

Roll No.

Total Pages : 03

BT-7/D-19

37153

BIG DATA AND ANALYTICS

CSE-423N

Time : Three Hours]

[Maximum Marks : 75

Note : Attempt *Five* questions in all, selecting at least *one* question from each Unit.

Unit I

1. (a) What is Big Data ? What are challenges in big data ? Explain. 7
(b) What are NoSQL Databases ? Explain their characteristics. 8
2. (a) What is Hadoop ? How is it used to process big data ? Explain. 7
(b) What is Data Centre ? Explain different tiers of data centre. How is it related to cloud computing ? 8

Unit II

3. (a) Define Big Data Analytics. What is Pig ? Explain key features of Pig. 7
(b) What is Flume ? Explain recoverability and multihop flows, reliability and failure handling in Flume. 8
4. (a) What are key characteristics of MongoDB ? Explain following in MongoDB : insert method, aggregate function and cursor. 8
(b) What is deduplication ? Explain fixed-size and variable-size based deduplication techniques. 7

Unit III

5. (a) Explain delta compression techniques used for big data. 7
(b) What is Secure Hash Algorithm ? Explain SHA-1 algorithm in detail. 8
6. Write short notes on the following :
(a) Incremental Secure Fingerprint based deduplication. 7
(b) Scalable Decentralized Deduplication Store (SDDS). 8

Unit IV

7. (a) What is YARN ? Explain its achitecture. How is it useful for processing big data ? 7
- (b) What is Cassandra ? Explain key characteristics of Cassandra. 8
8. (a) What is Hive ? Explain architecture of Apache Hive. 7
- (b) How can you setup single node and multinode Hadoop cluster ? Explain. 8