SIES College of Arts, Science and Commerce, Sion (W)

EMPOWERED AUTONOMOUS COLLEGE

RISE WITH EDUCATION
NAAC REACCREDITED - 'A' GRADE
BEST COLLEGE AWARD - UNIVERSITY OF MUMBAI
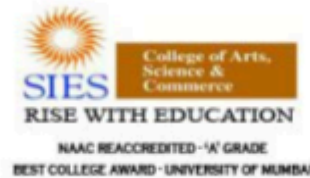
Syllabus for
Program: Second Year Masters of Science (NEP) Course: Computer Science
Semester: III & IV

With effect from
Academic Year 2024 -25

Preamble

The Master of Science in Computer Science (M.Sc. Computer Science) is an advanced program that combines academic research and industry standards, addressing the evolving needs of both the industry and research domains. The curriculum focuses on cutting-edge technologies and industry insights, ensuring students gain the necessary expertise to thrive in the current landscape. Throughout this program, students will delve into a wide range of relevant subjects.

In the first year, they study subjects including Machine Learning, Natural Learning process , IOT & Embedded System , Web Mining.

In the M.Sc. Second year program, students will engage with major mandatory subjects such as Social Network Analysis , Simulation & Modelling,Big Data Analytics & Visualization. They can also choose from major elective subjects like NoSQL or Spark Technologies . Research projects in both semesters provide practical experience and foster critical skills. This comprehensive curriculum equips students with the latest knowledge and prepares them for diverse opportunities in computer science. These courses are carefully designed to equip students with the skills required to tackle the challenges and opportunities in the rapidly expanding field of Computer Science.

The program is structured as a fusion of Major Mandatory and Major Elective courses, encompassing the latest trends and advancements in Computer Science. In each semester, students have the opportunity to choose one elective subject aligned with their interests from a selection of three options. The Major Mandatory courses establish a strong foundation in fundamental concepts of Computer Science and Research, while the electives enhance their knowledge for real-world applications. To further enhance the students; readiness for industry, the curriculum incorporates a mandatory On Job Training (OJT) component in Semester II.

This intensive training, equivalent to a full course, provides invaluable exposure to real-world scenarios within IT or IT-related organisations. By applying their theoretical knowledge in practical settings, students gain firsthand experience and develop the necessary skills to thrive in the professional world. In addition to technical skills, this program also focuses on cultivating research ethics and promoting a research-oriented mindset among learners. The inclusion of a Research Methodology Course helps students develop a strong research attitude, enabling them to contribute meaningfully to the advancement of Computer Science.

The comprehensive education provided by the M.Sc. in Computer Science program equips students with the confidence to adapt and excel in an ever evolving industry and academic landscape. The curriculum's continuous refinement has been made possible through the valuable inputs, suggestions, and observations of colleagues at the University of Mumbai, experts from premier institutions, and industry professionals. We extend our gratitude to all those who have directly or indirectly contributed to the development of this program. With these combined efforts, the M.Sc. in Computer Science program aims to empower students with the skills and knowledge necessary to thrive in the digital world.

2. Aims and Objectives

The aims and objectives of the M.Sc. Computer Science programs collectively aim to develop well-rounded computer science professionals who are not only technically competent but also capable of contributing to research, innovation, and the overall advancement of the field.

Program Objectives:

❖ **Provide In-depth Knowledge:** The program aims to provide students with a comprehensive understanding of the key concepts, theories, and methodologies in Computer Science. It covers a range of topics including machine learning, data mining, data visualization, and data management, enabling students to develop a deep knowledge base in these areas.

❖ **Develop Programming Skills:** The program aims to equip students with strong programming skills by providing hands-on experience with different tools and technologies. Students will gain proficiency in designing front-end and back-end solutions, enhancing their ability to develop robust and scalable applications.

❖ **Foster Problem-solving Abilities:** The program aims to enhance students' problem solving abilities by training them to approach real-world data challenges critically and creatively. Students will learn to identify problems, design appropriate data analysis strategies, and develop innovative solutions using their knowledge of Computer Science.

❖ **Encourage Collaboration and Teamwork:** The program aims to foster collaboration and teamwork skills among students, recognizing that computer science projects often require interdisciplinary collaboration. Students will learn to effectively communicate, collaborate, and contribute as part of a team, preparing them for collaborative work environments.

❖ **Foster Industry Relevance:** The program aims to stay up-to-date with industry trends and technologies to ensure graduates are well-prepared for the demands of the Computer Science job market. Through industry partnerships and internships, students will have the opportunity to gain practical experience and stay in touch with the latest advancements in the field.

❖ **Professional Development:** The program aims to prepare students for successful careers in the field of Computer Science. In addition to technical skills, students will develop professional skills such as teamwork, project management, and leadership. The program may also provide networking opportunities, internships, or collaborations with industry partners to enhance students' industry readiness and employability.

❖ **Cultivate Research Skills:** The program aims to cultivate research skills among students by providing a Research Methodology Course and encouraging participation in research projects. Students will learn to conduct literature reviews, design experiments, analyse data, and present their findings, fostering a research-oriented mindset and contributing to the advancement of Computer Science.

By incorporating these objectives, the M.Sc. Computer Science program aims to produce graduates who possess a strong foundation in Computer Science, are adept at problem solving and collaboration, have industry-relevant skills, and are well-prepared for both research and professional roles in the field.

3. Learning Outcomes

● Develop a solid foundation in fundamental concepts, theories, and methodologies of Computer Science.
● Offer opportunities for specialisation in a chosen area of Computer Science.
● Foster a research-oriented mindset and contribute to the advancement of Computer Science. Prepare learners for lifelong learning, adapting to emerging technologies and industry requirements.
● Inculcate professional attitudes, leadership qualities, and social responsibility.
● Equip students with industry-relevant skills and experiences for successful careers.
● Enhance critical thinking and innovative problem-solving abilities.

M.Sc. Computer Science Syllabus
Credit Based System and Grading System
Academic year 2024-2025

Semester – III						
Course Code	**Course Type**	**Course Title**	**Credits**	**Lectures/Week**		
				Theory	**Practical (2 lectures)**	**Total**
SIPCSCC611	Major Subject	Social Network Analysis	6	4		4
SIPCSCCP611	Major Practical	Practicals of Social Network Analysis			2	2
SIPCSCC612	Major Subject	Simulation & Modelling	6	4		4
SIPCSCCP612	Major Practical	Practicals of Simulation & Modelling			2	2
SIPCSEL611	Electives Subject	NoSQL Technologies	4	3		3
SIPCSELP611	Elective Subject	Practicals of NoSQL Technologies			1	1
SIPCSRP611	Research Project	—	6			6
Total						**22**

Semester – IV						
Course Code	**Course Type**	**Course Title**	**Credits**	**Lectures/Week**		
				Theory	**Practical (2 lectures)**	**Total**
SIPCSCC621	Major Subject	Big Data Analytics	6	4		4
SIPCSCCP621	Major Practical	Practicals of Big Data Analytics			2	2
SIPCSCC622	Major Subject	Data Analysis and Visualization	6	4		4
SIPCSCCP622	Major Practical	Practicals of Data Analysis and Visualization			2	2
SIPCSEL621	Electives Subject	Spark Technologies	4	3		3
SIPCSELP621	Elective Subject	Practical of Spark Technologies			1	1
SIPCSRP621	Research Project	—	6			6
Total						**22**

Course Code	Course Title	Credits	Lectures /Week
SIPCSCC611	**Social Network Analysis**	**4**	**4**

About the Course:
Social network perspectives prioritize the interdependence among social actors, view the social world as patterns or regularities in relationships, and focus on how relationships affect networked actors' behavior and outcomes.

Course Outcomes:
- ❖ To introduce students to the fundamental concepts of Social Network Analysis.
- ❖ To provide students with a comprehensive understanding of Network using Graph
- ❖ To familiarise students with different Networks Concepts , Centrality ,analyzing Network Architecture and Approaches to network positions and social roles.

