



COMP9321 Data Services Engineering

Term1, 2022

Week 1: Course Overview

Disclaimer

- Parts of the slides presented in the course are taken from previous offering for COMP9321 where the course material was prepared by Dr. Helen Paik and Dr. Lina Yao
- No services and Data-sets were hurt during the preparation of the course slides

Teaching Team

- Lecturer-in-Charge (LiC)
 - Morty Al-Banna (m.al-banna@unsw.edu.au) feel free to schedule consultations 6-7PM Every Tuesday.
 - Office: K17 401 (Desk 29) /Consultations via Microsoft Teams
- Course Administrator
 - Mohammad Ali Yaghub Zade Fard (m.yaghoubzadehfard@unsw.edu.au)
- Tutors
 - Mohammad Ali Yaghub Zade Fard
 - Dylan Sanusi-Goh
 - Alireza Tabebordbar
 - May Altulyan
- Course Web site
 - <http://www.cse.unsw.edu.au/~cs9321>
- Course Contact
 - E-mail: cs9321@cse.unsw.edu.au

COMP9321 Evolution

Previously known as Web applications engineering.

What was taught and why needed revision

- How to build Web sites using Java
- Standardised frameworks for Web apps (plenty)

Many Web apps are now data-oriented or utilise data heavily

–functionality requires combining or processing complex data from multiple sources

So COMP9321 became Data Service Engineering:

- How to work with data
- How to make the design and implementation of data-oriented service easy (i.e., an approach/technique)

So what is this course about?

Data Services Engineering

Data = is the problem we want to deal with, understanding the problems and possible ways to work with Data (e.g., “get” data, “publish” data, discover or manage multiple data sources, etc).

Services = is the proposed solution/design approach to make our problem “manageable”.

Engineering = (best practices, weighing options, we will think about these all throughout, at least try to) - obtain conceptual ideas as well as practical skills

Course Aims

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 visualize the data to communicate effectively*
- *How to apply available analytics to the data?*

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.

Assumed Knowledge

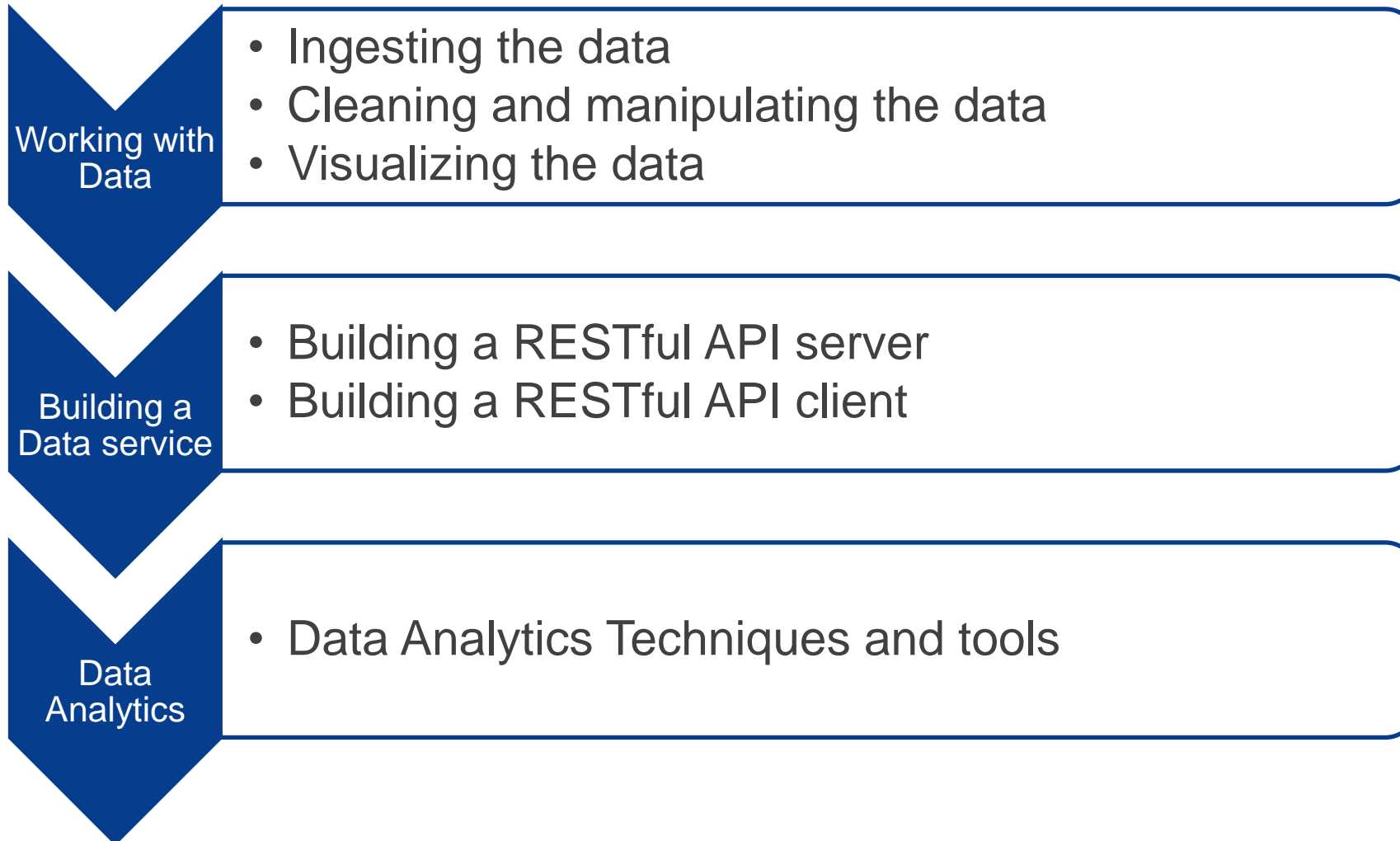
Before commencing this course, we will assume that students have:

- 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.

NOTE: This course is not meant to be an advanced course ...

Course Structure



Assessment (Tentative)

Assessment:

- 40% formal online exam: individual assessment.
- 50% on Individual assignment work
 - Assgn1 on Data ingestion, manipulation and visualization (individual) 15%
 - Assgn2 on building a service 15%
 - Assgn3 on building a data analytics service 20%
- 10% on 5 online quizzes (WebCMS-based quiz system, 'open' test)

Final Mark = quizzes + assignments + exam

Assignments Tentative

We have three individual assignments

Assignment 1: Data ingestion, cleaning manipulation and Visualization:

- 15 marks
- Release Week3, due on the end of week 5.

Assignment 2: Data Service (REST API):

- 15 marks
- Release on week 5, due on the end week 7.

Assignment 3: Data Analytics Service:

- 20 marks
- Release on week 7, due on the end week 10.

Bonus Mark

We have 5 bonus marks on the assignments work overall mark.

Bonus Mark

- 5 marks added to the assignments over all
- Assignment over all= assignment1 + assignment2 + assignment3 + Bonus
- Assignment overall cannot be more than 50%
- The weight of Bonus vary according to the contribution.

How?

- Interesting ideas about doing the same activity with less complexity (fewer lines of codes and more efficient)
- Improving the code (finding bugs, documentations, etc.)
- Adding new relevant activities or projects.
- **Making a video for an activity** and describing activities in detail
- Solving **challenges** announced during the lectures.

Consultation Labs

- A self-guided lab exercise is released every week.
- You can do them in your own time and come to the consultation Labs if needed.
- Use the forum. Share what you have learned/found

Tentative Schedule

Week	Lectures	Tutorials/Labs	Assignments
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.	-
3	Data Cleansing and Manipulation	Cleansing data with Python Pandas and Open refine	Assgn1 release
4	Data Visualization	Using matplotlib library for charts and plots	
5	Building a Data service (part1)	Build a simple Flask REST API	Assgn1 due Release Ass2
6	---	---	---
7	Building a Data service (part2)	RESTful Client	Assgn2 Due Release Assgn3
8	Data Analytics Applied Techniques and Tools part1	Classification example	
9	Data Analytics Applied Techniques and Tools part2	Clustering example	-
10	Final wrap-up	-	Assgn3 due

Supplementary Exam Policy

Supp Exam is only available to students who:

- DID NOT attend the final exam
- Have a good excuse for not attending
- Have documentation for the excuse

Submit special consideration within 72 hours (via myUNSW with supporting docs)

Everybody gets exactly one chance to pass the final exam. For CSE supplementary assessment policy, follow the link in the course outline.

Student Conduct

Please check: <https://student.unsw.edu.au/conduct>

Questions?