Slide 1: Title Slide

- Title: Advanced Database Concepts
- Subtitle: Clustering, Indexing, Denormalization, Data Control, Tuning, and Security
- Your Name/Organization
- Date

Slide 2: Database Clustering

- Concept of Database Clustering
  - Connecting multiple database instances/servers.
  - Managed by a master server.
  - Improves handling of high-volume requests.
- Necessity in Large Systems
  - Single server limitations.
  - Parallel processing with multiple servers.
- Complexity Management
  - Need for a higher-level server.

Slide 3: Database Cluster Architecture

- Diagram of a Database Cluster (from the document)
- SAN (Storage Area Network)
  - Provides access to consolidated, block-level storage.
  - Used to connect database servers.
- Cloud Database Storage
  - Third-party services.
  - Cost savings on maintenance.

Slide 4: Shared-Nothing Architecture

- Description
  - Each node is independent.
  - No shared resources.
  - No central master node.
- Scalability
  - Offers great horizontal scalability.
- Diagram of Shared-Nothing Architecture (from the document)

Slide 5: Shared-Disk Architecture

- Description
  - All nodes share access to all database servers.
  - Interconnection network between CPUs and database servers.
- Scalability

    - Limited scalability compared to shared-nothing.
- Diagram of Shared-Disk Architecture (from the document)

Slide 6: Advantages of Database Clustering

- Improved Performance
    - Distributes workload across nodes
- High Availability
    - Data replication across nodes
- Scalability
    - Easy addition of nodes
- Fault Tolerance
    - Redundancy ensures data is not lost
- Cost Savings
    - Use of commodity hardware
    - Reduced need for specialized personnel

Slide 7: Indexing in DBMS

- Purpose of Indexing
    - Optimizes database performance.
    - Minimizes disk accesses.
- Index as a Data Structure
    - Used to quickly locate and access data.
- Index Structure
    - Search key (copy of primary/candidate key).
    - Data reference (pointers to disk blocks).
- Diagram of Index Structure

Slide 8: Indexing Methods

- Types of Indexing Methods
    - Ordered indices
    - Primary Index
    - Clustering Index
    - Secondary Index
    - Dense index
    - Sparse index

Slide 9: Ordered Indices

- Description
    - Indices are sorted for faster searching.
- Example

    - Searching for a student record with ID-543.
    - Comparison of search with and without an index.

Slide 10: Primary Index

- Description
    - Index based on the primary key.
    - Primary keys are unique.
    - Efficient searching.
- Types of Primary Index
    - Dense index
    - Sparse index

Slide 11: Dense Index

- Description
    - Index record for every search key value.
    - Faster searching.
    - More space required.
- Example
    - Table with UP, USA, Nepal, UK entries.

Slide 12: Sparse Index

- Description
    - Index record for only some items.
    - Each item points to a block.
- Example
    - Table with UP, USA, Nepal, UK entries.

Slide 13: Clustering Index

- Description
    - Ordered data file.
    - Index on non-primary key columns.
    - Group records with similar characteristics
- Example
    - Employees in a company, grouped by Dept_ID.

Slide 14: Clustering Index (Continued)

- Confusing Schema
    - One disk block shared by records
- Better Technique
    - Use separate disk block for separate clusters

- Diagram of Clustering Index

Slide 15: Secondary Index

- Problem with Sparse Indexing
  - Mapping size grows with table size.
  - Slower address fetching.
- Solution: Secondary Indexing
  - Multiple levels of indexing.
  - Smaller mapping size.
  - Faster address fetching.

Slide 16: Secondary Index Example

- Example
  - Finding record with roll 111.
  - Search process through first and second-level indices.
  - Process of inserting, updating, or deleting.

Slide 17: Guidelines for Primary Index Selection

- Most Frequently Used Columns
- Unique or Highly Singular Columns
- Equality Predicate Conditions
- Even Distribution Across AMPs

Slide 18: Guidelines for Primary Index Selection (Continued)

- Non-Volatile Columns
- Many Distinct Values
- Exclusion of Certain Data Types (BLOB, CLOB, etc.)
- Non-Aggregated Columns in Join Index

Slide 19: General Indexing Guidelines

- Usefulness: Speed up queries
- Clustering: Keep related records together
- Scattering: Keep unrelated records apart
- Table Size: Avoid indexing small tables

Slide 20: Table Columns and Indexing

- Optimal Column Data Types
  - Exact numeric types (INT, BIGINT).
  - UNIQUE and NOT NULL columns.
- Column Order and Sorting

    - Order in query predicates and join conditions.
    - Keep index key short.

Slide 21: Denormalization

- Normalization
    - Breaking tables into smaller tables.
    - Join operations for data retrieval.
- Denormalization
    - Adding redundant data to a normalized database.
    - Optimizes database efficiency.
    - Alleviates issues with database queries.
- Denormalization as Optimization
    - Used *after* normalization.

Slide 22: Denormalization Example

- Student and Branch tables.
- JOIN operation to retrieve student and branch names.
- Problem with large tables: slow joins.
- Solution: add branch name to the student table.

Slide 23: Advantages of Denormalization

- Enhance Query Performance
    - Reduces number of joins.
- Make Database More Convenient
    - Avoids on-the-fly calculations.
- Facilitate and Accelerate Reporting
    - Faster statistics generation.

Slide 24: Disadvantages of Denormalization

- Increased storage space
- Data inconsistency
- Increased complexity
- Slower updates

Slide 25: Database Tuning

- Concept of Database Tuning
    - Optimizing database performance.
    - Improving query processing speed.
    - Enhancing overall system efficiency.
- Importance of Tuning

    - Faster response times.
    - Reduced resource consumption.
    - Increased throughput.

Slide 26: Subtopics of Database Tuning

- Query Optimization
    - Analyzing and improving SQL queries.
    - Using indexes effectively.
    - Avoiding inefficient constructs.
- Schema Optimization
    - Designing efficient database schemas.
    - Normalization and denormalization.
    - Choosing appropriate data types.
- Hardware Optimization
    - Optimizing server configuration.
    - Memory allocation and disk I/O.
    - Network configuration.
- Software Optimization
    - Database server settings.
    - Connection pooling.
    - Caching mechanisms.

Slide 27: Database Security

- Concept of Database Security
    - Protecting the database from threats.
    - Ensuring data confidentiality, integrity, and availability.
- Importance of Database Security
    - Preventing unauthorized access.
    - Maintaining data accuracy and reliability.
    - Complying with regulations (e.g., GDPR).
    - Avoiding financial loss and reputational damage

Slide 28: Types of Database Security

- Access Control
    - Authentication (verifying user identity).
    - Authorization (granting privileges).
    - Role-Based Access Control (RBAC).
- Data Encryption
    - Protecting data at rest and in transit.
    - Symmetric and asymmetric encryption.

    - Key management.
- Integrity Constraints
    - Ensuring data accuracy and consistency.
    - Primary keys, foreign keys, and constraints.
- Auditing
    - Tracking database activity.
    - Detecting suspicious behavior.
    - Compliance and accountability.
- Network Security
    - Firewalls
    - Intrusion Detection/Prevention Systems
    - Secure protocols (e.g., HTTPS, SSL/TLS)

Slide 29: Why Database Security Is Important

- Confidentiality
    - Preventing unauthorized disclosure
- Integrity
    - Maintaining data accuracy
- Availability
    - Ensuring continuous access
- Compliance
    - Meeting legal and regulatory requirements
- Business Continuity
    - Minimizing downtime and data loss

Slide 30: Common Threats and Challenges

- Internal Threats
    - Malicious or negligent employees.
    - Privilege escalation.
- External Threats
    - Hackers and cybercriminals.
    - SQL injection.
    - Denial-of-Service attacks.
- Data Breaches
    - Unauthorized access and disclosure.
- Data Loss
    - Accidental or intentional deletion.
    - Hardware or software failures.
- Evolving Threats

    - New attack vectors.
    - Increasingly sophisticated attacks.

Slide 31: Data Protection Tools and Platforms

- Database Management Systems (DBMS)
    - Built-in security features.
    - Access control, encryption, and auditing.
- Firewalls
    - Network security to prevent unauthorized access.
- Intrusion Detection/Prevention Systems (IDS/IPS)
    - Monitoring and blocking malicious activity.
- Encryption Tools
    - Software and hardware for data encryption.
- Access Control Systems
    - Solutions for managing user privileges.
- Vulnerability Scanners
    - Tools for identifying security weaknesses.
- Data Loss Prevention (DLP)
    - Preventing sensitive data from leaving the organization
- Security Information and Event Management (SIEM)
    - Centralized security monitoring and analysis