Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VI - THEORY EXAMINATION (2022-2023)

Subject: Big Data

Time: 3 Hours   Max. Marks: 100

General Instructions:

IMP: *Verify that you have received the question paper with the correct course, code, branch etc.*

1. *This Question paper comprises of* **three Sections -A, B, & C.** *It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.*

2. *Maximum marks for each question are indicated on right -hand side of each question.*

3. *Illustrate your answers with neat sketches wherever necessary.*

4. *Assume suitable data if necessary.*

5. *Preferably, write the answers in sequential order.*

6. *No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.*

<div align="center">

SECTION A   **20**

</div>

1. Attempt all parts:-

1-a.   Which platform is used to run hadoop? (CO1)   1

        (a) Debian

        (b) Cross platform

        (c) Bare metal

        (d) Unix like

1-b.   Which type of data transaction is used in bank? (CO1)   1

        (a) Unstructured data

        (b) Structured data

        (c) Both a and b

        (d) None of the above.

1-c.   YARN stands for (CO2)   1

        (a) Yahoo another resource name

        (b) Yet another resource negotiator

        (c) Yahoo archieved resource names

.

(d) Yet another resource need

1-d.   Every TaskTracker is configured with a _____ .(CO2)   1

(a) Maximum memory available in the node

(b) Not limited

(c) set of slots

(d) As decided by the job tracker

1-e.   Which one is taking a set of data and converted into another set of data? (CO3)   1

(a) Reduce stage

(b) Shuffling stage

(c) Map stage

(d) Shuffling and sorting

1-f.   Which one is name of the C++ interface to hadoop mapreduce? (CO3)   1

(a) Hadoop streaming

(b) Hadoop Pipes

(c) Mahaout

(d) Oozie

1-g.   Which of these an example of SaaS? (CO4)   1

(a) Google Workspace

(b) Dropbox

(c) Salesforce

(d) All of the above

1-h.   _____ is a paradigm of distributed computing to provide the   1
customers on-demand, utility based computing service. (CO4)

(a) Remote Sensing

(b) Remote Invocation

(c) Cloud Computing

(d) Private Computing

1-i.   _____ implies that all subnets obtained from the same subnet mask. (CO5)   1

(a) Static subnetting

(b) Dynamic subnetting

(c) Variable length subnetting

(d) Dynamic length subnetting

1-j.   The combination of _____ and _____ is often termed the local address of   1
.

the local portion of the IP address. (CO5)

      (a) Network number and host number

      (b) Network number and subnet number

      (c) Subnet number and host numbe

      (d) Host number

2. Attempt all parts:-

2.a.	Explain which is the most important challenges in big data? (CO1)	2
2.b.	Explain the limitation of traditional approach in big data process? (CO2)	2
2.c.	Summarize the HDFS goals. (CO3)	2
2.d.	Discuss IaaS. (CO4)	2
2.e.	How would you define the advantages of subnetting? (CO5)	2

SECTION B   30

3. Answer any <u>five</u> of the following:-

3-a.	Explain how big data is used in applications and describe its challenges? (CO1)	6
3-b.	Can you explain the digital data types and describe the data classification features? (CO1)	6
3-c.	How would you present the apache hadoop functions? (CO2)	6
3-d.	How can you describe the procedure of data analysis in hadoop? (CO2)	6
3.e.	Rank the importance of the hadoop ecosystem and Demonstrate the HDFS architecture. (CO3)	6
3.f.	How would you explain about (i) the Google App Engine (ii) the features Google App Engine (iii) the supporting languages of Google App Engine (CO4)	6
3.g.	Draw diagram and explain about the VPC architecture also elaborate its types? (CO5)	6

SECTION C   50

4. Answer any <u>one</u> of the following:-

4-a.	Explain in detail about (i) characteristics of cloud (ii) needs of cloud platform (iii) infrastructure development in cloud (CO1)	10
4-b.	Explain the different services of HDFS? (CO1)	10

5. Answer any <u>one</u> of the following:-

5-a.   Explain and draw the architecture of hadoop? (CO2)   10

5-b.   Explain HDFS. Also explain services of HDFS in detail.   10

6. Answer any <u>one</u> of the following:-

6-a.   What are the features of Apache Spark? CO3   10

6-b.   Brief out the features of hadoop HDFS and explain the functions of namenode   10
         and datanode. CO3

7. Answer any <u>one</u> of the following:-

7-a.   Write features and challenges of GCP. Also list down the features of any four   10
         srvices provided by GCP.

7-b.   Write and explain some popularly used cloud computing services.   10

8. Answer any <u>one</u> of the following:-

8-a.   Discuss the BigTable concept in Cloud.CO5   10

8-b.   Explain how machine learning is related to Bigquery. (CO5)   10