M.Tech. (Full Time) -Big Data Analytics
Curriculum & Syllabus
2015– 2016

FACULTY OF ENGINEERING AND TECHNOLOGY
SRM UNIVERSITY
SRM NAGAR, KATTANKULATHUR – 603 203

COURSE CODE	COURSE NAME	L	T	P	C
CORE COURSES - SEMESTER I & II					
DA2001	Foundations of Data Science	3	0	2	4
DA2002	Big Data Technology	2	0	3	4
DA_R2003	Computing for Data Analytics	3	0	2	4
DA2004	Programming for Data Analytics	2	0	3	4
DA_R2005	Marketing Analytics	3	0	2	4
DA_R2006	Algorithms for Advanced Analytics	3	0	2	4
CAC2001	Career Advancement Course For Engineers - I	1	0	1	1
CAC2002	Career Advancement Course For Engineers - II	1	0	1	1
TOTAL		18	0	16	26
CORE COURSES - SEMESTER III					
DA2047	Seminar	0	0	1	1
DA2049	Project work Phase I	0	0	12	6
CAC2003	Career Advancement Course For Engineers - II	1	0	1	1
TOTAL		1	0	14	8
CORE COURSE - SEMESTER IV					
DA2050	Project work Phase II	0	0	32	16
SUPPORTING COURSE					
DA2011	Multivariate Techniques for Data Analysis	3	0	0	3
PROGRAM ELECTIVES					
DA2101	Python Programming for Data Analytics	2	0	2	3
DA2102	Decision Management Systems	2	2	0	3
DA2103	Big Data Security	2	2	0	3
DA2104	Risk Analytics	3	0	0	3
IT2110	Information Storage Management	3	0	0	3
IT2111	Cloud Computing	2	0	2	3
DA2105	Cluster Computing	2	2	0	3
DA2106	Functional Programming	2	0	2	3
DA2107	Web Intelligence and Big Data Analytics	2	0	2	3
DA2108	Social Network Analytics	2	0	2	3
DA2109	Streaming Analytics	2	0	2	3
DA2110	Advanced Algorithms	2	0	2	3
DA2111	Natural Language Processing	2	2	0	3
CS2108	Pattern Recognition	3	0	0	3
DA2112	Deep Learning	2	0	2	3

Total number of credits to be earned for M.Tech Course = 74

NOTE:
Students have to register for the courses as per the following guidelines:

Sl. No.	Category	Credits				
		I Semester	**II Semester**	**III Semester**	**IV Semester**	**Category total**
1	Core courses	12 (3 courses)	12 (3 courses)	---	---	24
2	Program Elective courses	18 (in I to III semesters)			---	18
	Interdisciplinary elective courses (any one program elective from other programs)	3 (in I to III semesters)				3
3	Supportive courses - mandatory	3 (in I to III semesters)			---	3
4	Seminar	---	---	1	---	1
5	Career Advancement Courses	1	1	1	---	3
6	Project work	---	---	06	16	22
	Total					74

Course code	Course Title	L	T	P	C
DA2001	FOUNDATIONS OF DATA SCIENCE	3	0	2	4
	Total contact hours - 75				
	Prerequisite				
	Nil				

PURPOSE

Modern scientific, engineering, and business applications are increasingly dependent on data, existing traditional data analysis technologies were not designed for the complexity of the modern world. Data Science has emerged as a new, exciting, and fast-paced discipline that explores novel statistical, algorithmic, and implementation challenges that emerge in processing, storing, and extracting knowledge from Big Data.

INSTRUCTIONAL OBJECTIVES

1.	Able to apply fundamental algorithmic ideas to process data.
2.	Learn to apply hypotheses and data into actionable predictions.
3.	Document and transfer the results and effectively communicate the findings using visualization techniques.

UNIT I – INTRODUCTION TO DATA SCIENCE (9 hours)

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.

UNIT II – MODELING METHODS (9 hours)

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.

UNIT III – INTRODUCTION TO R (9 hours)

Reading and getting data into R – ordered and unordered factors – arrays and matrices – lists and data frames – reading data from files – probability distributions – statistical models in R - manipulating objects – data distribution.

UNIT IV – MAP REDUCE (9 hours)

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.

UNIT V- DELIVERING RESULTS (9 hours)

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.

Practical: (30 hours)

REFERENCES

1. Nina Zumel, John Mount, “Practical Data Science with R”, Manning Publications, 2014.
2. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, “Mining of Massive Datasets”, Cambridge University Press, 2014.
3. Mark Gardener, “Beginning R - The Statistical Pr ogramming Language”, John Wiley & Sons, Inc., 2012.
4. W. N. Venables, D. M. Smith and the R Core Team, “An Introduction to R”, 2013.
5. Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, Abhijit Dasgupta, “Practical Data Science Cookbook”, Packt Publishing Ltd., 2014.
6. Nathan Yau, “Visualize This: The FlowingData Guide to Design, Visualization, and Statistics”, Wiley, 2011.
7. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, “Professional Hadoop Solutions”, Wiley, ISBN: 9788126551071, 2015.
8. http://www.johndcook.com/R_language_for_programmers.html
9. http://bigdatauniversity.com/
10. http://home.ubalt.edu/ntsbarsh/stat-data/topics.htm#rintroduction

Course code	Course Title	L	T	P	C
DA2002	BIG DATA TECHNOLOGY	2	0	3	4
	Total contact hours – 75				
	Prerequisite				
	Nil				

PURPOSE

This course provides practical foundation level training that enables immediate and effective participation in big data projects. The course provides grounding in basic and advanced methods to big data technology and tools, including MapReduce and Hadoop and its ecosystem.

INSTRUCTIONAL OBJECTIVES

1.	Learn tips and tricks for Big Data use cases and solutions.
2.	Learn to build and maintain reliable, scalable, distributed systems with Apache Hadoop.
3.	Able to apply Hadoop ecosystem components.

UNIT I – INTRODUCTION TO BIG DATA (6 hours)

Introduction – distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce.

UNIT II – INTRODUCTION HADOOP (6 hours)

Big Data – Apache Hadoop & Hadoop EcoSystem – Moving Data in and out of Hadoop – Understanding inputs and outputs of MapReduce - Data Serialization.

UNIT- III HADOOP ARCHITECTURE (6 hours)

Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands , Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers - Cluster Setup – SSH & Hadoop Configuration – HDFS Administering –Monitoring & Maintenance.

UNIT-IV HADOOP ECOSYSTEM AND YARN (6 hours)

Hadoop ecosystem components - Schedulers - Fair and Capacity, Hadoop 2.0 New Features- NameNode High Availability, HDFS Federation, MRv2, YARN, Running MRv1 in YARN.

UNIT-V HIVE AND HIVEQL, HBASE (6 hours)

Hive Architecture and Installation, Comparison with Traditional Database, HiveQL - Querying Data - Sorting And Aggregating, Map Reduce Scripts, Joins & Subqueries, HBase concepts- Advanced Usage, Schema Design, Advance Indexing - PIG, Zookeeper - how it helps in monitoring a cluster, HBase uses Zookeeper and how to Build Applications with Zookeeper.

Practical: (45 hours)

REFERENCES

1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, “Professional Hadoop Solutions”, Wiley, ISBN: 9788126551071, 2015.
2. Chris Eaton, Dirk deroos et al. , “Understanding Big data ”, McGraw Hill, 2012.
3. Tom White, “HADOOP: The definitive Guide” , O Reilly 2012.

4. Vignesh Prajapati, “Big Data Analytics with R and Haoop”, Packet Publishing 2013.
5. Tom Plunkett, Brian Macdonald et al, “Oracle Big Data Handbook”, Oracle Press, 2014.
6. http://www.bigdatauniversity.com/
7. Jy Liebowitz, “Big Data and Business analytics”,CRC press, 2013.

Course Code	Course Title	L	T	P	C
DA_R2003	COMPUTING FOR DATA ANALYTICS	3	0	2	4
	Total contact hours – 75				
	Prerequisite				
	Nil				

PURPOSE	
This course teaches fundamental concepts and tools needed to understand the emerging role of business analytics in Organizations.	
INSTRUCTIONAL OBJECTIVES	
1.	Learn the Big Data in Technology Perspective.
2.	Understanding of the statistical procedures most often used by practicing engineers.
3.	Understand Forecasting methods and apply for business applications.

UNIT – I DATA ANALYTICS LIFE CYCLE (9 hours)

Introduction to Big data Business Analytics - State of the practice in analytics role of data scientists - Key roles for successful analytic project - Main phases of life cycle - Developing core deliverables for stakeholders.

UNIT – II STATISTICS (9 hours)

Sampling Techniques - Data classification, Tabulation, Frequency and Graphic representation - Measures of central value - Arithmetic mean, Geometric mean, Harmonic mean, Mode, Median, Quartiles, Deciles, Percentile - Measures of variation – Range, IQR, Quartile deviation, Mean deviation, standard deviation, coefficient variance, skewness, Moments & Kurtosis.

UNIT – III PROBABILITY AND HYPOTHESIS TESTING (9 hours)

Random variable, distributions, joint probability function, marginal density function. Random vectors - Some special probability distribution - Binomial, Poison, Geometric, uniform, exponential, normal, gamma and Erlang - Normal distribution.

UNIT – IV PREDICTIVE ANALYTICS (9 hours)

Sampling distribution – Estimation - point, confidence - Test of significance, 1& 2 tailed test, uses of t-distribution, F-distribution, $\chi 2$ distribution - Predictive modeling and Analysis - Regression Analysis, Correlation analysis, Rank correlation coefficient, Multiple correlation.

UNIT – V TIME SERIES FORECASTING AND DESIGN OF EXPERIMENTS (9 hours)

Forecasting Models for Time series : MA, SES, TS with trend, season - Design of Experiments, one way classification, two way classification, ANOVA, Latin square, Factorial Design.

Practical: (30 hours)

REFERENCES:

1. Chris Eaton, Dirk Deroos, Tom Deutsch et al., ―Understanding Big Data‖, Mc Graw Hill, 2012.
2. Alberto Cordoba, ―Understanding the Predictive Analytics Lifecycle‖, Wiley, 2014.
3. Eric Siegel, Thomas H. Davenport, ―Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die‖, Wiley, 2013.
4. James R Evans, ―Business Analytics – Methods, Models and Decisions‖, Pearson 2013.
5. R. N. Prasad, Seema Acharya, ―Fundamentals of Business Analytics‖, Wiley, 2015.

6. S M Ross, ―Introduction to Probability and Statistics for Engineers and Scientists‖, Academic Foundation, 2011.
7. David Hand, Heiki Mannila, Padhria Smyth, ―Principles of Data Mining‖, PHI 2013.
8. Spyros Makridakis, Steven C Wheelwright, Rob J Hyndman, ―Forecasting methods and applications‖, Wiley 2013(Reprint).
9. David Hand, HeikkiMannila, Padhraic Smyth, ―Principles of Data mining‖, PHI 2013.
10. http://cran.r-project.org/doc/manuals/R-intro.html 11. W.N. Venables, D.M Smith, ―An introduction to R‖, 12. R in Nutshell , O Reilly
11. R S N Pillai & Bagavathi, Practical Statistics, S Chand & Company Pvt Ltd, publishing date 2010-12-01

Course code	Course Title	L	T	P	C
DA2004	**PROGRAMMING FOR DATA ANALYTICS**	**2**	**0**	**3**	**4**
	Total contact hours – 75				
	Prerequisite				
	Programming in Java				

PURPOSE

The rate in which data is exponentially growing has led to the evolvement of many technologies to better utilize this data for timely and accurate decision making. Such data with huge variety, volume and velocity is coined as big data.The big data platform such as Hadoop is programmed in Java. This course aims at discussing the technical concepts which are the basic building blocks for most of the big data platforms.

INSTRUCTIONAL OBJECTIVES

1.	Understanding basic network and distributed programming.
2.	Constructing a real world application with data storage and retrieval
3.	Leveraging the benefits of reusable components
4.	Analyzing basic file modes and operations
5.	Applying MapReduce paradigm to solve problems

UNIT I – NETWORK PROGRAMMING & DISTRIBUTED OBJECTS (6 hours)

Connecting to a Server - Implementing Servers and Clients- Advanced Socket Programming – InetAddress - URL Connections – RMI Programming.

UNIT II – CONNECTING TO DATABASE (6 hours)

The Design of JDBC - Basic Concepts - Executing Queries – Prepared Statemets - Result Sets – Metadata -Transactions.

UNIT III – JAVABEANS (6 hours)

The Bean - Writing Process - Using Beans to Build an Application - Bean Property Types – Property Editors - Customizers.

UNIT IV – STREAMS AND FILES (6 hours)

Streams – Text Input and Output – Reading and Writing Binary Data – Zip Archives – Object Streams and Serialization – Memory Mapped Files.

UNIT V – PROGRAMMING MAP REDUCE (6 hours)

MapReduce program in Java – Map Reduce API – Progamming Examples- Combiner Functions - Distributed MapReduce Job.

Practical : **(45 hours)**

REFERENCES:

1. White, "Hadoop: The Definitive Guide", Third Edition - 2012 – O'Reilly – ISBN: 9789350237564.
2. Cay S. Horstmann, Gary Cornell, "Core Java™ 2: Volume II–Advanced Features", Prentice Hall, 9th edition, ISBN: 978-0137081608.
3. Jean Dollimore, Tim Kindberg, George Coulouris, "Distributed Systems Concepts and Design", 4th Edition, Jun 2005, Hardback, 944 pages, ISBN: 9780321263544.
4. Y. Daniel Liang, Introduction to Java Programming, Tenth Edition, Pearson, 2015.

Course code	Course Title	L	T	P	C
DA_R2005	MARKETING ANALYTICS	3	0	2	4
	Total contact hours – 75				
	Prerequisite				
	Nil				

PURPOSE

The objective of this course is to provide thorough knowledge required to address fundamental marketing decision problems. It will also train to view marketing processes and relationships systematically and analytically. The techniques discussed in this curse are useful in market segmentation, targeting, and mapping market structure and product design.

INSTRUCTIONAL OBJECTIVES

1.	Learn how to tap a simple and cost-effective tool, Microsoft Excel, to solve specific business problems using powerful analytic techniques.
2.	Helps to forecast sales and improve response rates for marketing campaigns.
3.	Explores how to optimize price points for products and services, optimize store layouts, and improve online advertising.

UNIT I – MARKETING DATA SUMMARIZATION **(9 hours)**

Slicing and Dicing Marketing Data with PivotTables - Using Excel Charts to Summarize Marketing Data - Using Excel Functions to Summarize Marketing Data.

UNIT II – FORECASTING TECHNIQUES **(9 hours)**

Simple Linear Regression and Correlation - Using Multiple Regression to Forecast Sales - Forecasting in the Presence of Special Events - Modeling Trend and Seasonality - Ratio to Moving Average Forecasting Method - Winter's Method - Using Neural Networks to Forecast Sales.

UNIT III – CUSTOMER NEEDS **(9 hours)**

Conjoint Analysis - Logistic Regression - Discrete Choice Analysis – Customer Value - Introduction to Customer value, Benefits.

UNIT IV – MARKET SEGMENTATION **(9 hours)**

Cluster Analysis - User-Based Collaborative Filtering - Collaborative Filtering - Using Classification Trees for Segmentation.

UNIT V – RETAILING AND MARKET RESEARCH TOOLS **(9 hours)**

Retailing - Introduction to retailing, Market Basket Analysis and Lift - Marketing Research Tools - Principal Components Analysis.

Tutorial: (30 hours)

REFERENCES

1. Wayne.L.Winston, "Marketing Analytics: Data driven techniques with MS-Excel", Wiley, 1st ed. 2014.
2. Stephan Sorger, "Marketing Analytics: Strategic models and metrics", CreateSpace Independent Publishing Platform, 1st ed., 2013.

Course Code	Course Title	L	T	P	C
DA_R2006	ALGORITHMS FOR ADVANCED ANALYTICS	3	0	2	4
	Total contact hours – 75				
	Prerequisite				
	Knowledge in basic analytical algorithms.				

PURPOSE

This course gives a comprehensive coverage of algorithms specially meant for analyzing data at an in-depth level. Decision trees, Support Vector machines and Neural networks are considered to be highly effective in analyzing complex data.

INSTRUCTIONAL OBJECTIVES

1.	Learn concepts and techniques and how to find useful knowledge.
2.	Understanding of the topics that can create an ideal analytic environment that is better suited to the challenges of today's analytics demands.
3.	Harness the power of high performance computing architectures and data mining, text analytics, and machine learning algorithms.

UNIT I – CLASSIFICATION ALGORITHMS (9 hours)

Issues regarding classification and prediction, Bayesian Classification, Classification by back propagation, Classification based on concepts from association rule mining, Other Classification Methods, Classification accuracy.

UNIT II – DECISION TREES (8 hours)

Introduction to Decision trees –Splitting approaches in decision tree - Classification by decision tree induction – Tree pruning methods (Cost complexity pruning - Chi square pruning) – Issues in decision trees –Extended Decision Trees (fuzzy decision trees)

UNIT III – CLUSTERING ALGORITHMS IN ANALYTICS (8 hours)

Distance Measures – Similarity Functions – Error measures – K-Means algorithm – Fuzzy C-Means algorithm– Hierarchical clustering algorithm– Mixture of Gaussian algorithm– Expectation & Maximization approach – Probabilistic clustering algorithms

UNIT IV– TEXT ANALYTICS (11 hours)

Introduction - Text mining operations -Border increment text mining algorithm - Preprocessing techniques – Feature selection using dimensionality reduction - Singular Value Decomposition – Hidden Markov Probabilistic Model for Information Extraction

UNIT V – NEURAL NETWORKS (9 hours)

Soft Computing Tools – Approximation of Multivariate functions - Support vector machines - VC Dimension – Structural Risk Minimization – linear maximal marginal classifier for SVM – Perceptron – Convergence Theorem - Activation functions – Adaptive Linear Neuron (Adaline)

Practical: **(30 hours)**

REFERENCES

1. Jiawei Han and Micheline Kamber, ―Data Mining: Concepts and Techniques‖, Morgan Kaufmann Publishers, 3rd ed, 2010.
2. Lior Rokach and Oded Maimon, ―Data Mining and Knowledge Discovery Handbook‖, Springer, 2nd edition, 2010.
3. Ronen Feldman and James Sanger, ―The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data‖, Cambridge University Press, 2006.
4. Vojislav Kecman, ―Learning and Soft Computing‖, MIT Press, 2010.
5. Jared Dean, ―Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners‖, Wiley India Private Limited, 2014.

Course code	Course Title	L	T	P	C
DA2011	**MULTIVARIATE TECHNIQUES FOR DATA ANALYSIS**	**3**	**0**	**0**	**3**
	Total contact hours - 75				
	Prerequisite				
	Nil				

PURPOSE

The purpose of this course is to introduce the students into the field of Multivariate Techniques for analyzing large volumes of data and to take decisions based on inference drawn.

INSTRUCTIONAL OBJECTIVES

1.	Data characteristics and form of Distribution of the Data Structures.
2.	Understanding the usage of multivariate techniques for the problem under the consideration.
3.	For drawing valid inferences and to plan for future investigations.

UNIT I - INTRODUCTION TO MULTIVARIATE ANALYSIS **(9 hours)**

Meaning of Multivariate Analysis, Measurements Scales - Metric measurement scales and Non-metric measurement scales, Classification of multivariate techniques (Dependence Techniques and Inter-dependence Techniques), Applications of Multivariate Techniques in different disciplines.

UNIT II – FACTOR ANALYSIS **(9 hours)**

Factor Analysis: Meanings, Objectives and Assumptions, Designing a factor analysis, Deriving factors and assessing overall factors, Interpreting the factors and validation of factor analysis.

UNIT III – CLUSTER ANALYSIS **(9 hours)**

Cluster Analysis: Objectives and Assumptions, Research design in cluster analysis, Deriving clusters and assessing overall fit (Hierarchical methods, Non Hierarchical Methods and Combinations), Interpretation of clusters and validation of profiling of the clusters.

UNIT IV - DISCRIMINANT ANALYSIS **(9 hours)**

Discriminant Analysis- concept, objective and applications. Procedure for conducting discriminant analysis. Stepwise discriminate analysis and Mahalanobis procedure. Logit model.

UNIT V – LINEAR PROGRAMMING (9 hours)

Linear Programming problem - Formulation, graphical method, simplex method. Integer Programming. Transportation and Assignment problem.

Practical: (30 hours)

REFERENCES:

1. Joseph F Hair, William C Black etal , "Multivariate Data Analysis" , Pearson Education, 7th edition, 2013.
2. T. W. Anderson , "An Introduction to Multivariate Statistical Analysis, 3rd Edition", Wiley, 2003.
3. William r Dillon, John Wiley & sons, "Multivariate Analysis methods and applications", Wiley, 1984.
4. Naresh K Malhotra, Satyabhusan Dash, "Marketing Research Anapplied Orientation", Pearson, 2011.
5. Hamdy A Taha, "Operations Research", Pearson, 2012.
6. S R Yaday, A K Malik, "Operations Research", Oxford, 2014.

Course code	Course Title	L	T	P	C
DA2101	**PYTHON PROGRAMMING FOR DATA ANALYTICS**	**2**	**0**	**2**	**3**
	Total contact hours - 60				
	Prerequisite				
	Programming knowledge in any OO language.				

PURPOSE

Data, which is available in abundance and in accessible forms, if analysed in an efficient manner unfolds many patterns and promising solutions. Data has to be pre-processed, converted to required format and fed to appropriately chosen algorithm to yield better results. This course aims at applying such techniques to raw data, using Python, to arrive at meaningful results.

INSTRUCTIONAL OBJECTIVES

1.	Understanding the basic concepts of Python
2.	Preparing and pre-processing data
3.	Understanding the data aggregation and grouping concepts
4.	Leveraging web scraping
5.	Visualizing the results of analytics effectively

UNIT I – PYTHON CONCEPTS , DATA STRUCTURES, CLASSES (6 hours)

Interpreter – Program Execution – Statements – Expressions – Flow Controls – Functions - Numeric Types – Sequences - Strings, Tuples, Lists and - Class Definition – Constructors – Inheritance – Overloading – Text & Binary Files - Reading and Writing.

UNIT II – DATA WRANGLING (4 hours)

Combining and Merging DataSets – Reshaping and Pivoting – Data Transformation – String Manipulation, Regular Expressions.

UNIT III – DATA AGGREGATION, GROUP OPERATIONS ,TIMESERIES (8 hours)

GoupBy Mechanics – Data Aggregation – Groupwise Operations and Transformations – Pivot Tables and Cross Tabulations – Date and Time Date Type tools – Time Series Basics – Data Ranges, Frequencies and Shifting.

UNIT IV – WEB SCRAPING **(6 hours)**

Data Acquisition by Scraping web applications –Submitting a form - Fetching web pages – Downloading web pages through form submission – CSS Selectors.

UNIT V – VISUALIZATION IN PYTHON **(6 hours)**

Matplotlib package – Plotting Graphs – Controlling Graph – Adding Text – More Graph Types – Getting and setting values – Patches.

Practical : **(30 hours)**

REFERENCES:

1. Mark Lutz, "Programming Python", O'Reilly Media, 4th edition, 2010.
2. Mark Lutz, "Learning Python", O'Reilly Media, 5th Edition, 2013.
3. Tim Hall and J-P Stacey, "Python 3 for Absolute Beginners", Apress, 1st edition, 2009.
4. Magnus Lie Hetland, "Beginning Python: From Novice to Professional", Apress, Second Edition, 2005.
5. Shai Vaingast, "Beginning Python Visualization Crafting Visual Transformation Scripts", Apress, 2nd edition, 2014.
6. Wes Mc Kinney, "Python for Data Analysis", O'Reilly Media, 2012.
7. White, "Hadoop: The Definitive Guide", Third Edition - O'Reilly , 2012.
8. Brandon Rhodes and John Goerzen, "Foundations of Python Network Programming: The Comprehensive Guide to Building Network Applications with Python",Apress, Second Edition, 2010.
9. http://blog.matthewrathbone.com/2013/11/17/python-map-reduce-on-hadoop---a-beginners-tutorial.html
10. http://www.michael-noll.com/tutorials/writing-an-hadoop-mapreduce-program-in-python/
11. http://allthingshadoop.com/category/python/

Course code	Course Title	L	T	P	C
DA2102	**DECISION MANAGEMENT SYSTEMS**	**2**	**2**	**0**	**3**
	Total contact hours – 60				
	Prerequisite				
	Nil				

PURPOSE

The course elaborates on the principles that guide the development of Decision Management System and lays out the framework for building them. It helps how to find suitable decisions and develop the understanding of those decisions that helps to automate them.

INSTRUCTIONAL OBJECTIVES

1.	Understanding how Decision Management Systems can transform the business.
2.	Planning the systems "with the decision in mind".
3.	Identifying, modeling, and prioritizing the decisions.
4.	Designing and implementing robust decision services.
5.	Monitoring ongoing decision-making and learning how to improve it.

UNIT I – PRINCIPLES OF DMS **(6 hours)**

Principles of Decision Management Systems - Begin with the Decision in Mind - Be Transparent and Agile - Be Predictive, Not Reactive - Test, Learn, and Continuously Improve.

UNIT II – BUILDING DECISION MANAGEMENT SYSTEMS (6 hours)

Building Decision Management Systems - Discover and Model Decisions - Characteristics of Suitable Decisions - A Decision Taxonomy - Finding Decisions - Documenting Decisions - Prioritizing Decisions.

UNIT III – DESIGN AND IMPLEMENT DECISION SERVICES (7 hours)

Design and Implement Decision Services - Build Decision Services - Integrate Decision Services - Best Practices for Decision Services Construction - Monitor and Improve Decisions - What Is Decision Analysis? - Monitor Decisions - Determine the Appropriate Response - Develop New Decision-Making Approaches - Confirm the Impact Is as Expected - Deploy the Change.

UNIT IV – ENABLERS FOR DECISION MANAGEMENT SYSTEMS (6 hours)

Enablers for Decision Management Systems - People Enablers - The Three-Legged Stool - A Decision Management Center of Excellence - Organizational Change - Process Enablers - Managing a Decision Inventory - Adapting the Software Development Lifecycle - Decision Service Integration Patterns - Moving to Fact-Based Decisioning - The OODA Loop - Technology Enablers.

UNIT V – BUSINESS RULES MANAGEMENT SYSTEMS (5 hours)

Business Rules Management Systems - Predictive Analytics Workbenches - Optimization Systems - Pre-Configured Decision Management Systems - Data Infrastructure - A Service-Oriented Platform.

Tutorial : (30 hours)

REFERENCES

1. James Taylor, "Decision Management Systems-A Practical guide to using Business rules and Predictive Analytics", IBM Press, 2012.
2. Efraim Turban , Jay E. Aronson , Ting-Peng Liang, "Decision Support Systems & Intelligent Systems", 9th edition, Prentice Hall, 2010.
3. Alberto Cordoba, "Understanding the Predictive Analytics Lifecycle", Wiley, 2014.
4. Eric Siegel, Thomas H. Davenport, "Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die", Wiley, 2013.
5. George M Marakas, "Decision support Systems", 2nd Edition, Pearson/Prentice Hall,2002
6. V.S. Janakiraman, K. Sarukesi, "Decision Support Systems",PHI, ISBN8120314441, 9788120314443, 2004.
7. Efrem G Mallach, "Decision Support systems and Data warehouse Systems", McGraw Hill, thirteenth reprint, 2008.

Course code	Course Title	L	T	P	C
DA2103	**BIG DATA SECURITY**	**2**	**2**	**0**	**3**
	Total contact hours - 60				
	Prerequisite				
	Knowledge in Big Data Technology				

PURPOSE

With the data generated from electronic devices growing exponentially, the need to analyse data on a large scale is important. Such data are of many types like financial, personal etc. Big Data environment also creates significant security challenges, when trying to make quick decisions. Data breach poses many complications. This course aims at introducing concepts related to big data security.

INSTRUCTIONAL OBJECTIVES	
1.	Understanding significance of privacy, ethics in big data environment
2.	Analyzing the steps to secure big data
3.	Building security in Hadoop environment and its ecosystem.
4.	Analyzing data security and event logging

UNIT I – BIG DATA PRIVACY, ETHICS AND SECURITY (6 hours)

Privacy – Reidentification of Anonymous People – Why Big Data Privacy is self regulating? – Ethics – Ownership – Ethical Guidelines – Big Data Security – Organizational Security.

UNIT II - SECURITY, COMPLIANCE, AUDITING, AND PROTECTION (6 hours)

Steps to secure big data – Classifying Data – Protecting – Big Data Compliance – Intellectual Property Challenge – Research Questions in Cloud Security – Open Problems.

UNIT III – HADOOP SECURITY DESIGN (6 hours)

Kerberos – Default Hadoop Model without security - Hadoop Kerberos Security Implementation & Configuration.

UNIT IV – HADOOP ECOSYSTEM SECURITY (6 hours)

Configuring Kerberos for Hadoop ecosystem components – Pig, Hive, Oozie, Flume, HBase, Sqoop.

UNIT V – DATA SECURITY & EVENT LOGGING (6 hours)

Integrating Hadoop with Enterprise Security Systems - Securing Sensitive Data in Hadoop – SIEM system – Setting up audit logging in hadoop cluster

Tutorial : (30 hours)

REFERENCES:

1. Mark Van Rijmenam, “Think Bigger: Developing a Successful Big Data Strategy for Your Business”, Amazon, 1 edition, 2014.
2. Frank Ohlhorst John Wiley & Sons, “Big Data Analytics: Turning Big Data into Big Money”, John Wiley & Sons, 2013.
3. Sherif Sakr, “Large Scale and Big Data: Processing and Management”, CRC Press, 2014.
4. Sudeesh Narayanan, “Securing Hadoop”, Packt Publishing, 2013.
5. Ben Spivey, Joey Echeverria, “Hadoop Security Protecting Your Big Data Problem”, O’Reilly Media, 2015.
6. Top Tips for Securing Big Data Environments: e-book (http://www.ibmbigdatahub.com/whitepaper/top-tips-securing-big-data-environments-e-book)
7. http://www.dataguise.com/?q=securing-hadoop-discovering-and-securing-sensitive-data-hadoop-data-stores
8. Gazzang for Hadoop http://www.cloudera.com/content/cloudera/en/solutions/enterprise-solutions/security-for-hadoop.html
9. eCryptfs for Hadoop https://launchpad.net/ecryptfs.
10. Project Rhino - https://github.com/intel-hadoop/project-rhino/

Course code	Course Title	L	T	P	C
DA2104	RISK ANALYTICS	3	0	0	3
	Total contact hours - 45				
	Prerequisite				
	Nil				

PURPOSE

With every organization leaning on to automated analysis tools to speed up the decision making process in an environment where only the fittest survive, it is imperative for students to understand the business process and risk involved. One has to make the analytic solution based on multitude of data available to him. People with domain expertise, statistical and data analytic skill can only provide a better solution.

INSTRUCTIONAL OBJECTIVES

1.	To understand the functioning of Banking and apply analytic techniques to mitigate risks
2.	To understand the operations of Insurance sector and apply analytic techniques to mitigate risks and improve profitability
3.	To understand the processes involved in Healthcare industry and use data analysis to improve patient care and optimize cost
4.	To understand human relationship management techniques for effective management of people

UNIT I - INTRODUCTION (9 hours)

Risk – Definitionand Examples, Components and Factors; Understanding Risk Assessment, Risk Mitigation and Risk Management; Risk Analytics- Definition and Objectives.

UNIT II - RISK ANALYTICS FOR BANKING DOMAIN (9 hours)

Introduction to Banking Sector; National and International laws; Credit Risk Analytics , Internal capital Adequacy Assessment Process related Risk Analytics , Limit Management , Risk-Adjusted Performance Management ,Fraud Risk; Case Studies

UNIT III - RISK ANALYTICS FOR INSURANCE DOMAIN (9 hours)

Introduction to Insurance Sector; Property & Causality Insurance Companies and Life Insurance Companies; Using Analytics for Customer Acquisition and Retention; Detecting, Preventing and Managing Fraud using Analytics; Case Studies

UNIT IV - RISK ANALYTICS FOR HEALTHCARE DOMAIN (9 hours)

Introduction to Healthcare Sector;HIPAA,Four Enterprise Disciplines of Health Analytics, Health Outcome Analysis, Health Value and Cost; Customer Insights, Actuary Services, Framework for Customer Analytics; Risk Management

UNIT V - WORKFORCE ANALYTICS (9 hours)

Workforce Environment and Psychology, HR Analytics and Talent Management- Understanding and Predicting Retention, BoostingEmployee Engagement, Sources of Hire and Quality of Hire, Profiling High Performers

References:

1. Clark Abrahams and Mingyuan Zhang, "*Credit Risk Assessment: The New Lending System for Borrowers, Lenders, and Investors*", ISBN 978-0-470-46168-6

2. Naeem Siddiqi, "*Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*", ISBN 978-0-471-75451-0

3. Laura B. Madsen, "*Data-Driven Healthcare: How Analytics and BI are Transforming the Industry*", M.S.ISBN 978-1-118-77221-8

4. Jason Burke, "Health Analytics: Gaining the Insights to Transform Health Care", John Wiley Sons Inc., 2013, ISBN: 978-1-118-38304-9

5. Jac Fitz-Enz , John R. Mattox II, "Predictive Analytics for Human Resources", ISBN-13: 978-8126552153.

6. James C. Sesil, "Applying Advanced Analytics to HR Management Decisions: Methods for Selection, Developing Incentives, and Improving Collaboration", ISBN-13: 978-0133064605

Weblink:

1. http://www.capgemini.com/resource-file access/resource/pdf/Analytics__A_Powerful_Tool_for_the_Life_Insurance_Industry.pdf

Course code	Course Title	L	T	P	C
IT2110	**INFORMATION STORAGE MANAGEMENT**	**3**	**0**	**0**	**3**
	Total contact hours - 45				
	Prerequisite				
	Knowledge of Database Management Systems, Computer Networks is preferred				

PURPOSE	
Information Storage and Management have highly developed into a sophisticated pillar of information technology, provides a variety of solutions for storing, managing, accessing, protecting, securing, sharing and optimizing information.	
INSTRUCTIONAL OBJECTIVES	
1	Identify the components of managing the data center and Understand logical and physical components of a storage infrastructure.
2	Evaluate storage architectures, including storage subsystems SAN, NAS, IPSAN,CAS
3	Understand thebusiness continuity, backup and recovery methods.

UNIT I-INTRODUCTION TO STORAGE AND MANAGEMENT (9 hours)

Introduction to Information Storage Management - Data Center Environment–Database Management System (DBMS) - Host - Connectivity –Storage-Disk Drive Components- Intelligent Storage System -Components of an Intelligent Storage System- Storage Provisioning- Types of Intelligent Storage Systems.

UNIT II-STORAGE NETWORKING (9 hours)

Fibre Channel: Overview - SAN and Its Evolution -Components of FC SAN -FC Connectivity-FC Architecture- IPSAN-FCOE-FCIP-Network-Attached Storage- General-Purpose Servers versus NAS Devices - Benefits of NAS- File Systems and Network File Sharing-Components of NAS - NAS I/O Operation -NAS Implementations -NAS File-Sharing Protocols-Object-Based Storage Devices-Content-Addressed Storage -CAS Use Cases.

