CS 63016/73016  BIG DATA ANALYTICS  3 Credit Hours

Instructor’s Name: Dr. Xiang Lian

Reference Books


Other Resources: Online Conference/Journal Papers from TODS, VLDBJ, TKDE, SIGMOD, PVLDB, ICDE, etc.

Course Content:
(Slashed with CS 73016) This course will cover a series of important Big-Data-related problems and their solutions. Specifically, we will introduce the characteristics and challenges of the Big Data, state-of-the-art computing paradigm sand platforms (e.g., MapReduce), big data programming tools (e.g., Hadoop and MongoDB), big data extraction and integration, big data storage, scalable indexing for big data, big graph processing, big data stream techniques and algorithms, big probabilistic data management, big data privacy, big data visualizations, and big data applications (e.g., spatial, finance, multimedia, medical, health, and social data).

Prerequisites or co-requisites: Graduate standing
Required, elective, or selected elective

Goals:
1. Understand the big data characteristics and challenges
2. Know the existing big data processing platforms/tools
3. Understand big data collection, integration and storage
4. Learn the big data indexing
5. Learn the basics of MapReduce paradigms
6. Learn various queries over big data
7. Learn the core techniques of processing big data
8. Understand different real applications and their techniques that involve big data

Outcomes:
1. Get familiar with big data characteristics and challenges
2. Be proficient with existing big data techniques/tools
3. Write a survey of existing big-data-related topics
4. Design and implement a research project on big data problems
5. Collaborate with team members to study the big data techniques
6. Present the project results to the class
Topics to be Covered:

1. Overview of Big Data
2. State-of-the-art computing paradigms/platforms
3. Big data programming tools (e.g., Hadoop, MongoDB, Spark, etc.)
4. Big data extraction and integration
5. Big data storage
6. Scalable big data indexing
7. Large-scale graph processing techniques
8. Big data stream techniques and algorithms
9. Large-scale probabilistic data analysis
10. Big data privacy
11. Big data visualizations
12. Problems in real applications of big spatial-temporal data (e.g., geographical databases)
   Problems in real applications of big financial data (e.g., time-series data)
13. Problems in real applications of big multimedia data (e.g., audios/videos)
14. Problems in real applications of big medical/health data
15. Problems in real applications of big social media data
16. Problems in real applications of big scientific data (e.g., bioinformatics data)

Abet Learning Outcomes:

- Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions.
- Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program’s discipline.
- Apply computer science theory and software development fundamentals to produce computing-based solutions.
- Communicate effectively in a variety of professional contexts.
- Recognize professional responsibilities and make informed judgments in computing practice based on legal and ethical principles.
- Function effectively as a member or leader of a team engaged in activities appropriate to the program’s discipline.