_____

Big Data Analytics

_____

Course Name: Big Data Analytics   Class: 70355

To be Held at: Main Campus   Room: MTC29

Course Instructor: Dr. Tariq Mahmood   Teaching Assistant: None

Lecture Timings: 830 am – 1115 pm (Every Tuesday)

Consultation Hours: 1130 am – 1230 pm (Tuesday)

_____

Scope: This course is centered on several dimensions of big data and its analysis. Along with theory, hands-on will be conducted mostly on the Docker platform. The plan is to introduce the basic concepts and definitions regarding big data along with several case studies. Then, we move to NoSQL technologies for solving big data problems in corporate sector, namely document, key value pairs, columnar stores and graph stores. We will conduct hands-on particularly with MongoDB (document), Redis and some other NoSQL offerings. We will also dive deep into Apache Hadoop and its ecosystem with Docker (Hive, HBase, Ambari, Cassandra etc.).

_____

Weekly Lecture Breakdown (3 hours/Week):

- Week 1:
    - Introduction to Big Data, NoSQL databases, Apache Hadoop Ecosystem
    - History, Evolution, Concepts, Definitions, Techniques, Big Data Landscapes, Columnar Warehouses
    - Cloud-based Solutions, Scaling Up, Caching, Scaling out
- Week 2:
    - Virtualization, Hypervisor, Hypervisor types, Containerization, Docker
    - Introduction to Docker Desktop
    - Hands-on with Docker Desktop - setting up
- Week 3:
    - Theory and Hands-on with Docker platform – History, evolution, architecture, process
    - Hands-on with Linux commands
    - Hands on with Linux Shell Scripting
- Week 4:
    - Theory and Hands-on with MongoDB on Docker platform – features, architecture, use cases, replication, cloud deployments, database functionalities, querying mechanism
- Week 5:
    - Theory and Hands-on with Redis on Docker platform – features, architecture, use cases, replication, cloud deployments, database functionalities, querying mechanism
- Week 6:
    - Theory and Hands-on: Hadoop on Docker – Hadoop architecture, HDFS, MapReduce concepts and exercises
- Week 7:

- o   Theory and Hands-on: Hive on Docker
- Week 8:
  - o   Theory and Hands-on: Hbase on Docker
- Week 9:
  - o   Theory and Hands-on: Neo4J on Docker
- Week 10:
  - o   Theory and Hands-on: Spark on Docker – Part I
- Week 11:
  - o   Theory and Hands-on: Spark on Docker – Part II
- Week 12:
  - o   Data Engineering and DevOps Pipelines – CI/CD - Theoretical (concepts, trends, techniques, usecases)
- Week 13:
  - o   Data Engineering and DevOps Pipelines – CI/CD - Setting up a pipeline on Docker
- Week 14:
  - o   Mid-Term Week (to be adjusted above)

_____

Grading Plan:

- Assignments: ~15%
- Quizzes: ~8%
- Mid-Term: 15%
- Final: 25%
- Project: ~35-40%

Text Books:

- The Guide to Big Data Analytics, Data Meer
- Big Data for Dummies, Wiley
- Big Data Analytics, TDWI EBook
- Mining Massive Datasets, Rajaraman, Leskovec, Ullman, Stanford University
- MongoDB Reference Manual (Online)
- RIAK Reference Manual (Online)
- Cassandra Reference Manual (Online)
- Hadoop: The Definitive Guide (OReilly)