**POORNIMA UNIVERSITY, JAIPUR**

**END SEMESTER EXAMINATION, November 2022**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **4BT7153** | Roll No. | Total Printed Pages: 2 |
| **4BT7153** |  |
| B. Tech. IV Year VII- Semester (Main/Back) End Semester Examination, November 2022  **(DS)** | |
| **BDS07102 : Advanced Big Data Analytics** | | | |

# Time: **3** Hours. Total Marks: **60**

Min. Passing Marks: **21**

Attempt **five** questions selecting one question from each Unit. There is an internal choice from Unit I to Unit V. Marks of each question or its parts are indicated against each question/part. Draw neat sketches wherever necessary to illustrate the answer. Assume missing data suitably (if any) and clearly indicate the same in the answer.

Use of the following supporting material is permitted during the examination for this subject.

# **1.--------------------------Nil--------------------** **2.------------------Nil-----------------------**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **UNIT-I (CO1)** | **Marks** | **Bloom Level** |
| **Q.1** | **(a)** | What Is Apache Sqoop? How does Apache Sqoop make data transfer so easy? Explain Sqoop Features. | **(6)** | **Knowledge** |
|  |  |  |  |  |
|  | **(b)** | What do you mean by Hive? Explain various steps involved in Importing the RDBMS Table in Hive using a suitable Sqoop Example. | **(6)** | **Create** |
|  |  |  |  |  |
|  |  | **OR** |  |  |
|  |  |  |  |  |
| **Q.2** | **(a)** | Elaborate on the Apache Sqoop Architecture and Internal Working. Describe the need and advantages of Apache Sqoop. | **(6)** | **Understanding** |
|  |  |  |  |  |
|  | **(b)** | Illustrate the various steps of Apache Sqoop Import flow of execution and Export data from Hadoop into traditional relational databases. | **(6)** | **Analyse** |
|  |  |  |  |  |
|  |  | **UNIT-II (CO2)** |  |  |
|  |  |  |  |  |
| **Q.3** | **(a)** | What do you mean by Apache Flume? Define its advantages and features. Also, specify its applications. | **(6)** | **Knowledge** |
|  |  |  |  |  |
|  | **(b)** | Define the following terms:  (i) HDFS Sink (ii) Partitioning (iii) Interceptors | **(6)** | **Understanding** |
|  |  |  |  |  |
|  |  | **OR** |  |  |
|  |  |  |  |  |
| **Q.4** | **(a)** | Illustrate the Apache Flume Architecture with suitable diagram and Flume Event, Agent. Define the components namely, source, channel, and sink. | **(6)** | **Analyze** |
|  |  |  |  |  |
|  | **(b)** | Define the following terms:  (i) File Formats (ii) FAN Out (iii) Transactions & Reliability | **(6)** | **Understanding** |
|  |  |  |  |  |
|  |  | **UNIT-III (CO3)** |  |  |
|  |  |  |  |  |
| **Q.5** | **(a)** | What is Apache Spark? Explain the Evolution and Features of Apache Spark. | **(6)** | **Knowledge** |
|  |  |  |  |  |
|  | **(b)** | Illustrate the Resilient Distributed Datasets (RDD) data structure of Spark. Define the terms parallelizing and MapReduce. | **(6)** | **Analyse** |
|  |  |  |  |  |
|  |  | **OR** |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| **Q.6** | **(a)** | Examine how Apache Spark works. Using a suitable diagram explain the Architecture of Apache Spark. | **(6)** | **Apply** |
|  |  |  |  |  |
|  | **(b)** | Explain the recommending steps involved in Apache Spark Deployment using a suitable example. | **(6)** | **Evaluate** |
|  |  |  |  |  |
|  |  | **UNIT-IV (CO4)** |  |  |
|  |  |  |  |  |
| **Q.7** | **(a)** | What is a Spark DataFrame? How does a DataFrame Work? Define the various Features of Spark DataFrames. | **(6)** | **Understanding** |
|  |  |  |  |  |
|  | **(b)** | Define the following:  (i) Classification with Naïve Bayes (ii) Clustering with K-Means  (iii) Artificial Neural Networks. | **(6)** | **Knowledge** |
|  |  |  |  |  |
|  |  | **OR** |  |  |
|  |  |  |  |  |
| **Q.8** | **(a)** | Demonstrate the various steps to create a Spark DataFrame. Explain the manipulation methods using domain-specific language and SQL queries. | **(6)** | **Apply** |
|  |  |  |  |  |
|  | **(b)** | What is Apache Spark Structured Streaming? What streaming sources and sinks does Databricks support? | **(6)** | **Understanding** |
|  |  |  |  |  |
|  |  | **UNIT V (CO5)** |  |  |
|  |  |  |  |  |
| **Q.9** | **(a)** | What is a streamingcontext? Define streamingcontext class in spark? What happens when a streamingcontext is stopped? | **(6)** | **Knowledge** |
|  |  |  |  |  |
|  | **(b)** | How do we import and process the data from any file format. Explain in brief the processing of data for text files, JSON files and parquet files. | **(6)** | **Create** |
|  |  |  |  |  |
|  |  | **OR** |  |  |
|  |  |  |  |  |
| **Q.10** | **(a)** | What is Discretized Streams (DStream) and how it works? What is the difference between Dstream and Structured Streaming? | **(6)** | **Evaluate** |
|  |  |  |  |  |
|  | **(b)** | Explain the following terms:  (i) TCP streams (ii) File streams (iii) SQL Context Datasets | **(6)** | **Analyse** |