

Subject Description Form

Subject Code	COMP5434
Subject Title	Big Data Computing
Credit Value	3
Level	5
Pre-requisites	Knowledge in database systems, machine learning and data analytics is preferred.
Objectives	<p>The objectives of this subject are to:</p> <ol style="list-style-type: none"> 1. introduce students the concept and challenge of big data; 2. teach students in applying skills and tools to manage and analyze the big data.
Intended Learning Outcomes	<p>Upon completion of the subject, students will be able to:</p> <ol style="list-style-type: none"> a. understand the concept and challenge of big data and why traditional technology is inadequate to analyze the big data; b. understand how to collect, manage, store, query, and analyze various form of big data; and c. familiar with large-scale analytics tools to solve some open big data problems; and d. understand the impact of big data for business decisions and strategy.
Subject Synopsis/ Indicative Syllabus	<ol style="list-style-type: none"> 1. Introduction to Big Data: Different V's, their challenges and application domains. 2. Cloud Computing Basics: Software as a service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), Desktop as a Service (DaaS), Public, Private and Enterprise Cloud. 3. Big Data Computing: Concepts, Platform, Service, and Tools 4. Large-Scale Programming Abstraction: MapReduce and its open source implementation of Hadoop 5. Large-Scale Data Processing Framework: Apache Spark and its Built-in Modules 6. Large-Scale Database Management: NoSQL and other tools, e.g. MongoDB, Google BigTable, etc. 7. Machine Learning Systems for Big Data: Methods and Tools 8. Big Data Visualization: Data types and dimensions; Visual encoding and perception 9. Big Data Case Studies

Teaching/Learning Methodology	<p>A mix of lectures, discussions and case studies.</p> <p>Class activities include lectures, tutorials, laboratory works and seminars.</p>																																									
Assessment Methods in Alignment with Intended Learning Outcomes	<table border="1" data-bbox="520 362 1474 945"> <thead> <tr> <th data-bbox="520 362 823 568" rowspan="2">Specific assessment methods/tasks</th> <th data-bbox="823 362 979 568" rowspan="2">% weighting</th> <th colspan="4" data-bbox="979 362 1474 501">Intended subject learning outcomes to be assessed (Please tick as appropriate)</th> </tr> <tr> <th data-bbox="979 501 1098 568">a</th> <th data-bbox="1098 501 1222 568">b</th> <th data-bbox="1222 501 1347 568">c</th> <th data-bbox="1347 501 1474 568">d</th> </tr> </thead> <tbody> <tr> <td data-bbox="520 568 823 672">1. Assignments or lab works</td> <td data-bbox="823 568 979 801" rowspan="3">55</td> <td data-bbox="979 568 1098 672">✓</td> <td data-bbox="1098 568 1222 672">✓</td> <td data-bbox="1222 568 1347 672">✓</td> <td data-bbox="1347 568 1474 672">✓</td> </tr> <tr> <td data-bbox="520 672 823 734">2. Project</td> <td data-bbox="979 672 1098 734">✓</td> <td data-bbox="1098 672 1222 734">✓</td> <td data-bbox="1222 672 1347 734">✓</td> <td data-bbox="1347 672 1474 734">✓</td> </tr> <tr> <td data-bbox="520 734 823 801">3. Quiz</td> <td data-bbox="979 734 1098 801">✓</td> <td data-bbox="1098 734 1222 801">✓</td> <td data-bbox="1222 734 1347 801">✓</td> <td data-bbox="1347 734 1474 801"></td> </tr> <tr> <td data-bbox="520 801 823 869">4. Examination</td> <td data-bbox="823 801 979 869">45</td> <td data-bbox="979 801 1098 869">✓</td> <td data-bbox="1098 801 1222 869">✓</td> <td data-bbox="1222 801 1347 869"></td> <td data-bbox="1347 801 1474 869">✓</td> </tr> <tr> <td data-bbox="520 869 823 945">Total</td> <td data-bbox="823 869 979 945">100</td> <td colspan="4" data-bbox="979 869 1474 945"></td> </tr> </tbody> </table> <p data-bbox="520 999 1474 1061">Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:</p> <p data-bbox="520 1084 1474 1357">Continuous assessments consist of a project, assignments, lab exercises, and quizzes, which are designed to facilitate students to achieve intended learning outcomes. Lab exercise is designed to encourage students to acquire good understanding of the relevant knowledge, practice in order to enrich their hands-on experience with various software tools. The project is designed to enhance students' ability to acquire the understanding and using different knowledge, principles, techniques, tools to solve a real problem through team. Quizzes are to ensure the students understand the concepts.</p> <p data-bbox="520 1388 1474 1451">Examination will evaluate student's understanding and usage of big data technologies.</p>				Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)				a	b	c	d	1. Assignments or lab works	55	✓	✓	✓	✓	2. Project	✓	✓	✓	✓	3. Quiz	✓	✓	✓		4. Examination	45	✓	✓		✓	Total	100				
Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)																																								
		a	b	c	d																																					
1. Assignments or lab works	55	✓	✓	✓	✓																																					
2. Project		✓	✓	✓	✓																																					
3. Quiz		✓	✓	✓																																						
4. Examination	45	✓	✓		✓																																					
Total	100																																									
Student Study Effort Expected	Class contact:																																									
	Class activities (lecture, tutorial, lab, etc.)		39 Hrs.																																							
	Other student study effort:																																									
	Assignments, Quizzes, Projects, Examination		65 Hrs.																																							
	Total student study effort		104 Hrs.																																							
Reading List and References	<ol style="list-style-type: none"> 1. Jared Dean, Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners. Wiley, 2014. 2. EMC Education Services (Editor), Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley, 2015. 3. Stonebraker et al., "MapReduce and Parallel DBMS's: Friends or Foes?", Communications of the ACM, January 2010. 4. How Vertica Was the Star of the Obama Campaign, and Other Revelations 																																									

- | | |
|--|--|
| | <ol style="list-style-type: none">5. Cohen et al. "MAD Skills: New Analysis Practices for Big Data", 20096. Dean and Ghemawat, "MapReduce: A Flexible Data Processing Tool", Communications of the ACM, January 2010.7. Rick Cattell, "Scalable SQL and NoSQL Data Stores", SIGMOD Record, December 2010 (39:4)8. Leskovec, Rajaraman, Ullman, Mining of Massive Datasets, 2nd Ed., Cambridge University Press, 2014.9. Pedro Domingos, A Few Useful Things to Know about Machine Learning, CACM 55(10), 2012 |
|--|--|