

MCA Syllabus:

Year 1:

Software Engineering

Chapter 1: Overview of System Analysis and Design

Introduction to System Concepts - Mapping Software with system definition - Categories of Information System - Concepts of Engineering - Software development Life Cycle (SDLC) - Different Approaches and Models for System including RAD and Spiral and - prototyping model - The Role and Task of System Analyst, Skills required for a Good Software Analyst

Chapter 2: Software Requirement Specification Technique

Introduction To Requirement Engineering - Requirement Anticipation - Requirement Investigation - Requirement Specification, - Solved Cases Study with IEEE standards for SRS document

Chapter 3: Information Requirement Analysis

Introduction to Decision Analysis - What is Functional Decomposition? - Process modeling with Data Flow Diagram - Entity Relationship Diagram - Normalization - Data Dictionary - Decision Tree - Decision Table

Chapter 4: Designing of Input, Output and Program

Introduction To System Design - Human Computer Interface and software design - Design of Input and Control - Design of Output - User Interface Design - Solved Case Study

Chapter 5: Software Maintenance

Introduction - Why We Need Software maintenance - Types of Maintenance - Maintenance Cost Introduction to Legacy - Reverse Engineering - Documentation

Chapter 6: Case Tools

Introduction to Case Tools - Types of CASE Tools used in various SDLC phases with example - Advantages and Disadvantages of Case Tools - Case Studies

Chapter 7: Current Trends in Software Engineering

Introduction - Software Engineering for Projects and Products - Scrum methodology - Web Engineering - Agile Process - CPM and PERT - Gantt Charts - SQA - SEI-CMM Introduction to Software Security

Object Oriented Programming with C++

Chapter 1: Principle of OOP's

Software crisis - Software Evaluation - POP (Procedure Oriented Programming) - OOP(Object Oriented Programming) - Basic concepts of OOP - Benefits of OOP - Object Oriented Language - Application of OOP

Chapter 2: Basics of C++

ABrief History of C and C++ - Difference between C and C++ - Features of C++ - Advantages and Disadvantages of C++ - Applications of C++ - Writing and Executing a C++ Program - Program Structure and Rules - Sample C++ Program - Comments - Return Type of MAIN()

Chapter 3: Expression

Introduction C++ Tokens - Understand different Data types in C++ - Understand Declaration of Variables - How to Initialization of Variables in C++ - Understand Reference Variables - Know the Operators and use of the Operators in C++ - Type Cast Operator in C++ - Understand Memory Management operators and use of the same - Mentioning of Expression - Understand Statement - Understand Symbolic Constant - Type Compatibility with C++ - Summary - SelfAssessment Questions

Chapter 4: Functions in C++

Introduction to Function, - Passing Information-Parameters, - Default Arguments, - Constant Arguments, - Function Overloading, - Inline Functions, - Recursive Functions - Summary - Self Assessment Questions

Chapter 5: Classes and Objects

Introduction of a Class, - Member Functions, - Making an Outside Function Inline, - Nesting Of Member Functions, - Private Member Function, - Arrays within a Class, - MemoryAllocation for Objects, - Arrays of Objects, - Objects as FunctionArguments, - Returning Objects, - Const Member Function, - Static Class Members - Pointer to Members, - Local classes, - Friend Functions, - Unions and classes, - Object Composition and Delegation - Summery - SelfAssessment Questions

Chapter 6: Constructor and Destructor

Introduction to Constructor - Constructor Definition, - Multiple Constructors in a class - Constructor with Default Arguments - Dynamic Initialization of Objects - Const Object, Destructor - Summary - Self Assessment Questions

Chapter 7: Operator Overloading and Type Conversion

Introduction, - Overloading Unary Operators, - Overloading Binary Operators, - Limitations of Operator Overloading, - This pointer, - Overloading<<and>>Operators, - Manipulation of String, - Types Conversion - Summary - SelfAssessment Questions

Chapter 8: Inheritance

Introduction OOPs – Inheritance - Single Inheritance, - Multiple Inheritance, - Multilevel Inheritance, - Hierarchical inheritance, - Hybrid Inheritance, - Container Classes, - Virtual Base Classes, - Construction in Derived classes, - Virtual Function, - Pure Virtual Functions, - Abstract Classes - Summary - Self Assessment Questions

Chapter 9: The C++ I/O System Basics

Introduction - C++ streams - C++ streams classes - Unformatted I/O Operations - Formatted console I/O Operations - Managing output with manipulators - Design Our Own Manipulators - Summary - SelfAssessment Questions

Chapter 10: Working with Files

Introduction Creating a Stream, - Opening a File, - Closing a File, - Checking For Failure with File Commands, - Detecting the End-of-file, - File Pointers and their Manipulation, - Reading/Writing a character From a File, - Write()and read() Functions, - .Buffers and Sychronization, - Other Functions, -

RandomAccess File Processing, - Updating a File :RandomAccess, - Command LineArguments. - Summary. - SelfAssessment Questions

Chapter 11: Template

Introduction - Class templates - Multiple parameters in class templates - Function templates - Multiple parameters in function templates - Overloading of template functions - Member function templates - Non-type template arguments - Summary - 0SelfAssessment Questions

Chapter 12: Exception Handling

Introduction - Principles of Exception handling - Exception handling mechanism - Throwing mechanism - Catching mechanism - Re-throwing an Exception - Specifying Exception - Summary - SelfAssessment Questions

Chapter 13: Introduction to Standard Template library

Introduction - The C++ Standard Template Library (STL) - .Algorithms - C++: Containers in STL - C++: Iterators in STL - Use and Application of STL - Classification of Containers in STL - Using Container Library in STL - 0Standard Exceptions in C++ - 1Dynamic MemoryAllocation in C++ - 2Summary. - 3SelfAssessment Questions

Chapter 14: Namespace

Introduction - Creating a Namespace - Rules to create Namespaces - Using a Namespace in C++ - Discontinuous Namespaces - Nested Namespaces - Summary. – Self-Assessment Questions

Chapter 15: New Style Caste and RTRI

Introduction - New-Style Casts, - Static_cast, - Dynamic_cast, - Const_cast, Reinterpret_cast, - Run-Time Type Information(RTTI), - ASimpleApplication of Run-Time Type ID, - Typied can be applied to Templates classes

Operating System Concepts

Chapter 1: COMPUTER SYSTEM AND OPERATING SYSTEM OVERVIEW

Overview of an Operating System - Computer System Structure - Four Components of a Computer System - Operating System Definition - Computer Startup - Computer System Organization - Computer-System Operation - Common Functions of Interrupts - I/O Structure - Direct MemoryAccess Structure - Storage Structure - Storage Hierarchy - Caching - Computer-SystemArchitecture - Working of a Modern Computer Works - Clustered Systems - Operating System Structure - Memory Layout for Multi-programmed System - Operating-System Operations - Transition from User to Kernel Mode - OPERATING SYSTEM FUNCTIONS - Performance of Various Levels of Storage - Migration of IntegerAfrom Disk to Register - Protection and Security - DISTRIBUTED SYSTEMS: Computing Environments Traditional computer - Peer-to-Peer Computing - Web-Based Computing - Open-Source Operating Systems - Operating System Services - AView of Operating System Service - Operating System Service - Bourne Shell Command Interpreter - The Mac OS X GUI - System Calls - API – System Call – OS Relationship - Types of System Calls - Operating System Environment - System Programs - Operating System Design and Implementation - Virtual Machines - Operating-System Debugging - Summary - Self-Assessment Question

