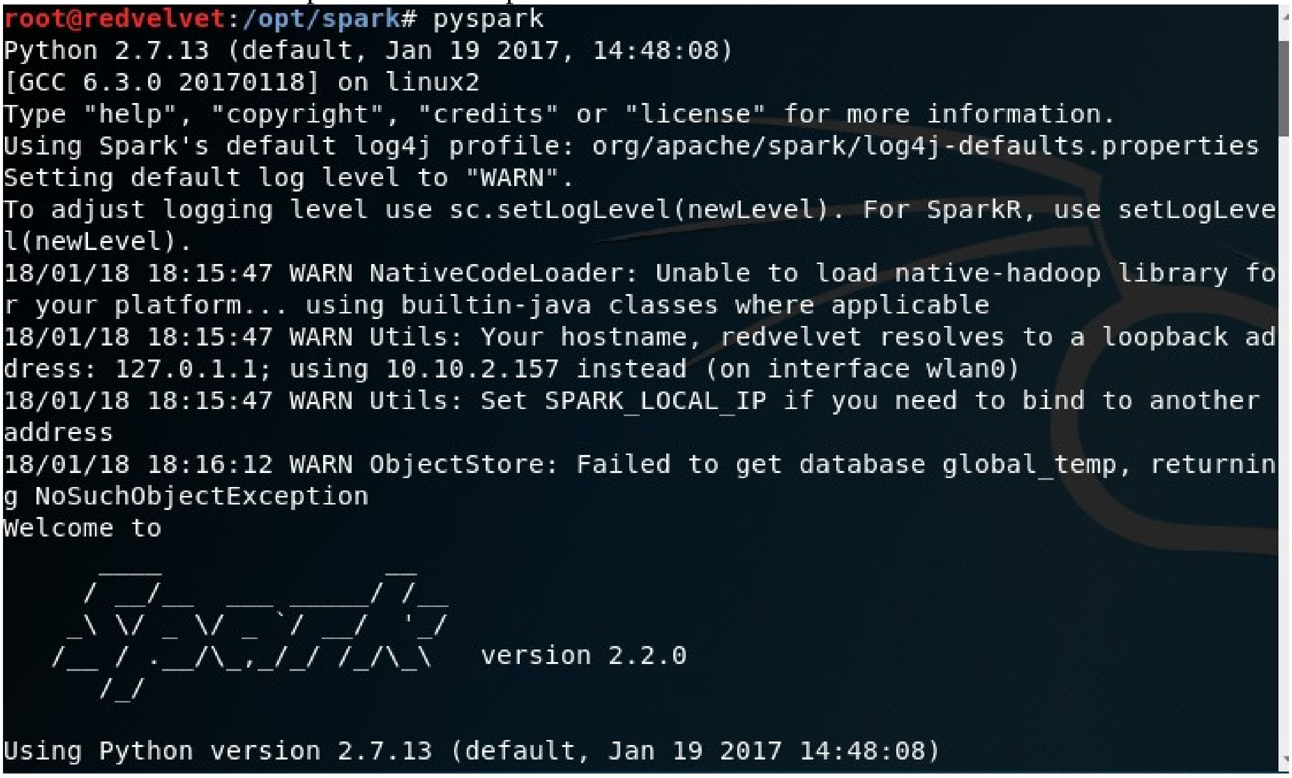
Tutorial Apache Spark Mencoba WordCount dengan map reduce pada Kali Linux Rolling 64 bit dengan PYSPARK

Chessa Pandu Aditirta
155610037
STMIK AKAKOM YOGYAKARTA
2018

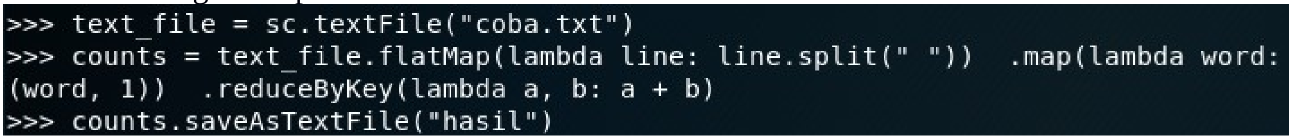
Menjalankan pyspark pada direktori root/opt/spark dimana target file yang akan di hitung contohnya menggunakan coba.txt yang berada di direktori coba.
