MIE1628: Cloud-based Data Analytics

Prerequisites:

APS1070, MIE1624H, ECE1513H, and CSC2515 (or equivalent) are strongly recommended but not required.

Given the wide range of programming languages deployed in data analytics, students will use Python as the main programming language and Java as a second programming language to implement assignments in this course. An understanding of Python and Java is expected.

Course Description:

This course covers Big Data fundamentals including an overview of Hadoop MapReduce and Spark. Covers Cloud fundamentals and Big Data Analytics on Cloud-based platforms including an introduction to a specific Cloud platform such as Microsoft Azure, Amazon Web Services, or Google Cloud Platform along with common practices for this platform. Covers Cloud technologies to store and process structured, unstructured, and semi-structured data. Covers Cloud-based implementation of Real-time Analytics and Machine Learning.

Grading: Assignment/Exam	**Weight (%)**	**Due Date / Time**
Assignment 1	10	Jan 30 @ 24:00
Assignment 2	10	Feb 13 @ 24:00
Midterm	25	Feb 25
Assignment 3	10	Mar 6 @ 24:00
Assignment 4	10	Mar 20 @ 24:00
Assignment 5	10	Apr 03 @ 24:00
Final Exam	25	Apr 08

Assignment submissions will be online through *Github/Quercus*. It is the student's responsibility to verify that the assignments are submitted. Assignments submitted up to 48h late will be given a 20% penalty. Assignments that are submitted 48h late will incur a mark of zero.

Academic honesty:
Do not submit code that you have not written yourself. Students suspected of plagiarism on a project, midterm or exam will be referred to the department for formal discipline for breaches of the Student Code of Conduct.

Student responsibilities:
It is the student's responsibility to attend lectures and ensure assignments are submitted on time.

Preliminary schedule of lecture topics:

No.	Week	Lecture	Assignment
1	Jan 9	Course Overview, Hadoop Framework	Self-Study
2	Jan 16	Hadoop in Detail	Assignment 1 (MapReduce)
3	Jan 23	Spark Framework	Assignment 1 (MapReduce)
4	Jan 30	Spark in Detail/Databricks	Assignment 2 (Spark)
5	Feb 06	Azure Cloud Fundamentals	Assignment 2 (Spark)
6	**Feb 13**	**No Class - Family Day Week**	Assignment 3 (Cloud Fundamentals)
7	**Feb 20**	**No Class - Mid Term**	Self-Study
8	Feb 27	Azure Big Data Platform Overview and ETL process	Assignment 3 (Cloud Fundamentals)
9	Mar 06	Data warehousing in the cloud	Assignment 4 (Data Integration and SQL)
10	Mar 13	Azure SQL Database and Cosmos DB	Assignment 4 (Data Integration and SQL)
11	Mar 20	Machine Learning/ Real-Stream Analytics in cloud	Assignment 5 (Real-Stream Analytics and Machine Learning)
12	Mar 27	Revision using Big Data Architecture (End to End Use Case)	Assignment 5 (Real-Stream Analytics and Machine Learning)
13	**Apr 03**	**No Class- Final Exam**	Self-Study

Assignments:

Assignment 1: Based on KMeans Clustering using MapReduce
Assignment 2: Based on the Recommender system using Spark
Assignment 3: Based on the Cloud Data Platform
Assignment 4: Based on Data Orchestration and SQL in the cloud
Assignment 5: Working with sensor data using Real Stream Analytics and Machine Learning in the cloud