Brainheaters

Viva & Orals BDA Lab

Set of VIVA Question & Answers selected from Bh.Questionbanks

Edition 1

Big Data Analysis

Sample preparation kit for VIVA exams

Q1. What is mean by Big Data characteristics?

Answer: Three characteristics define Big Data: volume, variety, and velocity. Together, these characteristics define "Big Data".

What are the characteristics of Big Data?

- The Volume of Data.
- The sheer volume of data being stored today is exploding.
- The Variety of Data. Source.
- The Velocity of Data.

Q2. Explain Types of Big Data

Answer: Types Of Big Data

- Structured Data.
- Unstructured Data.
- Semi-Structured Data.
- Subtypes of Data.
- Interacting with Data Through Programming.

Q4. Define Hadoop

Answer: Hadoop is an open-source, Java-based framework used for storing and processing big data. The data is stored on inexpensive commodity servers that run as clusters. Cafarella, Hadoop uses the MapReduce programming model for faster storage and retrieval of data from its nodes.

Q4. Explain Core Hadoop Component

Answer: The core of Apache Hadoop consists of a storage part, known as Hadoop Distributed File System (HDFS), and a processing part which is a MapReduce programming model.

Hadoop splits files into large blocks and distributes them across nodes in a cluster.

Q5. Explain Hadoop Ecosystem

Answer: Hadoop Ecosystem is a platform or a suite that provides various services to solve big data problems.

It includes Apache projects and various commercial tools and solutions.

There are four major elements of Hadoop i.e.

- HDFS,
- MapReduce,
- YARN,
- Hadoop Common.

Q6. Explain Physical Organization of Compute Nodes

Answer: Compute nodes are stored on racks, perhaps 8–64 on a rack.

The nodes on a single rack are connected by a network, typically Gigabit Ethernet.

There can be many racks of compute nodes, and racks are connected by another level of network or a switch.

Q7. Explain Large-Scale File-System Organization.

Answer: The new parallel-computing architecture, sometimes called cluster computing, is organized as follows.

Compute nodes are stored on racks, perhaps 8–64 on a rack.

The nodes on a single rack are connected by a network, typically Gigabit Ethernet.

Q8. Explain The Reduce Tasks

Answer: The reduce task is the actual instantiation of a reducer code that runs on a node in your cluster.

This task has a state machine and might fail.

In case of failure, another reduced task is spun up to restart the computation. This is called reduce task attempts.

Q9. Discuss Details of MapReduce Execution

Answer: MapReduce program executes in three stages, namely map stage, shuffle stage, and reduce stage.

Map stage - The map or mapper's job is to process the input data. Generally, the input data is in the form of a file or directory and is stored in the Hadoop file system (HDFS).

Q10. Explain Matrix-Vector Multiplication by MapReduce

Answer: Suppose we have an n×n matrix M, whose element in row i and column j will be denoted mij .

Suppose we also have a vector v of length n, whose jth element is vj.

Then the matrix-vector product is the vector x of length n, whose ith element xi is given by xi=∑nj=1mij×vj.

25+

More questions and answers are available in your StudyRoom at Brainheaters

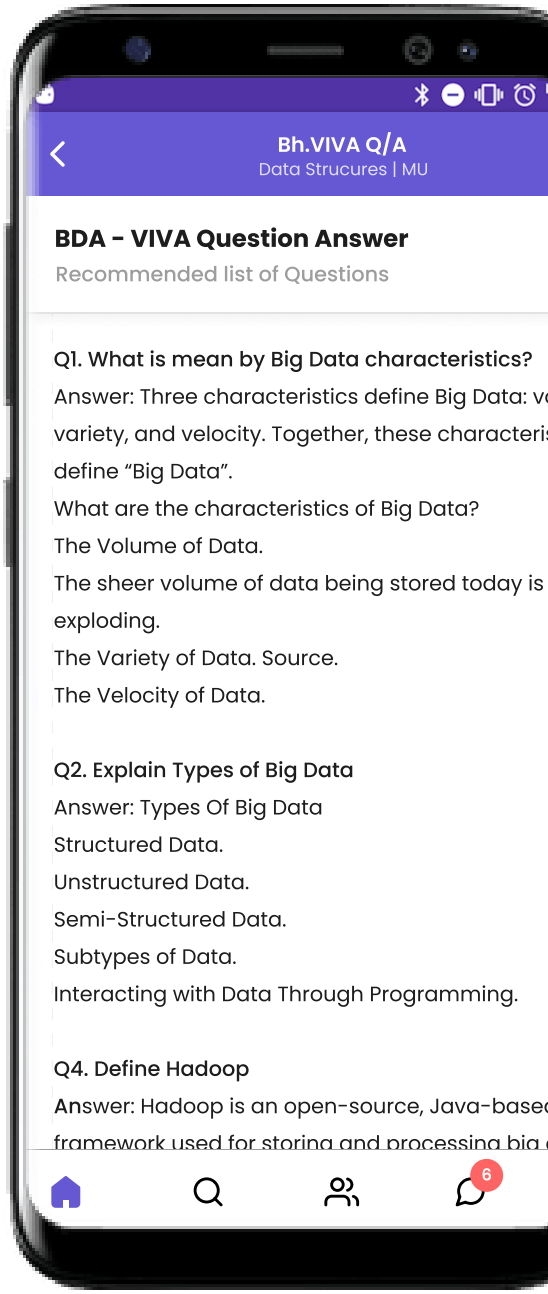
Download the App!

Activate your Smart Plan today!

Start your preparation with Brainheaters App

- Notes, Qbanks, IMPs, Paper Analysis & more...
- Offline access + Free Digital library

Download the App!



Brainheaters