Sl. No.

A9.1-R5 : BIG DATA ANALYTICS USING HADOOP

अवधि : 03 घंटे **अधिकतम अंक : 100**
DURATION : 03 Hours **MAXIMUM MARKS : 100**

ओएमआर शीट सं. : OMR Sheet No. :					

रोल नं. : **Roll No. :**

उत्तर-पुस्तिका सं. : **Answer Sheet No. :**

परीक्षार्थी का नाम : **परीक्षार्थी के हस्ताक्षर :**
Name of Candidate : _____ **Signature of Candidate :** _____

परीक्षार्थियों के लिए निर्देश :	**Instructions for Candidate :**
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 की भाषा अंग्रेजी है। परीक्षार्थी केवल अंग्रेजी भाषा में ही उत्तर दे सकता है।	Question Paper is in English language. Candidate can 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 है तथा **भाग दो** ''व्यक्तिपरक'' प्रकार का है और इसके कुल अंक 60 है।	**PART ONE** is Objective type and carries 40 Marks. **PART TWO** is Subjective type and carries 60 Marks.
भाग एक के उत्तर, **ओएमआर उत्तर-पुस्तिका** पर 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. **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 जाँच कर यह 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 the questions)

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 What does OLAP acronym stands for _____ ?

(A) Online Analyst Production

(B) Online Analytical Processing

(C) Online Level Active Provider

(D) Onstage Analytical Processing

1.2 The highest level in the hierarchy of data organization is called _____.

(A) Data bank

(B) Database

(C) Data file

(D) Data record

1.3 The information about data in a database is called _____.

(A) Metadata

(B) Hyper data

(C) Tera data

(D) None of the above

1.4 A table is a collection of relationships, there is a close correspondence between concept of _____.

(A) Tables and instances

(B) Tables and entries

(C) Tables and variables

(D) Tables and relations

1.5 Which of the following is used for Statistical analysis in R language ?

(A) RStudio

(B) Studio

(C) Heck

(D) KStudio

1.6 _____ is not the core components of Hadoop.

(A) HDFS

(B) MapReduce

(C) YARN

(D) Application

1.7 Applications can use the _____ to report progress and set application-level status.

(A) Partitioner

(B) OutputSplit

(C) Reporter

(D) InputSplit

1.8 RDBMS is used for OLTP (Online Transactional Processing) system while Hadoop is used for _____.

(A) Data discovery

(B) Data analytics

(C) OLAP system

(D) All of the above

1.9 NameNode is the master node in the distributed environment and it maintains the _____ for the blocks of data stored in HDFS.

(A) Metadata information

(B) Storing data

(C) Manages resources

(D) Task execution

1.10 _____ is a file-level computer data storage server connected to a computer network providing data access to a heterogeneous group of clients.

(A) Network-Attached Storage (NAS)

(B) Hadoop Distributed File System (HDFS)

(C) Network file system

(D) Shared disk file systems

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 Customer Relationship Management (CRM) entails all aspects of interaction that a company has with its customer.

2.2 In the relational model, relationships between relations or tables are created by using foreign keys.

2.3 Primary key must be made based on any one column and Primary keys must contain UNIQUE values.

2.4 If WHERE clause is missing in statement, all the records will be updated.

2.5 Business Intelligence help in gaining an advantage in terms of competitive markets with another close competitor.

2.6 Method overriding is combination of inheritance and polymorphism.

2.7 Public method can be accessed by calling object of the public class.

2.8 R has been reported to be running on modern tablets, phones, PDAs, and game consoles.

2.9 DataNodes are the slave nodes, which are responsible for storing data in the HDFS. NameNode manages all the DataNodes.

2.10 ResourceManager receives the processing requests, and then passes the parts of requests to corresponding NodeManagers.

3. **Match words and phrases in column X with the closest related meaning/ word(s)/phrase(s) in column Y. Enter your selection in the "OMR" answer sheet supplied with the question paper, following instructions therein.** **(1x10)**

X		Y	
3.1	rm command	**A.**	Encapsulation
3.2	cd command	**B.**	Rename files
3.3	.bashrc	**C.**	Not a Number
3.4	OOPS concept	**D.**	DataNode
3.5	NaN	**E.**	List files
3.6	NodeManager	**F.**	Shell configuration file
3.7	HDFS Data	**G.**	Dedicated
3.8	NAS data	**H.**	Remove files
3.9	ls command	**I.**	Distributed
3.10	pwd command	**J.**	Change directory
		K.	Not a Numeric
		L.	Display Directory
		M.	Program file

4. **Each statement below has a blank space to fit one of the word(s) or phrase(s) in the list below. Enter your choice in the "OMR" answer sheet attached to the question paper, following instructions therein.** **(1x10)**

A.	0 to 65535	**B.**	Apache Hadoop	**C.**	Box obj = new Box();
D.	RDBMS	**E.**	Same class	**F.**	DataNode
G.	SQL JOIN	**H.**	Apache Pig	**I.**	– 128 to 127
J.	SQL UNION	**K.**	chmod command	**L.**	touch command
M.	Display Directory				

4.1 _____ to combine rows from two or more tables based on a related column between them.

4.2 The _____ can be used to modify and change the access/modification timestamps of specific files.

4.3 To combine the result-set of two or more SELECT statements, SQL clause used is _____.

4.4 In order to set the permission for a file and a folder, we can use _____.

4.5 _____ is the numerical range of a char data type in Java.

4.6 _____ is a valid declaration of an object of class Box.

4.7 In R language, a vector is defined that it can only contain objects of the _____.

4.8 _____ helps in analyzing Big Data and making business decisions out of it, which can't be done efficiently and effectively using traditional systems.

4.9 _____ relies on the structured data and the schema of the data is always known.

4.10 _____ provides many built-in operators to support data operations like joins, filters, ordering, sorting etc.

PART TWO

(Answer any FOUR questions)

5. (a) What do you understand by Business Intelligent ? How is it useful for business ?

(b) Explain in brief about OLAP and its working. What are functionalities of OLAP ?

(c) Compare file-oriented approach versus Database-oriented approach for Data Management.

(5+5+5)

6. (a) What are the different components of DBMS ? Explain different types of databases.

(b) What is the definition of Hive ? Differentiate between Pig and Hive.

(c) What is HDFS ? Explain the differences between relational database and HDFS with examples.

(4+5+6)

7. (a) What do you understand about the term "Big Data" ? How big data and Hadoop are related to each other ?

(b) What is Hadoop Map Reduce ? Explain its working with examples.

(7+8)

8. (a) What is Java DataBase Connectivity (JDBC) ? Explain with examples its importance in steps using JAVA.

(b) What do you understand by Java ? What difference between constructor and method ?

(7+8)

9. (a) What are OLTP and OLAP ? Explain difference between OLTP and OLAP.

(b) Explain the difference between :

I. NameNode, Backup Node and Checkpoint NameNode.

II. NAS and HDFS.

(8+7)

- o O o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK