Midsem questionbank

1)A company wants to set up a Big Data solution to analyse customer behaviour. Apply your knowledge of the Big Data stack to suggest which components should be used for data storage, processing, and visualization, and justify your choices

2)Differentiate between ACID and BASE models in terms of consistency trade-off and evaluate which is more suitable for Big Data .

3)Examine the key features of Spark Streaming and analyze how they impact real-time data processing .

4)Apply YARN architecture with diagram to explain how Hadoop manages resources and job execution.

5) Explain the Map Reduce architecture in detail by using the Word Count example.

6)Which are the enabling technologies of Big Data? Explain with example

7)Define and explain HDFS. Explain its role in Big Data Ecosystem

8)Explain in detail the role of KAFKA in building real-time data pipeline

9)What is Data Streaming? Why is Real-time Big Data Pipeline So Important Nowadays?

10)Enlist and explain in brief the motivation behind emergence of Big Data