Shiromani Gurudwara Prabandhak Committee's

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)

Matunga, Mumbai – 400 019, Maharashtra

Program: Master of Science

Syllabus

Course: MSc-Big Data Analytics

Semester III

(Major paper- I)

Name of the paper: NOSQL Databases

(As per NEP guidelines-DSC model)

With effect from Academic Year 2024 - 2025)

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)
Program Structure

Semester-III

Course Code	Course Name	Teaching Hours		Credits Assigned		
		Theory	Practical	Theory	Practical	Total
GNKPSBDAMJ 1503	(Major) Paper-I NOSQL Databases	60	60	4	2	6
GNKPSBDAMJ 2503	(Major) Paper-II Generative AI	60	60	4	2	6
GNKPSBDAEL 1503 GNKPSBDAEL 2503	(Elective) Paper Operating systems and visualization OR Analytics of Things	45	30	3	1	4
GNKPSBDARP 503	Research project (RP) Research paper writing + Research Project Proposal	--	--	--	--	4
Total		**165**	**150**	**11**	**05**	**20**

Semester-IV

Course Code	Course Name	Teaching Hours		Credits Assigned		
		Theory	Practical	Theory	Practical	Total
GNKPSBDAMJ 1504	(Major) Paper-I Cloud Computing	60	60	4	2	6
GNKPSBDAMJ 2504	(Major) Paper-II Streaming data analytics	60	60	4	2	6
GNKPSBDAEL 1504 GNKPSBDAEL 2504	(Elective) Paper Domain Specific Predictive analytics OR Text, Web and Social Media Analytics	45	30	3	1	4
GNKPSBDARP 504	Research project (RP) Research project Implementation	--	--	--	--	4
Total		**165**	**150**	**11**	**05**	**20**

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)
Department of Computer Science

Course: MSc-II Big Data Analytics
Semester-III Paper-I
Course Title:NOSQL Databases
Course Code: GNKPSBDAMJ1503
Credits: 4
No of lectures (Hours): 60
Marks: 100 (75:25)

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	Explore the origins of NoSQL databases and the characteristics that distinguish them from traditional relational database management systems.
2	Understand the architectures and common features of the main types of NoSQL databases (key-value stores, document databases, column-family stores, graph databases)
3	Discuss the criteria that decision makers should consider when choosing between relational and non-relational databases and techniques for selecting the NoSQL database that best addresses specific use cases.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Explain the detailed architecture, Database properties and storage requirements.	PO1,PO7	PS01	U
CO 2	Differentiate and identify right database models for real time applications	PO1,PO6	PS01	U
CO 3	Outline Key value architecture and characteristics	PO1,PO6	PS01	U
CO 4	Design Schema and implement CRUD operations, distributed data operations	PO1,PO3	PS01, PSO7	U,Ap
CO 5	Compare data warehousing schemas and implement various column store internals	PO1,PO8	PS01	U
CO 6	Choose and implement Advanced columnar data model functions for the real time applications	PO1,PO2	PS05	U

CO 7	Develop Application with Graph Data model.	PO1,PO2	PS02,PS03	Ap,An

Unit		**Title**	**No. of lectures**	**CO Mapping**
Unit 1			**15**	
	1.1	NoSQL Data model: Aggregate Models- Document Data Model- Key-Value Data Model- Columnar Data Model, Graph Based Data Model Graph Data Model		**CO 1**
	1.2	NoSQL system ways to handle big data problems, Moving Queries to data, not data to the query, hash rings to distribute the data on clusters		**CO 2**
	1.3	Replication to scale reads, Database distributed queries to Data nodes.		**CO 3**
Unit 2			**15**	
	2.1	From array to key value databases, Essential features of key value Databases, Properties of keys, Characteristics of Values, Key-Value Database Data Modeling Terms,		**CO 4**
	2.2	Key-Value Architecture and implementation Terms, Designing Structured Values, Limitations of Key- Value Databases		**CO 5**
	2.3	Design Patterns for Key-Value Databases, Case Study: Key-Value Databases for Mobile Application Configuration		**CO 6**
Unit 3			**15**	
	3.1	Document, Collection, Naming, CRUD operation, querying, indexing, Replication, Sharding		**CO 7**
	3.2	Consistency Implementation: Distributed consistency, Eventual Consistency, Capped Collection		**CO 8**
	3.3	Case studies: document oriented database: Mongo DB and/or Cassandra		**C0 9**
Unit 4			**15**	
	4.1	Data warehousing schemas: Comparison of columnar and row-oriented storage, Column-store Architectures: C-Store and Vector-Wise, Column-store internals and, Inserts/updates/deletes		**CO 10**
	4.2	Indexing, Adaptive Indexing and Database Cracking Advanced techniques: Vectorized Processing, Compression, Write penalty, Operating Directly on Compressed Data Late Materialization Joins , Group-by, Aggregation and Arithmetic Operations, Case Studies		**CO 11**
	4.3	Property Graph Model Graph Analytics: Link analysis algorithm- Web as a graph, Page Rank- Markov chain, page rank computation, Topic specific page rank (Page Ranking Computation techniques: iterative processing,		**C0 12**

		Random walk distribution Querying Graphs: Introduction to Cypher, case study: Building a Graph Database Application- community detection		

References:

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

Examination:

- **Internal Examination (25 Marks):** 20 Marks exam (Presentation). And 5 Marks for Class Participation etc.
- **End Semester theory examination (75 Marks):** Weightage of each unit will be proportional to the number of lecture hours as mentioned in the syllabus. Duration of exam: 2hours 30mins
- **Combined passing of 40% with minimum 20% in Internal Component.**

Course: MSc Big Data Analytics Practical
Semester-III
Course Title: NOSQL Databases Practical Paper-I
Course Code: GNKPSBDAMJ1P503
Credits: 02
No of Practical (Hours): 60
Marks: 50

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	Explore the origins of NoSQL databases and the characteristics that distinguish them from traditional relational database management systems.
2	Understand the architectures and common features of the main types of NoSQL databases (key-value stores, document databases, column-family stores, graph databases)
3	Discuss the criteria that decision makers should consider when choosing between relational and non-relational databases and techniques for selecting the NoSQL database that best addresses specific use cases.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Explain the detailed architecture, Database properties and storage requirements.	PO1, PO7	PS01	U
CO 2	Differentiate and identify right database models for real time applications	PO1, PO6	PS01	U
CO 3	Outline Key value architecture and characteristics	PO1, PO6	PS01	U
CO 4	Design Schema and implement CRUD operations, distributed data operations	PO1, PO3	PS01, PSO7	U,Ap
CO 5	Compare data warehousing schemas and implement various column store internals	PO1, PO8	PS01	U
CO 6	Choose and implement Advanced columnar data model functions for the real time applications	PO1, PO2	PS05	U
CO 7	Develop Application with Graph Data model	PO1, PO2	PSO5	Ap,An

List of Experiments:

1. Import the Hubway data into Neo4j and configure Neo4j.Then, answer the following questions using the Cypher Query Language:
    a. List top 10 stations with most outbound trips (Show station name and number of trips)
    b. List top10 stations with most in bound trips (Show station name and number of trips)
    c. List top 5 routes with most trips (Show starting station name, ending station name and number of trips)
    d. List the hour number(forexample13means1pm-2pm) and number of trips which start 5 hours from the station ”B.U.Central”
    e. List the hour number(for example 13 means 1pm-2pm) and number of trips which end at the station ”B.U. Central”
2. Download a zip code dataset at http://media.mongodb.org/zips.json .Use mongo import to import the zip code dataset into MongoDB. After importing the data, answer the following questions by using aggregation pipelines:
    a. Find all the states that have a city called ”BOSTON”. Find all the states and cities whose names include the string ”BOST”. Each city has several zip codes.
    b. Find the city in each state with the most number of zip codes and rank those cities along with the states using the city populations. Mongo DB can query on spatial information.
3. Create a database that stores road cars. Cars have a manufacturer, a type. Each car has a maximum performance and a maximum torque value. Do the following: Test Cassandra's replication schema and Consistency models.
4. Master Data Management using Neo4j Manage your master data more effectively The world of master data is changing. Data architects and application developers are swapping their relational

databases with graph databases to store their master data. This switch enables them to use a data store optimized to discover new insights in existing data,provide a 360-degree view of master data and answer questions about data relationships in real time.

5. Shopping Mall case study using cassandra, where we have many customers ordering items from thermal land and we have suppliers who deliver them their ordered items.
6. provide a code snippet demonstrating how you might model data in Apache Cassandra and MongoDB for a social media platform scenario, along with some example queries to illustrate their differences.
7. Demonstrate the concept of dynamically updating indexing structures based on query patterns.
8. Provide code snippet demonstrating vectorized processing using the NumPy library, which provides efficient array operations leveraging SIMD instructions.
9. Provide code demonstrating a simplified version of PageRank computation using random walk simulation on a directed graph represented as an adjacency matrix.
10. Provide Python code demonstrating how the Louvain algorithm can be applied using Neo4j's Graph Data Science library.

