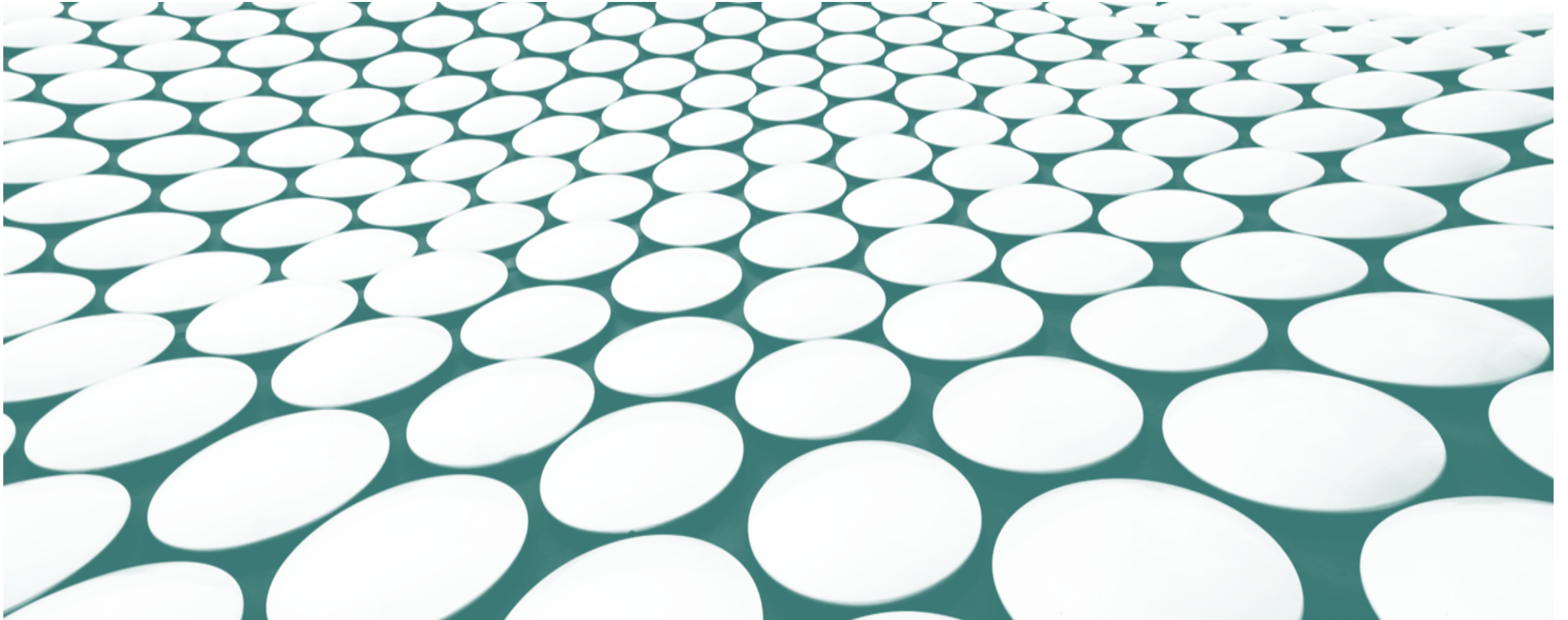


INTRODUCTION TO DDBMS

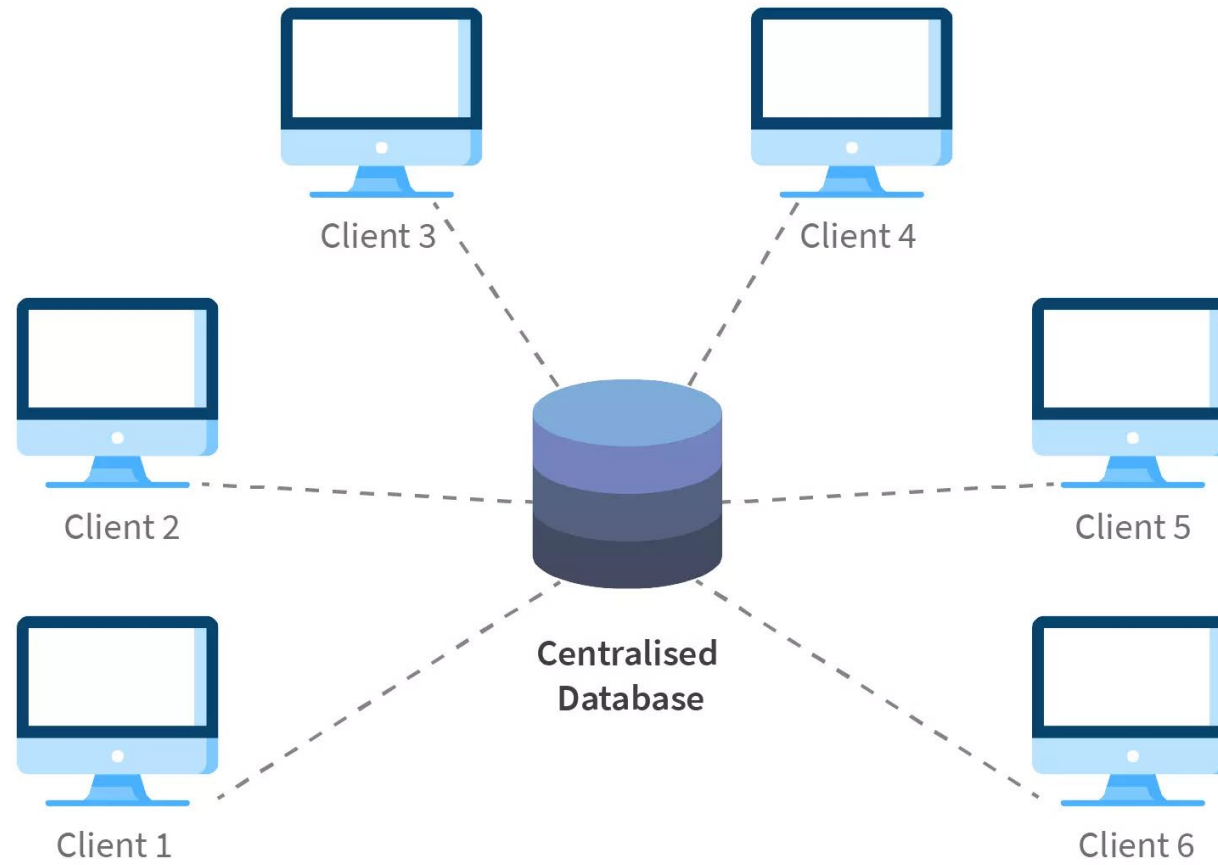
MOHAMMED D. JAWAD



INTRODUCTION TO DDBMS

- During recent times, we have seen the rapid developments in **network** and **data communication technology**, epitomized by the **Internet**, **mobile** and **wireless computing**, **intelligent devices**, and **grid computing**.
- Now, with the combination of these two technologies, **distributed database technology** may change the mode of working from **centralized** to **decentralized**.
- **Centralized Database** is a single logical database located at one site under the control of a single DBMS.

