UNIVERSITI TEKNOLOGI MARA
COURSE INFORMATION

Code	:	DSC650
Course	:	DATA TECHNOLOGY AND FUTURE EMERGENCE
Level	:	Bachelor Degree
Credit Unit	:	3
Contact Hour	:	4
Part	:	5
Lecturer	:	Dr Khairul Anwar Sedek (F136 Kompleks STAR; Mobile# 019 4746960)
Group	:	RCS2405B & RCS2406A

Objectives : At the end of the course, students should be able to:

1. Describe concepts and practices of big data technology and future emergence. (C2)
2. Display practical skills of data technology in a big data environment. (P5)
3. Analyse existing big data sets using data technology tools. (C4)

Course Description : The course will give the students to explore key data analysis and management techniques, which applied to massive datasets are the cornerstone that enables real-time decision making in distributed environments, business intelligence in the Web, and scientific discovery at large scale. In particular, the students will examine the map-reduce, parallel computing paradigm and associated technologies such as Apache Hadoop distributed file systems, Apache Hive, and NoSQL databases.

SYLLABUS CONTENTS

Week	Topics and Sub-topics	Remarks
1 24/3/2025	Introduction and briefing on topics and projects.	
	Lab Session – Briefing & team groupings	
CUTI PERAYAAN **31/3 – 6/4/2025**		

2
7/4/2025
	1. Overview of Data Technology
	   1.1 Overview of Data Technology Evolution
	   1.2 Introduction of Big Data
	   1.3 Big Data Ecosystem
	   1.4 Foundation of Big Data Technology
	   1.5 Career Related
		Entrance Survey

	Lab Activity 1	

3
14/4/2025
	2. Business Motivations and Drivers for Big Data Adoption
	   2.1 Marketplace Dynamic
	   2.2 Business Process Management
	   2.3 Data Analytics and Data Science
	   2.4 Digitization
	   2.5 Internet of Everything
		Entrance Survey

	Lab Activity 2	

4
21/4/2025
	3. Data Storage Technology
	   3.1 Evolution of Data Storage: On-Disk Storage, Distributed File System, RDBMS, NoSQL
	   3.2 Comparison between SQL and NoSQL Database
	   3.3 Hadoop Distributed File System (HDFS)	

	Lab Activity 3	

5
28/4/2025
	4. Data Processing
	   4.1 Different Type of Data Processing: Parallel, Distributed, Batch, Transactional, Cluster and etc
	   4.2 MapReduce Framework, Algorithm and Process Data	

	Lab Activity 5	

6 5/5/2025	4.3 Real-Time Data Analysis using Apache Spark 4.4 Scalability and Fault Tolerance 4.5 Optimization and Data Locality 4.6 Real World Cases	
	No Lab	
7 12/5/2025	**5. NoSQL** 5.1 Structured and Unstructured Data 5.2 Taxonomy and SQL Implementation	
	Lab Activity 6	
8 19/5/2025	5.3 Basic and Related Architecture: HBase, Cassandra, MongoDB and etc.	Assessment: Lab Assignment
	Lab Activity 7	
9 26/5/2025	**6. Searching and Indexing Big Data** 6.1 Full Text Indexing and Searching 6.2 Indexing with Lucene 6.3 Distributed Searching with Elastic Search	
	Lab Activity 8	
SEMESTER BREAK/CUTI PERAYAAN **30/5 – 8/6/2025**		
10 9/6/2025	**7. Big Data Technologies** 7.1 Introduction to Hadoop 7.2 Hadoop Ecosystem 7.3 Query Language for Hadoop	Assessment: Test (1-5 & Lab concepts)
	Lab Activity 9	
11 16/6/2025	7.4 Hadoop and Amazon Cloud 7.5 Migration to Other Big Data Platform	
	Lab Activity 10	

Week	Topic	Assessment
12 23/6/2025	**8. Trend in Data Technology** 8.1 Automated Data Discovery 8.2 Deep Learning 8.3 The Next Frontier	
	Lab Activity 11	
13 30/6/2025	**Group Project - Preparation**	**Exit Survey** **SuFo**
14 7/6/2025	**Group Project – Presentation & Report Submission**	**Assessment:** **Project Report** **Exit Survey** **SuFo**
14-20/7/2025	**REVISION WEEK**	
21/7 – 10/8/2025	**FINAL EXAMINATION**	

TEACHING METHODOLOGY
Methods of instruction - lectures and laboratory practice;
Physical – F2F Week 2 – 14
(Ufuture – SOW, notes, modules)

ASSESSMENT

Continuous Assessment: 60%

	Test	15%
	Assignment (group)	20%
	Project (group)	25% (refer Appendix)

Final Examination: 40%

Final Exam 40%

RECOMMENDED TEXT

Abirami, R. N., Kadry, S., Gandomi, A. H., & Balusamy, B. (2021). *Big data: Concepts, technology, and architecture*. John Wiley & Sons, Inc.

Luna, A. A. D. (2020). *Principles of big data*.

Erl, T., Khattak, W., & Buhler, P. (2016). Big Data Fundamentals: Concepts, Drivers & Techniques. Prentice Hall Press. ISBN: 9780132146326

REFERENCES

1. Krishnan, K. (2019). *Building big data applications.*

2. Santos, M. Y., & Costa, C. (2020). *Big data: Concepts, warehousing, and analytics.*

3. Kumar, N. (2021). *Big data using hadoop and hive.*

4. Aytas, Y. (2021). *Designing big data platforms : How to use, deploy, and maintain big data systems.*

5. Mendelevitch, O., Stella, C, & Eadline, D. (2017). Practical Data Science with Hadoop and Spark: Designing and Building Effective Analytics at Scale. Addison-Wesley Professional. ISBN-13: 978-1305657458, ISBN-10: 1305657454.

6. Mehrotra, S., & Grade, A. (2019). *Apache spark quick start guide : Quickly learn the art of writing efficient big data applications with apache spark.*

7. Mendelevitch, O., Stella, C, & Eadline, D. (2017). Practical Data Science with Hadoop and Spark: Designing and Building Effective Analytics at Scale. Addison-Wesley Professional. ISBN-13: 978-1305657458, ISBN-10: 1305657454.

****
Appendix

Group project Description;

Form a group of 3-4 members. Each group is to develop a big data project on any use case (public or private sector), ranging from understanding the scope of the project by identifying business problems and opportunities, to evaluating the big data project while providing insights into what worked well and/or what did not.
i.e. Who (audience), what (decision), how (method/tools) – data story telling/data visualization

Sources of data;
   i. www.data.gov.my
   ii. www.dosm.gov.my
   iii. https://ieee-dataport.org/datasets?t%5B%5D=open_access
   iv. https://datasetsearch.research.google.com/
   v. www.kaggle.com

Some project ideas, refer to https://data-flair.training/blogs/data-science-project-ideas/