

DS -644 Introduction to Big Data | Section -002, 004

Syllabus - Spring 2024

Instructor - Ravneet Kaur (rk956@njit.edu; ravneet.kaur@njit.edu)

Office Hours -

(W) 2.30 PM to 3.30 PM and (F) 12.30 to 1:00 PM in GITC 5711 or WebEx/Google Meet on request

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.

Attendance: You are supposed 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.

Tentative Course Topics (Subject to changes according to progress)

1. Introduction to Big Data Analytics
 - Introduction to Big Data
 - Scalability & Parallel Processing

- Designing Data Architecture
 - Data Source, Quality, Pre-processing & Storing
 - Data Storage & Analysis
 - Big Data Analytics Applications & Case Study
2. Introduction to Hadoop
- Introduction & Hadoop Installation
 - Hadoop & Its Ecosystem
 - Hadoop Distributed File System
 - MapReduce Framework & Programming Model
 - Hadoop Yarn
 - Hadoop Ecosystem Tools
3. NoSQL Big Data Management, MongoDB and Cassandra
- NoSQL Data Store
 - NoSQL Data Architecture Patterns
 - NoSQL to Manage Big Data
 - Shared–Nothing Architecture for Big Data Tasks
 - MongoDB Databases
 - Cassandra Databases
4. MapReduce, Hive and Pig
5. Spark and Big Data Analytics
6. Programming Examples in Analytics & Machine Learning

3 Assignments	15%
Mid Term Exam	30%
Exercise/ Presentation	20%
Final Exam	35%

<u>Grade</u>	<u>Marks</u>
A	100-90
B+	89-80
B	79-70
C+	69-60
C	59-45
F	44 and below