COMP733

Big Data Technologies

S2 Day 2019

Dept of Computing

Contents

General Information 2

Learning Outcomes 3

General Assessment Information 3

Assessment Tasks 3

Delivery and Resources 5

Unit Schedule 6

Policies and Procedures 7

Graduate Capabilities 8

Changes from Previous Offering 11

Grading Standards 11

Changes since First Published 12

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
Convenor, lecturer
Diego Molla-Aliod
diego.molla-aliod@mq.edu.au
Contact via diego.molla-aliod@mq.edu.au
4 Research Park Drive, 358
See: http://web.science.mq.edu.au/~diego/

Lecturer
Guanfeng Liu
guanfeng.liu@mq.edu.au
Contact via guanfeng.liu@mq.edu.au
4 Research Park Drive, 366
See http://web.science.mq.edu.au/~gliu

Tutor
Urvashi Khanna
urvashi.khanna@mq.edu.au
See iLearn

Credit points
4

Prerequisites
Admission to MRes

Corequisites

Co-badged status
ITEC874

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:

Obtain a high level of technical competency in standard and advanced methods for big data technologies

Understand 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

Develop a competency with emerging big data technologies, applications and tools

Communicate clearly and effectively

General Assessment Information

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.

Supplementary Exam

If you receive Special Consideration for the final exam, a supplementary exam will be scheduled after the normal exam period, following the release of marks. By making a special consideration application for the final exam you are declaring yourself available for a resit during the supplementary examination period and will not be eligible for a second special consideration approval based on pre-existing commitments. Please ensure you are familiar with the policy prior to submitting an application. Approved applicants will receive an individual notification one week prior to the exam with the exact date and time of their supplementary examination.

Assessment Tasks

Name	Weighting	Hurdle	Due
Assignment 1 - Data Lakes	10%	No	Week 3
Assignment 2 - Processing Data	20%	No	Week 8
Assignment 3 - Data Analysis	20%	No	Week 12
Final examination	50%	No	Examination period

Assignment 1 - Data Lakes

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:

- Obtain a high level of technical competency in standard and advanced methods for big data technologies
- Understand the current status of and recognize future trends in big data technologies
- Develop a competency with emerging big data technologies, applications and tools

Assignment 2 - Processing Data

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

In this assignment you will apply techniques to index, search and process high-dimensional data.

On successful completion you will be able to:

- Obtain a high level of technical competency in standard and advanced methods for big data technologies
- Understand the current status of and recognize future trends in big data technologies
- Develop a competency with emerging big data technologies, applications and tools

Assignment 3 - Data Analysis

Due: **Week 12**
Weighting: **20%**

In this assignment you will perform analysis of Big Data.

On successful completion you will be able to:

- Obtain a high level of technical competency in standard and advanced methods for big data technologies
- Understand 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
- Develop a competency with emerging big data technologies, applications and tools
- Communicate clearly and effectively

Final examination

Due: **Examination period**
Weighting: **50%**

The final exam will focus on the theoretical aspects of the unit.

On successful completion you will be able to:

- Obtain a high level of technical competency in standard and advanced methods for big data technologies
- Understand 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
- Develop a competency with emerging big data technologies, applications and tools
- Communicate clearly and effectively

Delivery and Resources

Required and Recommended Texts

All required and recommended readings will be provided as part of the lecture material.

Technology Used and Required

The following software is used in ITEC874:

- Java 8
  - Download: http://download.oracle.com/otn-pub/java/jdk/8u162-b12/0da788060d494f5095bf8624735fa2f1/jdk-8u162-windows-x64.exe
  - 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/
- VirtualBox
  - Download and installation: https://www.virtualbox.org/wiki/Downloads
- Hadoop
  - VirtualBox image: https://downloads.cloudera.com/demo_vm/virtualbox/cloudera-quickstart-vm-5.13.0-0-virtualbox.zip download and set up as a VBox virtual machine
- Python 3.7 (Anaconda version)
  - Download: https://www.anaconda.com/download

- MongoDB 4.0 Community Edition
  - Installation instructions: https://docs.mongodb.com/manual/installation/
- Robomongo
  - Download and installation: https://robomongo.org/
- SAS Viya for Learners
  - Online system; registration instructions will be provided. https://www.sas.com/en_us/software/viya-for-learners.html

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

Below is a tentative unit schedule.

Week	Lecture	Workshop
1	Introduction to Big Data	MongoDB
2	Organising Big Data -- From Relational to NoSQL	Data Lake Services
3	Big Data and Society	Big Data and Society
4	Indexing Big Data -- R-Tree	R-Tree
5	Searching Big Data -- NN Search and Skyline	NN Search
6	Processing Big Data -- Divide and Conquer	Divide and Conquer
7	Industry Talk -- (TBA)	Processing High-Dimensional Data
	RECESS	
8	Analysing Big Data	Azure Machine Learning Studio
9	Text Analytics	Azure ML Studio; SAS Enterprise Miner
10	Visualising Big Data	SAS Visual Analytics; Tableau
11	Analysing Streaming Data	SAS Visual Analytics; Tableau
12	Industry Talk -- (TBA)	TBA

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

Special Consideration

If you receive special consideration for the final exam, a supplementary exam will be scheduled in the week of December 17-21 2018. By making a special consideration application for the final exam you are declaring yourself available for a resit during the supplementary examination period and will not be eligible for a second special consideration approval based on pre-existing commitments. Please ensure you are familiar with the policy prior to submitting an application. Approved applicants will receive an individual notification one week prior to the exam with the exact date and time of their supplementary examination.