UNIT III-BACKUP AND RECOVERY (9 hours)

Business Continuity -Information Availability -BC Terminology-BC Planning Life Cycle - Failure Analysis -Business Impact Analysis-Backup and Archive - Backup Purpose -Backup Considerations -Backup Granularity - Recovery Considerations -Backup Methods -Backup Architecture - Backup and Restore Operations.

UNIT IV-CLOUD COMPUTING (9 hours)

Cloud Enabling Technologies -Characteristics of Cloud Computing -Benefits of Cloud Computing - Cloud Service Models-Cloud Deployment models-Cloud computing Infrastructure-Cloud Challenges.

UNIT V-SECURING AND MANAGING STORAGE INFRASTRUCTURE (9 hours)

Information Security Framework -Storage Security Domains-Security Implementations in Storage Networking - Monitoring the Storage Infrastructure -Storage Infrastructure Management Activities - Storage Infrastructure Management Challenges.

REFERENCES

1. EMC Corporation, *Information Storage and Management*, WileyIndia, 2nd Edition, 2011.
2. Robert Spalding, *"Storage Networks: The Complete Reference"*, Tata McGraw Hill, Osborne, 2003.
3. Marc Farley, *Building Storage Networks*, Tata McGraw Hill , Osborne,2nd Edition, 2001.
4. Meeta Gupta, *Storage Area Network Fundamentals*, Pearson Education Limited, 2002.

Course code	Course Title	L	T	P	C
IT2111	**CLOUD COMPUTING**	**2**	**0**	**2**	**3**
	Total contact hours – 60				
	Prerequisite				
	Knowledge of Computer Networks is preferred				

PURPOSE

Cloud Computing has drawn the attention of industries and researchers worldwide. Many applications that are being built nowadays were developed to suit the needs of cloud environment. Hence it becomes necessary to have course in cloud computing which deals with the basics of cloud, different services offered by cloud, and security issues in cloud. In a nutshell, this course on cloud computing provides information on fundamental aspects of the cloud environment.

INSTRUCTIONAL OBJECTIVES

1	Learn about different deployment models of cloud and different services offered by cloud
2	Understand the technique of virtualization through theoretical concepts and practical training
3	Become knowledgeable in the rudimentary aspects of cloud application development

UNIT I-CLOUD COMPUTING BASICS (4 hours)

Cloud computing components- Infrastructure-services- storage applications-database services – Deployment models of Cloud- Services offered by Cloud- Benefits and Limitations of Cloud Computing – Issues in Cloud security- Cloud security services and design principles.

UNIT II-VIRTUALIZATION FUNDAMENTALS (4 hours)

Virtualization – Enabling technology for cloud computing- Types of Virtualization- Server Virtualization- Desktop Virtualization – Memory Virtualization – Application and Storage Virtualization- Tools and Products available for Virtualization.

UNIT III-SaaS and PaaS (6 hours)

Getting started with SaaS- Understanding the multitenant nature of SaaS solutions- Understanding OpenSaaS Solutions- Understanding Service Oriented Architecture- PaaS- Benefits and Limitations of PaaS.

UNIT IV-IaaS AND CLOUD DATA STORAGE (6 hours)

Understanding IaaS- Improving performance through Load balancing- Server Types within IaaS solutions- Utilizing cloud based NAS devices – Understanding Cloud based data storage- Cloud based backup devices- Cloud based database solutions- Cloud based block storage.

UNIT V-CLOUD APPLICATION DEVELOPMENT (10 hours)

Client Server Distributed Architecture for cloud – Traditional apps vs. Cloud apps – Client side programming model: Web clients. Mobile clients- Server Side Programming Technologies : AJAX, JSON, Web Services (RPC, REST)- MVC Design Patterns for Cloud Application Development.

Practical (30 hours)

REFERENCES

1. Anthony T .Velte, Toby J.Velte, Robert Elsenpeter, "*Cloud Computing: A Practical Approach*", Tata McGraw Hill Edition, Fourth Reprint, 2010.
2. Kris Jamsa, "*Cloud Computing: SaaS, PaaS, IaaS, Virtualization, Business Models, Mobile, Security and more*", Jones & Bartlett Learning Company LLC, 2013.
3. Ronald L.Krutz, Russell vines, "*Cloud Security: A Comprehensive Guide to Secure Cloud Computing*", Wiley Publishing Inc., 2010.

Course code	Course Title	L	T	P	C
DA2105	**CLUSTER COMPUTING**	**2**	**2**	**0**	**3**
	Total contact hours – 60				
	Prerequisite				
	Nil				

PURPOSE

A computer cluster consists of a set of loosely or tightly connected computers that work together so that, in many respects, they can be viewed as a single system. Unlike grid computers, cluster computers have each node set to perform the same task, controlled and scheduled by software cluster environment. To improve the performance and availability over that of a single computer, cluster computers are essential.

INSTRUCTIONAL OBJECTIVES

1.	Able to Understand the Cluster installation and configuration
2.	Understand the Parallel programming models & paradigms
3.	Familiarize with Job management system and cluster scheduling process.

UNIT I - INTRODUCTION TO CLUSTER COMPUTING (6 hours)

Overview of cluster computing: Elements of cluster, requirements-classes of cluster- Hardware system structure- Node software- Node hardware: Mother board, Memory, Basic Input Output Stream-Peripheral Component Interconnect(PCI) bus, Node assembly.

UNIT II - CLUSTER NETWORKS (6 hours)

Network Hardware: Interconnecting Technologies, Pitfalls in cluster networks-Network software: sockets, Distributed File System, Remote command execution-Setting up clusters: Cluster

configuration, Installation of a node, Basic system administration-Security- NPACI Rocks (Rocks cluster Distribution) - OSCAR Toolkits-Load balancing.

UNIT III - PARALLEL PROGRAMMING (6 hours)

Parallel programming with Message Passing Interface (MPI): MPI compilation and running process, Implementation of MPI for clusters-Dynamic process management-Fault tolerance-RMA-Performance measurement-Parallel Virtual Machine (PVM): Overview, Setup, console details-Extended PVM.

UNIT IV - CLUSTER MANAGEMENT (5 hours)

Goal of workload management software- management activities-Distributed job scheduler-condor: features, architecture- Installation-Configuration –Administration tools.

UNIT V - CLUSTER SCHEDULING (7 hours)

Scheduling Parallel jobs on cluster-High performance cluster scheduler: Maui: overview, Installation-Configuration -Overview of Portable Batch System: Architecture, Features and PVFS: Parallel virtual File System-Mapping and scheduling on Heterogeneous system.

Tutorial : (30 hours)

REFERENCES:

1. Thomas Sterling, "Beowulf Cluster Computing with Linux", MIT Press, Second Edition, 2003.
2. Rajkumar Buyya , "High Performance Cluster Computing: Architectures and Systems", Vol. 1, Prentice Hall PTR, 2007.
3. Rajkumar Buyya, "High Performance Cluster Computing: Programming and Applications", Vol 2, Prentice Hall PTR, NJ, USA, 1999.

Course code	Course Title	L	T	P	C
DA2106	**FUNCTIONAL PROGRAMMING**	**2**	**0**	**2**	**3**
	Total contact hours - 30				
	Prerequisite				
	Nil				

PURPOSE	
Functional programming is a style of programming in which functions are basic building blocks of programs. It treats computing as mathematics functions and applies the function to parameters. This paradigm is getting popular because of reduced side effects and supports parallel code. This course aims at applying functional programming to problems using Scala.	
INSTRUCTIONAL OBJECTIVES	
1.	Functional programming makes it easier to write parallel code for today's and tomorrow's multiprocessors by replacing mutable variables and loops with powerful ways to define and compose functions
2.	To provide a powerful paradigm in which to tackle complex, real-world programming tasks.
3.	Able to understand the elements of the functional programming and learn how to apply them usefully in daily programming tasks.

UNIT I – INTRODUCTION TO FUNCTIONAL PROGRAMMING (6 hours)

Functional & Non- Functional – Benefits of Functional Programming in Scala – Referential Transparency, Purity and Substitution Model – Running a scala functional program – Modules,

Objects, Namespaces – Higher order functions – Polymorphic functions – Following types to implementations.

UNIT II – FUNCTIONAL DATA STRUCTURES & EXCEPTION HANDLING (6 hours)
Defining functional data structures – Pattern matching – Data Sharing – Recursion over lists – Generalizing to higher order functions – Trees – Alternative to exceptions – Option data type – Either data type.

UNIT III – PURELY FUNCTIONAL STATE AND PARALLELISM (6 hours)
Strict and non-strict functions – Lazy Lists Example – Infinite Steams and corecursion – Making stateful APIS pure – Data Types for parallel computations – Combining parallel computations – Explicit forking – Refining API - Algebra of API – Refining combinators.

