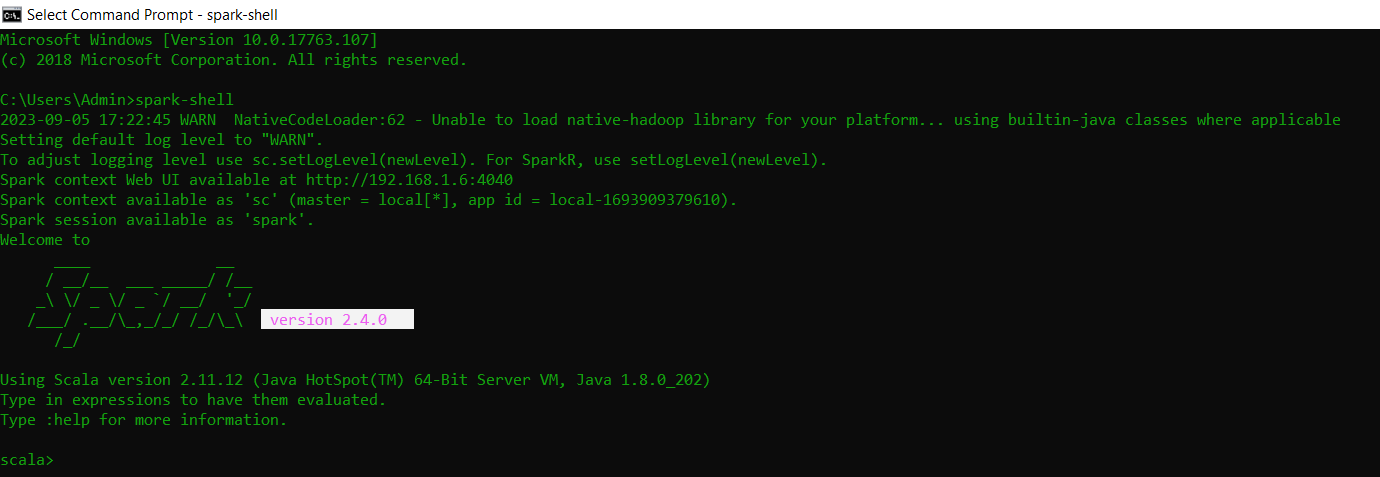
( xóa file main sacala )

2. Cấu hình

Vào file build.sbt kết quả hiển thị như sau :

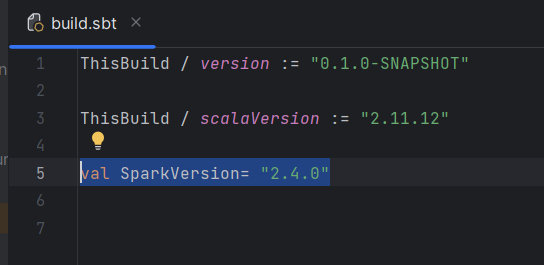


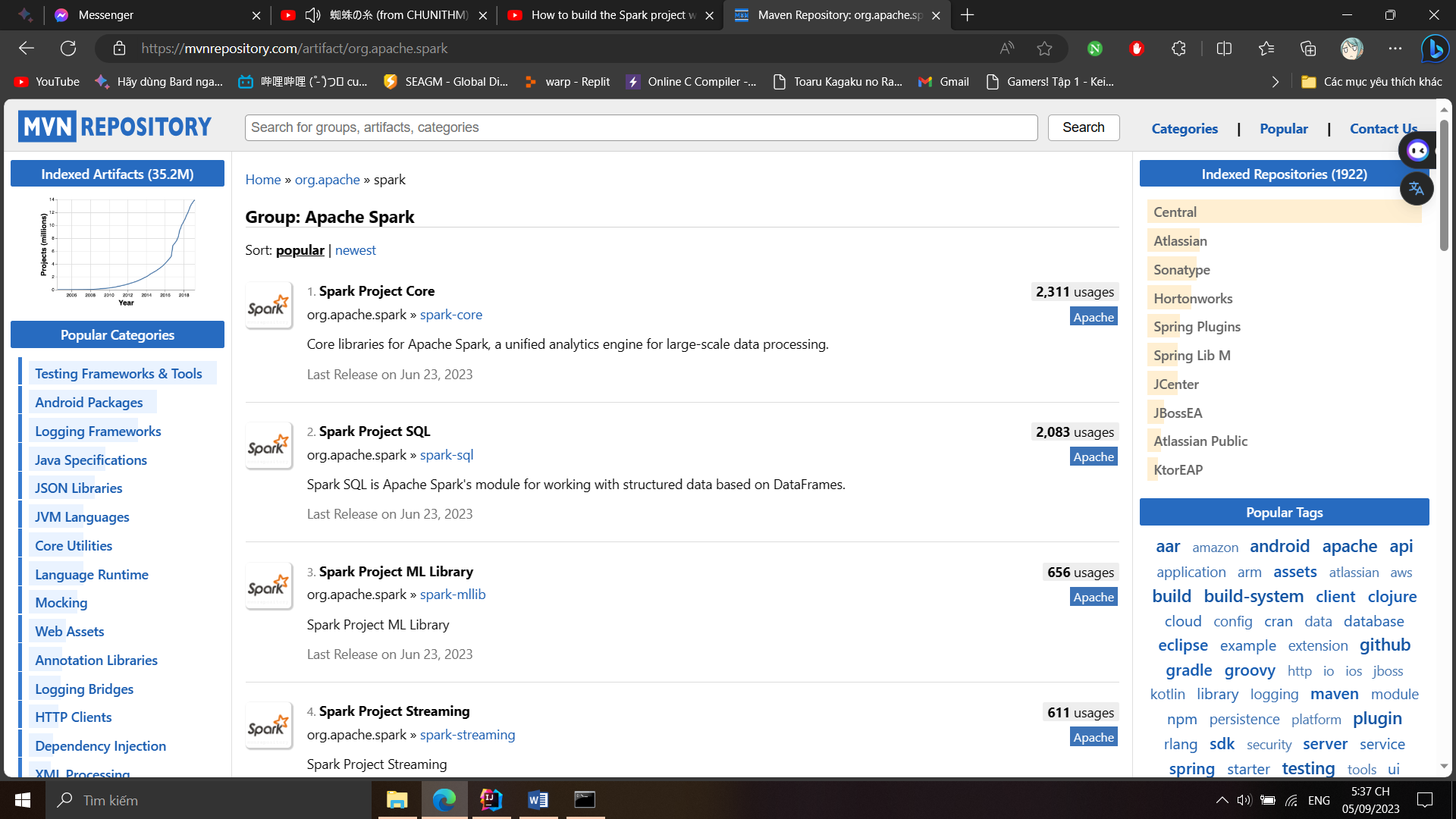
Quay lại cmd để ý version ( tùy phiên bản sử dụng của mỗi máy )