INTRODUCTION TO DDBMS

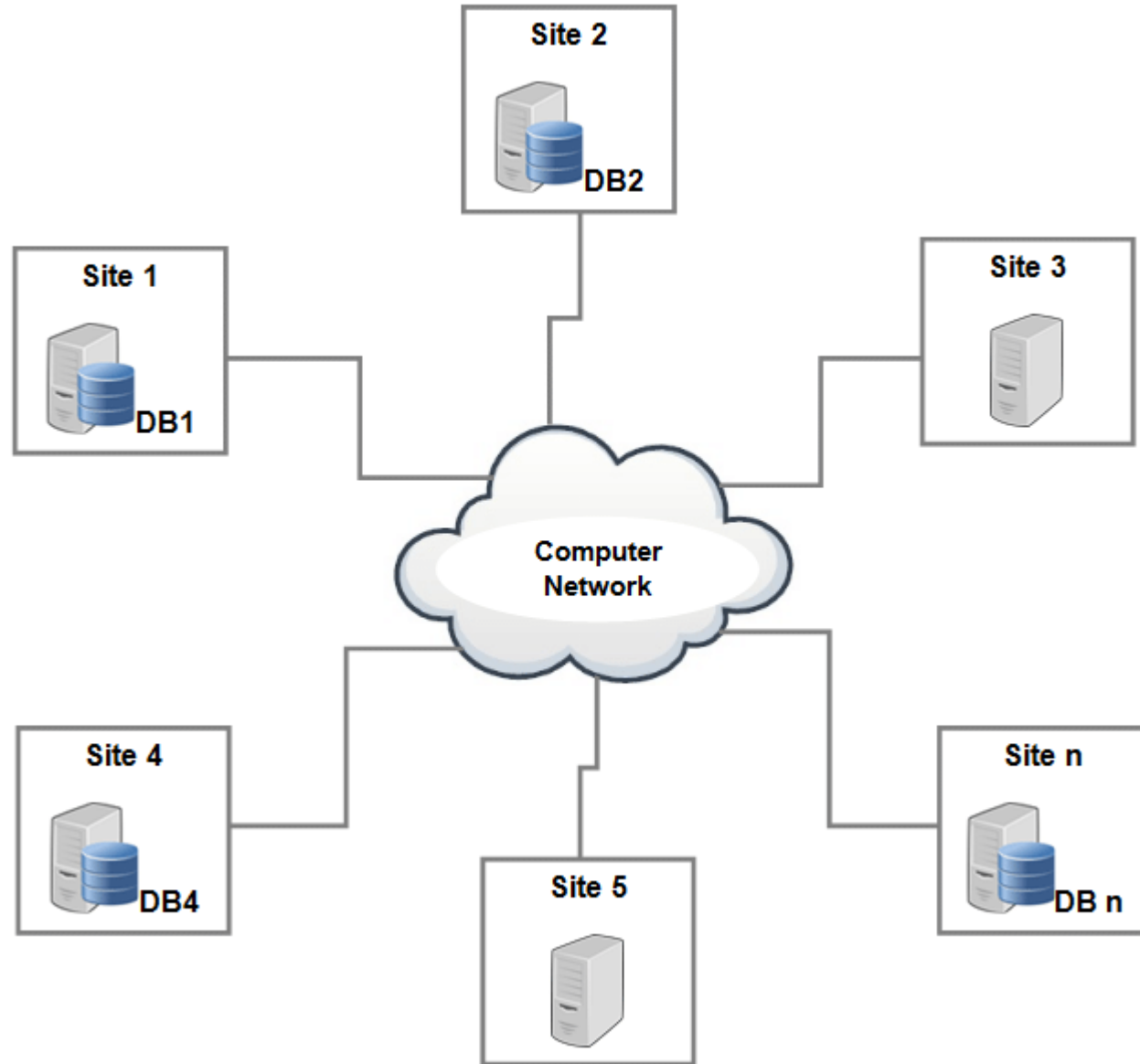


SCALER
Topics

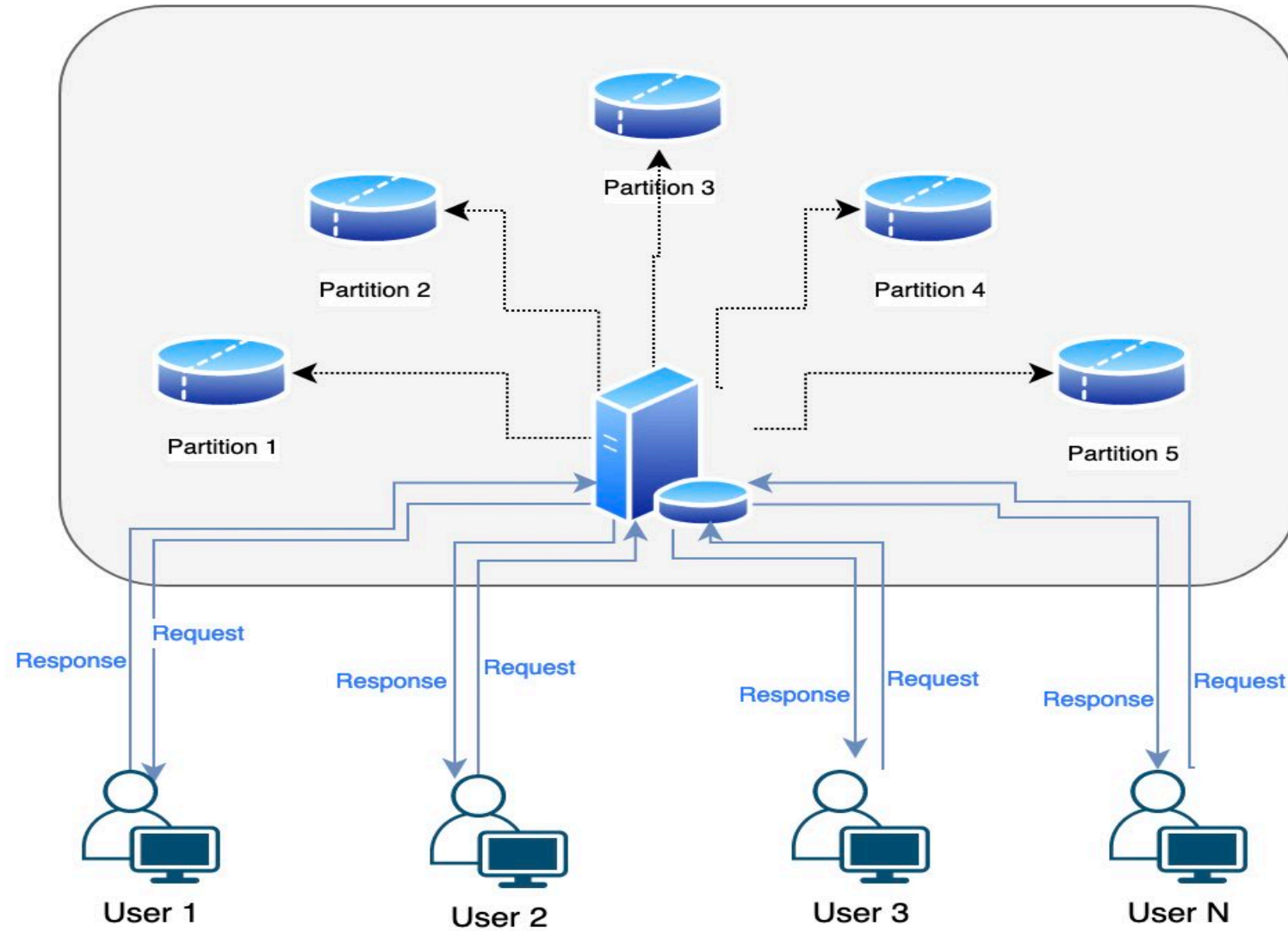
Centralized Database

- **Distributed database** A logically interrelated collection of shared data and physically distributed over a computer network.
- A distributed database is a database that is distributed across multiple nodes or servers. Each node or server stores a subset of the data, and the nodes communicate with each other to ensure that the data is consistent across all nodes.
- **Distributed DBMS** The software system that permits the management of the distributed database and makes the distribution transparent to users.

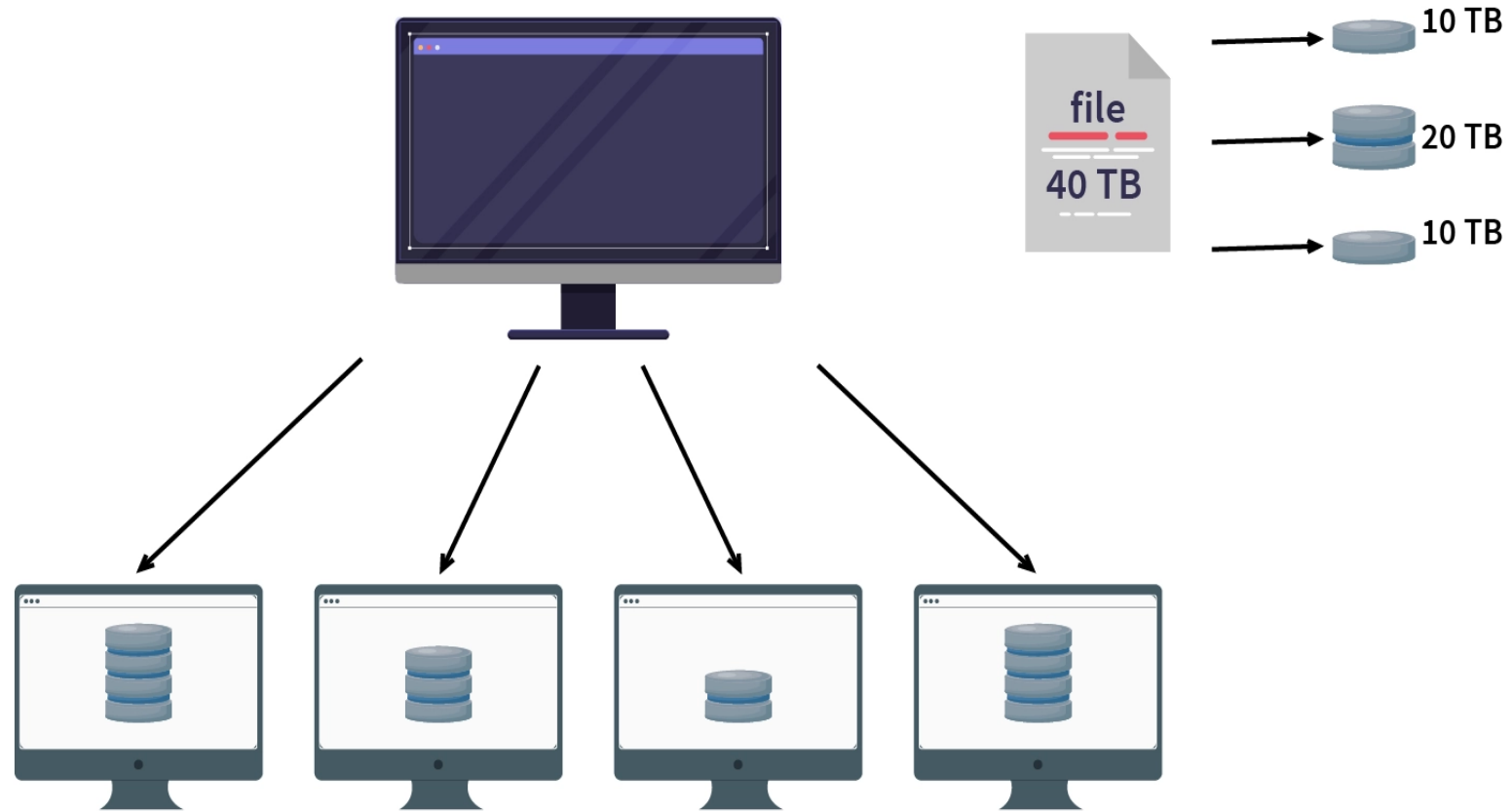
INTRODUCTION TO DDBMS



INTRODUCTION TO DDBBMS



INTRODUCTION TO DDBMS

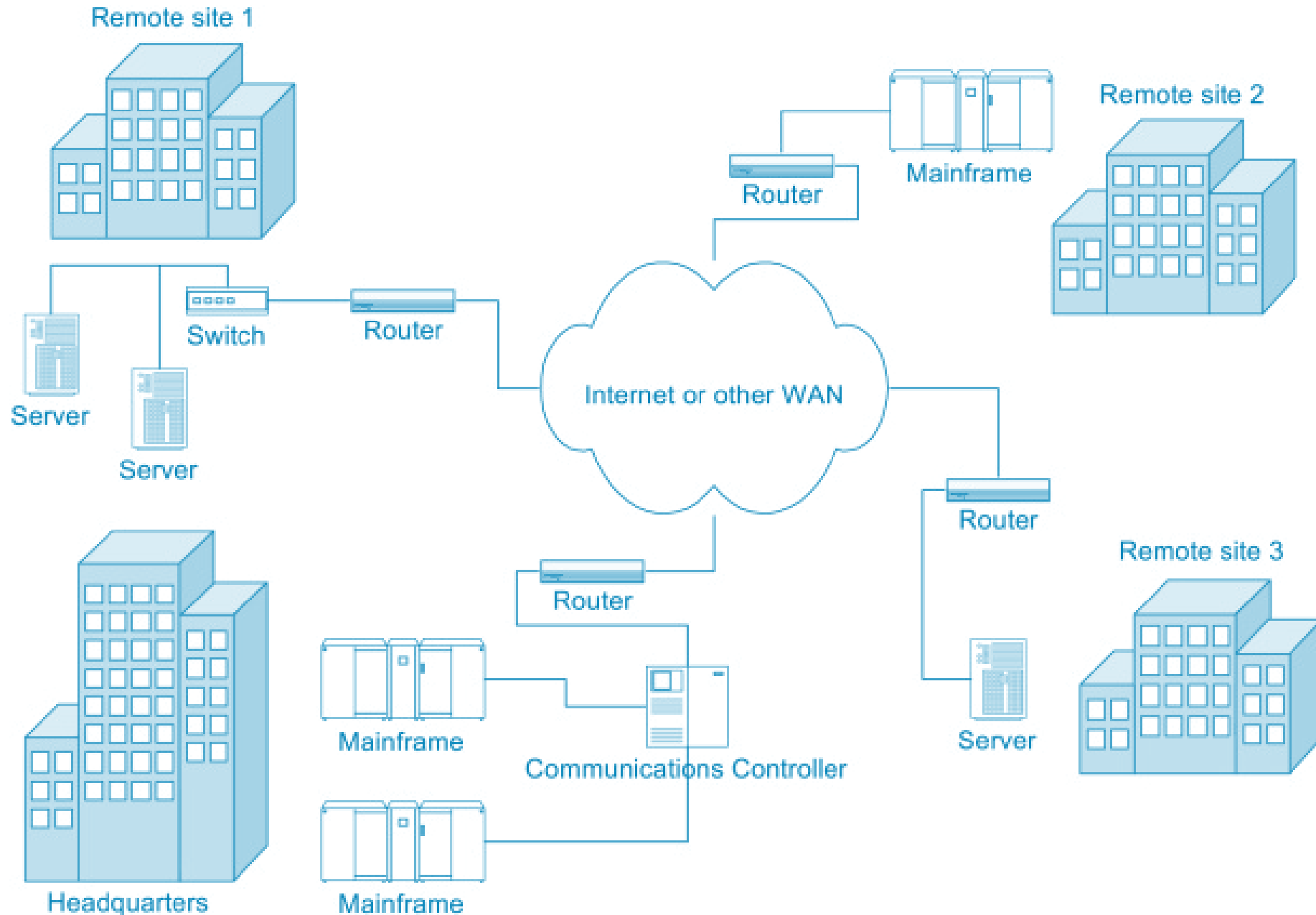


A distributed database is a database that is not limited to one computer system. It is like a database that consists of two or more files located in different computers or sites either on the same network or on an entirely different network.

INTRODUCTION TO DDBMS

- **Distributed Database Management System (DDBMS) (decentralized)** allows users to access not only the data at their own site but also data stored at **remote sites**.
- This decentralized approach **mirrors the organizational structure** of many companies, which are logically distributed into divisions, departments, projects, and so on, and physically distributed into offices, plants, factories, **where each unit maintains its own operational data**.
- The **shareability** of the data and **the efficiency of data access** should be improved by the development of a distributed database system that reflects this organizational structure, **makes the data in all units accessible, and stores data proximate to the location where it is most frequently used**.

INTRODUCTION TO DDBMS



GENERAL VIEW OF DDBMS

- DDBMS consists of a single logical database that is split into a number of **fragments**. Each fragment is stored on one or more computers under the control of a separate DBMS, with:
 - The computers connected by a communications network.
 - Each site is capable of **independently processing** user requests that require access to local data (that is, each site has some degree of local autonomy)
 - Also capable of processing data stored on other computers in the network.
 - Users access the distributed database via applications, which are classified as
 - Those that do not require data from other sites (**local applications**)
 - Those that do require data from other sites (**global applications**).
 - We require a DDBMS to have at least one global application.

GENERAL VIEW OF DDBMS

Vertical Fragmentation

id	Name	Sex	Age
1	Sebastian	M	3
2	John	M	5
3	Garry	M	8
4	Louis	F	7

id	Name
1	Sebastian
2	John
3	Garry
4	Louis



id	Name	Sex	Age
1	Sebastian	M	3
2	John	M	5
3	Garry	M	8
4	Louis	F	7

id	Sex
1	M
2	M
3	M
4	F

Horizontal Fragmentation

id	Name	Sex	Age
----	------	-----	-----

1	Sebastian	M	3
---	-----------	---	---

2	John	M	5
---	------	---	---

3	Garry	M	8
---	-------	---	---

4	Louis	F	7
---	-------	---	---



id	Name	Sex	Age
3	Garry	M	8
4	Louis	F	7

id	Name	Sex	Age
3	Garry	M	8

id	Name	Sex	Age
1	Sebastian	M	3
2	John	M	5

TYPES OF DDBMS

- Different types of distributed database are categorized based on **how data is distributed across multiple nodes**.

TYPES OF DDBMS

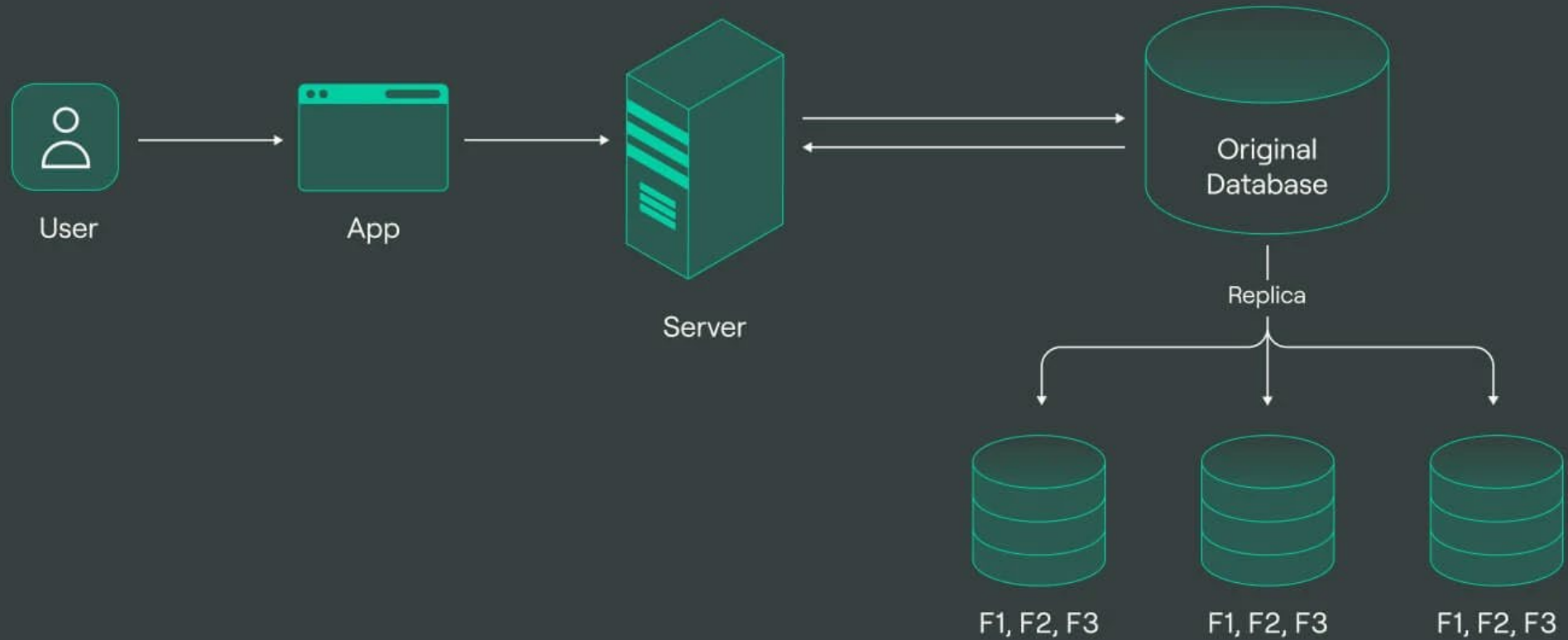


- Replicated databases
- data is replicated across multiple nodes.
- So that each node has a copy of the data.
- The data can be replicated across all nodes or a subset of nodes.
- Replication can improve data availability and reduce latency.
- However, it can also increase data inconsistency and storage overhead.
- Commons are Hadoop, Apache Spark, Cassandra.



TYPES OF DDBMS

Full replication in DBMS



TYPES OF DDBMS



Amazon
DynamoDB



mongoDB

- Partitioned databases
- data is divided into partitions.
- Each partition is stored on a separate node (may be a server).
- Partitioning can improve query performance and scalability.
- But, it can also increase **data inconsistency** and **complexity**.
- Commons are **MongoDB**, **Apache HBase**, **Amazon DynamoDB**.

TYPES OF DDBBMS



Data Node



A big database is split into smaller independent partitions for better management.



Partition 1



Partition 2



Partition 3

Horizontal Partitions

Invoice

customer_id	Invoice_id	creation_date	...
1	5101	2016-01-01	
2	5201	2017-10-10	
2	5301	2018-01-01	
3	5102	2017-01-30	
3	5202	2017-10-30	
4	5302	2017-07-30	

Partition key

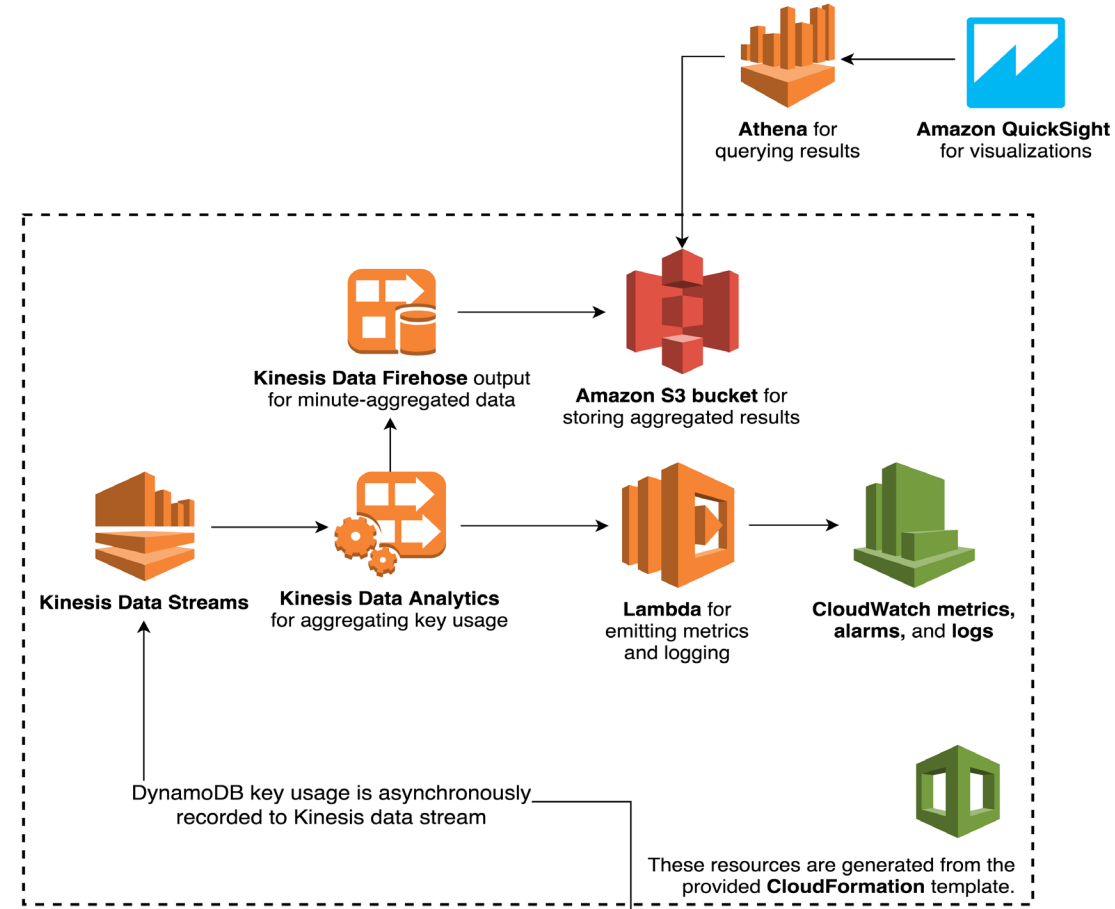
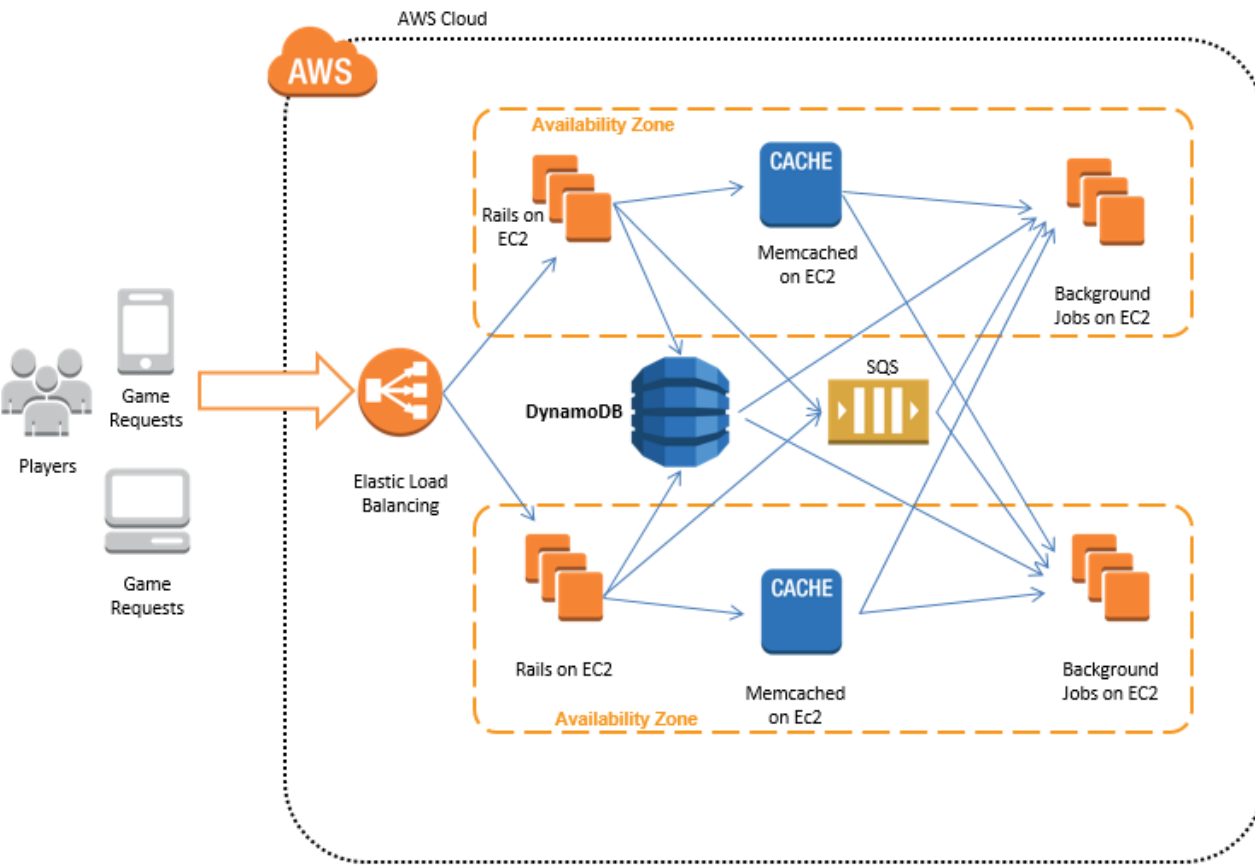
Sharding

Database Shard 1			
Invoice			
customer_id	Invoice_id	creation_date	...
1	5101	2016-01-01	
2	5201	2017-10-10	
2	5301	2018-01-01	

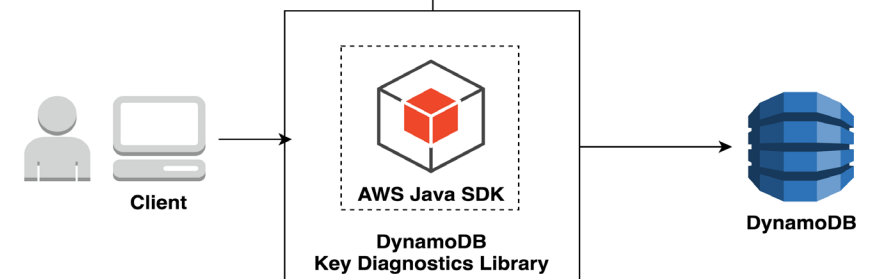
Database Shard 2			
Invoice			
customer_id	Invoice_id	creation_date	...
3	5102	2017-01-30	
3	5202	2017-10-30	
4	5302	2017-07-30	

TYPES OF DDBMS

DynamoDB Key Diagnostics Library architectural diagram



A Terrifying Students diagrams
رسومات لأثارة الرعب لدى الطالب



VOLTDDB

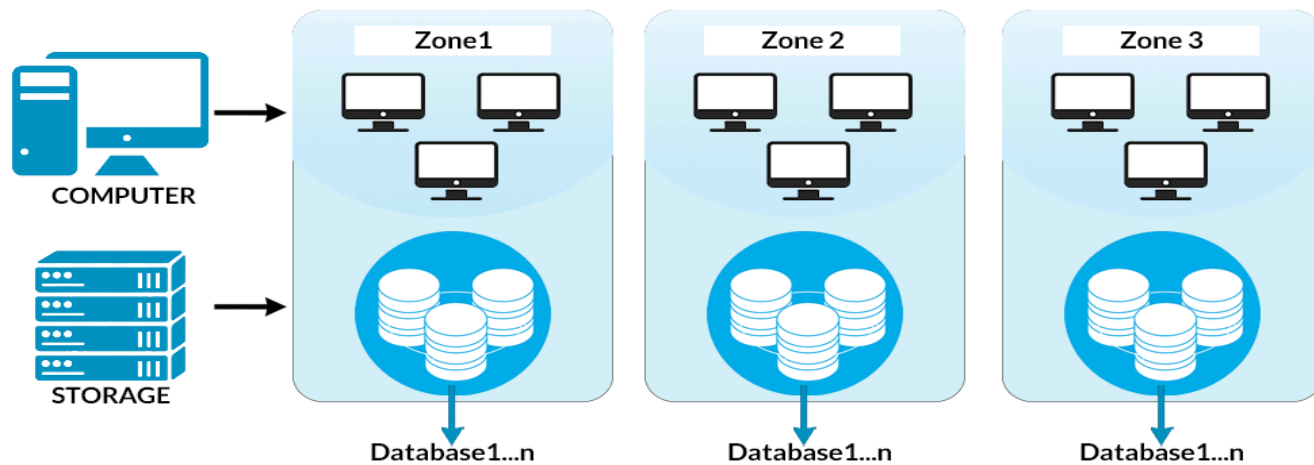
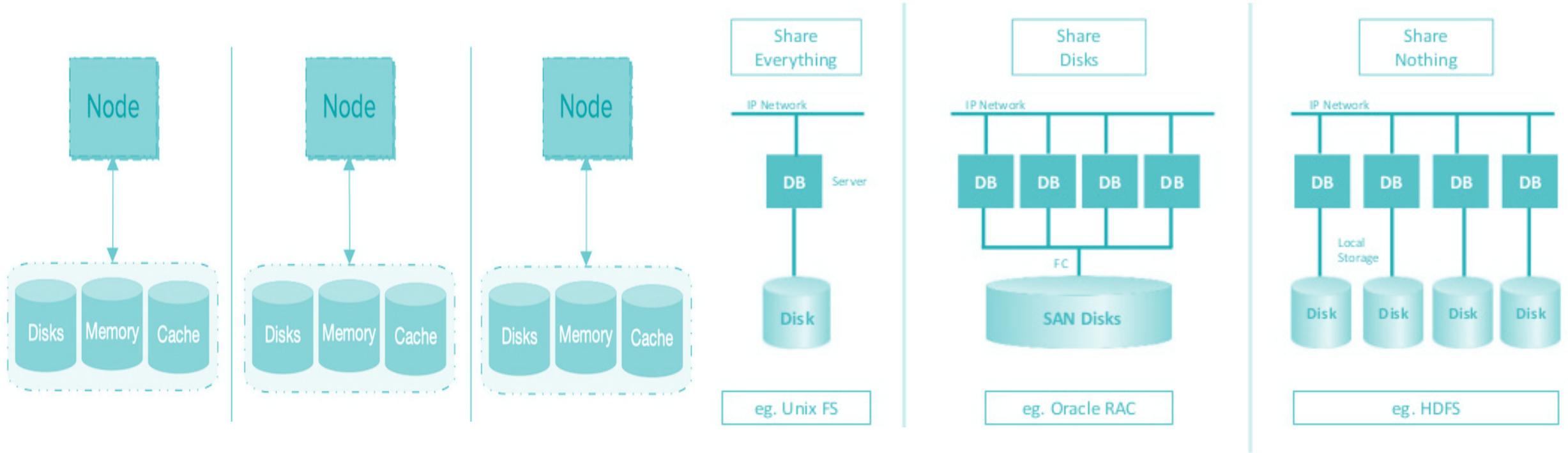
Google
Cloud
Spanner



