KKR & KSR INSTITUTE OF TECHNOLOGY AND SCIENCE(AUTONOMOUS)
Department of Computer Science and Engineering

Course Code	Course Name	L	T	P	C
20CS6T01	Cryptography and Network Security	3	0	0	3

Course Objectives:

- The concepts of classical encryption techniques and concepts of finite fields and number theory.
- Working principles and utilities of various cryptographic algorithms including secret key cryptography, hashes and message digests, and public key algorithms.
- Design issues and working principles of various authentication protocols, PKI standards.
- Various secure communication standards including Kerberos, IPsec, and SSL/TLS and email.
- Concepts of cryptographic utilities and authentication mechanisms to design secure applications.

Course Outcomes:

At the end of the course, the students will be able to:

CO-1: Identify information security goals, classical encryption techniques and acquire fundamental knowledge on the concepts of finite fields and number theory

CO-2: Compare and apply different encryption and decryption techniques to solve problems related to confidentiality and authentication

CO-3: Apply the knowledge of cryptographic checksums and evaluate the performance of different message digest algorithms for verifying the integrity of varying message sizes.

CO-4: Apply different digital signature algorithms to achieve authentication and create secure applications

CO-5: Apply network security basics, analyze different attacks on networks and evaluate the performance of firewalls and security protocols like SSL, IPSec, and PGP

UNIT I

ClassicalEncryptionTechniques: Classical EncryptionTechniques-Substitution techniques,Transposition techniques, Security Attacks, Services & Mechanisms, Symmetric Cipher Model, Cyber Threats, Phishing Attack, Web Based Attacks.

Block Ciphers: Traditional Block Cipher Structure, Block Cipher Design Principles.

UNIT II

Symmetric Key Cryptography: Data Encryption Standard (DES), Advanced Encryption Standard (AES), Blowfish, IDEA, Block Cipher Modes of Operations.

Number Theory: Prime and Relatively Prime Numbers, Modular Arithmetic, Fermat's and Euler's Theorems, the Chinese Remainder Theorem, Discrete Logarithms.

UNIT III

Public Key Cryptography: Principles, Public Key Cryptography Algorithms, RSA Algorithm, Diffie-Hellman Key Exchange, Elliptic Curve Cryptography. **CryptographicHashFunctions**: Application of Cryptographic Hash Functions, Requirements & Security, Secure Hash Algorithm, Message Authentication Functions, Requirements & Security, HMAC & CMAC.

Digital Signatures: NIST Digital Signature Algorithm, Key Management and Distribution.

UNIT IV

User Authentication: Remote User Authentication Principles, Kerberos. Electronic Mail Security: Pretty Good Privacy (PGP) And S/MIME.

IP Security: IP Security Overview, IP Security Architecture, Authentication Header, Encapsulating Security Payload, Combining Security Associations and Key Management.

UNIT V

Transport Level Security: Web Security Requirements, Secure Socket Layer (SSL) and Transport Layer Security (TLS), Secure Shell (SSH)

Firewalls: Characteristics, Types of Firewalls, Placement of Firewalls, Firewall Configuration, Trusted Systems.

Text Books:

1) Cryptography and Network Security- William Stallings, Pearson Education, 7th Edition.

2) Cryptography, Network Security and Cyber Laws – Bernard Menezes, Cengage Learning, 2010 edition.

Reference Books:

1) Cryptography and Network Security- Behrouz A Forouzan, DebdeepMukhopadhyaya, Mc-GrawHill, 3rd Edition, 2015.

2) Network Security Illustrated, Jason Albanese and Wes Sonnenreich, MGH Publishers, 2003.

E-Resources:

1)https://nptel.ac.in/courses/106/105/106105031/lecture by Dr.DebdeepMukhopadhyay IIT Kharagpur [Video Lecture]

2) https://nptel.ac.in/courses/106/105/106105162/
lecture by Dr. Sourav Mukhopadhyay IIT Kharagpur [Video Lecture]

3)https://www.mitel.com/articles/web-communication-cryptography-and-network-security[web articles by Mitel Power Connections].

DEPARTMENT OF CSE - DATA SCIENCE

SEMESTER - VI

Course Code	Course Name	L	T	P	C
20CD6T02	BIG DATA ANALYTICS	3	0	0	3

Pre- Requisites:
Students should have knowledge of one Programming Language (Java preferably), Practiceof SQL (queries and sub queries), exposure to Linux Environment.

Course Objectives:

1. Describe Big Data and its use cases from selected business domains.
2. Provide an overview of HDFS Architecture and its daemon services.
3. Perform Map Reduce analytics with YARN using Hadoop.
4. Understand the working of data ingestion tools and PIG Latin.
5. Use Hadoop related tools such as Hive and HBase for big data analytics.

Course Outcomes:

At the end of the Course, the student will be able to

1. Understand the concepts of Big Data and navigation of the Hadoop Ecosystem.
2. Illustrate the HDFS Architecture and the coordination service of Hadoop.
3. Implement distributed processing Map Reduce Paradigm with YARN.
4. Analyze importing and exporting data from Hadoop using Sqoop, Flume and workingwith PIG.
5. Examine the data stores - Hive and HBase on Hadoop

UNIT I

Introduction to Big Data and Hadoop:
Challenges of Traditional Decision Making, Solution with Big Data Analytics, Classification of Digital Data, Definition of Big Data, Characteristics of Big Data, Definition of Big Data Analytics, Features of Hadoop, History of Hadoop, RDBMS Vs. Hadoop, Hadoop Distributors, Ecosystems of Hadoop.

UNIT II

HDFS and Zoo Keeper:
HDFS: Concepts – Blocks, HDFS Components, Block Caching, Characteristics of HDFS, HDFS High Availability Architecture and its types, HDFS Command Line, Data Flow – Anatomy of File read and File write operations.
Zoo Keeper: Characteristics of Zoo Keeper, Zoo keeper Services, Zoo keeper Data Model.

UNIT III

Map Reduce and YARN
YARN: Elements of YARN Architecture, Map Reduce: Characteristics of Map Reduce, Phasesof Map Reduce with an Example, Anatomy of MR Job Run with YARN, Handling Failures, Task Execution, Map Reduce Input and Output Formats, Shuffle and Sort, Built - in Countersof MR, Joins in MR,

UNIT IV

Data Ingestion Tools and PIG

Data Ingestion Tools: Data Ingestion, Big Data Ingestion Tools, SQOOP - Benefits of SQOOP, SQOOP Connectors, Importing and Exporting to and from Hadoop using SQOOP, Limitationsof SQOOP, FLUME – Apache Flume, Data Sources for FLUME, Components of FLUME Architecture.

PIG: Introduction to PIG, Components of PIG, Data Types in PIG – Simple and Complex, PIG Execution Modes, PIG Interactive Modes, Comparison of PIG with databases, Data Processing Operators.

UNIT V

HIVE and HBASE

HIVE: Features of HIVE, HIVE Architecture, HIVE Meta store, Data types in HIVE,HIVEQL, Tables, File Format Types – Text, Sequence, AVRO, Parquet, Querying Data.

HBASE: NOSQL Database, Types of NOSQL Database, Characteristics of HBASE,Architecture, HBaseVs. RDBMS, HBASE Shell Commands.

Text Books:

1. Tom White "Hadoop: The Definitive Guide" 4thedition, O'reily Media,2012.
2. SeemaAcharya, SubhasiniChellappan, "Big Data Analytics" Wiley2015

Reference Books:

1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer,2007.
2. Jay Liebowitz, "Big Data and Business Analytics" Auerbach Publications, CRCpress(2013)
3. Tom Plunkett, Mark Hornick, "Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop", McGraw- Hill/Osborne Media (2013), Oracle press.
4. AnandRajaraman and Jefrey David Ulman, "Mining of Massive Datasets",Cambridge University Press, 2012.
5. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", John Wiley & sons,2012.
6. Glen J. Myat, "Making Sense of Data", John Wiley & Sons,2007
7. Pete Warden, "Big Data Glossary", O'Reily,2011.
8. Michael Mineli, Michele Chambers, AmbigaDhiraj, "Big Data, Big Analytics:Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley Publications, 2013.
9. ArvindSathi, "BigDataAnalytics: Disruptive Technologies for Changing the Game",MC Press,2012
10. Paul Zikopoulos ,Dirk DeRoos , Krishnan Parasuraman , Thomas Deutsch , James Giles, David Corigan , "Harness the Power of Big Data The IBM Big Data Platform ",Tata McGraw Hill Publications,2012.

SEMESTER - VI

Course Code	Course Name	L	T	P	C
20IT6T01	DATA MINING	3	0	0	3

Course Objectives:

- To understand data warehouse concepts, architecture, business analysis and tools
- To understand data pre-processing and data visualization techniques
- To study algorithms for finding hidden and interesting patterns in data
- To understand and apply various classification and clustering techniques using tools

Course Outcomes:

At the end of the course, the students will be able to:

CO 1. : Design a Data warehouse system and perform business analysis with OLAP tools

CO 2. : Apply suitable pre-processing and visualization techniques for data analysis

CO 3. : Apply frequent pattern and association rule mining techniques for data analysis

CO 4. : Apply appropriate classification techniques for data analysis

CO 5. : Apply appropriate clustering techniques for data analysis

UNIT I

Data Warehousing : Basic Concepts, Data Warehousing Components, Data Warehouse Architecture , Multidimensional Data Model, Data Warehouse Schemas, Concept Hierarchies, Characteristics of OLAP Systems, Typical OLAP Operations, OLAP vs OLTP.

UNIT II

Introduction to Data Mining Systems, Knowledge Discovery Process, Data Mining Techniques, Issues, applications, Data Objects and attribute types, Statistical description of data, Data Preprocessing Techniques, Data Visualization, Data similarity and dissimilarity measures.

UNIT III

Association Analysis: Basic Concepts and Algorithms, Frequent Item set Generation, The Apriori Principle, Frequent Item set Generation in the Apriori, Algorithm, Candidate Generation and Pruning, Support Counting, Rule Generation, Confidence-Based Pruning, Rule Generation in Apriori, Algorithm, Compact Representation of Frequent Item sets, Maximal Frequent Item sets, Closed Frequent Item sets, FP-Growth Algorithm, FP-tee Representation, Frequent Item set Generation in FP-Growth Algorithm

UNIT IV

Classification: Basic Concepts, Decision Trees, and Model Evaluation , General Approach to Solving a Classification Problem, Decision Tree Induction, How a Decision Tree Works, How to Build a Decision Tree, Methods for Expressing Attribute Test Conditions, Measures for Selecting the Best Split, Algorithm for Decision Tree Induction, Model Overfitting, Overfitting Due to Presence of Noise, Overfitting Due to Lack of Representative Samples, Evaluating the Performance of a Classifier, Methods for Comparing Classifiers

UNIT V

Cluster Analysis: Basic Concepts and Algorithms: Overview: What Is Cluster Analysis? Different Types of Clustering, Different Types of Clusters; K-means: The Basic K-means Algorithm, K-means

Additional Issues, Bisecting K-means, Strengths and Weaknesses; Agglomerative Hierarchical Clustering: Basic Agglomerative Hierarchical Clustering Algorithm DBSCAN: Traditional Density Center-Based Approach, DBSCAN Algorithm, Strengths and Weaknesses. (Tan &Vipin).

Text Books:

1) Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Third Edition, Elsevier, 2012.
2) Pang-Ning Tan, Michael Steinbach and Vipin Kumar, Introduction to Data Mining, Pearson, 2016.

Reference Books:

1) Alex Berson and Stephen J.Smith, —Data Warehousing, Data Mining & OLAP||, Tata McGraw – Hill Edition, 35th Reprint 2016.
2) K.P. Soman, Shyam Diwakar and V. Ajay, —Insight into Data Mining Theory and Practice||, Eastern Economy Edition, Prentice Hall of India, 2006.
3) Ian H.Witten and Eibe Frank, —Data Mining: Practical Machine Learning Tools and Techniques||, Elsevier, Second Edition.

e-Resources:

1) https://www.saedsayad.com/data_mining_map.htm
2) https://nptel.ac.in/courses/106/105/106105174/
3) (NPTEL course by Prof.Pabitra Mitra) http://onlinecourses.nptel.ac.in/noc17_mg24/preview
4) (NPTEL course by Dr. Nandan Sudarshanam & Dr. Balaraman Ravindran) http://www.saedsayad.com/data_mining_map.htm

KKR & KSR INSTITUTE OF TECHNOLOGY AND SCIENCE(AUTONOMOUS)
Department of Computer Science and Engineering

Course Code	Course Name	L	T	P	C
20CS6E01	PROFESSIONAL ELECTIVE- II 1. CLOUD COMPUTING	3	0	0	3

Course Objectives:

- The student will learn about the cloud environment, building software systems and components that scale to millions of users in modern internet cloud concepts capabilities across the various cloud service models including Iaas, Paas, Saas, and developing cloud based software applications on top of cloud platforms.

