

No. of Printed Pages : 8

A9.1-R5.1 : BIG DATA ANALYTICS USING HADOOP

DURATION : 03 Hours

MAXIMUM MARKS : 100

OMR Sheet No. :					
-----------------	--	--	--	--	--

Roll No. :

--	--	--	--	--	--

Answer Sheet No. :

--	--	--	--	--	--

Name of Candidate : _____ ; Signature of Candidate : _____

INSTRUCTIONS FOR CANDIDATES :

- Carefully read the instructions given on Question Paper, OMR Sheet and Answer Sheet.
- Question Paper is in English language. Candidate has to answer in English language only.
- There are **TWO PARTS** in this Module/Paper. **PART ONE** contains **FOUR** questions and **PART TWO** contains **FIVE** questions.
- **PART ONE** is Objective type and carries **40** Marks. **PART TWO** is Subjective type and carries **60** Marks.
- **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**.
- 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.
- **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.**
- After receiving the instruction to open the booklet and before answering the questions, the candidate should ensure that the Question Booklet is complete in all respects.

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 will be the output of below statements?

```
public class HelloWorld{  
public static void main(String []args){  
  
String s1 = "abc";  
String s2 = "def";  
  
System.out.println(s1.compareTo(s2));    }  
}
```

- (A) true
- (B) -3
- (C) 3
- (D) false

1.2 Assignment in R is provided by

- (A) <-
- (B) ==
- (C) =:
- (D) :=

1.3 A view in Hive can be seen by using

- (A) show tables
- (B) show views
- (C) describe views
- (D) (A) & (B) both

1.4 The number of maps is usually driven by the total size of :

- (A) Inputs
- (B) Outputs
- (C) Tasks
- (D) None of the above

1.5 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

1.6 Which of the join operations do not preserve non matched tuples.

- (A) Left outer join
- (B) Right outer join
- (C) Inner join
- (D) Cross Join

1.7 _____ maps input key/value pairs to a set of intermediate key/value pairs.

- (A) Mapper
- (B) Reducer
- (C) Both Mapper and Reducer
- (D) None of the mentioned

1.8 Out of Pig, Hive, and Jaql, which of the following attributes is specific to only Jaql ?

- (A) Translates high-level language to MapReduce jobs
- (B) Operates on JSON data structures
- (C) Dataflow language
- (D) Schema optional

1.9 Which one of the following is not a main component of HBase ?

- (A) Region Server
- (B) ZooKeeper
- (C) Nagios
- (D) Master Server

1.10 Which will be the output of following code in R ?

```
x <- -3
```

```
switch(6, 2 + 2, mean(1:10), rnorm(5))
```

- (A) 10
- (B) 1
- (C) NULL
- (D) All of the mentioned

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 A factor variable is a form of categorical variable that accepts either numeric or character string values.

2.2 rhive.desc.table function provides all the tables stored in hive.

2.3 Outer join is needed when you wish to include rows that do not have matching values.

2.4 StandBy Nodes and Active Nodes communicate with a group of light weight nodes to keep their state synchronized in Hadoop.

2.5 HBase does not support programmatic access using Java APIs.

2.6 A limited version of JSON Path is supported in Hive.

2.7 NoSQL belongs to the traditional database technology.

2.8 AWT stands for Ambient Window Toolkit in Java.

2.9 Map-side join is done in the map phase and done in memory.

2.10 The daemons associated with the MapReduce phase are map-tracker and task-trackers.

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	RHIVE	A	May or may not be NULL
3.2	YARN	B	HBase
3.3	MapFile class	C	Datawarehouse
3.4	NameNode	D	Interface in Hadoop
3.5	install.packages("<package_name>")	E	Data Flow Language
3.6	Pig Latin	F	Installs package in R
3.7	Writable	G	Metadata in Hadoop
3.8	Minor Compaction and Major Compaction.	H	Global resource management
3.9	Star schema	I	Always Not NULL
3.10	Foreign Key	J	HQL
		K	Perform a file-based mapping from keys to values
		L	Data Lake
		M	Data node

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.	primary	B.	COGROUP	C.	Protected
D.	Iterator	E.	REST and Thrift	F.	zookeeper
G.	Derby database	H.	rbind()	I.	Catalog
J.	FOREACH	K.	MapReduce	L.	Global
M.	Shuffling				

- 4.1 _____ traverses the elements into forward direction only in Java.
- 4.2 An attribute in a relation is a foreign key if the _____ key from one relation is used as an attribute in that relation.
- 4.3 The variables, methods and inner classes that are declared _____ are accessible to the sub-classes of the class in which they are declared.
- 4.4 _____ operator in Pig is used to work with multiple tuples.
- 4.5 HBase contains _____ and _____ interfaces.
- 4.6 Hive uses _____ to store metadata.
- 4.7 _____ mode of Pig is also known as the Hadoop mode.
- 4.8 The _____ is used to maintain the configuration information and communication between region servers and clients. It also provides distributed synchronization.
- 4.9 _____ tables are used to maintain the metadata information in HBase.
- 4.10 _____ operator is used for performing iteration in Pig.

PART TWO

(Answer any FOUR Questions)

5. (a) Differentiate between Network File System(NFS),Hadoop NameNode and JournalNode. Also mention that how is the distance between two nodes defined in Hadoop.

(b) Write a program in Java that shows the connectivity with the tables stored in Hive using create, drop and show commands Hive query language.

(5+10)

6. Consider the following tables and answer following queries in sql :

RESERVES (SID,BID,DAY), BOATS (BID, BNAME,COLOR),

SAILORS(SID,SNAME,RATING, AGE)

(i) Find the sids of all sailors who have reserved red boats but not green boats.

(ii) Find names of sailors who've reserved all boats.

(iii) Names of sailors who've reserved boat #103 :

(iv) Find sailors who've reserved at least one boat.

(v) Find the sid of all sailors who have reserved red boats or green boats.

(3+3+3+3+3)

7. (a) What is a data frame in R programming ? Create following data frame in R and extract 3rd and 5th rows with 1st and 3rd columns from created data frame. Also add a new column "country" in a given data frame.

Name	Score	Attempts	Qualify
Gandhi	12.5	1	yes
Dima	9.0	3	no
Modi	16.5	2	yes
Patel	12.0	3	no
Peehu	9.0	2	no
Sahu	20.0	3	yes
Siya	14.5	1	yes
Laura	13.5	1	no
Kevin	8.0	2	no
Jonas	19.0	1	yes

(b) Write R program to create a list containing a vector, a matrix and a list and give names to the elements in the list. Assign names to the elements as "Color", "Odd numbers", "Language(s)" respectively in the list. Access the first and second element of the list.

(9+6)

8. (a) Explain java swings and constructors. Write a sample program that creates a frame and button using constructor.

(b) What is overriding in Java? Explain with an example program.

(8+7)

9. Write a short Note on following:

(i) HIVE Data definition language (DDL) Commands

(ii) Big Data

(iii) Zookeeper

(5+5+5)

- o O o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK