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 attached to the question paper, following instructions therein. **(1x10)**

1.1 Which of the following is true about inheritance in Java ?

(1) Private methods are final.

(2) Protected members are accessible within a package and inherited classes outside the package.

(3) Protected methods are final.

(4) We cannot override private methods.

(A) (1), (2) and (4)

(B) Only (1) and (2)

(C) (1), (2) and (3)

(D) (2), (3) and (4)

1.2 What will be the output of the following Java code ?

```
int arr[ ] = new int [10];

System.out.println(arr);
```

(A) 0

(B) value stored in arr[0]

(C) 00000

(D) class name@ hashcode in hexadecimal form

1.3 Which one of the following attributes can be taken as a primary key ?

(A) Name

(B) Street

(C) Id Number

(D) Department

1.4 What is type of XML and JSON data ?

(A) Structured

(B) Semi-structured

(C) Unstructured

(D) None of the above

1.5 What are the advantages of 3x replication schema in Hadoop ?

(A) Fault Tolerance

(B) High Availability

(C) Reliability

(D) All of the above

1.6 When one of the join tables is small enough to fit into memory, the type of join used by Hive is ?

(A) Inner Join

(B) Map Join

(C) Reduce Join

(D) Sort Join

1.7 A View in Hive can be dropped by using :

(A) Drop table

(B) Drop view

(C) Delete view

(D) Remove view

1.8 What is the disadvantage of using too many partitions in Hive tables ?

(A) It slows down the namenode

(B) Storage space is wasted

(C) Join quires become slow

(D) All of the above

1.9 Which one of the vector types is not available in R programming ?

(A) double

(B) long

(C) complex

(D) character

1.10 Which command is used to remove the directory ?

(A) rdir

(B) rmdir

(C) remove

(D) rd

2. Each statement below is either TRUE or FALSE. Choose the most appropriate one and ENTER in the "OMR" sheet attached to the question paper, following instructions therein. **(1x10)**

2.1 chmod command is used to change the access mode of a file.

2.2 pkg is used to define packages in Java.

2.3 The core-site.xml file informs Hadoop daemon where NameNode runs in the cluster.

2.4 Hive cannot offer online transaction processing.

2.5 Structs is one of the complex data type in HIVE.

2.6 TRUNCATE Table command removes the metadata and data for a particular table.

2.7 A .class file contains bytecodes.

2.8 Data locality means moving computation to data instead of data to computation.

2.9 Garbage Collection is manual process in Java.

2.10 Referential Integrity in DBMS defines that a foreign key have a matching primary key.

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

X		Y	
3.1	To identify interesting patterns and relationships from the data	**A.**	limit
3.2	Operator that can be used to concatenate two or more String objects in java	**B.**	Hadoop Distributed File system
3.3	A drop-down list-user can select an option from that specified list	**C.**	Data Mining
3.4	To check the Hadoop services are up and running	**D.**	JComboBox
3.5	Hive SerDe stands for	**E.**	+
3.6	Bitwise XOR operator	**F.**	Hadoop Distributed Folder system
3.7	HDFS	**G.**	jps
3.8	To check the health of the file system	**H.**	hdfs fchk
3.9	Affordable devices that are generally compatible with other devices	**I.**	Serializer and Deserializer
3.10	To constrain the number of rows in hive	**J.**	Commodity Hardware
		K.	hdfs fsck
		L.	Serialize and Destruct
		M.	^

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

A.	primary	**B.**	column-oriented	**C.**	MapReduce
D.	data mart	**E.**	a directory	**F.**	SQL
G.	row-oriented	**H.**	Next	**I.**	TIMESTAMP
J.	catch	**K.**	max()	**L.**	a file
M.	JSON				

4.1 An attribute in a relation is a foreign key if the _____ key from one relation is used as an attribute in that relation.

4.2 A _____ is a subset of the data warehouse.

4.3 Exception generated in try block is caught in _____ block.

4.4 _____ is a programming model suitable for processing of huge data.

4.5 _____ is used to skip an iteration of a loop in R.

4.6 Each database created in hive is stored as _____.

4.7 Apache Pig supports pig latin language, which has _____ like command structure.

4.8 HBase is a distributed _____ database built on top of the Hadoop file system.

4.9 The _____ data types stores date in hive.

4.10 _____ function is used to find maximum value from given data frame in R.

PART TWO

(Answer any FOUR questions)

5. (a) What is Data Warehouse ? Explain difference between OLAP and OLTP.

(b) Explain any five methods of StringBuffer class with example.

(c) What is Big Data ? Explain characteristics of Big Data. **(5+5+5)**

6. (a) What are the advantages of Hadoop ? Explain Hadoop Architecture.

(b) Explain the use of throw using example.

(c) List any four R-Data types with brief explanation. **(6+5+4)**

7. (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) Explain write operation in HDFS. **(10+5)**

8. (a) Explain HBase Architecture.

(b) Explain the built-in functions of HIVE. **(6+9)**

9. (a) Explain the architecture of HIVE.

(b) How to write a UDF function in Hive ? Write a UDF in HIVE to remove leading and trailing white spaces from given string. **(6+9)**

- o O o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK