



# Sistem Basis Data Terdistribusi

**Arif Basofi**

Sumber: *Fundamentals of Database Systems, Third Edition ch.24*, Elmasri

Sumber Material: [tanzir.staff.gunadarma.ac.id](http://tanzir.staff.gunadarma.ac.id), T. Darmanto & Y. H. Chrisnanto, AmikBandung

# Konsep Basis Data Terdistribusi

- **Sistem Komputasi Terdistribusi** adalah sejumlah elemen proses yang terkoneksi melalui jaringan komputer dan saling bekerjasama dalam melakukan suatu tugas.
- **Basis Data Teridistribusi** adalah kumpulan basis-basis data yang saling berhubungan secara logika dan **tersebar** pada sebuah jaringan komputer.
- **Sistem Manajemen Basis Data** adalah sebuah sistem software yang mengelola basis data terdistribusi.

# Konsep Basis Data Terdistribusi...

- Pada sistem basis data terdistribusi, data di simpan tersebar di beberapa tempat.
- Setiap tempat penyimpanan dikelola oleh suatu DBMS yang mandiri.
- Agar tampilan view basis data terdistribusi transparan, maka harus memenuhi dua hal, yaitu:
  - **independensi** data terdistribusi dan
  - **atomisitas** transaksi terdistribusi.

# Konsep Basis Data Terdistribusi...

- Dengan independensi data terdistribusi, pengguna dapat melakukan query secara sederhana tanpa menyebutkan tempat data atau replika data atau fragmen data itu disimpan.
- Ini memenuhi prinsip independensi data fisik dan data logik atau data logik tidak tergantung data fisik.
- Lebih jauh lagi proses query juga harus sudah memperhitungkan biaya antara penyimpanan data fisik melalui komunikasi data atau disimpan sebagai data lokal (replika).

# Konsep Basis Data Terdistribusi...

- Dengan atomisitas transaksi terdistribusi pengguna harus dapat melakukan transaksi tulis, update atau akses data terhadap data terdistribusi, **seolah-olah data disimpan secara lokal**.
- **Efek transaksi terhadap data terdistribusi harus bersifat atomik**, yaitu:
  - perubahan secara persisten terhadap data remote dan data lokal, jika transaksi telah commit, atau
  - tidak terjadi perubahan sama sekali jika transaksi gagal (tidak dapat commit).
- Walaupun secara umum kedua hal tersebut harus dipenuhi, tetapi pada situasi **jika terjadi lalu lintas data yang padat dan terjadi kelambatan transmisi**, maka diperlukan **mekanisme khusus** untuk menanganinya yang berkaitan dengan **overhead administrasi dan performansi DBMS**.

# Jenis Basis Data Terdistribusi

- Jika DBMS yang menangani data untuk semua server sejenis, maka sistem basis data tersebut disebut sebagai sistem basis data terdistribusi homogen, tetapi
- sebaliknya jika DBMS yang menangani data terdistribusi beragam, maka disebut sistem basis data terdistribusi heterogen atau disebut juga sebagai sistem multi basis data.

# Jenis Basis Data Terdistribusi...

- Kunci keberhasilan membangun sistem yang heterogen adalah pada standarisasi **protokol gateway**.
- Protokol gateway adalah **API** (Application Programming Interface) yang memungkinkan DBMS berfungsi untuk aplikasi eksternal, sebagai contoh seperti : ODBC dan JDBC.
- Akses basis data terdistribusi melalui protokol gateway menjadi mungkin, karena perbedaan **format data** dan perbedaan antar **server** dijabatani.

# Arsitektur Basis Data Terdistribusi

- Terdapat tiga alternatif pendekatan untuk membedakan fungsi DBMS, yaitu: **client-server**, **kolaborasi server** dan **middleware**.
- Sistem **Client-Server** : memiliki satu atau banyak proses pada client dan satu atau banyak proses pada server. Client berurusan dengan antarmuka dengan user dan server mengelola data dan eksekusi transaksi.
- Arsitektur ini populer, karena relatif sederhana untuk diimplementasikan karena ada pemisahan fungsi yang jelas dan server yang tersentralisasi. Biaya yang tinggi hanya untuk satu server, selain pengguna akan lebih nyaman menggunakan antarmuka grafis pada client.
- Untuk memperlancar layanan server dan mengurangi overhead komunikasi, maka diperlukan caching pada server.



# Arsitektur Basis Data Terdistribusi...

## Sistem kolaborasi server:

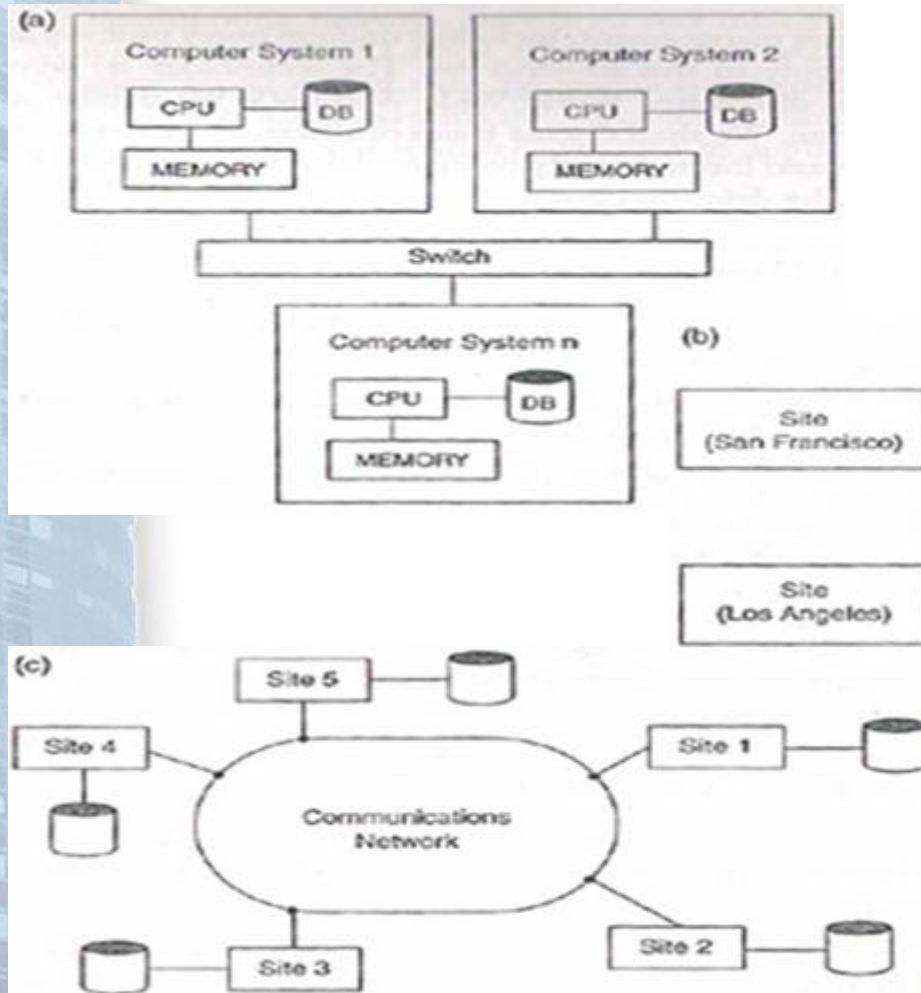
- Pada sistem client server tidak memungkinkan satu query pada client dijalankan oleh multipel server, karena tidak ada **mekanisme kolaborasi** antar server.
- Pada **sistem kolaborasi server** : query didekomposisi menjadi **sub-query** dan **disebar** ke server yang berbeda sesuai fungsinya.
- Secara idealnya dekomposisi sub-query harus memperhatikan **biaya komunikasi jaringan** dan **biaya pengolahan lokal**.

# Arsitektur Basis Data Terdistribusi...

## Sistem middleware:

- Sistem middleware dirancang untuk memungkinkan satu query dijalankan pada multipel server dengan server-server basis data **tidak mengelola eksekusi** pada multi tempat.
- Untuk **koordinasi sub-query dan eksekusi join** dilakukan oleh perangkat lunak tersendiri yang disebut middleware.

# Arsitektur Basis Data Terdistribusi...



- **Gambar a** : Shared nothing Database
- **Gambar b** : Centralized Database
- **Gambar c** : Distributed Database

# Keuntungan Basis Data Terdistribusi

- Manajemen data terdistribusi dengan tingkat transparansi yang berbeda
- Keandalan dan ketersediaan
- Peningkatan performa
- Ekspansi yang lebih mudah

# Keuntungan Basis Data Terdistribusi...

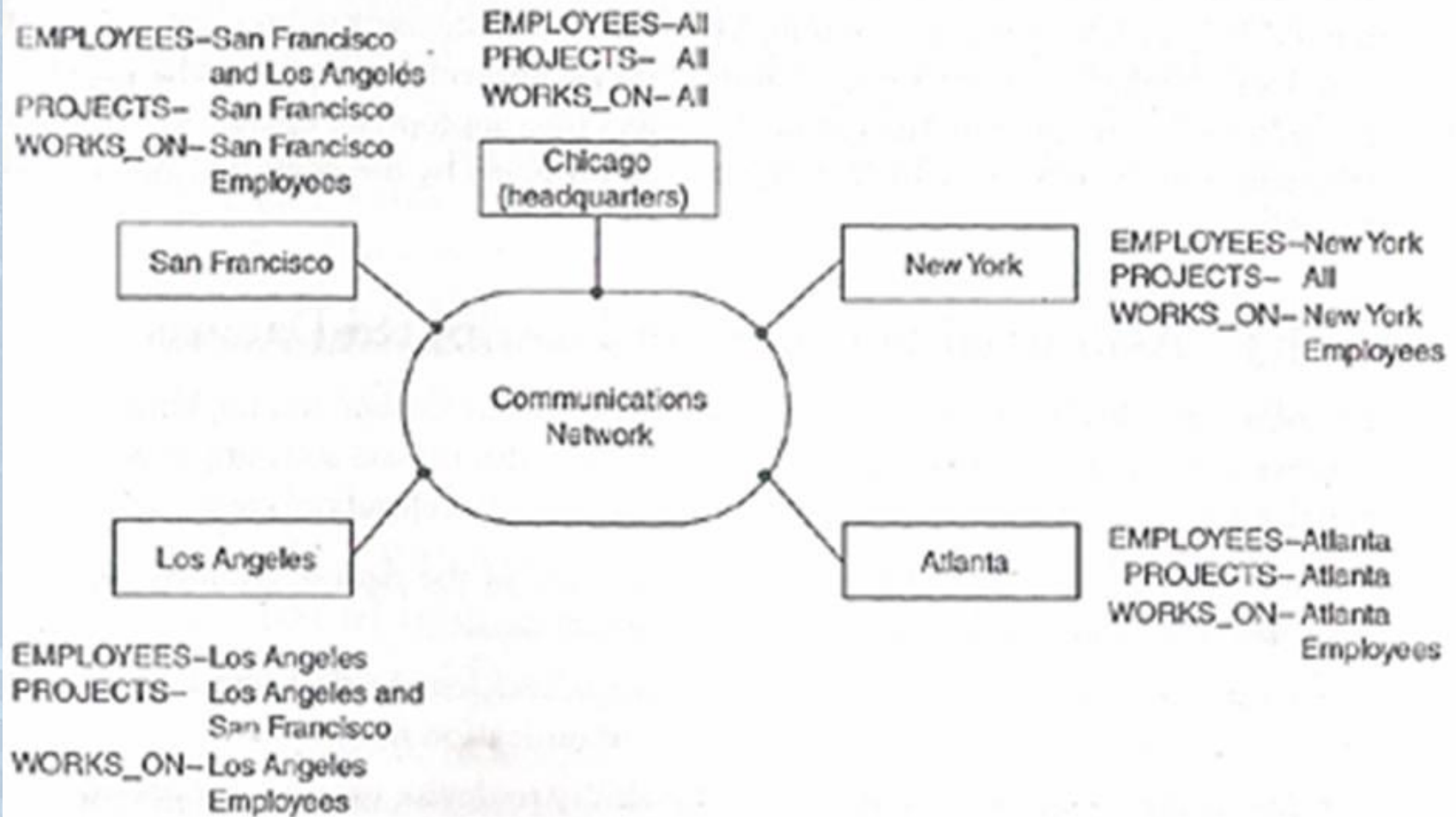


Figure 24.2 Data distribution and replication among distributed databases

# Fungsi Tambahan Basis Data Terdistribusi

- Keeping track of data
- Proses query yang terdistribusi
- Manajemen transaksi yang terdistribusi
- Manajemen replikasi data
- Pemulihan basis data terdistribusi
- Keamanan
- Manajemen direktori (katalog) terdistribusi

# Perbedaan DDBMS dengan Sistem Central : Hardware

- Ada banyak komputer yang disebut sites atau nodes.
- Site-site ini terhubung oleh jaringan komunikasi untuk mengirim data dan perintah-perintah di antara site-site tersebut.

# Penyimpanan Data pada DDBMS

- Pada **DBMS terdistribusi** relasi disimpan pada beberapa tempat yang berbeda (**sites / nodes**).
- Akses ke relasi secara remote menimbulkan biaya kirim pesan, oleh karena itu untuk mengurangi biaya satu relasi **dipartisi** atau dibuat **fragmen** atau **replika** yang di-*assign* tersebar pada beberapa site sedemikian sehingga penyimpanan dilakukan berdasarkan **frekuensi** penggunaan lokal, agar **biaya overhead** komunikasi berkurang.
- Informasi yang berhubungan dengan **fragmentasi**, **alokasi** dan **replikasi** data disimpan dalam sebuah global directory yang diakses oleh aplikasi **DDBMS**.



# Penyimpanan Data pada DDBMS...

## Fragmentasi :

- Fragmentasi dari data dalam bentuk fragmen data yang disimpan pada tempat yang berbeda, ada dua kemungkinan, yaitu **fragmentasi horizontal** dan **fragmentasi vertikal**.
- **Fragmen vertikal** :
  - Terdiri dari fragmen atribut (**kolom**) untuk semua record
  - Frekuensi akses data : dibuat berdasarkan jenis atribut (kolom) tertentu
- **Fragmen horizontal** :
  - Terdiri dari fragmen record (**baris**) untuk semua atribut
  - Frekuensi akses data : dibuat berdasarkan salah satu atribut (kolom) dengan nilai tertentu

# Penyimpanan Data pada DDBMS...

- Contoh Fragmentasi Vertikal & Horizontal :

Tid	pid	nama	kota	umur	saldo
T1	23412	Joni	Bogor	18	4300
T2	43252	Susi	Jakarta	18	3450
T3	43211	Susi	Jakarta	19	4220
T4	36123	Maya	Bandung	21	5340
T5	36056	Gani	Bandung	22	2500

Fragmen vertikal

Fragmen horizontal

# Penyimpanan Data pada DDBMS...

## Replikasi :

- **Replika** adalah hasil replikasi satu relasi data atau fragmen relasi yang dapat disimpan pada lebih dari satu tempat, jumlah replika fragmen relasi tidak harus sama untuk satu relasi.
- **Contoh:** jika relasi R dijadikan tiga fragmen R1, R2, R3, mungkin R1 tidak dibuat replikanya, tetapi R2 dibuat replika di satu tempat lain dan R3 dibuat replika di semua tempat.
- **Tujuan replikasi** ada dua motivasi, yaitu:
  - Meningkatkan availabilitas data, dan
  - Mempercepat evaluasi query jika ada replika fragmen atau satu relasi pada tempat lokal.

