Big Data and Hadoop Development

A C A D G I L D

Page

1

Session

7: Pig

Assignment

1

Big Data and Hadoop Development

A C A D G I L D

Page

2

Session 7

:

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

................................

................................

................................

................................

....

3

7.

Approximate Time to Complete Task

................................

................................

................................

...

4

Big Data and Hadoop Development

A C A D G I L D

Page

3

1.

Introduction

You will work on the concepts of Pig.

2.

Objective

This assignment will help you to understand pig concepts.

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

Give a brief

answers

to

the

questions

below

:

1.

Why Map-reduce program is needed in Pig Programming?

Pig is an application that works on top of MapReduce, Yarn or Tez. Pig is written in Java and compiles Pig Latin scripts into to MapReduce jobs.

Pig has a relational operator that is mapreduce that allows you t plug in Java mapreduce job.Pig is nothing but a higher level of abstraction,and mapreduce is a lower level of abstraction .So finally mapreduce program is needed in pig programming.

2.

What are advantages of pig over MapReduce?

Not everyone is a Java expert.Pig provides easier way to code.

Pig could be used for ETL(Extraction Transformation Load) tasks naturally as it can handle unstructured data.

MapReduce requires programmers:

Programmers must think in terms of map and reduce functions.

Most probably Java programmers are required.

Pig provides high-level language that can be used by:

Data Analysts

Data Scientists

Pig got its name because it’s omnivorous – it will happily consume any data you feed it: structured, semi-structured, or unstructured.

3.

What is pig engine and what is its importance?

Pig engine parses ,optimizes, and automatically executes Pig latin script.It is an important component of pig

Complex tasks comprised of multiple interrelated data transformations are explicitly encoded as data flow sequences, making them easy to write, understand, and maintain.

4.

What are the modes of Pig execution?

Local mode and mapreduce mode.

5.

What is grunt shell in Pig?

Grunt shell is an interactive Shell for executing Pig Commands.It is used

when script file is not provided.It can execute scripts from Grunt via run or exec commands.

6.

What are the features of Pig Latin language?

Pig Latin is the language used to analyze data in Hadoop using Apache Pig.Features are:Pig latin has many built in functions like Eval.math etc.which are very useful. This language provides various operators using which programmers can develop their own functions for reading, writing, and processing data.

7.

Is Pig latin commands case sensitive?

Yes.

8.

What is a data flow language?

In computer programming, dataflow programming is a programming paradigm that models a program as a directed graph of the data flowing between operations, thus implementing dataflow principles and architecture.

6.

Expected output

N/A