Distributed Database

Presented by Dr . Yasmine Medjadba
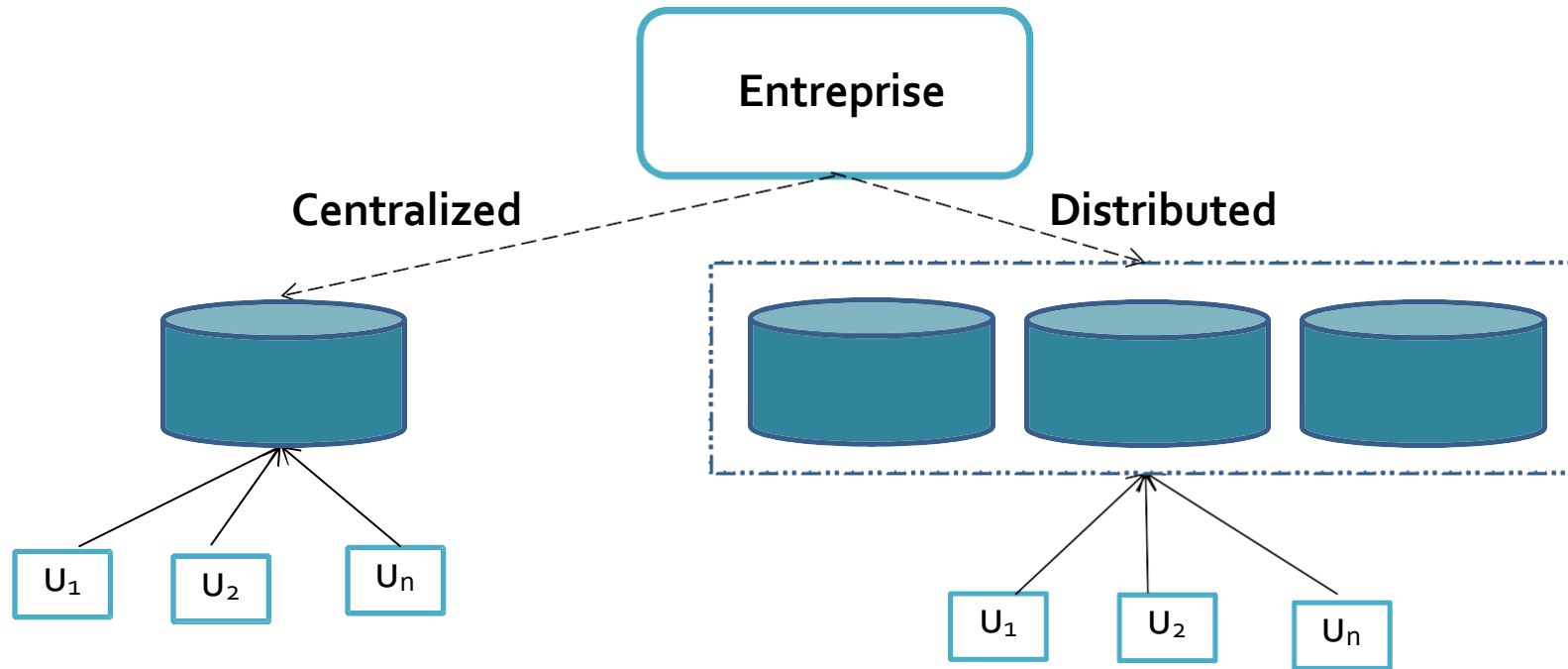
Outline

Chapter I: Introduction.

❖ Centralized or distributed database.

❖ What is a Distributed Database.

❖ Types of Distributed Databases.

❖ Distributed Data Storage.

❖ Advantages and Disadvantages of Distributed database.

Introduction

Centralized or distributed database?

❖ A company, depending on its needs, can choose between using a centralized or distributed database.

Introduction

Centralized or distributed database?

centralized database enviroment.

Centralized database problem

Centralized or distributed database?

❖ Increase in data volume.

❖ Increase in the volume of treatments.

❖ Increase in transaction volume

**Multiply data sources, and make them communicate through a network**

distributed database

Distributed Database

What is a Distributed Database?

