



# COMP8210

## Big Data Technologies

Session 2, Special circumstances 2021

*School of Computing*

### Contents

---

<a href="#"><u>General Information</u></a>	2
<a href="#"><u>Learning Outcomes</u></a>	2
<a href="#"><u>General Assessment Information</u></a>	3
<a href="#"><u>Assessment Tasks</u></a>	3
<a href="#"><u>Delivery and Resources</u></a>	6
<a href="#"><u>Unit Schedule</u></a>	7
<a href="#"><u>Policies and Procedures</u></a>	8
<a href="#"><u>Changes from Previous Offering</u></a>	9

#### **Disclaimer**

Macquarie University has taken all reasonable measures to ensure the information in this publication is accurate and up-to-date. However, the information may change or become out-dated as a result of change in University policies, procedures or rules. The University reserves the right to make changes to any information in this publication without notice. Users of this publication are advised to check the website version of this publication [or the relevant faculty or department] before acting on any information in this publication.

#### **Session 2 Learning and Teaching Update**

The decision has been made to conduct study online for the remainder of Session 2 for all units WITHOUT mandatory on-campus learning activities. Exams for Session 2 will also be online where possible to do so.

This is due to the extension of the lockdown orders and to provide certainty around arrangements for the remainder of Session 2. We hope to return to campus beyond Session 2 as soon as it is safe and appropriate to do so.

Some classes/teaching activities cannot be moved online and must be taught on campus. You should already know if you are in one of these classes/teaching activities and your unit convenor will provide you with more information via iLearn. If you want to confirm, see the list of [units with mandatory on-campus classes/teaching activities](#).

Visit the [MQ COVID-19 information page](#) for more detail.

## General Information

### Unit convenor and teaching staff

Convener

Amin Beheshti

[amin.beheshti@mq.edu.au](mailto:amin.beheshti@mq.edu.au)

Lecturer

Diego Molla-Aliod

[diego.molla-aliod@mq.edu.au](mailto:diego.molla-aliod@mq.edu.au)

Tutor

Urvashi Khanna

[urvashi.khanna@mq.edu.au](mailto:urvashi.khanna@mq.edu.au)

Tutor

Sam Khadivizand

[sam.khadivizand@mq.edu.au](mailto:sam.khadivizand@mq.edu.au)

Credit points

10

Prerequisites

COMP6210

Corequisites

Co-badged status

Unit description

This unit introduces students to the specialised technologies required for big data applications in business, organisations and scientific research. It covers specialised methods for storing, manipulating, analysing and exploiting the ever-increasing amounts of data that are encountered in practical applications, and provides hands-on training in advanced topics such as distributed computing clusters and 'cloud computing'.

## 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:** Demonstrate a high level of technical competency in standard and advanced methods for big data technologies

**ULO2:** Describe the current status of and recognize future trends in big data technologies

**ULO3:** Reflect on the changes the big data technologies bring to businesses, organisations and society, and critically analyse future trends

**ULO4:** Demonstrate a competency with emerging big data technologies, applications and tools

**ULO5:** Communicate clearly and effectively

## General Assessment Information

### Important Academic Dates

Information about important academic dates including deadlines for withdrawing from units are available at <https://students.mq.edu.au/important-dates>

### General Assessment Information

All assignments will be submitted using iLearn. The results of all assignments will be available via iLearn.

### Late Submission

No extensions will be granted without an approved application for Special Consideration. There will be a deduction of 10% of the total available marks made from the total awarded mark for each 24 hour period or part thereof that the submission is late. For example, 25 hours late in submission for an assignment worth 10 marks – 20% penalty or 2 marks deducted from the total. No submission will be accepted after solutions have been posted.

The final exam is not a **hurdle assessment**.

The final mark of the unit will be obtained by summing the marks of all the assessment tasks for a total mark of 100. In order to pass the unit:

- The sum of all assessed tasks must be at least 50.

## Assessment Tasks

Name	Weighting	Hurdle	Due
<a href="#">Assignment 1 - Data Lakes</a>	10%	No	Week 4
<a href="#">Assignment 2 - Processing Data</a>	25%	No	Week 7
<a href="#">Assignment 3 - Data Analysis</a>	25%	No	Week 12
<a href="#">Final examination</a>	40%	No	Week 13

## Assignment 1 - Data Lakes

Assessment Type <sup>1</sup>: Practice-based task

Indicative Time on Task <sup>2</sup>: 10 hours

Due: **Week 4**

Weighting: **10%**

In this assignment you will explore the management of big data using data lake technology.

On successful completion you will be able to:

- Demonstrate a high level of technical competency in standard and advanced methods for big data technologies
- Describe the current status of and recognize future trends in big data technologies
- Demonstrate a competency with emerging big data technologies, applications and tools

## Assignment 2 - Processing Data

Assessment Type <sup>1</sup>: Practice-based task

Indicative Time on Task <sup>2</sup>: 20 hours

Due: **Week 7**

Weighting: **25%**

In this assignment you will apply techniques to index, search and process high-dimensional data.

On successful completion you will be able to:

- Demonstrate a high level of technical competency in standard and advanced methods for big data technologies
- Describe the current status of and recognize future trends in big data technologies
- Demonstrate a competency with emerging big data technologies, applications and tools

## Assignment 3 - Data Analysis

Assessment Type <sup>1</sup>: Practice-based task

Indicative Time on Task <sup>2</sup>: 20 hours

Due: **Week 12**

Weighting: **25%**

In this assignment you will perform analysis of Big Data.

On successful completion you will be able to:

- Demonstrate a high level of technical competency in standard and advanced methods for big data technologies
- Describe the current status of and recognize future trends in big data technologies
- Reflect on the changes the big data technologies bring to businesses, organisations and society, and critically analyse future trends
- Demonstrate a competency with emerging big data technologies, applications and tools
- Communicate clearly and effectively

## Final examination

Assessment Type <sup>1</sup>: Case study/analysis

Indicative Time on Task <sup>2</sup>: 25 hours

Due: **Week 13**

Weighting: **40%**

Problem Analysis Report, to assess students' understanding of the learning outcomes.

On successful completion you will be able to:

- Demonstrate a high level of technical competency in standard and advanced methods for big data technologies
- Describe the current status of and recognize future trends in big data technologies
- Reflect on the changes the big data technologies bring to businesses, organisations and society, and critically analyse future trends
- Demonstrate a competency with emerging big data technologies, applications and tools
- Communicate clearly and effectively

---

<sup>1</sup> 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
- the [Writing Centre](#) for academic skills support.

<sup>2</sup> Indicative time-on-task is an estimate of the time required for completion of the assessment

task and is subject to individual variation

## Delivery and Resources

For details of days, times and rooms consult the [timetables webpage](#).

### Required and Recommended Texts

Much 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.mmms.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 COMP336:

- Java 8
  - Download: <https://www.oracle.com/technetwork/java/javase/downloads/jre10-downloads-4417026.html>
  - Installation instructions to set JAVA\_HOME:
    - [https://www.java.com/en/download/help/download\\_options.xml](https://www.java.com/en/download/help/download_options.xml)
    - [https://docs.oracle.com/cd/E19182-01/820-7851/inst\\_cli\\_jdk\\_javahome\\_t/](https://docs.oracle.com/cd/E19182-01/820-7851/inst_cli_jdk_javahome_t/)
- Mongo DB
  - <https://docs.mongodb.com/manual/tutorial/>
- Apache Druid
  - <https://druid.apache.org/>
- Hadoop
  - Download: <https://hadoop.apache.org/releases.html>
  - Installation instructions: <https://wiki.apache.org/hadoop/Hadoop2OnWindows>
- Python 3.8 (Anaconda version)
  - Download: <https://www.anaconda.com/download>
- <https://studio3t.com/> Here is an online tool that includes MongoDB and MapReduce, it

has a 30 day Trial but if you need more time you can also apply for a student license.

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.

## Unit Schedule

	Lecture	Workshop/Practical
Week 01	Intro to Big Data	Data Lake
Week 02	Bringing data together for analysis (Data Lake)   Microsoft Data Lake	
Week 03	Azure ML	Azure Labs
Week 04	Visualisation of data/PowerBI	PowerBI
Week 05	Principles of Architecture/Azure Labs	Azure Labs
Week 06	Applications of Big Data Analytics	Knowledge Lake
Week 07	Analysing Big Data	Data Analytics
Week 08	Text Analytics	Text Analytics
Week 09	Text Analytics (II)	Text Analytics
Week 10	Visualising Big Data	Visual Analytics

---

Week 11 | Analysing Streaming Data

| Visual and Text Analytics

---

Week 12 | Big Data and Society

| Assignment Demonstration

---

Week 13 | Unit Review + Final Assessment

| NA

---

## Policies and Procedures

Macquarie University policies and procedures are accessible from [Policy Central \(https://policies.mq.edu.au\)](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](#)
- [Grade Appeal Policy](#)
- [Complaint Management Procedure for Students and Members of the Public](#)
- [Special Consideration Policy](#)

Students seeking more policy resources can visit [Student Policies \(https://students.mq.edu.au/support/study/policies\)](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 \(https://policies.mq.edu.au\)](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 [ask.mq.edu.au](https://ask.mq.edu.au) or if you are a Global MBA student contact [globalmba.support@mq.edu.au](mailto:globalmba.support@mq.edu.au)



## Student Support

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

### Learning Skills

Learning Skills ([mq.edu.au/learningskills](http://mq.edu.au/learningskills)) provides academic writing resources and study strategies to help you improve your marks and take control of your study.

- [Getting help with your assignment](#)
- [Workshops](#)
- [StudyWise](#)
- [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

Students with a disability are encouraged to contact the [Disability Service](#) who can provide appropriate help with any issues that arise during their studies.

## Student Enquiries

For all student enquiries, visit Student Connect at [ask.mq.edu.au](http://ask.mq.edu.au)

If you are a Global MBA student contact [globalmba.support@mq.edu.au](mailto:globalmba.support@mq.edu.au)

## IT Help

For help with University computer systems and technology, visit [http://www.mq.edu.au/about\\_us/offices\\_and\\_units/information\\_technology/help/](http://www.mq.edu.au/about_us/offices_and_units/information_technology/help/).

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

In this offering, we work closely with Microsoft Team to learn about Big Data Technologies used in Microsoft to Organize, Analyze and Visualize Big Data.