CISC 5950 - Big Data Programming
Department of Computer and Information Science
Luigi Patruno, Fall 2017

Class Times: Wednesday, 7:55 - 10:10 PM in LL 514

Instructor: Luigi Patruno

Email: lpatruno@fordham.edu

Office Hours: Wednesday 6:55 - 7:55 pm, 3rd Floor Lounge. Please email me to schedule additional office hours!

Text: Big Data: Principles and best practices of scalable realtime data systems by Nathan Marz and James Warren.

Course Website: https://lpatruno.github.io/bigdata-fall17/

Course Description: Services like social networks, web analytics, and intelligent e-commerce often need to manage data at a scale too big for a traditional database. Complexity increases with scale and demand, and handling Big Data is not as simple as just adding more relational database servers. Big Data systems use many machines working in parallel to store and process data, which introduces fundamental challenges unfamiliar to most developers. This course will teach you to build these systems using an architecture that takes advantage of clustered hardware along with new tools designed specifically to capture and analyze web-scale data. We will discuss the theory of Big Data systems and how to implement these in practice.

Objectives: The goal of this course is to learn how to architect data systems when managing data that stretch the limits of relational database management systems. In order to do this, we will first discuss the traditional web service architecture and how this architecture fails to meet the needs of today's web services like social networks and e-commerce. After motivating these problems, we will discuss how to design architectures that can handle the complexity and scale of big data services. Along the way, we will introduce the theory and tools being used to solve these problems in industry today.

Attendance and Class Participation: It is important to attend and be prepared for every class. For this course, being prepared means completing the weekly reading assignments. I encourage you to take notes and write down questions while reading the chapter. The classes will be collaborative and discussion based in order facilitate and foster ideas. If you are going to miss a class, please let me know beforehand.

Academic Honesty: All work produced in this course should be your own unless it is specifically stated that you may work with others. You may discuss the homework problems with other students generally, but may not provide complete solutions to one another; copying of homework solutions is always unacceptable and will be considered a violation of Fordham's academic integrity policy. Violations of this policy will be handled in accordance with university policy which can include automatic failure of the assignment and/or failure of the course.

Grading: The percentages given below are guidelines for both the student and the instructor and minor changes may be made during the course. Students will be informed promptly of any such changes.

Homework & Quizzes	15%
Midterm Exam	30%
Final Project	40%
Participation	15%

Projects: The course will include a group final project where you will be asked to present a tutorial of a big data software system. Groups will be asked to research a big data technology and discuss the problem that the tool is solving. This project will also include a programming component i.e. a tutorial to the class of how the tool works. Note that students will be responsible for all aspects of the project including downloading and configuring the necessary software. I encourage you all to get started on this project as soon as possible. Ask me questions!