**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 | 1.5 hours |

\* 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:** Types of Digital data – Structured, Semi-structured, Unstructured data, characteristics of Data, Evolution of Big Data, Definition of Big Data, Challenges with Big Data, what is Big Data – Volume, Velocity, Variety, Veracity, Why Big Data, Common software/tool which deals with Big Data.

**Databases:** Introduction to Data, Database; Characteristics of the database approach, Database Management Software, Introduction to Relation DBMS–Tables, Attributes, Tuples. Introduction to MySQL tool, MySQL tool Installation, Getting familiar with MySQL windows.

**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:** Introduction to R, features of R, Installation of R & RStudio, Getting started – Window sections of RStudio, First interaction, Command line versus Scripts, Comments, help in R, Directory, Output in R, Packages, Functions (call, arguments, return values), Data types of R - Vectors, Matrices, Lists, DataFrame.

**Importing data into R**: Text, CSV, Excel, SPSS, SAS files, and from the web. Viewing data.

**Hands-on sessions and exercises using R-studio:** Descriptive Statistics, Data Visualization, Correlation and Regression (simple & multiple linear regression).

**UNIT 4 (7 Hours)**

**Introduction to Python with Practical Sessions**

**Programming essentials:** Introduction to Python, Python Stack for Data Analysis, Installation and working with Python, working with Jupyter notebook, Import statement, PIP utilities, download package from internet, Declaring Variables, Basic operators, Conditional Statements (if..else..else if statements), Generating Sequence Numbers, Control flow statements (loops – for, while), Functions – Call, Arguments and return values.

**Working with Python collections:** Lists, Tuples, Set, Dictionary

**UNIT 5 (9 Hours)**

**Python: Hands-on sessions**

**Descriptive Statistics – Structured Data:** Introduction to Pandas package, Working with tabular dataset - Reading & Writing CSV/Excel/Text files, Displaying records, Quick summary of data, Indexing, Value counts and Cross tabulations, Sorting, Creating new columns, Grouping, Aggregating, Joining various dataset, Remove columns/rows, Filtering records.

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

**Text analysis – Unstructured Data:** Introduction to Text Analysis, loading & exploring the textual dataset.

**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
* Data Analytics with R, Dr. Bharti Motwani ISBN: 9788126576463, Wiley
* Machine Learning using Python, by Manaranjan Pradhan, U Dinesh Kumar

**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
* 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.
* 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