Introduction & Overview of Database Systems

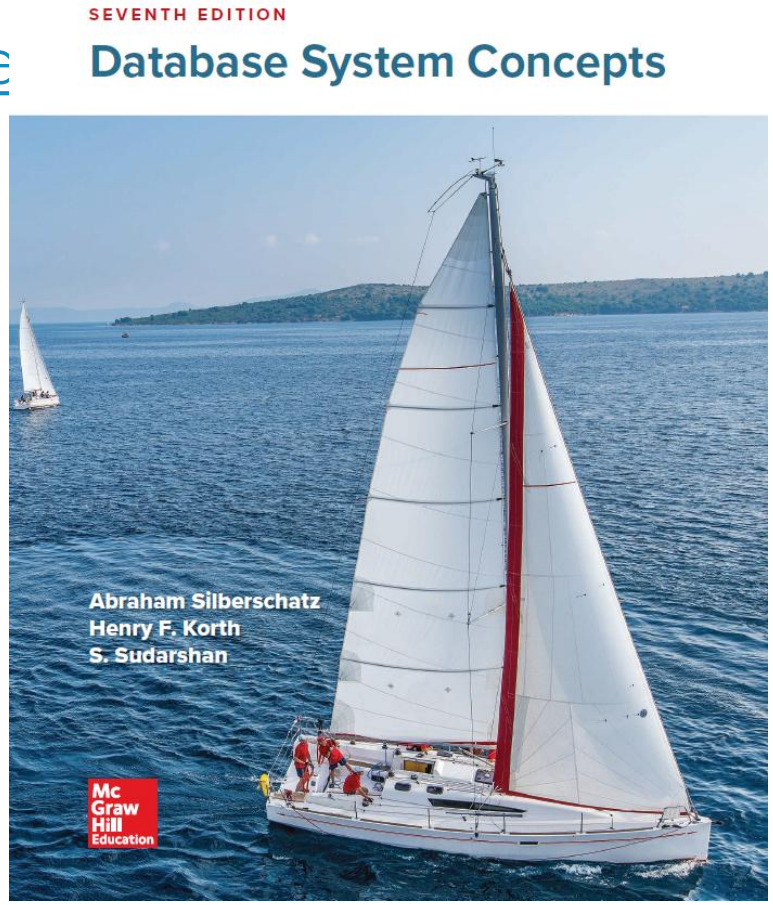
CE384: Database Design

Maryam Ramezani
Sharif University of Technology
maryam.ramezani@sharif.edu

Welcome!

- https://sut-ce-courses.github.io/database
- https://quera.org/course/add_to_course/course
  (Password: DB#SP2025)

- https://vc.sharif.edu/ch/maryam.ramezani



Course overview

Date	Session	Part	Lecture Topic
1403/11/28	S01	1	Introduction and Overview of Database Systems
1403/11/30	S02		Relations and related concepts, Types of keys
1403/12/05	S03		Relational Operations
1403/12/07	S04		Relational Algebra
1403/12/12	S05		Semantic modeling, Entity, attribute, relationship
1403/12/14	S06		ER & EER, Aggregation and Generalization
1403/12/19	S07		Converting semantic models to logical design
1403/12/21	S08	2	Schema, Data Types
1403/12/26	S09		DDL
1404/01/17	S10		DML
1404/01/19	S11		Join, DCL
1404/01/24	S12	3	View & Materialized View
1404/01/26	S13		Integrity, Assertion
1404/01/31	S14		Procedure, Function, Trigger
1404/02/02	S15		Output Control + Redirection, Window Functions, Nested Queries
1404/02/07	S16		Lateral Joins, Common Table Expressions (CTE), Recursive Queries
1404/02/09	S17	4	Functional Dependencies
1404/02/14	S18		Normalization
1404/02/16	S19		Anomalies
1404/02/21	Midterm Exam		
1404/02/23	S20	4	Query Execution Plan, Query Profiling, Indexes Clustered & Non-Clustered
1404/02/28	S21		B-Tree، B+-Tree, Hash
1404/02/30	S22		Tuning, Purging, Partitioning
1404/03/04	S23	5	Data warehouse, Data mart, Logical Design
1404/03/06	S24		Distributed Databases
1404/03/11	S25		NOSQL
1404/03/13	S26		Graph and Document based Databases, Vector Databases and AI
1404/03/22	Final Exam		
1404/04/10	DB Session		

Course overview

- Lectures
  - Goal: To introduce concepts in designing database, and motivate their use and importance.
  - Note: We try to cover useful materials in class, but we recommend you reading more!
- Assignments
  - Purpose: To give you a chance to exercise your mind, and to solidify the concepts introduced to you in class.
  - Outline: Homeworks have practical part.
  - Importance: Not important unless you want to learn the material and get a good grade!

Grading

- Final grade (22 Points) will be composed of:
  - Quizzes: 2 points (6 quizzes each 0.4 points; getting a sum of 2 points from quizzes is enough)
  - Assignments: 6 points
  - Midterm Exam: 5 points (Lectures part 1, 2, 3)
  - Final Exam: 7 points
  - Project and Presentation: 2 points (bonus; there will be a session after the final exam for presentation)

Item	Points	Description
Quizes	2	6 quizes, 10 points each, max required 50
Assignments	6	5 series
Midterm Exam	5	Lectures part 1, 2, 3
Final Exam	7	All Topics
Project and Presentation	2	Optional
Total	22	

Course overview

- Lecture slide will be uploaded.

- Many times we will write on board, in real-time, during lecture to prove a theory or answer a question or add some additional explanations. It will be your responsibility to take notes.

- Slides links will be provided on site.

Assignments

Assignment submission are not accepted after the solution is released. If you are late, submit what you have and start working on the next assignment. You have free 9 late days for assignments. Once you have used all late days, the penalty is 10% for each additional late day. You can use late days till release of the answer, remember no submission is accepted afterwards.

Assignment	Release	Deadline	Solution Release
HW1	1403/12/12	1403/12/26	1403/12/30
HW2	1403/12/28	1404/01/29	1404/02/02
HW3	1404/01/29	1404/02/17	1404/02/21
HW4	1404/02/21	1404/03/09	1404/03/13
HW5	1404/03/04	1403/03/21	1403/03/22

Exams

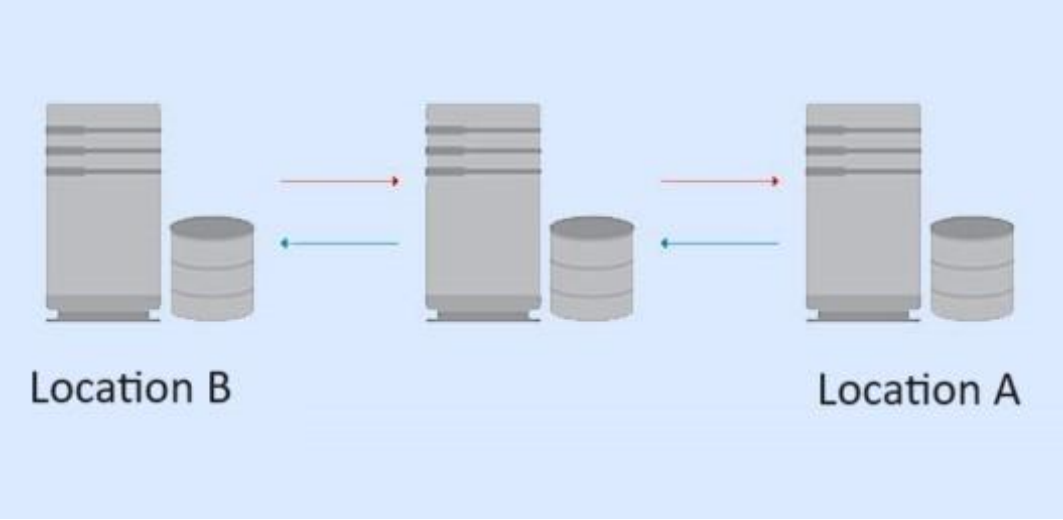
Exam	Time
Quiz 1	1403/12/12 15:00
Quiz 2	1403/12/21 15:00
Quiz 3	1404/01/24 15:00
Quiz 4	1404/02/07 15:00
Midterm	1404/02/21 15:00
Quiz 5	1404/02/28 15:00
Quiz 6	1404/03/11 15:00
Final	1404/03/22 14:30

