STQD6324 Course
Data Management

Semester 2 2024/2025

Course code and name	:	STQD6324 Data Management
Credit	:	4
Course status	:	Elective
Lecturer & Course coordinator	:	Dr. Bernard Lee Kok Bang
		Department of Mathematical Sciences,
		FST UKM
		Email: *bernardlkb@ukm.edu.my*
Course assessments	:	Assignment 1: 20%
		Assignment 2: 20%
		Assignment 3: 20%
		Final Exam: 40%
Pre-requirement	:	None
Lecture schedule	:	Saturday / Sunday
		(9.00 am – 1.00 pm,
		2.00 pm – 6.00 pm)

Sinopsis kursus

Kursus ini bertujuan untuk menyediakan asas dan teknologi terkini yang digunakan dalam pengurusan data dan penyelesaian data raya. Pelajar akan diperkenalkan kepada model data, pangkalan data (pangkalan data berhubungan seperti jadual HIVE, dan pangkalan data NoSQL seperti MongoDB, Cassandra, dan HBase), pertanyaan dan teknik pemprosesan data raya. Topik termasuk keselamatan data, pusat data, ekosistem Hadoop (MapReduce dan HDFS), dan seni bina Apache Spark, pengagihan data dan pemprosesan selari. Pelajar akan belajar mengoptimumkan pemprosesan data raya menggunakan Spark dan akan diperkenalkan kepada alat pengaturcaraan seperti Pig dan Hive. Pada akhir kursus, pelajar akan mahir dalam mengurus, memproses, dan mengoptimumkan set data raya menggunakan teknologi data moden.

This course aims to prepare the basics and current technologies used for data management and big data solutions. Students will be introduced to data models, databases (both relational database like HIVE table, and NoSQL databases like MongoDB, Cassandra, and HBase), queries, and big data processing technique. Topics include data security, data centre, the Hadoop ecosystem (MapReduce and HDFS), and Apache Spark's architecture, data

distribution, and parallel processing. Students will learn to optimize big data processing using Spark and will be introduced to programming tools like Pig and Hive. By the end of the course, students will be proficient in managing, processing, and optimizing large datasets using modern data technologies.

Reference

- Dimitoglou, G., Deligiannidis, L. & Arabnia, H.R. 2024. Big Data, Data Mining and Data Science: Algorithms, Infrastructures, Management and Security (Intelligent Computing). 1st Ed. German: De Gruyter.
- Hermans, K. 2023. Mastering Apache Hadoop: A Comprehensive Guide to Learn Apache Hadoop. 1st Ed. Independently published.
- Kaufmann, M. & Meier, A. 2023. SQL and NoSQL Databases: Modeling, Languages, Security and Architectures for Big Data Management. 2nd Ed. Switzerland: Springer.
- Geng, H. 2021. DATA CENTER HANDBOOK Plan, Design, Build, and Operations of a Smart Data Center. 2nd Ed. Hoboken: John Wiley & Sons.
- Damji, J., Lee, D., Wenig, B. & Das, T. 2020. Learning Spark: Lightning-Fast Data Analytics. 2nd Ed. Sebastopol: O'Reilly Media.

Course Learning Outcomes

1. Berkebolehan membincangkan keperluan kepada pengurusan data dan teknologi data raya
   Able to discuss the need for data management and big data technology

2. Berkebolehan memasang dan menguruskan pangkalan data NoSQL
   Able to install and manage NoSQL database

3. Berkebolehan menganalisis situasi dan menggunakan teknologi data yang berpatutan
   Able to analyse the situation by using appropriate data technology

4. Berkebolehan membangunkan sistem untk menangani situasi data yang kompleks
   Able to develop a system to tackle complex data situation

Lesson Plan

Lectures	Contents	Notes
Week 2 **(2025-04-06)** **9.00 am – 1.00 pm**	Introduction to Data Management	
Week 2 **(2025-04-06)** **2.00 pm – 6.00 pm**	Data Model & Data Centre	
Week 5 **(2025-04-27)** **9.00 am – 1.00 pm**	Hadoop Distributed File System (HDFS) & MapReduce	
Week 5 **(2025-04-27)** **2.00 pm – 6.00 pm**	Apache Pig & Apache Spark	**Assignment 1**
Week 8 **(2025-05-18)**	**MID SEMESTER BREAK**	
Week 9 **(2025-05-25)** **9.00 am – 1.00 pm**	Apache Hive & Sqoop	**Assignment 2**
Week 9 **(2025-05-25)** **2.00 pm – 6.00 pm**	HBase & Cassandra	
Week 12 **(2025-06-15)** **9.00 am – 1.00 pm**	MongoDB	**Assignment 3**
Week 12 **(2025-06-15)** **2.00 pm – 6.00 pm**	Data Querying: Drill, Phoenix, Presto	

Week 14 **(2025-06-28)** **9.00 am – 1.00 pm**	YARN, Tez, Oozie, Zookeeper	
Week 14 **(2025-06-28)** **2.00 pm – 6.00 pm**	Apache Zeppelin	
2025-07-14 **-** **2025-08-03**	**FINAL EXAMINATION**	