

COMP 3002 APPLICATIONS OF BIG DATA

Credit Points 10

Legacy Code 301110

Coordinator Yi Guo ([https://directory.westernsydney.edu.au/search/name/Yi Guo/](https://directory.westernsydney.edu.au/search/name/Yi%20Guo/))

Description Many techniques and tools have been developed over the past decade to cope with the ever-growing needs for the processing and analysis of big data. This unit will cover the key techniques that have been widely used in big data applications, such as relational and Not Only Structured Query Language (NoSQL) databases, Web Services, parallel and cloud computing, MapReduce, Hadoop and its eco-system. It aims to introduce the emerging technologies and applications in big data to students, and keep up with the latest trends in the industry.

School Computer, Data & Math Sciences

Discipline Computer Science

Student Contribution Band HECS Band 2 10cp

Check your HECS Band contribution amount via the Fees (https://www.westernsydney.edu.au/currentstudents/current_students/fees/) page.

Level Undergraduate Level 3 subject

Pre-requisite(s) COMP 1013 OR COMP 1005

Assumed Knowledge

Knowledge of computer software, databases, and entry-level statistics.

Learning Outcomes

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

1. Explain the major trends and latest development in big data technology
2. Describe a selection of major techniques in use today for big data storage and processing, including NoSQL, MapReduce, cloud computing, web services
3. Build tools to obtain data from various sources and in different formats
4. Evaluate the relative strengths and limitations of relational and NoSQL database systems and recommend the most appropriate solution to data storage and access for different application scenarios
5. Employ HDFS and Hadoop for data storage and manipulation on parallel platforms
6. Select and apply appropriate tools from the Hadoop eco-system for big data management tasks

Subject Content

Sources and formats of big data

Relational Databases and SQL

NoSQL Databases

Web Scraping

Cloud computing platforms for big data

Data parallelism and the MapReduce framework

Data storage and processing with Hadoop Distributed File Systems (HDFS) and Hadoop

The Hadoop eco-system, including Pig, Hive, Spark, etc.

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
Quiz	5 x 30 mins; 6% each quiz	30	N	Individual
Report	Report: 2,000 words (or a 500-1,000 line program)	30	N	Individual
Final Exam	2 hours	40	N	Individual

Teaching Periods

Autumn

Parramatta - Victoria Rd

Day

Subject Contact Yi Guo ([https://directory.westernsydney.edu.au/search/name/Yi Guo/](https://directory.westernsydney.edu.au/search/name/Yi%20Guo/))

View timetable (https://classregistration.westernsydney.edu.au/even/timetable/?subject_code=COMP3002_22-AUT_PS_D#subjects)

Sydney City Campus - Term 1

Sydney City

Day

Subject Contact Antoinette Cevenini ([https://directory.westernsydney.edu.au/search/name/Antoinette Cevenini/](https://directory.westernsydney.edu.au/search/name/Antoinette%20Cevenini/))

View timetable (https://classregistration.westernsydney.edu.au/even/timetable/?subject_code=COMP3002_22-SC1_SC_D#subjects)

Sydney City Campus - Term 3

Sydney City

Day

Subject Contact Antoinette Cevenini ([https://directory.westernsydney.edu.au/search/name/Antoinette Cevenini/](https://directory.westernsydney.edu.au/search/name/Antoinette%20Cevenini/))

View timetable (https://classregistration.westernsydney.edu.au/even/timetable/?subject_code=COMP3002_22-SC3_SC_D#subjects)