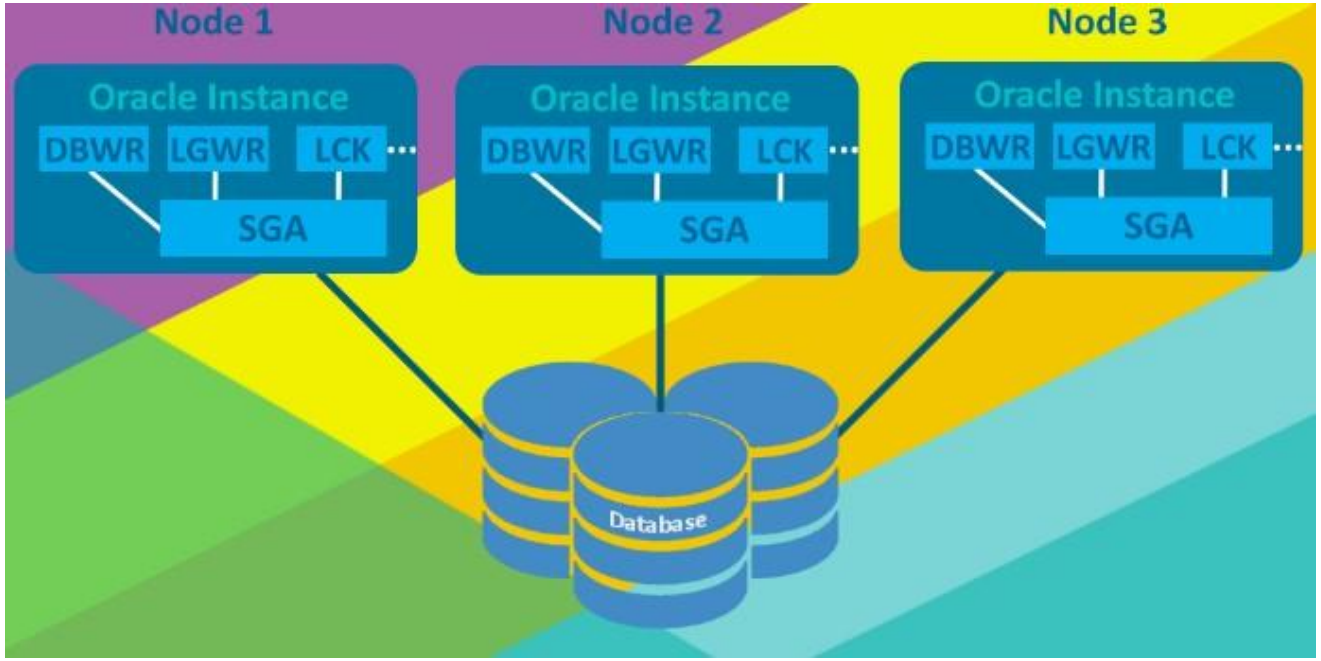
Workshops & TA Classes

Date	Title
1403/12/28	Workshop 1
1404/01/29	Workshop 2
1404/02/18	Midterm
1404/03/04	Workshop 3
1404/03/19	Final

Database Architecture

- Components and their relations
- Types:
  - Centralized architecture
  - Decentralized architecture
    - Client-server
    - Distributed
    - Architecture with parallel processing

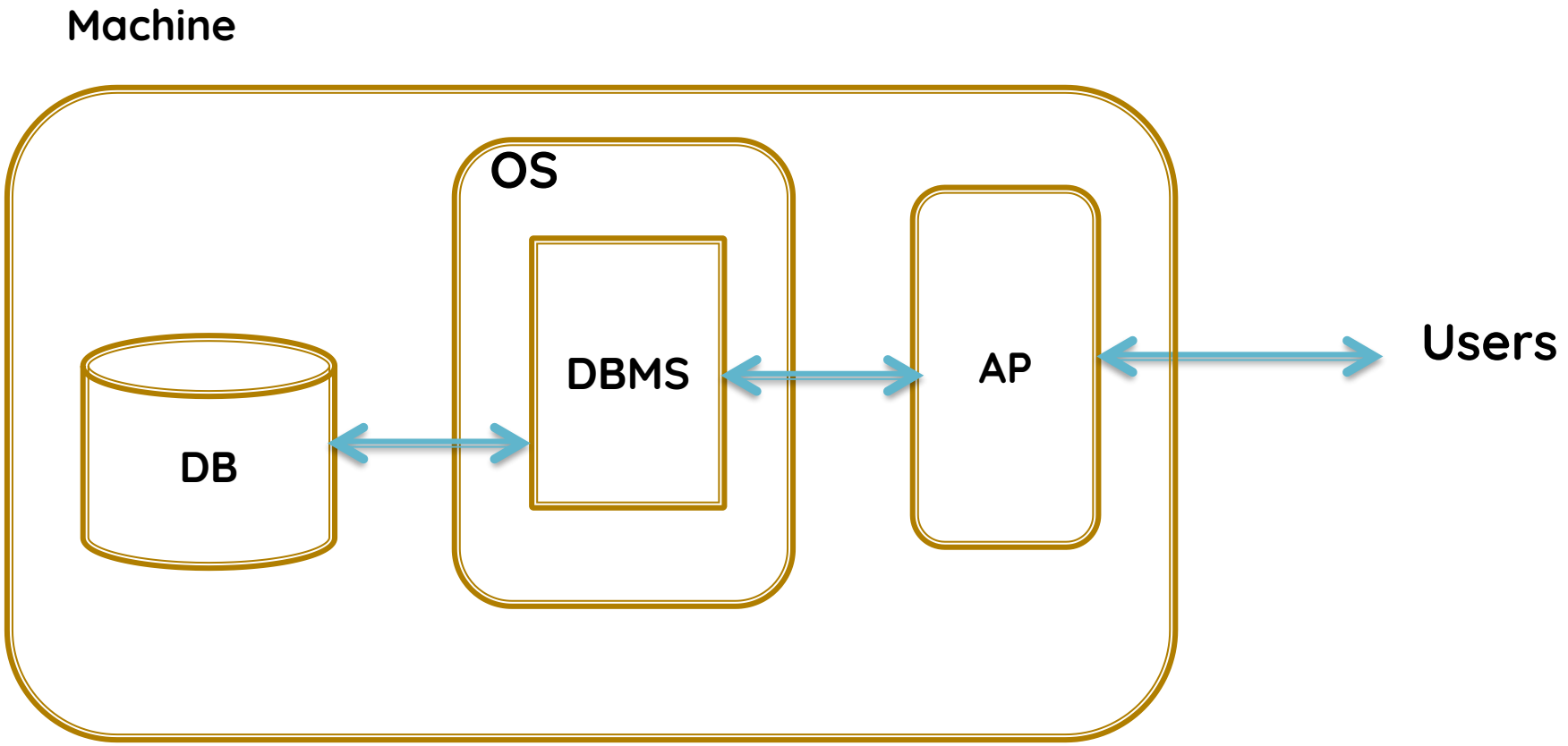




01

Centralized Architecture

Centralized Architecture



Centralized Architecture

- **Advantages**
  - Since all data is stored at a single location only thus it is easier to access and coordinate data.
  - The centralized database has very minimal data redundancy since all data is stored in a single place.
  - It is cheaper in comparison to all other databases available.
- **Disadvantages**
  - The data traffic in the case of a centralized database is more.
  - If any kind of system failure occurs in the centralized system then the entire data will be destroyed.

02

Client-server Architecture

Traditional two-tier Client–Server Architecture

First Tier

Client

Tasks

- User interface
- Main business and data processing logic

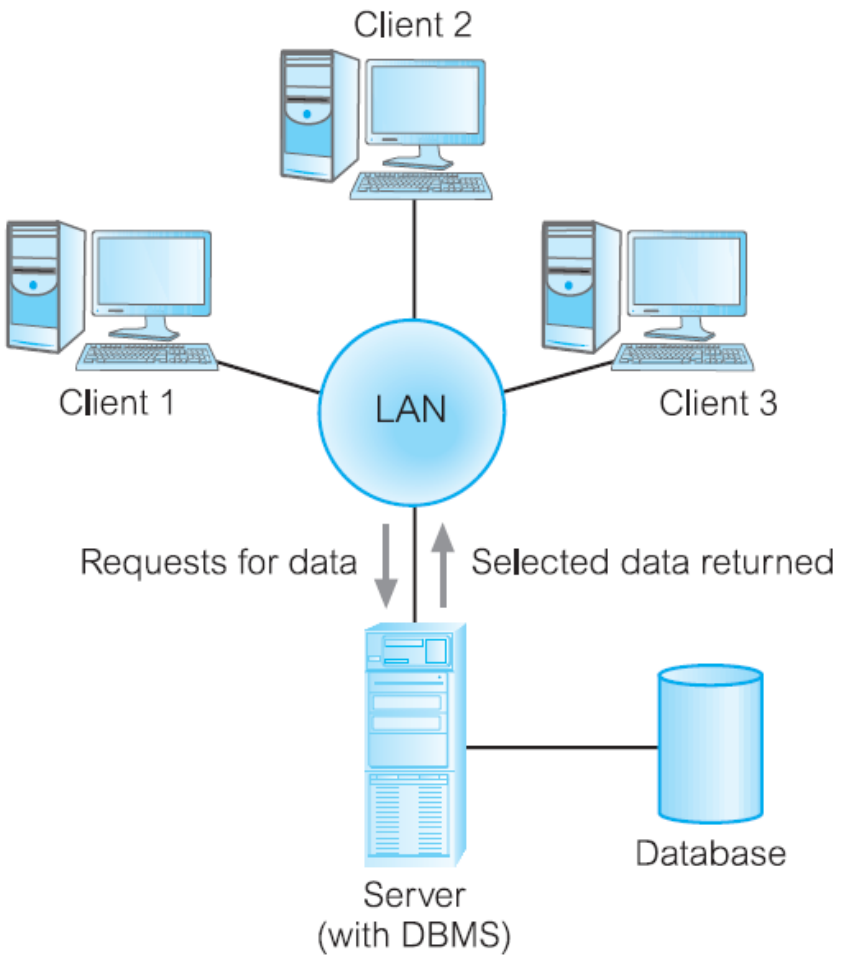
Second Tier

Database server

Tasks

- Server-side validation
- Database access

2-tier Client-server Architecture



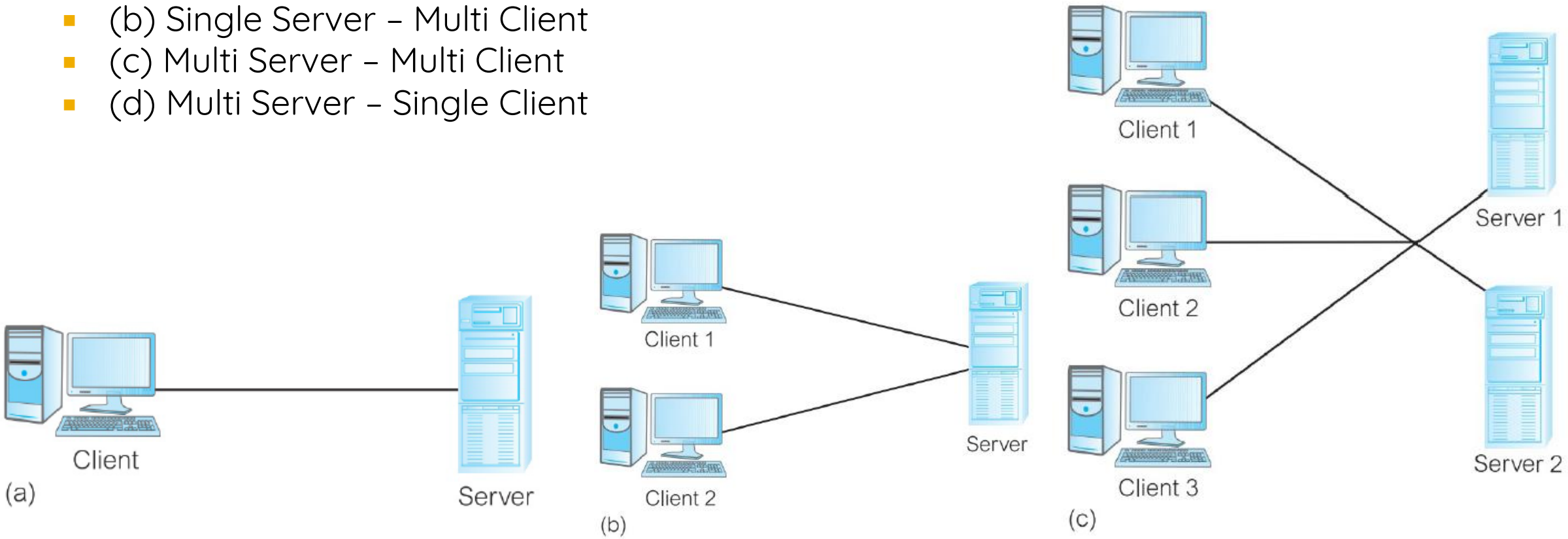
Traditional 2-tier Client–Server Architecture

