**BIG DATA TOOLS FOR MANAGERS**

|  |  |  |  |
| --- | --- | --- | --- |
| Subject Code | **N2MBA07** | CIE Marks\* | 50 |
| Credits (Lecture Hrs/week) | 3 | SEE Marks | 50 |
| Total no. of Lecture Hrs | 39 | Exam Duration | 01 Hour |

\* CIE will be based on the lab exercises

**Course Objectives:**

* To familiarize the basic concepts, evolution, technology and applications of Big Data.
* To educate the design of data Model and use MYSQL for designing, querying and manipulation applications.
* To build the requisite skills in R-studio.
* To introduce the basics of Python programming language
* To solve statistical and managerial problems using different tools.

**Pedagogy:** Lectures, Case Studies, Data analysis through hands-on sessions, exercises and activities.

**UNIT 1 (7 Hours)**

**Overview of Big Data:** Meaning of Big data, History of Data, Management – Evolution of Big Data, Structuring Big Data, Types of Data, Elements of Big Data, Volume, Velocity, Variety, Veracity, Big Data Analytics, Advantages of Big Data Analytics, Future of Big Data.

**Databases and Data users:** Introduction; characteristics of the database approach; actors on the scene(Users); workers behind the scene; advantages of using the DBMS approach. Concepts of data models and schemas. E-R (Entity-Relationship) Model.

**UNIT 2 (8 Hours)**

**Data Querying and Retrieval using SQL (Structured Query Language)**

Concepts of Data Definition Language (DDL) and Data Manipulation Language (DML), Data Dictionary, SQL, SQL Data Definition and Data Types, Specifying Basic Constraints in SQL, Schema Change Statements in SQL (DROP, ALTER command); Basic Queries in SQL; Insert, Delete and Update Statements in SQL; Additional Features of SQL; Views (Virtual Tables) in SQL. (Includes Hands-on sessions on Databases SQL commands and queries using MYSQL tool).

**UNIT 3 (8 Hours)**

**Introduction to R**

R-Studio interface. Importing data into R – text files, Excel, from other statistical software packages, from databases, and from the web. Viewing data. Basic data types in R. Vectors, Matrices, Data frames and Lists. Categorical data – factors, discretizing variables.

**Hands-on sessions and exercises using R-studio:** Descriptive Statistics, Data Visualization, Correlation and Regression and Statistical tests (t-test, Chi-Square test, one-way ANOVA).

**UNIT 4 (7 Hours)**

**Introduction to Python with Practical Sessions**

Programming essentials: Types of programming, Execution process of a program, Installation and working with Python – input, processing, and output; Python script files; correcting syntax errors; data types and expressions – strings, variables, assignment, operators, logical operators, Boolean expressions and type conversions; Control statements: for loops – count-controlled, augmented assignment, steps; if-else statements – one-way, multiway (elif); while loops – break, loop logic, errors and testing; Using functions and modules – arguments and return values.

**UNIT 5 (9 Hours)**

**Python: Hands-on sessions**

**Operators: Boolean** and logical operators

**String and text files:** string concatenation, subscript operator, indexing, slicing a string; string methods, manipulating files and directories; text files: reading/writing text and numbers from/to a file.

**Lists, Dictionaries and Tuples:** Basic operations and commands on Lists, Dictionaries and Tuples.

**Time series analysis and forecasting: Time** series analysis & forecasting model, Time Data Visualization

**Text analysis**

**Course Outcomes:**

CO1: Explain basics of Big-data and Databases.

CO2: Retrieve information using MYSQL.

CO3: Employ R for data analysis and visualization.

CO4: Explain the fundamental concepts of python

CO5: Employ python for managerial or mining applications.

**RECOMMENDED BOOK**

* BIG DATA Management and Analytics, Nitin Upadhyay, Cengage Publication,2018 Edition
* Elmasri and Navathe. Fundamentals of Database Systems. Pearson Education, 7th Edition,2016
* Python for Everybody: Exploring Data Using Python 3. Charles R. Severance Create Space Independent Publishing Platform 1 st Edition, 2016
* Wickham H., Grolemund G. (2016). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O’Reilly Media.

**Reference Book:**

* BIG DATA Black Book D T Editorial Services, Dreamtech press 2016 Edition
* Big Data and Analytics Seema Acharya, Subhashini Chellappan, Wiley India Publications, May 2015
* Gillenson, M. L., Ponniah, P., Kriegel, A., Trukhov, B. M., Taylor, A. G., Powell, G., & Miller, F. (2013). Introduction to Database Management. Sahibabad: Wiley India Pvt. Ltd.
* Abraham Silberschatz, Henry F Korth, Sudharshan.Data base System Concepts. Mc GrawHil (Indian edition) 6th Edition,2013
* Leon, A., & Leon, M. (2010). Fundamentals of Database Management Systems, McGraw Hill Education (India) Pvt. Ltd.
* Raghu Ramakrishnan and Johannes Gehrke Database Management Systems, McGraw-Hill Education, 3rd Edition,2014
* Lambert KA., Juneja BL. (2015). Fundamentals of Python. Cengage Learning.
* McKinney W (2018). Python for Data Analysis. 2nd Edition. O’Reilly Media.
* Cotton, R. (2013). Learning R: A Step-by-Step Function Guide to Data Analysis 1st Edition [Kindle Version]. Retrieved from http://www.amazon.in.
* Knell, R. (2013) Introductory R: A Beginner's Guide to Data Visualization, Statistical Analysis and Programming in R. [Kindle Version]. Retrieved from http://www.amazon.in.
* Murray, S. (2013) Learn R in a Day. [Kindle Version]. Retrieved from http://www.amazon.in