

# Big Data



## Big Data Engineering with Hadoop & Spark

Kafka Introduction

**Spark**  
**Streaming**



## Session 24: Assignment 24.1

---

This assignment is aimed at consolidating the concepts that was learnt during the Apache Spark session on Spark Streaming of the course.

# Task 1:

Read a stream of Strings, fetch the words which can be converted to numbers. Filter out the rows, where the sum of numbers in that line is odd. Provide the sum of all the remaining numbers in that batch.

## Solution:

**Note:** Source code file is provided along with this assignment report.

## Output:

- Before running the application, start “netcat” in terminal using the command “nc -lk 9999”
- Run the spark application
- Put some string with numbers on the shell as shown below

```
[acadgild@localhost ~]$ nc -lk 9999
Hello234 Everyone24
How35 are27 you23
manage45 all56
do24 like45 this22
```

- In screenshot below, we are able to see that lines containing sum of odd numbers are filter, even number is displayed, and sum of the number is displayed in the next line.

```
<terminated> OddStringSum$ [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Oct 9, 2018, 7:48:27 PM)
Time: 1539094740000 ms
-----
18/10/09 19:49:14 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/10/09 19:49:14 WARN BlockManager: Block input-0-1539094754000 replicated to only 0 peer(s) instead of 1 peers
18/10/09 19:49:15 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/10/09 19:49:15 WARN BlockManager: Block input-0-1539094755000 replicated to only 0 peer(s) instead of 1 peers
Time: 1539094760000 ms
-----
How35 are27 you23
manage45 all56
Time: 1539094760000 ms
42.0
Time: 1539094780000 ms
Time: 1539094780000 ms
```

## Task 2:

Read two streams

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

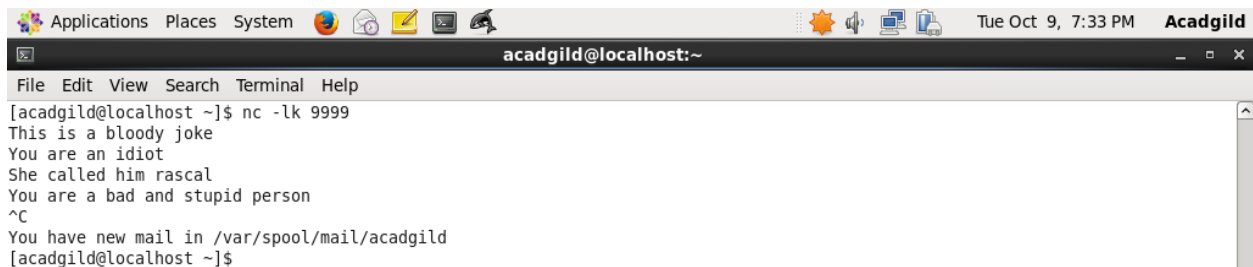
Find the word count of the offensive words inputted by the user as per the real-time set of offensive words

### **Solution:**

**Note:** Source code file is provided along with this assignment report.

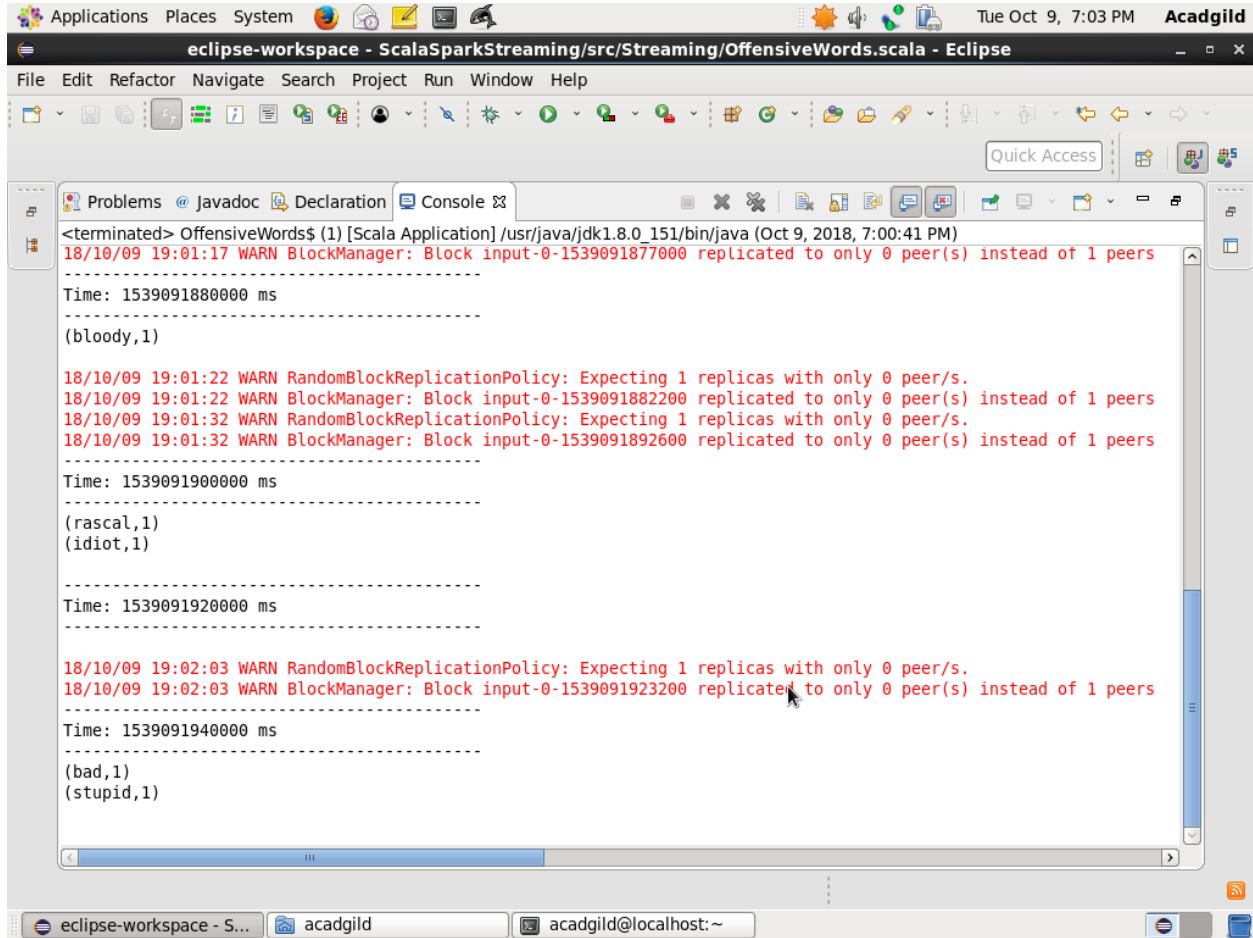
### **Output:**

- In the spark application, we have a set of words that we considered as offensive words *"idiot", "fool", "bad", "nonsense", "shit", "damn", "stupid", "dash", "bloody", "rascal"*
- Before running the application, start "netcat" in terminal using the command `"nc -lk 9999"`
- Run the spark application
- Put some string with numbers on the shell as shown below



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ nc -lk 9999  
This is a bloody joke  
You are an idiot  
She called him rascal  
You are a bad and stupid person  
^C  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$
```

- In screenshot below, we are able to see that spark application has counted the number of offensive words occur as per the input string provided by the user.



The screenshot shows the Eclipse IDE interface with the console window open. The console displays the following output:

```
<terminated> OffensiveWords$ (1) [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Oct 9, 2018, 7:00:41 PM)
18/10/09 19:01:17 WARN BlockManager: Block input-0-1539091877000 replicated to only 0 peer(s) instead of 1 peers
-----
Time: 1539091880000 ms
-----
(bloody,1)

18/10/09 19:01:22 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/10/09 19:01:22 WARN BlockManager: Block input-0-1539091882200 replicated to only 0 peer(s) instead of 1 peers
18/10/09 19:01:32 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/10/09 19:01:32 WARN BlockManager: Block input-0-1539091892600 replicated to only 0 peer(s) instead of 1 peers
-----
Time: 1539091900000 ms
-----
(rascal,1)
(idiot,1)

-----
Time: 1539091920000 ms
-----

18/10/09 19:02:03 WARN RandomBlockReplicationPolicy: Expecting 1 replicas with only 0 peer/s.
18/10/09 19:02:03 WARN BlockManager: Block input-0-1539091923200 replicated to only 0 peer(s) instead of 1 peers
-----
Time: 1539091940000 ms
-----
(bad,1)
(stupid,1)
```

The bottom of the IDE shows the status bar with the text "eclipse-workspace - S...", "acadgild", and "acadgild@localhost:~".