ourdiu.com | Shared by MD. GOLAM MUSTAKIM NAEEM

Daffodil International University (DIU)

Department of Computing and Information System (CIS)

Course: Big Data Analytics and Applications (CIS312) **Exam:** Class Test 03 (Set B)
Time: 30 Minutes **Total Marks:** 15

Answer all questions. Be concise and provide architectural justifications where required.

1. **[Marks: 3 + 2 = 5]**

   a) Explain how YARN splits the monolithic responsibilities of the legacy `JobTracker` between the global **ResourceManager** and the per-application **ApplicationMaster**.

   b) When running Apache Spark in **YARN Cluster Mode**, where does the Spark **Driver** execute, and what role do the YARN **NodeManagers** play in hosting Spark **Executors**?

2. ***[Marks: 2 + 3 = 5]***

   a) *What is* ***Lazy Evaluation*** *in Apache Spark, and how does it allow the Spark engine to optimize the overall execution plan before any actual data processing begins?*

   b) *Differentiate between the roles of the* **DAGScheduler** *and the* **TaskScheduler** *when converting a user's Spark job into physical execution across the worker nodes.*

3. ***[Marks: 2 + 3 = 5]***

   a) *Spark's Unified Memory Manager divides the Executor's RAM primarily into* **Execution Memory** *and* **Storage Memory**. *State the core purpose of each memory region.*

   b) *An iterative machine learning algorithm needs to process the same intermediate RDD 50 times. Explain how using the* `.cache()` *function improves performance for this workload, and describe how Spark recovers the data if the Executor holding the cached RDD suddenly crashes.*