

# UNIT-IV

## Distributed DBMS Reliability

## Parallel Database Systems

# Reliability Concepts and Measures

- ▶ Reliability: Probability that a system performs correctly over a specified time.
- ▶ Measures:
  - ▶ – Availability
  - ▶ – Mean Time To Failure (MTTF)
  - ▶ – Mean Time To Repair (MTTR)
  - ▶ – Mean Time Between Failures (MTBF)

# Fault Tolerance in Distributed Systems

- ▶ Ability to continue operation despite failures.
- ▶ Techniques:
  - ▶ – Redundancy
  - ▶ – Replication
  - ▶ – Recovery mechanisms
  - ▶ – Checkpointing

# Failures in Distributed DBMS

- ▶ Transaction Failures
- ▶ Site Failures
- ▶ Media Failures
- ▶ Communication Failures
- ▶ Network Partition Failures

# Local Reliability Protocols

- ▶ Local Recovery Manager
- ▶ Write-Ahead Logging (WAL)
- ▶ Checkpointing
- ▶ Undo and Redo Operations

# Distributed Reliability Protocols

- ▶ Coordinate recovery among multiple sites.
- ▶ Maintain consistency across distributed databases.

# Two-Phase Commit (2PC) Protocol

- ▶ Phase 1: Prepare/Vote
- ▶ Phase 2: Commit/Abort
- ▶ Ensures atomicity of distributed transactions.

# Three-Phase Commit (3PC) Protocol

- ▶ CanCommit Phase
- ▶ PreCommit Phase
- ▶ DoCommit Phase
- ▶ Reduces blocking problems of 2PC.

# Site Failures and Network Partitioning

- ▶ Site Failure: Complete node crash.
- ▶ Network Partitioning: Communication links break, splitting network into partitions.

# Handling Site Failures

- ▶ Failure Detection
- ▶ Recovery Procedures
- ▶ Replication Support
- ▶ Transaction Restart

# Parallel Database Systems

- ▶ Database systems that execute operations simultaneously using multiple processors and disks.

# Parallel Database Architectures

- ▶ Shared Memory Architecture
- ▶ Shared Disk Architecture
- ▶ Shared Nothing Architecture
- ▶ Hybrid Architecture

# Parallel Data Placement

- ▶ Round-Robin Partitioning
- ▶ Hash Partitioning
- ▶ Range Partitioning
- ▶ Composite Partitioning

# Parallel Query Processing

- ▶ Inter-query Parallelism
- ▶ Intra-query Parallelism
- ▶ Pipelined Parallelism
- ▶ Partitioned Parallelism

# Load Balancing

- ▶ Distributes workload evenly among processors.
- ▶ Improves resource utilization and performance.

# Database Clusters

- ▶ Group of interconnected servers working as a single database system.
- ▶ Provides scalability, availability, and fault tolerance.

# Advantages and Challenges

- ▶ Advantages:
  - ▶ – High Performance
  - ▶ – Scalability
  - ▶ – Reliability
  
- ▶ Challenges:
  - ▶ – Synchronization
  - ▶ – Communication Overhead
  - ▶ – Complex Recovery

# Conclusion

- ▶ Reliability mechanisms and parallel database techniques are essential for high-performance distributed database systems.