Subject Description Form

Subject Code	COMP5434
Subject Title	Big Data Computing
Credit Value	3
Level	5
Pre-requisites	Background in Database Computing
Objectives	The objectives of this subject are to: 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	Upon completion of the subject, students will be able to: (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	1. Introduction to Big Data: Different V's, their challenges and application domains. 2. Big Data Computing: Concepts, Platform, Service, and Tools 3. Large-Scale Programming Abstraction: MapReduce and its open source implementation of Hadoop 4. Large-Scale Data Processing Framework: Apache Spark and its Built-in Modules 5. Large-Scale Database Management: NoSQL and other tools, e.g. MongoDB, Google BigTable, etc. 6. Machine Learning Systems for Big Data: Methods and Tools 7. Big Data Visualization: Data types and dimensions; Visual encoding and perception 8. Big Data Case Studies

<table>
<tr><td>Teaching/Learning Methodology</td><td>A mix of lectures, discussions and case studies.

Class activities include lectures, tutorials, laboratory works and seminars.</td></tr>
<tr><td>Assessment Methods in Alignment with Intended Learning Outcomes</td><td>
<table>
<tr><td rowspan="2">Specific assessment methods/tasks</td><td rowspan="2">% weighting</td><td colspan="6">Intended subject learning outcomes to be assessed (Please tick as appropriate)</td></tr>
<tr><td>a</td><td>b</td><td>c</td><td>d</td><td></td><td></td></tr>
<tr><td>1. Assignments or lab works</td><td rowspan="3">55%</td><td>x</td><td>x</td><td>x</td><td>x</td><td></td><td></td></tr>
<tr><td>2. Project</td><td>x</td><td>x</td><td>x</td><td>x</td><td></td><td></td></tr>
<tr><td>3. Quiz</td><td>x</td><td>x</td><td>x</td><td></td><td></td><td></td></tr>
<tr><td>4. Examination</td><td>45%</td><td>x</td><td>x</td><td></td><td>x</td><td></td><td></td></tr>
<tr><td>Total</td><td>100 %</td><td colspan="6"></td></tr>
</table>

Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:

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.

Examination will evaluate student's understanding and usage of big data technologies.</td></tr>
<tr><td>Student Study Effort Expected</td><td>
<table>
<tr><td>Class contact:</td><td></td></tr>
<tr><td>Class activities (lecture, tutorial, lab, etc.)</td><td>39 Hrs.</td></tr>
<tr><td>Other student study effort:</td><td></td></tr>
<tr><td>Assigments, Quizzes, Projects, Examination</td><td>71 Hrs.</td></tr>
<tr><td>Total student study effort</td><td>110 Hrs.</td></tr>
</table>
</td></tr>
<tr><td>Reading List and References</td><td>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</td></tr>
</table>

	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 5. Cohen et al."MAD Skills: New Analysis Practices for Big Data", 2009 6. 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