



COIT12209 Data Science

Term 3 - 2023

Profile information current as at 26/04/2024 01:37 am

All details in this unit profile for COIT12209 have been officially approved by CQUniversity and represent a learning partnership between the University and you (our student). The information will not be changed unless absolutely necessary and any change will be clearly indicated by an approved correction included in the profile.

General Information

Overview

This unit focuses on the foundational concepts of data science. Digital data is growing at a very fast rate with data being the underlying driver of the knowledge economy. This unit will prepare you with foundational knowledge and practical skills about data collection, representation, storage, retrieval, management, analysis, and visualisation through the exploration of data-related challenges. You will also learn the impact of big data and business analytics on business performance to cater for the development of useful information and knowledge in an attempt to achieve data-driven decision making.

Details

Career Level: *Undergraduate*

Unit Level: *Level 2*

Credit Points: 6

Student Contribution Band: 8

Fraction of Full-Time Student Load: 0.125

Pre-requisites or Co-requisites

Prerequisite: COIT11226 Systems Analysis Co-requisite: COIT11237 Database Design & Implementation

Important note: Students enrolled in a subsequent unit who failed their pre-requisite unit, should drop the subsequent unit before the census date or within 10 working days of Fail grade notification. Students who do not drop the unit in this timeframe cannot later drop the unit without academic and financial liability. See details in the [Assessment Policy and Procedure \(Higher Education Coursework\)](#).

Offerings For Term 3 - 2023

- Online

Attendance Requirements

All on-campus students are expected to attend scheduled classes - in some units, these classes are identified as a mandatory (pass/fail) component and attendance is compulsory. International students, on a student visa, must maintain a full time study load and meet both attendance and academic progress requirements in each study period (satisfactory attendance for International students is defined as maintaining at least an 80% attendance record).

Website

[This unit has a website, within the Moodle system, which is available two weeks before the start of term. It is important that you visit your Moodle site throughout the term. Please visit Moodle for more information.](#)

Class and Assessment Overview

Recommended Student Time Commitment

Each 6-credit Undergraduate unit at CQUniversity requires an overall time commitment of an average of 12.5 hours of study per week, making a total of 150 hours for the unit.

Class Timetable

[Regional Campuses](#)

Bundaberg, Cairns, Emerald, Gladstone, Mackay, Rockhampton, Townsville

[Metropolitan Campuses](#)

Adelaide, Brisbane, Melbourne, Perth, Sydney

Assessment Overview

1. **Practical Assessment**

Weighting: 40%

2. **Written Assessment**

Weighting: 40%

3. **Presentation**

Weighting: 20%

Assessment Grading

This is a graded unit: your overall grade will be calculated from the marks or grades for each assessment task, based on the relative weightings shown in the table above. You must obtain an overall mark for the unit of at least 50%, or an overall grade of 'pass' in order to pass the unit. If any 'pass/fail' tasks are shown in the table above they must also be completed successfully ('pass' grade). You must also meet any minimum mark requirements specified for a particular assessment task, as detailed in the 'assessment task' section (note that in some instances, the minimum mark for a task may be greater than 50%). Consult the [University's Grades and Results Policy](#) for more details of interim results and final grades.

CQUniversity Policies

All University policies are available on the [CQUniversity Policy site](#).

You may wish to view these policies:

- Grades and Results Policy
- Assessment Policy and Procedure (Higher Education Coursework)
- Review of Grade Procedure
- Student Academic Integrity Policy and Procedure
- Monitoring Academic Progress (MAP) Policy and Procedure – Domestic Students
- Monitoring Academic Progress (MAP) Policy and Procedure – International Students
- Student Refund and Credit Balance Policy and Procedure
- Student Feedback – Compliments and Complaints Policy and Procedure
- Information and Communications Technology Acceptable Use Policy and Procedure

This list is not an exhaustive list of all University policies. The full list of University policies are available on the [CQUniversity Policy site](#).

Previous Student Feedback

Feedback, Recommendations and Responses

Every unit is reviewed for enhancement each year. At the most recent review, the following staff and student feedback items were identified and recommendations were made.

Feedback from Unit Evaluation

Feedback

Practice materials on the R language were insufficient.

Recommendation

Provide additional materials on R language.

Feedback from Unit Evaluation

Feedback

Big data lecture materials (i.e., Hadoop) were not updated.

Recommendation

Review the lecture slides on big data, and revamp the lecture materials.

Unit Learning Outcomes

On successful completion of this unit, you will be able to:

1. Discuss and demonstrate data science foundational concepts
2. Investigate and evaluate applications for data storage, management, retrieval, and analysis and visualisation
3. Apply knowledge to process data for data driven decision making
4. Analyse and generate solutions to solve data-related challenges
5. Demonstrate the knowledge required in using data science skills to solve business problems.

Australian Computer Society (ACS) recognises the Skills Framework for the Information Age (SFIA). SFIA is in use in over 100 countries and provides a widely used and consistent definition of ICT skills. SFIA is increasingly being used when developing job descriptions and role profiles.

ACS members can use the tool MySFIA to build a skills profile at

<https://www.acs.org.au/professionalrecognition/mysfia-b2c.html>

This unit contributes to the following workplace skills as defined by SFIA. The SFIA code is included:

Data Management (DATM)

Business Analysis (BUAN)

Data Analysis (DTAN)

IT Operation (ITOP)

Alignment of Learning Outcomes, Assessment and Graduate Attributes



Alignment of Assessment Tasks to Learning Outcomes

Assessment Tasks	Learning Outcomes				
	1	2	3	4	5
1 - Practical Assessment - 40%	•	•			
2 - Written Assessment - 40%		•	•	•	•

Assessment Tasks	Learning Outcomes				
	1	2	3	4	5
3 - Presentation - 20%			•	•	•

Alignment of Graduate Attributes to Learning Outcomes

Graduate Attributes	Learning Outcomes				
	1	2	3	4	5
1 - Communication	•	•	•	•	•
2 - Problem Solving		•	•	•	•
3 - Critical Thinking	•	•	•	•	•
4 - Information Literacy			•		•
5 - Team Work				•	
6 - Information Technology Competence				•	•
7 - Cross Cultural Competence					•
8 - Ethical practice	•	•			
9 - Social Innovation			•		•
10 - Aboriginal and Torres Strait Islander Cultures					

Alignment of Assessment Tasks to Graduate Attributes

Assessment Tasks	Graduate Attributes									
	1	2	3	4	5	6	7	8	9	10
1 - Practical Assessment - 40%	•		•	•		•				
2 - Written Assessment - 40%	•	•	•	•		•		•	•	
3 - Presentation - 20%	•	•		•						

Textbooks and Resources

Textbooks

COIT12209

Supplementary

Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking
(2013)

Authors: Foster Provost and Tom Fawcett

O'Reilly Media

Binding: eBook

COIT12209

Supplementary

Introduction to Data Science

Edition: Creative Commons Attribution- Non Commercial-ShareAlike 3.0 license (2013)

Authors: Jeffrey Stanton & Jeffrey Saltz

Binding: eBook

COIT12209

Supplementary

Practical Data Science with Hadoop and Spark: Designing and Building Effective Analytics at Scale
(2016)

Authors: Mendelevitch, O, Stella, C & Eadline, D

Pearson

Upper Saddle River , NJ , USA

ISBN: 9780134024141

Binding: Paperback

COIT12209

Supplementary

R Programming for Data Science

(2015)

Authors: Roger D. Peng

Leanpub

Binding: eBook

[View textbooks at the CQUniversity Bookshop](#)

IT Resources

You will need access to the following IT resources:

- CQUniversity Student Email
- Internet
- Unit Website (Moodle)
- R statistical program available as a free download online (download R studio as well - also free). Compatible on Mac and PC.
- R Studio and R
- Zoom capacity (webcam and microphone) will be required for online students

Referencing Style

All submissions for this unit must use the referencing style: [Harvard \(author-date\)](#)

For further information, see the Assessment Tasks.

