UNSW Engineering

School of Computer Science and Engineering

ZZEN9313

Big Data Management

Hexamester 4, 2022

Course Outline

Description

This course will help you understand modern large-scale data analytics systems. You will learn query languages (such as SQL, HiveQL and Spark SQL) to access data in different platforms such as a MySQL database and data warehousing. You will experience using Hadoop and how the data is created and stored in these platforms. Additionally, you will also learn the MapReduce programming and Spark programming for big data management.

Units of Credit

The course is worth 6 units of credit.

Pre-requisites and Assumed Knowledge

The pre-requisite for this course is that you have completed ZZEN9021 Principles of Programming.

Learning Outcomes

On successful completion of the course, you should be able to:

1. Describe and critique the important characteristics of Big Data.
2. Identify and evaluate an appropriate storage structure for a Big Data repository.
3. Use query languages to manipulate Big Data.
4. Utilise the map/reduce paradigm to manipulate Big Data.
5. Utilise the Spark platform to manipulate Big Data.
6. Select and apply appropriate tools to develop efficient solutions for analytical problems involving Big Data.

Schedule

Date	Week	Topic	Assessment Due
27 June	0	Orientation Week	
04 July	1	Introduction to Relational Databases and Big Data	
11 July	2	Querying data with SQL	Assessment 1
18 July	3	Hadoop and MapReduce	Assessment 2
25 July	4	MRJob	Assessment 3
01 August	5	Spark	Assessment 4
08 August	6	Hive and SparkSQL	Assessment 5
15 August	7	Examination Week	Assessment 6

Assessment

In order to pass the course, you must:

- achieve a total mark of at least 50;
- meet any additional requirements of the assessment tasks.

The assessment tasks are:

No.	Type	Weight	Due*	Learning Outcomes
1	Database diagrams	16%	Week 2, Monday 5pm 11 July	2
2	SQL queries	16%	Week 3, Monday 5pm 18 July	3
3	MapReduce using Hadoop Streaming	17%	Week 4, Monday 5pm 25 July	4,6
4	MRJob	17%	Week 5, Monday 5pm 01 August	4,6
5	Spark	17%	Week 6, Monday 5pm 08 August	5,6
6	Hive and Spark SQL	17%	Week 7, Monday 5pm 15 August	3,6

*All dates and times are Sydney (Australia) dates and times

If you submit after the due date, you will be penalised 5% per day capped at five days (120 hours) from the assessment deadline, after which you cannot submit an assessment unless you have an approved extension through Special Consideration or Equitable Learning Services.

Teaching Staff

Your teaching staff are:

- Xin Cao, xin.cao@unsw.edu.au (Course Convenor)

Teachers are your main point of contact. Their consultation times will be advised on the course website.

Resources

The website for the course is on Moodle, at: http://moodle.telt.unsw.edu.au. All readings and activities will be available there - no additional materials are required.

Technical Requirements

The course is fully online. You will need:

- A fast and reliable computer (or equivalent device), with an up-to-date operating system
- A fast and reliable internet connection
- The latest version of a modern browser (e.g. Edge, Chrome, Firefox or Safari)
- A reliable way to store your files - either on your computer with a backup routine, or in the cloud (e.g. using Dropbox)

UNSW
SYDNEY

Learning and Teaching Activities

The course contains a variety of resources and activities that are carefully designed to enhance your learning.

Some activities require you to work and think alone, by reading some text, listening to a recording or watching a video. You might be asked to engage with the material and explore interactive elements by clicking to reveal content, to help you better absorb and process the concepts. Some activities require you to produce work of your own. You might be answering a question, writing code to solve a problem, or posting to a forum, for example. Some activities are assessment tasks, which have been carefully designed to measure how well you have achieved the learning outcomes of the course. Typically, you will get feedback on your work, either from yourself (by checking your work with models that are provided), or from an automatic marking process, or from your peers, or from your teacher.

You also have access to a variety of ways to communicate with your peers and with the teaching staff. The general discussion forums are a place for you to ask and answer questions, to interact with your peers, and to be challenged by your teachers. Getting involved in these forums will enhance your learning experience and make it more enjoyable. Your course may include Webinars, which provide an opportunity to hear directly from your Online Lecturers and ask questions in real time. All webinars are recorded so you can access them at any time. Online Lecturers are available for consultations and will post information about how to access consultations on the course website. You can also contact your Online Lecturer by email using the email address in the teaching staff section of this outline.

It is up to you how much work you do. The more time and effort that you can dedicate to the course, the better will be your learning and your results.

Special Consideration

If illness or other circumstances beyond your control interfere with your assessment performance then you can apply for special consideration, to get an extra opportunity to demonstrate your level of performance.

You must make your application online, through the [Special Consideration portal on myUNSW](#). Do not apply to your course teaching staff - they will be notified automatically.

You must apply before the assessment task is due or the exam is held - if you submit the assessment or sit the exam then you are declaring yourself well enough to do so and are unable to subsequently apply for special consideration. If illness or misadventure prevent you from applying in advance, then you must apply as soon as possible, and provide evidence that you could not apply sooner. If you become unwell on the day of the exam, you must provide evidence dated within 24 hours of the exam, with your application.

Your application will be considered centrally, by a case review team, and they will notify you of the outcome. If your application is successful, then an alternative means of fulfilling the assessment requirements of the course will be provided.

You can read more about special consideration at: https://student.unsw.edu.au/special-consideration.

Academic Integrity

UNSW values academic integrity and has strict rules against cheating. In particular, it has strict rules against trying to get credit that you don't deserve. Thus, you should not plagiarise - i.e. present someone else's work as if it's your own. This could be the work of an academic, or a peer, or a contract writer, and it includes work of all kinds - exact words, general ideas, designs, drawings, software, and so on. Nor should you recycle your own work - i.e. submit it for credit in multiple courses. UNSW also has strict rules against helping others to cheat - e.g. by giving someone your work to copy, or doing someone's work for them, and so on.

For further information about academic integrity and plagiarism at UNSW go to:
https://student.unsw.edu.au/plagiarism

For information about acknowledging your sources and referencing go to:
https://student.unsw.edu.au/referencing. If you are not sure what referencing style to use in this course, you should ask your Online Lecturer.

Evaluation and Development

Toward the end of the hexamester you will be asked to give feedback about the course, via UNSW's MyExperience survey. Your feedback will be used, along with feedback from other stakeholders, to help improve the course. You can also contact your Course Convenor any time you have suggestions or other feedback.

Quality Assurance

UNSW actively monitors student learning and quality of the student experience in its programs. A random selection of completed assessment tasks may be used for quality assurance, such as determining the extent to which program learning goals are being achieved. The information is required for accreditation purposes, and aggregated findings will be used to inform changes aimed at improving the quality of programs. All material used for such processes will be treated as confidential.