**Session 4: SCHEDULERS IN YARN & INTRODUCTION TO PIG**

Assignment 4.3

Student Name: Abarajithan SA

Course: Big Data Hadoop & Spark Training

Start Date:  2017-09-09

End Date:  2017-11-26

**Assignment 4.3** – Write a program to implement wordcount using Pig.

Share the screenshots of the commands used with its associated output.

Contents

[Problem Statement 2](#_Toc496195845)

[Introduction 2](#_Toc496195846)

[Input Texts 2](#_Toc496195847)

[PIG Commands – Word Count Example 2](#_Toc496195848)

[1. Load the data from HDFS 2](#_Toc496195849)

[2. Convert the Sentence into words and Convert Column into Rows 2](#_Toc496195850)

[3. Apply GROUP BY 3](#_Toc496195851)

[4. Generate Word Count 4](#_Toc496195852)

[Expected Output 5](#_Toc496195853)

[Complete PIG Code 5](#_Toc496195854)

# Problem Statement

Write a program to implement word count using Pig. Share the screenshots of the commands used with its associated output.

# Introduction

In this assignment we are going to write **word count** program using Pig Latin.

# Input Texts

We have the below words in the text file wordcountpig.txt which is placed at /home/acadgild/hadoop.

***“Write a program to implement wordcount using Pig. Share the screenshots of the commands used with its associated output.***

***Here we will write a simple pig script for the word count program.”***

# PIG Commands – Word Count Example

## Load the data from HDFS

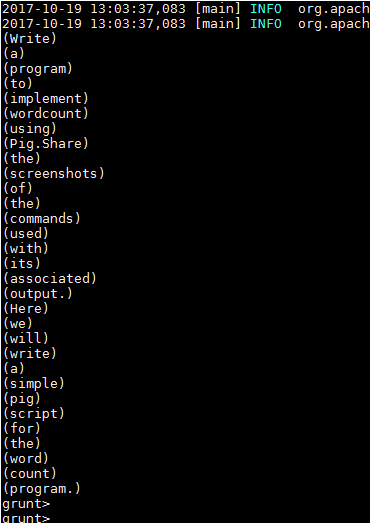
**“Word\_Count = LOAD '/home/acadgild/hadoop/wordcountpig.txt' USING PigStorage(',') AS (line:chararray);”**



## Convert the Sentence into words and Convert Column into Rows

**“Convert\_Words = FOREACH Word\_Count GENERATE FLATTEN(TOKENIZE(line,' ')) AS word;”**

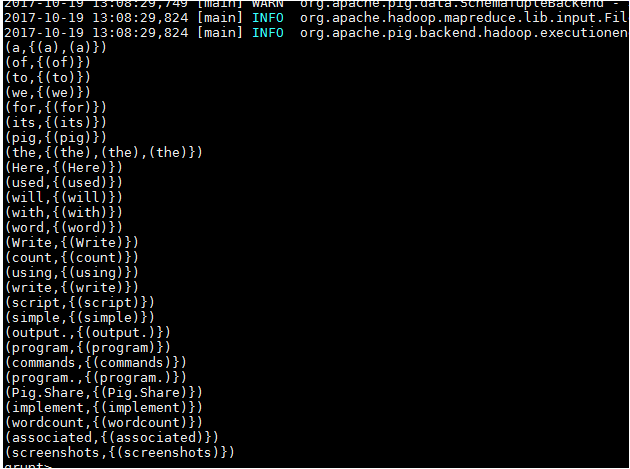




## Apply GROUP BY

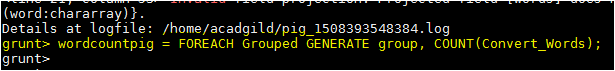
**“Grouped = GROUP Convert\_Words BY word;”**





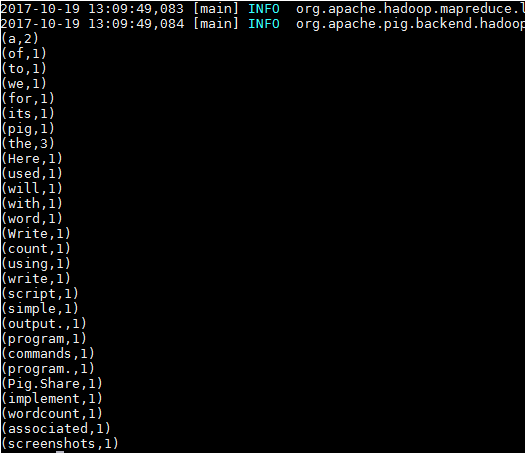
## Generate Word Count

**“wordcountpig = FOREACH Grouped GENERATE group, COUNT(Convert\_Words);**



## Expected Output

**DUMP wordcountpig;**



### Complete PIG Code

**Word\_Count = LOAD '/home/acadgild/hadoop/wordcountpig.txt' USING PigStorage(',') AS (line:chararray);**

**Convert\_Words = FOREACH Word\_Count GENERATE FLATTEN(TOKENIZE(line,' ')) AS word;**

**Grouped = GROUP Convert\_Words BY word;**

**wordcountpig = FOREACH Grouped GENERATE group, COUNT(Convert\_Words);**

**DUMP wordcountpig;**

**xxxxxx--------------------COMPLETED------------------xxxxxx**