COMP 7003 BIG DATA

Credit Points 10

Legacy Code 301046

Coordinator Yi Guo (https://directory.westernsydney.edu.au/search/name/Yi Guo/)

Description "Big data" is the label for the ever-increasing gigantic amount of data with which humanity has to cope. The availability of data and the development of cloud computing architectures to process and analyse these data have made data analytics a central tool in our endeavours. This unit will introduce students to the realm of "big data", covering the important principles and technologies of retrieving, processing and managing massive real-world data sets. It is designed to provide the basic techniques required by any discipline that needs to make sense out of the growing amount of data, and to equip students with the knowledge and key set of skills set to be competitive in the growing job market in the analytics field.

School Computer, Data & Math Sciences

Discipline Computer Science

Student Contribution Band HECS Band 2 10cp

Level Postgraduate Coursework Level 7 subject

Assumed Knowledge

It is expected that students enrolled in this subject should have basic programming skills in any programming language and working knowledge in elementary probability and statistics, including the concepts of random variables, basic probability distributions, expectations, mean and variance.

Learning Outcomes

On successful completion of this subject, students should be able to:

1. Explain the major trends in technology, business, and science behind big data
2. Analyse and compare a selection of major big data management techniques in use today, including parallel databases, NoSQL, MapReduce, cloud services
3. Evaluate the relative strengths and weaknesses of MapReduce and parallel database systems and apply the appropriate technique to tackle relevant big data problems
4. Apply proper methods of data pre-processing and cleaning for big data analysis

Subject Content

1. Foundations and recent trends of big data
2. Parallel database management systems
3. Data parallelism and the MapReduce framework
4. NoSQL databases and cloud services
5. Data processing and manipulation for big data analysis

Assessment

The following table summarises the standard assessment tasks for this subject. Please note this is a guide only. Assessment tasks are regularly updated, where there is a difference your Learning Guide takes precedence.

Item	Length	Percent	Threshold	Individual/ Group Task
Practical	2 hours per Practical (10 in total at 2% each)	20	N	Individual
Report	A report of 2000 words + final presentation 10 minutes including question time. Final mark is made up of 70% (or 28 out of 40) for the written component and 30% (or 12 out of 40) for the presentation	40	N	Individual
Final Exam	90 minutes open book exam including reading time	40	Y	Individual

Teaching Periods

Autumn
Parramatta - Victoria Rd
Evening
Subject Contact Yi Guo (https://directory.westernsydney.edu.au/search/name/Yi Guo/)

View timetable (https://classregistration.westernsydney.edu.au/even/timetable/?subject_code=COMP7003_22-AUT_PS_E#subjects)