| **INSTRUCTIONS:** | |
| --- | --- |
| 1. | Figures to the right indicate maximum marks for that question. |
| 2. | The symbols used carry their usual meanings. |
| 3. | Assume suitable data, if required & mention them clearly. |
| 4. | Draw neat sketches wherever necessary. |

| Examination | : Sessional II | Seat No. | : |
| --- | --- | --- | --- |
| Date | : 08/04/2022 | Day | : Friday |
| Time | : 9:00 a.m. to 10:15 a.m. | Max. Marks | : 36 |

| Q.1 | Do as directed: |  |
| --- | --- | --- |
| (a) | Fill in the blanks:  (i) Hadoop can be run in \_\_\_\_ , \_\_\_\_ or fully \_\_\_\_ mode.  (ii) In MongoDB terminology tables are called \_\_\_\_\_ and records are called \_\_\_\_\_\_ having fields key-value JSON. | [02] |
| (b) | State true or false.  (i) If you are importing package like “mapred”, you are using MR V1 and from “mapreduce” you are using MR V2.  (ii) To start MongoDB we run the background service named mongo and to connect we use mongod CLI. | [02] |
| (c) | Compare and contrast client-server to master-worker architecture. | [02] |
| (d) | Match the following:   | Column A | Column B | | --- | --- | | HDFS master component | ResourceManager | | HDFS worker component | DataNode | | YARN master component | NodeManager | | YARN worker component | NameNode | | [02] |
| (e) | Explain any four Hadoop EcoSystem components in brief. | [02] |
| (f) | How to monitor whether Hadoop components are running or not? CLI/GUI. | [02] |
|  |  |  |
| Q.2 | Attempt the following: |  |
| (a) | Develop a MongoDB supported solution to achieve below requirements:   * Create collection named “Students” * Insert documents as per dataset. * Display subjectName and marks of all records belonging to studentId 1. Do not show inbuilt \_id. * Group on “studentId” and compute the sum of marks as TotalMarks. * Group on “subjectName” and compute the avg of marks as SubjectAvgMarks. * Display records sorted by subjectName ascending and followed marks descending.   Data set :  {"studentId" : 1, "subjectName" : "Augmented Reality", "marks" : 25 }  {"studentId" : 1, "subjectName" : "Data Engineering", "marks" : 15 }  {"studentId" : 2, "subjectName" : "Data Engineering", "marks" : 22 }  {"studentId" : 3, "subjectName" : "Image Calculus", "marks" : 28 }  {"studentId" : 1, "subjectName" : "Visual Recognition", "marks" : 20 }  {"studentId" : 2, "subjectName" : "Cloud Engineering", "marks" : 20 }  {“studentId" : 2, "subjectName" : "Visual Presentation", "marks" : 24 }  {"studentId" : 1, "subjectName" : "Visual Presentation", "marks" : 20 }  {"studentId" : 3, "subjectName" : "Image Processing", "marks" : 18 } | [06] |
| (b) | Provide a solution using MongoDB map-reduce library and javascript/JSON support to count words from collection named input and generate results into collection output in the form of unique words as \_id with their frequencies line by line as value. | [06] |
| Q.3 | Attempt the following: |  |
| (a) | Perform below actions with respect to hadoop distributed file system:   * List the content of HDFS folder /user/hadoop/demo assuming that it exists. * Make directories day1, day2 and day3 in HDFS folder /user/hadoop/demo * From the current working directory of the host operating system, dump all text files to HDFS folder /user/hadoop/demo/day1. * Display content of the text files from /user/hadoop/demo/day1. * Copy files starting with ‘h’ letter in its name from HDFS folder /user/hadoop/demo/day1 to /user/hadoop/demo/day2. * Make a directory named “Downloads” in the host operating system within the current working directory. Retrieve all files from HDFS folder /user/hadoop/demo/day2 to host os inside “Downloads”. | [06] |
| (b) | Explain hadoop map-reduce framework with example or program java library to count words from input files and generate output in the form of unique words with their frequencies line by line. | [06] |