# Table of Contents

**FIT5043 Distributed and big data processing - Semester 2, 2015**

- Mode of Delivery ................................................................. 1
- Workload Requirements ...................................................... 1
- Unit Relationships ............................................................... 1
  - Prohibitions ........................................................................ 1
  - Prerequisites ...................................................................... 1
- Chief Examiner ..................................................................... 1
- Campus Lecturer ................................................................. 1
  - Caulfield ........................................................................... 1
- Your feedback to Us ............................................................ 2
- Previous Student Evaluations of this Unit .......................... 2

**Academic Overview** ................................................................ 3
  - Learning Outcomes ......................................................... 3

**Unit Schedule** ..................................................................... 4
  - Teaching Approach .......................................................... 4
  - Assessment Summary ....................................................... 4

**Assessment Requirements** ................................................ 5
  - Assessment Policy ............................................................ 5
  - Assessment Tasks ............................................................. 5
    - Participation .................................................................... 5
  - Examinations .................................................................... 6
    - Examination 1 ............................................................... 6
  - Learning resources ............................................................. 6
  - Reading list ........................................................................ 6
  - Feedback to you ............................................................... 6
  - Extensions and penalties .................................................. 6
  - Returning assignments ..................................................... 6
  - Referencing requirements ................................................ 7
  - Assignment submission ................................................... 7
  - Online submission ........................................................... 7
  - Required Resources .......................................................... 7

**Other Information** ............................................................ 8
  - Policies ............................................................................. 8
    - Faculty resources and policies ........................................ 8
      - Graduate Attributes Policy ........................................... 8
  - Student Charter ............................................................... 8
  - Student services .............................................................. 8
  - Monash University Library ................................................ 8
  - Disability Liaison Unit ...................................................... 8
FIT5043 Distributed and big data processing - Semester 2, 2015

This unit focuses on designing, developing and deploying distributed data management systems. The unit introduces various contemporary issues in managing large data in distributed environment such as data model, fragmentation, replication and concurrency control. The issues will be discussed using several contemporary approaches of distributed data management.

Mode of Delivery

Caulfield (Day)

Workload Requirements

Minimum total expected workload equals 12 hours per week comprising:

(a.) Contact hours for on-campus students:

• Two hours of lectures
• One 2-hour laboratory

(b.) Additional requirements (all students):

• A minimum of 8 hours independent study per week for completing lab and project work, private study and revision.

See also Unit timetable information

Unit Relationships

Prohibitions

CSE5200

Prerequisites

((FIT9131 or FIT5131 or FIT9017) and (FIT9132 or FIT5132 or FIT9003 or FIT9019)) or equivalent

Chief Examiner

Dr Pari Delir Haghighi

Campus Lecturer

Caulfield

Pari Delir Haghighi

Joseph Liu
Your feedback to Us

Monash is committed to excellence in education and regularly seeks feedback from students, employers and staff. One of the key formal ways students have to provide feedback is through the Student Evaluation of Teaching and Units (SETU) survey. The University’s student evaluation policy requires that every unit is evaluated each year. Students are strongly encouraged to complete the surveys. The feedback is anonymous and provides the Faculty with evidence of aspects that students are satisfied and areas for improvement.

For more information on Monash’s educational strategy, see:

www.monash.edu.au/about/monash-directions/ and on student evaluations, see:
www.policy.monash.edu/policy-bank/academic/education/quality/student-evaluation-policy.html

Previous Student Evaluations of this Unit

Previous feedback has been mainly encouraging and positive and to improve the unit further last year we rebranded the unit and introduced the big data topic as the second part of this unit where it covers the big data topic and current key technologies including Map Reduce and Hadoop. Students find this very useful for their future career.

If you wish to view how previous students rated this unit, please go to https://emuapps.monash.edu.au/unitevaluations/index.jsp
Academic Overview

Learning Outcomes

At the completion of this unit, students should be able to:

- describe the principles of distributed data management such as partition, replication and concurrency;
- identify the current research directions in the field;
- describe and contrast the current available models, technologies for and approaches to building distributed data management systems;
- develop a system that applies appropriate models and technologies to solve practical data management problems.
# Unit Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Activities</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Students need to allocate themselves to tutorials and visit the unit website</td>
<td>No formal assessment or activities are undertaken in week 0</td>
</tr>
<tr>
<td>1</td>
<td>Introduction to distributed database systems and computing</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Distributed database design</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Distributed query processing and optimization</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Transaction Management in Distributed DB</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Distributed Concurrency Control</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Reliability and Recovery, and Replication</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Big Data</td>
<td>Practical Assignment Due</td>
</tr>
<tr>
<td>8</td>
<td>GFS and Map Reduce</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Hadoop</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Hadoop Ecosystem</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Security in Distributed and Big Data Processing</td>
<td>Big Data Assignment and Analysis Report Due</td>
</tr>
<tr>
<td>12</td>
<td>Review and Exam Questions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SWOT VAC</td>
<td>No formal assessment is undertaken in SWOT VAC</td>
</tr>
</tbody>
</table>

*Unit Schedule details will be maintained and communicated to you via your learning system.

## Teaching Approach

### Lecture and tutorials or problem classes

The teaching and learning approach provides facilitated learning, practical exploration and peer learning.

