Introduction to Big Data

Description

This 3-day residential DISCnet event, given by Julie Weeds (Sussex) will provide an introduction to Big Data.

Aims:

The aims of this course are to explore how big data techniques can be used to solve massive scale data analysis problems. The course will aim to introduce students to both the theoretical background of cloud computing as well as the practical applications. The processing of large datasets using Big Data techniques, map-reduce and other techniques will be a large focus.

The objectives of the course include students leaving equipped to design and implement Big Data processing pipelines, as well as being able to use exploratory approaches to analyse large datasets.

Learning Objectives:

The attendees should understand the theoretical approaches to big data analysis and the design of modern big data processing pipelines. Students should be able to design a big data processing system as well as successfully analyse large datasets using Python and Spark.

Prerequisites:

Practicals will require programming in Python, as well as the use of the UNIX command line / bash shell (e.g., skill learnt during the DISC6001 Software Carpentry course, or equivalent). While students do not need significant experience in Python itself, some serious programming experience is required as the course exercises will require you to write big data analytics code. This course is not suitable for students who have zero practical experience in writing code.

Students who are not confident in Python are expected to use the resources on Python to gain experience before the class. All students are expected to complete the pre-study exercise which looks at lambda expressions in Python.

Important Information

The course is mandatory for DISCnet core students and also open to non-core DISCnet and GRADnet students. The course is aimed at students in Year 1 or 2 of their PhD. The course may also be suitable for students in later years, depending on their computing and programming experience.

You will need to bring a laptop computer with you for the course.

Accommodation and full-board is included with the course.