COMP336

Big Data

S1 Day 2019

Dept of Computing

Contents

General Information

Unit convenor and teaching staff
Amin Beheshti
amin.beheshti@mq.edu.au

Jia Wu
jia.wu@mq.edu.au

Credit points
3

Prerequisites
39cp at 100 level or above including COMP257

Corequisites
ISYS358

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:

Explain the key Big Data concepts and techniques.

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.

General Assessment Information

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

via iLearn.

Late submission to the assignments will be penalised with the following deductions:

- **Assignment 1**: 1 mark per day late.
- **Assignment 2**: 4 marks per day late.
- **Assignment 3**: 3 marks per day late.

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

Name	Weighting	Hurdle	Due
Assignment 1	5%	No	Week 3
Assignment 2	20%	No	Week 8
Assignment 3	15%	No	Week 12
Final Exam	60%	Yes	Examination period

Assignment 1

Due: **Week 3**
Weighting: **5%**

Due: **Week 3** Weighting: **5%**

In this assignment you will acquire hands-on experience in designing, implementing and querying a NoSQL database, i.e. MongoDB. This Assessment Task relates to the following Learning Outcomes:

- Apply techniques for storing large volumes of data.

On successful completion you will be able to:

- Explain the key Big Data concepts and techniques.

- 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 2

Due: **Week 8**
Weighting: **20%**

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 (i.e. an open-source version of MapReduce written in Java).

This Assessment Task relates to the following Learning Outcomes:

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

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

Due: **Week 12**
Weighting: **15%**

Due: **Week 12** Weighting: **15%**

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

This Assessment Task relates to the following Learning Outcomes:

- Apply Map-reduce techniques to a number of problems that involve Big Data.
- Apply Big Data techniques to data mining.

On successful completion you will be able to:

- Apply Big Data techniques to data mining.
- Apply techniques for storing large volumes of data.

Final Exam

Due: **Examination period**
Weighting: **60%**
This is a hurdle assessment task (see assessment policy for more information on hurdle

assessment tasks)

Final Exam

Due: **Examination period** Weighting: **60% This is a hurdle assessment task (see assessment policy for more information on hurdle assessment tasks)**

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

This is a **hurdle assessment**. This means that you need to pass the exam in order to pass the unit.

This Assessment Task relates to the following Learning Outcomes:

- Explain the key Big Data concepts and techniques.
- 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.

On successful completion you will be able to:

- Explain the key Big Data concepts and techniques.
- 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.

Delivery and Resources

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/
- 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/

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-Part I)

Week 6 - Processing Big Data (MapReduce-Part II)

Week 7: Big Data Mining with High Dimensions

Week 8: Big Data Mining with Large Instances

Week 9: Deep Learning Model

Week 10: Fast Mining Models

Week 11: Handling Uncertainty

Week 12: Big Data Mining Applications

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.)*

Undergraduate 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 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.

Graduate Capabilities

Creative and Innovative

Our graduates will also be capable of creative thinking and of creating knowledge. They will be imaginative and open to experience and capable of innovation at work and in the community. We want them to be engaged in applying their critical, creative thinking.

This graduate capability is supported by:

Learning outcomes

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

Assessment tasks

- Assignment 1
- Assignment 2
- Assignment 3
- Final Exam

Discipline Specific Knowledge and Skills

Our graduates will take with them the intellectual development, depth and breadth of knowledge, scholarly understanding, and specific subject content in their chosen fields to make them competent and confident in their subject or profession. They will be able to demonstrate, where relevant, professional technical competence and meet professional standards. They will be able to articulate the structure of knowledge of their discipline, be able to adapt discipline-specific knowledge to novel situations, and be able to contribute from their discipline to inter-disciplinary solutions to problems.

This graduate capability is supported by:

Learning outcomes

- Explain the key Big Data concepts and techniques.
- Apply techniques for storing large volumes of data.

Assessment tasks

- Assignment 1
- Assignment 2
- Assignment 3
- Final Exam

Critical, Analytical and Integrative Thinking

We want our graduates to be capable of reasoning, questioning and analysing, and to integrate and synthesise learning and knowledge from a range of sources and environments; to be able to critique constraints, assumptions and limitations; to be able to think independently and systemically in relation to scholarly activity, in the workplace, and in the world. We want them to have a level of scientific and information technology literacy.

This graduate capability is supported by:

Learning outcomes

- Explain the key Big Data concepts and techniques.
- Apply Map-reduce techniques to a number of problems that involve Big Data.

Assessment tasks

- Assignment 1

- Assignment 2
- Final Exam

Problem Solving and Research Capability

Our graduates should be capable of researching; of analysing, and interpreting and assessing data and information in various forms; of drawing connections across fields of knowledge; and they should be able to relate their knowledge to complex situations at work or in the world, in order to diagnose and solve problems. We want them to have the confidence to take the initiative in doing so, within an awareness of their own limitations.

This graduate capability is supported by:

Learning outcomes

- Explain the key Big Data concepts and techniques.
- 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.

Assessment tasks

- Assignment 1
- Assignment 2
- Assignment 3
- Final Exam

Changes from Previous Offering

The Big Data domain is advancing very fast. Accordingly, the content proposed in 2018 has been reviewed and updated for this offering. Particularly, we have offered new and trending topics in:

- Big Data Mining with High Dimensions

- Deep Learning Model

- Big Data Mining Applications