**CS5543 Real-Time Big Data Analytics**

**LAB ASSIGNMENT #2**

**Akhilesh Gattu**

1. **Question**

**Spark Programming:**

**Write a spark program with an interesting use case using text data as the input and program should have at least Two Spark Transformations and Two Spark Actions.**

**Transformations:**

**Map:**

Map function returns a new distributed dataset where each element is given as input through a function.

**SortByKey:**

SortByKey function pairs up the key value pairs where in K is ordered. This transformation returns the key value pairs in ascending or descending order as specified.

**ReduceByKey**

This transformation returns key value pairs where all the values for each key are reduced. Number of reduce functions could be specified in the augment.

**Actions:**

**Take(n):**

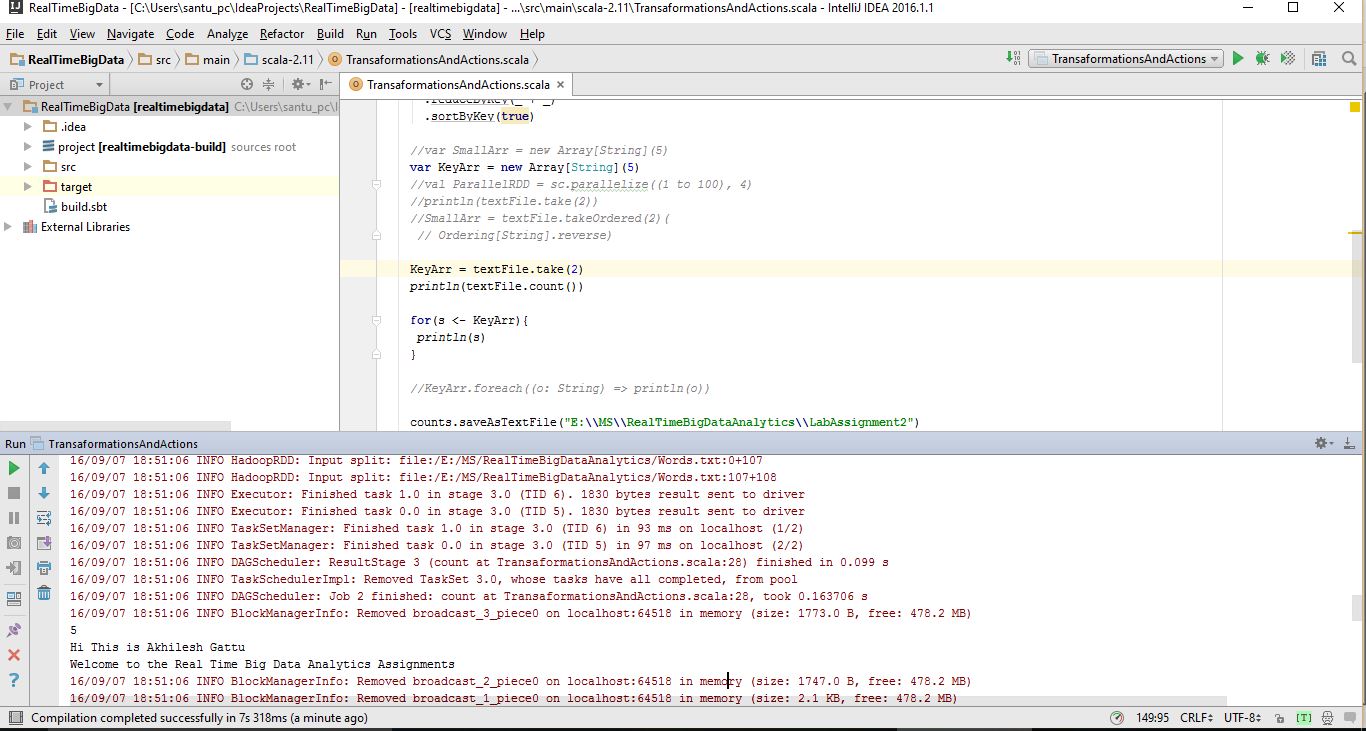
The Take action in spark returns the array with first n elements in the dataset.

**Count():**

The count action in spark returns the number of elements in the dataset.

**Screenshots:**

The actions used are count and take



The input given is

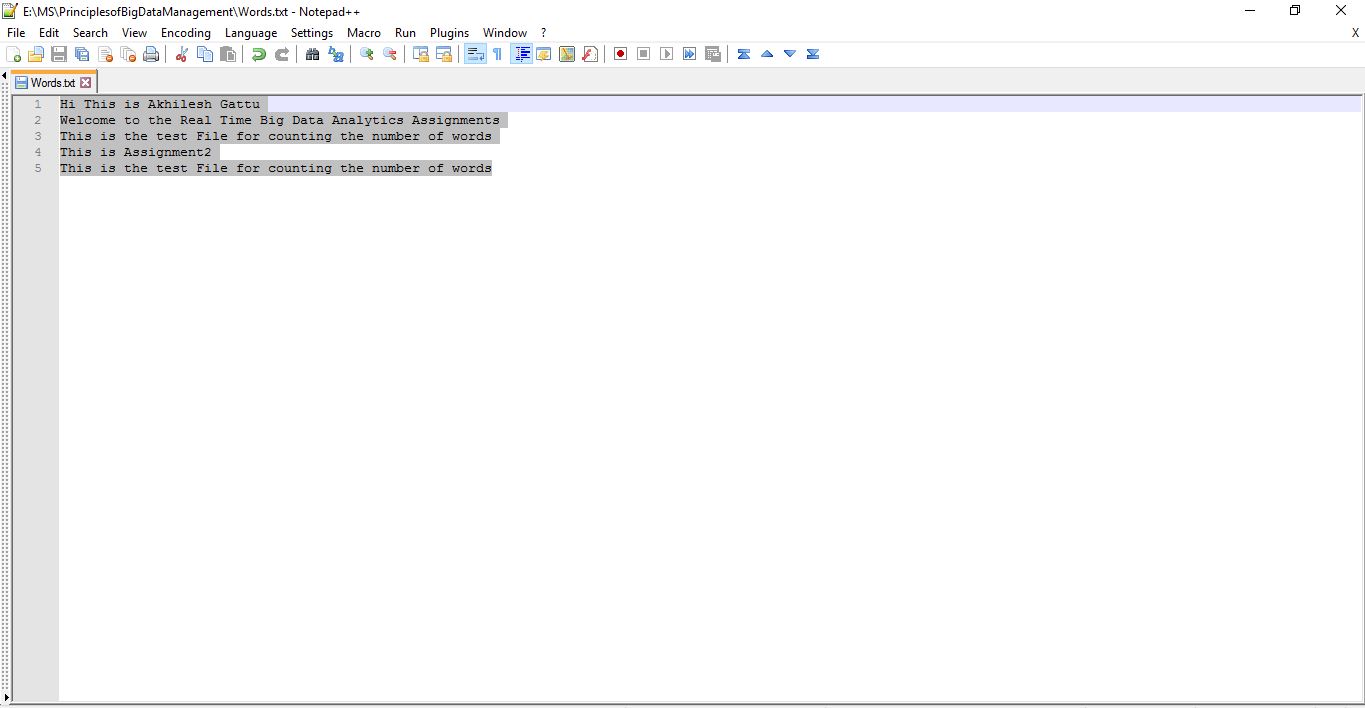
Hi This is Akhilesh Gattu

Welcome to Big Data Analytics Assignments

This is the text File for counting the number of words

This is Assignment2

This is the text File for counting the number of words



**Map Reduce Diagram:**

(Hi This is Akhilesh Gattu,1)

,

(Hi This is Akhilesh Gattu,1)

Hi This is Akhilesh Gattu

(Welcome to the Real Time Big Data Analytics Assignments,1)

(Welcome to the Real Time Big Data Analytics Assignments,1)

Welcome to the Real Time Big Data Analytics Assignments

Hi This is Akhilesh Gattu.

Welcome to the Real Time Big Data Analytics Assignments.

This is the test File for counting the number of words.

This is Assignment2.

This is the test File for counting the number of words.

(Hi This is Akhilesh Gattu,1)

(Welcome to the Real Time Big Data Analytics Assignments,1)

(This is the test File for counting the number of words,2)

(This is Assignment2,1)

(This is the test File for counting the number of words,2)

(This is the test File for counting the number of words,1)

This is the test File for counting the number of words

(This is Assignment2,1)

This is Assignment2

(This is Assignment2,1)

(This is the test File for counting the number of words,1)

This is the test File for counting the number of words