MSDS610 - Data Engineering: Syllabus

Instructor Information

Refer to Discussion Forum, Facilitator Introduction and Expectations

Course Title

MSDS610 - Data Engineering

Course Description

Presents techniques for designing, building, and managing information with relational databases, NoSQL databases, and big data infrastructure. Data Engineers are responsible for the creation and maintenance of the data infrastructure that facilitates the analysis of data by Data Scientists. Provides a hands-on experience running MongoDB, ProstgreSQL, Hadoop ecosystem and Spark.

Prerequisite Courses

MSDS600

Course Outcomes

Upon completion of this course, learners should be able to:

- Evaluate Data Engineering as a discipline of study and differentiate it from Data Science.
- Summarize cloud computing capabilities and compare cloud computing with on-site implementations.
- Utilize Linux and the command line to perform computing tasks and explain how Linux is used.
- Describe Hadoop's and Spark's role in big data and explain batch versus in-memory processing of big data.
- Describe NoSQL databases, explain the four types, and operate a MongoDB database.
- Summarize pros and cons of relational databases and SQL and implement a PostgreSQL database.
- Collect data from the web and store it in SQL and NoSQL databases.

Course Materials

Required Texts

Reading material is provided in the Worldclass online shell for both the classroom and online courses.

Required Resources

"From the Expert" presentations linked within each week and provided in Course Resources folder.

A PC that has no less than 8 Gig of RAM (16 Gig of RAM is preferred) and 100 GB of free hard drive space. (a PC with 4 Gig of RAM won't work)

Technology Tools

technical specifications

Access to CentOS Linux servers. This may be accomplished using remote access to services on the CC&IS Academic Research Network or using virtualization on your personal computer.

Pre-Assignment

See Course Assignments and Activities table below.

Course Assignments and Activities

Assignments for Classroom and Online Course		
Week	**Readings**	**Graded Assignments or Assessments (Percentage)**
1: Overview of Data Engineering; Basic Linux commands, Google Cloud Platform (GCP) and VM	*From the Expert	Lab Assignment – 2.5 pts Discussion – 2.5 pts Quiz – 3.75 pts

2: Overview of Hadoop. Install, Configure, and Test Hadoop on a 3 Node Cluster on GCP.	*From the Expert	Lab Assignment – 2.5 pts Discussion – 2.5 pts Quiz – 3.75 pts
3: Relational Databases. Use SQL & PostgreSQL.	*From the Expert	Lab Assignment – 2.5 pts Discussion – 2.5 pts Quiz – 3.75 pts
4: Data Architecture, Data Governance. NoSQL Databases using MongoDB.	*From the Expert	Lab Assignment – 2.5 pts Discussion – 2.5 pts Quiz – 3.75 pts
5: Python, Anaconda Python, and API data sources.	*From the Expert	Lab Assignment – 2.5 pts Discussion – 2.5 pts Quiz – 3.75 pts
6: Python and Pandas. Begin final Project.	*From the Expert	Lab Assignment – 2.5 pts Discussion – 2.5 pts Quiz – 3.75 pts
7: Spark. Project status report.	*From the Expert	Lab Assignment – 2.5 pts Discussion – 2.5 pts Quiz – 3.75 pts
8: Complete Project and Project report; finish outstanding labs.	*From the Expert	Lab Assignment – 2.5 pts Discussion – 2.5 pts Quiz – 3.75 pts Course Project – 30 pts

Summary of Assignments and Percentage Weight:

Assignments	Weighted Percentage
Weekly Lab assignments	20%
Weekly Discussions	20%
Course Project	30%
Weekly Quizzes	30%
TOTAL	**100 %**

Regis University Policies

Review the Regis University Policies on the Regis University website.

Attendance policy for in-person courses

Students taking in-person courses are expected to attend all classes for a course during the term. Missing classes may result in a failing grade or substantial grade penalties, at the discretion of the course instructor. Class absences should be discussed in advance with the course instructor.

OTHER INFORMATION

NOTE TO LEARNERS: On occasion, the course facilitator may, at his or her discretion, alter the Learning Activities shown in this Syllabus. The alteration of Learning Activities may not, in any way, change the Learner Outcomes or the grading scale for this course as contained in this syllabus. Examples of circumstances that could justify alterations in Learning Activities could include number of learners in the course; compelling current events; special facilitator experience or expertise; or unanticipated disruptions to class session schedule.