✅ **Distributed Computing – Short Notes (Copy Format)**

Unit-1: Concept & Overview of Distributed Database System (DDBS)

What is Distributed Database System (DDBS)?

A DDBS is a collection of logically interrelated databases distributed across multiple sites connected via a network. Each site has local autonomy but works as a single system.

Features of DDBS

- **Distribution** of data across multiple locations
- **Transparency** (location, fragmentation, replication)
- **Scalability**
- **Fault tolerance**
- **Data sharing** across sites
- **Improved performance** due to local processing

Promises of DDBS

- Higher availability
- Faster query processing
- Reliability and robustness
- Modular growth

Design Issues in DDBS

- Data distribution strategy
- Fragmentation and allocation
- Synchronization of distributed transactions
- Concurrency control
- Recovery from failures
- Query optimization

Distributed DBMS Architecture

1. **Client–Server System**
    - Clients request services; servers store & process data.
    - Centralized control with distributed access.
2. **Peer-to-Peer (P2P)**
    - All nodes have equal responsibility and share resources.
3. **Multi-Database System (MDBS)**
    - Independent databases connected.
    - No global schema; focus on interoperability.

Unit-2: Distributed Database Design

Distributed Database Design Concept

Designing how database is divided and allocated across sites.

Objective of Data Distribution

- Minimize communication cost
- Maximize local access
- Improve reliability
- Achieve transparency

Data Fragmentation

Breaking the database into smaller pieces (fragments).

Types:

1. **Horizontal Fragmentation** – splitting rows
2. **Vertical Fragmentation** – splitting columns
3. **Hybrid Fragmentation** – combination of both

Allocation of Fragments

Assigning fragments to sites based on usage, cost, and performance considerations.

Transparencies in Distributed Database

- **Fragmentation transparency**
- **Location transparency**
- **Replication transparency**
- **Local mapping transparency**

Unit-3: Distributed Transaction & Concurrency Control

Transaction Management Basics

A transaction is a sequence of operations that must satisfy ACID properties.

Objective of Distributed Transaction Management

- Ensure atomicity across multiple sites
- Synchronize distributed operations
- Handle failures at different nodes

Model for Transaction Management

- Coordinator site controls global transaction
- Participant sites perform local sub-transactions

Concurrency Control Objective

- Ensure correct synchronized execution
- Prevent conflicts in a distributed environment

Concurrency Control Anomalies

1. **Dirty Read**
2. **Unrepeatable Read**
3. **Lost Update**
4. **Phantom Read**

Distributed Serializability

Ensures the global schedule is equivalent to some serial order.

Locking-Based Algorithms

- **Two-Phase Locking (2PL)**
- **Distributed 2PL**
- Lock managers coordinate lock requests across sites

Unit-4: Distributed Deadlock & Recovery

Deadlock Introduction

Deadlock occurs when two or more transactions wait indefinitely for each other's locks.

Distributed Deadlock Prevention

Techniques to ensure system avoids deadlock (ordering resources, timestamping).

Distributed Deadlock Avoidance

System detects potential deadlock using wait-for graphs before it happens.

Distributed Deadlock Detection

- Local WFGs combined into a **global wait-for graph**
- Deadlock is detected when a cycle is found

Recovery from Deadlock

- Abort one or more transactions
- Restart them later

Commit Protocols

1. **Two-Phase Commit (2PC)**
    - Phase-1: Prepare
    - Phase-2: Commit/Abort
    - Ensures atomic commit across sites.
2. **Three-Phase Commit (3PC)**
    - Adds an extra "pre-commit" phase
    - Non-blocking and safer in failures

Unit-5: Distributed Query Processing & Optimization

Concepts

DQP deals with executing queries over distributed and fragmented data.

Objectives

- Reduce communication cost
- Minimize response time
- Forward only required data

Phases of DQP

1. **Query decomposition**
2. **Data localization**
3. **Global optimization**
4. **Plan generation and execution**

Join Strategies for Fragmented Relations

- **Semi-join**
- **Hash join**
- **Nested loop join**
- **Sort-merge join**

Global Query Optimization

Choosing the best execution plan considering:

- Network cost
- Processing cost
- Fragment allocation

Unit-6: Heterogeneous Database

Architecture of Heterogeneous Database

Databases with different models, query languages, or platforms connected to work as one system.

Components:

- Local DBMS
- Global schema
- Wrappers/mediators

Database Integration

Integrating multiple independent databases into a unified system.

Schema Translation

Converting local schemas into a common data model.

Schema Integration

Combining translated schemas by removing conflicts:

- Naming conflicts
- Structural conflicts
- Semantic conflicts

Query Processing Issues

- Different query languages
- Different data models
- Data type incompatibility
- Semantic differences
- High communication overhead

✅ UNIT–1: Distributed Database System – MCQs (10)

1. A Distributed Database System is defined as:

A. A centralized database

B. A collection of multiple databases connected via a network

C. Only replication of data

D. Only cloud-based storage

✓ Answer: B

2. Which is NOT a feature of DDBS?

A. Transparency

B. Autonomy

C. Scalability

D. Manual data sharing

✓ Answer: D

3. The main promise of DDBS is:

A. Higher cost

B. Less availability

C. Improved performance

D. No data consistency

✓ Answer: C

4. Which is NOT a DDBS architecture?

A. Client–Server

B. Peer-to-Peer

C. Multi-database system

D. Neural network system

✓ Answer: D

5. In DDBS, transparency means:

A. Hiding system details from users

B. Showing all fragments to users

C. Manually selecting data

D. Only replication of data

✓ Answer: A

6. Peer-to-Peer architecture means:

A. One node is master

B. All nodes are equal

C. All nodes are servers
D. All nodes are clients
✓ **Answer: B**

7. MDBS stands for:
A. Multi-Database System
B. Major Distributed Banking System
C. Multi-Data Sharing System
D. None
✓ **Answer: A**

8. Distributed database improves:
A. Rendering
B. Query performance
C. Transport
D. Encryption
✓ **Answer: B**

9. A key design issue in DDBS is:
A. Painting rules
B. Fragmentation and allocation
C. Display settings
D. Color coding
✓ **Answer: B**

10. DDBS provides improved reliability due to:
A. Centralization
B. Replication
C. Disconnection
D. Manual access
✓ **Answer: B**

✅ UNIT–2: Distributed Database Design – MCQs (10)

1. Fragmentation means:
A. Combining tables
B. Breaking database into pieces
C. Removing data

D. Encrypting data

✓ **Answer: B**

2. Horizontal fragmentation splits:

A. Columns

B. Rows

C. Database into files

D. Indexes

✓ **Answer: B**

3. Vertical fragmentation splits:

A. Rows

B. Columns

C. Tables

D. Databases

✓ **Answer: B**

4. The main objective of data distribution is:

A. Increase cost

B. Reduce performance

C. Improve local access

D. Lose transparency

✓ **Answer: C**

5. Allocation of fragments means:

A. Deleting fragments

B. Storing fragments at specific sites

C. Encrypting fragments

D. Combining fragments

✓ **Answer: B**

6. Fragmentation transparency hides:

A. Location of data

B. Number of fragments

C. Data model

D. User interface

✓ **Answer: B**

7. Which transparency hides WHERE data is stored?

A. Fragmentation

B. Replication
C. Location
D. Mapping
✓ **Answer: C**

8. Hybrid fragmentation is combination of:
A. Vertical + Replication
B. Horizontal + Vertical
C. Mapping + Querying
D. None
✓ **Answer: B**

9. Which is NOT a fragmentation type?
A. Horizontal
B. Vertical
C. Derived
D. Circular
✓ **Answer: D**

10. Distributed design aims to reduce:
A. Communication cost
B. Security
C. Redundancy
D. Transactions
✓ **Answer: A**

✅ UNIT–3: Distributed Transaction & Concurrency Control – MCQs (10)

1. A transaction must satisfy:
A. ACID
B. HTML
C. TCP/IP
D. WiFi
✓ **Answer: A**

2. Distributed transaction management ensures:
A. Local atomicity only
B. Global atomicity
C. No atomicity

D. Manual control

✓ **Answer: B**

3. Which is NOT a concurrency anomaly?

A. Lost Update

B. Dirty Read

C. Unrepeatable Read

D. Normalization

✓ **Answer: D**

4. Distributed serializability ensures:

A. Random execution

B. Equivalent to serial schedule

C. High cost

D. Data loss

✓ **Answer: B**

5. 2PL stands for:

A. Two Phase Locking

B. Two Page Logging

C. Two Phase Loading

D. None

✓ **Answer: A**

6. Which locking method is used in distributed systems?

A. Distributed 2PL

B. Vertical locking

C. Hardware locking

D. Physical lock

✓ **Answer: A**

7. Dirty read occurs when:

A. Reading committed data

B. Reading uncommitted data

C. Writing committed data

D. Writing uncommitted data

✓ **Answer: B**

8. Concurrency control avoids:

A. Parallel processing

B. Conflicts
C. Data storage
D. Compilation
✓ **Answer: B**

9. Coordinator in distributed transaction:
A. Executes all operations
B. Controls global commit
C. Deletes logs
D. Blocks communication
✓ **Answer: B**

10. Unrepeatable reads happen when:
A. Same data changes between reads
B. Data deleted permanently
C. Lock is held constantly
D. None
✓ **Answer: A**

✅ UNIT–4: Distributed Deadlock & Recovery – MCQs (10)

1. Deadlock occurs when:
A. CPU overheats
B. Processes wait for each other indefinitely
C. System restarts
D. Data updates
✓ **Answer: B**

2. Distributed deadlock detection uses:
A. Global wait-for graph
B. Commit logs
C. Encryption
D. CPU scheduling
✓ **Answer: A**

3. Deadlock prevention works by:
A. Avoiding resource ordering
B. Using timestamps to stop deadlock
C. Killing systems

D. Deleting files

✓ **Answer: B**

4. Deadlock recovery is done by:

A. Aborting transactions

B. Deleting tables

C. Resetting DBMS

D. Reinstalling OS

✓ **Answer: A**

5. 2PC stands for:

A. Two Phase Communication

B. Two Phase Commit

C. Two Page Commit

D. None

✓ **Answer: B**

6. In 2PC, Phase 1 is:

A. Abort

B. Prepare

C. Pre-commit

D. Lock

✓ **Answer: B**

7. 3PC prevents:

A. Data replication

B. Blocking

C. Query execution

D. Storage issues

✓ **Answer: B**

8. Deadlock avoidance predicts:

A. Deadlock before it occurs

B. Deadlock after it occurs

C. Data retrieval

D. Query optimization

✓ **Answer: A**

9. Cycles in WFG indicate:

A. Parallelism

B. Deadlock
C. Data mining
D. Concurrency
✓ **Answer: B**

10. Distributed recovery ensures:
A. Partial rollback
B. Database consistency
C. Data deletion
D. Timeout
✓ **Answer: B**

✅ UNIT–5: Distributed Query Processing & Optimization – MCQs (10)

1. DQP stands for:
A. Distributed Query Processing
B. Data Quality Protocol
C. Dual Query Processing
D. Distributed Quick Program
✓ **Answer: A**

2. Main objective of DQP:
A. Increase cost
B. Reduce response time
C. Reduce transparency
D. None
✓ **Answer: B**

3. Which is NOT a join strategy?
A. Semi-join
B. Sort-merge join
C. Hash join
D. Binary join
✓ **Answer: D**

4. Global query optimization selects:
A. Best execution plan
B. Random plan
C. Only local plan

D. Manual plan

✓ **Answer: A**

5. Semi-join reduces:

A. Communication cost

B. Fragmentation

C. Local autonomy

D. Replication

✓ **Answer: A**

6. Query decomposition means:

A. Combining queries

B. Breaking query into smaller parts

C. Deleting queries

D. Encrypting queries

✓ **Answer: B**

7. Data localization focuses on:

A. Identifying where data is stored

B. Removing data

C. Encrypting data

D. None

✓ **Answer: A**

8. Query processing phases include:

A. Decomposition

B. Localization

C. Optimization

D. All

✓ **Answer: D**

9. The biggest cost in DQP is:

A. CPU cost

B. Disk cost

C. Communication cost

D. Memory cost

✓ **Answer: C**

10. Query optimization techniques are used to:

A. Slow down system

B. Improve efficiency
C. Delete fragments
D. Modify schemas
✓ **Answer: B**

✅ UNIT–6: Heterogeneous Database – MCQs (10)

1. Heterogeneous database means:
A. Similar DBMSs
B. Different models or languages
C. Only SQL DBMS
D. Same schema everywhere
✓ **Answer: B**

2. Main challenge in heterogeneous DB:
A. High battery usage
B. Schema differences
C. Hardware heating
D. Display errors
✓ **Answer: B**

3. Schema translation converts:
A. Local schema to common model
B. Global schema to local
C. Queries to fragments
D. Data types to numbers
✓ **Answer: A**

4. Architecture component of heterogeneous DB:
A. Local DBMS
B. Wrapper
C. Global schema
D. All
✓ **Answer: D**

5. A mediator is used to:
A. Delete data
B. Integrate queries from different databases
C. Encrypt system

D. Lock records

✓ **Answer: B**

6. Schema integration solves:

A. Naming conflicts

B. Structural conflicts

C. Semantic conflicts

D. All

✓ **Answer: D**

7. Query processing issue in heterogeneous DB:

A. Data model incompatibility

B. Same query language everywhere

C. No network

D. Zero fragmentation

✓ **Answer: A**

8. Heterogeneous databases allow:

A. Multiple independent DBs to work together

B. Only one DBMS

C. No integration

D. Only SQL queries

✓ **Answer: A**

9. Wrapper architecture is used in:

A. Mediator systems

B. 3-tier web systems

C. Hardware caching

D. Encryption

✓ **Answer: A**

10. Heterogeneous DBMS challenge:

A. Query transformation

B. Deadlock

C. Fragment allocation

D. 2PC

✓ **Answer: A**