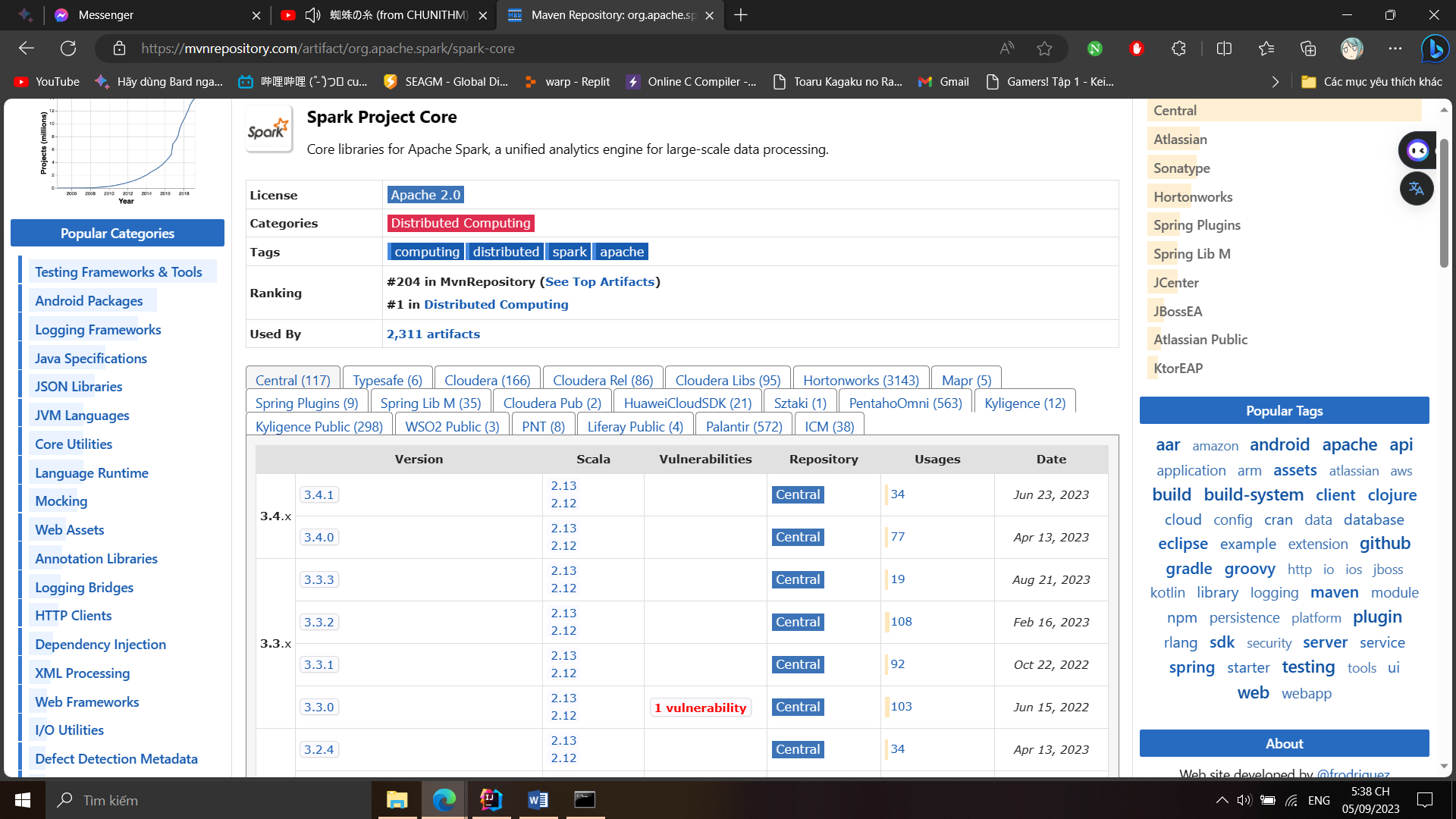
Và gõ: val SparkVersion= "<phiên bản được phôi trắng trên ảnh> "

Mình dùng 2.4.0 kết quả :



Sau khi để hãy lên web [Maven Repository: org.apache.spark (mvnrepository.com)](https://mvnrepository.com/artifact/org.apache.spark) : 

Chọn Spark Core :

