Coursework Assignment Brief

Postgraduate

***Academic Year 2021***

|  |  |  |  |
| --- | --- | --- | --- |
| **Module Title:** | Big Data Management | | |
| **Module Code:** | CMP7203 | | |
| **Assessment Title:** | Design and implementation of a big data solution | | |
| **Assessment Identifier:** | Coursework |  | Weighting: 100 |
| **School:** | School of Computing and Digital Technology | | |
| **Module Co-ordinator:** | Besher Alhalabi | | |
| **Hand in deadline date:** | Refer to Moodle page | | |
| **Return of Feedback date and format** | Refer to Moodle page | | |
| **Re-assessment hand in deadline date:** | Refer to Moodle page | | |
| **Support available for students required to submit a re-assessment:** | Timetabled revisions sessions will be arranged for the period immediately preceding the hand in date | | |
| **NOTE:** | At the first assessment attempt, the full range of marks is available. At the re-assessment attempt the mark is capped and the maximum mark that can be achieved is 50%. | | |
| **Assessment Summary** | The student will build a big data ecosystem using the tools and methods that discussed during the module; to do this the student will analyse a data set simulating big data generated from a large number of users who are playing an imaginary game called "Catch the Pink Flamingo". The student needs to achieve the following:   1. Critically evaluate the three big data processing paradigms as discussed during the course. 2. Develop and implement a big data solution that covers the following:  * Data Exploration: including acquiring, cleaning, exploring and preparing the data for analysis; any relevant data exploration tool/visualisation that could fit the purpose is allowed to be used. * Machine learning with big data: applying classification & clustering techniques on the proposed dataset. * Graph Analysis: Using Neo4j/ Gephi to perform graph analytics of the simulated chat data to find ways of improving the game.  1. Discuss and visualise the resulting data insights. 2. Evaluate the role of ethics on data storage and processing.   The report will be submitted as **one** deliverable in the form of a written report. The standard of academic writing should be excellent. (Maximum words: **4000 words**, excluding tables, figures and references). | | |

**IMPORTANT STATEMENTS**

***Standard Postgraduate Regulations***

Your studies will be governed by the BCU Academic Regulations on Assessment, Progression and Awards. Copies of regulations can be found at <https://icity.bcu.ac.uk/Academic-Services/Information-for-Students/Academic-Regulations-2018-19>

For courses accredited by professional bodies such as the IET (Institution of Engineering and Technology) there are some exemptions from the standard regulations and these are detailed in your Programme Handbook

***Cheating and Plagiarism***

Both cheating and plagiarism are totally unacceptable and the University maintains a strict policy against them. It is YOUR responsibility to be aware of this policy and to act accordingly. Please refer to the Academic Registry Guidance at <https://icity.bcu.ac.uk/Academic-Registry/Information-for-Students/Assessment/Avoiding-Allegations-of-Cheating>

The basic principles are:

* Don’t pass off anyone else’s work as your own, including work from “essay banks”. This is plagiarism and is viewed extremely seriously by the University.
* Don’t submit a piece of work in whole or in part that has already been submitted for assessment elsewhere. This is called duplication and, like plagiarism, is viewed extremely seriously by the University.
* Always acknowledge all of the sources that you have used in your coursework assignment or project.
* If you are using the exact words of another person, always put them in quotation marks.
* Check that you know whether the coursework is to be produced individually or whether you can work with others.
* If you are doing group work, be sure about what you are supposed to do on your own.
* Never make up or falsify data to prove your point.
* Never allow others to copy your work.
* Never lend disks, memory sticks or copies of your coursework to any other student in the University; this may lead you being accused of collusion.