UNIT IV – FUNCTIONAL DESIGN & FUNCTIONAL DESIGN PATTERNS (6 hours)
Parser Combinators: Design Algebra – Handle context sensitivity – Error Reporting – Implementing algebra **Monoids**: Folding Lists with monoids - Associativity and Parallelism - Foldable Data Structures – Composing Monoids **Monads:** Functors – Monads – Monadic Combinators – Monad Laws.

UNIT V – APPLICATIVE TRAVERSABLE FUNCTORS & IO (6 hours)
Generalizing monads – Applicative trait – Monads vs Applicative functors – Traversable functors – Uses of Traverse - Factoring Effects – Simple IO type – Avoiding Stack Overflow – Non blocking and asynchronous – General purpose IO type.

Practical : (30 hours)

REFERENCES:

1. Paul Chiusano and Rúnar Bjarnason, "Functional Programming in Scala", Manning Publishers, 2014.
2. Dean Wampler, Alex Payne, "Programming Scala", O'Reilly Media, 2009.
3. Scala by Example www.**scala**-lang.org/docu/files/**ScalaByExample**.pdf
4. Martin Odersky, "Scala Language Specification", 2008, http://www.scala-lang.org/docu/files/ScalaReference.pdf
5. Scala Library Documentation : http://www.scala-lang.org/docu/files/api/index.html

Course code	Course Title	L	T	P	C
DA2107	**WEB INTELLIGENCE AND BIG DATA ANALYTICS**	**2**	**0**	**2**	**3**
	Total contact hours - 60				
	Prerequisite				
	Nil				

PURPOSE

The incredible growth of the internet in recent years, along with the vast volumes of 'big data' it holds, has also resulted in a rather significant confluence of ideas from diverse fields of computing and AI. Many of the recent successes in each of these arenas have come through the deployment of many known but disparate techniques working together, and most importantly their deployment at scale, on large volumes of 'big data'; the successes are better described as those of 'web intelligence'.

INSTRUCTIONAL OBJECTIVES

1.	Leveraging intelligent web applications.
2.	Analyzing web applications in the real world.
3.	Building intelligence in your web.

UNIT I – INTRODUCTION TO INTELLIGENT WEB (6 hours)

Inside the search engine - Examples of intelligent web applications - Basic elements of intelligent applications - Machine learning, data mining – Searching, Reading, indexing, and searching.

UNIT II – LISTEN AND LOAD (6 hours)

Streams, Information and Language, - Statistics of Text - Analyzing Sentiment and Intent – Load - Databases and their Evolution, Big data Technology and Trends.

UNIT III – CLUSTERING AND CLASSIFICATION (8 hours)

An overview of clustering algorithms - Clustering issues in very large datasets - The need for classification - Automatic categorization of emails and spam filtering - Classification with very large datasets - Comparing multiple classifiers on the same data.

UNIT IV – REASONING (4 hours)

Reasoning: Logic and its Limits, Dealing with Uncertainty - Mechanical Logic - The Semantic Web - Limits of Logic - Description and Resolution - Collective Reasoning.

UNIT V- PREDICTING (6 hours)

Statistical Forecasting - Neural Networks - Predictive Analytics - Sparse Memories - Sequence Memory - Network Science – Data Analysis: Regression and Feature Selection - Case Study - set of retrieved and processed news stories.

Practical : (30 hours)

REFERENCES

1. Gautam Shroff, “Intelligent Web - Search, Smart Algorithms, and Big Data”, Oxford University Press, 2013.
2. Haralambos Marmanis, Dmitry Babenko, “Algorithms of the Intelligent Web”, Manning publications, 2009.
3. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, “An Introduction to Information Retrieval”, Cambridge University Press, 2009.
4. Mark Gardener, “Beginning R - The Statistical Pr ogramming Language”, John Wiley & Sons, Inc., 2012.
5. W. N. Venables, D. M. Smith and the R Core Team, “An Introduction to R”, 2013.
6. http://www.coursetalk.com/coursera/web-intelligence-and-big-data

Course code	Course Title	L	T	P	C
DA2108	**SOCIAL NETWORK ANALYTICS**	**2**	**0**	**2**	**3**
	Total contact hours - 60				
	Prerequisite				
	Nil				
PURPOSE					
This course will be used for social network analysis, both its theory and computational tools, to make sense of the social and information networks that have been fueled and rendered					

accessible by the internet.	
INSTRUCTIONAL OBJECTIVES	
1.	Analyze the structure and evolution of networks
2.	Able to gain knowledge from disciplines as diverse as sociology, mathematics, computer science.
3.	Understand the Online interactive demonstrations and hands-on analysis of real-world data sets.

UNIT I - INTRODUCTION (6 hours)

Overview: Social network data-Formal methods- Paths and Connectivity-Graphs to represent social relations-Working with network data- Network Datasets-Strong and weak ties - Closure, Structural Holes, and Social Capital.

UNIT II - SOCIAL INFLUENCE (6 hours)

Homophily: Mechanisms Underlying Homophily, Selection and Social Influence, Affiliation,Tracking Link Formation in OnLine Data, Spatial Model of Segregation - Positive and Negative Relationships - Structural Balance - Applications of Structural Balance, Weaker Form of Structural Balance.

UNIT III - INFORMATION NETWORKS AND THE WORLD WIDE WEB (6 hours)

The Structure of the Web- World Wide Web- Information Networks, Hypertext, and Associative Memory- Web as a Directed Graph, Bow-Tie Structure of the Web- Link Analysis and Web Search-Searching the Web: Ranking,Link Analysis using Hubs and Authorities- Page Rank- Link Analysis in Modern Web Search,Applications, Spectral Analysis, Random Walks, and Web Search.

UNIT IV- SOCIAL NETWORK MINING (6 hours)

Clustering of Social Network graphs: Betweenness, Girvan newman algorithm-Discovery of communities- Cliques and Bipartite graphs-Graph partitioning methods-Matrices-Eigen values-Simrank.

UNIT V - NETWORK DYNAMICS (6hours)

Cascading Behavior in Networks:Diffusion in Networks,Modeling Diffusion - Cascades and Cluster, Thresholds, Extensions of the Basic Cascade Model- Six Degrees of Separation-Structure and Randomness, Decentralized Search- Empirical Analysis and Generalized Models- Analysis of Decentralized Search.

Practical : (30 hours)

REFERENCES:

1. Easley and Kleinberg, "Networks, Crowds, and Markets: Reasoning about a highly connected world", Cambridge Univ. Press, 2010.
2. Robert A. Hanneman and Mark Riddle, "Introduction to social network methods", University of California, 2005.
3. Jure Leskovec,Stanford Univ.Anand Rajaraman,Milliway Labs, Jeffrey D. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2 edition, 2014.
4. Wasserman, S., & Faust, K, "Social Network Analysis: Methods and Applications", Cambridge University Press; 1 edition, 1994.
5. Borgatti, S. P., Everett, M. G., & Johnson, J. C., "Analyzing social networks", SAGE Publications Ltd; 1 edition, 2013.
6. John Scott , "Social Network Analysis: A Handbook" , SAGE Publications Ltd; 2nd edition, 2000.

Course code	Course Title	L	T	P	C
DA2109	STREAMING ANALYTICS	2	0	2	3
	Total contact hours - 60				
	Prerequisite				
	Nil				

PURPOSE

In recent days more and more live and real time data is generated. Hence rather than static offline analysis and reporting, the need for realtime analytics of data and decision making has acquired significant importance. Such data analytics is called streaming data analytics. This course aims at introducing the basics of streaming analytics.

INSTRUCTIONAL OBJECTIVES

1.	Understanding the need for stream computing
2.	Comprehend the architecture of stream analytics
3.	Building data flow management pipelines for streams.
4.	Processing streaming data
5.	Delivering the results of streaming analytics

UNIT I - INTRODUCTION TO STREAM COMPUTING (6 hours)

Streaming Data – Sources – Difference between Streaming Data and Static Data. Overview of Large Scale Stream Processing Engines – Issues in Stream Processing.

UNIT II - STREAMING ANALYTICS ARCHITECTURE (6 hours)

Phases in Streaming Analytics Architecture - Vital Attributes - High Availability – Low Latency – Horizontal Scalability-Fault Tolerance - Service Configuration and Management - Apache ZooKeeper.

UNIT III - DATA FLOW MANAGEMENT (6 hours)

Distributed Data Flows – At Least One Delivery – Apache Kafka – Apache Flume – Zero MQ - Messages, Events, Tasks & File Passing.

UNIT IV - PROCESSING & STORING STREAMING DATA (6 hours)

Distributed Stream Data Processing: Co-ordination, Partition and Merges, Transactions. Duplication Detection using Bloom Filters - Apache Spark Streaming Examples Choosing a storage system – NoSQL Storage Systems.

UNIT V - DELIVERING STREAMING METRICS (6 hours)

Visualizing Data – Mobile Streaming Apps –Times Counting and Summation - Stochastic Optimization – Delivering Time Series Data.

Practical : (30 hours)

REFERENCES:

1. Byron Ellis, “Real-Time Analytics: Techniques to Analyze and Visualize Streaming Data”, Wiley, 1st edition, 2014.
2. Sherif Sakr, “Large Scale and Big Data: Processing and Management”, CRC Press, 2014. 2014.
3. Bill Franks, “Taming The Big Data Tidal Wave Finding Opportunities In Huge Data Streams With Advanced Analytics”, Wiley, 2012.

4. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, “Mining of Massive Datasets”, Cambridge University Press, 2014.
5. Paul C Zikopoulos, Chris Eaton, Paul Zikopoulos, “Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data”, McGraw-Hil, 1st edition, 2011.
6. **kafka**.**apache**.org
7. **flume**.**apache**.org
8. **zookeeper**.**apache**.org
9. **spark**.**apache**.org
10. **zeromq**.org

Course code	Course Title	L	T	P	C
DA2110	ADVANCED ALGORITHMS	2	0	2	3
	Total contact hours - 60				
	Prerequisite				
	Nil				

PURPOSE

A major thrust of computer science is the design, analysis, implementation, and scientific evaluation of algorithms to solve critical problems. As parallel computing continues to merge into the mainstream of computing, it becomes more and more important for students and scientists to understand the application and analysis of algorithmic paradigms to both the (traditional) sequential model of computing and to a variety of parallel models.

INSTRUCTIONAL OBJECTIVES

1.	To integrate the parallel and sequential algorithms.
2.	Design and analysis of paradigms for sequential and parallel models.

UNIT I - INTRODUCTION TO ALGORITHMS (6 hours)

Introduction to Preliminaries - Design and Analysis Fundamentals - Mathematical Tools for Algorithm Analysis - Trees and Applications to Algorithms - More on Sorting Algorithms - Probability and Average Complexity of Algorithms.

UNIT II - DESIGN STRATEGIES (6 hours)

Major Design Strategies - The Greedy Method – Divide and Conquer - Dynamic Programming - Backtracking and Branch and Bound.

UNIT III - GRAPH AND NETWORK ALGORITHMS (6 hours)

Graph and Network Algorithms - Graphs and Digraphs - Minimum Spanning Tree and Shortest-Path Algorithms - Graph Connectivity and Fault-Tolerance of Networks - Matching and Network Flow Algorithms.

UNIT IV - PARALLEL AND DISTRIBUTED ALGORITHMS (6 hours)

Parallel and Distributed Algorithms - Introduction to Parallel Algorithms and Architectures - Parallel Design Strategies - Internet Algorithms -Distributed Computation Algorithms - Distributed Network Algorithms.

UNIT V – SEARCH ALGORITHMS (6 hours)

String Matching and Document Processing - Balanced Search Trees - The Fast Fourier Transform - Heuristic Search Strategies: A* - Search and Game Trees 24 - Probabilistic and Randomized Algorithms - Lower-Bound Theory - NP-Complete Problems - Approximation Algorithms.

Practical : (30 hours)

REFERENCES

1. Kenneth A. Berman, Jerome L. Paul , “Algorithms: Sequential, Parallel, and Distributed”, Amazon Bestsellers, 2004.
2. Russ Miller, Laurence Boxer, “Algorithms Sequential and Parallel: A Unified Approach”, Prentice Hall, 1 edition, 1999.
3. Dimitri P. Bertsekas and John N. Tsitsiklis, “Parallel and Distributed Computation: Numerical Methods”, Prentice Hall, 1989.

Course code	Course Title	L	T	P	C
DA2111	NATURAL LANGUAGE PROCESSING	2	2	0	3
	Total contact hours - 60				
	Prerequisite				
	Nil				

PURPOSE

An explosion of Web-based language techniques, merging of distinct fields, availability of phone-based dialogue systems, and much more make this an exciting time in speech and language processing. The first of its kind to thoroughly cover language technology – at all levels and with all modern technologies, based on applying statistical and other machine-learning algorithms to large corporations.

INSTRUCTIONAL OBJECTIVES

1.	Able to illustrate the relative strengths and weaknesses of various approaches.
2.	Understanding statistical sequence labeling, information extraction, question answering and summarization, advanced topics in speech recognition, speech synthesis.
3.	Learn the language modeling, formal grammars, statistical parsing, machine translation, and dialog processing.

UNIT I - INTRODUCTION (6 hours)

Overview: Origins and challenges of NLP- Theory of Language -Features of Indian Languages – Issues in Font –Models and Algorithms- NLP Applications.

UNIT II - MORPHOLOGY AND PARTS-OF-SPEECH (6 hours)

Phonology – Computational Phonology - Words and Morphemes – Segmentation – Categorization and Lemmatisation – Word Form Recognition – Valency - Agreement - Regular Expressions – Finite State Automata – Morphology- Morphological issues of Indian Languages – Transliteration.

UNIT III - PROBABILISTIC MODELS (6 hours)

Probabilistic Models of Pronunciation and Spelling – Weighted Automata – N- Grams – Corpus Analysis – Smoothing – Entropy - Parts-of-Speech – Taggers – Rule based – Hidden Markov Models – Speech Recognition.

UNIT IV - SYNTAX (6 hours)

Basic Concepts of Syntax – Parsing Techniques – General Grammar rules for Indian Languages – Context Free Grammar – Parsing with Context Free Grammars – Top Down Parser – Earley Algorithm – Features and Unification - Lexicalised and Probabilistic Parsing.

UNIT V - SEMANTICS AND PRAGMATICS (6 hours)

Representing Meaning – Computational Representation – Meaning Structure of Language – Semantic Analysis – Lexical Semantics – WordNet – Pragmatics – Discourse – Reference Resolution – Text Coherence – Dialogue Conversational Agents.

TUTORIAL : (**30 hours)**

REFERENCES

1. Daniel Jurafskey and James H. Martin "Speech and Language Processing", Prentice Hall, 2009.
2. Christopher D.Manning and Hinrich Schutze, "Foundation of Statistical Natural Language Processing", MIT Press, 1999.
3. Ronald Hausser, "Foundations of Computational Linguistics", Springer-Verleg, 1999.
4. James Allen, "Natural Language Understanding", Benjamin/Cummings Publishing Co. 1995.
5. Steve Young and Gerrit Bloothooft, "Corpus – Based Methods in Language and Speech Processing", Kluwer Academic Publishers, 1997.

Course code	Course Title	L	T	P	C
CS2108	**PATTERN RECOGNITION TECHNIQUES**	**3**	**0**	**0**	**3**
	Total contact hours - 60				
	Prerequisite				
	Nil				

PURPOSE	
To study the Pattern Recognition techniques and its applications.	
INSTRUCTIONAL OBJECTIVES	
1.	To learn the fundamentals of Pattern Recognition techniques
2.	To learn the various Statistical Pattern recognition techniques
3.	To learn the various Syntactical Pattern recognition techniques
4.	To learn the Neural Pattern recognition techniques

UNIT I – PATTERN RECOGNITION OVERVIEW (9 hours)

Pattern recognition, Classification and Description—Patterns and feature extraction with Examples—Training and Learning in PR systems—Pattern recognition Approaches.

UNIT II – STATISTICAL PATTERN RECOGNITION (9 hours)

Introduction to statistical Pattern Recognition—supervised Learning using Parametric and Non Parametric Approaches.

UNIT III – LINEAR DISCRIMINANT FUNCTIONS AND UNSUPERVISED LEARNING AND CLUSTERING (9 hours)

Introduction—Discrete and binary Classification problems—Techniques to directly obtain linear Classifiers -- Formulation of Unsupervised Learning Problems—Clustering for unsupervised learning and classification.

UNIT IV– SYNTACTIC PATTERN RECOGNITION (9 hours)

Overview of Syntactic Pattern Recognition—Syntactic recognition via parsing and other grammars– Graphical Approaches to syntactic pattern recognition—Learning via grammatical inference.

UNIT V– NEURAL PATTERN RECOGNITION (9 hours)

Introduction to Neural networks—Feedforward Networks and training by Back Propagation—Content Addressable Memory Approaches and Unsupervised Learning in Neural PR.

REFERENCES

1. Robert Schalkoff, "*Pattern Recognition: Statistical Structural and NeuralApproaches*", John wiley & sons , Inc,1992.
2. Earl Gose, Richard johnsonbaugh, Steve Jost, "*Pattern Recognition andImage Analysis*", Prentice Hall of India,.Pvt Ltd, New Delhi, 1996.
3. Duda R.O., P.E.Hart & D.G Stork, " *Pattern Classification*", 2nd Edition, J.Wiley Inc 2001.
4. Duda R.O.& Hart P.E., "*Pattern Classification and Scene Analysis*", J.wiley Inc, 1973.
5. Bishop C.M., "*Neural Networks for Pattern Recognition*", Oxford University Press, 1995.

Course Code	Course Title	L	T	P	C
DA2112	**Deep Learning**	**2**	**0**	**2**	**3**
	Total contact hours - 60				
	Prerequisite				
	Knowledge in Linear Algebra and Probability is preferred.				

PURPOSE

This course will enable the students to study fundamental concepts of Deep Learning and its applications useful in modern data processing Students will have knowledge on feed-forward deep neural networks and selected generative models. Students will know methods and algorithms for unsupervised and supervised training of deep neural networks.

INSTRUCTIONAL OBJECTIVES

1.	Understanding a very broad collection of machine learning algorithms and problems.
2.	To learn algorithmic topics of machine learning and mathematically deep enough to introduce the required theory.

UNIT-I FUNDAMENTALS CONCEPTS OF MACHINE LEARNING (6 Hours)

Historical Trends in Deep Learning-Machine Learning Basics: Learning Algorithms-Supervised and Unsupervised Training, Linear Algebra for machine Learning, Testing, Cross-Validation, Dimensionality reduction, Over/Under-fitting, Hyper parameters and validation sets, Estimators, Bias, Variance, Regularization-Introduction to a simple DNN, Platform for deep learning, Deep learning software libraries.

UNIT-II DEEP FEED FORWARD NETWORKS (6 Hours)

Deep feed forward networks-Introduction- Learning XOR- Gradient-Based Learning- Various Activation Functions, error functions- Architecture Design-differentiation algorithms-Regularization for Deep learning-Early Stopping, Drop out.

UNIT III CONVOLUTIONAL NEURAL NETWORKS AND SEQUENCE MODELING (6 Hours)

Convolutional Networks: Convolutional operation- Motivation- Pooling- Normalization, Applications in Computer Vision: Imagenet- Sequence Modeling: Recurrent Neural Networks-Difficulty in Training RNN- Encoder-Decoder.

UNIT IV AUTO ENCODERS (6 Hours)

Auto encoders - Auto encoders: under complete, regularized, stochastic, denoising, contractive, applications – dimensionality reduction, classification, recommendation, Optimization for Deep Learning: optimizers. RMS Prop for RNNs, SGD for CNNs

UNIT V DEEP ARCHITECTURES IN VISION (6 Hours)

Deep Architectures in Vision - Alexnet to ResNet, Transfer learning, Siamese Networks, Metric Learning, Ranking/Triplet loss, RCNNs, CNN-RNN, Applications in captioning and video tasks, 3D CNNs

Practical: (30 hours)

REFERENCES

1. Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning", MIT Press, 2016 (available at *http://www.deeplearningbook.org*)
2. Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012
3. Michael Nielsen, "Neural Networks and Deep Learning", Online book, 2016 (http://neuralnetworksanddeeplearning.com/)
4. Li Deng, Dong Yu, "Deep Learning: Methods and Applications", Foundations and Trends in Signal Processing.
5. Christopher and M. Bishop, "Pattern Recognition and Machine Learning", Springer Science Business Media, 2006.
6. Jason Brownlee , "Deep Learning with Python" , ebook, 2016
7. N. D. Lewis , "Deep Learning Step by Step with Python: A Very Gentle Introduction to Deep Neural Networks for Practical Data Science

SEMESTER I

		L	T	P	C
CAC2001	**CARRET ADVANCEMENT COURSE FOR ENGINEERS - I**	**1**	**0**	**1**	**1**
	Total contact hours – 30				
	Prerequisite				
	NIL				

PURPOSE	
To enhance holistic development of students and improve their employability skills.	
INSTRUCTIONAL OBJECTIVES	
1	To improve aptitude, problem solving skills and reasoning ability of the student.
2	To collectively solve problems in teams & group.
3	Understand the importance of verbal and written communication in the workplace
4	Understand the significance of oral presentations, and when they may be used.
5	Practice verbal communication by making a technical presentation to the class
6	Develop time management Skills

UNIT I–BASIC NUMERACY

- Types and Properties of Numbers, LCM, GCD, Fractions and decimals, Surds

UNIT II-ARITHMETIC – I

- Percentages, Profit & Loss, Equations

UNIT III-REASONING - I

- Logical Reasoning

UNIT IV-SOFT SKILLS - I

- Presentation skills, E-mail Etiquette

UNIT V-SOFT SKILLS - II

- Goal Setting and Prioritizing

ASSESSMENT

Soft Skills (Internal)

Assessment of presentation and writing skills.

Quantitative Aptitude (External)

Objective Questions- 60 marks
Descriptive case lets- 40 marks*
Duration: 3 hours

*Engineering problems will be given as descriptive case lets.

REFERENCE:

1. Quantitative Aptitude by Dinesh Khattar – Pearsons Publicaitons
2. Quantitative Aptitude and Reasoning by RV Praveen – EEE Publications
3. Quantitative Aptitude by Abijith Guha – TATA Mc GRAW Hill Publications
4. Soft Skills for Everyone by Jeff Butterfield – Cengage Learning India Private Limited
5. Six Thinking Hats is a book by Edward de Bono - Little Brown and Company
6. IBPS PO - CWE Success Master by Arihant - Arihant Publications(I) Pvt.Ltd – Meerut

SEMESTER II

		L	T	P	C
CAC2001	**CARRET ADVANCEMENT COURSE FOR ENGINEERS - II**	**1**	**0**	**1**	**1**
	Total contact hours – 30				
	Prerequisite				
	NIL				

PURPOSE	
To enhance holistic development of students and improve their employability skills.	
INSTRUCTIONAL OBJECTIVES	
1	To improve aptitude, problem solving skills and reasoning ability of the student.
2	To collectively solve problems in teams & group.
3	Understand the importance of verbal and written communication in the workplace
4	Understand the significance of oral presentations, and when they may be used.
5	Understand the fundamentals of listening and how one can present in a group discussion
6	Prepare or update resume according to the tips presented in class.

UNIT I-ARITHMETIC – II

- Ratios & Proportions, Mixtures & Solutions

UNIT II - MODERN MATHEMATICS

- Sets & Functions, Data Interpretation, Data Sufficiency

UNIT III – REASONING - II

- Analytical Reasoning

UNIT IV – COMMUNICATION - I

- Group discussion, Personal interview

UNIT V - COMMUNICATION - II

- Verbal Reasoning test papers

ASSESSMENT

Communication (Internal)

- Individuals are put through formal GD and personal interviews.
- Comprehensive assessment of individuals' performance in GD & PI will be carried out.

Quantitative Aptitude (External)

Objective Questions- 60 marks (30 Verbal +30 Quants)
Descriptive case lets- 40 marks*
Duration: 3 hours

*Engineering problems will be given as descriptive case lets.

REFERENCES

1. Quantitative Aptitude by Dinesh Khattar – Pearsons Publicaitons
2. Quantitative Aptitude and Reasoning by RV Praveen – EEE Publications
3. Quantitative Aptitude by Abijith Guha – TATA Mc GRAW Hill Publications
4. General English for Competitive Examination by A.P. Bharadwaj – Pearson Educaiton
5. English for Competitive Examination by Showick Thorpe - Pearson Educaiton
6. IBPS PO - CWE Success Master by Arihant - Arihant Publications(I) Pvt.Ltd - Meerut
7. Verbal Ability for CAT by Sujith Kumar - Pearson India
8. Verbal Ability & Reading Comprehension by Arun Sharma - Tata McGraw - Hill Education

SEMESTER III

		L	T	P	C
CAC2003	**CARRET ADVANCEMENT COURSE FOR ENGINEERS - III**	**1**	**0**	**1**	**1**
	Total contact hours – 30				
	Prerequisite				
	NIL				
PURPOSE					
To develop professional skills abreast with contemporary teaching learning methodologies.					
INSTRUCTIONAL OBJECTIVES					
1	Acquire knowledge on planning, preparing and designing a learning program				
2	Prepare effective learning resources for active practice sessions				
3	Facilitate active learning with new methodologies and approaches				
4	Create balanced assessment tools				
5	Hone teaching skills for further enrichment				

UNIT I- DESIGN **(2 hrs)**

- Planning &Preparing a learning program.
- Planning & Preparing a learning session

UNIT II – PRACTICE **(2 hrs)**

- Facilitating active learning
- Engaging learners

UNIT III – ASSESSMENT **(2 hrs)**

- Assessing learner's progress
- Assessing learner's achievement

UNIT IV – HANDS ON TRAINING **(10 hrs)**

- Group activities – designing learning session
- Designing teaching learning resources
- Designing assessment tools
- Mock teaching session

UNIT V – TEACHING IN ACTION **(14 hrs)**

- Live teaching sessions
- Assessments

ASSESSMENT (Internal)
Weightage:
Design - 40%
Practice – 40%
Quiz – 10%
Assessment – 10%

REFERENCES

Cambridge International Diploma for Teachers and Trainers Text book by Ian Barker - Foundation books

Whitehead, Creating a Living Educational Theory from Questions of the kind: How do I improve my Practice? Cambridge J. of Education