BIG DATA AND IOT
(CSEN 3236)

Time Allotted : 2½ hrs **Full Marks : 60**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 4 (four) from Group B to E, taking one from each group.

Candidates are required to give answer in their own words as far as practicable.

Group – A

1. Answer any twelve: **12 × 1 = 12**

Choose the correct alternative for the following

(i) The "PageRank" application that ranks web pages by importance, involves an iterated _____ where the dimension is many billions.
(a) matrix addition (b) matrix vector multiplication
(c) matrix multiplication (d) vector addition

(ii) What is the primary data type of the MapReduce model?
(a) string (b) integer
(c) character (d) <key, value> pairs

(iii) In MapReduce, a Reduce function is _____.
(a) non-associative and non-commutative
(b) non-associative and commutative
(c) associative and commutative
(d) associative and non-commutative

(iv) Choose the primary characteristics of big data among the following
(a) Volume (b) Value (c) Variety (d) All of these

(v) All of the following accurately describe Hadoop, except
(a) Java-based (b) Open source
(c) Real time (d) Distributed computing approach

(vi) PIG is an implementation of _____ on top of Hadoop.
(a) relational algebra (b) relational tuples
(c) natural join (d) propositional calculus

(vii) REST services follow a _____ IoT communication model between client-server
(a) Request-Response (b) Publish-Subscribe
(c) Push-Pull (d) Exclusive Pair

(viii) The relation between IoT and M2M is
(a) IoT is part of M2M (b) M2M is the part of IoT
(c) Both are same (d) None

(ix) The full form of YARN is
(a) Yet Another Resource Navigator (b) Yet Another Resource Negotiator
(c) You Are Really Near (d) Yarn Aims to Reuse Nodes

(x) WAMP stands for Web Application _____ Protocol.
(a) Managing (b) Measuring
(c) Mapping (d) Messaging

Fill in the blanks with the correct word

(xi) Big data's velocity component relates to the _____ at which data is generated and processed.

(xii) The term which describes the process of creating multiple virtual networks on a physical network infrastructure is _____.

(xiii) The MapReduce framework utilizes the _____ function to break down large datasets into smaller chunks that can be processed in parallel.

(xiv) WebSocket-based communication APIs allow _____ duplex communication between clients and servers.

(xv) The programming model used by Hadoop for processing large data sets is _____.

Group - B

2. Design MapReduce algorithms that take a very large file of integers and produce as output:
(i) the smallest integer
(ii) the mean of all the integers
(iii) the same set of integers, but with each integer appearing only once.

[(CO3)(Create/IOCQ)]
(4 + 4 + 4) = 12

3. (a) What is DFS? How does it help the users of physically distributed systems?
[(CO1)(Understand/LOCQ)]

(b) Explain the local aggregation techniques used in MapReduce jobs.
[(CO1)(Understand/LOCQ)]

(c) Write the algorithm to solve the word count problem using MapReduce.
[(CO1)(Apply/IOCQ)]
(2 + 2) + 4 + 4 = 12

Group - C

4. (a) How can MapReduce be effectively utilized to perform operations such as Union, Intersection, and Difference on large datasets distributed across multiple nodes?
[(CO3)(Remember/LOCQ)]

(b) What are the key components of the Hadoop ecosystem and how do they contribute to its overall functionality? *[(CO4)(Remember/LOCQ)]*
6 + 6 = 12

5. (a) Consider the following matrix A and matrix B. Apply MapReduce technique to multiply A and B.

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \text{ and } B = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$$ *[(CO2)(Apply/HOCQ)]*

(b) Write the algorithm to perform matrix-vector multiplication using MapReduce. *[(CO2)(Understand/LOCQ)]*

(c) How can MapReduce impact on the performance of finding the shortest path in a graph? *[(CO3)(Understand/IOCQ)]*

5 + 4 + 3 = 12

Group - D

6. (a) Discuss the physical design of the Internet of Things (IoT). *[(CO4)(Remember/LOCQ)]*

(b) What is M2M? Compare IoT and M2M. *[(CO4)(Analyze/LOCQ)]*

(c) Explain different components of the IoT deployment template. *[(CO5)(Understand/IOCQ)]*

4 + (2 + 2) + 4 = 12

7. (a) Explain why IoT system should be self-adapting and self-configuring. *[(CO5)(Analyse/IOCQ)]*

(b) Illustrate a scenario of IoT service where publish-subscribe communication model is used. *[(CO5)(Remember/LOCQ)]*

(c) Is REST-based IoT communication stateless? Explain. *[(CO5)(Apply/IOCQ)]*

4 + 6 + 2 = 12

Group - E

8. (a) Explain the working principle of Apache Oozie as a workflow schedular system in managing Hadoop jobs. *[(CO6)(Analyse/HOCQ)]*

(b) Apache Spark promises to be faster than Hadoop. Justify in brief. *[(CO6)(Evaluate/IOCQ)]*

6 + 6 = 12

9. (a) Define cloud storage models and communication APIs. *[(CO6)(Remember/LOCQ)]*

(b) Explain the Xively cloud for IoT. *[(CO6)(Understand/LOCQ)]*

(c) Describe HBase and Avro in the Hadoop ecosystem. *[(CO1)(Understand/IOCQ)]*

(2 + 2) + 4 + (2 + 2) = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	47.91	40.63	11.46

Course Outcome (CO):

After the completion of the course students will be able to:

CSEN 3236.1. Develop understanding of the MapReduce paradigm.
CSEN 3236.2. Solve Matrix-Vector problems using the MapReduce paradigm.
CSEN 3236.3. Solve basic algorithmic problems in Graph Theory using the MapReduce paradigm IOT.

CSEN 3236.4. Understand the architecture of Internet of Things and connected world.
CSEN 3236.5. Explore the use of various hardware and sensing technologies to build IoT applications.
CSEN 3236.6. Understand the available cloud services and communication API„s for developing smart cities.

**LOCQ: Lower Order Cognitive Question; IOCQ: Intermediate Order Cognitive Question; HOCQ: Higher Order Cognitive Question.*