Sl. No.

A9.1-R5 : 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 respect.

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

PART ONE

(Answer all the 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. Which command is used to show the current HBase user ?
(A) whoami
(B) ls
(C) pwd
(D) ps

1.2. Apache PIG provides to bags namely :
(A) inner and outer bag
(B) lower and upper bag
(C) left and right bag
(D) All of the above

1.3. Which one of the following is not an open source ?
(A) Hive
(B) PIG
(C) JAQL
(D) Linux

1.4. Jaql is proprietary language of :
(A) Yahoo
(B) Facebook
(C) IBM
(D) Google

1.5. Materialized assignment in Jaql is provided by :
(A) =
(B) = =
(C) = :
(D) : =

1.6. What will be the output of below statements ?

```
public class Test
{
    public static void main (String args[]) {
System.out.print (10 + 20 + "Javatpoint\t");
System.out.println("Javatpoint" + 10 + 20);
    } }
```

(A) 30Javatpoint Javatpoint1020
(B) -3 30Javatpoint Javatpoint30
(C) 1020Javatpoint Javatpoint1020
(D) 30Javatpoint Javatpoint30

1.7. If missing values in a csv file is represented by an exclamation mark ("!") and a question mark ("?"). Which of the codes below will read csv file correctly into R ?
(A) csv('Dataframe.csv')
(B) csv('Dataframe.csv',header=FALSE, sep=',',na.strings=c('?'))
(C) csv2 ('Dataframe. csv', header = FALSE, sep=',', na.strings=c('?','!'))
(D) dataframe('Dataframe.csv')

1.8. Java compiles to :
(A) Machine code
(B) Assembly code
(C) Byte code
(D) Source code

1.9. Which node refers to Master node in HADOOP ?
(A) inode
(B) DataNode
(C) NameNode
(D) Header Node

1.10. What will be the output of below statements ?

```
public class Main
{
    public static void main(String[] args) {
    String s1 = "Cat";
    String s2 = "Cat";
    String s3 = new String("Cat");

    System.out.print(s1==s2);
    System.out.print(s1==s3);}
}
```

(A) truetrue
(B) truefalse
(C) falsetrue
(D) falsefalse

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 HBASE is a non-relational database systems.

2.2 HABSE is key/value store.

2.3 Hive does not store metadata information using HDFS.

2.4 PIG is not a high-level data flow platform as compared to MapReduce.

2.5 MapReduce do not provide nested data types.

2.6 PIG does not provide nested data types like tuples, bags, and maps.

2.7 A unique key in relational database table can have null values.

2.8 Apache PIG provide bag data model.

2.9 Jaql has no support for record data type.

2.10 rhive.list.tables function provides the description of table stored in hive.

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	HBASE	**A.**	R
3.2	Hive	**B.**	Map-Reduce
3.3	JAQL	**C.**	Data warehouse
3.4	RMarkdown	**D.**	NoSQL based
3.5	Time Variant	**E.**	Handles JSON
3.6	Safe Mode	**F.**	To change permission
3.7	mv	**G.**	Displays the processes
3.8	ps	**H.**	Moves a file
3.9	Hadoop	**I.**	SQL based
3.10	chmod	**J.**	No modifications to NameNode in HDFS
		K.	Displays all users
		L.	Renames a file
		M.	Unique Key

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.	Datawarehouse	**B.**	Grunt Shell	**C.**	{}	**D.**	who
E.	HBase	**F.**	Sharding	**G.**	Sqoop	**H.**	rbind()
I.	Shiny	**J.**	NameNode	**K.**	Node	**L.**	ListIterator
M.	Relational						

4.1 _____ is Apache Pig interactive shell.

4.2 A bag in Pig is implemented using _____.

4.3 Hive is SQL _____ engine.

4.4 _____ can be used to join two data frames in R.

4.5 _____ is an R package that makes it easy to build interactive web apps. from R.

4.6 _____ is a non-relational column-oriented distributed database.

4.7 Hive uses _____ method for partition.

4.8 _____ stores all the file location information in Hadoop Distributed File System(HDFS).

4.9 _____ transfers the data between Relational database management and Hadoop HDFS.

4.10 _____ traverses the elements into forward and backward direction in Java.

PART TWO
(Answer any FOUR Questions)

5. Answer following queries in SQL :

PEAK (NAME, ELEV, DIFF, REGION), **CLIMBER** (NAME, SEX) **PARTICIPATED** (TRIP_ID, NAME), **CLIMBED** (TRIP_ID, PEAK, WHEN)

- PEAK lists the name of each peak, elevation, difficulty level for climbers, and the region where it is located in. CLIMBER gives climber name and gender.
- PARTICIPATED gives the set of climbers who participated in various climbing trips. The number of participants in each trip varies.
- CLIMBED tells which peaks were climbed on each climbing trips, along with the date that each peak was climbed.

(i) Who has gone on a climbing trip with "John" ?

(ii) Which females have climbed a peak having elevation of 14,000 ?

(iii) Who climbed their first peak after Jan 1, 2016 ?

(iv) Which peaks of difficulty level 5 (if any) remain unclimbed ?

(v) Who has climbed every single peak that "Gagan" climbed ?

(3x5)

6. (A) What are two core concepts of Hadoop ? Discuss the features of these core concepts in detail.

(B) Suppose a Sales datawarehouse in Hive has four dimensions: time, location, customer, and product, and two measures: Price (MRP) and Selling Cost. Answer the following queries using HiveQL.

(i) Find combination of month, city & customer which contain at least 50 items and whose average item price is at least Rs 800.

(ii) Find combination of month, city & customer which contain at least 200 items and whose profit is more than Rs. 6000.

(iii) Display the product earning second highest selling cost in Delhi in 2016.

(6+9)

7. (A) Write a Java program that takes a decimal number which converts to its binary equivalent and finds the length of the longest sequence of zeros in binary representation of that decimal number. E.g.

Number 7 has binary representation 111 and has no binary gaps.

Number 8 has binary representation 1000 and contains a binary gap of length 0.

Number 457 has binary representation 111001001 and contains a binary gap of length 2.

Number 40 has binary representation 101000 and contains one binary gap of length 1.

Number 12546 has binary representation 11000100000010 and contains highest binary gap of length 6.

(B) What is inheritance in Java. Explain with a example program.

(10+5)

8. (A) Write a R program to get all prime numbers up to N number.

(B) Write a R program to extract first 10 English letter in lower case and last 10 letters in upper case and extract letters between 22nd to 24th letters in upper case. (Use built-in datasets)

(10+5)

9. Write a short Note on following :

(A) RHIVE

(B) Exception Handling-in Java

(C) PIG

(5+5+5)

- o 0 o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK