School of Computing Science and Engineering

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

Program Educational Objectives

- Graduates will be prepared to review and understand foundational Concepts in Computer Science, Data Science and Engineering
- Graduates will be empowered to critically analyze current trends and learn future issues from a system perspective at multiple levels of detail and abstraction
- Graduates will be enabled to apply the interaction between theory and practice for problem solving based on case studies
- Graduates will be enabled to pursue lifelong multidisciplinary learning as professional engineers and scientists to effectively communicate technical information, function effectively on teams, and apply computer engineering solutions within a global, societal, and environmental context
- Graduates will be prepared to critically analyze existing literature, identify the gaps in the existing literature, map the existing problems as Big Data and propose innovative and research oriented solutions.
- Graduates will be enabled to process, model, analyze and visualize humongous data and uncover the data-driven latent insights

About Program

Full Time, Two years of duration, Master Degree Program

Curriculum

Comprehensive list of Courses and Credit requirements

Syllabi

Outline and summary of topics to be covered in M.Tech - CSE with Specialization in Big Data Analytics program

Curriculum

A. University Core

Course Code	Course Name	L	T	P	J	C
MAT6001	Advanced Statistical Methods	2	0	2	0	3
EFL5097/ENG5001	Scientific English / Foreign Language	1	0	2	0	2
STS5001	Soft Skills - I	0	0	0	0	1
STS5001	Soft Skills - II	0	0	0	0	1
SET5001	SET Project -I	0	0	0	0	2
SET5002	SET Project - II	0	0	0	0	2
CSE6099	Master's Thesis	0	0	0	0	16
Total Credits - 27						

L T P J C: **L**ecture, **T**utorial, **P**ractical, Pro**J**ect, Total **C**redits

B. University Elective

Course Code	Course Name	L T P J C
	University Elective I	
	University Elective II	
Minimum of 6 Credits to be earned		

C. Programme Core

Course Code	Course Name	L	T	P	J	C
CSE5001	Algorithms: Design and Implementations	2	0	2	0	3
CSE5003	Database Systems: Design and implementation	2	0	2	4	4
CSE5007	Exploratory Data Analysis	2	0	0	4	3
CSE6001	Big Data Frameworks	2	0	2	4	4
CSE6005	Machine Learning	2	0	2	4	4
Total Credits - 18						

D. Programme Electives

Course Code	Course Name	L	T	P	J	C
CSE5002	Operating Systems & Virtualization	2	0	2	0	3
CSE5006	Multi-core Architectures	2	0	2	0	3
CSE6006	NoSQL Databases	2	0	2	4	4
CSE6014	Programming for Data Science	0	0	4	0	2
CSE6016	Information Visualization	2	0	2	4	4
CSE6017	Mining Massive Data	2	0	2	4	4
CSE6018	Streaming Data Analytics	2	0	2	4	4
CSE6019	Text, Web and Social Media Analytic	3	0	0	4	4
CSE6020	Big Data Technologies	2	0	2	4	4
CSE6021	Domain Specific Predictive Analytics	3	0	0	4	4
CSE6022	Soft Computing	3	0	0	4	4
CSE6023	Cloud Computing Fundamentals	2	0	2	4	4
CSE6025	Analytics of Things	3	0	0	4	4

Minimum of 19 Credits to be earned

E. Credits Summary

Minimum Qualifying Credits	70
University Core	27
University Electives	06
Programme Core	18
Programme Electives	19

MAT6001	ADVANCED STATISTICAL METHODS	L T P J C 2 0 2 0 3	
Prerequisites	None		
Course Objectives	• To provide students with a framework that will help them choose the appropriate descriptive statistics in various data analysis situations. • To analyse distributions and relationships of real-time data. • To apply estimation and testing methods to make inference and modelling techniques for decision making using various techniques including multivariate analysis.		
Module	**Topic**	**Lecture Hours**	**SLO**
1	**Basic Statistical Tools for Analysis:** Summary Statistics, Correlation and Regression, Concept of R^2 and Adjusted R^2 and and Partial and Multiple Correlation, Fitting of simple and Multiple Linear regression, Explanation and Assumptions of Regression Diagnostics.	4	1,2,7,9
2	**Statistical inference :** 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.	9	1,2,7,9
3	**Modelling and Forecasting Methods:** 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.	9	1,2,7,9
4	**Design of Experiments:** 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	6	1,2,7,9
5	**Expert Guest Lecture:** Applications of Statistical methods in Science, Engineering and industry	2	1,2,7,9
	Total Lecture Hours	30	
Tutorial	Minimum of 10 problems per module	Non-contact hours	1,2,7,9
Laboratory	• Use any statistical software (R ,SAS, MATLAB, EXCEL) data numerically and visually, and to perform data analysis. • Experiments. 1. Computing Summary Statistics /plotting and visualizing data using Tabulation and Graphical Representations. 2. Applying simple linear and multiple linear regression models to real dataset; computing and interpreting the coefficient of determination for scale data. 3. Fitting of Normal distribution 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 and Chi-square test	30	1, 2, 7,9,12, 18, 20

	7. Applying Time series analysis-Trends. Growth ,Logistic, Exponential models 8. Applying Time series model AR ,ARMA and ARIMA and testing Forecasting accuracy tests. 9. Performing ANOVA (one-way and two-way), CRD, RBD and LSD for real dataset. 10. Performing 2^2 and 2^3 factorial experiments with real time Applications		
Text Books	1. Douglas C. Montgomery, George C. Runger, Applied Statistics and Probability for Engineers, (2010), John Wiley & Sons; 5th Edition. 2. Joseph F. Hair and William C. Black, Multivariate Data Analysis, (2014), 7th Edition Pearson India.		
Reference Books	1. J. L. Devore, Probability and Statistics, 8th Edition, Brooks/Cole, Cengage Learning (2012). 2. R. A. Johnson, Miller & Freund's, Probability and Statistics for Engineers, 8th edition, Prentice Hall India (2010).		
Mode of Teaching	• Class Room Teaching • Introducing Modules Through Realistic Applications • Minimum of 1 Lecture by Experts		
Mode of Evaluation	• Digital Assignments (Solutions by using soft skills) • Continuous Assessment Tests • Final Assessment Test		
Proposed by	Dr. G. Mokesh Rayalu		
Recommended by the board of studies	25-02-2017		
Date of approval by the academic council	16-03-2017		

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

Course Objectives:

To enable learners learn basic communication skills - Listening, Speaking, Reading and Writing and apply them for various purposes in academic and social contexts

Expected Course Outcome:

Ability to communicate effectively in social and academic contexts

Student Learning Outcomes (SLO): 16,18

Module:1	Listening	**8 hours**	**SLO: 16**
Understanding Conversation Listening to Speeches Listening for Specific Information			
Module:2	Speaking	**4 hours**	**SLO: 16**
Exchanging Information Describing Activities, Events and Quantity			
Module:3	Reading	**6 hours**	**SLO: 16,18**
Identifying Information Inferring Meaning Interpreting text			
Module:4	Writing: Sentence	**8 hours**	**SLO: 16**
Basic Sentence Structure Connectives Transformation of Sentences Synthesis of Sentences			
Module:5	Writing: Discourse	**4 hours**	**SLO: 16,18**
Instructions Paragraph Transcoding			
	Total Practical 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.

Mode of Evaluation: Online Quizzes, Presentation, Role play, Group Discussions, Assignments, Mini Project

List of Challenging Experiments (Indicative)		**SLO: 16,18**
1.	Familiarizing 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.	Making students identify their peer who lack Pace, Clarity and Volume during presentation and respond using Symbols.	4 hours
3.	Using Picture as a tool to enhance learners speaking and writing skills	2 hours
4.	Using 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.	Discussing 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 Practical Hours	30 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	24-08-2017

Course code	Course title	L	T	P	J	C
STS5001	Essentials of Business etiquettes	3	0	0	0	1
Pre-requisite		Syllabus version				
		2				

Course Objectives:

- Having Computational thinking (Ability to translate vast data into abstract concepts and to understand database reasoning)[SLO 7]
- Having problem solving ability- solving social issues and engineering problems [SLO 9]

Expected Course Outcome:

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

Student Learning Outcomes (SLO):	7, 9

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	SLO: 7

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	SLO: 9

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	SLO: 7

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	SLO: 9

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**	**SLO: 9**
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**	**SLO: 7,9**
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.

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 45	Date	15/06/2017

CSE 5001	Algorithms: Design and Implementation	L,T,P,J,C 2,0,2,0,3
Preamble	This course is a core course focusing on the design and implementation of algorithms.	
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 work in algorithms-sufficient to give some context for formulating and seeking known solutions to an algorithmic problem.	
Expected Outcomes	After successfully completing the course the student should be able to 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	
Student Learning Outcomes	1. Having an ability to apply mathematics and science in engineering applications 2. Having a clear understanding of the subject related concepts and of contemporary issues 14. Having an ability to design and conduct experiments, as well as to analyze and interpret data	

Module	Topics	L Hrs	SLO
1	**Introduction** Algorithm design techniques: Divide and Conquer, Brute force, Greedy, Dynamic Programming. Time complexity (asymptotic notation, recurrence relations)	5	1, 2
2	**Network Flows** Maximum Flows, Min-cost Flows, Max-Flow Min-Cut Theorem, Cycle Cancelling Algorithms, Strongly Polynomial-time Analysis, Minimum Cuts without Flows	6	2

3	**Tractable and Intractable Problems** Class complexity: P, NP, NP-Hard, NP-Complete	3	2
4	**Approximation Algorithms** Limits to Approximability, Vertex Cover problem, Set cover problem, Euclidean TSP	3	1, 2
5	**Search Algorithms for Graphs and Trees** Overview of fundamental algorithms, Dijkstra's algorithm, A* search algorithm	4	1, 2
6	**Computational Geometry** Line Segments, Convex hull finding algorithms	4	1, 2
7	**Linear Programming** Representing problems - shortest paths, maximum flow, and minimum-cost flow as linear programming problems. Simplex algorithm	3	1, 2
8	**Recent Trends**	2	2
List of Lab Experiments 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. 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 3) Implementation of Dinic's strongly polynomial algorithm for computing the maximum flow in a flow network and applying it for solving typical problems 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 5) Applying linear programming for solving maximum flow problem 6) Applying network flow algorithms for baseball elimination and airline scheduling 7) Given a flow network $G = (V, E, s, t)$,where V is the vertex set, E is		30 Hours	14

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.

8) Implementation of solution techniques for the minimum-cost flow problem

9) Design a polynomial time algorithm to compute the solution of a linear programming problem in two dimensions. Your algorithm should convert each constraint of 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 2 hours on machine M1 and 6 hours on machine M2 . A table requires 5 hours on machine M1 and no time on machine M2. There are 16 hours of time per day available on machine M1 and 30 hours on machine M2 . Profits gained by manufacturer from a chair and a table are Rs.1 and Rs. 5 respectively . The problem is to maximize the profit for the manufacturer

10) Implementation of algorithms for the vertex cover problem, set cover problem, TSP

11) Implementation of search algorithms for graphs and trees: fundamental algorithms, Dijkstra's algorithm

12) Consider the problem of barricading n 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)

13) A simple polygon is defined as a flat shape consisting of straight non-intersecting line segments or sides that are joined pair –wise to from a

closed path. Let $P = \{p_1, p_2, p_3, \ldots p_n\}$ 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. Students are free to implement the algorithms using programming languages of their choice.		

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.

Algorithms and Complexity

Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	Total Hours of Coverage (Theory + Practical)
CS:AL/Basic Analysis	1
CS:AL/Algorithmic Strategies	4 + 8
CS:SDF/Algorithms and Design	5
CS:AL/Basic Automata Computability and Complexity	3
CS:AL/Fundamental Data Structures and Algorithms	2 + 3
CS:IS/Basic Search Strategies	2
CS:AL/Advanced Data Structures Algorithms and Analysis	17 + 19

Body of Knowledge coverage

KA	Knowledge Unit	Topics Covered	Hours
CS:AL/Basic Analysis CS:AL/Algorithmic Strategies CS:SDF/Algorithms and Design	**Introduction**	Algorithm design techniques: Divide and Conquer, Brute force, Greedy, Dynamic Programming. Time complexity (asymptotic notation, recurrence relations)	5
CS:AL/Advanced Data Structures Algorithms and Analysis	**Network Flows**	Maximum Flows, Min-cost Flows, Max-Flow Min-Cut Theorem, Cycle Cancelling Algorithms, Strongly Polynomial-time Analysis, Minimum Cuts without Flows	6
CS:AL/Basic Automata Computability and Complexity	**Tractable and Intractable Problems**	Class complexity: P, NP, NP-Hard, NP-Complete	3
CS:AL/Advanced Data Structures Algorithms and Analysis	**Approximation Algorithms**	Limits to Approximability, Vertex Cover problem, Set cover problem, Euclidean TSP	3

CS:AL/Fundamental Data Structures and Algorithms CS:IS/Basic Search Strategies	**Search Algorithms for Graphs and Trees**	Overview of fundamental algorithms, Dijkstra's algorithm, A* search algorithm	4
CS:AL/Advanced Data Structures Algorithms and Analysis	**Computational Geometry**	Line Segments, Convex hull finding algorithms	4
CS:AL/Advanced Data Structures Algorithms and Analysis	**Linear Programming**	Representing problems such as shortest paths, maximum flow, and minimum-cost flow as linear programming problems. Simplex algorithm	3
	Recent Trends		2
		Total hours	30

Where does the course fit in the curriculum?

This course

- Is a core course. It is compulsory.
- Is opted by all students.

What is covered in the course?

The course is a core course on algorithms and complexity. It has eight modules. This course includes introduction, network flows, tractable and intractable problems, approximation algorithms, search algorithms for graphs and trees, computational geometry, linear programming and recent trends. The course lays emphasis on techniques used to design algorithms. The goal is for students to be able to apply all of the above to designing solutions for real-world problems.

Module 1: Introduction

This module introduces some algorithm design techniques: Divide and Conquer, Brute force, Greedy, Dynamic Programming. Time complexity is explained by means of asymptotic notation, and recurrence relations.

Module 2: Network Flows

This module discusses maximum flows, min-cost flows, the Max-Flow Min-Cut theorem, cycle cancelling algorithms, strongly polynomial-time analysis, and minimum cuts without flows.

Module 3: Tractable and Intractable Problems

This module discusses tractable and intractable problems by introducing the complexity classes P, NP, NP-Hard, and NP-Complete.

Module 4: Approximation Algorithms

This module discusses the limits to approximability, the vertex cover problem, the set cover problem, and the Euclidean Travelling Salesman Problem.

Module 5: Search Algorithms for Graphs and Trees

This module provides an overview of fundamental algorithms (breadth first search and depth first search), Dijkstra's algorithm, and the A* search algorithm.

Module 6: Computational Geometry

This module discusses line segment properties and some convex hull finding algorithms.

Module 7: Linear Programming

This module discusses representation of problems such as shortest paths, maximum flow, and minimum-cost flow as linear programming problems. It looks at the Simplex algorithm.

Module 8: Recent Trends

This module discusses some recent trends in algorithms.

What is the format of the course?

This course is a face to face course. The number of lecture hours is 30. The number of lab hours is 30. The course will have a combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading, quizzes, assignments, and tests. The course will have lab sessions. The course will attempt to include video lectures, flipped classroom strategy, virtual classes, and encourage MOOC learning,

How are students assessed?

- The students will be assessed on the basis of classroom discussion, quizzes, continuous assessment, and the final assessment test. Adequate time will be given to the students to complete their assignments, exercises, lab exercises, quizzes, and tests.

- For the programming assignments, students are strongly encouraged to use programming languages of their choice.

Session wise plan

Student Learning Outcomes Covered: 1,2,14

Sl. No	Topic Covered	Class Hour	Lab Hour	Levels of mastery	Reference Book	Remarks
1	**Introduction** Algorithm design techniques: Divide and Conquer, Brute force	2	4	Familiarity Usage	1, 2, 3	
2	Greedy, Dynamic Programming	2	4	Familiarity Usage	1, 2, 3	
3	Time complexity (asymptotic notation, recurrence relations)	1		Familiarity Usage	1, 2, 3	
4	**Network Flows** Maximum Flows, Min-cost Flows	2	4	Familiarity Usage	1, 2, 4	
5	Max-Flow Min-Cut Theorem	1		Familiarity	1, 4	
6	Cycle Cancelling Algorithms	1	2	Familiarity Usage	4	
7	Strongly Polynomial-time Analysis	1	2	Familiarity Usage	4	
8	Minimum Cuts without Flows	1		Familiarity	4	
9	**Tractable and Intractable Problems** Class complexity: P, NP	1		Familiarity	1, 2, 3, 4	
10	NP-Hard, NP-Complete	2		Familiarity	1, 2, 3, 4	
11	**Approximation Algorithms** Limits to Approximability, Vertex Cover problem	1	1	Familiarity Usage	1, 2	
12	Set cover problem	1	1	Familiarity Usage	1, 2	
13	Euclidean TSP	1	1	Familiarity Usage	1	
14	**Search Algorithms for Graphs and Trees** Overview of fundamental algorithms	1	2	Familiarity Usage	1, 2, 3, 5	
15	Dijkstra's algorithm	1	1	Familiarity Usage	1, 3	
16	A* search algorithm	2		Familiarity	5	
17	**Computational Geometry** Line Segments	2		Familiarity	1	
18	Convex hull finding algorithms	2	2	Familiarity Usage	1, 5	

19	**Linear Programming** Representing problems such as shortest paths, maximum flow, and minimum-cost flow as linear programming problems	1	2	Familiarity Usage	1	
20	Simplex algorithm	2	4	Familiarity Usage	1, 4	
21	**Recent Trends**	2		Familiarity		
Total hours covered		30	30			

CSE5003

DATABASE SYSTEMS: DESIGN AND IMPLEMENTATION

L,T,P,J,C

2,0,2,4,4

Objectives:

- To understand the underlying principles of Relational Database Management System.
- To focus on the modeling and design of databases and usage of advanced data models.
- To implement and maintain the structured, semi structured and unstructured data in an efficient database system using emerging trends.
- To identify the different threats to databases

Expected Outcome:

After successfully completing the course the student should be able to

- Design and implement a database depending on the business requirements, considering various design issues.
- Categorize and design the structured, semi structured and unstructured databases
- Characterize the database threats and its countermeasures.

