Department of Computer Science

CS5710/CS5810 High Performance Computational Infrastructures

Coursework Assessment for 2022/23

Table of Contents

Assessment Title	Individual Project Development
Module Leader	Alaa Marshan
Distribution Date	Week 20 - 4th week of Term 2
Submission Deadline	11 AM Monday April 24th 2023
Feedback by	3 weeks after the submission deadline
Contribution to overall module assessment	5 Credits out of 15
Indicative student time working on assessment	20 Hours
Word or Page Limit (if applicable)	There is no word/page limit for this assessment, but the best effort should be made to ensure the submission is as concise as possible.
Assessment Type (individual or group)	Individual

Main Objective of the Assessment

In this assessment, you are required to demonstrate the appropriate practical skills and abilities to implement solutions using modern large-scale data storage and processing infrastructures, and to critically reflect on the concepts, theory and use of high-performance computational infrastructures.

Description of the Assessment

You are required to identify and analyse a real-world problem, design and implement a solution to the problem using Hadoop, and evaluate your implementation. The problem can be a simplified version from its original scale, extent or level of difficulties etc.

Learning Outcomes and Marking Criteria

LO1: Demonstrate the appropriate practical skills/abilities required to implement solutions using modern large-scale data storage and processing infrastructures.

LO2: Reflect critically on the concepts, theory and appropriate use of large-scale data storage and processing infrastructures (commonly used in modern organisational environments).

Marking Criteria

The coursework will be marked for 2 main criteria:

1. **Identifying a real data analytics problem with strong motivation for using distributed processing methods (LO1)**
2. **Designing, Implementing and Evaluating a working solution using distributed analytical techniques (MapReduce Algorithms) (LO1 & LO2)**

Grade Band E and F (E+, E, E-, F)
The candidate fails to meet the minimum requirements as outlined in the learning outcomes.

Grade Band D (D+, D, D-)
The work demonstrates significant weaknesses, but all of the learning outcomes have been met at the minimum requirement level. The work provides evidence of some critical understanding of large-scale data storage and processing infrastructures and demonstrates some abilities and skills to implement solutions using these technologies. The characteristics of a coursework in this grade band are:

- The research question, which is based on a chosen dataset, is not very well set (vague and the purpose is not clear)
- The proposed MapReduce solution is the same as the analyses and programs used in the lab sheets but on different dataset/s
- The proposed MapReduce solution is partially working (this means that the MapReduce program is working and producing results but not really the results that answer the research question)
- Basic/No reflection on the output (results)

Grade Band C (C+, C, C-)
In addition to the requirements for a grade in D-band, it demonstrates the ability to develop an independent, systematic, logical and effective solution to the problems identified. It also demonstrates a significant degree of competence in the appropriate use of the relevant methodologies, practices, and tools, etc., to analyse the problems and evaluate the solutions. The characteristics of a coursework in this grade band are:

- Clear definition of the research question based on the chosen dataset
- The proposed MapReduce solution is similar to the analyses and programs used in the lab sheets but on different dataset/s
- The proposed MapReduce solution is working, and the MapReduce program is producing results that help answer the research question
- Basic reflection on the output (results) is provided
- No meaningful insights are discovered from the results

Grade Band B (B+, B, B-)
In addition to the requirements for a grade in C-band, it clearly demonstrates the ability to develop an independent, systematic, logical and effective solution to the problems identified. It also demonstrates a high degree of competence in the appropriate use of the methodologies, practices, and tools, etc., to analyse the problems and evaluate the solutions. The characteristics of a coursework in this grade band are:

- Clear and important research question, which is based on the chosen dataset, is set at the beginning (clear justification of using distributed computation techniques)
- The proposed MapReduce solution is slightly different compared to the analyses and programs in the lab sheets and on different dataset/s
- The proposed MapReduce solution is working, and the MapReduce program is producing results that help answer the research question
- Good reflection on the output (results) is provided
- Some insights are discovered from the results

Grade Band A (A*, A+, A, A-)
In addition to the requirements for a grade in B-band, the work clearly demonstrates a sophisticated, critical and thorough understanding and ability to process large-scale data using High-Performance Computational Infrastructures. It provides evidence of originality of thought and clearly demonstrates the ability to develop an independent, systematic, logical and effective solution to the problems identified. It also demonstrates excellence in the appropriate use of methodologies, practices, and tools, etc., to analyse the problems and evaluate the solutions. The characteristics of a coursework in this grade band are:

- Clear and important research question, which is based on the chosen dataset, is set at the beginning (clear justification of using distributed computation techniques)
- The proposed MapReduce solution is fundamentally different compared to the analyses and programs in the lab sheets and on different dataset/s
- The proposed MapReduce solution is working, and the MapReduce program is producing results that help answer the research question
- Excellent reflection on the output (results) is provided
- Meaningful and actionable insights are discovered from the results

Format of the Assessment

There is no word/page limit for this assessment, but the best effort should be made to ensure the submission is as concise as possible. You should include sections on (percentage of overall mark):

Introduction, Problem description & associated dataset (25%)
Design & Implementation (50%)
Results & Evaluation (25%)

Note 1: you do NOT need to include the dataset as part of the submission. Just refer to the source in the references.
Note 2: the full code should be included in a separate text file and not in the appendix

A mark will be assigned to each section to form an overall percentage. This will then be converted into your final grade.

Submission Instructions

You must submit your coursework as a PDF file on Wiseflow by 11 AM Monday April 24th 2023. You can follow the link to Wiseflow through the module's section on Brightspace. The name of your file should follow the normal convention set out in the student handbook and must therefore include your student ID number (e.g., 0612345.pdf). It can also include the module code (e.g., CS2001_0612345.pdf).

Avoiding Academic Misconduct

Before working on and then submitting your coursework, please ensure that you understand the meaning of plagiarism, collusion, and cheating (including contract cheating) and the seriousness of these offences. Academic misconduct is serious and being found guilty of it results in penalties that can reduce the class of your degree and may lead to you being expelled from the University. Information on what constitutes academic misconduct and the potential consequences for students can be found in Senate Regulation 6.

You may also find it useful to read this PowerPoint presentation which explains, in plain English, the different kinds of misconduct, how to avoid (even accidently) committing them, how we detect misconduct, and the common reasons that students give for engaging in such activities.

If you are experiencing difficulties with any part of your studies, remember there is always help available:

- Speak to your personal tutor. If you're not sure who your tutor is, please ask the Taught Programmes Office (TPOcomputerscience@brunel.ac.uk).
- Alternatively, if you prefer to speak to someone outside of the Department you can contact the Student Support and Welfare team.

Late Coursework

The clear expectation is that you will submit your coursework by the submission deadline stated in the study guide. In line with the University's policy on the late submission of coursework (revised in July 2016), coursework submitted up to 48 hours late will be accepted but capped at a threshold pass (D- for undergraduate or C- for postgraduate). Work submitted over 48 hours after the stated deadline will automatically be given a fail grade (F).

Please refer to the Computer Science student information pages and the Coursework Submission Procedure pages for information on submitting late work, penalties applied and procedures in the case of Extenuating circumstances.