UNSW Course Outline

COMP9313 Big Data Management - 2024

Published on the 10 Sep 2024

General Course Information

Course Code : COMP9313
Year : 2024
Term : Term 3
Teaching Period : T3
Is a multi-term course? : No
Faculty : Faculty of Engineering
Academic Unit : School of Computer Science and Engineering
Delivery Mode : In Person
Delivery Format : Standard
Delivery Location : Kensington
Campus : Sydney
Study Level : Undergraduate, Postgraduate
Units of Credit : 6

Useful Links

Handbook Class Timetable

Course Details & Outcomes

Course Description

This course introduces the core concepts and technologies involved in managing Big Data. It will first introduce the characteristics of big data and big data analysis. Then, we will learn the open-source big data management framework Hadoop. We will mainly focus on Hadoop MapReduce

programming. YARN, HDFS, HBase, and Hive will be briefly introduced as well. We will also learn an open-source memory-based distributed computing framework Spark. Another major focus of this course is algorithm design on large-scale data sets based on big data management frameworks, in various domains such as data stream mining, graph data processing, and finding similar items.

Course Aims

This course aims to introduce students to the concepts behind Big Data, the core technologies used in managing large-scale data sets, and a range of technologies for developing solutions to large-scale data analytics problems.

This course is intended for students who want to understand modern large-scale data analytics systems. It covers a wide range of topics and technologies. It will prepare students to be able to build such systems as well as use them efficiently and effectively to address challenges in big data management.

Course Learning Outcomes

Course Learning Outcomes
CLO1 : Describe the important characteristics of Big Data
CLO2 : Develop an appropriate storage structure for a Big Data repository
CLO3 : Utilise the map/reduce paradigm and the Spark platform to manipulate Big Data
CLO4 : Use a high-level query language to manipulate Big Data
CLO5 : Develop efficient solutions for analytical problems involving Big Data

Course Learning Outcomes	Assessment Item
CLO1 : Describe the important characteristics of Big Data	• Final Exam
CLO2 : Develop an appropriate storage structure for a Big Data repository	• Coding Project 1 • Coding Project 3 • Final Exam
CLO3 : Utilise the map/reduce paradigm and the Spark platform to manipulate Big Data	• Coding Project 2 • Coding Project 1 • Coding Project 3 • Final Exam
CLO4 : Use a high-level query language to manipulate Big Data	• Coding Project 2
CLO5 : Develop efficient solutions for analytical problems involving Big Data	• Coding Project 2 • Coding Project 3 • Final Exam

Learning and Teaching Technologies

Moodle - Learning Management System | Blackboard Collaborate | EdStem

Assessments

Assessment Structure

Assessment Item	Weight	Relevant Dates
Coding Project 1 Assessment Format: Individual	12%	Start Date: Not Applicable Due Date: Not Applicable
Coding Project 2 Assessment Format: Individual	16%	
Coding Project 3 Assessment Format: Individual	22%	
Final Exam Assessment Format: Individual	50%	

Assessment Details

Coding Project 1

Assessment Overview

This coding project assesses the student's MapReduce programming skills. It will be assessed manually by course tutors according to a rubric. The feedback will be provided in Moodle to students in the format of comments on the students' submissions.

Course Learning Outcomes

- CLO2 : Develop an appropriate storage structure for a Big Data repository
- CLO3 : Utilise the map/reduce paradigm and the Spark platform to manipulate Big Data

Generative AI Permission Level

No Assistance

This assessment is designed for you to complete without the use of any generative AI. You are not permitted to use any generative AI tools, software or service to search for or generate information or answers.

For more information on Generative AI and permitted use please see here.

Coding Project 2

Assessment Overview

This coding project assesses the student's Spark programming skills. It will be assessed manually by course tutors according to a rubric. The feedback will be provided in Moodle to students in the format of comments on the students' submissions.

