DS -644 Introduction to Big Data

Syllabus - Spring 2025 | Section - 002,004

Required Background:

Programming Skills

Java, Python, or C/C++ in Linux

Prerequisite Courses

- CS 610: Data Structures and Algorithms
- Or permission of instructor

Textbooks: Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph. By David Loshin, Elsevier, August 23, 2013.

Textbooks and reference books -



Attendance: You are expected to attend all the classes. Participation is highly encouraged to make the class more interactive. In general, students who attend class regularly perform much better than those who come only occasionally. If you miss one class, be sure to consult one of your classmates about the content of the lecture and use canvas to get notes, exercises, assignments, deadlines, and announcements.

Students who have special/emergency reasons to reschedule exams must apply and get approval of testing accommodation from the Office of Accessibility Resources and Services before the exams. https://www.njit.edu/accessibility/requesting-testing-accommodations

Course Overview and Learning Outcomes:

This course provides an in-depth coverage of various topics in big data from data generation, storage, management, transfer, to analytics, with focus on the state-of-the-art technologies, tools, architectures, and systems that constitute big-data computing solutions in high-performance networks. Real-life big-data applications and workflows in various domains (particularly in the sciences) are introduced as use cases to illustrate the development, deployment, and execution of a wide spectrum of emerging big-data solutions. When you have completed this course, you should be familiar with big data tools, techniques, and systems, and be able to analyze/solve big data problems.

Week	Торіс			
1	Introduction to Big Data Analytics, 4Vs			
2	Trends of Computing for Big Data			
	High-performance Computing (Supercomputers and Clusters)			
	Grid Computing			
	Continuum Computing: from Edge to Cloud			
	Mobile Computing			
3,4	Big Data Overview			
	Drivers of Big Data			
	Big Data Attributes and Data Structures			
	Big Data Ecosystem			
	Big Data Use Cases			
	Introduction to Hadoop, HDFS			
5	Map Reduce and Yarn, Yarn Schedulers, Discussion			
6	Review & Mid Term1			
7	Big Data Tools, Techniques, and Systems			
	HBase, and NoSQL (Document Store, Graph DB, etc.)			
8	Oozie, Tez, Hive			
	Hadoop 1 and Hadoop 2 (YARN)			
9	NJIT Spring Break - No Classes Schedule			
10	Spark, Pig, etc			
11	Analytical Theories and Methods for Big Data			
	Hadoop/Mahout			
	Machine Learning			
12	Review and mid term2			
13	Recommendation			
	Clustering			
	Classification			
	Regression			
14	Advanced Topics			
	Big Data Volume and Information Visualization			
	High-performance Networking for Big Data Movement			

Tentative Course Topics (Subject to changes according to class progress)

	Big Data Scientific Workflow Management and Optimization	
15,16	Review and Project presentation	

3 Homework's	15%
Mid Term Exam 1,2	30% -(15% each)
Project	15%
Final Exam	20%
Quiz	10%
Discussions	10%

Grade	Marks
А	100-90
B+	89-80
В	79-70
C+	69-60
С	59-45, anything below 45 is F

Note – The grades will not be curved unless necessary.