NYU SCHOOL OF PROFESSIONAL STUDIES

Special Topics: Managing Big Data

MASY1-GC 5000 | 103 | Spring 2024 | 1/23/2024 - 4/30/2024 | 3 Credit
Modality: In-person
Course Site URL: https://brightspace.nyu.edu/

General Course Information

Name/Title: Dr. Tom Schmidt, Professor, Computer Science
NYU Email: tms493@nyu.edu
Class Meeting Schedule: 1/23/2024 - 4/30/2024 | Tuesday | 06:20pm -- 08:55pm
Class Location: 7 East 12th, Room 124
Office Hours: Before Class, arrange at least one day in advance. Will do Zoom Tu/Thu 2-4PM by pre-arrangement.

Description

This seminar will enhance curriculum by identification, analysis, and application of special topics pertinent to the Management and Systems degree. The specific titles and content of each seminar will change to reflect emerging areas of interest, which can only be determined at the time of offering. The course may be used to satisfy the elective degree requirement. Applicability to specific concentrations will be noted in the course schedule and is at the department's discretion.

Prerequisites

1240 – Information Technology and Data Analytics

Learning Outcomes

At the conclusion of this course, students will be able to:

- Categorize a data structure using the four V's of Big Data (volume, velocity, variety, veracity)
- Select the appropriate architectural components and programming models used for scalable big data operational systems and big data analysis
- Analyze data projects for opportunities to apply big data and data science methodologies
- Create actionable insights and business value from massive amounts of seemingly unrelated data sets
- Analyze the types of analytics required for a project

Communication Methods

In addition to in-class participation, students will be submitting assignments, discussion responses, and final project materials via Brightspace. For personal topics and questions, credit students must use their NYU email to communicate. Non-degree students do not have NYU email addresses. Brightspace course-mail supports student privacy and FERPA guidelines.
All emails will be answered within 48 hours. For urgent matters, students can call and text the faculty.

Students have the opportunity to add their pronouns, as well as the pronunciation of their names, into Albert. Students can have this information displayed to faculty in Albert, Brightspace, and other NYU systems. Students can also opt out of having their pronouns viewed by their instructors.

https://www.nyu.edu/students/student-information-and-resources/registration-records-and-graduation/forms-policies-procedures/change-of-student-information/pronouns-and-name-pronunciation.html

Structure | Method | Modality

This course is In-person and will meet once a week on Tuesday. Brightspace is the learning management system we will use. Zoom is the online instruction platform used at NYU.
This special topic course will address the challenges and opportunities related to Big Data.

- What exactly is “big data” and where does it come from?
- How is big data gathered, stored, searched, shared, analyzed, and visualized?
- How can organizations create value from Big Data?

The course will include lectures and discussions. Students are expected to have read the assignments, and to actively participate during class discussions. Assignments will be provided and collected via Brightspace before/after each class meeting. In addition, each team of students will present one class session on a topic related to Big Data, and one management-level analysis. Assignments will be mostly team and final course project (group of 5) will be a technical demonstration case study of a topic pertaining to managing big data. The course project requires clear documentation of contributions towards the project.

Expectations

Learning Environment

As graduate students, you are expected to conduct yourselves in a professional manner and engage and collaborate with your classmates. SPS classrooms are diverse and include students who range in age, culture, learning styles, and levels of professional experience. To maintain an inclusive environment that ensures all students can equally participate with and learn from each other, as well as receive feedback and instruction from faculty during group discussions in the classroom, all course-based discussions and group projects should occur in a language that is shared among all participants.

Participation

You are integral to the learning experience in this class. Be prepared to actively contribute to class activities, group discussions, and work outside of class.

Assignments and Deadlines

- Weekly Readings (Case Studies, Readings)
  - The Cases are due on the specific day as noted in Brightspace and in the course outline.
  - Example: “Data Science: The Sexiest Job of the Century” is due Week 3
  - Assignments will be posted in Brightspace at least a week before they are due.

- Late assignments may receive half grade maximum for that assignment.
- Assignments submitted later than a week won't be counted.

Case Study Related Assignments:
You are required to read the case study assigned: In addition to the posted responses, a digital presentation (4-10 slides) summarizing the main points and answering the provided case study assignment questions with appropriate audio narration corresponding to the slides. In-Class Case Study Presentations is the oral in person presentation of the case study assignment.

Course Technology Use
We will utilize multiple technologies to achieve the course goals. I expect you to use technology in ways that enhance the learning environment for all students. All class sessions require use of technology (e.g., laptop, computer lab) for learning purposes.

Feedback and Viewing Grades
I will provide timely meaningful feedback on all your work via our course site in NYU Brightspace. You can access your grades on the course site Gradebook.

Attendance
I expect you to attend all class sessions. Attendance will be taken into consideration when determining your final grade.

