

Roll No.

Total Pages : 02

BT-7/M-20

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 features of big data ?
Explain. 7
- (b) What is NoSQL ? Explain different NoSQL data
models. 8
2. (a) What are Hadoop and IoT ? How are these related
to big data ? Explain. 7
- (b) Explain about big data generation and acquisition
in detail. 8

Unit II

3. (a) What do you mean by big data analytics ? How
Hive is used for this purpose ? Explain. 7

(3)L-37153

- (b) What is GFS ? How is it different from HDFS ?
Explain. **8**
- 4. (a) What is Cassandra ? Explain key features of
Cassandra. **8**
- (b) What is content defined chunking ? How is it
different from frequency based chunking ? Explain.
7

Unit III

- 5. (a) Explain Gear Hash and Tiger Hash with examples. **7**
- (b) What is SHA-1 algorithm ? How is it different
from MD5 ? Explain. **8**
- 6. Write short notes on the following :
 - (a) Compression using PCOMPRESS **7**
 - (b) Parallel deduplication. **8**

Unit IV

- 7. (a) What is SPARK ? Explain the features of SPARK. **7**
- (b) What is MongoDB ? Explain key characteristics of
MongoDB. **8**
- 8. (a) What is Pig ? Explain architecture of Apache Pig. **7**
- (b) Explain programming model of working of
MapReduce. **8**

(3)L-37153

2

—