# Replikasi dan Alokasi Data

- Berguna untuk meningkatkan ketersediaan data.
- Kasus paling ekstrim: replikasi keseluruhan basis data pada setiap site dalam sistem terdistribusi → **fully replicated distributed database**
  - **Keuntungan:** sistem terus beroperasi selama paling tidak satu site tetap berjalan
  - **Kelemahan:** operasi update menjadi lambat
- Kasus paling ekstrim lainnya: **tidak ada replikasi**

# Replikasi dan Alokasi Data...

- Contoh kasus khusus replikasi partial:
  - Untuk pekerja yang mobile -seperti sales rep, financial planners dan claim adjustors- mereka membawa basis data replikasi dalam laptop atau PDA dan melakukan sinkronisasi secara periodik dengan server basis data.
- Masing-masing fragmen -atau salinan dari sebuah fragmen- harus *diassign* ke site tertentu dalam sistem terdistribusi. Proses ini disebut distribusi data (atau alokasi data)

# Replikasi dan Alokasi Data...

## Contoh Fragmentasi, Alokasi dan Replikasi:

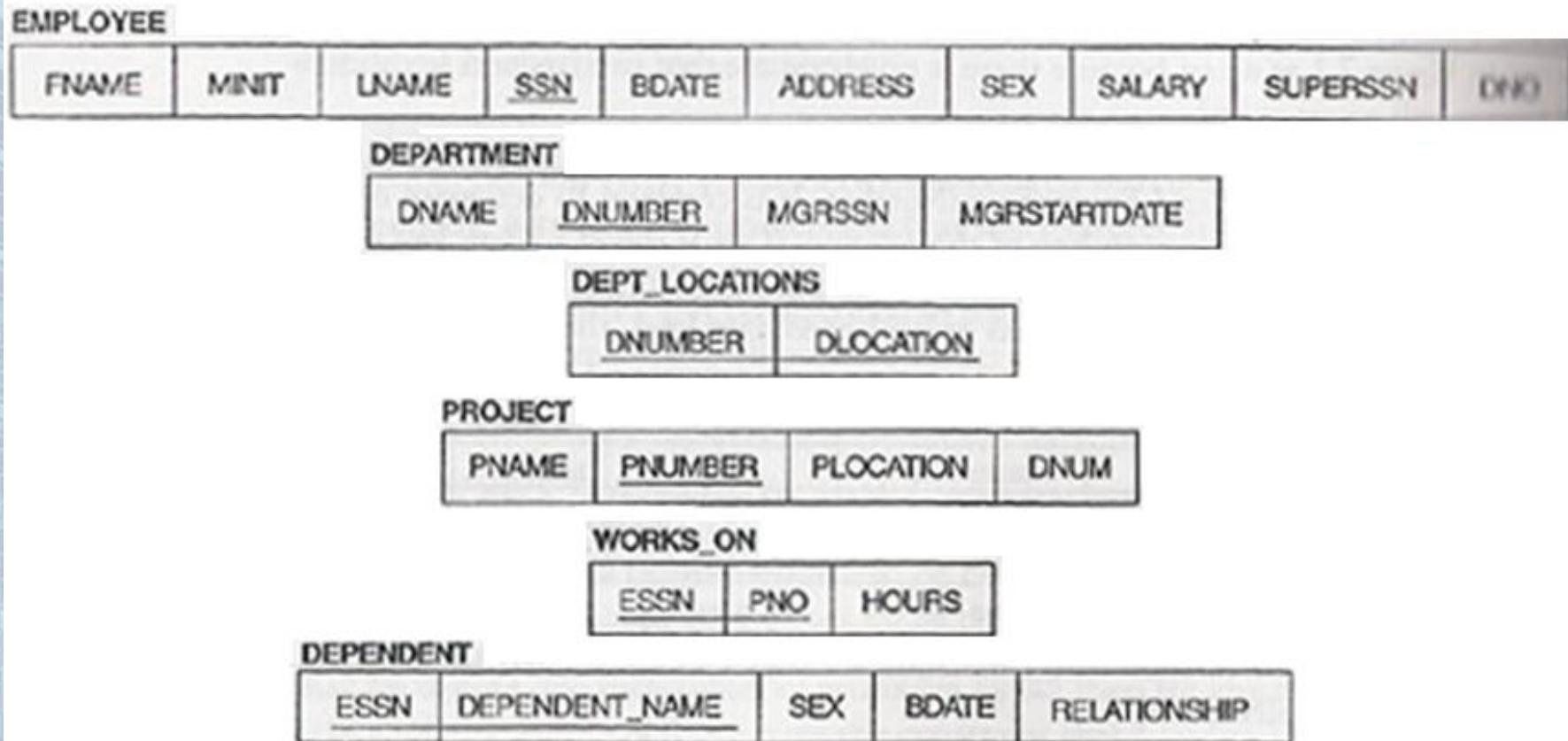


Figure 7.5 Schema diagram for the COMPANY relational database schema; the primary keys are underlined.

# Replikasi dan Alokasi Data...

Contoh Fragmentasi,  
Alokasi dan Replikasi:

EMPLOYEE	FNAME	MINIT	LNAME	SSN	BOATE	ADDRESS	SEX	SALARY	SUPERSSN	ONO
John	B	Smith	123456789	1965-01-09	731 Forden, Houston, TX	M	30000	333445555	5	
Franklin	T	Wong	333445555	1955-12-01	638 Voss, Houston, TX	M	40000	888665555	5	
Alda	J	Zelays	999887777	1968-07-19	3321 Costa, Spring, TX	F	25000	987654321	4	