Teaching Contacts

Chetanpal Singh Unit Coordinator

c.singh2@cqu.edu.au

Schedule

Week 1 - 06 Nov 2023

Module/Topic	Chapter	Events and Submissions/Topic
Introduction to Data Science: What is data science; data domination; innovation from internet giants; data science history; data science in modern enterprises; soft skills of a data scientist; data science project life cycle; types of data; big data; how is big data different.	Chapter 1; Practical Data Science with Hadoop and Spark Designing and Building Effective Analytics at Scale Authors: O Mendelevitch; C Stella; and D Eadline. Chapter 1; Data Science Author: J Stanton	

Week 2 - 13 Nov 2023

Module/Topic	Chapter	Events and Submissions/Topic
Identifying Data Problems: From business problems to data mining tasks; data mining tasks; data collection; business use cases; sampling; data mining process.	Chapter 2; Practical Data Science with Hadoop and Spark Designing and Building Effective Analytics at Scale Authors: O Mendelevitch; C Stella; and D Eadline. Chapter 2; Data Science Author: J Stanton	

Week 3 - 20 Nov 2023

Module/Topic	Chapter	Events and Submissions/Topic
Hadoop and Data Science: Storage requirements; what is Hadoop; Hadoop's evolution; Hadoop tools for data science; spark, R for data science; R package.	Chapter 3; Practical Data Science with Hadoop and Spark Designing and Building Effective Analytics at Scale Authors: O Mendelevitch; C Stella; and D Eadline. Chapter 3, 4 and 8; Data Science Author: J Stanton	

Week 4 - 27 Nov 2023

Module/Topic	Chapter	Events and Submissions/Topic
Data Presentation: Understand different ways of summarizing data; choose the right table/ graph for the right data and audience; self explanatory graphics; attractive graphs and tables.	CRO on Moodle Chapter 5, 6 and 9 ; Data Science Author: J Stanton	

Vacation Week - 04 Dec 2023

Module/Topic	Chapter	Events and Submissions/Topic
Revise all previous lecture slides and tutorial works	Revise all previous lecture slides and tutorial works	

Week 5 - 11 Dec 2023

Module/Topic	Chapter	Events and Submissions/Topic
Data Analytics: Why analytics; different types of analytics; delivery methods for the operational users; holistic approach to expand enterprise analytics; value of integration and data quality to analytics.	Chapter 7; Data Science Author: J Stanton and supplementary readings	

Week 6 - 18 Dec 2023

Module/Topic	Chapter	Events and Submissions/Topic
--------------	---------	------------------------------

Exploratory Analysis: trend analysis;
Box plot; pairs plot; time series
decomposition; geographical analysis.

Supplementary Readings

**Assessment 1 - Practical
Assessment (40% weighting) -
Due week 6, Friday, 11:45 pm
AEDT**

Practical Assessment Due: Week 6
Friday (22 Dec 2023) 11:45 pm AEST

Vacation Week - 25 Dec 2023

Module/Topic	Chapter	Events and Submissions/Topic
Revise all previous lecture slides and tutorial works	Revise all previous lecture slides and tutorial works	

Week 7 - 01 Jan 2024

Module/Topic	Chapter	Events and Submissions/Topic
Data Discovery and Data Mining: Data driven decisions; enabling data driven innovations; knowledge discovery process; data cleaning; data integration; data selection; data transformation; knowledge based systems; data mining and its goals; data mining operation and process.	Chapter 10; Data Science Author: J Stanton and supplementary readings	

Week 8 - 08 Jan 2024

Module/Topic	Chapter	Events and Submissions/Topic
Analytic Algorithms: clustering analysis; regression analysis; classifier analysis; association analysis; cohort analysis; graph analysis; traverse pattern analysis.	Supplementary Readings	

Week 9 - 15 Jan 2024

Module/Topic	Chapter	Events and Submissions/Topic
Data Integration: Analytic data integration; challenges in data integration; technologies in data integration; data mapping; data staging; data extraction; data transformation; data loading; need for integration; data integration approaches.	Chapter 18; Data Science Author: J Stanton and supplementary readings	

Week 10 - 22 Jan 2024

Module/Topic	Chapter	Events and Submissions/Topic
Data Security and Privacy: protection of personal data; data collection and significant risks; challenges of big data for data protection; confidentiality; integrity; availability; middleware security concerns; built in database protection; privacy issues; data security and storage; identification and authentication.	Chapter CRO provided on Moodle and supplementary readings.	

Week 11 - 29 Jan 2024

Module/Topic	Chapter	Events and Submissions/Topic
--------------	---------	------------------------------

System Design: An overview of ML systems in the real world

Chapter Chapter 1; Designing Machine Learning Systems by Chip Huyen

Written Assessment (40% weighting) - Due week 11, Friday, 11:45 pm AEDT

Written Assessment Due: Week 11 Friday (2 Feb 2024) 11:59 pm AEST

Week 12 - 05 Feb 2024

Module/Topic	Chapter	Events and Submissions/Topic
Cloud Computing for Data Processing		Assessment 3 - Presentation (20% weighting) - Due week 12, Friday, 11:45 pm AEDT Presentation Due: Week 12 Friday (9 Feb 2024) 11:59 pm AEST

Exam Week - 12 Feb 2024

Module/Topic	Chapter	Events and Submissions/Topic
--------------	---------	------------------------------

Term Specific Information

Unit Coordinator:

Chetanpal Singh
Address: 120 Spencer St Melbourne VIC 3000
Email: c.singh2@cqu.edu.au

Assessment Tasks

1 Practical Assessment

Assessment Type

Practical Assessment

Task Description

This assessment is designed to reinforce the contents taught in Week 1 to Week 5. This assessment relates to Learning Outcomes 1 and 2. This assessment is an individual assessment and should be submitted in Week 6. You will submit work on the data processing exercise using the R language. This will provide you with an opportunity to learn data storage, processing using R language. Each week you will be presented with a data-related challenge, and you will use computer tools to manipulate data to solve that challenge. This task will help to build your knowledge of data formats, retrieval, and analysis techniques. R language question details will be provided to you through Moodle in Week 2. This assessment contributes to 40% of the total marks.

Assessment Due Date

Week 6 Friday (22 Dec 2023) 11:45 pm AEST

Return Date to Students

Week 8 Friday (12 Jan 2024)
Within two weeks of submission

Weighting

40%

Assessment Criteria

Assessment 1 will be marked based on the following criteria.

Working R source code provided 12 marks
Submitted screen shot of all questions 12 marks
Analysis presented on the generated output 12 marks
Well-structured and coherent report 4 marks
Total 40 marks

Referencing Style

- [Harvard \(author-date\)](#)

Submission

Online

Submission Instructions

All files must be submitted to Moodle for marking by the due date.

Learning Outcomes Assessed

- Discuss and demonstrate data science foundational concepts
- Investigate and evaluate applications for data storage, management, retrieval, and analysis and visualisation

Graduate Attributes

- Communication
- Critical Thinking
- Information Literacy
- Information Technology Competence

2 Written Assessment

Assessment Type

Written Assessment

Task Description

This assessment is based on a case study to be provided to you in teaching Week 6. You are required to write a report of 2000 words. This is an individual assessment and contributes to Learning Outcomes 2, 3, 4 and 5. This report will follow a standard business report format. You will investigate and advise an organisation, whose details are given in the case study, on data storage, retrieval, and analysis mechanisms. You will also develop an analytic dashboard using R language for the organisation. The assignment will be marked out of a total of 100 marks and form 40% of the total marks for the unit.

Assessment Due Date

Week 11 Friday (2 Feb 2024) 11:59 pm AEST
Within two weeks of submission

Return Date to Students

Exam Week Friday (16 Feb 2024)
Feedback and marks for this assessment will be released after the certification date as this unit does not have an exam.

Weighting

40%

Assessment Criteria

Assessment 2 will be marked based on the following criteria:

Report formatting (font, header and footer, table of contents, numbering, referencing) 5 marks

Professional communication (correct spelling, grammar, formal business language used) 5 marks

Executive summary 10 marks
Report introduction 10 marks
Data collection and storage 15 marks
Data in action 25 marks
Model design and implementation 20 marks
Conclusion and recommendations 10 marks
Total 100 marks

Referencing Style

- [Harvard \(author-date\)](#)

Submission

Online

Submission Instructions

The assignment must be submitted to Moodle for marking by the due date.

