

**Academic Council Meeting No. and Date : September 04, 2023**

**Agenda Number : 02**

**Resolution Number : 34, 35 / 2.20, 2.41**



**Vidya Prasarak Mandal's  
B. N. Bandodkar College of  
Science (Autonomous), Thane**



**Syllabus for**  
**Programme : Master of Science**  
**Specific Programme : Data Science**

**[M.Sc. (Data Science) Part I]**

**Revised under NEP**  
**From academic year 2023 - 2024**

<b>Sr. No.</b>	<b>Heading</b>	<b>Particulars</b>
1.	Title of the Course	M.Sc. (Data Science)
2.	Eligibility for Admission	Students who have Completed B.Sc. with Data Science, Information Technology, Computer Science, Physics, Mathematics, Statistics from recognized University / Board / Institutions. Candidate should have programming background / concepts learned.
3.	Passing Marks	40%
4.	Ordinances / Regulations (if, any)	Existing ordinances and regulations.
5.	Number of years / Semesters	Two years – Four Semesters
6.	Level	PG
7.	Pattern	Semester
8.	To be implemented from Academic year	2023 – 2024
9.	Mode of conduct	Offline / Online Lectures / Practicals

## **Preamble**

Data Science is considered one of the trendiest courses in the world. It leads to the right skills and knowledge required to become a successful **Data Analyst** or **Scientist**. Data science is a “concept to unify statistics, data analysis, machine learning and their related methods” in order to “understand and analyze actual phenomena” with data. It’s study of the flow of information from structured and unstructured data available with an organization. It involves obtaining the meaningful insights from the data which is processed through analytical study. One needs to understand the language of data through analytical skill. Hence, it is absolutely necessary nowadays, to develop manpower with a skill to perform data analysis to get meaningful information from the data of different domains such as banking and finance, insurance, agriculture, healthcare, retail, education, social media, manufacturing, transportation, entertainment and so on.

**Credit Distribution Structure for Two Years/ One Year PG / M.Sc. ( Data Science )**

Year	Level	Sem	Major				RM	OJT/FP	RP	Cum. Cr.	Degree
			Mandatory			Electives					
1	6.0	Sem I	3*4+1*2			4	4	-	-	22	PG (after 2 Years Degree)
			Fundamentals of Data Science	TH	4	Python Programming (OR) Java Programming	Research Methodology				
			Databases and Data Warehousing	TH	4						
			Introduction to Data Analysis and Statistical Computing	TH	4						
			Fundamentals of Data Science / Database and Data Warehousing Practicals	PR	2						
		Sem II	3*4+1*2			4	-	4	-	22	
			Big Data Analytics	TH	4	Data Mining (OR) Cloud Management					
			Next Generation Databases	TH	4						
			High Performance Computing	TH	4						
			Next Generation Databases / High Performance Computing Practical	PR	2						
Cum. Cr. For PG Diploma			28			8	4	4		44	
Exit Option: PG Diploma (44 credits) after Three Year UG Degree											

# Semester I

Course Code	Course Title	Credits	No. of lectures
23BPDS1T1	Fundamentals of Data Science	04	
<b>Course Objectives (COs):</b> To enable the students to: CO1: Develop in depth understanding of the key technologies in data science and business analytics: data mining, machine learning, visualization techniques, predictive modelling and statistics. CO2: Practice problem analysis and decision-making. CO3: Gain practical, hands-on experience with statistics programming languages and big data tools through coursework and applied research experiences.			
Unit I	<b>INTRODUCTION TO DATA SCIENCE:</b> Data science process – roles, stages in data science project – working with data from files – working with relational databases – exploring data – managing data – cleaning and sampling for modeling and validation – introduction to NoSQL.	15 Hrs [OC1, OC2, OC3]	
Unit II	<b>MODELING METHODS</b> Choosing and evaluating models – mapping problems to machine learning, evaluating clustering models, validating models – cluster analysis – K-means algorithm, Naïve Bayes – Memorization Methods – Linear and logistic regression – unsupervised methods	15 Hrs [OC4, OC5, OC6]	
Unit III	<b>MAP REDUCE:</b> Introduction – distributed file system – algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce – Hadoop - Understanding the Map Reduce architecture - Writing Hadoop MapReduce Programs - Loading data into HDFS - Executing the Map phase - Shuffling and sorting - Reducing phase execution	15 Hrs [OC7, OC8, OC9]	
Unit IV	<b>DELIVERING RESULTS:</b> Documentation and deployment – producing effective presentations – Introduction to graphical analysis – plot() function – displaying multivariate data – matrix plots – multiple plots in one window - exporting graph - using graphics parameters. Case studies	15 Hrs [OC7, OC8, OC9]	

<b>Books and References:</b>					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Practical Data Science	Andreas François Vermeulen	APress		2018
2	Principles of Data Science	Sinan Ozdemir	PACKT		2016
3	Data Science from Scratch	Joel Grus	O'Reilly		2015
4	Data Science from Scratch first Principle in python	Joel Grus	Shroff Publishers		2017
5	Experimental Design in Data science with Least Resources	N C Das	Shroff Publishers		2018

## **Course Outcomes(OCs)**

Upon completing this course, the student will be able to:

OC1: Apply quantitative modeling and data analysis techniques to the solution of real-world business problems, communicate findings, and effectively present results using data visualization techniques.

OC2: Recognize and analyze ethical issues in business related to intellectual property, data security, integrity, and privacy.

OC3: Apply ethical practices in everyday business activities and make well-reasoned ethical business and data management decisions.

OC4: Demonstrate knowledge of statistical data analysis techniques utilized in business decision making.

OC5: Apply principles of Data Science to the analysis of business problems.

OC6: Use data mining software to solve real-world problems.

OC7: Employ cutting edge tools and technologies to analyze Big Data.

OC8: Apply algorithms to build machine intelligence.

OC9: Demonstrate use of team work, leadership skills, decision making and organization theory.

Course Code 23BPDS1T2	Course Title Databases and Data Warehousing	Credits 04	No. of lectures
<b>Course Objectives(COs):</b> <ol style="list-style-type: none"> <li>1. Develop and deploy the database management system to solve some real-world problems</li> <li>2. Describe the database and data warehouse design</li> <li>3. Understand warehousing architectures and tools.</li> </ol>			
<b>Unit I</b>	Database System Architectures: Centralized and Client-Server Architectures – Server System Architectures – Parallel Systems- Distributed Systems – Parallel Databases: I/O Parallelism – Inter and Intra Query Parallelism – Inter and Intra operation Parallelism – Design of Parallel Systems Distributed Database Concepts - Distributed Data Storage – Distributed Transactions – Commit Protocols – Concurrency Control – Distributed Query Processing – Case Studies INTELLIGENT DATABASES Active Databases: Syntax and Semantics (Starburst, Oracle, DB2)- Taxonomy- Applications Design Principles for Active Rules- Temporal Databases: Overview of Temporal Databases TSQL2- Deductive Databases-Recursive Queries in SQL- Spatial Databases- Spatial Data Types	15 Hrs [OC1-OC2]	
<b>Unit II</b>	XML DATABASES XML Databases: XML Data Model, DTD, XML Schema, XML Querying, Web Databases, Open Database Connectivity. MOBILE DATABASES Mobile Databases: Location and Handoff Management, Effect of Mobility on Data Management, Location Dependent Data Distribution, Mobile Transaction Models, Concurrency Control, Transaction Commit Protocols MULTIMEDIA DATABASES Multidimensional Data Structures, Image Databases, Text / Document Databases, Video Databases, Audio Databases	15 Hrs [OC1-OC2]	
<b>Unit III</b>	Introduction to Data Warehousing: Introduction, Necessity, Framework of the data warehouse, options, developing data warehouses, end points. Data Warehousing Design Consideration and Dimensional Modeling: Defining Dimensional Model, Granularity of Facts, Additivity of Facts, Functional dependency of the Data, Helper Tables, Implementation many to-many relationships between fact and dimensional modeling.	15 Hrs [OC1-OC3]	
<b>Unit IV</b>	Data warehousing and OLAP: Defining OLAP, The Value of Multidimensional data, OLAP terminologies, Multidimensional architectures, Multidimensional views of relational data, Physical Multidimensional databases, Data Explosion, Integrated relational OLAP, Data sparsity and data explosion.	15 Hrs [OC3]	

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Database System Concepts	Abraham Silberschatz, Henry Korth	McGraw Hill		
2.	DW2.0	W.H.Inmon, Derek Strauss	Morgan Kaufmann Publication		

### Course Outcomes(COs)

Upon completing this course, the student will be able to:

OC1: Understand the functionality of the various data mining and data warehousing component

OC2 : Understand the strengths and limitations of various data mining and data warehousing models

OC3: Analyzing techniques of various data

Course Code 23BPDS1T3	Course Title <b>Introduction to Data Analysis and Statistical Computing</b>	Credits <b>04</b>	No. of lectures
<b>Course Objectives (COs):</b> To enable the students to: CO1: Explore fundamental statistical data science concepts CO2: Create visual and numerical representations of data CO3: Interpret data analysis outputs CO4: Discuss interpretations and limitations of data			
<b>Unit I</b>	Introduction: What is Data analysis, importance, future Probability Theory: Sample Spaces, Events, Axioms, Counting, Conditional Probability and Bayes’ Theorem, The Binomial Theorem, Random variable and distributions : Mean and Variance of a Random variable-Binomial-Poisson-Exponential and Normal distributions. Curve Fitting and Principles of Least Squares- Regression and correlation.	15 Hrs [OC1-OC2]	
<b>Unit II</b>	Tabular data, Power and the computation of sample size, Advanced data handling, Multiple regression, Linear models, Logistic regression, Rates and Poisson regression, Nonlinear curve fitting.	15 Hrs [OC1-OC2]	
<b>Unit III</b>	Density Estimation, Recursive Partitioning, Smoothers and Generalized Additive Models, Survivals Analysis, Analyzing Longitudinal Data, Simultaneous Inference and Multiple Comparisons, Meta Analysis, Principal Component Analysis, Multidimensional Scaling Cluster Analysis.	15 Hrs [OC3-OC4]	
<b>Unit IV</b>	Introduction to R Packages, Scientific Calculator, Inspecting Variables, Vectors Matrices and Arrays, Lists and Data Frames Functions, Strings and Factors, Flow Control and Loops, Advanced Looping, Date and Times.	15 Hrs [OC3-OC4]	