Jennifer	S	Wallace	987654321	1941-06-20	291 Barry, Beilake, TX	F	43000	888665555	4	
Ramesh	K	Narayan	666884444	1982-09-15	975 First Oak, Humble, TX	M	38000	333445555	5	
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5	
Ahmad	V	Jabber	987987987	1989-03-29	990 Dallas, Houston, TX	M	25000	987654321	4	
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	null	1	

DEPARTMENT	DNAME	DNUMBER	MGRSSN	MGRSTARTDATE
Research		5	333445555	1988-05-22
Administration		4	987654321	1995-01-01
Headquarters		1	888665555	1981-05-19

DEPT_LOCATIONS	DNUMBER	DLOCATION
	1	Houston
	4	Stafford
	5	Beilake
	5	Sugarland
	5	Houston

WORKS_ON	ESSN	PNO	HOURS
	123456789	1	32.5
	123456789	2	7.5
	666884444	3	40.0
	453453453	1	20.0
	453453453	2	20.0
	333445555	2	10.0
	333445555	3	10.0
	333445555	10	10.0
	333445555	20	10.0
	999887777	30	30.0
	999887777	10	10.0
	987987987	10	35.0
	987987987	30	5.0
	987654321	30	20.0
	987654321	20	15.0
	888665555	20	null

PROJECT	PNAME	PNUMBER	PLOCATION	DNUM
ProductX		1	Beilake	5
ProductY		2	Sugarland	5
ProductZ		3	Houston	5
Computerization		10	Stafford	4
Reorganization		20	Houston	5
Newbenefits		30	Stafford	4

DEPENDENT	ESSN	DEPENDENT_NAME	SEX	BOATE	RELATIONSHIP
	333445555	Alice	F	1995-04-05	DAUGHTER
	333445555	Theodore	M	1993-10-25	SON
	333445555	Joy	F	1998-05-03	SPOUSE
	987654321	Ahmed	M	1942-02-28	SPOUSE
	123456789	Michael	M	1968-01-04	SON
	123456789	Alice	F	1968-12-30	DAUGHTER
	123456789	Elizabeth	F	1967-05-05	SPOUSE

Figure 7.6 One possible relational database state corresponding to the COMPANY schema.

