Subject Code: KCA022

PAPER ID-311116

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

MCA
(SEM III) THEORY EXAMINATION 2024-25
BIG DATA

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

Note: Attempt all Sections. In case of any missing data, choose suitably.

SECTION A

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

Q no.	Question	CO	Level
a.	Define Big Data.	1	K2
b.	Write any 4 technologies used for the implementation of Big Data.	1	K2
c.	Differentiate between NoSQL and relational Database.	2	K2
d.	Discuss the role of map function in the map-reduce framework.	2	K1
e.	Discuss role of Flume with its advantage.	3	K2
f.	Explain the role of NameNode.	3	K1
g.	Discuss the role of YARN in Hadoop.	4	K2
h.	Discuss the use of Spark.	4	K2
i.	Differentiate between Hive and Hbase.	5	K1
j.	What are the 2 different modes of PIG.	5	K1

SECTION B

2. Attempt any *three* of the following: **10 x 3 = 20**

Q no.	Question	CO	Level
a.	Emphasize the types of data used to actionable insights in big data.	1	K2
b.	Discuss the history of Hadoop and its impact on Big Data processing.	2	K2
c.	Illustrate the rack awareness algorithm of HDFS used in writing the data.	3	K3
d.	Discuss major differences between Hadoop 1.0 and Hadoop 2.0. Explain with examples.	4	K2
e.	Discuss the role of zookeeper with its architecture.	5	K2

SECTION C

3. Attempt any *one* part of the following: **10 x 1 = 10**

Q no.	Question	CO	Level
a.	Explain the concept of Big Data Analytics with its classification.	1	K2
b.	Discuss the role of Big Data in healthcare, highlighting its applications, benefits, and challenges with relevant examples.	1	K2

4. Attempt any *one* part of the following: **10 x 1 = 10**

Q no.	Question	CO	Level
a.	Describe the MapReduce framework in Big Data processing, explaining its working, key components.	2	K2
b.	Explain the different types of MapReduce jobs in Big Data.	2	K2

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

MCA
(SEM III) THEORY EXAMINATION 2024-25
BIG DATA

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

5. Attempt any *one* part of the following: 10 x 1 = 10

Q no.	Question	CO	Level
a.	Illustrate the components of the Hadoop Distributed File System and key features.	3	K2
b.	Explain the process to write the block in Hadoop.	3	K2

6. Attempt any *one* part of the following: 10 x 1 = 10

Q no.	Question	CO	Level
a.	Explain the architecture and working of YARN highlighting its advantages.	4	K2
b.	Explain the key features of Scala, its benefits, and its applications in Big Data processing.	4	K2

7. Attempt any *one* part of the following: 10 x 1 = 10

Q no.	Question	CO	Level
a.	Describe the key features of Apache Pig, its architecture, and its advantages in simplifying Big Data processing.	5	K2
b.	Explain the architecture of Hive with the help of a neat diagram.	5	K2

PAPER ID-311786

Subject Code: KCA022

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

MCA
(SEM III) THEORY EXAMINATION 2023-24
BIG DATA

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

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

SECTION A

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

Q no.	Question	Marks	CO
a.	Define the term Big Data.	2	1
b.	Which type of data can be used for the Big Data?	2	1
c.	How does Apache Hadoop help process the data?	2	2
d.	List the 2 components of Hadoop.	2	2
e.	Write down the default block size of HDFS.	2	3
f.	Discuss two compression formats in Hadoop.	2	3
g.	Which component was introduced newly in Hadoop 2.0?	2	4
h.	List the responsibilities division of Job Tracker in Hadoop 2.0.	2	4
i.	Discuss the execution modes of the Pig	2	5
j.	Write 2 differences between Hbase and RDBMS.	2	5

SECTION B

2. Attempt any *three* of the following: **10 x 3 = 30**

a.	Discuss the various types of analytics used in Big Data.	10	1
b.	Describe the Hadoop Ecosystem.	10	2
c.	What do you mean by scaling out? Discuss the various reasons for scaling out.	10	3
d.	Compare the NoSQL databases with Relational databases. Describe the various types of NoSQL databases.	10	4
e.	Explain the architecture of Pig with the help of a neat diagram.	10	5

SECTION C

3. Attempt any *one* part of the following: **10 x 1 = 10**

a.	Discuss the 5 Vs used in Big Data with the help of an example.	10	1
b.	Illustrate the use of cloud computing to manage the data.	10	1

4. Attempt any *one* part of the following: **10 x 1 = 10**

a.	Discuss the role of Hadoop Pipes in the map-reduce process.	10	2
b.	Illustrate the map-reduce architecture with the help of a neat diagram.	10	2

5. Attempt any *one* part of the following: **10 x 1 = 10**

a.	Illustrate the rack awareness algorithm for writing the data in HDFS.	10	3
b.	Illustrate the process of reading a block in Hadoop.	10	3

6. Attempt any *one* part of the following: **10 x 1 = 10**

a.	Explain the components of features and components of Apache Spark.	10	4
b.	Discuss the role of MapReduce in Hadoop. Give the comparison between MapReduce V1 and V2.	10	4

7. Attempt any *one* part of the following: **10 x 1 = 10**

a.	Describe the Hive Architecture with the help of a neat diagram.	10	5
b.	Discuss the requirements and architecture of zookeeper.	10	5

Printed Pages:02 **Sub Code: KCA-022**

Paper Id: | 2 | 3 | 2 | 3 | 4 | 2 | **Roll No.** | | | | | | | | | | | | |

MCA
(SEM III) THEORY EXAMINATION 2022-23
BIG DATA

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

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

SECTION A

1. Attempt *all* questions in brief. 2x10 = 20

a) Define types of digital data.
b) What is big data privacy and ethics?
c) Write history of hadoop.
d) What is real world Map Reduce?
e) What is HDFS concept?
f) What is Avro and file-based data structures?
g) Describe Hadoop ecosystem components.
h) State the need of NoSQL databases?
i) Describe the comparison of Pig with databases.
j) What is a zookeeper?

SECTION B

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

a) Explain the concept big data with its architecture, and technological components.
b) Define anatomy of map reduce job run.
c) How does hdfs store, read, and write files.
d) Differentiate in following in reference with proper examples-
   (i) SQL and NOSQL
   (ii) MapReduce and YARN
   (iii)HBase and HDFS
   (iv)Spark and SCALA
e) Describe applications on big data using Pig, Hive, and HBase.

SECTION C

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

a) Analyze nature of data, and challenges of convolutional systems over intelligent data analysis? Also differentiate between analysis and reporting.
b) Describe 5 v's of data and big data features in detail.

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

a) Describe Apache Hadoop, Hadoop streaming, and Hadoop pipes.
b) Explain how does map reduce works. Also explain unit tests with MR unit, test data and local tests.

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

a) Explain the procedure of setting up a Hadoop cluster, cluster specification, cluster setup and installation.

b) Define data ingest with flume and scoop. Also, explain the concept of Hadoop in the cloud.

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

a) What are spark applications and how to install spark.

b) Describe inheritance, functions and closures concept of scale.

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

a) What do you mean by apache hive architecture? Explain hive installation.

b) What do you mean by infosphere, BigInsights and big sheets?