**Detailed Syllabus**

Lecture-wise Breakup

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Course Code** | 21B12CS314 | **Semester** Even  **(specify Odd/Even)** | | **Semester** VI **Session** 2023 - 2024 Month from Jan 2024 to June 2024 | |
| **Course Name** | Introduction to Large Scale Database Systems | | | | |
| **Credits** | **3** | | **Contact Hours** | | **3-0-0** |

|  |  |  |
| --- | --- | --- |
| **Faculty (Names)** | **Coordinator(s)** | Dr. Devpriya Soni |
| **Teacher(s) (Alphabetically)** | Dr. Devpriya Soni, Dr. Payal Khurana Batra |

|  |  |  |
| --- | --- | --- |
| **COURSE OUTCOMES** | | **COGNITIVE LEVELS** |
| **C331-3.1** | Infer the background processes involved in queries and transactions, and explain how these impact on database operation and design | Understand level  (Level 2) |
| **C331-3.2** | Choose appropriate ways of storing data and optimize queries. | Analyze level  (Level4) |
| **C331-3.3** | Explain the concept and challenge of big data and demonstrate the comparison of relational database systems with NoSQL databases | Understand level  (Level 2) |
| **C331-3.4** | Compare and discover the suitability of appropriate large databases to manage, store, query, and analyze various form of big data | Analyze level  (Level4) |
| **C331-3.5** | Apply techniques for data fragmentation, replication, and allocation to design a distributed or parallel database system | Apply Level  (Level3) |

|  |  |  |  |
| --- | --- | --- | --- |
| **Module No.** | **Title of the Module** | **Topics in the Module** | **No. of Lectures for the module** |
| **1.** | Introduction to large scale Databases | Review of database systems, Data sources and join processing, modelling and query languages | 2 |
| **2.** | Transaction management | Transaction processingconcepts, Concurrency control techniques and protocols | 4 |
| **3.** | Data Storage and Indexing | Data storage and indexing of massive databases in databases and data warehouses. Introduction to technologies for handling big data | 7 |
| **4.** | Query processing and Optimization | Measures of query cost, Evaluation of expressions, Query planning, evaluation and optimization | 5 |
| **5.** | Big data Tools and Technologies | Review of Big data, CAP Theorem (consistency, availability, partition tolerance), Using big data in businesses, Data visualization for data analysis, NoSQL databases | 7 |
| **6.** | Hadoop and its Ecosystem | Hadoop core components, Hadoop Ecosystem components, Data storage and processing in Hadoop framework | 5 |
| **7.** | Application-driven databases | Parallel and Distributed databases, Distributed Database Design, Architecture of Distributed DBMS | 8 |
| **8.** | Advanced databases | Graph databases, spatial and temporal databases | 4 |
| **Total number of Lectures** | | | **42** |
| **Evaluation Criteria**  **Components Maximum Marks**  T1 20  T2 20  End Semester Examination 35  TA 25Attendance (10 Marks), Assignment/Quiz/Mini-project (15 Marks)  **Total 100** | | | |

Project based Learning: Each student in a group of two or three student will explore a large database from the domain of their choice. For real time applicability of subject, they will explore and choose one visualization tool available. The chosen visualization tool will be used for analyzing the database. Understanding the data visualization process, will help in their employability in big data analysis organizations.

|  |  |
| --- | --- |
| **Recommended Reading material:** Author(s), Title, Edition, Publisher, Year of Publication etc. (Text books, Reference Books, Journals, Reports, Websites etc. in the IEEE format) | |
| **Text Books** | |
| **1.** | [AviSilberschatz](http://www.cs.yale.edu/homes/avi), [Henry F. Korth](http://www.cse.lehigh.edu/~korth), [S. Sudarshan](http://www.cse.iitb.ac.in/~sudarsha), Database System Concepts, Seventh Edition, [McGraw-Hill](http://www.mhcollege.com/), March 2019. |
| **2.** | [RamezElmasri](https://www.amazon.com/Ramez-Elmasri/e/B000APV0OK/ref=dp_byline_cont_book_1), [Shamkant B. Navathe](https://www.amazon.com/s/ref=dp_byline_sr_book_2?ie=UTF8&field-author=Shamkant+B.+Navathe&text=Shamkant+B.+Navathe&sort=relevancerank&search-alias=books), Fundamentals of Database Systems (7th Edition) 7th Edition, Pearson Education (June 18, 2015), ISBN-10: 0133970779, ISBN-13: 978-0133970777. |
| **3.** | Sadalage, P.J.  &Foowlwer, M. 2013.  NoSQL distilled: a brief guide to the emerging world of polygot persistence.  Addison-Wesley |
| **4.** | White, Tom. Hadoop: The definitive guide. " O'Reilly Media, Inc.", 2012. |
| **5.** | Zikopoulos, Paul, and Chris Eaton. Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media, 2011. |
| **6.** | Shashank Tiwari, Professional NoSQL, Wiley, 2011 |
| **Reference Books** | |
| **1.** | Rick, Smolan, and Jennifer Erwitt. "The human face of big data." Against All Odds Production (2012). |
| **2.** | Prajapati, Vignesh. Big data analytics with R and Hadoop. Packt Publishing Ltd, 2013. |
| **3.** | Provost, Foster, and Tom Fawcett. Data Science for Business: What you need to know about data mining and data-analytic thinking. " O'Reilly Media, Inc.", 2013. |
| **4.** | DeRoos, Dirk. Hadoop for dummies. John Wiley & Sons, 2014. |
| **5.** | Mayer-Schönberger, Viktor, and Kenneth Cukier. Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, 2013. |