



**CCT College Dublin**

**Assessment Cover Page**

| **Module Title:** | HDip AI Concepts - Feb 2023 - SB+HCI cohort  Cohort: FT/PT  Storage Solutions for Big Data, Data Visualization Techniques |
| --- | --- |
| **Assessment Title:** | Integrated Continuous Assessment 2 |
| **Lecturer Name:** | Lecturer(s): Muhammad Iqbal, James Garza |
| **Student Full Name:** | Ciaran Quinlan |
| **Student Number:** | Programme Title: Module Title(s): Assignment Type: Individual Weighting(s): 40%, 40%  Assignment Title:Lecturer(s): Muhammad Iqbal, James Garza  Issue Date: 19th October 2023  Submission  Deadline Date:  Ciaran Quinlan (sbs23098@student.cct.ie) |
| **Assessment Due Date:** | 26th Nov 2023 |
| **Date of Submission:** | 22/11/2023 |

**Ciaran Quinlan**

**Declaration**

| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |
| --- |

# Table of Contents

[Table of Contents 3](#_heading=h.gjdgxs)

[1. Introduction 5](#_heading=h.30j0zll)

[2. SSBD Assessment Task 5](#_heading=h.1fob9te)

[Question 1: 5](#_heading=h.aoees4tqhce7)

[Question 2: 5](#_heading=h.aejjy3kfdtho)

[Question 3: 5](#_heading=h.9jrv21liu0os)

[Question 4: 5](#_heading=h.ypmaj1omb772)

[Question 5: 6](#_heading=h.jkzhc4wg9gob)

[7. Conclusion 6](#_heading=h.2afmg28)

[8. References 7](#_heading=h.pkwqa1)

[Appendix 1 7](#_heading=h.ru1zxpow32dx)

[Filenames and directories 7](#_heading=h.swzu1ky1ay4k)

# Introduction

This is the integrated assessment for Muhammad and James.

See the Appendix for list of file

See Git @ for any images, code or markdown.

# SSBD Assessment Task

## Question 1:

Can you define Big Data? Explain major characteristics of Big Data. Can banks enhance their profits with the support of big data processing and analysis? Research and name top three businesses that have obtained the benefits of big data storage solutions in the recent past.

Big Data is……………..

## Question 2:

What is HDFS, and how does it differ from a traditional file system? Describe and explain different layers of Hadoop framework. Explain five important characteristics of the Hadoop framework. Show the deployment of Hadoop on your virtual machine (VM) by providing the screenshots of (Namenode, Datanode etc.) and your username clearly shows your VM.

| git config --global user.email "sbs23098@student.cct.ie"  git config --global user.name "CIARAN QUINLAN CCT" |
| --- |

## Question 3:

Demonstrate a comparison of MySQL and Apache Hive based on the architecture and performance.

Consider a dataset and perform a query on both systems with at least 5,000 rows and at least 5 features. Show the duration of query execution by displaying screenshots obtained from a virtual machine (VM).

## Question 4:

Explain Apache Flink architecture and illustrate with your own conceptual diagram (Use of online/ book images is prohibited, Use draw.io to create the image). What is Apache Storm, and how does it differ from other distributed computing systems? Consider a text file comprising at least 20,000 words and write a wordcount program (Java/ Python) to count the frequency of words and related aggregation functions.

## Question 5:

Why is Apache Storm useful for Stream processing specifically? Distinguish the characteristics of Apache storm as compared to Hadoop. What is the role of Apache Zookeeper in Apache Storm deployment. Provide the screenshot of your VM to show working of Storm UI including Cluster, Nimbus and Owner summary.

# Conclusion

In conclusion, the go

# References

## Appendix 1

### Filenames and directories