CLOUD COMPUTING

Practical 7: Analyze Big Data with Hadoop on AWS

School of Computer Science
University College Dublin

To be GRADED: YES

Description

Amazon Elastic MapReduce (EMR) is a managed service to run Apache Hadoop to process vast amounts of data. It is an AWS tool for big data processing and analysis on Cloud. Besides, it is often used for data analysis in log analysis, web indexing, data warehousing, machine learning, financial analysis, scientific simulation, bioinformatics and more.

Suggestion: Students can use the guide document of AWS at link:
https://aws.amazon.com/getting-started/projects/analyze-big-data/

Questions

In this practical, students will deploy a fully functional Hadoop cluster, ready to analyze log data. Students can create or download a log file for yourself to practical. The students should do by followings steps:

Step 1: Set Up Prerequisites

Step 2: Launch The Cluster

Step 3: Allow SSH Access

Step 4: Run a Hive Script to Process Data

Step 5: Clean Up Resources

After finishing every steps, students need to capture the screen to input into your answer file.

Submission Instructions

All submissions must be done via **Brightspace** with deadline: **15 November 2021, 23:55**. Your submission should consist of one file (PDF), which contains the answers, to the above questions. The submitted file should be named following the format:

*COMPxxxxx_Surname_FirstName_StudentNo_ **Pracitcal**07*. pdf Example: *(COMP47780_Smith_John_12345 _**Practical**06.pdf)*