

# Tutorial memasang Apache Spark dan Konfigurasinya pada Kali Linux Rolling 64 bit

Chessa Pandu Aditirta  
155610037  
STMIK AKAKOM YOGYAKARTA  
2018

## Install Scala

Membuat direktori scala di dalam direktori opt (langkah ini optional, membuat di usr/local juga boleh)

```
root@redvelvet:~# mkdir /opt/scala
```

direktori scala telah terbuat

```
root@redvelvet:~# ls /opt
Teeth          libreoffice5.4
cobaGradle     projectone
findbugs       scala
findbugs-3.0.1.tar.gz spark
genymobile     spark-workcount
gradle         virtualbox-5.2_5.2.0-118431-Debian~jessie_amd64.deb
hadoop
```

mengarahkan ke direktori scala

```
root@redvelvet:~# cd /opt/scala/
root@redvelvet:/opt/scala#
```

Mengunduh Scala dengan menambahkan wget di direktori tersebut atau unduh pada repository saya

```
root@redvelvet:~# wget http://downloads.lightbend.com/scala/2.12.1/scala-2.12.1.deb
```

Melakukan ekstrak pada file yang telah di download

```
root@redvelvet:/opt/scala# dpkg -i scala-2.12.1.deb
```

NB:

setelah dilakukan ekstrak pindah semua file yang ada di dalam direktori scala-2.12.1 ke direktori scala.

## Konfigurasi Scala

Menambahkan Environment scala

```
root@redvelvet:~# export SCALA_HOME=/opt/scala
root@redvelvet:~# export PATH=$PATH:$SCALA_HOME/bin
```

Mengecek Instalasi scala dengan perintah scala -version

```
root@redvelvet:~# scala -version
Scala code runner version 2.12.1 -- Copyright 2002-2016, LAMP/EPFL and Lightbend, Inc.
```

## Install apache Spark

Mengunduh file dengan perintah berikut atau dapat langsung mengunduh file tersebut pada repository saya

```
root@redvelvet:~# wget http://d3kbcqa49mib13.cloudfront.net/spark-2.0.2-bin-hadoop2.7.tgz
```

melakukan ekstrak file tersebut dan memindahkannya pada direktori opt/spark.

```
root@redvelvet:~# tar -xvf spark-2.0.2-bin-hadoop2.7.tgz
```

```
root@redvelvet:~# cp -rv spark-2.0.2-bin-hadoop2.7/* /opt/spark
```

NB: sama seperti scala sebelumnya perlu memindahkan semua yang ada pada direktori spark-2.0.2-bin-hadoop2.7 ke direktori spark

## Konfigurasi Apache Spark

Menambahkan envirointment seperti berikut :

```
root@redvelvet:~# export SPARK_HOME=/opt/spark
```

```
root@redvelvet:~# export PATH=$PATH:$SPARK_HOME/bin
```

## Menjalankan Spark shell

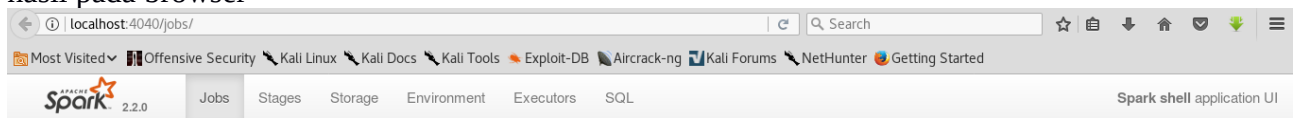
```
root@redvelvet:~# spark-shell
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
18/01/18 16:33:51 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/01/18 16:33:51 WARN Utils: Your hostname, redvelvet resolves to a loopback address: 127.0.1.1; using 10.10.27.67 instead (on interface wlan0)
18/01/18 16:33:51 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
18/01/18 16:34:11 WARN ObjectStore: Failed to get database global_temp, returning NoSuchObjectException
Spark context Web UI available at http://10.10.27.67:4040
Spark context available as 'sc' (master = local[*], app id = local-1516268035481).
Spark session available as 'spark'.
Welcome to

    ____
   / __ \_____/ _ \/ ___/ 
  / / / //___/\_ \/_/___/  
 /_/ /_//___/\_\/___/_____/_/ version 2.2.0

Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_151)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

hasil pada browser



## Spark Jobs (?)

User: root  
Total Uptime: 1.3 min  
Scheduling Mode: FIFO  
[▶ Event Timeline](#)

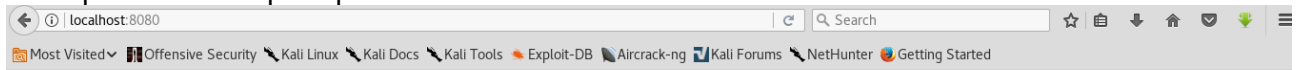
## Percobaan

## Menjalankan StandAlone Cluster

```
root@redvelvet:/opt/spark# sbin/start-all.sh
starting org.apache.spark.deploy.master.Master, logging to /opt/spark/logs/spark-
root-org.apache.spark.deploy.master.Master-1-redvelvet.out
localhost: ssh: connect to host localhost port 22: Connection refused
```

NB: untuk mematikan standalone cluster tersebut cukup berikan perintah `sbin/stop-all.sh`

## Hasil pada Browser pada port 8080



### Spark Master at spark://redvelvet.misterc:7077

URL: spark://redvelvet.misterc:7077

REST URL: spark://redvelvet.misterc:6066 *(cluster mode)*

Alive Workers: 0

Cores in use: 0 Total, 0 Used

Memory in use: 0.0 B Total, 0.0 B Used

Applications: 0 [Running](#), 0 [Completed](#)

Drivers: 0 Running, 0 Completed

Status: ALIVE

#### Workers

Worker Id	Address	State	Cores	Memory
-----------	---------	-------	-------	--------

#### Running Applications

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	----------------	------	-------	----------

#### Completed Applications

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	----------------	------	-------	----------