NIRMA UNIVERSITY

Institute of Technology

M Tech Computer Science and Engineering (Data Science)

Semester – I

L	T	P	C
3	0	2	4

Course Code	6CS301
Course Title	Introduction to Scalable Systems

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to –

1. comprehendthe distributed computing models for scalable systems
2. analyse the scalable systems in the context of various performance parameters
3. apply concepts of scalable systems in designing data intensive applications

Syllabus:

Teaching hours:

Unit I **5**

Introduction and Architectures: Features and types of distributed systems, Distributed Models of Computation, Architectures: Regular graphs, random graphs, power-law, and small-world networks, middleware, and self-management in distributed systems, causality & logical time

Unit II **5**

Distributed System Models and Enabling Technologies: Scalable Computing over the Internet, Technologies for Network-Based Systems, System Models for Distributed and Cloud Computing, Software Environments for Distributed Systems and Clouds, Performance, Security, and Energy Efficiency

Unit III **5**

Consistency and Replication: Consistency models: strong and weak, Scalable Causal Consistency, Highly reliable distributed coordination with Zookeeper, Replication management, Distributed Replication, PAXOS and RAFT Algorithms

Unit IV **6**

Computer Clusters for Scalable Parallel Computing: Clustering for Massive Parallelism, Computer Clusters and MPP Architectures, Design Principles of Computer Clusters, Cluster Job and Resource Management, Case Studies of Top Supercomputer Systems

Unit V **7**

Cloud Platform Architecture over Virtualized Data Centers: Cloud Computing and Service Models, Data-Center Design and Interconnection Networks, Architectural Design of Compute and Storage Clouds, Public Cloud Platforms: GAE,

AWS, Azure, Inter-Cloud Resource Management, Cloud Security and Trust Management

Unit VI **6**

Cloud Programming and Software Environments: Features of Cloud and Grid Platforms, Parallel and Distributed Programming Paradigms, Programming Support of Google App Engine, Programming on Amazon AWS and Microsoft Azure, Emerging Cloud Software Environments

Unit VII **6**

Fault Tolerance and Security: Failure models, failure detection, algorithms for fault tolerance, and recovery from failure in distributed systems, Authentication in Distributed Systems, Distribution of security mechanisms, access control, and security management

Unit VIII **5**

File Systems and Distributed Storage Systems: Network File System, Andrew File system, Google File System, Hadoop Distributed File System

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

1. Kai Hwang, Geoffrey C. Fox and Jack J. Dongarra, Distributed and Cloud Computing from Parallel Processing to the Internet of Things, Elsevier
2. Andrew S. Tanenbaum and Maarten van Steen, Distributed Systems: Principles and Paradigms, Createspace
3. George Coulouris, Jean Dollimore, Tim Kindberg, Gordon Blair, Distributed Systems: Concepts and Design, Addison Wesley
4. Ajay D. Kshemkalyani and MukeshSinghal, Distributed Computing: Principles, Algorithms, and Systems, Cambridge University Press
5. Kenneth P Birman, Guide to Reliable Distributed Systems: Building High-Assurance Applications and Cloud-Hosted Services, Springer
6. Relevant research papers in the area of distributed systems

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list