Savitribai Phule Pune University Fourth Year of Artificial Intelligence and Data Science (2020 Course) 417530: Distributed Computing		
Teaching Scheme: TH: 03 Hours/Week	Credit 03	Examination Scheme: In-Sem (Paper): 30 Marks End-Sem (Paper): 70 Marks

Prerequisites Courses: Computer Network (317521), Data Science (317529)

Companion Course: Computer Laboratory III (417533)

Course Objectives:

- To understand the fundamentals and knowledge of the architectures of distributed systems
- To gain knowledge of working components and fault tolerance of distributed systems
- To make students aware about security issues and protection mechanisms for distributed environments

Course Outcomes:

After completion of the course, learners should be able to-

CO1: Understand the features and properties of Distributed computing system with integration of AI

CO2: Analyze the Concept of data management and storage in distributed computing

CO3: Understand the algorithm used in distributed computing by applying artificial intelligence

CO4: Understand the integration of machine learning algorithm and advanced tools used in distributed computing

CO5: Analyze how big data is processed in distributed computing

CO6: Identify Security and privacy issues of distributed computing and apply on specific application

Course Contents

Unit I	Introduction to Distributed Computing	07 Hours

Fundamentals of distributed computing: Characteristics of Distributed Systems: Issues, Goals, and Types of distributed systems, Distributed System Models

Introduction to Artificial Intelligence and Data Science in distributed computing:

Distributing computational tasks, handling large volumes of data, and leveraging parallel processing capabilities, issues related to data storage and retrieval, data consistency, communication overhead, synchronization, and fault tolerance.

Use cases and applications of integrating AI and data science in distributed systems:

Predictive Maintenance, Fraud Detection, Intelligent Transportation Systems, Supply Chain Optimization, Energy Management, Healthcare and Medical Diagnostics, Customer Behavior Analysis and Natural Language Processing (NLP)

#Exemplar/Case Studies	Introduction to Distributed Computing in E-commerce
*Mapping of Course Outcomes for Unit I	CO1

Unit II	Distributed Data Management and Storage	07 Hours

Overview of Distributed Computing Frameworks and Technologies
Parallel Computing, Distributed Computing Models, Message Passing, Distributed File Systems: Hadoop Distributed File System (HDFS) and Google File System (GFS), Cluster Computing: (AWS), Microsoft Azure, and Google Cloud Platform (GCP), Message Brokers and Stream Processing, Edge Computing

Data Replication and Consistency Model: Eager Replication, Lazy Replication, Quorum-Based Replication, Consensus-Based Replication, Selective Replication, Strong Consistency, Eventual Consistency, Read-your-writes Consistency, Consistent Prefix Consistency, Causal Consistency

Distributed data indexing and retrieval techniques: Distributed Hash Tables (DHTs), Distributed Inverted Indexing, Range-based Partitioning, Content-based Indexing, Peer-to-Peer (P2P) Indexing, Hybrid Approaches

#Exemplar/Case Studies	Distributed Data Management and Storage in Healthcare
*Mapping of Course Outcomes for Unit II	CO2

Unit III	Distributed Computing Algorithms	07 Hours

Distributed Computing Algorithms: Communication and coordination in distributed systems
Distributed consensus algorithms (Other consensus algorithms ● Viewstamped Replication ● RAFT ● ZAB ● Mencius ● Many variants of Paxos (Fast Paxos, Egalitarian Paxos etc)
Fault tolerance and recovery in distributed systems,

Load balancing and resource allocation strategies: Weighted Round Robin, Least Connection, Randomized Load Balancing, Dynamic Load Balancing, Centralized Load Balancing, Distributed Load Balancing, Predictive Load Balancing

Applying AI techniques to optimize distributed computing algorithms: Machine Learning for Resource Allocation, Reinforcement Learning for Dynamic Load Balancing, Genetic Algorithms for Task Scheduling, Swarm Intelligence for Distributed Optimization

#Exemplar/Case Studies	Distributed Computing Algorithms in Weather Prediction
*Mapping of Course Outcomes for Unit III	CO3

Unit IV	Distributed Machine Learning and AI	07 Hours

Introduction to distributed machine learning algorithms: Types of Distributed Machine Learning: Data Parallelism and Model Parallelism, Distributed Gradient Descent, Federated Learning, All-Reduce, Hogwild, Elastic Averaging SGD

Software to implement Distributed ML: Spark, GraphLab, Google TensorFlow, Parallel ML System (Formerly Petuum), Systems and Architectures for Distributed Machine Learning

Integration of AI algorithms in distributed systems: Intelligent Resource Management, Anomaly Detection and Fault Tolerance, Predictive Analytics, Intelligent Task Offloading

#Exemplar/Case Studies	Distributed Machine Learning and AI in Fraud Detection
*Mapping of Course Outcomes for Unit IV	CO4

Unit V	Big Data Processing in Distributed Systems	07 Hours

Big data processing frameworks in distributed computing: Hadoop, Apache Spark, Apache Storm, Samza, Flink

Parallel and distributed data processing techniques: Single Instruction Single Data (SISD), Multiple Instruction Single Data (MISD), Single Instruction Multiple Data (SIMD), Multiple Instruction Multiple Data (MIMD), Single program multiple data (SPMD), Massively parallel processing (MPP)

Scalable data ingestion: types of data ingestion, Benefits, challenges, tools, transformation in distributed systems

Real-time analytics and Streaming analytics: types of real time analytics, types of streaming analytics, Comparison of real time analytics and streaming analytics, Applying AI and data science for large-scale data processing and analytics.

#Exemplar/Case Studies	Big Data Processing in Distributed Systems for Social Media Analytics
*Mapping of Course Outcomes for Unit V	CO5

Unit VI	Distributed Systems Security and Privacy	07 Hours

Security Challenges in Distributed Systems, Insider Threats, Encryption and Secure Communication: TLS/SSL, PKI, VPN, AMQP, Privacy Preservation Techniques: Differential Privacy, Homomorphic Encryption, Secure Multi-Party Computation (SMPC), Federated Learning, Anonymization and Pseudonymization, Access Control and Data Minimization, AI-based Intrusion Detection and Threat Mitigation Techniques: Anomaly Detection, Behavior-based Detection, Threat Intelligence and Analysis, Real-time Response and Mitigation, Adaptive Security, User and Entity Behavior Analytics (UEBA), Threat Hunting and Visualization.

#Exemplar/Case Studies	Distributed Systems Security and Privacy in Healthcare
*Mapping of Course Outcomes for Unit VI	CO6

Learning Resources

Text Books:

1. Distributed Computing and Artificial Intelligence, 12th International Conference: 373 (Advances in Intelligent Systems and Computing) Paperback by Sigeru Omatu (Editor), Qutaibah M. Malluhi (Editor), Sara Rodríguez Gonzalez (Editor), Grzegorz Bocewicz (Editor), Edgardo Bucciarelli (Editor), Gianfranco Giulioni (Editor), Farkhund Iqba (Editor)
2. Distributed Systems, Concepts and Design, George Coulouris, J Dollimore and Tim Kindberg, Pearson Education, 5th Edition
3. Distributed Systems, Principles and paradigms, Andrew S.Tanenbaum, Maarten Van Steen, Second Edition, PHI
4. Distributed Artificial Intelligence by Michael Huhns Volume I 1st Edition - January 1, 1987

Reference Books:

1. Distributed OS by Pradeep K. Sinha (PHI)
2. Tanenbaum S.: Distributed Operating Systems, Pearson Education
3. George Coulouris, Jean Dollimore. Tim Kindberg: Distributed Systems concepts and design
4. Distributed Systems, An Algorithm Approach, Sikumar Ghosh, Chapman & Hall/CRC, Taylor & Francis Group, 2007.
5. Distributed Computing: Principles, Algorithms, and Systems by Ajay D. Kshemkalyani, Mukesh Singhal

e-Resources:
1. George-Coulouris-Distributed Systems-5th-Edition.pdf (google.com)
2. ds-solutions.pdf (distributed-systems.net)
3. Distributed Systems: Principles and Paradigms (fsinf.at)
4. Distributed Systems (wordpress.com)
5. https://eclass.uoa.gr/modules/document/file.php/D245/2015/DistrComp.pdf

MOOC Courses:
1. NPTEL: https://archive.nptel.ac.in/courses/106/106/106106168/
2. Distributed Computing with Spark SQL | Coursera
3. Distributed Systems for Practitioners | Educative

The CO-PO Mapping Matrix

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	1	-	1	1	-	-	-	-	-	-	1	-
CO2	1	-	1	2	-	-	-	-	-	-	-	-
CO3	-	-	1	1	-	-	-	-	-	-	1	-
CO4	2	-	2	1	1	-	-	-	-	-	1	-
CO5	1	-	1	2	2	-	-	-	-	-	-	-
CO6	1	-	2	2	3	-	-	-	-	-	1	-

.