

No. of Printed Pages : 8

**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.**

**(1x10)**

**1.1** Data warehouse is used commonly in which of the following systems ?

- (A) Transaction
- (B) Logical
- (C) Decision Support
- (D) None of Above

**1.2** The information about data in a database is called \_\_\_\_\_.

- (A) Metadata
- (B) Hyper data
- (C) Tera data
- (D) None of the above

**1.3** Which of the following class is containing all of the classes for creating user interfaces and for painting graphics and images ?

- (A) Java.awt
- (B) Java.applet
- (C) Java.string
- (D) Java.lang

**1.4** What is the default interval time for the heartbeat in the Hadoop frame work ?

- (A) 3 seconds
- (B) 5 seconds
- (C) 3 minutes
- (D) 5 minutes

**1.5** \_\_\_\_\_ is the duty of DataNode in HDFS.

- (A) Maintain the file system tree and metadata for all files and directories.
- (B) Control the execution of an individual map task or a reduce task.
- (C) Manage the file system namespace.
- (D) Store and retrieve blocks when told to by clients or the NameNode.

**1.6** The language which is used for Hive is \_\_\_\_\_.

- (A) HQL
- (B) SQL
- (C) HDFS
- (D) DFS

- 1.7 UDF stands for \_\_\_\_\_.
- (A) User Data Function
  - (B) User Data Flow
  - (C) User Define Function
  - (D) User Define Flow
- 1.8 Which of the following is the actual physical database that is associated with the hive service ?
- (A) Hive Server
  - (B) JDBC
  - (C) Hive Metadata
  - (D) Client Gateway
- 1.9 Which of the following function is used to call the code if it is stored in external file ?
- (A) exec()
  - (B) execute()
  - (C) source()
  - (D) sourcefile()
- 1.10 To read the data from PIG which of the following command is used ?
- (A) Read
  - (B) Run
  - (C) Write
  - (D) Load
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 OLTP uses array-based multidimensional storage engines for multidimensional views of data.
- 2.2 Try block is used to examine the Exception in java code.
- 2.3 MapTask part of the MapReduce is responsible for processing one or more chunks of data and producing the output results.
- 2.4 Apache Hive is a data warehouse software project built on top of Apache Hadoop for providing data query and analysis.
- 2.5 Business Intelligence help in gaining an advantage in terms of competitive markets with another close competitor.
- 2.6 Data is found in structural format only.
- 2.7 Cat command is used to view compressed text file contents.
- 2.8 There are four namespaces for variables in Hive.
- 2.9 Pig is a relational database with SQL support.
- 2.10 Beeline is a JDBC client that is based on the SQLLine CLI.

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	Load	A	function creates a regular sequence of values to form a vector in R
3.2	show tables	B	Deletes the file from HDFS recursively
3.3	dfsadmin	C	Primitive data type in Hive
3.4	rmr	D	Used to check Hadoop services are running or not
3.5	chgrp	E	Keyword Used in interface in java
3.6	Library()	F	Hive and JDBC
3.7	seq	G	To read the data in PIG
3.8	jps	H	Command changes a file's group owner
3.9	implements	I	A view in which Hive can be seen
3.10	Miscellaneous	J	Runs a DFS admin client
		K	Used to watch for all available package in library
		L	Name Node
		M	Command load the 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.	JobTracker	B.	set	C.	NameNode
D.	script	E.	scan	F.	Column
G.	255	H.	512	I.	commas
J.	Mapreduce	K.	touchz	L.	Functional
M.	FS shell				

- 4.1 By Default, \_\_\_\_\_ is the mode of Hive.
- 4.2 HDFS provides a command line interface called \_\_\_\_\_ used to interact with HDFS.
- 4.3 The maximum file size in the linux is \_\_\_\_\_ Bytes.
- 4.4 \_\_\_\_\_ function is used to enter in data at the terminal.
- 4.5 \_\_\_\_\_ command is used to show values to keys used in Pig.
- 4.6 To run PIG in batch mode, PIG \_\_\_\_\_ is used.
- 4.7 \_\_\_\_\_ is the command use to create an empty file.
- 4.8 Data is separated by \_\_\_\_\_ in JSON syntax.
- 4.9 JAQL is \_\_\_\_\_ data processing language.
- 4.10 \_\_\_\_\_ is responsible to schedules jobs and tracks the assign jobs to Task tracker.

**PART TWO**

**(Answer any FOUR Questions)**

5. (a) What do you understand by Business Intelligent ? How is it useful for business ?
- (b) What is inheritance ? How many types of inheritance is supported in java?
- (c) Explain any five basic commands in Linux.

**(5+5+5)**

6. (a) What are the basic data types in Hive ? Explain them.
- (b) What is Hadoop ? Explain some of the HDFS command.
- (c) Write down the differences between Apache Pig and Map-Reduce.

**(3+6+6)**

7. (a) Explain working of Hive with proper steps and diagram.
- (b) Discuss in brief that How Pig data model will help in effective data flow ?
- (c) List any four R-Data types with brief explanation.

**(5+6+4)**

8. (a) How data is stored in HBase? How to create table in HBase? What is row-oriented and column-oriented database ? Give difference between them.
- (b) What is JDBC-ODBC Bridge ? Explain JDBC Drivers in brief.

**(8+7)**

9. (a) Explain relational operators in Pig.
- (b) What is the difference between file oriented approach and database oriented approach?
- (c) Draw HDFS architecture and explain the components of Hadoop.

**(5+5+5)**

**- o O o -**

---

**SPACE FOR ROUGH WORK**

---

SPACE FOR ROUGH WORK