

**Bachelor of Technology in Computer Science Engineering (B.TECH CSE)/
Bachelor of Computer Application(BCA)**

**Re-appear Examination
Dec, 2024
ETCS318A- Big Data Analytics**

Time: 3 Hours

Maximum Marks: 50

Instructions:

- 1) Question No. 1 under Section-A is compulsory.
- 2) Attempt any four from Q. No. 2 to Q. No. 6 under Section-B
- 3) Attempt any two from Q. No. 7 to Q. No.9 under Section-C

Section-A(Compulsory)

1. (a) Fill in the blanks with appropriate words: 3
[CO-1] [L-1]
 - (i) Big Data analytics deals with processing and analyzing _____ volumes of data to extract valuable insights and patterns.
 - (ii) In Hadoop Distributed File System (HDFS), data is stored in blocks, typically with a default size of _____ megabytes.
 - (iii) Apache _____ is a high-level data warehousing infrastructure built on top of Hadoop for providing data summarization, query, and analysis.
1. (b) Give the term used for the following: 2
[CO-1] [L-2]
 - (i) Collection of data happens in this component of hadoop
 - (ii) Component that is an abstraction over mapreduce.
1. (c) State one function of each of the following: 2
[CO-2] [L-4]
 - (i) SQOOP
 - (ii) HBASE
1. (d) Match the items given in Column-A with those in Column-B: 3
[CO-3] [L-1]

Column-A	Column-B
(i) YARN	a) Data Flow Language
(ii) MYSQL	b) Resource manager
(iii) PIG	c) RDBMS

Section-B

2. (a) Define Hadoop ecosystem.
b) Explain the components of Hadoop ecosystem.

5[CO-2] [L-1]
3. Differentiate between HDFS and HBase in the context of Big Data storage.

5[CO-3] [L-2]
4. Explain the role of Sqoop in Big Data analytics.

5[CO-3] [L-1]
5. Discuss the advantages and disadvantages of using Hive in Big Data processing.

5[CO-2] [L-4]
6. (a) Describe the MapReduce programming model.
b) Explain how it processes data in Hadoop.

5[CO-2] [L-4]

Section-C

7. Explain the working of Apache Pig in Big Data analytics. Discuss its various operators and use cases.

10[CO-4] [L-5]
 8. (a) Differentiate between Hadoop 1.x and Hadoop 2.x.
(b) Discuss the significance of YARN in Hadoop 2.x

10[CO-3] [L-2]
 9. Explain the architecture of Apache HBase. Discuss its suitability for handling real-time data and its integration with Hadoop ecosystem.

10[CO-4] [L-5]
-