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 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 attached to the question paper, following instructions therein.**

(1 x 10)

1.1 Data cube is also called as _____ when the generalization of cross tab represented visually.

(A) Two dimensional cube

(B) Multi-dimensional cube

(C) N-dimensional cube

(D) Cuboid

1.2 An attribute in a relation is which of the following key if the primary key from one relation is used as an attribute in that relation.

(A) Super

(B) Foreign

(C) Candidate

(D) Sub

1.3 Which of the following is not consider as a primitive data type in java ?

(A) Array

(B) Short

(C) Boolean

(D) Double

1.4 Which statement from the following is correct ?

(A) Hadoop do need specialized hardware to process the data

(B) Hadoop 2.0 allows live stream processing of real time data

(C) In Hadoop programming framework output files are divided into lines or records

(D) None of the above

1.5 Which of the following operator is used to execute a command from the Hive shell ?

(A) |

(B) !

(C) ^

(D) +

1.6 Which Package contains most fundamental functions to run R ?

(A) root

(B) child

(C) base

(D) parent

1.7 Pig operates in how many number of nodes ?

(A) 2

(B) 3

(C) 4

(D) 5

1.8 What is the default number of versions of a cell value which Hbase can store ?

(A) 2

(B) 3

(C) 4

(D) 5

1.9 Which of the following is not the type of meta store ?

(A) Embedded Metastore

(B) Local Metastore

(C) Remote Metastore

(D) Distributed Metastore

1.10 Jaql web server is allowing to executing Jaql scripts via which of the following calls ?

(A) REST API

(B) API

(C) JSON

(D) HIVE

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 The pivot operation is also known as Rotation.

2.2 A Field is a property of the entire relation, rather than of the individual tuples in which each tuple is unique.

2.3 Erase is the command to remove files.

2.4 HDFS works in a Master Slave Architecture.

2.5 Hive is developed by Facebook.

2.6 HDFS is implemented in Java language.

2.7 Hive Supports Subqueries.

2.8 Set is the command which is used to show values to keys in Pig.

2.9 HBase is Row oriented defines only column families.

2.10 The class which is at the top of the exception hierarchy is Throwable.

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 attached to the question paper, following instructions therein.**

(1x10)

	X		Y
3.1	ls	**A**	Actual storage space in HDFS
3.2	Namenode	**B**	single key or a group of multiple keys that can uniquely identify tuples in a table
3.3	Candidate Key	**C**	Complex Data Type of Hive
3.4	get	**D**	Fetches the contents of row or a cell in HBase.
3.5	Datanode	**E**	List Computer Files
3.6	Matrix	**F**	Mathematical Calculation
3.7	Throws	**G**	Manages the set of java database connectivity
3.8	Metadata	**H**	Master Node
3.9	Driver Manager	**I**	Shell for running pig command
3.10	Grunt shell	**J**	Command for Executing Hive
		K	To identify each unique record independently of any other data
		L	Data about data
		M	The Exception explicitly in the code

4. Each statement below has a blank space to fit one of the word(s) of phrases in the list below. Enter your choice in the "OMR" answer sheet attached to the question paper, following instructions therein. (1 x 10)

A.	echo	**B.**	Array	**C.**	Base
D.	Root	**E.**	ODBC	**F.**	Slicing
G.	void	**H.**	Data node	**I.**	Statistical
J.	Null	**K.**	msg	**L.**	JDBC
M.	JSON				

4.1 To Display the message on the terminal we use _____ command.

4.2 A _____ acts as the slave and is responsible for executing a task assigned to it by the JobTracker.

4.3 _____ is the return type of a method that does not return any value.

4.4 The process of viewing the cross-tab (Single dimensional) with a fixed value of one attribute is _____.

4.5 _____ method is used by data analysist to get insights form data.

4.6 Specific Commands of hive is run from baseline when _____ driver is used in Hive.

4.7 Most fundamental functions to run R is contained in _____ package.

4.8 The "function" and " var" are known as _____ in JavaScript.

4.9 _____ is one of the complex data type in Hive.

4.10 _____ is a text format for sorting and transporting of a data.

PART TWO

(Answer any FOUR questions)

5. (a) What is distributed file system ? Define HDFS and its main components.

(b) What is Exception ? List and explain the keywords that are used to handle the exception in a program.

(c) What is Relational Model ? Explain the various components of Relational model in brief.

(5+5+5)

6. (a) Define the term Big Data in brief.

(b) Why we are using swing ? Discuss any four components of it.

(c) What is the role of Keys in DBMS ? Explain primary, foreign, alternate, candidate key in DBMS.

(3+6+6)

7. (a) Explain DDL commands in HIVE.

(b) What is Apache Pig and why do we need it ? Draw and explain its architecture.

(c) Differentiate the OLAP and OLTP.

(5+6+4)

8. (a) What is Hive ? How Hive interact with Hadoop ? Draw diagram which depicts the workflow between Hive and Hadoop.

(b) What is the use of MapReduce ? Explain the phases of MapReduce with example. **(8+7)**

9. (a) Write short note on JAQL.

(b) What are the various merits of R Programming ? List out some of the functions of R Programming.

(c) Compare Business Analytics and Business Intelligence. **(5+5+5)**

- o O o -

SPACE FOR ROUGH WORK