## Assessment Summary

Examination (3 hours): 50%, In-semester assessment: 50%

<table>
<thead>
<tr>
<th>Assessment Task</th>
<th>Value</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practical Assignment</td>
<td>30%</td>
<td>Week 7</td>
</tr>
<tr>
<td>Big Data Assignment and Analysis Report</td>
<td>20%</td>
<td>Week 11</td>
</tr>
<tr>
<td>Examination 1</td>
<td>50%</td>
<td>To be advised</td>
</tr>
</tbody>
</table>
**Assessment Requirements**

**Assessment Policy**

Faculty Policy - Unit Assessment Hurdles

Academic Integrity - Please see resources and tutorials at
http://www.monash.edu/library/skills/resources/tutorials/academic-integrity/

**Assessment Tasks**

**Participation**

- **Assessment task 1**
  
  **Title:** Practical Assignment
  
  **Description:** Learning outcomes: 1 and 4. Students will design and build a distributed relational database that its data integrity and distribution is maintained. They will also create a client-server, distributed database application as the front end using NetBeans connecting to Oracle server.
  
  **Weighting:** 30%
  
  **Criteria for assessment:**
  
  How well the database is designed and implemented.
  
  How well the PL/SQL and Java are written.
  
  How well the constraints, triggers, procedures and functions are designed and implemented.
  
  **Due date:** Week 7
  
- **Assessment task 2**
  
  **Title:** Big Data Assignment and Analysis Report
  
  **Description:** Learning outcomes: 2 and 3. This task requires students to conduct simple data processing and queries using Big Data technologies (Hadoop ecosystem in Hortonworks), compare them, and write a short report.
  
  **Weighting:** 20%
  
  **Criteria for assessment:**
  
  Group Assignment (groups of 2)
  
  Students will be assessed based on:
  
  ♦ How well the data processing and query tasks are completed.
  ♦ How well the technologies are understood, compared and analysed.
The assignment has a group mark (50%), and an individual component (50%) which is determined based on the Task Allocation and peer review responses.

**Due date:**
Week 11

**Examinations**

- **Examination 1**
  - **Weighting:**
    - 50%
  - **Length:**
    - 3 hours
  - **Type (open/closed book):**
    - Closed book
  - **Electronic devices allowed in the exam:**
    - None

**Learning resources**

**Reading list**


Monash Library Unit Reading List (if applicable to the unit)
http://readinglists.lib.monash.edu/index.html

**Feedback to you**

Types of feedback you can expect to receive in this unit are:

- Informal feedback on progress in labs/tutes
- Graded assignments with comments

**Extensions and penalties**

Submission must be made by the due date otherwise penalties will be enforced.

You must negotiate any extensions formally with your campus unit leader via the in-semester special consideration process: http://www.monash.edu.au/exams/special-consideration.html

**Returning assignments**

Students can expect assignments to be returned within two weeks of the submission date or after receipt, whichever is later.
Assessment Requirements

Referencing requirements

APA style is recommended.

Refer to:


Assignment submission

It is a University requirement (http://www.policy.monash.edu/policy-bank/academic/education/conduct/student-academic-integrity-managing-plagiarism.html) for students to submit an assignment coversheet for each assessment item. Faculty Assignment coversheets can be found at http://www.infotech.monash.edu.au/resources/student/forms/. Please check with your Lecturer on the submission method for your assignment coversheet (e.g. attach a file to the online assignment submission, hand-in a hard copy, or use an electronic submission). Please note that it is your responsibility to retain copies of your assessments.

Online submission

If Electronic Submission has been approved for your unit, please submit your work via the learning system for this unit, which you can access via links in the my.monash portal.

Required Resources

Please check with your lecturer before purchasing any Required Resources. Limited copies of prescribed texts are available for you to borrow in the library, and prescribed software is available in student labs.

Netbeans 8.0.2 and Hortonworks (with VMPlayer)

Software will be available in computer labs
Other Information

Policies

Monash has educational policies, procedures and guidelines, which are designed to ensure that staff and students are aware of the University’s academic standards, and to provide advice on how they might uphold them. You can find Monash’s Education Policies at:

Faculty resources and policies

Important student resources including Faculty policies are located at
http://intranet.monash.edu.au/infotech/resources/students/

Graduate Attributes Policy

http://www.policy.monash.edu/policy-bank/academic/education/management/monash-graduate-attributes-policy.html

Student Charter


Student services

The University provides many different kinds of support services for you. Contact your tutor if you need advice and see the range of services available at http://www.monash.edu.au/students. For Malaysia see http://www.monash.edu.my/Student-services, and for South Africa see http://www.monash.ac.za/current/.

Monash University Library

The Monash University Library provides a range of services, resources and programs that enable you to save time and be more effective in your learning and research. Go to www.lib.monash.edu.au or the library tab in my.monash portal for more information. At Malaysia, visit the Library and Learning Commons at http://www.lib.monash.edu.my/. At South Africa visit http://www.lib.monash.ac.za/

Disability Liaison Unit

Students who have a disability or medical condition are welcome to contact the Disability Liaison Unit to discuss academic support services. Disability Liaison Officers (DLOs) visit all Victorian campuses on a regular basis.

- Website: http://www.monash.edu/equity-diversity/disability/index.html
- Telephone: 03 9905 5704 to book an appointment with a DLO; or contact the Student Advisor, Student Community Services at 03 55146018 at Malaysia
- Email: dlu@monash.edu
- Drop In: Equity and Diversity Centre, Level 1, Building 55, Clayton Campus, or Student Community Services Department, Level 2, Building 2, Monash University, Malaysia Campus