CLO - Cloud Computing and Big Data: Pre-study

Paul Fremantle, paul@fremantle.org, June 2019

The ability to control servers remotely and virtualization technology has led to the creation of cloud computing. In turn, the availability of multiple low-cost servers has dramatically shifted the way in which data is processed. The result is an explosion of tools, technologies and approaches for handling very large datasets using cloud technologies.

In this course, we will study a range of cloud computing and big data technologies. We will primarily look at the Amazon EC2 cloud infrastructure and Apache Spark data processing model, but we will also explore other systems.

The learning objectives of this course are that you should be able to create systems that process large quantities of data in the cloud, including datasets that are too large to satisfactorily manage on any single system.

The textbook ***Big Data: Principles and best practices of scalable real-time data systems*** is a strong proponent of an approach called the Lambda Architecture, which was pioneered by Nathan Marz, the author. Although the Lambda Architecture is an important part of this course, we will not be focusing on it as much as the book does! In addition we will not be using all of the projects that the book does. However, it is one of the best currently available books on the subject, and I enjoy the technical level of detail. **Please read at least Chapter 1 of the book.**

The **practicals** will be using the Linux / Unix shell command line. We will mainly be using **Python 2.7**.

If you have never programmed in Python, please look at the following tutorial:

https://docs.python.org/2/tutorial/ (especially sections 1-5)

Big data programming makes significant use of a functional programming model called the **Lambda Calculus.** If you have never used lambdas and higher-order functions, please try the attached **Lambda Exercise.** It assumes you understand basic Python, so please do the Python exercises first if necessary.

If you have never used a Linux Command Shell, please read:
http://linuxcommand.org/lc3_learning_the_shell.php

Thanks, Paul