Unit	Topics	No of Lectures
I	**Introduction to Social Network and their Relationships:** Basics , Social Network, Types of ties. **Analysing relationships to understand people and groups:** Binary and valued relationships , Symmetric and Asymmetric relationships, Multimode Relationship. **Using Graph theory for social networks analysis:** Basics , Adjacency matrices, Edge List , Adjacency List. **Graph Traversal:** Depth-First Traversal, Breadth First Traversal Implementing Dijkstra's Algorithm , Graph Distance and Diameter,Social Network VS Link Analysis , Ego-Centric and Socio-Centric Density.	15L
II	**Understanding Networks:** Network Density, Reachability , Connectivity , Reciprocity, Group-external and group-internal ties in networks, Ego networks ,Triads , Transitivity, Triad census, Bridges, Neighborhood Overlap, Structural holes. **Centrality :** Local Centrality- Degree Centrality, Closeness Centrality, Betweenness Centrality. **Global Centrality:** Google Page Rank Algorithm, Eigenvector Centrality. **Analyzing network structure - Bottom-up approach :** Clique , N Clique , N- Clan , K- Plexes , K- Core , F Group. **Top- Down approach :** Components, Block And Cut-points, Lambda set and Bridges.	15L
III	**Approaches to network positions and social roles:**Defining equivalence or similarity. **Equivalences in SNA:** Structural, Automorphic and Regular Equivalence. **Measuring Similarity/ Dissimilarity:**Valued Relationships , Pearson correlations covariance and cross-product in measuring similarity /dissimilarity. **Understanding Clustering:** Types of Hierarchical Clustering, Measuring Clustering: Euclidean Distance ,Manhattan Distance,	15L

		Jaccard Similarity Coefficient,Cosine Similarity,Hamming Distance	
IV		**Understanding mode networks:** Bipartite Data Structure, Visualising two-mode data. **Quantitative analysis using two-mode :** Two Mode SVD , Two mode Factor analysis, Two mode correspondence analysis. **Qualitative analysis using two-mode :** Two mode Core- Periphery analysis, Affiliation and Attribute network.	

Textbook(s):
1. Introduction to Social Network Methods: Robert A. Hanneman, Mark Riddle, University of California, 2005 [Published in digital form and available at http://faculty.ucr.edu/~hanneman/nettext/index.html].
2. Social Network Analysis for Startups- Finding connections on the social web: Maksim Tsvetovat, Alexander Kouznetsov, O'Reilly Media, 2011.
3. Social Network Analysis- 3rd edition, John Scott, SAGE Publications, 2012.

Additional Reference(s):
1. Exploratory Social Network Analysis with Pajek, Second edition: Wouter deNooy, Andrej Mrvar, Vladimir Batagelj, Cambridge University Press, 2011.
2. Analyzing Social Networks, Stephen P Borgatti, Martin G. Everett, Jeffrey C.Johnson, SAGE Publications, 2013.
3. Statistical Analysis of Network Data with R: Eric D. Kolaczyk, Gábor Csárdi, Springer, 2014.
4. Network Analysis: Methodological Foundations, (Editors) Ulrik Brandes, ThomasErlebach. Springer, 2005.
5. Models and Methods in Social Network Analysis: (Editors) Peter J. Carrington, John Scott,Stanley Wasserman, Cambridge University Press, 2005.

Course Code	Course Title	Credits	Lectures /Week
SIPCSCCP611	**Social Network Analysis -Practicals**	**2**	**2**
1	Write a program to compute the following for a given network a. Number of edges b. Number of vertices c. Degree of nodes d. Node with the lowest Degree e. Adjacency Matrix		
2	View Data collection and import data from excel. Create a social graph of imported data.		
3	Write a program to find a. Density b. Degree c. Reciprocity d. Transitivity Index e. Transitivity Census		
4	Write a program to find out a. Degree Centrality b. Closeness Centrality c. Betweenness Centrality for a given graph.		
5	For a given network, find the following a. Length of shortest path from given node to another node. b. Density of the graph		
6	Write a program to distinguish a network as a matrix, a network as an edge list, and a network as a sociogram(network graph) using three distinct network representations.		
7	Write a program to create the following sample graph: a. Full graph b. Ring graph c. Star graph d.Random graph		
8	Write a program to create the following sample graph: a. Cliques in a network b. Components in a network c. Random walk on a graph d. Change the colour of vertices and edges e. Save the graph in the form of an edge list on your computer.		
9	For a given data Write a program to find the following a.Hamming distance b.Euclidean Distance c.Manhattan distance		
10	For a given dataset Write a program to perform clustering.		

