COMP9321 Data Services Engineering

Term 1, 2019

Course Details

Course Code	COMP9321
Course Title	Data Services Engineering
Units of Credit	6
Course Website	http://www.cse.unsw.edu.au/~cs9321
Handbook Entry	https://www.handbook.unsw.edu.au/postgraduate/courses/2019/COMP9321

Course Summary

Software engineering has advanced rapidly in recent years. The knowledge-, service-, and cloud-based economy in parallel with the continuous improvement in connectivity, storage and data processing capabilities allow access to a data deluge from sensors, social-media, news, user-generated, government and private data sources. Accordingly, in a modern data-oriented landscape, data-driven applications may need to deal with a collection of data sets - from unstructured, relational to NoSQL - that holds a vast amount of information gathered from various private/open data sources. Therefore, well-engineered service-oriented functionalities are critical for ingesting, organizing and querying the growing volume of data in modern web-based applications, as well as recommender systems that can help making decisions.

This course aims to introduce the student to core concepts and practical skills for engineering the data in service-oriented data-driven applications. Specifically, the course aims to answer these questions:

- How to access and ingest data from various external sources?
- How to process and store the data for applications?
- How to curate (e.g. Extract, Transform, Correct, Aggregate, and Merge/Split) and publish the data?
- How to apply available analytics to the data?
- How to visualize the data to communicate effectively
- How to utilize recent recommender systems to help making decisions?

Fundamentally, we will look at these questions through the lens of 'service-oriented' software design and implementation principles. At each topic, we will learn some core concepts, and how to implement the concepts in software through services.

The course uses the Python Programming Language as the practical basis for its modules. However, the concepts taught are universal and can be applied to any other web development framework.

Assumed Knowledge

Before commencing this course, we will assume that students have:

- completed one programming course (expected to be in Python)
- basic linear algebra knowledge
- basic data modelling and relational database knowledge

These are assumed to have been acquired in the following courses: For Postgrad - COMP9021 and COMP9311. For Undergrad - COMP1531 and COMP2041.

Assessment

Item	*Topics*	*Due*	*Marks*
Quizzes	All topics	Week 2-9	10%
Assignment 1	Data ingestion and manipulation	Week 3	10%
Assignment 2	Data publication as a RESTful service API	Week 5	10%
Assignment 3	Data Analytics project	Week 8	20%
Final Exam	All topics	Exam period	50%

Note: to pass this course you have to get more than 50% marks in final exam.

Course Schedule

Week	*Lectures*	*Labs*	*Assignments*	*Quizzes*	*Notes*
1	Course intro Data Access and ingestion		Release Assn1		
2	Data Cleansing and Manipulation	Accessing NoSQL DB, API data sourced, CSV files, text files.		Quiz 1	
3	Data Visualization	Cleansing data with Python Pandas and Open refine	Assn1 due Release Assn2	Quiz 2	
4	Building a Data service	Using matplotlib library for charts and plots		Quiz 3	
5	RESTful web clients	Build a RESTful service with flask	Assn2 due Release Assn3	Quiz 4	

6	Data Analytics Overview	Build a RESTful client		Quiz 5	
7	Data Analytics Applied Techniques and Tools	Introducing sci-kit learn toolkit		Quiz 6	
8	Introduction to recommender systems	Data classification and clustering example	Assn3 due	Quiz 7	Consultation
9	Demo week	Simple recommender systems with MovieLens dataset			Consultation
10	Final wrap-up	Consultation labs			

Note: although the quizzes are planned for every week, the actual number may vary depending on the progress of the weekly topics.

Student Conduct

The Student Code of Conduct (Information , Policy) sets out what the University expects from students as members of the UNSW community. As well as the learning, teaching and research environment, the University aims to provide an environment that enables students to achieve their full potential and to provide an experience consistent with the University's values and guiding principles. A condition of enrolment is that students inform themselves of the University's rules and policies affecting them, and conduct themselves accordingly.

In particular, students have the responsibility to observe standards of equity and respect in dealing with every member of the University community. This applies to all activities on UNSW premises and all external activities related to study and research. This includes behaviour in person as well as behaviour on social media, for example Facebook groups set up for the purpose of discussing UNSW courses or course work. Behaviour that is considered in breach of the Student Code Policy as discriminatory, sexually inappropriate, bullying, harassing, invading another's privacy or causing any person to fear for their personal safety is serious misconduct and can lead to severe penalties, including suspension or exclusion from UNSW.

If you have any concerns, you may raise them with your lecturer, or approach the School Ethics Officer , Grievance Officer , or one of the student representatives.
Plagiarism is defined as using the words or ideas of others and presenting them as your own. UNSW and CSE treat plagiarism as academic misconduct, which means that it carries penalties as severe as being excluded from further study at UNSW. There are several on-line sources to help you understand what plagiarism is and how it is dealt with at UNSW:

- Plagiarism and Academic Integrity
- UNSW Plagiarism Procedure

Make sure that you read and understand these. Ignorance is not accepted as an excuse for plagiarism. In particular, you are also responsible that your assignment files are not accessible by anyone but you by setting the correct permissions in your CSE directory and code repository, if using. Note also that plagiarism includes paying or asking another person to do a piece of work for you and then submitting it as your own work.

UNSW has an ongoing commitment to fostering a culture of learning informed by academic integrity.
All UNSW staff and 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.

If you haven't done so yet, please take the time to read the full text of

- UNSW's policy regarding academic honesty and plagiarism

The pages below describe the policies and procedures in more detail:

- Student Code Policy
- Student Misconduct Procedure
- Plagiarism Policy Statement
- Plagiarism Procedure

You should also read the following page which describes your rights and responsibilities in the CSE context:

- Essential Advice for CSE Students

Resources for Students

See Student Resources from the course site menu.

Course Evaluation and Development

This course is being improved, and to align 2025 Strategy for UNSW, a survey through UNSW's myExperience is to be conducted at the end of session, aka from week 11 until the exam, to obtain feedback on the quality of various aspects of this course. Your participation in the survey is volunteered, while encouraged and appreciated. Informal feedback is always welcome during the session, lecturer-in-charge (LiC) and course admin will be happy to help with any problems arisen.
Students will receive an email to their UNSW email account with a personalised web-link to their surveys. This page will contain surveys for each course in which they are enrolled. If students do not complete all surveys they will receive an email reminder and follow up email while the evaluation period is open.