Enrollment No......................................

Faculty of Engineering
End Sem Examination May-2023
CS3ET05 Big Data Analytics
Programme: B.Tech.   Branch/Specialisation: CSE All

Duration: 3 Hrs.   **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

Q.1 i. Transaction data of the bank is-   1
(a) Structured data   (b) Unstructured data
(c) Both (a) and (b)   (d) None of these

ii. How many V's of Big Data?   1
(a) 2   (b) 3   (c) 4   (d) 5

iii. Who was the developer of Hadoop language?   1
(a) Apache Software Foundation
(b) Hadoop Software Foundation
(c) Sun Microsystems
(d) Bell Labs

iv. What was the Initial release date of Hadoop?   1
(a) 1st April 2007   (b) 1st April 2006
(c) 1st April 2008   (d) 1st April 2005

v. HDFS works in a _____ fashion.   1
(a) Worker-master fashion   (b) Master-slave fashion
(c) Master-worker fashion   (d) Slave-master fashion

vi. _____ NameNode is used when the Primary NameNode goes down.   1
(a) Rack   (b) Data   (c) Secondary   (d) Both (a) and (b)

vii. Point out the correct statement.   1
(a) MapReduce tries to place the data and the compute as close as possible
(b) Map Task in MapReduce is performed using the Mapper() function
(c) Reduce Task in MapReduce is performed using the Map() function
(d) None of these

P.T.O.

viii. The number of maps is usually driven by the total size of _____.   1
(a) Inputs   (b) Output   (c) Task   (d) None of these

ix. What is the full form of YARN?   1
(a) Yet Another Resource Network
(b) Yet Another Relational Negotiator
(c) Yet Another Resource Negotiator
(d) Yet Another Relational Network

x. Yarn commands are invoked by the _____ script.   1
(a) Hive   (b) Bin   (c) Hadoop   (d) Home

Q.2 i. What is the difference between an RDBMS and Hadoop?   2
ii. Explain types of data.   3
iii. Explain V's of big data with example.   5
OR iv. Define the significance of big data and what are the causes of big data?   5

Q.3 i. What do you mean by metadata in Hadoop?   2
ii. What is Hadoop distributed file system? Also explain Hadoop ecosystem.   8
OR iii. What are the advantages of Hadoop? Explain Apache Hadoop framework and master slave architecture.   8

Q.4 i. Explain what is heartbeat in HDFS? What happens when a data node fails?   3
ii. What are the main components of Hadoop distributed file system? Explain how files are stored in Hadoop.   7
OR iii. How read and write operations can be done and how data integrity can be maintained in Hadoop distributed file system? Explain.   7

Q.5 i. Mention what are the main configuration parameters that user need to specify to run MapReduce Job?   4
ii. How MapReduce works? Explain with example.   6
OR iii. Explain what is distributed Cache in MapReduce Framework.   6

Q.6 Attempt any two:
i. Differentiate between Hadoop 1.x and 2.x architecture.   5
ii. What are the key components of YARN? Explain each of them.   5
iii. Explain scheduling policies available in YARN.   5

******

Marking Scheme
CS3ET05 - Big Data Analytics

Q.1 i)   Transaction data of the bank is?   1
          a) structured data

     ii)   How many V's of Big Data   1
          d) 5

     iii) Who was the developer of Hadoop language?   1
          a) Apache Software Foundation

     iv)   What was the Initial release date of hadoop?   1
          b) 1st April 2006

     v)   HDFS works in a _____ fashion.   1
          b) master-slave fashion

     vi)   _____ NameNode is used when the Primary NameNode goes   1
          c) Secondary

     vii) Point out the correct statement.   1
          a) MapReduce tries to place the data and the compute as close as
          possible

     viii) The number of maps is usually driven by the total size of _____   1
          a) Inputs

     ix)   What is the full form of YARN?   1
          c) Yet Another Resource Negotiator

     x)   Yarn commands are invoked by the _____ script.   1
          b) bin

Q.2 i.   What is the difference between an RDBMS and Hadoop?   2
          Each difference   -1 Mark

     ii.   Explain types of data.   3
          Each type (there are 3 types of data)   -1 Mark

     iii. Explain V's of big data with example.   5
          5V's explanation of each   -1 Mark

OR   iv.   Define the significance of big data and what are the causes of big   5
          data?
          Explanation   -5 Marks

Q.3 i.   What do you mean by metadata in Hadoop?   2

     Explanation   -2 Marks

     ii.   What is Hadoop Distributed File System? Also explain Hadoop   8
          ecosystem.
          Hadoop Distributed File System explanation   -4 Marks
          Hadoop ecosystem explanation   -2 Mraks
          Hadoop ecosystem diagram   -2 Marks

OR   iii. What are the advantages of Hadoop? Explain Apache Hadoop   8
          framework and master slave architecture.
          Advantages   -2 Marks
          Hadoop framework   -3 Marks
          Master slave architecture   -3 Marks

Q.4 i.   Explain what is heartbeat in HDFS? What happens when a data   3
          node fails?
          Explanation   -3 Marks

     ii.   What are the main components of Hadoop distributed file system?   7
          Explain how files are stored in Hadoop.
          Components   -3 Marks
          File storage process   -4 Marks

OR   iii. How read and write operations can be done and how data integrity   7
          can be maintained in Hadoop distributed file system, explain.
          Read and write operations explanation   -4 Marks
          Data integrity explanation   -3 Marks

Q.5 i.   Mention what are the main configuration parameters that user   4
          need to specify to run MapReduce Job?
          Explanation   -4 Marks

     ii.   How MapReduce works? Explain with example.   6
          MapReduce Explanation   -4 Marks
          Example   -2 Marks

OR   iii. Explain what is distributed Cache in MapReduce Framework?   6
          Explanation   -4 Marks
          Diagram   -2 Marks

Q.6   Attempt any two:
     i.   Differentiate between Hadoop 1.x and 2.x architecture.   5

	Each difference	-1 Mark
ii.	What are the key components of YARN? Explain each of them.	**5**
	Explanation	-3 Marks
	Diagram	-2 Marks
iii.	Explain scheduling policies available in YARN?	**5**
	Explanation	-5 Marks

******