Cloud Computing for Data Analysis

DSBA 6190-U91 (CRN 27657)

Instructor Information

Instructor	Email	Office Location & Hours
Colby T. Ford, Ph.D.	colby.ford@uncc.edu	Bioinformatics 302A (By Appt.)

General Information

Description

Introduction to the basic principles of cloud computing for data intensive applications. Covers a broad range of technologies and solutions from data platform architecture to data analytics. Focuses on the scalable deployment of cloud resources and the integration between individual services. Topics covered will include data architecture such as SQL databases and data lakes, parallel computing using clusters such as Hadoop and Spark, machine learning using common classification, clustering, and regression algorithms, and deep learning using GPU-based infrastructure.

Restriction(s): Familiarity with R and/or Python, SQL, Unix, Data Structures, Machine Learning Algorithms, and Statistics; good programming skills and a solid mathematical background.

Learning Outcomes:

1. Understand the benefits of cloud-based architecture
2. Architect end-to-end solutions based on user/organizational requirements
3. Recognize the differences in data platform options on-premise versus in the cloud
4. Discuss the cloud and on-premise machine learning approaches and the benefits therein

Course Materials

Cloud Data Design, Orchestration, and Management Using Microsoft Azure – [Link]

(Other materials will be distributed throughout the course as needed.)

Grading:

The final course grade will be determined by the student's total number of points earned in the class out of the total possible points.

Exercise	Points		Final Grading Ranges	
Data Platform Lab	100		≥900/1000pts	A
Machine Learning Lab	150		800-899/1000pts	B
Midterm Exam	200		700-799/1000pts	C
Parallel Computing Lab	150		<700/1000pts	D or Inc.
Deep Learning Lab	100		Academic Dishonesty	F
Final Exam	300			
Total	1000			

Tentative Course Schedule

Date	Section	Topic(s)	Lab
January 9th	Intro	- Introduction to the course - Review syllabus	

January 16th (REMOTE CLASS)	Data Platform	- Structured Data Stores
		    o Databases
		    o Data Warehouses
			Begin Data Platform Lab

January 23rd	Data Platform	- Azure Data Factory
		- Unstructured Data Stores
		    o NoSQL Databases
		    o Data Lakes	

January 30th	Machine Learning	- Intro to Machine Learning - Intro to IDEs - Regression	Data Platform Lab Due Begin Machine Learning Lab
February 6th	Machine Learning	- Regression - Classification - Clustering	
February 13th	Machine Learning	- Cross Validation - Parameter Tuning - Intro to Training Parallelization	
February 20th	Review	- Review Data Platform - Review Machine Learning	
February 27th	Midterm Exam		Machine Learning Lab Due
March 6th	*NO CLASS*		
March 13th	Parallel Computing	- Intro to Parallel Computing - MPI/SNOW-based Clusters	Begin Parallel Computing Lab
March 20th	Parallel Computing	- Kubernetes and Docker	
March 27th	Parallel Computing	- Hadoop and Spark - Streaming Data (IoT) Overview	
April 3rd	Deep Learning	- Intro to Deep Learning - Intro to Neural Networks	Parallel Computing Lab Due Begin Deep Learning Lab
April 10th	Deep Learning	- Training Deep Learning Models with GPUs	
April 17th	Deep Learning	- Deep Learning as a Service	
April 24th	Review	- Review Parallel Computing - Review Deep Learning	Deep Learning Lab Due
May 1st	*NO CLASS*		
May 8th-ish	Final Exam		

Academic Integrity and Honesty:
Students are required to read and abide by the Code of Student Academic Integrity available from Dean of Students Office. This code forbids cheating, fabrication or falsification of information, multiple submissions of academic work, plagiarism (including viewing others work without instructor permission), abuse of academic materials, and complicity of academic dishonesty. Violations of the Code of Student Academic Integrity, including plagiarism, result in disciplinary action as provided by the Code.

Civility:
We are concerned with a positive learning experience. This course strives to create an inclusive academic climate in which the dignity of all individuals is respected and maintained. We value diversity that is beneficial to both employers and society at large. Students are encouraged to actively and appropriately share their views in class discussions.

Inclement Weather:
University Policy Statement #13 states the University is open unless the Chancellor announces that the University is closed. The inclement weather hotline number to call is 704-687-1900. In the event of inclement weather, check your e-mail, and Canvas. The instructor will post a message on Canvas, and through e-mail. The instructor will use their best judgment as to whether class should be held.

Disability:
UNC Charlotte is committed to access to education. If you have a disability and need academic accommodations, please provide a letter of accommodation from Disability Services early in the semester. For more information on accommodations, contact the Office of Disability Services at 704-687-0040 or visit their office in Fretwell 230.

Withdrawal:
The University policy on Course Withdrawal allows students a limited number of opportunities available to withdraw from courses. There are financial and academic consequences that may result from course withdrawal. If a student is concerned about his / her ability to succeed in this course, it is important to make an appointment to speak with the instructor as soon as possible.

Syllabus Revision:
The instructor may modify the class schedule and syllabus throughout the semester. Changes will appear on Canvas. Students are responsible for refreshing their syllabus.

E-Mail Communication:
Students are responsible for *all* announcements made in class and on the class online resources. Students should check the online class resources throughout the semester. The Instructor and Teaching Assistants send occasional e-mails with important information. We send this information to the student's UNC Charlotte e-mail address listed on Banner system. If a student is not checking his / her UNC Charlotte e-mail address (ex. userName@uncc.edu) please be sure to access this e-mail and check it regularly during this course.