COP5711 – Parallel and Distributed Databases

Instructor: *Kien A. Hua*

Office: HEC 229   Email: kienhua@cs.ucf.edu

Course Description: Storage manager, implementation techniques for parallel DBMSs, distributed DBMS architectures, distributed database design, query processing, multidatabase systems.

Objectives: The purpose of this course is to prepare students for working in fields that involves using high-performance computing technology for data-intensive applications.

Outcomes:

- The student will have good understanding of database design and computation techniques for parallel and distributed database technology, not only at the algorithmic level but also in a system setting.

- The student will be able to read technical papers and write critiques to illustrate their ability to conduct independent study on related topics.

Topics to be covered: One of the major motivations behind the use of database systems is the desire to integrate the operational data of an enterprise and to provide centralized, thus controlled access to that data. The technology of computer networks, on the other hand, promotes a mode of work that goes against all centralization efforts. At first glance it might be difficult to understand how these two contrasting approaches can possibly be synthesized to produce a technology that is more powerful and more promising than either one alone. The key to this understanding is the realization that the most important objective of the database technology is integration, not centralization. It is important to realize that either one of these terms does not necessarily imply the other. It is possible to achieve integration without centralization, and that is what the distributed database technology attempts to achieve. In this course, we will discuss various such integration techniques. Another topic for this course is parallel database technology. For applications that require a system capable of sustaining trillions of operations per second on very large data sets (e.g., 'Big Data' and cloud database applications), parallel processing is the only solution.   We will examine techniques available for implementing such systems, and discuss the following topics:

- SQL Review
- DBMS Internal
- Parallel Architectures for DBMSs
- Data Placement Strategies
- Parallel Algorithms
- Parallel DBMS Implementation Techniques

- Distributed DBMS Architectures
- Distributed Database Design
- Distributed Query Processing
- Multidatabase Systems
- Peer-to-Peer Systems
- Mobile Computing
- Internet of Things

Prerequisite:	COP4710 or working knowledge of DBMSs.
Text:	A text is not required. Lecture slides are available at http://www.cs.ucf.edu/~kienhua/classes/
Class Time:	Tuesday, 6:00 - 9:50pm
Class Location:	ZOOM (virtual classroom)
Office Hours:	Email kienhua@cs.ucf.edu
Grading Policy:	Test 1: 30% Critical Reviews - 20% Test 2: 30% Project - 20%

Important Dates:

- TEST 1: June 9, 2020
- TEST 2: July 28, 2020
- Withdrawal deadline is Monday, June 29, 2020
- Class ends on July 28, 2019
- Holidays are 5/25/2020 (Memorial Day), 7/3/2020 (Independence Day).