

JAMIA HAMDARD

Department of Computer Science & Engineering
School of Engineering Sciences & Technology.
New Delhi-110062

BCA 5th Semester Examination 2022

Paper Name: Introduction to Data Science and Big Data

Paper Code: BCA-PE 523

Time: 3 Hours

Maximum Marks: 75

Note: Mobile Phones, programmable calculator, or any devices with memory are not allowed inside the Examination Hall. Follow proper numbering for answering the questions.

SECTION: A (35 Marks)

35x1=35

All questions are compulsory

Q.No.	Questions	CO	BL
1	Moore's Law basically says that technology always gets cheaper True/False	1	BL1
2	Predictive analysis technique is (a) Classification (b) Univariate distribution (c) Central Tendency (d) Association Rules	3	BL2
3	----- Lifecycle Marketing really starts with the capture of customer permission.	2	BL4
4	Years of customer records and transactions can be categorised under <i>high volume</i> . True/False	2	BL2
5	Another approach to solving ----- with Big Data is social network analysis (SNA).	3	BL1
6	A highly sophisticated set of processes in which "insights" are made "actionable" via automated "decisions" is called -----.	1	BL4
7	Quantitative risk analysts help banks develop trading rules and implement these rules using modern technology True/False	4	BL4
8	Parallel processing can be provided by -----.	4	BL5
9	MapReduce works by breaking the processing into : (a) Jump (b) Grant (c) Revoke (d) None	3	BL6
10	Map and Reduce has key-value pairs as input and output True/False	4	BL3
11	if splits are too small, then the overhead of managing the splits and of map task decrease the total job execution time True/False	2	BL4
12	A disk has a block size, which is the maximum amount of data that it can read or write True/False	5	BL5
13	HDFS, too, has the concept of a block, but it is a much larger unit ----- MB by default	3	BL6
14	----- keeps a copy of the merged namespace image, which can be used in the event of the namenode failing	5	BL2
15	HDFS Federation, introduced in the ----- release series	1	BL3
16	On large clusters with many files and blocks, the time it takes for a namenode to start from cold can be ----- minutes or more.	2	BL2
17	The transition from the active namenode to the standby is managed by a new entity in the system called the ----- controller	3	BL1
18	Failover controllers are pluggable, but the first implementation uses ----- to ensure that only one namenode is active.	1	BL2
19	Define <i>graceful failover</i> .	2	BL4
20	Define <i>ungraceful failover</i> .	4	BL5
21	Define <i>fencing</i> .	5	BL6
22	Define checksum	3	BL3
23	Define CRC	3	BL2
24	Define Hamming code	2	BL4
25	Why bzip2 is preferred in Hadoop	1	BL5
26	JobSubmitter asks the jobtracker for a new job ID by calling-----.	1	BL6
27	JobSubmitter checks the ----- directory has not been specified or it already exists.	2	BL3
28	JobSubmitter computes the ----- splits for the job.	5	BL3
29	For very large clusters in the region of 4000 nodes and higher ----- can be used.	5	BL2
30	In 2010 a group at ----- began to design the next generation of MapReduce named as YARN.	4	BL2
31	(a) Google (b) Apache (c) Yahoo! (d) Facebook	2	BL4
32	----- is used to import and export data to and from between HDFS and RDBMS.	3	BL5
33	----- is a procedural language platform used to develop a script for MapReduce operations.	5	BL4
34	----- is a platform used to develop SQL type scripts to do MapReduce operations.	1	BL1
35	HBase is a column-oriented database and data is stored in tables True/False	5	BL3

SECTION-B

Attempt any **FOUR** Questions.

(5x4=20 Marks)

Q.No.	Questions	CO	BL
1	What is big data and why it is important?	1	BL1
2	Explain the input and output format of Map and Reduce functions using example.	2	BL5
3	What is the effect of large number of small size files in the Hadoop Ecosystem. Define rack awareness.	2	BL4
4	Draw the architecture of Hive and explain	5	BL1
5	Define data integrity and data compression the context of Hadoop I/O.	3	BL3
6	Define the components of Cassandra. What is Gossip protocol?	1	BL2

SECTION-C

Attempt any **TWO** Questions.

(10x2=20 Marks)

Q.No.	Questions	CO	BL
1	What do you understand by Fraud Management in the context of big data. Support your answer by a case study. Explain big data and algorithmic trading.	2	BL4
2	Explain the components of HDFS. What is the role of HDFS Federation?	1	BL1
3	Explain MapReduce 1 with the help of block diagram. What is the difference between classic MapReduce and YARN.	3	BL3
4	Write short notes on a) Combiner function b) Hadoop Streaming and Hadoop Pipes	2	BL2

Course Outcomes (CO): from syllabus, BL* - Bloom's Taxonomy Levels (1: Remembering, 2: Understanding, 3: Applying, 4: Analyzing, 5: Evaluating, 6: Creating)

END