

Duration: 3hrs

[Max Marks: 80]

- N.B. :** (1) Question No 1 is Compulsory.
(2) Attempt any three questions out of the remaining five.
(3) All questions carry equal marks.
(4) Assume suitable data, if required and state it clearly.

- 1 Attempt any FOUR [20]**
- a** Why has Big Data become increasingly important in recent years? [5]
 - b** What is the Hadoop ecosystem, and why is it essential for Hadoop's functionality? [5]
 - c** What is NoSQL, and how does it differ from traditional SQL databases? [5]
 - d** What are combiners, and how do they contribute to the optimization of MapReduce jobs? [5]
 - e** What are the key advantages of using Bloom filters in stream processing applications? [5]
- 2**
- a** How can businesses address issues such as data security, privacy, and scalability in Big Data projects? [10]
 - b** What is Oozie, and how does it help in managing Hadoop workflow jobs? [10]
- 3**
- a** What were the specific challenges they faced, and how did NoSQL databases help address those challenges? [10]
 - b** What is the MapReduce WordCount program, and how does it showcase the MapReduce paradigm? [10]
- 4**
- a** What is the Market-Basket Model, and how does it relate to frequent itemset mining? [10]
 - b** What are recommendation systems, and why are they important in various online platforms and services? [10]
- 5**
- a** What are the key NoSQL data architecture patterns, including key-value stores, graph stores, column-family stores (Bigtable), and document stores? [10]
 - b** What is the significance of MapReduce in the context of the new software stack for distributed computing? [10]
- 6**
- a** What techniques are used for the efficient computation of PageRank, and how is it implemented using MapReduce? [10]
 - b** How can standard clustering techniques be applied to social graphs for analysis? [10]
