Big Data (MHI222956/MHI225101)

1.1 Module Introduction

Module Contacts

Module Leader: Dr Yan Zhang

Room: M207

Email: yan.zhang@gcu.ac.uk

- To email the module leader, please start your email subject with: “Big Data 21/22A”

Module Overview

- This module will provide an introduction to the challenges and possible solutions around big data.
- The size and complexity of big data creates challenges in terms of storage, search, sharing, transfer, analysis and visualization.
- Students will be introduced to the problems of big data and technologies that help address these issues.

- Module descriptor

- Module handbook

Module Format

Activity	Hours per week
Independent Learning	10
Lectures	2
Labs	2

- Lectures - Flipped classroom
  - Full time students: online live sessions
  - GA students: online + F2F sessions
  - Online sessions to be accessed via ‘Collaborate Ultra’ link on GCU Learn
- Labs
  - Full time students: on campus as timetabled
  - GA students: self-arrange

Assessment

- Coursework Only!
- Coursework One (40%)
  - Submission deadline: 3rd Nov 2021
- Coursework Two (60%)
  - Submission deadline: 15th Dec 2021
- Threshold for each part: 35%

Teaching Schedule

Week #	Commencing	Lecture Topic	Lab Topic
1	27/09/2020	Big Data Introduction	---
2	4/10/2020	Foundations for Big Data Systems	Working with Python
3	11/10/2020	Big Data Modelling and Management System	Advanced program with Python A simple MapReduce task
4	18/10/2020	MongoDB Schema Design	Working with MongoDB
5	25/10/2020	Big Data Integration and Processing Pipeline	Python and MongoDB
6	01/11/2020	Data Exploring and Data Preparing	Coursework I relevant
7	08/11/2020	Machine Learning I	Data exploring and date preparing
8	15/11/2020	Machine Learning II	Classification and Regression
9	22/11/2020	Model Evaluation and Data Analysis Recap	Clustering and Association
10	29/11/2020	Big Data Framework	Coursework II relevant
11	06/12/2020	Big Data Security	A simple Spark task Coursework relevant

Software Requirements

- Lab Virtual Machine (VM)

  Three ways to run it:

  1. Run on a lab PC on campus

  2. Run on your computer using VMWare Player (Windows) or Virtual Box (Windows/Mac)

     Download and unzip a copy of the VM from: https://edshare.gcu.ac.uk/id/eprint/5215

  3. Run on a lab PC using remote access

     Need to enable MFA to assess computer lab remotely

Note: Refer to “Software Requirements for Labs and Coursework.pdf” posted on GCU-Learn for further details.

Software Requirements

- Python / Jupyter notebook

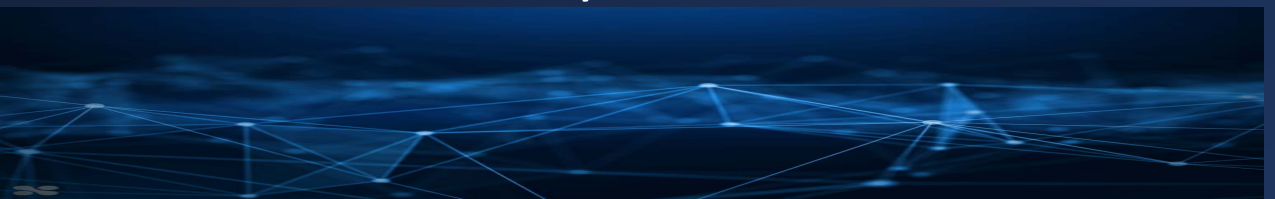
  To set these up on your own computer it is recommended that you install Anaconda.

- Google Colab

  Jupyter notebook runs in the cloud

- MongoDB

  You can download and install MongoDB Community Server



Module Feedback

Feedback is an essential part of your learning. It helps you to maximize your potential by raising your awareness of strengths and areas for improvement and identifying actions you can take to improve performance.

In this module you will receive feedback in a number of ways:

- **Lab exercises**

  You will attempt lab activities each week, and solutions will generally be made available to allow you to reflect on your own work

- **Assessments**

  You will receive feedback along with your marks for the summative

  assessments

- **Feedback from classes**

  You should attend all classes as these give you opportunities to get informal but valuable feedback on your understanding from staff and peers