

Course Name: DATA821 – Data Architecture**Term Year: 2017**

DATA821: Data Architecture

Academic Requirements Fulfilled: Analytics Certificate

March 2017 through June 2017 (8 weeks)

Scott A. Valcourt

Scott.Valcourt@unh.edu

603-862-4489 (office), 603-380-2860 (cell/text)

Slack: @svalcourt

Course Overview

This course exposes students to the current and changing landscape of data architecture. It begins with relational database design and querying using structured query language (SQL) and continues to high volume computing platforms such as Hadoop and Spark, NoSQL tools and also issues of Big Data streaming. Current case base data will be used as examples.

Course Description

In this class, students will learn the foundations of databases and large datasets: upon completion, students will be able to explore out-of-ram datasets through traditional SQL databases and NoSQL databases. Students will also be introduced to distributed computing. All learning objectives are achieved through active application of these techniques to world datasets. Prereq: DATA 800; DATA 820.

Learning Objectives

- Students will use and apply systems theory and concepts to construct Entity Relationship Models for up to 3rd Normal Form databases based upon the Context Model.
- Students will learn the core commands of Structured Query Language (SQL) and apply those skills in building queries using sample databases.
- Students will work with a basic web server (Apache) to demonstrate at least one method towards integrating a RDBMS with a web-based interface.
- Students will be introduced to Hadoop and Spark as two tools among many that could be used to process data extraction and utilization processing.
- The concept of Data-as-a-Service (DaaS) will be introduced using examples of large public data sets that students will need to access and process using Hadoop and Spark.
- The introduction of Hive as an interface tool will also be introduced.
- Students will work with a non-relational database management system (TinyDB) to observe the differences between relational and non-relational databases. A discussion on the benefits and issues in the use of either type of database will be considered.
- Students will build a NoSQL database to collect data as well as a parallel SQL database and compare the relative ease and challenges the exercise offers.

Course Structure

Canvas is the learning management tool we use for this course. The course is online and asynchronous and organized by weeks. The overall course navigation includes:

Course Name: DATA821 – Data Architecture
Term Year: 2017

<i>Course Navigation</i>	<i>Description</i>
Home Page	Landing Page for the course. Links to individual modules.
Syllabus	The syllabus, course schedule and other key class documents are located here.
Modules	This area contains the learning Modules. The majority of course content, activities, assignments and participation is located in this area. See below for structure
Resources (optional)	Supplemental Resources.

<i>Module Structure</i>	
Lecture	The primary lecture(s) or media and key concepts.
Resources (or Review)	Assigned readings or media for review.
Participation	Every module will have a participation component. This could take the shape of a discussion forum, wiki, blog, collaborate event, etc.
Assignments	Activities and assignments related to the module.

Course Schedule

<i>Week</i>	<i>Date(approx.)</i>	<i>Topics Covered</i>
One	March 20 – March 24	Requirements Gathering and Entity-Relationship Modeling
Two	March 27 – March 31	Schema Development, Normalization, and Metadata
Three	April 3 – April 7	SQL Programming including SELECT, JOIN
Four	April 10 – April 14	Integration of RDBMS with Web and other Interfaces
Five	April 17 – April 21	Hadoop Theory and Usage
Six	April 24 –	Integration of Hadoop

Course Name: DATA821 – Data Architecture
Term Year: 2017

	April 28	with RDBMS and Data Storage
Seven	May 1 – May 5	Data Streaming Using Spark
Eight	May 8 – May 12	NoSQL Concepts using TinyDB

Grades

The course grade will be based on the following breakdown of the individual grade opportunities throughout the online course period.

Item	% or points	Requirements
Quizzes	25%	Each module will have at least one (1) quiz to take online. When to take the quiz will be at the discretion of the student.
Homework Assignments	30%	Each module will have at least one (1) homework assignment to complete and submit for grading. The due date of the assignment is at the discretion of the student.
Mid-term Exam	15 %	The mid-term exam is a summation exam of the modules contained in the first four weeks of the course. This exam will be taken by the student when coursework is completed, but prior to the end of the formal course period.
Class Participation	10 %	Class participation will be judged by the engagement that the student has in online assignments and posting requirements.
Final Exam	20%	The final exam is a summation of the entire course (all eight weeks). Students will determine when they will be ready to complete the online final exam, but not after the end date of the formal course window.

Assignment Details

Each assignment will contain a complete set of instructions and details on how to complete the assignment, as well as the objectives that should be achieved as a result of completing the assignment.

Policy on Late Submissions and Quizzes

My intention is to provide as much time as is necessary to complete the assignments and review the materials in the course. Should you have difficulty remaining current or understanding the materials presented, please let me know as soon as possible so that we can work out any arrangements to address the issues.

Class Participation

Course Name: DATA821 – Data Architecture
Term Year: 2017

Weekly participation is required. For discussion and blog postings, the participation will be graded based on the following criteria and values. Review this carefully.

Quality of postings

Below are desired attributes for a posting.

- Ability to synthesize the main concepts from instructor, course content, external resources and class community.
- Use of proper grammar.
- Ideas are organized, persuasive and elevate the overall dialogue.
- Opinions are substantiated.
- Demonstration of critical or creative thinking.
- Evidence of preparation.

Quantity of postings

Per weekly module forum, I would expect to see one original posting based on the reading for that module and at least one response posting to another student's post. Try to spread your postings over the duration of the week to allow for responses from classmates to be entered. Posting 5 times in one sitting is not acceptable. When you contribute over several days, you are more able to synthesize other perspectives and contribute to an evolving discussion. A good rule of thumb is to post early in the week after you have reviewed the related materials and reading. Allow others to respond and then post again after the dialogue has developed.

	0-60 Points	61-80 Points	81-100 Points
Promptness and frequency	Does not post; posts marginally.	Postings are late and random.	Postings are early, frequent and between postings.
Grammar	Grammar has not been checked.	Poor grammar or typos are frequent in postings.	Proper grammar is consistently applied.
Relevance	Postings have no relevance on the topic or assignment.	Some postings are pertinent to topic and assignment.	Postings are consistently connected to the topic and assignment and reflect student's preparation.
Expression & Organization of Ideas	Ideas and or opinions are not organized.	Some postings convey clear ideas and opinions. Other postings are lacking in organization or original expression.	Contributions are clearly articulated; reflect understanding of topic, substantive insights and persuasive logic.
Building Community	Does not participate in learning	Irregularly demonstrates initiative in adding to the dialogue and creating	Demonstrates consistent initiative in building community through

Course Name: DATA821 – Data Architecture

Term Year: 2017

	community.	community.	discussion. Demonstrates 'listening' to peers and elevates the overall quality of the discussion.
--	------------	------------	---

Etiquette

- Divergent opinions. Academic debate and differences are embraced in higher education and the forums in this course. Be mindful and respectful of how you articulate a difference or divergent opinion.
- 'I agree' statements. Unsubstantiated 'I agree' posts will not count to final participation grade without articulated rationale to support opinion.
- Off-topic postings. Discussions occasional veer off-topic. This is normal. These posts will not count and students are asked to stay on-topic. However, related topical discussions are encouraged, especially topics that are current and relevant to the subject matter of the module or overall course.
- Long responses. Grades will be influenced by an ability to demonstrate an understanding of the topic or question and on one's ability to be concise.

Student to Instructor Communication Expectations

My Schedule

I will be active in the Canvas class area daily, Monday through Friday. I login early in the morning and in the evening. If you post a question for me in a Forum, anticipate a response within 24-hours or sooner. On Saturday and Sunday, I may not login at a regular time. If you post late on Friday or anytime Saturday, I might not respond until Sunday evening. Sunday I have blocked out 2 hours for course work.

How to Reach Me

Questions related to assignments or learning should first be asked, if possible, in the respective discussion forums or blog. The use of email is reserved for questions of a private nature and or if you would like to schedule virtual office hours or send me an email and I will respond promptly. My contact information is as follows:

Email: Scott.Valcourt@unh.edu – use title header of [DATA821] to help the mail system flag your request quicker.

Phone: 603-862-4489 (office), 603-380-2860 (cell)

Physical office hours are M-F 9am-5pm, though calling or texting ahead will help to determine if I am in the office that day or offsite

Technical Requirements and Technical Support

See website listings for current recommendations and requirements related to this course - <http://unh.edu/eunh/technical-requirements> For technical assistance please call (603) 862-2431 or fill out an online support form at <https://itsupport.unh.edu/onlinelearning/>

University Disability Accommodations

Course Name: DATA821 – Data Architecture**Term Year: 2017**

The University is committed to providing students with documented disabilities equal access to all university programs and facilities. If you think you have a disability requiring accommodations, you must register with Disability Services for Students (DSS). Contact DSS at (603) 862-2607 or disability.office@unh.edu. If you have received Accommodation Letters for this course from DSS, please provide me with that information privately in my office so that we can review those accommodations.

Academic Honesty

Students are required to abide by the UNH Academic Honesty policy located in the [Student Rights, Rules, and Responsibilities Handbook](#).

Plagiarism of any type may be grounds for receiving an “F” in an assignment or an “F” in the overall course. Plagiarism is defined as “the unattributed use of the ideas, evidence, or words of another person, or the conveying the false impression that the arguments and writing in a paper are your own.” (UNH Academic Honesty Policy, 09.3) Incidents are reported to the school dean and may be grounds for further action. If you have questions about proper citation refer to your department’s writing guidelines. You can contact me at any time on this issue. Additional resources are located below:

<http://libraryguides.unh.edu/unhmcitingsources>

<http://www.library.unh.edu/reference/citation.shtml>

Note: This syllabus is subject to change. Students will be promptly notified of any changes.