



BIG DATA
DEVELOPMENT

ACADGILD

Session 8: Pig

Assignment 2

Session 8: Pig

Assignment –PIG

Table of Contents

1. Introduction	3
2. Objective	3
3. Prerequisites	3
4. Associated Data Files	3
5. Problem Statement.....	3
6. Expected output.....	4
7. Approximate Time to Complete Task	4

Big Data and Hadoop Development

1. Introduction

In this assignment you write the MapReduce code for analyzing the Olympic data.

2. Objective

This assignment will help you to understand basic pig commands.

3. Prerequisites

Acadgild's VM, or Linux operating system with Hadoop and Pig installed in it

4. Associated Data Files

N/A

5. Problem Statement

Create a sample dataset and implement the below Pig commands on the same dataset.

- 1) Concat
- 2) Tokenize
- 3) Sum
- 4) Min
- 5) Max
- 6) Limit
- 7) Store
- 8) Distinct
- 9) Flatten
- 10) IsEmpty

6. Expected output

N/A

7. Approximate Time to Complete Task

60 min