**Session 20: SPARK STREAMING**

Assignment 20.2

Student Name: Abarajithan SA

Course: Big Data Hadoop & Spark Training

Start Date:  2017-09-09

End Date:  2017-11-26

**Assignment 20.2**– Introduction to Spark Streaming.

Contents

[Introduction 2](#_Toc503949306)

[Problem Statement 2](#_Toc503949307)

[Read Two Streams – Spark Streaming 2](#_Toc503949308)

[Step – 1 - Import all the spark streaming packages 2](#_Toc503949309)

# 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.

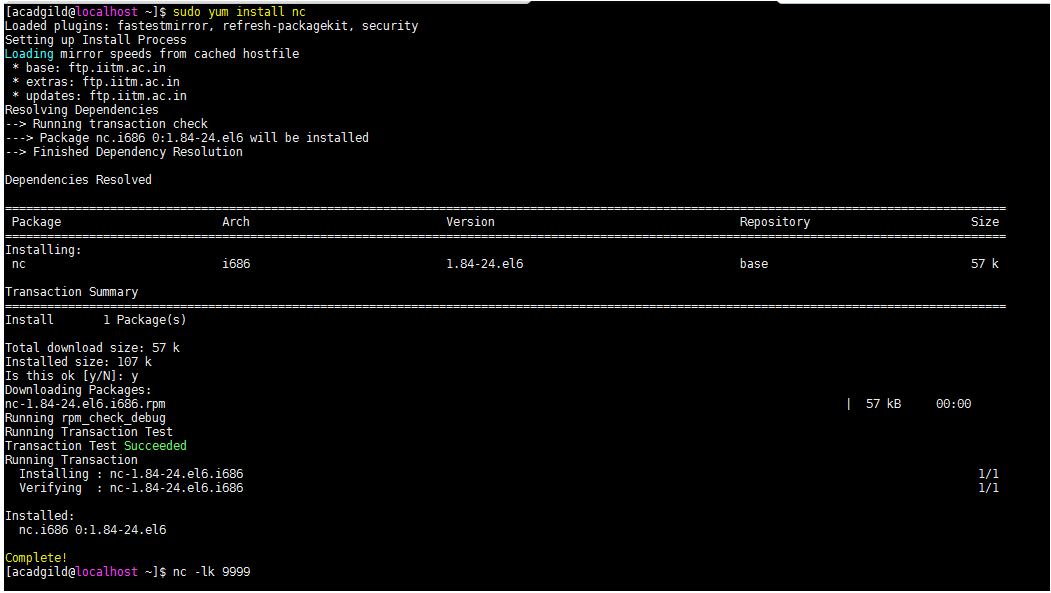
# Pre requisite

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.***



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

Command,

* **nc –lk 9999**



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.\_***

