PVP 20

Code: 20DS3602

III B.Tech - II Semester – Regular Examinations - APRIL 2025

BIG DATA ANALYTICS
(DATA SCIENCE)

Duration: 3 hours Max. Marks: 70

Note: 1. This paper contains questions from 5 units of Syllabus. Each unit carries 14 marks and have an internal choice of Questions.
2. All parts of Question must be answered in one place.

BL – Blooms Level CO – Course Outcome

			BL	CO	Max. Marks
UNIT-I					
1	a)	Illustrate the working principle of Data Analytics Life cycle with neat diagram.	L2	CO1	7 M
	b)	Explain the concept of data processing in Hadoop.	L2	CO1	7 M
OR					
2	a)	Explain the concept of HDFS and relate it with any one real time example.	L2	CO1	7 M
	b)	Illustrate the classification of Data Analytics.	L2	CO1	7 M
UNIT-II					
3	a)	Apply Map-Reduce and explain how data is processed in Hadoop.	L3	CO2	7 M
	b)	Demonstrate the architecture of Map-Reduce and explain it.	L2	CO2	7 M

OR					
4	a)	Differentiate Map task and Reduce task.	L2	CO2	7 M
	b)	Classify the types of Map-Reduce and explain their usage.	L4	CO4	7 M
UNIT-III					
5	a)	Differentiate PIG latin and HIVE. Explain how it relates with real time examples.	L3	CO4	7 M
	b)	Analyze the HIVE services and relate it with real time example.	L4	CO4	7 M
OR					
6	a)	Demonstrate the application of user defined functions in HIVE.	L3	CO2	7 M
	b)	Explain the concept of Tables in HIVE using DDL.	L2	CO2	7 M
UNIT-IV					
7	a)	Demonstrate the features of SPARK.	L2	CO3	7 M
	b)	Explain the concept of Spark Ecosystem.	L1	CO3	7 M
OR					
8	a)	Illustrate the need of Transformations and actions in Resilient Distributed Dataset with an example.	L3	CO3	7 M
	b)	Explain the concept of Persistence and caching.	L1	CO3	7 M
UNIT-V					
9	a)	Analyze the Operations Performed on Data Frames.	L4	CO3	7 M

	b)	Explain how the data frames are created and their transforming.	L2	CO3	7 M
OR					
10	a)	Analyze the Role of Spark SQL in the Spark ecosystem.	L4	CO4	7 M
	b)	Explain the file formats in Spark with example.	L2	CO3	7 M