Rutgers, The State University of New Jersey
Ph.D. Program – IT Major
Special Topics in Information Systems: Big Data: Management, Analysis, and Applications
26:198:685
Spring 2014
Tuesday 2:30 – 5:20 pm, Newark

Instructor Prof. Nabil Adam
Office : 1029 1 Washington Park (Newark)
E-mail' : adam@cimic.rutgers.edu; Homepage : http://cimic.rutgers.edu/~adam

Purpose: The purpose of this course is to present advanced topics and delve into research challenges related to the area of "Big Data".

Data is growing at an exponential rate and recently, the term "Big Data" has been commonly used in the research community, industry, and government. Big data is typically characterized by not only volume but also variety (of various information sources), and velocity (of complex analytical processing).

Big data cuts across a wide spectrum of domains such as Web, social networks, cloud, mobile, sensor networks, and multimedia/streaming with several application verticals such as business, government, healthcare, transportation, biosciences, energy utilities, and scientific computing. At the data acquisition phase, there is a need to deal with reliable storage of the incoming mass streams of data together with the appropriate metadata. Some of these data, are unstructured, others such as tweets and blogs are "weakly" structured fragments of text, and yet others, such as images and videos although are structured for storage and display they are not structured for semantic content and search. A major challenge then is transforming such data in a structured format for later analysis.

Examples of big data systems that address the needs for structured and unstructured data include the NoSQL storage systems, Hadoop Map-Reduce, data analytics platforms, search and indexing platforms, and messaging infrastructures. These systems are still evolving.

The value of data grows exponentially when linked and integrated with other data. Given the large variety of information sources combined with the large volume of data to be handled and the velocity to handle these data, automated and scalable data integration is a challenging task. -- Automatic wrappers

The large volume of data combined with its complexity makes the scalability of data analysis algorithms an important challenge. Visualization of these complex relationships in a way that is intuitive for nontechnical subject matter experts an additional challenge.

Expected Work
1. Homework and In Class Practice Exercises addressing some recent related published work 30%. Student are required to upload their work to the Black Board (Solution to Homework Assignment and solution to the In Class Practice Exercises).
2. Research Paper/Project and Presentation 35%

      a. Either, write a comprehensive survey paper on one of the topics as agreed upon with the instructor.

      b. Or, work on a related development project.

3. Final Examinations 35%

Topics
1. Introduction
2. Application Areas
   a. Government
   b. Businesses
   c. Healthcare
   d. Environment
   e. Communities
3. Sources of Big Data
   a. Social media
   b. Smart devices
   c. Sensor network
   d. Cloud computing
   e. Mobile Computing
4. Big Data Applications
   a. Complex Big Data Applications in Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication, Science, and Engineering
   b. Big Data Analytics in Small Business Enterprises (SMEs)
   c. Big Data Analytics in Government, Public Sector and Society in General
   d. Real-life Case Studies of Value Creation through Big Data Analytics
5. Big Data Infrastructure
   a. Cloud/Grid/Stream Computing for Big Data
   b. Autonomic Computing and Cyber-infrastructure, System Architectures, Design and Deployment
   c. Energy-efficient Computing for Big Data
   d. New Programming Models for Big Data beyond Hadoop/MapReduce, STORM
6. Big Data Management
   a. Advanced database and Web Applications
   b. Data Provenance
   c. Interfaces to Database Systems and Analytics Software Systems
   d. Information Integration and Heterogeneous and Multi-structured Data Integration
   e. Spatiotemporal and Stream Data Management
   f. NoSQL database management system
   g. Database Management Challenges: Architecture, Storage, and Query
7. Big Data Search and Mining
   a. Social Web Search and Mining
   b. Data Acquisition, Integration, Cleaning, and Best Practices
   c. Algorithms and Systems for Big Data mining
   d. Big Data Search Architectures, Scalability and Efficiency
   e. Visualization Analytics for Big Data
   f. Large-scale Recommendation Systems and Social Media Systems

8. Big Data Security & Privacy

Reference

See most recent reference list posted on the BB. The list will be updated throughout the semester.