References:

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

Examination (Total Marks): 50 (Two practical's of 25 marks each)
Experiment Marks: 15
Journal Marks: 05
Viva Marks: 05

Shiromani Gurudwara Prabandhak Committee's

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)

Matunga, Mumbai – 400 019, Maharashtra

Program: Master of Science

Syllabus

Course: MSc-Big Data Analytics

Semester III

(Major paper- II)

Name of the paper: Generative AI

(As per NEP guidelines-DSC model)

With effect from Academic Year 2024 - 2025)

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)
Department of Computer Science

Course: MSc-II Big Data Analytics
Semester-III Paper-II
Course Title: Generative AI
Course Code: GNKPSBDAMJ2503
Credits: 4
No of lectures (Hours): 60
Marks: 100 (75:25)

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	Understanding Generative AI: ● Define and differentiate between generative models, including probabilistic models, autoencoders, VAEs, and GANs. ● Explain the fundamental principles behind generative modeling techniques. ● Identify real-world applications where generative AI can be applied effectively.
2	Applications of Generative AI: ● Identify and describe specific use cases for generative AI across various domains, such as image generation, text generation, and audio generation. ● Analyze case studies to understand how generative AI techniques have been applied in different industries.
3	Proficiency in Autoencoders: ● Understand the architecture and functionality of autoencoders, including VAEs. ● Implement basic autoencoder models and interpret their output. ● Explore practical applications of autoencoders in generative tasks.
4	Mastery of GANs: ● Describe the architecture and training process of GANs. ● Compare and contrast different architectural variations of GANs. ● Demonstrate proficiency in applying GANs for tasks such as image generation and text generation.
5	Text Generation Techniques: ● Explain the concepts of RNNs, LSTMs, and transformer models for text generation. ● Implement text generation models using the mentioned techniques. ● Evaluate the performance of text generation models using appropriate metrics.
6	Image Generation Techniques: ● Understand techniques such as conditional image generation, style transfer, and image-to-image translation.

	● Implement image generation models using these techniques. ● Assess the quality of generated images using evaluation metrics.
7	Audio Generation: ● Describe waveform generation with neural networks and music generation using recurrent models. ● Implement audio generation models. ● Evaluate the quality of generated audio using appropriate metrics.
8	Evaluation and Metrics: ● Define evaluation criteria and metrics for assessing the quality of generated content. ● Implement evaluation methods to measure the performance of generative AI models. ● Interpret evaluation results and make recommendations for model improvement.
9	Ethical Considerations: ● Identify ethical considerations and biases associated with generative AI. ● Critically analyze the implications of these considerations in real-world applications. ● Propose strategies for mitigating ethical concerns and biases in generative AI development and deployment.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Understanding Generative AI: Students will gain a comprehensive understanding of generative models, including probabilistic models, autoencoders, variational autoencoders (VAEs), and Generative Adversarial Networks (GANs).	PO1, PO2, PO3, PO6	PS01	U, Ap, An
CO 2	Applications of Generative AI: Students will explore various applications of generative AI across different domains, including image generation, text generation, audio generation, and more.	PO2, PO3, PO4, PO5, PO6, PO7, PO8	PS01	Ap, An, E
CO 3	Proficiency in Autoencoders: Students will grasp the basics of autoencoders, including variational autoencoders (VAEs), and understand their applications in generative AI.	PO2, PO3, PO5, PO6	PS01	U, Ap
CO 4	Mastery of GANs: Students will be proficient in understanding GANs, their training process, architectural variations, and applications in image generation, text generation, and other areas.	PO1, PO2, PO3, PO4, PO5, PO6, PO7, PO8	PS01	U, Ap, An, E E

CO 5	Text Generation Techniques: Students will learn about recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) networks, transformer models, and their applications in text generation.	PO1, PO2, PO3, PO6, PO7	PS01	U, Ap, An
CO 6	Image Generation Techniques: Students will understand techniques for conditional image generation, style transfer, image-to-image translation, and evaluation metrics for generated content quality.	PO1, PO2, PO3, PO6, PO7, PO8	PS01	U, Ap, An E
CO 7	Audio Generation: Students will explore waveform generation with neural networks, music generation using recurrent models, speech synthesis techniques, and evaluation metrics for audio generation.	PO1, PO2, PO3, PO6, PO7	PS01	U, Ap, An E
CO 8	Evaluation and Metrics: Students will be equipped with evaluation criteria and metrics for assessing the quality of generated content, including attention mechanisms and meta-learning for generative AI.Ethical Considerations: Students will critically analyze ethical considerations and biases in generative AI, understanding their implications and potential consequences in real-world applications.	PO1, PO2, PO3, PO6, PO7, PO8	PS01	U, Ap, An, E

Unit		**Title**	**No. of lectures**	**CO Mapping**
Unit 1			**15**	
	1.1	Introduction to Generative AI Overview of generative models, Applications of generative AI, Probabilistic Models		**CO 1**
	1.2	Overview of generative models Applications of generative AI, Probabilistic Models		**CO 2**
	1.3	Basics of autoencoders, Variational autoencoders (VAEs) Applications of autoencoders in generative AI Generative Adversarial Networks (GANs)		**CO 3**
Unit 2			**15**	
	2.1	Introduction to GANs Training GANs, Architectural variations of GANs, Applications of GANs in image generation, Text Generation		**CO 4**
	2.2	Recurrent Neural Networks (RNNs) for text generation Long Short-Term Memory (LSTM) networks		**CO 5**
	2.3	Transformer models for text generation Image Generation		**CO 6**

Unit 3			**15**	
	3.1	Conditional image generation, Style transfer techniques Image-to-image translation, Audio Generation		**CO 7**
	3.2	Waveform generation with neural networks Music generation using recurrent models		**CO 8**
	3.3	Speech synthesis techniques Evaluation and Metrics		**C0 9**
Unit 4			**15**	
	4.1	Evaluation criteria for generative AI models Metrics for assessing generated content quality		**CO 10**
	4.2	Attention mechanisms in generative models		**CO 11**
	4.3	Meta-learning for generative AI Ethical considerations and biases in generative AI		**C0 12**

References:

1. Artificial Intelligence & Generative AI for Beginners by David M. Patel.
2. Generative Deep Learning, 2nd Edition, O'Reilly Media, Inc.
3. AI Made Simple, A Beginner's Guide to Generative Intelligence, By Rajeev Kapur by David Foster

Examination:

- **Internal Examination (25 Marks):** 20 Marks exam (Presentation). And 5 Marks for Class Participation etc.
- **End Semester theory examination (75 Marks):** Weightage of each unit will be proportional to the number of lecture hours as mentioned in the syllabus. Duration of exam: 2hours 30mins
- **Combined passing of 40% with minimum 20% in Internal Component.**

Course: MSc Big Data Analytics Practical
Semester-III
Course Title: Generative AI Practical Paper-II
Course Code: GNKPSBDAMJ2P503
Credits: 02
No of Practical (Hours): 60
Marks: 50

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	Understand the architecture and principles of autoencoder networks.
2	Implement GANs for generating realistic images and data samples.
3	Apply Long Short-Term Memory (LSTM) networks to overcome the vanishing gradient problem in RNNs.
4	Apply image-to-image translation models for tasks like colorization and semantic segmentation.
5	Audio Generation: ● Evaluate the quality of generated audio using objective metrics and subjective listening tests.
6	Music Generation: ● Explore techniques such as reinforcement learning for generating diverse and coherent music sequences.
7	Speech Synthesis: ● Apply linguistic features and prosody modeling for natural-sounding speech synthesis.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Understand the architecture and functioning of autoencoders.	PO1,PO2	PS01	U
CO 2	Implement GANs for generating realistic images and data samples.	PO2,PO6	PS06	U,Ap
CO 3	Apply Long Short-Term Memory (LSTM) networks to overcome the vanishing gradient problem in RNNs.	PO1 ,PO2,PO3	PSO5	Ap
CO 4	Apply image-to-image translation models for tasks like colorization and semantic segmentation.	PO1 ,PO2,PO6	PSO6	Ap
CO 5	Evaluate the quality of generated audio using objective metrics and subjective listening tests.	PO2,PO7, PO8	PSO7	E
CO 6	Explore techniques such as reinforcement learning for generating diverse and coherent music sequences.	PO2,PO6, PO8	PSO6	U,Ap
CO 7	Apply linguistic features and prosody modeling for natural-sounding speech synthesis.	PO1 ,PO2,PO6, PO7	PSO7	Ap

List of Experiments:

Implement Case study /Practicals on

1. Autoencoders

2. Generative Adversarial Networks (GANs)
3. Recurrent Neural Networks (RNNs) for text generation
4. Long Short-Term Memory (LSTM)
5. Image generation
6. Audio generation
7. Music generation
8. speech synthesis

References:

1. Artificial Intelligence & Generative AI for Beginners by David M. Patel.
2. Generative Deep Learning, 2nd Edition, O'Reilly Media, Inc.
3. AI Made Simple, A Beginner's Guide to Generative Intelligence, By Rajeev Kapur by David Foster

Examination (Total Marks): 50 (Two practical's of 25 marks each)
Experiment Marks: 15
Journal Marks: 05
Viva Marks: 05

Shiromani Gurudwara Prabandhak Committee's

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)

Matunga, Mumbai – 400 019, Maharashtra

Program: Master of Science

Syllabus

Course: MSc-Big Data Analytics

Semester III

(Elective paper- I)

Name of the paper: Operating Systems And Virtualization

(As per NEP guidelines-DSC model)

With effect from Academic Year 2024 - 2025)

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)
Department of Computer Science

Course: MSc-II Big Data Analytics
Semester-III Elective Paper
Course Title: Operating Systems And Virtualization
Course Code: GNKPSBDAEL1503
Credits: 3
No of lectures (Hours): 45
Marks: 100 (75:25)

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	To introduce Virtualization, operating systems fundamental concepts and its technologies
2	To provide skills to write programs that interact with operating system components such as processes, thread, memory during concurrent execution.
3	To provide the skills and knowledge necessary to implement, provisioning and administer server and desktop virtualization.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Study operating system layers and kernel architectures.	PO3	PSO3	An, E
CO 2	Design various techniques for process management.			
CO 3	Construct various address translation mechanism.	PO3	PSO3	Ap ,An
CO 4	Perform process threading and synchronization.	PO3	PSO3	Ap ,An
CO 5	Study various methods of virtualization and perform desktop and server virtualization.	PO3, PO4	PSO3	U, Ap
CO 6	Classify the light-weight virtual machines with dockers and containers.	PO3	PSO4	U, Ap
CO 7	Develop programs related to the simulations of operating systems and virtualization concepts.	PO3	PSO3	Ap, C

Unit		Title	No. of lectures	CO Mapping
Unit 1			**15**	
	1.1	**History of OS** - Computer system architecture a layered view with interfaces, Glenford Myer		**CO 1**
	1.2	Monolithic Linux Hybrid Windows 10 kernels Layered architecture of operating system and core functionalists.		**CO 2**
	1.3	Introduction, Process Operations, States, Context switching, Data Structures (Process Control Block (PCB)		**CO 3**
Unit 2			**15**	
	2.1	**Process Scheduling:** Multi-Level Feedback Queue, Multi-processor Scheduling, Deadlocks and its detection.		**CO 4**
	2.2	Introduction, Thread Models, Thread API, Building Evaluating a Lock, Test And Set, Classical problems handling using semaphore, Monitors		**CO 5**
	2.3	**Persistence - File Organization:** The i-node, Crash Consistency file security.		**CO 6**
Unit 3			**15**	
	3.1	**VIRTUAL MACHINES :** Process and System VMs Taxonomy of VMs.		**CO 7**
	3.2	Hardware Emulation, Full Virtualization with binary translation, Hardware assisted, Operating System Virtualization, OS assisted /Para virtualization.		**CO 8**
	3.3	**HYPERVISOR :** Type 1, Type 2, Para-virtualization, Server Virtualization, Desktop Virtualization, Overview VM portability - Clones, Templates, Snapshots, OVF, Hot and Cold Cloning Protecting Increasing Availability, Light Weight Virtual machine: Container / Docker.		**C0 9**

References:

1. Silberschatz, Abraham, Greg Gagne, and Peter B. Galvin, “Operating system concepts”, 10th Edition, Wiley Publishers, 2018.
2. Matthew Portnoy, “Virtualization Essentials”, John Wiley Sons Inc; 2ndEdition Edition, 2016.
3. Thomas Anderson, Michael Dahlin, “Operating Systems: Principles and Practice”, 2nd Edition, Recursive Books, 2014.

4. William Stallings, “Operating Systems: Internals and Design Principles”, 8th Edition, 2014. Smith, Nair, “Virtual Machines: Versatile Platforms for Systems and Processes”, 1st Edition, Morgan Kaufmann Publishers, 2005.

Examination:

- **Internal Examination (25 Marks):** 20 Marks exam (Presentation). And 5 Marks for Class Participation etc.
- **End Semester theory examination (75 Marks):** Weightage of each unit will be proportional to the number of lecture hours as mentioned in the syllabus. Duration of exam: 2hours 30mins
- **Combined passing of 40% with minimum 20% in Internal Component.**

Course: MSc-II Big Data Analytics Practical
Semester-III
Course Title: OPERATING SYSTEMS AND VIRTUALIZATION Practical
Course Code: GNKPSBDAEL1P503
Credits: 01
No of Practical (Hours): 30
Marks: 50

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	Understand the fundamentals of Unix/Linux operating systems and command-line interfaces, including file management, process handling, and system administration tasks, through hands-on practice and demonstrations
2	Develop proficiency in shell programming concepts, including input/output redirection, decision-making structures, looping mechanisms, and multi-level branching, by writing and executing scripts to automate tasks and solve problems.
3	Explore process management in Unix-like systems, including the creation of child processes using the fork() system call, understanding the concepts of orphan and zombie processes, and implementing process control mechanisms to ensure efficient system resource utilization.
4	Gain practical experience in simulating various CPU scheduling algorithms, such as First Come First Serve (FCFS), Shortest Job First (SJF), Priority Scheduling, and Round Robin, and analyze their performance characteristics through experimentation and performance metrics evaluation.
5	Develop a deep understanding of resource allocation and deadlock avoidance strategies in operating systems by implementing and analyzing the Banker's algorithm, including verifying system safety and determining the feasibility of granting additional resource requests, thereby gaining insights into real-world challenges in OS design and resource management.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Students will be able to demonstrate proficiency in using basic Linux commands for file manipulation, directory navigation, process management, and system administration tasks.	P01	PSO 3	U
CO 2	Students will gain a comprehensive understanding of shell programming, including input/output redirection, decision-making constructs (such as if-else statements), looping structures (such as for and while loops), and multi-level branching (nested conditions and loops).	P01	PSO 3	Ap,An
CO 3	Students will develop the ability to create and manage processes in a Unix-like environment, including understanding the concepts of parent-child relationships, orphan processes, and zombie processes through practical implementation using the fork() system call.	P02	PSO 3	U,An
CO 4	Students will be able to simulate various CPU scheduling algorithms (such as First Come First Serve, Shortest Job First, Priority Scheduling, and Round Robin) and analyze their performance characteristics in terms of throughput, turnaround time, waiting time, and response time.	PO2	PSO 3	U,An
CO 5	Students will gain proficiency in implementing and analyzing the Banker's algorithm for resource allocation in an operating system, including checking system safety and determining the immediate grantability of additional resource requests, fostering a deeper understanding of resource allocation and deadlock avoidance strategies in OS design.	PO3	PSO 3	U,An

List of Experiments:

1. Study of Basic Linux Commands.
2. Shell Programming (I/O, Decision making, Looping, Multi-level branching).
3. Crating child process using fork() system call, Orphan and Zombie process creation.
4. Simulation of CPU scheduling algorithms (FCFS, SJF, Priority and Round Robin).
5. Simulation of Bankers algorithm to check weather given system is in safe state or not. Also check whether addition resource requested can be granted immediately.
6. Parallel Thread management using pthread library. Implement a data parallelism using multi-threading.
7. Dynamic memory allocation algorithms - first-fit, best-fit, worst-fit algorithms.

8. Page Replacement Algorithms FIFO, LRU and Optimal.
9. Virtualization Setup: Type-1, Type-2 Hypervisor.
10. Implementation of OS / Server Virtualization.

References:

1.Silberschatz, Abraham, Greg Gagne, and Peter B. Galvin, "Operating system concepts", 10th Edition, Wiley Publishers, 2018.

2.Matthew Portnoy, "Virtualization Essentials", John Wiley Sons Inc; 2ndEdition Edition, 2016.

Thomas Anderson, Michael Dahlin, "Operating Systems: Principles and Practice", 2nd Edition, Recursive Books, 2014.

3.William Stallings, "Operating Systems: Internals and Design Principles", 8th Edition, 2014. Smith, Nair, "Virtual Machines: Versatile Platforms for Systems and Processes", 1st Edition, Morgan Kaufmann Publishers, 2005.

Examination (Total Marks): 50
Experiment Marks: 40
Journal Marks: 05
Viva Marks: 05

Shiromani Gurudwara Prabandhak Committee's

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)

Matunga, Mumbai – 400 019, Maharashtra

Program: Master of Science

Syllabus

Course: MSc-Big Data Analytics

Semester III

(Elective paper- II)

Name of the paper: Analytics of Things

(As per NEP guidelines-DSC model)

With effect from Academic Year 2024 - 2025)

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)
Department of Computer Science

Course: MSc-II Big Data Analytics
Semester-III Elective Paper
Course Title: Analytics of Things
Course Code: GNKPSBDAEL2503
Credits: 3
No of lectures (Hours): 45
Marks: 100 (75:25)

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	To introduce the technology that enables IoT, application of IoT, cloud support for IoT and access data using mobile computing devices. This will serve as the foundation for the cyber physical systems, Internet of services leading to Industry 4.0 changes.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Identify the technologies that enables IoT.	PO1	-	U,Ap,An
CO 2	Able to use Hardware and software required to design and build IoT	PO1,PO6	PS01	U,Ap
CO 3	Develop programs for interfacing with sensors and actuators and other IoT devices	PO2,PO6	PSO2	U,Ap,An
CO 4	Set up the servers to upload IoT data to cloud for further analysis	PO1,PO3	PS04	U,Ap,An

Unit		Title	No. of lectures	CO Mapping
Unit 1			**15**	
	1.1	Introduction to IoT Introduction, Characteristics of IoT, Difference between IoT and M2M, Applications of IoT		**CO 1**
	1.2	Physical and logical design of IoT, IoT levels and deployment templates		**CO 2**
	1.3	IoT enabling technologies: Wireless Sensor Networks, RFID, GPS		**CO 3**

Unit 2			**15**	
	2.1	IOT Hardware platforms Overview of IoT supported Hardware Platforms: Raspberry pi, Arduino, Intel Galileo		**CO 4**
	2.2	Communication in IOT Interface protocol, Serial, SPI, I2C, 6LoWPAN, 802.11wifi, 802.15 Bluetooth, 802.15.4 Zigbee, RTLS, GPS, CoAp – Constrained application protocol, RPL – routing protocol for lossy networks.		**CO 5**
	2.3	IOT Software development Linux, Networking configurations in Linux, Accessing Hardware & Device Files interactions, Python packages: JSON, XML, HTTPLib, URLLib, SMTPLib, XMPP, Contiki OS		**CO 6**
Unit 3			**15**	
	3.1	IoT Physical Servers & Cloud Offerings Introduction to Cloud Storage Models & Communication APIs, Cloud of things, Xively Cloud for IOT, PHP & MySQL for data processing, WAMP, Designing a RESTful Web API, MQTT, Amazon Web Services for IoT		**CO 7**
	3.2	Data Analytics for IoT Configuring and using Apache Storm for Real-time Data Analysis		**CO 8**
	3.3	Case Studies illustrating IoT Design Smart Home, Smart Parking, weather reporting and monitoring		**C0 9**

References:

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

Examination:

- **Internal Examination (25 Marks):** 20 Marks exam (Presentation). And 5 Marks for Class Participation etc.

- **End Semester theory examination (75 Marks):** Weightage of each unit will be proportional to the number of lecture hours as mentioned in the syllabus. Duration of exam: 2hours 30mins
- **Combined passing of 40% with minimum 20% in Internal Component.**

Course: MSc-II Big Data Analytics Practical
Semester-III
Course Title: Analytics of Things Practical
Course Code: GNKPSBDAEL2P503
Credits: 01
No of Practical (Hours): 30
Marks: 50

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	Understand basic hardware interaction and programming concepts: Students will be able to set up a Raspberry Pi or Arduino board and write a program to blink an LED, demonstrating fundamental hardware interaction and programming skills.
2	Learn about Wireless Sensor Networks (WSNs) and Zigbee communication: Students will configure a WSN using Zigbee modules to collect temperature data from multiple sensor nodes and transmit it to a central coordinator node, gaining hands-on experience with WSN technologies.
3	Develop proficiency in sensor data acquisition and cloud communication: Students will develop Python scripts to read data from sensors connected to IoT devices and publish this data to a cloud server using the MQTT protocol, learning practical skills in data acquisition and cloud integration.Acquire knowledge of web development and API design: Students will design and implement a RESTful Web API using Python Flask or Django framework to interact with IoT devices, allowing users to query sensor data or control actuators remotely, enhancing their understanding of web development and API design principles.
4	Gain expertise in real-time data processing and analytics: Students will implement an Apache Storm pipeline to process incoming sensor data streams and calculate simple statistics, enabling them to gain proficiency in real-time data processing and analytics techniques, essential for IoT applications.
5	Acquire knowledge of web development and API design: Students will design and implement a RESTful Web API using Python Flask or Django framework to interact with IoT devices, allowing users to query sensor data or control actuators remotely, enhancing their understanding of web development and API design principles.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed

CO 1	Demonstrate proficiency in basic hardware interaction and programming with Raspberry Pi or Arduino boards. ● This CO corresponds to the task of setting up a Raspberry Pi or Arduino board and writing a program to blink an LED, demonstrating basic hardware interaction.	PO1,PO2	PSO3	U,Ap,An
CO 2	Design and implement a Wireless Sensor Network (WSN) using Zigbee modules for data collection and transmission. ● This CO aligns with the task of configuring a WSN using Zigbee modules to collect temperature data from multiple sensor nodes and transmit it to a central coordinator node.	PO1,PO2, PO3,PO5	PSO5	Ap,An
CO 3	Develop skills in data acquisition and communication by creating Python scripts to read sensor data and publish it to a cloud server using MQTT protocol. ● This CO relates to the task of developing a Python script to read data from a sensor connected to a Raspberry Pi or Arduino board and publish it to a cloud server using MQTT protocol.	PO2, PO3,PO7	PSO4	Ap,An,E
CO 4	Design and implement a RESTful Web API using Python Flask or Django for interacting with IoT devices. ● This CO corresponds to the task of designing a RESTful Web API using Python Flask or Django framework to interact with IoT devices, allowing users to query sensor data or control actuators remotely.	PO4,PO5, PO7	PS01	Ap,An
CO 5	Gain proficiency in real-time data processing and analytics by implementing an Apache Storm pipeline for processing sensor data streams. ● This CO aligns with the task of implementing a real-time data analytics pipeline using Apache Storm to process incoming sensor data streams and calculate simple statistics.	PO2, PO3,PO6	PS05, PS07	Ap,An,E

List of Experiments:

1. Set up a Raspberry Pi or Arduino board and write a simple program to blink an LED connected to it, demonstrating basic hardware interaction.
2. Configure a Wireless Sensor Network (WSN) using Zigbee modules and collect temperature data from multiple sensor nodes, then transmit this data to a central coordinator node.
3. Develop a Python script to read data from a sensor (e.g., temperature sensor or motion sensor) connected to your Raspberry Pi or Arduino board and publish this data to a cloud server using MQTT protocol.
4. Design a RESTful Web API using Python Flask or Django framework to interact with IoT devices, allowing users to query sensor data or control actuators remotely.
5. Implement a real-time data analytics pipeline using Apache Storm to process incoming sensor data streams and calculate simple statistics (e.g., average temperature over time).
6. Set up an MQTT broker (e.g., Mosquitto) on a Raspberry Pi or a cloud server and configure IoT devices to publish and subscribe to sensor data using MQTT protocol.
7. Deploy a simple Smart Home system using IoT devices such as smart plugs, sensors, and actuators, and develop a web interface to monitor and control these devices remotely.
8. Create a data visualization dashboard using tools like Grafana or Plotly to display real-time sensor data collected from IoT devices.
9. Configure a GPS module to track the location of a mobile IoT device (e.g., Raspberry Pi with GPS module) and visualize its movement on a map in real-time.
10. Develop a prototype of a weather reporting and monitoring system using IoT sensors to collect weather data (e.g., temperature, humidity, air pressure) and store this data in a cloud database for further analysis.

References:

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

Examination (Total Marks): 50

Experiment Marks: 40

Journal Marks: 05

Viva Marks: 05

Shiromani Gurudwara Prabandhak Committee's

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)

Matunga, Mumbai – 400 019, Maharashtra

Program: Master of Science

Syllabus

Course: MSc-Big Data Analytics

Semester III

(Research Project)

Name of the paper: Research Paper Writing + Research Project Proposal

(As per NEP guidelines-DSC model)

With effect from Academic Year 2024 - 2025)

Course: MSc-II Big Data Analytics
Semester-III
Course Title: Research Paper Writing
Course Code: GNKPSBDARP503
Credits: 02
No of Lecture (Hours): 30
Marks: 100

Course Objectives:

Sr. No.	Course objectives
The course *aims at:*	
1	Understanding Research Paper Structure: To familiarize students with the structure and components of a research paper in computer science, enabling them to identify and understand the purpose of each section.
2	Literature Review Skills: To develop students' ability to conduct thorough literature reviews, critically evaluate existing research, and identify gaps in the literature that can form the basis for their own research.
3	Research Methodology Proficiency: To provide students with the knowledge and skills necessary to design and conduct research studies in computer science, including selecting appropriate methodologies, designing experiments, and addressing ethical considerations.
4	Data Analysis and Presentation: To equip students with the skills to analyze research data using appropriate techniques and present their findings effectively using tables, charts, graphs, and other visualization methods.
5	Critical Thinking and Analytical Skills: To foster students' critical thinking skills, enabling them to evaluate research findings, identify strengths and weaknesses, and propose meaningful interpretations and conclusions.

6	Effective Academic Writing: To enhance students' academic writing skills, including grammar, punctuation, clarity, and coherence, enabling them to communicate their ideas and arguments effectively in academic prose.
7	Publication and Presentation Skills: To familiarize students with the academic publishing process, including selecting suitable venues for publication, writing abstracts and manuscripts, and presenting research findings at conferences.
8	Peer Review and Feedback: To engage students in peer review activities, providing them with opportunities to receive constructive feedback on their work, develop their ability to provide feedback to others, and refine their research papers based on peer input.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Students will demonstrate an understanding of the structure and components of a research paper in the field of computer science, including the introduction, literature review, methodology, results, discussion, and conclusion sections.	PO1,PO2	PSO3	Ap,An
CO 2	Students will be able to conduct comprehensive literature reviews using academic databases and search engines, critically evaluate the credibility and relevance of sources, and synthesize existing literature to identify research gaps.	PO2,PO3	PSO3	U,An

CO 3	Students will acquire proficiency in research methodology techniques specific to computer science, including experimental design, data collection, ethical considerations, and writing the methodology section of a research paper.	PO3,PO5	PSO3	Ap,An,C
CO 4	Students will develop skills in data analysis techniques relevant to computer science research, including statistical analysis and visualization methods, and effectively present research findings using tables, charts, and graphs.	PO2,PO5 , PO6	PSO3	Ap,An,C
CO 5	Students will cultivate critical thinking and analytical skills necessary for evaluating research findings, identifying limitations, proposing future research directions, and engaging in scholarly discourse.	PO6,PO7 , PO8	PSO3	Ap,An,C
CO 6	Students will enhance their academic writing skills, including writing style, grammar, punctuation, and sentence structure, and learn to effectively communicate their ideas and arguments in clear, concise, and coherent academic prose.	PO1,PO2 , PO6	PSO3	Ap,An,C
CO 7	Students will gain knowledge of the academic publishing process, including selecting appropriate journals and conferences for submission, writing abstracts and manuscripts, and presenting research findings effectively at academic conferences.	PO7, PO8	PSO3	Ap,An,C

CO 8	Students will engage in peer review activities to provide constructive feedback on their classmates' research papers, learn from their peers' perspectives, and refine their own research papers based on peer feedback.	PO5, PO8	PSO3	Ap,An,C

Week wise schedule	**Title**	**No. of lectures**	**CO Mapping**
Week 1	● Introduction to Research Paper Writing ● Overview of academic writing in computer science ● Understanding the purpose and significance of research papers ● Types of research papers and their structures ● Identifying research topics and formulating research questions.	4	CO 1

Week 2	• Literature Review • Searching for relevant literature using academic databases and search engines • Evaluating the credibility and relevance of sources • Synthesizing literature and identifying research gaps • Proper citation and referencing techniques	4	CO 2
Week 3	• Research Methodology • Introduction to research methodologies in computer science • Designing experiments and collecting data • Ethical considerations in research • Writing the methodology section of a research paper	4	CO 3
Week 4	• Data Analysis and Results • Techniques for data analysis in computer science research • Presenting data using tables, charts, and graphs • Interpreting results and drawing conclusions • Writing the results section of a research paper	4	CO 4
Week 5	• Discussion and Conclusion • Analyzing and discussing research findings • Addressing limitations and future research directions • Writing the discussion and conclusion sections of a research paper	4	CO 5

	● Strategies for effective argumentation and persuasion		
Week 6	● Academic Writing Skills ● Writing style and tone in academic writing ● Grammar, punctuation, and sentence structure ● Editing and revising drafts ● Peer review and feedback	4	CO 6
Week 7	● Publishing and Presentation ● Understanding the academic publishing process ● Selecting appropriate journals and conferences for submission ● Writing abstracts and preparing manuscripts ● Presentation skills for academic conferences	4	CO 7
Week 8	● Final Project and Presentation ● Finalizing research paper drafts ● Presenting research findings to the class ● Peer feedback and evaluation	2	CO 8

Assessment:

Literature review assignment

Research methodology proposal

Data analysis and results presentation

Drafts of discussion and conclusion sections

Final research paper submission

Presentation of research findings

Course Code	Course Title	Credits
GNKPSBDRP503	**Research Project Proposal**	**02**

Guidelines for Project Proposal in Semester - III

- **Students should take a topic related to the specialization he or she is planning to take in Semester IV.**
- **Should have studied the related topics in the elective he or she has chosen in semester-II and semester- III**
- **A student is expected to devote at least 2 to 3 months of study as part of topic selection and its documentation.**
- **The student should be comfortable to implement the proposal in the semester – IV.**

Guidelines for Documentation of Project Proposal in Semester –III

Student is expected to make a project proposal documentation which should contain the following:

- **Title:** A suitable title giving the idea about what work is proposed.
- **Introduction:** An introduction to the topic of around 3-5 pages, giving proper back ground of the topic discussed.

- **Related works:** A detailed survey of the relevant works done by others in the domain. Student is expected to refer at least 5 research papers in addition to text books and web-links in the relevant topic. It may be around 7 to 10 pages.

- **Objective:** A detailed objective of the proposal is needed. It may be of 1 to 2 pages.

- **Methodology:** A proper and detailed procedure of how to solve the problem discussed. It shall contain the techniques, tools, software and data to be used. It shall be of around 3 to 5 pages.

The report may be of around 20 pages, which needs to be signed by the teacher in charge and head of the Department. Students should submit the signed project proposal documentation at the time of viva as part of the University examination

Shiromani Gurudwara Prabandhak Committee's

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)

Matunga, Mumbai – 400 019, Maharashtra

Program: Master of Science

Syllabus

Course: MSc-Big Data Analytics

Semester IV

(Major Paper- I)
Name of the paper: Cloud Computing

(As per NEP guidelines-DSC model)

With effect from Academic Year 2024 - 2025)

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)
Department of Computer Science

Course: MSc-II Big Data Analytics
Semester-IV Paper-I
Course Title: Cloud Computing
Course Code: GNKPSBDAMJ1504
Credits: 4
No of lectures (Hours): 60
Marks: 100 (75:25)

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	To provide students with the fundamentals and essentials of Cloud Computing.
2	To provide students a sound foundation of the Cloud computing so that they are able to start using and adopting Cloud Computing services and tools in their real life scenarios.
3	To enable students exploring some important cloud computing driven commercial systems such as Google Apps, Microsoft Azure and Amazon Web Services and other businesses cloud applications.
4	To impart knowledge in applications of cloud computing

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Design, Develop & Demonstrate real-world applications from the Cloud Computing	PO4	PSO4, PSO7	Ap,An
CO 2	Understand the subtle architectural difference in Public and PrivateClouds.	PO1	PS04	U,An,E
CO 3	Appreciate the requirements of various service paradigms in Cloud Computing.	PO1,PO5	PS04	U,Ap,E
CO 4	Describe the methods of processing multimedia elements and other information presentation concepts during multimedia communications.	PO2,PO3	PS06	U,An

Unit		Title	No. of lectures	CO Mapping

Unit 1			**15**	
	1.1	**Cloud Computing Overview:** Characteristics – challenges, benefits, limitations, Evolution of Cloud Computing, Cloud computing architecture, Cloud Reference Model (NIST Architecture)		**CO 1**
	1.2	**Infrastructure as a Service:** Service Model, Characteristics, Benefits, Enabling Technologies Case Study : AWS, OpenStack		**CO 2**
	1.3	**Platform as a Service:** Service Model, Characteristics, Benefits, Enabling Technologies Case Studies : IBM Bluemix,GAE, Microsoft Azure		**CO 3**
Unit 2			**15**	
	2.1	**Software as a Service:** Service Model, Characteristics, Benefits, Enabling Technologies Case Study : Salesforce.com, CRM, Online Collaboration Services		**CO 4**
	2.2	**Data Analytics as a Service:** Hadoop as a service, MapReduce on Cloud, Chubby locking Service		**CO 5**
	2.3	**Introduction to Public and Private Clouds:** Shared Resources – Resource Pool – Usage and Administration Portal – Usage Monitor – Resource Management– Cloud Security – Workload Distribution – Dynamic provisioning.		**CO 6**
Unit 3			**15**	
	3.1	**Storage as a service:** Historical Perspective, Datacenter Components, Design Considerations,		**CO 7**
	3.2	Power Calculations, Evolution of Data Centers,		**CO 8**
	3.3	Cloud data storage – CloudTM		**C0 9**
Unit 4			**15**	
	4.1	**Cloud Simulators- CloudSim and GreenCloud** Introduction to Simulator, understanding CloudSim simulator, CloudSim Architecture(User code, CloudSim, GridSim, SimJava) Understanding Working platform for CloudSim, Introduction to GreenCloud		**CO 10**
	4.2	**Introduction to VMWare Simulator** Basics of VMWare, advantages of VMware virtualization, using Vmware workstation		**CO 11**
	4.3	creating virtual machines-understanding virtual machines, create a new virtual machine on local host, cloning virtual machines, virtualize a physical machine, starting and stopping a virtual machine.		**C0 12**

References:

1. Kai Hwang, Geoffrey Fox, Jack J. Dongarra, Morgan Kaufmann, "Distributed and Cloud Computing: From Parallel Processing to the Internet of Things," 1st Edition, 2011.
2. Gautham Shroff, "Enterprise Cloud Computing: Technology, Architecture, Applications", Cambridge press, 2010.
3. Kris Jamsa, "Cloud Computing", Jones & Barlett Learning, 2013.
4. Rajkumar Buyya, James Broberg, Andrzej Goscinski, "Cloud Computing Principles and Paradigms", John Wiley & Sons, 2011.
5. Cloud computing a practical approach - Anthony T.Velte , Toby J. Velte Robert Elsenpeter, TATA McGraw- Hill , New Delhi – 2010
6. Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online - Michael Miller - Que 2008

Examination:

- **Internal Examination (25 Marks):** 20 Marks exam (Presentation). And 5 Marks for Class Participation etc.
- **End Semester theory examination (75 Marks):** Weightage of each unit will be proportional to the number of lecture hours as mentioned in the syllabus. Duration of exam: 2hours 30mins
- **Combined passing of 40% with minimum 20% in Internal Component.**

Course: MSc Big Data Analytics Practical
Semester-IV
Course Title: Cloud Computing Practical Paper-I
Course Code: GNKPSBDAMJ1P504
Credits: 02
No of Practical (Hours): 60
Marks: 50

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	Understand Cloud Computing Fundamentals: ● Define cloud computing and its key characteristics. ● Explain the differences between Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). ● Compare and contrast various cloud computing platforms such as AWS, IBM Blue Mix, and VirtualBox.
2	Master Network Design and Configuration: ● Design and implement VLANs for efficient network segmentation. ● Configure routing protocols to enable communication between different network segments. ● Perform subnetting and gateway configuration to optimize network performance and security.

3 **4**	Deploy and Manage Cloud Resources: ● Create and manage virtual machines using VirtualBox and AWS EC2 instances. ● Deploy static websites on AWS S3 buckets and configure access permissions. ● Develop and deploy web applications using AWS Beanstalk, ensuring scalability and availability. Automate Infrastructure Deployment and Management: ● Utilize configuration management tools such as Chef or Puppet to automate the deployment and configuration of OpenStack clusters. ● Implement DevOps practices to streamline software development and deployment processes. ● Design and implement auto-scaling solutions for applications to handle varying workloads efficiently.
5	Utilize Cloud Services for Data Management and Analytics: ● Deploy and manage Hadoop clusters as a service for big data processing and analysis. ● Implement Cloud TM for efficient data storage, retrieval, and management. ● Develop mobile applications on IBM Blue Mix, integrating backend services and data analytics capabilities.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Proficiency in Cloud Computing Technologies: Upon completing the course, students will demonstrate proficiency in deploying and managing various cloud computing services, including virtual machines on VirtualBox, AWS EC2 instances, and IBM Blue Mix applications.	PO5	PSO4	U,Ap
CO 2	Networking and Infrastructure Design Skills: Students will develop skills in designing and implementing complex network infrastructures, including VLANs, routing configurations, subnetting, gateway configurations, and load balancing, both locally and on cloud platforms like AWS.	PO4,PO8	PSO3	U,An
CO 3	Application Deployment and Management: Learners will gain hands-on experience in deploying and managing applications across different platforms, such as deploying static websites on AWS S3 buckets, web applications using AWS Beanstalk, and mobile applications on IBM Blue Mix.	PO3	PSO7	Ap,An
CO 4	Automation and DevOps Practices: Students will learn automation techniques using tools like	PO1,PO7	PS04	Ap

	Chef or Puppet to automate the deployment and configuration of infrastructure components, such as OpenStack clusters, and gain insights into DevOps practices for efficient software development and deployment.			
CO 5	Data Management and Analytics in the Cloud: Participants will understand how to leverage cloud-based services for data management and analytics, including deploying Hadoop as a Service for big data processing and utilizing Cloud TM for efficient data storage and retrieval.	PO2,PO6	PSO1,PSO 2,PSO5,PS 06	Ap,An

List of Experiments:

1. Cisco simulator – VLAN design, Routing, Sub netting, Gateway configuration
2. Virtual box based Webserver creation, Images/Snapshots access webpage from 2nd VM on another subnet work
3. EC2 AWS – S3 bucket based static webpages.
4. EC2 AWS – Instance Creation, Migration
5. EC2 AWS – Web application using Beanstalk.
6. AWS – Local balancing and auto scaling.
7. IBM Blue Mix - Mobile Application development
8. DaaS – Deployment of a basic web app and add additional functionality(Java scripts based)
9. PaaS – IOT – Mobile sensor based IOT application hosted via PaaS environment
10. SaaS – Deployment of any SaaS application for a online collaborative tool

References:

1. John Rhoton and Risto Haukiojal, “Cloud Computing Architectured: Solution Design Handbook”, Recursive Press, 2013.
2. George Recse, “Cloud Application Architectures: Building Application and Infrastructure in the Cloud” , O’ Reilly Media, First Edition, 2009.
3. Dinkar Sitaram, Geetha Manjunathan, “Moving to the Cloud: Developing Apps in the new world of Cloud Computing”, Syngress, 2012.
4. Samee. U. Khan, Albert. Y. Zomaya, “Handbook on Data Centers”, Springer,2015.

Examination (Total Marks): 50 (Two practical’s of 25 marks each)
Experiment Marks: 15
Journal Marks: 05
Viva Marks: 05

Shiromani Gurudwara Prabandhak Committee's

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)

Matunga, Mumbai – 400 019, Maharashtra

Program: Master of Science

Syllabus

Course: MSc-Big Data Analytics

Semester IV

(Major Paper- II)
Name of the paper: Streaming Data Analytics

(As per NEP guidelines-DSC model)

With effect from Academic Year 2024 - 2025)

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)
Department of Computer Science

Course: MSc-II Big Data Analytics
Semester-III Paper-II
Course Title: Streaming Data Analytics
Course Code: GNKPSBDAMJ2504
Credits: 4
No of lectures (Hours): 60
Marks: 100 (75:25)

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	It introduces theoretical foundations, algorithms, methodologies, and applications of streaming data and also provide practical knowledge for handling and analyzing streaming data.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Recognize the characteristics of data streams that make it useful to solve real-world problems.	PO1,PO2	PSO1,PSO 2	An,Ap
CO 2	Identify and apply appropriate algorithms for analyzing the data streams for variety of problems.	PO2	PSO4	U,Ap
CO 3	Implement different algorithms for analyzing the data streams	PO2,PO3	PSO5	An,Ap
CO 4	Identify the metrics and procedures to evaluate a model	PO6,PO7	PSO6,pso 7	Ap,C

Unit		Title	No. of lectures	CO Mapping
Unit 1			**15**	
	1.1	**Introduction:** Characteristics of the data streams, Challenges in mining data streams Requirements and		**CO 1**

		principles for real time processing, Concept drift Incremental learning.		
	1.2	**Data Streams** Basic Streaming Methods, Counting the Number of Occurrence of the Elements in a Stream, Counting the Number of Distinct Values in a Stream, Bounds of Random Variables,Poisson Processes		**CO 2**
	1.3	Maintaining Simple Statistics from Data Streams, Sliding Windows, Data Synopsis, Change Detection: Tracking Drifting Concepts, Monitoring the Learning Process		**CO 3**
Unit 2			**15**	
	2.1	**Decision Trees:** Very Fast Decision Tree Algorithm (VFDT), The Base Algorithm, Analysis of the VFDT Algorithm,		**CO 4**
	2.2	**Extensions to the Basic Algorithm:** Processing Continuous Attributes, Functional Tree Leaves, Concept Drift.		**CO 5**
	2.3	**Clustering from Data Streams** Clustering Examples: Basic Concepts, Partitioning Clustering - The Leader Algorithm, Single Pass k-Means, Micro Clustering, Clustering Variables: A Hierarchical Approach		**CO 6**
Unit 3			**15**	
	3.1	**Frequent Pattern Mining** Mining Frequent Item sets from Data Streams-Landmark Windows		**CO 7**
	3.2	Mining Recent Frequent Item sets, Frequent Item sets at Multiple Time Granularities		**CO 8**
	3.3	Sequence Pattern Mining- Reservoir Sampling for Sequential Pattern Mining over data streams		**C0 9**
Unit 4			**15**	
	4.1	**Evaluating Streaming Algorithms** Evaluation Issues, Design of Evaluation Experiments, Evaluation Metrics, Error Estimators using a Single Algorithm and a Single Dataset, Comparative Assessment, The 0-1 loss function		**CO 10**
	4.2	Evaluation Methodology in Non-Stationary Environments, The Page-Hinkley Algorithm.		**CO 11**
	4.3	**Complex Event Processing:** Introduction, Features of CEP, Need for CEP, CEP Architectural Layers, Scaling CEP, Events, Timing and Causality, Event Patterns, Rules and Constraint, STRAW- EPL, Complex Events and Event Hierarchies		**C0 12**

References:

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

Examination:

- **Internal Examination (25 Marks):** 20 Marks exam (Presentation). And 5 Marks for Class Participation etc.
- **End Semester theory examination (75 Marks):** Weightage of each unit will be proportional to the number of lecture hours as mentioned in the syllabus. Duration of exam: 2hours 30mins
- **Combined passing of 40% with minimum 20% in Internal Component.**

Course: MSc Big Data Analytics Practical

Semester-III

Course Title: Streaming Data Analytics Practical Paper-I

Course Code: GNKPSBDAMJ2P504

Credits: 02

No of Practical (Hours): 60

Marks: 50

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	It introduces theoretical foundations, algorithms, methodologies, and applications of streaming data and also provide practical knowledge for handling and analyzing streaming data.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Recognize the characteristics of data streams that make it useful to solve real-world problems.	PO1,PO2	PSO1,PSO 2	An,Ap
CO 2	Identify and apply appropriate algorithms for analyzing the data streams for variety of problems.	PO2	PSO4	U,Ap

CO 3	Implement different algorithms for analyzing the data streams	PO2,PO3	PSO5	An,Ap

List of Experiments:

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

References:

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

Examination (Total Marks): 50 (Two practical’s of 25 marks each)
Experiment Marks: 15
Journal Marks: 05
Viva Marks: 05

Shiromani Gurudwara Prabandhak Committee's

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)

Matunga, Mumbai – 400 019, Maharashtra

Program: Master of Science

Syllabus

Course: MSc-Big Data Analytics

Semester IV

(Elective Paper- I)
Name of the paper: Domain Specific Predictive Analytics

(As per NEP guidelines-DSC model)

With effect from Academic Year 2024 - 2025)

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)
Department of Computer Science

Course: MSc-II Big Data Analytics
Semester-IV Elective Paper
Course Title: Domain Specific Predictive Analytics
Course Code: GNKPSBDAEL1504
Credits: 3
No of lectures (Hours): 45
Marks: 100 (75:25)

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	It introduces theoretical foundations, algorithms, methodologies for analysing data in various domains such Retail, Finance, Risk and Healthcare.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Recognize challenges in dealing with data sets in domains such as finance, risk and healthcare.	PO2,PO3	PSO1,PSO 2	U,An,Ap
CO 2	Identify real-world applications of machine learning in domains such as finance, risk and healthcare.	PO1,PO6, PO7	PSO3	Ap,E
CO 3	Identify and apply appropriate algorithms for analyzing the data for variety of problems in finance, risk and healthcare.	PO3,PO7	PSO4	An,Ap,E
CO 4	Make choices for a model for new machine learning tasks based on reasoned argument	PO3,PO7	PSO5	Ap,E,C

Unit		Title	No. of lectures	CO Mapping
Unit 1			**15**	
	1.1	**Retail Analytics:** **Understanding Customer: Profiling and Segmentation, Modelling Churn. Modelling Lifetime Value, Modelling Risk, Market Basket Analysis.**		**CO 1**
	1.2	**Risk Analytics:** Risk Management and Operational Hedging: An Overview, Supply Chain Risk Management		**CO 2**
	1.3	A Bayesian Framework for Supply Chain Risk Management, Credit Scoring and Bankruptcy Prediction		**CO 3**
Unit 2			**15**	
	2.1	**Financial Data Analytics** Financial News analytics: Framework, techniques, and metrics, News events impact market sentiment, Relating news analytics to stock returns		**CO 4**
	2.2	**Financial Time Series Analytics** Financial Time Series and Their Characteristics, Common Financial Time Series models,Long term forecasting		**CO 5**
	2.3	**Social Media Analytics for Healthcare:** Tracking of Infectious Disease Outbreaks, Readmission risk Prediction		**CO 6**
Unit 3			**15**	
	3.1	**Healthcare Analytics:** Introduction to Healthcare Data Analytics, Electronic Health Records, Privacy-Preserving Data Publishing Methods in Healthcare, Clinical Decision Support Systems		**CO 7**
	3.2	**Healthcare Data Analytics:** Natural Language Processing and Data Mining for Clinical Text: Core NLP Components, Information Extraction and Named Entity Recognition		**CO 8**
	3.3	**Genomic Data Analytics:** Microarray Data, Microarray Data Analysis, Genomic Data Analysis for Personalized Medicine, Patient Survival Prediction from Gene Expression Data, Genome Sequence Analysis		**C0 9**

References:

1. Chris Chapman, Elea McDonnell Feit "R for Marketing Research and Analytics", Springer, 2015.
2. Olivia Parr Rud “Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management”, Wiley, 2001.

3. Chandan K. Reddy, Charu C. Aggarwal "Healthcare Data Analytics", CRC Press,2015.
4. Rene Carmona "Statistical Analysis of Financial Data in R", Springer, 2014.
5. James B. Ayers “Handbook Of Supply Chain Management” Auerbach Publications, 2006.
6. Panos Kouvelis, Ling xiu Dong, Onur Boyabatli, Rong Li "The Handbook of Integrated Risk Management in Global Supply Chains", Wiley, 2012.

Examination:

- **Internal Examination (25 Marks):** 20 Marks exam (Presentation). And 5 Marks for Class Participation etc.
- **End Semester theory examination (75 Marks):** Weightage of each unit will be proportional to the number of lecture hours as mentioned in the syllabus. Duration of exam: 2hours 30mins
- **Combined passing of 40% with minimum 20% in Internal Component.**

Course: MSc-II Big Data Analytics Practical
Semester-IV Elective Paper
Course Title: Domain Specific Predictive Analytics Practical
Course Code: GNKPSBDAEL1P504
Credits: 01
No of Practical (Hours): 30
Marks: 50

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	It introduces theoretical foundations, algorithms, methodologies for analysing data in various domains such Retail, Finance, Risk and Healthcare.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Recognize challenges in dealing with data sets in domains such as finance, risk and healthcare.	PO2,PO3	PSO1,PSO 2	U,An,Ap
CO 2	Identify real-world applications of machine learning in domains such as finance, risk and healthcare.	PO1,PO6, PO7	PSO3	Ap,E
CO 3	Identify and apply appropriate algorithms for analyzing the data for variety of problems in	PO3,PO7	PSO4	An,Ap,E

	finance, risk and healthcare.			
CO 4	Make choices for a model for new machine learning tasks based on reasoned argument	PO3,PO7	PSO5	Ap,E,C

List of Experiments:

Perform case study on

1. Retail Analytics
2. Risk Analytics
3. Financial Data Analytics
4. Financial Time Series Analytics
5. Social Media Analytics for Healthcare
6. Healthcare Analytics
7. Healthcare Data Analytics
8. Genomic Data Analytics

References:

1. Rene Carmona "Statistical Analysis of Financial Data in R", Springer, 2014.
2. James B. Ayers “Handbook Of Supply Chain Management” Auerbach Publications, 2006.
3. Panos Kouvelis, Ling xiu Dong, Onur Boyabatli, Rong Li "The Handbook of Integrated Risk Management in Global Supply Chains", Wiley, 2012.

Examination (Total Marks): 50

Experiment Marks: 40

Journal Marks: 05

Viva Marks: 05

Shiromani Gurudwara Prabandhak Committee's

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)

Matunga, Mumbai – 400 019, Maharashtra

Program: Master of Science

Syllabus

Course: MSc-Big Data Analytics

Semester IV

(Elective Paper- II)
Name of the paper: Text, Web and Social Media Analytics

(As per NEP guidelines-DSC model)

With effect from Academic Year 2024 - 2025)

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)
Department of Computer Science

Course: MSc-II Big Data Analytics
Semester-IV Elective Paper
Course Title: Text, Web and Social Media Analytics
Course Code: GNKPSBDAEL2504
Credits: 3
No of lectures (Hours): 45
Marks: 100 (75:25)

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	To provide an overview of common text mining and social media data analytic activities.
2	To understand the complexities of processing text and network data from different data sources.
3	To enable students to solve complex real-world problems for sentiment analysis and Recommendation systems.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Interpret the terminologies, metaphors and perspectives of social media analytics.	PO1,PO2	PSO2,PSO 3	R,U
CO 2	Apply a wide range of classification, clustering, estimation and prediction algorithms on Textual data.	PO3	PSO3	U,An,Ap
CO 3	Perform social network analysis to identify important social actors, subgroups and network properties in social media sites.	PO2,PO3	PSO5	Ap,E
CO 4	Apply state of the art web mining tools and libraries on realistic datasets as a basis for business decisions and applications.	PO6	PSO7,PSO 8	Ap,E,C
CO 5	Provide solutions to the emerging problems with social media such as behavior analytics and recommendation systems.	PO7	PSO7	An,Ap,E,C

CO 6	Design new solutions to opinion extraction, sentiment classification and data summarization problems.	PO6,PO7	PSO8,PSO 9	Ap,E,C

Unit		**Title**	**No. of lectures**	**CO Mapping**
Unit 1			**15**	
	1.1	**Introduction to Text Mining** Text Representation- tokenization, stemming, stop words, TF-IDF		**CO 1**
	1.2	Feature Vector Representation, NER, N-gram modeling.		**CO 2**
	1.3	**Mining Textual Data** Text Clustering, Text Classification, Topic Modeling-LDA, HDP		**CO 3**
Unit 2			**15**	
	2.1	**Introduction to Web-Mining** Inverted indices and Boolean queries. PLSI, Query optimization, page ranking.		**CO 4**
	2.2	**Web Usage Web content Mining** Web Crawling-Crawler Algorithms, Implementation Issues, Evaluation, Session & visitor Analysis, Visitor Segmentation		**CO 5**
	2.3	Analysis of Sequential & Navigational Patterns, Predictions based on web user transactions. Web content Mining		**CO 6**
Unit 3			**15**	
	3.1	**Introduction to Social Media Network** Essentials of Social graphs, Social Networks, Models, Information Diffusion in Social Media.		**CO 7**
	3.2	**Mining Social Media** Behavioral Analytics, Influence and Homophily, Recommendation in Social Media.		**CO 8**
	3.3	**Sentimental Mining** Sentiment classification feature based opinion mining, comparative sentence and relational mining, Opinion spam.		**C0 9**

References:

1. Bing Liu, “Web Data Mining-Exploring Hyperlinks, Contents, and Usage Data”, Springer, Second Edition, 2011.
2. Reza Zafarani, Mohammad Ali Abbasi and Huan Liu, “Social Media Mining - An Introduction”, Cambridge University Press, 2014.
3. Bing Liu, “Sentiment Analysis and Opinion Mining”, Morgan & Claypool Publishers, 2012.
4. Nitin Indurkhya, Fred J Damerau, “Handbook of Natural Language Process”, 2nd Edition, CRC Press, 2010.
5. Matthew A.Russell, “Mining the social web”, 2nd edition- O'Reilly Media, 2013.

Examination:

- **Internal Examination (25 Marks):** 20 Marks exam (Presentation). And 5 Marks for Class Participation etc.
- **End Semester theory examination (75 Marks):** Weightage of each unit will be proportional to the number of lecture hours as mentioned in the syllabus. Duration of exam: 2hours 30mins
- **Combined passing of 40% with minimum 20% in Internal Component.**

Course: MSc-II Big Data Analytics Practical

Semester-IV Elective Paper

Course Title: Text, Web and Social Media Analytics Practical

Course Code: GNKPSBDAEL2P504

Credits: 01

No of Practical (Hours): 30

Marks: 50

Course Objectives:

Sr. No.	Course objectives
The course aims at:	
1	It introduces theoretical foundations, algorithms, methodologies for analysing data in various domains such Retail, Finance, Risk and Healthcare.

Course Outcomes (COs):

Sr. No.	On completing the course, the student will be able to:	POs addressed	PSOs addressed	Cognitive Levels addressed
CO 1	Interpret the terminologies, metaphors and perspectives of social media analytics.	PO1,PO2	PSO2,PSO3	R,U
CO 2	Apply a wide range of classification, clustering, estimation and prediction algorithms on Textual data.	PO3	PSO3	U,An,Ap
CO 3	Perform social network analysis to identify important social actors, subgroups and network properties in social media sites.	PO2,PO3	PSO5	Ap,E
CO 4	Apply state of the art web mining tools and libraries on realistic datasets as a basis for business decisions and applications.	PO6	PSO7,PSO8	Ap,E,C
CO 5	Provide solutions to the emerging problems with social media such as behavior analytics and recommendation systems.	PO7	PSO7	An,Ap,E,C
CO6	Design new solutions to opinion extraction, sentiment classification and data summarization problems.	PO6,PO7	PSO8,PSO9	Ap,E,C

List of Experiments:

Perform case study on

1. Text mining
2. Web Mining
3. Web Content Mining
4. Predictions based on web user transactions.
5. Social media mining
6. Sentimental Mining
7. Analysis of Sequential Patterns

8. Analysis of Navigational Patterns

References:

1. Bing Liu, “Web Data Mining-Exploring Hyperlinks, Contents, and Usage Data”, Springer,
    a. Second Edition, 2011.
2. Reza Zafarani, Mohammad Ali Abbasi and Huan Liu, “Social Media Mining - An
    a. Introduction”, Cambridge University Press, 2014.
3. Bing Liu, “Sentiment Analysis and Opinion Mining”, Morgan & Claypool Publishers, 2012.
4. Nitin Indurkhya, Fred J Damerau, “Handbook of Natural Language Process”, 2nd Edition,
    a. CRC Press, 2010.
5. Matthew A.Russell, “Mining the social web”, 2nd edition- O'Reilly Media, 2013.

Examination (Total Marks): 50
Experiment Marks: 40
Journal Marks: 05
Viva Marks: 05

Shiromani Gurudwara Prabandhak Committee's

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)

Matunga, Mumbai – 400 019, Maharashtra

Program: Master of Science

Syllabus

Course: MSc-Big Data Analytics

Semester IV

(Research Project)

Name of the paper: Research Project Implementation

(As per NEP guidelines-DSC model)

With effect from Academic Year 2024 - 2025)

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)
Department of Computer Science

Course: MSc Big Data Analytics

Semester-IV

Course Title: Research Project Implementation (RP)

Course Code: GNKPSBDAERP504

Credits: 4

No of lectures (Hours): 120

Marks: 100

Course Code	Course Title	Credits
GNKPSBDARP504	**Research Project Implementation**	**04**
	Guidelines for Research Project Implementation ● A student is expected to devote at least 3 to 4 months of effort to the implementation. ● Students should submit a detailed project implementation report at the time of viva.	

Guidelines for Documentation of Research Project Implementation

A student should submit a project implementation report with the following details:

- Title: Title of the project.
- Objective: A detailed objective of the proposal is needed.
- Related works: A detailed survey of the relevant works done by others in the domain. The student is expected to refer to at least 15 recent (last five years) research papers in addition to textbooks and web links in the relevant topic.
- Methodology: A proper and detailed procedure of how to solve the problem discussed. It shall contain the techniques, tools, software, and data to be used.
- Implementation details: A description of how the project has been implemented.
- Experimental setup and results: A detailed explanation of how experiments were conducted, what software was used, and the results obtained. Details like screenshots, tables, and graphs can come here.
- Analysis of the results: A description of what the results mean and how they have been arrived at. Different performing measures or statistical tools used etc may be part of this.
- Conclusion: A conclusion of the project performed in terms of its outcome

	● Future enhancement: A small description of what enhancement can be done when more time and resources are available ● Program code: The program code may be given as an appendix. The project documentation needs to be signed by the teacher in charge and head of the Department. Students should also attach the certified copy of the internal evaluation report (Appendix I) at the time of Project evaluation and viva as part of the University examination.	

Examination:

- **Internal Examination 50 Marks): Per week discussion with internal guide + Documentation.**
- **End Semester Project Viva examination (50 Marks): A learner will be evaluated based on innovation in research, implementation uniqueness and working of project.**
- **Combined passing of 40% with minimum 20% in both Components.**

Appendix-I

Maintain the weekly online diary for each week in the following format.

	Day	Date	Name of the Topic/Module Completed	Remarks
WEEK No	MONDAY			
	TUESDAY			
	WEDNESDAY			
	THRUSDAY			
	FRIDAY			
	SATURDAY			

Signature of the Faculty mentor:

Seal of the University/College

xxxx

Course Outcomes

Bloom's Taxonomy categorizes educational objectives into six cognitive levels, listed from the simplest to the most complex.

The course outcomes (CO) are mapped to the revised Bloom's Taxonomy using the following Cognitive levels along with abbreviations: **Remembering-R, Understanding-U, Applying-Ap, Analyzing-An, Evaluating-E, Creating-C.**

R-Remembering: This level involves recalling facts, terms, basic concepts, and answers without necessarily understanding the meaning or implications.

U-Understanding: Understanding involves explaining ideas or concepts and interpreting them in one's own words to demonstrate comprehension.

Ap-Applying: Applying knowledge involves using acquired knowledge in new situations or applying it in different ways.

An-Analyzing: Analyzing involves breaking down information into parts to understand its organizational structure, recognizing patterns, and identifying relationships between components.

E-Evaluating: Evaluating involves making judgments based on criteria and standards, assessing the value of theories, presentations, or materials.

C-Creating: Creating involves putting elements together to form a coherent or functional whole, reorganizing elements into a new pattern or structure.

Use the following action verbs for Blooms taxonomy levels to prepare Course outcomes:

No	Levels	Action verbs
1	**Remember**	Choose, Describe, Define, Label List, Locate, Match, Memorize, Name, Omit, Recite, Select, State, Count, Draw, Outline, Point, Quote, Recall, Recognize, Repeat, Reproduce, Recall, Arrange, Duplicate, Tabulate.
2	**Understand**	

		Restate, Discuss, Clarify, Locate, Recognise, Classify, Translate, Explain, Express, Review, Interpret, Select, Summarise, Contrast, Predict, Associate, Estimate, Extend
3	**Apply**	Demonstrate, Schedule, Operate, Dramatize, Apply, Employ, Use, Practise, Illustrate, Choose, Solve, Write, Calculate, Complete, Show, Examine, Modify, Relate, Classify, Experiment.
4	**Analyse**	Distinguish, Differentiate, Investigate, Categorise, Appraise, Inspect, Test, Debate, Compare, Contrast, Question
5	**Evaluate**	Judge, Score, Select, Evaluate, Choose, Rate, Assess, Compare, Estimate, Value, Measure, Discriminate, Argue, Defend, Support, Conclude, Summarize, Appraise, Revise.
6	**Create**	Compose, Assemble, Organise, Plan, Collect, Propose, Construct, Design, Create, Formulate, Arrange, Devise, Modify, Derive, Develop, Integrate, Rearrange, Substitute, Invent, Generalise.

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)

PROGRAMME OUTCOMES (PO)

MASTERS IN SCIENCE (MSc)

The undergraduate and postgraduate Programmes in Sciences will instil in the students a scientific temper and impart a holistic education through the following outcomes:

Postgraduate Science Program Outcomes:

PO1	**Advanced Understanding:** Attain an advanced understanding of specialized scientific areas and theories, building upon undergraduate knowledge.
PO2	**Advanced Analytical Skills:** Develop advanced analytical and problem-solving skills for complex scientific issues.
PO3	**Research Proficiency:** Acquire proficiency in scientific research, contributing significantly to the advancement of knowledge.
PO4	**Leadership Skills:** Develop leadership skills to lead scientific initiatives and contribute to the scientific community.
PO5	**Interdisciplinary Collaboration:** Collaborate effectively with interdisciplinary teams for comprehensive scientific solutions.
PO6	**Innovation and Creativity:** Foster innovation and creativity in scientific research and practice.
PO7	**Professionalism:** Demonstrate professionalism and excellence in all scientific endeavours.
PO8	**Quality Improvement:** Embrace self-evaluation and continuous improvement for achieving excellence in scientific pursuits.

Guru Nanak Khalsa College of Arts, Science and Commerce (Autonomous)
Department of Computer Science

Programme: MSc Big Data Analytics

Programme Specific Outcomes (PSOs) for MSc in MSc Big Data Analytics.

Sr. No.	A student completing MSc Big Data Analytics will be able to:
PSO 1	Proficiency in NoSQL Databases ● Understand the fundamentals of NoSQL databases including their types, architectures, and use cases. ● Demonstrate the ability to design, implement, and manage NoSQL databases to store and retrieve large volumes of unstructured and semi-structured data efficiently.
PSO 2	Data Analytics Proficiency with R ● Apply advanced statistical techniques using R programming language for data analysis and visualization. ● Develop and implement data analytics solutions to solve real-world problems using R and relevant packages.
PSO 3	Competence in Operating Systems and Visualization ● Understand the concepts and principles of operating systems and their role in managing resources for big data processing. ● Apply visualization techniques to interpret and communicate insights derived from large datasets effectively.
PSO 4	Proficiency in Cloud Computing ● Understand the principles and architectures of cloud computing and its role in big data analytics. ● Demonstrate the ability to deploy, manage, and scale big data applications in cloud environments using relevant platforms and services.
PSO 5	Proficiency in Streaming Data Analytics ● Understand the challenges and techniques involved in analyzing streaming data in real-time. ● Develop and implement streaming data analytics solutions to extract actionable insights from continuous data streams.

PSO 6	Proficiency in Domain-Specific Predictive Analytics OR Text, Web, and Media Analytics OR Mining Massive Data ● Domain-Specific Predictive Analytics: Apply predictive modeling techniques to specific domains such as finance, healthcare, or marketing to make informed decisions. ● Text, Web, and Media Analytics: Analyze and derive insights from text data, web content, and multimedia sources using advanced analytics techniques. ● Mining Massive Data: Apply scalable data mining techniques to extract patterns and knowledge from massive datasets efficiently.
PSO 7	● Conduct independent research projects in the field of big data analytics. ● Demonstrate the ability to critically analyze existing literature, propose novel solutions, and present findings effectively.