



Raj Kumar Goel Institute of Technology
5 –Km. Stone, Delhi – Meerut Road, Ghaziabad (U.P. – 201017)
Tel: (0120) 2788409, Fax: (0120) 2788350

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Big Data Question Bank (2023-24)

UNIT-1 INTRODUCTION TO BIG DATA

- A)What is Big Data? Explain characteristics of Big Data.
- B)What are the benefits of Big Data? Discuss challenges under Big Data.
- C)How Big Data Analytics can be useful in the development of smart cities.
(Discuss one application) ?
- D)Explain working of following phases of Map Reduce with one common example.
 - (i) Map Phase (ii) Combiner Phase
 - (iii) Shuffle and Sort Phase (iv)Reducer Phase
- E)Discuss big data in healthcare, transportation and medicine.
- F)Explain concept of 4 V's in detail?
- G)Explain categories of Data with example?

Unit-2 Introduction to Hadoop and its Architecture

- A)What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with proper diagram.
- B)Write Short note on Hadoop Ecosystem also explain various elements of hadoop.
- C)What is Zookeeper? List the benefits of it.
- D)How the data move into the hadoop? Explain with the help of example?
- E)Explain the following terms

1. Flume.
2. Chukwa
3. Scribe

F) Using sqoop to export data to MySQL?

G) Explain input and output format of MapReduce?

H) Show the comparison among SequenceFiles, Protocol Buffer, Thrift and Avro?

Unit-3 HDFS,HIVE,HIVEQL and HBASE

A) What do you mean by HiveQL data definition Language?

B) Explain any three HiveQL commands with their syntax and example?

C) Explain Metastore in Hive.

D) Explain storage mechanism in HBase. Compare Row oriented and Column Oriented database structures.

E) What is Hbase? Discuss in detail the data model and Implementation Aspect.

F) Elaborate on HiveQL data manipulation queries in detail

G) Discuss the concept of regions in HBase and storing Big data with HBase.

H) Draw and discuss the architecture of Hive in detail.

I) Explain how HBase uses Zookeeper to build applications with Zookeeper.

J) Discuss how Pig data model will help in effective data flow. Compare

K) Row oriented and Column Oriented database structures.

Unit-4. Spark

A) Explain Spark components in detail. Also list the features of Spark.

B) What is RDD? Explain about transformations and actions in the context of RDDs. State and explain RDD operations in brief.

C) Write a brief short note on: Spark Unified Stack

D) Give overview of MLlib in detail?

Unit-5 NoSQL

A) Write a short note on NoSQL database. Explain advantages and uses of NoSQL?

B) Explain types of NoSQL database?

C) List the difference between NoSQL and relational database?

D) Explain CRUD operations in MongoDB.

E) Explain Key features of MongoDB?

F) Explain following for MongoDB

1. Indexing.
2. Aggregation