<b>Books and References:</b>					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Statistics	Murray R. Spiegel, Larry J. Stephens.	McGRAW – HILL		
2.	Fundamental of Mathematical Statistics	S.C. Gupta and V.K. Kapoor	Sultan Chand and Sons	11 <sup>th</sup>	2011
3.	Mathematical Statistics	J.N. Kapur and H.C. Saxena	S. Chand	20 <sup>th</sup>	2005

### Course Outcomes(OCs)

Upon completing this course, the student will be able to:

- CO1: Understand the have the basic knowledge on data collection and various statistical elementary tools.
- CO2: Have the critical thinking in the theory of probability and its applications in real life problems.
- CO3: Apply the statistical tools in business, economical and commercial areas with the help of time series, index numbers, etc.
- CO4: Analyze such problems and to make better decisions for future in their fields.

<b>Course Code</b> 23BPDS1P1	<b>Course Title</b> <b>Fundamentals of Data Science / Database and Data Warehousing Practicals</b>	<b>Credits</b> 02	<b>No. of lectures</b>
---------------------------------	---	----------------------	----------------------------

<b>Practical 1</b>	Practical of Clustering
<b>Practical 2</b>	Practical of Time-series forecasting
<b>Practical 3</b>	Practical of Logistics Regression
<b>Practical 4</b>	Practical of Analysis of Variance
<b>Practical 5</b>	Practical of Decision Tree
<b>Practical 6</b>	Data definition language(DDL)
<b>Practical 7</b>	Data manipulation language (DML) commands in RDBMS
<b>Practical 8</b>	Performing Insertion, Deletion, Modifying, Altering, Updating and Viewing records based on conditions.
<b>Practical 9</b>	To create the view, execute and verify the various operations on views.
<b>Practical 10</b>	Creating an Employee database to set various constraints in RDBMS
<b>Practical 11</b>	Creating relationship between the databases in RDBMS
<b>Practical 12</b>	Case study on Data Science Techniques

Course Code 23BPDS1T4	Course Title <b>Elective I: Python Programming</b>	Credits <b>04</b>	No. of lectures
<b>Course Objectives (COs):</b> To enable the students to: <b>CO1:</b> The technologies related to Data Science <b>CO2:</b> Major features of the Python Language <b>CO3:</b> Python libraries to solve problems			
<b>Unit I</b>	Introduction: The Python Programming Language, History, features, Installing Python, Running Python program, Debugging : Syntax Errors, Runtime Errors, Semantic Errors, Experimental Debugging, Formal and Natural Languages, The Difference Between Brackets, Braces, and Parentheses, Variables and Expressions Values and Types, Variables, Variable Names and Keywords, Type conversion, Operators and Operands, Expressions, Interactive Mode and Script Mode, Order of Operations. Conditional Statements: if, if-else, nested if –else Looping: for, while, nested loops Control statements: Terminating loops, skipping specific conditions Functions: Function Calls, Type Conversion Functions, Math Functions, Composition, Adding New Functions, Definitions and Uses, Flow of Execution, Parameters and Arguments, Variables and Parameters Are Local, Stack Diagrams, Fruitful Functions and Void Functions, Why Functions? Importing with from, Return Values, Incremental Development, Composition, Boolean Functions, More Recursion, Leap of Faith, Checking Types	15 Hrs [OC1-OC2]	
<b>Unit II</b>	Strings: A String Is a Sequence, Traversal with a for Loop, String Slices, Strings Are Immutable, Searching, Looping and Counting, String Methods, The in Operator, String Comparison, String Operations Lists: Values and Accessing Elements, Lists are mutable, traversing a List, Deleting elements from List, Built-in List Operators, Concatenation, Repetition, In Operator, Built-in List functions and methods Tuples and Dictionaries: Tuples, Accessing values in Tuples, Tuple Assignment, Tuples as return values, Variable-length argument tuples, Basic tuples operations, Concatenation, Repetition, in Operator, Iteration, Built-in Tuple Functions Creating a Dictionary, Accessing Values in a dictionary, Updating Dictionary, Deleting Elements from Dictionary, Properties of Dictionary keys, Operations in Dictionary, Built-In Dictionary Functions, Built-in Dictionary Methods Files: Text Files, The File Object Attributes, Directories Exceptions: Built-in Exceptions, Handling Exceptions, Exception with Arguments	15 Hrs [OC3-OC4]	

<b>Books and References:</b>					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	An Introduction to Computer Science using Python 3	Jason Montojo, Jennifer Campbell, Paul Gries	SPD	1 <sup>st</sup>	2014
2.	Introduction to Problem Solving with Python	E. Balagurusamy	TMH	1 <sup>st</sup>	2016
3.	Murach's Python programming	Joel Murach, Michael Urban	SPD	1 <sup>st</sup>	2017
4.	Object-oriented Programming in Python	Michael H. Goldwasser, David Letscher	Pearson Prentice Hall	1 <sup>st</sup>	2008
5.	Exploring Python	Budd	TMH	1 <sup>st</sup>	2016

Course Code 23BPDS1P2	Course Title <b>Python Programming Practicals</b>	Credits 02	No. of lectures
<b>Practical 1</b>	A. Create a program that asks the user to enter their name and their age. Print out a message addressed to them that tells them the year that they will turn 100 years old. B. Enter the number from the user and depending on whether the number is even or odd, print out an appropriate message to the user. C. Write a program to generate the Fibonacci series. D. Write a function that reverses the user defined value. E. Write a function to check the input value is Armstrong and also write the function for Palindrome. F. Write a recursive function to print the factorial for a given number.		
<b>Practical 2</b>	A. Write a function that takes a character (i.e. a string of length 1) and returns True if it is a vowel, False otherwise. B. Define a function that computes the length of a given list or string.		
<b>Practical 3</b>	A. A <i>pangram</i> is a sentence that contains all the letters of the English alphabet at least once, for example: <i>The quick brown fox jumps over the lazy dog</i> . Your task here is to write a function to check a sentence to see if it is a pangram or not. B. Take a list, say for example this one: a=[1,1,2,3,5,8,13,21,34,55,89] and write a program that prints out all the elements of the list that are less than 5.		
<b>Practical 4</b>	A. Write a program that takes two lists and returns True if they have at least one common member. B. Write a Python program to print a specified list after removing the 0th, 2nd, 4th and 5th elements. C. Write a Python program to clone or copy a list		
<b>Practical 5</b>	A. Write a Python script to sort (ascending and descending) a dictionary by value. B. Write a Python script to concatenate following dictionaries to create a new one. C. Write a Python program to sum all the items in a dictionary.		
<b>Practical 6</b>	A. Write a Python program to read an entire text file. B. Write a Python program to append text to a file and display the text. C. Write a Python program to read last n lines of a file.		

### Course Outcomes:

- OC1: Interpret the fundamental Python syntax and semantics and be fluent in the use of Python control flow statements.
- OC2: Express proficiency in the handling of strings and functions.
- OC3: Determine the methods to create and manipulate Python programs by utilizing the data structures like lists, dictionaries, tuples and sets
- OC4: Identify the commonly used operations involving file systems and regular expressions.

Course Code 23BPDS1T5	Course Title <b>Elective II: JAVA Programming</b>	Credits 04	No. of lectures
<b>Course Objectives (COs)</b> <ul style="list-style-type: none"> <li>To learn why Java is useful for the design of desktop and web applications.</li> <li>To learn how to implement object-oriented designs with Java.</li> <li>To identify Java language components and how they work together in applications.</li> <li>To design and program stand-alone Java applications.</li> </ul>			
<b>Unit I</b>	<b>The Java Language:</b> Features of Java, Java programming format, Java Tokens, Java Statements, Java Data Types, Typecasting, Arrays <b>OOPS:</b> Introduction, Class, Object, Static Keywords, Constructors, this Key Word, Inheritance, super Key Word, Polymorphism (overloading and overriding), Abstraction, Encapsulation, Abstract Classes, Interfaces <b>String Manipulations:</b> String, String Buffer, String Tokenizer <b>Packages:</b> Introduction to predefined packages (java.lang, java.util, java.io, java.sql, java.swing), User Defined Packages, Access specifiers <b>Exception Handling:</b> Introduction, Pre-Defined Exceptions, Try-Catch-Finally, Throws, throw, User Defined Exception examples <b>Multithreading:</b> Thread Creations, Thread Life Cycle, Life Cycle Methods, Synchronization, Wait() notify() notify all() methods	15 [OC1-OC2]	
<b>Unit II</b>	<b>Networking:</b> Introduction, Socket, Server socket, Client – Server Communication <b>Wrapper Classes:</b> Introduction, Byte, Short, Integer, Long, Float, Double, Character, Boolean classes <b>Collection Framework:</b> Introduction, util Package interfaces, List, Set, Map, List interface & its classes, Set interface & its classes, Map interface & its classes <b>Inner Classes:</b> Introduction, Member inner class, Static inner class, Local inner class, Anonymous inner class <b>AWT:</b> Introduction, Components, Event-Delegation-Model, Listeners, Layouts, Individual components Label, Button, CheckBox, Radio Button, Choice, List, Menu, Text Field, Text Area	15 [OC3-OC4]	

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Herbert Schildt, Java The Complete Reference		McGraw-Hill Education,	Ninth Edition	2014
2.	E. Balagurusamy, Programming with Java		McGraw-Hill Education	002C	2014
3.	Programming in JAVA,	Sachin Malhotra,	Oxford Press	2nd Ed	

Course Code 23BPIT1P5	Course Title <b>Elective III: JAVA Programming Practical</b>	Credits 04	No. of lectures
<b>List of Practicals</b> <ol style="list-style-type: none"> <li>Accept the coefficients of quadratic equation. Find the solution of quadratic equation.</li> <li>Accept two n x m matrices. Write a Java program to find addition of these matrices.</li> <li>Accept n strings. Sort names in ascending order.</li> <li>Demonstrate Java inheritance using extends keyword.</li> <li>Demonstrate method overloading and method overriding in Java.</li> <li>Write a Java List example and demonstrate methods of Java List interface.</li> <li>Design simple calculator GUI application using AWT components.</li> </ol>			

**Course Outcome :-**

- CO1 :Object oriented programming concepts using Java.  
 CO2: Knowledge of input, its processing and getting suitable output.  
 CO3 : Understand, design, implement and evaluate classes and applets.  
 CO4 : Knowledge and implementation of AWT package

<b>Course Code</b> 23BPRM1T7	<b>Course Title</b> <b>Research Methodology</b>	<b>Credits</b> <b>04</b>	<b>No. of lectures</b>
<b>Course Objectives(COs)</b> CO1. To be able to conduct business research with an understanding of all the latest theories CO2. To develop the ability to explore research techniques used for solving any real world or innovate problem			
<b>Unit I</b>	<b>Introduction:</b> Role of Business Research, Information Systems and Knowledge Management, Theory Building, Organization ethics and Issues <b>Beginning Stages of Research Process:</b> Problem definition, Qualitative research tools, Secondary data research	<b>15</b> <b>[OC1-OC2]</b>	
<b>Unit II</b>	<b>Research Methods and Data Collection:</b> Survey research, communicating with respondents, Observation methods, Experimental research	<b>15</b> <b>[OC3-OC4]</b>	
<b>Unit III</b>	<b>Measurement Concepts, Sampling and Field work:</b> Levels of Scale measurement, attitude measurement, questionnaire design, sampling designs and procedures, determination of sample size	<b>15</b> <b>[OC5-OC6]</b>	
<b>Unit IV</b>	<b>Data Analysis and Presentation:</b> Editing and Coding, Basic Data Analysis, Univariate Statistical Analysis and Bivariate Statistical analysis and differences between two variables. Multivariate Statistical Analysis	<b>15</b> <b>[OC7-OC8]</b>	

<b>Books and References:</b>					
<b>Sr. No.</b>	<b>Title</b>	<b>Author/s</b>	<b>Publisher</b>	<b>Edition</b>	<b>Year</b>
1.	Business Research Methods	William G.Zikmund, B.J Babin, J.C. Carr, Atanu Adhikari, M.Griffin	Cengage	8e	2016
2.	Business Analytics	Albright Winston	Cengage	5e	2015
3.	Research Methods for Business Students Fifth Edition	Mark Saunders			2011
4.	Multivariate Data Analysis	Hair	Pearson	7e	2014

**Course Outcomes(OCs):** A learner will be able to:

OC 1: Solve real world problems with scientific approach.

OC 2: Develop analytical skills by applying scientific methods.

OC 3: Recognize, understand and apply the language, theory and models of the field of business analytics

OC 4: Foster an ability to critically analyze, synthesize and solve complex unstructured business problems

OC 5: Understand and critically apply the concepts and methods of business analytics

OC 6: Identify, model and solve decision problems in different settings

OC 7: Interpret results/solutions and identify appropriate courses of action for a given managerial situation whether a problem or an opportunity

OC 8: Create viable solutions to decision making problems

# **Semester II**

Course Code	Course Title	Credits	No. of lectures
23BPDS2T1	<b>Big Data Analytics</b>	<b>04</b>	
<b>Course Objectives:</b> <ul style="list-style-type: none"> <li>To provide an overview of an exciting growing field of big data analytics.</li> <li>To introduce the tools required to manage and analyze big data like Hadoop, NoSql MapReduce.</li> <li>To teach the fundamental techniques and principles in achieving big data analytics with scalability and streaming capability.</li> <li>To enable students to have skills that will help them to solve complex real-world problems in for decision support</li> </ul>			
<b>Unit I</b>	<ul style="list-style-type: none"> <li>Introduction to Big Data, Characteristics of Data, and Big Data Evolution of Big Data, Definition of Big Data, Challenges with big data, Why Big data? Data Warehouse environment, Traditional Business Intelligence versus Big Data. State of Practice in Analytics, Key roles for New Big Data Ecosystems, Examples of big Data Analytics.</li> <li>Big Data Analytics, Introduction to big data analytics, Classification of Analytics, Challenges of Big Data, Importance of Big Data, Big Data Technologies, Data Science, Responsibilities, Soft state eventual consistency. Data Analytics Life Cycle</li> </ul>	<b>15</b> <b>[OC1-OC2]</b>	
<b>Unit II</b>	<ul style="list-style-type: none"> <li>Analytical Theory and Methods: Clustering and Associated Algorithms, Association Rules, Apriori Algorithm, Candidate Rules, Applications of Association Rules, Validation and Testing, Diagnostics, Regression, Linear Regression, Logistic Regression, Additional Regression Models.</li> <li>Analytical Theory and Methods: Classification, Decision Trees, Naïve Bayes, Diagnostics of Classifiers, Additional Classification Methods, Time Series Analysis, Box Jenkins methodology, ARIMA Model, Additional methods. Text Analysis, Steps, Text Analysis Example, Collecting Raw Text, Representing Text, Term Frequency-Inverse Document Frequency (TFIDF), Categorizing Documents by Topics, Determining Sentiments</li> </ul>	<b>15</b> <b>[OC3-OC4]</b>	
<b>Unit III</b>	<ul style="list-style-type: none"> <li>Data Product, Building Data Products at Scale with Hadoop, Data Science Pipeline and Hadoop Ecosystem, Operating System for Big Data, Concepts, Hadoop Architecture, Working with Distributed file system, Working with Distributed Computation,</li> <li>Framework for Python and Hadoop Streaming, Hadoop Streaming, MapReduce with Python, Advanced MapReduce. In-Memory Computing with Spark, Spark Basics, Interactive Spark with PySpark, Writing Spark Applications</li> </ul>	<b>15</b> <b>[OC5-OC6]</b>	
<b>Unit IV</b>	<ul style="list-style-type: none"> <li>Distributed Analysis and Patterns, Computing with Keys, Design Patterns, Last-Mile Analytics, Data Mining and Warehousing, Structured Data Queries with Hive, HBase, Data Ingestion</li> <li>Importing Relational data with Sqoop, Injesting stream data with flume. Analytics with higher level APIs, Pig, Spark's higher level APIs</li> </ul>	<b>15</b> <b>OC7</b>	

<b>Books and References:</b>					
<b>Sr. No.</b>	<b>Title</b>	<b>Author/s</b>	<b>Publisher</b>	<b>Edition</b>	<b>Year</b>
1.	Big Data and Analytics	Subhashini Chellappan Seema Acharya	Wiley	First	
2.	Data Analytics with Hadoop <i>An Introduction for Data Scientists</i>	<i>Benjamin Bengfort and Jenny Kim</i>	O'Reilly		2016
3.	Big Data and Hadoop	V.K Jain	Khanna Publishing	First	2018

### **Course Outcomes(OCs)**

Upon completion of this course the Students will be able to:

- OC1 Understand Big Data Concepts
- OC2 Do Data Collection and Integration
- OC3 Develop Data Storage and Management
- OC4 Perform Data Preprocessing and Cleaning
- OC5 Understand Data Transformation and Feature Engineering
- OC6 Perform Exploratory Data Analysis (EDA)
- OC7 Use Big Data Analytics Tools

Course Code 23BPDS2T2	Course Title Next Generation Databases	Credits 02	No. of lectures
<b>Course Objectives (COs)</b> <ul style="list-style-type: none"> <li>Understand Concepts of No-SQL Databases</li> <li>The types No-SQL databases</li> <li>The features of MongoDB</li> <li>The use of jQuery</li> </ul>			
<b>Unit I</b>	<b>NoSQL:</b> SQL, NoSQL, Definition, A Brief History of NoSQL, ACID vs. BASE, CAP Theorem (Brewer's Theorem), The BASE, NoSQL Advantages and Disadvantages, Advantages of NoSQL, Disadvantages of NoSQL, SQL vs. NoSQL Databases, Categories of NoSQL Databases	15	
<b>Unit II</b>	<b>Introducing MongoDB:</b> History, MongoDB Design Philosophy, Speed, Scalability, and Agility, Non-Relational Approach, JSON-Based Document Store, Performance vs. Features, Running the Database Anywhere, SQL Comparison <b>The MongoDB Data Model:</b> The Data Model, JSON and BSON, The Identifier (_id), Capped Collection, Polymorphic Schemas, Object Oriented Programming, Schema Evolution	15	
<b>Unit III</b>	<b>Using MongoDB Shell :</b> Basic Querying, Create and Insert, Explicitly Creating Collections, Inserting Documents Using Loop, Inserting by Explicitly Specifying _id, Update, Delete, Read, Using Indexes, Stepping Beyond the Basics, Using Conditional Operators, Regular Expressions, MapReduce, aggregate(), Designing an Applications Data Model, Relational Data Modeling and Normalization, MongoDB Document Data Model Approach		
<b>Unit IV</b>	SSD and In-Memory Databases: End of Disk, Solid State Disk, The Economics of Disk, SSD-Enabled Databases, In-Memory Databases, TimesTen, Redis, <b>jQuery:</b> Introduction, Traversing the DOM, DOM Manipulation with jQuery, Events, Ajax with jQuery, jQuery Plug-ins, jQuery Image Slider		

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Next Generation Database: NoSQL and big data	Guy Harrison	Apress		
2	Practical MongoDB	Shakuntala Gupta Edward NavinSabharwal	Apress		
3	Beginning jQuery	Jack Franklin Russ Ferguson	Apress		

After completion of the course, students will be able to

OC1: Describe the features of No-SQL Databases

OC2: Construct simple queries using MongoDB

OC3: Apply advanced MongoDB features

OC4: Use of jQuery.

Course Code 23BPDS2T3	Course Title High Performance Computing	Credits 02	No. of lectures
Course Objectives (COs): <ul style="list-style-type: none"><li>Understand High Performance Computing (HPC) system architectures and various computational models.</li><li>Apply parallel execution models and methodologies for parallel programming and parallel applications development.</li><li>Design and implement compute intensive applications on HPC platform.</li></ul>			
Unit I	Graph Algorithms, Minimum Spanning Tree, Prim's Algorithm, Single, Source Shortest Paths, Dijkstra's Algorithm, All-Pairs Shortest Paths, Algorithms for Sparse Graphs, Search Algorithms for Discrete Optimization, Problems, Sequential Search Algorithms, Parallel Depth, First Search, Parallel Breadth, First Search	15 [OC1]	
Unit II	Message passing, MPI example, Messages and point-to-point communication, Collective communication, Non blocking point-to-point communication, Virtual topologies, Example, MPI parallelization of Jacobi solver, Communication parameters	15 [OC1-OC2]	
Unit III	Synchronization serialization, contention, Implicit serialization and synchronization Contention, Reducing communication overhead, Optimal domain decomposition, Aggregating messages, Non blocking vs. asynchronous communication	15 [OC2]	
Unit IV	<b>Neural Networks:</b> Introduction to Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks <b>Healthcare :</b> Understanding the Healthcare Domain, Provider Analytics, Payer Analytics, Analytics in the Pharmaceutical Industries <b>Banking and Financial Services :</b> Introduction to Banking and Financial Services, Acquisition Analytics, Engagement Analytics, Risk Analytics	15 [OC3]	

<b>Books and References:</b>					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Parallel and High Performance Computing	Robert Robey, Zamora Yuliana	Manning		
2	High Performance Cluster Computing: Architectures and Systems,	Rajkumar	Pearson Education		
3	High Performance Computing for Big Data Methodologies and Application	Chao Wang	CRC Press		

OC1: Understand High Performance Computing (HPC) system architectures and various computational models.

OC2: Apply parallel execution models and methodologies for parallel programming and parallel applications development.

OC3: Design and implement compute intensive applications on HPC platform.

<b>Course Code</b> 23BPDS2P1	<b>Course Title</b> <b>Big Data Analytics / Next Generation Database Practical</b>	<b>Credits</b> <b>02</b>	<b>No. of lectures</b>
---------------------------------	---	-----------------------------	------------------------

### Next Generation Database Practical

<b>Practical 1</b>	<b>MongoDB Basics</b> a. Write a MongoDB query to create and drop database. b. Write a MongoDB query to create, display and drop collection c. Write a MongoDB query to insert, query, update and delete a document.
<b>Practical 2</b>	<b>Simple Queries with MongoDB</b>
<b>Practical 3</b>	<b>Implementing Aggregation</b> a. Write a MongoDB query to use sum, avg, min and max expression. b. Write a MongoDB query to use push and addToSet expression. c. Write a MongoDB query to use first and last expression.
<b>Practical 4</b>	<b>Python and MongoDB</b> Connecting Python with MongoDB and inserting, retrieving, updating and deleting.
<b>Practical 5</b>	<b>Programs on Basic jQuery</b> a. jQuery Basic, jQuery Events b. jQuery Selectors, jQuery Hide and Show effects c. jQuery fading effects, jQuery Sliding effects
<b>Practical 6</b>	<b>jQuery Advanced</b> a. jQuery Animation effects, jQuery Chaining b. jQuery Callback, jQuery Get and Set Contents c. jQuery Insert Content, jQuery Remove Elements and Attribute

### High Performance Computing Practical

<b>Sr. No</b>	<b>Details</b>	
<b>1</b>	Write a program to implement prim's algorithm using C/C++/Python/Java/C# language.	30 Hrs [OC1-OC2]
<b>2</b>	Write a program to implement linear search using C/C++/Python/Java/C# language.	
<b>3</b>	Write a program to implement Divide and Concur algorithm using C/C++/Python/Java/C# language.	
<b>4</b>	Write a program to implement DFS algorithm using C/C++/Python/Java/C# language.	
<b>5</b>	Write a program to implement BFS algorithm using C/C++/Python/Java/C# language.	
<b>6</b>	Write a program to implement Dijkstra's shortest path algorithm using C/C++/Python/Java/C# language.	
<b>7</b>	Implementation of convolutional neural network to predict numbers	
<b>8</b>	Implementing regularization to avoid overfitting in binary classification.	

Course Code 23BPDS2T4	Course Title <b>Elective I : Data Mining</b>	Credits <b>04</b>	No. of lectures
<b>Course Objectives (COs)</b> <ul style="list-style-type: none"> <li>To introduce students to basic applications, concepts, and techniques of data mining.</li> <li>To develop skills for using recent data mining software (eg. R,weka) to solve practical problems in a variety of disciplines.</li> <li>To gain experience doing independent study and research.</li> <li>Be familiar with mathematical foundations of data mining tools.</li> <li>Understand and implement classical models and algorithms in data warehouses and data mining</li> </ul>			
<b>Unit I</b>	<b>Introduction:</b> Basics of data mining, related concepts, Data mining techniques. <b>Data:</b> Introduction, Attributes, Data Sets, and Data Storage, Issues Concerning the Amount and Quality of Data, <b>Knowledge Representation:</b> Data Representation and their Categories: General Insights, Categories of Knowledge Representation, Granularity of Data and Knowledge Representation Schemes, Sets and Interval Analysis, Fuzzy Sets as Human-Centric Information Granules, Shadowed Sets, Rough Sets, Characterization of Knowledge Representation Schemes, Levels of Granularity and Perception Perspectives, The Concept of Granularity in Rules.	15 [OC1-OC2]	
<b>Unit II</b>	<b>Mining Frequent Patterns, Associations, and Correlations:</b> Basic Concepts, Efficient and Scalable Frequent Item set Mining Methods, Mining Various Kinds of Association Rules, From Association Mining to Correlation Analysis, Constraint-Based Association Mining <b>Graph Mining, Social Network Analysis, and Multirelational Data Mining:</b> Graph Mining, Social Network Analysis, Multirelational Data Mining. <b>Mining Object, Spatial, Multimedia, Text, and Web Data:</b> Multidimensional Analysis and Descriptive Mining of Complex Data Objects, Spatial Data Mining, Multimedia Data Mining, Text Mining, Mining the World Wide Web.	15 [OC3-OC4]	

<b>Books and References:</b>					
Sr.No.	Title	Author/s	Publisher	Edition	Year
1.	Data Mining: Introductory and Advanced Topics.	M. H. Dunham.	Pearson Education.		2010
2.	Data Mining A Knowledge Discovery Approach”,	Krzysztof J. Cios, W. Pedrycz,R. W. Swiniarski, L.A. Kurgan	Springer		
3.	Data Mining: Concepts and Techniques	J. Han and M. Kamber,	Elsevier	2nd	2008
4.	Principles of Data Mining	D. Hand, H. Mannila and P. Smyth	Prentice-Hall		2001
5.	Data Mining with SQL Server 2005	Z.Tang and J MacLennan,	Wiley		

Course Code 23BPDS2P4	Course Title <b>Elective I : Data Mining Practical</b>		No. of lectures
<b>List of Practicals</b> <ol style="list-style-type: none"> <li>1. Show the implementation of Naïve Bayes algorithm.</li> <li>2. Show the implementation of Decision Tree.</li> <li>3. Show the implementation of Clustering Algorithm.</li> <li>4. Show the implementation of Apriori Algorithm</li> <li>5. Show the implementation of Time Series Algorithm.</li> </ol>			

**CO1:** Identify the scope and necessity of Data Mining for the society

**CO2:** To understand various tools of Data Mining and their techniques to solve the real time problems. .

**CO3:** To develop ability to design various algorithms based on data mining tools.

**CO4:** To develop further interest in research and design of new Data Mining techniques.

Course Code 23BPDS2T5	Course Title <b>Elective II: Cloud Management</b>	Credits <b>04</b>	No. of lectures
<b>Course Objectives (COs)</b> <ul style="list-style-type: none"> <li>To provide students with the fundamentals and essentials of Cloud Management.</li> <li>To provide students a sound foundation of the Cloud Management so that they are able to start using and adopting Cloud Management services and tools in their real life scenarios.</li> <li>To enable students exploring some important cloud Management driven commercial systems and applications.</li> </ul>			
<b>Unit I</b>	Virtualized Data Center Architecture: Cloud infrastructures; public, private, hybrid. Service provider interfaces; Saas, Paas, Iaas. VDC environments; concept, planning and design, business continuity and disaster recovery principles. Managing VDC and cloud environments and infrastructures Cloud Management: System Center 2012 and Cloud OS, Provisioning Infrastructure: Provisioning Infrastructure with Virtual Machine Designing, Planning and Implementing.	15	
<b>Unit II</b>	Managing Hyper-V Environment with VMM 2012. Provisioning self-service with AppController, AppController essentials, Managing Private, Public, Hybrid clouds. AppControllercmdlets. Managing and maintaining with Configuration Manager 2012, Design, Planning, Implementation, Administration, Distributing Applications, Updates, Deploying Operating Systems, Asset Management and reporting. Backup and recovery with Data Protection Manager. Design, Planning, Implementation and Administration.	15	

<b>Books and References:</b>					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Introducing Microsoft System Center 2012, Technical Overview	Mitch Tulloch, Symon Perriman and Symon Perriman	Microsoft		
2.	Microsoft System Center 2012 Unleashed	Chris Amaris, Rand orimoto, Pete Handley, David E. Ross, Technical Editby Yardeni	Pearson Education		
3.	Automating vSphere with Mware vCenter Orchestrator				

<b>Course Code</b> 23BPDS2P5	<b>Course Title</b> <b>Elective I: Cloud Management Practical</b>	<b>Credits</b> <b>04</b>	<b>No. of lectures</b>
List of Practicals 1. Create and Manage Cloud using SCVMM 2019, Deploy a guarded host fabric using Microsoft SCVMM 2019, Deploy and manage SDN Infra structure using SCVMM 2019, Deploy and Manage Storage Space Direct (S2D) using SCVMM 2019  2. Deploy Service Manager 2019 and install on 4 Computer Scenario, Setup SQL Server reporting Service using Service Manager 2019, User Connectors to import data: Import data from Active Directory Domain Services, Import data and alerts from Operations Manager, Import data from Configuration Manager, Import runbooks from Orchestrator, Import data from VMM, Use a CSV file to import data  3. Install Orchestrator, Create and test a monitor runbook, Manage Orchestrator Servers – 1, Runbook permissions, Back up Orchestrator, Bench mark , Optimize performance of .Net activities, Configure runbook throttling, Recover a database  4. Managing devices with Configuration Manager, Design a hierarchy of sites using Microsoft End Point Configuration manager, Data transfers between sites, Types of data transfer, File-based replication, Database replication			

OC1: Explain the core concepts of the cloud management paradigm: how and why this paradigm shift came about, the characteristics, advantages and challenges brought about by the various models and services in cloud management.

OC2: Apply the fundamental concepts in datacenters to understand the tradeoffs in power, efficiency and cost.

OC3: Identify resource management fundamentals, i.e. resource abstraction, sharing and sandboxing and outline their role in managing infrastructure in cloud management.

OC4: Analyze various cloud programming models and apply them to solve problems on the cloud.