Course Code	Course Title	Credits	Lectures /Week
SIPCSCC612	**Simulation & Modelling**	**4**	**4**

About the Course:

In Simulation & Modelling, Modelling is the process of representing a model which includes its construction and working. This model is similar to a real system, which helps the analyst predict the effect of changes to the system. Simulation of a system is the operation of a model in terms of time or space, which helps analyse the performance of an existing or a proposed system. In this course, we will discuss the concept and classification of Modelling & Simulation, different statistical models, applications, and other key ideas.

Course Outcome:
- ❖ To introduce students to the fundamental concepts and components of computer simulation and modelling.
- ❖ To provide students with a comprehensive understanding of statistical models commonly used in simulation studies.
- ❖ To familiarise students with different queueing models and their characteristics for analysing system performance.
- ❖ To develop students' skills in generating random numbers and random variates for simulation experiments.
- ❖ To equip students with the knowledge and techniques for input modelling and selecting appropriate input models for simulations.

Unit	Topics	No of Lectures
I	**Introduction to Simulation:** System and System environment, Components of system, Type of systems, Type of models, Steps in simulation study, Advantages and Disadvantages of simulation. **General Principles:** Concepts of discrete event simulation, List processing Statistical Models in Simulation: Useful statistical model, Discrete distribution, Continuous distribution, Poisson process, Empirical distribution.	15L
II	**Queueing Models:** Characteristics of Queueing systems, Queueing notations, Long run measures of performance of Queueing systems, Steady state behaviour of infinite population Markovian models, Steady state behaviour finite population model, Network of Queues Random Number Generation, Random Variate Generation, Input Modelling, and Output Analysis **Random Number Generation:** Properties of random numbers, Generation of pseudo random numbers, Techniques for generating random numbers, Tests for random numbers.	15L
III	**Random Variate Generation:** Inverse transform technique, Convolution method, Acceptance rejection techniques **Input Modelling:** Data Collection, Identifying the Distribution of data, Parameter estimation, Goodness of fit tests, Selection input model without data, Multivariate and Time series input models. **Verification and Validation of Simulation Model:** Model building, Verification, and Validation, Verification of simulation models, Calibration and Validation of models.	15L
IV	**Output Analysis for a Single Model:** Types of simulations with respect to output analysis, Stochastic nature of output data,	15 L

Measure of performance and their estimation, Output analysis of terminating simulators, Output analysis for steady state simulation
Designing state-based behaviour: Statecharts, State transitions, Viewing and debugging Statecharts at runtime, Statecharts for dynamic objects.
Discrete events and Event model object: Discrete event, Event-the simplest low level model object, Dynamic events, and Exchanging data with external world. Presentation and animation: Working with shapes, groups and colours, Designing interactive models: using controls, Dynamic properties of controls, 3D Animation.

Textbooks:
1. Discrete Event System Simulation, Jerry Banks, John Carson, Barry Nelson, David Nicol, 3rd Edition, Pearson, 2013
2. Simulation Modeling and Analysis, 5th Edition, McGRAW- HILL, Averill Law, W. David Kelton, 2012
3. Simulation: The Practice of Model Development and Use by Stewart Robinson, John Wiley and Sons, Ltd, 2004.
4. The Big Book of Simulation Modeling: Multi Method Modeling by Andrei Borshchev, 2013

Reference Books:
1. System Simulation, Geoffrey Gordon, Pearson, 2007
2. Theory of Modeling and Simulation, Bernard P. Zeigler, Tag Gon Kim , Herbert Praehofer, Academic Press, 2011
3. System Simulation with Digital Computer, NarsinghDeo, PHI

Course Code	Course Title	Credits	Lectures /Week
SIPCSCCP612	**Simulation & Modelling**	**2**	**2**
1	Design and develop agent based model by - Creating the agent population - Defining the agent behaviour - Add a chart to visualise the model output. [Use a case scenario like grocery store, telephone call centre etc for the purpose]		
2	Design and develop agent based model by - Creating the agent population - Defining the agent behavior - Adding a chart to visualise the model output - Adding word of mouth effect - Considering product discards Considering delivery time [Use a case scenario like a restaurant].		
3	Design and develop agent based model by - Creating the agent population - Defining the agent behavior		

	- Adding a chart to visualize the model output - Adding word of mouth effect - Considering product discards - Consider delivery time - Simulating agent impatience - Comparing model runs with different parameter values [Use a scenario like market model]
4	Design and develop System Dynamic model by - Creating a stock and flow diagram - Adding a plot to visualize dynamics - Parameter Variation Calibration [Use a case scenario like spread of contagious disease for the purpose]
5	Design and develop a discrete-event model that will simulate process by: - Creating a simple model - Adding resources - Creating 3D animation - Modeling delivery [Use a case situation like a company"s manufacturing and shipping].
6	Design and develop time-slice simulation for a scenario like airport model to design how passengers move within a small airport that hosts two airlines, each with their own gate. Passengers arrive at the airport, check in, pass the security checkpoint and then go to the waiting area. After boarding starts, each airline"s representatives check their passengers" tickets before they allow them to board
7	Verify and validate a model developed like bank model or manufacturing model
8	Create defense model to stimulate aircraft behavior
9	Stimulate the travelling sales man problem to compute the shortest path
10	Stimulate the Urban dynamics to address the scenarios like: (a) The problem of public transport line (b) To compute the time taken for train to enter the station

Course Code	Course Title	Credits	Lectures /Week
SIPCSEL611	NoSQL Technologies	3	3

About the Course:
NoSQL databases diverge from traditional relational databases by leveraging non-tabular storage methods, such as document databases tailored for storing data in document formats.This course provides an introduction to the evolution of database technology, specifically focusing on the transition from traditional DBMS to NoSQL databases.

Course Outcome:
- ❖ To introduce students to the fundamental concepts of NoSQL
- ❖ To provide students with a comprehensive understanding of interfacing & interaction with NoSQL.
- ❖ To provide students with a comprehensive understanding of Storage Architecture and CRUD operations.

Unit	Topics	No of Lectures
I	**Introduction to NoSQL and Interfacing with NoSQL Data Stores** **Basics Introduction to NoSQL:**Understanding the need for NoSQL databases,Comparison with traditional relational databases, Characteristics of NoSQL,Overview of different types of NoSQL databases(Document, Key-Value, Column-family, Graph),Advantages and Drawbacks, NoSQL Products	15L
II	**Document Databases :** Introduction to document-oriented databases Case study: MongoDB ,Data modeling with documents,CRUD operations in MongoDB , Querying data in MongoDB, Indexing and ordering in MongoDB. **Querying NoSQL Stores:** Similarities Between SQL and MongoDB Query Features.	15L
III	**Key-Value Stores:** Introduction to key-value stores, Case study: Redis,Data modeling with key-value pairs,Data structures and commands in Redis. **Managing Transactions and Data Integrity:** RDBMS and ACID, Distributed ACID Systems, Upholding CAP, Consistency**.** **CouchDB:** Installation, CouchDB Databases, CouchDB Document, Connectivity.	15L

Textbooks:
1. QL & NoSQL Databases, Andreas Meier · Michael Kaufmann, Springer Vieweg, 2019
2. Professional NoSQL by Shashank Tiwari, Wrox-John Wiley & Sons, Inc, 2011
3. SQL & NoSQL Databases, Andreas Meier · Michael Kaufmann, Springer Vieweg, 2019
4. NoSQL: Database for Storage and Retrieval of Data in Cloud, Ganesh Chandra Deka, CRC Press, 2017
5. Demystifying NoSQL by Seema Acharya, Wiley, 2020

Course Code	Course Title	Credits	Lectures /Week
SIPCSELP611	**NoSQL Technologies**	**1**	**1**
1	Setting up and Exploring MongoDB a) Install MongoDB on your local machine or lab server. b) Create a new MongoDB database and collection. c) Insert sample data into the collection. d) Retrieve and display data from the collection using MongoDB queries.		
2	Interacting with Redis a) Install Redis on your lab server or local machine. b) Store and retrieve data in Redis using various data structures like strings, lists, and sets. c) Implement basic Redis commands for data manipulation and retrieval		
3	Querying MongoDB a) Write and execute MongoDB queries to retrieve specific data from a collection.		
4	Redis Data Manipulation a) Use Redis commands to manipulate and modify data stored in different data structures. b) Retrieve specific data using Redis query operations		
5	Implementing Indexing in MongoDB a) Create an index on a specific field in a MongoDB collection. b) Measure the impact of indexing on query performance		
6	Data Storage in Redis a) Implement caching functionality using Redis as a cache store. b) Store and retrieve data from Redis cache using appropriate commands.		

7	CouchDB:
	    a) Installation and Working with Databases
	    b) Working on CouchDB Documents		

Course Code	Course Title	Credits	Lectures /Week
SIPCSCC621	**Big Data Analytics**	4	4

About the Course: This course provides the background of big data and cloud computing: their properties, characteristics, source of data, applications ,vector model, distributed programming using MapReduce, and database systems such as SQL, NoSQL and column-based for big data applications. It focuses more on hands-on experience with storage systems, big data processing on Hadoop platform, HDFS, YARN, Spark - Spark Streaming, Spark SQL.It also introduces public clouds deployment solutions for big data applications on clouds.

Course Outcomes:
● Understand the concept of Big data Analytics Technologies with cloud Computing implementation.
● Understand the concept of Cloud Infrastructures.
● Understand the implementation of programming Models , Big data Deployments & Virtualization .

Unit	Topics	No of Lectures
I	**Big Data Introduction**: Where does big data come from? what makes big data valuable, Characteristics of big data **Introduction to Cloud computing**: Definition and evolution of cloud computing. Enabling technologies, Service and Deployment models, Popular cloud stacks and use cases, Cloud computing-benefits, risks and challenges, Economic models and SLA's **Cloud Infrastructures:** Historical perspective of data centers, Datacenter components- IT equipment and Facilities, Design considerations- Requirements. Power, Efficiency and Redundancy, Power calculations- PUE and challenges in Cloud datacenters, Cloud management and Cloud Software Deployment considerations.	15L
II	**Cloud for Big data storage**: Introduction to storage systems, Cloud storage concepts, Distributed file systems(HDFS, Ceph FS), Cloud databases (HBase, MongoDB, Cassandra, DynamoDB), Cloud Object Storage(Amazon S3, OpenStack Swift. Ceph), NoSQL (Document-based , Key-value, Column-based, Graph-based) ,Vector Model.	15L
III	**Programming Models**: Distributed programming for the cloud, Data-Parallel analysis with Hadoop MapReduce(YARN); iterative, Data-Parallel iterative Analytics(Spark), Graph-Parallel Analytics with GraphLab 2.0(PowerGraph), Stream Processing(Samza) Spark SQL, Big data collection, Data lake, Spark streaming, Big data analytics, Spark Machine learning,	15L
IV	**Big data Deployment**: Hadoop solution , cloud solution Virtualization: Virtualization(CPU,Memory, I/O), Case Study: Amazon EC2, Software Defined Networks(SDN), Software Defined Storage(SDS)	

Course Code	Course Title	Credits	Lectures /Week
SIPCSCCP621	Big Data Analytics	2	2
1	Assignment Choice #1: Advantages of Cloud Computing		
2	Assignment Choice #2: Cloud-Based Organizations		
3	Assignment Choice #3: Facebook as a SaaS Solution		
4	Assignment Choice #4: IDaaS Service Model Presentation		
5	Assignment Choice #5: Streaming Video from the Cloud		
6	Assignment Choice #6: Working in the Cloud Presentation		
7	Assignment Choice #7: Utilizing a Web Service		
8	Assignment Choice #8: Cloud Based Disaster Recovery Plan		
9	Assignment Choice #9: Migrating an Application to the Cloud		
10	Assignment Choice #10: Evaluating Return on Investment (ROI)		

Course Code	Course Title	Credits	Lectures /Week
SIPCSCC622	**Data Analysis and Visualization**	4	4

About the Course:
This course provides the background of Data Analysis,Data Wrangling ,Tableau implementations, Concept of Storytelling through PowerBI .

Course Outcomes:
- Understand the concept of Data Analysis , Pipelines ,Data Wrangling.
- Understand and implement the visualisation techniques through the use of PowerBi and tableau
- Solve real world data analysis problems.

Unit	Topics	No of Lectures
I	**Introduction to Data Analysis:** Data Analysis - Exploratory Data Analysis and Data Science Process - Responsibilities of a Data Analyst - Data Analytics vs. Data Analysis - Types of Data - Understanding Different Types of File Formats - Sources of Data - Languages for Data Professionals - Overview of Data Repositories - Data Marts, Data Lakes, ETL, and Data Pipelines - Foundations of Big Data - Identifying Data for Analysis	15L
II	**Data Wrangling** :Data Sources - How to gather and Import Data - Data Loading, Storage and File Formats - Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, interacting with Web APIs, Interacting with Databases – Data Wrangling - Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting - Tools for Data Wrangling - Data Cleaning and Preparation - Handling Missing Data, Data Transformation, String Manipulation.	15L
III	**Data Visualization:** Intro to data visualization - Introduction to Visualization and Dashboarding Software - Visualization Tools - Getting started with Tableau Desktop – Connecting to the dataset - Creating charts – Creating common visualizations (bar charts, line charts etc.) - Filtering and sorting data - Adding Titles, Labels, and descriptions - Publish your work to Tableau Cloud - Interactivity with text and visual tooltips - Interactivity with actions (filter, highlight, URL) – Assembling dashboards from multiple charts.	15L
IV	**Story Telling:** Introduction to Power BI - Understanding Desktop - Understanding Power BI Report Designer - Report Canvas, Report Pages: Creation, Renames - Report Visuals, Fields and UI Options - Experimenting Visual Interactions, Advantages - Reports with Multiple Pages and Advantages - Pages with Multiple Visualizations - PUBLISH Options and Report Verification in Cloud - Adding Report Titles. Report Format Options - Introduction to data storytelling - Creating a data story.	15 L

Textbook(s):

1. Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython by McKinney, W., 2nd edition. O"Reilly Media, 2017
2. Doing Data Science: Straight Talk from the Frontline by O"Neil, C., & Schutt, R, O"Reilly Media, 2013
3. The Big Book of Dashboards by Steve Wexler, Jeffrey Shaffer, Andy Cotgreave, John Wiley & Sons, 2017
4. Practical Tableau by Ryan Sleeper, O"Reilly Media, 2018
5. Power BI. Book-1, Business Intelligence Clinic: Create and Learn by Roger F Silva, 2018
6. Introducing Microsoft Power BI by Alberto Ferrari and Marco Russo, Microsoft Press, Washington, 2016

Course Code	Course Title	Credits	Lectures /Week
SIPCSCCP622	**Data Analysis and Visualization (Power BI & Tableau)**	**2**	**2**
1	Implement Data Loading, Storage and File Formats. Read data and store them in text format.		
2	Implement the code to interact with Web APIs and to perform web scrapping.		
3	Demonstrate Data Cleaning and Preparation.		
4	Implement Data wrangling on a data set.		
5	Demonstrate the handling of missing data and string manipulation		
6	Create common charts with title, labels and descriptions using Tableau		
7	Perform sorting and filtering using tableau, create visualizations and publish it on Tableau Cloud.		
8	Perform data visualization using Power BI.		
9	Create reports using Power BI.		
10	Create a data story in Tableau or power BI.		

Course Code	Course Title	Credits	Lectures /Week
SIPCSEL621	SPARK Technologies	3	3

About the Course: This course provides the background and Basics of Spark Technologies , Resilient Distributed Datasets ,Working with Statistics , Spark SQL, GraphX.

Course outcomes:
- Understand the concept of SPARK Technologies and its implementation
- Understand the concept of RDD
- Understand the implementation of SPARK SQL, GraphX, Performance Tuning.
- Use Sparks Resilient Distributed Datasets to process and analyse large data sets across many CPUs Understand how the GraphX library helps with network analysis problems
- Understand how Spark SQL lets you work with structured data

Unit	Topics	No of Lectures
I	**Introduction to SPARK Technologies**- Components of the Spark unified stack, Features of Spark, Spark Web UI, an introduction to RDDs - Resilient Distributed Datasets, Launching and using Spark"s Scala and Python shell, Spark Context, Spark Ecosystem	15L
II	**In-Memory data** – Spark, Creating,Loading and Saving RDD, Transformations in RDD, Actions in RDD, Key-Value Pair RDD, Map Reduce and Pair RDD operations RDD Partitions	15L
III	**Implementation of SPARK Technologies** : Spark Applications vs. Spark Shell, Creating SparkContext, Building a Spark Application, Spark and Hadoop Integration-HDFS, Handling Sequence File, Spark RDD-RDD Lineage, RDD Persistence Overview, Distributed Persistence. Spark Streaming, ML library for Spark, Working with Statistics, SPARK SQL, GraphX, Performance Tuning.	15L

Textbooks:
1. Learning Spark: Lightning-Fast Data Analytics 2nd Edition, by Jules S. Damji, Brooke Wenig , Tathagata Das, Denny Lee, O"Reilly , 2020
2. Apache Spark Machine Learning Blueprints 1st Edition, Kindle Edition by Alex Liu, Packt Publishing, 2016
3. Apache Spark 2.x Cookbook: Cloud-ready recipes for analytics and data science 2nd Edition, by Rishi Yadav, Packt Publishing, 2017

Course Code	Course Title	Credits	Lectures /Week
SIPCSELP621	**SPARK Technologies**	1	1

	Basics Practicals
1	Spark Installation and Setup : Windows , Jupyter Notebook

2	Spark RDD -
	     a) Parallelize b)Read text file & Read CSV
	     c) Create RDD d) Actions e) Pair Functions

3	Spark RDD -
	     a) Repartition and Coalesce , Shuffle Partitions
	     b) Broadcast Variables , Accumulator Variables
	     c) Convert RDD to DataFrame

4)	SPARK Data Frames - createDataFrame() , where() & filter() , withColumn() ,withColumnRenamed() ,drop(), distinct() , groupBy() ,join() , map() vs mapPartitions(),foreach() vs foreachPartition() ,pivot() , union(),collect(),cache() & persist() ,udf()
3	Spark SQL- String Functions & Date and Timestamp Functions
4	Spark SQL- Array Functions & Map Functions
5	Spark SQL Sort Functions ,Aggregate Functions , Window Functions ,JSON Functions
	Advanced Practicals
6	Implement graph algorithms like PageRank or connected components using Spark GraphX.
7	Implement Clustering Techniques Using SPARK.
8	Build a system to detect anomalies in time-series data using Spark.

****

NEP -Second Year MSc 2024-25
Examination Pattern
Sr No	Semester	Course Category	Course Title	Course Code	Internal Evaluation Marks Out Of	Semester End Theory Marks Out Of	Semester End Practical Marks Out Of
MSc-II Semester III
1	III	Core Course-1	Social Network Analysis	SIPCSCC611/ SIPCSCCP611	40	60	50
2	III	Core Course-2	Simulation & Modelling	SIPCSCC612/ SIPCSCCP612	40	60	50
3	III	Discipline-specific Elective	NoSQL Technologies	SIPCSEL611/ SIPCSELP611	25	50	25
4	III	RP	Research Project	SIPCSRP611	50	Nil	100
	**MSc -II Semester IV**
6	IV	Core Course-1	Big Data Analytics	SIPCSCC621/ SIPCSCCP621	40	60	50
7	IV	Core Course-2	Data Analysis and Visualization	SIPCSCC622/ SIPCSCCP622	40	60	50
9	IV	Discipline-specific Elective	Spark Technologies	SIPCSEL621/ SIPCSELP621	25	50	25
10	IV	Research Project	Research Project	SIPCSRP621	50	Nil	100

****