COURSE DESCRIPTION FORM: CS-4109: Big Data Analytics

INSTITUTION FAST School of Computing, National University of Computer and Emerging Sciences, Islamabad

PROGRAM TO BE EVALUATED	BS-CS: **Spring-2026**

Course Description

Course Code	CS-4109
Course Title	Big Data Analytics
Credit Hours	3+0
Prerequisites by Course(s) and Topics	
Grading policy	Absolute Grading
Policy about missed assessment items in the course	Retake of missed assessment items (other than sessional/ final exam) will not be held. Student who misses an assessment item (other than sessional / final exam) is awarded zero marks in that assessment item i.e. late submission will not be accepted. For missed sessional/ final exam, exam retake/ pretake application along with necessary evidence are required to be submitted to the department secretary. The examination assessment and retake committee decides the exam retake/ pretake cases.
Course Plagiarism Policy	Plagiarism in project or sessional/ final exam will result in F grade in the course. Plagiarism in an assignment will result in zero marks in the whole assignments category.
Assessment Instruments with Weights (homeworks, quizzes, sessional exams, final exam, assignments, etc.)	Assessment with the tentative weight.

Assessment Item	**Number**	**Weight (%)**
Assignments	>3	12
Quizzes	>4	13
Project/Research	1	10
Sessional I	1	10
Sessional II	1	15
Final Exam	1	40

Course Coordinator	
Course Instructor	Dr. Asif Muhammad, Ms. Kainat Iqbal
Lab Instructor	
URL (if any)	**Google Classroom Link:** https://classroom.google.com/c/ODQwMTA0MDk5Nzlw?cjc=rnnqxezk **Google Classroom Code: rnnqxezk**
Current Catalog Description	Introduction, Algorithm for Finding Similar Items in Large Datasets at Scale, Hadoop Introduction, HDFS, MapReduce Programming Model, Page Rank, Page Rank Computation using Map Reduce, Hive, Introduction to Spark, Spark Programming Model, RDDs, Transformations and Actions, SparkSQL DataFrames, Spark Job Execution, Frequent Pattern Mining, Stream Data Processing using Apache Kafka, Publish-Subscribe Model, SQL vs NoSQL, Disc-level data processing and query execution for SQL (Row-Oriented) and No-SQL (Column-Oriented) Databases
Textbook	• **"Mining of Massive Datasets",** Jure Lesovec, Anand Rajaraman, Jeff Ullman • **"Big Data Systems. A360-degree Approach",** Jawwad Ahmad Shamsi, Muhammad Ali Khojaya • **"Hadoop: The Definitive Guide 4th edition",** Tom White, O'Reilly • **"Learning Spark for Lightning-Fast Big Data Analysis",** Matei Zaharia, Holden Karau, Andy Konwinski, Patrick Wendell
Reference Material	Lecture Slides and Other Material referred and provided during the class

Course Goals

A. Course Learning Outcomes (CLOs)	Bloom Taxonomy Level
After course completion, the students shall be able to: 1. Understand the fundamental concepts of Big Data and its programming paradigm. 2. Apply Big Data Concepts using Hadoop/MapReduce Programming Framework, and Apache Spark Ecosystem	2 5

B. PLO Summary

PLO 1	Academic Education	
PLO 2	Knowledge for Solving Computing Problems	✔
PLO 3	Problem Analysis	
PLO 4	Design/ Development of Solutions	
PLO 5	Modern Tool Usage	✔
PLO 6	Individual and Team Work	
PLO 7	Communication	
PLO 8	Computing Professionalism and Society	
PLO 9	Ethics	
PLO 10	Lifelong Learning	

C. Mapping of CLOs on PLOs
(CLO: Course Learning Outcome, PLOs: Program Learning Outcomes)

		PLOs									
		1	2	3	4	5	6	7	8	9	10
CLOs	1		✔								
	2					✔					

Topics Covered in the Course, with Number of Lectures on Each Topic

Topics to be covered:

List of Topics	No. of Weeks	Contact Hours	CLO(s)

<table>
<tr><td>(assume 15-week instruction and one-hour lectures)</td><td>
<table>
<tr><td>Introduction to Big Data Analytics</td><td>1</td><td>3</td><td>1</td></tr>
<tr><td>Hadoop, Map Reduce</td><td>3</td><td>9</td><td>2</td></tr>
<tr><td>Finding Similar Items at Scale</td><td>2</td><td>6</td><td>2</td></tr>
<tr><td>Apache Kafka, Streaming Data Processing</td><td>3</td><td>9</td><td>2</td></tr>
<tr><td>Frequent Pattern Mining</td><td>3</td><td>9</td><td>2</td></tr>
<tr><td>Apache Spark</td><td>1</td><td>3</td><td>2</td></tr>
<tr><td>Apache Storm</td><td>1</td><td>3</td><td>2</td></tr>
<tr><td>Machine Learning for Big Data (Spark MLlib)</td><td>1</td><td>3</td><td>2</td></tr>
<tr><td>Data Processing in SQL and NoSQL Databases</td><td>1</td><td>3</td><td>1</td></tr>
<tr><td>Total</td><td>16</td><td>48</td><td></td></tr>
</table>
</td></tr>
<tr><td>Laboratory Projects/Experiments Done in the Course</td><td colspan="4">Hands-on for all the topics stated above (except Frequent Pattern Mining and Locality Sensitive Hashing) will be performed from installation to writing code for various big data problems.</td></tr>
<tr><td>Programming Assignments/ Lab Done in the Course</td><td colspan="4">Python, Hadoop, Map Reduce, Apache Spark, MongoDB, Apache Kafka, Apache Storm</td></tr>
<tr><td>Class Time Spent (in percentage)</td><td>Theory</td><td>Problem Analysis</td><td>Solution Design</td><td>Social and Ethical Issues</td></tr>
<tr><td></td><td>50</td><td>25</td><td>20</td><td>5</td></tr>
<tr><td>Oral and Written Communications</td><td colspan="4">Every student is required to submit at least __5___ written reports of typically ___5____ pages each and to make __1___ oral presentation of typically ____10___ minutes' duration.</td></tr>
</table>