Course Learning Outcomes

- CLO3 : Utilise the map/reduce paradigm and the Spark platform to manipulate Big Data
- CLO4 : Use a high-level query language to manipulate Big Data
- CLO5 : Develop efficient solutions for analytical problems involving Big Data

Generative AI Permission Level

No Assistance

This assessment is designed for you to complete without the use of any generative AI. You are not permitted to use any generative AI tools, software or service to search for or generate information or answers.
For more information on Generative AI and permitted use please see here.

Coding Project 3

Assessment Overview

This coding project assesses the student's Spark programming skills, using a real cloud computing platform such as Google Dataproc. It will be assessed manually by course tutors according to a rubric. The feedback will be provided in Moodle to students in the format of comments on the students' submissions.

Course Learning Outcomes

- CLO2 : Develop an appropriate storage structure for a Big Data repository
- CLO3 : Utilise the map/reduce paradigm and the Spark platform to manipulate Big Data
- CLO5 : Develop efficient solutions for analytical problems involving Big Data

Generative AI Permission Level

No Assistance

This assessment is designed for you to complete without the use of any generative AI. You are not permitted to use any generative AI tools, software or service to search for or generate information or answers.
For more information on Generative AI and permitted use please see here.

Final Exam

Assessment Overview

The final exam assesses the students' MapReduce and Spark programming skills, as well as algorithm design for big data analytics. The exam will be marked by course tutors manually according to a rubric. The feedback is provided upon students' request.

Course Learning Outcomes

- CLO1 : Describe the important characteristics of Big Data
- CLO2 : Develop an appropriate storage structure for a Big Data repository
- CLO3 : Utilise the map/reduce paradigm and the Spark platform to manipulate Big Data
- CLO5 : Develop efficient solutions for analytical problems involving Big Data

Hurdle rules

The student has to obtain at least 20 marks (out of 50) in the final exam.

Generative AI Permission Level

No Assistance

This assessment is designed for you to complete without the use of any generative AI. You are not permitted to use any generative AI tools, software or service to search for or generate information or answers.
For more information on Generative AI and permitted use please see here.

General Assessment Information

Grading Basis

Standard

Course Schedule

Attendance Requirements

Students are strongly encouraged to attend all classes and review lecture recordings.

Staff Details

Position	Name	Email	Location	Phone	Availability	Equitable Learning Services Contact	Primary Contact
Convenor	Xin Cao					No	Yes

Other Useful Information

Academic Information

I. Special consideration and supplementary assessment

If you have experienced an illness or misadventure beyond your control that will interfere with your assessment performance, you are eligible to apply for Special Consideration prior to, or within 3 working days of, submitting an assessment or sitting an exam.

Please note that UNSW has a Fit to Sit rule, which means that if you sit an exam, you are declaring yourself fit enough to do so and cannot later apply for Special Consideration.

For details of applying for Special Consideration and conditions for the award of supplementary assessment, please see the information on UNSW's Special Consideration page.

II. Administrative matters and links

All students are expected to read and be familiar with UNSW guidelines and polices. In particular, students should be familiar with the following:

- Attendance
- UNSW Email Address
- Special Consideration
- Exams
- Approved Calculators
- Academic Honesty and Plagiarism
- Equitable Learning Services

III. Equity and diversity

Those students who have a disability that requires some adjustment in their teaching or learning environment are encouraged to discuss their study needs with the course convener prior to, or at the commencement of, their course, or with the Equity Officer (Disability) in the Equitable Learning Services. Issues to be discussed may include access to materials, signers or note-takers, the provision of services and additional exam and assessment arrangements. Early notification is essential to enable any necessary adjustments to be made.

IV. Professional Outcomes and Program Design

Students are able to review the relevant professional outcomes and program designs for their streams by going to the following link: https://www.unsw.edu.au/engineering/student-life/student-resources/program-design.

Note: This course outline sets out the description of classes at the date the Course Outline is published. The nature of classes may change during the Term after the Course Outline is published. Moodle or your primary learning management system (LMS) should be consulted for the up-to-date class descriptions. If there is any inconsistency in the description of activities between the University timetable and the Course Outline/Moodle/LMS, the description in the Course Outline/Moodle/LMS applies.