Chapter 2: Process Management

Process Concept - Process Control Block - Process Scheduling - Process Operations - Interprocess Communication - Communication in Client-Server – RTOS (Real Time Operating System) - Summary - Self-Assessment Questions

Chapter 3: CPU Scheduling

CPU Scheduling - Scheduling Criteria - Implementation - Summary - Self-Assessment Questions

Chapter 4: Process Synchronization and Deadlock - Process Synchronization - Deadlock and Starvation - Monitors - System Model - Checkpoints - Serializability - Deadlock. - Summary - Self-Assessment Question

Chapter 5: Memory Management

Concept Memory Management Techniques: - Dynamic relocation using a relocation register - Swapping - Contiguous Allocation - Hardware Support for Relocation and Limit Registers - Paging - Address Translation Scheme - Paging Hardware With TLB - Memory Protection - Hashed Page Tables - Inverted Page Table - Segmentation - Virtual Memory - Page Replacement Algorithm - FIFO Algorithm - LRU Algorithm - Page Buffering Algorithm - Thrashing. - Summary - Self-Assessment Questions

Chapter 6: File Management

Introduction File Structure. - ACCESS METHODS - Directory Structures - File System Mounting - File Sharing - Protection - File System Implementation - FAT(File Allocation Table) - Free space management - Directory Implementation - Permissions In A Unix System - Windows - Summary - Self-Assessment Questions

Chapter 7: Disk Management

Overview of Mass Storage Structure - Disk Structure - Disk Attachment - Disk Scheduling - SCAN - Disk Management - Swap-Space Management - Data Structures for Swapping on Linux Systems - Tertiary Storage Devices - Operating System Support - Swap-Space Management - RAID Structure (redundant arrays of independent disks) - Summary - Self-Assessment Questions

Chapter 8: Distributed Operating System - Introduction to Distributed Systems - Centralized vs Decentralized vs Distributed - Network Operating Systems (NOS): - Network Structure. - Communication Structure: - Packet Strategies: - Connection Strategies: - Communication Protocols: - Summary - Self-Assessment Questions

Chapter 9: Case Study of Window OS and Non-Windows OS

Introduction. - Feature Migration - An Introduction to Modern Mobile Operating Systems.

SUBJECT: DATA COMMUNICATION & COMPUTER NETWORK

Chapter 1: Introduction, Basics of Data Communication, Connection oriented N/Ws Vs Connection less N/Ws, Ethernet, Wireless LAN, Protocol Layering, TCP/IP Protocol Suite, The OSI Model, OSI vs TCP/IP, Summary, Self-Assessment Questions.

Chapter 2: Physical Communication: Introduction, Hardware Architecture, Transmission Media, Communication devices, Switching and switching techniques, Summary, Self-Assessment Questions.

Chapter 3: Link Layer Communication: Introduction, Error Detection and Correction techniques, Framing and Its Types, Flow and error Control, HDLC Protocol, P2pProtocol, Summary, Self-Assessment Questions.

Chapter 4: IP Addressing and Routing: Introduction, Internet protocol, IPV4, Addresses, Network Part, Host Part and Network Mask, IPV4 Address Classes, Network Addresses and Special Addresses, Routing, TCP Connections, IPV6, IPV4 Vs IPV6, Summary, Self-Assessment Questions.

Chapter 5: Domain Network Services (DNS): Introduction, Domain Space Name, DNS in the Internet, DNS Resolution and caching, Resource Records, DNS message and DNS security, DHCP and Scope Resolutions, Summary, Self –Assessment Questions.

Chapter 6: Network Applications; Introduction, Hyper Text Transfer Protocol (HTTP), Email – sending & receiving Email, Addressing Message structure, MIME- Multipurpose Internet Mail extension, SMTP- Simple Mail transfer protocol, POP-Post Office Protocol, IMAP- Internet Message Access Protocol, Summary, Self-Assessment Questions.

Chapter 7: Network Security: Introduction, Threat: Active attack, Passive attack, Cryptography: Symmetric and Asymmetric Key Cryptography, Digital signature, VPN and VPN protocols, Firewall: Packet filter, Application gateway, Summary, Self-assessment Questions.

Chapter 8: Advanced Network Technologies: Introduction, WiMAX, LTE, VOIP, Unified Communication, Introduction to IOT and Sensor Networks, Summary, Self-Assessment Questions.

SUBJECT-DATA STRUCTURE USING C++

Chapter 1: An Introduction to Data Structure: Introduction: - Concept of Data structure- Overview of Data structure, Implementation of Data structure, Array: Definition- Terminology- One dimensional array Multidimensional arrays pointer arrays. Linked lists: Single linked lists – Circular listed list- Double linked lists- Circular Double linked list - Application of linked list-Memory representation - Boundary Tap system - DE allocation strategy - Buddy system – Compaction, Summary Self-Assessment Questions.

Chapter 2 Stack & Queue: Introduction, Stack definition representation of stack-operation of stack-application of stack, Queue - definition- representation of queue- operation of queue- application of queue, Summary, Self-Assessment Questions

Chapter 3: TREE Introduction, Definition and concepts, Representation of binary tree, Operations on binary tree- types of binary tree, Trees and forests- B- tree- Indexing tric tree indexing Summary Self-Assessment Questions

CHAPTER 4 GRAPH Introduction, Terminology, representation of graph, operations and graphs, Application of graphs structure, Summary, Self-Assessment Questions

Chapter 5: SET: Introduction, Types of Sets, Subset, Operations on Sets, Intersection of Sets, Venn Diagrams, The Disjoint Set ADT, Application of set, Summary, Self-Assessment Questions

Chapter 6 Searching and Sorting Introduction, Linear Search, Binary Search, Introduction To Sorting, Sorting on Multiple Keys 6.6 Bubble Sort, Insertion Sort, Selection Sort, Merge Sort, Quick Sort, Heap Sort, Tree Sort, Summary, Self-Assessment Questions

SUBJECT: ADVANCE DATABASE MANAGEMENT SYSTEM

Chapter 1- ADBMS: Concepts and Architectures: Centralized Database System, Client Server Model, Transaction servers, Data servers, Parallel Systems, Interconnection Network Architectures, Parallel database architecture, Distributed systems, Cloud based servers, Two Tier architecture, Three Tier architecture, N-Tier Architecture, SOAP (Simple Object Access protocol), Summary, Self-Assessment questions.

