This document provides you with information about the requirements for assessment. Detailed instructions and resources are included to help you to complete and submit the task. The Criterion Reference Assessment (CRA) Rubric that markers use to grade the assessment task is included and should be used as a guide when working on the assessment task.

## Task Overview

| | |
|---|---|
| **Assessment name:** | Project (applied) |
| **Context:** | Big data analytics is an area of growing significance. Many companies realise the potential of data they have to make better and informed decisions. Big data is of huge volume, high variety and high velocity. There were concerns raised regarding the storage and analysis of big data using traditional databases. It was then NoSQL databases came into existence. MongoDB is a popular NoSQL database widely used for big data analytics. Big data analytic skills are beneficial whether you work as a data analyst, business analyst, software engineer, project manager, or in any other role.<br>This assessment exposes you to the capabilities of MongoDB to conduct big data analytics. |
| **Task description:** | In this assessment you will use MongoDB to analyse a big dataset, 'tweets', which will be provided to you.<br><br>You are expected to use aggregation commands, a special feature of MongoDB, to analyse information in the dataset. These are also the commands which are used in industry to conduct big data analytics and consider this assessment as if you were working on a real project in industry.<br><br>You will also create a map-reduce function which is used to parallelize the processing of huge datasets, which in this case is 'tweets'.<br><br>For further detail, please see the specific items noted in the **Task details** section of this document. |
| **Learning outcomes addressed:** | 1. Describe key issues for modern data management and identify corresponding technologies to address the issues, independent of any specific platform or framework.<br>2. Evaluate various capabilities of modern data management technologies.<br>3. Design and develop solutions to manage data using modern data management technologies.<br>4. Work collaboratively with others to efficiently manage and deliver on projects related to the development of modern data management solutions for a client. |

| | |
|---|---|
| **Due Date** | Sunday 30th October 2022 (Week 13) at 23:59 (AEST), via Blackboard. |
| **Report:** | Technical report, 12-point font |
| **Weighting:** | 40% |
| **Individual or Group:** | Group and Individual Components |
| **Authentic Assessment:** | Yes |
| **Formative/Summative:** | Summative |
| **How will I be assessed:** | 7-point grading scale using a rubric. The outlined tasks and technical report will relate to the first four criteria on the rubric, while the Peer Evaluation ratings will contribute to the "Team collaboration" criterion. Group work contributes towards 30% of this assessment. The group section in the rubric will be used to mark group work.<br><br>Individual work contributes to remaining 10% of this assessment. The individual section of rubric will be used to mark individual work. |

## Task details

| | |
|---|---|
| **What you will do:** | 1. Read the Criterion Reference Assessment Rubric.<br><br>2. Complete Tasks 1-15 **(see Appendix 1)**<br>In each task (apart from Task 1 and Task 15) teams are required to think of a specific scenario that is of value to an individual, group, or business.<br><br>3. The scenario should be more from analytics perspective. E.g. How many people from the same country tweet on similar topic areas?<br><br>4. Then you need to clearly explain why you think the query is of value to an individual, group, or business.<br><br>5. You will then write and execute the query using the aggregation framework of MongoDB to address the question raised in your scenario. You need to write the scenario, value proposition, and submit a screenshot of the query as well as the output of the query. You need to use Mongo Shell for the entire assessment. |

| | |
|---|---|
| | 6. You will also rate yourself and your peers, in terms of how you have contributed to the group submissions. |

| | |
|---|---|
| **Presentation requirements:** | The technical report needs to written, with:<br>• 12-point font<br>• Single linespacing<br>• Use of graphs or other forms of graphics to complement your findings is encouraged.<br><br>A form for the peer evaluation of group work will be provided and you need to fill in the gaps.<br><br>All documents including individual work saved as a single zip file for submission.<br><br>Use Vancouver (preferred) or APA referencing if you have any sources to cite, ensuring you are consistent with the style you use. |
| **Resources needed to complete task:** | • IAB206 Blackboard site<br>• Appendix 1: Tasks 1 - 15<br>• Appendix 2: Technical Report Template (Suggested)<br>• Appendix 3: Peer Evaluation Form<br>• QUT Cite\|Write APA guide |

## Submission Information

| | |
|---|---|
| **What you will submit:** | You need to submit a **zip file** containing:<br><br>(i) A brief technical report **(Appendix 2)** answering all questions (Tasks 1-15) and containing the screenshots and explanations (where required) for Tasks 2-14. **Please note** you need to have a screenshot of the query and the document(s) before and after executing the query. You need to use **Mongo Shell.**<br>(ii) Document from each group member which comprises the individual task of this assessment.<br>(iii) The script for each of the queries you create in Tasks 2 -14.<br>(iv) Completed peer-evaluation of group work form **(Appendix 3).** |

| | |
|---|---|
| **How to submit:** | Only **one submission per team** is to be submitted and the submission is due on **30st October 2022.**<br><br>You will see a submission link in the assessment section of the IAB206 page in Blackboard by 25th October 2022.<br><br>Note that the submission deadline set in Blackboard is a hard deadline and the submission link will be disabled by the system once the deadline is reached.<br>(i) Each team should register their team details on Blackboard. The assignment file name should be IAB206_Team#_A3.zip. (Replace ## with your team number. Example: IAB206_Team5_A3.zip)<br><br>(ii) Each submission must contain a declaration, signed by all group members, stating that they have viewed the final version of the assignment that is to be submitted and that it is their original work.<br><br>(iii) Each submission must consist of individual work done by each group member.<br><br>Late submissions without accompanying eligible extension will be given a grade of 0 as per the QUT policy for late assessment, MOPP 6.3.5. |
| **What feedback will I receive?** | Under normal circumstances, you will receive individual written feedback on this assessment task within 15 days of submission.<br><br>The teaching team will be available to answer specific questions about the assessment, but not to pre-mark assessments. Teaching staff will NOT read report drafts and review detailed models prior to the submission.<br><br>Students are welcome to discuss any queries about the feedback they will receive. |
| **Moderation:** | All staff who are assessing your work meet to discuss and compare their judgements before marks or grades are finalised. |

## Academic Integrity

As a student of the QUT academic community, you are asked to work to uphold the principles of academic integrity during your course of study. QUT sets expectations and responsibilities of students, more specifically it states that students "adopt an ethical approach to academic work and assessment in accordance with this policy and the Student Code of Conduct. E/2.1 (MOPP C/5.3 Academic Integrity).

At university, students are expected to demonstrate their own understanding and thinking using the ideas provided by 'others' to support and inform their work, always making due acknowledgement to the source. While we encourage peer learning, it is not appropriate to share assignments with other students unless your assessment piece has been stated as being a group assignment. If you do share your assignment with another student, and they copy part of or all of your assignment for their submission, this is considered collusion and you may also be reported for academic misconduct. If you are unsure and need further information you can find this at: http://www.mopp.qut.edu.au/C/C_05_03.jsp#C_05_03.03.mdoc.

# Appendix 1: Tasks 1 – 15

**Task 1:** Consider your team to be experts in data analytics. You can assume your team to be working on a project provided by an individual, group, or organization. You need to decide that. The project primarily involved analysing the 'tweets' dataset. Familiarise yourself with the dataset. Have a look at the fields, understand what they mean, and the data they contain. Based on your evaluation of the dataset, provide a one-page report that is expected to be presented to key stakeholders of your project so that they are familiar with the dataset when you present the analysis to them (see Task 15).

***Tasks 2-14 should be completed using Mongo Shell***

**Task 2:** Describe a scenario and write a query to summarise (summarise would mean finding the mean, median, and other value of interest) a field of your choice from the entire dataset.

**Task 3:** Describe a scenario and write a query that uses $match operator to match documents against two conditions (e.g., age greater than 20 and year less than 2010) and then uses $project operator to project any four fields.

**Task 4:** Describe a scenario and write a query that has the following elements:

- Matches the documents against a given condition (e.g. number of employees greater than 500)
- Groups the documents using a logical _id field
- Provides aggregated information for each group (you may use $max, $min, $avg, $sum)

**Task 5:** Describe a scenario and write a query that skips through some documents in the third stage of the aggregation pipeline.

**Task 6:** Describe a scenario and write a query that sorts documents in the third stage of the aggregation pipeline.

**Task 7:** Describe a scenario and write a query that uses $bucket operator and limits results to a certain number of documents in the aggregation pipeline.

**Task 8:** Describe a scenario and write a query to reshape a document in your dataset such that the names of two fields is displayed as a fieldname of your choice in the output.

**Task 9:** Describe a scenario and write a query that uses any two of these functions: $concat, $substr, $toLower, $toUpper

**Task 10:** Describe a scenario and write a query that uses any two of these functions: $add, $divide, $mod, $multiply, $subtract

**Task 11:** Describe a scenario and write a query that uses $redact, $$descend and $$prune command

**Task 12:** Describe a scenario and write a query that uses the $graphlookup operator. Limit your search to 20 documents.

**Task 13:** Think from the perspective of a data analyst. You want to present some interesting information from the dataset, which may be useful for the key stakeholders. Describe a scenario and write a query using the aggregation framework of MongoDB that outputs this information. The query should not be similar to any queries you have written in Tasks 2-12 but may include any operator used in the prior tasks.

