

A9.1-R5 : BIG DATA ANALYTICS USING HADOOP

अवधि : 03 घंटे

DURATION : 03 Hours

अधिकतम अंक : 100

MAXIMUM MARKS : 100

ओएमआर शीट सं. :					
OMR Sheet No. :					

रोल नं. :

Roll No. :

उत्तर-पुस्तिका सं. :

Answer Sheet No. :

परीक्षार्थी का नाम :

Name of Candidate :

परीक्षार्थी के हस्ताक्षर :

Signature of Candidate :

परीक्षार्थियों के लिए निर्देश :

Instructions for Candidate :

कृपया प्रश्न-पुस्तिका, ओएमआर शीट एवं उत्तर-पुस्तिका में दिये गए निर्देशों को ध्यानपूर्वक पढ़ें।	Carefully read the instructions given on Question Paper, OMR Sheet and Answer Sheet.
प्रश्न-पुस्तिका की भाषा अंग्रेजी है। परीक्षार्थी केवल अंग्रेजी भाषा में ही उत्तर दे सकता है।	Question Paper is in English language. Candidate can answer in English language only.
इस मॉड्यूल/पेपर के दो भाग हैं। भाग एक में चार प्रश्न और भाग दो में पाँच प्रश्न हैं।	There are TWO PARTS in this Module/Paper. PART ONE contains FOUR questions and PART TWO contains FIVE questions.
भाग एक "वैकल्पिक" प्रकार का है जिसके कुल अंक 40 हैं तथा भाग दो "व्यक्तिपरक" प्रकार का है और इसके कुल अंक 60 हैं।	PART ONE is Objective type and carries 40 Marks. PART TWO is Subjective type and carries 60 Marks.
भाग एक के उत्तर, ओएमआर उत्तर-पुस्तिका पर ही दिये जाने हैं। भाग दो की उत्तर-पुस्तिका में भाग एक के उत्तर नहीं दिये जाने चाहिए।	PART ONE is to be answered in the OMR ANSWER SHEET only. PART ONE is NOT to be answered in the answer book for PART TWO.
भाग एक के लिए अधिकतम समय सीमा एक घण्टा निर्धारित की गई है। भाग दो की उत्तर-पुस्तिका, भाग एक की उत्तर-पुस्तिका जमा कराने के पश्चात् दी जाएगी। तथापि, निर्धारित एक घंटे से पहले भाग एक पूरा करने वाले परीक्षार्थी भाग एक की उत्तर-पुस्तिका निरीक्षक को सौंपने के तुरंत बाद, भाग दो की उत्तर-पुस्तिका ले सकते हैं।	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.
परीक्षार्थी, उपस्थिति-पत्रिका पर हस्ताक्षर किए बिना और अपनी उत्तर-पुस्तिका, निरीक्षक को सौंपे बिना, परीक्षा हॉल/कमरा नहीं छोड़ सकते हैं। ऐसा नहीं करने पर, परीक्षार्थी को इस मॉड्यूल/पेपर में अयोग्य घोषित कर दिया जाएगा।	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.
प्रश्न-पुस्तिका को खोलने के निर्देश मिलने के पश्चात् एवं उत्तर लिखना आरम्भ करने से पहले उम्मीदवार जाँच कर यह सुनिश्चित कर लें कि प्रश्न-पुस्तिका प्रत्येक दृष्टि से संपूर्ण है।	After receiving the instruction to open the booklet and before starting to answer the questions, the candidate should ensure that the Question Booklet is complete in all respect.

जब तक आपसे कहा न जाए, तब तक प्रश्न-पुस्तिका न खोलें।

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** Hadoop is a framework that works with a variety of related tools. Common cohorts include :
- (A) MapReduce, Hive and HBase
 - (B) MapReduce, MySQL and Google Apps
 - (C) MapReduce, Hummer and Iguana
 - (D) MapReduce, Heron and Trumpet
- 1.2** All of the following accurately describe Hadoop, except :
- (A) Open source
 - (B) Real-time
 - (C) Java-based
 - (D) Distributed computing approach
- 1.3** The number of maps is usually driven by the total size of :
- (A) Inputs
 - (B) Outputs
 - (C) Tasks
 - (D) None of the above

- 1.4** Point out the wrong statement.

- (A) A MapReduce job usually splits the input data-set into independent chunks which are processed by the map tasks in a completely parallel manner.
- (B) The MapReduce framework operates exclusively on <key, value> pairs.
- (C) Applications typically implement the Mapper and Reducer interfaces to provide the map and reduce methods.
- (D) None of the above.

- 1.5** _____ 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

- 1.6** Which of the following phases occur simultaneously ?

