

Distributes Databases

Introduction

A distributed computing system consists of a number of processing elements, not necessarily homogeneous, that are interconnected by a computer network, and that cooperate in performing certain assigned tasks.

As a general goal, distributed computing systems partition a big, unmanageable problem into smaller pieces and solve it efficiently in a coordinated manner.

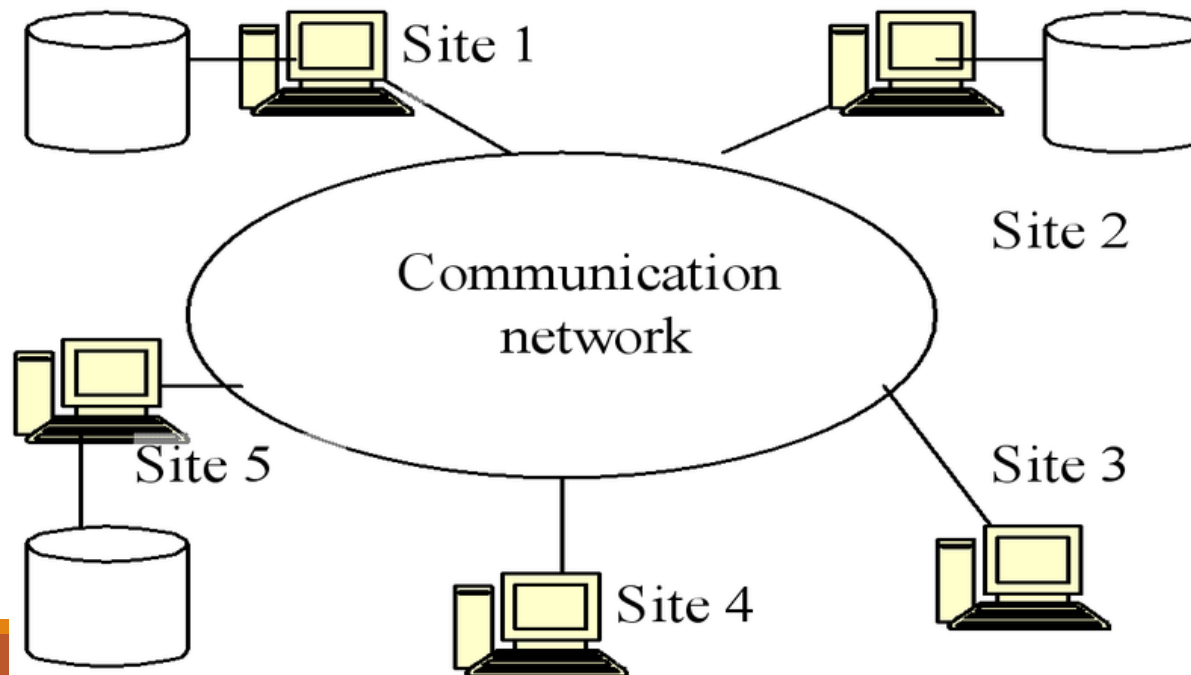
The economic viability of this approach stems from two reasons:

- more computing power is harnessed to solve a complex task,
- each autonomous processing element can be managed independently to develop its own applications.

Introduction

Distributed database (DDB) can be defined as a collection of multiple logically interrelated databases distributed over a computer network.

Distributed database management system (DDBMS) can be defined as a software system that manages a distributed database while making the distribution transparent to the user.



Introduction

Distributed database (DDB) can be defined as a collection of multiple logically interrelated databases distributed over a computer network.

Advantages of Distributed Databases:

1. Improved ease and flexibility of application development:

- Developing and maintaining applications at geographically distributed sites of an organization is facilitated owing to transparency of data distribution and control.

2. Increased reliability and availability:

This is achieved by the isolation of faults to their site of origin without affecting the other databases connected to the network.

When the data and DDBMS software are distributed over several sites, one site may fail while other sites continue to operate.

Introduction

Only the data and software that exist at the failed site cannot be accessed.

This improves both reliability and availability.

Further improvement is achieved by judiciously replicating data and software at more than one site.

In a centralized system, failure at a single site makes the whole system unavailable to all users.

In a distributed database, some of the data may be unreachable, but users may still be able to access other parts of the database.

If the data in the failed site had been replicated at another site prior to the failure, then the user will not be affected at all.

Introduction

3. Improved performance:

A distributed DBMS fragments the database by keeping the data closer to where it is needed most.

Data localization reduces the contention for CPU and I/O services and simultaneously reduces access delays involved in wide area networks.

When a large database is distributed over multiple sites, smaller databases exist at each site.

As a result, local queries and transactions accessing data at a single site have better performance because of the smaller local database.

Each site has a smaller number of transactions executing than if all transactions are submitted to a single centralized database.

Introduction

4. Easier expansion.

In a distributed environment, expansion of the system in terms of adding more data, increasing database sizes, or adding more processors is much easier.

Types of Distributed Database System.

The term distributed database management system can describe various systems that differ from one another in many respects.

The main thing that all such systems have in common is the fact that data and software are distributed over multiple sites connected by some form of communication network.

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.

Hence, they're easy to manage.

Types of Distributed Database System.

2. Heterogeneous Database:

In a heterogeneous distributed database, different sites can use different schema and software that can lead to problems in query processing and transactions.

A particular site might be completely unaware of the other sites.

Different computers may use a different operating system, different database application.

They may even use different data models for the database.

Hence, translations are required for different sites to communicate.

Distributed Database Architectures

A Distributed Database System is a kind of database that is present or divided in more than one location, which means it is not limited to any single computer system.

It is divided over the network of various systems.

The Distributed Database System is physically present on the different systems in different locations.

This can be necessary when different users from all over the world need to access a specific database.

For a user, it should be handled in such a way that it seems like a single database.

Distributed Database Architectures

Different architectures are:

1. Shared nothing architecture.
2. A networked architecture with a centralized database at one of the sites.
3. A truly distributed database architecture.

Distributed Database Architectures

1. Shared nothing architecture.

In this architecture, every processor has its own primary and secondary (disk) memory.

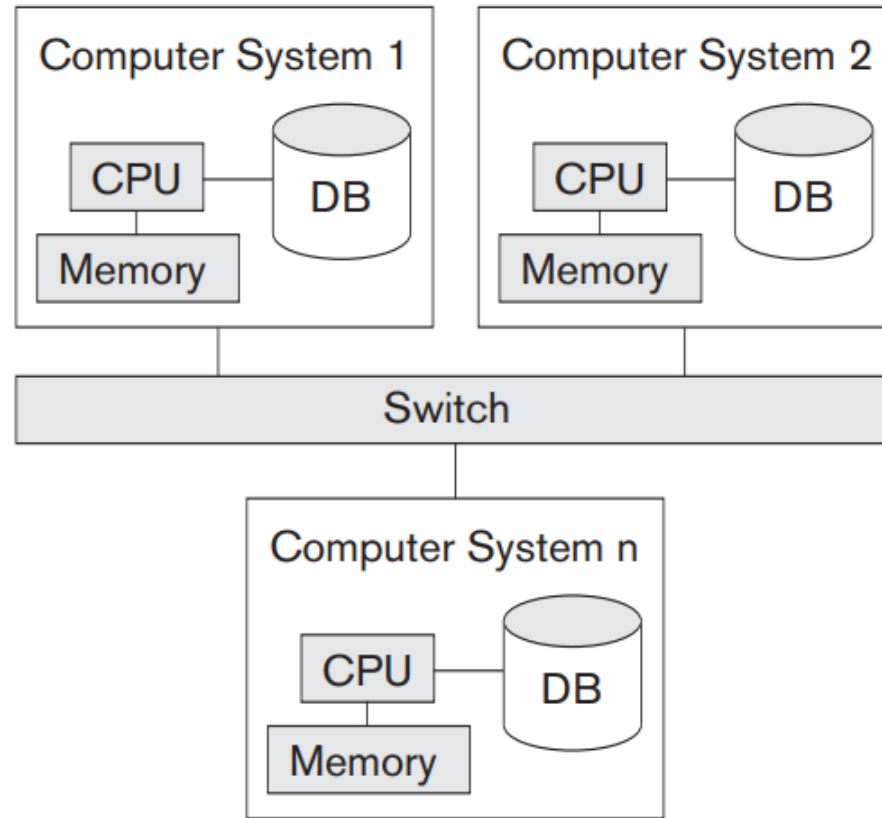
No common memory exists, and the processors communicate over a high speed interconnection network (bus or switch).

Although the shared nothing architecture resembles a distributed database computing environment, major differences exist in the mode of operation.

In shared nothing multiprocessor systems, there is symmetry and homogeneity of nodes; this is not true of the distributed database environment where heterogeneity of hardware and operating system at each node is very common.

Shared nothing architecture is also considered as an environment for parallel databases.

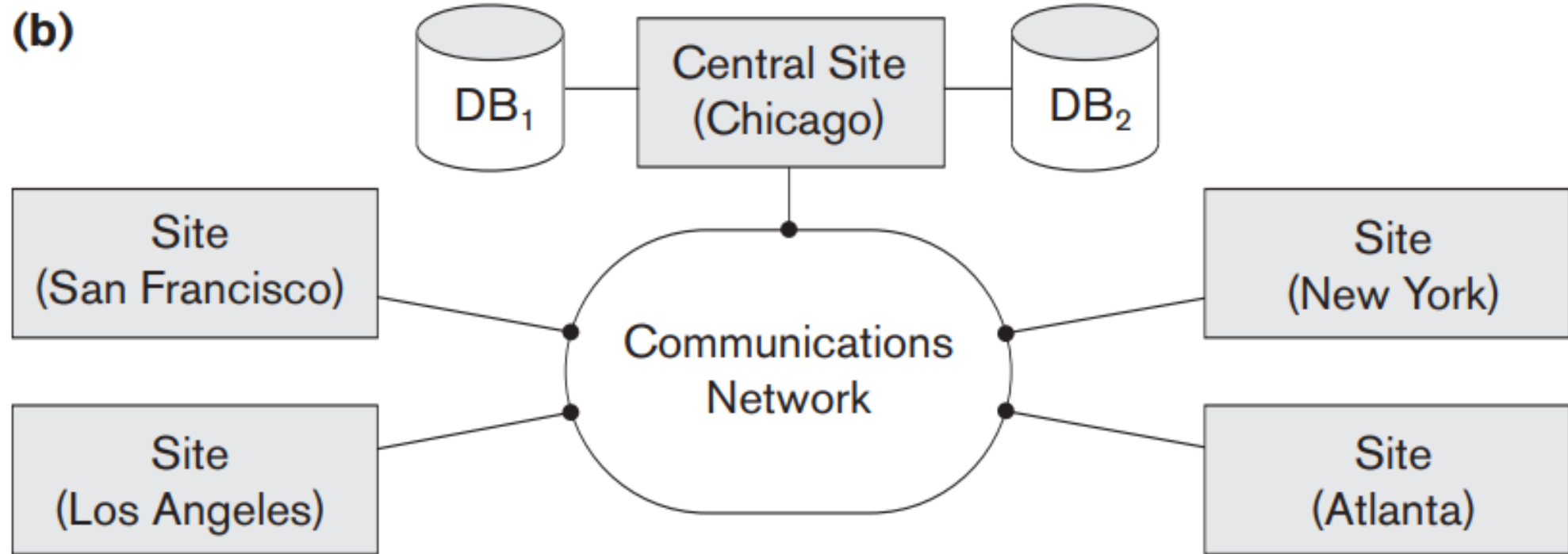
Distributed Database Architectures



Shared nothing architecture

Distributed Database Architectures

2. A networked architecture with a centralized database at one of the sites.



Distributed Database Architectures

1. Shared nothing architecture.

In this architecture, every processor has its own primary and secondary (disk) memory.

No common memory exists, and the processors communicate over a high speed interconnection network (bus or switch).

Although the shared nothing architecture resembles a distributed database computing environment, major differences exist in the mode of operation.

In shared nothing multiprocessor systems, there is symmetry and homogeneity of nodes; this is not true of the distributed database environment where heterogeneity of hardware and operating system at each node is very common.

Shared nothing architecture is also considered as an environment for parallel databases.

Distributed Database Architectures

3. A truly distributed database architecture.