Student Learning Outcome:

5. Having design thinking capability
7. Having computational thinking (Ability to translate vast data in to abstract concepts and to understand database reasoning)
11. Having interest in lifelong learning

Module	Topics	L Hrs	SLO
1	**Relational Model** Database System Architecture – EER Modeling- Indexing – Normalization –Query processing and optimization – Transaction Processing	6	7
2	**Parallel Databases** Architecture, Data partitioning strategy, Interquery and Intraquery Parallelism –Parallel Query Optimization	4	5,7

3	**Distributed Databases** Features – Distributed Database Architecture –Fragmentation – Replication- Distributed Query Processing – Distributed Transactions Processing	5	5,7
4	**Spatial and Mobile Databases** Spatial databases -Type of spatial data– Indexing in spatial databases, Mobile Databases – Transaction Model in MDS	3	5
5	**Semi Structured Databases:** Semi Structured databases – XML –Schema-DTD- XPath- XQuery. Semantic Web –RDF –RDFS	4	7
6	**Database Security** Introduction to Database Security Issues –Security Models – Different Threats to databases – Counter measures to deal with these problems	3	7
7	**Emerging Technologies** Cloud databases – Streaming Databases - Graph Databases-New SQL databases	3	7
8	**Recent Trends**	2	11
Total Lecture Hours		30	

Reference Books

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.
3. S.K.Singh, "Database Systems: Concepts, Design & Applications", 2nd edition, Pearson education, 2011.
4. Joe Fawcett, Danny Ayers, Liam R. E. Quin: "Beginning XML", Wiley India Private Limited, 5th Edition, 2012.
5. Thomas M. Connolly and Carolyn Begg "Database Systems: A Practical Approach to Design, Implementation, and Management", 6th edition, Pearson India, 2015.

List of Lab Experiments

1. Model any given scenario into ER/EER Model using any tool (ERDPlus,ERWin,OracleSQL developer)
2. Creating applications with RDBMS

   a) Table creation with constraints, alter schema ,insert values, aggregate functions,
   simple and complex queries with joins

   b) PLSQL-PROCEDURES,CURSORS,FUNCTIONS,TRIGGERS
3. Partition a given database based on the type of query and compares the execution speed of the query with/without parallelism.
4. Create an XML document and validate it against an XML Schema/DTD. Use XQuery to query and view the contents of the database.
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

14,17

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 XQuery engine4.

6. To implement parallel join and parallel sort algorithms to get marks from different colleges of the university and publish 10 ranks for each discipline.

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 at least 5 suitable queries on Employee fragments. Add relations to the database as per your requirements.

9. Download a spatial dataset based on any specific theme (containing layer information) from Quantum GIS and import it into PostgresSQL(PostGIS) and Query and view the database.

10. 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 geocoded from a list of addresses provided by the EPA

11. Use sample datasets from health care domain ,Visualize and interpret the results

12. Import the Hubway data into Neo4j and configure Neo4j. 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 (for example 13 means 1pm -2pm) and number of trips which end at the station "B.U. Central"

Sample Projects

1. Developing applications by **creating** and modeling data into different databases as given below and creating an interface for Querying/ Viewing the database for various functionalities.
    - Relational database
    - Distributed database
    - Spatial Database
    - XML Database
    - Mobile databases
    -
2. Analyzing and Visualizing social networks like facebook , twitter etc. using NoSQL Databases.

3. Using Sample datasets from http://www.rdatamining.com/resources/data, UCLA Repository, kaggle dataset etc. and analyzing them using NoSQL databases.

Knowledge area	**Total Hours of coverage**
CS: SF (System Fundamentals)	4
CS:IM :Information Management/ CE:DBS Database Systems	14
CS:PD: Parallel and Distributed Computing	9
CS: IAS Information Assurance and	3
Security	
Total	30

Body of Knowledge Coverage

KA	Knowledge Unit	Topics Covered	Hours
CS:IM CE:DBS	Relational Databases Data Modeling	**Relational Model** Database System Architecture – EER Modeling- Indexing – Normalization –Query processing and optimization – Transaction Processing	6
CS:PD	Parallel and Distributed Computing	**Parallel Databases** Architecture, Data partitioning strategy, Interquery and Intraquery Parallelism –Parallel Query Optimization **Distributed Databases** Features – Distributed Database Architecture –Fragmentation – Replication- Distributed Query Processing – Distributed Transactions Processing	9
CS:IM	Spatial and Mobile databases	**Spatial and Mobile Databases** Spatial databases -Type of spatial data– Indexing in spatial databases, Mobile Databases – Transaction Model in MDS	3
CS:IM	Semi Structured databases	**Semi Structured Databases:** Semi Structured databases – XML	4

		–XPath- XQuery ,Semantic Web, RDF /RDFs	
CS: IAS	Information Assurance and Security	Introduction to Database Security Issues –Security Models – Different Threats to databases – Counter measures to deal with these problems	3
CS:IM	Emerging technologies	**Emerging Technologies-** Cloud databases – Streaming Databases - Graph Databases-New SQL databases and Recent Trends	5
			30

Where does the course fit in the curriculum?

This course is a

- Core Course.
- Suitable from 1st semester onwards.

What is covered in the course?

Module I - Relational Model –It introduces the Database System Architecture and also discusses how to model data using EER Modeling. Different Indexing techniques and the ways to tune the database using Normalization is also discussed. This module also discusses on how to process transactions and optimize the Queries using heuristics query processing.

Module II - Parallel Databases – This module discusses about the different types of Parallel Architecture, Data partitioning strategy, Interquery and Intraquery Parallelism. It also throws light on Parallel Query Optimization.

Module III - Distributed Databases -This segment discusses about Distributed Database Architecture, its features and the techniques called Fragmentation and Replication. Distributed Query Processing is elaborated here to find the optimum transfer cost of the query. Distributed Transactions Processing covers the techniques to handle a transaction in that kind of environment.

Module IV –Spatial and Mobile Databases –This module gives an introduction to Spatial and Mobile databases. It also discusses the types of spatial data, indexing in spatial databases and also the Transaction Model in MDS

Module V - Semi Structured Databases – This Segment introduces the concept of Semi Structured databases, XML, How to validate the XML file against XML Schema. How to Query the database using XQuery. It also discusses about the role of semantic web and how to model data using RDF.

Module VI - Introduction to Database Security Issues – This module discusses the Security Models, the different Threats to databases and Counter measures to deal with these problems

Module VII - Emerging Technologies- This module discusses about the emerging and current technologies like Cloud databases, Streaming Databases, Graph Databases, New SQL databases and Recent Trends and a few case studies.

What is the format of the course?

This Course is designed with 100 minutes of in-classroom sessions per week, 100 minutes of lab hours per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course should have the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, and assignment.

How are students assessed?

- Students are assessed on a combination of group activities, classroom discussion, projects and continuous, final assessment tests.
- Additional weightage will be given to students working with projects based on different databases, and competitions and projects handling with large databases.
- Students can earn additional weightage based on certificate of completion of a related MOOC course.

Session wise plan

Class Hour	Lab Hour	Topics Covered	Level of mastery	Text/Reference Book	Remarks
1		Data base system Architecture	Familiarity	1,2	
1	2	ER /EER Modeling	Usage, Assessment	1,2	
1		Indexing and its types	Familiarity	1,2	
1		Normalization	Familiarity, Assessment	1,2	
1	5	SQL Query Processing and Optimization	Familiarity, Assessment	1,2	
1		Transaction Processing	Familiarity	1,2	
1		Introduction to parallel databases, Architecture	Familiarity	1,2	
1	2	Data partitioning Strategy	Usage	1,2	
1		Interquery and Intraquery parallelism	Familiarity	1,2	
1		Introduction to Distributed databases, Architecture	Familiarity	1,2	
1		Fragmentation	Familiarity	1,2	
1		Replication	Familiarity	1,2	
1	3	Distributed Query Processing	Usage, Assessment	1,2	
1		Distributed Transaction Processing	Usage, Assessment	1,2	
1	4	Introduction to spatial databases, Types of spatial data	Usage	1,2	
1		Indexing in spatial databases	Familiarity	1,2	
1	2	Mobile databases , transaction in MDS	Familiarity	1,2	
1	3	XML –introduction, DTD		1,2	
1		Schema and validating XML with it, Xquery	Usage, Assessment	1,2	
1		Semantic web, Ontologies	Familiarity	1,2	
1	3	RDF, RDFS	Usage, Assessment	1,2	
1		Introduction to Database Security Issues	Familiarity	1,2	
1		Security Models	Familiarity	1,2	
1		Different threats to databases	Familiarity,	1,2	

			1Assessments		
1		Counter Measures to deal with problems	Familiarity	1,2	
1		Cloud databases	Familiarity, Usage		
1		Streaming Databases	Familiarity		
1	6	Graph Databases ,New SQL Databases	Familiarity, Usage		
2		Recent Trends	Familiarity		
30 Hours/2 hours per week	30 Hours/2 Hours per week				

CSE5007	Exploratory Data Analysis	L,T,P,J,C 2,0,0,4,3
Objective	This course introduces the methods for data preparation and data understanding. It covers essential exploratory techniques for understanding multivariate data by summarizing it through statistical methods and graphical methods.	
Expected Outcomes	After successfully completing the course the student should be able to 1. Handle missing data in the real world data sets by choosing appropriate methods. 2. Summarize the data using basic statistics. 3. Visualize the data using basic graphs and plots. 4. Identify the outliers if any in the data set. 5. Choose appropriate feature selection and dimensionality reduction techniques for handling multi-dimensional data.	
SLO's	2,7,17	

Module	Topics	LHrs	SLO
1	**INTRODUCTION TO EXPLORATORY DATA ANALYSIS** Data Analytics life cycle, Exploratory Data Analysis (EDA) –Definition, Motivation, Steps in data exploration, The basic data types, Data Type Portability	3	2
2	**PREPROCESSING - TRADITIONAL METHODS AND MAXIMUM LIKELIHOOD ESTIMATION** Introduction to Missing data, Traditional methods for dealing with missing data, Maximum Likelihood Estimation – Basics, Missing data handling, Improving the accuracy of analysis	4	7
3	**PREPROCESSING - BAYESIAN ESTIMATION** Introduction to Bayesian Estimation, Multiple Imputation - Imputation Phase, Analysis and Pooling Phase, Practical Issues in Multiple Imputation, Models for Missing Not at Random Data	4	7
4	**DATA SUMMARIZATION & VISUALIZATION** Statistical data elaboration, 1-D Statistical data analysis, 2-D Statistical data Analysis, N-D Statistical data analysis	4	7
5	**OUTLIER ANALYSIS** Introduction, Extreme Value Analysis, Clustering based, Distance Based and Density Based outlier analysis, Outlier Detection in Categorical Data	3	7
6	**FEATURE SUBSET SELECTION** Feature selection algorithms: filter methods, wrapper methods and embedded methods, Forward selection, backward elimination, Relief, greedy selection, genetic algorithms for feature selection	4	7

7	**DIMENSIONALITY REDUCTION** Introduction, Principal Component Analysis (PCA), Kernel PCA, Canonical Correlation Analysis, Factor Analysis, Multidimensional scaling, Correspondence Analysis	6	7
8	**Recent Trends**	2	2
Project# Generally a team project[5to10 members] # Concepts studied in this course should have been used # 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 of3 reviews. Projects may be given as group projects The following is the sample project that can be given to students to be implemented: 1. Exploring the data sets for Data Science problems from Kaggle website 2. Applying exploratory data analysis in the field of biometrics for reliable and robust identification of humans from their personal traits, mainly for security and authentication purposes 3. Analyze the dataset for Fraud Detection, Customer segmentation etc. Note: Students can down load real-time data sets for different Machine Learning Tasks fromhttps://archive.ics.uci.edu/ml/datasets.htmland http://sci2s.ugr.es/keel/datasets.php#sub1 and do the projects		**60**[Non Contact hrs]	17

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

Exploratory Data Analysis

Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	**Total Hours of Coverage**
CS: IS (Intelligent Systems)	26
CS:CN (Computational Science)	4

Body of Knowledge coverage

[List the Knowledge Units covered in whole or in part in the course. If in part, please indicate which topics and/or learning outcomes are covered. For those not covered, you might want to indicate whether they are covered in another course or not covered in your curriculum at all. This section will likely be the most time-consuming to complete, but is the most valuable for educators planning to adopt the CS2013 guidelines.]

KA	**Knowledge Unit**	**Topics Covered**	**Hours**
CS: IS	Basic concepts	**Introduction to Exploratory Data Analysis**	3
CS: IS	Data Preparation	**Preprocessing - Traditional methods and Maximum Likelihood Estimation** **Preprocessing - Bayesian Estimation**	8
CS: CN	Data Understanding	**Data Summarization & Visualization**	4
CS: IS	Advanced concepts	**Outlier Analysis, Feature Subset Selection, Dimensionality Reduction**	13
		Recent Trends	2
		Total hours	**30**

What is covered in the course?

[A short description, and/or a concise list of topics - possibly from your course syllabus.(This is likely to be your longest answer)]

Part1: Introduction to Exploratory Data Analysis

It introduces the data analytics life cycle and the role of Exploratory data analysis, various steps in elaborating data, basic types of data and its portability.

Part II: Preprocessing Traditional Methods And Maximum Likelihood Estimation

It introduces the data preparation on how to handle missing data with traditional methods, maximum likelihood estimation.

Part III: Preprocessing - Bayesian Estimation

It introduces the data preparation on how to handle missing data with Bayesian estimation, , Multiple Imputation - Imputation Phase, Analysis and Pooling Phase, Practical Issues in Multiple Imputation, Models for Missing Not at Random Data.

Part IV: Data Summarization and Visualization

It introduces the process of understanding the data through basic statistics – 1-D statistical analysis, 2-D statistical analysis, N-D statistical analysis and the visualization of data.

Part V: Outlier Analysis

It introduces the concept of outlier analysis and discusses various methods on outlier analysis: Extreme Value Analysis, Clustering based, Distance Based and Density Based outlier analysis, Outlier Detection in Categorical Data

Part VI: Feature Subset Selection

It introduces various feature subset selection techniques like filter methods, wrapper methods, embedded methods, forward selection, backward elimination, Relief, greedy selection, genetic algorithms for feature selection

Part VII: Dimensionality Reduction

It introduces dimensionality reduction techniques like Principal Component Analysis (PCA), Kernel PCA, Canonical Correlation Analysis, Factor Analysis, Multidimensional scaling, Correspondence Analysis

Part VIII: Recent Trends

What is the format of the course?

[*Is it face to face, online or blended? How many contact hours? Does it have lectures, lab sessions, and discussion classes?*]

This Course is designed with 100 minutes of in-class room sessions per week as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course should have the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, quizzes.

How are students assessed?
[*What type, and number, of assignments are students are expected to do? (papers, problem sets, programming projects, etc.). How long do you expect students to spend on completing assessed work?]*

- Students are assessed on a combination group activities, classroom discussion, projects, and continuous, final assessment tests.
- Additional weightage will be given based on their rank in crowd sourced projects/ Kaggle like competitions.
- Students can earn additional weightage based on certificate of completion of a related MOOC course.

Additional topics
[List notable topics covered in the course that you do not find in the CS2013 Body of Knowledge]

Other comments
[optional]

Session wise plan

Student Learning Outcomes Covered: 2, 7, 17

Class Hour	Topic Covered	levels of mastery	Reference Book	Remarks
1	Data Analytics life cycle, Exploratory Data Analysis (EDA) –Definition, Motivation	Familiarity	1,4	
2	Steps in data exploration, The basic data types, Data Type Portability	Familiarity	1, 4	
2	Introduction to Missing data, Traditional methods for dealing with missing data	Usage	2	

2	Maximum Likelihood Estimation – *Basics, Missing data handling, Improving the accuracy of analysis*	Usage	2	
2	Introduction to Bayesian Estimation, Multiple Imputation - *Imputation Phase, Analysis and Pooling Phase*,	Usage	2	
2	Practical Issues in Multiple Imputation, Models for Missing Not at Random Data	Familiarity	2	
1	Statistical data elaboration, 1-D Statistical data analysis	Usage	4	
2	2-D Statistical data Analysis, N-D Statistical data analysis	Usage	4	
1	Extreme Value Analysis, Clustering based, Distance Based	Usage	1, 4	
2	Density Based outlier analysis, Outlier Detection in Categorical Data	Usage	1, 4	
2	Feature selection algorithms: filter methods, wrapper methods and embedded methods for feature selection	Usage	5	
2	Forward selection, backward elimination, Relief, greedy selection, genetic algorithms for feature selection	Usage	5	
2	Principal Component Analysis, Kernel PCA	Usage	3	
2	Multidimensional scaling, Correspondence Analysis	Familiarity		
1	Canonical Correlation Analysis	Familiarity	3	
1	Factor Analysis	Familiarity	3	
2	Recent Trends			

CSE6001	**Big Data Frameworks**	**L,T,P,J,C** **2,0,2,4,4**
Objective	The course objective is to impart an understanding of the challenges in storing and processing big data and how to use different big data frameworks effectively to store and process big data.	
Expected Outcomes	After successfully completing the course, the student should be able to a) Discuss the challenges in Big Data. b) Describe the need of different big data frameworks. c) Write MapReduce programming in both Hadoop and Spark Framework. d) Write programs in Spark Streaming, SPARK SQL and GraphX	
SLO's	2. Having a clear understanding of the subject related concepts and of contemporary issues 9. Having problem solving ability- solving social issues and engineering problems 14. Having an ability to design and conduct experiments, as well as to analyze and interpret data	

Module	**Topics**	**LHrs**	**SLO**
1	**INTRODUCTION TO BIG DATA** 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	3	2
2	**Hadoop Framework** 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	6	2,9
3	**Hadoop Ecosystem** Introduction to Hadoop ecosystem technologies: Serialization: AVRO, Co-ordination: Zookeeper, Databases: HBase, Hive, Scripting language: Pig, Streaming: Flink, Storm	3	2,9

4	**Spark Framework** 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	5	2,9
5	**Interactive Data Analysis with Spark Shell** Writing Spark Application - Spark Programming in Scala, Python, R, Java - Application Execution	3	2,9
6	**Spark SQL and GraphX** SQL Context – Importing and Saving data – Data frames – using SQL – GraphX overview – Creating Graph – Graph Algorithms	5	2,9
7	**Spark Streaming** Overview – Errors and Recovery – Streaming Source – Streaming live data with spark	3	2,9
8	**Recent Trends**	2	2
Lab (Indicative List of Experiments (in the areas of) 1. HDFS Commends 2. MapReduce Program to show the need of Combiner 3. MapReduce I/O Formats –Text, key- value 4. MapReduce I/O Formats - NLine, Multiline 5. Sequence file Input / Output Formats 6. Secondary sorting 7. Distributed Cache & Map Side Join, Reduce side Join 8. Building and Running a Spark Application 9. Wordcount in Hadoop and Spark 10. Manipulating RDD 11. Inverted Indexing in Spark 12. Sequence alignment problem in Spark 13. Implementation of Matrix algorithms in Spark 14. Spark Sql programming 15. Building Spark Streaming application		**30**	14

Project # Generally a team project[5 to10 members] # Concepts studied in this course should have been used # 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 of3 reviews. Projects may be given as group projects The following is the sample project that can be given to students to be implemented: 1. Predicting forest cover 2. Anomaly detection 3. Text Analytics 4. Co-occurrence of terms in social networks using GraphX 5. HITS algorithm 6. Geospatial and Temporal data analytics	**60**[Non Contact hrs]	17

Reference Books

1. Mike Frampton, "Mastering Apache Spark", Packt Publishing, 2015.
2. Tom White, "Hadoop: The Definitive Guide", O'Reilly, 4th Edition, 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, "MapReduce Design Pattern", O'Reilly, 2012

Big Data Frameworks

Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	**Total Hours of Coverage**
CS: AL(Algorithms and Complexity) / CE: CAO	3
CS: PL(Programming Languages) / CE: CAO	24
CS: DS / CE: DSC	3

Body of Knowledge coverage

[List the Knowledge Units covered in whole or in part in the course. If in part, please indicate which topics and/or learning outcomes are covered. For those not covered, you might want to indicate whether they are covered in another course or not covered in your curriculum at all. This section will likely be the most time-consuming to complete, but is the most valuable for educators planning to adopt the CS2013 guidelines.]

KA	Knowledge Unit	Topics Covered	Hours
CE: AR	Memory System Organization and Architecture	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	3
CE: PD CS: PL / CE: PRF	Parallel algorithms, Analysis and Programming Language Pragmatics	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 SQL Context – Importing and Saving data – Data frames Writing Spark Application - Spark Programming in Scala, Python, R, Java - Application Execution	12
CS: PL / CE: PRF	Advanced Programming Constructs	Map side join, Reduce Side Join, Secondary sorting, Pipelining MapReduce jobs Overview – Errors and Recovery – Streaming Source – Streaming live data with spark Introduction to Hadoop ecosystem technologies: Serialization: AVRO, Co-ordination: Zookeeper, Databases: HBase, Hive, Scripting language: Pig, Streaming: Flink, Storm	12

CE/NC CS: DS / CE: DSC	Social networking Graphs and Trees	using SQL – GraphX overview – Creating Graph – Graph Algorithms	3
		Total Hours	30

What is covered in the course?
[A short description, and/or a concise list of topics - possibly from your course syllabus.(This is likely to be your longest answer)]

Part 1: Introduction to Big Data
This part of the course gives introduction to the basics of big data, characteristics of big data, challenges involved and the need for big data frameworks

Part II: Hadoop Framework
Describes the Hadoop Architecture and compares it with legacy distributed computing. This part of the course also introduces data storage in Hadoop and writing MapReduce code. The essential ecosystems of Hadoop are introduced in this part.

Part III: Spark and Streaming
This part of the course, introduces Spark tool, Graph algorithms and streaming. Spark will lead to interactive data analysis and supports streaming.

What is the format of the course?
[*Is it face to face, online or blended? How many contact hours? Does it have lectures, lab sessions, and discussion classes?*]

This Course is designed with 100 minutes of in-classroom sessions per week, 100 minutes of lab hours per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course has the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, quizzes.

How are students assessed?
[*What type, and number, of assignments are students are expected to do? (papers, problem sets, programming projects, etc.). How long do you expect students to spend on completing assessed work?]*

- Students are assessed on a combination group activities, classroom discussion, projects, & continuous and final assessment tests.

- Additional weightage will be given based on their rank in crowd sourced projects.

- Students can earn additional weightage based on certificate of completion of a related MOOC course or any online course completion.

Session wise plan

Student learning Outcomes Covered : 2, 9, 14

S.No	Topic Covered	Class Hour	Levels of mastery	Reference Book
1	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	3	Familiarity	2
2	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	6	Usage	2, 5
3	Introduction to Hadoop ecosystem technologies: Serialization: AVRO, Co-ordination: Zookeeper, Databases: HBase, Hive, Scripting language: Pig, Streaming: Flink, Storm	3	Familiarity	2, 3
4	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	5	Familiarity	1, 3, 4
5	Writing Spark Application - Spark Programming in Scala, Python, R, Java - Application Execution	3	Usage	1, 3, 4
6	SQL Context – Importing and Saving data – Data frames – using SQL – GraphX overview – Creating Graph – Graph Algorithms	5	Usage	1, 3, 4
7	Overview – Errors and Recovery – Streaming Source – Streaming live data with spark	3	Usage	1, 3, 4
8	**Recent Trends**	2		
Total hours		*30*		

CSE6005	Machine Learning	L, T, P, J, C 2,0,2,4,4
Objective	It introduces theoretical foundations, algorithms, methodologies, and applications of Machine Learning and also provide practical knowledge for handling and analysing data sets covering a variety of real-world applications.	
Expected Outcomes	After successfully completing the course the student should be able to 1. Recognize the characteristics of machine learning that make it useful to solve real-world problems. 2. Identify real-world applications of machine learning. 3. Identify and apply appropriate machine learning algorithms for analyzing the data for variety of problems. 4. Implement different machine learning algorithms for analyzing the data 5. Design test procedures in order to evaluate a model 6. Combine several models in order to gain better results 7. Make choices for a model for new machine learning tasks based on reasoned argument	
SLO's	7. Having computational thinking (Ability to translate vast data in to abstract concepts and to understand database reasoning) 14. Having an ability to design and conduct experiments, as well as to analyze and interpret data 17. Having an ability to use techniques, skills and modern engineering tools necessary for engineering practice	

Module	Topics	L Hrs	SLO
1	**INTRODUCTION TO MACHINE LEARNING** Introduction, Examples of Various Learning Paradigms, Perspectives and Issues, Version Spaces, Finite and Infinite Hypothesis Spaces, PAC Learning, VC Dimension.	3	7
2	**Supervised Learning** 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	9	7
3	**Ensemble Learning** Model Combination Schemes, Voting, Error-Correcting Output Codes, Bagging: Random Forest Trees, Boosting: Adaboost, Stacking	3	7

4	**Unsupervised Learning** Introduction to clustering, Hierarchical: AGNES, DIANA, Partitional: K-means clustering, K-Mode Clustering, Expectation Maximization, Gaussian Mixture Models	5	7
5	**Probabilistic Learning** Bayesian Learning, Bayes Optimal Classifier, Naïve Bayes Classifier, Bayesian Belief Networks	3	7
6	**Learning Association Rules** Mining Frequent Patterns - basic concepts -Apriori algorithm, FP-Growth algorithm, Association-based Decision Trees	3	7
7	**Machine Learning in Practice** Design, Analysis and Evaluation of Machine Learning Experiments, Other Issues: Handling imbalanced data sets	2	7
8	**Recent Trends**	2	7
Lab (Indicative List of Experiments (in the areas of) 1. Implement Decision Tree learning 2. Implement Logistic Regression 3. Implement classification using Multilayer perceptron 4. Implement classification using SVM 5. Implement Adaboost 6. Implement Bagging using Random Forests 7. Implement K-means Clustering to Find Natural Patterns in Data 8. Implement Hierarchical clustering 9. Implement K-mode clustering 10. Implement Association Rule Mining using FP Growth 11. Classification based on association rules 12. Implement Gaussian Mixture Model Using the Expectation Maximization 13. Evaluating ML algorithm with balanced and unbalanced datasets 14. Comparison of Machine Learning algorithms 15. Implement k-nearest neighbours algorithm		**30**	14
Project# Generally a team project [5 to 10 members] # Concepts studied in this course should have been used # 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.		**60** [Non Contact hrs]	17

Projects may be given as group projects The following is the sample project that can be given to students to be implemented: 1. Solving Data Science problems from Kaggle website 2. Applying Machine Learning algorithms in the field of biometrics for reliable and robust identification of humans from their personal traits, mainly for security and authentication purposes 3. Applying Machine Learning for OCR, Video Analytics 4. Applying Machine Learning algorithms in the field of Natural Language Processing for document clustering and sentiment analysis 5. Applying Machine Learning for Fraud Detection, Customer segmentation etc. Note: Students can down load real time data sets for different Machine Learning Tasks from https://archive.ics.uci.edu/ml/datasets.html and http://sci2s.ugr.es/keel/datasets.php#sub1 and do the projects		

Reference Books

1. Ethem Alpaydin,"Introduction to Machine Learning", MIT Press, Prentice Hall of India, Third Edition 2014.
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 Han and Micheline Kambers and Jian Pei, "Data Mining –Concepts and Techniques", 3rd edition, Morgan Kaufman Publications, 2012.

Machine Learning

Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	**Total Hours of Coverage**
CS: IS(Intelligent System)	30

Body of Knowledge coverage

[List the Knowledge Units covered in whole or in part in the course. If in part, please indicate which topics and/or learning outcomes are covered. For those not covered, you might want to indicate whether they are covered in another course or not covered in your curriculum at all. This section will likely be the most time-consuming to complete, but is the most valuable for educators planning to adopt the CS2013 guidelines.]

KA	**Knowledge Unit**	**Topics Covered**	**Hours**
CS: IS	IS/Basic Machine Learning	**Introduction to Machine Learning**	3
CS: IS	IS/Advanced Machine Learning	Supervised Learning Ensemble Learning Unsupervised Learning Probabilistic Learning Learning Association Rules Machine Learning in Practice Recent Trends	27
		Total hours	**30**

What is covered in the course?
[A short description, and/or a concise list of topics - possibly from your course syllabus.(This is likely to be your longest answer)]

Part 1: Introduction to Machine Learning

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

Part II: Supervised Learning

This chapter covers supervised learning algorithms for classification tasks. The algorithms covered are the following: 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

Part III: Ensemble Learning

This chapter covers ensemble learning algorithms for classification tasks. Model Combination Schemes, Voting, Error-Correcting Output Codes, Bagging: Random Forest Trees, Boosting: Adaboost, Stacking

Part IV: Unsupervised Learning

This chapter covers unsupervised learning algorithms for clustering tasks. The algorithms covered are the following: Introduction to clustering, Hierarchical: *AGNES*, *DIANA*, Partitional: *K-means clustering*, *K-Mode Clustering*, Expectation Maximization, Gaussian Mixture Models

Part V: Probabilistic Learning

This chapter covers learning algorithms based on Bayesian theory.Bayesian Learning, Bayes Optimal Classifier, Naïve Bayes Classifier, Bayesian Belief Networks

Part VI: Learning Association Rules

This chapter covers learning association rules from the data. The algorithms covered are the following: Mining Frequent Patterns - basic concepts -Apriori algorithm, FP-Growthalgorithm, Association-based Decision Trees

Part VII: Machine Learning in Practice

This chapter covers necessary points to be taken when applying machine learning algorithms on the data. Also discuss about evaluation metrics and methods for comparison of Machine learning algorithms.

Part VIII: Recent Trends

What is the format of the course?
[*Is it face to face, online or blended? How many contact hours? Does it have lectures, lab sessions, discussion classes?*]

This Course is designed with 100 minutes of in-classroom sessions per week, 100 minutes of lab hours per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally, this course should have the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, quizzes.

How are students assessed?
[*What type, and number, of assignments are students are expected to do? (papers, problem sets, programming projects, etc.). How long do you expect students to spend on completing assessed work?]*

- Students are assessed on a combination group activity, classroom discussion, projects, and continuous, final assessment tests.

- Additional weightage will be given based on their rank in crowd sourced projects/ Kaggle like competitions.

- Students can earn additional weightage based on certificate of completion of a related MOOC course.