Course Outcomes:

At the end of the course, students will be able to:

CO-1: Understanding the key dimensions of the challenge of Cloud Computing

CO-2: Assessment of the economics, financial, and technological implications for selecting cloud computing for own organization

CO-3: Assessing the financial, technological, and organizational capacity of employer's for actively initiating and installing cloud-based applications.

CO-4: Assessment of own organizations' needs for capacity building and training in cloud computing-related IT areas.

CO-5: Outline the cloud resource management and scheduling policies.

UNIT -I:

Systems modeling, Clustering and virtualization Scalable Computing over the Internet, Technologies for Network based systems, System models for Distributed and Cloud Computing, Software environments for distributed systems and clouds, Performance, Security And Energy Efficiency

UNIT- II:

Virtual Machines and Virtualization of Clusters and Data Centers Implementation Levels of Virtualization, Virtualization Structures/ Tools and mechanisms, Virtualization of CPU, Memory and I/O Devices, Virtual Clusters and Resource Management, Virtualization for Data Center Automation.

UNIT- III:

Cloud Platform Architecture Cloud Computing and service Models, Architectural Design of Compute and Storage Clouds, Public Cloud Platforms, Inter Cloud Resource Management, Cloud Security and Trust Management. Service Oriented Architecture, Message Oriented Middleware.

UNIT -IV:

Cloud Programming and Software Environments Features of Cloud and Grid Platforms, Parallel & Distributed Programming Paradigms, Programming Support of Google App Engine, Programming on Amazon AWS and Microsoft Azure, Emerging Cloud Software Environments.

UNIT- V:

Cloud Resource Management and Scheduling Policies and Mechanisms for Resource Management Applications of Control Theory to Task Scheduling on a Cloud, Stability of a Two Level Resource Allocation Architecture, Feedback Control Based on Dynamic Thresholds. Coordination of Specialized Autonomic Performance Managers, Resource Bundling, Scheduling Algorithms for Computing Clouds, Fair Queuing, Start Time Fair Queuing, Borrowed Virtual Time, Cloud Scheduling Subject to Deadlines, Scheduling MapReduce Applications Subject to Deadlines.

TEXT BOOKS:

1) Distributed and Cloud Computing, Kai Hwang, Geoffry C. Fox, Jack J. Dongarra MK Elsevier.
2) Cloud Computing, Theory and Practice, Dan C Marinescu, MK Elsevier.
3) Cloud Computing, A Hands on approach, ArshadeepBahga, Vijay Madisetti, University Press

REFERNCE BOOKS:

1) Cloud Computing, A Practical Approach, Anthony T Velte, Toby J Velte, Robert Elsenpeter, TMH
2) Mastering Cloud Computing, Foundations and Application Programming, Raj Kumar Buyya, Christen vecctiola, S Tammaraiselvi, TMH

E-Resources:

1) https://archive.nptel.ac.in/courses/106/105/106105167/

KKR & KSRINSTITUTE OF TECHNOLOGY AND SCIENCES

(Autonomous)

Vinjanampadu, Vatticherukuru Mandal, Guntur, Andhra Pradesh 522017

Course Code	Course Name	L	T	P	C
20ECXO04	**Design of IoT Systems**	3	0	0	3

ELECTRONICS AND COMMUNICATION ENGINEERING
COURSE STRUCTURE

COURSE OUTCOMES:

After successful completion of this course, students should be able to

CO1 :Explain Internet of Things and its hardware.
CO2 :Explain Components of Internet of Things.
CO3 :Interface I/O Devices, Sensors, and Communication modules.
CO4 :Monitor data and control devices.
CO5 :Implement real time IoT based applications.

SYLLABUS

UNIT-I : Introduction to IoT

Introduction to IoT, Architectural Overview, Design principles and needed capabilities, Basics of Networking, M2M and IoT Technology Fundamentals- Devices and gate ways, Data management, Everything as a Service (XaaS), Role of Cloud in IoT, Security aspects in IoT.

UNIT-II : Elements of IoT

Hardware Computing Components- Arduino, Raspberry Pi, ARM Cortex-A class processor, Embedded Devices – ARM Cortex-M class processor, Arm Cortex-M0 Processor Architecture, Block Diagram, Cortex-M0 Processor Instruction Set, ARM and Thumb Instruction Set.

UNIT-III : IoT Application Development

Communication, IoT Applications, Sensing, Actuation, I/O interfaces. Software Components- Programming API's (using Python/Node.js/Arduino) for Communication Protocols-MQTT, ZigBee, CoAP, UDP, TCP, Bluetooth. Bluetooth Smart Connectivity, Bluetooth overview, Bluetooth Key Versions, Bluetooth Low Energy (BLE) Protocol, Bluetooth, Low Energy Architecture, PSoC4 BLE architecture and

Component Overview.

UNIT-IV : Solution framework for IoT applications

Implementation of Device integration, Data acquisition and integration, Device data storage- Unstructured data storage on cloud/local server, Authentication, authorization of devices.

UNIT-V : IoT Case Studies

IoT case studies and mini projects based on Industrial automation, Transportation, Agriculture, Healthcare, Home Automation.

TEXTBOOKS:

1. Raj Kamal, "Internet of Things: Architecture and Design Principles", 1st Edition, McGraw Hill Education,2017.
2. The Definitive Guide to the ARM Cortex-M0 by JosephYiu,2011
3. Vijay Madisetti, Arshdeep Bahga, Internet of Things, "A Hands-on Approach", UniversityPress,2015

REFERENCEBOOKS:

1. Cypress Semiconductor/PSoC4 BLE (Bluetooth Low Energy) Product TrainingModules.
2. Pethuru Raj and Anupama C. Raman, "The Internet of Things: Enabling Technologies,Platforms, and Use Cases", CRC Press,2017

SEMESTER –VI

Course Code	Course Name	L	T	P	C
20SH5M01	**Mandatory Course** **Professional Ethics and Human Values**	**2**	**0**	**0**	**0**

Course Objectives:
- To create an awareness on Engineering Ethics and Human Values.
- To instill Moral and Social Values and Loyalty
- To appreciate the rights of others
- To create awareness on assessment of safety and risk

Course outcomes:
Students will be able to:
CO-1: Identify and analyze an ethical issue in the subject matter under investigation or in a relevant field
CO-2: Identify the multiple ethical interests at stake in a real-world situation or practice and articulate what makes a particular course of action ethically defensible
CO-3: Assess their own ethical values and the social context of problems and Identify ethical concerns in research and intellectual contexts, including academic integrity, use and citation of sources, the objective presentation of data, and the treatment of human subjects
CO-4: Demonstrate knowledge of ethical values in non-classroom activities, such as service learning, internships, and field work
CO-5: Integrate, synthesize, and apply knowledge of ethical dilemmas and resolutions in academic settings, including focused and interdisciplinary research.

UNIT I:
Human Values & Principles for Harmony: Morals, Values and Ethics – Integrity – Work Ethics–Service Learning–Civic Virtue –Respect for others–Living Peacefully–Caring –Sharing–Honesty–Courage–Value Time –Co-operation–Commitment–Self-confidence–Spirituality-Character. Truthfulness–Customs and Traditions-Value Education –Human Rights – Fundamental Duties - Aspirations and Harmony (I, We & Nature) – Gender Bias - Emotional Intelligence.

UNITII:
Engineering Ethics and Social Experimentation: History of Ethics - Need of Engineering Ethics - Senses of Engineering Ethics- Profession and Professionalism ––Self Interest - Moral Autonomy – Utilitarianism – Virtue Theory -Uses of Ethical Theories - Deontology- Types of Inquiry –Kohlberg's Theory - Gilligan's Argument –Heinz's Dilemma - Comparison with Standard Experiments –– Learning from the Past –Engineers as Managers – Consultants and Leaders – Balanced Outlook on Law -Role of Codes–Codes and Experimental Nature of Engineering.

UNITIII:
Engineers' Responsibilities towards Safety and Risk: Concept of Safety - Safety and Risk –

Types of Risks – Voluntary v/s Involuntary Risk –Consequences-RiskAssessment–Accountability–Liability-ReversibleEffects-ThresholdLevelsof Risk-Delayedv/s Immediate Risk-Safety andtheEngineer–DesigningforSafety–Risk-BenefitAnalysis-Accidents.

UNIT IV:

Engineers'DutiesandRights: ConceptofDuty-ProfessionalDuties–Collegiality-TechniquesforAchievingCollegiality – Professional and Individual Rights –Confidential and Proprietary Information - Conflict of Interest-Ethical egoism - Collective Bargaining – Confidentiality - Gifts and Bribes - Problem solving-Occupational Crimes –Industrial Espionage-Price Fixing-Whistle Blowing.

UNITV:

GlobalIssues: GlobalizationandMNCs–CrossCultureIssues-BusinessEthics–MediaEthics-Environmental Ethics – Endangering Lives - Bio Ethics - Computer Ethics - War Ethics – Research Ethics-Intellectual Property Rights.

References:

1. Professional Ethics, R .Subramaniam–Oxford Publications, New Delhi.
2. Ethics in Engineering Mike W. Martin and Roland Schinzinger - Tata McGraw-Hill–2003.
3. Professional Ethics and Morals, A. R. Aryasri, Dharanikota Suyodhana – Maruthi Publications.
4. Engineering Ethics, Harris , Pritchardand Rabins, Cengage Learning, New Delhi.
5. Human Values & Professional Ethics, S.B.Gogate, Vikas Publishing House Pvt. Ltd., Noida.
6. Engineering Ethics & Human Values, M. Govindarajan,S. NatarajanandV.S.SenthilKumar-PHI Learning Pvt. Ltd–2009.
7. Professional Ethics and Human Values, A. Alavudeen, R. Kalil Rahman and M.Jayakumaran–niversity Science Press.
8. Professional Ethics and Human Values, D. R.Kiran-TataMcGraw-Hill-2013
9. Human Values andProfessionalEthics,JayshreeSureshandB.S.Raghavan,S.ChandPublications

SEMESTER -VI

Course Code	Course Name	L	T	P	C
20CS6L01	**CRYPTOGRAPHY AND NETWORK SECURITY LAB**	**0**	**0**	**3**	**1.5**

Course Objectives:

- The concepts of classical encryption techniques and concepts of finite fields and number theory.
- Working principles and utilities of various cryptographic algorithms including secret key cryptography, hashes and message digests, and public key algorithms.
- Design issues and working principles of various authentication protocols, PKI standards.

Course Outcomes:

- Apply the knowledge of cryptographic checksums and evaluate the performance of different message digest algorithms for verifying the integrity of varying message sizes.
- Apply different digital signature algorithms to achieve authentication and create secure applications

List of Experiments:

1) Write a C program that contains a string (char pointer) with a value 'Hello world'. The program should XOR each character in this string with 0 and displays the result.
2) Write a C program that contains a string (char pointer) with a value 'Hello world'. The program should AND or and XOR each character in this string with 127 and display the result.
3) Write a Java program to perform encryption and decryption using the following algorithms:
   a. Ceaser cipher b. Substitution cipher
4) Write a C/JAVA program to implement the DES algorithm logic.
5) Write a C/JAVA program to implement the Blowfish algorithm logic.
6) Write the RC4 logic in Java Using Java cryptography; encrypt the text "Hello world" using Blowfish. Create your own key using Java key tool.
7) Write a Java program to implement RSA algorithm.
8) Implement the Diffie-Hellman Key Exchange mechanism using HTML and JavaScript.
9) Calculate the message digest of a text using the SHA-1 algorithm in JAVA.
10) Calculate the message digest of a text using the MD5 algorithm in JAVA.

E-Resources:

1) https://www.geeksforgeeks.org/caesar-cipher-in-cryptography/
2) https://www.javatpoint.com/hill-cipher-program-in-java

DEPARTMENT OF CSE - DATA SCIENCE
SEMESTER - VI

Course Code	Course Name	L	T	P	C
20CD6L02	BIG DATA & HADOOP LAB	0	0	3	1.5

Course Objectives:

1. Provide the knowledge to setup a Hadoop Cluster.
2. Impart knowledge to develop programs using MapReduce.
3. Discuss Pig, PigLatin and HiveQL to process bigdata.
4. Present latest big data frameworks and applications using Spark
5. Integrate Hadoop with R (RHadoop) to process and visualize.

