UNIVERSITI TEKNOLOGI MARA
DSC650: DATA TECHNOLOGY AND FUTURE EMERGENCE

Course Name (English)	DATA TECHNOLOGY AND FUTURE EMERGENCE `APPROVED`
Course Code	DSC650

MQF Credit	3

Course Description	The course will give the students to explore key data analysis and management techniques, which applied to massive data sets 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 Hadoop distributed file systems, and no sql databases.

Transferable Skills	Strong interpersonal, oral and written communication and presentation skills, ability to communicate complex findings.

Teaching Methodologies	Lectures, Lab Work

CLO	CLO1 Demonstrate an understanding on the basic concepts and practices of big data technology. (C3) CLO2 Display information management skills related to data technology and future emergence. (A5) CLO3 Demonstrate the use of data technology in big data environment. (P5)

Pre-Requisite Courses	No course recommendations

Topics

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

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

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)

4. Data Munging
4.1) Different Type of Data Processing: Parallel, Distributed, Batch, Transactional, Cluster and etc
4.2) MapReduce Framework, Algorithm and Process Data
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

5. NoSQL
5.1) Structured and Unstructured Data
5.2) Taxonomy and SQL Implementation
5.3) Basic and Related Architecture: HBase, Cassandra, MongoDB and etc

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
7. Big Data Technologies 7.1) Introduction to Hadoop 7.2) Hadoop Ecosystem 7.3) Query Language for Hadoop 7.4) Hadoop and Amazon Cloud 7.5) Migration to Other Big Data Platform
8. Trend in Data Technology 8.1) Automated Data Discovery 8.2) Deep Learning 8.3) The Next Frontier

Assessment Breakdown	%
Continuous Assessment	60.00%
Final Assessment	40.00%

Details of Continuous Assessment	Assessment Type	Assessment Description	% of Total Mark	CLO
	Assignment	n/a	20%	CLO2
	Group Project	n/a	25%	CLO3
	Test	n/a	15%	CLO1

Reading List	Recommended Text	**Erl, T., Khattak, W., & Buhler, P. (2016). 2016, *Big Data Fundamentals: Concepts, Drivers & Techniques.*, Prentice Hall [ISBN: 9780132146326]**
	Reference Book Resources	**Mendelevitch, O., Stella, C, & Eadline, D. (2017) 2017, *Practical Data Science with Hadoop and Spark: Designing and Building Effective Analytics at Scale.*, Addison-Wesley Professional [ISBN: 978130565745]**
		Tom White 2015, *Hadoop: the Definitive Guide ; Storage and Analysis at Internet Scale*, 14 Ed., O'Reilly Media [ISBN: 1491901632]
		Nathan Marz,James Warren 2015, *Big Data*, Manning Publications Company [ISBN: 1617290343]
		Albert Y. Zomaya, Sherif Sakr 2017, *Handbook of Big Data Technologies*, Springer [ISBN: 3319493398]
		Alex Gorelik 2019, *The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science 1st Edition*, 1 Ed., O'Reilly [ISBN: 1491931554]
Article/Paper List	This Course does not have any article/paper resources	
Other References	This Course does not have any other resources	