			Marks	BL	CO	PO	PSO
OR	iii	Explain fundamental difference between column as databases & other types (like row-based relational databases on documents databases).	6	3	4	1	1,3

Q.6   Attempt any two-

			Marks	BL	CO	PO	PSO
	i.	List the types of search engines and state the major differences between them in terms of use cases.	5	2	5	3	1,3
	ii.	In the context of search engines, how does reverse indexing improve search efficiency-	5	2	5	1	1,3
	iii.	Web crawlers play a crucial role in search engines. List and explain the steps a web crawler follows to gather and update information from websites.	5	3	5	2	1,3

*****

Enrollment No.....................................

Faculty of Engineering
End Sem Examination Dec 2024
CS3EL17 No SQL Database

Programme: B.Tech.   Branch/Specialisation: CSE All

Duration: 3 Hrs.   **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

			Marks	BL	CO	PO	PSO
Q.1	i.	Which of the following is not a characteristic feature of NoSQL databases? (a) Horizontal scalability (b) Schema flexibility (c) Support for ACID transactions by default (d) Key-value storage mechanism	1	1	1	1,2,3	1
	ii.	What is the primary focus of the BASE model in NoSQL databases? (a) Ensuring strong consistency (b) Achieving scalability with eventual consistency (c) Strict data partitioning rules (d) Handling complex joins between data	1	1	1	1	1
	iii.	Why might a business choose a commercial NoSQL solution over open source? (a) Lower costs (b) Better security and support (c) Faster deployment (d) More flexibility	1	2	2	1	1
	iv.	Which of the following is a key factor in ensuring data safety in NoSQL databases? (a) Horizontal partitioning (b) ACID transaction support (c) Data replication (d) Schema flexibility	1	2	2	1	1,2

v. A Key-Value database replicates data across 3 servers for fault tolerance. If a single value is 50 KB, how much total storage will be required for 1,000 values, including replication. **1** 3 3 1 1,3
(a) 50 MB   (b) 100 MB
(c) 150 MB   (d) 200 MB

vi. Which feature of Key-Value databases ensures data remains accessible during partitioning or replication processes? **1** 4 3 1 1,3
(a) Hash functions   (b) Strong consistency
(c) Sharding   (d) Eventual consistency

vii. A graph consists of 10 nodes and forms a complete bipartite graph with two sets of nodes (5 nodes in each set). How many edges does this graph contain? **1** 3 4 1 1,3
(a) 5   (b) 10   (c) 25   (d) 50

viii. Which of the following is a scaling technique used in column family databases? **1** 4 4 1 1
(a) Sharding   (b) Vertical scaling
(c) Hash partitioning   (d) Both (a) & (c)

ix. In search engines, indexing refers to- **1** 3 5 1 1
(a) Storing and organizing web page data to facilitate faster search queries
(b) Ranking web pages based on user behavior
(c) Crawling the web to gather new data
(d) Displaying relevant ads to users

x. Reverse indexing in search engines is primarily used for- **1** 2 5 1 1
(a) Matching user queries with relevant web content
(b) Improving the crawling speed of the search engine
(c) Identifying and deleting duplicate content
(d) Reversing the ranking order of the search results

Q.2 i. Explain the concept of eventual consistency in NoSQL databases. how does it differ from strong consistency in traditional RDBMS systems? **2** 4 1 7 3

ii. Describe polyglot persistence and discuss how it supports the use of multiple database models in a **3** 5 1 10 4

single application. provide a scenario where polyglot persistence would be advantageous.

iii. The CAP theorem states that distributed databases cannot simultaneously guarantee consistency, availability, and partition tolerance. Explain and analyse this trade-off with specific example. **5** 3 1 1 3

OR iv. Explain the significance of BASE in NoSQL databases. How does this model provide flexibility in distributed systems? Compare it with the ACID model. **5** 5 1 2 1

Q.3 i. How does data visualization help in evaluating the performance of NoSQL databases? **2** 2 2 3 1

ii. Compare the pros and cons of open-source versus commercial NoSQL databases in terms of cost, support, and business impact. **8** 2 2 3 1

OR iii. Explain the challenges of scaling NoSQL databases while maintaining data safety and performance. How can businesses address these challenges? **8** 3 2 2 1,2

Q.4 i. A Key-Value store uses consistent hashing to distribute 8,000 keys across 4 nodes. If the system adds 2 more nodes to the cluster, calculate how the distribution of keys changes and how many keys will, on average, be stored on each node after the addition **3** 3 3 2 1,2

ii. Describe the role of hash functions in Key-Value databases. How do they facilitate data retrieval and storage? Discuss the advantages and limitations of using hash functions in this context. **7** 4 3 1 1,3

OR iii. Compare and contrast the consistency models used in Key-Value stores and Document databases. What are the implications of each model on data reliability and availability? **7** 3 3 3 1

Q.5 i. State the properties of a weighted graph. **4** 2 4 2 1
ii. How do graph databases help in solving network related problems? Provide a relevant use case. **6** 4 4 4 1

Marking Scheme
CS3EL17 No SQL Database

Q.1

1. Which of the following is not a characteristic feature of NoSQL databases?

Correct Answer: C) Support for ACID transactions by default
Explanation: NoSQL databases typically focus on flexibility and scalability over strict adherence to ACID transactions.

2. What is the primary focus of the BASE model in NoSQL databases?

Correct Answer: B) Achieving Scalability with Eventual Consistency
Explanation: BASE (Basically Available, Soft state, Eventual consistency) prioritizes availability and eventual consistency over strict consistency.

3. Why might a business choose a commercial NoSQL solution over open-source?

Correct Answer: B) Better security and support
Explanation: Commercial solutions offer enterprise-grade support and security, which may be important for businesses with critical applications.

4. Which of the following is a key factor in ensuring data safety in NoSQL databases?

Correct Answer: C) Data replication
Explanation: Replication ensures that multiple copies of data are stored, enhancing fault tolerance and data safety.

5. A Key-Value database replicates data across 3 servers for fault tolerance. If a single value is 50 KB, how much total storage will be required for 1,000 values, including replication?

Correct Answer: C) 150 MB
Explanation: 1,000 values × 50 KB = 50,000 KB = 50 MB. With replication across 3 servers: 50 MB × 3 = 150 MB.

6. Which feature of Key-Value databases ensures data remains accessible during partitioning or replication processes?

Correct Answer: D) Eventual consistency
Explanation: Eventual consistency ensures data availability even during replication and partitioning events, with consistency being achieved later.

7. A graph consists of 10 nodes and forms a complete bipartite graph with two sets of nodes (5 nodes in each set). How many edges does this graph contain?

Correct Answer: C) 25
Explanation: The number of edges in a complete bipartite graph is the product of the sizes of the two node sets, i.e., 5 × 5 = 25.

8. Which of the following is a scaling technique used in Column Family Databases?

Correct Answer: D) Both A and C (Sharding and Hash partitioning)
Explanation: Sharding and hash partitioning are common methods for distributing data across multiple servers in column-family databases.

9. In search engines, indexing refers to:

Correct Answer: A) Storing and organizing web page data to facilitate faster search queries
Explanation: Indexing is the process of organizing data so that search queries can be processed efficiently.

10. Reverse indexing in search engines is primarily used for:

Correct Answer: A) Matching user queries with relevant web content
Explanation: Reverse indexing helps match search terms with relevant content more efficiently by creating an inverted index.

Q.2 i. Explain the concept of eventual consistency in NoSQL databases. How does it differ from strong consistency in traditional RDBMS systems? **2**
Answer :

☐ Eventual consistency allows for temporary inconsistencies in data, but guarantees that all updates will eventually propagate to all nodes.**1M**
☐ Strong consistency, typical in RDBMS systems, ensures that all updates are immediately visible to all users, but at the cost of higher latency and reduced availability. **1M**

ii. Describe Polyglot Persistence and discuss how it supports the use of multiple database models in a single application. Provide a scenario where Polyglot Persistence would be advantageous. **3**
☐ Polyglot Persistence refers to using different types of databases in a single application to leverage the strengths of each model. **1M**
☐ Example: An e-commerce site using a document database for product catalogs and a graph database for social recommendations, optimizing both performance and scalability for specific use cases. **2M**

iii. The CAP theorem states that distributed databases cannot simultaneously guarantee consistency, availability, and partition tolerance. Explain and analyse this trade-off with specific example. **5**
Answer CAP diagram **2M**+explanation **2M**+tradeoffs **1M**

OR iv. Explain the significance of BASE in NoSQL databases. How does this model provide flexibility in distributed systems? Compare it with the ACID model. **5**
Answer : ACID /BASE **2M**+ flexibility **1M**+comparision **2M**

Q.3 i. How does data visualization help in evaluating the performance of **2**

NoSQL databases?
Define-**2M**
Answer: Data visualization helps in identifying trends, performance bottlenecks, and anomalies by graphically representing metrics such as latency, throughput, and resource utilization.

ii. Compare the pros and cons of open-source versus commercial NoSQL databases in terms of cost, support, and business impact. **8**
Answer : Pros+Cons **2M** + cost **2M**, support **2M**, business impact **2M**
☐ Open-source: Low cost, large community support, but limited official support. Suitable for small to medium businesses.
☐ Commercial: Higher cost but provides enterprise-level support, security, and compliance. Ideal for large enterprises with mission-critical workloads.

OR iii. Explain the challenges of scaling NoSQL databases while maintaining data safety and performance. How can businesses address these challenges? **8**
Answer:
☐ Challenges: Maintaining consistency during scaling, avoiding performance bottlenecks, ensuring data replication and fault tolerance. **4M**
☐ Solutions: Employing sharding, replication strategies, eventual consistency models, and monitoring tools to balance performance and safety. **4M**

Q.4 i. A Key-Value store uses consistent hashing to distribute 8,000 keys across 4 nodes. If the system adds 2 more nodes to the cluster, calculate how the distribution of keys changes and how many keys will, on average, be stored on each node after the addition **3**
Answer : After adding 2 more nodes, there are 6 nodes total. The number of keys per node will be $8{,}000 / 6 \approx 1{,}333$ keys per node.

ii. Describe the role of hash functions in Key-Value databases. How do they facilitate data retrieval and storage? Discuss the advantages and limitations of using hash functions in this context. **7**
Answer : retrieval and storage-**2M**
☐ Hash functions map keys to locations on servers, facilitating

P.T.O.

quick data storage and retrieval-**2M**

☐ Advantages: Fast access, easy partitioning. **3M**

☐ Limitations: Hash collisions, inability to perform range queries.

OR iii. Compare and contrast the consistency models used in Key-Value stores and Document databases. What are the implications of each model on data reliability and availability? **7**

Answer :

☐ Key-Value stores often use eventual consistency, allowing for higher availability but at the risk of temporary data inconsistencies. **3.5M**

☐ Document databases may use stronger consistency models but may sacrifice availability under network partitions. **3.5M or Compare 7 marks for 7 point**

Q.5 i. Define Weighted Graph. State the properties of a weighted graph. **4**

Answer : properties **4 points 4 mark**

ii. You are designing a distributed system using a graph database for a logistics company where nodes represent locations and edges represent transportation routes. The company needs to minimize the overall cost of transportation across multiple routes. How would you use graph algorithms to optimize this, and which algorithm would you select? Justify your choice. **6**

Answer:

☐ Define flexibility, define scalability, Define any graph, algorithm (DJ or MST). **3M**

☐ Any Use Case-**3M**

OR iii. Six difference 1 mark each **6**

Q.6

i. List the types of search engines and state the major differences between them in terms of use cases. **3 type 3M** **5**

Use case **2M**

Answer: types(vertical and general) +differences+ Use case

ii. In the context of search engines, how does reverse indexing improve search efficiency? **5**

Answer: index vs reverse index **3M**+ how **2M**

iii. Web crawlers play a crucial role in search engines. List and **5**

explain the steps a web crawler follows to gather and update information from websites.

Answer: working of web crwler **3M** +list+process **2M**