Course Outcomes:

1. Understand Hadoop working environment.
2. Apply Map Reduce programs for real world problems.
3. Implement scripts using Pig to solve real world problems.
4. Analyze queries using Hive to analyze the datasets
5. Understand spark working environment and integration with R

TASK 1: a) Understanding and using basic HDFS commands
b) Run a basic word count Map Reduce program to understand Map Reduce Paradigm.
TASK 2: Write a Map Reduce program that mines weather data
TASK 3: Implement matrix multiplication with Hadoop Map Reduce.
TASK 4: Working with files in Hadoop file system: Reading, Writing and Copying
TASK-5: Write Pig Latin scripts sort, group, join, project, and filter your data.
TASK 6: Run the Pig Latin Scripts to find Word Count and max. temp for each and every year.
TASK-7: Writing User Defined Functions/Eval functions for filtering unwanted data in Pig
TASK-8: Working with Hive QL, Use Hive to create, alter, and drop databases, tables, views, functions, and indexes
TASK 9: Writing User Defined Functions in Hive
TASK 10: Understanding the processing of large dataset on Spark framework.
TASK 11: Ingesting structured and unstructured data using Sqoop, Flume

Text Books:

1. Tom White, "Hadoop: The Definitive Guide", 4th Edition, O'Reilly Inc,2015.
2. Tanmay Deshpande, "Hadoop Real-World Solutions Cookbook", 2ndEdition, Packt Publishing, 2016

Reference Books:

1. Edward Capriolo, Dean Wampler, and Jason Rutherglen, "Programming Hive", O'Reilly Inc,2012.
2. Vignesh Prajapati, "Big data Analytics with R and Hadoop", Packt Publishing,2013

SEMESTER -VI

Course Code	Course Name	L	T	P	C
20IT6L01	**DATA MINING LAB**	**0**	**0**	**3**	**1.5**

Course Objectives:

- To understand the mathematical basics quickly and covers each and every condition of data mining in order to prepare for real-world problems
- The various classes of algorithms will be covered to give a foundation to further apply knowledge to dive deeper into the different flavours of algorithms
- Students should aware of packages and libraries of R and also familiar with functions used in R for visualization
- To enable students to use R to conduct analytics on large real life datasets
- To familiarize students with how various statistics like mean median etc

Course Outcomes:

At the end of the course, student will be able to

CO 1. : Extend the functionality of R by using add-on packages

CO 2. : Examine data from files and other sources and perform various data manipulation tasks

CO 3. : Code statistical functions in R

CO 4. : Use R Graphics and Tables to visualize results of various statistical operations on data

CO 5. : Apply the knowledge of R gained to data Analytics for real life applications

List of Experiments:

1. Implement all basic R commands.
2. Interact data through .csv files (Import from and export to .csv files).
3. Get and Clean data using swirl exercises. (Use 'swirl' package, library and install that topic from swirl).
4. Visualize all Statistical measures (Mean, Mode, Median, Range, Inter Quartile Range etc., using Histograms, Boxplots and Scatter Plots).
5. Create a data frame with the following structure

.

EMP ID	EMP NAME	SALARY	START DATE
1	Satish	5000	01-11-2013
2	Vani	7500	05-06-2011
3	Ramesh	10000	21-09-1999
4	Praveen	9500	13-09-2005
5	Pallavi	4500	23-10-2000

a) Extract two column names using column name.

b) Extract the first two rows and then all columns.

c) Extract 3rd and 5th row with 2nd and 4th column.

6. Write R Program using 'apply' group of functions to create and apply normalization function on each of the numeric variables/columns of iris dataset to transform them into

i) 0 to 1 range with min-max normalization.

ii) a value around 0 with z-score normalization.

Create a data frame with 10 observations and 3 variables and add new rows and columns to it using 'rbind' and 'cbind' function.

8. Write R program to implement linear and multiple regression on 'mtcars' dataset to estimate the value of 'mpg' variable, with best R^2 and plot the original values in 'green' and predicted values in 'red'.
9. Implement k-means clustering using R.
10. Implement k-medoids clustering using R.
11. implement density based clustering on iris dataset.
12. implement decision trees using 'readingSkills' dataset.
13. Implement decision trees using 'iris' dataset using package party and 'rpart'.
14. Use a Corpus() function to create a data corpus then Build a term Matrix and Reveal word frequencies.

Text Books:

1) R and Data Mining: Examples and Case Studies, 1st ed, Yanchang Zhao, Sprnger, 2012.
2) R for Everyone, Advanced Analytics and Graphics, 2nd ed, Jared Lander, Pearson, 2018.

e-Resources:

1) www.r-tutor.com