

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; 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 the instructions therein. (1x10)

1.1. Business Intelligence (BI) includes which of the following application programs ?

- (A) Decision support
- (B) Data mining
- (C) Online Analytical Processing (OLAP)
- (D) All of the above

1.2 Collection of files is commonly referred to as _____.

- (A) Fields
- (B) Records
- (C) Database
- (D) Sectors

1.3. Combining methods and attribute into a class in JAVA is achieved by _____.

- (A) Encapsulation
- (B) Inheritance
- (C) Polymorphism
- (D) Abstraction

1.4 What is the primary role of Pig Latin's 'LOAD' operator ?

- (A) To store data in a Pig relation
- (B) To load data into HBase tables
- (C) To load data from HDFS into a Pig relation
- (D) To perform a load-balancing operation

1.5 Which of the following components is responsible for handling data storage and retrieval in HBase ?

- (A) Region Server
- (B) NameNode
- (C) DataNode
- (D) ResourceManager

1.6 XML and JSON data is of which category ?

- (A) Structured
- (B) Semi-structured
- (C) Unstructured
- (D) None of the above

1.7 What is the primary function of Pig in the Hadoop ecosystem ?

- (A) Data visualization
- (B) Data integration
- (C) Data storage
- (D) Data processing

- | | |
|---|--|
| <p>1.8. A specific instance of a relation in relational model is referred as :</p> <p>(A) Relation instance</p> <p>(B) Relation entity</p> <p>(C) Relation tuple</p> <p>(D) Relation attribute</p> | <p>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 the instructions therein. (1x10)</p> |
| <p>1.9 HIVE-JDBC is primarily used for what purpose ?</p> <p>(A) Real-time data processing</p> <p>(B) Batch data processing</p> <p>(C) Interactive querying</p> <p>(D) Data storage</p> | <p>2.1 pkg is used to define packages in Java.</p> <p>2.2 A .class file contains bytecodes.</p> <p>2.3 Garbage Collection is manual process in Java.</p> <p>2.4 Referential Integrity in DBMS defines that a foreign key have a matching primary key.</p> <p>2.5 chmod command is used to change the access mode of a file.</p> <p>2.6 MapReduce is a distributed computing model for processing large datasets in parallel across a cluster of nodes.</p> <p>2.7 Apache Spark is the storage layer of the Hadoop ecosystem, designed to store large datasets reliably and in a distributed manner across a cluster of commodity hardware.</p> <p>2.8 Map function in MapReduce processes each input key-value pair and applies a user-defined operation to produce intermediate key-value pairs, typically involving filtering, transformation or aggregation.</p> <p>2.9 Shuffle phase in MapReduce is responsible for processing and combining the intermediate key-value pairs produced by the Map function.</p> <p>2.10 remove command is used to remove the directory.</p> |
| <p>1.10 What is the primary function of RHive ?</p> <p>(A) Data visualization</p> <p>(B) Data ingestion</p> <p>(C) Data analytics</p> <p>(D) Data integration</p> | |

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 the instructions therein. (1x10)

	X		Y
3.1	Operator that can be used to concatenate two or more String objects in java	A	Start, stop and restart server
3.2	To check the Hadoop services are up and running	B	Encapsulation
3.3	HDFS	C	plus (+) operator
3.4	Node Manager	D	read and write table rows
3.5	Bundle of fields and methods	E	Data mining
3.6	NaN	F	Slave node
3.7	Name node	G	Data Warehouse
3.8	Data Analysis	H	jps
3.9	Hive SerDe stands for	I	a distributed file system
3.10	To identify interesting patterns and relationships from the data	J	Master node
		K	Inspection, Cleaning, Transforming and Reporting
		L	DataNode
		M	Not a Number

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 supplied with the question paper, following the instructions therein. (1x10)

A.	Primary	B.	Data Mart	C.	Mkdir command
D.	Candidate key	E.	YARN	F.	Apache Hadoop
G.	Batch processing	H.	Garbage Collection	I.	Hive
J.	Select*from table name	K.	Catch	L.	Database
M.	MapReduce				

- 4.1 A _____ is a subset of the data warehouse.
- 4.2 _____ is a programming model suitable for processing of huge data.
- 4.3 _____ is the resource management layer of Hadoop.
- 4.4 Hadoop MapReduce is optimized for _____ of large datasets.
- 4.5 _____ process reclaim the clean up the storage in JAVA.
- 4.6 _____ query will display all data from the table.
- 4.7 Exception generated in try Block is caught in _____ Block.
- 4.8 _____ is a framework which provides us various services or tools to store and process Big Data.
- 4.9 We will create a new directory named "abc" in the currently working directory using _____.
- 4.10 An attribute in a relation is a foreign key if the _____ key from one relation is used as an attribute in that relation.

PART TWO

(Answer any FOUR questions)

5. (a) Explain the components of the architecture of HIVE.
(b) Explain UDF function in Hive. Write a UDF in HIVE to remove leading and trailing white spaces from given string.
(7+8)
6. (a) Write a program in JAVA that accepts employee data such as Employee No, age, first name and last name. Store the data in table using JDBC and also display data from the table.
(b) What is JAVA database connectivity (JDBC) ? Explain steps involved into making JDBC connectivity using JAVA.
(7+8)
7. (a) Write down the steps involved into deploy a Big Data solution. Explain some important features of Hadoop.
(b) Explain any five methods of String Buffer class with example.
(8+7)

8. (a) Discuss roll of Apache Hadoop as a solution to Big Data.
(b) What is MapReduce ? How is it different from Apache Pig ?
(7+8)
9. (a) How is Data Mining and Data Warehouse related with each other ? Explain difference between OLAP and OLTP.
(b) What is Big Data ? Explain characteristics of Big Data. Explain the applications of Big Data.
(7+8)

- o O o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK