Printed Pages:02 **Sub Code:KOE-097**

Paper Id: 236036 **Roll No.** | | | | | | | | | | |

B.TECH
(SEM VIII) THEORY EXAMINATION 2022-2023
BIG DATA

Time: 3 Hours ***Total Marks: 100***

Note: Attempt all Sections. If require any missing data; then choose suitably.

SECTION A

1. Attempt *all* questions in brief. **2 x 10 = 20**

(a) Explain benefits of HDFS over NFS.

(b) Differentiate between structured, semi-structured and unstructured data.

(c) Explain sources of data in big data.

(d) Define Metadata in HDFS.

(e) Differentiate between Map & Reduce.

(f) Define indexing.

(g)

(h) Explain TF-IDF.

(i) Explain name node, data node, job tracker and task tracker.

(j) Define file name and block size for Windows, Linux and Hadoop.

SECTION B

2. Attempt any *three* of the following: **10x3=30**

(a) Explain the 5 Vs of Big Data. Also discuss their importance in the context of Big Data?
(b) Illustrate the history of Hadoop and its evolution over time into the Apache Hadoop platform that is widely used today?
(c) Explain the concept of data replication in HDFS and its benefits and challenges.
(d) Compare and contrast the fair and capacity schedulers used in the Hadoop YARN framework.
(e) Explain Pig and its execution modes. Compare pig with databases.

SECTION C

3. Attempt any *one* part of the following: **10x1=10**

(a) Discuss the role of security, compliance, auditing and protection in Big Data. Also discuss are the key features of Big Data in terms of security and privacy.
(b) Explain the challenges of conventional data systems. Discuss the process of providing a solution to these challenges by Big Data?

4. Attempt any *one* part of the following: **10x1=10**

(a) Discuss the Hadoop Distributed File System, and discuss its role to allow for the storage and processing of large data sets across distributed computing

clusters.

(b) What is the anatomy of a Map Reduce job run?

5. Attempt any *one* part of the following: 10x1=10

(a) Illustrate the data flows and data ingest methods in Hadoop, including Flume and Scoop?

(b) Discuss the support provided by Hadoop for compression, serialization, Avro, and file-based data structures in Hadoop I/O.

6. Attempt any *one* part of the following: 10x1=10

(a) Explain the basics of NoSQL databases and the MongoDB database in particular, including data types, document creation and manipulation, querying, and indexing.

(b) Discuss SCALA and its basic features, including classes and objects, basic types and operators, control structures, functions, and closures?

7. Attempt any *one* part of the following: 10x1=10

(a) Illustrate HBase and how its difference with RDBMS? Discuss the advantages of HBase's advanced indexing and schema design?

(b) Discuss the role of Zoo Keeper in monitoring a cluster. Also discuss the process of building applications with Zoo Keeper.