**Session 16: INTRODUCTION TO APACHE SPARK**

Assignment 16.1

Student Name: Abarajithan SA

Course: Big Data Hadoop & Spark Training

Start Date:  2017-09-09

End Date:  2017-11-26

**Assignment 16.1**– Perform Apache Spark operations.

Contents

[Introduction 1](#_Toc502178309)

[Problem Statement 1](#_Toc502178310)

[Task1 - Find the sum of all numbers 2](#_Toc502178311)

[Task2 - find the total elements in the list 2](#_Toc502178312)

[Task3 - calculate the average of the numbers in the list 2](#_Toc502178313)

[Task4 - find the sum of all the even numbers in the list 2](#_Toc502178314)

[Task5 - find the total number of elements in the list divisible by both 5 and 3. 2](#_Toc502178315)

# Introduction

In this assignment, we are going to perform some Spark RDD operation with the given problem statement.

# Problem Statement

1. Given a list of numbers - List[Int] (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)

- find the sum of all numbers

- find the total elements in the list

- calculate the average of the numbers in the list

- find the sum of all the even numbers in the list

- find the total number of elements in the list divisible by both 5 and 3.

# Task1 - Find the sum of all numbers

RDD,

***val num=sc.parallelize(list(1, 2, 3, 4, 5, 6, 7, 8, 9, 10))***

***val sum=num.sum()***

# Task2 - find the total elements in the list

***val count=num.count()***

# Task3 - calculate the average of the numbers in the list

***val average=num.mean()***

# Task4 - find the sum of all the even numbers in the list

***val even=num.filter(i=>(i%2==0))***

***val sum\_even=even.sum()***

# Task5 - find the total number of elements in the list divisible by both 5 and 3.