B.E./Sem-VII/COMP+CSE/C-Scheme/NOV-2025/14-11-2025

(03 HOURS) (MAX. MARKS : 80)

Note:

1. Question **No. 1** is compulsory.
2. Attempt **any three** questions out of remaining **five** questions.
3. Assume suitable data wherever necessary.
4. Figures to right indicate full marks.

KONKAN GYANPEETH COLLEGE OF ENGINEERING KARJAT Dist. Raigad

Q. No.	Question	Max. Marks
Q 1	Answer the following	
a)	What is the role of a combiner in the map reduce framework? Explain with the help of an example.	05
b)	Which algorithm can be used for finding distinct elements in a long data stream ? Give two real life applications for the same.	05
c)	Describe terms TF and IDF with respect to content based filtering	05
d)	Why recall is 100% for blooms filter, explain in detail.	05
Q2.a)	Explain Hadoop Ecosystem in detail.	10
b)	Write map-reduce pseudo code/algorithm to solve wordcount problem. Apply map-reduce working on the following document: "Big data is powerful. Big data drives decisions."	10
Q3. a)	Discuss Blooms filter in detail and insert elements 10 and 7 in Blooms filter of size 5 using following hash functions: 1-h1(x)= x mod 5 and 2-h2(x)=(2 x+ 6) mod 5	10
b)	Summarize how business problems have been successfully solved faster,cheaper and more effectively considering NOSQL Google's bigtable case study. Also, illustrate the business drivers and findings in it.	10

muquestionpapers.com

Q4.a) Predict rating of user 2 for Movie 3 using Pearson's Correlation coefficient. **10**

users	Ratings				
	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User1	2	3	2	1	1
User2	3	4	-	4	3
User 3	4	3	4	3	3

KONKAN GYANPEETH COLLEGE OF ENGINEERING KARJAT Dist. Raigad

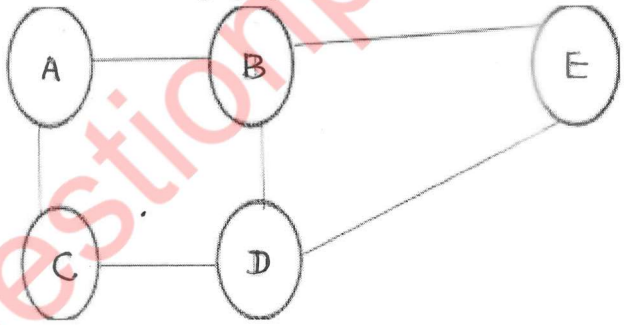
b) Write a Map Reduce pseudo code to Vector-Matrix multiplication. Demonstrate all the steps for following matrix and vector **10**

$$M = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad X \quad V = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Q5.a) Suppose our stream consists of the integers 3, 1, 4, 1, 5, 9, 2,6, 5. Determine the number of distinct elements if the hash function is: $h(x) = 3x+ 7 \bmod 8$. **10**

b) Mention four characteristics of big data. Elaborate these characteristics with respect to social media websites. **10**

Q6 a) Demonstrate the different communities using Girwan-Newman algorithm for the following graph **10**



b) Explain different data visualization techniques in R with examples. **10**

****

muquestionpapers.com

Gmail

Exam KGCE <exam@kgce.edu.in>

Correction in Q.P.Code: 10097245

1 message

support@muapps.in <support@muapps.in> Fri, Nov 14, 2025 at 11:33 AM
Reply-To: support@muapps.in
To: exam@kgce.edu.in

University of Mumbai

BIG DATA ANALYTICS
Correction in Q.P.Code: 10097245
The correction for 3(a) is as follows:
Discuss Blooms filter in detail and insert elements 10 and 7 in Blooms filter of size 5 using following hash functions:1) h1(x) = x mod 5 2) h2(x) = (2 x+ 6) mod 5

University of Mumbai
https://qpd.qpd.muapps.in
support@muapps.in

You have received this email because you are registered with us.
To unsubscribe; please reply to this mail with subject "Unsubscribe"

muquestionpapers.com

Gmail

Exam KGCE <exam@kgce.edu.in>

Correction in Q.P.Code: 10097245

support@muapps.in <support@muapps.in> Fri, Nov 14, 2025 at 12:41 PM
Reply-To: support@muapps.in
To: exam@kgce.edu.in

University of Mumbai

BIG DATA ANALYTICS
Correction in Q.P.Code: 10097245
The correction for 5(a) is as follows:
Suppose our stream consists of the integers 3, 1, 4, 1, 5, 9, 2,6, 5.
Determine the number of distinct elements if the hash function is: h(x) = 3x+ 7 mod 8 for some a and b, treat the result as a 5-bit binary integer.
[Quoted text hidden]

muquestionpapers.com

B.E./Sem-VII/COMP+CSE/C-Scheme/NOV-2025/14-11-2025

(03 HOURS) (MAX. MARKS : 80)

Note:

1. Question **No. 1** is compulsory.
2. Attempt **any three** questions out of remaining **five** questions.
3. Assume suitable data wherever necessary.
4. Figures to right indicate full marks.

KONKAN GYANPEETH COLLEGE OF ENGINEERING KARJAT Dist. Raigad

Q. No.	Question	Max. Marks
Q 1	Answer the following	
a)	What is the role of a combiner in the map reduce framework? Explain with the help of an example.	05
b)	Which algorithm can be used for finding distinct elements in a long data stream ? Give two real life applications for the same.	05
c)	Describe terms TF and IDF with respect to content based filtering	05
d)	Why recall is 100% for blooms filter, explain in detail.	05
Q2.a)	Explain Hadoop Ecosystem in detail.	10
b)	Write map-reduce pseudo code/algorithm to solve wordcount problem. Apply map-reduce working on the following document: "Big data is powerful. Big data drives decisions."	10
Q3. a)	Discuss Blooms filter in detail and insert elements 10 and 7 in Blooms filter of size 5 using following hash functions: 1-h1(x)= x mod 5 and 2-h2(x)=(2 x+ 6) mod 5	10
b)	Summarize how business problems have been successfully solved faster,cheaper and more effectively considering NOSQL Google's bigtable case study. Also, illustrate the business drivers and findings in it.	10

muquestionpapers.com

Q4.a) Predict rating of user 2 for Movie 3 using Pearson's Correlation coefficient. **10**

users	Ratings				
	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User1	2	3	2	1	1
User2	3	4	-	4	3
User 3	4	3	4	3	3

KONKAN GYANPEETH COLLEGE OF ENGINEERING KARJAT Dist. Raigad

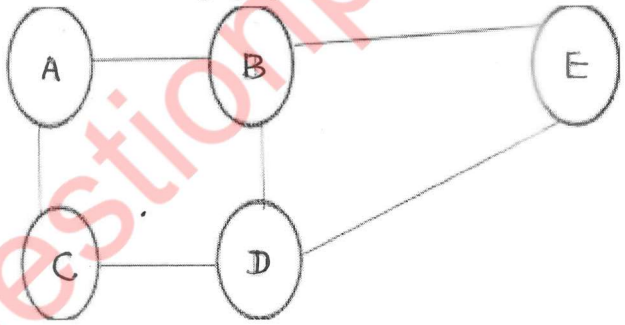
b) Write a Map Reduce pseudo code to Vector-Matrix multiplication. Demonstrate all the steps for following matrix and vector **10**

$$M = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad X \quad V = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Q5.a) Suppose our stream consists of the integers 3, 1, 4, 1, 5, 9, 2,6, 5. Determine the number of distinct elements if the hash function is: $h(x) = 3x+ 7 \bmod 8$. **10**

b) Mention four characteristics of big data. Elaborate these characteristics with respect to social media websites. **10**

Q6 a) Demonstrate the different communities using Girwan-Newman algorithm for the following graph **10**



b) Explain different data visualization techniques in R with examples. **10**

****

muquestionpapers.com