Learning Outcomes Assessed

- Investigate and evaluate applications for data storage, management, retrieval, and analysis and visualisation
- Apply knowledge to process data for data driven decision making
- Analyse and generate solutions to solve data-related challenges
- Demonstrate the knowledge required in using data science skills to solve business problems.

Graduate Attributes

- Communication
- Problem Solving
- Critical Thinking
- Information Literacy
- Information Technology Competence
- Ethical practice
- Social Innovation

3 Presentation

Assessment Type

Presentation

Task Description

This assessment contributes to the Learning Outcomes 3, 4 and 5. This is an individual presentation. The presentation topic is based on your report from assignment 2 and learning outcomes from this unit. All students are required to deliver the presentation. Please ensure you adhere to appropriate and professional dress codes for your zoom/face-to-face presentation.

Assessment Due Date

Week 12 Friday (9 Feb 2024) 11:59 pm AEST
Presentation will be held on Week 12 lecture time

Return Date to Students

Exam Week Friday (16 Feb 2024)
Feedback and marks for this assessment will be released after the certification date as this unit does not have an exam.

Weighting

20%

Assessment Criteria

Assessment 3 will be marked based on the following criteria:
Stay on topic 3 marks
Fulfill requirements of topic 3 marks
Slide style 4 marks
Presentation style 5 marks

Valid information presented 5 marks
Total 20 marks

Referencing Style

- [Harvard \(author-date\)](#)

Submission

Online

Submission Instructions

The presentation file must be submitted to Moodle by the due date.

Learning Outcomes Assessed

- Apply knowledge to process data for data driven decision making
- Analyse and generate solutions to solve data-related challenges
- Demonstrate the knowledge required in using data science skills to solve business problems.

Graduate Attributes

- Communication
- Problem Solving
- Information Literacy

Academic Integrity Statement

As a CQUniversity student you are expected to act honestly in all aspects of your academic work.

Any assessable work undertaken or submitted for review or assessment must be your own work. Assessable work is any type of work you do to meet the assessment requirements in the unit, including draft work submitted for review and feedback and final work to be assessed.

When you use the ideas, words or data of others in your assessment, you must thoroughly and clearly acknowledge the source of this information by using the correct referencing style for your unit. Using others' work without proper acknowledgement may be considered a form of intellectual dishonesty.

Participating honestly, respectfully, responsibly, and fairly in your university study ensures the CQUniversity qualification you earn will be valued as a true indication of your individual academic achievement and will continue to receive the respect and recognition it deserves.

As a student, you are responsible for reading and following CQUniversity's policies, including the [Student Academic Integrity Policy and Procedure](#). This policy sets out CQUniversity's expectations of you to act with integrity, examples of academic integrity breaches to avoid, the processes used to address alleged breaches of academic integrity, and potential penalties.

What is a breach of academic integrity?

A breach of academic integrity includes but is not limited to plagiarism, self-plagiarism, collusion, cheating, contract cheating, and academic misconduct. The Student Academic Integrity Policy and Procedure defines what these terms mean and gives examples.

Why is academic integrity important?

A breach of academic integrity may result in one or more penalties, including suspension or even expulsion from the University. It can also have negative implications for student visas and future enrolment at CQUniversity or elsewhere. Students who engage in contract cheating also risk being blackmailed by contract cheating services.

Where can I get assistance?

For academic advice and guidance, the [Academic Learning Centre \(ALC\)](#) can support you in becoming confident in completing assessments with integrity and of high standard.

What can you do to act with integrity?



Be Honest

If your assessment task is done by someone else, it would be dishonest of you to claim it as your own



Seek Help

If you are not sure about how to cite or reference in essays, reports etc, then seek help from your lecturer, the library or the Academic Learning Centre (ALC)



Produce Original Work

Originality comes from your ability to read widely, think critically, and apply your gained knowledge to address a question or problem