**Coursework Overview and Assessment Criteria**

**Module Title:** Big Data & Infrastructure

**Module Code:** COM 745

**Module Coordinator:** Dr Joseph Rafferty

**QAHE Teaching Staff Responsible:** Dr Adediji Fakoya

**Semester (s) Taught:** Three - QAHE

**Course / Year Group**: MSc Internet of Things/Computer Science

**Coursework / Exam Weighting**: 100/0

**Coursework Assessment Overview**

# This module is assessed by two pieces of coursework.

# Coursework 1 consists of a single in class examination which will have a time limit of 60 minutes. Coursework 1 contributes to 25% of the overall mark for this module.

# Coursework 2 is a practical skills assessment wherein students need to develop a solution and create a related presentation plus demonstrative video. Coursework 2 contributes to 75% of the overall mark for this module.

The university has a number of rules and regulations surrounding assessment, late submissions and illness. These are in the student guide [1] - ensure you read this and understand the impact of these rules and regulations.

These coursework assignments are detailed below.

Note: Students who submit coursework are declaring the following.

*“I declare that this is all my own work. Any material I have referred to has been accurately referenced and any contribution of Artificial Intelligence technology has been fully acknowledged. I have read the University’s policy on academic misconduct and understand the different forms of academic misconduct. If it is shown that material has been falsified, plagiarised, or I have otherwise attempted to obtain an unfair advantage for myself or others, I understand that I may face sanctions in accordance with the policies and procedures of the University. A mark of zero may be awarded and the reason for that mark will be recorded on my file.”*

*Also note:*

*You will receive feedback as per University Guidance which is currently set at 20 working days after submission.*

**Coursework 1 – Practical Skills Assessment [25%]**

# *Occurs:*

# Week commencing 21/07/2025, depending on your group.

# *Feedback Date:*

**27/08/2025**

**Related learning outcomes:**

1. Demonstrate a comprehensive understanding of what is meant by big data and how a variety of database/data storage paradigms may be applied to address the challenges it presents.

During the delivery course of the module, students will be expected to complete a 60-minute, online test. This test will assess understanding of concepts which have been introduced and detailed until that point.

This exam will be set in the in the middle of the semester and will incorporate the following topics:

* General Database Concepts
* Relational Databases
* NoSQL Concepts
* Document Databases
* Timeseries Databases
* Graph Databases

Coursework 1 will be delivered, submitted and assessed through the Blackboard online learning environment.

**This is a closed book examination and will occur ON CAMPUS in UU Belfast.**

*N.B. unforeseen circumstances, such as inclement weather, may require rescheduling of this exam. In this eventuality, students and the course director will be consulted.*

**Coursework 2 – a set exercise [75%]**

# *Released:*

**20/06/2025**

*Submission Deadline:*

**27/08/2025 at Noon.**

*Feedback Date:*

**26/09/2025**

**Related Learning Outcomes**

1. Appraise the concepts behind a range of database/data storage paradigms and critically evaluate when to apply these paradigms to big data problems.

1. Autonomously and independently investigate deficiencies when interacting with a range of technologies and leveraging knowledge of these deficiencies to improve future practice.
2. Examine, select and autonomously apply skills to leverage data stored in a range of database/data storage paradigms.

The exercise will assess understanding of further concepts and demonstrate practical skills related to a data lake type environment, as taught within the module (such as **Hadoop**, Amazon EMR or Azure Data Lakes).

Students will be set an exercise where they will be expected to:

1. Identify and evaluate a number of publicly available datasets related to **Social Deprivation and related deaths.**

These may be from open sources such as kaggle.com, data.gov.in, data.gov, data.stats.gov.cn, or data.gov.uk

1. Select appropriate datasets, as informed by their interests and the topic issued.
2. Integrate and import these datasets into a suitable data lake system, as covered within the module, while providing rationale for their choice.
3. Perform meaningful analysis of the data to derive some simple useful information, as can be obtained by the dataset selected.
4. Provide visualisation of the analysis through any data lake-associated technologies which the students deem suitable.

Once the solution is produced, students are required to produce presentation which incorporates a 5-minute [indicative] video capture demonstrating the solution (NOTE: The video is to demonstrate the solution. It is not to be a recording of the slides being presented).

**Note: Students will submit a PowerPoint Slide deck containing an embedded video demonstrating the solution.**

The presentation element, without video is worth 70% of this assessment and 52.5% of the overall module credit.

The video embedded in the presentation is worth 30% of this assessment and 22.5% of the overall module credit.

It is recommended to have 15 content slides which incorporates the below outline:

**Slide 0.** Title Slide. (0%)

**Slides 1 - 3.** Discussion of the problem and justification of the dataset (15%).

**Slides 4 - 8.** Overview of the technical solution developed (25%).

**Slides 9 - 12.** The analysis performed, and insight obtained (20%).

**Slide 13.** Functionality of the recorded demonstration (30%) [5-minute video].

**Slide 14.** Concluding comments (5%).

**Slide 15.** References (5%).

The assessment criteria for coursework 2 is presented as an appendix to this document.

***N.B. Students should be aware of the plagiarism policy of the University and submit their coursework in accordance to this.***

**References**

[1] “Ulster University Student Guide.” [Online]. Available: https://www.ulster.ac.uk/connect/guide.

[2] IEEE, “Manuscript Templates for Conference Proceedings.” [Online]. Available: https://www.ieee.org/conferences\_events/conferences/publishing/templates.html.

[3] IEEE, “IEEE Citation Reference.” [Online]. Available: https://www.ieee.org/documents/ieeecitationref.pdf.

[4] Mendeley Ltd, “Mendeley Citation Manager.” [Online]. Available: https://www.mendeley.com/.

**Appendix I – assessment criteria coursework 2 COM745 – assessment criteria coursework 2**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Criteria  (100%) | Fail  (0-49%) | Pass  (50-59%) | Commendation  (60-69%) | Distinction  (70-100%) |
| *Problem analysis/Selection of Dataset*  *(15%)* | The core problem is not discussed in a meaningful matter.  Datasets were selected with little or insufficient justification or reasoning.  The insight that analysis of this data could provide was not elaborated upon. | A problem area was faintly identified and discussed with sound justification.  Datasets were selected, justification may have been slight but clear  The avenue to insight that the datasets would provide was inadequately considered. | A problem area was correctly identified, discussed and evidenced in an; informed manner.  Datasets were appropriately selected and examined.  The insight that could be extracted from these datasets was reasoned upon and justified. | A problem area was correctly identified, discussed and evidenced; in an extensive manner.  Datasets were appropriately selected and examined. Excellent rationale was applied to their inclusion.  The insight that could be extracted from these datasets was reasoned upon in an effective manner. |
| *Solution produced*  *(25%)* | Justification for the choice of technology applied to the problem was very limited.  The solution was documented inadequately.  Scalability was not justified or examined, and alternative technologies were not explored. | The technology used to produce the solution was sound given the requirements of the scenario.  The solution was documented adequately incorporating of control flow diagrams.  The utility of the solution was clearly examined and justified. | The technology used to produce the solution was carefully examined and logically chosen in an informed manner – considering the requirements of the scenario.  Alternative technologies were examined and excluded accordingly.  The solution was documented well incorporating control flows and incorporation of software architecture diagrams.  Scalability of the solution was examined and justified. | The technology used to produce the solution was carefully examined and logically chosen in an excellent manner.  The approach considered the requirements of the scenario and scope beyond the stated immediate issue.  An extensive range of alternative technologies were examined and excluded/including appropriately.  The solution was documented in an exemplary manner; incorporating control flows, psueocode/script snippets and incorporation of software architecture diagrams.  Scalability of the solution was examined and justified. |
| *Analysis/Insight from data*  *(20%)* | Some limited analysis was attempted.  Outputs may have been weak or inadequate.  There may have been little or no justification of the analysis applied. | A clear approach was taken to analysis.  Analysis was performed across a large dataset (or a representative.  The analysis was adequate and some insight was provided. The analysis produced tabular output or basic graphs. | A logical and informed approach to analysis was taken.  Analysis was performed across a number of large datasets or a single appropriately massive dataset.  The analysis was complex in nature and provided nuanced insight. The analysis produced tabular output and simple graphical representations. | An excellent approach to analysis was taken.  Analysis was performed across a number of large datasets or a single appropriately massive dataset.  The analysis was complex in nature and provided nuanced insight. The analysis produced tabular output and advanced graphical Visualisation, such as maps. |
| *Concluding comments*  *(5%)* | Very limited reflection was applied to the solution and analysis. | Solid reflection was applied to the solution and analysis. | Thorough reflection was applied to the solution and analysis.  Weaknesses were identified and improvements were suggested. | Excellent reflection was applied to the solution, analysis and potential alternative approaches.  Weaknesses were identified and improvements were suggested with a . |
| *Referencing*  *(5%)* | Very limited referencing. | Inadequate or incorrect referencing. | Correct and appropriate referencing. | Excellent, correct and appropriate referencing was applied to a high standard. |
| *Video Demonstration: Insight offered into data*  *(10%)* | Limited metrics are produced – providing little or no insight. | Good metrics are produced - providing clear insight. | Very good metrics are produced in addition to incorporation of visualisation, providing informed insight. | Excellent metrics are produced in addition to incorporation of visualisation, providing excellent insight across a range of medium. |
| *Video Demonstration: Functionality*  *(20%)* | The solution functioned in a very limited fashions or didn’t operate at all. | The solution functioned moderately well.  Implementation issues may have been present but were deemed acceptable. | The solution performed well and had some minor implementation issues.  Advanced techniques or functionality, such as visualisation, has been incorporated. | The solution performed well and had minimal implementation issues.  Advanced techniques or functionality, such as interactive visualisation, has been incorporated. |