VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

CURRICULUM AND SYLLABI

(2018-2019)

M.Tech (CSE) - Specialisation in Big Data Analytics

VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

VISION STATEMENT OF VELLORE INSTITUTE OF TECHNOLOGY

Transforming life through excellence in education and research.

MISSION STATEMENT OF VELLORE INSTITUTE OF TECHNOLOGY

World class Education: Excellence in education, grounded in ethics and critical thinking, for improvement of life.

Cutting edge Research: An innovation ecosystem to extend knowledge and solve critical problems.

Impactful People: Happy, accountable, caring and effective workforce and students.

Rewarding Co-creations: Active collaboration with national & international industries & universities for productivity and economic development.

Service to Society: Service to the region and world through knowledge and compassion.

VISION STATEMENT OF THE SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

To be a world-renowned centre of education, research and service in computing and allied domains.

MISSION STATEMENT OF THE SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

- To offer computing education programs with the goal that the students become technically competent and develop lifelong learning skill.
- To undertake path-breaking research that creates new computing technologies and solutions for industry and society at large.
- To foster vibrant outreach programs for industry, research organizations, academia and society.

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

M.Tech (CSE) - Specialisation in Big Data Analytics

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

1. Graduates will be engineering professionals who will engage in technology development and deployment with social awareness and responsibility.

2. Graduates will function as successful practicing engineer / researcher / teacher / entrepreneur in the chosen domain of study.

3. Graduates will have holistic approach addressing technological, societal, economic and sustainability dimensions of problems and contribute to economic growth of the country.

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

M. Tech Computer Science and Engineering
Specialization in Big Data Analytics

PROGRAMME OUTCOMES (POs)

PO_1 Having an ability to apply mathematics and science in engineering applications

PO_2 Having an ability to design a component or a product applying all the relevant standards and with realistic constraints

PO_3 Having an ability to design and conduct experiments, as well as to analyze and interpret data

PO_4 Having an ability to use techniques, skills and modern engineering tools necessary for engineering practice

PO_5 Having problem solving ability- solving social issues and engineering problems

PO_6 Having adaptive thinking and adaptability

PO_7 Having a clear understanding of professional and ethical responsibility

PO_8 Having a good cognitive load management [discriminate and filter the available data] skills

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

M.Tech (CSE) - Specialisation in Big Data Analytics

PROGRAMME SPECIFIC OUTCOMES (PSOs)

1. Ability to design and develop computer programs/computer-based systems in the advanced level of areas including algorithms design and analysis, networking, operating systems design, etc.

2. Ability to apply the advanced concepts of Big Data that pave the way to create a platform to gain analytical skills which impacts business decisions and strategies.

3. Ability to bring out the capabilities for research and development in contemporary issues and to exhibit the outcomes as technical report.

VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

M. Tech Computer Science and Engineering
Specialization in Big Data Analytics

CREDIT STRUCTURE

Category-wise Credit distribution

Category	Credits
University Core (UC)	**27**
Programme Core (PC)	**18**
Programme Elective (PE)	**19**
University Elective (UE)	**06**
Bridge Course (BC)	-
Total Credits	**70**

CURRICULUM

M.Tech.-CSE (Spl. in Big Data Analytics) - (2018)

Programme Core	Programme Elective	University Core	University Elective	Total Credits
18	19	27	6	70

Course Code	Course Title	Course Type	L	T	P	J	C
PROGRAMME CORE							
CSE5001	Algorithms: Design and Implementation	ETL	2	0	2	0	3
CSE5003	Database Systems: Design and Implementation	ETLP	2	0	2	4	4
CSE5007	Exploratory Data Analysis	ETP	2	0	0	4	3
CSE6001	Bigdata Frameworks	ETLP	2	0	2	4	4
CSE6005	Machine Learning	ETLP	2	0	2	4	4
Course Code	**Course Title**	**Course Type**	**L**	**T**	**P**	**J**	**C**
PROGRAMME ELECTIVE							
CSE5002	Operating Systems and Virtualization	ETL	2	0	2	0	3
CSE6006	NoSQL Databases	ETLP	2	0	2	4	4
CSE6014	Programming for Data Science	LO	0	0	4	0	2
CSE6016	Information Visualization	ETLP	2	0	2	4	4
CSE6017	Mining Massive Data	ETLP	2	0	2	4	4
CSE6018	Streaming Data Analytics	ETLP	2	0	2	4	4
CSE6019	Text, Web and Social Media Analytic	ETP	3	0	0	4	4
CSE6020	Big Data Technologies	ETLP	2	0	2	4	4
CSE6021	Domain Specific Predictive Analytics	ETP	3	0	0	4	4
CSE6022	Soft Computing	ETP	3	0	0	4	4
CSE6023	Cloud Computing Fundamentals	ETLP	2	0	2	4	4
CSE6025	Analytics of Things	ETP	3	0	0	4	4
Course Code	**Course Title**	**Course Type**	**L**	**T**	**P**	**J**	**C**
UNIVERSITY CORE							
CSE6099	Masters Thesis	PJT	0	0	0	0	16
MAT6001	Advanced Statistical Methods	ETL	2	0	2	0	3
SET5001	Science, Engineering and Technology Project - I	PJT	0	0	0	0	2
SET5002	Science, Engineering and Technology Project - II	PJT	0	0	0	0	2
EFL5097	English and Foreign Language	CDB	0	0	0	0	2
ENG5001 - Fundamentals of Communication Skills - LO							
ENG5002 - Professional and Communication Skills - LO							
FRE5001 - Francais fonctionnel – TH							
GER5001 - Deutsch fuer Anfaenger – TH							
STS6777	Soft Skills M.Tech.	CDB	0	0	0	0	2
STS5001 - Essentials of Business Etiquettes - SS							
STS5001 - Essentials of Business Etiquette and Problem Solving - SS							
STS5002 - Preparing for Industry – SS							
STS5102 - Programming and Problem Solving Skills - SS							

Course Code	Course Title	Course Type	L	T	P	J	C
Course Code	Course Title	Course Type	L	T	P	J	C
BRIDGE COURSE							
Course Code	Course Title	Course Type	L	T	P	J	C
NON CREDIT COURSE							

CSE5001	ALGORITHMS: DESIGN AND IMPLEMENTATION	L	T	P	J	C
		2	0	2	0	3
Pre-requisite	**NIL**				**Syllabus version**	
					1.0	

Course Objectives:

1. To focus on the design of algorithms in various domains
2. To provide a foundation for designing efficient algorithms.
3. To provide familiarity with main thrusts of working algorithms-sufficient to gives context for formulating and seeking known solutions to an algorithmic problem.

Expected Course Outcome:

1. Solve a problem using Algorithms and design techniques
2. Solve complexities of problems in various domains
3. Implement algorithm, compare their performance characteristics, and estimate their potential effectiveness in applications
4. Solve optimization problems using simplex algorithm
5. Designing approximate algorithms for graph theoretical problems
6. Application of appropriate search algorithms for graphs and trees
7. Application of computational geometry method on optimization problems

Module:1	**Introduction**	**5 hours**

Algorithm design techniques: Divide and Conquer, Brute force, Greedy, Dynamic Programming. Time complexity (asymptotic notation, recurrence relations)

Module:2	**Network Flows**	**5 hours**

Maximum Flows, Min-cost Flows, Max-Flow Min-Cut Theorem, Cycle Canceling Algorithms, Strongly Polynomial-time Analysis, Minimum Cuts without Flows

Module:3	Tractable and Intractable Problems	**4 hours**

Class complexity: P, NP, NP-Hard, NP-Complete Approximation Algorithms

Module:4	Approximation Algorithms	**4 hours**

Limits to Approximability, Vertex Cover problem, Set cover problem, Euclidean TSP

Module:5	Search Algorithms for Graphs and Trees	**4 hours**

Overview of fundamental algorithms, Dijkstra's algorithm, A*search algorithm

Module:6	Computational Geometry	**4 hours**

Line Segments, Convex hull finding algorithms

Module:7	Linear Programming	**2 hours**

Representing problems-shortest paths, maximum flow, and minimum-cost flow as linear programming problems. Simplex algorithm

Module:8	Recent Trends	**2 hours**
	Total Lecture hours:	**30 hours**

Text Book(s)

Reference Books

1. Cormen, Leiserson, Rivest and Stein, Introduction to Algorithms, 3rd edition, McGraw-Hill, 2009.
2. J.Kleinberg and E.Tardos. Algorithm Design, Pearson Education, 2009.
3. E.Horowitz, S.Sahni, S.Rajasekaran, Fundamentals of Computer Algorithms, 2nd edition, Universities Press, 2011.
4. Ravindra K.Ahuja, Thomas L. Magnanti, and James B.Orlin, Network Flows: Theory, Algorithms, and Applications, Pearson Education, 2014.
5. George T. Heineman, Gary Pollice, Stanley Selkow, Algorithms in a nutshell, O'Reilly Media, 2nd edition, 2016.

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar

List of Challenging Experiments (Indicative)

1.	Implementation of algorithms for problems that can be solved by one or more of the following strategies: Divide and Conquer, Brute force, Greedy, Dynamic Programming.	3 hours
2.	Implementation of Ford Fulkerson method, Edmonds-Karp algorithm for finding maximum flow in a flow network and applying them for solving typical problems such as railway network flow, maximum bipartite matching	2 hours
3.	Implementation of Dinics strongly polynomial algorithm for computing them maximum flow in a flow network and applying it for solving typical problems	2 hours
4.	Implementation of push-relabel algorithm of Goldberg and Tarjan for finding maximum flow in a flow network and applying it for solving typical problems.	2 hours
5.	Applying linear programming for solving maximum flow problem	2 Hours
6.	Applying network flow algorithms for baseball elimination and airline scheduling	2 Hours
7.	Given a flow network G=(V,E,s,t) ,where V is the vertex set, E is the edge set, s and t are source and destination. An edge of the flow network is called critical if a decrease in the flow over that edge results in a decrease in the total flow of the flow network. An edge of the flow network is called a bottleneck edge if an increase in the flow over that edge results in an increase in the total flow of the flow network. Assume that you are using to compute the maximum flow of the network. (a) Write a program (any language) to identify all the critical edges. (b) Write a program (any language) to identify all bottleneck edges in the network.	3 Hours
8.	Implementation of solution techniques for the minimum-cost flow problem.	2 hours

9.	Design a polynomial time algorithm to compute the solution of a linear programming problem in two dimensions. Your algorithm should convert each constrain to f the problem, into a planar region. Use that algorithm to compute the solution of the following problem. Implement your algorithm in any programming language. A manufacturer of furniture makes two products: chairs and tables. Processing of these products is done on two machines M1 and M2. A chair requires 2hours on machine M1and 6hours on machine M2. A table requires5 hours on machine M1and no time on machine M2.There are 16 hours of time per day available on machine M1and30 hours on machine M2. Profits gained by manufacturer from a chair and a table are Rs.1and Rs.5 respectively. The problem is to maximize the profit for the manufacturer.	2 hours
10.	Implementation of algorithms for the vertex cover problem, set cover problem, TSP	2 hours
11.	Implementation of search algorithms for graphs and trees: fundamental algorithms, Dijkstra's algorithm	2 hours
12.	Consider the problem of barricading sleeping tigers by a fence of shortest length. Forest officials have tranquilized each tiger. Suggest an algorithm for the purpose. You are allowed to assume any information required for your algorithm. Implement your algorithm in any programming language (using convex hull)	3 hours
13.	A simple polygon is defined as a flat shape consisting of straight non-intersecting line segments or sides that are joined pairwise tofromaclosedpath.Letp1,p2,....,pn be a set of points in the two dimensional plane. (a) Write a program to find the simple polygon of P. (b) Write a program (linear time) to convert that the simple polygon of P to a Convex Hull.	3 hours
Total Laboratory Hours		**30 hours**

Mode of assessment:			
Recommended by Board of Studies	**13.05.2016**		
Approved by Academic Council	41	**Date**	**17.06.2016**

CSE5003	DATABASE SYSTEMS: DESIGN AND IMPLEMENTATION	L	T	P	J	C
		2	0	2	4	4
Pre-requisite	**NIL**	**Syllabus version**				
		1.0				

Course Objectives:

1. To emphasize the underlying principles of Relational Database Management System.
2. To model and design advanced data models to handle threat issues and counter measures.
3. To implement and maintain the structured, semi-structured and unstructured data in an efficient database system using emerging trends.

Expected Course Outcome:

1. Design and implement database depending on the business requirements and considering various design issues.
2. Select and construct appropriate parallel and distributed database architecture and formulate the cost of queries accordingly.
3. Understand the requirements of data and transaction management in mobile and spatial database and differentiate those with RDBMS.
4. Categorize and design the structured, semi-structured and unstructured databases.
5. Characterize the database threats and its counter measures.
6. Review cloud, streaming and graph databases.
7. Comprehend, design and query the database management system.

Module:1	Relational Model	6 hours

Database System Architecture–EER Modeling-Indexing–Normalization–Query processing and optimization – Transaction Processing

Module:2	Parallel Databases	4 hours

Architecture, Data partitioning strategy, Interquery and Intraquery Parallelism –Parallel Query Optimization

Module:3	Distributed Databases	5 hours

Features – Distributed Database Architecture –Fragmentation –Replication- Distributed Query Processing – Distributed Transactions Processing

Module:4	Spatial and Mobile Databases	3 hours

Spatial databases-Type of spatial data–Indexing in spatial databases, Mobile Databases– Transaction Model in MDS

Module:5	Semi-Structured Databases	4 hours

Semi Structured databases – XML –Schema-DTD- XPath- XQuery, Semantic Web –RDF–RDFS

Module:6	Database Security	3 hours

Introduction to Database Security Issues–Security Models–Different Threats to databases– Counter measures to deal with these problems		
Module:7	**Emerging Technologies**	**3 hours**
Cloud databases – Streaming Databases - Graph Databases-New SQL		
Module:8	**Recent Trends**	**2 hours**
	Total Lecture hours:	**30 hours**

Text Book(s)

1. Avi Silberschatz, Hank Korth, and S.Sudarshan,"Database System Concepts", 6th Ed. McGraw Hill, 2010.
2. Ramez Elmasri B.Navathe: "Fundamentals of database systems", 7th edition, Addison Wesley,2014

Reference Books

1. S.K.Singh, "Database Systems: Concepts, Design Applications", 2nd edition, Pearson education, 2011.
2. Joe Fawcett, Danny Ayers, Liam R. E. Quin: "Beginning XML", Wiley India Private Limited5th Edition, 2012.
3. Thomas M. Connolly and Carolyn Begg "Database Systems: A Practical Approach to Design, Implementation, and Management", 6th edition, Pearson India, 2015.

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar

List of Challenging Experiments (Indicative)

1.	Model any given scenario into ER/EER Model using any tool ERD Plus, ER Win, Oracle SQL developer)	3 hours
2.	Creating applications with RDBMS Table creation with constraints, alter schema, insert values, aggregate functions, simple and complex queries with joins PLSQL-PROCEDURES, CURSORS, FUNCTIONS, TRIGGERS	3 hours
3.	Partition a given database based on the type of query and compares the execution speed of the query with/without parallelism.	3 hours
4.	Create an XML document and validate it against an XML Schema/DTD. Use XQuery to query and view the contents of the database.	3 hours
5.	Consider an application in which the results of football games are to be represented in XML,DTD and Xquery. For each game, we want to be able to represent the two teams involved ,which one was playing at home, which players scored goals(some of which may have been penalties)and the time when each was scored, and which players were shown yellow or red cards. You might use some attributes. You can check your solutions with the online demo of the Zorba XQueryengine4.	3 hours
6.	To implement parallel join and parallel sort algorithms to get marks from different colleges of the university and publish10 ranks for each discipline.	2 hours

7.	Create a distributed database scenario, insert values, fragment the database and query the database.	
8.	Consider a schema that contains the following table with the key underlined: Employee (Eno, Ename, Desg, Dno). Assume that we horizontally fragment the table as follows: Employee1(Eno;Ename;Desg;Dno), where 1¡= Dno ¡=10 Employee2(Eno;Ename; Desg; Dno), where 11 ¡= Dno ¡=20 Employee3 (Eno;Ename; Desg;Dno),where 21¡=Dno¡=30 In addition, assume we have 4 sites that contain the following fragments: Site1 has Employee1 Site2 has Employee2 Site3 has Employee2 and Employee3 Site4 has Employee1 Implement atleast 5 suitable queries on Employee fragments. Add relations to the database as per your requirements.	3 hours
9.	Download a spatial dataset based on any specific theme (containing layer information) from Quantum GIS and import it into Postgres SQL(PostGIS) and Query and view the database.	2 hours
10.	To investigation of some spatial analysis techniques using Toxic Release Inventory (www.epa.gov/triexplorer/) data for Massachusetts from the Environmental Protection Agency (EPA),which indicate the magnitude of the releases of toxic core chemicals into land, water and air at a site in the state. Note that these TRI locations were geo coded from a list of addresses provided by the EPA	3 hours
11.	Use sample datasets from health care domain, Visualize and interpret the results	3 hours
12.	Import the Hubway data intoNeo4jandconfigureNeo4j.Then, answer the following questions using the Cypher Query Language: a) List top 10 stations with most outbound trips (Show station name and number of trips) b) List top 10 stations with most inbound trips (Show station name and number of trips) c) List top 5 routes with most trips (Show starting station name, ending station name and number of trips) (4) List the hour number (for example 13 means 1pm -2pm) and number of trips which start from the station" B.U.Central" d)List the hour number(forexample13means1pm-2pm)and number of trips which end at the station "B.U. Central"	2 hours
	Total Laboratory Hours	**30 hours**

Mode of assessment: ***Project/Activity***			
Recommended by Board of Studies	**13.05.2016**		
Approved by Academic Council	**41**	**Date**	**17.06.2016**

CSE5007	Exploratory Data Analysis	L	T	P	J	C
		2	0	0	4	3
Pre-requisite	**Nil**	**Syllabus version**				
		1.0				

Course Objectives:

1. This course introduces the methods for data preparation and data understanding.
2. It covers essential exploratory techniques for understanding multivariate data by summarizing it through statistical methods and graphical methods.
3. Supports to Summarize the insurers use of predictive analytics, data science and Data Visualization

Expected Course Outcome:

1. Handle missing data in the real world data sets by choosing appropriate methods.
2. Summarize the data using basic statistics. Visualize the data using basic graphs and plots.
3. Identify the outliers if any in the data set.
4. Choose appropriate feature selection and dimensionality reduction
5. Techniques for handling multi-dimensional data

Module:1	**Introduction To Exploratory Data Analysis**	**3 hours**

Data Analytics lifecycle, Exploratory Data Analysis (EDA)– Definition, Motivation, Steps in data exploration, The basic data types Data Type Portability

Module:2	**Preprocessing-Traditional Methods and Maximum Likelihood Estimation**	**4 hours**

Introduction to Missing data, Traditional methods for dealing with missing data, Maximum Likelihood Estimation – Basics, Missing data handling, Improving the accuracy of analysis

Module:3	**Preprocessing Bayesian Estimation**	**4 hours**

Introduction to Bayesian Estimation ,Multiple Imputation-Imputation Phase, Analysis and Pooling Phase, Practical Issues in Multiple Imputation, Models for Missing Notation Random Data

Module:4	**Data Summarization & Visualization**	**4 hours**

Statistical data elaboration, 1-D Statistical data analysis, 2-D Statistical data Analysis, N- D Statistical data analysis

Module:5	**Outlier Analysis**	**3 hours**

Introduction, Extreme Value Analysis, Clustering based, Distance Based and Density Based outlier analysis, Outlier Detection in Categorical Data

Module:6	**Feature Subset Selection**	**4 hours**

Feature selection algorithms: filter methods, wrapper methods and embedded methods, Forward selection backward elimination, Relief, greedy selection, genetic algorithms for features selection

Module:7	**Dimensionality Reduction**	**6 hours**

Introduction, Principal Component Analysis(PCA), Kernel PCA, Canonical Correlation Analysis, Factor Analysis, Multi-dimensional scaling, Correspondence Analysis

Module:8	**Recent Trends**	**2 hours**

Recent Trends

	Total Lecture hours:	**30 hours**	

Text Book(s)	
Reference Books	
1	Charu C. Aggarwal, “Data Mining The Text book”, Springer, 2015.
2	Craig K. Enders, “Applied Missing Data Analysis”, The Guilford Press, 2010.
3.	Inge Koch, “Analysis of Multivariate and High dimensional data”, Cambridge University Press, 2014.
4.	Michael Jambu, “Exploratory and multivariate data analysis”, Academic Press Inc. 1990.
5.	Charu C. Aggarwal, “Data Classification Algorithms and Applications”, CRC press, 2015

Mode of assessment:			
Recommended by Board of Studies	**13-05-2016**		
Approved by Academic Council	**No. 41**	**Date**	**17-06-2016**

CSE6001	BIG DATA FRAMEWORKS	L	T	P	J	C
		2	0	2	4	4
Pre-requisite	**NIL**					**Syllabus version**
						1.0

Course Objectives:

1. To understand the need of Big Data, challenges and different analytical architectures
2. Installation and understanding of Hadoop Architecture and its ecosystems
3. Processing of Big Data with Advanced architectures like park.
4. Describe graphs and streaming data in Spark

Expected Course Outcome:

1. Discuss the challenges and their solutions in Big Data
2. Understand and work on Hadoop Framework and eco systems.
3. Explain and Analyse the Big Data using Map-reduce programming in Both Hadoop and Spark framework.
4. Demonstrate spark programming with different programming languages.
5. Demonstrate the graph algorithms and live streaming data in Spark
6. Analyse and implement different frame work tools by taking sample data sets.
7. Illustrate and implement the concepts by taking an application problem.

Module:1	**Introduction to Big Data**	**3 hours**

Data Storage and Analysis - Characteristics of Big Data – Big Data Analytics - Typical Analytical Architecture – Requirement for new analytical architecture – Challenges in Big Data Analytics – Need of big data frameworks

Module:2	**Hadoop Framework**	**6 hours**

Hadoop – Requirement of Hadoop Framework - Design principle of Hadoop –Comparison with other system - Hadoop Components – Hadoop 1 vs Hadoop 2 – Hadoop Daemon's – HDFS Commands – Map Reduce Programming: I/O formats, Map side join, Reduce Side Join, Secondary sorting, Pipelining MapReduce jobs

Module:3	**Hadoop Ecosystem**	**3 hours**

Introduction to Hadoop ecosystem technologies: Serialization: AVRO, Co-ordination: Zookeeper, Databases: HBase, Hive, Scripting language: Pig, Streaming: Flink, Storm.

Module:4	**Spark Framework**	**4 hours**

Overview of Spark – Hadoop vs Spark – Cluster Design – Cluster Management – performance,Application Programming interface (API): Spark Context, Resilient Distributed Datasets, Creating RDD, RDD Operations, Saving RDD - Lazy Operation – Spark Jobs.

Module:5	**Data Analysis with Spark Shell**	**4 hours**
Writing Spark Application - Spark Programming in Scala, Python, R, Java - Application Execution.		
Module:6	**Spark SQL and GraphX**	**5 hours**
SQL Context – Importing and Saving data – Data frames – using SQL – GraphX overview – Creating Graph – Graph Algorithms.		
Module:7	**Spark Streaming**	**3 hours**
Overview – Errors and Recovery – Streaming Source – Streaming live data with spark		
Module:8	**Recent Trends**	**2 hours**
	Total Lecture hours:	**30 hours**

Reference Books

1. Mike Frampton, "Mastering Apache Spark", Packt Publishing,2015.
2. TomWhite,"Hadoop:TheDefinitiveGuide",O'Reilly,4thEdition,2015.
3. Nick Pentreath, Machine Learning with Spark, Packt Publishing, 2015.
4. Mohammed Guller, Big Data Analytics with Spark, Apress,2015
5. Donald Miner, Adam Shook, "Map Reduce Design Pattern", O'Reilly, 2012

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar

List of Challenging Experiments (Indicative)		
1.	HDFS Commends Map Reduce Program to show the need of Combiner	4 hours
2.	Map Reduce I/O Formats-Text, key-value Map Reduce I/O Formats – Nline, Multiline	5 hours
3.	Sequence file Input/Output Formats Secondary sorting	5 hours
4.	Distributed Cache & Map Side Join, Reduce side Join Building and Running a Spark Application Word count in Hadoop and Spark Manipulating RDD	8 hours
5.	Inverted Indexing in Spark Sequence alignment problem in Spark Implementation of Matrix algorithms in Spark Spark Sql programming, Building Spark Streaming application	8 hours
	Total Laboratory Hours	**30 hours**

Mode of assessment: ***Project/Activity***

Recommended by Board of Studies	**13.05.2016**		
Approved by Academic Council	**No. 41**	**Date**	**17.06.2016**

CSE6005	MACHINE LEARNING	L	T	P	J	C
		2	0	2	4	4
Pre-requisite	**NIL**			**Syllabus version**		
				1.0		

Course Objectives:

1. Acquire theoretical Knowledge on setting hypothesis for pattern recognition
2. Apply suitable machine learning techniques for data handling and to gain knowledge from it
3. Evaluate the performance of algorithms and to provide solution for various real- world applications

Expected Course Outcome:

1. Recognize the characteristics of Machine Learning techniques that enable to solve real world problems
2. Recognize the characteristics of machine learning strategies
3. Apply various supervised learning methods to appropriate problems
4. Identify and integrate more than one techniques to enhance the performance of learning
5. Create probabilistic and unsupervised learning models for handling unknown pattern
6. Analyze the co-occurrence of data to find interesting frequent patterns

Module:1	**INTRODUCTION TO MACHINE LEARNING**	**3hours**

Introduction, Examples of Various Learning Paradigms, Perspectives and Issues, Version Spaces, Finite and Infinite Hypothesis Spaces, PAC Learning, VC Dimension.

Module:2	**Supervised Learning**	**9 hours**

Decision Trees: ID3, Classification and Regression Trees, Regression: Linear Regression, Multiple Linear Regression, Logistic Regression, Neural Networks: Introduction, Perceptron, Multilayer Perceptron, Support vector machines: Linear and Non-Linear, Kernel Functions, K-Nearest Neighbours

Module:3	**Ensemble Learning**	**3 hours**

Model Combination Schemes, Voting, Error-Correcting Output Codes, Bagging: Random Forest Trees, Boosting: Adaboost, Stacking

Module:4	**Unsupervised Learning**	**5hours**

Introduction to clustering, Hierarchical: AGNES, DIANA, Partitional: K-means clustering, K-Mode Clustering, Expectation Maximization, Gaussian Mixture Models

Module:5	**Probabilistic Learning**	**3 hours**

Bayesian Learning, Bayes Optimal Classifier, Naıve Bayes Classifier, Bayesian Belief Networks

Module:6	**Learning Association Rules**	**3hours**

Mining Frequent Patterns - basic concepts -Apriori algorithm, FP- Growth algorithm, Association-based Decision Trees

Module:7	**Machine Learning in Practice**	**2 hours**

Design, Analysis and Evaluation of Machine Learning Experiments, Other Issues: Handling imbalanced data sets			
Module:8	**Recent Trends in Big Data Analytics**		**2 hours**
	Total Lecture hours:	**30 hours**	

Text Book(s)

Reference Books

1. Ethem Alpaydin,"Introduction to Machine Learning", MIT Press, Prentice Hall of India, Third Edition2014.
2. Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, "Foundations of Machine Learning", MIT Press,2012.
3. Tom Mitchell, "Machine Learning", McGraw Hill, 3rd Edition, 1997.
4. Charu C. Aggarwal, "Data Classification Algorithms and Applications", CRC Press, 2014.
5. Charu C. Aggarwal, "DATA CLUSTERING Algorithms and Applications", CRC Press, 2014.
6. Kevin P. Murphy "Machine Learning: A Probabilistic Perspective", The MIT Press, 2012
7. Jiawei Hanand Micheline Kambers and Jian Pei, "Data Mining Concepts and Techniques", 3rd edition, Morgan Kaufman Publications, 2012.

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar

List of Challenging Experiments (Indicative)

1.	Implement Decision Tree learning	2 hours
2.	Implement Logistic Regression	2 hours
3.	Implement classification using Multilayer perceptron	2 hours
4.	Implement classification using SVM	2 hours
5.	Implement Adaboost	2 hours
6.	Implement Bagging using Random Forests	2 hours
7.	Implement K-means Clustering to Find Natural Patterns in Data	2 hours
8.	Implement Hierarchical clustering	2 hours
9.	Implement K-mode clustering	2 hours
10.	Implement Association Rule Mining using FP Growth	2 hours
11.	Classification based on association rules	2 hours
12.	Implement Gaussian Mixture Model Using the Expectation Maximization	2 hours
13.	Evaluating ML algorithm with balanced and unbalanced datasets	2 hours
14.	Comparison of Machine Learning algorithms	2 hours
15.	Implement k-nearest neighbours algorithm	2 hours

		Total Laboratory Hours	**30 hours**
Mode of assessment: ***Project/Activity***			
Recommended by Board of Studies	**13.05.2016**		
Approved by Academic Council	**No. 41**	**Date**	**17.06.2016**

CSE5002	OPERATING SYSTEMS AND VIRTUALIZATION	L	T	P	J	C
		2	0	2	0	3
Pre-requisite	**Nil**	**Syllabus version**				
		1.0				

Course Objectives:

1. To introduce Virtualization, operating systems fundamental concepts and its technologies.
2. To provides skills to write programs that interact with operating system components such as processes, thread, memory during concurrent execution.
3. To provide the skills and knowledge necessary to implement, provisioning and administer server and desktop virtualization.

Expected Course Outcomes:

Upon completion of the course, the students will be able to

1. Study operating system layers and kernel architectures.
2. Design various techniques for process management.
3. Construct various address translation mechanism.
4. Perform process threading and synchronization.
5. Study various methods of virtualization and perform desktop and server virtualization.
6. Classify the light-weight virtual machines with dockers and containers.
7. Develop programs related to the simulations of operating systems and virtualization concepts.

Module:1	**INTRODUCTION**	**2 hours**

History of OS - Computer system architecture a layered view with interfaces, Glenford Myer, Monolithic Linux Hybrid Windows 10 kernels Layered architecture of operating system and core functionalists.

Module:2	**PROCESS**	**5 hours**

Introduction, Process Operations, States, Context switching, Data Structures (Process Control Block (PCB), Process Scheduling: Multi-Level Feedback Queue, Multi-processor Scheduling, Deadlocks and its detection.

Module:3	**MEMORY**	**4 hours**

Introduction, Address Spaces, Memory API, Address Translation, Paging - Faster Translations (TLB), Smaller Tables. Virtual Memory System in x86.

Module:4	**CONCURRENCY**	**6 hours**

Introduction, Thread Models, Thread API, Building Evaluating a Lock, Test And Set, Classical problems handling using semaphore, Monitors, Persistence - File Organization: The i-node, Crash Consistency file security.

Module:5	**VIRTUAL MACHINES**	**2 hours**

Process and System VMs Taxonomy of VMs.

Module:6	**TYPES OF VIRTUALIZATION**	**4 hours**

Hardware Emulation, Full Virtualization with binary translation, Hardware assisted, Operating System Virtualization, OS assisted /Para virtualization.

Module:7	**HYPERVISOR**	**5 hours**

Type 1, Type 2, Para-virtualization, Server Virtualization, Desktop Virtualization, Overview VM portability - Clones, Templates, Snapshots, OVF, Hot and Cold Cloning Protecting Increasing Availability, Light Weight Virtual machine: Container / Docker.

Module:8	**RECENT TRENDS**	**2 hours**

	Total Lecture hours:	**30 hours**	

Text Book(s)

1.	Silberschatz, Abraham, Greg Gagne, and Peter B. Galvin, "*Operating system concepts"*, 10th Edition, Wiley Publishers, 2018.
2.	Matthew Portnoy, "*Virtualization Essentials*", John Wiley Sons Inc; 2nd Edition Edition, 2016.
Reference Books	
1.	Thomas Anderson, Michael Dahlin, "*Operating Systems: Principles and Practice*", 2nd Edition, Recursive Books, 2014.
2.	William Stallings, *"Operating Systems: Internals and Design Principles",* 8th Edition, 2014.
3.	Smith, Nair, *"Virtual Machines: Versatile Platforms for Systems and Processes"*, 1st Edition, Morgan Kaufmann Publishers, 2005.
Mode of Evaluation: CAT / Assignment / Quiz / FAT / LAB / Seminar	
List of Indicative Experiments	
1.	Study of Basic Linux Commands.
2.	Shell Programming (I/O, Decision making, Looping, Multi-level branching).
3.	Crating child process using fork() system call, Orphan and Zombie process creation.
4.	Simulation of CPU scheduling algorithms (FCFS, SJF, Priority and Round Robin).
5.	Simulation of Bankers algorithm to check weather given system is in safe state or not. Also check whether addition resource requested can be granted immediately.
6.	Parallel Thread management using pthread library. Implement a data parallelism using multi-threading.
7.	Dynamic memory allocation algorithms - first-fit, best-fit, worst-fit algorithms.
8.	Page Replacement Algorithms FIFO, LRU and Optimal.
9.	Virtualization Setup: Type-1, Type-2 Hypervisor.
10.	Implementation of OS / Server Virtualization.
Total Laboratory Hours	**30 hours**
Mode of assessment: CAT / Assignment / Quiz / FAT / Seminar	
Recommended by Board of Studies	**13-05-2016**
Approved by Academic Council	**No. 41** **Date** **17-06-2016**

CSE6006	NOSQL Databases	L	T	P	J	C
		2	0	2	4	4
Pre-requisite	**NIL**	**Syllabus version**				
		1.0				

Course Objectives:

1. Explore the origins of NoSQL databases and the characteristics that distinguish them from traditional relational database management systems.
2. Understand the architectures and common features of the main types of NoSQL databases (key-value stores, document databases, column-family stores, graph databases)
3. Discuss the criteria that decision makers should consider when choosing between relational and non-relational databases and techniques for selecting the NoSQL database that best addresses specific use cases.

Expected Course Outcome:

1. Explain the detailed architecture, Database properties and storage requirements
2. Differentiate and identify right database models for real time applications
3. Outline Key value architecture and characteristics
4. Design Schema and implement CRUD operations, distributed data operations
5. Compare data ware housing schemas and implement various column store internals
6. Choose and implement Advanced columnar data model functions for the real time applications
7. Develop Application with Graph Data model

Module:1	INTRODUCTION TO NOSQL CONCEPTS	4hours

Data base revolutions: First generation, second generation, third generation, Managing Transactions and Data Integrity, ACID and BASE for reliable database transactions, Speeding performance by strategic use of RAM, SSD, and disk, Achieving horizontal scalability with Data base sharding, Brewers CAP theorem.

Module:2	NOSQL DATA ARCHITECTURE PATTERNS	4 hours

NoSQL Data model: Aggregate Models- Document Data Model- Key-Value Data Model- Columnar Data Model, Graph Based Data Model Graph Data Model, NoSQL system ways to handle big data problems, Moving Queries to data, not data to the query, hash rings to distribute the data on clusters, replication to scale reads, Database distributed queries to Data nodes.

Module:3	KEY VALUE DATA STORES	5 hours

From array to key value databases, Essential features of key value Databases, Properties of

keys, Characteristics of Values, Key-Value Database Data Modeling Terms, Key-Value Architecture and implementation Terms, Designing Structured Values, Limitations of Key-Value Databases, Design Patterns for Key-Value Databases, Case Study: Key-Value Databases for Mobile Application Configuration

Module:4	**DOCUMENT ORIENTED DATABASE**	**4 hours**

Document, Collection, Naming, CRUD operation, querying, indexing, Replication, Sharding, Consistency Implementation: Distributed consistency, Eventual Consistency, Capped Collection, Case studies: document oriented database: Mongo DB and/or Cassandra

Module:5	**COLUMNAR DATA MODEL - I**	**4 hours**

Data warehousing schemas: Comparison of columnar and row-oriented storage, Column-store Architectures: C-Store and Vector-Wise, Column-store internals and, Inserts/updates/deletes, Indexing, Adaptive Indexing and Database Cracking.

Module:6	**COLUMNAR DATA MODEL – II**	**3hours**

Advanced techniques: Vectorized Processing, Compression, Write penalty, Operating Directly on Compressed Data Late Materialization Joins , Group-by, Aggregation and Arithmetic Operations, Case Studies

Module:7	**DATA MODELING WITH GRAPH**	**4 hours**

Comparison of Relational and Graph Modeling, Property Graph Model Graph Analytics: Link analysis algorithm- Web as a graph, Page Rank- Markov chain, page rank computation, Topic specific page rank (Page Ranking Computation techniques: iterative processing, Random walk distribution Querying Graphs: Introduction to Cypher, case study: Building a Graph Database Application- community detection

Module:8	**Recent trends**	**1 hours**

	Total Lecture hours:	**30 hours**	

Reference Books

1. Christopher D.manning, Prabhakar Raghavan, Hinrich Schutze, An introduction to Information Retrieval, Cambridge University Press
2. Daniel Abadi, Peter Boncz and Stavros Harizopoulas, The Design and Implementation of Modern Column-Oriented Database Systems, Now Publishers.
3. Guy Harrison, Next Generation Database: NoSQL and big data, Apress.

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar

List of Challenging Experiments (Indicative)

1.	ImporttheHubwaydataintoNeo4jandconfigureNeo4j.Then, answer the following questions using the Cypher Query Language: a) List top 10 stations with most outbound trips (Show station name and number of trips) b) List top10 stations with most in bound trips (Show station name and number of trips) c) List top 5 routes with most trips (Show starting station name, ending station name and number of trips) (4) List the hour number(forexample13means1pm-2pm) and number of trips which start	5 hours

	from the station ”B.U.Central” d) List the hour number(for example 13 means 1pm-2pm) and number of trips which end at the station ”B.U. Central”	
2.	Download a zip code dataset at http://media.mongodb.org/zips.json .Use mongo import to import the zip code dataset into Mongo DB. After importing the data, answer the following questions by using aggregation pipelines: (1) Find all the states that have a city called ”BOSTON”. Find all the states and cities whose names include the string ”BOST”. Each city has several zip codes. Find the city in each state with the most number of zip codes and rank those cities along with the states using the city populations. Mongo DB can query on spatial information.	5 hours
3.	Create a database that stores road cars. Cars have a manufacturer, a type. Each car has a maximum performance and a maximum torque value. Do the following: Test Cassandras replication schema and Consistency models.	5 hours
4.	Master Data Management using Neo4j Manage your master data more effectively The world of master data is changing. Data architects and application developers are swapping their relational databases with graph databases to store their master data. This switch enables them to use a data store optimized to discover new insights in existing data,providea360-degree view of master data and answer questions about data relationships in real time	5 hours
5.	Shopping Mall case study using cassendra, where we have many customers ordering items from themal land we have suppliers who deliver them their ordered items.	5 hours
Total Laboratory Hours		**30 hours**
Mode of assessment: ***Project/Activity***		
Recommended by Board of Studies	**13.05.2016**	
Approved by Academic Council	**No. 41**	**Date** **17.06.2016**

CSE6014	Programming for Data Science	L	T	P	J	C
		0	0	4	0	2
Pre-requisite	**NIL**	**Syllabus version**				
		1.0				

Course Objectives:

1. To provide necessary knowledge on how to manipulate data objects, produce graphics, analyse data using common statistical methods and generate reproducible statistical reports with programming in Python and R

Expected Course Outcome:

1. Ability to solve the analytical problems using Python and R
2. Develop competency in the Python programming language and a number of data- related Python libraries such as Pandas, Numpy, and Scipy
3. Ability to communicate results of analysis effectively using visualizations in Python and R
4. Import, export and manipulate data and produce statistical summaries of continuous and categorical data in Python and R
5. Ability to perform exploratory data analysis using Python and R

Module:1	Expressions, Operators, matrices, Decision Statements in python	**2 hours**
Module:2	Control Flow and Functions in python	**2 hours**
Module:3	Classes, Objects, Packages and Files in python	**2 hours**
Module:4	Tuple, Lists, Sequences, Dictionaries, Comprehensions	**2 hours**
Module:5	Numpy Arrays objects, Creating Arrays, basic operations, Indexing, Slicing and iterating, copying arrays, shape manipulation, Identity array, eye function, Universal function	**2 hours**
Module:6	Linear algebra with Numpy, eigen values and eigen vectors with Numpy	**2 hours**
Module:7	Aggregation and Joining, Pandas Object: Concatenating and appending data frames, index objects	**2 hours**
Module:8	Handling Time series data using pandas Handling missing values using pandas	**2 hours**
Module:9	Reading and writing the data including JSON data	**2 hours**

