[NUS Internal]

Tentative Syllabus IS3107 (AY21/22 Sem 2): Data Engineering

Lecturer: Prof Frank Xing, xing@nus.edu.sg
Time and mode of instruction: Friday 18:30-20:30, online no recording.
Modular credits: 4MCs
Size: To ensure teaching quality, enrolment approval requests (e.g. co-taking w/ ATAP etc.) after 2021 Dec 7th 23:59PM, unfortunately, will NOT be entertained.

This IS3107 module covers the main concepts of data engineering, which include data pipeline design, data transformation, move and process, advanced data storage and query techniques with an emphasis on financial applications.

Prerequisite:
BT2102 “Data Management and Visualisation”
or CS2102 “Database Systems”.

Some knowledge of Python programming.

ILO (Intended Learning Objectives):
Be familiar with concepts and skills of a data engineer.
Be able to design data pipelines per data available and the desired task.
Understand the value of data engineering in business practices.

Assessment:
One mid-term quiz (30%)
One course project (60%)
Involvement - attendance or active participation (10%)
No final exam.

Structure:
This module has 10 hrs per week workload:
2-hr lecture + 1-hr tutorial + 3-hr preparation + 4-hr project.

13 lectures and 8 tutorials in total.

Other details:

Week 1 Introduction to Data Engineering

Week 2 Basics of Data Pipeline

Week 3 Data Pipeline Design

Week 4 Data Types and Storage
Tutorial 1 Your Minimum ETL Working Example in Python

Week 5 Cloud Service and Data Lake
Tutorial 2 Apache Airflow

Week 6 Distributed Data Processing
Tutorial 3 Your own UI -Scheduler and DAG

***
Recess Week
***

Week 7 NoSQL
Tutorial 4 [Time Slot for Mid-term Quiz]

Week 8 Stream Data Processing
Tutorial 5 ETL in Airflow

Week 9 Scalable Machine Learning
Tutorial 6 Accessing MongoDB

Week 10 Financial Data Engineering (Part 1)
Tutorial 7 Querying Time Series Data

Week 11 Financial Data Engineering (Part 2)
Tutorial 8 QA Session for Course Project

Week 12 Data Valuation

Week13 In-class Project Presentation