PAPER ID-410302

Subject Code: KDS601

Roll No: | | | | | | | | | | | | | |

BTECH
(SEM VI) THEORY EXAMINATION 2023-24
BIG DATA AND ANALYTICS

TIME: 3 HRS **M.MARKS: 100**

Note: 1. Attempt all Sections. If require any missing data; then choose suitably.

SECTION A

1. Attempt *all* questions in brief.

a.	List the differences between structured, semi-structured, and unstructured data.	2
b.	Write short note on Drivers of Big Data.	2
c.	List the core functionalities of Apache Hadoop.	2
d.	Explain how the data format of Hadoop is important.	2
e.	List the steps involved in using Sqoop to import data from RDBMS into Hadoop.	2
f.	Define how file system Works.	2
g.	List the key differences between the fair scheduler and the capacity scheduler in Hadoop.	2
h.	List the Data Type used in Mango DB.	2
i.	Define HiveQL and its key features.	2
j.	List out the Data Processing Operators used in Pig.	2

SECTION B

2. Attempt any *three* of the following:

a.	Discuss why Big Data is crucial for modern businesses and industries.	10
b.	Discuss Hadoop Distributed File System (HDFS), and how does it work.	10
c.	Discuss the benefits and challenges of using HDFS for big data storage and processing.	10
d.	Differentiate NoSQL databases from traditional RDBMS. Also list the key characteristics and use cases for NoSQL databases.	10
e.	Discuss the role of Zookeeper in monitoring a Hadoop cluster.	10

SECTION C

3. Attempt any *one* part of the following:

a.	Distinguishes Big Data analytics from traditional data analytics. Also list the techniques and tools used in Big Data analytics.	10
b.	Discuss the various Big Data Features in terms of Security, Protection and Auditing.	10

4. Attempt any *one* part of the following:

a.	Discuss the roles of additional components like HBase, Pig, and Hive in the Hadoop ecosystem.	10
b.	Discuss the Working of Map Reduce and its Characteristics	10

5. Attempt any *one* part of the following:

a.	Explain how HDFS implement data replication and list its advantages.	10
b.	Discuss the Security issues in Hadoop and why it is important for Data analysis.	10

6. Attempt any *one* part of the following:

a.	Explain how documents can be created, updated, and deleted in MongoDB.	10
b.	Explain the Various Ecosystem Components used in Hadoop with proper example.	10

7. Attempt any *one* part of the following:

a.	Explain the IBM's overall strategy for Big Data, and its key components.	10
b.	Discuss Big SQL, and how it extends SQL capabilities to Big Data environments.	10