

IM.Sc. DS Sem.-8 Examination

CC-413

Big Data Analysis

April-2025

Time : 2-30 Hours]

[Max. Marks : 70

Instructions: All questions are compulsory. Use of non-programmable scientific calculator is allowed.

- Q.1**
- (a) What are the benefits of Big Data? Discuss challenges under Big Data. How Big Data Analytics can be useful in the development of smart cities. (07)
- (b) Explain (i) descriptive analytics (ii) predictive analytics and (iii) prescriptive analytics in the context of big data analytics. (07)
- OR**
- (a) Differentiate between big data and big data analytics. (03)
Differentiate between unstructured and semi-structured data with examples of each. (04)
- (b) What is Big Data? Discuss Five Vs and applications of Big Data. (07)
- Q.2**
- (a) Explain NameNode, Secondary NameNode, and DataNode with example. (07)
- (b) i) What do you mean by HiveQL Data Definition Language? Explain any three HiveQL DDL command with its syntax and example. (04)
ii) Write a short note on Apache Pig. (03)
- OR**
- (a) With suitable block diagram explain architecture of HDFS. Give following commands for HDFS. (07)
- Starting HDFS server
 - Listing files in HDFS
 - Inserting data into HDFS
- (b) i) Write HBASE command (04)
- To create a table as student with two column field.
 - Insert data in to the table.
 - Get data from the table.
 - Delete data from the table.
- ii) Explain structure of zookeeper and list its benefits. (03)
- Q.3**
- (a) What is HiveQL? Explain various statements in HiveQL with example. (07)
- (b) Explain the advantages of using Apache Spark over Hadoop and brief four major libraries of Apache Spark. (07)
- OR**
- (a) (i) Write a Pig Script for word count. (03)
(ii) Difference Between Hbase and Hive. (04)

- (b) Write a short note on Spark stack. Give brief explanation of each component. (07)
- Q.4** (a) What is MongoDB? Explain in brief key features of MongoDB. Show basic CRUD operations in MongoDB with proper example. (07)
- (b) Explain the Replication Architecture in MongoDB. (04)
Explain scaling feature of MongoDB. (03)
- OR**
- (a) What is NoSQL database? Discuss key characteristics and advantages of NoSQL database. (07)
- (b) Requirement specification of blog application in social networking is as follows: (07)
- i) Every post has a unique title, description & URL.
 - ii) Every post can have one or more tags.
 - iii) Every post has name of its publisher and total number of likes.
 - iv) Every post has comments given by users along with their name, message, data-time and likes.
 - v) On each post, there can be zero or more comments.
- For this set of requirements, design a MongoDB schema.
- Q.5** Attempt any **SEVEN** out of **TWELVE**: (14)
- (1) What will be the output of \$jps command if all the Hadoop services are up and running?
 - (2) What is the Mongo Shell? How to start and run the shell?
 - (3) Explain any two commands of HDFS with example:
(i) copyToLocal (ii) fsck (iii)put
 - (4) List various configuration files used in Hadoop Installation. What is use of hdfs-site.xml?
 - (5) List the different commands used in Hbase operations?
 - (6) Write the following HIVE commands
• UPDATE
• EXPORT
 - (7) Write equivalent MongoDB queries of following SQL queries.
 - i. Select first name, salary from employee where designation ="Manager".
 - ii. Select * from employee where designation ="Manager".
 - (8) What is Partitioning and bucketing in Hive?
 - (9) What is RDD in spark? what is the importance of Lazy evaluation in spark?
 - (10) Suppose we have a tuple in Student collection is of the following form:
{sem:"3", Name: 'Bob', Branch: 'IT', year: 2019, Sub:'Big Data'}
 - i. Create NoSQL query to display the students those have opted for the subject Big Data.
 - ii. Insert a data of student as sem:3, Alice, IT, 2019, Cyber Security.
 - (11) What data types are supported by pig?
 - (12) What is the use of \$hadoop namenode -format command?