Additional topics
[List notable topics covered in the course that you do not find in the CS2013 Body of Knowledge]

Other comments
[optional]

Session wise plan

Student Outcomes Covered: 7, 14, 17

Class Hour	Lab Hour	Topic Covered	levels of mastery	Reference Book	Remarks
1		Introduction, Examples of Various Learning Paradigms	Familiarity	1,2	
1		Perspectives and Issues	Familiarity	1, 2	
1		Version Spaces, Finite and Infinite Hypothesis Spaces, PAC Learning, VC Dimension	Familiarity	1,2	
2		Decision Trees: ID3, Classification and Regression Trees	Usage	1	
2		Regression: *Linear Regression, Multiple Linear Regression, Logistic Regression*	Usage	1	
1		Neural Networks: *Introduction, Perceptron*		3	
1		Multi-layer Perceptron	Usage	3	
1		Support vector Machines - Linear	Usage	1,4	
1		Support vector Machines – Non-Linear, kernel functions		1,4	

1		K-nearest neighbour	Usage	3	
1		Model Combination Schemes, Voting, Error-Correcting Output Codes, Stacking	Usage	1,4	
1		Bagging: Random Forest Trees	Usage	1,4	
1		Boosting: Adaboost	Usage	1,4	
2		Introduction to clustering, Hierarchical Clustering: *AGNES, DIANA*	Usage	5	
2		Partitional K-means clustering, K-mode Clustering	Usage	5	
1		Expectation Maximization, Gaussian Mixture Models	Usage	5	
2		Bayesian Learning, Bayes Optimal Classifier, Naïve Bayes Classifier	Usage	3	
1		Bayesian Belief Networks	usage	3	
1		Mining Frequent Patterns - basic concepts –Apriori algorithm	Usage	7	
1		FP-Growth algorithm	Usage	7	
1		Association-based Decision Trees	Usage	1,6	
1		Design, Analysis and Evaluation of Machine Learning Experiments	Usage	6	

1		Comparison of Machine Learning algorithms, Other Issues: Handling imbalanced data sets		6	
2		Recent Trends			

CSE5002	Operating systems and Virtualization	L,T,P,J,C 2,0,2,0,3
Objectives	This course introduces to virtualization, operating systems fundamental concepts and its technologies. This course provides the skills to write programs that interact with operating systems components such as Processes, Thread, Memory during concurrent execution. This course provides the skills and knowledge necessary to implement, provisioning and administer server and desktop virtualization.	
Expected Outcomes	After successfully completing the course the student should be able to • Categorize the levels of abstraction in a computer system correspond to implementation layers in both hardware and software. • Apply and design the procedure used for concurrency and memory management. • Comprehend the basics of virtualization and to differentiate types of Virtualization. • Develop and provision server and desktop virtualization • Analyze the inner-working of a Virtual Machine and its Management	
Student Learning Outcomes	• Having a clear understanding of the subject related concepts and of contemporary issues (2) • Having design thinking capability(5) • Having an ability to use techniques, skills and modern engineering tools necessary for engineering practice(17)	

Module	Topics	L Hrs	SLO
1	Computer system architecture a layered view with interfaces - Glenford Myer, Monolithic Linux & Hybrid Windows 10 kernels Layered architecture of operating system and core functionalities	2	2
2	Process 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	4	5
3	Memory Introduction, Address Spaces, Memory API, Address Translation, Paging - Faster Translations (TLB), Smaller Tables. Virtual Memory System in x86	4	5
4	Concurrency	4	2

	Introduction, Thread Models, Thread API, Building & Evaluating a Lock, Test And Set, Two phase lock, Classical problems handling using semaphore		
4	Persistence - File Organization: The i-node, Crash Consistency & file security	2	5
5	Virtual Machines Process and System VMs Taxonomy of VMs	2	2
6	Types of Virtualization- Hardware Emulation Full Virtualization with binary translation, Hardware assisted, Operating System Virtualization OS assisted /Para virtualization	4	17
7	Hypervisor Type 1, Type 2	1	17
	Para-virtualization Server Virtualization, Desktop Virtualization	3	17
	Overview VM portability Clones, Templates, Snapshots, OVF, Hot and Cold Cloning Protecting & Increasing Availability Light Weight Virtual machine: Container / Docker	2	2
8	Recent Trends	2	2
		30	
List of Lab Experiments		30 hrs	14

Each experiment should require the student to submit a system analysis & design document that describes the details of the experiment. The experiment may be submitted before the next lab if not completed within class hours. Collaboration and discussion with co-students on the experiments is encouraged. How ever plagiarism will be penalized severely as per University regulations.

1. Retrieval of System data file and its information
2. Write a program to simulate multi-level queue scheduling algorithm

(ex) All the processes in the system are divided into two categories – system processes and user processes. System processes are to be given higher priority than user processes. Use FCFS scheduling for the processes in each queue.

3. Program to demonstrate Process Life Cycle
4. Process Synchronization using Binary and Counting Semaphore
5. Deadlock Avoidance Using Semaphores: Implement the deadlock-free solution to Dining Philosophers problem to illustrate the problem of deadlock and/or starvation that can occur when many synchronized threads are competing for limited resources.
6. Implementation of memory management scheme - Paging / Segmentation

7. Parallel Thread management using pthread library.

Implement a data parallelism using multi threading. Application should have Thread create, Thread synchronization, Thread termination. In the program, every thread must return the value and must be synchronized in the main function. Final consolidation should be done by main thread (main function)

8. Type II Hypervisor setup and configuration Create Virtual Machines (VM)
   Create Snapshot of all Virtual Machines Create clone of VM
   Configure Networking
9. Type I Hypervisor setup and configuration Virtual machines on Xen
   Migration of VMs Create a Virtual Network Switch
   Create Differencing Hard Disks for use by Virtual Machines Associate Hard Disks with Virtual Machines
10. VM migration and Backup and recovery virtual machines in type I
11. Implementation of Server virtualization – multiple servers on single system
12. Implementation of Desktop virtualization

Reference Books

1. Thomas Anderson, Michael Dahlin , "Operating Systems: Principles and Practice" , Second Edition, Recursive Books,2014
2. William Stallings," Operating Systems: Internals and Design Principles", 8th Edition, 2014.
3. Remzi H. Arpaci-Dusseau and Andrea C, "Operating Systems: Three Easy Pieces", . Arpaci- Dusseau , University of Wisconsin – Madison,2015
4. A. Silberschatz and P. Galvin. "Operating System Concepts". Eight Edition, John Wiley & Sons, 2008
5. Matthew Portnoy, "Virtualization Essentials", John Wiley & Sons Inc; 2nd Edition edition (8 September 2016)
6. Smith, Nair, "Virtual Machines: Versatile Platforms for Systems and Processes", Morgan Kaufmann Publishers (2005)

Knowledge Areas that contain topics and learning outcomes covered in the

course Knowledge Area of Coverage	Total Hours
CS-OS	30

Body of Knowledge coverage

KA	Knowledge Unit	Topics Covered	Hours
CS	OS/Overview of Operating Systems	Computer system architecture a layered view with interfaces - Glenford Myer, Monolithic Linux & Hybrid Windows 10 kernels. Layered architecture of operating system and core function	**2**
CS	OS/Operating System Principles	Introduction, Process Operations, States, Context switching, Data Structures(Process Control Block (PCB),	**2**
CS	OS/Scheduling and Dispatch	Process Scheduling: Multi-Level Feedback Queue, Multi-processor Scheduling, Deadlocks and its detection	**2**
CS	OS/Memory Management	Address Spaces, Memory API, Address Translation, Paging - Faster Translations (TLB), Smaller Tables. Virtual Memory System in x86	**4**
CS	OS/Concurrency	Thread Models, Thread API, Building & Evaluating a Lock, Test And Set, Two phase lock, Classical problems handling using semaphore	**4**
CS	OS/File Systems	Persistence - File Organization: The i-node, Crash Consistency & file security	**2**
CS	OS/Virtual Machines	Virtual Machines, Types of Virtualization, Hypervisor - Type 1, Type 2, Server Virtualization, Desktop Virtualization, Overview VM portability, Light Weight Virtual machine	**14**

Where does the course fit in the curriculum?

[*In what year do students commonly take the course? Is it compulsory? Does it have pre- requisites, required following courses? How many students take it?*]
This course is a

- An Elective Course
- Suitable from first semester onwards.

What is the format of the course?

[*Is it face to face, online or blended? How many contact hours? Does it have lectures, lab sessions, discussion classes?*]

This Course is designed with 100 minutes of in-classroom sessions per week, 60 minutes of video/reading instructional material per week, 100 minutes of lab hours per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course should have the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory of pre-class reading material, quizzes.

How are students assessed?

[*What type, and number, of assignments are students are expected to do? (papers, problem sets, programming projects, etc.). How long do you expect students to spend on completing assessed work?]*

- Students are assessed on a combination group activities, classroom discussion, and continuous, final assessment tests.
- Additional weightage will be given based on their rank in crowd sourced projects/ Kaggle like competitions and certificate of completion of a related MOOC course.

Session wise Plan

Students Outcome coverage : 2,5,17

Class Hour	Topics Coverage	Levels of Mastery	Referenc e Book	Remarks
2	Computer system architecture a layered view with interfaces - Glenford Myer, Monolithic Linux & Hybrid Windows 10 kernels. Layered architecture of operating system and core function	Familiarity	6	
2	Introduction, Process Operations, States, Context switching, Data Structures(Process Control Block (PCB),	Familiarity	3	
2	Process Scheduling: Multi-Level Feedback Queue, Multi-processor Scheduling, Deadlocks and its detection	Usage	3	
4	Address Spaces, Memory API, Address Translation, Paging - Faster Translations (TLB), Smaller Tables. Virtual Memory System in x86	Familiarity	3	
4	Thread Models, Thread API, Building & Evaluating a Lock, Test And Set, Two phase lock, Classical problems handling using semaphore	Usage	3	
2	Persistence - File Organization: The i-node, Crash Consistency & file security	Familiarity	4	
12	Virtual Machines, Types of Virtualization, Hypervisor - Type 1, Type 2, Server Virtualization, Desktop Virtualization, Overview VM portability, Light Weight Virtual machine	Usage	5,6	

2	Recent Trends			

	Theory : 30 Hours (2 Credit hours /week D 15 Weeks schedule) Laboratory : 30 Hours (1 Credit hours / week)			

Multicore Architectures

CSE5006	**Multicore Architectures**	**L,T,P,J,C** **2,0,2,0,3**
Preamble	This course is to provide knowledge on multicore architectures that lays the foundation for the development of High Performance Applications through OpenMP, CUDA parallel programming platforms. It enables to analyse the performance of HPC applications using various profiling tools	
Objective of the course	• To provide knowledge on basics of Multicore architectures • To understand concepts of parallel computers and its programming models • To design and develop parallel programs • To practice parallel programming using OpenMP, CUDA parallel programming platforms • To apply program optimizations on parallel programs • To analyse the performance using profiling tools • To explore various contemporary tools and recent trends in field of multicore architectures	
Expected Outcome	After successfully completing the course the student should be able to 1) Describe various parallel programming models 2) Design and develop High Performance Applications using contemporary tools 3) Improve performance of applications through program optimizations 4) Analyse performance of parallel applications	
Student Learning Outcome	2. Having a clear understanding of the subject related concepts and of contemporary issues 11. Having interest in lifelong learning 14. Having an ability to design and conduct experiments, as well as to analyze and interpret data	

Module	Topics	L Hrs	SLO
1	**Introduction to Multi-Core Architectures** Evolution of multicores through Moor's Law, Comparisons of single core, multi-core, multi-processing and hyper threading	2	2
2	**Parallel Computers and programming** Threading Concepts, Communication Architectures and Communication Costs, Thread Level Parallelism(TLP), Instruction Level Parallelism(ILP), Comparisons, Cache Hierarchy and Memory-level Parallelism, Cache Coherence, Parallel programming models, Shared Memory and Message Passing, Vectorization	5	2
3	**OpenMP programming (Open multi-processing)** Introduction to OpenMP, Parallel constructs, Runtime Library routines, Work-sharing constructs, Scheduling clauses, Data environment clauses, atomic, master Nowait Clause, Barrier Construct	5	2
4	**CUDA Programming(Compute Unified Device Architecture)** Introduction to GPU Computing, CUDA Programming Model, CUDA API, Simple Matrix, Multiplication in CUDA , CUDA Memory Model, Shared Memory Matrix Multiplication, Additional CUDA API Features	6	2
5	**Performance Analysers** Trace analyzer and collector (ITAC), VTune Amplifier XE, Energy Efficient Performance, Integrated Performance Primitives (IPP)	4	14
6	**Contemporary tools** MKL (Math Kernel Library), Threading Building Blocks, CUDA Tools	3	14
7	**HTC and MTC** HTC (High Throughput Computing), MTC (Many Task Computing), Top 500 Super computers in the world, Top 10 Super Computer architectural details, Exploring Linpack.	3	14
8	**Recent Trends**	2	11
		30	
Project		**60** [Non Contact hrs]	17

Projects may be given as group projects Design and development of High Performance applications through parallel programming platforms in the following areas Network Security Data Compression Image Processing Bio-Medical Information retrieval Natural Language Processing Health care Applications		

Reference Books

1. Rob Farber, "CUDA Application Design and Development", Morgan Kaufmann Publishers, 2013
2. Shameem Akhter and Jason Roberts, "Multi-Core Programming", 1st edition, Intel Press, 2012
3. Cameron Hughes, Tracey Hughes, "Professional Multicore Programming Design and Implementation for C++ Developers", Wiley, 2008
4. Robert Oshana, "Multicore Software Development Techniques: Applications, Tips, and Tricks", Newnes,1 edition, 2015
5. David B. Kirk , Wen-mei W. Hwu, "Programming Massively Parallel Processors: A Hands-on Approach (Applications of GPU Computing Series)", 1st edition, Morgan Kaufmann, 2010.

Knowledge areas that contain topics and learning outcomes covered in the course

Knowledge Area	Total Hours of Coverage
Systems Fundamentals (SF)	30

Body of Knowledge coverage

[List the Knowledge Units covered in whole or in part in the course. If in part, please indicate which topics and/or learning outcomes are covered. For those not covered, you might want to indicate whether they are covered in another course or not covered in your curriculum at all. This section will likely be the most time-consuming to complete, but is the most valuable for educators planning to adopt the CS2013 guidelines.]

KA	Knowledge Unit	Topics Covered	Hours
SF	Computer Organization	Evolution of multicores through Moor's Law, Comparisons of single core, multi-core, multi-processing and hyper threading.	2
SF	Parallelism	Threading Concepts, Communication Architectures and Communication Costs, TLP, ILP, Comparisons, Cache Hierarchy and Memory-level Parallelism, Cache Coherence, Parallel programming models, Shared Memory and Message Passing, Vectorization	5
SF	Parallel Programming Language	Introduction to OpenMP, Parallel constructs, Runtime Library routines, Work-sharing constructs, Scheduling clauses, Data environment clauses, atomic, master Nowait Clause, Barrier Construct	5
	Heterogeneous architecture and its programming	Introduction to GPU Computing, CUDA Programming Model, CUDA API, Simple Matrix, Multiplication in CUDA , CUDA Memory Model, Shared Memory Matrix Multiplication, Additional CUDA API Features	6
SF	Program Analyzer	Trace analyzer and collector (ITAC), VTune Amplifier XE, Energy Efficient Performance, Integrated Performance Primitives (IPP)	4

SF	Contemporary Tools	MKL (Math Kernel Library), Threading Building Blocks, CUDA Tools	3
SF	HTC and MTC	HTC (High Throughput Computing), MTC (Many Task Computing), Top 500 Super computers in the world, Top 10 Super Computer architectural details, Exploring Linpack.	3
SF	Recent Trends	HTC (High Throughput Computing), MTC (Many Task Computing), Top 500 Super computers in the world, Top 10 Super Computer architectural details, Exploring Linpack.	2
		Total hours	**30**

Where does the course fit in the curriculum?
[*In what year do students commonly take the course? Is it compulsory? Does it have pre-requisites, required following courses? How many students take it?*]

This course is a

- An elective Course.
- Suitable from first semester onwards.

What is covered in the course?
[A short description, and/or a concise list of topics - possibly from your course syllabus.(This is likely to be your longest answer)]

Module 1: Introduction to Multi-Core architecture

Evolution of multi-cores through Moor's Law, Comparisons of single core, multi-core, multi-processing and hyper threading.

Module 2: Parallel Computers & its programming

Threading Concepts, Communication Architectures and Communication Costs, TLP, ILP, Comparisons, Cache Hierarchy and Memory-level Parallelism, Cache Coherence, Parallel programming models, Shared Memory and Message Passing, Vectorization

Module 3: OpenMP Programming
Introduction to OpenMP, Parallel constructs, Runtime Library routines, Work-sharing constructs, Scheduling clauses, Data environment clauses, atomic, master No-wait Clause, Barrier Construct

Module 4: CUDA Programming
Introduction to GPU Computing, CUDA Programming Model, CUDA API, Simple Matrix, Multiplication in CUDA , CUDA Memory Model, Shared Memory Matrix Multiplication, Additional CUDA API Features

Module 5: **Performance Analysers**
Trace analyzer and collector (ITAC), VTune Amplifier XE, Energy Efficient Performance, Integrated Performance Primitives (IPP)

Module 6: Contemporary Tools
MKL (Math Kernel Library), Threading Building Blocks, CUDA Tools

Module 7: HTC and MTC
HTC (High Throughput Computing), MTC (Many Task Computing), Top 500 Super computers in the world, Top 10 Super Computer architectural details, Exploring Linpack.

Module 8: Recent trends

What is the format of the course?
[*Is it face to face, online or blended? How many contact hours? Does it have lectures, lab sessions, discussion classes?*]

This Course is designed with 100 minutes of in-classroom sessions per week. Generally this course should have the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, assignments and quizzes.

How are students assessed?
[*What type, and number, of assignments are students are expected to do? (papers, problem sets, programming projects, etc.). How long do you expect students to spend on completing assessed work?]*

- Students are assessed based on group activities, classroom discussion, assignments, quiz, projects, continuous (CAT) assessment test, and final assessment test.

Additional topics
[List notable topics covered in the course that you do not find in the CS2013 Body of Knowledge]

Cuda Programming, Top 10 Super Computers in the world and Benchmarks

Other comments

Nil

Session wise plan

Student Outcomes Covered: 2, 5, 9, 17

Sl. No.	Class Hour	Topic Covered	Levels of mastery	Reference Book	Remarks
1	2	Evolution of multi-cores through Moor's Law, Comparisons of single core, multi-core, multi-processing and hyper threading.	Usage	2	
2	5	Threading Concepts, Communication Architectures and Communication Costs, TLP, ILP, Comparisons, Cache Hierarchy and Memory-level Parallelism, Cache Coherence, Parallel programming models, Shared Memory and	Usage	2	

		Message Passing, Vectorization			
3	5	Introduction to OpenMP, Parallel constructs, Runtime Library routines, Work-sharing constructs, Scheduling clauses, Data environment clauses, atomic, master Nowait Clause, Barrier Construct	Usage	2	Assignments
4	6	Introduction to GPU Computing, CUDA Programming Model, CUDA API, Simple Matrix, Multiplication in CUDA , CUDA Memory Model, Shared Memory Matrix Multiplication, Additional CUDA API Features	Usage	1	Assignments
5	4	Trace analyzer and collector (ITAC), VTune Amplifier XE, Energy Efficient Performance,	Usage	1	

		Integrated Performance Primitives (IPP)			
6	3	MKL (Math Kernel Library), Threading Building Blocks, CUDA Tools	Usage	2	
7	2	HTC (High Throughput Computing), MTC (Many Task Computing), Top 500 Super computers in the world, Top 10 Super Computer architectural details, Exploring Linpack.	Familiarity		
8		Recent Trends			
	30 Hours (2 Credit hours /week ✍ 15 Weeks schedule)				

CSE6006	NoSQL Databases	L,T,P,J,C 2,0,2,4,4
Objective	This course will explore the origins of NoSQL databases and the characteristics that distinguish them from traditional relational database management systems. This covers the architectures and common features of the main types of NoSQL databases (key-value stores, document databases, column-family stores, graph databases) Finally, 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 Outcomes	After successfully completing the course the student should be able to 1. Explain the detailed architecture, define objects, load data, query data and performance tune NoSQL databases 2. Define NoSQL, its characteristics, history and primary benefits using NoSQL Databases. 3. Define the major types of NoSQL databases including a primary use case and advantages/disadvantages of each type. 4. Analyze semi-structured data and choose an appropriate storage structure	
SLO's	2,7,12	

Module	Topics	L Hrs	SLO
1	**INTRODUCTION TO NOSQL CONCEPTS** 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 database sharing, Brewer's CAP theorem.	4	2
2	**NOSQL DATA ARCHITECTURE PATTERNS** 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.	4	12
3	**KEY –VALUE DATA STORES** 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	5	7

4	**DOCUMENT ORIENTED DATABASE** Document, Collection, Naming, CRUD operation, querying, indexing, Replication, Sharing, Consistency Implementation: Distributed consistency, Eventual Consistency, Capped Collection, Case studies: document oriented database: MongoDB and/or Cassandra	5	7
5	**COLUMNAR DATA MODEL - I** 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.	3	7
6	**COLUMNAR DATA MODEL - II** Advanced techniques: Vectorized Processing, Compression, Write penalty, Operating Directly on Compressed Data Late Materialization Joins , Group-by, Aggregation and Arithmetic Operations, Case Studies	3	7
7	**DATA MODELING WITH GRAPH** Comparison of Relational and Graph Modeling, Property Graph Model Graph Analytics: Link analysis algorithm- Web as a graph, PageRank- 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	4	7
8	**Recent Trends**	2	2

Lab (Indicative List of Experiments (in the areas of)	**30**	14

1. Import the Hubway data into Neo4j and configure Neo4j. 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 (for example 13 means 1pm -2pm) and number of trips which end at the station "B.U. Central"

2. The flight data can be found at http://stat-computing.org/dataexpo/2009/the-data.html . You need to download just one year and from there you can sample a subset of at least 10000 records. You can use the data from a full year if you want but we recommend using a smaller dataset for simplicity.

    Hint: If you need to unzip the data file, you can use the command: bzip2 –d data file from a terminal. For example, for the 2008, you download the file and unzip it using: bzip2 -d 1987.csv.bz2. The airport data can be found at http://stat-computing.org/dataexpo/2009/supplemental-data.html .

    1) Download the flight dataset and airport dataset.
    (2) Clean the dataset (for example: remove columns you do not need, remove records with missing information, remove duplicate records and so on).
    (3) Give the header to csv files
    (4) Import the data into Neo4j.
    (5) Write the queries to answer following questions:
        (5.1) List top 10 airports with most outbound flights.
        (5.2) List top 10 airports with most inbound flights.
        (5.3) List top 5 routes with most flights in weekdays.
        (5.4) List top 5 routes with most flights in weekends.
            List the hour number (for example 13 means 1pm -2pm) and number of flights, which depart from a specific airport in your data (e.g., Boston Logan Airport).
            List the hour number (for example 13 means 1pm -2pm) and number of flights, which arrive at specific airport in your data (e.g., Boston Logan Airport).

    In your report, you should answer the following questions:
    (a) List the year of the flights that you downloaded and prepared for this assignment. You can get a sample set from one-year data. However, the number of flights cannot be smaller than 10k.
    (b) Describe how you clean the data (Which columns you remove and why? Which rows you remove and why?). Hint: You can clean your data by writing a small program in Java, Python, C, Matlab or any kind of programming language.
    (c) Describe the header you give to the csv files.
    (d) Write down the command for importing data.
    (e) Write and execute the queries from step (5) above.

3. Download a zip code dataset at http://media.mongodb.org/zips.json. Use mongoimport to import the zip code dataset into MongoDB.

After importing the data, answer the following questions by using aggregation pipelines:

(1) Find all the states that have a city called "BOSTON".

(2) Find all the states and cities whose names include the string "BOST".

(3) 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.

(4) MongoDB can query on spatial information.

Assume we have a spatial position as [-72, 42], and in the range of 2 (it can be [-71.5, 41.5] or [-72.5, 42.5] or somewhere else), there may exist a number of zip codes . Try to find the states in that range. You should return the total populations and the number of cities of each state in that range. Rank the states based on the number of cities.

(5) Consider a certain rectangular area, in which the vertices are [-80 , 30] , [-90 , 30] , [-90 , 40] and [-80 , 40]. Find and report the top 10 largest cities (by population) in this area.

4. Create a database that stores road cars. Cars have a manufacturer, a type. Each car has a maximum performance and a maximum torque value. Use ifconfig to determine a machine's IP address. To check if Cassandra is running in the background, run: **ps aux | grep cassandr[a]**

Do the following:

5. Test Cassandra's replication schema and consistency models.
6. Network Partition without Replication
7. Network Partition with Replication and Weak Consistency
8. Network Partition with Replication and Quorum Consistency
9. Cars have different powertrains. Each type can be described with different parameters:
10. Internal combustion engine: fuel type, displacement, maximum torque, maximum power
11. Electric motor: maximum torque, maximum power
12. Both: all of the above and the combined maximum torque and power values
13. The class hierarchy for different powertrain types
14. Extend the cars column family to store the powertrain of each car.
15. Write a query that collects the cars with an internal combustion engine.
16. Write a query that collects the cars with an internal combustion engine or an electric motor.

Project Projects may be given as group projects The following is the sample project that can be given to students to be implemented: 1. Analyzing and Visualizing social networks like Facebook and twitter using NoSQL Databases. 2. Using Sample datasets from http://www.rdatamining.com/resources/data, UCLA Repository, kaggle dataset etc., and analyzing those using NoSQL databases. 3. Twitter provides a fire hose of data. Automatically filtering, aggregating, analyzing such data can allow a way to harness the full value of the data, extracting valuable information. The idea of this project is investigating stream processing technology to operate on social streams. 4. Project on Combining Database management and Cloud storage system. 5. CarTel. In the CarTel project, we are building a system for collecting and managing data from automobiles. There are several possible CarTel related projects: 6. One of the features of CarTel is a GUI for browsing geo- spatial data collected from cars. We currently have a primitive interface for retrieving parts of the data that are of interest, but developing a more sophisticated interface or query language for browsing and exploring this data would make a great project. 7. One of the dangers with building a system like CarTel is that it collects relatively sensitive personal information about user's location and driving habits. Protecting this information from casual browsers, insurance companies, or other undesired users is important. However, it is also important to be able to combine different user's data together to do things like intelligent route planning or vehicle anomaly detection. The goal of this project would be to find a way to securely perform certain types of aggregate queries over CarTel data without exposing personally identifiable information. **Reference Books** 1. Guy Harrison, "Next Generation database: NoSQL New SQL and Big Data", Apress, I[st] Edition, 2015 2. Daniel G. McCreary and Ann M. Kelly "Making Sense of NoSQL" Manning publisher, Edition illustrated, 2013 3. Shanshak Tiwari, "Professional NoSQL", Wrox, I[st] Edition, 2011 4. Christopher D. manning, Prabhakar Raghavan, Hinrich Schutze, "An introduction to Information Retrieval", Cambridge University Press, 2008 5. Daniel Abadi, Peter Boncz, Stavros Harizopoulos, "The Design and Implementation of Modern Column-Oriented Database Systems", Now Publisher, 2013. 6. Kristina Chodorow, "Mongo DB the Definitive Guide" O'Reilly Media, 2013.	60	17

2. Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	Theory Hours
CS: IM (Distributed Databases)	30
Total	30

2.1 Body of Knowledge coverage

KA	Knowledge Unit	Topics Covered	Hours
CS: IM	Distributed Databases	**Introduction to NoSQL Concepts** **NoSQL Data Architecture** **Patterns Key –Value Data Stores** **Document Oriented Database** **Columnar Data Model - I** **Columnar Data Model - II** **Data Modeling With Graph** **Recent Trends**	30
		Total hours	30

What is covered in the course?

Module I - INTRODUCTION TO NOSQL CONCEPTS

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 database sharing, Brewer's CAP theorem.

Module II - NOSQL DATA ARCHITECTURE PATTERNS

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 III - KEY –VALUE DATA STORES

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 IV – DOCUMENT ORIENTED DATABASE

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

Module V - COLUMNAR DATA MODEL-I

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 VI - COLUMNAR DATA MODEL-II

Advanced techniques: Vectorized Processing, Compression, Write penalty, Operating Directly on Compressed Data Late Materialization Joins , Group-by, Aggregation and Arithmetic Operations, Inserts/updates/deletes, Indexing, Adaptive Indexing and Database Cracking, Case Studies

Module VII - DATA MODELING WITH GRAPH

Comparison of Relational and Graph Modeling, Property Graph Model

Graph Analytics: Link analysis algorithm- Web as a graph, PageRank- 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 VIII – Recent Trends

What is the format of the course?

This Course is designed with 100 minutes of in-classroom sessions per week, 100 minutes of lab hours per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course should have the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, and assignment.

How are students assessed?

- Students are assessed on a combination of group activities, classroom discussion, projects and continuous, final assessment tests.
- Additional weightage will be given to students working with projects based on different databases, and competitions and projects handling with large databases.
- Students can earn additional weightage based on certificate of completion of a related MOOC course.

Session wise plan

Class Hour	Lab Hour	Topics Covered	Level of mastery	Text/Reference Book	Remarks
2		Data base revolutions : First generation, second generation, third generation, Managing Transactions and Data Integrity ACID and BASE for reliable database transactions	Familiarity	1,2	
2	2	Speeding performance by strategic use of RAM, SSD, and disk Achieving horizontal scalability with database sharing, Brewer's CAP theorem	Familiarity	1,2	
2		NoSQL Data model: Aggregate Models, Document Data model, Key-value Data model, Columnar Data model, Graph Based Data model Graph data model	Familiarity	1,2	
2	2	Ways that NoSQL systems handle big data problems Moving queries to the data, not data to the queries Using hash rings to evenly distribute data on a cluster Using replication to scale reads Letting the database distribute queries evenly to data nodes	Usage, Assessment	1,2	
2		Key-value data stores From Arrays 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 Terms	Familiarity,	1,2	
3	4	Key-Value Implementation Terms Key Design and Partitioning Designing Structured Values Limitations of Key-Value Databases Design Patterns for Key-Value Databases Case Study: Key-Value Databases	Usage, Assessment	1,2	

		for Mobile Application Configuration			
3	4	Document Oriented Database Document, Collection, Naming, CRUD operation, querying, indexing, Replication, Sharing, Consistency Implementation: Distributed consistency, Eventual Consistency	Usage, Assessment	1,2	
2	4	Advanced Topics: Capped Collection, Case studies : document oriented database: MongoDB and/or Cassandra	Usage, Assessment	1,2	
2		Data warehousing schemas: Comparison of columnar and row-oriented storage, Column-store Architectures : C-Store and Vector-Wise	Familiarity	1,2	
4	6	Column-store internals and advanced techniques Vectorized Processing Compression ,Write penalty Operating Directly on Compressed Data Late Materialization Joins , Group-by, Aggregation and Arithmetic Operations , Inserts/updates/deletes Indexing, Adaptive Indexing and Database Cracking	Usage, Assessment	1,2	
2	4	Data modeling with Graph : Comparison of Relational and Graph Modeling, Property Graph Model, Graph Analytics: Link analysis algorithm- Web as a graph, PageRank- Markov chain, page rank computation, Topic specific page rank.	Usage, Assessment	1,2	
2	4	Building a Graph Database Application- community detection Querying Graphs: An Introduction to Cypher	Usage, Assessment	1,2	

2		Recent trends	Familiarly		
30 Hours/2 hours per week	30 Hours/2 Hours per week				

CSE6014	Programming for Data Science	L, T, P, J, C 0, 0, 4, 0, 2
Objective:	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 Outcomes	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.bility to perform exploratory data analysis using Python and R	
SLO	17	

S.No	Topics	Lab Hrs	S LO's	Reference Books
1	Expressions, Operators, matrices, Decision Statements in python	2	17	1,2
2	Control Flow and Functions in python	2		1,2
3	Classes, Objects, Packages and Files in python	2		1,2
4	Tuple, Lists, Sequences, Dictionaries, Comprehensions.	2		1,2
5	Numpy Arrays objects, Creating Arrays, basic operations, Indexing, Slicing and iterating, copying arrays, shape manipulation, Identity array, eye function, Universal function	2		4,7,8
6	Linear algebra with Numpy, eigen values and eigen vectors with Numpy	2		4,7,8
7	Pandas series Object, Pandas data Frame Objects: Data Aggregation and Joining,	2		3,4,5
8	Pandas Object: Concatenating and appending data frames, index objects	2		3,4,5
9	Handling Time series data using pandas	2		3,4,5
10	Handling missing values using pandas	2		3,4,5
11	Reading and writing the data including JSON data	2		4,5
12	Web scraping using python, Combining and merging datasets	2		4,5,6
13	Data transformations	2		4,5,6
14	Basic matplotlib plots, common plots used in statistical analysis in python	2		4
15	common plots used in statistical analysis in python	2		4

16	Data types in R	2		9,10
17	Sequence generation, Vector and subscript, Random number generation in R	2		9,10
18	Data frames and R functions	2		9,10,11
19	Data manipulation and Data Reshaping using plyr, dplyr, reshape2	2		14
20	Parametric statistics and Non-parametric statistics	2		10,14
21	Continuous and Discrete Probability distribution using R	2		9,10,11
22	Correlation and covariance, contingency tables	2		9,10,11
23	Overview of Sampling, different sampling techniques	2		9,10,11
24	R and data base connectivity	2		9
25	Web application development with R using Shiny	2		12
26	Approaches to dealing with missing data in R	2		
27	Exploratory data analysis with simple visualizations using R	2		14
28	Feature or Attribute selection using R	2		15
29	Dimensionality Reduction with R	2		16
30	Time series data analysis with R	2		9,10,11

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. Goldwasser, "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 MakeYour 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. TorstenHothorn, 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. PawelCichosz, "Data Mining Algorithms: Explained Using R", wiley, 2014
16. BaterMakhabel, "Learning Data Mining with R", Packt Publication, 2015

CSE6016	Information Visualization	L,T,P,J,C 2,0,2,4,4
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 Outcome	After successfully completing the course the student should be able to 1. Identify the different data types, visualization types to bring out the insight. 2. Relate the visualization towards the problem based on the dataset to analyze and bring out valuable insight on large dataset. 3. Design visualization dashboard to support the decision making on large scale data. 4. Demonstrate the analysis of large dataset using various visualization techniques and tools.	
SLOs	• Having a clear understanding of the subject related concepts and of contemporary issues –2 • Having Sense-Making Skills of creating unique insights in what is being seen or observed (Higher level thinking skills which cannot be codified) --4 • Having computational thinking (Ability to translate vast data in to abstract concepts and to understand database reasoning) --7	

Module	Topics	L Hrs	SLO
1	**Introduction to Data Visualization** Overview of data visualization - Data Abstraction - Task Abstraction - Analysis: Four Levels for Validation, Human Visual Perception	4	2, 4
2	**Visualization Techniques - I** Scalar and point techniques – vector visualization techniques – matrix visualization	3	4

3	**Visualization Techniques - II** Visualization Techniques for Trees, Graphs, and Networks, Multidimensional data	6	7
4	**Visual Analysis of data from various domains - I** Time-oriented data visualization – Spatial data visualization and case studies	5	4
5	**Visual Analysis of data from various domains - II** Text data visualization – Multivariate data visualization, and case studies	5	7
6	**Designing Effective Visualizations** Guidelines for designing successful visualizations, Data visualization dos and don'ts	2	4
7	**Dashboard Creation and Visual Story Telling** Dashboard Design principles, Effective Dashboard Display Media, Dashboard creation using visualization tools for the use cases: Finance-marketing-insurance-healthcare etc.,	3	7
8	**Recent Trends**	2	2
Lab (Indicative List of Experiments (in the areas of) 1. Association Rule Mining and Clustering using R 2. Visualization on KNN or Naive Bayes Classification using R 3. Financial analysis using Clustering, Histogram and HeatMap 4. Time-series analysis – stock market 5. Visualization of various massive dataset - Finance - Healthcare - Census - Geospatial 6. Visualization on Streaming dataset (Stock market dataset, weather forecasting) 7. Market-Basket Data analysis-visualization 8. Text visualization using web analytics 9. Hadoop and R integration in Tableau using Hortonworks 10. Google API with maps 11. Visualization using D3.js 12. Visualization using Zeppelin 13. Network Visualization using Gephi 14. Visualization of reconstruction network using Qlickview 15. Dash Board Creation using Tableau		**30**	
Project # Generally a team project [2 to 4 members] Projects may be given as group projects The following is the sample project that can be given to students to be implemented using appropriate visualization tools. 1. Analysis of social media data using visualization (Sentiment Analysis, Opinion Mining, Recommender Systems)		**60** [Non Contact hrs]	14

2. Visualization of Fraudulent Behaviour in finance and insurance sectors 3. Creating dashboard using visualization to enable quick decision making on IOT (data will be received from different sensors & stored in data centers. It will help us to identify & take quick decision according to the real time and historic data trends and alarms displayed in the dashboard)		

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. From Web http://www.fusioncharts.com/whitepapers/

Information Visualization

2.Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	**Total Hours of Coverage**
CS: GV - Graphics and Visualization	30

2.1.Body of Knowledge coverage

[List the Knowledge Units covered in whole or in part in the course. If in part, please indicate which topics and/or learning outcomes are covered. For those not covered, you might want to indicate whether they are covered in another course or not covered in your curriculum at all. This section will likely be the most time-consuming to complete, but is the most valuable for educators planning to adopt the CS2013 guidelines.]