Module:10	Web scraping using python, Combining and merging	**2 hours**
Module:11	Datasets Data transformations Basic matplotlib plots, common plots used in statistical analysis in python	**3 hours**
Module:12	Common plots used in statistical analysis in python Datatypes in R2. Sequence generation, Vector and subscript, Random2 number generation in R Data frames and R functions2 Data manipulation and Data Reshaping using plyr, dplyr,2 reshape2 Parametric statistics and Non-parametric statistics2 Continuous and Discrete Probability distribution using R2	**2 hours**
Module:13	Correlation and covariance, contingency tables2 Overview of Sampling, different sampling techniques2 R and data base connectivity2	**3 hours**
Module:14	Web application development with R using Shiny2 Approaches to dealing with missing data in R2 Exploratory data analysis with simple visualizations using R 2 Feature or Attribute selection using R2 Dimensionality Reduction with R2 Time series data analysis with R2	**2 hours**
	Total Lecture hours: **30 hours**	

Reference Books

1. James Payne, "Beginning Python: Using Python 2.6 and Python 3.1" Wrox, Ist Edition, 2010
2. Michael T. Goodrich, Roberto Tamassia, Michael H. Gold wasser, "Data Structures and Algorithms in Python", John Wiley & sons, 2013.
3. Ivan Idris, "Python Data Analysis", Packt Publishing Limited, 2014
4. Wes McKinney, "Python for Data Analysis Data Wrangling with Pandas, NumPy, and IPython", O'Reilly Media, Ist Edition, 2012
5. Michael Heydt, "Learning Pandas - Python Data Discovery and Analysis Made Easy", Packt Publishing Limited, 2015.
6. Jacqueline Kazil, Katharine Jarmul, "Data Wrangling with Python: Tips and Tools to Make Your Life Easier", O'Reilly Media, Ist Edition, 2016.
7. https://docs.scipy.org/doc/numpy-dev/reference/index.html#reference
8. http://www.python-course.eu/numpy.php
9. Michael J. Crawley, "The R Book", Wiley, 2nd Edition, 2012.
10. Robert Kabacoff, "R in Action", Manning Publication, Ist Edition, 2011.
11. Torsten Hothorn, Brian S. Everitt, "A Handbook of Statistical Analyses Using R", Chapman and Hall_CRC, 2nd Edition, 2009.
12. Chris Beeley "Web Application Development with R Using Shiny", Pact Publishing, 2013.
13. Phil Spector, "Data Manipulation with R", Springer, 2008.

	14. Prabhanjan N. Tattar, Suresh Ramaiah, B. G. Manjunath, " A Course in Statistics with R", wiley, 2016 15. Pawel Cichosz, "Data Mining Algorithms: Explained Using R", wiley, 2014 16. Bater Makhabel, "Learning Data Mining with R", Packt Publication, 2015			
Mode of assessment: ***Project/Activity***				
Recommended by Board of Studies		13.05.2016		
Approved by Academic Council		**No. 41**	**Date**	17.05.2016

CSE6016	INFORMATION VISUALIZATION	L	T	P	J	C
		2	0	2	4	4
Pre-requisite	**NIL**			**Syllabus version**		
				1.0		

Course Objectives:

1. To understand the various types of data, apply and evaluate the principles of data visualization.
2. Acquire skills to apply visualization techniques to a problem and its associated dataset.
3. To apply structured approach to create effective visualizations.
4. To learn how to bring valuable insight from the massive dataset using visualization.
5. To learn how to build visualization dashboard to support decision making.
6. To create interactive visualization for better insight using various visualization tools.

Expected Course Outcome:

1. Identify the data types and its associated visualization mechanisms.
2. Apply the various scalar and vector visualization techniques to create suitable visualization for real life applications.
3. Handle and analyse multidimensional data and hierarchical data for visualization.
4. Perform multivariate data analysis and visualization.
5. Apply the visualization guidelines for effective information visualization.
6. Demonstrate the concept of visualization through dashboard creation for various applications.
7. Choose appropriate methods for the given real world problems and produce meaningful visualization.

Module:1	**Introduction to Data Visualization**	**4 hours**

Overview of data visualization - Data Abstraction - Task Abstraction - Analysis: Four Levels for Validation, Human Visual Perception

Module:2	**Visualization Techniques – I**	**3 hours**

Scalar and point techniques – vector visualization techniques – matrix visualization

Module:3	**Visualization Techniques – II**	**6 hours**

Visualization Techniques for Trees, Graphs, and Networks, Multidimensional data

Module:4	**Visual Analysis of data from various domains – I**	**5 hours**

Time-oriented data visualization – Spatial data visualization and case studies

Module:5	**Visual Analysis of data from various domains – II**	**5 hours**

Text data visualization – Multivariate data visualization, and case studies

Module:6	**Designing Effective Visualizations**	**2 hours**

Guidelines for designing successful visualizations, Data visualization dos and don’ts

Module:7	**Dashboard Creation and Visual Story Telling**	**3 hours**

Dashboard Design principles, Effective Dashboard Display Media, Dashboard creation using visualization tools for the use cases: Finance- marketing-insurance-healthcare etc.,			
Module:8	**Recent Trends**		**2 hours**
	Total Lecture hours:	**30 hours**	

Reference Books

1. Tamara Munzer, "Visualization Analysis and Design", CRC Press, 2014.
2. Stephen Few, "Now You See It", Analytics Press, 2009.
3. Stephen Few, "Information Dashboard Design: the effective visual communication of data", Oreilly, 2006.
4. Matthew O. Ward, Georges Grinstein, Daniel Keim, "Interactive Data Visualization: Foundations, Techniques, and Applications", CRC Press, Second Edition, 2015.
5. Dr. Chun-hauh Chen, W.K.Hardle, A. Unwin, "Handbook of Data Visualization", Springer publication, 2008.
6. Ben Fry, "Visualizing Data", O'Reilly Media, 2008
7. Winston Chang, "R Graphics Cookbook", O'Reilly, 2012.
8. http://www.fusioncharts.com/whitepapers/

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar

List of Challenging Experiments (Indicative)		
1.	Association Rule Mining and Clustering using R	3 hours
2.	Visualization on KNN or Naïve Bayes Classification using R	3 hours
3.	Financial analysis using Clustering, Histogram and HeatMap	2 hours
4.	Time-series analysis –stock market	2 hours
5.	Visualization of various massive dataset-Finance-Healthcare- Census -Geospatial	2 hours
6.	Market-Basket Data analysis-visualization	2 hours
7.	Text visualization using web analytics	2 hours
8.	Hadoop and R integration in Table au using Hortonworks	2 hours
9.	Google API with maps	2 hours
10.	VisualizationusingD3.js	2 hours
11.	Visualization using Zeppelin	2 hours

12.	Network Visualization using Gephi			2 hours
13.	Visualization of reconstruction network using Qlickview			2 hours
14.	Dash Board Creation using Tableau			2 hours
			Total Laboratory Hours	**30 hours**
Mode of assessment: ***Project/Activity***				
Recommended by Board of Studies	**13.05.2016**			
Approved by Academic Council	**No. 41**	**Date**	**17.06.2016**	

CSE6017	MINING MASSIVE DATA	L	T	P	J	C
		2	0	2	4	4
Pre-requisite	**Nil**	**Syllabus version**				
		1.0				

Course Objectives:

1. To provide comprehensive knowledge on developing and applying machine learning algorithms for massive real-world datasets in distributed frameworks.
2. To demonstrate the use of big data analytics tools like Spark and Mahout for mining massive datasets.
3. To impart in depth knowledge on Deep Learning and Extreme Learning concepts.

Expected Course Outcome:

1.Identify right machine learning / mining algorithm for handling massive data
2.Apply classification and regression models with Spark and Mahout
3. Implement clustering models using Spark and Mahout
4.Mine social Network graphs using MapReduce
5.Apply semi supervised learning for clustering and classification
6.Use deep learning to solve real-life problem
7. Use Extreme Learning Machine for classification and regression.
8. Use big data analytics tools such as Spark, Mahout and H2O in solving problems based on Machine learning

Module:1	**MapReduce Based Machine Learning**	**7 hours**

K-Means, PLANET, Parallel SVM, Association Rule Mining in MapReduce, Inverted Index, Page Ranking, Expectation Maximization, Bayesian Networks

Module:2	**Classification and Regression models with Spark and Mahout**	**5 hours**

Linear support vector machines - Naive Bayes model- Decision Trees – Least square regression- Decision trees for regression.

Module:3	Clustering in Spark and Mahout	**4 hours**

Hierarchical Clustering in a Euclidean and Non-Euclidean Space - The Algorithm of Bradley, Fayyad, and Reina - A variant of K-means algorithm - Processing Data in BFR Algorithm CURE algorithm - Clustering models with Spark - Spectral clustering using Mahout

Module:4	**Mining Social-Network Graphs**	**3 hours**

Clustering of Social-Network Graphs - Direct Discovery of Communities - Partitioning of Graphs Finding Overlapping Communities - Counting Triangles using MapReduce Neighborhood Properties of Graphs

Module:5	**Semi-Supervised Learning**	**3 hours**

Introduction to Semi-Supervised Learning, Semi-Supervised Clustering, Transductive Support Vector Machines

Module:6	**Deep Learning**	**4 hours**

Introduction, Deep Neural Networks, Deep Belief Networks, Auto Encoders, Recurrent Networks

Module:7	**Extreme Learning**	**2 hours**
Extreme Learning Machines (ELM), ELM auto encoder, Extreme Support Vector Regression		
Module:8	**Recent Trends:**	**2 hours**

	Total Lecture hours:	**30 hours**	

Text Book(s)

1.Jure Leskovec, Anand Rajaraman, Jeffrey Ullman, "Mining of Massive Datasets", Standford Press,2011.
2. Nick Pentreath, "Machine Learning with Spark", Packt Publishing,
3. Olivier Chapelle, Bernhard Scholkopf, Alexander Zien "Semi-Supervised Learning", The MIT Press, 2006.

Reference Books

1. Ron Bekkerman, Mikhail Bilenko, John Langford "Scaling Up Machine Learning: Parallel and
Distributed Approaches", Cambridge University Press, 2012.
2. Jimmy Lin, Chris Dyer, "Data-Intensive Text Processing with MapReduce", Morgan Claypool Publishers, 2010.
3. Hennessy, J.L. and Patterson, D.A., 2011. Computer architecture: a quantitative approach. Elsevier.
4. Chandramani Tiwary "Learning Apache Mahout", Packt Publishing, 2015.
5. Fuchen Sun, Kar-Ann Toh, Manuel Grana Romay, KezhiMao,"Extreme Learning Machines 2013: Algorithms and Applications", Springer, 2014.

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar

List of Challenging Experiments (Indicative)		
1.	K-means implementation in MapReduce	2 hours
2.	Association Rule Mining with MapReduce	2 hours
3.	Decision trees in Spark	2 hours
4.	Nave bayes classification using Spark	2 hours
5.	Advanced text processing with Spark	2 hours
6.	Clustering models with Spark	2 hours
7.	Building a recommendation engine with Spark	2 hours
8.	Representing social-network data using Graphs	2 hours
9.	Implementing Semi-supervised Clustering	2 hours
10.	Deep Learning using H2O	2 hours
11.	Predictive analysis using H2O tool	2 hours
12.	SVM Classification using Mahout	2 hours
13.	Spectral clustering using Mahout	2 hours
14.	Building a recommendation engine with Sparkling water	2 hours
15.	Deep Learning using DL4J	2 hours
	Total Laboratory Hours	**30 hours**

Mode of assessment:			
Recommended by Board of Studies	**13-05-2016**		
Approved by Academic Council	**No. 41**	**Date**	**17-06-2016**

CSE6018	Streaming Data Analytics	L	T	P	J	C
		2	0	2	4	4
Pre-requisite	**Nil**			**Syllabus version**		
				1.0		

Course Objectives:

1. It introduces theoretical foundations, algorithms, methodologies, and applications of streaming data and also provide practical knowledge for handling and analyzing streaming data.

Expected Course Outcome:

1. Recognize the characteristics of data streams that make it useful to solve real-world problems.
2. Identify and apply appropriate algorithms for analyzing the data streams for variety of problems.
3. Implement different algorithms for analyzing the data streams
4. Identify the metrics and procedures to evaluate a model

Module:1	Introduction	2 hours

Characteristics of the data streams, Challenges in mining data streams Requirements and principles for real time processing, Concept drift Incremental learning.

Module:2	Data Streams	5 hours

Basic Streaming Methods, Counting the Number of Occurrence of the Elements in a Stream, Counting the Number of Distinct Values in a Stream, Bounds of Random Variables, Poisson Processes, Maintaining Simple Statistics from Data Streams, Sliding Windows, Data Synopsis,

Change Detection: Tracking Drifting Concepts, Monitoring the Learning Process

Module:3	Decision Trees	4 hours

Very Fast Decision Tree Algorithm (VFDT), The Base Algorithm, Analysis of the VFDT Algorithm, Extensions to the Basic Algorithm: Processing Continuous Attributes, Functional Tree Leaves, Concept Drift.

Module:4	Clustering from Data Streams	5 hours

Clustering Examples: Basic Concepts, Partitioning Clustering - The Leader Algorithm, Single Pass k-Means, Micro Clustering, Clustering Variables: A Hierarchical Approach

Module:5	Frequent Pattern Mining	4 hours

Mining Frequent Item sets from Data Streams- Landmark Windows, Mining Recent Frequent Item sets, Frequent Item sets at Multiple Time Granularities

Sequence Pattern Mining- Reservoir Sampling for Sequential Pattern Mining over data streams

Module: 6	**Evaluating Streaming Algorithms**	**4 hours**
Evaluation Issues, Design of Evaluation Experiments, Evaluation Metrics, Error Estimators using a Single Algorithm and a Single Dataset, Comparative Assessment, The 0-1 loss function, Evaluation Methodology in Non-Stationary Environments, The Page-Hinkley Algorithm.		
Module: 7	**Complex Event Processing**	**4 hours**
Introduction to Complex Event Processing, Features of CEP, Need for CEP, CEP Architectural Layers, Scaling CEP, Events, Timing and Causality, Event Patterns, Rules and Constraint, STRAW-EPL, Complex Events and Event Hierarchies		
Module: 8	**RECENT TRENDS**	**2 hours**
	Total Lecture hours:	**30 hours**

Text Book(s)

Reference Books

1. Joao Gama, "Knowledge Discovery from Data Streams", CRC Press, 2010.
2. David Luckham, "The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems", Addison Wesley, 2002.
3. Charu C. Aggarwal, "Data Streams: Models And Algorithms", Kluwer Academic Publishers, 2007

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar

List of Challenging Experiments (Indicative)	
1. Exploring one stream processing engine like storm or STREAM etc (**2 classes**) 2. Implementation of algorithms for example : VFDT, CVFDT(**2 classes**) 3. Implementation of Clustering 4. Implementation of Frequent pattern mining 5. Exploring one CEP engine like ESPER or DROOLS(**2 classes**) 6. Exercise with continuous queries Logical operations on single stream 7. Exercise with continuous queries Logical operations on multiple streams 8. Exercise with continuous queries temporal operators on single stream 9. Exercise with continuous queries temporal operators on multiple streams 10. Exercise with complex continuous queries with logical, relational & temporal operators on multiple streams	3 hours each
Total Laboratory Hours	**30 hours**

Mode of assessment:

Recommended by Board of Studies	**13.05.2016**		
Approved by Academic Council	**No. 41**	**Date**	**17.06.2016**

CSE6019	Text, Web and Social Media Analytics	L	T	P	J	C
		3	0	0	4	4
Pre-requisite	**Nil**	**Syllabus version**				
		1.0				

Course Objectives:

1. To provide an overview of common text mining and social media data analytic activities.
2. To understand the complexities of processing text and network data from different data sources.
3. To enable students to solve complex real-world problems for sentiment analysis and Recommendation systems.

Expected Course Outcome:

1. Interpret the terminologies, metaphors and perspectives of social media analytics.
2. Apply a wide range of classification, clustering, estimation and prediction algorithms on Textual data.
3. Perform social network analysis to identify important social actors, subgroups and network properties in social media sites.
4. Apply state of the art web mining tools and libraries on realistic data sets as a basis for business decisions and applications.
5. Provide solutions to the emerging problems with social media such as behavior analytics and recommendation systems.
6. Design new solutions to opinion extraction, sentiment classification and data summarization problems.

Module:1	**Introduction to Text Mining**	**6 hours**

Text Representation- tokenization, stemming, stop words, TF-IDF, Feature Vector Representation, NER, N-gram modeling.

Module:2	**Mining Textual Data**	**6 hours**

Text Clustering, Text Classification, Topic Modeling-LDA, HDP

Module:3	**Introduction to Web-Mining**	**6 hours**

Inverted indices and Boolean queries. PLSI, Query optimization, page ranking.

Module:4	**Web Usage Web content Mining**	**7 hours**

Web Crawling-Crawler Algorithms, Implementation Issues, Evaluation, Session & visitor Analysis, Visitor Segmentation, Analysis of Sequential & Navigational Patterns, Predictions based on web user transactions.

Module:5	**Introduction to Social Media Network**	**6 hours**

Essentials of Social graphs, Social Networks, Models, Information Diffusion in Social Media.

Module:6	**Mining Social Media**	**6 hours**

Behavioral Analytics, Influence and Homophily, Recommendation in Social Media.

Module:7	**Sentimental Mining**	**6 Hours**

Sentiment classification feature based opinion mining, comparative sentence and relational mining, Opinion spam.

Module:8	**Recent Threads**	**2 hours**
	Total Lecture hours:	**45 hours**

Reference Books			
1. Bing Liu, "Web Data Mining-Exploring Hyperlinks, Contents, and Usage Data", Springer, Second Edition, 2011. 2. Reza Zafarani, Mohammad Ali Abbasi and Huan Liu, "Social Media Mining - An Introduction", Cambridge University Press, 2014. 3. Bing Liu, "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, 2012. 4. Nitin Indurkhya, Fred J Damerau, "Handbook of Natural Language Process", 2nd Edition, CRC Press, 2010. 5. Matthew A.Russell, "Mining the social web", 2nd edition- O'Reilly Media, 2013.			
Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar			
Recommended by Board of Studies	**13-05-2016**		
Approved by Academic Council	**No. 41**	**Date**	**17-06-2016**

CSE6020	BIG DATA TECHNOLOGIES	L	T	P	J	C
		2	0	2	4	4
Pre-requisite	**NIL**	**Syllabus version**				
		1.0				

Course Objectives:

1. To have knowledge on accessing, storing and manipulating the huge data from different resources.
2. To understand the working environment of Pig and Hive for processing the structured and unstructured data.
3. To differentiate the RDBMS and Hive architectures and implement queries to process the data using sqoop.
4. To have a knowledge on searching mechanisms using solr.

Expected Course Outcome:

1. Illustrate the usage of data on different Big data ecosystems.
2. Demonstrate the Pig architecture and evaluation of pig scripts.
3. Describe the Hive architecture and execute SQL queries on sample data sets.
4. Understand the process of transferring data between different file systems and to execute operations using sqoop.
5. Understand the concepts of indexing and use these concepts in solr search engine.
6. Implement and evaluate the data manipulation procedures using pig, hive, sqoop and solr.
7. Develop an application using different eco system tools by taking standard sample data set.

Module:1	**Introduction**	**4 hours**

Big data- Concepts, Needs and Challenges of big data. Types and source of big data. Components of Hadoop Eco System- Data Access and storage, Data Intelligence, Data Integration, Data Serialization, Monitoring, Indexing.

Module:2	**Apache Pig**	**6 hours**

Introduction, Parallel processing using Pig, Pig Architecture, Grunt, Pig Data Model-scalar and complex types. Pig Latin- Input and output, Relational operators, User defined functions. Working with scripts.

Module:3	**Apache Hive Fundamentals**	**3 hours**

Introduction-Hive modules, Data types and file formats, Hive QL-Data Definition and Data Manipulation.

Module:4	**Apache Hive Advanced Concepts**	**4 hours**

Hive QL queries, Hive QL views- reduce query complexity. Hive scripts. Hive QL Indexes-create, Show drop. Aggregate functions. Bucketing vs Partitioning.

Module:5	**Importing and Handling Relational Data in Hadoop using Sqoop**	**3 hours**

Relational database management in Hadoop: Bi directional data transfer between Hadoop and external database. Import data- Transfer an entire table, import subset data, use different file format. Incremental import new data, incrementally import data, preserving the value

Module:6	**Sqoop**	**4 hours**
Export transfer data from Hadoop, update the data, update at the same time, export subset of columns. Hadoop ecosystem integration- import data to hive, using partitioned hive tables, replace special delimiters.		
Module:7	**Solr**	**4 hours**
Introduction. Information retrieval search engine, categories of data, inverted index. Design- field attributes and types. Indexing- indexing tool. Indexing operations using CSV documents. Searching data- parameters, default query.		
Module:8	**Recent Trends**	**2 hours**
	Total Lecture hours:	**30 Hours**

Reference Books

1. Alan Gates, Programming Pig Data flow Scripting with Hadoop, O'Reilly Media, Inc,2011.
2. Jason Rutherglen, Dean Wampler, Edward Caprialo, Programming Hive, O'Reilly Media Inc,2012
3. Kathleen Ting, Jarek Jarcec Cecho, Apache Sqoop Cook book, O'Reilly Media Inc, 2013.
4. Dikshant Shahi, Apache Solr: A Practical approach to enterprise search, A press, 2015.
5. Chuck Lam, Hadoop in Action, Manning Publications, 2010.
6. Andrea Gazzarini, Apache Solr Essentials, PACKT Publications, 2015.

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar

List of Challenging Experiments (Indicative)		
1.	Implement a program using Pig Latin operators and user defined functions Implement a program using operators and Pig Latin scripts Program using Hive manipulation and data definition languages. Implement a program using Hive queries with partitioning.	6 hours
2.	Implement a program using Hive indexes. Implement a program using Hive views Implement a program using Hive external table by accessing the external file created by Pig or any other tool. Program using Hive scripts and aggregate functions	7 hours
3.	Implement a program using Hive queries with bucketing and clustering. Implement a program for data transfer between Hadoop and external Data base using sqoop. Program to import data and incremental data in sqoop.	6 hours
4.	Program to preserve the value in sqoop Program to export data from Hadoop using sqoop Program to import data to hive and using partitioned hive tables	6 hours
5.	Program for inverted index using solr Program for indexing operations using CSV files in solr. Program to search data using solr	5 hours
	Total Laboratory Hours	**30 hours**

Mode of assessment: ***Project/Activity***

Recommended by Board of Studies	13.05.2016		
Approved by Academic Council	**No. 41**	**Date**	17.06.2016

CSE6021	Domain Specific Predictive Analytics	L	T	P	J	C
		3	0	0	4	4
Pre-requisite	**NIL**			**Syllabus version**		
				1.0		

Course Objectives:

1. It introduces theoretical foundations, algorithms, methodologies for analysing data in various domains such Retail, Finance, Risk and Healthcare.

Expected Course Outcome:

1. Recognize challenges in dealing with data sets in domains such as finance, risk and healthcare.
2. Identify real-world applications of machine learning in domains such as finance, risk and healthcare.
3. Identify and apply appropriate algorithms for analyzing the data for variety of problems in finance, risk and healthcare.
4. Make choices for a model for new machine learning tasks based on reasoned argument

Module:1	**Retail Analytics**	**7 hours**

Understanding Customer: Profiling and Segmentation, Modelling Churn. Modelling Lifetime Value, Modelling Risk, Market Basket Analysis.

Module:2	**Risk Analytics**	**5 hours**

Risk Management and Operational Hedging: An Overview, Supply Chain Risk Management, A Bayesian Framework for Supply Chain Risk Management, Credit Scoring and Bankruptcy Prediction

Module:3	**Financial Data Analytics**	**5 hours**

Financial News analytics: Framework, techniques, and metrics, News events impact market sentiment, Relating news analytics to stock returns

Module:4	**Financial Time Series Analytics**	**6 hours**

Financial Time Series and Their Characteristics, Common Financial Time Series models, Autoregressive models, Markov chain models, Time series models with leading indicators, Long term forecasting

Module:5	**Health care Analytics**	**6 hours**

Introduction to Healthcare Data Analytics, Electronic Health Records, Privacy-Preserving Data Publishing Methods in Healthcare, Clinical Decision Support Systems

Module:6	**Healthcare Data Analytics**	**7 hours**

Natural Language Processing and Data Mining for Clinical Text: Core NLP Components, Information Extraction and Named Entity Recognition, Social Media Analytics for Healthcare: Tracking of Infectious Disease Outbreaks, Readmission risk Prediction

Module:7	**Genomic Data Analytics**	**7 hours**

Microarray Data, Microarray Data Analysis, Genomic Data Analysis for Personalized Medicine, Patient Survival Prediction from Gene Expression Data, Genome Sequence Analysis			
Module:8	**RECENT TRENDS**		**2 hours**
	Total Lecture hours:	**45 hours**	
Text Book(s)			
Reference Books			

1. Chris Chapman, Elea McDonnell Feit "R for Marketing Research and Analytics", Springer, 2015.
2. Olivia Parr Rud “Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management”, Wiley, 2001.
3. Chandan K. Reddy, Charu C. Aggarwal "Healthcare Data Analytics", CRC Press, 2015.
4. Rene Carmona "Statistical Analysis of Financial Data in R", Springer, 2014.
5. James B. Ayers “Handbook Of Supply Chain Management” Auerbach Publications, 2006.
6. Panos Kouvelis, Ling xiu Dong, Onur Boyabatli, Rong Li "The Handbook of Integrated Risk Management in Global Supply Chains", Wiley, 2012.

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar			
Mode of assessment:			
Recommended by Board of Studies	**13.05.2016**		
Approved by Academic Council	**No. 41**	**Date**	**17.06.2016**

CSE6022	Soft Computing	L	T	P	J	C
		2	0	2	4	4
Pre-requisite	**NIL**			**Syllabus version**		
				1.0		

Course Objectives:

The objective of this course is to introduce methods for handling imprecise and uncertain data using Rough sets, Neuro Fuzzy Systems and foster their abilities in designing and implementing optimal solutions for real-world and engineering problems using derivative free optimization techniques.

Expected Course Outcome:

After successfully completing the course the student should be able to

1. Have a general understanding of soft computing methodologies, to deal with imprecise and uncertain data
2. Develop computational neural network models for some simple biological systems;
3. Develop fuzzy models for engineering systems, particularly for control systems;
4. Apply derivative free optimization methods to solve real world problems
5. Demonstrate some applications of computational intelligence

Module:1	**Introduction to Soft Computing**	**3 hours**

Soft Computing Overview – Uncertainty in data, Hard vs Soft Computing

Module:2	**Neural Networks**	**7 hours**

Introduction, RBF Networks, Self-Organizing Map, Boltzmann Machines, Convolutional Neural Networks

Module:3	**Fuzzy Systems**	**3 hours**

Fuzzy Sets, Fuzzy Relations, and Membership functions, Properties of Membership functions, Fuzzification and Defuzzification.

Module:4	**Fuzzy logic**	**4 hours**

Fuzzy Rule based systems, Fuzzy Decision making, Fuzzy Classification, Fuzzy C-Means Clustering.

Module:5	**Rough Sets**	**3 hours**

Rough Sets – Definition, Upper and Lower Approximations, Boundary Region, Decision Tables and Decision Algorithms. Properties of Rough Sets. Rough K-means clustering, Rough support vector clustering

Module:6	**Optimization Techniques**	**4 hours**
Introduction, Genetic Algorithm, Memetic Algorithms, Particle Swarm Optimization, Ant Colony Optimization, Frog-Leaping.		
Module:7	**Hybrid Systems**	**4 hours**
GA Based Back Propagation Networks, Fuzzy Back Propagation Networks, Evolutionary Ensembles		
Module:8	**Recent Trends**	**2 hours**
	Total Lecture hours:	**30 hours**

Reference Books

1. S.N. Sivanandham and S.N.Deepa, “Principles of Soft Computing”, 2nd Edition, Wiley Publications.
2. Andries P. Engelbrecht, "Computational Intelligence: An Introduction", John Wiley & Sons,2007
3. Laurene V. Fausett “Fundamentals of Neural Networks: Architectures, Algorithms And Applications”, Pearson,1993
4. Simon Haykin "Neural Networks and Learning Machines" Prentice Hall, 2008.
5. Timothy Ross, “Fuzzy Logic with Engineering Applications”, Third Edition, Wiley,

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar

List of Challenging Experiments (Indicative)

Project

Generally a team project consists of four to six members

Down to earth application and innovative idea should have been attempted # Report in Digital format with all drawings using software package to be submitted.

Assessment on a continuous basis with a min of 3 reviews.

The following is the sample project that can be given to students to be implemented in any programming languages.

- Develop Fuzzy Decision-Making for Job Assignment Problem
- Implement TSP using Optimization Techniques
- Develop a suitable method for Health Care Application using Neuro- Fuzzy systems
- Develop a suitable method for Face Recognition System
- Layout Optimization using Genetic Algorithms
- Fault Diagnosis using rough set theory
- Software safety analysis using rough sets

A Neuro-Fuzzy Approach to Bad Debt Recovery in Healthcare

Total Laboratory Hours	**30 hours**

Mode of assessment: ***Project/Activity***

Recommended by Board of Studies	**13.05.2016**		
Approved by Academic Council	**No. 41**	**Date**	**17.06.2016**

CSE6023	Cloud Computing Fundamentals	L	T	P	J	C
		2	0	2	4	4
Pre-requisite	**Nil**	**Syllabus version**				
		1.0				

Course Objectives:

1. To provide students with the fundamentals and essentials of Cloud Computing.
2. To provide students a sound foundation of the Cloud computing so that they are able to start using and adopting Cloud Computing services and tools in their real life scenarios.
3. To enable students exploring some important cloud computing driven commercial systems such as Google Apps, Microsoft Azure and Amazon Web Services and other businesses cloud applications.
4. To impart knowledge in applications of cloud computing

Expected Course Outcome:

1. Design, Develop & Demonstrate real-world applications from the Cloud Computing
2. Understand the subtle architectural difference in Public and Private Clouds.
3. Appreciate the requirements of various service paradigms in Cloud Computing.
4. Describe the methods of processing multimedia elements and other information presentation concepts during multimedia communications.

Module:1	**Introduction to Cloud Computing**	**4 hours**

Cloud Computing Overview: Characteristics – challenges, benefits, limitations, Evolution of Cloud Computing, Cloud computing architecture, Cloud Reference Model (NIST Architecture)

Module:2	**Infrastructure as a Service**	**4 hours**

Service Model, Characteristics, Benefits, Enabling Technologies Case Study : AWS, OpenStack

Module:3	**Platform as a Service**	**4 hours**

Service Model, Characteristics, Benefits, Enabling Technologies Case Studies : IBM Bluemix, GAE, Microsoft Azure

Module:4	**Software as a Service**	**4 hours**

Service Model, Characteristics, Benefits, Enabling Technologies Case Study : Salesforce.com, CRM, Online Collaboration Services

Module:5	**Data Analytics as a Service**	**4 hours**

Hadoop as a service, MapReduce on Cloud, Chubby locking Service

Module:6	**Introduction to Public and Private Clouds**	**5 hours**

Shared Resources – Resource Pool – Usage and Administration Portal – Usage Monitor – Resource Management– Cloud Security – Workload Distribution – Dynamic provisioning.

Module:7	**Storage as a service**	**3 hours**

Historical Perspective, Datacenter Components, Design Considerations, Power Calculations, Evolution of Data Centers, Cloud data storage – CloudTM			
Module:8	**Recent Trends**		**2 hours**
	Total Lecture hours:	**30 hours**	
Text Book(s)			
Reference Books			

1) Kai Hwang, Geoffrey Fox, Jack J. Dongarra, Morgan Kaufmann, "Distributed and Cloud Computing: From Parallel Processing to the Internet of Things," 1st Edition, 2011.
2) Gautham Shroff, "Enterprise Cloud Computing: Technology, Architecture, Applications", Cambridge press, 2010.
3) Kris Jamsa, "Cloud Computing", Jones & Barlett Learning, 2013.
4) Rajkumar Buyya, James Broberg, Andrzej Goscinski, "Cloud Computing Principles and Paradigms", John Wiley & Sons, 2011.
5) John Rhoton and Risto Haukiojal, "Cloud Computing Architectured: Solution Design Handbook", Recursive Press, 2013.
6) George Recse, "Cloud Application Architectures: Building Application and Infrastructure in the Cloud" , O' Reilly Media, First Edition, 2009.
7) Dinkar Sitaram, Geetha Manjunathan, "Moving to the Cloud: Developing Apps in the new world of Cloud Computing", Syngress, 2012.
8) Samee. U. Khan, Albert. Y. Zomaya, "Handbook on Data Centers", Springer, 2015.

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar		
List of Challenging Experiments (Indicative)		
	1) Cisco simulator – VLAN design, Routing, Sub netting, Gateway configuration 2) Virtual box based Webserver creation, Images/Snapshots access webpage from 2nd VM on another subnet work 3) EC2 AWS – S3 bucket based static webpages. 4) EC2 AWS – Instance Creation, Migration 5) EC2 AWS – Web application using Beanstalk. 6) AWS – Local balancing and auto scaling. 7) IBM Blue Mix - Mobile Application development 8) DaaS – Deployment of a basic web app and add additional functionality(Java scripts based) 9) PaaS – IOT – Mobile sensor based IOT application hosted via PaaS environment 10) SaaS – Deployment of any SaaS application for a online collaborative tool 11) Deployment of Open stack or Virtual box from the scratch 12) Automating Open stack deployment using Chef/Puppet configuration for 4 node/ 5 node/ HA clusters 13) Hadoop as a Service 14) Cloud TM 15) Online Collaboration Services (User Defined Applications)	30 Hours

Total Laboratory 30Hours		
Mode of assessment:		
Recommended by Board of Studies	**13-05-2016**	
Approved by Academic Council	**No. 41**	**Date** **17-06-2016**

CSE6025	Analytics of Things	L	T	P	J	C
		3	0	0	4	4
Pre-requisite	Nil	Syllabus version				
		1.0				

Course Objectives:

1. To introduce the technology that enables IoT, application of IoT, cloud support for IoT and access data using mobile computing devices. This will serve as foundation for the cyber physical systems, Internet of services leading to Industry 4.0 changes.

Expected Course Outcome:

1. Identify the technologies that enables IoT.
2. Able to use Hardware and software required to design and build IoT
3. Develop programs for interfacing with sensors and actuators and other IoT devices
4. Set up the servers to upload IoT data to cloud for further analysis

Module:1	Introduction to IoT	6 hours

Introduction, Characteristics of IoT, Difference between IoT and M2M, Applications of IoT, Physical and logical design of IoT, IoT levels and deployment templates, IoT enabling technologies: Wireless Sensor Networks, RFID, GPS

Module:2	IOT Hardware platforms	9 hours

Overview of IoT supported Hardware Platforms: Raspberry pi, Arduino, Intel Galileo

Module:3	Communication in IOT	5 hours

Interface protocol, Serial, SPI, I2C, 6LoWPAN, 802.11wifi, 802.15 Bluetooth, 802.15.4 Zigbee, RTLS, GPS, CoAp – Constrained application protocol, RPL – routing protocol for lossy networks.

Module:4	IOT Software development	7 hours

Linux, Networking configurations in Linux, Accessing Hardware & Device Files interactions, Python packages: JSON, XML, HTTPLib, URLLib, SMTPLib, XMPP, Contiki OS

Module:5	IoT Physical Servers & Cloud Offerings	6 hours

Introduction to Cloud Storage Models & Communication APIs, Cloud of things, Xively Cloud for IOT, PHP & MySQL for data processing, WAMP, Designing a RESTful Web API, MQTT, Amazon Web Services for IoT

Module:6	Data Analytics for IoT	5 hours

Configuring and using Apache Storm for Real-time Data Analysis

Module:7	Case Studies illustrating IoT Design	5 hours

Smart Home, Smart Parking, weather reporting and monitoring

Module:8	Recent Trends	2 hours

	Total Lecture hours:	**45 hours**

Text Book(s)

Reference Books

1. Arshdeep Bahga, Vijay Madisetti, “Internet of Things: A hands-on Approach”, University Press, 2015.
2. Adrian McEwen & Hakim Cassimally, “Designing the Internet of Things” Wiley, 2014.
3. Nik Bessis, Ciprian Dobre "Big Data and Internet of Things: A Roadmap for Smart Environments", Springer, 2014.
4. Maik Schmidt "Arduino: A Quick-Start Guide", The Pragmatic Bookshelf, 2011.
5. Dirk Slama, Frank Puhlmann, Jim Morrish, Rishi M Bhatnagar "Enterprise IoT: Strategies and Best Practices for Connected Products and Services", O'Reilly Media, 2015.
6. Honbo Zhou, "The Internet of Things in the Cloud: A Middleware Perspective", CRC Press, 2012.
7. Quinton Anderson "Storm Real-time Processing Cookbook", PACKT Publishers, 2013.
8. Onur Dundar,"Home Automation with Intel Galileo", Packt Publishing, 2015

Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar

Mode of assessment:

Recommended by Board of Studies	**13-05-2016**		
Approved by Academic Council	**No. 41**	**Date**	**17-06-2016**

Course code	Masters Thesis	L	T	P	J	C
CSE6099		0	0	0	0	16
Pre-requisite	As per the academic regulations	Syllabus version				
		1.0				

Course Objectives:

To provide sufficient hands-on learning experience related to the design, development and analysis of suitable product / process so as to enhance the technical skill sets in the chosen field and also to give research orientation.

Expected Course Outcome:

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

1. Formulate specific problem statements for ill-defined real life problems with reasonable assumptions and constraints.
2. Perform literature search and / or patent search in the area of interest.
3. Conduct experiments / Design and Analysis / solution iterations and document the results.
4. Perform error analysis / benchmarking / costing
5. Synthesis the results and arrive at scientific conclusions / products / solution
6. Document the results in the form of technical report / presentation

Contents

1. Capstone Project may be a theoretical analysis, modeling & simulation, experimentation & analysis, prototype design, fabrication of new equipment, correlation and analysis of data, software development, applied research and any other related activities.
2. Project can be for two semesters based on the completion of required number of credits as per the academic regulations.
3. Should be individual work.
4. Carried out inside or outside the university, in any relevant industry or research institution.
5. Publications in the peer reviewed journals / International Conferences will be an added advantage

Mode of Evaluation: Periodic reviews, Presentation, Final oral viva, Poster submission

Recommended by Board of Studies	13.05.2016		
Approved by Academic Council	41st AC	Date	17.06.2016

MAT6001	ADVANCED STATISTICAL METHODS	L	T	P	J	C
		2	0	2	0	3
Pre-requisite	**Nil**	**Syllabus version**				
		2.0				

Course Objectives:

1. To provide students with a framework that will help them choose the appropriate descriptive statistics in various data analysis situations.
2. To analyze distributions and relationships of real-time data.
3. To apply estimation and testing methods to make inference and modeling techniques for decision making using various techniques including multivariate analysis.

Expected Course Outcome:

1. Understand the value of statistics as a discipline and its relevance for Engineering
2. Analyze data using appropriate graphical methods and numerical summaries
3. Interpret and communicate the outcomes of estimation and hypothesis tests in the context of a problem
4. Perform large sample test and small sample testing of Hypothesis as well as calculate confidence interval for a population parameter for real time data.
5. describe and verify mathematical considerations for analyzing time series, including concepts of white noise, stationary, auto covariance, autocorrelation ; apply various techniques of time series models, including the regression with ARMA models

Module:1	Basic Statistical Tools for Analysis:	4hours

Summary Statistics, Correlation and Regression, Concept of R^2 and Adjusted R^2 and Partial and Multiple Correlation, Fitting of simple and Multiple Linear regression, Explanation and Assumptions of Regression Diagnostics

Module:2	Statistical inference :	9 hours

Basic Concepts, Normal distribution-Area properties, Steps in tests of significance –large sample tests-Z tests for Means and Proportions, Small sample tests –t-test for Means, F test for Equality of Variances, Chi-square test for independence of Attributes.

Module:3	Modelling and Forecasting Methods:	9hours

Introduction: Concept of Linear and Non Liner Forecasting model ,Concepts of Trend, Exponential Smoothing, Linear and Compound Growth model, Fitting of Logistic curve and their Applications, Moving Averages, Forecasting accuracy tests.

