COMP20008 Elements of Data Processing

Credit Points:	12.5
Level:	2 (Undergraduate)
Dates & Locations:	2016, Parkville This subject commences in the following study period/s: Semester 1, Parkville - Taught on campus.
Time Commitment:	Contact Hours: 48 hours, comprising of two 1-hour lectures and one 2-hour workshop per week Total Time Commitment: 170 hours
Prerequisites:	**Subject** / **Study Period Commencement:** / **Credit Points:** COMP10001 Foundations of Computing / Semester 1, Semester 2 / 12.5 And **Subject** / **Study Period Commencement:** / **Credit Points:** COMP10002 Foundations of Algorithms / Semester 1, Semester 2 / 12.5
Corequisites:	None
Recommended Background Knowledge:	None
Non Allowed Subjects:	None
Core Participation Requirements:	For the purposes of considering request for Reasonable Adjustments under the Disability Standards for Education (Cwth 2005), and Student Support and Engagement Policy, academic requirements for this subject are articulated in the Subject Overview, Learning Outcomes, Assessment and Generic Skills sections of this entry. It is University policy to take all reasonable steps to minimise the impact of disability upon academic study, and reasonable adjustments will be made to enhance a student's participation in the University's programs. Students who feel their disability may impact on meeting the requirements of this subject are encouraged to discuss this matter with a Faculty Student Adviser and Student Equity and Disability Support: http://services.unimelb.edu.au/disability
Coordinator:	Prof James Bailey
Contact:	Prof James Bailey email: **baileyj@unimelb.edu.au (mailto:baileyj@unimelb.edu.au)**
Subject Overview:	**AIMS** Data processing is fundamental to computing and data science. This subject gives an introduction to various aspects of data processing including database management, representation and analysis of data, information retrieval, visualisation and reporting, and cloud computing. This subject introduces students to the area, with an emphasis on both tools and underlying foundations. **INDICATIVE CONTENT** The subject's focus is on the data pipeline, and activities known colloquially as 'data wrangling'. Indicative topics covered include: # Capturing data (data ingress) # Data representation and storage

	# Cleaning, normalization and filling in missing data (imputation) # Combing multiple sources of data (data integration) # Query languages and processing # Scripting to support the data pipeline # Distributing a database over multiple nodes (sharding), cloud computing file systems Visualisation and presentation
Learning Outcomes:	**INTENDED LEARNING OUTCOME (ILO)** Having completed this subject the student is expected to: 1 Be familiar with the relationship of the data pipeline to data science 2 Be able to develop and critically evaluate alternative approaches to components of typical data pipelines 3 Apply data processing methodologies to preparing data while managing data quality, system scalability, and usability for decision making
Assessment:	Project work during semester, applying data processing to datasets, requiring approximately 45-50 hours of work in total, due in approximately week 6 and week 11, (40%). Addresses Intended Learning Outcomes, (ILO) 1, 2 and 3. One 5-minute workshop presentation, requiring approximately 10-12 hours of work in total, presented during semester, (10%). Addresses ILO 3. One 2-hour end-of-semester examination,(50%). Addresses ILO 1 and 2. Hurdle requirement. To pass the subject, students must obtain at least: 20 / 50 in the continuous assessment 20 / 50 in the end-of-semester written examination
Prescribed Texts:	None
Recommended Texts:	None
Breadth Options:	This subject is not available as a breadth subject.
Fees Information:	Subject EFTSL, Level, Discipline & Census Date, http://enrolment.unimelb.edu.au/fees
Generic Skills:	On completion of this subject, students should have developed the following generic skills: # An ability to apply knowledge of basic science and engineering fundamentals # An ability to undertake problem identification, formulation and solution # The capacity to solve problems, including the collection and evaluation of information # The capacity for critical and independent thought and reflection # Profound respect for truth and intellectual integrity, and for the ethics of scholarship An expectation of the need to undertake lifelong learning, and the capacity to do so.
Notes:	**EARNING AND TEACHING METHODS** **INDICATIVE KEY LEARNING RESOURCES** **CAREERS / INDUSTRY LINKS**
Related Majors/Minors/ Specialisations:	Science-credited subjects - new generation B-SCI and B-ENG.