COMP6210

Big Data

Session 1, Weekday attendance, North Ryde 2020

Dept of Computing

Contents

General Information 2
Learning Outcomes 2
General Assessment Information 3
Assessment Tasks 3
Delivery and Resources 5
Unit Schedule 6
Policies and Procedures 7
Changes from Previous Offering 8

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
Contact via +61 (2) 9850 6344
Room 365, BD Building
By Appointment

Guanfeng Liu
guanfeng.liu@mq.edu.au
Contact via +61-2-9850-9542
366, BD Building
By Appointment

Credit points
10

Prerequisites
Admission to MDataSc or MSclnnoationIT

Corequisites
COMP6200 or ITEC657

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://students.mq.edu.au/important-dates

Learning Outcomes

ULO1: Apply Map-reduce techniques to a number of problems that involve Big Data.

ULO2: Apply Big Data techniques to data mining.
ULO3: Explain the key Big Data concepts and techniques.
ULO4: Apply techniques for storing large volumes of 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 exam is a hurdle assessment. This means that:

- If the exam mark is between 24 and 30 (out of a maximum of 60), you will be given a second opportunity to sit at the exam.
- If the final exam mark is less than 30 out of 60 (after the second opportunity if given), you will fail the unit.

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.
- The final mark of the exam must be at least 30 out of 60.

Assessment Tasks

<table>
<thead>
<tr>
<th>Name</th>
<th>Weighting</th>
<th>Hurdle</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment 1</td>
<td>5%</td>
<td>No</td>
<td>Week 3</td>
</tr>
<tr>
<td>Assignment 2</td>
<td>20%</td>
<td>No</td>
<td>Week 8</td>
</tr>
<tr>
<td>Assignment 3</td>
<td>15%</td>
<td>No</td>
<td>Week 12</td>
</tr>
<tr>
<td>Final examination</td>
<td>60%</td>
<td>No</td>
<td>Examination period</td>
</tr>
</tbody>
</table>
Assignment 1
Assessment Type: Project
Indicative Time on Task: 10 hours
Due: Week 3
Weighting: 5%

In this assignment you will acquire hands-on experience in designing, implementing and querying a NoSQL database.

On successful completion you will be able to:
  • Apply techniques for storing large volumes of data.

Assignment 2
Assessment Type: Project
Indicative Time on Task: 25 hours
Due: Week 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:
  • Apply Map-reduce techniques to a number of problems that involve Big Data.
  • Apply Big Data techniques to data mining.
  • Apply techniques for storing large volumes of data.

Assignment 3
Assessment Type: Project
Indicative Time on Task: 25 hours
Due: Week 12
Weighting: 15%

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

On successful completion you will be able to:
  • Apply Big Data techniques to data mining.
  • Explain the key Big Data concepts and techniques.

Final examination
Assessment Type: Examination
Indicative Time on Task: 15 hours
Due: **Examination period**  
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:

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

---

1 If you need guidance or support to understand or complete this type of assessment, please contact the Learning Skills Team

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](https://unitguides.mq.edu.au/unit_offerings/122604/unit_guide/print).

### Required and Recommended Texts

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

- 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
Installation instructions to set JAVA_HOME:

- https://docs.oracle.com/cd/E19182-01/820-7851/inst_cli_jdk_javahome_t/

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

MongoDB 3.6.2

- Installation instructions: https://docs.mongodb.com/v3.2/tutorial/install-mongodb-on-windows/
- 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**

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 - AWS/Microsoft/IBM)
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 and Exam Review

**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). 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). 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 (https://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 improve your marks and take control of your study.

- Workshops
- StudyWise
- Academic Integrity Module for Students
- Ask a Learning Adviser

Student Enquiry Service

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

Equity 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.

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

The Big Data domain is advancing very fast. Accordingly, the content proposed in 2019 has been reviewed and updated for this offering. Particularly, we have offered new and trending topics in Big Data Platforms and Advanced Topic in Big Data.