Chapter 2- Parallel Databases – Parallel database, Client Server Model, I/O (Input / Output) Parallelism, Horizontal Partitioning, Handling of skew, Evaluating parallel Query in Parallel databases, Parallel Processing of Relational Operations, Parallel Sort, Fragment and Replicate Join, Partitioned Parallel Hash-Join, Interoperation parallelism, Query optimization, Design of Parallel Systems, Virtualization, Summary, Self-Assessment Questions.

Chapter 3- Distributed Database Management System (DDBMS) – Types of Distributed databases, Distributed DBMS Architectures, Architectural Models, Design alternatives, Design Strategies, Distribution, Transparency, Combination of Transparencies, Database Control, Relational algebra for Query Optimization, query processing, Relational Algebra, Computation of Relational Algebra, Operators, Computation of Selection, Computation of Joins, Transaction Processing Systems, Schedules and Conflicts, Serial ability, Controlling concurrency, Concurrency in Control in distributed systems, Conflict graphs, Distributed Optimistic Concurrency Control Algorithm Failure and Commit, Commit Protocol, Transaction log, Database recovery, Summary, Self –Assessment Questions.

Chapter 4- Specialty Database and Applications: Object – Relational database Management System (ORDBMS), Object Oriented Database, What is Object, Object Identity, Attributes, Database design for ORDBMS, Non-Temporal database, Temporal Database, Spatial Data, Geographic Data, Multimedia Data, Mobility and Personal database, A Model of Mobile Computing, Dis-connectivity and Consistency, Version Vector Scheme, summary, Self-Assessment Questions.

Chapter 5- Database Warehousing : Data Warehouse, Why a data Warehouse is separated from Operational Database, Features, Data warehouse Application, Types of Data Warehouse, Architecture, Three Tier Data Warehouse Architecture, Data Warehouse Development, Data Warehouse back –end Tools and Utilities, Metadata, Dimensional data modelling, Dimension Normalization, Star Schema, Snowflake Schema, Difference between snowflake and start schema, Fact Constellation, OLAP, Data Cubes, Data pre-Processing, Data Integration, Data Transformation, Data Reduction, Summary, Self-Assessment Questions.

Chapter 6 – Knowledge Base System and Data Mining: Knowledge Discovery, data Mining, data Mining and Business Intelligence, Data Mining: Confluence of Multiple Disciplines, machine Learning, Association rules, Market Basket Analysis, Corporate Analysis and Risk Management, Decision tree Induction,

Bayesian Classification, Clustering, Regression, Pattern Discovery, Visual Data Mining, Artificial Neural Networks (ANN), Genetic Algorithm (GAs), Text Mining, Summary, Self-Assessment Questions.

Chapter 7 – Data Exchange Through XML: XML, Well-Formed, Document Type Definition (DTD), Entities, XML Schema, DOM Parser, SAX (Simple API for XML), XSLT, Summary, Self-Assessment Questions.

SUBJECT: DESIGN AND ANALYSIS OF ALGORITHM

Chapter 1: Introduction: Algorithm, analysis, Time complexity and space complexity, O-notation, Omega notation and Theta notation, Heaps and Heap sort, Sets and disjoint set, union and find algorithms, Sorting in linear time, Tower of Hanoi, Summary, Self-Assessment Questions.

Chapter 2: Divide and Conquer: Divide and Conquer, General Strategy, Exponentiation. Binary Search, Quick Sort, Merge Sort, Summary, Self-Assessment Questions.

Chapter 3: Greedy Method: General Strategy, Knapsack problem, Job sequencing with Deadlines, Optimal merge patterns, Minimal Spanning Trees, Dijkstra’s algorithm, Summary, Self-Assessment Questions.

Chapter 4: Dynamic Programming: General Strategy, Multistage graphs, OBST, 0/1 Knapsack, Traveling Salesperson Problem, Flow Shop Scheduling, Summary, Self-Assessment Questions.

Chapter 5: Back Tracking: Backtracking: General Strategy, N- Queen’s problem, Graph Coloring, Hamiltonian Cycles, 0/1 Knapsack, Summary, Self-Assessment Questions.

Chapter 6: Branch and Bound: General Strategy, 0/1 Knapsack, Traveling Salesperson Problem, Summary, Self-Assessment Questions.

Year 2:

Elective: Data Science

Essential of Data Science

Chapter 1 Introduction to Data Science

Structure:

- 1.1 Data Science: Meaning
- 1.2 The Ubiquity of Data Opportunities
- 1.3 What Do Data Science Do?
- 1.4. Doing Data Science
- 1.5 A Practical Definition
- 1.6 A Data Science, Engineering, and Data-Driven Decision Making
- 1.7 Data Processing and “Big Data”
- 1.8 From Big Data 1.0 to Big Data 2.0
- 1.9 Data and Data Science Capability as a Strategic Asset
- 1.10 Data-Analytic Thinking
- 1.11 Data Mining and Data Science, Revisited

- 1.12 Chemistry Is Not About Test Tubes: Data Science VS the Work of the Data Scientist
- 1.13 Summary
- 1.14 Self Assessment Questions

Chapter 2 Probability And Statistics

Structure:

- 1.1 Introduction
- 1.2 General Notions of the Probability Theory
- 1.3 Approaches to Probability
- 1.4 Relative Frequency Interpretation of Probability
- 1.5 Basic Rules for Finding Probabilities
- 1.6 Statistics Introduction
- 1.7 Measures of central tendency & Measures of Dispersion
- 1.8 Introduction to Statistical Inference
- 1.9 Hypothesis Testing
- 1.10 Summary
- 1.11 Self-Assessment Questions

Chapter 3 Simulation And Hypothesis Testing

Structure:

- 1.1 Introduction: Simulation
- 1.2 Simulation Methodology
- 1.3 Simulating Proportions
- 1.4 Simulating Difference of Means
- 1.5 Simulating the Null
- 1.6 The Role of Technology in Simulation
- 1.7 Simulation Input and Output Data Analysis
- 1.8 Verification and Validation of Simulation Models
- 1.9 Capturing Randomness in Input Data
- 1.10 Explorations in hypothesis testing, error, and statistical power
- 1.11 Randomized Simulation - Simulating the Null Hypothesis
- 1.12 Summary
- 1.13 Self-Assessment Questions

Chapter 4 Data Exploration And Visualization

Structure:

- 1.1 Introduction Data Exploration and Visualization
- 1.2 Features of Data Exploration
- 1.3 Visualization
- 1.4. Data Visualization and Summary Statistics
- 1.5 Variable Types
- 1.6 OLAP Operations
- 1.7 Summary
- 1.8 Self-Assessment Questions

Chapter 5 Data Cleaning And Transformation

Structure:

- 1.1 Introduction Data Cleaning and Transformation
- 1.2 Role, and requirement of data cleaning and transformation at front end
- 1.3 Data Preprocessing
- 1.4 Forms of data preprocessing
- 1.5 Strategies of Data Mining
- 1.6 Closed Loop Data Mining
- 1.7 The Data Mining Process
- 1.8 Cleaning data for data mining purposes usually requires the following steps
- 1.8 Summary
- 1.9 Self Assessment Questions

