IT 775
Database Technology

DataStores

Distributed and Parallel Databases

DISTRIBUTED DATABASES

Distributed database system (DDBS)

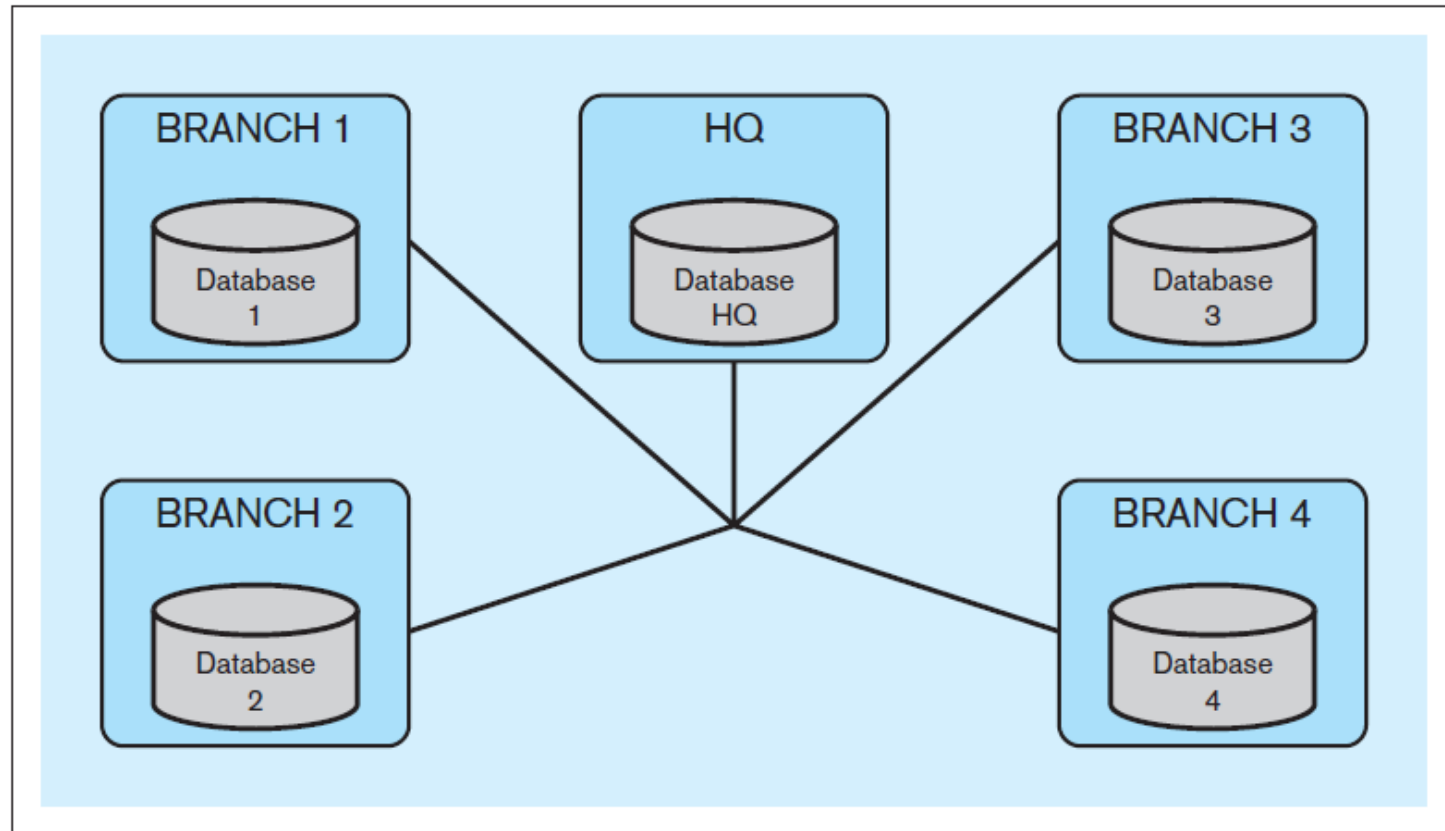Database system in which the database is distributed among separate computers

DISTRIBUTED DATABASES

Distribution transparency

Ensures that the end users can use the distributed database in the same fashion as if the database is not distributed

DISTRIBUTED DATABASES

Distributed database system (DDBS) - example

DISTRIBUTED DATABASES

Homogeneous DDBS,

All the computers run the same DBMS

Heterogeneous DDBS

Different DBMS can run on different computers

DISTRIBUTED DATABASES

Database fragmentation

Strategy for distributing the data across different locations in a distributed database

- **Horizontal fragmentation**
- **Vertical fragmentation**
- **Mixed fragmentation**

DISTRIBUTED DATABASES

Horizontal fragmentation

Subsets of the records from a table stored at different locations

Horizontal fragmentation - Example

EMPLOYEE (Not Fragmented)

EmpID	EmpName	EmpGender	EmpPhone	EmpBdate
0001	Joe	M	x234	1/11/1985
0002	Sue	F	x345	2/7/1983
0003	Amy	F	x456	8/4/1990
0004	Pat	F	x567	3/8/1971
0005	Mike	M	x678	5/5/1965

EMPLOYEE (Horizontally Fragmented - Location A)

EmpID	EmpName	EmpGender	EmpPhone	EmpBdate
0001	Joe	M	x234	1/11/1985
0002	Sue	F	x345	2/7/1983
0003	Amy	F	x456	8/4/1990

EMPLOYEE (Horizontally Fragmented - Location B)

EmpID	EmpName	EmpGender	EmpPhone	EmpBdate
0001	Joe	M	x234	1/11/1985
0004	Pat	F	x567	3/8/1971
0005	Mike	M	x678	5/5/1965

Horizontal fragmentation - Example

Query:

```
SELECT * FROM employee_location_a
UNION
SELECT * FROM employee_location_b;
```

DDBS transparency allows this query to be issued as:

```
SELECT * FROM employee
```

DISTRIBUTED DATABASES

Vertical fragmentation

Subsets of the columns of a table are stored at different

Vertical fragmentation - Example

EMPLOYEE (Not Fragmented)

EmpID	EmpName	EmpGender	EmpPhone	EmpBdate
0001	Joe	M	x234	1/11/1985
0002	Sue	F	x345	2/7/1983
0003	Amy	F	x456	8/4/1990
0004	Pat	F	x567	3/8/1971
0005	Mike	M	x678	5/5/1965

EMPLOYEE (Vertically Fragmented - Location A)

EmpID	EmpName	EmpPhone
0001	Joe	x234
0002	Sue	x345
0003	Amy	x456
0004	Pat	x567
0005	Mike	x678

EMPLOYEE (Vertically Fragmented - Location B)

EmpID	EmpGender	EmpBdate
0001	M	1/11/1985
0002	F	2/7/1983
0003	F	8/4/1990
0004	F	3/8/1971
0005	M	5/5/1965

Vertical fragmentation - Example

Query:

```
SELECT    a.empid, a.empname, b.empgender, a.empphone, b.empbdate
FROM      employee_location_a a, employee_location_b b
WHERE     a.empid = b.empid;
```

DDBS transparency allows this query to be issued as:

```
SELECT * FROM employee
```

DISTRIBUTED DATABASES

Mixed fragmentation

A combination of horizontal and vertical fragmentation

Mixed fragmentation - Example

EMPLOYEE (Not Fragmented)

EmpID	EmpName	EmpGender	EmpPhone	EmpBdate
0001	Joe	M	x234	1/11/1985
0002	Sue	F	x345	2/7/1983
0003	Amy	F	x456	8/4/1990
0004	Pat	F	x567	3/8/1971
0005	Mike	M	x678	5/5/1965

EMPLOYEE (Location A)

EmpID	EmpName	EmpPhone
0001	Joe	x234
0002	Sue	x345
0003	Amy	x456

EMPLOYEE (Location B)

EmpID	EmpName	EmpPhone
0001	Joe	x234
0004	Pat	x567
0005	Mike	x678

EMPLOYEE (Location C)

EmpID	EmpGender	EmpBdate
0001	M	1/11/1985
0002	F	2/7/1983
0003	F	8/4/1990
0004	F	3/8/1971
0005	M	5/5/1965

Mixed fragmentation - Example

Query:

```
SELECT    a.empid, a.empname, c.empgender, a.empphone, c.empbdate
FROM      employee_location_a a, employee_location_c c
WHERE     a.empid = c.empid
UNION
SELECT    b.empid, b.empname, c.empgender, b.empphone, c.empbdate
FROM      employee_location_b b, employee_location_c c
WHERE     b.empid = c.empid;
```

DDBS transparency allows this query to be issued as:

```
SELECT * FROM employee
```

DISTRIBUTED DATABASES

Data replication

Storing more than one copy of the data at different locations in a distributed database

Fully replicated distributed database

- The entire database replicated at each location in the distributed system

Partially replicated distributed database

- Some parts of database replicated at multiple locations
- Other parts are not replicated

DISTRIBUTED DATABASES

Federated database

A collection of preexisting databases that are connected into one system

PARALLEL DATABASES

Parallel databases

Uses multiple computers to work on the same task

Multiple computers perform the same operations simultaneously on different portions of the same data set

PARALLEL DATABASES

Massively parallel processing (MPP)

Uses a large number of separate computer processors running in parallel to execute a single program

Shared-nothing MPP architecture

A parallel system in which each processor has its own memory and disk storage