VIT®

Vellore Institute of Technology

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

School of Computer Science and Engineering

CURRICULUM AND SYLLABI

(2023-2024)

M.Tech (CSE) – (Big Data Analytics)

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

VISION STATEMENT OF VELLORE INSTITUTE OF TECHNOLOGY

Transforming life through excellence in education and research.

MISSION STATEMENT OF VELLORE INSTITUTE OF TECHNOLOGY

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

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

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

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

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

VISION STATEMENT OF THE SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

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

MISSION STATEMENT OF THE SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

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

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

School of Computer Science and Engineering

M.Tech (CSE) – (Big Data Analytics)

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

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

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

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

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

M. Tech Computer Science and Engineering
(Big Data Analytics)

PROGRAMME OUTCOMES (POs)

PO_01: Having an ability to apply mathematics and science in engineering applications.

PO_03: Having an ability to design a component or a product applying all the relevant standards and with realistic constraints, including public health, safety, culture, society and environment

PO_04: Having an ability to design and conduct experiments, as well as to analyze and interpret data, and synthesis of information

PO_05: Having an ability to use techniques, skills, resources and modern engineering and IT tools necessary for engineering practice

PO_06: Having problem solving ability- to assess social issues (societal, health, safety, legal and cultural) and engineering problems

PO_07: Having adaptive thinking and adaptability in relation to environmental context and sustainable development

PO_08: Having a clear understanding of professional and ethical responsibility

PO_11: Having a good cognitive load management skills related to project management and finance

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

M. Tech Computer Science and Engineering
(Big Data Analytics)

ADDITIONAL PROGRAMME OUTCOMES (APOs)

APO_02: Having Sense-Making Skills of creating unique insights in what is being seen or observed (Higher level thinking skills which cannot be codified)

APO_03: Having design thinking capability

APO_04: Having computational thinking (Ability to translate vast data in to abstract concepts and to understand database reasoning

APO_07: Having critical thinking and innovative skills

APO_08: Having a good digital footprint

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

School of Computer Science and Engineering

M.Tech (CSE) – (Big Data Analytics)

PROGRAMME SPECIFIC OUTCOMES (PSOs)

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

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

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

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

M. Tech Computer Science and Engineering (Big Data Analytics)

CREDIT STRUCTURE

Category-wise Credit distribution

S.no	Catagory	Credits
1	Discipline Core	24
2	Specialization Elective	12
3	Projects and Internship	26
4	Open Elective	3
5	Skill Enhancement	5
Total Credits		70

Discipline Core									
sl.no	**Course Code**	**Course Title**	**Course Type**	**Version**	**L**	**T**	**P**	**J**	**Credits**
1	MCSE501L	Data Structures and Algorithms	Theory Only	1.0	3	0	0	0	3.0
2	MCSE501P	Data Structures and Algorithms Lab	Lab Only	1.0	0	0	2	0	1.0
3	MCSE502L	Design and Analysis of Algorithms	Theory Only	1.0	3	0	0	0	3.0
4	MCSE502P	Design and Analysis of Algorithms Lab	Lab Only	1.0	0	0	2	0	1.0
5	MCSE503L	Computer Architecture and Organisation	Theory Only	1.0	3	0	0	0	3.0
6	MCSE503P	Computer Architecture and Organisation Lab	Lab Only	1.0	0	0	2	0	1.0
7	MCSE504L	Operating Systems	Theory Only	1.0	3	0	0	0	3.0
8	MCSE504P	Operating Systems Lab	Lab Only	1.0	0	0	2	0	1.0
9	MCSE505L	Computer Networks	Theory Only	1.0	3	0	0	0	3.0
10	MCSE505P	Computer Networks Lab	Lab Only	1.0	0	0	2	0	1.0
11	MCSE506L	Database Systems	Theory Only	1.0	3	0	0	0	3.0
12	MCSE506P	Database Systems Lab	Lab Only	1.0	0	0	2	0	1.0

Specialization Elective									
sl.no	**Course Code**	**Course Title**	**Course Type**	**Version**	**L**	**T**	**P**	**J**	**Credits**
1	MCSE614L	Big Data Frameworks and Technologies	Theory Only	1.0	2	0	0	0	2.0
2	MCSE614P	Big Data Frameworks and Technologies Lab	Lab Only	1.0	0	0	2	0	1.0
3	MCSE615L	Data Analytics	Theory Only	1.0	2	0	0	0	2.0
4	MCSE615P	Data Analytics Lab	Lab Only	1.0	0	0	2	0	1.0
5	MCSE616L	Data Visualization	Theory Only	1.0	2	0	0	0	2.0
6	MCSE616P	Data Visualization Lab	Lab Only	1.0	0	0	2	0	1.0
7	MCSE617L	Domain Specific Predictive Analytics	Theory Only	1.0	2	0	0	0	2.0
8	MCSE617P	Domain Specific Predictive Analytics Lab	Lab Only	1.0	0	0	2	0	1.0
9	MCSE618L	Social Network Analytics	Theory Only	1.0	2	0	0	0	2.0
10	MCSE618P	Social Network Analytics Lab	Lab Only	1.0	0	0	2	0	1.0
11	MCSE619L	Text and Speech Analytics	Theory Only	1.0	2	0	0	0	2.0
12	MCSE619P	Text and Speech Analytics Lab	Lab Only	1.0	0	0	2	0	1.0
13	MCSE620L	Analytics for Internet of Things	Theory Only	1.0	2	0	0	0	2.0

Specialization Elective									
14	MCSE620P	Analytics for Internet of Things Lab	Lab Only	1.0	0	0	2	0	1.0

Projects and Internship									
sl.no	**Course Code**	**Course Title**	**Course Type**	**Version**	**L**	**T**	**P**	**J**	**Credits**
1	MCSE696J	Study Oriented Project	Project	1.0	0	0	0	0	2.0
2	MCSE697J	Design Project	Project	1.0	0	0	0	0	2.0
3	MCSE698J	Internship I/ Dissertation I	Project	1.0	0	0	0	0	10.0
4	MCSE699J	Internship II/ Dissertation II	Project	1.0	0	0	0	0	12.0

Open Elective									
sl.no	**Course Code**	**Course Title**	**Course Type**	**Version**	**L**	**T**	**P**	**J**	**Credits**
1	MFRE501L	Francais Fonctionnel	Theory Only	1.0	3	0	0	0	3.0
2	MGER501L	Deutsch fuer Anfaenger	Theory Only	1.0	3	0	0	0	3.0

Skill Enhancement									
sl.no	**Course Code**	**Course Title**	**Course Type**	**Version**	**L**	**T**	**P**	**J**	**Credits**
1	MENG501P	Technical Report Writing	Lab Only	1.0	0	0	4	0	2.0
2	MSTS501P	Qualitative Skills Practice	Soft Skill	1.0	0	0	3	0	1.5
3	MSTS502P	Quantitative Skills Practice	Soft Skill	1.0	0	0	3	0	1.5

Course Code	Course title	L	T	P	C
MCSE501L	Data Structures and Algorithms	3	0	0	3
Pre-requisite	NIL			Syllabus version	
				v.1.0	

Course Objectives

1. To familiarize the concepts of data structures and algorithms focusing on space and time complexity.
2. To provide a deeper insight into the basic and advanced data structures.
3. To develop the knowledge for the application of advanced trees and graphs in real- world scenarios.

Course Outcomes

Upon completion of the course the student will be able to

1. Understand and analyze the space and time complexity of the algorithms.
2. Identification of suitable data structure for a given problem.
3. Implementation of graph algorithms in various real-life applications.
4. Implementation of heaps and trees for querying and searching.
5. Use of basic data structures in advanced data structure operations.
6. Use of searching and sorting in various real-life applications.

Module:1	**Growth of Functions**	**3 hours**

Overview and importance of algorithms and data structures- Algorithm specification, Recursion, Performance analysis, Asymptotic Notation - The Big-O, Omega and Theta notation, Programming Style, Refinement of Coding - Time-Space Trade Off, Testing, Data Abstraction.

Module:2	**Elementary Data Structures**	**6 hours**

Array, Stack, Queue, Linked-list and its types, Various Representations, Operations & Applications of Linear Data Structures

Module:3	**Sorting and Searching**	**7 hours**

Insertion sort, merge sort, sorting in linear Time-Lower bounds for sorting, Radix sort, Bitonic sort, Cocktail sort, Medians and Order Statistics-Minimum and maximum, Selection in expected linear time, Selection in worst-case linear time, linear search, Interpolation search, Exponential search.

Module:4	**Trees**	**6 hours**

Binary trees- Properties of Binary trees, B-tree, B-Tree definition- Operations on B-Tree: Searching a B-tree, Creating, Splitting, Inserting and Deleting, B+-tree.

Module:5	**Advanced Trees**	**8 hours**

Threaded binary trees, Leftist trees, Tournament trees, 2-3 tree, Splay tree, Red-black trees, Range trees.

Module:6	**Graphs**	**7 hours**

Representation of graphs, Topological sorting, Shortest path algorithms- Dijkstra's algorithm, Floyd-Warshall algorithm, Minimum spanning trees - Reverse delete algorithm, Boruvka's algorithm.

Module:7	**Heap and Hashing**	**6 hours**

Heaps as priority queues, Binary heaps, binomial and Fibonacci heaps, Heaps in Huffman coding, Extendible hashing.

Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours:	**45 hours**

Text Book(s)

1. Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. Introduction to algorithms. MIT press, 2022.

Reference Books

1. Skiena, Steven S. "The Algorithm Design Manual (Texts in Computer Science)." 3rd edition, 2020,

	Springer.		
2.	Brass, Peter. Advanced data structures. Vol. 193. Cambridge: Cambridge University Press, 2008.		
Mode of Evaluation: CAT / Written Assignment / Quiz / FAT			
Recommended by Board of Studies	26-07-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Course title	L	T	P	C
MCSE501P	Data Structures and Algorithms LAB	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v. 1.0			

Course Objectives

1. To familiarize the concepts of data structures and algorithm focusing on space and time complexity.
2. To provide a deeper insight on the basic and advanced data structures.
3. To develop the knowledge for application of the advanced trees and graphs in real world scenarios.

Course Outcome

Upon completion of the course the student will be able to

1. Understand and analyze the space and time complexity of the algorithms.
2. Identification of suitable data structure for a given problem.
3. Implementation of graph algorithms in various real-life applications.
4. Implementation of heaps and trees for querying and searching.
5. Use of basic data structures in advanced data structure operations.
6. Use of searching and sorting in various real-life applications.

Indicative Experiments

1.	Analyzing the complexity of iterative and recursive algorithms
2.	Implement Linear data structures (Stacks, Queues, Linked Lists)
3.	Linear time sorting techniques
4.	Interpolation search & Exponential search
5.	Binary tree & Tree traversals
6.	B-trees & B+ trees
7.	Advanced Trees: 2-3 tree, splay tree, red black tree etc.
8.	Advanced Trees: Threaded Binary trees, tournament trees
9.	Graph traversals (BFS, DFS, Topological sorting)
10.	Determining the Shortest path between pair of nodes in the given graph
11.	Minimum Spanning trees- reverse delete & Boruvka's algorithm
12.	Heaps & Hashing
	Total Laboratory Hours | 30 hours

Text Book(s)

1.	Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. Introduction to algorithms. MIT press, 2022.

Reference Books

1.	Skiena, Steven S. "The Algorithm Design Manual (Texts in Computer Science)." 3rd edition, 2020, Springer.
2.	Brass, Peter. Advanced data structures. Vol. 193. Cambridge: Cambridge University Press, 2008.

Mode of Evaluation: CAT / Mid-Term Lab/ FAT

Recommended by Board of Studies	26-07-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Course title	L	T	P	C
MCSE502L	Design and Analysis of Algorithms	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		v. 1.0			

Course Objectives

1. To provide a mathematical framework for the design and analysis of algorithms.
2. To disseminate knowledge on how to create strategies for dealing with real-world problems.
3. To develop efficient algorithms for use in a variety of engineering design settings.

Course Outcomes

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

1. Apply knowledge of computing and mathematics to algorithm design.
2. Apply various algorithm paradigms to solve scientific and real-life problems.
3. Demonstrate the string matching and network flow algorithms relating to real-life problems.
4. Understand and apply geometric algorithms.
5. Apply linear optimization techniques to various real-world linear optimization problems.
6. Explain the hardness of real-world problems with respect to algorithmic design.

Module:1	Greedy, Divide and Conquer Techniques Introduction	6 hours

Overview and Importance of Algorithms - Stages of algorithm development: Describing the problem, Identifying a suitable technique, Design of an algorithm, Illustration of Design Stages - Greedy techniques: Graph Coloring Problem, Job Sequencing Problem with Deadlines- Divide and Conquer: Karatsuba's fast multiplication method, the Strassen algorithm for matrix multiplication

Module:2	Dynamic Programming, Backtracking and Branch & Bound Techniques	9 hours

Dynamic programming: Matrix Chain Multiplication, Longest Common Subsequence. Backtracking: N-Queens problem, Subset Sum, Graph Coloring- Branch & Bound: A-Star, LIFO-BB and FIFO BB methods.

Module:3	Amortized analysis and String Matching Algorithms	6 hours

Stack operation and Incrementing Binary counter -The aggregate method, the accounting method, the potential method, and Dynamic tables. Naïve String matching Algorithms, KMP algorithm, Rabin-Karp Algorithm, String matching with Finite Automata.

Module:4	Network Flow Algorithms	6 hours

Flow Networks, Maximum Flows: Ford-Fulkerson, Edmond-Karp, Push relabel Algorithm, The relabel-to-front algorithm, Minimum Cost flows – Cycle Cancelling Algorithm.

Module:5	Computational Geometry	5 hours

Line Segments – properties, intersection; Convex Hull finding algorithms- Graham's Scan, Jarvis's March Algorithm.

Module:6	Linear Optimization and Randomized algorithms	5 hours

Linear Programming problem - Simplex Method-Big M Method, LP Duality- The hiring problem, Finding the global Minimum Cut.

Module:7	NP Completeness and Approximation Algorithms	6 hours

The Class P - The Class NP - Reducibility and NP-completeness - Circuit Satisfiability problem-SAT 3CNF, Independent Set, Clique, Approximation Algorithm: Vertex Cover, Set Cover and Travelling salesman.

Module:8	Contemporary Issues	2 hours

	Total Lecture hours:	**45 hours**

Text Book(s)	
1.	Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. Introduction to algorithms. MIT press, 2022.
Reference Books	
1.	Rajeev Motwani, Prabhakar Raghavan; "Randomized Algorithms, Cambridge University Press, 1995 (Online Print — 2013).
2.	Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin, Network Flows: Theory, Algorithms, and Applications, 1st Edition, Pearson Education, 2014.
3.	Jon Kleinberg and EvaTardos, Algorithm Design, Pearson Education, 1"Edition, 2014.

Mode of Evaluation: CAT / Written Assignment / Quiz / FAT			
Recommended by Board of Studies	26-07-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Course title	L	T	P	C
MCSE502P	Design and Analysis of Algorithms Lab	0	0	2	1
Pre-requisite	NIL			Syllabus version	
				v.1.0	

Course Objectives

1. To provide a mathematical framework for the design and analysis of algorithms.
2. To disseminate knowledge on how to create strategies for dealing with real-world problems.
3. To develop efficient algorithms for use in a variety of engineering design settings.

Course Outcome

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

1. Apply knowledge of computing and mathematics to algorithm design.
2. Apply various algorithm paradigms to solve scientific and real-life problems.
3. Demonstrate the string matching and network flow algorithms relating to real-life problems.
4. Understand and apply geometric algorithms.
5. Apply linear optimization techniques to various real-world linear optimization problems.
6. Explain the hardness of real-world problems with respect to algorithmic design.

Indicative Experiments

1.	Greedy Strategy : Graph Coloring Problem, Job Sequencing Problem with Deadlines
2.	Divide and Conquer : Karatsuba's fast multiplication method, the Strassen algorithm for matrix multiplication
3.	Dynamic Programming: Matrix Chain Multiplication, Longest Common Subsequence, 0-1 Knapsack
4.	Backtracking: N-queens, Subset sum
5.	Branch and Bound: Job selection
6.	String Matching Algorithms: Rabin Karp Algorithm, KMP Algorithm
7.	Network Flows : Ford -Fulkerson and Edmond – Karp, Cycle cancelling algorithm
8.	Minimum Cost flows – Cycle Cancelling Algorithm
9.	Linear programming: Simplex method
10.	Randomized Algorithms: Las Vegas and Monte carlo
11.	Polynomial time algorithm for verification of NPC problems
12.	Approximation Algorithm: Vertex cover ,Set cover and TSP
	Total Laboratory Hours | 30 hours

Text Book(s)

1.	Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. Introduction to algorithms. MIT press, 2022.

Reference Books

1.	Rajeev Motwani, Prabhakar Raghavan; Randomized Algorithms, Cambridge University Press, 1995 (Online Print — 2013).
2	Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin, Network Flows: Theory, Algorithms, and Applications, 1st Edition, Pearson Education, 2014.
3	Jon Kleinberg and EvaTardos, Algorithm Design, Pearson Education, 1"Edition, 2014.

Mode of Evaluation: CAT / Mid-Term Lab/ FAT

Recommended by Board of Studies	26-07-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Course title	L	T	P	C
MCSE503L	Computer Architecture and Organization	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		v. 1.0			

Course Objectives

1. To provide knowledge on the basics of computer architectures and organization that lays the foundation to study high-performance architectures
2. To design and develop parallel programs using parallel computing platforms such as OpenMP, CUDA
3. To evaluate the performance using profiling tools and optimize parallel codes using various optimization techniques

Course Outcomes

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

1. Outline the developments in the evolution of computer architectures and parallel programming paradigms
2. Comprehend the various programming languages and libraries for parallel computing platforms
3. Use of profiling tools to analyze the performance of applications by interpreting the given data
4. Evaluate efficiency trade-offs among alternative parallel computing architectures for an efficient parallel application design
5. Develop parallel programs using OpenMP and CUDA and analyze performance parameters such as speed-up, and efficiency for parallel programs against serial programs

Module:1	Computer Evolution And Performance	5 hours

Defining Computer Architecture and Organization, Overview of Computer Components, Von Neumann architecture, Harvard Architecture CISC & RISC, Flynn's Classification of Computers, Moore's Law, Multi-threading, Comparisons of Single Core, Multi Processors, and Multi-Core architectures, Metrics for Performance Measurement

Module:2	Memory Hierarchy	8 hours

Key Characteristics of Memory systems, Memory Hierarchy, Cache Design policies, Cache Performance, Cache Coherence, Snoopy Protocols, Cache coherence protocols, MSI, MESI, MOESI

Module:3	Parallel Computers	8 hours

Instruction Level Parallelism(ILP), Compiler Techniques for ILP & Branch Prediction, Thread Level Parallelism (TLP), Threading Concepts, Shared Memory, Message Passing, Vectorization

Module:4	Multithreaded Programming using OpenMP	7 hours

Introduction to OpenMP, Parallel constructs, Runtime Library routines, Work-sharing constructs, Scheduling clauses, Data environment clauses, atomic, master Nowait Clause, Barrier Construct

Module:5	Programming for GPU	6 hours

Introduction to GPU Computing, CUDA Concepts, CUDA Programming Model, Program Structure of CUDA & Execution, Methods for operations on Device Memory, Thread Organization, Examples

Module:6	Performance Analyzers	6 hours

Performance Evaluation, performance bottlenecks, Profiling categories; Profiling tools: Trace analyzer and collector (ITAC), VTune Amplifier XE, Energy Efficient Performance, Integrated Performance Primitives (IPP)

Module:7	Energy Efficient Architectures	5 hours

Overview of power issues, CMOS Device-level Power dissipation basics, Sources of energy Consumption, Strategies to save power or Energy, Low power designs, Power management techniques

Module:8	**Contemporary Issues**	1 hours
	Total Lecture hours:	45 hours

Text Book(s)	
1.	William Stallings, Computer Organization and Architecture: Designing for Performance, Pearson, 2022, 11th Edition, Pearson
2	Gerassimos Barlas, Multicore and GPU Programming: An Integrated Approach, 2022, 2nd edition, Morgan Kaufmann
Reference Books	
1.	J.L. Hennessy and D.A. Patterson. Computer Architecture: A Quantitative Approach. 5th Edition, 2012, Morgan Kauffmann Publishers.
2.	Shameem Akhter, Jason Roberts, Multi-core Programming: Increasing Performance Through Software Multi-threading, 2010, Intel Press, BPB Publications

Mode of Evaluation: CAT / Written Assignment / Quiz / FAT			
Recommended by Board of Studies	26-07-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Course title	L	T	P	C
MCSE503P	Computer Architecture and Organization LAB	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. To provide knowledge on basics of computer architectures and organization that lays foundation to study high performance architectures
2. To design and develop parallel programs using parallel computing platforms such as OpenMP, CUDA
3. To evaluate the performance using profiling tools and optimize parallel codes using various optimization techniques

Course Outcome

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

1. Outline the developments in the evolution of computer architectures and parallel programming paradigms
2. Comprehend the various programming languages and libraries for parallel computing platforms
3. Use of profiling tools to analyze the performance of applications by interpreting the given data
4. Evaluate efficiency trade-offs among alternative parallel computing architectures for an efficient parallel Application design.
5. Develop parallel programs using OpenMP and CUDA and analyze performance parameters such as speed-up, efficiency for parallel programs against serial programs

Indicative Experiments

1.	Set-up an environment for OpenMP Programming: Activities: create a Project using Visual Studio, Writing Sample OpenMp Program, Setting up properties, compile & Execute OpenMP program, OpenMP manual study, Creation of Login credential on Intel for Intel Parallel Studio
2.	OpenMP program using following construct and describe scenario for the need of construct Use of Parallel Construct, Determine the Number of processors in a parallel Region, Find the thread ID of each processor
3.	Computation of Execution Time Using OpenMP clock, Using windows clock
4.	OpenMP Program using various Environment Routines to access the processor run-time information and write interesting observations by comparing various routines
5.	OpenMP program using following Worksharing Constructs and describe scenario for the need of construct loop construct, sections construct, single construct
6.	OpenMP program using following schedule clauses and describe scenario for the need of clause Static, Dynamic, Guided
7.	Develop parallel programs for given serial programs and profile the program using Vtune Analysis tool Matrix-Matrix multiplication, Matrix-Vector multiplication
8.	Develop parallel programs for given serial programs and profile the program using Vtune Analysis tool Quicksort, Minimum Spanning Tree
9.	CUDA-platform setup on NVIDIA / Google Colab
10.	Write a CUDA C/C++ program that add two array of elements and store the result in third array
11.	Write a CUDA C/C++ program that Reverses Single Block in an Array; CUDA C/C++
12.	Write a CUDA C program for Matrix addition and Multiplication using Shared memory
	Total Laboratory Hours | **30 hours**

Text Book(s)

1.	Gerassimos Barlas, Multicore and GPU Programming: An Integrated Approach, 2022, 2nd edition, Morgan Kaufmann

Reference Books			
1.	Shameem Akhter, Jason Roberts, Multi-core Programming: Increasing Performance Through Software Multi-threading, 2010, Intel Press, BPB Publications		
Mode of Evaluation: CAT / Mid-Term Lab/ FAT			
Recommended by Board of Studies	26-07-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Course title	L	T	P	C
MCSE504L	OPERATING SYSTEMS	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		v. 1.0			

Course Objectives

1. To focus the core functionalities required to develop and manage operating systems.
2. To encompass process management, synchronization strategies, memory management, file systems, device management, and virtualization.
3. To introduce the concepts and features of real-time operating systems as well as virtualization.

Course Outcomes

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

1. Understand the fundamental operating system abstractions, including processes, threads, semaphores, and file systems.
2. Implement scheduling, devising and addressing synchronization issues.
3. Gain an understanding of memory management tasks.
4. Develop real-time working prototypes of different small-scale and medium-scale embedded systems.
5. Comprehend the basics of virtualization and differentiate types of virtualization.

Module:1	Introduction to Operating Systems	4 hours

Computer Organization and Architecture - OS definition – OS history – OS Operations – OS design issues - Operating systems structures - Library files - Systems calls – Interrupts - Kernel approaches – Building and booting an OS.

Module:2	Process and Scheduling	6 hours

Process states – State transitions with suspend and resume - Process control block - Context-switching - Processes operations - Process scheduling - CPU scheduling: Non-preemptive, preemptive - Multi-queue scheduling - Multi-level feedback queue scheduling.

Module:3	Synchronization	9 hours

IPC: Shred memory, message passing - Race condition – Critical section problem - Peterson's solution – Bakery Algorithm - Mutex locks - Semaphores – Classical synchronization problems – Monitors - Thread synchronization – Multi-threading Models, Deadlocks – Resource allocation graphs – Deadlock: prevention, avoidance, detection and recovery.

Module:4	Memory Management	5 hours

Address binding – Fragmentation - Pinning Memory – Paging – Structure of the page table – Swapping - Segmentation - Demand Paging – Copy-on-write - Replacement – Thrashing – Working set – Memory compression – Allocating kernel memory.

Module:5	Managing Devices, Files, Security and Protection	9 hours

I/O Management – DMA - Delayed write - Disk scheduling algorithms: Seek-time and rotational latency based - File control block – Inode – Access method – Directory structure - Directory implementation – File allocation methods - Free space management – Program and network threats – Cryptography as a security tool – Domains of protection – Access matrix – Capability based systems

Module:6	Real-time Operating Systems	5 hours

RTOS Internals - Real-Time Scheduling - Task Specifications - Performance Metrics of RTOS - Schedulability Analysis – RTOS Programming Tools.

Module:7	Virtualization	5 hours

Need for virtualization - Virtual machines and architectures – Hypervisors - Virtualization Technologies: Para Virtualization, Full Virtualization - Virtualization types: Server virtualization, Application virtualization, Storage virtualization.

Module:8	Contemporary Issues	2 hours

	Total Lecture hours:	**45 hours**

Text Book(s)	
1.	Abraham Silberschatz, Peter B. Galvin, Greg Gagne, "Operating System Concepts", 2018, 10th Edition, Wiley, United States.
Reference Books	
1.	Arpaci-Dusseau, R. H., & Arpaci-Dusseau, A. C, "Operating Systems: Three easy pieces, 2018, 1st Edition, Boston: Arpaci-Dusseau Books LLC.
2.	Kamal, R, Embedded Systems: Architecture, Programming and Design, 2011, 1st Edition, Tata McGraw-Hill Education.
3.	Portnoy, M, "Virtualization Essentials", 2012, 2nd Edition, John Wiley & Sons, New Jersey, USA.

Mode of Evaluation: CAT / Written Assignment / Quiz / FAT			
Recommended by Board of Studies	26-07-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Course title	L	T	P	C
MCSE504P	OPERATING SYSTEMS LAB	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. To encompass process management, synchronization strategies, memory management, file systems, device management, and virtualization.
2. To introduce the concepts and features of real-time operating systems as well as virtualization.

Course Outcome

On completion of this course, student should be able to

1. Implement scheduling, devising and addressing synchronization issues.
2. Gain an understanding of memory management tasks.
3. Develop real-time working prototypes of different small-scale and medium-scale embedded systems.
4. Comprehend the basics of virtualization and differentiate types of virtualization.

Indicative Experiments

1.	Investigate the fundamental Unix/Linux commands.
2.	Obtaining the OS system data file and its associated information.
3.	Shell Programming.
4.	Create utility programs that use I/O system calls to simulate operations such as ls, cp, grep, and others.
5.	Create child, Orphan and Zombie processes using suitable system calls such as fork(), exec(), wait(), kill(), sleep() and exit() system calls.
6.	Create a program that mimics the CPU Scheduling algorithms including multi-level queue scheduling algorithm. Ex: Assume that all 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.
7.	Implement the deadlock-free solution to Dining Philosophers problem using Semaphore.
8.	Simulation of Bankers algorithm to check whether the given system is in safe state or not. Also check whether addition resource requested can be granted immediately.
9.	Parallel Thread management using Pthreads library. Implement a data parallelism using multi-threading. Ex: An application should have a thread created with synchronization and thread termination. Every thread in the sub-program must return the value and must be synchronized with the main function. Final consolidation should be done by the main (main function).
10.	Dynamic memory allocation algorithms – First-fit, Best-fit, Worst-fit algorithms.
11.	Page Replacement Algorithms FIFO, LRU and Optimal
12.	Implement a file locking mechanism.
13.	RTOS Based Parameter Monitoring and Controlling System – Monitoring: Collecting data from sensors and interface display devices/actuators using a microcontroller. Controlling: Provide an alert when the received data reaches a certain threshold value.
14.	Virtualization Setup: Type-1, Type-2 Hypervisor (Detailed Study Report).
	Total Laboratory Hours | 30 hours

Text Book(s)

1.	Vijay Mukhi, "The C Odyssey: UNIX: v. 3", 2004, 3rd Edition, BPB Publications, New Delhi, India.

Reference Books

1.	Stevens, W. R., & Rago, S. A. (2013). Advanced Programming in the UNIX Environment: Advanc Progra UNIX Envir_p3. Addison-Wesley.
2.	Love, Robert, "Linux System Programming: talking directly to the kernel and C library", 2013, 2nd Edition, O'Reilly Media, Inc, United States.

Mode of Evaluation: CAT / Mid-Term Lab/ FAT

Recommended by Board of Studies	26-07-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Course title	L	T	P	C
MCSE505L	Computer Networks	3	0	0	3
Pre-requisite	NIL			Syllabus version	
				v.1.0	

Course Objectives

1. To learn various network models, layers and their protocols.
2. To gain a fundamental understanding of routing algorithms.
3. To comprehend the basics of wireless as well as mobile networks and their characteristics.