- The client (tier 1) is primarily responsible for the presentation of data to the user
  - handle user interface actions and the main business and data application logic
- The server (tier 2) is primarily responsible for supplying data services to the client
  - provide limited business application logic, typically validation that the client is unable to carry out due to lack of information, and access to the requested data, independent of its location

CLIENT	SERVER
Manages the user interface	Accepts and processes database requests from clients
Accepts and checks syntax of user input	Checks authorization
Processes application logic	Ensures integrity constraints not violated
Generates database requests and transmits to server	Performs query/update processing and transmits response to client
Passes response back to user	Maintains system catalog Provides concurrent database access Provides recovery control

Client-server Architecture

- (a) Single Server – Single Client
- (b) Single Server – Multi Client
- (c) Multi Server – Multi Client
- (d) Multi Server – Single Client



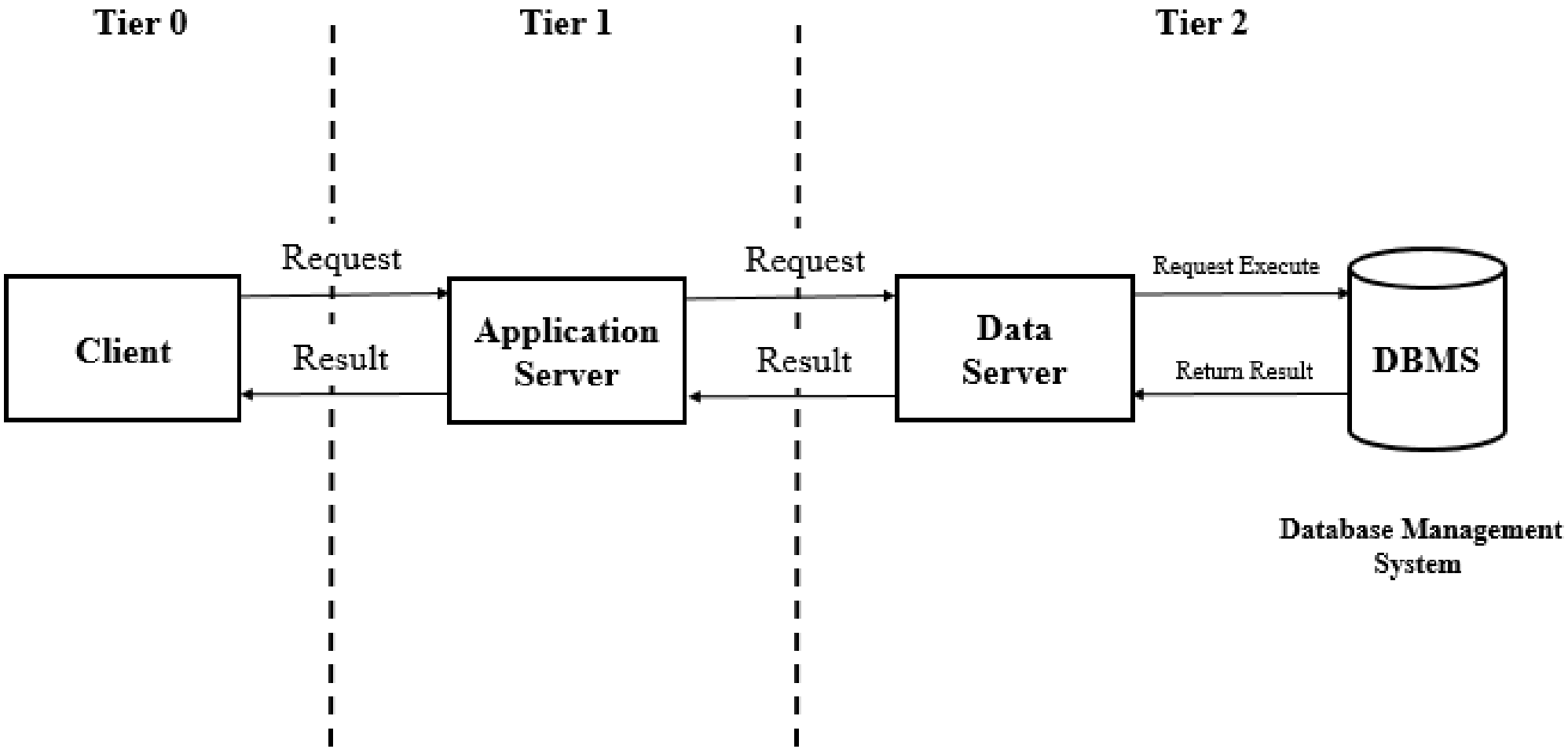
Traditional two-tier Client–Server Architecture

- Problems
  - A “fat” client, requiring considerable resources on the client’s computer to run effectively. This includes disk space, RAM, and CPU power.
  - A significant client-side administration overhead.

Three-Tier Client–Server Architecture

- The user interface layer, which runs on the end-user's computer (the thin client).
- The business logic and data processing layer. This middle tier runs on a server and is often called the application server.
- A DBMS, which stores the data required by the middle tier. This tier may run on a separate server called the database server.

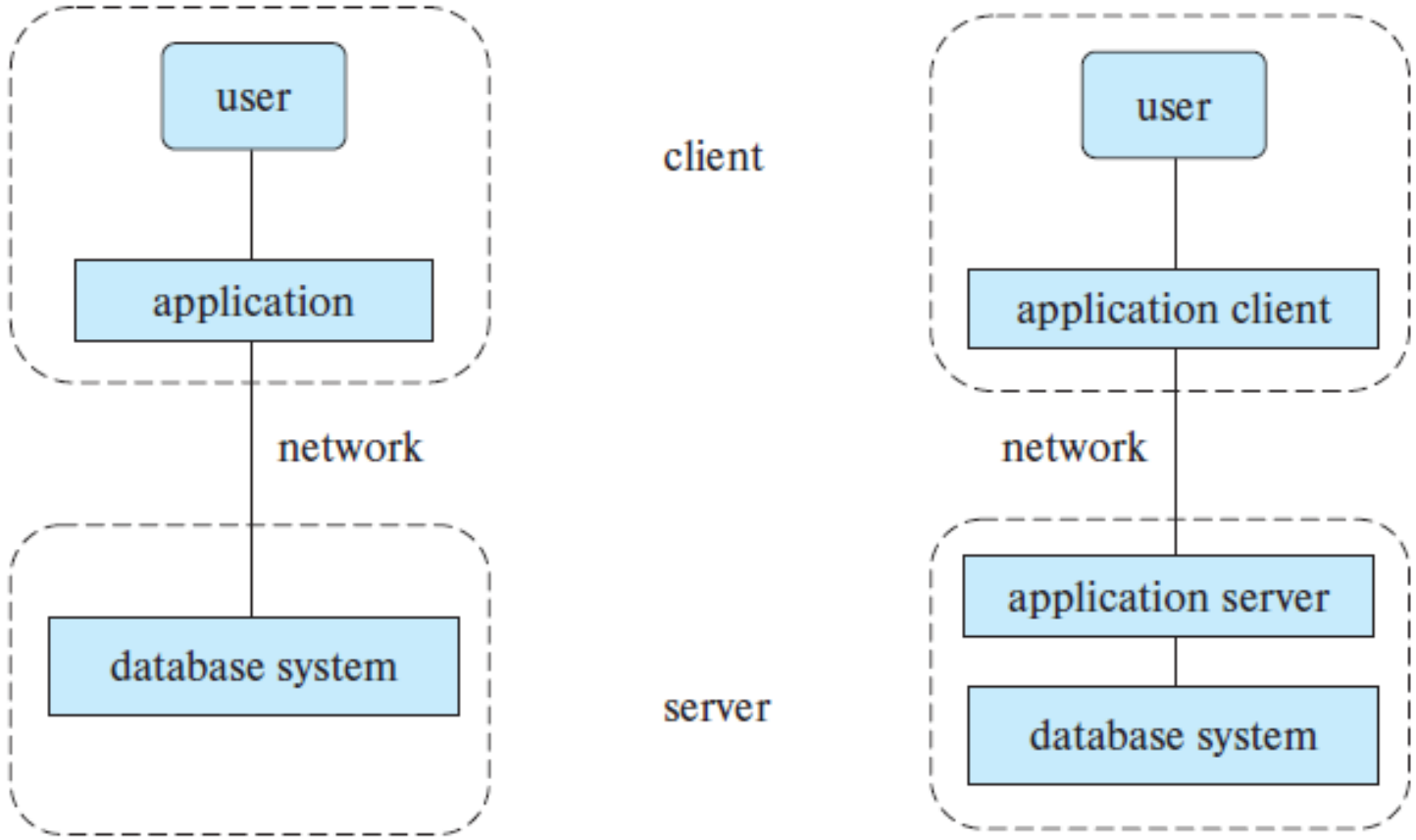
Three-Tier Client–Server Architecture



Three-Tier Client–Server Architecture

- Advantages
  - The need for less expensive hardware because the client is “thin.”
  - Application maintenance is centralized with the transfer of the business logic for many end-users into a single application server. This eliminates the concerns of software distribution that are problematic in the traditional two-tier client–server model.
  - The added modularity makes it easier to modify or replace one tier without affecting the other tiers.
  - Load balancing is easier with the separation of the core business logic from the database functions.

