Chapter 1

INTRODUCTION TO DISTRIBUTED DATABASE SYSTEMS

1. Why a Distributed Database?

Databases come in many flavors. We use databases in our daily activities such as airline reservations, banking transactions via an ATM machine, searching information in the Internet by using a search engine, borrowing books from a library, etc. Have you ever thought about where and how the data are accessed and managed?

Database systems evolved as the need for shared data keep increasing. A very good example to realize the truth behind this statement is a Video-on-Demand (VOD) system to which several users subscribe to access movies on a regular basis. In this case, the database is a movie database comprising movies stored in a particular style. In particular, the rapid proliferation in computer hardware technology coupled with the underlying communications technology made a grand success for the advent of database systems. The service offering capabilities were further enhanced by the use of modern day computer architectures such as SISD (single instruction stream over a single data stream) to MIMD (multiple instruction streams over multiple data streams)[KH93], together with the use of sophisticated operating systems exclusively developed for architectures with multiple CPUs\(^1\)[PJF97]. As the size of corporate offices, government organizations, educational sectors, etc, grow with time, need for sharing the available computer and communication facilities becomes a clever solution without incurring additional investment overheads. In turn, this demands for devising efficient technologies and methods to disseminate the required data to the users at the required times. Con-

\(^1\)These are also referred to as Multiprocessor architectures.

sequently, data organization, management, delivery, storage, etc, form imperative components in rendering services to several application domains mentioned above. In order to simplify the control mechanism in such sophisticated systems, the computer and communication facilities were traditionally biased towards a *centralized* style of operation. In a centralized database system, all the data are collected into a single database, as shown in Figure 1.1, where the *Database Management System* (DBMS) is essentially a software component that provides services for accessing the data stored in the database.

![Figure 1.1. A centralized database system architecture](image-url)

Obviously, the use of centralized database systems makes sense if the user domain is somewhat smaller in size and possibly confined to a smaller geographical area. However, industrial organizations, educational bodies with multi-campuses, transactions between several banks, etc, require a decentralized way of operational style due to geographic separation. Thus, exercising a single point control to coordinate and store all the required data for such systems will be highly inefficient. For instance, users may undergo long waiting times to access the centralized database. Managing and using a single large size database for such diversified applications are prohibitively complex and long communication time between the database and the users, etc, are some of the disadvantages we can immediately comprehend. Essentially, the motivation to