

Question Paper

Exam Date & Time: 10-Feb-2021 (10:00 AM - 01:15 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

MANIPAL SCHOOL OF INFORMATION SCIENCES, MANIPAL
FIRST SEMESTER MASTER OF ENGINEERING- ME (BIG DATA AND DATA ANALYTICS / MACHINE LEARNING)
DEGREE EXAMINATION - FEBRUARY 2021
Architecture of Big Data Systems [BDA 623]

Marks: 100

Duration: 180 mins.

WEDNESDAY, FEBRUARY 10, 2021

Answer all the questions.

- 1) Explain five major challenges associated with Big Data systems. (TLO:1.1) (10 marks) (10)
- 2) Explain the typical problems encountered when scaling a traditional database. (TLO:1.2) (10 marks) (10)
- 3) Explain lambda architecture with different layers. (TLO:1.2) (10 marks) (10)
- 4) With sketch explain write operations in HDFS. (TLO:2.1) (10 marks) (10)
- 5) What are the different failure modes when running MapReduce jobs? Briefly explain each. (TLO:2.2) (10 marks) (10)
- 6) What are the responsibilities of YARN? Name the different components of YARN and state their tasks. (TLO:2.2) (3+7 marks) (10)
- 7) What are transformation and Actions in Spark? What is Lazy evaluation in Spark? Explain with an example. (TLO:3.1) (5+3+2 marks) (10)
- 8) Explain Continuous and Micro-Batch Stream Processing methods. What are their advantages and disadvantages? (TLO:3.2) (6+4 marks) (10)
- 9) Write Sqoop / Hive commands for the following operations (TLO:4.1) (2.5 x 4 marks) (10)
 - a. Importing multiple tables into HDFS, excluding specified table (s)
 - b. Export the table from HDFS to local database
 - c. Create managed table in HIVE and load data from HDFS
 - d. Sqoop command to transfer data from table in local dB to HiveSpark
- 10) Write PySpark code to perform the following tasks for Movie dataset. (TLO:4.1) (2.5 x 4 marks) (10)

Movie ID	Title	Type
1	Toy Story (1995)	Children

- a. Load data from csv file to Spark DataFrame.
- b. Add additional column "Year" extracting from column "Title".
- c. Create a UDF (user defined function) to add additional column "Decade". Input value for this column comes from "Year".
- d. List Movie title and Type for selected "Decade".

-----End-----