Course Number	CS 504-004 - The course objective is to organize your thinking about the principles related to all things data in a way that you can approach data science problems with a structured problem solving methodology that is domain general, but applicable to specific problem areas. You will be able to perform six key data analysis tasks by the end of the course (framing, clustering, connecting, relating, measuring, and comparing). These skills do require any background in mathematics or statistics.	Instructor Background: Dr. Conroy works as a Data Scientist in the US Intelligence Community. He has degrees from GWU, the University of Southern California, and Virginia Tech. He is a former Marine helicopter pilot and worked for a variety of organizations. His technical background is in c/c++/java and currently supports development is geospatial related technologies including Cython/Python/R/ArcGis and other languages.	Teaching Philosophy - This is a concepts course. I am a cognitive scientist and believe that you construct knowledge and understanding through your engagement with the real world. For this reason we will engage in a variety of modeling activities in class. All the tests and exams are online and are T/F questions based on a controlled vocabulary of terms and relationships between terms.
Course Title	Principles of Data Management and Mining		
Location	Arlington: Arlington: Founders Hall 466		
Instructor	Arthur Conroy, PhD		
Instructor Email	aconroy2@gmu.edu		
Alt Email	conroyat@vt.edu		
Day of Week/Time	Thursday (430pm - 710pm)		
Textbook (Optional)	Data Science from Scratch: First Principles with Python by Joel Grus Paperback $26.57		
Course Description	Techniques to store, manage, and use data including databases, relational model, schemas, queries and transactions. On Line Transaction Processing, Data Warehousing, star schema, On Line Analytical Processing. MOLAP, HOLAP, and hybrid systems. Overview of Data Mining principles, models, supervised and unsupervised learning, pattern finding. Massively parallel architectures and Hadoop.	Hours of Lecture or Seminar per week: 3 This course cannot be taken for credit by students of the MS CS, MS ISA, MS SWE, CS PhD or IT PhD programs. Prerequisites: Graduate Standing.	You will be reading seminal paper across a number of disciplines during the course which will be accessible via Blackboard

Class	Class Topic	Day of Week	Notes
1	Course Overview - Administration, Rules of the Road, Success Formula, Student Expectations, Course Outcomes	Thursday, January 26, 2017	First Day of Class
2	Conceptual Foundations and Interrelationships - Data, Big Data, Data Science, Data Analysis, Predictive Analytics	Thursday, February 02, 2017	
3	Top Down Conceptual, Logical, and Physical Modeling	Thursday, February 09, 2017	
4	Bottom Up Conceptual, Logical, and Physical Modeling	Thursday, February 16, 2017	
5	Deep Learning, AI, and the Internet of Things	Thursday, February 23, 2017	
6	Procedural Foundations and Interrelationships - Data, Big Data, Data Science, Data Analysis, Predictive Analytics	Thursday, March 02, 2017	Test - Conceptual Knowledge, Project Assignment
7	Sourcing Data - Data Wrangling	Thursday, March 09, 2017	
8	Transforming Data - Extract Transform Loading	Thursday, March 16, 2017	Spring Recess
9	Loading Data - Streaming, Near Real Time, Batch	Thursday, March 23, 2017	
10	Full Stack Development (Models - Views - Controllers & Frameworks)	Thursday, March 30, 2017	
11	Declarative Foundations and Interrelationships - Data, Big Data, Data Science, Data Analysis, Predictive Analytics	Thursday, April 06, 2017	Test - Procedural Knowledge
12	The Open Source Ecosystem - languages, Tools, Vendors, and Technologies	Thursday, April 13, 2017	
13	Algorithms/Data Mining/Predictive Analytics	Thursday, April 20, 2017	
14	Emerging Technologies - SQL/NoSQL/In-Memory/EDW/Parallelism	Thursday, April 27, 2017	Test - Declarative Knowledge
15	Last Day of Class - Course Conclusion	Thursday, May 04, 2017	Projects Due and Final Exam