**SAVITIBAI PHULE UNIVERSITY OF PUNE**
**M.Sc. (Computer Science) Part II Semester III**
**Practical Examination 2022-2023**
**SUBJECT: Big Data Analytics Practical (CSDP234A)**

**Time: 2 hours**                                                                                   **Max. Marks: 35**

| Marks ____ + ____ + ____ = ____ | Examiner 1 Sign: |
|---|---|
| 10      15      10      35 | Examiner 2 Sign: |

Q.1     Execute the following HDFS commands on Hadoop environment.          [10]

   a)  Remove a demo directory from Hadoop environment. (Create a demo directory).

   b)  Create a student.txt file in Hadoop environment and copy this file in root directory.

   c)  display the content of student.txt file.

   d)  expunge command

   e)  df command

Q.2 Write basic Word Count Map Reduce program using python/java to understand Map
        Reduce Paradigm.                                                              [15]

Q.3     Viva                                                                           [10]

**Time: 2 hours**                                                                                    **Max. Marks: 35**

| Marks ____ + ____ + ____ = ____ | Examiner 1 Sign: |
| 10      15      10      35 | Examiner 2 Sign: |

Q.1.    A) Show practical examples to list files, Insert data, retrieving data,append to file  and shutting down HDFS.                                                          [10]

Q.2    Write a MapReduce program to analyze weather data set and print whether the day is shinny or cool day.Hint: Weather sensors collecting data every hour at many locations across the globe gather a large volume of log data, which is a good candidate for analysis with Map Reduce, since it is semi structured and record-oriented.                    [15]

 Q.3 Viva                                                                                                         [10]

**Time: 2 hours**                                                                    **Max. Marks: 35**

| Marks ____ + ____ + ____ = ____ | Examiner 1 Sign: |
| 10      15      10      35 | Examiner 2 Sign: |

Q.1 Execute the following HDFS commands on Hadoop environment.                [10]

   a) Create a emp.txt file in root directory and move this file in Hadoop environment.

   b) display the statistics about the file.(use default format).

   c) change the permission of the file.

   d) implement checksum command

   e) delete emp.txt file.

Q.2    Weather Report POC-Map Reduce Program to analyse time-temperature statistics and generate report with max/min temperature.Problem Statement is as follows                [15]

   1. The system receives temperatures of various cities(Austin, Boston,etc) of USA captured at regular intervals of time on each day in an input file.

   2. System will process the input data file and generates a report with Maximum and Minimum temperatures of each day along with time.

   3. Generates a separate output report for each city.

   Ex: Austin-r-00000

   Boston-r-00000

   Newjersy-r-00000

   Baltimore-r-00000

   California-r-00000

   Newyork-r-00000

   Expected output:- In each output file record should be like this:

   25-Jan-2014 Time: 12:34:542 MinTemp: -22.3 Time: 05:12:345 MaxTemp: 35.7

Q.3    Viva                                                                            [10]

**Time: 2 hours**                                                                    **Max. Marks: 35**

Marks _____ + _____ + _____ = _____        Examiner 1 Sign:
         10          15          10          35          Examiner 2 Sign:

Q.1  Execute the following HDFS commands on Hadoop environment.          [10]

    a) Create a emp.txt file in root directory and move this file in Hadoop environment.
    b) Display last few lines of the above emp.txt file.
    c) du command
    d) df command
    e) fsck

Q.2  It contains the monthly electrical consumption and the annual average for various years.

To find the maximum number of electrical consumption and minimum number of electrical

consumption in the year.                                                              [15]

|      | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Avg |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1979 | 23  | 23  | 2   | 43  | 24  | 25  | 26  | 26  | 26  | 26  | 25  | 26  | 25  |
| 1980 | 26  | 27  | 28  | 28  | 28  | 30  | 31  | 31  | 31  | 30  | 30  | 30  | 29  |
| 1981 | 31  | 32  | 32  | 32  | 33  | 34  | 35  | 36  | 36  | 34  | 34  | 34  | 34  |
| 1984 | 39  | 38  | 39  | 39  | 39  | 41  | 42  | 43  | 40  | 39  | 38  | 38  | 40  |
| 1985 | 38  | 39  | 39  | 39  | 39  | 41  | 41  | 41  | 00  | 40  | 39  | 39  | 45  |

Q.3 Viva                                                                              [10]

**Time: 2 hours**                                                                                          **Max. Marks: 35**

| Marks ____ + ____ + ____ = ____ | Examiner 1 Sign: |
|---|---|
| 10      15      10      35 | Examiner 2 Sign: |

Q. 1 Execute the following HDFS commands on Hadoop environment.                    [10]

    a.  count the number of files and directories in HDFS.(Use options for the command)
    b.  find
    c.  getmerge
    d.  usage
    e.  test

