CS670: Distrubuted Database Management Systems

Class meets on Mons. and Weds. from 4:00 to 5:15PM in room FH560L

Ph: 235 2366. Office Hrs: 12:30-1:00PM on MW. Office: RFH 550J

Vijay Kumar

Prerequisite: CS570 (Advanced database systems) or Instructor's permission.

Course description: This course deals with all aspects of Distributed Database Management Systems (DDBMS). It is assumed that the students enrolled in this course have a very good understanding of centralized systems (database and operating system), therefore, it begins from the point where centralized database system ends. In the past few years there have been significant advances in distributed architecture infrastructure and many new components have been added the basic infrastructure. This includes web, mobility, etc. This course does refer to mobility also but does not cover any topic. All aspects of mobility in information management is covered in CS572 (Mobile Computing).

It is obvious that operating systems, database systems, etc., are based on some type of distributed platform. Such strong presence of distributed discipline demands efficient management of application level and system level activities. This course, therefore, covers theoretical as well as applied aspects of distributed platform and analyzes a number of working systems (case studies) and discusses the future infrastructure.

Text Book: Principles of Distributed Database Systems. Ozsu and Valduriez. Prentice Hall.

The following reference books must also be consulted

1. Distributed Systems: Concept and Design. Coulouris, Dollimore, and Kindberg. AW.
2. Distributed Database Principles and Systems. Ceri and Pelagatti. McGraw Hill.
3. Recovery Mechanisms in Database Systems. Kumar and Hsu, Prentice Hall.
4. Concurrency Control and Recovery in Database Systems. Bernstein, Hadzilacos and Goodman, AW.

Research papers and Tutorials

In addition to these books, a number of research papers will be referred to. A complete list of research references will be provided in the class.

The course will cover nearly all chapters of the text book. A sample topic list is as follows:

1. **Architecture of distributed systems**: A detailed review of distributed system architecture (network operating system, distributed operating systems, etc.) will be presented leading to distributed database systems. This will then be categorized into (a) federated database systems, (b) multidatabase systems, and (c) Client/Server systems. This will help us to understand the general as well as unique problems (solved and yet to be solved) of these systems.
2. **Advanced transaction model**: For managing data processing on distributed platform the conventional transaction model needs some improvements. In this course we will discuss

some advanced transaction models suitable for different types of distributed database systems.

3. **Workflow**: It is a unit of business processing. From conventional viewpoint it is a set of tightly linked atomic processing units which requires special concurrency control and commit protocols. We will discuss existing ways of handling workflows.
4. **Query processing and Optimization:** On distributed systems a query may be fragmented for processing on multiple nodes. This give rise to the problem of query fragmentation and distribution which must be addressed for improving performance. This topic will be discussed in detail.
5. **Application distribution**: To support parallel and concurrent processing of transactions processing application have to be distributed. This gives rise to application recovery problem. This course will explore new ways of managing application recovery which is more complex than database recovery.
6. **Transaction management, commit protocol and database recovery**: These are system related issues. We will discuss commonly used schemes and advanced protocols for managing these activities.
7. **Buffer management**: Database maintains their own buffer for processing transactions. We will discuss the buffer architecture and buffer management schemes (replacement, allocation, etc.)

Course Management

The entire course work will be organized as follows:

Project: There will be one research project where each student has to solve a research problem and develop efficient solution. I emphasize that the research report must be of conference or journal quality. A list of useful projects (unsolved problem) will be presented to the class and students are free to select a topic from this list. A student is free to select a research topic out side of this list but the topic must be related to the course material and must be approved by the instructor before students begins his/her research.

Seminar: Each student must present a seminar on the topic of their research. They can select the same topic for seminar and research paper. Each seminar will be an hour long and 15 minutes for question and answer and the seminar participation is mandatory for each student of this course. These seminars will be presented near the end of the semester. In case of high enrollment two students may be assigned to a project and seminar.

Grading: The final grade will be assigned on the basis of the quality of the seminar, participation in the class discussion, and the quality of the research report.

Students are encouraged to join in the class discussion and present their thoughts and ideas on the all distributed system problems.