CIT650: Introduction to Big Data

(Tuesday 5:30 PM)

Dr. Tamer Arafa

SCHOOL OF
INFORMATION TECHNOLOGY
&COMPUTER SCIENCE

Our Vision

To be a world-class school recognized as one of the top information technology and computer science schools in the region for research, education, and entrepreneurship

Our Mission

The school is committed to preparing scientifically and professionally distinguished graduates in many information technology and computer science disciplines. It strives to: strongly contribute to society's prosperity, achieve sustainable development goals; and support the information technology industry through multidisciplinary scientific research, innovation and enhancement of entrepreneurial capabilities

Nile University
جامعة النيل

Course Description

- **Big Data Explosion:** Coined to express the surge in global digital data, Big Data originates from diverse sources and formats.
- **Universal Significance:** Big Data is a core theme in industries, research, and society, impacting sectors like automotive, finance, healthcare, and manufacturing.
- **Industry Advancements:** Industries benefit from faster data processing, with automotive, finance, healthcare, and manufacturing experiencing notable improvements.
- **Tech Boost:** Big Data's progress is powered by affordable, high-powered computing platforms, enabling fault-tolerant storage and processing in large clusters with thousands of processors and terabytes of memory.

Course Aim

- **Course Objectives:** This course aims to familiarize students with advanced principles and methods for managing and processing data effectively.
- **Data Handling Techniques:** Students will explore storage and processing techniques for various data types, including structured, semi-structured, and unstructured data.
- **Cutting-edge Topics:** The course will delve into the latest advancements in big data processing systems, covering areas such as
  - Batch processing
  - Stream processing.

Course Outcomes

On successful completion of this course, students should be able to:

- **Recognize Scalable Data Needs:** Understand the escalating demand for scalable data storage and processing in diverse domains.
- **Evaluate Solutions:** Assess advanced data management solutions, choosing systems for specific challenges.
- **Implement Cutting-edge Systems:** Apply state-of-the-art data processing for scalable solutions in diverse domains.
- **Performance Analysis:** Use qualitative and quantitative methods to analyze and compare system performance.
- **Build Data Pipelines:** Demonstrate skill in constructing complex data processing pipelines for diverse data types.

Course Topics

- Principles of Big Data
- Batch Processing Systems for Big Data
  - Hadoop
  - Spark
- Big SQL Systems
  - Hive
  - Impala
  - Spark Data Frames/SQL
- Big Stream Processing
  - Storm
  - Spark Streaming
  - Flink

Grade Distribution

3 Quizzes	15%
2 Assignments	15%
Midterm	20%
1 Project	20%
Final exam	30%

References

- Sherif Sakr and Mohamed Gaber. ”Large Scale and Big Data: Processing and Management”, CRC Press, 2014.
- Sherif Sakr. ”Big Data 2.0 Processing Systems”, Springer, 2016
- Albert Zomaya and Sherif Sakr. ”Handbook of Big Data Technologies”, Springer, 2017
- Sherif Sakr and Albert Zomaya. ”Encyclopedia of Big Data Technologies”, Springer, 2018
- Sherif Sakr et al. ”Large Scale Graph Processing Using Apache Giraph”, Springer, 2016