Q.2 The table includes the monthly visitors of website page and annual average of five years. To find the maximum number of visitors and minimum number of visitors in the year.  [15]

|  | JAN | FEB | MAR | APR | MAY | JUN | JULY | AUG | SEP | OCT | NOV | DEC | AVG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2008 | 23 | 23 | 2 | 43 | 24 | 25 | 26 | 26 | 26 | 25 | 26 | 26 | 25 |
| 2009 | 26 | 27 | 28 | 28 | 28 | 30 | 31 | 31 | 31 | 30 | 30 | 30 | 29 |
| 2010 | 31 | 32 | 32 | 32 | 33 | 34 | 35 | 36 | 36 | 34 | 34 | 34 | 34 |
| 2014 | 39 | 38 | 39 | 39 | 39 | 41 | 42 | 43 | 40 | 39 | 39 | 38 | 40 |
| 2016 | 38 | 39 | 39 | 39 | 39 | 41 | 41 | 41 | 00 | 40 | 40 | 39 | 45 |

Q.3 Viva                                                                                                                    [10]

**Time: 2 hours**                                               **Max. Marks: 35**

Marks _____ + _____ + _____ = _____        Examiner 1 Sign:
             10          15          10          35        Examiner 2 Sign:

Q.1 Execute the following HDFS commands on Hadoop environment.                    [10]

   a) Create a studentdata.txt file in Hadoop environment and move this file to root
      directory.
   b) put
   c) tail
   d) fsck
   e) display the list of files in specified directory.(Create files and directory accordingly.)

Q.2    Write a MapReduce program to analyze Sales related information.The input data used
is SalesJan2009.csv. It contains Sales related information like Product name, price, payment
mode, city, country of client etc. The goal is to Find out Number of Products Sold in Each
Country.                                                                          [15]

Q.3 Viva                                                                           [10]

**Time: 2 hours**                                                                 **Max. Marks: 35**

Marks _____ + _____ + _____ = _____          Examiner 1 Sign:
        10        15        10        35        Examiner 2 Sign:

Q. 1) Execute the following HDFS commands on Hadoop environment.`          [10]

    a) mkdir

    b) copy the demo.txt in hadoop environment.

    c) move the directory into another directory.

    d) tail

    e) delete file created in hadoop environment.

(Note: you may create a file/directory depending on the command requirements.)

Q.2) Write a basic Word Count Map Reduce program using python/java to understand Map Reduce Paradigm. create file called sample.txt whose contents are as follows:

    **Dear, Bear, River, Car, Car, River, Deer, Car and Bear**          [15]

Q.3 Viva                                                                 [10]

**SAVITIBAI PHULE UNIVERSITY OF PUNE**
**M.Sc. (Computer Science) Part II Semester III**
**Practical Examination 2022-2023**
**SUBJECT: Big Data Analytics Practical (CSDP234A)**

**Time: 2 hours**                                                        **Max. Marks: 35**

Marks _____ + _____ + _____ = _____          Examiner 1 Sign:
           10        15        10        35              Examiner 2 Sign:

Q.1 Execute the following HDFS commands on Hadoop environment.          [10]

   a) changing the group of 'sample.zip' file of the HDFS file system.
   b) changing the owner of a file name sample.
   c) prints a summary of the amount of disk usage of all files.
   d) show last modified time of directory
   e) implement test command

(Note: you may create a file/directory depending on the command requirements.)

Q.2 Implement matrix multiplication with Hadoop Map Reduce.          [15]

Q.3 Viva          [10]

**Time: 2 hours**                                                    **Max. Marks: 35**

| Marks ___ + ___ + ___ = ___ | Examiner 1 Sign: |
| 10      15      10      35 | Examiner 2 Sign: |

Q. 1 Execute the following HDFS commands on Hadoop environment.          [10]

   a) create empty file in hdfs.
   b) copy sample.txt from local environment to HDFS.
   c) print the content of sample.txt file.
   d) display the total size of file.
   e) chmod command

 (Note: you may create a file/directory depending on the command requirements.)

Q.2 Write MapReduce program using python/java to find out how many reviews for each rating. consider this  dataset that consists of hotel reviews and ratings.

                                                                          [15]


Q.3     Viva                                                              [10]

**Time: 2 hours**                                                                 **Max. Marks: 35**

| Marks ____ + ____ + ____ = ____ | Examiner 1 Sign: |
|---|---|
| 10      15      10      35 | Examiner 2 Sign: |

Q.1 Execute the following HDFS commands on Hadoop environment.            [10]

    a) copy one file from one directory to another within HDFS.
    b) show the statistics about the directory in the specified format.(%b,%g,%u,%n)
    c) implement text command.
    d) fsck
    e) cat command

(Note: you may create a file/directory depending on the command requirements.)

Q.2 Write mapreduce program to analyze employee details and find the maximum and minimum salary in each department.employee details consists of list of employees with there department and salary.            [15]

Q.3 Viva            [10]

**Time: 2 hours**                                                              **Max. Marks: 35**

| Marks ____ + ____ + ____ = ____ | Examiner 1 Sign: |
|---|---|
| 10      15      10      35 | Examiner 2 Sign: |

Q.1) Execute the following HDFS commands on Hadoop environment.          [10]

   a) display summary of the amount of disk usage of all files.

   b) change the permission of the file.

   c) Create an employee.txt file with some content and Moves this file from local file system to the Hadoop file system.

   d) copy one file from one directory to another within HDFS.

   e) display last 1kb data of employee.txt file.

(Note: you may create a file/directory depending on the command requirements.)

Q.2) Implement matrix multiplication with Hadoop Map Reduce.          [15]

Q.3) Viva          [10]

**Time: 2 hours**                                                                   **Max. Marks: 35**

**Marks** _____ + _____ + _____ = _____          **Examiner 1 Sign:**
          10        15          10          35          **Examiner 2 Sign:**

Q.1) Execute the following HDFS commands on Hadoop environment.          [10]

   a. create patient_detalis.txt file in Hospital directory and move this file in another directory within hdfs.
   b. display the total size of Hospital directory.
   c. show the statistics about the directory in the specified format.(%o,%r,%u,%y)
   d. list of all files in hospital directory.(add and list a minimum 4 files).
   e. delete hospital directory.

(Note: you may create a file/directory depending on the command requirements.)

Q.2) Write mapreduce program to analyze employee details and find the count of employees in each department.employee details consists of list of employees with there department.

                                                                                   [15]

Q.3)Viva                                                                           [10]

**Time: 2 hours**                                                                                          **Max. Marks: 35**

| Marks ____ + ____ + ____ = ____ | Examiner 1 Sign: |
|---|---|
| 10      15      10      35 | Examiner 2 Sign: |

Q.1) Execute the following HDFS commands on Hadoop environment.                    [10]

   a) Create a college.txt file in root directory and move this file in Hadoop environment.
   b) Count the number of files and directories in HDFS.(use specified format as option -v,-h,-q)
   c) du command
   d) implement find command for college.txt
   e) Remove a bank_details directory from Hadoop environment. (Create a bank_details directory).

(Note: you may create a file/directory depending on the command requirements.)

Q.2) Write a MapReduce program to analyze Sales related information.The input data used is SalesJan2009.csv. It contains Sales related information like Product name, price, payment mode, city, country of client etc. The goal is to Find out Number of Products Sold in Each Country.                                                                                          [15]

Q.3)Viva                                                                                          [10]

**Time: 2 hours**                                                                 **Max. Marks: 35**

| Marks | ____ + ____ + ____ = ____ | Examiner 1 Sign: |
|-------|---------------------------|------------------|
|       | 10    15    10    35      | Examiner 2 Sign: |

Q.1) Execute the following HDFS commands on Hadoop environment.          [10]

  a)  Create a teacher.txt file in Hadoop environment and copy this file in root directory.

  b)  display the content of teacher.txt file.

  c)  display the list of files in specified directory.(Create files and directory accordingly.)
  d)  implement tail command on teacher.txt file.
  e)  remove teacher.txt file from directory.

(Note: you may create a file/directory depending on the command requirements.)

Q.2) Write a MapReduce program to analyze weather data set and print whether the day is shinny or cool day.Hint: Weather sensors collecting data every hour at many locations across the globe gather a large volume of log data, which is a good candidate for analysis with Map Reduce, since it is semi structured and record-oriented.          [15]

Q.3)Viva          [10]

**SAVITIBAI PHULE UNIVERSITY OF PUNE**
**M.Sc. (Computer Science) Part II Semester III**
**Practical Examination 2022-2023**
**SUBJECT: Big Data Analytics Practical (CSDP234A)**

**Time: 2 hours**                                                        **Max. Marks: 35**

**Marks** _____ + _____ + _____ = _____          **Examiner 1 Sign:**
          10        15         10        35          **Examiner 2 Sign:**

Q.1) Execute the following HDFS commands on Hadoop environment.          [10]

   a)  create and move the bank directory into another directory within hadoop environment.
   b)  change the permission of the file.
   c)  Display last few lines of the above customer.txt file.
   d)  change the replication factor of customer.txt file in HDFS.
   e)  delete bank directory from HDFS.

(Note: you may create a file/directory depending on the command requirements.)

Q.2) Write basic Word Count Map Reduce program using python/java to understand Map Reduce Paradigm.                                                        [15]

Q.3)Viva                                                        [10]