JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

M.Tech. (DS) – II SEMESTER | T1 EXAM - FEBRUARY 2026

COURSE CODE (CREDITS): 22M11CI213 (3) MAX. MARKS: 15

COURSE NAME: Big Data Analytics (BDA) MAX. TIME: 1 Hour

COURSE INSTRUCTORS: D. Gupta

***Note:** 1) All questions are compulsory. Marks and COs for each question are indicated. 2) **Answer the questions in the given order.** 3) **Be concise and write neatly.** 4) Use of calculators is allowed.*

Q. No.	Question	CO	Marks
Q. 1	An HDFS system stores a 1 TB file with block size = 128 MB and replication factor = 3. Calculate: (a) Number of blocks created (b) Total storage space required after replication Describe all steps clearly.	1	3
Q. 2	Compare and contrast Hadoop and Apache Spark with respect to architecture, processing speed, type of processing, and typical use cases. Present your answer in a tabular form.	1	3
Q. 3	Draw a neat labeled diagram of Apache Spark architecture showing Driver (SparkContext), Cluster Manager, Worker Nodes, Executors, and Tasks. Briefly state the role of each component in executing a Word Count job, highlighting lazy transformations, actions, and cache usage.	1	3
Q. 4	A web graph contains 5 pages with initial PageRank = 1/5 and damping factor = 0.85. Perform one iteration of the PageRank algorithm and compute the updated rank of each page. Show your calculations clearly.	2	3
Q. 5	Design a MapReduce solution to count the number of students in each department from a large dataset. Specify the Mapper input–output pairs and Reducer logic.	2	3