EEL 4932 Massive Storage and I/O for Big Data Computing

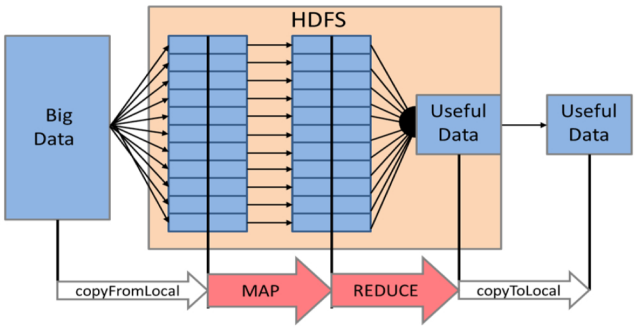
COURSE OVERVIEW:

This course provides a broad introduction to the fundamentals of massive file storage systems and I/O architectures in big data computing and its enabling systems infrastructure such as MapReduce and storage, with a focus on system architecture, file storage, distributed file systems and parallel file systems, programming models, application development and evaluation. Selected scientific applications will be used as case studies.

HADOOP FILE SYSTEM (HDFS): Massive data is split into

blocks and distributed across multiple nodes in the cluster. Each block is replicated multiple times.

MAPREDUCE: splits data sets into independent pieces to run the computations faster on parallel machines.



TOPICS:

- **PART I: Fundamentals of Computing**
- **PART II: Warehouse Scale Computers**
- **PART III: Hadoop and MapReduce**
- **PART IV: Storage systems**
- **PART V: In class discussion**

Visit http://www.csit-team.org/proposed-courses/ **for course syllabus**

ABOUT DR. JUN WANG

Associate professor Jun Wang joined Department of Electrical Engineering and Computer Science in University of Central Florida in 2006. Prior to that, he was a faculty in Computer Science and Engineering Department of University of Nebraska, Lincoln. He received his Ph.D. from University of Cincinnati in 2002.

Recently, he has won 2013 Dean's Research Professorship Award, Charles N. Millican Faculty Fellow 2010-2012, and University of Central Florida Research Incentive Award 2010. He has authored over 80 publications in premier journals.

UCF
COLLEGE OF ENGINEERING AND COMPUTER SCIENCE

UNIVERSITY OF CENTRAL FLORIDA

EEL 4932 Massive Storage and I/O for Big Data Computing

TENTATIVE COURSE OUTLINE: WEEKLY SCHEDULE

PART I: FUNDAMENTALS OF COMPUTING IN THE PAST AND PRESENT — WEEK 1-WEEK 4

1. Data intensive Scalable computing and cloud, and HPC
2. Quantitative principles of computing
3. Gustafson's Law
4. Power, energy efficiency, scalability

PART II: WAREHOUS SCALE COMPUTERS — WEEK 5-WEEK 6

1. Architecture design of warehouse scale computers
2. Evaluation of warehouse scale computers
3. Infrastructure and costs of warehouse scale computers

PART III: HADOOP AND MAPREDUCE — WEEK 7-WEEK 10

1. Introduction of Hadoop, such as how to launch MapReduce , Hadoop and write/read data to /from HDFS
2. Introduction about MapReduce flow
3. How to write a MaprReduce program using different programing language such as JAVA, #C, etc.

PART IV: INTRODUCTION OF SORAGE SYSTEMS — WEEK 11-WEEK 12

1. Hard drive, firmware, storage metrics, High performance file system such as parallel, Cluster, etc.

PART V: IN CLASS DISCUSSION OF VARIOUS RESEARCH TOPICS — WEEK 12-WEEK 15

1. ***Analytics:*** Decoupled Analytics for Shared storage systems (Fast'13)
2. ***Social network***: An analysis of Facebook Photo Caching (SOSP'13)
3. ***Geo-distributed***: Achieving Serializability with low latency in Geo-Distributed Storage Systems (SOSP'13)
4. ***Load Balancing***: How to Rebalance load in Optimizing MapReduce on Heterogeneous Clusters (Fast'13)
5. ***Cloud Storage***: Optimal Flash Provisioning for Cloud Storage Workloads (Fast'13)