KA	**Knowledge Unit**	**Topics Covered**	**Hours**
CS:GV	Graphic Visualization	**Introduction to Data Visualization** **Visualization Techniques - I** **Visualization Techniques - II** **Visual Analysis of data from various domains - I** **Visual Analysis of data from various domains - II** **Designing Effective Visualizations** **Dashboard Creation and Visual Story Telling** **Recent Trends**	30
		Total hours	**30**

What is covered in the course?

[A short description, and/or a concise list of topics - possibly from your course syllabus.(This is likely to be your longest answer)]

Part 1: Introduction to Data Visualization
Overview of data visualization - Data Abstraction - Task Abstraction - Analysis: Four Levels for Validation, Human Visual Perception

Part II: Visualization Techniques - I
Scalar and point techniques – vector visualization techniques – matrix visualization

Part III: Visualization Techniques - II
Visualization Techniques for Trees, Graphs, and Networks, Multidimensional data

Part IV: Visual Analysis of data from various domains - I
Time-oriented data visualization – Spatial data visualization and case studies

Part V: Visual Analysis of data from various domains - II
Text data visualization – Multivariate data visualization and case studies

Part VI: Designing Effective Visualizations
Guidelines for designing successful visualizations, Data visualization dos and don'ts

Part VII: Dashboard Creation and Visual Story Telling
Dashboard Design principles, Effective Dashboard Display Media, Dashboard creation using visualization tools for the use cases: Finance-marketing-insurance-healthcare etc.

Part VIII: Recent Trends

What is the format of the course?
[*Is it face to face, online or blended? How many contact hours? Does it have lectures, lab sessions, and discussion classes?*]

This Course is designed with 100 minutes of in-classroom - 2 sessions per week, 100 minutes of lab hours per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course should have the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material.

How are students assessed?
[*What type, and number, of assignments are students are expected to do? (Papers, problem sets, programming projects, etc.). How long do you expect students to spend on completing assessed Work?]*

- Students are assessed on a combination group activities, classroom discussion, projects, and continuous, final assessment tests.

- Students can earn additional weightage based on certificate of completion of a related MOOC course.

Additional topics
[List notable topics covered in the course that you do not find in the CS2013 Body of Knowledge]

Other comments
[optional]

Session wise plan

Student Outcomes Covered: 2, 4, 7

S.No	Topic Covered	Class Hour	Levels of mastery	Reference Book
1	Overview of data visualization - Data Abstraction - Task Abstraction - Analysis: Four Levels for Validation, Human Visual Perception	4	Familiarity	1,8
2	Scalar and point techniques – vector visualization techniques – matrix visualization	3	Usage	5
3	Visualization Techniques for Trees, Graphs, and Networks, Multidimensional data	6	Usage	4
4	Time-oriented data visualization – Spatial data visualization and case studies	5	Usage	4
5	Text data visualization – Multivariate data visualization, and case studies	5	Usage	4
6	Guidelines for designing successful visualizations, Data visualization dos and don'ts	2	Usage	8
7	Dashboard Design principles, Effective Dashboard Display Media, Dashboard creation using visualization tools for the use cases: Finance-marketing-insurance-healthcare etc.,	3	Usage	3
8	**Recent Trends**	2	Usage	
Total hours		*30*		

CSE6017	**Mining Massive Data**	**L,T,P,J,C** 2, 0 , 2 , 4 , 4
Prerequisite	**Big Data Frameworks, Machine Learning**	
Objective	To provide comprehensive knowledge on developing and applying machine learning algorithms for massive real-world datasets in distributed frameworks	
Expected Outcomes	After successfully completing the course the student should be able to 1. Identify right machine learning / mining algorithm for handling massive data 2. Implement machine learning algorithms in distributed frameworks such as MapReduce and Spark 3. Use deep learning and extreme learning to solve real-life problems having multifarious complexities 4. Use big data analytics tools such as Spark, Mahout and H_2O in solving problems based on Machine learning	
SLO's	• Having a clear understanding of the subject related concepts and of contemporary issues –2 • Having computational thinking (Ability to translate vast data in to abstract concepts and to understand database reasoning) --7 • Having an ability to design and conduct experiments, as well as to analyze and interpret data --14	

Module	Topics	LHrs	SLO
1	**MapReduce Based Machine Learning** K-Means, PLANET, Parallel SVM, Association Rule Mining in MapReduce, Inverted Index, Page Ranking, Expectation Maximization, Bayesian Networks	7	7
2	**Classification & Regression models with Spark & Mahout** Linear support vector machines – Naive Bayes model – Decision Trees – Least square regression – Decision trees for regression	5	7

3	**Clustering in Spark and Mahout** 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	4	7
4	**Mining Social-Network Graphs** Clustering of Social-Network Graphs – Direct Discovery of Communities – Partitioning of Graphs – Finding Overlapping Communities – Counting Triangles using MapReduce – Neighborhood Properties of Graphs	3	7
5	**Semi-Supervised Learning** Introduction to Semi-Supervised Learning, Semi-Supervised Clustering, Transductive Support Vector Machines	3	7
6	**Deep Learning** Introduction, Deep Neural Networks, Deep Belief Networks, Auto Encoders, Recurrent Networks	4	7
7	**Extreme Learning** Extreme Learning Machines (ELM), ELM auto encoder, Extreme Support Vector Regression	2	7
8	**Recent Trends**	2	2

Lab(Indicative List of Experiments(in the areas of) 1. K-means implementation in MapReduce 2. Association Rule Mining with MapReduce 3. Decision trees in Spark 4. Naïve bayes classification using Spark 5. Advanced text processing with Spark 6. Clustering models with Spark 7. Building a recommendation engine with Spark 8. Representing social-network data using Graphs 9. Implementing Semi-supervised Clustering 10. Deep Learning using H_2O 11. Predictive analysis using H_2O tool 12. SVM Classification using Mahout 13. Spectral clustering using Mahout 14. Building a recommendation engine with Sparkling water 15. Deep Learning using DL4J	**30**	14

Project# Generally a team project [5to10 members] # Concepts studied in XXXX should have been used # 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 of3 reviews. Projects may be given as group projects The following is the sample project that can be given to students to be implemented: Students may be given projects to 1. Design a recommendation system 2. Design a text summarization system 3. Inference from social – network data 4. Bankruptcy prediction 5. Classify, Cluster or Predict from massive data sets from various domains: images, video, sound, text and DNA	**60**[Non Contact hrs]	17

Reference Books

1. Jure Leskovec, Anand Rajaraman, Jeff Ullman, “Mining of Massive Datasets”, Stanford Press, 2011.
2. Nick Pentreath, “Machine Learning with Spark”, Packt Publishing, 2015**.**
3. Olivier Chapelle, Bernhard Scholkopf, Alexander Zien "Semi-Supervised Learning", The MIT Press, 2006.
4. Ron Bekkerman, Mikhail Bilenko, John Langford "Scaling Up Machine Learning: Parallel and Distributed Approaches", Cambridge University Press, 2012.
5. Jimmy Lin, Chris Dyer, "Data-Intensive Text Processing with MapReduce", Morgan & Claypool Publishers, 2010.
6. Li Deng, Dong Yu, “Deep Learning: Methods and Applications”, Now Publisher, 2014.
7. Chandramani Tiwary "Learning Apache Mahout", Packt Publishing, 2015.
8. Fuchen Sun, Kar-Ann Toh, Manuel Grana Romay, KezhiMao,"Extreme Learning Machines 2013: Algorithms and Applications", Springer, 2014.

2. Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	Theory Hours
CS:PD (Parallel and Distributed Computing)/IS (Intelligent Systems)	30
Total	30

2.1 Body of Knowledge coverage

KA	Knowledge Unit	Topics Covered	Hours
CS: PD/IS	**Parallel Algorithms, Analysis, and Programming/Advanced Machine Learning**	**MapReduce Based Machine Learning,** **Classification & Regression models with Spark & Mahout,** **Clustering in Spark and Mahout,** **Mining Social-Network Graphs, Semi-** **Supervised Learning,** **Deep Learning** **Extreme Learning**	30
		Total hours	30

What is covered in the course?

Part I: MapReduce Based Machine Learning

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

Part II: Classification & Regression models with Spark & Mahout

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

Part III: Clustering in Spark and Mahout

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

Part IV: Mining Social-Network Graphs

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

Part V: Semi-Supervised Learning

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

Part VI: Deep Learning

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

Part VII: Extreme Learning

Extreme Learning Machines (ELM), ELM auto encoder, Extreme Support Vector Regression

Part VIII: Recent Trends

What is the format of the course?

This Course is designed with 100 minutes of in-classroom sessions per week, 100 minutes of lab hours per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course has the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, quizzes.

How are students assessed?

- Students are assessed on a combination group activities, classroom discussion, projects, & continuous and final assessment tests.
- Additional weightage will be given based on their rank in crowd sourced projects.
- Students can earn additional weightage based on certificate of completion of a related MOOC course or any online course completion.

Session wise plan

S.No	Topic Covered	Class Hour	Levels of mastery	Reference Book
1	K-Means, PLANET, Parallel SVM, Association Rule Mining in MapReduce, Inverted Index, Page Ranking, Expectation Maximization, Bayesian Networks	7	Usage	1,5

2	Linear support vector machines – Naive Bayes model – Decision Trees – Least square regression – Decision trees for regression	5	Usage	7,2
3	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	4	Usage	7,2
4	Clustering of Social-Network Graphs – Direct Discovery of Communities – Partitioning of Graphs – Finding Overlapping Communities – Counting Triangles using MapReduce – Neighborhood Properties of Graphs	3	Usage	1, 5
5	Introduction to Semi-Supervised Learning, Semi-Supervised Clustering, Transductive Support Vector Machines	3	Usage	3
6	Introduction, Deep Neural Networks, Deep Belief Networks, Auto Encoders, Recurrent Networks	4	Usage	6
7	Extreme Learning Machines (ELM), ELM auto encoder, Extreme Support Vector Regression	2	Usage	8
8	**Recent Trends**	2		
Total hours		***30***		

CSE6018	Streaming Data Analytics	L,T,P,J,C 2,0,2,4,4
Objective	It introduces theoretical foundations, algorithms, methodologies, and Applications of streaming data and also provide practical knowledge for handling and analyzing streaming data.	
Expected Outcomes	After successfully completing the course the student should be able to 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	
SLO's	• Having a clear understanding of the subject related concepts and of contemporary issues –2 • Having computational thinking (Ability to translate vast data in to abstract concepts and to understand database reasoning) --7 • Having an ability to use techniques, skills and modern engineering tools necessary for engineering practice --17	

Module	Topics	LHrs	SLO
1	**Introduction** Characteristics of the data streams, Challenges in mining data streams Requirements and principles for real time processing, Concept drift Incremental learning.	2	2
2	**Data Streams** 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	5	7
3	**Decision Trees** The 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.	4	7

4	**Clustering from Data Streams** Clustering Examples: Basic Concepts, Partitioning Clustering - The Leader Algorithm, Single Pass k-Means, Micro Clustering, Clustering Variables: A Hierarchical Approach	5	7

5	**Frequent Pattern Mining** 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	4	7
6	**Evaluating Streaming Algorithms** 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.	4	7
7	**Complex Event Processing** 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	4	7
8	**Recent Trends**	2	2
Lab(Indicative List of Experiments (in the areas of) 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		**30**	17

Project# Generally a team project [5to10 members] # Concepts studied in this course should have been used # 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 of3 reviews. Projects may be given as group projects The following is the sample project that can be given to students to be implemented: 1. Solving Data Science problems from Kaggle website 2. Applying stream algorithms in the field of biometrics for reliable and robust identification of humans from their personal traits, mainly for security and authentication purposes 3. Applying stream algorithms Video Analytics 4. Applying CEP for Fraud Detection, Healthcare etc.	**60**[Non Contact hrs]	17

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

Stream data analytics

Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	**Total Hours of Coverage**
CS: IS(Intelligent System)	30

Body of Knowledge coverage
[List the Knowledge Units covered in whole or in part in the course. If in part, please indicate which topics and/or learning outcomes are covered. For those not covered, you might want to indicate whether they are covered in another course or not covered in your curriculum at all. This section will likely be the most time-consuming to complete, but is the most valuable for educators planning to adopt the CS2013 guidelines.]

KA	**Knowledge Unit**	**Topics Covered**	**Hours**
CS: IS	IS/Advanced Machine Learning	Introduction Data Streams Decision Trees Clustering from Data Streams Frequent Pattern Mining Evaluating Streaming Algorithms Complex Event Processing Recent Trends	30
		Total hours	**30**

What is covered in the course?
[A short description, and/or a concise list of topics- possibly from your course syllabus.(This is likely to be your longest answer)]

What is the format of the course? [*Is it face to face, online or blended? How many contact hours? Does it have lectures, lab sessions, and discussion classes?*]

This Course is designed with 100 minutes of in-classroom sessions per week, 100minutes of lab hours per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course should have the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, quizzes.

How are students assessed?
[*What type, and number of assignments are students are expected to do? (papers, problem sets, programming projects, etc.). How long do you expect students to spend on completing assessed work?]*

- Students are assessed on a combination group activities, classroom discussion, projects, and continuous, final assessment tests.

Additional topics *[List not able topics covered in the course that you do not find in the CS2013Bodyof Knowledge]*

Other comments
[optional]

Session wise plan

Student Outcomes Covered: 2, 7, 17

Class Hour	Lab Hour	Topic Covered	levels of mastery	Ref eren
1		Characteristics of the data streams, Challenges in mining data streams, Requirements and principles for real time processing	Familiarity	1,3
1		Concept drift Incremental learning.	Familiarity	1, 3
5		Basic Streaming Methods, Counting the Number of Occurrence of the Elementsin 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	Familiarity	1

2		The Very Fast Decision Tree Algorithm (VFDT), The Base Algorithm, Analysis of the VFDT Algorithm	Familiarity, usage	1
2		Extensions to the Basic Algorithm: Processing Continuous Attributes, Functional Tree Leaves, Concept Drift.	Familiarity, Usage	1
2		Clustering Examples: Basic Concepts, Partitioning Clustering - The Leader Algorithm	Usage	1,3
2		Single Pass k-Means, Micro Clustering	Familiarity	1,3
1		Clustering Variables: A Hierarchical Approach	Familiarity	1,3
2		Mining Frequent Itemsets from Data Streams	Usage	1
2		Sequence Pattern Mining- Reservoir Sampling for Sequential Pattern Mining over data streams	Usage	1
4		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	Usage	1
4		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	Usage	2
2		Recent trends	Familiarity	

CSE6019	Text, Web and Social Media Analytics	L,T,P,J,C 3,0,0,4,4
Objectives	To Gain experience with both theoretical and practical aspects of text, web, social Media mining.	
Expected Outcome	After successfully completing the course the student should be able to 1. Build and evaluate the computer programs that generate new knowledge from natural language text. 2. Discover useful knowledge from the web. 3. Perform social network analysis to identify important social actors, subgroups (i.e., clusters), and network properties in social media sites such as Twitter, Facebook, and YouTube. 4. Provide solutions to the emerging problems with social media such as sentiment analysis and recommendation systems.	
SLOs	• Having computational thinking (Ability to translate vast data in to abstract concepts and to understand database reasoning) --7 • Having problem solving ability- solving social issues and engineering problems --9 • Having an ability to use techniques, skills and modern engineering tools necessary for engineering practice --17	

Module	Topics	L Hrs	SLO
1	**Introduction to Text Mining** Text Representation- **tokenization**, stemming, stop words, TF-IDF, Feature Vector Representation, NER, N-gram modeling.	6	7
2	**Mining Textual Data** Text Clustering, Text Classification , Topic Modeling-LDA, HDP	6	7
3	**Introduction to Web-mining** Inverted indices and Boolean queries. PLSI, Query optimization, page ranking.	6	7
4	**Web usage & Web content Mining** Web Crawling-Crawler Algorithms, Implementation Issues, Evaluation, Session & visitor Analysis, Visitor Segmentation, Analysis of Sequential & Navigational Patterns, Predictions based on web user transactions.	7	9
5	**Introduction Social Media Networks** Essentials of Social graphs, Social Networks, Models, Information Diffusion in Social Media.	6	9
6	**Mining Social Media** Behavioral Analytics, Influence and Homophily, Recommendation in Social Media	6	7,9
7	**Sentiment Mining** Sentiment Classification, feature based opinion mining, comparative sentence and relational mining, Opinion spam.	6	7,9
8	**RECENT TRENDS** Recent Trends in Text, Web and Social Media Analytics	2	9

Projects may be given as group projects 1. Sentiment Analysis. 2. Recommendation Systems 3. Fake Review Identification 4. Author Profiling		17

5. Personality Trait Recognition from Micro-blogs 6. Extracting a Knowledge Graph from Wikipedia 7. Finding the Social Roots of Controversy in Wikipedia 8. Techniques to improve detection of trending topics on Twitter 9. Mining Hospital Records for Predicting Patient Drop-off		

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, "Hand book of Natural Language Process", 2nd Edition, CRC Press, 2010.
5. Matthew A. Russell,"Mining the social web", 2nd edition - O'Reilly Media, 2013.

Text, Web and Social Media Analytics

Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	**Total Hours of Coverage**
CS: IS(Intelligent Systems)&HCI(Human Computer Interaction), IM (Information Management)	45

Body of Knowledge coverage

[List the Knowledge Units covered in whole or in part in the course. If in part, please indicate which topics and/or learning outcomes are covered. For those not covered, you might want to indicate whether they are covered in another course or not covered in your curriculum at all. This section will likely be the most time-consuming to complete, but is the most valuable for educators planning to adopt the CS2013 guidelines.]

KA	**Knowledge Unit**	**Topics Covered**	**Hours**
CS: IS	HCI	Introduction to Text Mining	6
CS: IS	HCI	Mining Textual Data	6

CS: IS	Web Mining	Introduction to Web-mining	6
CS: IS	Web Mining	Web usage & Web Content Mining	7
CS:IS	HCI	Introduction to Social Media Networks	6
CS:IS	HCI	Mining Social Media	6
CS:IS	HCI	Sentiment Mining	6
CS:IS		Recent Trends	2
		Total hours	**45**

What is covered in the course?
[A short description, and/or a concise list of topics - possibly from your course syllabus. (This is likely to be your longest answer)]

Part 1: Introduction to Text Mining
Text Representation- tokenization, stemming, stop words, TF-IDF, Feature Vector Representation, NER, N-gram modeling.

Part II: Mining Textual Data
Text Clustering, Text Classification, Topic Modeling-LDA, HDP

Part III: Introduction to Web-mining
Inverted indices and Boolean queries. PLSI, Query optimization, page ranking.

Part IV: Web usage & Web Content Mining

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

Part V: Introduction to Social Media Networks
Essentials of Social graphs, Social Networks, Models, Information Diffusion in Social Media.

Part VI: Mining Social Media
Behavioural Analytics, Influence and Homophily, Recommendation in Social Media

Part VII: Sentiment Mining

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

Part VIII: Recent Trends
Recent Trends in Text, Web and Social Media Analytics

What is the format of the course?
[*Is it face to face, online or blended? How many contact hours? Does it have lectures, lab sessions, discussion classes?*]

This Course is designed with 150 minutes of in-classroom sessions per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course should have the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, quizzes.

How are students assessed?
[*What type, and number, of assignments are students are expected to do? (papers, problem sets, programming projects, etc.). How long do you expect students to spend on completing assessed work?]*

- Students are assessed on a combination group activities, classroom discussion, projects, and continuous, final assessment tests.

- Additional weightage will be given based on their rank in crowd sourced projects/ Kaggle like competitions.
- Students can earn additional weightage based on certificate of completion of a related MOOC course.

Additional topics
[List notable topics covered in the course that you do not find in the CS2013 Body of Knowledge]

Other comments *[optional]*

Session wise plan

Student Outcomes Covered: 7, 9, and 17.

Class Hour	Topic Covered	levels of mastery	Reference Book	Remarks
6	Text Representation- tokenization, stemming, stop words, TF-IDF, Feature Vector Representation, NER, N-gram modeling.	Familiarity	4	

6	Text Clustering, Text Classification, Topic Modeling- LDA, HDP.	Usage	4	
6	Inverted indices and Boolean queries. PLSI, Query optimization, page ranking.	Usage	1	
7	Web Crawling-Crawler Algorithms, Implementation Issues, Evaluation, Session & visitor Analysis, Visitor Segmentation, Analysis of Sequential & Navigational Patterns, Predictions based on web user transactions.	Usage	1	
6	Essentials of Social graphs, Social Networks, Models, Information Diffusion in Social Media.	Usage	2	
6	Behavioural Analytics, Influence and Homophily, Recommendation in Social Media	Usage	2	
6	Sentiment Classification, feature based opinion mining, Comparative sentence and relational mining, Opinion spam.	Usage	3	
2	Recent Trends			
45 Hrs (3 Credit Hrs/week 15 Weeks schedule)				

CSE6020	Big Data Technologies	L,T,P,J,C 2,0,2,4,4
Objective	To focus on various tools and technologies to handle the humongous data sets for a variety of real-world applications.	
Expected Outcomes	At the end of the course, student will 1. Possess the skills necessary for utilizing tools to handle a variety of big data analytics. 2. Select the suitable technologies and apply the technologies on a variety of applications in solving problems aroused due to big data.	
SLO's	• Having a clear understanding of the subject related concepts and of contemporary issues –2 • Having an ability to design and conduct experiments, as well as to analyze and interpret data --14 • Having an ability to use techniques, skills and modern engineering tools ecessary for engineering practice --17	

Module	Topics	L Hrs	SLO
1	**Introduction** 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.	3	2
2	**Apache Pig** 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.	6	2
3	**Apache Hive Fundamentals** Introduction-Hive modules, Data types and file formats, Hive QL-Data Definition and Data Manipulation.	3	2, 14
4	**Apache Hive Advanced Concepts** Hive QL queries, Hive QL views- reduce query complexity. Hive scripts. Hive QL Indexes- create, show, drop. Aggregate functions. Bucketing vs Partitioning	4	2, 14

5	**Importing and Handling Relational Data in Hadoop using Sqoop** 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 – import new data, incrementally import data, preserving the value.	3	2
6	**Sqoop** 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.	4	2, 14
7	**Solr** 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.	5	2, 14
8	**Recent Trends**	2	2
Lab (Indicative List of Experiments) **1.** Implement a program using Pig latin operators and user defined functions **2.** Implement a program using operators and Pig latin scripts **3.** Program using Hive manipulation and data definition languages. **4.** Implement a program using Hive queries with partitioning. **5.** Implement a program using Hive indexes. **6.** Implement a program using Hive views **7.** Implement a program using Hive external table by accessing the external file created by Pig or any other tool. **8.** Program using Hive scripts and aggregate functions. **9.** Implement a program using Hive queries with bucketing and clustering. **10.** Implement a program for data transfer between Hadoop and external database using sqoop. **11.** Program to import data and incremental data in sqoop. **12.** Program to preserve the value in sqoop **13.** Program to export data from Hadoop using sqoop **14.** Program to import data to hive and using partitioned hive tables **15.** Program for inverted index using solr **16.** Program for indexing operations using csv files in solr. **17.** Program to search data using solr.		**30**	14

Project # Generally a team project [3 to 4 members] # Concepts studied in XXXX should have been used # 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. Projects may be given as group projects The following is the sample project that can be given to students to be implemented: 1. Solving Data Science problems from Kaggle website 2. Apply the big data technologies in various applications like Social Network data Page ranking Airlines data Networking data Fraud detection data Agriculture data , etc.	**60** [Non Contact hrs]	17

Reference Books

1. Alan Gates, "Programming Pig Dataflow 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 Cookbook", O'Reilly Media Inc, 2013.
4. Dikshant Shahi, "Apache Solr: A Practical approach to enterprise search", Apress, 2015.
5. Chuck Lam, "Hadoop in Action", Manning Publications, 2010.
6. Andrea Gazzarini, "Apache Solr Essentials", PACKT Publications, 2015.

Big Data Technologies

Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	**Total Hours of Coverage**
CS: IS(Intelligent System)/PD(Parallel Distributed Computing)	30

Body of Knowledge coverage

[List the Knowledge Units covered in whole or in part in the course. If in part, please indicate which topics and/or learning outcomes are covered. For those not covered, you might want to indicate whether they are covered in another course or not covered in your curriculum at all. This section will likely be the most time-consuming to complete, but is the most valuable for educators planning to adopt the CS2013 guidelines.]

KA	**Knowledge Unit**	**Topics Covered**	**Hours**
CS: IS/PD	Intelligent System/ Parallel Distributed Computing	Introduction to Big data Technologies Apache Pig, Hive, Sqoop, Solr, Recent Trends	30
		Total hours	**30**

What is covered in the course?

Part 1: Introduction

It introduces the concepts of big data. 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.

Part II: Apache Pig

It introduces 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.

Part III: Apache Hive Fundamentals

It introduces the Hive data warehouse. Hive modules, Data types and file formats, Hive QL-Data Definition and Data Manipulation

Part IV: Apache Hive Advanced Concepts

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

Part V: Importing and Handling Relational Data in Hadoop using Sqoop

It introduces 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 – import new data, incrementally import data, preserving the value.

Part VI: Sqoop

It discussed about data 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.

Part VII: Solr

It introduces the 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.

Part VIII: Recent Trends

What is the format of the course?

This Course is designed with 100 minutes of in-classroom sessions per week, 60 minutes of video/reading instructional material per week, 100 minutes of lab hours per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course should have the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, quizzes/assignments.

How are students assessed?

- Students are assessed on a combination group activities, classroom discussion, projects, and continuous, final assessment tests.
- Additional weightage will be given based on their rank in crowd sourced projects/ Kaggle like competitions.
- Students can earn additional weightage based on certificate of completion of a related MOOC course.

Session wise plan

Student Outcomes Covered: 2,7 ,9

Class	Lab Hour	Topic Covered	levels of mastery	Reference Book	Remarks
1		Introduction, challenges, source of big data	familiarity	1,online material	
1		Data access, intelligence, integration	familiarity	1,2,online material	
1		Data Serialization, Monitoring, Indexing.	familiarity	1,2, online material	
2		Parallel processing using Pig, Pig Architecture, Grunt	usage	1	
2		Pig Data Model-scalar and complex types. Pig Latin- Input and output, Relational operators	usage	1	
2		User defined function, Working with scripts.	usage	1	
1		Hive data ware house. Hive modules	familiarity	2	
2		Data types and file formats, Hive QL-Data Definition and Data Manipulation	Usage	2	

2		Hive QL queries, Hive QL views- reduce query complexity. Hive scripts.	Usage	2	
2		Hive QL Indexes- create, show, drop. Aggregate functions. Bucketing vs Partitioning.	usage	2	
1		Relational database management in hadoop: Bi directional data transfer between hadoop and external database	Usage	3	
2		Import data- Transfer an entire table, import subset data, use different file format. Incremental import – import new data, incrementally import data, preserving the value.	Usage	3	
2		Data export – transfer data from Hadoop, update the data, update at the same time, export subset of columns.	Usage	3	
2		Hadoop ecosystem integration- import data to hive, using partitioned hive tables, replace special delimiters.	Usage	3	
1		Information retrieval – search engine, categories of data	familiarity	4	
2		Inverted index. Design- field attributes and types. Indexing- indexing tool	Usage	4	
2		Indexing operations using csv documents. Searching data- parameters, default query.	Usage	4	
2		Recent Trends			

CSE6021	Domain Specific Predictive Analytics	L, T, P, J, C 3,0,0,4,4
Objective	It introduces theoretical foundations, algorithms, methodologies for analysing data in various domains such Retail, Finance, Risk and Healthcare.	
Expected Outcomes	After successfully completing the course the student should be able to 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	
SLO's	• Having a clear understanding of the subject related concepts and of contemporary issues –2 • Having computational thinking (Ability to translate vast data in to abstract concepts and to understand database reasoning) --7 • Having problem solving ability- solving social issues and engineering problems --9	

Module	Topics	L Hrs	SLO
1	**Retail Analytics** Understanding Customer: Profiling and Segmentation, Modeling Churn. Modeling Lifetime Value, Modeling Risk, Market Basket Analysis	7	2
2	**Risk Analytics** Risk Management and Operational Hedging: An Overview, Supply Chain Risk Management, A Bayesian Framework for Supply Chain Risk Management, Credit Scoring and Bankruptcy Prediction	5	7,9
3	**Financial Data Analytics** Financial News analytics: Framework, techniques, and metrics, News events impact market sentiment, Relating news analytics to stock returns	5	7,9

4	**Financial Time Series Analytics** 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	6	7,9
5	**Introduction Healthcare Analytics**	6	7,9
	An Introduction to Healthcare Data Analytics, Electronic Health Records, Privacy-Preserving Data Publishing Methods in Healthcare, Clinical Decision Support Systems		
6	**Healthcare Data Analytics** 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	7	7,9
7	**Genomic Data Analytics** Microarray Data, Microarray Data Analysis, Genomic Data Analysis for Personalized Medicine, Patient Survival Prediction from Gene Expression Data, Genome Sequence Analysis	7	7,9
8	**Recent Trends**	2	2
Project# Generally a team project [5 to 10 members] # Concepts studied in XXXX should have been used # 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. Projects may be given as group projects The following is the sample project that can be given to students to be implemented: 1. Event and Behaviour based targeting 2. Fraud detection 3. Abnormal trading pattern analysis 4. Social graph analysis and profile segmentation 5. Disease pattern and out break analysis 6. Drug discovery for personalized medicine 7. Customer churn analysis 8. Medical hospital readmission prediction		**60** [Non Contact hrs]	17

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, Lingxiu Dong, Onur Boyabatli, Rong Li "The Handbook of Integrated Risk Management in Global Supply Chains", Wiley, 2012.

Domain Specific Predictive Analytics

Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	**Total Hours of Coverage**
CS: IS(Intelligent System)	45

Body of Knowledge coverage

[List the Knowledge Units covered in whole or in part in the course. If in part, please indicate which topics and/or learning outcomes are covered. For those not covered, you might want to indicate whether they are covered in another course or not covered in your curriculum at all. This section will likely be the most time-consuming to complete, but is the most valuable for educators planning to adopt the CS2013 guidelines.]

KA	**Knowledge Unit**	**Topics Covered**	**Hours**
CS: IS	IS/Advanced Machine Learning	**Retail Analytics** **Risk Analytics** **Financial Data Analytics** **Financial Time Series Analytics** **Introduction Healthcare** **Analytics Healthcare Data** **Analytics Genomic Data** **Analytics**	45
		Total hours	**45**

What is covered in the course?
[A short description, and/or a concise list of topics - possibly from your course syllabus. (This is likely to be your longest answer)]

Part I: Retail Analytics
Understanding Customer: Profiling and Segmentation, Modeling Churn. Modeling Lifetime Value, Modeling Risk, Market Basket Analysis

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

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

Part IV: Financial Time Series Analytics
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

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

Part VI: Healthcare Data Analytics
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

Part VII: Genomic Data Analytics
Microarray Data, Microarray Data Analysis, Genomic Data Analysis for Personalized Medicine, Patient Survival Prediction from Gene Expression Data, Genome Sequence Analysis

Part VIII: Recent Trends

What is the format of the course?
[*Is it face to face, online or blended? How many contact hours? Does it have lectures, lab sessions, discussion classes?*]

This Course is designed with 150 minutes of in-classroom sessions per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course has the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, quizzes.

How are students assessed?
[*What type, and number, of assignments are students are expected to do? (papers, problem sets, programming projects, etc.). How long do you expect students to spend on completing assessed work?]*

- Students are assessed on a combination group activities, classroom discussion, projects, and continuous, final assessment tests.

- Additional weightage will be given based on their rank in crowd sourced projects/ Kaggle like competitions.

- Students can earn additional weightage based on certificate of completion of a related MOOC course.