Chapter 6 Introduction To Machine Learning

Structure:

- 1.1 Machine Learning: Introduction and Meaning
- 1.2 Zoo of ML Problems
- 1.3 The importance of a good representation
- 1.4 What is Machine Learning?
- 1.5 Summary

1.6 Self Assessment Questions

Chapter 7 Introduction To R-Programming For Data Science

Structure:

- 1.1 Meaning
- 1.2 R Overview
- 1.3 R- Basics
- 1.4 Control structures
- 1.5 Matrix operations
- 1.6 Reading & writing Data
- 1.7 Importing data files using the scan function
- 1.8 Working with Lists
- 1.9 Strengths and Weaknesses
- 1.10 Summary
- 1.11 Self-Assessment Questions

Machine Learning concepts using Python

Chapter 1: Machine Learning and Classification Algorithms

Structure:

- 1.1 Building intelligent machines to transform data into knowledge
- 1.2 Types of machine learning
- 1.3 An introduction to the basic terminology and notations
- 1.4 A roadmap for building machine learning systems
- 1.5 Using Python for machine learning
- 1.6 Artificial neurons – a brief glimpse into the early history of machine learning
- 1.7 Implementing a perceptron learning algorithm in Python
- 1.8 Adaptive linear neurons and the convergence of learning
- 1.9 Summary
- 1.10 Self Assessment Questions

Chapter 2: A Tour of Machine Learning Classifiers Using Scikit-learn and Data Preprocessing

Structure:

- 2.1 Choosing a classification algorithm
- 2.2 Modeling class probabilities via logistic regression
- 2.3 Maximum margin classification with support vector machines
- 2.4 Solving nonlinear problems using a kernel SVM
- 2.5 Decision tree learning
- 2.6 K-nearest neighbors – a lazy learning algorithm
- 2.7 Building Good Training Sets – Data Preprocessing
- 2.8 Handling categorical data
- 2.9 Partitioning a dataset in training and test sets
- 2.10 Bringing features onto the same scale
- 2.11 Selecting meaningful features
- 2.12 Summary
- 2.13 Self Assessment Questions

Chapter 3: Compressing Data via Dimensionality Reduction

Structure:

- 3.1 Unsupervised dimensionality reduction via principal component analysis
- 3.2 Principal component analysis in scikit-learn
- 3.3 Supervised data compression via linear discriminant analysis
- 3.4 Using kernel principal component analysis for nonlinear mappings
- 3.5 Kernel principal component analysis in scikit-learn
- 3.6 Summary
- 3.7 Self Assessment Questions

Chapter 4: Combining Different Models for Ensemble Learning

Structure:

- 4.1 Learning with ensembles
- 4.2 Implementing a simple majority vote classifier
- 4.3 Combining different algorithms for classification with majority vote
- 4.4 Evaluating and tuning the ensemble classifier
- 4.5 Bagging – building an ensemble of classifiers from
- 4.6 Leveraging weak learners via adaptive boosting
- 4.7 Summary
- 48 Self Assessment Questions

Chapter 5: Applying Machine Learning to Sentiment Analysis

Structure:

- 5.1 Obtaining the IMDb movie review dataset
- 5.2 Introducing the bag-of-words model
- 5.3 Assessing word relevancy via term frequency-inverse
- 5.4 Cleaning text data
- 5.5 Processing documents into tokens
- 5.6 Training a logistic regression model for document classification
- 5.7 Working with bigger data – online algorithms and out-of-core learning
- 5.8 Summary
- 5.9 Self Assessment Questions

Chapter 6: Predicting Continuous Target Variables with Regression Analysis

Structure:

- 6.1 Introducing a simple linear regression model
- 6.2 Exploring the Housing Dataset
- 6.3 Visualizing the important characteristics of a dataset
- 6.4 Implementing an ordinary least squares linear regression model
- 6.5 Estimating the coefficient of a regression model via scikit-learn
- 6.6 Evaluating the performance of linear regression models
- 6.7 Turning a linear regression model into a curve – polynomial regression
- 6.8 Modeling nonlinear relationships in the Housing Dataset
- 6.9 Dealing with nonlinear relationships using random forests
- 6.10 Summary
- 6.11 Self Assessment Questions

Chapter 7: Working with Unlabeled Data – Clustering Analysis

Structure:

- 7.1 Grouping objects by similarity using k-means
- 7.2 K-means++
- 7.3 Hard versus soft clustering
- 7.4 Using the elbow method to find the optimal number of clusters
- 7.5 Quantifying the quality of clustering via silhouette plots
- 7.6 Organizing clusters as a hierarchical tree

- 7.7 Applying agglomerative clustering via scikit-learn
- 7.8 Summary
- 7.9 Self Assessment Questions

Big Data Hadoop

Chapter 1: Big Data Basics

Structure

- 1.1 Big data Introduction
- 1.2 Human generated data Vs Machine generated data
- 1.3 Challenges of big data (Structured, Semi structured, Un-structured data)
- 1.4 Technologies for Big data support
- 1.5 Examples of Big data in the real world
- 1.6 Computing systems
- 1.7 Big data open-source solution – Apache Hadoop
- 1.8 How Hadoop solves Big data problem
- 1.9 RDBMS Vs Hadoop
- 1.10 Big data terms
- 1.11 Summary
- 1.12 Self-Assessment Questions

Chapter 2: Big data Applications

Structure

- 2.1 Introduction
- 2.2 Big data analytics technologies and tools
- 2.3 Marketing applications
- 2.4 Fraud Detections applications
- 2.5 Risk analysis applications
- 2.6 Health care applications
- 2.7 Summary
- 2.8 Self-Assessment Questions

Chapter 3: Hadoop Distributed File System (HDFS)

Structure

- 3.1 Introduction to Apache Hadoop
- 3.2 Apache Hadoop ecosystem
- 3.3 Hadoop Vs Java

- 3.4 HDFS: Hadoop Distributed File System
- 3.5 Features of HDFS
- 3.6 Daemons of Hadoop
- 3.7 Replication and data organization in HDFS
- 3.9 Accessing HDFS
 - 3.9.1 Unix commands
 - 3.9.2 Read/Write steps in HDFS
- 3.10 Installation and Set-up of Hadoop
- 3.11 HDFS Access
 - Hadoop serialization programs
- 3.12 HDFS read/write architecture
- 3.13 Summary
- 3.14 Self-Assessment Questions.

Chapter 4: Map Reduce Programming Model

Structure

- 4.1 Map Reduce introduction
- 4.2 Applications of Map Reduce
- 4.3 Map Reduce Process steps
- 4.4 Map Reduce Architecture
- 4.5 Map Reduce working Examples
- 4.6 Built in MapReduce Algorithms.
- 4.7 Writing MapReduce Programs.
- 4.8 Summary
- 4.9 Self-Assessment Questions

CHAPTER 5: PIG

Structure

- 5.1 PIG Introduction
- 5.2 Map Reduce Vs. Apache Pig
- 5.3 SQL vs. Apache Pig
- 5.4 Installation of PIG
- 5.5 Different data types in Pig
- 5.6 Modes of Execution in Pig
- 5.7 Loading data
- 5.8 Exploring Pig
- 5.9 Latin commands
- 5.10 Sample PIG programs

CHAPTER 6: HIVE

Structure

- 6.1** Hive introduction
- 6.2** Hive architecture
- 6.3** Hive Vs RDBMS
- 6.4** Installation of HIVE
- 6.5** HiveQL and the shell
- 6.6** Managing tables (external Vs managed)
- 6.7** Data types and schemas
- 6.8** Partitions and buckets
- 6.9** HIVE commands
- 6.10** Sample HIVE programs
- 6.11** Summary
- 6.12** Self-Assessment Questions

Chapter 7: HBASE

Structure

- 7.1.** Introduction
- 7.2 Hbase Architecture
- 7.3 HBase vs. RDBMS
- 7.4 Hbase Installation
- 7.5 Hbase commands
- 7.6 Read/write steps in Hbase
- 7.7 Sample HBASE program
- 7.8 Summary
- 7.9 Self-Assessment Questions

Business Application of Machine Learning

Chapter 1 Introduction to Artificial Intelligence & Machine Learning

Structure:

- 1.1 Artificial Intelligence: Basic Concepts
- 1.2 AI, Big Data, Data Science, Machine Learning: Relationship
- 1.3 Beyond the AI Hype
- 1.4 Summary
- 1.5 Self-Assessment Questions

Chapter 2 Machine Learning Application in Customer Service

Structure:

- 2.1 Customer Experience
- 2.2 Chatbots
- 2.3 Call Center Automation
- 2.4 Summary
- 2.5 Self-Assessment Questions

Chapter 3 Machine Learning Application in Hospitality

Structure:

- 3.1 Best Fit / Dynamic Pricing
- 3.2 Demand Forecasting
- 3.3 Summary
- 3.4 Self-Assessment Questions

Chapter 4 Machine Learning Application in Banking & Financial Services

Structure:

- 4.1 Robo-Advisers based Portfolio Management
- 4.2 Loan & Insurance Underwriting
- 4.3 Summary
- 4.4 Self-Assessment Questions

Chapter 5 Machine Learning Application in Retail

Structure:

- 5.1 Online Recommendation
- 5.2 Inventory Optimization
- 5.3 Summary
- 5.4 Self-Assessment Questions

Chapter 6 Future Potential of Machine Learning

Structure:

- 6.1 Technological Advances in Machine Learning
- 6.2 Future Possibilities Unlimited
- 6.3 Summary

6.4 Self-Assessment Questions

Chapter 7 Summary

Structure:

7.1 Machine Learning Application Recap

7.2 Self Assessment Questions

Statistics and Python in Machine Learning

Chapter 1 Getting Started With Python

Structure:

- 1.1 Introduction to python
- 1.2 Python Programming Domains
- 1.3 Python installation
- 1.4 Setting path in python
- 1.5 Execute modes in python
- 1.6 Python keywords, identifiers and variables
- 1.7 Python statement
- 1.8 Structuring with indentation
- 1.9 Comments in python
- 1.10 Python data types
- 1.11 Python lists
- 1.12 Python tuples
- 1.13 Shallow and deep copy
- 1.14 Python dictionaries
- 1.15 Python sets and types
- 1.16 Python operators
- 1.17 Python control structures
- 1.18 Summary
- 1.19 Self assessment questions

Chapter 2 Functions, File Handling and Exceptions

Structure:

- 2.1 Python Functions
- 2.2 Definition of Recursion

- 2.3 Python Input And Output
- 2.4 Python File Handling
- 2.5 With Statement
- 2.6 Exception Handling
- 2.7 User-defined Exceptions
- 2.8 Built-in Exceptions
- 2.9 Summary
- 2.10 Self assessment questions

Chapter 3 Data Manipulation Using NumPy and Pandas

Structure:

- 3.1 Python Libraries
- 3.2. NumPy: Arrayys and Metrics
- 3.3 Pandas: Data Manipulation
- 3.4 File I/O
- 3.5 The Fisher's F- distribution
- 3.6 Case Study
- 3.7 Summary
- 3.8 Self assessment questions

Chapter 4 Introduction to Statistics and Probability

Structure:

- 4.1 Statistics in python
- 4.2 Probability distributions
- 4.3 9 most commonly used probability distributions
- 4.4 Hypothesis testing (T Test)
- 4.5 Linear regression
- 4.6 Statistical Data Modeling
- 4.7 Introduction to Bayesian Thinking
- 4.8 Markov Chains in Python
- 4.9 Summary
- 4.10 Self Assessment Questions

Chapter 5 Mathematical and Statistics Functions

Structure:

- 5.1 Mathematical Functions in Python
- 5.2 Statistical functions in python
- 5.3 Histogram (hist) function with multiple data sets
- 5.4 Box plot vs. violin plot comparison
- 5.5 Producing multiple histograms side by side
- 5.6 Barchart
- 5.7 Scatter plots
- 5.8 Line-style reference
- 5.9 Scatter Hist
- 5.10 Pie Chart
- 5.11 Polar Demo
- 5.12 Summary
- 5.13 Self assessment questions

Chapter 6 Data Visualization in Python

Structure:

- 6.1 Introduction of data Visualization
- 6.2 Advance Visualization Methods
- 6.3 Visualization libraries in python
- 6.4 Plots in Python
- 6.5 Colormap in Plots
- 6.6 Style sheets reference
- 6.7 Summary
- 6.8 Self Assessment Questions

TIME SERIES ANALYSIS

| Unit-1 | |
|-------------------|--------------------------------------|
| Chapter 1: | Introduction to Time series analysis |

| | |
|--------------------------|---|
| Contents in the chapter: | <ul style="list-style-type: none"> 1.1 Meaning of time series analysis 1.2 Definitions 1.3 Characteristics of Time series analysis 1.4 Important Concepts (Missing Data, Sample Size, Representativeness, Stationarity etc.) 1.5 Scope of Time series analysis 1.6 Components of Time series 1.7 Summary |
| Chapter 2: | Forecasting Overview |
| Contents in the chapter: | <ul style="list-style-type: none"> 2.1 Introduction to Forecasting 2.2 Time Series Modeling and Forecasting 2.3 Examples of Time Series and Forecasting, 2.4 Forecasting Data, 2.5 Steps in Forecasting 2.6 Conclusion |
| Chapter 3: | Decomposition of Time Series |
| Contents in the chapter: | <ul style="list-style-type: none"> 3.1 Introduction 3.2 Goodness of Fit Indicators 3.3 Averaging techniques for decomposition 3.4 Wold's Decomposition Theorem 3.5 Conclusion |
| Chapter 4: | Visualization of Time Series Data |
| Contents in the chapter: | <ul style="list-style-type: none"> 4.1 Introduction 4.2 Use of Graphical Tools in Time Series Analysis 4.3 Types of Graphical Tools used in Time Series Analysis 4.4 Time Series Plots 4.5 Demerits of Graphical Visualization 4.6 Summary |
| Unit-2 | |
| Chapter 5: | Forecasting Methods and Exponential Smoothing |
| Contents in the chapter: | <ul style="list-style-type: none"> 5.1 Introduction 5.2 Important terms used in Forecasting Methods (Series, Observed Vs. Expected, Method, Error, SSE and MSE) 5.3 Methods of Forecasting (Naïve Method, Simple Average, Moving Average, Weighted Moving Average, Exponential Smoothing) 5.4 Simple Exponential Smoothing(SES) 5.5 Double Exponential Smoothing 5.6 Triple Exponential Smoothing 5.7 Evaluation |
| Chapter 6: | ARMA(Autoregressive Moving Average) Models |

| | |
|--------------------------|--|
| Contents in the chapter: | <ul style="list-style-type: none"> 6.1 Introduction 6.2 The Autoregressive Moving Average Models 6.3 The Auto Correlation Function(ACF) and Partial Auto Corelation Function(PACF) 6.4 The Auto Covariance Function (ACF) 6.5 Conclusion |
| Chapter 7: | Modeling and Forecasting with ARMA Processes |
| Contents in the chapter: | <ul style="list-style-type: none"> 7.1 Preliminary Estimation (Yule–Walker Estimation; Burg’s Algorithm; The Innovations Algorithm; The Hannan–Rissanen Algorithm) 7.2 Maximum Likelihood Estimation 7.3 Diagnostic Checking 7.4 Forecasting 7.5 Order Selection 7.6 The FPE Criterion(Final Prediction Error) 7.7 The AICC Criterion(The Corrected Akaike Information Criterion) 7.8 Conclusion |
| Chapter 8: | Non stationary and Seasonal Time Series Models |
| Contents in the chapter: | <ul style="list-style-type: none"> 8.1 Introduction 8.2 ARIMA Models 8.3 Identification of ARIMA Models 8.4 Unit Roots in Time Series Models 8.5 Steps of Forecasting from ARIMA Models 8.6 Seasonal ARIMA Models (SARIMA) 8.7 Conclusion |
| Unit-3 | |
| Chapter 9: | Box–Jenkins Model of Forecasting |
| Contents in the chapter: | <ul style="list-style-type: none"> 9.1 Introduction 9.2 Relevance of Box Jenkins Method in Forecasting 9.3 The Three Stage Modeling Approach 9.4 Steps of Modeling 9.5 Conclusion |
| Chapter 10: | Multivariate Time Series |
| Contents in the chapter: | <ul style="list-style-type: none"> 10.1 Multivariate Time Series: Introduction 10.2 Second-Order Properties of Multivariate Time Series. 10.3 Estimation of the Mean and Covariance Function 10.4 Multivariate ARMA Processes 10.5 Best Linear Predictors of Second-Order Random Vectors 10.6 Modeling and Forecasting with Multivariate AR Processes 10.7 Cointegration |

| | |
|--------------------------|---|
| | |
| Chapter 11: | Non Linear Time Series Models |
| Contents in the chapter: | 11.1 Introduction 11.2 Examples of Non Linear Time Series Models 11.3 ARCH Model 11.4 The GARCH Model 11.5 Bilinear Model |
| Chapter 12: | Transfer Function Models |
| Contents in the chapter: | 12.1 Introduction 12.2 Transfer Function Models 12.3 Transfer Function-Noise Models 12.4 Cross Correlation Function 12.5 Model checking 12.6 Forecasting |
| Unit-4 | |
| Chapter 13: | Intervention Analysis |
| Contents in the chapter: | 13.1 Introduction 13.2 Assumptions of the Event Intervention (Impact) Model 13.3 Analysis Theory 13.4 Advantages of Intervention Analysis 13.5 Limitations of Intervention Analysis |
| Chapter 14: | Other Forecasting Techniques |
| | 14.1 The ARAR Algorithm 14.2 The Holt–Winters Algorithm 14.3 The Holt–Winters Seasonal Algorithm 14.4 State Space Models 14.5 Spectral Analysis 14.6 Choosing a Forecasting Algorithm 14.7 Conclusion |

Deep Learning and NLP

Chapter 1 Artificial Intelligence Overview

Structure:

- 1.1 Artificial Intelligence
- 1.2 Machine Learning
- 1.3 Deep Learning
- 1.4 Natural Language Processing

1.5 Summary

1.6 Self-Assessment Questions

Chapter 2 Review of Mathematical Concepts

Structure:

2.1 Overview

2.2 Linear Algebra

2.2 Probability & Statistics

2.3 Calculus

2.4 Summary

2.5 Self-Assessment Questions

Chapter 3 Introduction to Artificial Neural Networks

Structure:

3.1 Context

3.2 Research into ANN

3.3 Brain & Neurons

3.4 Initial concepts in ANN

3.5 Summary

3.6 Self-Assessment Questions

Chapter 4 Modelling a Neuron

Structure:

4.1 Network Topology

4.2 Neural Network Learning

4.3 Activation Functions

4.4 Summary

4.5 Self-Assessment Questions

Chapter 5 Deep Learning Algorithms

Structure:

- 5.1 Perceptron Learning Algorithm
- 5.2 Back Propagation Neural Network Algorithm
- 5.3 Deep Boltzmann Machine (DBM) Learning Algorithm
- 5.4 Deep Belief Network (DBN) Learning Algorithm
- 5.5 Convolutional Neural Network (CNN) Learning Algorithm
- 5.6 Stacked Autoencoder Learning Algorithm
- 5.7 Summary
- 5.8 Self-Assessment Questions

Chapter 6 Deep Learning Platforms and Frameworks

Structure:

- 6.1 Overview
- 6.2 TensorFlow
- 6.3 Keras
- 6.4 Microsoft Cognitive Toolkit (CNTK)
- 6.5 PyTorch
- 6.6 Others
- 6.7 Summary
- 6.8 Self-Assessment Questions

Chapter 7 Deep Learning Project

Structure:

- 7.1 API Environment Overview
- 7.2 Project Scenario and Solution
- 7.3 Summary
- 7.4 Self-Assessment Questions

Chapter 8 NLP Introduction

Structure:

- 8.1 Overview and History of NLP
- 8.2 Business Applications of NLP
- 8.3 Summary
- 8.4 Self-Assessment Questions

Chapter 9 Approach to NLP

Structure:

- 9.1 Introduction to NLP and NLTK
- 9.2 Text Pre-processing
- 9.3 Feature Engineering on Text
- 9.4 NLP use cases
- 9.5 Summary
- 9.6 Self-Assessment Questions

IOT Cloud and Watson Analytics

Chapter 1:

Structure:

- 1.1 Introduction to Networking
 - 1.1.1 Wired and Wireless Networks
 - 1.1.2 Devices used in Networking
- 1.2 Introduction to IoT
 - 1.2.1 History of IoT
 - 1.2.2 Advantages of IoT
 - 1.2.3 Overview

- 1.2.4 Characteristics of IoT
- 1.2.5 IoT Definitions
 - 1.2.6.1 IoT Detailed Architecture
 - 1.2.6.2 Three and five layer architectures
 - 1.2.6.3 Cloud and Fog Based Architectures
 - 1.2.6.4 Representative Architecture
 - 1.2.6.5 Service Oriented architecture (SOA)
 - 1.2.6.6 API Oriented Architecture
- 1.2.7 IoT Framework
- 1.2.8 Big Data Analytics
- 1.2.9 Smart Objects
- 1.2.10 Smart Applications

Chapter 2: Fundamental IOT Mechanism and Key Technologies

Structure:

- 2.1 Internet Principles
- 2.2 Identification of IoT Objects and Services:
- 2.3 Device Intelligence:
- 2.4 Mobility Support
- 2.5 Device Power
- 2.6 Sensor Technology
- 2.7 RFID Technology
- 2.8 Satellite Technology
- 2.9 Raspberry Pi & Arduino device

Chapter 3: IoT Cloud and Watson Analytics

Structure:

- 3.1 Introduction to cloud computing
- 3.2 The internet of Things and Cloud Computing
- 3.3 Mobile Cloud Computing
- 3.4 Integration of Cloud computing and Internet of Things(CloudIoT):
- 3.5 Main CloudIoT drivers
- 3.6 Internet of Things to Smart IoT Through Semantic, Cognitive, and Perceptual Computing

- 3.7 Cognitive computing and the Internet of Things
- 3.8 Introduction to Watson IoT Platform
- 3.9 Important concepts in the Watson IoT Platform
- 3.10 Watson IoT Platform Feature Overview
- 3.11. Standards and requirements
- 3.12 IBM Cloud and the Watson IoT Platform
- 3.13 Introduction to data management
- 3.14 Introduction to cognitive computing
- 3.14 Introduction to Watson Analytics
- 3.15 Watson APIs: Build with Watson
- 3.16 IBM Watson applied to industries, businesses, and science
- 3.17 Watson use cases

Resource Management in Internet of Things

Structure:

- 4.1 Clustering
- 4.2 Software Agents
- 4.3 Clustering Principles in an Internet of Things Architecture
- 4.4 Design Guidelines
- 4.5 Software Agents for Object Representation
- 4.6 Data Synchronization
- 4.7 Identity Portrayal
- 4.8 Identity Management
- 4.8 Federated Identity Management Model
- 4.9 User-Centric Identity Management
- 4.10 Device Centric Identity Management and Hybrid-Identity Management

Internet of Things: Privacy, Security and Governance

Structure:

- 5.1 Issues and Challenges in IoT Security

- 5.2 Vulnerabilities of IoT
- 5.3 Security Requirements
- 5.4 Threat Analysis
- 5.5 Use Cases and misuse cases
- 5.6 Activity Modeling and Threats
- 5.7 Security Threats at different layers of IoT Architecture
- 5.8 Identity Establishment
- 5.9 Access control
- 5.10 Non-Repudiation and Availability
- 5.11 Security model for IoT
- 5.12 Providing security on layers to defend IoT
- 5.13 Security Measures for IoT Platforms/Devices
- 5.14 IoT Governance

IOT APPLICATIONS AND CASE STUDIES

Structure:

- 6.1 Connected Life
- 6.2 Consumer and socio-economic impact
- 6.3 Distinctive features of IOT
- 6.4 Application-Cloud with IoT in HealthCare
- 6.4 Application- Smart cities and communities
- 6.4 Application- Smart home and smart metering
- 6.4 Application- Automotive and Smart mobility
- 6.5 IoT Some Examples
- 6.6 Patras: Internet of Things Case Study:
- 6.7 AirQ: Air Quality Internet of Things Case Study
- 6.8 Case Study: Transforming to an engaged and connected city(The City of Mississauga)
- 6.9 Internet of Things in Agriculture: a Case Study of Smart Dairy Farming in Ontario, Canada
- 6.10 HP Turns Printers into IoT Hubs in the Home

Social Media, Web Analytics and Visualization

Chapter 1 Introduction to Web Analytics

Structure:

- 1.1 Web analytics: introduction
- 1.2 The purpose of web analytics
- 1.3 The chronological journey of web analytics
- 1.4 Benefits of web analytics
- 1.5 Challenges of web analytics
- 1.6 Importance of web analytics
- 1.7 Unlock your website's potential through web analytics
- 1.8 Grow your business with web analytics
- 1.9 Using web analytics basics
- 1.10 The five 'whs' of web analytics
- 1.11 Most commonly used terms
- 1.12 Basic criteria's to choose a web analytic tool
- 1.13 Top 7 web analytics tools
- 1.14 Categories of web analytics software
- 1.15 Web analytic dashboard
- 1.16 Metrics categories
- 1.17 Site referrers
- 1.18 Identifying your audience
- 1.19 How to identify the most important pages?
- 1.20 Key performance indicators
- 1.21 Ten most common web analytics mistakes and pitfalls
- 1.22 Web analytics: best practices
- 1.23 How to know if your analytic data is performing well.
- 1.24 Minimizing cost using testing software
- 1.25 Web analytics methods
- 1.26 Summary
- 1.27 Self assessment questions

Chapter 2 Web Analytics Tools

Structure:

- 2.1 Web analytics: objectives and process
- 2.2 Web analytics tools
- 2.3 Traffic & trends of websites
- 2.4 Marketing automation & conversion metrics

- 2.5 Keywords, seo, ppc & competitor research
- 2.6 User behaviour & flow
- 2.7 Google analytics
- 2.8 Analyzing data through google analytics
- 2.9 Optimizely
- 2.10 Kissmetrics
- 2.11 Crazy egg
- 2.12 Data source
- 2.13 Dashboard
- 2.14 Goals of analytics
- 2.15 Web traffic data analysis
- 2.16 The top 5 pillars of web analytics
- 2.17 Increasing your site's visibility through web analytics
- 2.18 Summary
- 2.19 Self assessment questions

Chapter 3 Introduction to Social Media Analytics

Structure:

- 3.1 Social media
- 3.2 From Social Media to Social Media Analytics
- 3.3 Social media analytics
- 3.4 Importance of tracking Social Media Analytics
- 3.5 Benefits of social media
- 3.6 Social media communities
- 3.7 Data analytics
- 3.8 Data analytics in social media
- 3.9 Tools for social media analytics
- 3.10 Structured and Unstructured Data
- 3.11 How social-media analytics help individuals and organizations
- 3.12 How to use social data analytics to improve your marketing strategies
- 3.13 Popular Social Media Tools and Platforms
- 3.14 Benefits of adopting Social Media Strategy
- 3.15 Organizational Goals for Social Media
- 3.16 Social Media Impact on Business

- 3.17 Summary
- 3.18 Self assessment questions

Chapter 4 Social Media Platforms and Analytics

Structure:

- 4.1 Social Media Platform
- 4.2 Facebook Insights
- 4.3 Twitter Analytics
- 4.4 Pinterest Analytics
- 4.5 LinkedIn Analytics
- 4.6 Google+ Analytics
- 4.7 Instagram Analytics
- 4.8 YouTube Analytics
- 4.9 Blogging Analytics
- 4.10 How to Turn Social Media Analytical Data Into Actionable Insights
- 4.11 Gaining an Advantage from Social Media Trends
- 4.12 Summary
- 4.13 Self assessment questions

Chapter 5 Social Media Metrics, ROI and Risks

Structure:

- 5.1 Social Media Metrics: Introduction
- 5.2 Tools to track Social media metric
- 5.3 Social media ROI (Return on Investment)
- 5.4 Importance of measuring ROI
- 5.5 Social media ROI: objectives, goals, and metrics
- 5.6 Social media ROI tools
- 5.7 Report your social media ROI
- 5.8 Big Data Analytics on Social Media
- 5.9 Social Media Mining
- 5.10 Issues in Mining Social Media
- 5.11 Social Media Optimization & Search Engine Optimization
- 5.12 Strategies for Social Media Optimization

- 5.13 Social Media risks
- 5.14 Social Media Risk Assessment
- 5.15 Types of Business Risks
- 5.16 Steps to Effectively Managing Social Media Risk
- 5.17 Legal & Regulatory Challenges with Social Media
- 5.18 Summary
- 5.19 Self assessment questions

Chapter 6 Social Media Benchmarking, Reporting And Visualization

Structure:

- 6.1 Improving performance through Social Media Benchmarking
- 6.2 Metrics to Benchmark
- 6.3 Things can be benchmarked on social media
- 6.4 Social media analytics reports
- 6.5 Visualize your Social Media Analytics
- 6.6 Tips for visualizing social media data
- 6.7 Best Data Visualization Tools
- 6.8 Summary
- 6.9 Self assessment questions

Spark on Azure HD

Chapter 1 Introduction to Spark

Structure:

- 1.1 Overview about Data Science
- 1.2 Introduction of Hadoop
- 1.3 Introduction of MapReduce
- 1.4 Spark Overview
- 1.5 Evolution of Apache Spark
- 1.6 Spark Vs Traditional MapReduce

1.7 Features of Apache Spark

1.8 Components of Spark

1.9 Summary

1.10 Self Assessment Questions

Chapter 2 Exploring Data with Spark

Structure:

2.1 Introduction.

2.2 HDInsight Spark Offering.

2.3 Apache Spark Terminologies

2.4 Spark Architecture

2.5 RDD(Resilient Distributed Datasets)

2.6 Spark SQL

2.7 Spark Streaming

2.8 Summary

2.9 Self Assessment Questions

Chapter 3 Machine Learning Models

Structure:

3.1 What is Machine Learning

3.2 Importance Machine Learning

3.3 Understanding Machine Learning

3.4 Types of Machine Learning Model

3.5 Machine Learning Algorithms

3.6 Summary

3.7 Self Assessment Questions

Chapter 4 Machine Learning in Spark & Data Preparation

Structure:

4.1 Why to use Spark for Data Preparation and Machine Learning

4.2 How Machine learning Works

4.3.MLlib Library

4.4 Data Preparation

4.5 Building Blocks of Data Preparation

4.6 Data Preparation Methodologies

4.7 Summary

4.8 **Self Assessment Questions**

Chapter 5 Build Machine Learning Solutions in Spark

Structure:

5.1 MLlib Data types

5.2 Basic statistics

5.3 Classification and regression

5.4 Collaborative filtering

5.5 Clustering

5.6 Dimensionality reduction

5.7 Feature extraction and transformation

5.8 Optimization (developer)

5.9 Build a Spark machine learning application in HDInsight

5.10 Summary

5.11 Self Assessment Questions

Chapter 6 R Server on Spark

Structure:

6.1 What is Spark R

6.2 SparkR Architecture

6.3. Why choose R Server in HDInsight

6.4 Key features of R Server on HDInsight

6.5 Data storage options for R Server on HDInsight

6.6 Creating the R Server on Spark Cluster

6.7 Installing RStudio with R Server on HDInsight

6.8 Install an HDInsight application

6.9 List installed HDInsight apps and properties.

6.10 Summary

6.11 Self Assessment Questions

Chapter 7 Machine Learning with R-Server on Spark

Structure:

7.1 Setting Your R Profile

7.2 Starting Your Machine Learning

7.3 Exploratory Data Analysis with SparkR

7.4 Data Manipulation with SparkR

7.5 Modeling with Microsoft R Server

7.6 Use Case – Movie Recommendation System

Basics of R Programming

Chapter 1 Introduction R programming

Structure - 1.1 Introduction - 1.2 R-Environment - 1.3 R-Packages - 1.4 Name spaces - 1.5 Statistics with R - 1.6 R-Studio - 1.7 Analytics with R - 1.8 Business analytics with R - 1.9 Data visualization - 1.10 Import Data into R system - 1.11 Data Manipulations - 1.12 Data Science & Machine Learning with R - 1.13 Installation steps – Windows platform - 1.14 Basic operations - 1.15 Companies using R - 1.16 Self-Assessment Questions - 1.17 Summary

CHAPTER2: Analytics Environment on MS Azure Machine

Structure - 2.1 Introduction - 2.2 Machine Learning Algorithms - 2.3 Azure Machine Learning - 2.4 Azure Machine Learning on Cloud Environment - 2.5 Machine Learning Process - 2.6 Azure Management - 2.7 Azure ML Components - 2.8 Azure Analytical Services - 2.9 Linear Regression using R - 2.10 Summary - 2.11 Self-Assessment Questions

Chapter 3: Basic R Constructs

Structure - - 3.1 Introduction - 3.2 R- Variables - 3.3 Types of Operators - 3.4 Control Structures - 3.5 String operations - 3.6 Vectors - 3.7 Missing values - 3.8 Summary - 3.9 Self-Assessment Questions

Chapter 4: Data Management Techniques in R

Structure - 4.1 Introduction - 4.2 Variable management - 4.3 Built-in functions - 4.4. Vectors - 4.5 Matrix operations - 4.6 Data frame management - 4.7 List Management - 4.8 Arrays - 4.9 Factors - 4.10 Function management in R - 4.11 Summary - 4.12 Self-Assessment Questions - 4.13 References

CHAPTER 5: Data Preparing Techniques

Structure - 5.1 Introduction - 5.2 Raw data extractions - 5.3 Combining datasets - 5.4 Data cleaning - 5.5 Sorting, Ranking and Ordering - 5.6 Data Binning - 5.7 Type conversions - 5.8 Summary - 5.9 Self-Assessment Questions

Chapter 6: Statistical Techniques

Structure - 6.1 Introduction - 6.2 Types of data - 6.3 Measures of Central Tendency - 6.4 Measures of Dispersion - 6.5 Statistical tests - 6.6 Predictive modeling – Linear Regression - 6.7 Summary - 6.8 Self-Assessment Questions