**Task 14:** Describe a scenario and write a query to create a simple map-reduce function for the dataset provided to you.

**Task 15*:** Write a short report summarising your findings of Tasks 2-14. Based on the summary, provide at least 5 recommendations for consideration to the key stakeholders. Your findings and recommendations should be understandable by the stakeholders.

**Please note that for this question each member in the group needs to have their own individual response. In other words, this question will be marked on an individual basis. Please provide separate responses for this question when submitting your assessment.*

# Appendix 2: Technical Report Template (Suggested)

**Important! Remember! For each task you need to include: description of scenario, value proposition, screenshot of query, screenshot of result. Screenshot should capture the entire screen with the query editor and output.**

| | |
|---|---|
| Task no | ….. |
| Scenario Description | ….. |
| Value Proposition | ….. |
| Query screenshot | ….. |
| Output screenshot | ….. |

# Appendix 3: Peer Evaluation Form

Write the name and student number of each of your group members in a separate column.
For each person, indicate the extent to which your team agrees with the statement on the left.

SA – Strongly Agree; A – Agree; D – Disagree; SD – Strongly Disagree

| **Evaluation Criteria** | **Group member** | **Group member** | **Group member** | **Group Member** |
|---|---|---|---|---|
| Attends group meetings and contributes meaningfully to group discussions. | SA/A/D/SD | SA/A/D/SD | SA/A/D/SD | SA/A/D/SD |
| Completes assigned tasks on time. | SA/A/D/SD | SA/A/D/SD | SA/A/D/SD | SA/A/D/SD |
| Prepares high-quality work. | SA/A/D/SD | SA/A/D/SD | SA/A/D/SD | SA/A/D/SD |
| Demonstrates a cooperative and supportive attitude. | SA/A/D/SD | SA/A/D/SD | SA/A/D/SD | SA/A/D/SD |
| Contributes significantly to the success of the project. | SA/A/D/SD | SA/A/D/SD | SA/A/D/SD | SA/A/D/SD |
| Based on these considerations, state a peer mark that each team member should receive out of 10. | /10 | /10 | /10 | /10 |

Adapted from a peer evaluation form developed at Johns Hopkins University (October, 2006)
Adapted from
https://www.cmu.edu/teaching/designteach/design/instructionalstrategies/groupprojects/tools/index.html

# IAB206 | MODERN DATA MANAGEMENT | ASSESSMENT TASK 3 RUBRIC | 40% WEIGHT

| Criteria | High Distinction 100 – 85% | Distinction 84 – 75% | Credit 74 – 65% | Pass 64 – 50% | Marginal Fail 49 – 40% | Fail / Low Fail 39 – 1% | No Evidence |
|---|---|---|---|---|---|---|---|
| **Group (30% of overall task weighting) – Tasks 1 - 14** | | | | | | | |
| **Analysis of client dataset**<br><br>(10%) | Thorough and well-reasoned critical analysis of client dataset demonstrates in-depth understanding of the requirements. | Critical analysis of client dataset demonstrates understanding of the specific requirements with a few minor errors/ omissions or misconceptions. | Analysis of client dataset demonstrates a broad understanding of the requirements but with some errors/omissions or misconceptions. | Analysis identifies key aspects of client dataset demonstrating an understanding of the requirements. | Client dataset is described with relations between some parts of the data identified. Demonstrates limited understanding of the requirements with significant errors/ omissions or misconceptions. | Basic review of client dataset conducted. Listed evidence is unrelated to focus and reveals a significant misinterpretation of the requirements. | No evidence of addressing this criterion. |
| **Problem solving and generation of solution for clients**<br><br>(35%) | Innovative and insightful scenario and value proposition is developed.<br><br>Solution generated demonstrates deep insight and thoroughly addresses multiple factors of the problem.<br><br>Could be used as a model solution. | Logical scenario and value proposition is developed.<br><br>Generates a realistic, comprehensive and effective solution for the client.<br><br>Could be used as a model solution after minor corrections. | Appropriate and efficient strategies were chosen for solving the problem.<br><br>Solution addresses factors of the problem in a proficient manner.<br><br>Could be used as a model solution after making few corrections. | Basic scenario and value proposition is developed. Most are similar.<br><br>Solution addresses the problem but overlooks some relevant factors. | Limited problem solving evidenced.<br><br>Solution for client is superficial and does not directly address the problem and/or solution is incorrect. | Problem is not clearly articulated, and no methodology is evident in addressing the problem.<br><br>Needs assistance to generate simple solutions. | No evidence of addressing this criterion. |
| **Execution of aggregation framework queries**<br><br>(40%) | All aspects of the queries were accurate and complete. Right operators were used.<br><br>Could be used as a model solution. | All the queries work without any issue retrieving data as intended. Not all operators are correct.<br><br>Could be used as a model solution after minor corrections. | Most queries were accurate. Most operators were correct.<br><br>Most of the queries work as intended retrieving data with some minor corrections. | Most of the queries work as intended to retrieve data with some correction required. Operators are not correctly used. | Some queries work as intended to retrieve data as required while others make an unsuccessful attempt. | None of queries work as intended to retrieve data as required. | No evidence of addressing this criterion. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Written Communication**<br><br>Technical Report Requirements<br><br>(10%) | The technical report demonstrates a professional standard of written communication skills in all aspects. Effective use of discipline-based terminology and notation and attends to all report requirements outlined.<br><br>Selection of supporting graphical imagery contributes significantly to the overall professional standard of the report. | The technical report demonstrates a high standard of written communication skills in all aspects. Discipline-based terminology and notation is used effectively within the report format.<br><br>Selection of supporting graphical imagery contributes significantly to the standard of the report. | The technical report demonstrates a competent standard of written communication skills.<br><br>Discipline-based terminology and notation are utilized within the technical report format outlined.<br><br>Selection of supporting graphical imagery contributes to the standard of the report. | The technical report meets a satisfactory standard of written communication skills.<br><br>Uses language that conveys meaning with some clarity.<br><br>Use of graphical imagery complements the report. | The technical report demonstrates an unsatisfactory standard of written communication skills.<br><br>Information presented lacks organisation resulting in an unclear central message. | The technical report submitted is incomplete.<br><br>Information presented is unstructured and disconnected to the report requirements outlined. | No evidence of addressing this criterion. |
| **Team Collaboration**<br>(5%) | Peer evaluation provides clear evidence of team collaboration and unity including effective planning of the tasks and equitable individual contributions. | Peer evaluation provides clear evidence of team collaboration and unity including effective planning of the tasks and an equitable split of workload. | Peer evaluation provides evidence of team collaboration and unity including effective planning of the tasks and a mostly equitable split of workload. | Peer evaluation provides some evidence of team collaboration and unity including planning of the tasks and a mostly equitable split of workload.<br><br>At times, individual contributions are of variable quality. | Peer evaluation provides evidence of infrequent collaboration and an inequitable split of workload.<br><br>Individual contributions are of variable quality and quantity.<br><br>Team required redirection at times. | Peer evaluation is incomplete and/or provides little evidence of team collaboration and unity.<br><br>Team required close monitoring. | No evidence of addressing this criterion. |
| **Individual (10% of overall task weighting) – Task 15** | | | | | | | |
| **Problem solving and generation of solution for clients**<br><br>(60%) | Innovative and insightful scenario and value proposition is developed.<br><br>Solution generated demonstrates deep insight and thoroughly addresses multiple factors of the problem.<br><br>Could be used as a model solution. | Logical scenario and value proposition is developed.<br><br>Generates a realistic, comprehensive and effective solution for the client.<br><br>Could be used as a model solution after minor corrections. | Appropriate and efficient strategies were chosen for solving the problem.<br><br>Solution addresses factors of the problem in a proficient manner.<br><br>Could be used as a model solution after making few corrections. | Basic scenario and value proposition is developed. Most are similar.<br><br>Solution addresses the problem but overlooks some relevant factors. | Limited problem solving evidenced.<br><br>Solution for client is superficial and does not directly address the problem and/or solution is incorrect. | Problem is not clearly articulated, and no methodology is evident in addressing the problem.<br><br>Needs assistance to generate simple solutions. | No evidence of addressing this criterion. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Written Communication**<br><br>*Technical Report Requirements*<br><br>(40%) | The technical report demonstrates a professional standard of written communication skills in all aspects. Effective use of discipline-based terminology and notation and attends to all report requirements outlined.<br><br>Selection of supporting graphical imagery contributes significantly to the overall professional standard of the report. | The technical report demonstrates a high standard of written communication skills in all aspects. Discipline-based terminology and notation is used effectively within the report format.<br><br>Selection of supporting graphical imagery contributes significantly to the standard of the report. | The technical report demonstrates a competent standard of written communication skills.<br><br>Discipline-based terminology and notation are utilized within the technical report format outlined.<br><br>Selection of supporting graphical imagery contributes to the standard of the report. | The technical report meets a satisfactory standard of written communication skills.<br><br>Uses language that conveys meaning with some clarity.<br><br>Use of graphical imagery complements the report. | The technical report demonstrates an unsatisfactory standard of written communication skills.<br><br>Information presented lacks organisation resulting in an unclear central message. | The technical report submitted is incomplete.<br><br>Information presented is unstructured and disconnected to the report requirements outlined. | No evidence of addressing this criterion. |