Higher Diploma in Data Analytics

Programming for BIG Data Project Description

Semester 1

2013 – 2014

You are required to carry out a series of analyses of two datasets utilising appropriate programming languages and programming environments. For each of the chosen datasets you are required to compile a report of the analysis. For each of the reports you should provide the following:

- a description of the underlying datasets

  [2 x 5%]

- a description of the objective of the analysis

  [2 x 5%]

- the data processing activities carried out

  [2 x 5%]

- the presentation of the final analysis results along with any associated annotations and supplementary information

  [2 x 15%]

Additionally, you should also identify a particular dataset and analysis that can utilise the MapReduce programming paradigm for processing. You must then:

- implement algorithms to process the dataset in a MapReduce environment

  [20%]

- present your results via an appropriate report

  [10%]

All deliverables should be compiled into an accompanying project report document for submission along with any programming code elements. Your project report (3000 words) should discuss the challenges that you encountered whilst handling your chosen datasets and the means and mechanisms you implemented to overcome these challenges.

  [10%]

Note: the project contributes towards a maximum of 50% of the marks for the module.