

**PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)**

**BCA DEGREE EXAMINATION MAY 2022
(Sixth Semester)**

Branch – **COMPUTER APPLICATIONS**

BIG DATA ANALYTICS

Time: Three Hours

Maximum: 75 Marks

SECTION-A (10 Marks)

Answer **ALL** questions

ALL questions carry **EQUAL** marks (10 x 1 = 10)

- 1 Which of the following is a correct statement?
 - (i) Machine learning emphasizes on prediction, based on well-known properties learned from the training data
 - (ii) Data Cleaning emphasizes on prediction, based on well-known properties learned from the training data
 - (iii) Both i and ii
 - (iv) None of these
- 2 Which storage subsystem can support massive data volumes of increasing size
 - (i) Extensibility
 - (ii) Fault tolerance
 - (iii) Scalability
 - (iv) High-speed I/O capacity
- 3 NOSQL means _____
 - (i) Not SQL
 - (ii) No Usage of SQL
 - (iii) Not Only SQL
 - (iv) Not for SQL
- 4 In Data Science project data acquisition step involves _____
 - (i) Acquiring data from various sources
 - (ii) Selecting dataset
 - (iii) Data preprocessing
 - (iv) Data modeling
- 5 _____ is an open source framework for storing data and running application on clusters of commodity hardware.
 - (i) HDFS
 - (ii) Hadoop
 - (iii) Map Reduce
 - (iv) Cloud
- 6 The data node and name node in HADOOP are _____.
 - (i) Worker Node and Master Node respectively
 - (ii) Master Node and Worker Node respectively
 - (iii) Both Worker Nodes
 - (iv) Both Master Nodes
- 7 _____ is a programming model for writing applications that can process Big Data in parallel on multiple nodes
 - (i) HDFS
 - (ii) Map Reduce
 - (iii) Hadoop
 - (iv) HIVE
- 8 _____ is a type of local reducer that groups similar data from the map phase to identifiable sets.
 - (i) Mapper
 - (ii) Reducer
 - (iii) Combiner
 - (iv) Partitioner
- 9 HIVE also support custom extensions written in _____.
 - (i) C
 - (ii) C++
 - (iii) Java
 - (iv) c#
- 10 Which of the following company has developed PIG?
 - (i) Google
 - (ii) Yahoo
 - (iii) Microsoft
 - (iv) Apple

Cont...

SECTION - B (25 Marks)

Answer ALL questions
ALL questions carry EQUAL Marks (5 x 5 = 25)

- 11 a Describe characteristics of Big Data.
OR
b Summarize about technology challenges for Big Data.
- 12 a Explain about the challenges that prevent business from capitalizing on Big Data.
OR
b Write in brief about advantages of NOSQL.
- 13 a Differentiate between RDBMS and Hadoop.
OR
b State the reasons for using Mongo DB.
- 14 a Explain the Features of CASSANDRA.
OR
b Write in brief about CQL DATA types.
- 15 a Write in brief about the HIVE file formats.
OR
b Explain in short about PIG Latin.

SECTION -C (40 Marks)

Answer ALL questions
ALL questions carry EQUAL Marks (5 x 8 = 40)

- 16 a Elucidate Big Data architecture in detail.
OR
b Explain in detail about Big Data applications.
- 17 a Define Data Science. Explain in detail about data scientist and their skills.
OR
b Elucidate about terminologies used in Big Data environment.
- 18 a Elucidate Distributed computing challenges in Hadoop.
OR
b Explicate about managing resources and applications with Hadoop YARN.
- 19 a Explain in detail about Combiner and Partitioner in Map reduce program.
OR
b Explain in detail about key spaces in CASSANDARA.
- 20 a Explain in detail about Architecture of HIVE.
OR
b Explain in detail about Relational operators in PIG.

Z-Z-Z

END