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

PG - Capable of Professional and Personal Judgment and Initiative

Our postgraduates will demonstrate a high standard of discernment and common sense in their professional and personal judgment. They will have the ability to make informed choices and decisions that reflect both the nature of their professional work and their personal perspectives.

This graduate capability is supported by:

Learning outcomes

- Obtain a high level of technical competency in standard and advanced methods for big data technologies
- Understand 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
- Develop a competency with emerging big data technologies, applications and tools

Assessment tasks

- Assignment 1 - Data Lakes
- Assignment 2 - Processing Data
- Assignment 3 - Data Analysis
- Final examination

PG - Discipline Knowledge and Skills

Our postgraduates will be able to demonstrate a significantly enhanced depth and breadth of knowledge, scholarly understanding, and specific subject content knowledge in their chosen fields.

This graduate capability is supported by:

Learning outcomes

- Obtain a high level of technical competency in standard and advanced methods for big data technologies
- Understand 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
- Develop a competency with emerging big data technologies, applications and tools

Assessment tasks

- Assignment 1 - Data Lakes
- Assignment 2 - Processing Data
- Assignment 3 - Data Analysis
- Final examination

PG - Critical, Analytical and Integrative Thinking

Our postgraduates will be capable of utilising and reflecting on prior knowledge and experience, of applying higher level critical thinking skills, and of integrating and synthesising learning and knowledge from a range of sources and environments. A characteristic of this form of thinking is the generation of new, professionally oriented knowledge through personal or group-based critique of practice and theory.

This graduate capability is supported by:

Learning outcomes

- Understand 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
- Develop a competency with emerging big data technologies, applications and tools

Assessment tasks

- Assignment 1 - Data Lakes
- Assignment 2 - Processing Data
- Assignment 3 - Data Analysis
- Final examination

PG - Research and Problem Solving Capability

Our postgraduates will be capable of systematic enquiry; able to use research skills to create new knowledge that can be applied to real world issues, or contribute to a field of study or practice to enhance society. They will be capable of creative questioning, problem finding and problem solving.

This graduate capability is supported by:

Learning outcomes

- Understand 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
- Develop a competency with emerging big data technologies, applications and tools

Assessment tasks

- Assignment 1 - Data Lakes
- Assignment 2 - Processing Data
- Assignment 3 - Data Analysis
- Final examination

PG - Effective Communication

Our postgraduates will be able to communicate effectively and convey their views to different social, cultural, and professional audiences. They will be able to use a variety of technologically supported media to communicate with empathy using a range of written, spoken or visual formats.

This graduate capability is supported by:

Learning outcomes

- Reflect on the changes the big data technologies bring to businesses, organisations and society, and critically analyse future trends
- Communicate clearly and effectively

Assessment tasks

- Assignment 3 - Data Analysis
- Final examination

PG - Engaged and Responsible, Active and Ethical Citizens

Our postgraduates will be ethically aware and capable of confident transformative action in relation to their professional responsibilities and the wider community. They will have a sense of connectedness with others and country and have a sense of mutual obligation. They will be able to appreciate the impact of their professional roles for social justice and inclusion related to national and global issues

This graduate capability is supported by:

Learning outcome

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

Assessment task

- Final examination

Changes from Previous Offering

There are no significant changes from the previous offering.

Grading Standards

This unit does not have hurdle assessments. The final mark of the unit will be obtained by summing the marks of all the assessment tasks for a total mark of 100. The final grade will be determined based on the final mark according to the thresholds established by Macquarie University. As per July 2019, the thresholds are:

- **High Distinction: 85 - 100** You are an exceptional student who shows consistent evidence of deep and critical understanding in relation to the learning outcomes. Assignment solutions are outstanding and they show substantial originality and insight. You provide a critical evaluation of problems, their solutions and their implications.
- **Distinction: 75 - 84** There is evidence of learning that goes beyond replication of content knowledge or skills relevant to the learning outcomes. You are able to apply the technologies covered to new problems or in new ways. Assignment solutions are well

presented, implementing extended features or displaying high quality work. You show good skills in the use of means of communication appropriate to the discipline and the audience.

- **Credit: 65 - 74** There is evidence of learning that goes beyond replication of content knowledge or skills relevant to the learning outcomes. You communicate ideas fluently and clearly in terms of the conventions of the discipline.
- **Pass: 50 - 64** There is sufficient evidence that you attained the learning outcomes. You demonstrate an understanding of fundamental concepts covered in the unit. You are able to apply core big data technology.
- **Fail: 0 - 49** There is no evidence you attained the learning outcomes. There is missing or partial or superficial or faulty understanding and application of the fundamental concepts covered in the unit.

Criteria for passing the different assessment tasks will reflect the above standards and be made clear in the guidelines distributed with the task descriptions.

Changes since First Published

Date	Description
29/07/2019	The updated guide includes the name of the tutor (Urvashi Khanna). Policies on late submission and supplementary exam have been updated. The section on grading standards has now more detail.