Lab: Spark Streaming with Windows

About This Lab

Objective:	Create a Spark Streaming utilizing a `window` function to find words read in the previous 10 seconds
File locations:	No files
Successful outcome:	Developer will use the `window` function to create a windowed wordcount.
Before you begin	You should be logged in to your ssh

Lab Steps

Perform the following steps:

1. Close the REPL
2. Start a new REPL specifying the following information:

```
#pyspark --master local[2]
```

3. Create a Spark Streaming application that performs a wordcount on a socket text stream using the window function `reduceByKeyAndWindow`. Set a 10 second window with a 2 second sliding duration
    a. Import the Streaming library:

```
>>>from pyspark.streaming import StreamingContext
```

    b. Create the streaming context, with a 2 second batch duration:

```
>>>ssc = StreamingContext(sc, 2)
```

    c. Create the `Dstream` using `sandbox` and port 9999

```
>>>inputDS = ssc.socketTextStream("sandbox",9999)
```

    d. For this lab, enable checkpointing the lazy way:

```
>>>ssc.checkpoint("hdfs:///user/root/checkpointDir")
```

    e. Transform the `inputDS` to use a `window` and then a `reducebykey`:

```
>>>windowDS = inputDS.window(10,2).flatMap(lambda line:
 line.split(" ")).map(lambda word: \
 (word,1)).reduceByKey(lambda a,b: a+b)
```

    f. Print the output out:

```
>>>windowDS.pprint()
```

    g. To avoid cluttering the output, set the `loglevel` to `ERROR`:

```
>>>sc.setLogLevel("ERROR")
```

h. Start the streaming application:

```
>>>ssc.start()
```

4. In a new terminal, run the following command to start outputting to the stream:

```
#nc -lkv 9999
```

a. Start typing words separated by space, press return occasionally to submit them.

b. Look at the other terminal where the streaming application is running.

c. While the application is running, navigate to the web UI in Firefox and explore the web UI tabs:

```
sandbox:4040
```

d. To quit the streaming application, press `control-d`, `control-c` for the terminal running NC.

RESULT

You have now successfully created an application that utilizes the `window` function.