

ADVANCED DATABASE QUESTION BANK

PART A – 2 Mark Questions with Answers | PART B – 16 Mark Questions

UNIT I – DISTRIBUTED DATABASES

PART A – 2 Mark Questions with Answers

Q.No	Question	Answer
1	What is a Distributed Database?	A distributed database is a collection of multiple, logically interrelated databases spread over a computer network, managed by a Distributed DBMS.
2	Define Distributed DBMS.	A DDBMS is the software system that manages the distributed database and makes the distribution transparent to users.
3	What is Data Fragmentation?	Data fragmentation is the process of dividing a relation into smaller pieces (fragments) stored at different sites. Types: Horizontal, Vertical, Hybrid.
4	Define Horizontal Fragmentation.	Horizontal fragmentation divides a relation into subsets of tuples (rows), where each subset is stored at a different site.
5	What is Vertical Fragmentation?	Vertical fragmentation divides a relation by columns (attributes), with each fragment containing a subset of columns plus the primary key.
6	What is a Commit Protocol?	A commit protocol ensures all-or-nothing execution of distributed transactions. The most common is Two-Phase Commit (2PC).
7	Define Two-Phase Commit (2PC).	2PC is a distributed protocol with two phases — Prepare phase (coordinator asks all participants to vote) and Commit phase (coordinator sends final decision based on votes).
8	What is Distributed Concurrency Control?	It is a mechanism to ensure serializability of concurrent transactions in a distributed database, using techniques like distributed locking or timestamping.
9	Define Data Replication.	Data replication is the process of storing copies of data at multiple sites to improve availability and reliability.
10	What is Distributed Query Processing?	It involves translating a high-level query into an efficient execution plan that minimizes data transfer across the network in a distributed system.

11	What is a global schema?	A global schema describes the logical structure of the entire distributed database as if it were a single centralized database.
12	Define Site Transparency.	Site transparency means users need not know the physical location of data; queries are processed regardless of where data is stored.

PART B – 16 Mark Questions

Q.No	Question (16 Marks)
1	Explain the architecture of a Distributed Database System with a neat diagram. Discuss the advantages and disadvantages.
2	Describe the different types of data fragmentation (Horizontal, Vertical, Hybrid) with examples for each.
3	Explain the Two-Phase Commit Protocol (2PC) in detail. What are the problems associated with it?
4	Discuss Distributed Concurrency Control techniques: Distributed Locking and Timestamp Ordering.
5	Explain Distributed Query Processing in detail. How is a query decomposed and optimized in a distributed environment?

UNIT II – SPATIAL AND TEMPORAL DATABASES

PART A – 2 Mark Questions with Answers

Q.No	Question	Answer
1	What is a Temporal Database?	A temporal database stores and manages time-varying data by maintaining historical and future states of data using valid time and transaction time.
2	Define Valid Time and Transaction Time.	Valid time is when a fact is true in the real world. Transaction time is when a fact is stored in the database.
3	What is a Spatial Database?	A spatial database stores and queries data related to objects defined in a geometric space, such as maps, GIS data, and location-based data.
4	What are Spatial Data Types?	Spatial data types include Point (single location), Line (sequence of points), Polygon (closed region), and Multi-variants of each.
5	Define Spatial Indexing.	Spatial indexing organizes spatial data to allow efficient querying of geographic data. Common structures include R-Tree and Quad-Tree.
6	What is an Active Database?	An active database includes Event-Condition-Action (ECA) rules that automatically trigger actions when specific events occur and conditions are met.
7	What is a Mobile Database?	A mobile database is a database that can be connected to a mobile computing environment, supporting disconnected operations and location-aware queries.
8	Define Handoff Management.	Handoff management in mobile databases is the process of transferring a mobile user's connection from one base station to another without interrupting service.
9	What is a Deductive Database?	A deductive database uses logic rules to derive new facts from existing data, combining the features of relational databases and logic programming.
10	What is a Multimedia Database?	A multimedia database stores and retrieves multiple types of media such as text, images, audio, video, and animation in an integrated manner.
11	What is Temporal Querying?	Temporal querying allows users to query data across different time periods, retrieving historical, current, or future states of data.
12	What is a Mobile Transaction Model?	A mobile transaction model handles transactions in wireless environments with frequent

	disconnections using models like Kangaroo, Clustering, or MDBS.
--	---

PART B – 16 Mark Questions

Q.No	Question (16 Marks)
1	Explain Temporal Databases in detail. Discuss valid time, transaction time, and bi-temporal models with examples.
2	Describe Spatial Databases: spatial data types, spatial operators, spatial queries, and spatial indexing techniques.
3	Explain the concept of Active Databases and ECA (Event-Condition-Action) rules with examples.
4	Discuss Mobile Databases: characteristics, location management, handoff management, and mobile transaction models.
5	Write about Multimedia Databases: data types, storage, retrieval techniques, and challenges.

UNIT III – NOSQL DATABASES

PART A – 2 Mark Questions with Answers

Q.No	Question	Answer
1	What is NoSQL?	NoSQL (Not Only SQL) refers to a category of databases that are non-relational, distributed, and designed to handle large volumes of unstructured or semi-structured data.
2	State the CAP Theorem.	The CAP Theorem states that a distributed system can only guarantee two of three properties simultaneously: Consistency, Availability, and Partition Tolerance.
3	What is Sharding?	Sharding is a database partitioning technique that splits data across multiple machines to improve performance and scalability.
4	What is a Document-Based Database?	A document-based database stores data as semi-structured documents (e.g., JSON, BSON). MongoDB is a popular example.
5	What is MongoDB?	MongoDB is an open-source, document-oriented NoSQL database that stores data in BSON format and supports dynamic schemas.
6	Define Key Space in Cassandra.	A Key Space in Cassandra is the outermost container for data, similar to a schema in relational databases, holding column families.
7	What is CQL?	CQL (Cassandra Query Language) is the query language used to interact with Apache Cassandra databases, syntactically similar to SQL.
8	What is HIVE?	Apache Hive is a data warehouse infrastructure built on Hadoop that allows querying and managing large datasets using a SQL-like language called HiveQL.
9	What is a Graph Database?	A graph database uses graph structures (nodes, edges, properties) to store, map, and query relationships between data. OrientDB is an example.
10	Define Replication in MongoDB.	Replication in MongoDB is the process of synchronizing data across multiple servers using a replica set for redundancy and high availability.
11	What is HiveQL?	HiveQL is the SQL-like query language used in Apache Hive to query data stored in Hadoop Distributed File System (HDFS).
12	What are OrientDB features?	OrientDB supports multi-model data (graph, document, key-value, object), ACID transactions, SQL-like queries, and distributed architecture.

PART B – 16 Mark Questions

Q.No	Question (16 Marks)
1	Explain the CAP Theorem and discuss how different NoSQL databases handle its trade-offs.
2	Describe MongoDB in detail: data model, CRUD operations, indexing, replication, and sharding.
3	Explain Apache Cassandra: data model, Key Space, column families, CRUD operations, and CQL with examples.
4	Discuss Apache HIVE: architecture, data types, database operations, partitioning, and HiveQL with examples.
5	Explain OrientDB as a graph database. Discuss its data model, graph features, and OrientDB query language.

UNIT IV – XML DATABASES

PART A – 2 Mark Questions with Answers

Q.No	Question	Answer
1	What is XML?	XML (Extensible Markup Language) is a markup language designed to store and transport data in a human-readable and machine-readable format.
2	Differentiate structured and semi-structured data.	Structured data has a fixed schema (e.g., relational tables). Semi-structured data has flexible, self-describing structure (e.g., XML, JSON).
3	What is a Well-Formed XML Document?	A well-formed XML document follows XML syntax rules: single root element, properly nested tags, attribute values in quotes, and case-sensitive tags.
4	What is a Valid XML Document?	A valid XML document is well-formed AND conforms to a schema defined by a DTD (Document Type Definition) or XML Schema (XSD).
5	What is DTD?	DTD (Document Type Definition) defines the legal building blocks of an XML document, specifying elements, attributes, and their structure.
6	What is XML Schema (XSD)?	XML Schema is a more powerful alternative to DTD that defines the structure, data types, and constraints of an XML document using XML syntax itself.
7	What is XPath?	XPath (XML Path Language) is a language used to navigate through elements and attributes in an XML document using path expressions.
8	What is XQuery?	XQuery is a query language designed to query XML data, similar to SQL for relational databases.
9	What is the XML Hierarchical Data Model?	The XML Hierarchical Data Model represents XML data as a tree structure with a root node, element nodes, attribute nodes, and text nodes.
10	What is an XML Document?	An XML document is a text file containing user-defined tags that describe and carry data in a structured, self-describing format.
11	How is XML used with Databases?	XML can store relational data, exchange data between databases, and be queried using XQuery or stored natively in XML databases.
12	What is an XML element vs attribute?	An XML element is defined by start/end tags and may contain content. An attribute provides additional information about an element within its start tag.

PART B – 16 Mark Questions

Q.No	Question (16 Marks)
1	Explain the XML Hierarchical Data Model. Discuss elements, attributes, and the tree representation with examples.
2	What is DTD? Explain DTD syntax, element declarations, and attribute declarations with examples.
3	Explain XML Schema (XSD) in detail with examples. How does it differ from DTD?
4	Discuss XML Querying: explain XPath expressions and XQuery language with detailed examples.
5	Explain how XML integrates with databases. Discuss XML document storage, retrieval, and the mapping between relational and XML data.

UNIT V – INFORMATION RETRIEVAL AND WEB SEARCH

PART A – 2 Mark Questions with Answers

Q.No	Question	Answer
1	What is Information Retrieval (IR)?	Information Retrieval is the process of obtaining relevant information resources from a collection of resources in response to a user query.
2	What is an Inverted Index?	An inverted index maps each term/word to the list of documents in which it appears, enabling fast full-text search.
3	Define Precision and Recall.	Precision is the ratio of relevant retrieved documents to total retrieved documents. Recall is the ratio of relevant retrieved documents to total relevant documents.
4	What is the Boolean Retrieval Model?	The Boolean Retrieval Model retrieves documents based on exact matching using Boolean operators (AND, OR, NOT) applied to query terms.
5	What is the Vector Space Model?	The Vector Space Model represents documents and queries as vectors in a multi-dimensional term space and uses cosine similarity to measure relevance.
6	What is TF-IDF?	TF-IDF (Term Frequency–Inverse Document Frequency) is a weighting scheme that assigns importance to terms based on their frequency in a document relative to the entire corpus.
7	What is Text Pre-processing?	Text pre-processing involves tokenization, stop-word removal, stemming, and lemmatization to normalize text before indexing or retrieval.
8	What is Stemming?	Stemming is the process of reducing words to their root/base form (e.g., 'running' → 'run') to improve retrieval by matching morphological variants.
9	What is Web Search?	Web search is the process of using search engines (e.g., Google) to retrieve relevant web pages from the internet based on user queries.
10	What is a Web Crawler?	A web crawler (spider/bot) is an automated program that browses the web systematically to index web page content for search engines.
11	What is PageRank?	PageRank is Google's algorithm that ranks web pages based on the number and quality of links pointing to them, used to measure page importance.
12	What is the Probabilistic Retrieval Model?	The Probabilistic Model ranks documents by the probability that they are relevant to a query, based on statistical analysis of term occurrences.

PART B – 16 Mark Questions

Q.No	Question (16 Marks)
1	Explain Information Retrieval concepts: retrieval models (Boolean, Vector Space, Probabilistic) with examples.
2	Discuss the Vector Space Model in detail. Explain TF-IDF weighting and cosine similarity with examples.
3	Explain Text Pre-processing techniques: tokenization, stop-word removal, stemming, and lemmatization with examples.
4	Describe Inverted Indexing: structure, construction, and how it is used in query processing.
5	Explain Web Search and Analytics: discuss web crawling, indexing, PageRank algorithm, and current trends in web search.