teradata.

- Shared-nothing databases:
- Each node has its own processor, memory, and disk storage, and there is no shared memory or disk between nodes.
- Shared-nothing architectures can provide high scalability and availability.
- But they can also increase complexity and cost.
- Commons are Google Spanner, Teradata, VoltDB

TYPES OF DDBMS





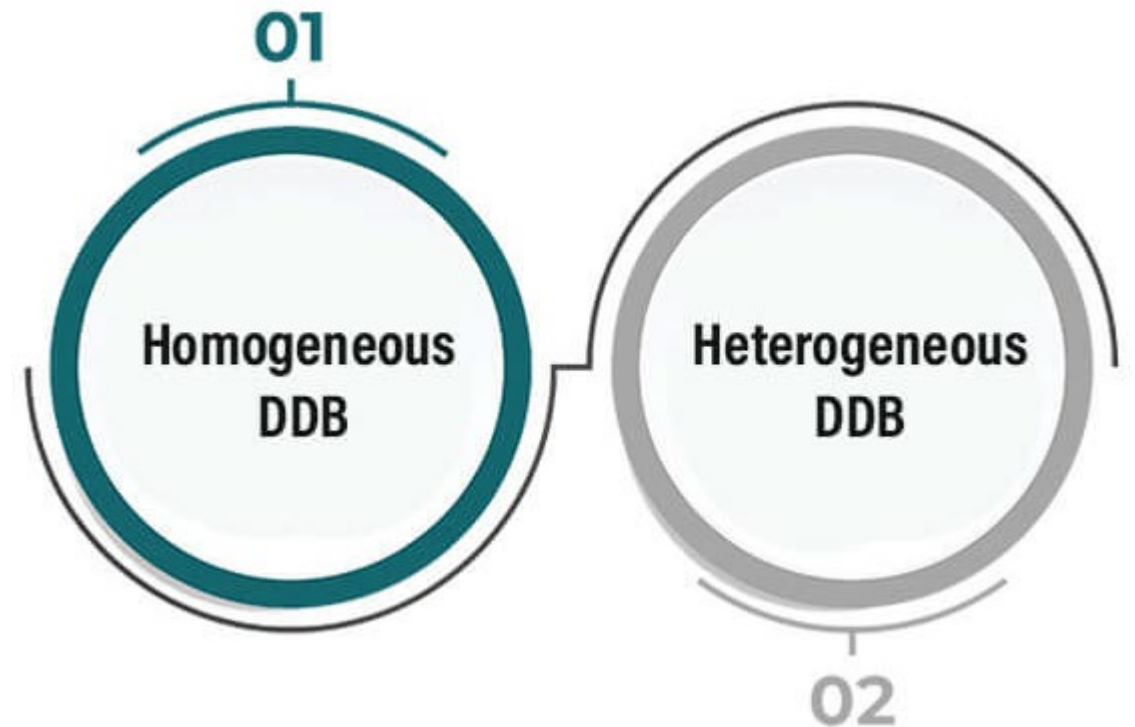
- Hybrid databases
- Combines two or more of the above architectures, such as partitioning and replication.
- Hybrid databases can provide a **balance** between **performance**, **scalability**, and **consistency**,
- but they can also increase complexity and cost.

TYPES OF DDBMS

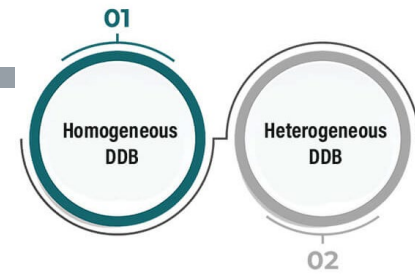
- Different types of distributed database are categorized based on
- The internal data structure
- data storage mechanism of different nodes

TYPES OF DDBMS

- Different types of distributed database are categorized based on
- The internal data structure
- data storage mechanism of different nodes

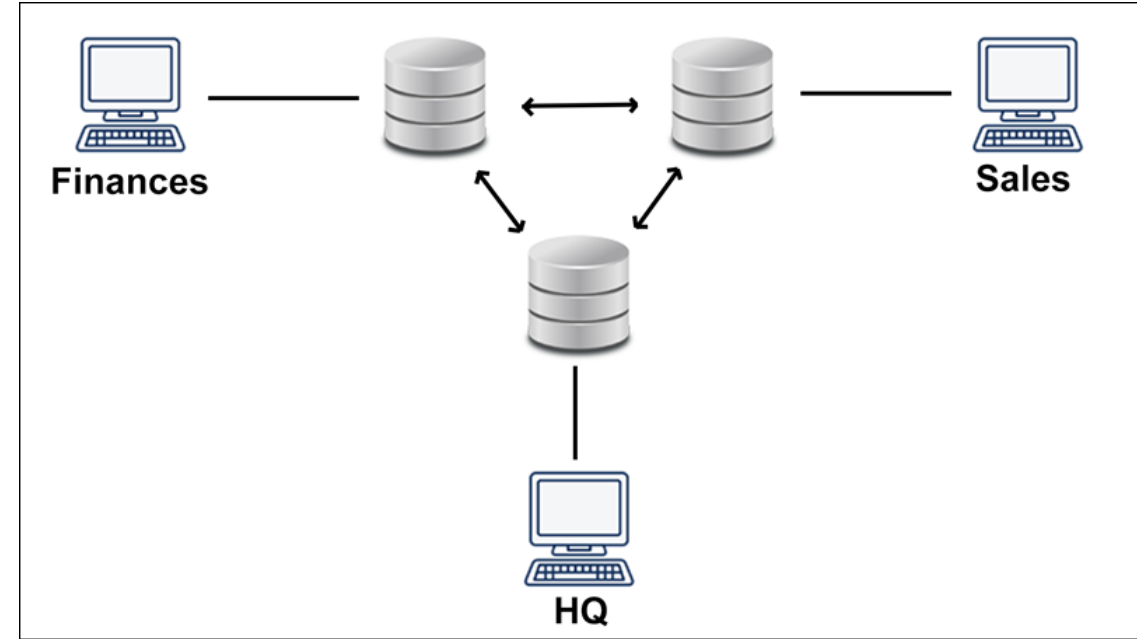
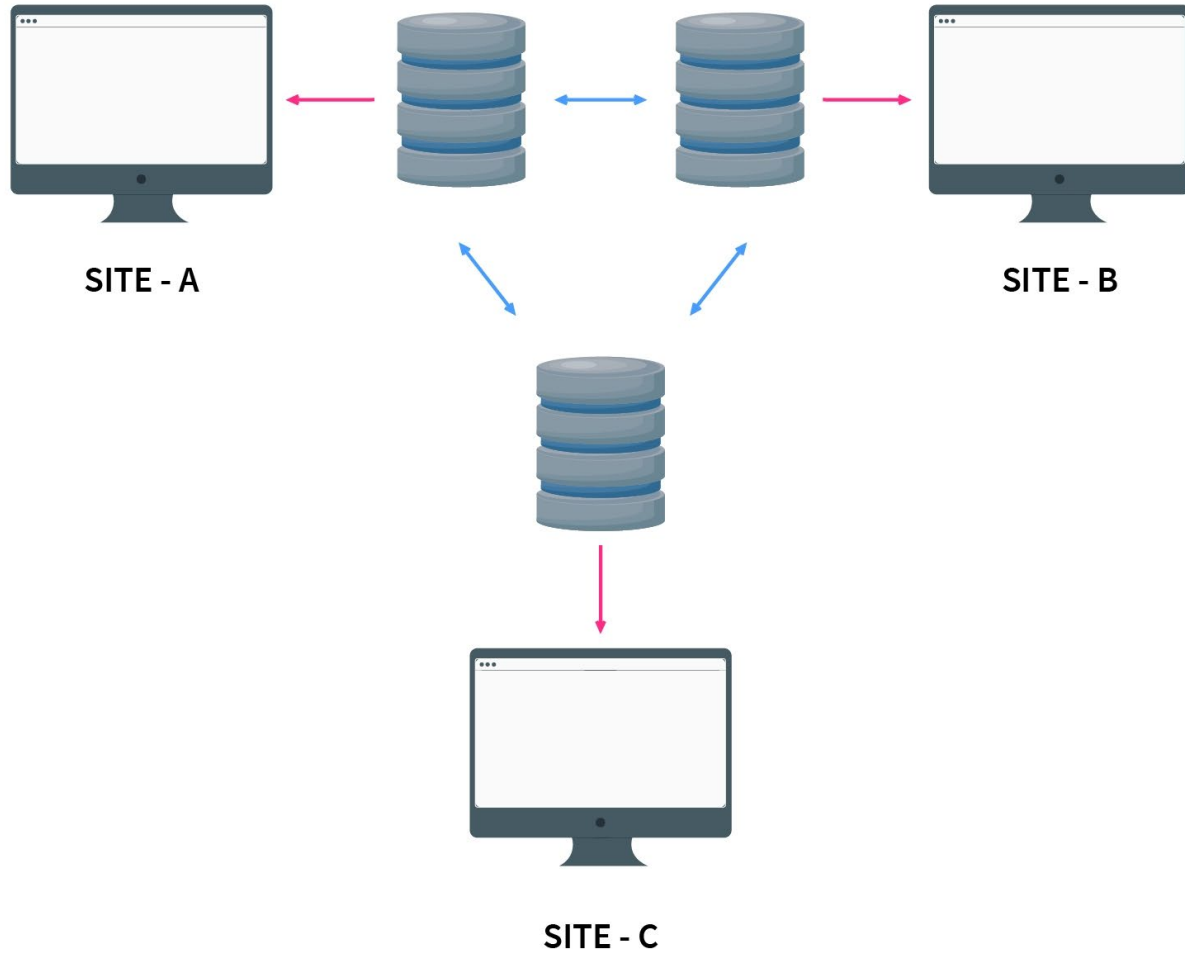
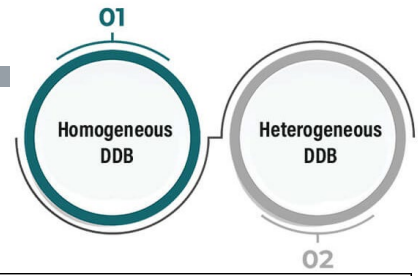


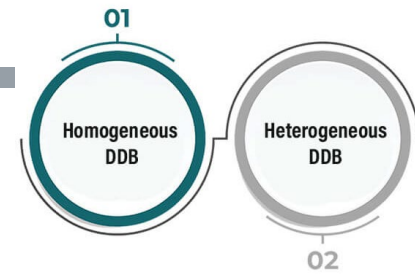
TYPES OF DDBMS



- A Homogenous distributed database is a network of identical databases stored on multiple sites.
- All databases stores data identically, the operating system, DDBMS and the data structures used – all are same at all sites, making them easy to manage.
- Same software and have the same data structure
- Homogeneous distributed databases are easier to manage and provide better consistency, but they can be less flexible and scalable.

TYPES OF DDBMS





- **Heterogeneous Distributed Database**
- It is the opposite of a Homogenous distributed database.
- It uses different schemas, operating systems, DDBMS, and different data models causing it difficult to manage.
- In the case of a Heterogeneous distributed database, a particular site can be completely unaware of other sites.
- This causes limited cooperation in processing user requests, this is why translations are required to establish communication between sites.

TYPES OF DDBMS