Pindahkan file tersebut pada direktori spark anda



```
root@redvelvet:/opt/spark# pyspark
Python 2.7.13 (default, Jan 19 2017, 14:48:08)
[GCC 6.3.0 20170118] on linux2
Type "help", "copyright", "credits" or "license" for more information.
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLeve
l(newLevel).
18/01/18 18:15:47 WARN NativeCodeLoader: Unable to load native-hadoop library fo
r your platform... using builtin-java classes where applicable
18/01/18 18:15:47 WARN Utils: Your hostname, redvelvet resolves to a loopback ad
dress: 127.0.1.1; using 10.10.2.157 instead (on interface wlan0)
18/01/18 18:15:47 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another
address
18/01/18 18:16:12 WARN ObjectStore: Failed to get database global_temp, returnin
g NoSuchObjectException
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 2.2.0
      /_/

Using Python version 2.7.13 (default, Jan 19 2017 14:48:08)
```

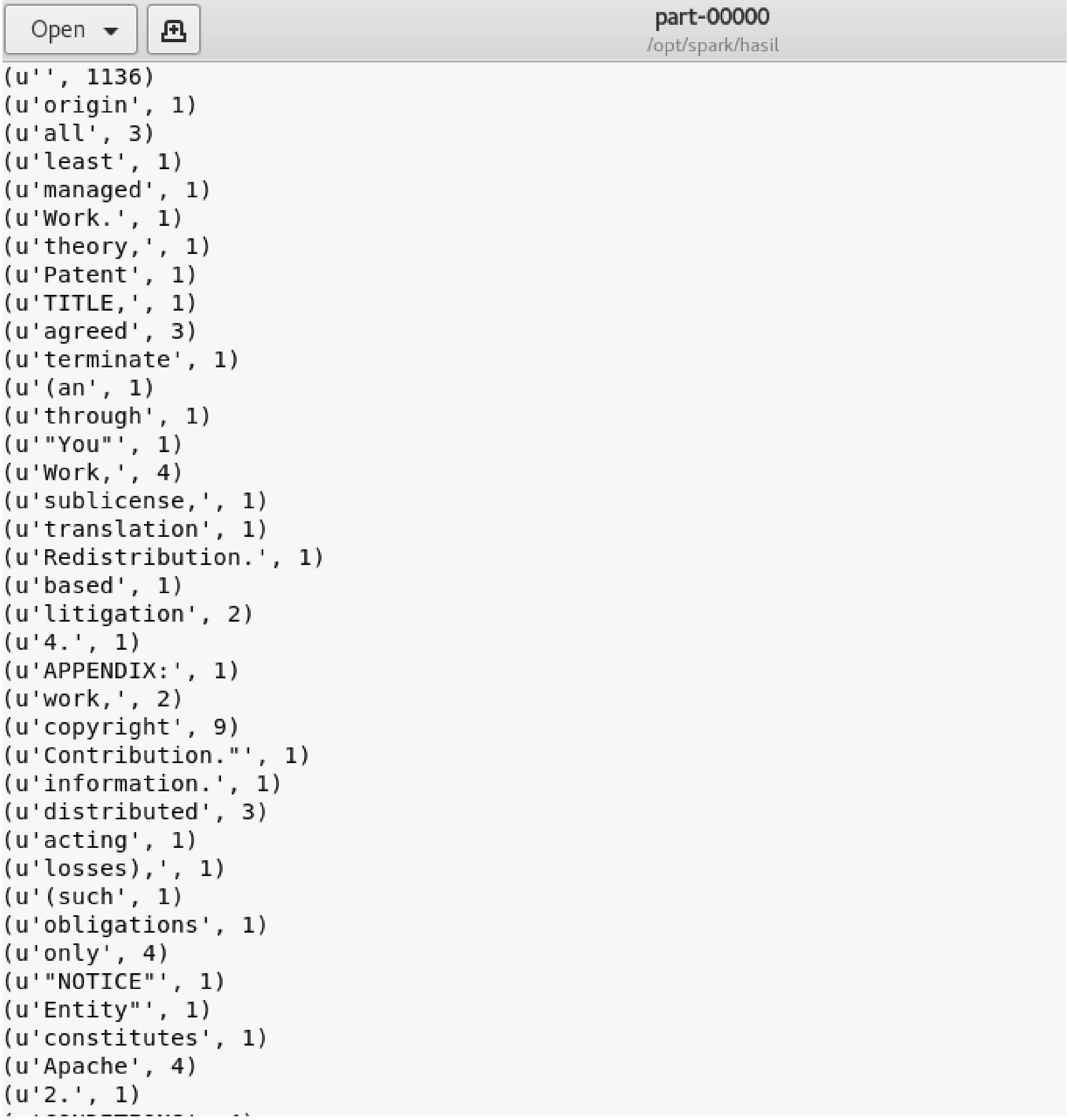
kemudian mengetikan perintah



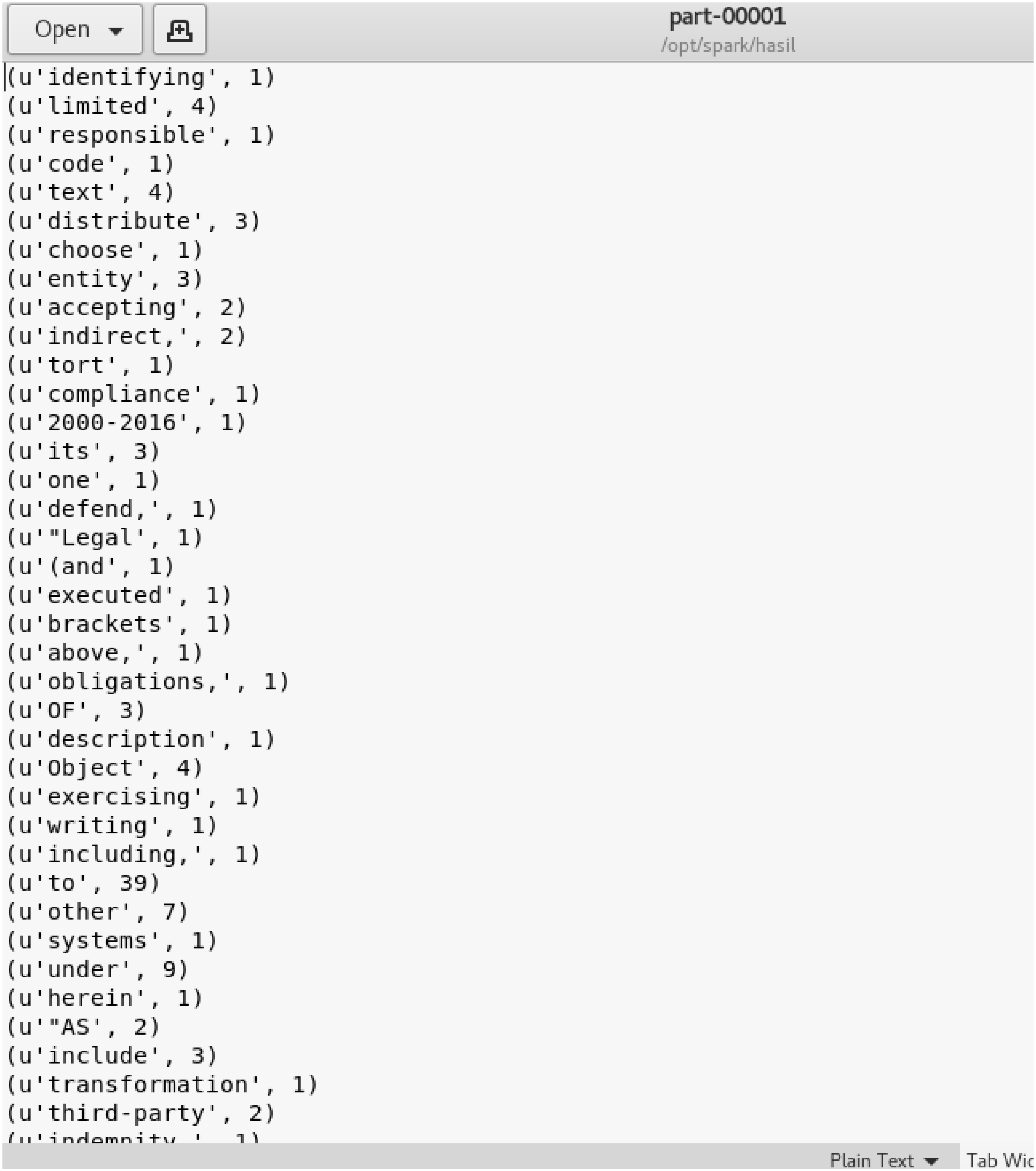
```
>>> text_file = sc.textFile("coba.txt")
>>> counts = text_file.flatMap(lambda line: line.split(" "))  .map(lambda word:
(word, 1))  .reduceByKey(lambda a, b: a + b)
>>> counts.saveAsTextFile("hasil")
```

source file coba.txt berisi lisensi apache yang menjadi target percobaan untuk melakukan wordcount spark python.

Perintah counts membuat setiap kata dipisah dan dihitung tanpa memperdulikan seberapa panjang sepasi dan pada counts.saveTextFile(“hasil”) membuat hasil perhitungan wordcount tersaji dalam beberapa file berikut lapiran filenya



```
(u'', 1136)
(u'origin', 1)
(u'all', 3)
(u'least', 1)
(u'managed', 1)
(u'Work.', 1)
(u'theory,', 1)
(u'Patent', 1)
(u'TITLE,', 1)
(u'agreed', 3)
(u'terminate', 1)
(u'(an', 1)
(u'through', 1)
(u'"You"', 1)
(u'Work,', 4)
(u'sublicense,', 1)
(u'translation', 1)
(u'Redistribution.', 1)
