Macao Polytechnic University

Faculty of Applied Sciences

Master of Science in Big Data and Internet of Things

Module Outline

Academic Year <u>2022/2023</u>   Semester <u>1</u>

Learning Module	Introduction to Big Data			Class Code	COMP6122
Pre-requisite(s)	Nil				
Medium of Instruction	English			Credit	3
Lecture Hours	30 hrs	**Lab/Practice Hours**	15 hrs	**Total Hours**	45 hrs
Instructor	Wuman Luo		**E-mail**	luowuman@mpu.edu.mo	
Office	Rm. A323, Chi Un Building		**Telephone**	8599-6321	

Description

This learning module covers the characteristics of Big Data, the sources of massive data in enterprises and sensor networks, and the challenges in data ingestion, data storage and analytic processing. The students will acquire skills and working knowledge of the Big Data tools and technologies. This course focuses on the planning, designing and implementing Big Data solutions. Examples and exercises of Big Data systems are used to provide hands-on experiences in the workings of major components in Big Data solutions. The students will also be able to integrate the Big Data tools to form coherent solutions for business problems. Finally, additional related topics in the area of Big Data, such as alternative large-scale processing platforms, non-relational data stores, and Cloud Computing execution infrastructure are presented.

Learning Outcomes

After completing the learning module, students will be able to:

1. Identify the characteristics and challenges of Big Data in a range of complex application domains; (SM2fl, ET2fl)

2. Model and implement efficient Big Data solutions by integrating various advanced technologies; (D2fl, EP1fl)

3. Analyze the trade-offs in Big Data processing system design in complex infrastructures to handle a wide range of application problems; (EA1fl)

4. Demonstrate an understanding of non-relational databases in Big Data analytical pipelines. (EP2fl, EA2fl)

Content

1. Overview   (6 hours)

    1.1. Sources of Big Data

    1.2. Characteristics of Big Data

    1.3. Data Science and Data Analytics

2. Scalable Computing Systems   (9 hours)

    2.1. Distributed File Systems

    2.2. Commodity Clusters and Hadoop Architecture

    2.3. HDFS: Blocks and Replication, Read / write operations

    2.4. YARN: Cluster Resource Management and Distributed Processing

    2.5. Big Data Architecture and Data Pipeline

3. Large-scale Data Processing with Spark   (12 hours)

    3.1. The Scala Programming Language

    3.2. The Apache Spark Analytics Engine

    3.3. Resilient Distributed Datasets (RDD)

    3.4. RDD Operations: Transformations and Actions

4. Structured Data Analytics   (9 hours)

    4.1. Spark SQL

    4.2. Handling Structured and Semi-Structured Data

    4.3. Big Data File Formats

5. NoSQL Databases   (9 hours)

    5.1. Features and Design Goals

    5.2. Data Modelling and Querying

Teaching Method
Lectures and tutorial

Attendance
Attendance requirements are governed by the "Academic Regulations Governing Master's Degree Programmes" of Macao Polytechnic University. Students who do not meet the attendance requirements for the module shall be awarded an 'F' grade.

Assessment
This learning module is graded on a 100 point scale, with 100 being the highest possible score and 50 the passing score.

	Item	Description	AHEP3 LO	Percentage
1.	Assignment(s)	Home-based exercises	EA1fl, EA2fl, D2fl, EP1fl, EP2fl	40%
2.	Test(s)	Knowledge assessment	SM2fl, EA1fl, D2fl, ET2fl, EP1fl	20%
3.	Examination	3-hour written examination	SM2fl, EA1fl, D2fl, ET2fl, EP1fl	40%
			Total Percentage:	100%

Students with an overall score of less than 35 in the coursework will fail the module even if the overall score for the module is 50 or above.

Students with a score of less than 35 in the final examination will fail the module even if the overall score for the module is 50 or above.

Teaching Material(s)
Textbook(s)
There is no official text for this course. Course notes are distributed in the class.

Reference
Reference book(s)

1. T. Erl, W Khattak (2016). Big Data Fundamentals: Concepts, Drivers & Techniques, ServiceTech

2. T. White (2015). Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale, O'Reilly

3. M. Guller (2015). Big Data Analytics with Spark, Apress