Printed Page:- 04	Subject Code:- ACSDS0601	
	Roll. No:	
NOIDA INSTITUTE OF ENGINEERING	AND TECHNOLOGY, GREATER NOIDA	
(An Autonomous Institute Affiliated to AKTU, Lucknow)		
В.Т	ech	
SEM: VI - THEORY EXAM	IINATION (2023 - 2024)	

Subject: Big Data Analytics

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.

SECTION A 20

Max. Marks: 100

## 1. Attempt all parts:-

Time: 3 Hours

- 1-a. Which of the following is the most popular NoSQL database for scalable big data store with Hadoop? [CO1]
  - (a) MONGODB
  - (b) HBASE
  - (c) Cassandra
  - (d) ORACLE
- 1-b. Point out the correct statement? [CO1]
  - (a) The sequence file also can contain a "secondary" key-value list that can be used as file Metadata
  - (b) SequenceFile formats share a header that contains some information which allows the reader to recognize is format
  - (c) There're Key and Value Class Name's that allow the reader to instantiate those classes, via reflection, for reading
  - (d) All of the mentioned
- 1-c. Among the following options which component deals with ingesting streaming

1

1

	data into Hadoop? [CO2]	
	(a) Oozie	
	(b) Hive	
	(c) Kafka	
	(d) Flume	
1-d.	Identify the term used to define the multidimensional model of the data warehouse? [CO2]	
	(a) Table	
	(b) Data Cube	
	(c) Tree	
	(d) Data Structure	
1-e.	Which command is used to copy files from the local file system to HDFS? [CO3]	
	(a) put	
	(b) get	
	(c) ls	
	(d) rm	
1-f.	Which command is used to remove a file or directory from HDFS? [CO3]	
	(a) put	
	(b) get	
	(c) ls	
	(d) rm	
1-g.	Which of the following command provides information about the user? [CO4]	,
	(a) Status	
	(b) Version	
	(c) Whoami	
	(d) User	
1-h.	Hive usesstyle escaping within the strings? [CO4]	
	(a) C	
	(b) Java	
	(c) Python	
	(d) Scala	
1-i.	what option best describes this statement - "A dominant supervision platform for resources responsible for delivering scalable processes"? [CO5]	

	(a) SparkContext	
	(b) YARN	
	(c) RDD	
	(d) SparkStreaming	
1-j.	What component is present on the top of Spark Core? [CO5]	1
	(a) Spark RDD	
	(b) Spark Streaming	
	(c) Worker Nodes	
	(d) Spark SQL	
2. Atte	empt all parts:-	
2.a.	Distinguish between Hadoop 1.x, Hadoop2.x and Hadoop 3.x in detail? [CO1]	2
2.b.	Differentiate between data analytics and big data analytics? [CO2]	2
2.c.	What are the features of Massive parallel processing system? [CO3]	2
2.d.	What is the purpose of the "JOIN" command in HiveQL? [CO4]	2
2.e.	Create a scenario where big Data Analytics be used as a Decision Making tool? [CO5]	2
	SECTION B	30
3. Ans	wer any <u>five</u> of the following:-	
3-a.	What is Name node & Data node in Hadoop Architecture, discuss in detail? [CO1]	6
3-b.	What are major security challenges in Big Data? [CO2]	6
3-c.	Explain the following (i) HDFS safe mode (ii) Namenode high availability [CO2]	6
3-d.	Explain Job Scheduling in Map Reduce ? [CO2]	6
3.e.	Write and draw anatomy of a file write with HDFS? [CO3]	6
3.f.	Comparison between Hive vs Pig vs SQL? [CO4]	6
3.g.	Write a step-by-step procedure for installation of APACHE SPARK? [CO5]	6
	SECTION C	50
4. Ans	wer any <u>one</u> of the following:-	
4-a.	What are 5 Principles of Big Data Ethics? Compare between analysis and reporting? [CO1]	10
4-b.	i) Discuss the features of Hadoop and explain the functionalities of Hadoop cluster? (7) (ii) Give brief about Hadoop input and output and also write a note on data integrity? [CO1]	10

J. A11544	er uny <u>one</u> or the following.	
5-a.	Describe the input and output formats of MapReduce. CO2	10
5-b.	What are the Hadoop daemons? Explain their roles and also draw a neat diagram of Hadoop Architecture? [CO2]	10
6. Answ	er any <u>one</u> of the following:-	
6-a.	What are the stable versions of Hadoop? What are the additional benefits YARN brings in to Hadoop? [CO3]	10
6-b.	Specify the role of job tracker and task tracker in HDFS. CO3	10
7. Answ	er any <u>one</u> of the following:-	
7-a.	What are the important components of the Spark ecosystem, Explain how Spark runs applications with the help of its architecture? [CO4]	10
7-b.	How can you aggregate data in a Hive query? With examples. [CO4]	10
8. Answ	er any <u>one</u> of the following:-	
8-a.	What are the components of Apache HBase and Region Server in Hbase? Mention the differences between "HBase" and "Relational Databases"? [CO5]	10
8-b.	What is Apache ZooKeeper and Apache Oozie, How do you configure an "Oozie" job in Hadoop, Explain in Detail? [CO5]	10