



Session : SPARK STREAMING

Assignment

Student Name: Subham Vishal

Course: Big Data Hadoop & Spark Training

Assignment – Introduction to Spark Streaming.

Contents

Introduction	1
Problem Statement	1
Read Two Streams – Spark Streaming.....	2
Step – 1 - Import all the spark streaming packages	2

Introduction

In this assignment, we are going to read 2 streams. One from the user and we have real-time set of offensive words.

Problem Statement

Read two streams

- List of strings input by user
- Real-time set of offensive words

Find the word count of the offensive words inputted by the user as per the real-time set of offensive words.

Pre requisite

In this assignment we are going to read the strings which is captured in a port using **netcat** and hence we are installing the **netcat** in our server,



Command,

```
 sudo yum install nc
```

The below screen shot shows the successful installation of **netcat** using **yum**.

```
[acadgild@localhost ~]$ sudo yum install nc
Loaded plugins: fastestmirror, refresh-packagekit, security
Setting up Install Process
Loading mirror speeds from cached hostfile
 * base: ftp.iitm.ac.in
 * extras: ftp.iitm.ac.in
 * updates: ftp.iitm.ac.in
Resolving Dependencies
--> Running transaction check
--> Package nc.i686 0:1.84-24.el6 will be installed
--> Finished Dependency Resolution

Dependencies Resolved

=====================================================================================================================================
 Package                               Arch                               Version                               Repository                           Size
=====================================================================================================================================
Installing:
 nc                                   i686                               1.84-24.el6                           base                                  57 k
Transaction Summary
-----
Install      1 Package(s)


Total download size: 57 k
Installed size: 107 k
Is this ok [y/N]: y
Downloading Packages:
nc-1.84-24.el6.i686.rpm                               | 57 kB    00:00
Running rpm_check_debug
Running Transaction Test
Transaction Test Succeeded
Running Transaction
  Installing : nc-1.84-24.el6.i686                               1/1
  Verifying  : nc-1.84-24.el6.i686                               1/1

Installed:
 nc.i686 0:1.84-24.el6

Complete!
[acadgild@localhost ~]$ nc -lk 9999
```

Start listening the port **9999** using the below command,

Command,

```
 nc -lk 9999
```

```
[acadgild@localhost ~]$ nc -lk 9999
HT
```




Start Spark shell with multi threads, in this case we are taking 4 multi threads.

Command,

```
 /home/acadgild/spark-2.2.1-bin-hadoop2.7/bin/spark-shell --master local[4]
```

Read Two Streams – Spark Streaming

Step – 1 - Import all the spark streaming packages

```
 import org.apache.spark._
 import org.apache.spark.streaming._
 import org.apache.spark.streaming.StreamingContext._
```



```
scala> import org.apache.spark._  
import org.apache.spark._  
  
scala> import org.apache.spark.streaming._  
import org.apache.spark.streaming._  
  
scala> import org.apache.spark.streaming.StreamingContext._  
import org.apache.spark.streaming.StreamingContext._
```