BIA 678 A & EMT 678 A
Spring 2023
(Tuesday | 6:30 PM – 9 PM)

School	School of Business
Course Title	Big Data Technologies
Program(s)	BI&A & EMT
Instruction	Venu Guntupalli, vguntup1@stevens.edu

Description

This 3-course credit will focus on utilizing classic techniques from Business Intelligence & Analysis (BI&A), along with new tools and processes to deal with the volume, velocity, and variety associated with Big Data. As students enter the workforce, a significant percentage will be directly involved with Big Data either as technologists, managers, or users. This course will build on their understanding of the basic concepts of BI&A to provide them with the background to succeed in the evolving data-centric world, not only from the point of view of the technologies required, but in terms of management, governance, and organization.

Prerequisites

Students taking the course will be expected to have some background in areas such as multivariate statistics, data mining, data management, and programming – Python, SQL and PySpark will be used extensively on vendor cloud platforms (ex: Databricks, AWS, GCP) to create data pipelines and machine learning models.

Course Outcomes

1. Understand and discuss what Big Data is, and how it differs from traditional approaches to BI&A
2. Plan and use the primary tools associated with Big Data in creating systems to take advantage of big data.
3. Extract knowledge and intelligence from datasets which exhibit high volume, velocity, and/or variety.
4. Plan and execute a project that includes the use of at least one Big Data dataset.
5. Understand and discuss the meta issues around big data such as governance, security, privacy, and OAM&P.
6. Understand and be able to execute analyses oriented to streaming data.
7. Have a framework with which to understand new advances in the field; distinguish hype from reality.
8. Understand and discuss organizational issues related to Big Data.

Assignment and Grading

Assignment Type	Assignments Per Semester	Total Points	% of Final Grade
Writing (Individual)	10	100	10%
Programming (Individual)	10	100	20%
Mid-term (Individual)	1	50	25%
Finals Proposal (Team)	1	10	5%
Finals (Individual)	1	10	5%
Finals (Team)	1	50	30%
Participation (Individual)	Present for no less than 12 lectures Engagement		5%
Total	**24**	**320**	**100%**

Writing Assignment: Students are expected to read /two papers relevant to topics covered in the class and submit a write up summarizing the paper, including key takeaways.
Programming Assignment: Programming assignments will include coding in Python, SQL and PySpark
Mid-term: Mid-term typically includes 50 multiple-choice questions to be taken in class. Questionnaire covers topics and programming concepts covered through the mid-term week.
Finals Proposal: All teams should submit proposal covering the details of the final project. Teams will receive feedback to align with objectives of the final project and any changes recommended should be confirmed before receiving grade
Finals – Individual: Finals will include presentations by all members of the team, clearly detailing contributions by all individuals. All members of the team must code and contribute to the final report.
Finals – Team: Team presentation and a detailed final report.
Participation: Engagement with instructor and peers is encouraged to further learning. Students are expected to attend all lectures in person.

Grading Scheme

Follows the Graduate Grading Scheme at Stevens.

Grade	Range
A	100% to 94.0%
A-	< 94.0% to 90.0%
B+	< 90.0% to 87.0%
B	< 87.0% to 84.0%
B-	< 84.0% to 80.0%
C+	< 80.0% to 77.0%
C	< 77.0% to 70.0%
F	< 70.0% to 0.0%

Textbook(s) or References

1. Content from lectures – slides and code notebooks – in Canvas shell
2. Case studies and papers posted in Canvas shell

Syllabus & Schedule

Week	Date	Notes	Lecture	Assignment	Assignment Due
1	1/24		Introduction I		
2	1/31		Introduction II		
3	2/7	Attendance mandatory starting this week	Distributed Programming - MapReduce	Writing A1 Programming A1	2/14
4	2/14	Databricks CE Spark setup	Distributed Programming - Spark I	Writing A2 Programming A2	2/21
5	2/21		Distributed Programming - Spark II	Writing A3 Programming A3	2/28
6	2/28		Data Stream Management I	Writing A4 Programming A4	3/7
7	3/7	Team formation due	Data Stream Management II	Writing A5 Programming A5	3/14

	3/14	Spring Recess	-	-	
9	3/21	Mid-term			
10	3/28	Confirm AWS/GCP access	Database Management I	Writing A6 Programming A6	4/4
11	4/4	Final project proposal due	Database Management II	Writing A7 Programming A7	4/11
12	4/11	Survey open (subject to change)	Recommender Systems	Writing A8 Programming A8	4/18
13	4/18		Ensembles - Boosting	Writing A9 Programming A9	4/25
14	4/25		Analytics & Governance I	Writing A10 Programming A10	5/2
15	5/2	Last day of instruction	Analytics & Governance II		
16	5/9	Finals			
17	5/16	Finals			5/16