❖ A **distributed database** is a collection of multiple, logically interrelated databases distributed over a computer network.

❖ A distributed database management system (**distributed DBMS**) is then defined as the software system that permits the management of the distributed database and makes the distribution transparent to the users.

❖ Sometimes "**distributed database system**" (**DDBS**) is used to refer jointly to the distributed database and the distributed DBMS.
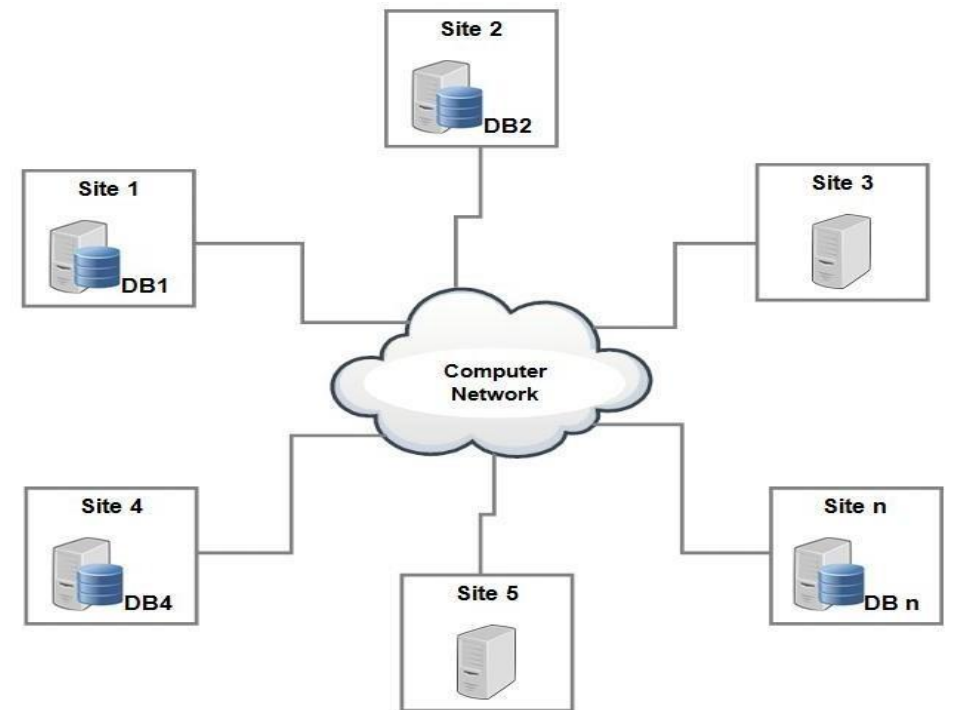
Distributed Database

Types of Distributed Databases:

1. **Homogeneous Database**

   ❖ In a homogeneous database, all different sites store database identically (the operating system, database management system, and the data structures used),

   ❖ All are the same at all sites. Therefore, they're easy to manage.

2. **Heterogeneous Database:**

   ❖ In a heterogeneous distributed database, different sites can use different schema and software (**operating system, different database application. They may even use different data models for the database.**)

   ❖ It can lead to problems in query processing and transactions.

   ❖ In addition, a particular site might be completely unaware of the other sites. Hence, translations are required for different sites to communicate.

Distributed Database

Distributed Data Storage

There are 2 ways in which data can be stored on different sites. These are:

1. **Replication**

    ❖ In this approach, the entire relationship is stored redundantly at 2 or more sites.

    ❖ If the entire database is available at all sites, it is a fully redundant database.

    ❖ In replication, systems maintain copies of data.

    ❖ However, Data needs to be constantly updated.

Distributed Database

Distributed Data Storage

2. **Fragmentation**

    In this approach, the relations are fragmented (i.e., they're divided into smaller parts) and each of the fragments is stored in different sites where they're required. It must be made sure that the fragments are such that they can be used to reconstruct the original relation (i.e, there isn't any loss of data).

Distributed Database

Advantages of Distributed database:

Many advantages of DDBSs have been cited in literature, ranging from sociological reasons for decentralization to better economics. In addition, it can be summarised into four fundamentals which may also be seen as promises of DDBS technology:

1. Management of database with different level of transparency

   Transparency refers to separation of the higher-level semantics of a system from lower-level implementation issues. In other words, a transparent system "hides" the implementation details from users.

   ❖ **Replication transparencies :**

   It basically made user unaware of the existence of copies as we know that copies of data may be stored at multiple sites for better availability performance and reliability.

   ❖ **Fragmentation transparency:**

   It basically made user unaware about the existence of fragments it may be the vertical fragment or horizontal fragmentation.

Distributed Database

❖ **Data Independence**

data definition occurs at two levels. At one level the logical structure of the data are specified, and at the other level its physical structure.

  a. **Logical data independence**

   Refers to the immunity of user applications to changes in the logical structure (i.e., schema) of the database.

  b. **Physical data independence**

   Deals with hiding the details of the storage structure from user applications

Distributed Database

❖ **Network Transparency**

    In centralized database systems, the only available resource that needs to be shielded from the user is the data (i.e., the storage system). In a distributed database environment, however, there is a second resource that needs to be managed in much the same manner: **the network**.

    Preferably, the user should be protected from the operational details of the network; possibly even hiding the existence of the network.

    From a DBMS perspective, distribution transparency requires that users do not have to specify where data are located

Distributed Database

2. This database is more secure in comparison to a centralized database.

3. Increased Reliability and availability.

4. Improved Performance.

Distributed Database

Disadvantages

❖ This database is very costly and difficult to maintain because of its complexity.

❖ It is difficult to provide a uniform view to users since it is spread across different

physical locations.

Distributed Database

Exemple:

Consider an engineering company that has offices in Algiers, Batna and Oran. They run projects at each of these sites and would like to maintain a database of their employees, the projects and other related data.

The information can be stored as follow:

EMP(EId, Ename, title)

PROJ(PId, Pname, Budget, City).

SAL(title, salary) // stores salary information

ASG(EId, PId, Resp, Dur)   // indicates which employees have been assigned to which projects for what duration with what responsibility

Distributed Database

EMP

EId	Ename	title
E1	Sm	accounting
E2	Md	Ing.
E3	Hm	Elec.Ing
E4	Ad	Programmer

PROJ

PId	Pname	Budget	City
P1	Instrumentati-on	15000	Alger
P2	DB develop.	135000	Batna
P3	Construction	250000	Batna
P4	Entretien	310000	Oran

SAL

Title	Salary
accounting	40000
Ing.	34000
Elec.Ing	27000
Programmer	24000

ASG

EId	PId	Resp	Dur
E1	P1	Director	12
E2	P1	Analyst	24
E2	P2	Analyst	6
E3	P3	Consultant	10
E3	P4	Engineer	48
E4	P2	Programmer	18

Distributed Database

Exemple:

If all of this data were stored in a centralized DBMS, and we wanted to find out the **names** of **employees and their salaries** who worked on a project for **more than 12 months**, we would specify this using the following SQL query:

SELECT Ename, salary

FROM EMP, ASG, SAL

WHERE ASG.DUR > 12

AND EMP.EId = ASG.EId

AND SAL.title = EMP.title

Distributed Database

❖ In the case where the database is distributed, here the user asks the same question to the system without paying attention to the distribution of data (transparency),

❖ It is the system that manages to collect essential information.

Distributed Database

❖ Now, given the distributed nature of this company, it is preferable, to localize data such that data, project and salary information in Algeries office are stored in Algeries, those in the Batna office are stored in Batna, and so forth.

❖ Thus, this can be defined as a process where we partition each of the relations and store each partition at a different site. This is known as **fragmentation.**

Distributed Database

References

1. .	M. Tamer Özsu and Patrick Valduriez, "Principles of Distributed Database Systems", Third Edition, Springer, 2011

2. .	IBM, "Distributed database programming", Version 6 Release 1, 2008.

3. Georges Gardarin, « Bases de données réparties », Cours BDD Web.

THANK YOU