LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION – **DATA SCIENCE**

SECOND SEMESTER – **APRIL 2022**

PDS 2510 – BIG DATA ANALYTICS

Date: 20-06-2022   Dept. No.   Max. : 100 Marks
Time: 09:00 AM - 12:00 NOON

PART – A

Q. No	**Answer ALL Questions**	**(10 x 2 = 20 Marks)**

1.   What is Map Reduce Paradigm?
2.   What is mahout in Hadoop?
3.   How Apache Spark is different from Hadoop?
4.   What are the Language Supported by Spark?
5.   What are the data formats supported by Spark?
6.   What is Spark Session?
7.   What is the difference between repartition and coalesce
8.   What is Consumer and producer in Spark streaming
9.   How can you achieve machine learning in Spark?
10. What is Silhouette in K-means clustering

PART – B

Answer FIVE Questions   **(5 x 8 = 40 Marks)**

11.   i)Explain Big Hadoop Distributed File System

Or

ii) Explainis Big Data Analytics and its Applications.

12.   i)Write short notes on Spark Architecture

Or

ii)Explain Local Variables , Broadcast Variables and Accumulators

13.   i)Explain Data Frames operations Filter, Join, GroupBy, Agg operations

Or

ii)Explain Operations on Data Frame

14.   i)Write notes on Flume and Kafka

Or

ii) Explain DStreams operations.

15.   i)Explain Principal Component Analysis in Spark Machine Learning

Or

ii)Explains Recommendation Engines in Spark Machine Learning

PART – C

Answer any TWO Question   **(2 x 20 = 40 Marks)**

16.   i)Explain Apache Hadoop Ecosystem
      ii) Explain RDD and its Operations
17.   i)Explain Temporary Tables, Viewing and Querying Temporary Tables.
      ii) Explain Dstream RDDs and Dstream Processing
18.   i)Explain Linear Regression in Spark
      ii)Explain Random Forest Classification

@@@@@@@

1