



COMP7210

Big Data Technologies

Session 2, Special circumstance 2020

Department 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>	6
<u>Unit Schedule</u>	7
<u>Policies and Procedures</u>	7

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.

Notice

As part of [Phase 3 of our return to campus plan](#), most units will now run tutorials, seminars and other small group learning activities on campus for the second half-year, while keeping an online version available for those students unable to return or those who choose to continue their studies online.

To check the availability of face-to-face and online activities for your unit, please go to [timetable viewer](#). To check detailed information on unit assessments visit your unit's iLearn space or consult your unit convenor.

General Information

Unit convenor and teaching staff

Amin Beheshti

amin.beheshti@mq.edu.au

Diego Molla-Aliod

diego.molla-aliod@mq.edu.au

Credit points

10

Prerequisites

Admission to MRes

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

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

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

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**.

We will replace the final exam with a Problem Analysis Report, to assess students' understanding of the learning outcomes in the Big Data TEchnologies. It will be a Problem Analysis Report that will be made available to the students online on week 12, with the submission deadline in week 14. The final exam report should be submitted on iLearn (Turnitin), by the deadline.

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
Assignment 1 - Data Lakes	10%	No	Week 4
Assignment 2 - Processing Data	25%	No	Week 8
Assignment 3 - Data Analysis	25%	No	Week 12
Final examination	40%	No	Week 13-14

Assignment 1 - Data Lakes

Assessment Type ¹: Essay

Indicative Time on Task ²: 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 ¹: Essay

Indicative Time on Task ²: 20 hours

Due: **Week 8**

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: Essay

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

Final examination

Assessment Type ¹: Case study/analysis

Indicative Time on Task ²: 25 hours

Due: **Week 13-14**

Weighting: **40%**

We will replace the final exam with a Problem Analysis Report, to assess students' understanding of the learning outcomes in the Big Data Problems. The final exam will no longer be a Hurdle. It will be a Problem Analysis Report that will be made available to the students online on week 12, with the submission deadline in week 14. The final exam report should be submitted on iLearn (Turnitin), by the deadline.

On successful completion you will be able to:

- Demonstrate a high level of technical competency in standard and advanced methods for big data technologies
- Reflect on the changes the big data technologies bring to businesses, organisations and society, and critically analyse future trends
- Describe the current status of and recognize future trends in big data technologies
- 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](#).

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.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 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/>
- Apache Druid
 - <https://druid.apache.org/>
- Hadoop
 - Download: <https://hadoop.apache.org/releases.html>
 - Installation instructions: <https://wiki.apache.org/hadoop/Hadoop2OnWindows>
- Python 3.6 (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 licence.

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 login 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

1. Lecture: Introduction to Big Data Workshop: Relational and NoSQL DBs
2. Lecture: Organizing Big Data (NoSQL - Mongo DB) Workshop: Mongo DB
3. Lecture: Organizing Big Data (Apache Cassandra) Workshop: Mongo DB
4. Lecture: Apache Druid (cloud-native, stream-native) Workshop: Apache Druid
5. Lecture: Apache Druid (Analytics) Workshop: Apache Druid
6. IBM Big Data & AI Services Workshop: Assignment Demonstration

7. Lecture: Analysing Big Data Workshop: Data Analytics
8. Lecture: Text Analytics Workshop: Text Analytics
9. Lecture: Text Analytics (II) Workshop: Text Analytics
10. Lecture: Visualising Big Data Workshop: Visual Analytics
11. Lecture: Analysing Streaming Data Workshop: Visual and Text Analytics
12. Lecture: Big Data and Society Workshop: Assignment Demonstration

Policies and Procedures

Macquarie University policies and procedures are accessible from [Policy Central \(https://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central\)](https://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central). 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](#) (**Note:** *The Special Consideration Policy is effective from 4 December 2017 and replaces the Disruption to Studies Policy.*)

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

If you would like to see all the policies relevant to Learning and Teaching visit [Policy Central](http://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central) (<http://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central>).

Student Code of Conduct

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: <https://students.mq.edu.au/study/getting-started/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

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) 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

If you are a Global MBA student contact 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/.

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.