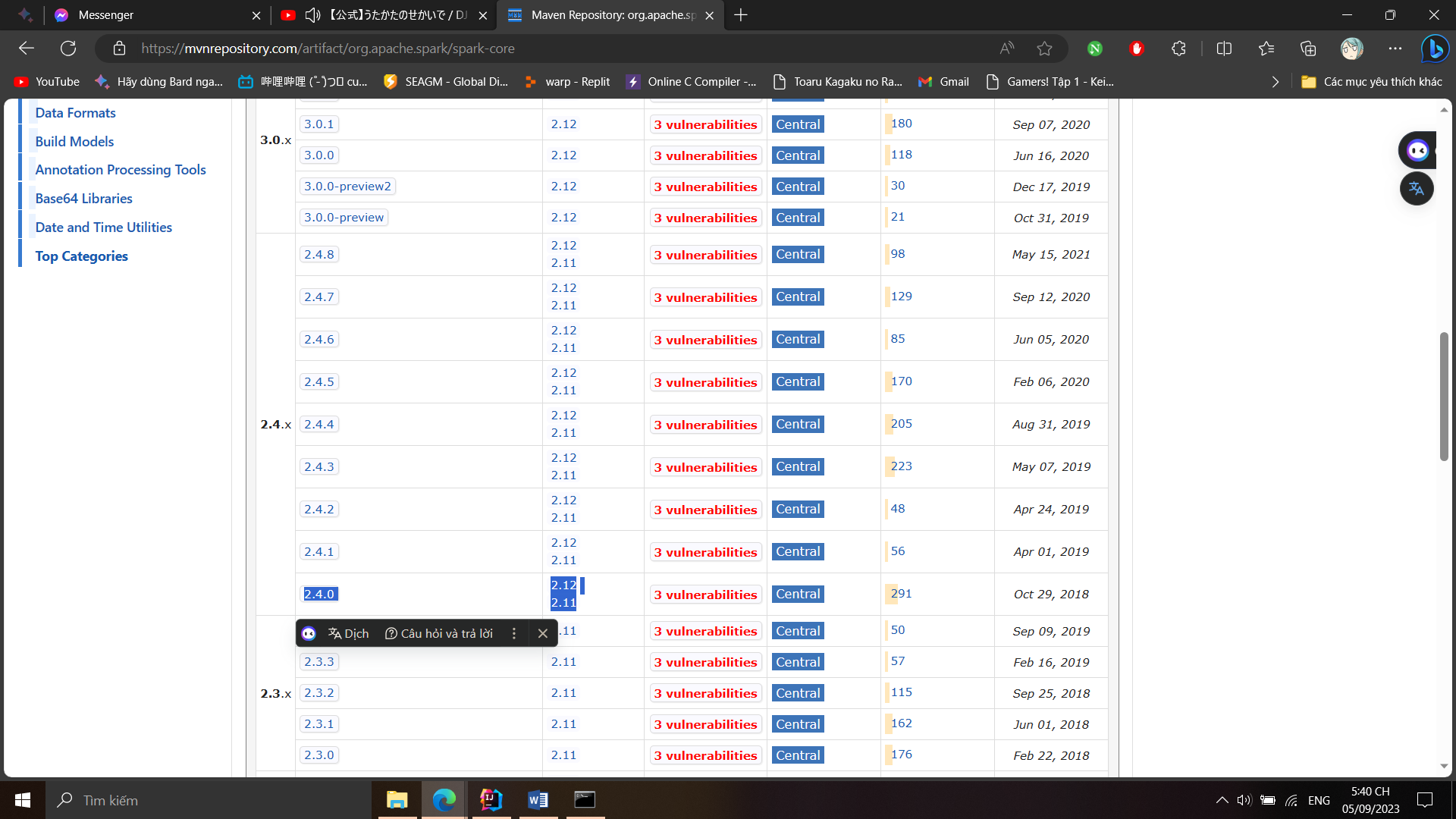


Dựa vào cmd như trên hãy tìm bản phù hợp: ( kết quả này tùy mỗi máy )

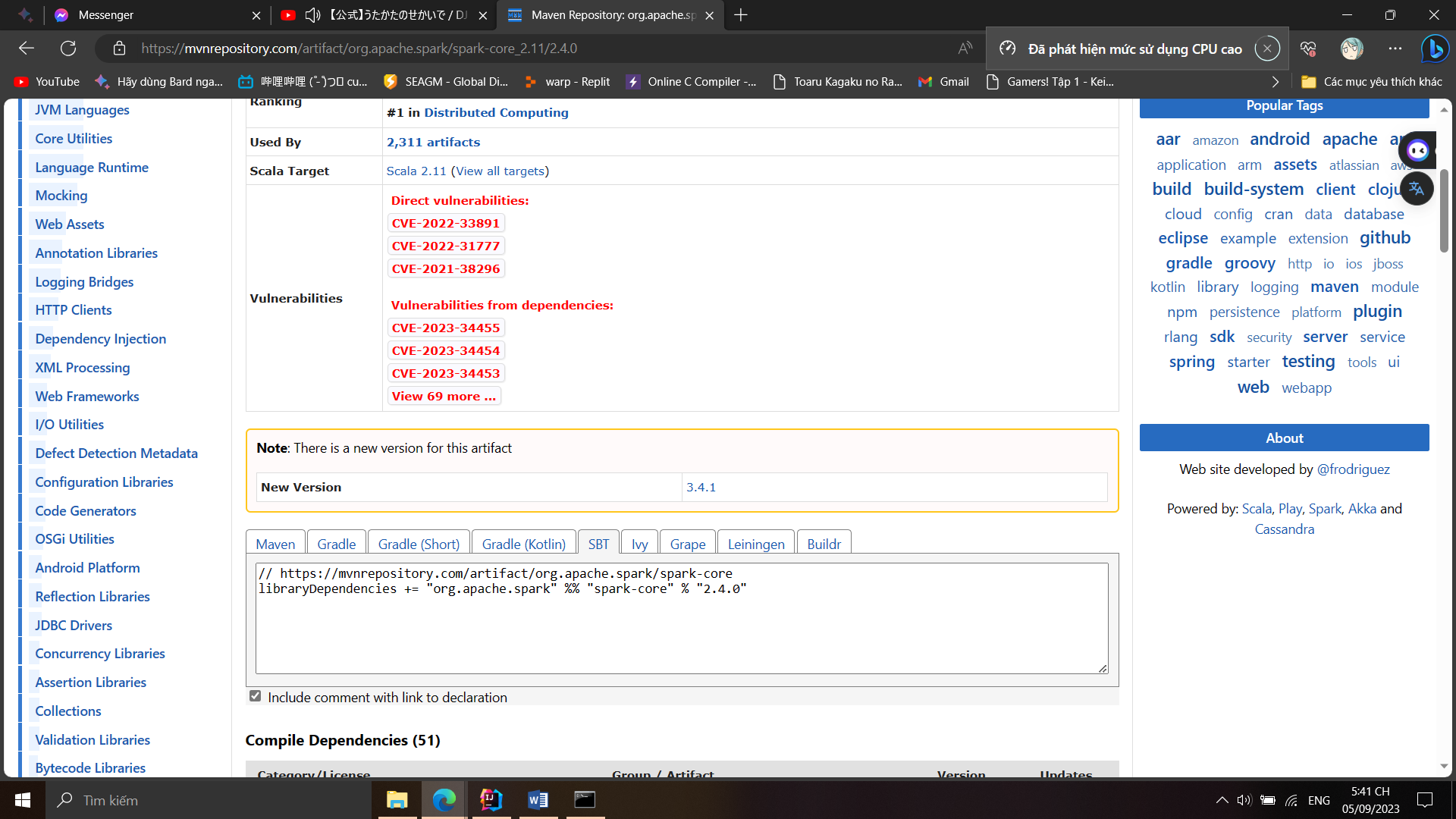
Version : 2.4.0 ( theo cmd được hiển thị trên lap )

Scala :2.11.12 ( theo cmd được hiển thị trên lap )

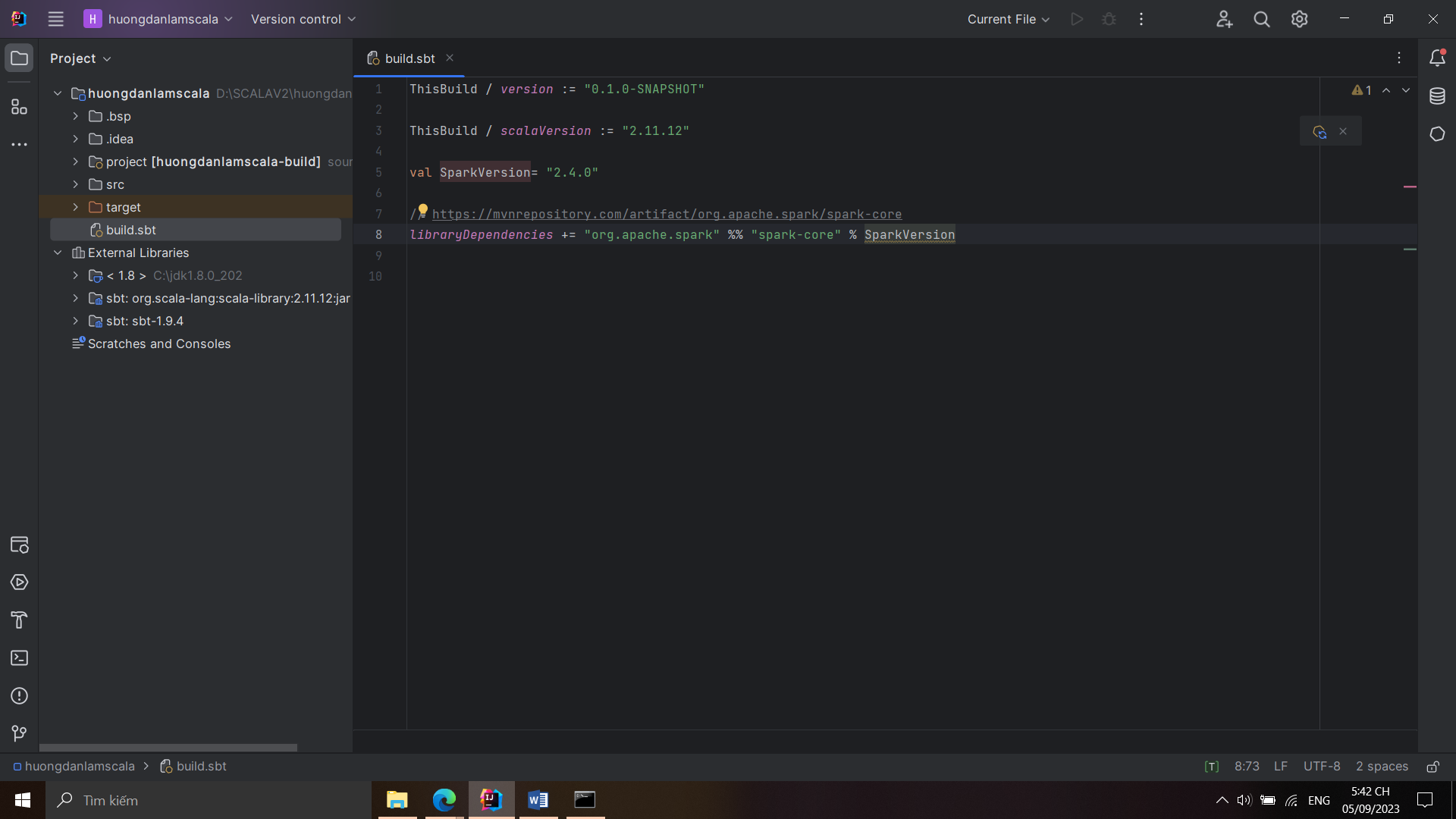
Lựa chọn version tương ứng :



Sau khi chọn hãy vào mục sbt



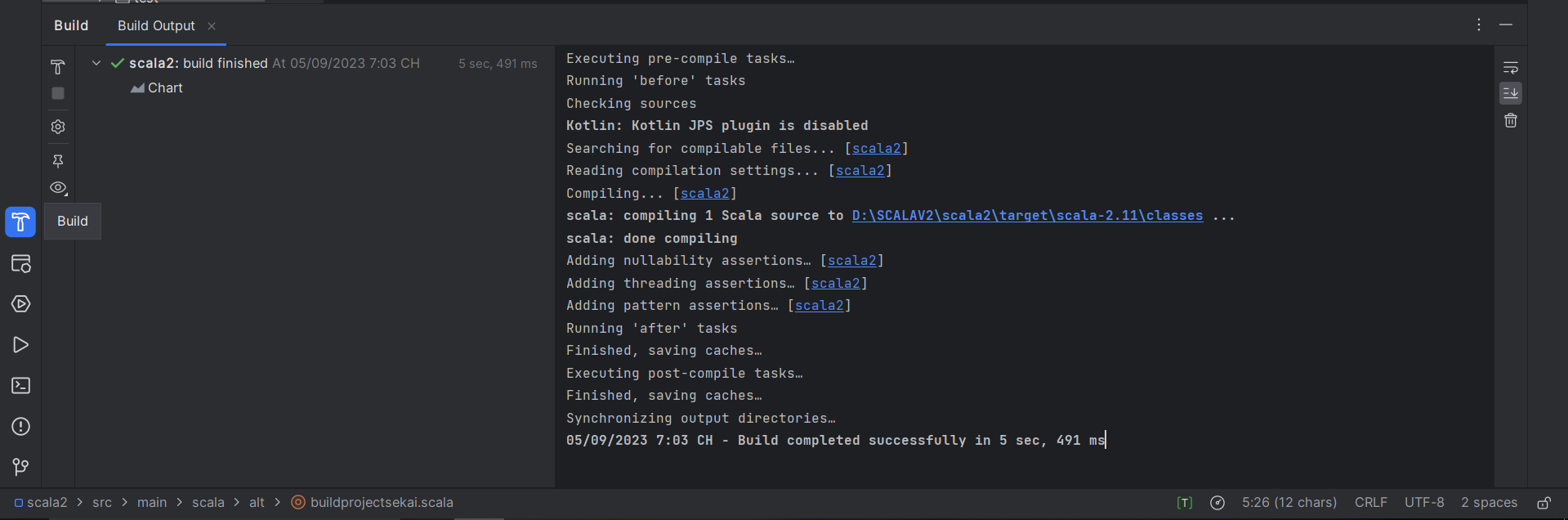
Copy code paste vào file build sbt:



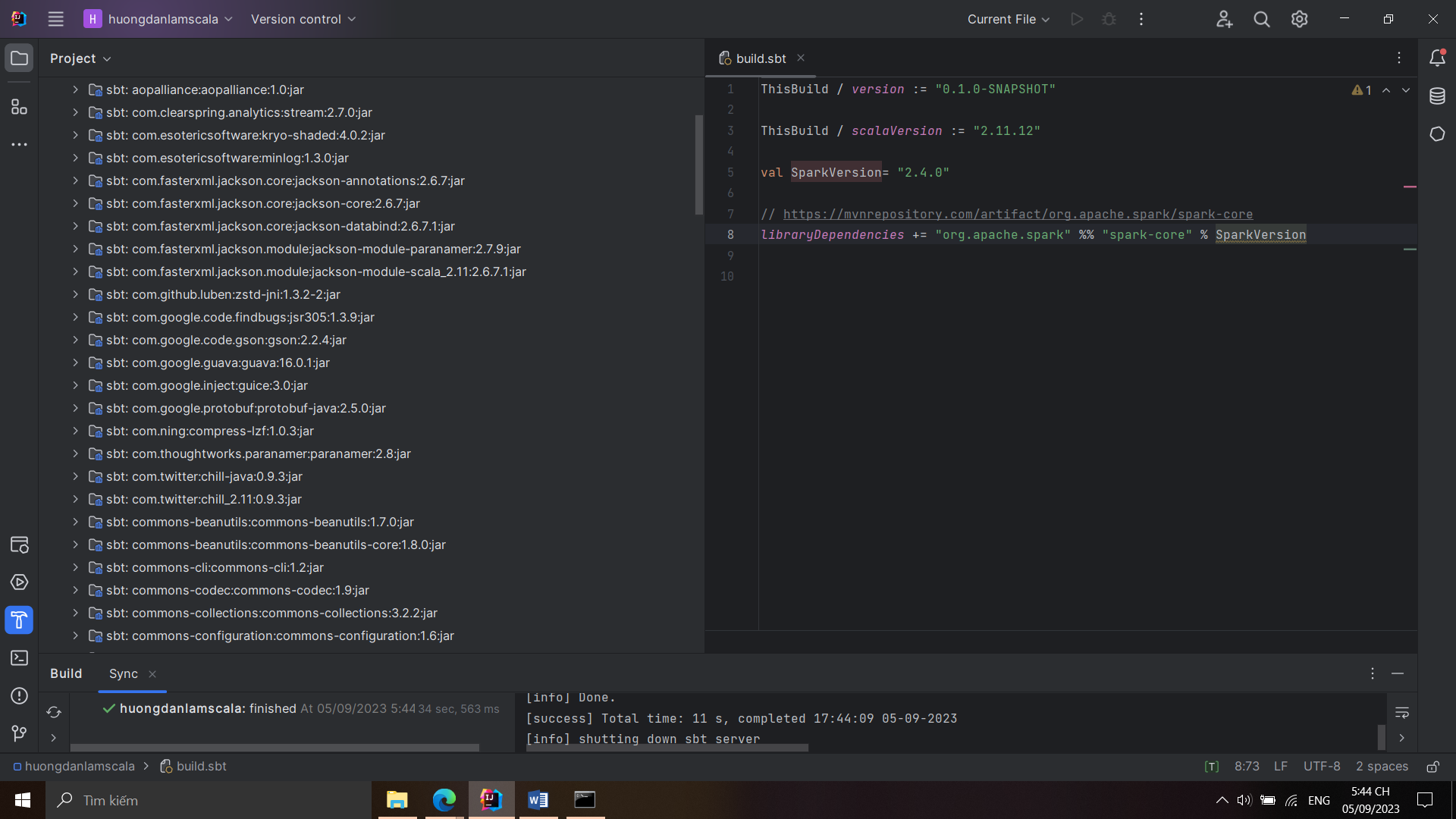
( thay vì ghi 2.4.0 như trên ảnh web thì ghi sparkversion )

Ấn build ( hình cái búa ) và đợi, nếu ko được nó sẽ có nút refresh để chạy lại build:

:logo cái búa là cái đầu tiên

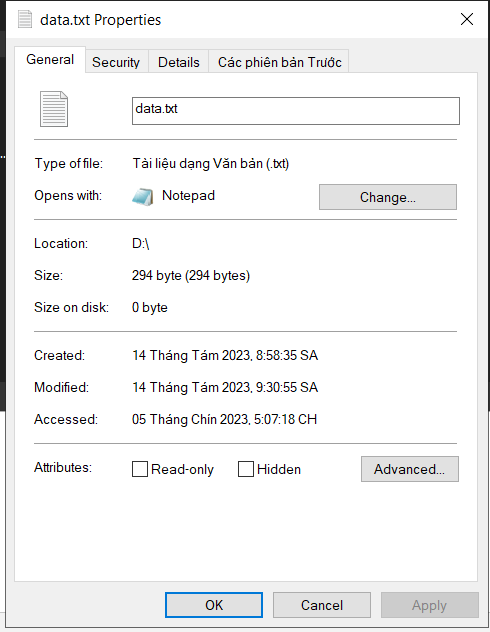


Kết quả :



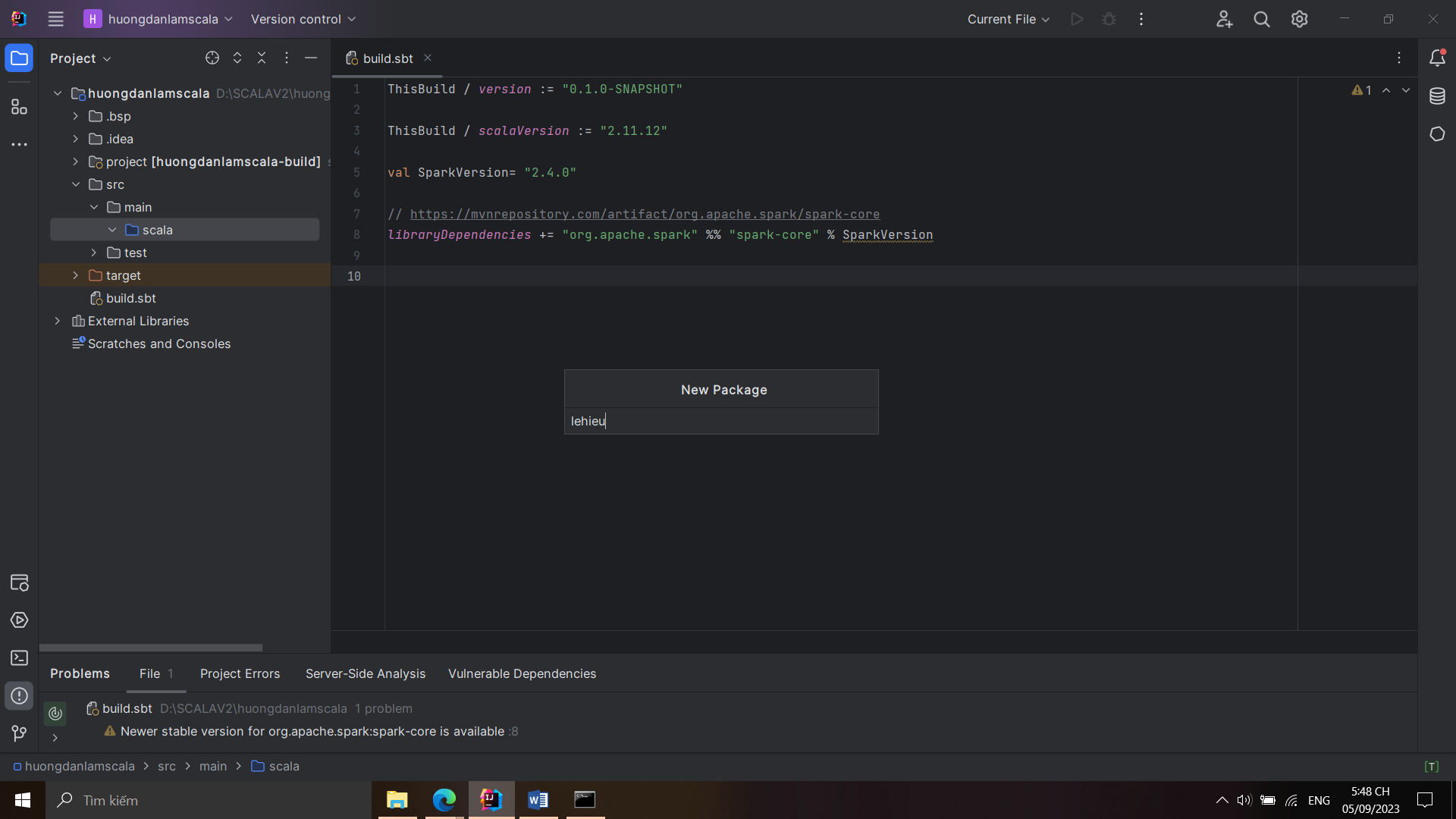
3. Chạy demo:

Chuẩn bị 1 file txt ( mình tạo 1 file tên data.txt trong ổ D ) :

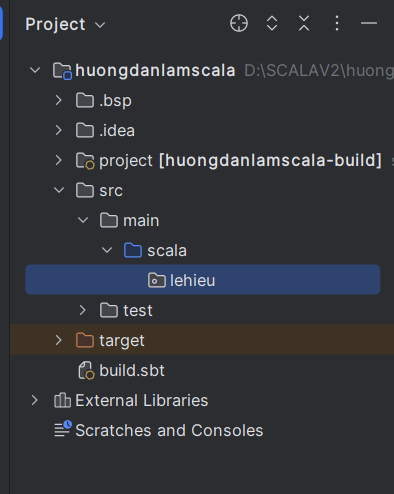


Nội dung ghi gì cũng được

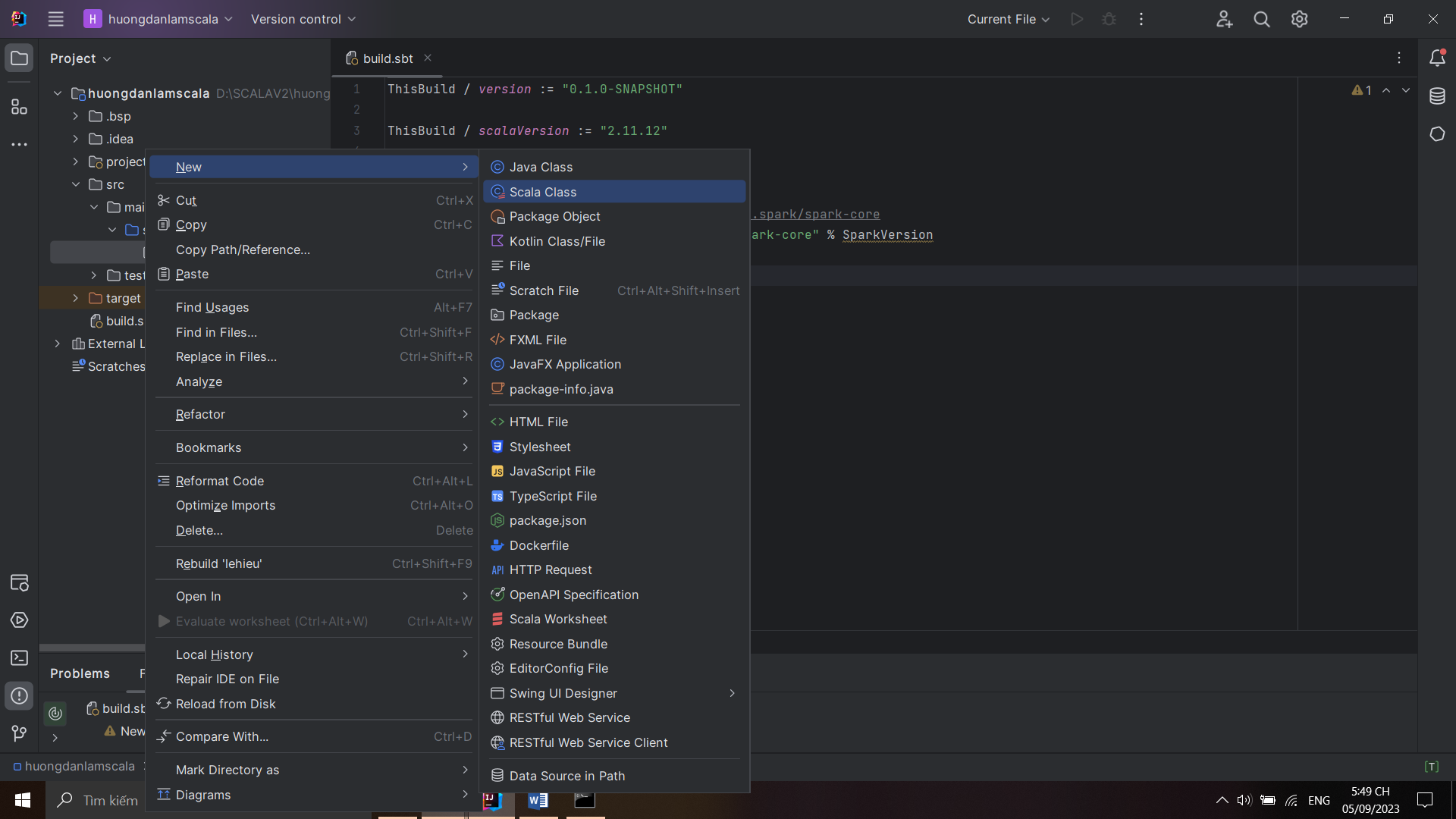
Tạo 1 packgae với tên tùy chọn



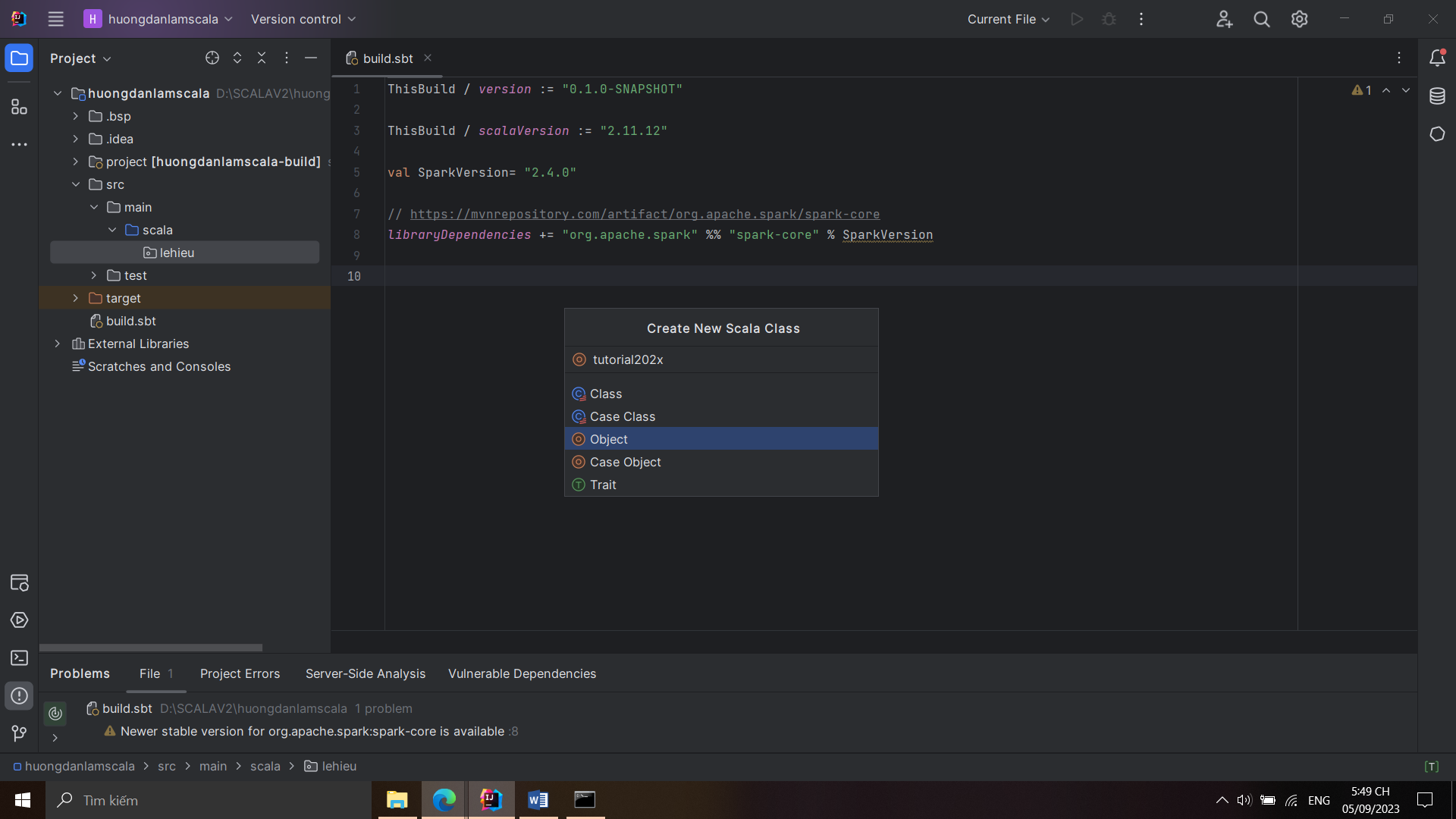
Kết quả:



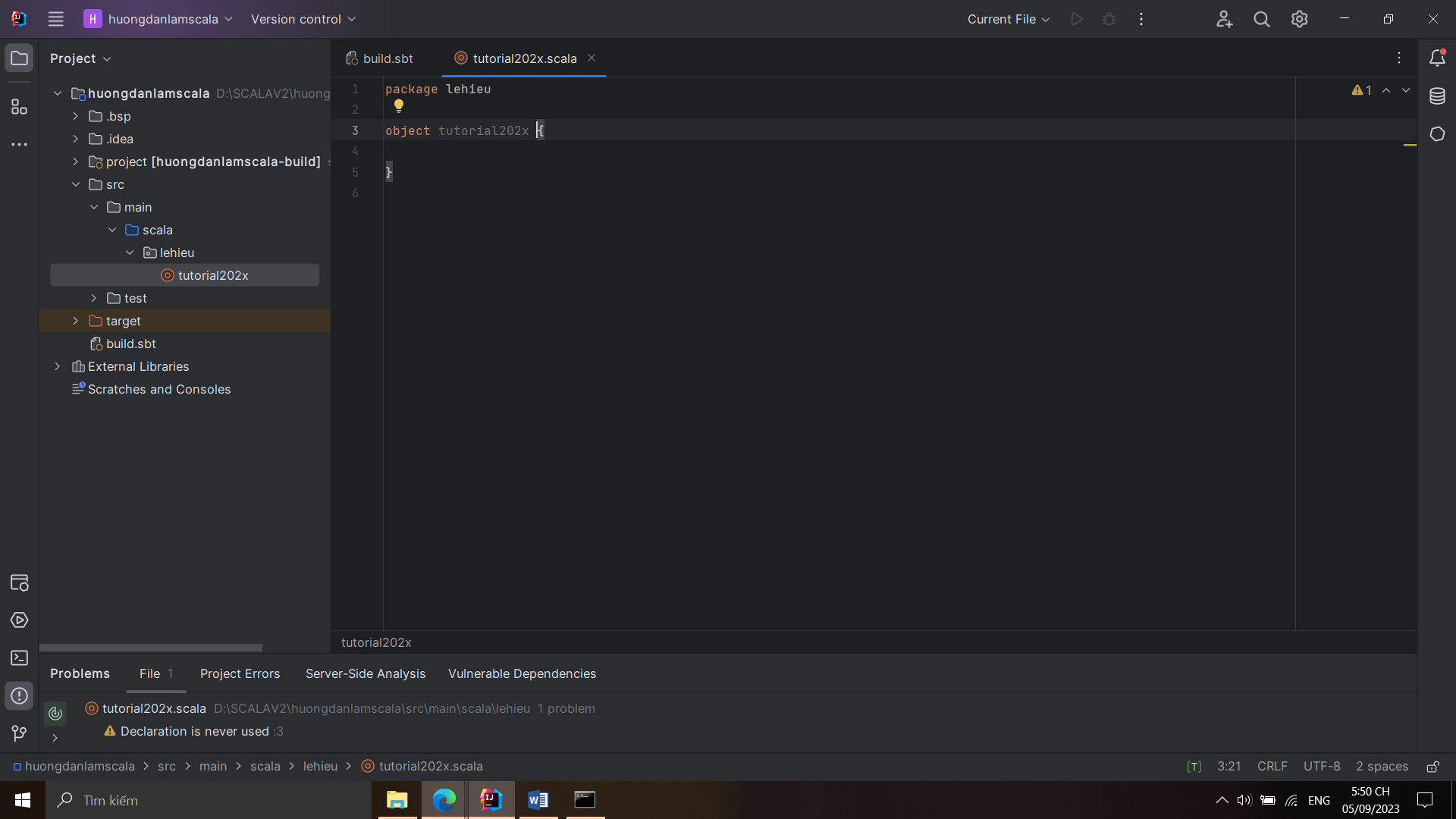
Tạo 1 class mới bằng cách trỏ chuột phải vào packge => new => scala class



Chọn object gõ tên bất kì:



Kết quả



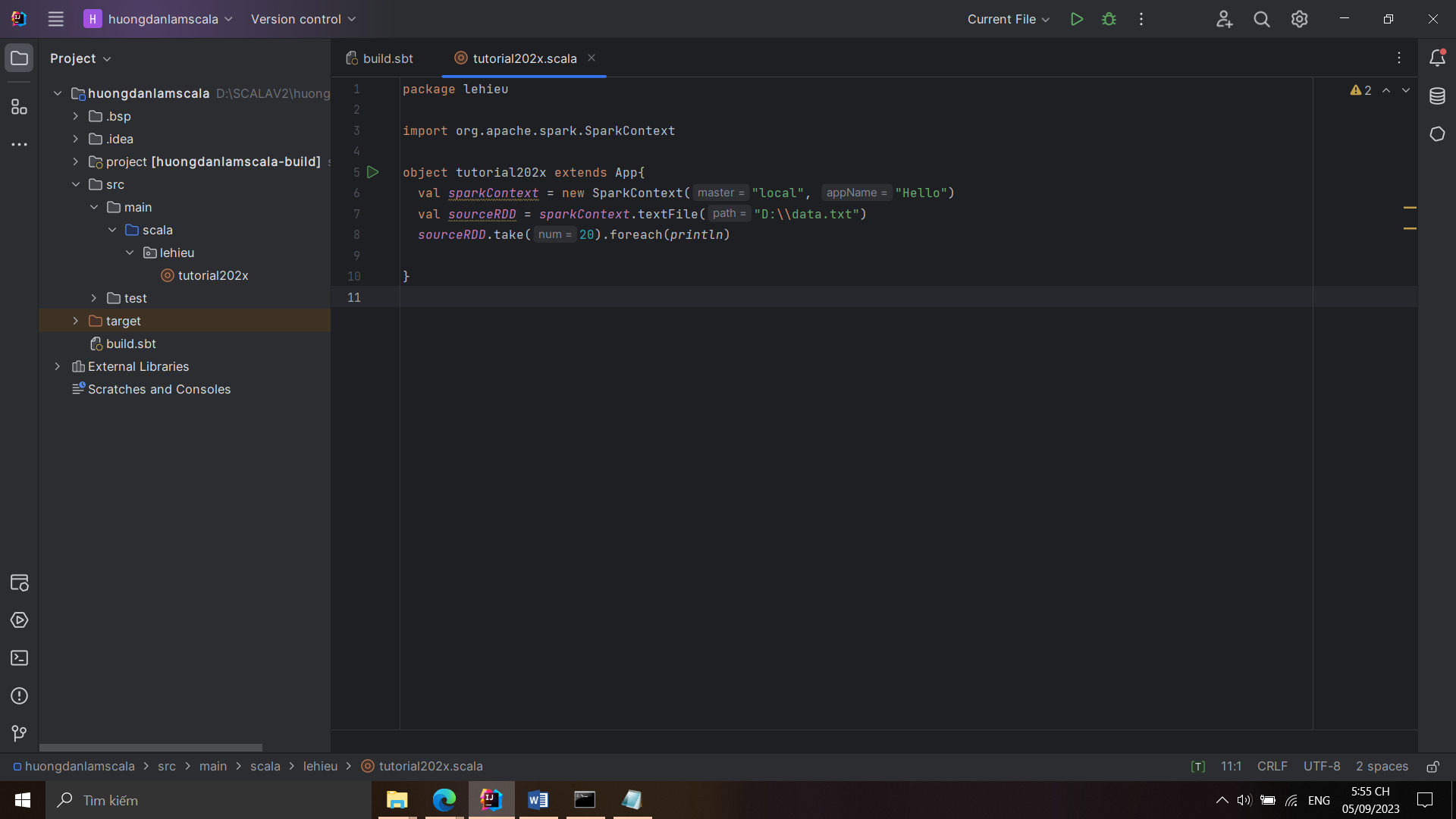
Code :

object tutorial202x extends App {  
 val *sparkContext* = new SparkContext("local", "Hello")  
 val *sourceRDD* = *sparkContext*.textFile("D:\\data.txt")  
 *sourceRDD*.take(10).foreach(*println*)  
  
}

( tùy đường dẫn file để đâu mà dòng source địa chỉ link sẽ khác

Tùy số lượng dòng nó đọc theo mà số 10 kia tùy chọn )

Nhập xong :



Ấn build để chạy : kết quả cuối cùng:

