Csci5980: Storage Systems for Supporting Big Data Applications

*Basic Storage Technologies, Storage Systems, Data Management, New Storage Architecture for Big Data, AI & ML for Data Storage, Software Defined Storage and Software Defined Network, Hyperconverged Infrastructure*

Course Syllabus

**Class Meeting Time and Venue:**

       Monday and Wednesday 1:00 to 2:15 pm ; Vincent Hall Room 364

**Instructor**

**Professor David H.C. Du**  
Email: du@umn.edu

Office: Keller 4-225B   
Office Hours: Monday 3 to 4 pm and Wednesday 4 to 5 pm

**Course Description:**

 The Internet today has grown to an enormously large scale. Devices large and small are connected globally from anywhere on the earth. With the rapid advancement of technology, we now also have cheap and small devices with high computing power and large storage capacity. These devices are designed to improve our daily life by monitoring our environment, collecting critical data, and executing special instructions. These devices have gradually become a dominant part of our Internet. Many imaging, audio and video data are converted from analog to digital and digital data are generated at an alarming rate. As a result, unprecedented amount of data is available and generated daily. We can certainly say that we are in a big data era. In this course, we will examine the challenges and research trends in storage systems for supporting this new big data era. Several other major developments like cloud computing, mobile computing, new memory/storage technologies, AI and machine learning are certainly also creating big impact in this new era.

In this course, we will first introduce the new emerging memory/storage technologies. We will then examine the data storage requirements for big data applications. Finally, we will discuss the new memory/storage hierarchies that can potentially address the following challenges:

* How to store and access the desired information from this sea of data?
* How to manage these data?
* How to process these data (in-memory and big graph processing) ?
* How to preserve data for a very long time (100 years)?
* What are the emerging memory/storage technologies like SSD (Solid State Drives), Non-Volatile memory, SMR (Shingle Magnetic Recording Drives), Kinetic Drives, active storage devices, and OSD (Object Storage Devices)?
* What are the emerging memory/storage hierarchies?
* What are the storage systems for mobile computing, cloud computing and big data applications (Hadoop, MapReduce, SPARK, OpenStack, RamCloud, VM, Docker Container, etc.)?
* What is the emerging SDS (Software Defined Storage)?
* What are the future file systems?

The students in this class are required to read a set of papers. The class will focus on discussing the problems, the challenges and possible solutions defined in these papers. The goal is to develop thinking and solution techniques for solving big data problems via new and emerging storage systems.

This is an ***advanced graduate***course which is primarily targeted at MS/Ph.D graduate students *who would like to pursue research in storage systems, cloud computing, networking and related areas.*  Please note that this is *not* a regular lecture-based course, but *more a seminar- and project-oriented*course, with student presentations, classroom discussions as well as student research projects. More specifically, we will operate the course in *two "parallel"*tracks: In one track, we will read, study and discuss research papers on related areas. This track will consist of some presentations by the instructor, and by the students. Active participation in class discussion is required! The presenter functions primarily as the lead to facilitate discussion! In parallel, each student, possibly teamed with another student, will study and investigate a selected research topic of interest. With the help of the instructor, each student will identify one research problem he/she would like to study, write up a project proposal, make their case in front of the class; and throughout the course, make presentations on their "findings" and proposed solutions, and so forth. Amble time will be reserved for these purposes. Each student/team is required to submit a final project report in the format of a "workshop" paper by the end of the course. In addition to the presentation and term project, each student is expected to read every research paper, write up paper reviews, occasionally answer a few questions/solve problems related to the papers, participate in classroom discussion, and take part in the "peer review" process.

**Class Readings:**

We will provide handouts and recommend reading materials from recent journals/conference proceedings and other sources.   
A tentative course schedule (for the first few weeks) can be found in class website (with additional readings and other information continuously updated throughout the course).

**Coursework and Evaluation:**

*In addition to study the required reading materials and participating in class discussions*, all students enrolled in this course for credit will be required to complete the following tasks. They will be evaluated based on their performance of these tasks.

* **Paper Presentation:**each student will study a number of research papers (2-3) on a selected topic in details, prepare and make two presentations, and lead the classroom discussion.
* **Research Project:**Each student will be expected to conduct a "research project" to address one or two research issues from a list of selected topics. Depending on the number of the students in the class, it is possible for some of you to work as a team of 2 students. The "research topics" will be chosen in consultation with the instructor. A project might consist of doing an "in-depth analysis'' of one or two research topics and writing a summary paper; it might consist of delving much more deeply into a specific research area by performing some simulations/experiments to verify existing ideas/approaches; it might consist of proposing original ideas or conducting original work which extends the ideas or approaches of other researchers. The following procedure will be observed:
  1. By the fourth week, you should select a research topic in consultation with the instructor.
  2. By the sixth week, a "draft research proposal" (1-2 pages) must be submitted which describes the scope of the project, lists the issues to be addressed and outlines the approaches to be taken. Several recommended papers related to the project must also be provided.
  3. By the end of Week 8, a detailed project proposal and initial progress report (5 pages) must be submitted.
  4. By the last day of class, the final project report (10 page maximum) in the format of a workshop paper is due.
* **Peer Review:**Each student/team will also be responsible for "peer reviewing" the project of another student/team. This includes 1) reading the recommended papers, project proposal, progress report and the presentation outline of the other student/team, writing a review report and posing a list of suggested questions for "panel" discussions; 2) reviewing and evaluating the final project report of the other student/team.