Duration 3 Hours   **Maximum marks 80**

NB:

1) Question 1 is compulsory
2) Attempt any three from the remaining questions
3) All questions carry equal marks.
4) Assume suitable data if necessary.

Q.1.	Attempt any **four** from the following questions	
a	List and explain the characteristics of Big Data.	**5**
b	Distinguish between Name node and Data node in HDFS.	**5**
c	Compare Content based recommendation system with collaborative filtering system.	**5**
d	Explain Cosine distance and Edit distance with example.	**5**
e	Give the pseudo code for implementing relational operation union and intersection using MapReduce.	**5**
Q.2.a	Draw the architecture of Hive and explain the working principles of Hive.	**10**
b	Explain in detail the core components of Hadoop.	**10**
Q.3.a	Elaborate the four ways that NoSQL system handles Big Data Problems.	**10**
b.	Explain cloaking technique for spamdexing.	**10**
Q.4.a	Explain the three distinct regions of web structure and dead ends and spider traps.	**10**
b.	Write MapReduce pseudo code for word count. Illustrate with an example how the mapper and reducer work to perform word counting.	**10**
Q.5 a	Explain how Bloom's filter block the blacklisted url's.	**10**
b	Explain the Park-Chen-Yu (PCY) algorithm for finding frequent item pairs. Illustrate with an example.	**10**
Q.6 a	Explain the typical characteristics of social networks. and explain the different types of social networks	**10**
b	Explain any two NoSQL data architecture patterns.	**10**

_____

X126Y465E29X126Y465E29X126Y465E29X126Y465E29