**Unit 3**

1. Define distributed databases and explain how they differ from centralized databases. Discuss the advantages and benefits of using distributed databases in contrast to traditional centralized databases. Also, explain the challenges and potential issues associated with distributed databases.
2. Discuss the architecture of DDBMS. Within the context of a centralized DBMS, briefly explain new components introduced by the distribution of data.
3. What are the main software modules of a DDBMS? Discuss the main functions of each of these modules in the context of the client/server architecture.
4. Describe the concept of data fragmentation and its significance in distributing data across multiple nodes in a distributed database. Compare and contrast horizontal and vertical data fragmentation techniques and provide examples of scenarios where each method is suitable. Also explain why fragmentation is a useful concept in distributed database design.
5. Explain the concept of data replication and its role in enhancing data availability and fault tolerance in distributed databases.
6. Discuss various data allocation techniques for placing data fragments and replicason different sites to optimize performance and minimize communication overhead.
7. Differentiate homogenous and heterogeneous distributed database systems and explain their respective advantages and challenges. Provide examples of distributed database systems that fall into each category and discuss their use cases.
8. Describe the different distributed database architectures, including client-server and peer-to-peer.Discuss the pros and cons of each distributed database architecture and their suitability forspecificapplication scenarios.
9. What are the different stages of processing a query in a DDBMS?
10. Discuss the different techniques for executing an equijoin of two files located at different sites. What main factors affect the cost of data transfer?
11. Discuss the two-phase commit protocol used for transaction management in a DDBMS. List its limitations and explain how they are overcome using the three-phase commit protocol.
12. When are voting and elections used in distributed databases?
13. Discuss catalog management in distributed databases.

**Unit 4**

1. Define NoSQL databases and explain the main motivations behind their development. Compareand contrast NoSQL databases with traditional relational databases. Also, mention the scenariosanduse caseswhere NoSQLdatabases arepreferred over relationaldatabases.
2. ExplaintheCAPtheorem(Consistency,Availability,andPartitionTolerance)anditssignificancein the design of distributed systems. Discuss the trade-offs among consistency, availability, andpartitiontolerance andhowNoSQL databases addressthese trade-offs.
3. Describe the data models and characteristics of document-based, key-value stores, column-based,andgraph-based NoSQL databases.
4. Compare the traditional ACID properties of relational databases with the BASE principles(Basically Available, Soft state, Eventually consistent) followed by some NoSQL databases.Discuss how BASE principles enable high availability, fault tolerance, and scalability indistributedNoSQL systems.
5. What are the main categories of NOSQL systems? List a few of the NOSQL systems in each category.
6. What are the data modeling concepts used in MongoDB? What are the main CRUD operations of MongoDB?
7. What are the data modeling concepts used in the graph-oriented NOSQL system Neo4j?
8. What is the query language for Neo4j?
9. What are the data modeling concepts used in column-based NOSQL systems and Hbase?
10. What are the main CRUD operations in Hbase?

**Unit 5**

1. Define Big Data and explain the 5 V's of Big Data. Discuss the challenges and opportunitiespresented by Big Data in various industries and domains. Also, provide examples of applicationswheretheanalysis of BigDatahas made a significantimpact.
2. Explain the MapReduce programming model and its role in processing large-scale data. Describethe key components of the MapReduce model: Mapper, Reducer, and Combiner. Also, provide astep-by-stepexampleof howMapReduceworkstoprocessandanalyzedata(analgorithm).
3. Describe the architecture of Hadoop with the Hadoop Distributed File System (HDFS). Also,discussthebenefitsof Hadoop'sdistributedandfault-tolerantdesignforhandlingBigData.
4. List and explain some of the key components in the Hadoop ecosystem, such as Apache Hive,Apache Pig, and Apache HBase.Describe the role of these components in processing, querying,andanalyzingBigData usingHadoop.
5. Give an overview of the YARN architecture. How does Resource Manager work in YARN?