Probability models for time series: Concepts of AR, ARMA and ARIMA models.

Module:4	Design of Experiments:	6hours

Analysis of variance – one and two way classifications – Principle of design of experiments, CRD – RBD – LSD, Concepts of 2^2 and 2^3 factorial experiments

Module:5	**Contemporary issues:**	**2hours**
Lecture by Industry Experts		
	Total Lecture hours:	**30 hours**

Text Book(s)

1. Applied Statistics and Probability for Engineers, 6ed, (2016),Douglas C. Montgomery George C. Runger, John Wiley & Sons

2. Time Series Analysis and Its Applications With R Examples (2017), by Shumway, Robert H., Stoffer, David S. Springer publications

Reference Books

1.Trevor Hastie and Robert Tibshirani , "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Second Edition -Springer Series in Statistics, (2017.

2. J. Susan Milton and Jesse Arnold, "Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences" Mc.Graw Hill education, 2017

Mode of Evaluation: Digital Assignments, Quiz, Continuous Assessments, Final Assessment Test

List of Challenging Experiments (Indicative)

1.	Computing Summary Statistics using real time data
2	Lotting and visualizing data using Tabulation and Graphical Representations.
3	Applying simple linear and multiple linear regression models to real dataset; computing and interpreting the coefficient of determination for scale data.
4.	Testing of hypothesis for Large sample tests for real-time problems.
5.	Testing of hypothesis for Small sample tests for One and Two Sample mean and paired comparison (Pre-test and Post-test)
6.	Testing of hypothesis for Small Sample tests for F-test
7	Testing of hypothesis for Small Sample tests for Chi-square test
8	Applying Time series analysis-Trends. Growth ,Logistic, Exponential models
9	Applying Time series model AR , ARMA and ARIMA and testing Forecasting accuracy tests.
10	Performing ANOVA (one-way and two-way), CRD, RBD and LSD for real dataset.
11	Performing 2^2 factorial experiments with real time Applications
12	Performing 2^3 factorial experiments with real time Applications
	Total Laboratory 24 Hours

Mode of assessment:

Recommended by Board of Studies	**11.08.2017**		
Approved by Academic Council	**No. 46**	**Date**	**24.08.17**

SET5001	SCIENCE, ENGINEERING AND TECHNOLOGY PROJECT– I	L	T	P	J	C
						2
Pre-requisite		**Syllabus Version**				
Anti-requisite		1.0				

Course Objectives:

- To provide opportunity to involve in research related to science / engineering
- To inculcate research culture
- To enhance the rational and innovative thinking capabilities

Expected Course Outcome:

On completion of this course, the student should be able to:

1. Identify problems that have relevance to societal / industrial needs
2. Exhibit independent thinking and analysis skills
3. Demonstrate the application of relevant science / engineering principles

Modalities / Requirements

1. Individual or group projects can be taken up
2. Involve in literature survey in the chosen field
3. Use Science/Engineering principles to solve identified issues
4. Adopt relevant and well-defined / innovative methodologies to fulfill the specified objective
5. Submission of scientific report in a specified format (after plagiarism check)

Student Assessment : Periodical reviews, oral/poster presentation

Recommended by Board of Studies	17-08-2017		
Approved by Academic Council	No. 47	Date	05-10-2017

SET5002	SCIENCE, ENGINEERING AND TECHNOLOGY PROJECT– II	L	T	P	J	C
						2
Pre-requisite		**Syllabus Version**				
Anti-requisite		1.0				

Course Objectives:

- To provide opportunity to involve in research related to science / engineering
- To inculcate research culture
- To enhance the rational and innovative thinking capabilities

Expected Course Outcome:

On completion of this course, the student should be able to:

1. Identify problems that have relevance to societal / industrial needs
2. Exhibit independent thinking and analysis skills
3. Demonstrate the application of relevant science / engineering principles

Modalities / Requirements

1. Individual or group projects can be taken up
2. Involve in literature survey in the chosen field
3. Use Science/Engineering principles to solve identified issues
4. Adopt relevant and well-defined / innovative methodologies to fulfill the specified objective
5. Submission of scientific report in a specified format (after plagiarism check)

Student Assessment : Periodical reviews, oral/poster presentation

Recommended by Board of Studies	17-08-2017		
Approved by Academic Council	No. 47	Date	05-10-2017

ENG5001	Fundamentals of Communication Skills	L	T	P	J	C
		0	0	2	0	1
Pre-requisite	Not cleared EPT (English Proficiency Test)	**Syllabus version**				
		1.0				

Course Objectives:

1. To enable learners learn basic communication skills - Listening, Speaking, Reading and Writing
2. To help learners apply effective communication in social and academic context
3. To make students comprehend complex English language through listening and reading

Expected Course Outcome:

1. Enhance the listening and comprehension skills of the learners
2. Acquire speaking skills to express their thoughts freely and fluently
3. Learn strategies for effective reading
4. Write grammatically correct sentences in general and academic writing
5. Develop technical writing skills like writing instructions, transcoding etc.,

Module:1	Listening	**8 hours**

Understanding Conversation
Listening to Speeches
Listening for Specific Information

Module:2	Speaking	**4 hours**

Exchanging Information
Describing Activities, Events and Quantity

Module:3	Reading	**6 hours**

Identifying Information
Inferring Meaning
Interpreting text

Module:4	Writing: Sentence	**8hours**

Basic Sentence Structure
Connectives
Transformation of Sentences
Synthesis of Sentences

Module:5	Writing: Discourse	**4hours**

Instructions
Paragraph
Transcoding

	Total Lecture hours:	**30 hours**

Text Book(s)

1. Redston, Chris, Theresa Clementson, and Gillie Cunningham. *Face2face Upper Intermediate Student's Book.* 2013, Cambridge University Press.

Reference Books

1. Chris Juzwiak .*Stepping Stones: A guided approach to writing sentences and Paragraphs (Second Edition)*, 2012, Library of Congress.
2. Clifford A Whitcomb & Leslie E Whitcomb, *Effective Interpersonal and Team Communication Skills for Engineers*, 2013, John Wiley & Sons, Inc., Hoboken: New Jersey.
3. Arun Patil, Henk Eijkman &Ena Bhattacharya, *New Media Communication Skills for Engineers and IT Professionals,*2012, IGI Global, Hershey PA.
4. Judi Brownell, *Listening: Attitudes, Principles and Skills*, 2016, 5th Edition, Routledge: USA
5. John Langan, Ten Steps to Improving College Reading Skills, 2014, 6th Edition, Townsend Press:USA

6.	Redston, Chris, Theresa Clementson, and Gillie Cunningham. *Face2face Upper Intermediate Teacher's Book*. 2013, Cambridge University Press.			
	Authors, book title, year of publication, edition number, press, place			
Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar				
List of Challenging Experiments (Indicative)				
1.	amiliarizing students to adjectives through brainstorming adjectives with all letters of the English alphabet and asking them to add an adjective that starts with the first letter of their name as a prefix.			2 hours
2.	aking students identify their peer who lack Pace, Clarity and Volume during presentation and respond using Symbols.			4 hours
3.	sing Picture as a tool to enhance learners speaking and writing skills			2 hours
4.	sing Music and Songs as tools to enhance pronunciation in the target language / Activities through VIT Community Radio			2 hours
5.	Making students upload their Self- introduction videos in Vimeo.com			4 hours
6.	Brainstorming idiomatic expressions and making them use those in to their writings and day to day conversation			4 hours
7.	Making students Narrate events by adding more descriptive adjectives and add flavor to their language / Activities through VIT Community Radio			4 hours
8	Identifying the root cause of stage fear in learners and providing remedies to make their presentation better			4 hours
9	Identifying common Spelling & Sentence errors in Letter Writing and other day to day conversations			2 hours
10.	iscussing FAQ's in interviews with answers so that the learner gets a better insight in to interviews / Activities through VIT Community Radio			2 hours
			Total Laboratory Hours	**32 hours**
Mode of evaluation: Online Quizzes, Presentation, Role play, Group Discussions, Assignments, Mini Project				
Recommended by Board of Studies	22-07-2017			
Approved by Academic Council	No. 46	Date	24-8-2017	

ENG5002	Professional and Communication Skills	L	T	P	J	C
		0	0	2	0	1
Pre-requisite	ENG5001	**Syllabus version**				
		1.1				

Course Objectives:

1. To enable students to develop effective Language and Communication Skills
2. To enhance students' Personal and Professional skills
3. To equip the students to create an active digital footprint

Expected Course Outcome:

1. Improve inter-personal communication skills
2. Develop problem solving and negotiation skills
3. Learn the styles and mechanics of writing research reports
4. Cultivate better public speaking and presentation skills
5. Apply the acquired skills and excel in a professional environment

Module:1	**Personal Interaction**	**2hours**
Introducing Oneself- one's career goals Activity: SWOT Analysis		
Module:2	**Interpersonal Interaction**	**2 hours**
Interpersonal Communication with the team leader and colleagues at the workplace Activity: Role Plays/Mime/Skit		
Module:3	**Social Interaction**	**2 hours**
Use of Social Media, Social Networking, gender challenges Activity: Creating LinkedIn profile, blogs		
Module:4	**Résumé Writing**	**4 hours**
Identifying job requirement and key skills Activity: Prepare an Electronic Résumé		
Module:5	**Interview Skills**	**4 hours**
Placement/Job Interview, Group Discussions Activity: Mock Interview and mock group discussion		
Module:6	**Report Writing**	**4 hours**
Language and Mechanics of Writing Activity: Writing a Report		
Module:7	**Study Skills: Note making**	**2hours**
Summarizing the report Activity: Abstract, Executive Summary, Synopsis		
Module:8	**Interpreting skills**	**2 hours**
Interpret data in tables and graphs Activity: Transcoding		
Module:9	**Presentation Skills**	**4 hours**
Oral Presentation using Digital Tools Activity: Oral presentation on the given topic using appropriate non-verbal cues		
Module:10	**Problem Solving Skills**	**4 hours**
Problem Solving & Conflict Resolution Activity: Case Analysis of a Challenging Scenario		
	Total Lecture hours:	**30hours**

Text Book(s)

1	Bhatnagar Nitin and Mamta Bhatnagar, *Communicative English For Engineers And Professionals*, 2010, Dorling Kindersley (India) Pvt. Ltd.

Reference Books				
1	Jon Kirkman and Christopher Turk, *Effective Writing: Improving Scientific, Technical and Business Communication*, 2015, Routledge			
2	Diana Bairaktarova and Michele Eodice, *Creative Ways of Knowing in Engineering*, 2017, Springer International Publishing			
3	Clifford A Whitcomb & Leslie E Whitcomb, *Effective Interpersonal and Team Communication Skills for Engineers*, 2013, John Wiley & Sons, Inc., Hoboken: New Jersey.			
4	Arun Patil, Henk Eijkman &Ena Bhattacharya, *New Media Communication Skills for Engineers and IT Professionals*, 2012, IGI Global, Hershey PA.			
Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar				
List of Challenging Experiments (Indicative)				
1.	WOT Analysis – Focus specially on describing two strengths and two Weaknesses			2 hours
2.	Role Plays/Mime/Skit -- Workplace Situations			4 hours
3.	Use of Social Media – Create a LinkedIn Profile and also write a page or two on areas of interest			2 hours
4.	Prepare an Electronic Résumé and upload the same in vimeo			2 hours
5.	Group discussion on latest topics			4 hours
6	Report Writing – Real-time reports			2 hours
7	Writing an Abstract, Executive Summary on short scientific or research Articles			4 hours
8	Transcoding – Interpret the given graph, chart or diagram			2 hours
9	Oral presentation on the given topic using appropriate non-verbal cues			4 hours
10	Problem Solving -- Case Analysis of a Challenging Scenario			4 hours
			Total Laboratory Hours	**32 hours**
Mode of evaluation: : Online Quizzes, Presentation, Role play, Group Discussions, Assignments, Mini Project				
Recommended by Board of Studies	22-07-2017			
Approved by Academic Council	No. 47	Date	05-10-2017	

FRE5001	FRANCAIS FONCTIONNEL	L	T	P	J	C
		2	0	0	0	2
Pre-requisite	**Nil**	**Syllabus version**				
		1.0				

Course Objectives:

The course gives students the necessary background to:

1. Demonstrate competence in reading, writing, and speaking basic French, including knowledge of vocabulary (related to profession, emotions, food, workplace, sports/hobbies, classroom and family).
2. Achieve proficiency in French culture oriented view point.

Expected Course Outcome:

The students will be able to

1. Remember the daily life communicative situations via personal pronouns, emphatic pronouns, salutations, negations, interrogations etc.
2. Create communicative skill effectively in French language via regular / irregular verbs.
3. Demonstrate comprehension of the spoken / written language in translating simple sentences.
4. Understand and demonstrate the comprehension of some particular new range of unseen written materials.
5. Demonstrate a clear understanding of the French culture through the language studied.

Module:1	Saluer, Se présenter, Etablir des contacts	**3 hours**

Les Salutations, Les nombres (1-100), Les jours de la semaine, Les mois de l'année, Les Pronoms Sujets, Les Pronoms Toniques, La conjugaison des verbes réguliers, La conjugaison des verbes irréguliers- avoir / être / aller / venir / faire etc.

Module:2	**Présenter quelqu'un, Chercher un(e) correspondant(e), Demander des nouvelles d'une personne.**	**3 hours**

La conjugaison des verbes Pronominaux, La Négation, L'interrogation avec *'Est-ce que ou sans Est-ce que'*.

Module:3	**Situer un objet ou un lieu, Poser des questions**	**4 hours**

L'article (défini/ indéfini), Les prépositions (à/en/au/aux/sur/dans/avec etc.), L'article contracté, Les heures en français, La Nationalité du Pays, L'adjectif (La Couleur, l'adjectif possessif, l'adjectif démonstratif/ l'adjectif interrogatif (quel/quelles/quelle/quelles), L'accord des adjectifs avec le nom, L'interrogation avec Comment/ Combien / Où etc.,

Module:4	**Faire des achats, Comprendre un texte court, Demander et indiquer le chemin.**	**6 hours**

La traduction simple :(français-anglais / anglais –français)

Module:5	**Trouver les questions, Répondre aux questions générales en français.**	**5 hours**

L'article Partitif, Mettez les phrases aux pluriels, Faites une phrase avec les mots donnés, Exprimez les phrases données au Masculin ou Féminin, Associez les phrases.

Module:6	**Comment ecrire un passage**	**3 hours**

Décrivez :
La Famille /La Maison, /L'université /Les Loisirs/ La Vie quotidienne etc.

Module:7	**Comment ecrire un dialogue**	**4 hours**

Dialogue:

a) Réserver un billet de train b) Entre deux amis qui se rencontrent au café c) Parmi les membres de la famille d) Entre le client et le médecin			
Module:8	**Invited Talk: Native speakers**		**2 hours**
	Total Lecture hours:	**30 hours**	

Text Book(s)	
1.	Echo-1, Méthode de français, J. Girardet, J. Pécheur, Publisher CLE International, Paris 2010.
2	Echo-1, Cahier d'exercices, J. Girardet, J. Pécheur, Publisher CLE International, Paris 2010.
Reference Books	
1.	CONNEXIONS 1, Méthode de français, Régine Mérieux, Yves Loiseau,Les Éditions Didier, 2004.
2	CONNEXIONS 1, Le cahier d'exercices, Régine Mérieux, Yves Loiseau, Les Éditions Didier, 2004.
3	ALTER EGO 1, Méthode de français, Annie Berthet, Catherine Hugo, Véronique M. Kizirian, Béatrix Sampsonis, Monique Waendendries , Hachette livre 2006.

Mode of Evaluation: CAT / Assignment / Quiz / FAT			
Recommended by Board of Studies	13.05.2016		
Approved by Academic Council	No. 41	Date	17-06-2016

GER5001	Deutsch für Anfänger	L	T	P	J	C
		2	0	0	0	2
Pre-requisite	**NIL**	**Syllabus version**				
		1.0				

Course Objectives:

The course gives students the necessary background to:

1. Enable students to read and communicate in German in their day to day life
2. Become industry-ready
3. Make them understand the usage of grammar in the German Language.

Expected Course Outcome:

The students will be able to

1. Create the basics of German language in their day to day life.
2. Understand the conjugation of different forms of regular/irregular verbs.
3. Understand the rule to identify the gender of the Nouns and apply articles appropriately.
4. Apply the German language skill in writing corresponding letters, E-Mails etc.
5. Create the talent of translating passages from English-German and vice versa and To frame simple dialogues based on given situations.

Module:1		**3 hours**

Einleitung, Begrüssungsformen, Landeskunde, Alphabet, Personalpronomen, Verb Konjugation, Zahlen (1-100), W-fragen, Aussagesätze, Nomen – Singular und Plural

Lernziel:

Elementares Verständnis von Deutsch, Genus- Artikelwörter

Module:2		**3 hours**

Konjugation der Verben (regelmässig /unregelmässig) die Monate, die Wochentage, Hobbys, Berufe, Jahreszeiten, Artikel, Zahlen (Hundert bis eine Million), Ja-/Nein- Frage, Imperativ mit Sie

Lernziel :

Sätze schreiben, über Hobbys erzählen, über Berufe sprechen usw.

Module:3		**4 hours**

Possessivpronomen, Negation, Kasus- AkkusatitvundDativ (bestimmter, unbestimmterArtikel), trennnbare verben, Modalverben, Adjektive, Uhrzeit, Präpositionen, Mahlzeiten, Lebensmittel, Getränke

Lernziel :

Sätze mit Modalverben, Verwendung von Artikel, über Länder und Sprachen sprechen, über eine Wohnung beschreiben.

Module:4		**6 hours**

Übersetzungen : (Deutsch – Englisch / Englisch – Deutsch)

Lernziel :

Grammatik – Wortschatz – Übung

Module:5		**5 hours**

Leseverständnis,Mindmap machen,Korrespondenz- Briefe, Postkarten, E-Mail

Lernziel :

Wortschatzbildung und aktiver Sprach gebrauch			
Module:6	.		**3 hours**
Aufsätze : Meine Universität, Das Essen, mein Freund oder meine Freundin, meine Familie, ein Fest in Deutschland usw			
Module:7			**4 hours**
Dialoge: e) Gespräche mit Familienmitgliedern, Am Bahnhof, f) Gespräche beim Einkaufen ; in einem Supermarkt ; in einer Buchhandlung ; g) in einem Hotel - an der Rezeption ;ein Termin beim Arzt. Treffen im Café			
Module:8			**2 hours**
Guest Lectures/Native Speakers / Feinheiten der deutschen Sprache, Basisinformation über die deutschsprachigen Länder			
	Total Lecture hours:	**30 hours**	

Text Book(s)	
1.	**Studio d A1 Deutsch als Fremdsprache, Hermann Funk, Christina Kuhn, Silke Demme : 2012**
Reference Books	
1	etzwerk Deutsch als Fremdsprache A1, Stefanie Dengler, Paul Rusch, Helen Schmtiz, Tanja Sieber, 2013
2	Lagune ,Hartmut Aufderstrasse, Jutta Müller, Thomas Storz, 2012.
3	eutsche SprachlehrefürAUsländer, Heinz Griesbach, Dora Schulz, 2011
4	hemenAktuell 1, HartmurtAufderstrasse, Heiko Bock, MechthildGerdes, Jutta Müller und Helmut Müller, 2010
	ww.goethe.de irtschaftsdeutsch.de eber.de, klett-sprachen.de ww.deutschtraning.org

Mode of Evaluation: CAT / Assignment / Quiz / FAT			
Recommended by Board of Studies	13.05.2016		
Approved by Academic Council	No. 41	Date	17-06-2016

STS5001	Essentials of Business Etiquettes	L	T	P	J	C
		3	0	0	0	1
Pre-requisite		**Syllabus version**				
		2.0				

Course Objectives:

1. To develop the students' logical thinking skills
2. To learn the strategies of solving quantitative ability problems
3. To enrich the verbal ability of the students
4. To enhance critical thinking and innovative skills

Expected Course Outcome:

1. Enabling students to use relevant aptitude and appropriate language to express themselves
2. To communicate the message to the target audience clearly

Module:1	**Business Etiquette: Social and Cultural Etiquette and Writing Company Blogs and Internal Communications and Planning and Writing press release and meeting notes**	**9 hours**

Value, Manners, Customs, Language, Tradition, Building a blog, Developing brand message, FAQs', Assessing Competition, Open and objective Communication, Two way dialogue, Understanding the audience, Identifying, Gathering Information,. Analysis, Determining, Selecting plan, Progress check, Types of planning, Write a short, catchy headline, Get to the Point –summarize your subject in the first paragraph., Body – Make it relevant to your audience,

Module:2	**Study skills – Time management skills**	**3 hours**

Prioritization, Procrastination, Scheduling, Multitasking, Monitoring, Working under pressure and adhering to deadlines

Module:3	**Presentation skills – Preparing presentation and Organizing materials and Maintaining and preparing visual aids and Dealing with questions**	**7 hours**

10 Tips to prepare PowerPoint presentation, Outlining the content, Passing the Elevator Test, Blue sky thinking, Introduction , body and conclusion, Use of Font, Use of Color, Strategic presentation, Importance and types of visual aids, Animation to captivate your audience, Design of posters, Setting out the ground rules, Dealing with interruptions, Staying in control of the questions, Handling difficult questions

Module:4	**Quantitative Ability -L1 – Number properties and Averages and Progressions and Percentages and Ratios**	**11 hours**

Number of factors, Factorials, Remainder Theorem, Unit digit position, Tens digit position, Averages, Weighted Average, Arithmetic Progression, Geometric Progression, Harmonic Progression, Increase & Decrease or successive increase, Types of ratios and proportions

Module:5	**Reasoning Ability-L1 – Analytical Reasoning**	**8 hours**
Data Arrangement(Linear and circular & Cross Variable Relationship), Blood Relations, Ordering/ranking/grouping, Puzzle test, Selection Decision table		
Module:6	**Verbal Ability-L1 – Vocabulary Building**	**7 hours**
Synonyms & Antonyms, One word substitutes, Word Pairs, Spellings, Idioms, Sentence completion, Analogies		
	Total Lecture hours:	**45 hours**

Reference Books

1.	Kerry Patterson, Joseph Grenny, Ron McMillan, Al Switzler(2001) Crucial Conversations: Tools for Talking When Stakes are High. Bangalore. McGraw-Hill Contemporary
2.	Dale Carnegie,(1936) How to Win Friends and Influence People. New York. Gallery Books
3.	Scott Peck. M(1978) Road Less Travelled. New York City. M. Scott Peck.
4.	FACE(2016) Aptipedia Aptitude Encyclopedia. Delhi. Wiley publications
5.	ETHNUS(2013) Aptimithra. Bangalore. McGraw-Hill Education Pvt. Ltd.

Websites:

1.	**www.chalkstreet.com**
2.	**www.skillsyouneed.com**
3.	**www.mindtools.com**
4.	**www.thebalance.com**
5.	**www.eguru.ooo**

Mode of Evaluation: FAT, Assignments, Projects, Case studies, Role plays,
3 Assessments with Term End FAT (Computer Based Test)

Recommended by Board of Studies	09/06/2017		
Approved by Academic Council	No. 45th AC	Date	15/06/2017

STS5002	Preparing for Industry	L	T	P	J	C
		3	0	0	0	1
Pre-requisite		**Syllabus version**				
		2.0				

Course Objectives:

1. To develop the students' logical thinking skills
2. To learn the strategies of solving quantitative ability problems
3. To enrich the verbal ability of the students
4. To enhance critical thinking and innovative skills

Expected Course Outcome:

1. Enabling students to simplify, evaluate, analyze and use functions and expressions to simulate real situations to be industry ready.

Module:1	**Interview skills – Types of interview and Techniques to face remote interviews and Mock Interview**	**3 hours**

Structured and unstructured interview orientation, Closed questions and hypothetical questions, Interviewers' perspective, Questions to ask/not ask during an interview, Video interview, Recorded feedback, Phone interview preparation, Tips to customize preparation for personal interview, Practice rounds

Module:2	**Resume skills – Resume Template and Use of power verbs and Types of resume and Customizing resume**	**2 hours**

Structure of a standard resume, Content, color, font, Introduction to Power verbs and Write up, Quiz on types of resume, Frequent mistakes in customizing resume, Layout - Understanding different company's requirement, Digitizing career portfolio

Module:3	**Emotional Intelligence - L1 – Transactional Analysis and Brain storming and Psychometric Analysis and Rebus Puzzles/Problem Solving**	**12 hours**

Introduction, Contracting, ego states, Life positions, Individual Brainstorming, Group Brainstorming, Stepladder Technique, Brain writing, Crawford's Slip writing approach, Reverse brainstorming, Star bursting, Charlette procedure, Round robin brainstorming, Skill Test, Personality Test, More than one answer, Unique ways

Module:4	**Quantitative Ability-L3 – Permutation-Combinations and Probability and Geometry and mensuration and Trigonometry and Logarithms and Functions and Quadratic Equations and Set Theory**	**14 hours**

Counting, Grouping, Linear Arrangement, Circular Arrangements, Conditional Probability, Independent and Dependent Events, Properties of Polygon, 2D & 3D Figures, Area & Volumes, Heights and distances, Simple trigonometric functions, Introduction to logarithms, Basic rules of logarithms, Introduction to functions, Basic rules of functions, Understanding Quadratic Equations, Rules & probabilities of Quadratic Equations, Basic concepts of Venn Diagram

Module:5	**Reasoning ability-L3 – Logical reasoning and**	**7 hours**

	Data Analysis and Interpretation	
Syllogisms, Binary logic, Sequential output tracing, Crypto arithmetic, Data Sufficiency, Data interpretation-Advanced, Interpretation tables, pie charts & bar chats		
Module:6	**Verbal Ability-L3 – Comprehension and Logic**	**7 hours**
Reading comprehension, Para Jumbles, Critical Reasoning (a) Premise and Conclusion, (b) Assumption & Inference, (c) Strengthening & Weakening an Argument		
	Total Lecture hours:	**45 hours**

Reference Books

1.	Michael Farra and JIST Editors(2011) Quick Resume & Cover Letter Book: Write and Use an Effective Resume in Just One Day. Saint Paul, Minnesota. Jist Works
2.	Daniel Flage Ph.D(2003) The Art of Questioning: An Introduction to Critical Thinking. London. Pearson
3.	David Allen(2002) Getting Things done : The Art of Stress -Free productivity. New York City. Penguin Books.
4.	FACE(2016) Aptipedia Aptitude Encyclopedia.Delhi. Wiley publications
5.	ETHNUS(2013) Aptimithra. Bangalore. McGraw-Hill Education Pvt. Ltd.

Websites:

1.	**www.chalkstreet.com**
2.	**www.skillsyouneed.com**
3.	**www.mindtools.com**
4.	**www.thebalance.com**
5.	**www.eguru.ooo**

Mode of Evaluation: FAT, Assignments, Projects, Case studies, Role plays,
3 Assessments with Term End FAT (Computer Based Test)

Recommended by Board of Studies	09/06/2017		
Approved by Academic Council	No. 45th AC	Date	15/06/2017