Subject Description Form

Subject Code	DSAI5204
Subject Title	Efficient Data Processing
Credit Value	3
Level	5
Pre-requisite/ Co-requisite/ Exclusion	Nil
Objectives	The objectives of this subject are: 1. apply data structures, searching and sorting algorithms in developing computer programs; 2. use a database for efficient data storage and retrieval; 3. explore efficient data processing in distributed environments.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: a. demonstrate a comprehensive understanding of data structures, searching and sorting algorithms; b. use databases properly in real-life problems and apply optimization techniques to enhance query performance; c. apply the techniques for distributed database design and query processing in real-life problems.
Subject Synopsis/ Indicative Syllabus	**1. Introduction to Data Structures** • Data structures: stack, queue, linked-list, heap, tree. **2. Efficient Data Processing in Computer Programs** • Searching algorithms: linear search, binary search; • Sorting algorithms: bubble sort, insertion sort, selection sort, quick sort, merge sort, heap sort; • Big-O notation and time complexity analysis. **3. Introduction to Databases** • Relational model; relational algebra; SQL; integrity constraints. **4. Efficient Data Processing in Databases** • File organization; indexing; view materialization and maintenance; • Query processing and optimization; • Transaction management and concurrency control; two-phase locking protocol. **5. Introduction to Distributed Databases** • DDB architecture; DDB design: data fragmentation; data allocation; data replication; consistency issues. **6. Efficient Data Processing in Distributed Databases** • DDB query processing and optimization; query decomposition; data localization; join strategies in DDB; • DDB transaction management and concurrency control; two-phase commit protocol.

<table>
<tr><td>Teaching/Learning Methodology</td><td>This subject emphasizes the technical aspects of data structures and practical aspects of database systems. It is intended to equip students with knowledge and experience on solving real-life problems by using data structures, centralized databases, and distributed databases.

The lectures will be used to deliver course materials.

Labs and tutorials will be used to practice exercises.</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="3">Intended subject learning outcomes to be assessed (Please tick as appropriate)</td></tr>
<tr><td>a</td><td>b</td><td>c</td></tr>
<tr><td>1. Quizzes and assignments</td><td>60%</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>2. Exam</td><td>40%</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Total</td><td>100 %</td><td colspan="3"></td></tr>
</table>
Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:

Quizzes and assignments provide continuous assessment opportunities, allowing students to demonstrate their understanding of data structures and databases in a practical and iterative manner. These assessments encourage regular engagement with the subject material and enable students to apply the learned techniques to real-life problems, thereby directly addressing the learning outcomes. The exam serves as a comprehensive evaluation of the students' overall understanding and ability to synthesize and apply the knowledge gained throughout the course. This balance ensures that students are assessed on both their ongoing learning process and their cumulative knowledge, providing a holistic measure of their proficiency in the subject.</td></tr>
<tr><td>Student Study Effort Expected</td><td>
<table>
<tr><td>Class contact:</td><td></td></tr>
<tr><td>▪ Lecture</td><td>26 Hrs.</td></tr>
<tr><td>▪ Tutorial/Lab</td><td>13 Hrs.</td></tr>
<tr><td>Other student study effort:</td><td></td></tr>
<tr><td>▪ Assignments, reading book chapters</td><td>66 Hrs.</td></tr>
<tr><td>Total student study effort</td><td>105 Hrs.</td></tr>
</table>
</td></tr>
</table>

Reading List and References	1. Brad Miller and David Ranum, Problem Solving with Algorithms and Data Structures Using Python, 2nd Edition, 2011. 2. Frank M. Carrano, Data Abstraction & Problem Solving with C++: Walls & Mirrors, 7th Edition, Pearson, 2017. 3. Michael T. Goodrich, Roberto Tamassia, Michael H. Goldwasser, Data Structures and Algorithms in Java, 6th Edition, John Wiley, 2014. 4. A Silberschatz, H.F. Korth, S. Sudarshan, Database System Concepts, 6th Edition, McGraw Hill, 2011. 5. Hector Garcia-Molina, Jeffrey D. Ullman & Jennifer Widom, Database System Implementation, Prentice Hall, 3rd Edition, 2008. 6. M. Tamer Ozsu, Principles of Distributed Database Systems, 3rd Edition, Springer, 2011.