CSci 5980: Storage Systems for Big Data

Time: Monday and Wednesday 4:00 to 5:15 pm
Location: Keller Hall Room 6-212
Spring 2017, 3 Credits
Instructor

David Hung-Chang Du
Email: du@cs.umn.edu
Office: EE/CS 4-225B for Office Hours

Phone: 612-6252560
Office Hours: Monday and Wednesday 1:00 to 2:00 pm
Teaching Assistants

There will be no TA for this class.
What is CSci 5980 about?

- Introductory to storage systems
- Fundamental principles and general concept
- What are the new developments in storage technologies and devices?
- Focus on both software and hardware aspects
- Familiar with techniques to support big data
Course Materials

• No text book
• A set of papers will be provided
• One to three papers will be assigned for each lecture
• Each student is expected reading the paper before coming to the class
• Interactive discussion will be the format for each lecture
Class Information

http://www-users.cselabs.umn.edu/classes/Spring-2017/csci5980/

Grades will be based on class attendance (40%) and final exam (60%)

Class attendance include showing up in each lecture, reading the paper assigned and actively involved with discussion
Covered Subjects

• Introduction
  – What is storage?
  – Software and Hardware I/O Stacks
  – What Are New?
    – Emerging Applications: Mobile/Sensor Networks, IOT, Big Data Analytics

• Emerging storage devices: HDD, flash memory based SSD, NVRAM, SMR, OSD, Tape Drives, Active Storage, and Kinetic Drives
• Storage System Architectures: Direct Attached, Storage Area Networks, Tiered Storage, Storage Hierarchies, Cold Storage, Archive Storage and Cloud Storage

• File Systems: Journaling and Special File Systems
Subjects (Cont.)

• Applications: Data Deduplication, Big Graphs, Checkpointing, Data Provenance and Long-term data preservation

• In-Memory Processing: Hadoop, Spark, RamCloud

• Cloud Storage: Key-Value Store, OpenStack, VM vs. Container (Docker)

• New Storage Trends: Software Defined Storage, Tiered Storage, Object Storage