COMP6210

Big Data

Session 2, Special circumstances 2021

School of Computing

Contents

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
Unit convenor and lecturer
Yan Wang
yan.wang@mq.edu.au
Contact via +61-2-9850 9539
Room 354, BD Building
By Appointment

Tutor
Asim Adnan Eija
asim-adnan.eijaz@students.mq.edu.au

Yan Wang
yan.wang@mq.edu.au

Guanfeng Liu
guanfeng.liu@mq.edu.au

Credit points
10

Prerequisites
COMP6200 and Admission to MDataSc or MScInnovationIT or GradCertInfoTech or MBusAnalytics

Corequisites

Co-badged status

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.

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

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 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 raw mark needs to be 50 or above.

Assessment Tasks

Name	Weighting	Hurdle	Due
Assignment 1	20%	No	Week 7-8
Assignment 2	20%	No	Week 13
Final examination	60%	No	TBA

Assignment 1

Assessment Type [1]: Practice-based task
Indicative Time on Task [2]: 30 hours
Due: **Week 7-8**
Weighting: **20%**

In this assignment you will implement MapReduce techniques for the processing of Big Data. You will build your assignment on top of Hadoop.

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 techniques for handling high-dimensional big data.

Assignment 2

Assessment Type [1]: Practice-based task
Indicative Time on Task [2]: 30 hours
Due: **Week 13**
Weighting: **20%**

In this assignment you will implement a non-trivial problem that processes Big Data.

On successful completion you will be able to:

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

Final examination

Assessment Type [1]: Examination
Indicative Time on Task [2]: 15 hours
Due: **TBA**
Weighting: **60%**

The final exam will focus on the theoretical aspects of the unit, including algorithms and implementation issues.

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.

---

[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

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.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/
- Python 3.7 (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 licence.
    - Download: https://studio3t.com/download/
- Hadoop
    - Download: https://hadoop.apache.org/releases.html
    - Installation instructions: https://wiki.apache.org/hadoop/Hadoop2OnWindows

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

Note: Lectures will be online.

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
- 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). 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

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.

Changes from Previous Offering

Compared to Semester 1 2020, three assignments are reduced to two assignments. There is no hurdle any more.

Changes since First Published

Date	Description
13/07/2021	no change