No. of Printed Pages : **8**

# A9.1-R5.1 : 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** In IPv6, how does Stateless Address Autoconfiguration (SLAAC) assign IP addresses without a DHCP server ?

(A) By using the MAC address and EUI-48 format to generate a unique identifier.

(B) By randomly generating a 64-bit interface identifier.

(C) By using the network prefix and a link-local address to generate the full address.

(D) By communicating directly with the router to obtain an IP.

**1.2** Which HTML5 API allows a web page to retrieve data without refreshing the page, thus enabling asynchronous interaction ?

(A) WebSocket API

(B) XMLHttpRequest API

(C) Web Storage API

(D) Geolocation API

**1.3** Consider the following Python code :

```
def func (a, b=[]):
b. append(a) return b
print(func(1)) print(func(2))
```

What will be the output ?

(A) [1] [2]

(B) [1] [1, 2]

(C) Error

(D) [1] [2, 2]

**1.4** In an IoT system, which of the following is a major challenge for implementing IPv6 over Low Power Wireless Personal Area Networks (6LoWPAN) ?

(A) High data rate

(B) Power consumption

(C) Lack of routing protocols

(D) Inability to support large address spaces

**1.5** In a Binary Search Tree (BST), which traversal method returns the nodes in ascending order ?

(A) Preorder traversal

(B) Inorder traversal

(C) Postorder traversal

(D) Level-order traversal

**1.6** In a pipelined processor, which hazard occurs when the outcome of an instruction is not known before the next instruction is executed ?

(A) Structural hazard

(B) Data hazard

(C) Control hazard

(D) Memory hazard

**1.7** In a distributed database system, what is the primary reason for using Two-Phase Commit (2PC) protocol ?

(A) To ensure distributed database availability.

(B) To provide consistency in distributed transactions.

(C) To reduce the response time in distributed systems.

(D) To handle deadlock situations effectively.

**1.8** In Object-Oriented Design, which design pattern allows an object to alter its behaviour when its internal state changes, making it appear as if the object has changed its class ?

(A) Factory Pattern

(B) Observer Pattern

(C) State Pattern

(D) Command Pattern

**1.9** In the Hadoop ecosystem, which of the following ensures fault-tolerant processing by replicating blocks across multiple nodes ?

(A) HDFS

(B) MapReduce

(C) YARN

(D) Pig

**1.10** In Data Science, what is the significance of the cross_val_score() function in Python's scikit-learn library ?

(A) It splits the data into multiple random subsets for validation.

(B) It tunes hyperparameters using grid search.

(C) It computes cross-validated metrics to evaluate model performance.

(D) It generates predictions on new data.

**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)**

**2.1** In IPv6, a single network can have multiple unique addresses assigned to a single interface, but IPv4 strictly allows only one address per interface.

**2.2** In HTML5, the <canvas> element supports vector-based graphics rendering, making it ideal for creating the scalable image.

**2.3** CSS Grid Layout allows a developer to create complex layouts where items can overlap without specifying exact pixel positions.

**2.4** In Python, a lambda function can have multiple expressions in its body, similar to regular functions.

**2.5** In Python, the global keyword allows the modification of a variable defined in the global scope within a local function.

**2.6** Zigbee and LoRaWAN are both high-power, high-data-rate communication protocols widely used in IoT for long-range transmissions.

**2.7** In IoT systems, edge computing allows data to be processed closer to where it is generated, reducing latency and bandwidth usage.

**2.8** A splay tree is a type of binary search tree that self-adjusts so that frequently accessed nodes are moved closer to the root, improving access time over repeated operations.

**2.9** In a graph represented using an adjacency list, the space complexity for an undirected graph with n nodes and e edges is O(n + 2e).

**2.10** In relational databases, an index on a table's primary key guarantees that all SQL queries accessing the table will run faster.

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** | A system that manages resources and provides services to applications | **A** | Internet of Things (IoT) |
| **3.2** | Programming language primarily used for big data analysis and AI | **B** | MVC Framework |
| **3.3** | Technology that enables smart homes and automation | **C** | Web Development Framework |
| **3.4** | Relational database management systems example | **D** | Hadoop |
| **3.5** | Design methodology focusing on system components and their interactions | **E** | Operating System |
| **3.6** | Open-source framework used for managing and processing large data sets | **F** | Python |
| **3.7** | Conceptual modeling technique used to design database structures | **G** | SQL |
| **3.8** | A network protocol ensuring secure data transmission | **H** | Object-Oriented Programming |
| **3.9** | The language is primarily used for relational database queries | **I** | Entity-Relationship Model |
| **3.10** | Methodology for managing network resources and devices | **J** | Relational Databases |
| | | **K** | Information Security |
| | | **L** | Systems Analysis and Design |
| | | **M** | Network Management |

**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.** | A* (A-star) | **B.** | HDFS (Hadoop Distributed File System) | **C.** | pattern recognition |
| **D.** | UNIQUE | **E.** | opacity | **F.** | MQTT (Message Queuing Telemetry Transport) |
| **G.** | variable | **H.** | Secure Shell (SSH) | **I.** | kernel |
| **J.** | Queue | **K.** | node | **L.** | Decision Tree |
| **M.** | training | | | | |

**4.1** _____ is the protocol used for remote login to a network device, enabling text - based communication between a client and server over a secure channel.

**4.2** The _____ property in CSS is used to control the transparency of an element.

**4.3** In Python, a _____ is a named location used to store data in memory that can be accessed and modified during program execution.

**4.4** The _____ protocol is commonly used in IoT for lightweight communication between devices, especially in constrained environments.

**4.5** In a linked list, each element is known as a _____, which contains a data part and a reference to the next element.

**4.6** The _____ is the software layer responsible for managing hardware resources and providing services to applications in an operating system.

**4.7** In SQL, the _____ constraint is used to ensure that all values in a column are unique across a table.

**4.8** The _____ file system is used by Hadoop to distribute and store large data sets across a cluster of machines.

**4.9** The _____ algorithm is widely used in AI for searching through a graph and guarantees finding the shortest path if one exists.

**4.10** In machine learning, _____ is the process of finding patterns in data using algorithms that build models based on training data.

## PART TWO

**(Answer any FOUR questions)**

**5.** (a) Given a network where a file of 10 MB needs to be transmitted :

(i) In a circuit-switched network, a dedicated 1 Mbps link is established for the entire duration of the transfer.

(ii) In a packet-switched network, the file is divided into 1000 packets, each of size 10 KB, with a transmission rate of 2 Mbps and a latency of 20 ms per packet.

Calculate the total time required to transmit the file in both cases, and compare the performance of the two methods.

(b) What is Network Address Translation (NAT) ? How does it resolve the issue of IPv4 address exhaustion ? **(7+8)**

**6.** (a) Consider the following scenario for a Progressive Web App (PWA) implementation :

(i) A company has a web application that takes an average of 2 seconds to load each page, with 5 pages accessed during a typical user session.

(ii) After converting the app into a PWA with offline caching, each page load time is reduced to 0.5 seconds due to local storage of assets.

Compute the following questions based on the scenarios as mentioned above.

I. Calculate the total time saved per user session after implementing the PWA.

II. If the application has 10,000 daily users, then estimate the total time saved across all users in one day.

III. Discuss how these improvements in performance could impact user engagement and business outcomes.

(b) Design a CSS Grid layout to create a three-column responsive webpage. Provide code and explain how it works. **(7+8)**

**7.** (a) Write a Python program to solve the N-Queens problem using backtracking. Explain the recursive approach.

(b) Compare iterative and recursive solutions to solving complex problems. Which approach is more efficient for large datasets and why ? **(7+8)**

**8.** (a) Discuss the role of edge computing in the context of IoT. How does it differ from Cloud Computing in terms of processing and security?

(b) What are IoT data processing architectures? Discuss the importance of real-time analytics in IoT systems. **(7+8)**

**9.** (a) Write the pseudocode for Dijkstra's algorithm and implement it in Python to find the shortest path in a weighted graph.

(b) Explain the trade-offs between using binary search trees and hash tables for searching large datasets. **(7+8)**

**- o 0 o -**

**SPACE FOR ROUGH WORK**

**SPACE FOR ROUGH WORK**