



FIRST/ SECOND SEMESTER 2020-2021

Course Handout Part II

Date: 11-08-2020

In addition to part-I (General Handout for all courses appended to the time table) this portion gives further specific details regarding the course.

Course No. : CS G554
Course Title : Distributed Data Systems
Instructor-in-Charge : Dr. Manik Gupta

Description :

Distributed File Systems - File System Models; Replication and Synchronization - Caching; Failure & Recovery; File System Security. Distributed Databases - Distributed Data Sources and Updates; Database Connectivity; Concurrency Control and Distribution mechanism; Distributed indexing schemes. Database security. Data on the Web - Web as a distributed data repository. Data Collection and Use Crawlers, Search Engines, and Indexing Schemes. Information Retrieval Techniques.

Data Exchange - Hierarchical Data Models, XML, and query languages. Semi-structured / Unstructured data -querying and synchronization.

Pervasive Data - Data distribution and access for non-computing devices, small computing devices, embedded computing devices and sensory devices.

Scope and Objective of the Course:

With the ever-growing pervasive data and the subsequent computational requirements, distributed systems are becoming more and more widespread. They are a vast and complex field of study in computer science. In this course, the focus is particularly upon distributed systems wrt data distribution and processing.

The course aims at familiarizing the students with the concepts of Distributed Data Systems which includes topics like

1. Distributed Databases
2. Distributed File Systems
3. Data on the Web with Web serving as a distributed data repository
4. Hierarchical data models for semi-structured/unstructured data

The objectives of this course are as follows:

1. To gain an understanding of how data distribution is planned, designed and implemented for distributed database design.
2. To understand challenges in distributed database query processing and optimization, transaction processing and concurrency control.
3. To be able to understand the working of distributed file systems.
4. To gain knowledge on semi-structured and unstructured data model with focus on XML data model.
5. To gain knowledge in design and implementation of distributed data systems in the context of relational, semi-structured and unstructured data models.
6. To gain knowledge about new paradigms of web data systems.



7. To gain hands on experience in both practical as well as design aspects of distributed data systems.

Textbooks:

T1: M. Tamer Ozsu and Patrick Valduriez. *Principles of Distributed Database Systems*, Second Edition

Reference books:

R1: Ramez Elmasri and Shamkant B. Navathe. *Fundamentals of database systems*, Seventh edition

R2: M. van Steen and A.S. Tanenbaum. *Distributed Systems*, Third edition

Course Plan:

Lecture No.	Learning objectives	Topics to be covered	Chapter in the textbook
1-5 (5)	To understand the need and fundamentals of distributed systems	Introduction, architectures, challenges and problem areas	R2-Ch.1 & 2
	To have a recap of the concepts related to database systems	Overview of RDBMS	T1-Ch.2
	To understand the concepts related to computer networks in context of distributed systems	Review of computer networks	T1-Ch.3
6-14 (9)	To have a good understanding of distributed DBMS (DDBMS) architecture and be able to design a distributed database by applying the concepts learnt related to fragmentation and allocation	Distributed DBMS architecture, DDBMS design, fragmentation, allocation	T1-Ch.1, 4 & 5 R1-Ch.23
15-21 (7)	To be able to understand query processing issues in distributed databases and query optimization strategies	DDS query processing, query decomposition, data localization and optimization	T1-Ch.7 - 9
22-30 (9)	To understand the challenges in transaction management, distributed concurrency control and reliability	Transaction management, concurrency control and DDBMS reliability	T1-Ch.10 - 12
31-33 (3)	To get an overview of various concepts related to distributed file systems and case studies	File system models	*Paper Readings



34-36 (3)	To understand and case studies related to NoSQL systems	Semi structured and unstructured data, Hierarchical data models, XML and query languages	R1-Ch.13, 24 *Paper Readings
37-42 (6)	To understand the concepts in web search and analysis	Web as a data repository, crawlers, search engines	R1-Ch.27 *Paper Readings

* There will be additional research paper readings for some of the topics and more information will be disseminated during the course.

Evaluation Scheme:

Component	Duration	Weightage (%)	Date & Time	Nature of Component
Test 1	0.5 Hrs	10%	To be announced	Open Book
Test 2	0.5 Hrs	10%	To be announced	Open Book
Test 3	0.5 Hrs	15%	To be announced	Open Book
Lab based Assignments, presentations and exercises	-	30%	To be announced	Open Book
Comprehensive exam	2 Hrs	35%	02/12 AN	Open Book

Make-up-Policy:

- No make-up requests for lab demos, research presentation and tests will be catered to.
- Prior permission of Dean, AUGSD is required to get make-up for the comprehensive exam.
- Instructor-in-charge's/Dean's decision in the matter of granting make-up would be final.

Course Notices:

All notices pertaining to this course will be displayed on the CMS Course webpage and Google classroom.

Chamber Consultation:

Thursday 12:00pm to 1:00pm

Academic Honesty and Integrity Policy:

Academic honesty and integrity are to be maintained by all the students throughout the semester and no type of academic dishonesty is acceptable. Students who are guilty of using unfair means will be reported for disciplinary action.



Lab No.	Task Details
1	Assignment 1 - Introduction to SQL, Assignment 2 - Django based 3-tier architecture implementation
2	Assignment 1- Introduction to SQL, Assignment 2- Django based 3-tier architecture implementation
3	Lab Exercises on Fragmentation, Assignment 2- Django based 3-tier architecture implementation
4	Lab Exercises on Fragmentation, Assignment 2- Django based 3-tier architecture implementation
5	Lab Exercises on Query optimization, Presentation 1 - DDBMS case study
6	Lab Exercises on Query optimization, Presentation 1 -DDBMS case study
7	Lab Exercises on Query optimization, Presentation 1 -DDBMS case study
8	Lab Exercises on Query optimization, Presentation 1 -DDBMS case study
9	Assignment 3 - DDBMS system implementation, Presentation 2 - Research paper
10	Assignment 3-DDBMS system implementation, Presentation 2 -Research paper
11	Assignment 3-DDBMS system implementation, Presentation 2 -Research paper
12	Assignment 3-DDBMS system implementation, Presentation 2 -Research paper
13	Assignment 3- DDBMS system implementation, Presentation 2 -Research paper
14	Assignment 3-DDBMS system implementation, Presentation 2 -Research paper

**Lab details for
CS G554
Distributed
Data Systems**



INSTRUCTOR-IN-CHARGE
CS G554

