ITEC874 — Big Data Technologies

Week 1 Lecture 1: Housekeeping

Diego Mollá

Department of Computer Science
Macquarie University

ITEC874 2018H2

Programme

Reading

- Handbook entry
  http://handbook.mq.edu.au/2018/Units/PGUnit/ITEC874
- ITEC874 Unit Guide
  http://unitguides.mq.edu.au/units/search/2018?query=ITEC874

Welcome to ITEC874!

Handbook Entry

Big Data Technologies – ITEC874

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'.

Who are we?

Diego Molla Unit Convenor, Lecturer (weeks 7-12)
9WW 331, diego.molla-aliod@mq.edu.au

Amin Beheshti Lecturer (weeks 1-6)
9WW 340, amin.beheshti@mq.edu.au

Fred Amouzgar Workshops
fred.amouzgar@mq.edu.au

Who are you?

Please answer the question about your programming experience (iLearn, activity week 1).

Web Resources

- The unit is available in iLearn http://ilearn.mq.edu.au.
- All the administrative material presented in this lecture is also available at this site.
  - Unit Outline.
  - Administrative Information.
  - Lecture Notes
  - Pointers to Reading.
  - Other Useful Stuff.
- You are expected to keep up-to-date by using iLearn for:
  - Relevant news and information.
  - Discussions.
  - Submission of assignments.

Github

- Some of the material of this unit is available in a public github repository.
- https://github.com/dmollaaliod/itec874-2018
-
  - Lecture notes
  - Workshop tasks
  - Code
- If you know how to use git, this will be the best way to make sure you have the latest versions.
  - git is one of the most popular version control systems.
  - Search the Web for tutorials and additional information on git.
- You can use the github browser interface to download individual files.

Learning Outcomes

1. Obtain a high level of technical competency in standard and advanced methods for big data technologies.
2. Understand the current status of and recognize future trends in big data technologies.
3. Reflect on the changes the big data technologies bring to businesses, organisations and society, and critically analyse future trends.
4. Develop a competency with emerging big data technologies, applications and tools.
5. Communicate clearly and effectively.

Learning Outcomes

1. Obtain a high level of technical competency in standard and advanced methods for big data technologies.
2. Understand the current status of and recognize future trends in big data technologies.
3. Reflect on the changes the big data technologies bring to businesses, organisations and society, and critically analyse future trends.
4. Develop a competency with emerging big data technologies, applications and tools.
5. Communicate clearly and effectively.

Learning Outcomes

1. Obtain a high level of technical competency in standard and advanced methods for big data technologies.
2. Understand the current status of and recognize future trends in big data technologies.
3. Reflect on the changes the big data technologies bring to businesses, organisations and society, and critically analyse future trends.
4. Develop a competency with emerging big data technologies, applications and tools.
5. Communicate clearly and effectively.

Learning Outcomes

1. Obtain a high level of technical competency in standard and advanced methods for big data technologies.
2. Understand the current status of and recognize future trends in big data technologies.
3. Reflect on the changes the big data technologies bring to businesses, organisations and society, and critically analyse future trends.
4. Develop a competency with emerging big data technologies, applications and tools.
5. Communicate clearly and effectively.

Learning Outcomes

1. Obtain a high level of technical competency in standard and advanced methods for big data technologies.
2. Understand the current status of and recognize future trends in big data technologies.
3. Reflect on the changes the big data technologies bring to businesses, organisations and society, and critically analyse future trends.
4. Develop a competency with emerging big data technologies, applications and tools.
5. Communicate clearly and effectively.

Rooms and Times

Lectures

- Thursday 4pm-6pm (9 Wally's Walk, 131 Tutorial room)

Workshops

One of these; check your timetable!

- Thursday 6pm-8pm (9 Wallys Wlk, 112 Faculty Unix Lab)
- Monday 3pm-5pm (9 Wallys Wlk, 114 Faculty Unix Lab)

Please Note

- Workshops start from this week.
- Thursday workshop of week i will cover topics from week i.
- Monday workshop of week i will cover topics from week $i - 1$.

Workshops

Workshops

- Workshops will typically demonstrate key technology for Big Data applications.
- Tasks will typically cover practical assignment tasks and will help understand key concepts for the final examination.
- This is also your opportunity to discuss and clarify content issues.

Practical Assessed Assignments

1. Data Lakes (10%, due Week 3)
   In this assignment you will explore the management of big data using data lake technology.
2. Apache Hadoop (20%, due Week 8)
   In this assignment you will apply Apache Hadoop to solve a problem using Big Data.
3. Data Analysis (20%, due Week 12)
   In this assignment you will perform analysis of Big Data.

Practical Assessed Assignments

1. Data Lakes (10%, due Week 3)
   In this assignment you will explore the management of big data using data lake technology.
2. Apache Hadoop (20%, due Week 8)
   In this assignment you will apply Apache Hadoop to solve a problem using Big Data.
3. Data Analysis (20%, due Week 12)
   In this assignment you will perform analysis of Big Data.

Practical Assessed Assignments

1. Data Lakes (10%, due Week 3)
   In this assignment you will explore the management of big data using data lake technology.
2. Apache Hadoop (20%, due Week 8)
   In this assignment you will apply Apache Hadoop to solve a problem using Big Data.
3. Data Analysis (20%, due Week 12)
   In this assignment you will perform analysis of Big Data.

Submitting your Assignment

- Read the assignment specifications.
- Submit in iLearn.
- Hard deadlines:
  - 20% of the **maximum** mark off per day of delay.

Plagiarism

- You may discuss but not write together.
- Read the Academic Honesty Policy. https://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policies/academic-honesty

Assessment

Assessment Components

- Assignment 1: 10%
- Assignment 2: 20%
- Assignment 3: 20%
- Exam: 50%

Final Assessment

- Your final mark and grade are entirely determined by the sum of marks of the individual assessment tasks.
- To pass the unit, the sum of marks must be at least 50% of the total assessment marks.
- This unit does not have hurdle assessments.

Tentative Schedule — Amin

Week	Lecture	Workshop
1	Introduction to Big Data	Microsoft Azure
2	Organising Big Data	Data Lake Services
3	Curating Big Data	Knowledge Lake Services
4	Processing Big Data	Hadoop
5	Industry Talk	Cloudera
6	Cloud Computing	Cloud

Tentative Schedule — Diego

Week	Lecture	Workshop
7	Analysing Big Data	SAS, Microsoft Azure
8	Streams	TBA
9	Visualising Big Data	SAS Visual Statistics
10	Visualising Big Data	TBA
11	Big Data and Society	TBA
12	Industry Talk	TBA