COMP620

COMP620: Graduate Seminar in Distributed Computing - Big Data and Analytics Systems

Course Information (Spring 2015)

Instructor:	Prof. Faye Briggs, DH 2062	**Graduate TA:**	Deepak Majeti, DH 2069
Lectures:	Keck 107	**Lecture times:**	Fri 11:00 am - 11:50 am

Lecture Slides and Notes

	Lecture Name	Date	Slides	Notes
1	**Introduction to the course**	1/17/2015	PPT	
2	**Course outline and student presentation assignment**	1/23/2015	PPT	
3	**Big Data: Applications & Platform Architectures**	1/30/2015	PPT	
4	**Big Data and Analytics Systems: Computer System Architecture**	2/6/2015	PPT	
5	**More Applications of Big Data**	4/24/2015		

Schedule for Student Presentations

Note: The slides and related material for each topic will be provided. Hence, do not hesitate about the workload if you like a topic not in your domain.

	Topic	Student Name	Date	Presentation
1	**Distributed file systems and map-reduce as a tool for creating parallel algorithms that succeed on very large amounts of data**	**Yiting Xia**	2/13 /2015	PPT
2	**Similarity search, including the key techniques of min-hashing and locality sensitive hashing**	**Deepak Majeti**	2/20 /2015	PDF
3	**Data-stream processing and specialized algorithms for dealing with data that arrives so fast it must be processed immediately or lost**	**Wei-Cheng Xiao**	2/27 /2015	PPTX
4	**Anomaly Detection** ~~**Frequent-itemset mining, including association rules, market baskets, the A-Priori Algorithm and its improvements**~~	**Deepak Majeti**	3/13 /2015	PPT
5	**The technology of search engines, including Google's Page Rank, link-spam detection, and the hubs-and-authorities approach**	**Omid Pouya**	3/20 /2015	PPT1, PPT2
6	**Algorithms for clustering very large, high-dimensional datasets**	**Simbarashe Dzinamarira**	3/27 /2015	
7	**Two key problems for Web applications: managing advertising and recommendation systems**	**Lei Tang**	4/10 /2015	
8	**Algorithms for analyzing and mining the structure of very large graphs, especially social-network graphs**	**Shangyu Luo**	4/17 /2015	
	~~**Techniques for obtaining the important properties of a large dataset by dimensionality reduction, including singular-value decomposition and latent-semantic indexing**~~	~~**Zhipeng Wang**~~	~~4/24 /2015~~	

	~~Machine learning algorithms that can be applied to very large data, such as perceptrons, support vector machines, and gradient descent~~	~~Zhipeng Wang~~	~~4/30/2015~~	

~~R~~esources

Text Book: Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman

Systems Architecture for Big Data and Analytics : A Big Data Architecture for Large Scale Security Monitoring, by Samuel Marchal, Xiuyan Jiang, Radu State, Thomas Engel

Databases & Tools : Hadoop & HDFS, Hive, SPARK, Map-Reduce Google Big Table & GoogleFS, Google Cluster Experiences with MapReduce

Programming Approaches to Big Data Analytics : OpenMP, MPI, etc

Analytics Algorithms and Applications :

GraphX : Unifying Table and Graph Analytics, Joseph Gonzalez

Analytics for all : Challenges in analytics applications

Machine Learning Review : Machine Learning Foundation, By Jason Brownlee

Modeling and Detection Techniques for Counter-Terror Social Network Analysis and Intent Recognition, by Clifford Weinstein, William Campbell, Brian Delaney, Gerald O'Leary

Visualization Tools :

Cell Phone Mini Challenge Award : Intuitive Social Network Graphs Visual Analytics of Cell Phone Data using MobiVis and OntoVis, by Carlos D. Correa Tarik Crnovrsanin Christopher Muelder Zeqian Shen Ryan Armstrong James Shearer Kwan-Liu Ma