

Reg. No. :

--	--	--	--	--	--	--	--	--	--

Question Paper Code: 57203

B.E. / B.Tech. DEGREE EXAMINATION, DEC 2020

Seventh Semester

Computer Science and Engineering

15UCS703- DATA SCIENCE

(Regulation 2015)

Duration: One hour

Maximum: 30 Marks

PART A - (6 x 1 = 6 Marks)

(Answer any six of the following questions)

- Which of the following is one of the key data science skill ? CO1-R
(a) Statistics (b) Machine Learning
(c) Data Visualization (d) All of the Mentioned
- The most convenient way to use R is at a graphics workstation running a _____ system. CO1-R
(a) windowing (b) running
(c) interfacing (d) all of the mentioned
- The most commonly used measure of similarity is the _____ or its square CO2-R
(a) euclidean distance (b) city-block distance
(c) Chebychev's distance (d) Manhattan distance
- Decision Nodes are represented by _____. CO2-R
(a) Disks. (b) Squares (c) Circles (d) Triangles
- Which ecosystem project is ideal for use when we have multiple MapReduce and Pig programs to run in a sequence? CO3-R
(a) Oozie (b) Pig (c) Hive (d) Sqoop

6. What does Job Tracker do? CO3-R
 (a) Stores blocks of (b) Coordinates and schedules the job
 (c) Stores metadata (d) Acts as a mini reducer
7. Decision Nodes are represented by _____. CO4-R
 (a) Disks. (b) Squares. (c) Circles. (d) Triangles
8. The output of the _____ is not sorted in the Mapreduce CO4-R
 framework for Hadoop.
 (a) Mapper (b) Cascader (c) Scalding (d) None of the mentioned
9. Which of the following phases occur simultaneously? CO5-R
 (a) Shuffle and Sort (b) Reduce and Sort
 (c) Shuffle and Map (d) All of the mentioned
10. _____ are highly resilient and eliminate the single-point-of- CO5-R
 failure risk with traditional Hadoop deployments
 (a) EMR (b) Isilon solutions (c) AWS (d) All of the mentioned

PART – B (3 x 8= 24 Marks)

(Answer any three of the following questions)

11. Draw the pie chart for given data which consists of 21, 62, 10, CO1- U (8)
 53, "London", "New York", "Singapore", "Mumbai".
12. Explain in detail about various components of Hadoop ecosystem CO2- U (8)
13. Explain big data from business Perspective CO3-U (8)
14. What are the problems with Hadoop and where does it fail to deliver? CO4 -U (8)
15. Write a mapreduce program to sort data by student name(Value).input CO5- U (8)
 data 1001 , john ,45
 1002 , jack , 39
 1003 , alex ,44
 1004 , smith , 38
 1005, bop , 33