Understanding Big Data with Cloud Computing

FINTECH

Course outline

- Learn to process large big datasets using parallel and distributed computing with Dask
- Use Dask Dataframes to capture large amounts of data in tabular form
- Learn to provide Amazon Web Services (AWS) EC2 Virtual Machines for Distributed Computing with Dask

Course objectives

With more than half the world's data being generated in the decade, its availability becomes both a boon and bane as we are confronted with large datasets that can hardly fit into the memory of an average computer these days.

This programme explores ways to handle Big Data with Dask through parallelisation and distributed computing. Participants will get to leverage their understanding of Numpy and Pandas and utilise corresponding methods and functions provided by the Dask package to handle big data.

Visualise large datasets using Datashader

Use the Dask-ML API to perform Linear Regression on a big dataset

Course details
1 week

Certificated by Singapore Management University (SMU)

Who should attend

Aspiring data science professionals seeking to apply Python to real world data problems
Anyone with an interest in learning about the advanced data analysis techniques and putting it in practice
Managers looking into costs and performance hurdles through predictive modelling

Pre-requisites
Experience in Python programming (equivalent to that attained in Professional Certificate in Python Programming programme) is essential

Tools
Python

Model of training
Classroom,
Field trip

www.semeco.com.sg