WQD7007 Big Data Management

Introduction to the course

WQD7007 Big Data Management

- Course learning outcome (CLO):
    1. Explain the processes in data pipeline
    2. Discuss database concepts and technologies for big data storage and retrieval
    3. Apply appropriate models, tools, and technologies to implement storage, search and retrieval systems for large-scale structured and unstructured system
    4. Analyze data provenance and data trustworthiness, and its role in sharing and reuse of data

WQD7007 Big Data Management

- Method of assessment:
  - Continuous assessment (60%)
    - Group assignment (10%, submission on week 5)
    - Mid-term test (15%, 1 hour, week 7)
    - Lab test (15%, 1 hour, week 12)
    - Group project (20%, report and presentation on week 14)
  - Final exam (40%, 2 hours)

WQD7007 Big Data Management

- Topics to be covered:
    - Big data pipeline using Hadoop
    - Big data concepts
    - Big data technologies
    - Distributed computing for big data
    - Data provenance and data trustworthiness

WQD7007 Big Data Management

Lecturer: Dr Hoo Wai Lam

- Lecture:
  Tuesday, 6-8pm (DK1, Block B)
- Tutorial/Lab:
  Tuesday, 8-9pm (MM3, Block B)

WQD7007 Big Data Management

- Things to take note:
  - Online learning platform: https://spectrum.um.edu.my/
  - Software to be used:
    - OS: Ubuntu or CentOS, or
    - Virtualbox (if your system have more than 12 GB RAM)