- (A) Shuffle and Sort
- (B) Reduce and Sort
- (C) Shuffle and Map
- (D) All of the mentioned

1.7 The _____ can also be used to distribute both jars and native libraries for use in the map and/or reduce tasks.

- (A) Data Cache
- (B) Distributed Data
- (C) Distributed Cache
- (D) All of the mentioned

1.8 HBase provides _____ like capabilities on top of Hadoop and HDFS.

- (A) Top Table
- (B) BigTop
- (C) Bigtable
- (D) None of the mentioned

1.9 HDFS works in a _____ fashion.

- (A) peer-to-peer
- (B) master-slave
- (C) worker/slave
- (D) all of the mentioned

1.10 For YARN, the _____ Manager UI provides host and port information.

- (A) Data Node
- (B) Name Node
- (C) Resource
- (D) Replication

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 Data Node is the slave/worker node and holds the user data in the form of Data Blocks.

2.2 MapReduce tries to place the data and the compute as close as possible.

2.3 The MapReduce framework operates exclusively on <key, value> pairs.

2.4 The output of the Reducer is re-sorted.

2.5 Super keyword must be used to inherit a class.

2.6 Main () method must be made static.

2.7 Yarn commands are invoked by the bin script.

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

2.9 When object of class is declared, each object contains its own copy of static variables.

2.10 The shell waits for the command to complete and normally can't do any work while the command is executing.

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	Metadata related to HDFS is stored on node	A.	DataNode
3.2	Used to store data blocks	B.	Three
3.3	Default replication factor of HDFS	C.	Yet Another Resource Negotiator
3.4	Shell utility which can be used to run Hive queries	D.	Name Node
3.5	YARN stands for	E.	\$HIVE_HOME/bin/hive
3.6	Arbitrates resources among all the applications in the system	F.	Recursion
3.7	Process of defining a method in terms of itself	G.	Polymorphism
3.8	Method overriding	H.	Resource Manager
3.9	Command for listing files in a directory	I.	Ctrl-Z
3.10	Command is used to suspend a job	J.	ls
		K.	Yet Another Remote Navigator
		L.	Encapsulation
		M.	Distributed

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.	super	B.	core-site.xml	C.	Application Master
D.	\$	E.	chmod +x	F.	shuffling
G.	>	H.	Method overriding	I.	creating
J.	MapReduce, Hive and HBase	K.	Method Overloading	L.	<
M.	#				

- 4.1 Output of the mapper is first written on the local disk for sorting and _____ process.
- 4.2 Hadoop is a framework that works with a variety of related tools. Common cohorts include _____.
- 4.3 The _____ is a framework-specific entity that negotiates resources from the Resource Manager.
- 4.4 _____ gives site-specific configuration for a given Hadoop installation.
- 4.5 _____ keyword can be used in a subclass to call the constructor of superclass.
- 4.6 _____ process of defining a method in a subclass having same name & type signature as a method in its superclass.
- 4.7 To run the script, we should make it executable first by using _____.
- 4.8 We can use the _____ symbol with to redirect our output to a specified file.
- 4.9 Apart from displaying file contents, cat command is also used for _____ files.
- 4.10 _____ is the default symbol for command prompt in Bourne shell.

PART TWO

(Answer any FOUR questions)

5. (a) Differentiate between throw and throws in JAVA with the help of a suitable example. Write a program in JAVA to declare an array of 2 elements. Further, the program throws an exception when the user tries to access the third element.

- (b) Explain the concept of keys in RDBMS taking suitable example.

(9+6)

6. (a) Consider a scenario where Bank is a class that provides functionality to get the rate of interest. However, the rate of interest varies according to banks. For example, SBI, ICICI and AXIS banks could provide 8%, 7%, and 9% rate of interest. Write a JAVA Program to demonstrate the above scenario.

- (b) Explain method overloading in JAVA with suitable example.

(10+5)

7. (a) Assume you have five files, and each file contains two columns that represent a city and the corresponding temperature recorded in that city for the various measurement days. For example : (Toronto, 20). Out of all the collected data, you want to find the maximum temperature for each city across the data files (note that each file might have the same city represented multiple times). Write a pseudo code in Map-Reduce for the same.

- (b) Explain the architecture of HBASE. Compare HBASE with HDFS.

- (c) Differentiate between XML and JSON. Give one example of each.

(6+4+5)

8. Explain the architecture of Hive. Further, highlight the interaction of Hive with Hadoop framework. (15)

9. Explain the following terms :

- (i) JSON File

- (ii) HDFS architecture

- (iii) Polymorphism

(5+5+5)

- o O o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK