# COMP9321 Data Services Engineering

# Term3, 2019

## 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.

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*

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 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.

# Student Learning Outcomes

When you successfully complete this course, you should be able to:

- Understand how to work with data and the various technologies involved in accessing, curating, storing, and publishing data
- Understand how to apply existing analytics and visualisation techniques to data
- Describe the main requirements to design and implement APIs (specifically REST APIs).
- Describe the main architecture of a modern Web-based, data-oriented applications
- Design and develop non-trivial data services solutions. The solutions can be about building applications that utilise the above technologies, or about making the technologies accessible to potential consumers
- Understand the basic issues with scalability, security/privacy and using online data processing platforms

*Prospective students should note that this course does not aim to introduce you to the latest packages or products available in the market. Rather, it strives to teach students the basic concepts and the fundamental principles in their implementation technologies so that they are able to follow and absorb technological developments in this space.*

This course contributes to the development of the following graduate capabilities:

| Graduate Capability |
| --- |
| **Scholars** capable of independent and collaborative enquiry, rigorous in their analysis, critique and reflection, and able t innovate by applying their knowledge and skills to the solution of novel as well as routine problems |
| **Entrepreneurial leaders** capable of initiating and embracing innovation and change, as well as engaging and enabling others to contribute to change |
| **Professionals** capable of ethical, self- directed practice and independent lifelong learning |
| **Global citizens** who are culturally adept and capable of respecting diversity and acting in a socially just and responsible way |

# Teaching Strategies

- Lectures: introduce concepts, show examples
- Lab Work: introduce technology required for the assignments
- Online Quizzes: revision of the concepts introduced in Lectures and Lab Work
- Assignments: solve significant problems (individually, as well as through group work)

# Teaching Rationale

This course is taught the way it is because we believe in structured learning, also learning by doing. We provide timely feedback for learning via small, step-by-step assignments that gradually build up knowledge and practical skills.

# Assessment

| Item | Topics | Due | Marks |
|---|---|---|---|
| Quizzes | All topics | Throughout the session | 10% |
| Assignment 1 | Data ingestion, manipulation and publication as a RESTful service API | Week 6 | 20% |
| Assignment 2 | Data Analytics project | Week 10 | 30% |
| Final Exam | All topics | Exam period | 40% |

# Course Schedule

| Week | Lectures | Labs | Assignments | Quizzes | Notes |
|---|---|---|---|---|---|
| 1 | Course intro | (No Lab, student should start by the Setup Python, Flask, NumPy, Pandas) | - | - | - |
| 2 | Data Access and ingestion | Accessing NoSQL DB, API data sourced, CSV files, text files. | - | Quiz 1 | - |
| 3 | Data Cleansing and Manipulation | Cleansing data with Python Pandas and Open refine | | Quiz 2 | - |
| 4 | Data Visualization | Using matplotlib library for charts and plots | Ass2 release | Quiz 3 | - |
| 5 | Building a Data service part1 | Build a simple Flask application | | Quiz 4 | - |
| 6 | Building a Data service part2 | Build a RESTful service with flask | Ass1 due | Quiz 5 | - |

| | | | | | |
|---|---|---|---|---|---|
| 7 | Data Analytics Overview | Introducing scikit learn toolkit | Release Ass2 | Quiz 7 | - |
| 8 | Data Analytics Applied Techniques and Tools part1 | Classification example | | Quiz 8 | - |
| 9 | Data Analytics Applied Techniques and Tools part2 | Clustering example | - | Quiz 9 | |
| 10 | Final wrap-up | - | Ass2 due | - | |

*Note: the Course Schedule might change according to the progress of the topics and feedback throughout the course. 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 evaluated each session using the myExperience system.