

A9.1-R5 : BIG DATA ANALYTICS USING HADOOP

अवधि : 03 घंटे

DURATION : 03 Hours

अधिकतम अंक : 100

MAXIMUM MARKS : 100

ओएमआर शीट सं. :

OMR Sheet No. :

रोल नं. :

Roll No. :

उत्तर-पुस्तिका सं. :

Answer Sheet No. :

परीक्षार्थी का नाम :

Name of Candidate :

परीक्षार्थी के हस्ताक्षर :

; Signature of Candidate :

परीक्षार्थियों के लिए निर्देश :	Instructions for Candidates :
कृपया प्रश्न-पुस्तिका, ओएमआर शीट एवं उत्तर-पुस्तिका में दिये गए निर्देशों को ध्यानपूर्वक पढ़ें।	Carefully read the instructions given on Question Paper, OMR Sheet and Answer Sheet.
प्रश्न-पुस्तिका अंग्रेजी भाषा में है। परीक्षार्थी उत्तर लिखने के लिए केवल अंग्रेजी भाषा का ही प्रयोग कर सकते हैं।	Question Paper is in English language. Candidate has to answer in English language only.
इस मॉड्यूल/पेपर के दो भाग हैं। भाग एक में चार प्रश्न और भाग दो में पाँच प्रश्न हैं।	There are TWO PARTS in this Module/Paper. PART ONE contains FOUR questions and PART TWO contains FIVE questions.
भाग एक "वैकल्पिक" प्रकार का है जिसके कुल अंक 40 हैं तथा भाग दो "व्यक्तिपरक" प्रकार का है और इसके कुल अंक 60 हैं।	PART ONE is Objective type and carries 40 Marks. PART TWO is Subjective type and carries 60 Marks.
भाग एक के उत्तर, इस प्रश्न-पत्र के साथ दी गई ओएमआर उत्तर-पुस्तिका पर, उसमें दिये गए अनुदेशों के अनुसार ही दिये जाने हैं। भाग दो की उत्तर-पुस्तिका में भाग एक के उत्तर नहीं दिये जाने चाहिए।	PART ONE is to be answered in the OMR ANSWER SHEET only, supplied with the question paper, as per the instructions contained therein. PART ONE is NOT to be answered in the answer book for PART TWO.
भाग एक के लिए अधिकतम समय सीमा एक घण्टा निर्धारित की गई है। भाग दो की उत्तर-पुस्तिका, भाग एक की उत्तर-पुस्तिका जमा कराने के पश्चात् दी जाएगी। तथापि, निर्धारित एक घंटे से पहले भाग एक पूरा करने वाले परीक्षार्थी भाग एक की उत्तर-पुस्तिका निरीक्षक को सौंपने के तुरंत बाद, भाग दो की उत्तर-पुस्तिका ले सकते हैं।	Maximum time allotted for PART ONE is ONE HOUR. Answer book for PART TWO will be supplied at the table when the Answer Sheet for PART ONE is returned. However, Candidates who complete PART ONE earlier than one hour, can collect the answer book for PART TWO immediately after handing over the Answer Sheet for PART ONE to the Invigilator.
परीक्षार्थी, उपस्थिति-पत्रिका पर हस्ताक्षर किए बिना और अपनी उत्तर-पुस्तिका, निरीक्षक को सौंपे बिना, परीक्षा हॉल/कमरा नहीं छोड़ सकते हैं। ऐसा नहीं करने पर, परीक्षार्थी को इस मॉड्यूल/पेपर में अयोग्य घोषित कर दिया जाएगा।	Candidate cannot leave the examination hall/room without signing on the attendance sheet and handing over his/her Answer Sheet to the invigilator. Failing in doing so, will amount to disqualification of Candidate in this Module/Paper.
प्रश्न-पुस्तिका को खोलने के निर्देश मिलने के पश्चात् एवं उत्तर लिखना आरम्भ करने से पहले उम्मीदवार यह जाँच कर सुनिश्चित कर लें कि प्रश्न-पुस्तिका प्रत्येक दृष्टि से संपूर्ण है।	After receiving the instruction to open the booklet and before starting to answer the questions, the candidate should ensure that the Question Booklet is complete in all respect.

जब तक आपसे कहा न जाए, तब तक प्रश्न-पुस्तिका न खोलें।

DO NOT OPEN THE QUESTION BOOKLET UNTIL YOU ARE TOLD TO DO SO.

PART-ONE

(Answer all Questions. Each question carries ONE mark)

1. Each question below gives a multiple choice of answers. Choose the most appropriate one and enter in the "OMR" answer sheet supplied with the question paper, following instructions therein.

(1x10)

- 1.1. Business Intelligence (BI) is a broad category of application programs which includes :
- (A) Decision support
 - (B) Data mining
 - (C) Online Analytical Processing (OLAP)
 - (D) All of the above
- 1.2. What are the correct features of a distributed database ?
- (A) Is always connected to the internet
 - (B) Always requires more than three machines
 - (C) Users see the data in one global schema.
 - (D) Have to specify the physical location of the data when an update is done
- 1.3. Large collection of files are called_____.
- (A) Fields
 - (B) Records
 - (C) Database
 - (D) Sectors
- 1.4. To refer to a specific instance of a relation in relational model, term used is known as :
- (A) Relation instance
 - (B) Relation entity
 - (C) Relation tuple
 - (D) Relation attribute
- 1.5. Which of the following concept of JAVA is achieved by combining methods and attribute into a class ?
- (A) Encapsulation
 - (B) Inheritance
 - (C) Polymorphism
 - (D) Abstraction

- 1.6. Which of the following is not the component of Apache HBase ?
- (A) Region Server
 - (B) HMaster
 - (C) ZooKeeper
 - (D) HDFS
- 1.7. Which of the following is the processing framework in Hadoop, that manages resources and provides an execution environment to the processes ?
- (A) YARN
 - (B) Resource Manager
 - (C) Node Manager
 - (D) Datanode
- 1.8. Although the Hadoop framework is implemented in JAVA, MapReduce applications need not to be written in which of the following programming language ?
- (A) Java
 - (B) C
 - (C) C#
 - (D) Python
- 1.9. The Mapper implementation processes one line at a time via which of the following method ?
- (A) Map
 - (B) Reduce
 - (C) Mapper
 - (D) Reducer
- 1.10. What are the various configuration parameters required to run a MapReduce job ?
- (A) Job's input locations in the distributed file system
 - (B) Job's output location in the distributed file system
 - (C) JAR file containing the mapper, reducer and driver classes
 - (D) All of the Above.

2. Each statement below is either TRUE or FALSE. Choose the most appropriate one and enter your choice in the "OMR" answer sheet supplied with the question paper, following instructions therein.

(1x10)

- 2.1 Business Intelligent is a category of database software that provides an interface to help users quickly and interactively scrutinize the results in a variety of dimensions of the data.
- 2.2 Garbage collection is manual process in JAVA.
- 2.3 The PRIMARY KEY uniquely identifies each record in a SQL database table.
- 2.4 A table can have multiple primary keys with single or multiple fields and Primary keys can contain NULL values.
- 2.5 Multiple records can be updated at a time using WHERE clause.
- 2.6 Public method is accessible only to subclasses of its parent class.
- 2.7 R runs only on Windows computing platform and operating system.
- 2.8 Name Node is the master node in the distributed environment and it maintains the metadata information for the blocks of data stored in HDFS.
- 2.9 Node Manager is installed on every Data Node and it is responsible for the execution of the task on every single Data Node.
- 2.10 In HDFS Data Blocks are stored on single machine in a cluster. Where as in NAS data is stored on a dedicated hardware.

3. Match words and phrases in column X with the closest related meaning / words(s) / phrase(s) in column Y. Enter your selection in the "OMR" answer sheet supplied with the question paper, following instructions therein.

(1x10)

Column X		Column Y	
3.1	OLAP	A.	List files
3.2	ALTER TABLE	B.	Display Directory
3.3	ls command	C.	Working Directory
3.4	pwd command	D.	Shell script
3.5	Data Analysis	E.	Business Intelligence (BI)
3.6	.bashrc	F.	MapReduce
3.7	OOPS concept	G.	Delete columns
3.8	NameNode	H.	Resource Manager
3.9	HDFS	I.	Slave nodes
3.10	Data Nodes	J.	Useful knowledge
		K.	Polymorphism
		L.	Master node
		M.	Insert columns

4. Each statement below has a blank space to fit one of the word(s) or phrase(s) in the list below. Choose the most appropriate option, enter your choice in the "OMR" answer sheet supplied with the question paper, following instructions therein.

(1x10)

A.	Select * from employee	B.	- 32768 to 32767	C.	Catch	D.	MapReduce
E.	Different class	F.	OLAP	G.	Delete from table_name ;	H.	Apache Hadoop
I.	Shell script	J.	- 128 to 127	K.	mkdir command	L.	cp command
M.	SQL						

- 4.1 _____ query will display all data from the table.
- 4.2 SQL Query to delete all rows in a table without deleting the table _____.
- 4.3 Exception generated in try Block is caught in _____ Block.
- 4.4 We use _____ to copy file to another directory and to copy a file from source to destination.
- 4.5 We will create a new directory named "abc" in the currently working directory using _____.
- 4.6 _____ is the range of short data type in JAVA.
- 4.7 _____ is the range of byte data type in JAVA.
- 4.8 A list is represented as a vector but can contain objects of _____.
- 4.9 HDFS is designed to work with _____ paradigm, where computation is moved to the data.
- 4.10 _____ is a framework which provides us various services or tools to store and process Big Data.

PART TWO

(Answer any FOUR Questions)

5. (A) What are OLTP and OLAP ? Explain difference between OLTP and OLAP.
- (B) What are the characteristics of data in the database ?
- (C) What is DBMS ? Explain advantages and disadvantages of DBMS. Explain different types of keys available in Relational Model.

(5+5+5)

6. (A) What are the key components of HBase ? Define the difference between Hive and HBase.
- (B) What do you understand about Hive ? Explain its usage with example.
- (C) What is HDFS ? Why should you use Hadoop Distributed File System ?

(4+5+6)

7. (A) What is Big Data ? Discuss roll of Apache Hadoop as a solution to Big Data.
- (B) What is MapReduce ? How is it different from Apache Pig ?

(8+7)

8. (A) Write down the steps involved into deploy a Big Data solution. Explain some important features of Hadoop.
- (B) Explain any five methods of String Buffer class with example.

(7+8)

9. (A) Write a program in JAVA that accepts students data such as Roll No, age, first name and last name. store the data in table using JDBC and also display data from the table.
- (B) What is JAVA database connectivity (JDBC) ? Explain steps involved into making JDBC connectivity using JAVA.

(8+7)

- o o o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK