

Printed Page:-

Subject Code:- ACSDS0601

Roll. No:

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA
(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: VI - THEORY EXAMINATION (20.....- 20.....)

Subject: Big Data Analytics

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.

SECTION-A

20

1. Attempt all parts:-

- 1-a. State which is the correct statement? [CO1,K1] 1
- (a) Hadoop do need specialized hardware to process the data
 - (b) Hadoop 2.0 allows live stream processing of real-time data
 - (c) In the Hadoop programming framework output files are divided into lines or records
 - (d) None of the mentioned
- 1-b. State according to analysts, for what can traditional IT systems provide a foundation when they're integrated with big data technologies like Hadoop? [CO1,K1] 1
- (a) Big data management and data mining
 - (b) Data warehousing and business intelligence
 - (c) Management of Hadoop clusters
 - (d) Collecting and storing unstructured data
- 1-c. State whether Fact tables are _____ ? [CO2,K1] 1
- (a) HDFS
 - (b) MapReduce
 - (c) YARN
 - (d) All of the Above
- 1-d. Identify the correct definition of Reconciled data? [CO2,K1] 1
- (a) Reconcile data is a data scored in one operational system in the organization

- (b) Reconcile data is the data that has been selected and formatted for end-user support applications.
- (c) Reconcile data is the current data intended to be the single source for all decision support systems
- (d) None

- 1-e. State which of the following is not a responsibility of the Application Master in YARN? [CO3,K1] 1
- (a) Negotiating resources with the ResourceManager
 - (b) Monitoring the progress of the application
 - (c) Reporting application status to the ResourceManager
 - (d) Allocating resources to other applications
- 1-f. List which of the following is not a resource that can be requested by an application in YARN? [CO3,K1] 1
- (a) Memory
 - (b) CPU
 - (c) Disk space
 - (d) Network bandwidth
- 1-g. State to see the partitions present in a Hive table the command used is? [CO4,K1] 1
- (a) Describe
 - (b) show
 - (c) describe extended
 - (d) show extended
- 1-h. State to see the data types details of only a column not the table we should use the command? [CO4,K1] 1
- (a) DESCRIBE
 - (b) DESCRIBE EXTENDED
 - (c) DESCRIBE FORMATTED
 - (d) DESCRIBE COLUMN
- 1-i. State whether Sqoop written in? [CO5,K1] 1
- (a) C
 - (b) C++
 - (c) Java
 - (d) hadoop
- 1-j. List how Data processed by Scoop can be used for? [CO5,K1] 1
- (a) Hbase
 - (b) HDFS
 - (c) Mapreduce
 - (d) MahOut

2. Attempt all parts:-

- | | | |
|------|--|---|
| 2.a. | Discuss why is big data analytics important? [CO1,K1] | 2 |
| 2.b. | Identify, if a Hadoop Cluster has 10 nodes, each with 8 cores and 64 GB of RAM, If each node can process 100 MB of data per second, what is the total processing capacity of the cluster in GB/s? [CO2,K1] | 2 |
| 2.c. | List out the best application areas for the Big Data Analytics? [CO3,K1] | 2 |
| 2.d. | Discuss If a hive query has 3 stages, each taking 2 minutes to complete. What is the total execution time of the query? [CO4,K1] | 2 |
| 2.e. | State how does SPARK handle fault tolerance? [CO5,K1] | 2 |

SECTION-B

30

3. Answer any five of the following:-

- | | | |
|------|---|---|
| 3-a. | Discuss Big Data Analytics usage in a project as an example, Also explain 5V's of Big Data Industry, Briefly? [CO1,K2] | 6 |
| 3-b. | Explain the difference between Structured, Semi-Structures, Quasi-Structured and Unstructured data? [CO1,K1] | 6 |
| 3-c. | Explain how node failure can be handled in hadoop,Also state how can you configure " ZOOKEEPER" with Hbase? [CO2,K2] | 6 |
| 3-d. | State what are the core components of Hadoop. Explain in brief each of its components? [CO2,K1] | 6 |
| 3.e. | Discuss what is commodity hardware, Also explain what happens if during the PUT operation, HDFS block is assigned a replication factor 1 instead of default value 3? [CO3,K2] | 6 |
| 3.f. | Explain different execution modes available in Pig? [CO4,K1] | 6 |
| 3.g. | Explain and make a note on why RDD is better than Map Reduce data STORAGE? [CO5,K1] | 6 |

SECTION-C

50

4. Answer any one of the following:-

- | | | |
|------|--|----|
| 4-a. | Discuss the role of Master and Slave demons in detail, Also synthesize the role of machine learning in enhancing intelligent data analysis for Big Data? [CO1,K2] | 10 |
| 4-b. | Explain the role of following nodes in detail? [CO1,K2]
(i) Resource Manager
(ii) Application Manager
(iii) Job Tracker
(iv) Task Tracker
(v) Namenode
(vi) Datanode | 10 |

5. Answer any one of the following:-

- | | | |
|------|---|----|
| 5-a. | Discuss and write short notes on? [CO2,K1]
(i) Driver Code
(ii) Mapper Code | 10 |
|------|---|----|

(iii) Reducer Code

(iv) Combiner

- 5-b. Discuss and write short notes on Serialization and Byte Writable? [CO2,K2] 10
6. Answer any one of the following:-
- 6-a. Explain the concept of Partitioning in Hive, Also describe how Partitioning works including benefits and limitations of Partitioning? [CO3,K2] 10
- 6-b. Discuss how do you copy data from the local file system onto HDFS using commands ,Also state why is Map Reduce slower in processing data in comparison to other processing frameworks,Take a sample .txt file data as input and explain? [CO3,K2] 10
7. Answer any one of the following:-
- 7-a. Describe the different join types in Hive, Also describe how Hive can be used for data warehousing, including data modeling, ETL, and querying? [CO4,K2] 10
- 7-b. Explain the concept of column families in Hbase, Explain how column families work in Hbase including storage and retrieval of data? [CO4,K2] 10
8. Answer any one of the following:-
- 8-a. Apply Spark Core Processing RDD to run Word Count program with a sample data as input, write and explain all steps? [CO5,K3] 10
- 8-b. Explain the following in detail? [CO5,K2] 10
- (i). Installation of PIG.
- (ii).Write Pig Latin scripts sort, group, join, project, and filter your data.