

# ***STQD6324 Course Data Management***

## ***Semester 2 2024/2025***

<b>Course code and name</b>	: STQD6324 Data Management
<b>Credit</b>	: 4
<b>Course status</b>	: Elective
<b>Lecturer &amp; Course coordinator</b>	: Dr. Bernard Lee Kok Bang <i>Department of Mathematical Sciences, FST UKM</i> Email: <i>bernardlkb@ukm.edu.my</i>
<b>Course assessments</b>	: Assignment 1: 20% Assignment 2: 20% Assignment 3: 20% Final Exam: 40%
<b>Pre-requirement</b>	: None
<b>Lecture schedule</b>	: Thursday (9.00 am – 1.00 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 berhubungungan 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). 1<sup>st</sup> Ed. German: De Gruyter.
- Hermans, K. 2023. Mastering Apache Hadoop: A Comprehensive Guide to Learn Apache Hadoop. 1<sup>st</sup> Ed. Independently published.
- Kaufmann, M. & Meier, A. 2023. SQL and NoSQL Databases: Modeling, Languages, Security and Architectures for Big Data Management. 2<sup>nd</sup> Ed. Switzerland: Springer.
- Geng, H. 2021. DATA CENTER HANDBOOK Plan, Design, Build, and Operations of a Smart Data Center. 2<sup>nd</sup> Ed. Hoboken: John Wiley & Sons.
- Damji, J., Lee, D., Wenig, B. & Das, T. 2020. Learning Spark: Lightning-Fast Data Analytics. 2<sup>nd</sup> 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 untuk menangani situasi data yang kompleks  
*Able to develop a system to tackle complex data situation*

## ***Lesson Plan***

<b>Week</b>	<b>Date</b>	<b>Content</b>	<b>Remarks</b>
1.	2025-03-20	Introduction to Data Management	Online
2.	2025-03-27	Data Centre and Data Security	Online
3.	2025-04-03	Hadoop Ecosystem	Online
4.	2025-04-10	Hadoop Distributed File System (HDFS) & MapReduce	
5.	2025-04-17	Apache Pig	
6.	2025-04-24	Spark	
	<b>2025-05-01</b>	<b>Mid-semester break</b>	
7.	2025-05-08	Apache Hive & Sqoop	
8.	2025-05-15	NoSQL & CAP Theorem	
9.	2025-05-22	HBase & Cassandra	
10.	2025-05-29	MongoDB	
11.	2025-06-05	Data Querying: Drill, Phoenix, Presto	
12.	2025-06-12	YARN, Tez, Oozie, Zookeeper	
13.	2025-06-19	Apache Zeppelin	

Notes:

- **Revision Week: 2025-06-23 – 2025-06-29**
- **Final Exam: 2025-06-30 – 2025-07-20**