



# COIT12209 *Data Science*

## Term 1 - 2019

Profile information current as at 21/05/2024 01:25 pm

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 1 - 2019

- Brisbane
- Melbourne
- Online
- Sydney

#### 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 Have your Say Moodle Evaluation

##### Feedback

Data science is an interesting unit. The subject itself was the best part about this course along with the R language work.

##### Recommendation

Keep providing the support in the lecture and tutorials. Continue with the good practices of teaching.

#### Feedback from Have your Say Moodle Evaluation

##### Feedback

The lecturer organised weekly Zoom sessions that the students were not able to join because of other units.

##### Recommendation

Continue posting the recordings of the sessions for the benefit of students who could not attend live Zoom sessions.

#### Feedback from Have your Say Moodle Evaluation

##### Feedback

Assessment 1 is related to weekly tutorials. The mapping of assessment 1 questions to weekly tutorials is not clear.

##### Recommendation

Provide clear guidelines in reference to Assignment 1 and how it is mapped and related to weekly tutorials. Learning Resources will be improved to add more clarification on assignments.

## 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%		•	•	•	•
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 2013 (2013)**

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

Authors: Jeffrey Stanton

Binding: eBook

COIT12209

#### Supplementary

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

Edition: latest (2013)

Authors: Foster Provost and Tom Fawcett

O'Reilly Media

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 Latest (2015)**

(2015)

Authors: Roger D Peng

Leanpub

Binding: eBook

#### Additional Textbook Information

INTRODUCTION TO DATA SCIENCE A PDF version of this book and code examples used in the book are available at: <http://jsresearch.net/wiki/projects/teachdatascie>

Paper copies of Mendelovitch are available at the CQUni Bookshop here: <http://bookshop.cqu.edu.au> (search on the Unit code)

### 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

## Referencing Style

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

For further information, see the Assessment Tasks.

## Teaching Contacts

**Meena Jha** Unit Coordinator  
[m.jha@cqu.edu.au](mailto:m.jha@cqu.edu.au)

## Schedule

### Week 1 - 11 Mar 2019

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 - 18 Mar 2019

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 - 25 Mar 2019

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 - 01 Apr 2019

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	

### Week 5 - 08 Apr 2019

Module/Topic	Chapter	Events and Submissions/Topic
Data Dissemination and Use: The purpose of dissemination; dissemination issues and concerns; strengths and weaknesses of different communication formats; components of dissemination plans.	Chapter 7; Data Science Author: J Stanton and supplementary readings	

### Vacation Week - 15 Apr 2019

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

**Week 6 - 22 Apr 2019**

Module/Topic	Chapter	Events and Submissions/Topic
Data Retrieval: What is information retrieval; information retrieval (IR) vs question; ideal information retrieval; relevant answers; how is IR accomplished; information acquisition process; what is search; how IR systems work; search engines; structure of an IR system.	Chapter 10; Data Science Author: J Stanton and supplementary readings	Practical Assessment due  <b>Practical Assessment</b> Due: Week 6 Monday (22 Apr 2019) 11:45 pm AEST

**Week 7 - 29 Apr 2019**

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 11, 12; Data Science Author: J Stanton and supplementary readings	

**Week 8 - 06 May 2019**

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 13, and 14; Data Science Author: J Stanton and supplementary readings	

**Week 9 - 13 May 2019**

Module/Topic	Chapter	Events and Submissions/Topic
Semantic Analysis: What is semantic analysis; context sensitive analysis; approaches to semantic analysis; applications of semantic analysis; comparison with artificial intelligence; strategies for semantic analysis; symbol table and type checking.	Chapter 15, and 16; Data Science Author: J Stanton and supplementary readings	

**Week 10 - 20 May 2019**

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 17; Data Science Author: J Stanton and supplementary readings	

**Week 11 - 27 May 2019**

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	Written Assessment due  <b>Written Assessment</b> Due: Week 11 Monday (27 May 2019) 11:00 pm AEST

## Week 12 - 03 Jun 2019

Module/Topic	Chapter	Events and Submissions/Topic
Cloud Computing for Data Processing: What is cloud; what is cloud computing; deployment models of cloud; advantages of cloud; characteristics of cloud	CRO provided on Moodle and supplementary readings.	Presentation due  <b>Presentation</b> Due: Week 12 Monday (3 June 2019) 11:45 pm AEST

## Review/Exam Week - 10 Jun 2019

Module/Topic	Chapter	Events and Submissions/Topic
No Final Exam for this Unit		

## Exam Week - 17 Jun 2019

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

## Term Specific Information

Contact information for Meena Jha: Email: [m.jha@cqu.edu.au](mailto:m.jha@cqu.edu.au) Office: Level 2, 400 Kent Street, Sydney Campus; P +61 2 9324 5776 | X 55776. Please submit questions about the unit through the 'Q&A' discussion forum in Moodle - that way, everyone can benefit from the questions and answers. If you have any individual queries, please email me and I'll try to get back to you within a day or so

## Assessment Tasks

### 1 Practical Assessment

#### Assessment Type

Practical Assessment

#### Task Description

This assessment is designed to reinforce the content 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 R language on data processing exercises. 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 will use computer tools to manipulate data to solve that challenge. This task will help to build your knowledge of data formats, and retrieval and analysis techniques. R language questions detail will be provided to you through Moodle in Week 1. This assessment contributes to 40% of the total marks.

#### Assessment Due Date

Week 6 Monday (22 Apr 2019) 11:45 pm AEST

All files must be submitted on Moodle.

#### Return Date to Students

Week 8 Monday (6 May 2019)

Markers will do their best to return feedback on assignment 1 on due date.

#### Weighting

40%

#### Assessment Criteria

The total weight is 40% of this assessment. 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

Report nicely written = 4 mark

Total 40 marks

#### Referencing Style

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



**Submission**

Online

**Submission Instructions**

All files must be submitted to Moodle for marking.

**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 provided to you in teaching week 4. You are required to write a report of 2000 words. This is an individual report and contributes to Learning Outcome 2, 3, 4 and 5. This individual report will follow a standard business report format. You will be investigating how you might advise an organisation whose details are given in a case study on data storage, retrieval, and analysis mechanisms. You will also be developing an analytic dashboard using language R for the organisation. You will submit your assignment to Moodle. The assignment will be marked out of a total of 100 marks and forms 40% of the total assessment for the unit.

**Assessment Due Date**

Week 11 Monday (27 May 2019) 11:00 pm AEST

This assignment is due on Monday week 11 at 11:00 PM

**Return Date to Students**

Exam Week Monday (17 June 2019)

This assessment feedback will be released after certification date as this unit does not have exam.

**Weighting**

40%

**Assessment Criteria**

Marking criteria for assessment 2

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

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

Executive summary 10

Report introduction 10

Data Collection and Storage 15

Data in Action 20

Business continuity 5

Dashboard Design 20

Conclusion and Recommendations 10

Total = 100

**Referencing Style**

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

**Submission**

Online

**Submission Instructions**

All assignments must be submitted to Moodle for marking.

**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 Outcome 3, 4 and 5. This is an individual presentation and an extension of assessment 2. All students are required to give presentation. You are required to present the recommendations, dashboard design, and the outcome of the case study you have investigated and the advice you would like to give the organisation whose details are given in a case study on data storage, retrieval, and analysis mechanisms.

**For DISTANCE Students Only:** Distance students will have ZOOM presentation. The details of ZOOM will be provided to all distance students in week 10. The unit Coordinator will conduct this presentation.

### Assessment Due Date

Week 12 Monday (3 June 2019) 11:45 pm AEST

### Return Date to Students

Exam Week Monday (17 June 2019)

The feedback will be released after the certification date as this unit does not have exam.

### Weighting

20%

### Assessment Criteria

Marking Criteria for presentation:

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

### Referencing Style

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

### Submission

Online

### Submission Instructions

All ppt slides must be submitted to Moodle.

### 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