# Replikasi dan Alokasi Data...

(a)

EMP05	FNAME	MINIT	LNAME	SSN	SALARY	SUPERSSN	DNO
	John	B	Smith	123456789	30000	333445555	5
	Franklin	T	Wong	333445555	40000	888665555	5
	Ramesh	K	Narayan	666884444	30000	333445555	5
	Joyce	A	English	453453453	25000	333445555	5

DEP5	DNAME	DNUMBER	MGRSSN	MGRSTARTDATE
	Research	5	333445555	1983-05-22

DEP5_LOCS	DNUMBER	LOCATION
	5	Bellevue
	5	Sugarland
	5	Houston

WORKS_ON5	ESSN	PNO	HOURS
	123456789	1	32.5
	123456789	2	7.5
	666884444	3	40.0
	453453453	1	20.0
	453453453	2	20.0
	333445555	2	10.0
	333445555	3	10.0
	333445555	10	10.0
	333445555	20	10.0

PROJ5	PNAME	PNUMBER	PLOCATION	DNUM
	Product X	1	Bellevue	5
	Product Y	2	Sugarland	5
	Product Z	3	Houston	5

Data at Site 2

(b)

EMP04	FNAME	MINIT	LNAME	SSN	SALARY	SUPERSSN	DNO
	Alicia	J	Keyser	999887777	25000	987654321	4
	Jennifer	S	Walter	987654321	43000	888665555	4
	Ahmad	V	Jabbar	987987987	25000	987654321	4

DEP4	DNAME	DNUMBER	MGRSSN	MGRSTARTDATE
	Administration	4	987654321	1995-01-01

DEP4_LOCS	DNUMBER	LOCATION
	4	Stafford

WORKS_ON4	ESSN	PNO	HOURS
	333445555	10	10.0
	999887777	30	30.0
	999887777	10	10.0
	987987987	10	35.0
	987987987	30	5.0
	987654321	30	20.0
	987654321	20	15.0

PROJ4	PNAME	PNUMBER	PLOCATION	DNUM
	Computerization	10	Stafford	4
	Newbenefits	30	Stafford	4

Data at Site 3

Figure 24.3

Allocation of fragments to sites. (a) Relation fragments at site 2 corresponding to department 5. (b) Relation fragments at site 3 corresponding to department 4.

Contoh Fragmentasi,  
Alokasi dan Replikasi:



# Replikasi dan Alokasi Data...

## Contoh Fragmentasi, Alokasi dan Replikasi:

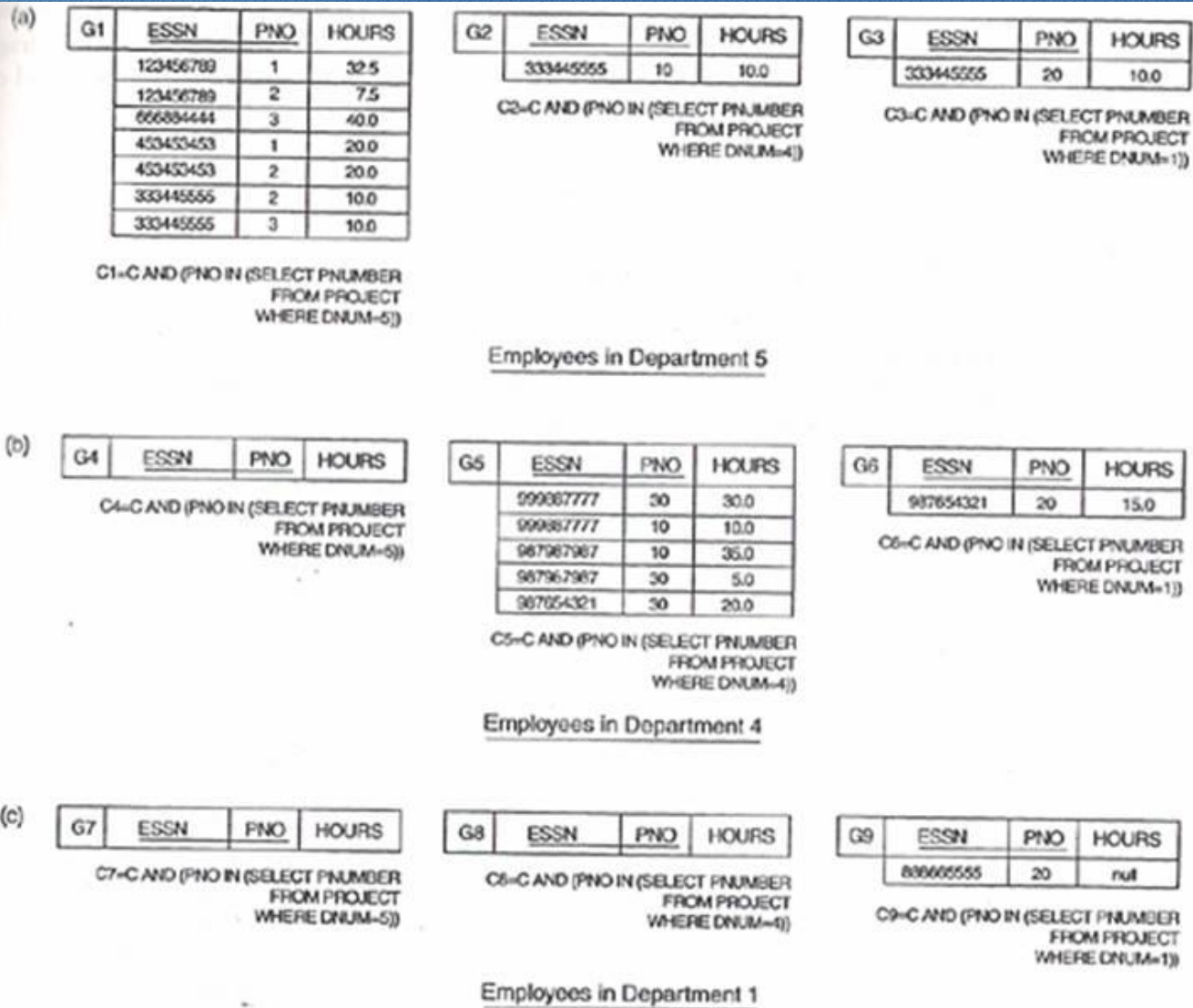


Figure 24.4 Complete and disjoint fragments of the WORKS\_ON relation. (a) Fragments of WORKS\_ON for employees working in department 5 ( $C=[\text{ESSN IN (SELECT SSN FROM EMPLOYEE WHERE DNO=5)}]$ ). (b) Fragments of WORKS\_ON for employees working in department 4 ( $C=[\text{ESSN IN (SELECT SSN FROM EMPLOYEE WHERE DNO=4)}]$ ). (c) Fragments of WORKS\_ON for employees working in department 1 ( $C=[\text{ESSN IN (SELECT SSN FROM EMPLOYEE WHERE DNO=1)}]$ ).

# Transaksi Teristribusi

- Pada sistem terdistribusi, transaksi dapat dilakukan pada suatu tempat tetapi dapat akses data di tempat lain.
- Setiap transaksi **dipecah** menjadi beberapa **subtransaksi** yang dijalankan secara tersebar melalui manajer transaksi pada setiap tempat sub-transaksi dijalankan untuk dikoordinasikan.
- Untuk kasus kontrol proses yang terjadi bersamaan (**konkuren**), maka ada mekanisme **penguncian** objek yang digunakan yang ada di tempat lain, juga bagaimana cara mendeteksi jika terjadi **deadlock**.

# Transaksi Teristribusi...

- Pengelolaan penguncian objek yang terdistribusi dapat dilakukan dengan beberapa cara, yaitu secara **terpusat (sentralisasi)**, **replika primer** dan **terdistribusi penuh**.
- Pada cara **terpusat**, penanganan penguncian dilakukan dari satu tempat.
- Pada cara **replika primer**, penanganan penguncian dilakukan pada tempat replika primer berada.
- Dan pada cara **terdistribusi penuh**, maka penanganan penguncian dilakukan pada tempat replika yang akan dikunci. Cara ketiga ini lebih **banyak digunakan**.

**S.E.K.I.A.N. . .**