Excused absences are granted in cases of documented serious illness, family emergency, religious observance, or civic obligation. In the case of religious observance or civic obligation, this should be reported in advance. Unexcused absences from sessions may have a negative impact on a student's final grade. Students are responsible for assignments given during any absence.
Each unexcused absence or being late may result in a student's grade being lowered by a fraction of a grade. A student who has three unexcused absences may earn a Fail grade.

Refer to the SPS Policies and Procedures page for additional information about attendance.

Textbooks and Course Materials
Students can purchase these items through the NYU Bookstore.

Required: McCoy, Scott. *Python for Data Analysis, 1st Edition*. 2021, Murach Associates, ISBN: 978-1-943872-76-3, $59.50

Recommended: Jake Vanderplas**, *Python Data Science Handbook***
Mayer-Schönberger, Viktor and Cukier, Kenneth. Big Data: A Revolution That Will Transform How We Live, Work, and Think. New York: Eamon Dolan/Houghton Mifflin, 2013. Hardcover.
Tufte, Edward R. The Visual Display of Quantitative Information. Cheshire, Conn. (Box 430, Cheshire 06410): Graphics, 1983. Hardcover.

Grading | Assessment

Technical Assignments 20% (individual grade)
- Will include R, Python, and AWS assignments, graded individually

Analysis Assignments 20% (team grades)
- Each Team will analyze 5 Big Data Case Studies and debate two; each team member will submit an individual evaluation of the case study

Participation 10%
- Graded individually

Quizzes 20%
- Offered at the start of class, based off reading in Brightspace
- Graded individually

Final Project 30% (team grade)
- Each team will present a final project demonstrating its integration of the five areas presented in adherence to the final project rubric

DESCRIPTION	**PERCENTAGE**
Analysis Assignments (6)	20%
Quizzes (best 9 of 12)	20%
Participation	10%
Technical Assignments (10)	20%
Final Project	30%
TOTAL POSSIBLE	100%

See the "Grades" section of Academic Policies for the complete grading policy, including the letter grade conversion, and the criteria for a grade of incomplete, taking a course on a pass/fail basis, and withdrawing from a course.

Course Outline
Start/End Dates: 1/23/2024 - 4/30/2024 | Tuesday
Time: 06:20pm -- 08:55pm
No Class Date(s): Spring Break - Tuesday, March 19, 2024
Special Notes:

Session 1 - 01/23/24
Topic Description:
- *Overview of the course and expectations of the class.*
- *Review the high-level concepts of where, what, why, and how Big Data and Business*
- *Discussion about course project and operation of the course*

Assignments:
- *On Brightspace*

Session 2 – 01/30/24

Topic description *– Overview of Big Data Analytics*

- *Discussion about Data Science, Scientists, and other roles that have impact in the Big*
- *Overview of concepts regarding fundamental analytical algorithms and how they leverage Big Data*
- *R Workshop: Exercise of R, utilizing parallel computing libraries and emulating distributed platform*

Assignments:

Reading: Data Science: The Sexiest Job of the Century

- *Assignments on Brightspace*

Session 3 – 02/06/24

Topic description *– Traditional Databases: Data Warehouse, Data Virtualization:*

- *Overview of Data Warehouses in Enterprises*
- *Overview of Data Virtualization*
- *How Data warehouses and Data Virtualization relates to Big Data environments and Business*
- *R Workshop*

Assignments:

- *Case Study Debate Due and presented by Team 1 and Team 2*

Session 4 – 02/13/24

Topic description – *Overview of how Data Visualizations, Analytics, and Big Data converge*

- *The business importance of Data Visualizations and its limitations*
- *1D, 2D, 3D, 4D Data visualizations, Linear and Non-Linear*
- *Visualizing Descriptive Analytics*
- *Visualizing Analytics*
- *Hands-on: Visualizing Data with Tableau*

Assignments:

- *Tell the Story of Your Data*
- *Others on Brightspace*

Session 5 – 02/20/24

Topic description – *Data Blending, ETL*

- *Overview of Data Blending and how it differs from ETL processes*
- *Advantages and Disadvantages of Data Blending analytics tools.*
- *Python Workshop*

Assignments:

*Project Proposal Due**

- *Case Study Debate Due and presented by Team 3 and Team 4*

Session 6 – 02/27/24

Topic description *– Structured and Unstructured Data*

- *Overview of Unstructured data*
- *How and why is it unstructured?*
- *Tools to structure the data in relation to Big Data*
- *Hands-on: Setting up AWS for loading Structured Data*

Assignments:

- *Create A Database Server on AWS and Load Data into it; Connect to it*

Session 7 – 03/05/24

Topic description *– NoSQL Databases: MongoDB, Couch DB*

- *Overview of non-columnar in-memory database architectures.*
- *Business and application advantages and disadvantages of NoSQL databases*
- *Does it make sense to use with Big Data?*
- *Case Study Debate Due and presented by Team 5 and Team 1*

Assignments:

-

Session 8 – 03/12/24

Topic description *– Cloud-based Solutions, MongoDB*

- *Overview of Cloud based Architecture in relation to Big Data*
- *The business advantageous and disadvantages of using a cloud-based solution.*
- *Cloud based solutions for big data management and analytics*
- *Hands-on: Setting Up RapidMiner*

Assignments:

- *Demonstrate MongoDB setup on BrightSpace*

Session 9 – 03/26/24

Topic description *– Database Appliances*

- *Overview of ," Out of the Box "database solutions*
- *Capabilities and Limitations of Database Appliances*
- *Business aspects of implementing Database Appliances.*

Assignments:

- *Case Study Debate Due and presented by Team 2 and Team 3*
- *Other assignments on BrightSpace*

Session 10 – 04/02/24

Topic description *– Machine Learning 1: RapidMiner*

- *The capabilities and limitations of Distributed Databases*
- *Business efficiency of Distributed Databases*
- *Hands-on: Scraping Data and Machine Learning with Rapidminer and Python*

Assignments:

- *Install, setup, and operate a RapidMiner learning environment. Integrate*

- *Other assignments on BrightSpace*

Session 11 – 04/09/24
Topic description *– Data Design Thinking: Design by Data*
- *Power of predictive consumer data analytics*
- *How to build a billion-dollar asset leveraging data?*
- *Positive financial impact of data design thinking*

Assignments:
- *Case Study Debate Due and presented by Team 4 and Team 5*

Session 12 – 04/16/24
Topic description *– Large Data Analytics*
- *Hands-on: Using Python and Amazon SageMaker for Analytics on large data sets*

Assignments:
- *Hands-on: Large Learning Models and Big Data: ChatGPT?*

Session 13 – 04/23/24
Topic description *– **Into the Metaverse***
- *VR, AR, and XR in Data Visualization*

Assignments:

Session 14 – 04/30/24
Topic description *– FINAL COURSE Project Presentations*
-

Assignments:
- *Present Projects*

Turn in Team Evaluation

NOTES:

The syllabus may be modified to better meet the needs of students and to achieve the learning outcomes.

The School of Professional Studies (SPS) and its faculty celebrate and are committed to inclusion, diversity, belonging, equity, and accessibility (IDBEA), and seek to embody the IDBEA values. The School of Professional Studies (SPS), its faculty, staff, and students are committed to creating a mutually respectful and safe environment (*from the SPS IDBEA Committee*).

New York University School of Professional Studies Policies

1. Policies - You are responsible for reading, understanding, and complying with University Policies and Guidelines, NYU SPS Policies and Procedures, and Student Affairs and Reporting.

2. Learning/Academic Accommodations - New York University is committed to providing equal educational opportunity and participation for students who disclose their dis/ability to the Moses Center for Student Accessibility. If you are interested in applying for academic accommodations, contact the Moses Center as early as possible in the semester. If you already receive accommodations through the Moses Center, request your accommodation letters through the Moses Center Portal as soon as possible (mosescsa@nyu.edu | 212-998-4980).

3. Health and Wellness - To access the University's extensive health and mental health resources, contact the NYU Wellness Exchange. You can call its private hotline (212-443-9999), available 24 hours a day, seven days a week, to reach out to a professional who can help to address day-to-day challenges as well as other health-related concerns.

4. Student Support Resources - There are a range of resources at SPS and NYU to support your learning and professional growth. For a complete list of resources and services available to SPS students, visit the NYU SPS Office of Student Affairs site.

5. Religious Observance - As a nonsectarian, inclusive institution, NYU policy permits members of any religious group to absent themselves from classes without penalty when required for compliance with their religious obligations. Refer to the University Calendar Policy on Religious Holidays for the complete policy.

6. Academic Integrity and Plagiarism - You are expected to be honest and ethical in all academic work. Moreover, you are expected to demonstrate how what you have learned incorporates an understanding of the research and expertise of scholars and other appropriate experts; and thus recognizing others' published work or teachings—whether that of authors, lecturers, or one's peers—is a required practice in all academic projects.

Plagiarism involves borrowing or using information from other sources without proper and full credit. You are subject to disciplinary actions for the following offenses which include but are not limited to cheating, plagiarism, forgery or unauthorized use of documents, and false form of identification

Turnitin, an originality detection service in NYU Brightspace, may be used in this course to check your work for plagiarism.

Read more about academic integrity policies at the NYU School of Professional Studies on the Academic Policies for NYU SPS Students page.

7. Use of Third-Party Tools - During this class, you may be required to use non-NYU apps/platforms/software as a part of course studies, and thus, will be required to agree to the “Terms of Use” (TOU) associated with such apps/platforms/software.

These services may require you to create an account but you can use a pseudonym (which may not identify you to the public community, but which may still identify you by IP address to the company and companies with whom it shares data).

You should carefully read those terms of use regarding the impact on your privacy rights and intellectual property rights. If you have any questions regarding those terms of use or the impact on the class, you are encouraged to ask the instructor prior to the add/drop deadline.