

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	

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

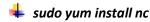
Find the word count of the offensive words inputted by the user as per the real-time set of offensive words.

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,

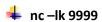


The below screen shot shows the successful installation of **netcat** using **yum**.

```
cadgild@localhost ~1$ sudo yum install nc
aded plugins: fastestmirror, refresh-packagekit, security
tting up Install Process
ading mirror speeds from cached hostfile
base: ftp.iitm.ac.in
extras: ftp.iitm.ac.in
updates: ftp.iitm.ac.in
solving Dependencies
> Running transaction check
> Package nc.i686 0:1.84-24.el6 will be installed
> Finished Dependency Resolution
    pendencies Resolved
 Package
                                                                                                                                                                               Version
                                                                                                                                                                                                                                                                                                     Repository
 installing:
                                                                                      i686
                                                                                                                                                                               1.84-24.el6
                                                                                                                                                                                                                                                                                                                                                                                                   57 k
 ransaction Summary
 nstall
                                    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
Running Transaction Test
                                                                                                                                                                                                                                                                                                                                                 | 57 kB
                                                                                                                                                                                                                                                                                                                                                                                    00:00
     mining fransaction lest
ansaction Test Succeeded
nning Transaction
Installing : nc-1.84-24.el6.i686
Verifying : nc-1.84-24.el6.i686
 installed:
nc.i686 0:1.84-24.el6
 Complete!
acadgild@localhost ~]$ nc -lk 9999
```

Start listening the port 9999 using the below command,

Command,



~C [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._
```