COP5711 – Parallel and Distributed Databases

Instructor: *Kien A. Hua*
Office: HEC 229   Email: kienhua@cs.ucf.edu

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 exactly 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. For both topics, we will focus on how the systems work (i.e., system internal), rather than on how to use some commercial systems. We will discuss the following subjects:

- **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**
- **Location-based Applications**

Prerequisite:	COP4710 or working knowledge of DBMSs.
Class Notes:	Available at http://www.cs.ucf.edu/~kienhua/classes/
Class Time:	Tuesday & Thursday, 6:00-7:15pm, Room HEC-117
Office Hours:	Tuesday & Thursday, 5:00 ~ 6:00pm or by appointments
Grading Policy:	Test 1 (Parallel DBMS's) - 30% Critical Reviews - 20%
	Test 2 (Distributed DBMS's) - 30% Project - 20% (+ extra credit)

Important Dates:

- Withdrawal deadline is Wednesday, March 24, 2015
- Class ends on April 27, 2015
- Midterm exam: March 5, 2015
- Final exam: http://registrar.ucf.edu/exam/2015/spring
- Spring holidays are January 19, 2015 (Martin Luther King Jr. Day) and March 9-14, 2015 (Spring Break)