



COMP8210

Big Data Technologies

Session 2, In person-scheduled-weekday, North Ryde 2023

School of Computing

Contents

<u>General Information</u>	2
<u>Learning Outcomes</u>	2
<u>General Assessment Information</u>	3
<u>Assessment Tasks</u>	3
<u>Delivery and Resources</u>	5
<u>Unit Schedule</u>	7
<u>Policies and Procedures</u>	7
<u>Changes from Previous Offering</u>	9
<u>Changes since First Published</u>	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.

General Information

Unit convenor and teaching staff

Amin Beheshti

amin.beheshti@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, is 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 Assessment Submission Penalty

No late submissions will be accepted in this unit unless a Special Consideration is Submitted before the assessment submission deadline and Granted.

Assessment Tasks

Name	Weighting	Hurdle	Due
Assignment 1 - Data Lakes	10%	No	Week 3
Assignment 2 - Processing Data	25%	No	Week 7
Assignment 3 - Data Analysis	25%	No	Week 12
Problem Analysis Report	40%	No	Week 13

Assignment 1 - Data Lakes

Assessment Type ¹: Practice-based task

Indicative Time on Task ²: 10 hours

Due: **Week 3**

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 ¹: Practice-based task

Indicative Time on Task ²: 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 ¹: Practice-based task

Indicative Time on Task ²: 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

Problem Analysis Report

Assessment Type ¹: Case study/analysis

Indicative Time on Task ²: 25 hours

Due: **Week 13**

Weighting: **40%**

A report on a major problem analysis on Big Data Technologies.

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

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
- the [Writing Centre](#) 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

Delivery and Resources

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

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.

Required and Recommended Texts

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

- A. Beheshti, S. Ghodratnama, M. Elahi, H. Farhood, "[Social Data Analytics](#)", ISBN 978-1-032-19627-5, CRC Press, 2022
- J. Leskovec, A. Rajaraman, J. Ullman, Mining of Massive Datasets. The book is free and available from <http://www.mmids.org/>, where you can also find links to MOOCs, 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 listing 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://docs.oracle.com/cd/E19182-01/820-7851/inst_cli_jdk_javahome_t/
- Mongo DB
 - <https://docs.mongodb.com/manual/tutorial/>
- Neo4j
 - <https://neo4j.com/>
- 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 you have working copies of all the above on your 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 must 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

Week 01 | Intro to Big Data

Week 02 | Organizing Big Data - NoSQL Database (MongoDB)

Week 03 | Organizing Big Data - Graph Database I (Neo4j)

Week 04 | Organizing Big Data - Graph Database II (Neo4j)

Week 05 | Data Lake (Snowflake)

Week 06 | Data Lake (Databricks)

Week 07 | Intro to ML at Scale

Week 08 | Analytics I (Microsoft - PowerBI/Synapse)

Week 09 | Analytics II (Google - BigQuery)

Week 10 | Distributed ML (Apache -Spark)

Week 11 | MLOps (Microsoft - Azure DevOps)

Week 12 | AI (Google - Bard)

Week 13 | Exam/Report

Policies and Procedures

Macquarie University policies and procedures are accessible from [Policy Central \(https://policies.smq.edu.au\)](https://policies.smq.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 \(https://students.smq.edu.au/su\)](https://students.smq.edu.au/su)

[pport/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 ask.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/>

The Writing Centre

[The Writing Centre](#) provides resources to develop your English language proficiency, academic writing, and communication skills.

- [Workshops](#)
- [Chat with a WriteWISE peer writing leader](#)
- [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 [AskMQ](#), or contact [Service Connect](#).

IT Help

For help with University computer systems and technology, visit 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

We value student feedback to be able to improve the way we offer our units continually. 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 on 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 the 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
19/07/2023	Dear Gaurav, I updated the Delivery and Resources and Changes to the unit from the last offering, as you suggested. Best, Amin