Course Outcomes

On completion of this course, student should be able to

1. Explore the basics of Computer Networks and various performance metrics.
2. Interpret the application layer services and their protocols.
3. Evaluate the requirements for reliable services and implications of congestion at the transport layer services.
4. Analyse various functionalities required in the control and data plane at network layer services.
5. Infer the characteristics of wireless as well as mobile networks and their security standards.

Module:1	Computer Networks and the Internet	7 hours

Internet: A Nuts-and-Bolts Description - Network Protocols - The Network Edge: Access Networks and Physical Media - The Network Core: Packet Switching, Circuit Switching - Network of Networks - Delay, Loss and Throughput in Packet-Switched Networks - Protocol Layers and Their Service Models

Module:2	Application Layer	5 hours

Principles of Network Applications: Architectures, Processes and Transport Services - The Web and HTTP - Electronic Mail in the Internet - DNS—The Internet's Directory Service - Peer-to-Peer File Distribution - Socket Programming: Creating Network Applications

Module:3	Transport Layer	7 hours

Relationship Between Transport and Network Layers - Overview of the Transport Layer in the Internet - Multiplexing and Demultiplexing - Connectionless Transport: UDP - Reliable Data Transfer: Go-Back-N (GBN) and Selective Repeat (SR) - Connection-Oriented Transport: TCP, Flow Control and Congestion Control

Module:4	Network Layer: Data Plane	5 hours

Network Layer – Router - The Internet Protocol (IP): IPv4, Addressing and IPv6 - Generalized Forwarding and SDN

Module:5	Network Layer: Control Plane	5 hours

Control Plane: Per-router control and logically centralized control - Routing Algorithms - Link-State (LS) Routing Algorithm, Distance-Vector (DV) Routing Algorithm, Intra-AS Routing in the Internet: OSPF and Routing Among the ISPs: BGP - SDN Control Plane

Module:6	Link Layer and LANs	8 hours

Overview of Link Layer Services - Error-Detection and -Correction Techniques: Parity Checks, Checksum and CRC - Multiple Access Links and Protocols: Channel Partitioning Protocols and Random-Access Protocols - Switched Local Area Networks: Link-Layer Addressing and ARP - Virtual Local Area Networks

Module:7	Wireless and Mobile Networks-Security	6 hours

Elements of a wireless network - Wireless Links and Network Characteristics - WiFi: 802.11 Wireless LANs - Mobility Management: Principles - Wireless and Mobility: Impact on Higher-Layer Protocol- Security in Computer Network- Message Integrity and Digital Signatures - Network-Layer Security: IPsec and Virtual Private Networks

Module:8	Contemporary Issues	2 hours

	Total Lecture hours:	45 hours

Text Book(s)

1. James F. Kurose, Keith W. Ross, "Computer Networking: A Top-Down Approach", 2022, 8th

	Edition(Paperback), Pearson, United Kingdom.			
Reference Books				
1.	Larry Peterson and Bruce Davie, “Computer Networks: A Systems Approach”, 2019, 6th Edition, Morgan Kaufmann, United States of America.			
2.	Andrew S. Tanenbaum, “Computer Networks”, 2013, 6th Edition, Pearson, Singapore.			
Mode of Evaluation: CAT / Written Assignment / Quiz / FAT				
Recommended by Board of Studies		26-07-2022		
Approved by Academic Council		No. xx	Date	DD-MM-YYYY

Course code	Course title	L	T	P	C
MCSE505P	Computer Networks Lab	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. To introduce the computer network concepts and provide skills required to trouble shoot the network devices.
2. To describe the basic knowledge of VLAN.
3. To develop the knowledge for application of software defined networks.

Course Outcome

On completion of this course, student should be able to

1. Understand the types of network cables and practical implementation of cross-wired and straight through cable.
2. Design and implementation of VLAN.
3. Analyze and apply network address translation using packet tracer and network simulators.
4. Design and develop software defined networks.

Indicative Experiments

1.	Hardware Demo(Demo session of all networking hardware and Functionalities) OS Commands(Network configuration commands)
2.	Error detection and correction mechanisms Flow control mechanisms
3.	IP addressing Classless addressing
4.	Network Packet Analysis using Wireshark i. Packet Capture Using Wire shark ii. Starting Wire shark iii. Viewing Captured Traffic iv. Analysis and Statistics & Filters.
5.	Socket programming(TCP and UDP) Multi client chatting
6.	Networking Simulation Tool –Wired and Wireless
7.	SDN Applications and Use Cases
8.	Security in Network- Use cases
9	Performance evaluation of routing protocols using simulation tools.

Reference Books

1.	James F. Kuross, Keith W. Ross, "Computer Networking, A Top-Down Approach", 8th Edition(Paperback), Pearson Education, 2022.

Mode of Evaluation: CAT / Mid-Term Lab/ FAT

Recommended by Board of Studies	26-07-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Course title	L	T	P	C
MCSE506L	DATABASE SYSTEMS	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. To understand the underlying principles of Relational Database Management Systems
2. To focus on the modeling and design of secured databases and usage of advanced data models
3. To implement and maintain the structured, semi-structured, and unstructured data in an efficient database system using emerging trends

Course Outcomes

On completion of this course, students must be able to

1. Design and implement a database depending on the business requirements, considering various design issues
2. Understand the concepts of Indexing, Query optimization, transaction management, concurrency control, and recovery mechanisms
3. Learn to apply parallel and distributed databases in Real-time scenarios
4. Categorize and design the structured, semi-structured, and unstructured databases
5. Characterize the database threats and their countermeasures

Module:1	Design and Implementation of Relational Model	6 hours

Database System Concepts and Architecture, Entity-Relationship (ER) Modelling, Relational Model-Keys, and Integrity Constraints, Mapping ER model to Relational Schema, Normalization, Boyce Codd Normal Form, Multi-valued dependency and Fourth Normal form

Module:2	Query Processing and Transaction Management	6 hours

Storage and File Structure, Indexing, Query processing, and Query Optimization, Transaction Management, Concurrency Control, Recovery

Module:3	Parallel Databases and Distributed Databases	8 hours

Parallel Database Architecture, Data partitioning strategy, Inter-Query, and Intra-Query Parallelism, Distributed Database Features, Distributed Database Architecture, Fragmentation, Replication, Distributed Query Processing, Distributed Transactions Processing

Module:4	Spatial and Multimedia Databases	6 hours

Spatial database concepts, Spatial data types, and models, Spatial operators and queries, Indexing in spatial databases, Multimedia database concepts, Automatic Analysis of Images, Object Recognition in Images, Semantic Tagging of Images

Module:5	Semi-Structured Databases	6 hours

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

Module:6	Cloud and NoSQL Databases	6 hours

Cloud databases- Data Storage Systems on the Cloud, Data Representation, Partitioning and Retrieving Data, Challenges with Cloud-Based Databases- NoSQL Data model: Aggregate Models, Document Data Model, Key-Value Data Model, Columnar Data Model, Graph-Based Data Model

Module:7	Database Security	5 hours

Database Security Issues, Security Models, Different threats to databases, Challenges to maintaining database security

Module:8	Contemporary Issues	2 hours

	Total Lecture hours:	**45 hours**
Text Book(s)		
1	Abraham Silberschatz, Henry F. Korth, and S. Sudharsan, "Database System Concepts", 7h Edition, McGraw Hill, 2019.	
2	R. Elmasri and S. Navathe, Fundamentals of Database Systems, 7th Edition, Addison-Wesley, 2016	
Reference Books		
1	Fawcett, Joe, Danny Ayers, and Liam RE Quin. "Beginning XML", Wiley India Private Ltd., 5th Edition, 2012	
2	Rigaux, Ph, Michel Scholl, and Agnes Voisard. "Spatial databases: with application to GIS". Morgan Kaufmann, 2002.	
3	Dunckley L. Multimedia databases: An object relational approach. Addison-Wesley Longman Publishing Co., Inc.; 2003 Jan 1.	
Mode of Evaluation: CAT / Written Assignment / Quiz / FAT		
Recommended by Board of Studies	26-07-2022	
Approved by Academic Council	No. xx	Date DD-MM-YYYY

Course code	Course title	L	T	P	C
MCSE506P	DATABASE SYSTEMS LAB	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. To understand the underlying principles of Relational Database Management System.
2. To focus on the modeling and design of secure databases and usage of advanced data models.
3. To implement and maintain the structured, semi structured and unstructured data.

Course Outcome

On completion of this course, student should be able to

1. Construct database queries using Structured Query Language (SQL)
2. Design and implement applications that make use of distributed fault-tolerant databases.
3. Apply Spatial and Multimedia Database concepts to solve real-world problems.
4. Implement applications that work with structured, semi-structured, and unstructured databases
5. Create applications that use cloud storage technologies and relevant distributed file systems

Indicative Experiments

1.	Study of Basic SQL Commands. Model any given scenario into ER/EER Model
2.	Table creation with constraints, alter schema, insert values, aggregate functions, simple and complex queries with joins, Views, Subqueries.
3.	PL/SQL-Procedures, Cursors, Functions, Triggers
4.	Partition a given database based on the type of query and compares the execution speed of the query with/without parallelism.
5.	Create a distributed database scenario, insert values, fragment and replicate the database Query the distributed database
6.	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.
7.	Plot points, lines, and polygons using Spatial Databases such as Oracle Spatial, PostgreSQL, Microsoft SQL Server etc
8.	• Use Spatial Databases to store data using Latitude and Longitude, find the distance between two spatial objects, find the area of a polygon • Store and retrieve images from a multimedia database
9.	Create an XML document and validate it against an XML Schema/DTD. Use XQuery to query and view the contents of the database
10.	Execute XPATH expressions on a database.
11.	Perform the following using a MongoDB Database

	• Create an Employee Collection and insert a few documents (sample document given below for reference) { "name" : "Satish", "salary" : 30000, "address" : "Vellore", "school" : "SCOPE“ } • Display all employees whose address is vellore and salary is greater than 30000 • Update the salary for an employee by name ‘Ram’ as 40000 • Display only name and salary for all employees in the collection • Display all employees who are not from ‘SCOPE’ school • Display only documents that contains the address property
12.	Create an application that interacts with a cloud database.
	Total Laboratory Hours | 30 hours
Text Book(s)	
1.	D Abraham Silberschatz, Henry F. Korth, S. Sudarshan “Database System Concepts” 7th Edition McGraw Hill, 2021
Reference Books	
1.	Elmasri and Navathe “Fundamentals of Database Systems”, 7th Edition Addison Wesley, 2014
2.	Thomas Connolly, Carolyn Begg “Database Systems: A Practical Approach to Design, Implementation and Management” 6th Edition, Pearson India, 2015
3.	Mishra, Sanjay, and Alan Beaulieu. Mastering Oracle SQL: Putting Oracle SQL to Work. O'Reilly Media, Inc., 2004.

Mode of Evaluation: CAT / Mid-Term Lab/ FAT			
Recommended by Board of Studies	26-07-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Big Data Frameworks and Technologies	L	T	P	C
MCSE614L		2	0	0	2
Pre-requisite	NIL				Syllabus version
					v. 1.0

Course Objectives

1. To understand the need of a framework to store and process the big data.
2. To have knowledge on the Big Data Technologies for processing the Different types of Data.
3. To understand the advanced frame work for faster accessing and processing of Big Data.

Course Outcomes

Upon completion of the course the student will be able to

1. Understand the need of new frame work to deal with huge amounts of Data.
2. Demonstrate the Hadoop framework Hadoop Distributed File System and MapReduce.
3. Demonstrate the Pig architecture and evaluation of pig scripts.
4. Describe the Hive architecture and execute SQL queries on sample data sets.
5. Demonstrate spark programming with different programming languages and graph algorithms.

Module:1	**Big Data**	**3 hours**
Understanding Big Data: Concepts and terminology, Big Data Characteristics, Different types of Data, Identifying Data Characteristics - Big Data Architecture - Big Data Storage: File system and Distributed File System, NoSQL, Sharding, Replication, Sharding and Replication, ACID and BASE Properties.		
Module:2	**Hadoop Framework**	**5 hours**
Hadoop Architecture - Hadoop Distributed File System (HDFS) –YARN – Hadoop I/O – Map Reduce: Developing a map-reduce application – Map-reduce working procedure – Types and Formats - Features of Map reduce: sorting and joins- Pipelining MapReduce jobs.		
Module:3	**Hadoop Technologies-PIG**	**4 hours**
Introduction, Parallel processing using Pig, Pig Architecture, Grunt, Pig Data Model-scalar and complex types. Pig Latin- Input and output, Relational operators, User defined functions -Working with scripts. Hadoop Operations.		
Module:4	**Hive**	**4 hours**
Introduction-Hive modules, Data types and file formats, Hive QL-Data Definition and Data Manipulation-Hive QL queries, Hive QL views- reduce query complexity. Hive scripts. Hive QL Indexes- Aggregate functions-Bucketing vs Partitioning.		
Module:5	**Spark**	**5 hours**
Overview of Spark – Hadoop Overview of Spark – Hadoop vs. Spark – Cluster Design – Cluster Management – performance, Application Programming interface (API): Spark Context, Resilient Distributed Datasets, Creating RDD, RDD Operations, and Saving RDD - Lazy Operation – Spark Jobs.		
Module:6	**Data Analysis with Spark Shell**	**4 hours**
Writing Spark Application - Spark Programming in Scala, Python, R, Java - Application Execution		
Module:7	**Spark SQL and GraphX**	**4 hours**
SQL Context – Importing and Saving data – Data frames – using SQL – GraphX overview – Creating Graph – Graph Algorithms.		
Module:8	**Contemporary Issues**	**1 hour**
	Total Lecture hours:	**30 hours**

Text Book(s)			
1.	Thomas Erl, Wajid Khattak, and Paul Buhler, Big Data Fundamentals: Concepts, Drivers &Techniques, Pearson India Education Service Pvt. Ltd., First Edition, 2016.		
2.	Tom White, Hadoop: The Definitive Guide, O'Reilly Media, Inc., Fourth Edition, 2015.		
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'ReillyMedia Inc,2012		
3.	Mike Frampton, "Mastering Apache Spark", Packt Publishing, 2015.		
Mode of Evaluation: CAT / written assignment / Quiz / FAT / Project / Seminar			
Recommended by Board of Studies	DD-MM-YYYY		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Big Data Frameworks and Technologies Lab	L	T	P	C
MCSE614P		0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. To understand the need of a framework to store and process the big data.
2. To have knowledge on the Big Data Technologies for processing the Different types of Data.
3. To understand the advanced frame work for faster accessing and processing of Big Data.

Course Outcome

Upon completion of the course the student will be able to

1. Implement and evaluate the data manipulation procedures using pig, hive and spark on Hadoop frame work.

Indicative Experiments	
1.	Installing and configuring the Hadoop frame work. HDFS Commands,
2.	Map Reduce Program to show the need of combiner
3.	Map Reduce I/O Formats – Text, Key – Value
4.	Map Reduce I/O Formats – NLine – Multiline
5.	Installing and Configuring Apache PIG and HIVE
6	Sequence File Input / Output Formats
7.	Distributed Cache & Map side Join, Reduce Side Join
8.	Building and Running Spark Application
9.	Word count in Hadoop and Spark
10.	Manipulation RDD
11.	Spark Implementation of Matrix algorithms in Spark Spark Sql programming, Building Spark Streaming application
	Total Laboratory Hours **30 hours**

Reference Books

1. **Mike Frampton "Mastering Apache Spark" – Pract Publishing 2015**
2. **Tom White, "Hadoop – The Definitive Guide", O'Relly 4th Edition 2015**
3. **Nick Pentreath, "Machine Learning with Spark" Pract Publishing 2015**
4. **Mohammed Gulle , "Big Data Analytics with Spark: A Practitioner's Guide to Using Spark for Large Scale Data Analysis" – Apress 2015**
5. **Adam Shook and Donald Mine,"MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems" - O'Relly 2012**

Mode of Assessment: Continuous Assessment / FAT / Oral examination and others

Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Data Analytics	L	T	P	C
MCSE615L		2	0	0	2
Pre-requisite	NIL				Syllabus version
					v.1.0

Course Objectives

1. Explicate how to design, construct, and quality check a dataset before using it to a build prediction model.
2. Understanding the importance about feature selection in data models.
3. Understanding how information theory, similarity score and Probability theory can be used to build prediction models.

Course Outcomes

Upon completion of the course the student will be able to

1. Students will understand the basic concept of data mining and life cycles of data analytics.
2. Analyze and Apply the different data preprocessing techniques.
3. Analyze the characteristics of the data and its important feature.
4. Apply the prediction model for decision making for a given set of problems.
5. Students will understand the concept of distributed machine learning.

Module:1	Introduction to Data Mining	4 hours

Introduction to Data Mining, Challenges in Data Mining, Data Mining Tasks, Machine Learning, Predictive Data Analytics Lifecycle, Predictive Data Analytics Tools

Module:2	Exploring Data	5 hours

Different types of data, Normal Distribution, Identifying Data Quality Issues, Missing Values, Irregular Cardinality, Outlier, Advanced Data Exploration, Visualizing Relationships Between Features, Measuring Covariance and Correlation, Data Preparation, Normalization, Binning, Sampling

Module:3	Feature Selection	3 hours

Feature Reduction- Feature Selection, Statistics for Feature Selection, Chi-Squared Test for Feature Selection, ANOVA F-test for Feature Selection, RFE feature selection, Dimensionality Reduction and PCA

Module:4	Decision Tree and Similarity-based Learning	5 hours

Decision Trees, Shannon's Entropy Model, Information Gain, Standard Approach: The ID3 Algorithm, Feature Space, Measuring Similarity Using Distance Metrics, Standard Approach: The Nearest Neighbor Algorithm, Extensions and Variations, Handling Noisy Data, Efficient Memory Search, Data Normalization, Predicting Continuous Targets

Module:5	Probability-based Learning	3 hours

Fundamentals, Bayes' Theorem, Bayesian Prediction, Conditional Independence and Factorization, Standard Approach: The Naive Bayes Model

Module:6	Error-based Learning	4 hours

Simple Linear Regression, Measuring Error, Error Surfaces, Standard Approach: Multivariable Linear Regression with Gradient Descent, Multivariable Linear Regression, Gradient Descent, Choosing Learning Rates and Initial Weights.

Module:7	Distributed Machine Learning	5 hours

Data Parallelism - Splitting Input Data, Parameter Server and All-Reduce - Building a Data Parallel Training and Serving Pipeline-Model Parallelism - Splitting the Model-Pipeline Input and Layer Split- Implementing Model Parallel Training and Serving Workflows - Federated Learning and Edge Devices

Module:8	Contemporary Issues	1 hour

	Total Lecture hours	30 hours

Text Book(s)

1. John D. Kelleher, Brian Mac Namee, Aoife D'Arcy -Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, MIT Press 2020 , 2nd Edition.

2.	Jason Brownlee -Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python, First Edition, 2020.
Reference Books	
1.	Pang-Ning Tan; Michael Steinbach; Anuj Karpatne; Vipin Kumar -Introduction to Data Mining. By: Publisher: Pearson, Edition: 2nd, 2019.
2.	Guanhua Wang-Distributed Machine Learning with Python, Packt Publishing, 2022.

Mode of Evaluation: CAT / written assignment / Quiz / FAT / Project / Seminar			
Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Data Analytics Lab	L	T	P	C
MCSE615P		0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. Understand and analyze how information theory, similarity score and Probability theory can be used to build prediction models.

Course Outcome

Upon completion of the course the student will be able to

1. Analyze the different data preprocessing techniques.
2. Apply the prediction model for decision making for a given set of problems.
3. Apply regression algorithms for finding relationships between data variables

Indicative Experiments	
1.	Find the statistical measures of central tendency and dispersion such as min(), max(), mean(), meadian(), quantile(), sd() ,var() and summary() for real world datasets.
2.	Demonstrate the different data visualization techniques. (Scatter Plot, Horizontal Bar Chart, Histogram, Visualization of Time Series data (Line Graphs) for applications such as weather analysis.
3.	Perform the chi-square test and ANOVA F-test on datasets.
4.	Implement the PCA method for dimensionality reduction on datasets.
5.	Implement the RFE method and show the importance of features
6.	Implement the Decision Tree for given datasets and compute the accuracy of model.
7.	Implement the K-Nearest Neighbor Algorithm for given datasets and analyze the results.
8.	Implement the Naïve Bayes method.
9.	Implement simple linear regression program to predict the future values and analyze the goodness of fit.
10.	Implement multivariate linear regression program to predict the future values analyze the goodness of fit.
11.	Implementation of Distributed Decision Trees
	Total Laboratory Hours **30 hours**

Text Book(s)

1. John D. Kelleher, Brian Mac Namee, Aoife D'Arcy -Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, MIT Press 2020 2nd Edition.

Reference Books

1. Jason Brownlee -Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python, First Edition, 2020.
2. Guanhua Wang-Distributed Machine Learning with Python, Packt Publishing, 2022.

Mode of Assessment: Continuous Assessment / FAT / Oral examination and others

Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Data Visualization	L	T	P	C
MCSE616L		2	0	0	2
Pre-requisite	NIL				Syllabus version
					v. 1.0

Course Objectives

1. To understand the various types of data, apply and evaluate the principles of data visualization.
2. Acquire skills to apply visualization techniques to a problem and its associated dataset.
3. To apply structured approach to create effective visualizations from the massive dataset using various visualization tools.

Course Outcomes

Upon completion of the course the student will be able to

1. Analyze 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.

Module:1	Introduction to Data Visualization	4 hours

Overview of data visualization - Data Abstraction - Task Abstraction - Dimensions and Measures - Analysis: Four Levels for Validation. Statistical charts (Bar Chart - stacked bar chart – Line Chart - Histogram - Pie chart - Frequency Polygon - Box plot - Scatter plot - Regression curves.)

Module:2	Visualization Techniques	4 hours

Introduction to various data visualization tools - Scalar and point techniques - vector visualization techniques - multidimensional techniques - visualizing cluster analysis – K-means and Hierarchical Cluster techniques.

Module:3	Spatio-temporal Data Visualization	4 hours

Time Series data visualization – Text data visualization – Spatial Data Visualization

Module:4	Visual Analytics	3 hours

Networks and Trees - Heat Map – Tree Map - Map Color and Other Channels Manipulate View - Visual Attributes

Module:5	Multivariate Data Visualization	5 hours

Multivariate data visualization – Geometric projection techniques - Icon-based techniques - Pixel-oriented techniques - Hierarchical techniques - Scatterplot matrix - Hyper box - Trellis display - Parallel coordinates

Module:6	Data Visualization Tools	5 hours

Tableau functions and logics: Marks and Channels-Arrange Tables- Arrange Spatial Data- Facets into multiple views

Module:7	Visualization Dashboard Creations	4 hours

Data Dashboard- Taxonomies- User Interaction- Organizational Functions-Dashboard Design – Worksheets - Workbooks – Workbook Optimization - Protection and common mistakes. Dashboard creation using visualization tool use cases: Finance-marketing-insurance-healthcare.

Module:8	Contemporary Issues	1 hour

	Total Lecture hours:	30 hours

Text Book(s)

1. Tamara Munzer, Visualization Analysis and Design, 1st edition, CRC Press, United States, 2015.
2. Michael Fry, Jeffrey Ohlmann, Jeffrey Camm, James Cochran, Data Visualization: Exploring and Explaining with Data, South-Western College Publishing, 2021

Reference Books

1. Dr. Chun-hauh Chen, W. K. Hardle, A. Unwin, Handbook of Data Visualization, 1st edition, Springer publication, Germany, 2008.

2.	Ben Fry, Visualizing Data, 1st edition, O'Reilly Media, United States, 2008.		
3.	Avril Coghlan, A little book of R for multivariate analysis, 1st edition, Welcome Trust Sanger Institute, United Kingdom, 2013.		
Mode of Evaluation: CAT / written assignment / Quiz / FAT / Project / Seminar			
Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Data Visualization Lab	L	T	P	C
MCSE616P		0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. Analyze and solve real time data visualization scenarios using Python/R integrating with Tableau.

Course Outcome

Upon completion of the course the student will be able to

1. Integrate with Tableau for various data visualization scenarios.
2. Design visualization dashboard to support the decision making on large scale data.
3. Demonstrate the analysis of large dataset using various visualization techniques and tools.

Indicative Experiments	
1.	Acquiring and plotting data
2.	Statistical Analysis
3.	K-means and Hierarchical Cluster techniques
4.	Multivariate Analysis, Correlation, regression and analysis of variance.
5.	Financial analysis Clustering, Histogram and Heat Map.
6.	Time-series analysis Stock Market.
7.	Visualization of various massive dataset Healthcare, Census, Geospatial.
8.	Visualization on Streaming dataset Stock market, weather forecasting.
9.	Market-Basket Data analysis-visualization
10.	Text visualization using web analytics
	Total Laboratory Hours **30 hours**

Text Book(s)

1. Tamara Munzer, Visualization Analysis and Design, 1st edition, CRC Press, United States, 2015.
2. Michael Fry, Jeffrey Ohlmann, Jeffrey Camm, James Cochran, Data Visualization: Exploring and Explaining with Data, South-Western College Publishing, 2021

Reference Books

1. Dr. Chun-hauh Chen, W. K. Hardle, A. Unwin, Handbook of Data Visualization, 1st edition, Springer publication, Germany, 2008.
2. Ben Fry, Visualizing Data, 1st edition, O'Reilly Media, United States, 2008.
3. Avril, A little book of R for multivariate analysis, 1st edition, Welcome Trust Sanger Institute, United Kingdom, 2013.

Mode of Assessment: Continuous Assessment / FAT / Oral examination and others

Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course Code	Domain Specific Predictive Analytics	L	T	P	C
MCSE617L		2	0	0	2
Pre-requisite	NIL				Syllabus version
					v.1.0

Course Objectives

1. To introduce the fundamental concepts of predictive analytics.
2. To impart the knowledge on various steps that are necessary before constructing the predictive model.
3. To gain knowledge on the assessment of predictive models for decision making.

Course Outcomes

Upon completion of the course the student will be able to

1. Understand the fundamental concepts of predictive analytics.
2. Define the problem and prepare the data for analysis.
3. Construct different predictive models for decision making.
4. Apply descriptive modeling techniques for the given data.
5. Assess and interpret different predictive models.
6. Understand and apply appropriate algorithms for analyzing the data in healthcare domain.

Module:1	Overview of Predictive Analytics	4 hours

Introduction to Analytics – Predictive Analytics – Parametric vs. Non-Parametric Models -Business Intelligence – Predictive Analytics vs. Business Intelligence – Predictive Analytics vs. Statistics – Predictive Analytics vs. Data Mining – Challenges in using Predictive Analytics - Obstacles with Data - Obstacles with Modeling .

Module:2	Problem Setting, Data understanding and Preparation	4 hours

Defining Data for Predictive Modeling – Defining Target Variable – Defining Measures of Success for Predictive Models - Single Variable and Multiple Variable Summaries – Data Visualization – Variable Cleaning – Feature Creation - Case study: Fraud Detection.

Module:3	Predictive Modeling	4 hours

Parameter Settings – Measures of Interesting Rules – Deploying Association Rules – Building Classification Rules from Association Rules – Neural Networks - Decision Trees – Linear Regression - Logistic Regression – K-Nearest Neighbor Classifier.

Module:4	Descriptive Modeling	4 hours

Data Preparation Issues with Descriptive Modeling - Principal Component Analysis (PCA) Algorithm - Applying PCA to New Data - PCA for Data Interpretation - Clustering Algorithms - The K-Means Algorithm - The Kohonen SOM Algorithm - Visualizing Kohonen Maps.

Module:5	Model Ensembles and Assessing Predictive Models	4 hours

Model Ensembles - The Wisdom of Crowds - Bias Variance Tradeoff - Bagging - Boosting - Random Forests - Stochastic Gradient Boosting - Heterogeneous Ensembles - Interpreting Model Ensembles - Batch Approach to Model Assessment - Percent Correct Classification - Rank-Ordered Approach to Model Assessment - Assessing Regression Models.

Module:6	Healthcare Analytics	4 hours

Introduction - Healthcare Data Sources and Basic Analytics - Electronic Health Records - Clinical Prediction Models - Privacy-Preserving Data Publishing - Temporal Data Mining for Healthcare Data - Association Analysis - Classical Methods - Temporal Methods - Temporal Pattern Mining - Sequential Pattern Mining - Time-Interval Pattern Mining - Medical Applications - Sensor Data Analysis - Convolutional Event Pattern Discovery - Patient Prognostic via Case-Based Reasoning - Disease Progression Modeling.

Module:7	**Visual Analytics for Healthcare Data**	**5 hours**
Visual Analytics and Medical Data Visualization - Clinical Data Types - Standard Techniques to Visualize Medical Data - High-Dimensional Data Visualization - Visualization of Imaging Data - Visual Analytics in Healthcare - Visual Analytics in Public Health and Population Research - Geospatial Analysis- Visual Analytics for Clinical Workflow - Visual Analytics for Clinicians - Patient Progress and Guidelines - Visual Analytics for Patients - Assisting Comprehension		
Module:8	**Contemporary Issues**	**1 hour**
	Total Lecture hours:	**30 hours**

Text Book(s)

1.	Dean Abbott, Applied Predictive Analytics: Principles and Techniques for the professional Data Analyst, John Wiley & Sons Inc. Publishers, First edition, 2014.
2.	Chandan K. Reddy, Charu C. Aggarwal, Healthcare Data Analytics, Chapman & Hall/CRC, Data Mining and Knowledge Discovery Series, 2015.

Reference Books

1.	Klimberg, Ron and B.D. McCullough, Fundamentals of Predictive Analytics with JMP®, Cary, NC: SAS Institute Inc., Second Edition, 2016.
2.	Eric Siegel, Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die, John Wiley & Sons Inc. Publishers, Second edition, 2016.
3.	Hui Yang, Eva K. Lee, Healthcare Analytics: From Data to Knowledge to Healthcare Improvement, John Wiley & Sons Inc. Publishers, 2016.

Mode of Evaluation: CAT / written assignment / Quiz / FAT / Project / Seminar			
Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Domain Specific Predictive Analytics Lab	L	T	P	C
MCSE617P		0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. To introduce the fundamental concepts of predictive analytics.
2. To impart the knowledge on various steps that are necessary for constructing the predictive model.
3. To gain knowledge on the assessment of predictive models for decision making.

Course Outcome

Upon completion of the course the student will be able to

1. Understand the fundamental concepts of predictive analytics.
2. Define the problem and prepare the data for analysis.
3. Construct different predictive models for decision making.
4. Apply descriptive modeling techniques for the given data.
5. Assess and interpret different predictive models.
6. Understand and apply appropriate algorithms for analyzing the data in healthcare domain.

Indicative Experiments

Experiments can be implemented using R/Python.

1.	Clustering based data analytics using R/Python. (K-Means, SOM algorithms)
2.	Demonstrate the statistics for a sample data like mean, standard deviation, normal/uniform distribution, variance and correlation.
3.	Demonstrate missing value analysis, fixing missing values and outlier analysis using Healthcare domain datasets.
4.	Demonstrate data visualization, histograms and multiple variable summaries.
5.	Demonstrate transformation, scaling, binning, fixing skewed values and sampling.
6	Demonstration of Apriori algorithm on transaction dataset to find association rules.
7.	Demonstration of Linear and Logistic regression using various domain datasets.
8.	Demonstration of predictive models such as Decision Tree, Neural network and K-Nearest Neighbor using various domain datasets.
9.	Demonstration of Temporal Mining Techniques
10.	Demonstration of predictive analytics using healthcare data and microarray data.
	Total Laboratory Hours **30 hours**

Text Book(s)

1. Dean Abbott, Applied Predictive Analytics: Principles and Techniques for the professional Data Analyst, John Wiley & Sons Inc. Publishers, First edition, 2014.
2. Chandan K. Reddy, Charu C. Aggarwal, Healthcare Data Analytics, Chapman & Hall/CRC, Data Mining and Knowledge Discovery Series, 2015.

Reference Books

1. Manohar Swamynathan, Mastering Machine Learning with Python in Six Steps, Apress Publishers, First edition, 2017.

Mode of Assessment: Continuous Assessment / FAT / Oral examination and others

Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Social Network Analytics	L	T	P	C
MCSE618L		2	0	0	2
Pre-requisite	NIL				Syllabus version
					v.1.0

Course Objectives

1. Understand the components and entities of the social network
2. Analyze social media data to comprehend user sentiments and recommend the essential information appropriately.
3. Model and visualize the social network

Course Outcomes

1. Illustrate the basic concepts of social network.
2. Analyse the networks to find prominent actors and relate social network models.
3. Develop social network applications using tools and techniques.
4. Detect and analyze the communities in social networks.
5. Design a system to assimilate information available on the web to model and build Social Network Application.

Module:1	Fundamentals of Social Network Analysis	4 hours

Social Network Perspective, Fundamentals concepts in Network Analysis: Sociogram, Sociometry. Social Network Data: Types of Networks: One-Mode, Two-Mode, Affiliation, Ego-centered and Special Dyadic Networks, Network Data, Measurement and Collection, Notations for Social Network Data: Graphs, Directed, Singed, Valued graphs, Multigraph, Relations and Matrices.

Module:2	Centrality and Prestige	4 hours

Prominence: Actor-Centrality, Prestige, Group-Centrality, Prestige, Non directional Relations-Degree, Closeness, Betweenness, Eigen Vector Centrality, Directional Relations-Centrality, Prestige.

Module:3	Structural Balance and Transitivity	3 hours

Structural Balance: Signed Non directional, Signed Directional Relations, Checking for Balance, Index for Balance, Clusterability-Theorems, Clustering Coefficient and Transitivity.

Module:4	Cohesive Subgroups	5 hours

Social Group and Subgroup-Notation, Subgroups Based on Complete Mutuality: Clique, Reachability and Diameter: n-cliques, n-clans and n-clubs, Subgroups Based on Nodal Degree: k-plexes, k-cores, Measures of Subgroup Cohesion, Community detection using Subgroups and Betweenness.

Module:5	Roles and Positions	4 hours

Structural Equivalence: Definition, Social Roles and , Positional Analysis, Measuring Structural Equivalence, Representation of Network Positions, Block Models-Introduction, Network Positions and roles-Introduction

Module:6	Dyadic and Triadic Methods	4 hours

Dyads: Definitions, Dyad Census, Index, Simple Distributions, Triads: Random Models and Substantive Hypotheses, Triad Census, Distribution of a Triad Census- Mean and Variance, Testing Structural Hypotheses.

Module:7	Models in Social Network	5 hours

Small world network- Watt Strogatz networks - statistical models for social networks - network evaluation model - Preferential attachment - power law - Random Model : Erdos -Renyi model - Barabasi Albert model - Epidemic model - Case study: Text and opinion Analysis

Module:8	Contemporary Issues	1 hour

	Total Lecture hours:	30 hours

Text Book(s)

1. Wasserman Stanley, and Katherine Faust, Social Network Analysis: Methods and Applications, Structural Analysis in the Social Sciences. Cambridge University Press, 2012 Online Edition.

2.	Albert-László Barabási, Network Science, Cambridge University Press, 1st edition, 2016.		
Reference Books			
1.	John Scott, "Social Network Analysis", Sage Publications Ltd., Fourth Edition, 2017.		
2.	David Knoke & Song Yang, "Social Network Analysis", Sage Publishing, Third Edition, 2020.		
Mode of Evaluation: CAT / written assignment / Quiz / FAT / Project / Seminar			
Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Social Network Analytics Lab	L	T	P	C
MCSE618P		0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. Understand the components of the social network
2. Analyze social media data to understand user sentiment and recommend the requisite information accordingly.
3. Model and visualize the social network

Course Outcome

Upon completion of the course the student will be able to

1. Demonstrate the basic properties of social network
2. Demonstrate of analysis of social networks to find prominent actors and apply social network models.
3. Develop social network applications using visualization tools.
4. Detect and analyze the communities in social networks.
5. Design a system to harvest information available on the web to model and build Social Network Application.

Indicative Experiments

1.	Study and demonstrate to find the basic properties of a Graph/Social Network.
2.	Demonstrate the calculation of Centrality measures.
3.	Demonstrate the ranking of web pages in a web graph.
4.	Find divisions in a Social Network.
5.	Implement Community Detection algorithms on a Social Network.
6.	Demonstrate modelling of Social Networks.
7.	Visualize multidimensional Social Network.
8.	Applications of Classification and Clustering on a Social Network.
9.	Design and implement a Sentiment Analyzer.
10.	Design and implement a Social Network.
	Total Laboratory Hours **30 hours**

Text Book(s)

1. Wasserman Stanley, and Katherine Faust, Social Network Analysis: Methods and Applications, Structural Analysis in the Social Sciences. Cambridge University Press, 2012 Online Edition.
2. Albert-László Barabási, Network Science, Cambridge University Press, 1st edition, 2016.

Reference Books

1. John Scott, "Social Network Analysis", Sage Publications Ltd., Fourth Edition, 2017.
2. David Knoke & Song Yang, "Social Network Analysis", Sage Publishing, Third Edition, 2020.

Mode of Assessment: Continuous Assessment / FAT / Oral examination and others

Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Text and Speech Analytics	L	T	P	C
MCSE619L		2	0	0	2
Pre-requisite	NIL				Syllabus version
					v.1.0

Course Objectives

1. To introduce the tools and techniques for performing text and speech analytics in diverse contexts.
2. To understand the tools and technologies involved in developing text and speech applications.
3. To demonstrate the use of computing for building applications in text and speech processing.

Course Outcomes

Upon completion of the course the student will be able to

1. Develop tools to analyse the syntax and semantics of a statement written in a natural language.
2. Apply machine learning and deep learning techniques to natural language processing.
3. Use signal processing techniques to analyze/represent speech.
4. Execute trials of speech systems.
5. Evaluate the performance of NLP & Speech systems.

Module:1	Introduction to Text Processing and Language Modeling	5 hours

Introduction to Natural Language Processing (NLP) and Levels of NLP - Regular Expression - Basic Text processing- Text normalization - Vector Semantics and embedding : Lexical Semantics , Vector Semantics , Words and Vectors - Pointwise Mutual Information, N-gram Language Models : N-grams, Smoothing.

Module:2	Parts of speech and Named entities	4 hours

Parts of Speech Tagging - Hidden Markov Model - Conditional Random Fields. Constituency Grammars: Constituency, Context Free Grammars, Dependency Parsing: Dependency Relations, Dependency Formalism, Neural Dependency Parser.

Module:3	Logical Representations of Sentence Meaning	4 hours

Logical Representations of Sentence Meaning, Word Sense and Word Net, Word Sense Disambiguation, Word Sense Induction.

Module:4	Applications of Text and NLP	4 hours

Naive Bayes and Sentiment Analysis: Naive Bayes for text classification, Information Extraction - Relation extraction. Learning Architectures for Sequence Processing: Recurrent Neural Networks for text classification- Long Short-Term Memory (LSTM).

Module:5	Phonetics	3 hours

Speech Sounds and Phonetic Transcription, Articulatory Phonetics – Prosody - Acoustic Phonetics and Signals - Phonetic Resources.

Module:6	Automatic Speech Recognition	4 hours

Automatic Speech Recognition (ASR) Task - Feature Extraction : Log Mel Spectrum - Speech Recognition Architecture - Connectionist Temporal Classification(CTC) - ASR Evaluation: Word Error Rate.

Module:7	Text-To-Speech	5 hours

Text-To-Speech(TTS) Preprocessing: Text normalization – TTS: Spectrogram Prediction – TTS: Vocoding - TTS Evaluation.

Module:8	Contemporary Issues	1 hour

	Total Lecture hours:	**30 hours**

Text Book(s)	
1.	Jurafsky, D. and J. H. Martin, Speech and language processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (3rd Draft), 2021.
Reference Books	
1.	John Atkinson-Abutridy, Text Analytics: An Introduction to the Science and Applications of Unstructured Information Analysis, CRC Press, 2022.
2.	Introduction to Voice Computing in Python, Jim Schwoebel, NeuroLex, 2018
3.	Theory and Applications of Digital Speech Processing, Lawrence R. Rabiner, Ronald W. Schafe, 1st Edn. Pearson, 2010.
4.	Srinivasa-Desikan, Bhargav. Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras. Packt Publishing Ltd, 2018.

Mode of Evaluation: CAT / written assignment / Quiz / FAT / Project / Seminar			
Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Text and Speech Analytics Lab	L	T	P	C
MCSE619P		0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. To introduce the tools and techniques for performing text and speech analytics in diverse contexts.
2. To understand the tools and technologies involved in developing text and speech applications.
3. To demonstrate the use of computing for building applications in text and speech processing.

Course Outcomes

Upon completion of the course the student will be able to

1. Develop tools to analyse the syntax and semantics of a statement written in a natural language.
2. Apply machine learning and deep learning techniques to natural language processing.
3. Use signal processing techniques to analyze/represent speech.
4. Execute trials of speech systems.
5. Evaluate the performance of NLP & Speech systems.

	Indicative Experiments
1.	Introduction to text processing packages in Python.
2.	Demonstration of Genism for Vectorizing Text, Transformations and n-grams.
3.	Demonstration of Part-of-Speech tagging using spaCy.
4.	Demonstration of text parsing, topic modeling, text clustering and text classification.
5.	Demonstration of Deep learning techniques for text classification and for designing a chatbot.
6	Analyze Speech signal - Fast Fourier Transform (FFT), spectrogram, Linear predictive coding, Mel-frequency Cepstral Coefficients (MFCC) features.
7.	Demonstration of Hidden Markov Model based Isolated word recognition.
8.	Demonstration of Continuous speech recognition using CTC.
9.	Demonstration of Alexa speech enabled application development system.
10	Demonstration of Google voice API based speech transcription system.
	Total Laboratory Hours **30 hours**

Text Book(s)

1. Jurafsky, D. and J. H. Martin, Speech and language processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (3rd Draft), 2021.

2. Srinivasa-Desikan, Bhargav. Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras. Packt Publishing Ltd, 2018.

Reference Books

1. John Atkinson-Abutridy, Text Analytics: An Introduction to the Science and Applications of Unstructured Information Analysis, CRC Press, 2022.

<table>
<tr><td colspan="4">2. Introduction to Voice Computing in Python, Jim Schwoebel, NeuroLex, 2018

3. Theory and Applications of Digital Speech Processing, Lawrence R. Rabiner, Ronald W. Schafe, 1st Edn. Pearson, 2010.</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4">Mode of Assessment: Continuous Assessment / FAT / Oral examination and others</td></tr>
<tr><td>Recommended by Board of Studies</td><td colspan="3">18-11-2022</td></tr>
<tr><td>Approved by Academic Council</td><td>No. xx</td><td>Date</td><td>DD-MM-YYYY</td></tr>
</table>

Course code	Analytics for Internet of Things	L	T	P	C
MCSE620L		2	0	0	2
Pre-requisite	NIL				Syllabus version
					v.1.0

Course Objectives

1. To introduce the fundamentals of IoT data analytics and major challenges in IoT data analytics.
2. To provide knowledge on IoT network architecture and design.
3. To understand smart objects and IoT networking protocols.

Course Outcomes

Upon completion of the course the student will be able to

1. Understand the specific challenges in applying data analytics techniques over IoT data.
2. Will know IoT network architecture and design.
3. Smart objects and connecting smart objects
4. Analyze various IoT networking protocols.
5. Apply IoT analytics for cloud and data science for IoT analytics.

Module:1	**IoT Analytics and Challenges (Ch1)**	**3 hours**
Defining IoT analytics: Defining Analytics, Defining Internet of Things, The concepts of constrained - IoT analytics challenges: the Data volume, Problem with time and space, Data quality, Analytics Challenges - Business value concerns.		
Module:2	**IoT Network Architecture and Design(T2:Ch2)**	**5 hours**
Drivers behind New Network Architectures, Comparing IoT Architectures, A Simplified IoT Architecture, The Core IoT Functional Stack, IoT Data Management and Compute Stack.		
Module:3	**Smart Objects: The Things in IoT(T2:Ch2)**	**3 hours**
Sensors, Actuators, and Smart Objects, Sensor Networks		
Module:4	**Connecting Smart Objects(T1:Ch2)**	**6 hours**
Communications Criteria, Range, Frequency Bands, Power Consumption, Topology, Constrained Devices, Constrained-Node Networks, IoT Access Technologies, IEEE 802.15.4, IEEE 802.15.4g and 802.15.4e, LoRaWAN.		
Module:5	**IoT Networking Protocols(T1:Ch2)**	**3 hours**
IoT networking data messaging protocols, Message Queue Telemetry Transport (MQTT), Hyper-Text Transport Protocol (HTTP), Constrained Application Protocol (CoAP), Data Distribution Service (DDS).		
Module:6	**IoT Analytics for the Cloud (T1:Ch3)**	**4 hours**
Building elastic analytics, Elastic analytics concepts, designing for scale, Cloud security and analytics, The AWS overview, Microsoft Azure overview.		
Module:7	**Data Science for IoT Analytics(T1:Ch10)**	**5 hours**
Machine learning (ML), Feature engineering with IoT data, Validation methods, Understanding the bias–variance tradeoff, Comparing different models to find the best fit using R, Random forest models using R, Anomaly detection using R.		
Module:8	**Contemporary Issues**	**1 hour**
	Total Lecture hours:	**30 hours**

Text Book(s)

1. Andrew Minteer , Analytics for the Internet of things, Packt publishing 2017.
2. David Hanes, Gonzalo Salgueiro, Patrick Grossetete, Rob Barton and Jerome Henry, IoT Fundamentals: Networking Technologies, Protocols and Use Cases for Internet of Things, Cisco Press, 2017.

Reference Books

1.	Pethuru Raj, Anupama C. Raman, The Internet of Things, Enabling Technologies, Platforms, and Use Cases, CRC Press, 2017.
2.	Rajkumar Buyya, Amir Vahid Dastjerdi, Internet of Things Principles and Paradigms, Morgan Kaufmann, 1st edition, 2016.
3.	Marco Schwartz, Internet of Things with Arduino Cookbook, Packt Publishing,2016
4.	Adeel Javed, "Building Arduino Projects for the Internet of Things: Experiments with Real-World Applications", 1st Edition, Apress, 2016.

Mode of Evaluation: CAT / written assignment / Quiz / FAT / Project / Seminar			
Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY

Course code	Analytics for Internet of Things	L	T	P	C
MCSE620P		0	0	2	1
Pre-requisite	NIL	Syllabus version			
		v.1.0			

Course Objectives

1. To introduce the fundamentals of IoT data analytics and major challenges in IoT data analytics.
2. To provide knowledge on IoT network architecture and design.
3. To understand smart objects and IoT networking protocols.

Course Outcome

Upon completion of the course the student will be able to

1. Understand the specific challenges in applying data analytics techniques over IoT data.
2. Will know IoT network architecture and design.
3. Smart objects and connecting smart objects
4. Analyze various IoT networking protocols.
5. Apply IoT analytics for cloud and data science for IoT analytics.

Indicative Experiments	
1.	Study different sensors, actuators, and their applications.
2.	Write a program using Arduino IDE for Blink LED.
3.	Write a program to interface the DHT11 sensor with Arduino/Raspberry to print temperature and humidity readings.
4.	Write an application to read temperature from the environment. If the temperature crosses the threshold value then it notifies with a buzzer.
5.	Study and implement MQTT protocol using Arduino.
6	Study and implement COAP protocol using Arduino.
7.	Write a program on Arduino/Raspberry Pi to upload temperature and humidity data to the ThingSpeak cloud.
8.	Write an application to send Light Sensor Values to the ThingSpeak cloud
9.	Write an application to send Temperature and Humidity Values to the ThingSpeak cloud
10.	Implementation of Machine learning approaches over IoT data.
	Total Laboratory Hours **30 hours**

Text Book(s)

1.	Andrew Minteer , Analytics for the Internet of things, Packt publishing 2017.
2.	David Hanes, Gonzalo Salgueiro, Patrick Grossetete, Rob Barton and Jerome Henry, IoT Fundamentals: Networking Technologies, Protocols and Use Cases for Internet of Things, Cisco Press, 2017.

Reference Books

1.	Pethuru Raj, Anupama C. Raman, The Internet of Things, Enabling Technologies, Platforms, and Use Cases, CRC Press, 2017.
2.	Rajkumar Buyya, Amir Vahid Dastjerdi, Internet of Things Principles and Paradigms, Morgan Kaufmann, 1st edition, 2016.
3.	Marco Schwartz, Internet of Things with Arduino Cookbook, Packt Publishing,2016
4.	Adeel Javed, "Building Arduino Projects for the Internet of Things: Experiments with Real-World Applications", 1st Edition, Apress, 2016.

Mode of Assessment: Continuous Assessment / FAT / Oral examination and others

Recommended by Board of Studies	18-11-2022		
Approved by Academic Council	No. xx	Date	DD-MM-YYYY