**Unit 1**

1. What is Big Data? Elaborate on the different Data Architecture principles.
2. What do you mean by Meta Data. Explain Metadata with respect to Big Data.
3. What do you mean by Data Governance and explain the fundamentals of Data Governance.
4. What is Data security, elaborate the concept of security and the different levels incorporated in a BigData system.
5. Explain Data as a service.
6. What is Hadoop and explain the evolution of Data Architecture with Hadoop.
7. Explain the Hadoop Data Architecture with a neat diagram.
8. What do you mean by Data Wrangling and explain the different phases involved in it.
9. What do you mean by Data Masking, list out the different methods that you are aware.
10. Elaborate on the different components that are part of the HDFS architecture.
11. What do you mean by YARN and explain its architecture.
12. What are the different hadoop file format that you are aware and which format is betterfor processing data.
13. What do you mean by Cluster. Elaborate Hadoop cluster with an example.
14. Elaborate on the various Installation modes of installing hadoop in the System.
15. What do you mean by Big Data Analytics. List out any five applications of Big Data Analytics.
16. Elaborate on the various V's associated with Big Data with real time examples.

**Unit 2**

1. Define Data Ingestion and what is the differences between Batch Processing and Real time processing.
2. What is the use of Sqoop and explain its architecture.
3. Explain Apache Flume architecture elaborately with a diagram. Describe each component as well.
4. What is the advantage of using Apache NiFi in Hadoop? Explain its architecture and list out its benefits.
5. What do you mean by Kafka Connect? Explain the concept of Fault Tolerance in Kafka Connect cluster distributed architecture.
6. What do you mean by Hive. Explore the differences between Hive and RDBMS.
7. Explain the following terms: i. HBase, ii. Sqoop, iii. Heart Beat, iv. Resource Manager
8. Explain the following terms: i. MongoDB, ii. Task, iii. Job, iv. NodeManager
9. Elaborate in detail about the various NoSQL storage types.
10. Illustrate any five commands used in Mongodb with example.
11. List out the differences between SQL and HiveQL
12. Explain Partitioning and bucketing in Hive with an example.
13. What do you mean by NoSql Databases? Explore the need of it and elaborate on the different types of NoSql databas
14. List out the different types of applications and the various NoSql databases that suits each applications.
15. You are developing a food delivery application called "Fwiggy". As a Database administrator, what type of database v you incorporate for this application and why? List out the reasons and justify it.
16. You are working for the KFC group as a database admin. The company wants to develop a billing application to be use in the retail outlet. Will you choose MongoDB? If yes elaborate the reasons, if not list out your reasons.

**Unit 3**

1. What is Hadoop? What are the two core components of Hadoop?
2. Compare Hadoop with RDBMS.
3. What is HDFS? What are the design consideration in using HDFS?
4. What are the features of HDFS?
5. What are the advantages of using HDFS in Hadoop?
6. Compare Name Node and Data Node
7. What is a Block in HDFS? Why do you think block size of HDFS is larger than that of Unix block size?
8. With an example, explain the following commands i) CopyFromLocal, ii) CopyToLocal
9. What is Map Reduce paradigm? With a neat block diagram explain the process of solving problem using map-reduce paradigm.
10. What is the necessity for using 'WritableComaprable' interface in Hadoop?
11. Why do you think Hadoop does not make use of Java primitive data types?
12. Write a short note on Primitive data type supported in Hadoop?
13. What is a SequenceFile? Illustrate the processes of reading a SequenceFile in Hadoop?
14. Explain the format of Sequence File format with no compression and with record compression
15. Illustrate word count program with a diagram and explain the purpose of Mapping and Reducing in solving the above problem.
16. Explain in detail about Sorting, Shuffling, Combining and Partitioning with respect to MApReduce.

**Unit 4**

1. Briefly describe the Hadoop Eco system
2. Describe the architecture of HDFS.
3. What is the significance of shuffle and sort step in map reduce?**#mr pg11**
4. How does Hadoop keep track of the progress and status of a given task?
5. Which are the different types of failures that may occur in map reduce job? How are they handled in Hadoop?
6. What are counters? What is the significance of task counters and job counters in Hadoop?
7. Compare classic map reduce with YARN
8. Describe the Anatomy of a YARN Application Run**# 155**
9. Explain in detail YARN architecture **# hdc u1 pg7**
10. Explain the working of FIFO, Capacity in YARN **#mp u iii pg 124**
11. Explain the working of Capacity, and Fair Schedulers in YARN **# 127**
12. Elaborate on K-Means Clustering method with respect to MapReduce. **#kmeans map reduce job**
13. Write any five HDFS commands that you know and give an example to its usage.
14. List out any Seven Linux commands that you are aware. Explain with an example.
15. Explain the following components of a MapReduce program:

Tool,

ToolRunner,

Job object, **#pg 82**

InputFormat **#pg 13**

Output

1. Elaborate the process of configuring a job by setting up your MapReduce to accept command line arguments

**Unit 5**

1. What is the significance of map-side join and reduce-side join?
2. What are the limitations of classic mapreduce?
3. Elaborate on a different Input formats that is used for a MapReduce program. **#pg 13**
4. Explain on the Writable interface of the MapReduce. **# pg 63**
5. Explain on Sorting and Partitioning of MapReduce. **#pg 30 & pg 40**
6. How will you run an SQL select with MapReduce.
7. Elaborate on a MapReduce program to count the occurrences of words in input text.**#pg 2**
8. What is the role of Inverted Index in a search engine, and the role of mapreduce in generating the inverted index.
9. List out the differences between PIG and Hive. **#hive pg 23**
10. Explain the architecture of Pig with a neat diagram
11. List out the various types of commands that are used in PIG.
12. Describe the various modes of execution, when you are execting a Pig Program.
13. Elaborate on the various output formats that is used for a MapReduce Program.
14. Explain Sorting and Partitioning in detail, with a neat diagram.
15. Define Replication. How does hadoop provide replication with a neat diagram.**# hdc u1 pg4**
16. What do you mean by Fault Tolerance System. How does Hadoop ensure this capability. Explain in detail.