Additional topics
[List notable topics covered in the course that you do not find in the CS2013 Body of Knowledge]

Other comments
[optional]

Session wise plan

Student Outcomes Covered: 2,7,9

S.No	Topic Covered	Class Hour	Levels of mastery	Reference Book
1	Understanding Customer: Profiling and Segmentation,	7	Usage	1
	Modeling Churn. Modeling Lifetime Value, Modeling Risk, Market Basket Analysis			
2	Risk Management and Operational Hedging: An Overview, Supply Chain Risk Management, A Bayesian Framework for Supply Chain Risk Management, Credit Scoring and Bankruptcy Prediction	5	Usage	3, 4
3	Financial News analytics: Framework, techniques, and metrics, News events impact market sentiment, Relating news analytics to stock returns	5	Usage	1,5,6
4	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	6	Usage	1, 5,6
5	An Introduction to Healthcare Data Analytics, Electronic Health Records, Privacy-Preserving Data Publishing Methods in Healthcare, Clinical Decision Support Systems	6	Usage	3
6	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	7	Usage	3,
7	Microarray Data, Microarray Data Analysis, Genomic Data Analysis for Personalized Medicine, Patient Survival Prediction from Gene Expression Data, Genome Sequence Analysis	7	Usage	3,
8	Recent Trends	2		
Total hours		*45*		

CSE6022	Soft Computing	L T P J C 3 0 0 4 4
Objective of the course	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 Outcome	After successfully completing the course the student should be able to • Have a general understanding of soft computing methodologies, to deal with imprecise and uncertain data • Develop computational neural network models for some simple biological systems; • Develop fuzzy models for engineering systems, particularly for control systems; • Apply derivative free optimization methods to solve real world problems • Demonstrate some applications of computational intelligence	
SLOs	• Having a clear understanding of the subject related concepts and of contemporary issues –2 • Having computational thinking (Ability to translate vast data in to abstract concepts and to understand database reasoning) --7 • Having problem solving ability- solving social issues and engineering problems --9	

Module	Topics	L Hrs	SLO
1	**Introduction to Soft Computing** Soft Computing Overview – Uncertainty in data, Hard vs. Soft Computing	2	2
2	**Neural Networks** Introduction, RBF Networks, Self-Organizing Map, Boltzmann Machines, Convolutional Neural Networks	7	7,9
3	**Fuzzy Systems** Fuzzy Sets, Fuzzy Relations, and Membership functions, Properties of Membership functions, Fuzzification and Defuzzification.	7	7,9

4	**Fuzzy logic** Fuzzy Rule based systems, Fuzzy Decision making, Fuzzy Classification, Fuzzy C-Means Clustering	7	7,9
5	**Rough Sets** Rough Sets – Definition, Upper and Lower Approximations, Boundary Region, Decision Tables and Decision Algorithms. Properties of Rough Sets. Rough K-means clustering, Rough	7	7,9
	Support Vector Clustering		
6	**Optimization Techniques** Introduction, Genetic Algorithm, Memetic Algorithms, Particle Swarm Optimization, Ant Colony Optimization, Frog-Leaping.	8	7,9
7	**Hybrid Systems:** GA Based Back Propagation Networks, Fuzzy Back Propagation Networks, Evolutionary Ensembles	5	2
8	**Recent Trends**	2	2
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		*60* Non-Contact Hours	17

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, 2010.
6. S. N. Sivanandam,S. Sumathi and S. N. Deepa, “Introduction to Fuzzy Logic using Matlab”–Springer, 2007.
7. Samir Roy, Udit Chakraborty, “Introduction to Soft Computing Neuro Fuzzy Genetic algorithms”, Pearson Education, 2013.
8. Witold Pedrycz, Andrzej Skowron, Vladik Kreinovich “Handbook of Granular Computing”, Wiley, 2008

2. Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	**Theory Hours**
CS: IS (Intelligent Systems)	30
CS: CN (Computational Science)	15
Total	45

2.1 Body of Knowledge coverage

KA	**Knowledge Unit**	**Topics Covered**	**Hours**
CS: IS	Intelligent Systems	Soft Computing Principles, NN, Fuzzy Logic & Fuzzy Systems, Rough Sets	30
CS: CN	Computational Science	Optimization Techniques & Hybrid Systems	15
		Total hours	45

What is covered in the course?

Part 1: Introduction to Soft Computing
Soft Computing Overview – Uncertainty in data, Hard vs., Soft Computing

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

Part III: Fuzzy Systems
Fuzzy Sets, Fuzzy Relations, and Membership functions, Properties of Membership functions, Fuzzification and Defuzzification.

Part IV: Fuzzy logic
Fuzzy Rule based systems, Fuzzy Decision making, Fuzzy Classification, Fuzzy C-Means Clustering

Part V: Rough Sets
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.

Part VI: Optimization Techniques
Introduction, Genetic Algorithm, Memetic Algorithms, Particle Swarm Optimization, Ant Colony Optimization, Frog-Leaping

Part VII: Hybrid Systems:
GA Based Back Propagation Networks, Fuzzy Back Propagation Networks, Evolutionary Ensembles

Part VIII: Recent Trends

What is the format of the course?

This Course is designed with 150 minutes of in-classroom sessions per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course has the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, quizzes.

How are students assessed?

- Students are assessed on a combination group activities, classroom discussion, projects, & continuous and final assessment tests.

- Additional weightage will be given based on their rank in crowd sourced projects.
- Students can earn additional weightage based on certificate of completion of a related MOOC course or any online course completion.

Session wise plan

S.No	Topic Covered	Class Hour	Levels of mastery	Reference Book
1	Soft Computing Overview – Hard vs, Soft Computing	2	Familiarity	1
2	Introduction, RBF Networks, Self-Organizing Map, Boltzmann Machines, Convolutional Neural Networks	7	Usage	3, 4
3	Fuzzy Sets, Fuzzy Relations, and Membership functions, Properties of Membership functions, Fuzzification and Defuzzification	7	Usage	1,5,6
4	Fuzzy Rule based systems, Fuzzy Decision making, Fuzzy Classification, Fuzzy C-Means Clustering	7	Usage	1, 5,6
5	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	7	Usage	7.8
6	Introduction, Genetic Algorithm, Memetic Algorithms, Particle Swarm Optimization, Ant Colony Optimization, Frog-Leaping.	8	Usage	1, 2
7	GA Based Back Propagation Networks, Fuzzy Back Propagation Networks, Evolutionary Ensembles	5	Assessment	1,2
8	Recent Trends	2		
Total hours		*45*		

CSE6023	Cloud Computing Fundamentals	L,T,P,J,C 2,0,2,4,4
Objectives	• To provide students with the fundamentals and essentials of Cloud Computing. • 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. • 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. • To impart knowledge in applications of cloud computing	
Expected Outcomes	After successfully completing the course the student should be able to • Design, Develop & Demonstrate real-world applications from the Cloud Computing • Understand the subtle architectural difference in Public and Private Clouds. • Appreciate the requirements of various service paradigms in Cloud Computing. • Describe the methods of processing multimedia elements and other information presentation concepts during multimedia communications.	

SLO's	• Having a clear understanding of the subject related concepts and of contemporary issues –2 • Having computational thinking (Ability to translate vast data in to abstract concepts and to understand database reasoning) --7 • Having an ability to use techniques, skills and modern engineering tools necessary for engineering practice --17		
Module	**Topics**	**L Hrs**	**SLO**
1	**Introduction to Cloud Computing** Cloud Computing Overview: Characteristics – challenges, benefits, limitations, Evolution of Cloud Computing, Cloud computing architecture, Cloud Reference Model (NIST Architecture)	**(4Hrs)**	2
2	**Infrastructure as a Service** Service Model, Characteristics, Benefits, Enabling Technologies Case Study : AWS, OpenStack,	**(4Hrs)**	7
3	**Platform as a Service** Service Model, Characteristics, Benefits, Enabling Technologies Case Studies : IBM Bluemix, GAE, Microsoft Azure	**(4Hrs)**	7
4	**Software as a Service** Service Model, Characteristics, Benefits, Enabling Technologies Case Study : Salesforce.com, CRM, Online Collaboration Services	**(4Hrs)**	7

5	**Data Analytics as a Service** Hadoop as a service, MapReduce on Cloud, Chubby locking Service	**(3Hrs)**	7

6	**Introduction to Public and Private Clouds** Shared Resources – Resource Pool – Usage and Administration Portal – Usage Monitor – Resource Management– Cloud Security – Workload Distribution – Dynamic provisioning.	**(5Hrs)**	7
7.	**Storage as a service** Historical Perspective, Datacenter Components, Design Considerations, Power Calculations, Evolution of Data Centers, Cloud data storage - CloudTM	**(3Hrs)**	7
8.	**Recent Trends**	**(2Hrs)**	2
		30	
Lab (Indicative List of Experiments (in the areas of) 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**	17
Projects: Understanding the Business aspects, and making them to develop simple cloud ecosystem.		**60**	17

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.

2. Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	**Theory Hours**
CS: PD (Cloud Computing)	30
Total	30

2.1 Body of Knowledge coverage

KA	**Knowledge Unit**	**Topics Covered**	**Hours**
CS:PD	Cloud Computing	**Introduction to Cloud Computing** **Infrastructure as a Service** **Platform as a Service** **Software as a Service** **Data Analytics as a Service** **Introduction to Public and Private Clouds** **Storage as a service**	30
		Total hours	30

What is covered in the course?

Part 1: Introduction to Cloud Computing
Cloud Computing Overview: Characteristics – challenges, benefits, limitations, Evolution of Cloud Computing, Cloud computing architecture, Cloud Reference Model (NIST Architecture)

Part II: Infrastructure as a Service
Service Model, Characteristics, Benefits, Enabling Technologies
Case Study : AWS, OpenStack

Part III: Platform as a Service
Service Model, Characteristics, Benefits, Enabling Technologies
Case Studies : IBM Bluemix, GAE, Microsoft Azure.
.

Part IV: Software as a Service
Service Model, Characteristics, Benefits, Enabling Technologies
Case Study : Salesforce.com, CRM, Online Collaboration Services.

Part V: Data Analytics as a Service
Hadoop as a service, MapReduce on Cloud, Chubby locking Service.

Part VI: Introduction to Public and Private Clouds
Shared Resources – Resource Pool – Usage and Administration Portal – Usage Monitor – Resource Management– Cloud Security – Workload Distribution – Dynamic provisioning.

Part VII: Storage as a service
Historical Perspective, Datacenter Components, Design Considerations, Power Calculations, Evolution of Data Centers, Cloud data storage – CloudTM.

Part VIII: Recent Trends

What is the format of the course?

This Course is designed with 100 minutes of in-classroom sessions per week, 100 minutes of lab hours per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally, this course should have the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, quizzes.

How are students assessed?

- Students are assessed on a combination group activities, classroom discussion, projects,
  & continuous and final assessment tests.
- Additional weightage will be given based on their rank in crowd sourced projects.
- Students can earn additional weightage based on certificate of completion of a related MOOC course or any online course completion.

Session wise plan

S.No	Topic Covered	Class Hour	Levels of mastery	Reference Book
1	Cloud Computing Overview: Characteristics – challenges, benefits, limitations, Evolution of Cloud Computing, Cloud computing architecture, Cloud Reference Model (NIST Architecture)	**4**	Familiarity	1
2	Service Model, Characteristics, Benefits, Enabling Technologies Case Study : AWS, OpenStack,	**4**	Familiarity	1, 2
3	Service Model, Characteristics, Benefits, Enabling Technologies Case Studies : IBM Bluemix, GAE, Microsoft Azure	**4**	Usage	1, 3
4	Service Model, Characteristics, Benefits, Enabling Technologies Case Study : Salesforce.com, CRM, Online Collaboration Services	**4**	Assessment	1, 4
5	Hadoop as a service, MapReduce on Cloud, Chubby locking Service	**3**	Usage	1, 2
6	Shared Resources – Resource Pool – Usage and Administration Portal – Usage Monitor – Resource Management– Cloud Security – Workload Distribution – Dynamic provisioning.	**5**	Usage	1, 2, 3
7	Historical Perspective, Datacenter Components, Design Considerations, Power Calculations, Evolution of Data Centers, Cloud data storage - CloudTM	**3**	Familiarity	1, 5, 6
8	**Recent Trends**	**2**	Usage	1, 5, 6
Total hours		*45*		

CSE6025	Analytics of Things	L,T,P,J,C 3,0,0,4,4
Objective of the course	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 Outcome	After successfully completing the course the student should be able to 1. Identify the technologies that enable 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	
SLO's	• Having a clear understanding of the subject related concepts and of contemporary issues –2 • Having design thinking capability -5 • Having an ability to use techniques, skills and modern engineering tools necessary for engineering practice --17	

Module	Topics	L Hrs	SLO
1	**Introduction to IoT** 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	6	2
2	**IOT Hardware platforms** Overview of IoT supported Hardware Platforms: Raspberry pi, Arduino, Intel Galileo	9	5
3	**Communication in IOT** 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.	5	5
4	**IOT Software development** Linux, Networking configurations in Linux, Accessing Hardware & Device Files interactions, Python packages: JSON, XML, HTTPLib, URLLib, SMTPLib, XMPP, Contiki OS	7	5
5	**IoT Physical Servers & Cloud Offerings** 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	6	2,5

6	**Data Analytics for IoT** Configuring and using Apache Storm for Real-time Data Analysis	5	2
7	**Case Studies illustrating IoT Design** Smart Home, Smart Parking, weather reporting and monitoring	5	2

8	**Recent Trends**	2	2
Project # Generally a team project [5 to 10 members] # Concepts studied in XXXX should have been used # Down to earth application and innovative idea should have been attempted # Report in Digital format with all drawings using software package to be submitted. [Ex. 1. Design of a traffic light system using sequential circuits OR 2. Design of digital clock] # Assessment on a continuous basis with a min of 3 reviews. **Projects may be given as group projects** 1. Sample project titles: 2. Smart grid 3. Vehicle charging using IOT 4. Energy measurement and storage at cloud 5. Water measurement and storage at cloud 6. Analysis and presentation IOT data stored at cloud 7. Smart Parking 8. Flash flood prediction 9. Real time monitoring of flood 10. Remote Monitoring & Sensing 11. Remote Controlling, Performance Analysis 12. IOT industries and what they are doing, selling 13. Healthcare application		**60** [Non Contact hrs]	17

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

Analytics of Things

2.Knowledge Areas that contain topics and learning outcomes covered in the course

Knowledge Area	**Total Hours of Coverage**
CS: OS/Real Time and Embedded Systems CE:CAO(Interfacing and communication, Device subsystems, Processor systems design)	45

2.1. Body of Knowledge coverage

[List the Knowledge Units covered in whole or in part in the course. If in part, please indicate which topics and/or learning outcomes are covered. For those not covered, you might want to indicate whether they are covered in another course or not covered in your curriculum at all. This section will likely be the most time-consuming to complete, but is the most valuable for educators planning to adopt the CS2013 guidelines.]

KA	**Knowledge Unit**	**Topics Covered**	**Hours**
CS: OS CE:CAO	Real Time and Embedded Systems Interfacing and communication, Device subsystems, Processor systems design	Introduction to IoT IOT Hardware platforms Communication in IOT IOT Software Development IoT Physical Servers & Cloud Offerings Data Analytics for IOT Case Studies illustrating IoT Design Recent Trends	45
		Total Hours	**45**

What is covered in the course?

Part 1: Introduction to IoT
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

Part II: IOT Hardware platforms

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

Part III: Communication in IOT
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.

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

Part V: IoT Physical Servers & Cloud Offerings
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

Part VI: Data Analytics for IOT
Configuring and using Apache Storm for Real-time Data Analysis

Part VII: Case studies illustrating IoT Design
Smart Home, Smart Parking, weather reporting and monitoring

Part VIII: Recent Trends

What is the format of the course?

This Course is designed with 150 minutes of in-classroom sessions per week, as well as 200 minutes of non-contact time spent on implementing course related project. Generally this course has the combination of lectures, in-class discussion, case studies, guest-lectures, mandatory off-class reading material, quizzes.

How are students assessed?

- Students are assessed on a combination group activities, classroom discussion, projects, and continuous, final assessment tests.
- Additional weightage will be given based on their rank in crowd sourced projects/ Kaggle like competitions.
- Students can earn additional weightage based on certificate of completion of a related MOOC course.

Session wise plan

Student Outcomes Covered: 2,5, 17

S.No	Topic Covered	Class Hour	Levels of mastery	Reference Book
1	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	6	Familiarity	1
2	Overview of IoT supported Hardware Platforms: Raspberry pi, Arduino, Intel Galileo	9	Usage	1,5,8
3	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.	5	Usage	1,2
4	Linux, Networking configurations in Linux, Accessing Hardware & Device Files interactions, Python packages: JSON, XML, HTTPLib, URLLib, SMTPLib, XMPP, Contiki OS	7	Usage	1, 2
5	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	6	Usage	1,6
6	Configuring and using Apache Storm for Real-time Data Analysis	5	Usage	1, 7
7	Smart Home, Smart Parking, weather reporting and monitoring	5	Assessment	1,5
8	Recent Trends	2		
Total hours		*45*		