CIS 8395 The Big Data Analytics Experience

Contents

Prerequisites: MSIS Big Data Analytics concentration and prior approval of the program director.

Catalog Description:

This course serves as the capstone course for the Big Data Management and Analytics concentration in the MS IS program and is intended to integrate the material that the students have learned in various courses in the concenration.

Detailed Course Description:

This project provides an opportunity for students to apply the knowledge and skills acquired from all courses in the Big Data Analytics concentration in the MSIS program. The project will be designed to perform work on specific topics relevant to the concentration. Students, individually or in teams, will develop a project plan with milestones like project charter, deliverables etc. Upon completion of the course, students will present both oral and written reports on their project.

Course Objectives:

This course is designed to provide the student experience in working on a project that will help assimilate all the experiences gained in the various courses in the Big Data Management and Analytics concentration. In doing this, the student will demonstrate his/her ability to:

- Understand how the various data components are being applied in global corporations.
- How to create business value using data
- Importance of Data Quality
- Why Data engineering and Data Governance critical for Big data solutions.
- How to apply advanced machine learnings and AI on big data.
- Prepare and present both written and oral reports

Syllabus for Big Data Experience

Students will get the experience of how big data analytics departments function in large global corporations. They will get a view of how the solutions are utilized, to improve revenue or reduce operating cost. Students will be forming groups and will be working on big data analytical projects utilizing large amounts of data. The solutions will be focused primarily on emerging technologies such as AI (Image, NLP, Voice), Recommendation engines, Forecasting, Classification, and Clustering. Students will be encouraged to use advanced solutions such as Generative Adversarial Network and Reinforcement Learning.

Day 1

Project – Forming groups, explaining the project, reviewing previous students work and providing guidance on project selection.

Theory – Students will learn how the data supply chain is implemented in global corporations and we will examine how large global corporations source, clean, and store data.

Data Strategy – We will explore how the data office creates data strategy in major organizations.

Data Sourcing – Review the importance of MSA, GDPR, CCPA, Third Party Agreement (TPA), and responsibilities between IT/business teams. Experience related to legal issues which arise due to TPA will be shared during this session.

Data Cleaning – Discuss in detail how the data is transformed in the ETL layer and how the data stewards play a key role.

Day 2

Project – Following up on team progress and reviewing the selected project.

Data Governance – Discuss the importance of MDM (Master Data Management) and its impact to data quality. In this session students will get to know about data catalog, data linage, and how these tools are critical for self-service data analytics solutions.

Data Engineering – In this section we will deep dive into the importance of data engineering, data reference architecture, different cloud technologies (AWS, Azure, Google).

Orchestration Layer – Students will understand how data acquisition pipelines are created and the benefits of the orchestration layer. The professor will be sharing the experience of issues faced in the absence of the orchestration layer.

ETL/ELT – Examine when to use extract transform and load vs extract load and transform. Discuss about the common tools used for ETL/ELT in the market. (Talend, Informatica, Databricks).

Day 3/ Day 4 - Project proposal presentation

Day 5

Storage – Explore the benefits/limitations of different data storage solutions. (Databases, Data Warehouses, Data Lakes, Data Mart)

Streaming data – Streaming data is increasing with the implementation of IOT and social media solutions. Students will learn how to handle streaming data for analytics.

Unstructured data – Unstructured data is key for AI solutions, this section we will discover how to handle image and voice data.

Extraction layer – In this section students will learn how to create an extraction layer for data science/ Tableau/MicroStrategy/Power BI.

Discuss regarding 1st paper due before Day 6.

Day 6.

Visualization layer - We will discuss in detail how major corporations generate and handle thousands of reports and explore how 10,000+ users using visualization tools are being managed in a global organization.

Self Service analytics trend – Investigate how global corporations are moving towards self-service analytics and how the data is being democratized within the organization. Students will also understand the semantic layer and its importance in regard to self-service analytics adoption.

Augmented analytics - Students will learn how global organizations are using augmented analytics. We will further analyze how augmented analytics AI is helping build visualization and helping identify the correlations/trends/outliers in the data.

Advanced analytics – We will examine various used cases implemented in organization using AI (Image, NLP, Voice), Recommendation engines, Forecasting, Classification, and Clustering. Students will understand emerging technologies such as Generative Adversarial Network, Reinforcement Learning and Transfer Learning.

Analytics by functionality – Marketing Analytics, Supply Chain analytics, Manufacturing analytics, HR analytics, Legal analytics, Finance analytics.

Day 7/8 – **Final Presentation of Projects** – Feedback provided on their presentation and quality of the projects.

Final Paper due within one week from last class (date will be announced during the program).

Method of Instruction:

Students will be grouped into teams. There will be interim project milestones and deliverables during the course to ensure that each project is progressing on schedule. On each due date, each student / team will submit a written report of the total project to date, and give a class presentation that updates the class on their progress.

After obtaining the scope of the audit project, the student /team will arrange a presentation of their scope definition, to make sure the team has a correct and complete understanding of project requirements ("Reality Check".)

At the end of the course, and after approval by the Course Director, the student /team will arrange a final presentation, in which they will present their deliverables. The final report will integrate the work performed during the project with all appropriate documentation.

Grades:

To earn a project grade of "A", a student /team must successfully accomplish all the requiremens, plus provide comprehensive documentation. Individual student grades will be allocated according to their contribution. If a team then the individual's contribution to the project will be determined by the team peer evaluations. However, a student may not get a grade higher than "one step above" the project grade earned by the team. For example, if the project grade is B, then the highest grade will be B+ for individuals on that project team.

FACTOR weights in grading:

- Deliverables:
  - Class Participation – 10%,
  - Peer Review – 5%
  - Proposal Presentation – 15%
  - Paper on Data source, Data Cleaning, ETL, Data Storage – 20%
  - Final Presentation - 20%,
  - Final Paper – 30%

Grade Distribution

Grade	**A+**	**A**	**A-**	**B+**	**B**	**B-**	**C+**	**C**	**C-**	**D**	**F**
Total Points	97.0 - 100	93.0 - 96.9	90.0- 92.9	87.0- 89.9	83.0- 86.9	80.0- 82.9	77.0- 79.9	73.0- 76.9	70.0- 72.9	60.0 - 69.9	Below 60.0

Policies:

General Class Policies

Student work submitted in fulfillment of course requirements and any student activity recorded is deemed to be granted in the public domain (copyright-free) for the purposes of use as instructional or research material or for examples of student work in future courses.

Students are expected to attend all classes and group meetings, except when precluded by emergencies, religious holidays or bona fide extenuating circumstances.

Students who, for non-academic reasons beyond their control, are unable to meet the full requirements of the course should notify the instructor. Incompletes may be given if a student has ONE AND ONLY ONE outstanding assignment.

Please see http://www.gsu.edu/es/20399.html for details regarding withdrawals.

Spirited team participation is encouraged and informed discussion in the team is expected. Unless specifically stated by the instructor, all exams and assignments are to be completed by the student alone.

Within group collaboration is allowed on project work. Collaboration between project groups will be considered cheating unless specifically allowed by an instructor.

Work copied from the Internet without a proper reference will be considered plagiarism and is subject to disciplinary action as delineated in the Student Handbook.

Any non-authorized collaboration will be considered cheating and the student(s) involved will have an Academic Dishonesty charge completed by the instructor and placed on file in the Deans office and the CIS Department. All instructors regardless of the type of assignment will apply this Academic Dishonesty policy equally to all students. See excerpt from the Student Handbook below:

Academic Honesty

(Abstracted from GSUs *Student Handbook* Student Code of Conduct Policy on Academic Honesty and Procedures for Resolving Matters of Academic Honesty - http://www2.gsu.edu/~wwwdos/codeofconduct.html .)

As members of the academic community, students are expected to recognize and uphold standards of intellectual and academic integrity. The University assumes as a basic and minimum standard of conduct in academic matters that students be honest and that they submit for credit only the products of their own efforts. Both the ideals of scholarship and the need for fairness require that all dishonest work be rejected as a basis for academic credit. They also require that students refrain from any and all forms of dishonorable or unethical conduct related to their academic work.

Students are expected to discuss with faculty the expectations regarding course assignments and standards of conduct. Here are some examples and definitions that clarify the standards by which academic honesty and academically honorable conduct are judged at GSU.

Plagiarism. Plagiarism is presenting another persons work as ones own. Plagiarism includes any paraphrasing or summarizing of the works of another person without acknowledgment, including the submitting of another students work as ones own. Plagiarism frequently involves a failure to acknowledge in the text, notes, or footnotes the quotation of the paragraphs, sentences, or even a few phrases written or spoken by someone else. The submission of research or completed papers or projects by someone else is plagiarism, as is the unacknowledged use of research sources gathered by someone else when that use is specifically forbidden by the faculty member. Failure to indicate the extent and nature of ones reliance on other sources is also a form of plagiarism. Failure to indicate the extent and nature of ones reliance on other sources is also a form of plagiarism. Any work, in whole or part, taken

from the internet or other computer based resource without properly referencing the source (for example, the URL) is considered plagiarism. A complete reference is required in order that all parties may locate and view the original source. Finally, there may be forms of plagiarism that are unique to an individual discipline or course, examples of which should be provided in advance by the faculty member. The student is responsible for understanding the legitimate use of sources, the appropriate ways of acknowledging academic, scholarly or creative indebtedness, and the consequences of violating this responsibility.

Cheating on Examinations. Cheating on examinations involves giving or receiving unauthorized help before, during, or after an examination. Examples of unauthorized help include the use of notes, texts, or crib sheets during an examination (unless specifically approved by the faculty member), or sharing information with another student during an examination (unless specifically approved by the faculty member). Other examples include intentionally allowing another student to view ones own examination and collaboration before or after an examination if such collaboration is specifically forbidden by the faculty member.

Unauthorized Collaboration. Submission for academic credit of a work product, or a part thereof, represented as its being ones own effort, which has been developed in substantial collaboration with assistance from another person or source, or computer honesty. It is also a violation of academic honesty knowingly to provide such assistance. Collaborative work specifically authorized by a faculty member is allowed.

Legal Requirements: Some project assignments may require legal agreements with the respective clients, and some projects may not require legal agreements. In cases where legal agreements are necessary, legal documents are available for student use. All students are expected to be familiar with the terms of the legal requirement documents furnished by the University.

All student work presented for course credit will identify the sources of information used, and any sources quoted *verbatim* will be further identified by including the material in quotations. Any material included in work submitted that is copied from other sources without giving credit for the source will result in a grade of F for the course.