

Liverpool John Moores University

Title: BIG DATA, TOOLS AND ANALYSIS
Status: Definitive
Code: **6124COMP** (121301)
Version Start Date: 01-08-2021

Owning School/Faculty: Computer Science and Mathematics
Teaching School/Faculty: Computer Science and Mathematics

Team	Leader
Yun Sheng	Y
Carl Chalmers	

Academic Level: FHEQ6 **Credit Value:** 20 **Total Delivered Hours:** 56.5
Total Learning Hours: 200 **Private Study:** 143.5

Delivery Options

Course typically offered: Semester 1

Component	Contact Hours
Lecture	22
Practical	33

Grading Basis: 40 %

Assessment Details

Category	Short Description	Description	Weighting (%)	Exam Duration
Report	AS1	Machine Learning Exercise	60	
Exam	AS2	Examination	40	1.5

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.

Learning Outcomes

After completing the module the student should be able to:

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

Learning Outcomes of Assessments

The assessment item list is assessed via the learning outcomes listed:

Machine Learning Exercise	3	
Examination	1	2

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

Learning Activities

Lectures will explore the underpinning theories of large scale data storage and

processing approaches, while practical sessions will implement those theories in a practical manner.

Notes

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.