By submitting coursework, either physically or electronically, you are confirming that it is your own work (or, in the case of a group submission, that it is the result of joint work undertaken by members of the group that you represent) and that you have read and understand the University’s guidance on plagiarism and cheating*.*

You should be aware that coursework may be submitted to an electronic detection system in order to help ascertain if any plagiarised material is present. You may check your own work prior to submission using Turnitin at the [Formative Moodle Site](https://moodle.bcu.ac.uk/enrol/index.php?id=715). If you have queries about what constitutes plagiarism, please speak to your module tutor or the Centre for Academic Success.

***Electronic Submission of Work***

It is your responsibility to ensure that work submitted in electronic format can be opened on a faculty computer and to check that any electronic submissions have been successfully uploaded. If it cannot be opened it will not be marked. Any required file formats will be specified in the assignment brief and failure to comply with these submission requirements will result in work not being marked. You must retain a copy of all electronic work you have submitted and re-submit if requested.

|  |
| --- |
| **Learning Outcomes to be Assessed:**   1. Critically evaluate modern big data processing paradigms. 2. Develop and implement a big data solution for a provided dataset. 3. Analyse use cases, visualise and report the results of a big data solution. 4. Assess how ethics govern the design choices in devising a Big Data enabled solution. |

**Assessment Details:**

|  |
| --- |
| **Title:** *Design and develop big data eco system*  **Type:** *Coursework*  **Style:** Report |
| **Rationale:**  Understanding big data solutions and their impact on business growth is an essential topic in the industry. This assignment will walk through the different stages of developing big data solution that gives an understanding of what insights big data can provide through hands-on experience with the tools and systems used by big data scientists and engineers. It will help the students to ask the right questions about data, communicate effectively with data scientists, and do basic exploration of large, complex datasets.  **Description:**  The student will use cutting edge big data technologies to walk through a typical big data science steps for acquiring, exploring, preparing, analysing, and reporting big data. Each student (individually) is tasked to:   * Identify a unique and acceptably **challenging data analysis problem** that can result in a factual insight for the proposed dataset. The student will deliver insights from the dataset and identify an appropriate method for the analysis process. * Apply machine learning techniques to explore and prepare data for modelling. * Identify the type of machine learning problem in order to apply the appropriate set of techniques * Analyse big data problems using scalable machine learning algorithms on Spark. * Model, store, retrieve and analyse graph-structured data. * Critically evaluate how ethics could govern the design of a Big Data solution. * Write reports.   **Additional information:**   * Students will be guided through the whole technical process in the lab sessions. * Students are going to demonstrate their progress for the assessment during week 8 and week 10. * The report must include:   + CoverPage(Student ID, StudentName).   + Section 1: Evaluation of big data processing paradigms.   + Section 2: Exploratory data analysis.   + Section 3: Classification results on the proposed data set.   + Section 4: Clustering results on the proposed data set.   + Section 5: Graph analysis including ways to improve the game.   + Section 6: Ethics, Findings & recommendations. * References (as per Harvard Referencing Style) * Students must use Harvard Referencing Style <https://icity.bcu.ac.uk/Library-and-Learning-Resources/Referencing/harvard-referencing> ). * Report format for the submissions is **DOCX**. * A video course on **Impactful Writing**is available at: [https://www.youtube.com/playlist?list=PLU98PJIZ0JfIj\_msDROdMU85wdV72ihks](https://owa.bcu.ac.uk/owa/redir.aspx?C=gjf6luDeFxr-C78IxyXUsgfOfyh7q9-9Aa-gFVsz-MuK3qsYsEzWCA..&URL=https%3a%2f%2fwww.youtube.com%2fplaylist%3flist%3dPLU98PJIZ0JfIj_msDROdMU85wdV72ihks)   For advice on writing style, referencing and academic skills, please make use of the Centre for Academic Success: <https://icity.bcu.ac.uk/celt/centre-for-academic-success>    For advice on writing style, referencing and academic skills, please make use of the Centre for Academic Success: <https://icity.bcu.ac.uk/celt/centre-for-academic-success> |
| **Workload:**  *A typical student will spend up to 90 hours to complete this work. The word count limit is the equivalent of 4000 words.* |
| **Transferable skills:**   * Problem solving * Programming skills * Analytical skills * Team work * Time management * Project management * Communicating scientific and technical knowledge * Presentation Skills |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Marking Criteria:**  *Marking criteria should be used as a general guide to the expectations at different grades. Ensure it is clear how the criteria will be used to derive the final grade for the component. Ensure it is clear how the component grade goes towards the final mod*  **Table of Assessment Criteria and Associated Grading Criteria**   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Assessment**  **Criteria**  **** | **1.**  **Critically evaluate modern big data processing paradigms.** | **2.**  **Develop and implement a big data solution** | **3.**  **Analyse use cases, visualise and report the results of a big data solution** | **4.**  **Assess how ethics govern the design choices in devising a Big Data enabled solution** | | **Weighting:** | **20%** | **35%** | **35%** | **10%** | | **Grading**  **Criteria**  **0 – 29%**  **F** | Insufficient evaluation of modern big data processing paradigms. Less than 3 different paradigms mentioned but not evaluated. | No or insufficient development of a big data solution. Dataset neither uploaded nor explored. | Insufficient analysis and problem selection. Problem is trivial or has less than 2 features to be analysed. | No understanding and demonstration of ethics and data governance. | | **30 – 39%**  **E** | Poor but sufficient evaluation of modern big data processing paradigms where more than 3 paradigms mentioned but not evaluated. | Poor but sufficient development of a big data solution. Dataset is uploaded but poorly explored. | Poor analysis and problem selection. Problem is not complex enough to allow the analysis of multiple different features. | Only superficial understanding but no real link to the coursework | | **40 – 49%**  **D** | Basic evaluation of modern big data processing paradigms where more than 3 paradigms evaluated. | Basic development of a big data solution. Dataset is uploaded with basic attempt to perform Exploratory Data Analysis (EDA). Only a few statistical techniques have been provided. | Basic analysis and problem selection. Problem is complex enough to allow the analysis of multiple different features but limited in scope. No visualisation. | Some link to the broad concept of ethics and data governance but no specific links. | | **50 – 59%**  **C** | A good evaluation of modern big data processing paradigms where more than 3 paradigms critically evaluated. | A good development of a big data solution. Dataset is uploaded and explored with different techniques (statistical and graphical). | A good analysis and problem selection. Problem is complex enough to allow the analysis of multiple different features.  Less than 2 types of relevant noninteractive visualisation techniques used. | Specific links to items from ethics and governance to some use cases. | | **60 – 69%**  **B** | A good evaluation of modern big data processing paradigms where more than 3 paradigms critically evaluated including distributed data processing engines. | A good development of a big data solution. Dataset is uploaded and explored with different techniques (statistical and graphical). Only one machine learning technique is applied to the proposed dataset. | A good analysis and problem selection. Problem is complex enough to allow the analysis of multiple different features.  More than 2 types of relevant noninteractive visualisation techniques used.  Visualisation for only one machine learning technique is provided. | A good attempt to cover all aspects/use cases for the issues arising from ethics and governance perspectives. | | **70 – 79%**  **A** | A very good evaluation of modern big data processing paradigms where more than 3 paradigms critically evaluated including distributed data processing engines and distributed streaming platforms. | A very good development of a big data solution. Dataset is uploaded and explored with different techniques (statistical and graphical). Two machine learning techniques are applied to the proposed dataset. | A very good analysis and problem selection. Problem is complex enough to allow the analysis of multiple different features.  More than 2 types of relevant noninteractive visualisation techniques used.  Visualisation for two machine learning techniques is provided. | An excellent attempt to cover all aspects/use cases for the issues arising from ethics and governance perspectives. | | **80 – 89%**  **A+** | Excellent critical evaluation of modern big data processing paradigms where more than 3 paradigms critically evaluated including distributed data processing engines and distributed streaming platform | Excellent development of a big data solution. Dataset is uploaded and explored with different techniques (statistical and graphical). Two machine learning techniques are applied to the proposed dataset. Graph analysis is performed | Excellent analysis and problem selection. Problem is complex enough to allow the analysis of multiple different features.  More than 2 types of relevant noninteractive visualisation techniques used.  Visualisation for two machine learning techniques and graph analysis is provided. | A critical view beyond the course material that focuses on some of the emerging issues in ethics/ data governance, and how that may affect the submitted work or related work in real life. Excellent Scientific Writing style. (where applies) | | **90 – 100%**  **A\*** | exceptional critical evaluation of modern big data processing paradigms where more than 3 paradigms critically evaluated including distributed data processing engines and distributed streaming platform, | Excellent development of a big data solution. Dataset is uploaded and explored with advanced techniques (statistical and graphical). Two machine learning techniques are applied to the proposed dataset. Graph analysis is performed. | excellent analysis and problem selection. Problem is complex enough to allow the analysis of multiple different features.  More than 2 types of relevant noninteractive visualisation techniques used.  Advanced Visualisation for two machine learning techniques and graph analysis is provided. | A critical view beyond the course material that focuses on most recent emerging issues in ethics/ data governance, and how that may affect the submitted work or related work in real life. Excellent Scientific Writing style. (where applies) | |

**Submission Details:**

|  |
| --- |
| **Format:**  *Upload MS Word file to Moodle* |
| **Regulations:**  If you submit an assessment late at the first attempt then you will be subject to one of the following penalties:     * if the submission is made **between 1 and** **24 hours** after the published deadline the original mark awarded will be reduced by **5%**. For example, a mark of 60% will be reduced by 3% so that the mark that the student will receive is 57%. ; * if the submission is made between **24** **hours** and **one week (5 working days)** after the published deadline the original mark awarded will be reduced by 10%. For example, a mark of 60% will be reduced by 6% so that the mark the student will receive is 54%. * **if the submission is made after 5 days following the deadline, your work will be deemed as a fail and returned to you unmarked.**   The reduction in the mark will not be applied in the following two cases:   * + the mark is below the pass mark for the assessment. In this case the mark achieved by the student will stand   + where a deduction will reduce the mark from a pass to a fail. In this case the mark awarded will be the threshold (i.e. 50%)   Please note:   * **If you submit a re-assessment late then it will be deemed as a fail and returned to you unmarked.** |

**Feedback:**

Feedback for the deliverable will be provided via Moodle. The students are also strongly encouraged to discuss their draft work with tutors in lecture sessions, whenever time permits.

Marks and Feedback on your work will normally be provided within 20 working days of its submission deadline.

**Where to get help:**

Students can get additional support from the library support for searching for information and finding academic sources. See their iCity page for more information: <http://libanswers.bcu.ac.uk/>

The Centre for Academic Success offers 1:1 advice and feedback on academic writing, referencing, study skills and maths/statistics/computing. See their iCity page for more information: <https://icity.bcu.ac.uk/celt/centre-for-academic-success>

Link to My Assignment Planner tool: <http://library.bcu.ac.uk/MAP2/freecalc-mail/>

**Fit to Submit:**

Are you ready to submit your assignment – review this assignment brief and consider whether you have met the criteria. Use any checklists provided to ensure that you have done everything needed.

|  |  |
| --- | --- |
| Item | Completed |
| Read the entire assignment brief, the required task and marking criteria? |  |
| Work through the checklist to know what you need to do and submit? |  |
| Clarify any points you are unsure of with the module coordinator? |  |
| Cloudera virtual machine is up and running? |  |
|  |  |
| Spark ML, Knime, Neo4J, Splunk is up and running ? |  |
| Dataset downloaded and been put on the big data eco system? |  |
| Included all the sections in ‘Assessment Details’? |  |