Academic Honesty and Plagarism

UNSW has an ongoing commitment to fostering a culture of learning informed by academic integrity. All UNSW students have a responsibility to adhere to this principle of academic integrity. Plagiarism undermines academic integrity and is not tolerated at UNSW. *Plagiarism at UNSW is defined as using the words or ideas of others and passing them off as your own.*

Plagiarism is a type of intellectual theft. It can take many forms, from deliberate cheating to accidentally copying from a source without acknowledgement. UNSW has produced a website with a wealth of resources to support students to understand and avoid plagiarism, visit: student.unsw.edu.au/plagiarism. The Learning Centre assists students with understanding academic integrity and how not to plagiarise. They also hold workshops and can help students one-on-one.

You are also reminded that careful time management is an important part of study and one of the identified causes of plagiarism is poor time management. Students should allow sufficient time for research, drafting and the proper referencing of sources in preparing all assessment tasks.

Repeated plagiarism (even in first year), plagiarism after first year, or serious instances, may also be investigated under the Student Misconduct Procedures. The penalties under the procedures can include a reduction in marks, failing a course or for the most serious matters (like plagiarism in an honours thesis or contract cheating) even suspension from the university. The Student Misconduct Procedures are available here:

www.gs.unsw.edu.au/policy/documents/studentmisconductprocedures.pdf

Submission of Assessment Tasks

Work submitted late without an approved extension by the course coordinator or delegated authority is subject to a late penalty of five percent (5%) of the maximum mark possible for that assessment item, per calendar day.

The late penalty is applied per calendar day (including weekends and public holidays) that the assessment is overdue. There is no pro-rata of the late penalty for submissions made part way through a day. This is for all assessments where a penalty applies.

Work submitted after five days (120 hours) will not be accepted and a mark of zero will be

awarded for that assessment item.

For some assessment items, a late penalty may not be appropriate. These will be clearly indicated in the course outline, and such assessments will receive a mark of zero if not completed by the specified date. Examples include:

- Weekly online tests or laboratory work worth a small proportion of the subject mark;
- Exams, peer feedback and team evaluation surveys;
- Online quizzes where answers are released to students on completion;
- Professional assessment tasks, where the intention is to create an authentic assessment that has an absolute submission date; and,
- Pass/Fail assessment tasks.

Faculty-specific Information

Engineering Student Support Services – The Nucleus - enrolment, progression checks, clash requests, course issues or program-related queries

Engineering Industrial Training – Industrial training questions

UNSW Study Abroad – study abroad student enquiries (for inbound students)

UNSW Exchange – student exchange enquiries (for inbound students)

UNSW Future Students – potential student enquiries e.g. admissions, fees, programs, credit transfer

Phone

(+61 2) 9385 8500 – Nucleus Student Hub

(+61 2) 9385 7661 – Engineering Industrial Training

(+61 2) 9385 3179 – UNSW Study Abroad and UNSW Exchange (for inbound students)

School Contact Information

CSE Help! - on the Ground Floor of K17

- For assistance with coursework assessments.

The Nucleus Student Hub - https://nucleus.unsw.edu.au/en/contact-us

- Course enrolment queries.

Grievance Officer - grievance-officer@cse.unsw.edu.au

- If the course convenor gives an inadequate response to a query or when the courses convenor does not respond to a query about assessment.

Student Reps - stureps@cse.unsw.edu.au

- If some aspect of a course needs urgent improvement. (e.g. Nobody responding to forum queries, cannot understand the lecturer)

You should **never** contact any of the following people directly:

- Vice Chancellor

- Pro-vice Chancellor Education (PVCE)

- Head of School

- CSE administrative staff

- CSE teaching support staff

They will simply bounce the email to one of the above, thereby creating an unnecessary level of indirection and a delay in the response.