COMP6210

Big Data

Session 2, In person-scheduled-weekday, North Ryde 2026

School of Computing

Contents

General Information

Unit convenor and teaching staff
Unit Convener and Lecturer
Yan Wang
yan.wang@mq.edu.au

Lecturer
Qingkai Xu
qiongkai.xu@mq.edu.au

Credit points
10

Prerequisites
COMP6200 or Admission to MInfoTechNetworking

Corequisites

Co-badged status
COMP3210

Unit description
Even simple tasks like counting elements can seem impossible when the amount of data to process is huge. This unit explores some of the key aspects related to processing and mining information from large volumes of data. We present technology commonly used in industry such as map-reduce, and show how a range of data processing methods can be realised using map-reduce. Especial emphasis will be placed in the adaptation of data mining techniques for large volumes of data and for data streaming.

Learning in this unit enhances student understanding of global challenges identified by the United Nations Sustainable Development Goals (UNSDGs) Industry, Innovation and Infrastructure

Important Academic Dates

Information about important academic dates including deadlines for withdrawing from units are available at https://www.mq.edu.au/study/calendar-of-dates

Learning Outcomes

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

ULO1: Explain the key Big Data concepts and techniques.

ULO2: Apply techniques for storing large volumes of data.

ULO3: Apply Map-reduce techniques to a number of problems that involve Big Data.

ULO4: Apply techniques for handling high-dimensional big data.

General Assessment Information

Requirements to Pass this Unit

To pass this unit you must:

- Achieve a total mark equal to or greater than 50%.

Release Dates

- Assessment 1 – To be released no later than week 4
- Assessment 2 – To be released no later than week 8

Late submission

Unless a Special Consideration request has been submitted and approved, a 5% penalty (of the total possible mark of the task) will be applied for each day a written report or presentation assessment is not submitted, up until the 7th day (including weekends). After the 7th day, a grade of '0' will be awarded even if the assessment is submitted. The submission time for all uploaded assessments is **11:55 pm**. A 1-hour grace period will be provided to students who experience a technical concern. The late submission rule was changed to align with the new Faculty policy.

For any late submission of time-sensitive tasks, such as scheduled tests/exams, performance assessments/presentations, and/or scheduled practical assessments/labs, please apply for Special Consideration. For example, if the assignment is worth 8 marks (of the entire unit) and your submission is late by 19 hours (or 23 hours 59 minutes 59 seconds), 0.4 marks (5% of 8 marks) will be deducted. If your submission is late by 24 hours (or 47 hours 59 minutes 59 seconds), 0.8 marks (10% of 8 marks) will be deducted, and so on.

Assessments where Late Submissions will be accepted

- Assessment 1 – YES, Standard Late Penalty applies
- Assessment 2 – YES, Standard Late Penalty applies

Special Consideration

The Special Consideration Policy aims to support students who have been impacted by short-term circumstances or events that are serious, unavoidable, and significantly disruptive, and which may affect their performance in assessment. If you experience circumstances or events that affect your ability to complete the assessments in this unit on time, please inform the convenor and submit a Special Consideration request through http://connect.mq.edu.au/.

Details for each assignment will be available via iLearn

You are encouraged to:

- Set your personal deadline earlier than the actual one
- Keep backups of all your important files
- Ensure that no-one else picks up your printouts

Generative AI

- **Open** These assessments do not restrict the application of AI tools. However, it is expected that the use of AI is acknowledged in an appropriate manner, similar to the expectation to provide references for other resources or tools that are used in producing your work. Any use of AI should comply with the University's Responsible and Ethical Use of Artificial Intelligence Policy.
- **Observed** These tasks are either fully or partly observed or invigilated. AI use will either not be permitted or will be restricted during the assessment. For more information, please see https://students.mq.edu.au/study/assessment-exams/assessments/ai-and-assessment

Assessment Tasks

Name	Weighting	Hurdle	Due	Groupwork/Individual	Short Extension	AI Approach
Assignment 1	30%	No	26/09/2026	Individual	Yes	Open
Assignment 2	40%	No	07/11/2026	Individual and Group	No	Open
Final exam	30%	No	TBA	Individual	No	Observed

Assignment 1

Assessment Type [1]: Experiential task
Indicative Time on Task [2]: 30 hours
Due: **26/09/2026**
Weighting: **30%**
Groupwork/Individual: Individual
Short extension [3]: Yes
AI Approach: Open

In this assignment, you need to implement and discuss scalable algorithms (e.g., using Hadoop and MapReduce) to process Big Data.

On successful completion you will be able to:

- Explain the key Big Data concepts and techniques.
- Apply techniques for storing large volumes of data.
- Apply Map-reduce techniques to a number of problems that involve Big Data.

Assignment 2

Assessment Type [1]: Experiential task
Indicative Time on Task [2]: 30 hours
Due: **07/11/2026**
Weighting: **40%**
Groupwork/Individual: Individual and Group
Short extension [3]: No
AI Approach: Open

In this group assignment, you will write a project plan, implement big data processing algorithms, and present the results of your work.

On successful completion you will be able to:

- Explain the key Big Data concepts and techniques.
- Apply techniques for storing large volumes of data.
- Apply Map-reduce techniques to a number of problems that involve Big Data.
- Apply techniques for handling high-dimensional big data.

Final exam

Assessment Type [1]: Examination
Indicative Time on Task [2]: 15 hours
Due: **TBA**
Weighting: **30%**
Groupwork/Individual: Individual
Short extension [3]: No
AI Approach: Observed

You will undertake a final examination during the formal examination period.

On successful completion you will be able to:

- Explain the key Big Data concepts and techniques.

---

[1] If you need help with your assignment, please contact:

- the academic teaching staff in your unit for guidance in understanding or completing this type of assessment
- Academic Success for academic skills support.

[2] Indicative time-on-task is an estimate of the time required for completion of the assessment task and is subject to individual variation.

[3] An automatic short extension is available for some assessments. Apply through the Service Connect Portal.

Delivery and Resources

Classes

Basically, each week has two hours of lectures and two hours of workshops. Lectures will start in **week 1**. Workshops will start in **week 2**. For details of days, times and rooms consult the timetables webpage.

Required and Recommended Texts

Some of the contents of the unit will be based on the following books:

- J. Leskovec, A. Rajaraman, J. Ullman, Mining of Massive Datasets. The book is free and available from http://www.mmds.org/, where you can also find links to a MOOC, slides, and videos.
- C.Coronel, S. Morris. Database Systems: Design, Implementation and Management. 13th edition. Chapter 14 is the most relevant chapter. This chapter will be made available to students attending the classes.

Additional material including lecture notes will be made available during the semester. See the unit schedule for a listing of the most relevant reading for each week.

Technology Used and Required

The following software is used in COMP6210:

- Java 8
  - Download: https://www.java.com/en/download/
  - Installation instructions to set JAVA_HOME:
    - https://www.java.com/en/download/help/download_options.xml
    - https://docs.oracle.com/cd/E19182-01/820-7851/inst_cli_jdk_javahome_t/
- Python 3.9 (Anaconda version)
  - Download: https://www.anaconda.com/download
  - Installation instructions: https://docs.anaconda.com/anaconda/install/
- MongoDB
  - Installation instructions: https://docs.mongodb.com/v3.2/tutorial/install-mongodb-on-windows/

- Studio 3T
    - Here is an online tool to access MongoDB and MapReduce. It has a 30 day Trial but if you need more time you can also apply for a student license.
    - Download: https://studio3t.com/download/
- Hadoop
    - Download: https://hadoop.apache.org/releases.html
    - Installation instructions: https://wiki.apache.org/hadoop/Hadoop2OnWindows
- DASK
    - Install: https://docs.dask.org/en/stable/install.html
    - Tutorial: https://tutorial.dask.org/00_overview.html

This software is installed in the labs; you should also ensure that you have working copies of all the above on your own machine. Note that some of this software requires internet access.

Many packages come in various versions; to avoid potential incompatibilities, you should install versions as close as possible to those used in the labs.

Unit Web Page

The unit web page will be hosted in iLearn, where you will need to log in using your Student One ID and password. The unit will make extensive use of discussion boards also hosted in iLearn. Please post questions there, they will be monitored by the staff on the unit.

Methods of Communication

We will communicate with you via your university email and through announcements on iLearn. Queries to convenors can either be placed on the iLearn discussion board or sent to the unit convenor via the contact email on iLearn.

COVID Information

For the latest information on the University's response to COVID-19, please refer to the Coronavirus infection page on the Macquarie website: https://www.mq.edu.au/about/coronavirus-faqs. Remember to check this page regularly in case the information and requirements change during the semester. If there are any changes to this unit in relation to COVID, these will be communicated via iLearn.

Unit Schedule

Unit Schedule

Week 1: Data and Big Data

Week 2: Organizing Big Data

Week 3: Curating Big Data

Week 4: Processing Big Data (Cloud Computing)

Week 5: Processing Big Data (MapReduce)

Week 6: Big Data Platforms (Guest Lecture)

Week 7: Big Data with High Dimensions

Week 8: Indexing Big Data

Week 9: Searching Big Data

Week 10: Multidimensional Divide and Conquer

Week 11: Grid Decomposition in Big Data

Week 12: Advanced Topic in Big Data (Guest Lecture)

Week 13: Unit Review

Policies and Procedures

Macquarie University policies and procedures are accessible from Policy Central (https://policies.mq.edu.au). Students should be aware of the following policies in particular with regard to Learning and Teaching:

- Academic Appeals Policy
- Academic Integrity Policy
- Academic Progression Policy
- Assessment Policy
- Fitness to Practice Procedure
- Assessment Procedure
- Complaints Resolution Procedure for Students and Members of the Public
- Special Consideration Policy

Students seeking more policy resources can visit Student Policies (https://students.mq.edu.au/support/study/policies). It is your one-stop-shop for the key policies you need to know about throughout your undergraduate student journey.

To find other policies relating to Teaching and Learning, visit Policy Central (https://policies.mq.edu.au) and use the search tool.

Student Code of Conduct

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: https://students.mq.edu.au/admin/other-resources/student-conduct

Results

Results published on platform other than eStudent, (eg. iLearn, Coursera etc.) or released directly by your Unit Convenor, are not confirmed as they are subject to final approval by the University. Once approved, final results will be sent to your student email address and will be made available in eStudent. For more information visit connect.mq.edu.au or if you are a Global

MBA student contact globalmba.support@mq.edu.au

Academic Integrity

At Macquarie, we believe academic integrity – honesty, respect, trust, responsibility, fairness and courage – is at the core of learning, teaching and research. We recognise that meeting the expectations required to complete your assessments can be challenging. So, we offer you a range of resources and services to help you reach your potential, including free online writing and maths support, academic skills development and wellbeing consultations.

Student Support

Macquarie University provides a range of support services for students. For details, visit http://students.mq.edu.au/support/

Academic Success

Academic Success provides resources to develop your English language proficiency, academic writing, and communication skills.

- Ask a Study Coach
- Access StudyWISE
- Upload an assignment to Studiosity
- Complete the Academic Integrity Module

The Library provides online and face to face support to help you find and use relevant information resources.

- Subject and Research Guides
- Ask a Librarian

Student Services and Support

Macquarie University offers a range of Student Support Services including:

- IT Support
- Accessibility and disability support with study
- Mental health support
- Safety support to respond to bullying, harassment, sexual harassment and sexual assault
- Social support including information about finances, tenancy and legal issues
- Student Advocacy provides independent advice on MQ policies, procedures, and processes

Student Enquiries

Got a question? Ask us via the Service Connect Portal, or contact Service Connect.

IT Help

For help with University computer systems and technology, visit https://students.mq.edu.au/support/technology/service-desk.

When using the University's IT, you must adhere to the Acceptable Use of IT Resources Policy. The policy applies to all who connect to the MQ network including students.

Changes from Previous Offering

We value student feedback to be able to continually improve the way we offer our units. As such we encourage students to provide constructive feedback via student surveys, to the teaching staff directly, or via the FSE Student Experience & Feedback link in the iLearn page.

Student feedback from the previous offering of this unit was very positive overall, with students pleased with the clarity around assessment requirements and the level of support from teaching staff. As such, no change to the delivery of the unit is planned, however we will continue to strive to improve the level of support and the level of student engagement.

Changes since First Published

Date	Description
27/07/2026	added 'TBA' for exam time; adjusted the order of three assessments

Unit information based on version 2026.02 of the Handbook