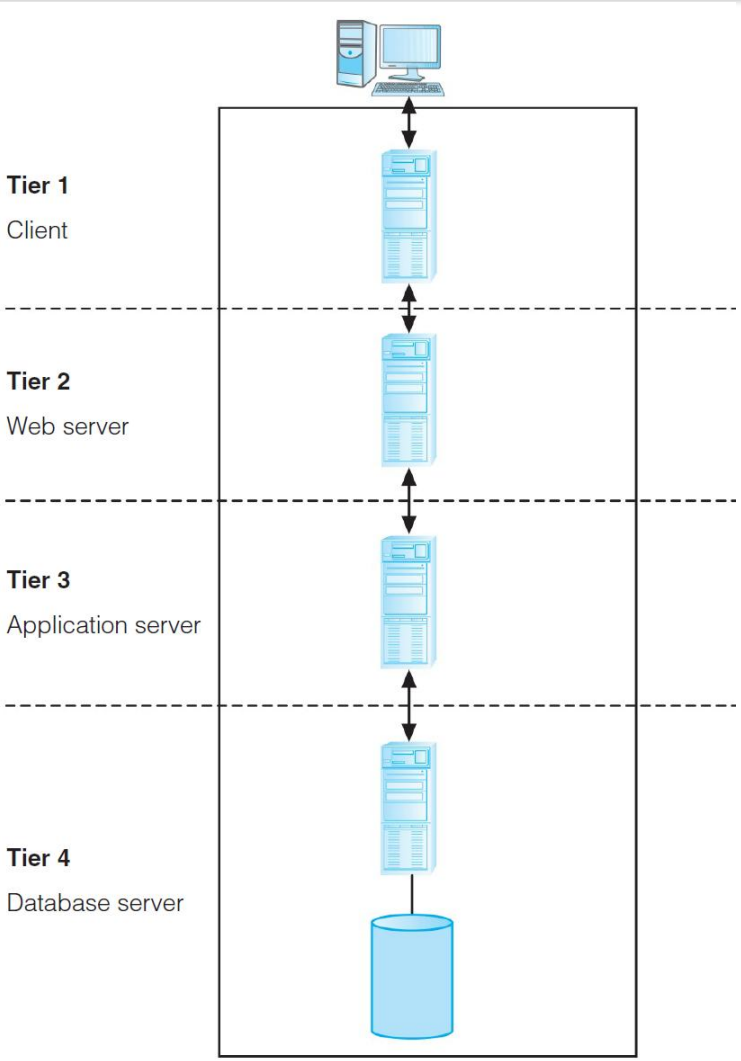
Two-tier and Three-tier Architectures

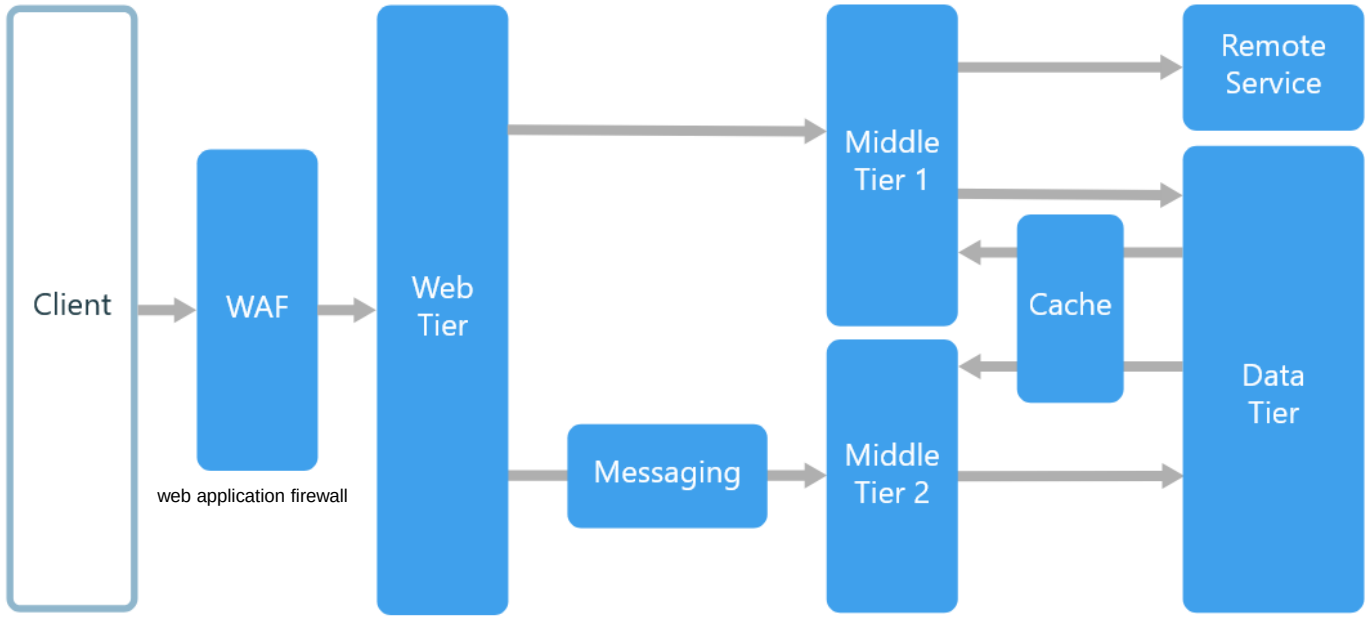


(a) Two-tier architecture

(b) Three-tier architecture

N-Tier Architectures



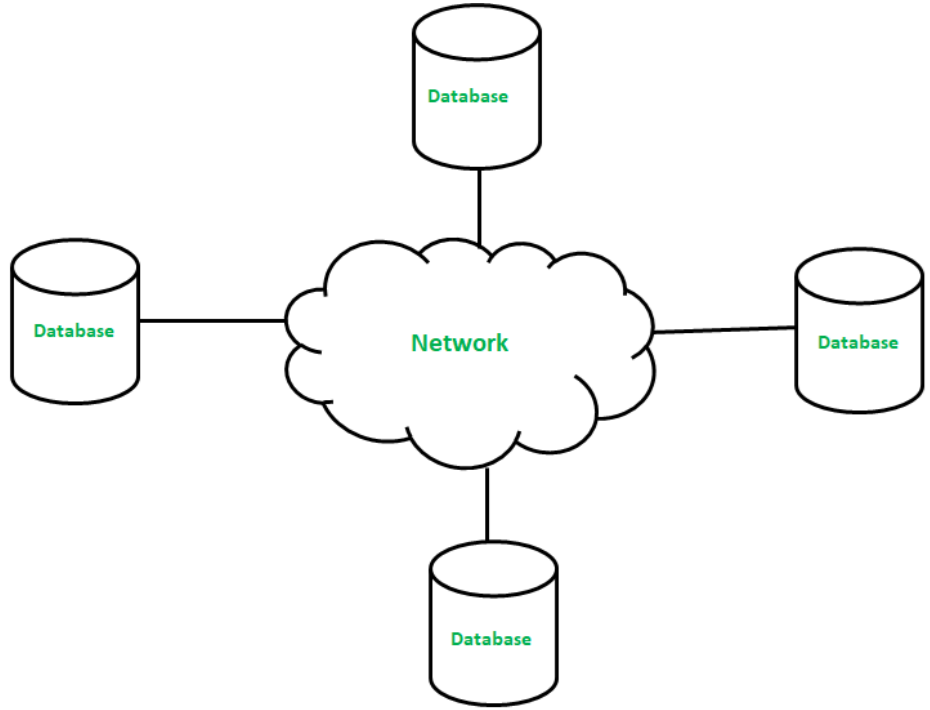


03

Distributed Architecture

Distributed Database

- A distributed database is a database that has multiple nodes distributed across the network.
- The database files can be located in different computers across the network.
- System administrators can allocate data in multiple locations.
- A distributed database is located on a network server, decentralized computers on the internet, corporate intranets, or extranets on other organization networks.
- Processes such as replication and duplication ensure that the distributed database is up to date.

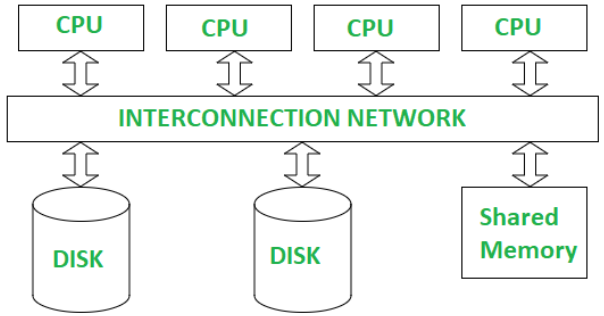


04

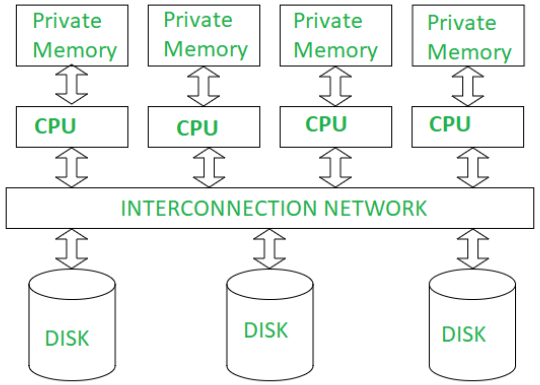
Parallel Architecture

Parallel Database

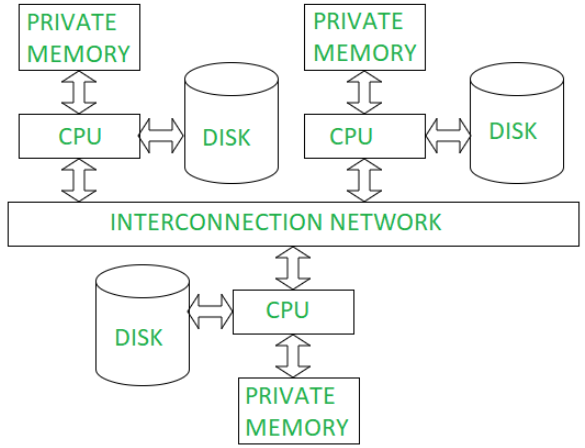
- There are multiple CPUs and disks. Multiple processors are connected in parallel.
- As they all work simultaneously, the performance of a parallel database is higher. Furthermore, the user can access the required data faster.
- Different Types:
  - Shared Memory Architecture
  - Shared Disk Architecture
  - Shared Nothing Architecture



Shared Memory



Shared Disk



Shared Nothing

Any Question?