

Big Data, Tools and Analysis

Module Information

2022.01, Approved

Summary Information

Module Code	6020DACOMP
Formal Module Title	Big Data, Tools and Analysis
Owning School	Computer Science and Mathematics
Career	Undergraduate
Credits	20
Academic level	FHEQ Level 6
Grading Schema	40

Teaching Responsibility

LJMU Schools involved in Delivery
Computer Science and Mathematics

Learning Methods

Learning Method Type	Hours
Lecture	22
Practical	33

Module Offering(s)

Display Name	Location	Start Month	Duration Number Duration Unit
SEP-CTY	CTY	September	12 Weeks

Aims and Outcomes

Aims	The aim of this module is to develop the knowledge and skills for working effectively with the large scale data storage and processing frameworks that underpin data science.
------	---

After completing the module the student should be able to:

Learning Outcomes

Code	Number	Description
MLO1	1	Differentiate between the functions of the components of big data storage and processing frameworks
MLO2	2	Appraise emerging trends in large scale data storage and processing
MLO3	3	Formulate a machine learning/analytics exercise for a given subject area

Module Content

Outline Syllabus	Big Data Volume – tracks what happens Velocity – real-time Variety – text, images, audio, video Big Data Difficulties Variability – inconsistency of data Veracity – quality of data Complexity – complex data management issues Big Data storage and Analysis Tools Apache Hadoop Hadoop provenance Apache Hadoop Framework Common Distributed File System (HDFS) YARN MapReduce Job Tracker Task Tracker Apache Hadoop Tools Pig (Pig Latin, ETL) Hive (data warehousing + SQL) in detail Apache Spark (in-memory analytics) in detail Apache Mahout (machine learning system) in detail Apache SOLR (scalable search tool) Hadoop in the Cloud - Amazon EC2/S3 Services Emerging Trends in Big Data storage and processing
Module Overview	
Additional Information	This module provides both theoretical and practical experience of large scale data storage considerations and the development of tools to support the processing of that data.

Assessments

Assignment Category	Assessment Name	Weight	Exam/Test Length (hours)	Module Learning Outcome Mapping
Technology	Machine Learning Exercise	60	0	MLO3
Centralised Exam	Examination	40	1.5	MLO1, MLO2

Module Contacts

Module Leader

Contact Name	Applies to all offerings	Offerings
Yun Sheng	Yes	N/A

Partner Module Team

Contact Name	Applies to all offerings	Offerings