(u'based', 1)
(u'litigation', 2)
(u'4.', 1)
(u'APPENDIX:', 1)
(u'work,', 2)
(u'copyright', 9)
(u'Contribution."', 1)
(u'information.', 1)
(u'distributed', 3)
(u'acting', 1)
(u'losses),', 1)
(u'(such', 1)
(u'obligations', 1)
(u'only', 4)
(u'"NOTICE"', 1)
(u'Entity"', 1)
(u'constitutes', 1)
(u'Apache', 4)
(u'2.', 1)
```



```
(u'identifying', 1)
(u'limited', 4)
(u'responsible', 1)
(u'code', 1)
(u'text', 4)
(u'distribute', 3)
(u'choose', 1)
(u'entity', 3)
(u'accepting', 2)
(u'indirect,', 2)
(u'tort', 1)
(u'compliance', 1)
(u'2000-2016', 1)
(u'its', 3)
(u'one', 1)
(u'defend,', 1)
(u'"Legal', 1)
(u'(and', 1)
(u'executed', 1)
(u'brackets', 1)
(u'above,', 1)
(u'obligations,', 1)
(u'OF', 3)
(u'description', 1)
(u'Object', 4)
(u'exercising', 1)
(u'writing', 1)
(u'including,', 1)
(u'to', 39)
(u'other', 7)
(u'systems', 1)
(u'under', 9)
(u'herein', 1)
(u'"AS', 2)
(u'include', 3)
(u'transformation', 1)
(u'third-party', 2)
(u'indemnity,', 1)
```