

# Big Data Analytics

## Question Bank

	<b>Introduction Big Data</b>
1	What is Big Data ? Explain Characteristics of Big Data.
2	Draw HDFS Architecture. Explain any two commands of HDFS from the following commands with syntax atleast one example of each. CopyFromLocal, setrep, checksum
3	What is big data analytics? Explain four 'V's of Big data. Briefly discuss applications of big data.
4	What is Map Reduce? Explain working of various phases of Map Reduce with appropriate example and diagram
5	Discuss Big Data in Healthcare, Transportation & Medicine.
6	What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with proper diagram.
7	Write Map Reduce steps for counting occurrences of specific numbers in the input text file(s). Also write the commands to compile and run the code.
8	Discuss Hadoop YARN in detail with failures in classic MapReduce.
9	What are the benefits of Big Data? Discuss challenges under Big Data. How Big Data Analytics can be useful in the development of smart cities.
10	Explain Job Scheduling in Map Reduce. How it is done in case of (i) The Fair Scheduler (ii) The Capacity Scheduler
11	What is Big data? Discuss it in terms of four dimensions, volume, velocity
12	List various application of big data. How it can be used to improve business for a superstore.
13	Explain Map-reduce framework in brief.
14	Explain the difference between structure and unstructured data.
15	Explain "Map Phase" and "Combiner Phase" in MapReduce.
16	What is Big Data? Explain how big data processing differs from distributed processing.
17	Explain Avro data serialization technique in MapReduce.
18	Explain following commands with syntax and at least one example of each. (1) copyFromLocal (2) showing the content of outputfile (3) setrep (4) checksum.
19	Explain advantages and disadvantages of big data analytics.

	<b>Introduction to Hadoop and Hadoop Architectur</b>
1	What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with proper diagram.
2	What is Hadoop Ecosystem? Discuss various components of Hadoop Ecosystem.
3	Explain core architecture of Hadoop with suitable block diagram. Discuss role of each component in detail.
4	What is data serialization? With proper examples discuss and differentiate structured, unstructured and semi-structured data. Make a note on how type of data affects data serialization.
5	List various configuration files used in Hadoop Installation. What is use of mapred- site.xml?
6	What is Name node & Data node in Hadoop Architecture.
	<b>Hdfs,Hive and HiveQL,Hbase</b>
1	Explain working of Hive with proper steps and diagram.
2	(i) What is Zookeeper? List the benefits of it. (ii) Differentiate: Apache pig Vs Map Reduce.
3	Define HDFS. Discuss the HDFS Architecture and HDFS Commands in brief.
4	What do you mean by HiveQL Data Definition Language? Explain any three HiveQL DDL command with its syntax and example.
5	(i) Explain Metastore in Hive. (ii) Explain Storage mechanism in HBase.
6	With suitable block diagram explain architecture of HDFS. Discuss role of Data node and Name node in HDFS. Give commands with appropriate arguments to perform data transfer between local file system and HDFS.
7	Explain the HiveQL-Select-Order By with suitable example.
8	Define join and explain types of join
9	Explain the concept of Blocks and Heartbeat Message in HDFS Architecture. What are the benefits of block transfer?
10	Explain Hive Data types
11	Explain HBase architecture.
12	Discuss role of Data node and Name node in HDFS.
13	Write a short note on Apache Pig. Enlist applications of Apache Pig.
14	Difference between Hive and RDBMS
15	Difference between HDFS and Hbase.
16	Compare Row oriented and Column Oriented database structures.
17	Explain the 5 P's of Data science in brief.

	<b>NoSQL</b>
1	What is NoSQL database? List the differences between NoSQL and relational databases. Explain in brief various types of NoSQL databases in practice.
2	Use of NoSQL in industry.
3	Define NewSQL and explain benefits and limitation of NewSQL.
4	Define NoSQL and where is it used? (b) i) Document Oriented Database ii) Graph based Database.
5	Write differences between NoSQL and SQL.
6	Explain NoSQL.
	<b>MapReduce</b>
1	How does the Hadoop Map reduce data flow work for a word count program? Give an example?
2	How do you run MapReduce and Message Passing interface (MPI) on YARN architecture
3	What is Map Reduce?
4	Illustrate a simple example of the working of MapReduce.
5	What are the main components of MapReduce Job?
6	What is Shuffling and Sorting in Map Reduce?
7	What is Identity Mapper and Chain Mapper?
8	What main configuration parameters are specified in MapReduce? Name Job control options specified by Map Reduce.