IS3107 Syllabus: Data Engineering

This IS3107 module covers the core concepts and intended goals (low latency, high throughput, high reliability, etc) of data engineering, which span the whole data engineering lifecycle (from ingestion to serving). Students will learn about topics including ETL, data pipeline, data formats, data architecture, and data moving, storage, and processing strategies to meet specific business/system requirements.

Instructor and tutor:

Frank XING, fxing@comp.nus.edu.sg

Liu Yan, yan.liu@nus.edu.sg

Course credits and logistics:

4 unit credits

Weekly in LR18, tutorials start from week 3 in the attached seminar room of LR19.

L1: Fri 12:00-14:00;   T1: Fri 14:00-15:00;   T2: Fri 17:00-18:00

Prerequisites:

- BT2102 "Data Management and Visualisation" or CS2102 "Database Systems"
- Some experience of database and Python programming (e.g. course project).

ILO (Intended learning objectives):

- Be able to apply data engineering concepts to analyze problems.
- Describe/understand challenges and propose strategies for corporate data storage and processing.

Assessment:

Participation in discussions and course activities (10%)

Assignment and Presentation (30%)

Quizzes (10% on week 6 + 10% on week 13 = 20%)

Course project (40%)

Course reference book:

Fundamentals of Data Engineering (2022) ISBN 978-1-09-810830-4

Tentative Lesson Plan:

Week and Date	Lecture Topic
Week 1	Introduction to Data Engineering
Week 2	Data Formats and Processing
Week 3	Data Pipeline and Orchestration
Week 4	Data Storage (Physical and Cloud)
Week 5	Data Organization
Week 6	Data Querying and Quiz 1
Recess Week	******
Week 7	Data Replication and Partitioning
Week 8	Data Architecture
Week 9	MapReduce and Hadoop
Week 10	MapReducible Algorithms
Week 11	Stream Data Processing
Week 12	Presentation
Week 13	Presentation and Quiz 2