CISC 5950 Big Data Programming Spring 2019

**Description:** This course covers Apache Hadoop and Spark technologies and their ecosystems in the context of mining big data. It provides students both theoretical background and hands-on computing techniques in big data analytics and its applications. The students will learn how to collect, query, and analyze data. Topics include Hadoop core technologies (HDFS, MapReduce, Yarn), Spark Streaming, MLlib, Clustering, and Spark SQL. Scala programming language will be taught as part of the Spark component.

**Instructor:** Dr. Ying Mao (JMH 328A / LL610H)

**Email:** ymao41 at fordham.edu

**Office Hours:** Tuesday 4:00 p.m. – 5:30 p.m. (other times by appointment).

**Textbook:** No textbook required. The reading assignments will be posted on the course website.

**Course Web Resources:** https://yingmao.github.io/cisc5950/

**Grading Policy:**

- Programming Labs (4) ---- 20%
- Projects (2) ---- 30%
- Midterm ---- 20%
- Final ---- 30%

- A ( >= 90 )
- B ( >= 80 )
- C ( >= 70 )
- D ( >= 60 )
- F ( < 60 )

**Policy:** Attendance is strongly encouraged and required.

**Notes:**

- **Late submission of labs/projects:** 20% reduction each day.
- Dispute on grading must be resolved within two weeks after receiving your score.
Tentative Course Schedule

Week 1  Introduction and Overview of Big Data Systems  
Week 2  Hadoop Distributed File Systems (HDFS)  
Week 3  MapReduce Framework  
Week 4  Resource Management in the cluster (YARN)  
Week 5  Apache Scala Basic  
Week 6  No class due to Monday schedule  
Week 7  Apache Scala Advances  
Week 8  Resilient Distributed Datasets (RDD)  
Week 9  Midterm  
Week 10  Spring Break  
Week 11  Apache Spark  
Week 12  Apache Spark SQL  
Week 13  Data analytics on Hadoop / Spark  
Week 14  Machine learning on Hadoop / Spark  
Week 15  Spark Streaming  
Week 16  Other Hadoop/Spark ecosystem components  
Week 17  Final

Additional Remarks

- 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. For more information, please refer to the Academic Integrity website.

- Makeup Exam
There will be no make-up exams given after the exam date. If you know in advance that you will have to miss an exam, you must check with me (in advance) to avoid getting a zero for that exam. In case of illness on an exam date, please contact me as soon as possible, so that appropriate arrangement can be made.