

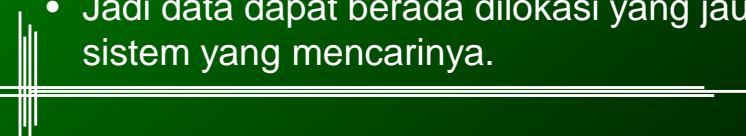


PENGANTAR DATA TERDISTRIBUSI

Database Terdistribusi

Pendahuluan



- Teknologi database cenderung membentuk suatu konfigurasi sistem pengolahan data yang terpusat.
 - Secara teoritis database dapat didistribusikan dimana datanya didistribusikan ke berbagai tipe konfigurasi
 - Sebelum mendistribusikan database perlu ditentukan lokasi dan mapping antara database akan ditanamkan atau dilaksanakan.
 - Jadi data dapat berada dilokasi yang jauh, lalu sistem yang mencarinya.
- 

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

Ciri Basis Data Terdistribusi

1. Data disimpan pada sejumlah tempat. Setiap tempat secara logic terdiri dari prosesor tunggal
2. Prosesor pada tempat yang berbeda tersebut dihubungkan dengan jaringan komputer
3. Sistem Basis data terdistribusi bukan sekumpulan file yang berada pada berbagai tempat tetapi sebuah database pada berbagai tempat
4. Setiap tempat mempunyai kemampuan untuk secara mandiri memproses permintaan user yang membutuhkan akses ke data di tempat tersebut, dan juga mampu untuk memproses data yang tersimpan di tempat lain

Teknologi DDBS (Distributed Database System)

- merupakan gabungan dari dua pendekatan pengolahan data yang sama sekali berlawanan :
 - Database System, dan
 - Teknologi Jaringan Komputer.
- Tujuan utama Database System
 - Mengintegrasikan data dan sentralisasi, sehingga akses (deskripsi, manipulasi dan kontrol) terhadap data sangat terkontrol
→ **Integration, not Centralization**
- Tujuan utama Jaringan Komputer
 - Membuat mode kerja yang benar-benar menghindari terjadinya sentralisasi beban kerja.
- Teknologi DDBS mengintegrasikan sistem database yang terdistribusi, untuk menghasilkan suatu teknologi yang powerful dan menjanjikan
→ Integration without Centralization

Distributed Processing (Distributed Computing)

- Apa itu Distributed Processing (Distributed Computing)?
- Beberapa sinonim dari Distributed Processing:
 - Distributed Function: Distribusi fungsi (CPU atau I/O) meskipun dalam sistem single-processor komputer.
 - Distributed Computing System: Interkoneksi sejumlah elemen-elemen pengolahan (computing device yang dapat mengeksekusi program di masing-masing) dalam suatu jaringan komputer dan bekerjasama dalam menjalankan suatu tugas.
 - Network
 - Multiprocessors
 - Timeshared systems
- Apa yang didistribusikan?
 - Processing logic, seperti definisi dari Distributed Computing System diatas.
 - Functions, pendelegasian ke bermacam-macam hardware/software (tergantung fungsi komputer)
 - Data, pendistribusian data ke sejumlah site pengolahan.
 - Control, kontrol eksekusi sejumlah task secara terdistribusi.

Distributed Processing (Distributed Computing) (lanj.)

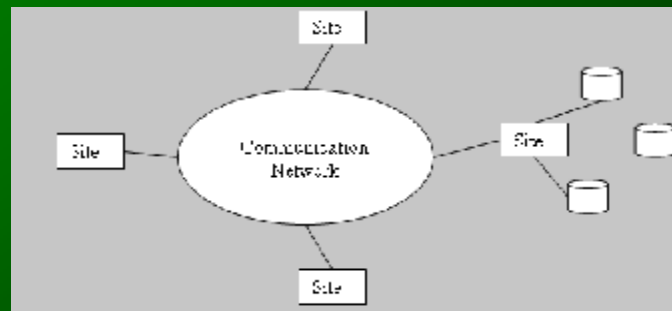
- Klasifikasi Distributed Computing Systems (by Bochmann):
 - Berdasarkan degree of coupling
 - Saling Tergantung
- Berdasarkan struktur interkoneksi
 - Berdasarkan ketergantungan komponen
 - Berdasarkan sinkronisasi antar komponen
- Berdasarkan degree of coupling (derajat konektivitas antar elemen pengolahan), yakni rasio jumlah pertukaran data per jumlah pengolahan lokal suatu task:
 - Weak Coupling: Computer Network
 - Strong Coupling: A Computer
- Berdasarkan struktur interkoneksi:
 - Point-to-point interconnection
 - Common interconnection channel
- Berdasarkan ketergantungan komponen:
 - Saling tergantung saat eksekusi suatu task
 - Komunikasi diawal sebelum eksekusi dan setelah eksekusi
- **Apa perlunya Distribusi?**
 - Respon terhadap struktur organisasi yang mengarah ke distributed enterprise
 - Lebih ekonomis, praktis dan terpercaya (reliable)

APA ITU BASIS DATA TERDISTRIBUSI?

- Basis Data Terdistribusi (DDB: Distributed Database)
 - Kumpulan sejumlah database yang saling berhubungan secara logis (logically interrelated) diatas suatu jaringan komputer (distributed over a computer network).
- Distributed DBMS adalah sistem software yang memberikan fasilitas untuk manajemen DDB dan menyediakan mekanisme agar distribusi tersebut transparent di hadapan user
- Sistem Basis data Terdistribusi (Distributed Database System / DDBS) adalah DDB yang menggunakan DBMS
- DDBS bukan koleksi file yang secara individual disimpan di setiap node pada suatu jaringan komputer.
- File tidak hanya saling berhubungan secara logis, tapi file-file tersebut harus terstruktur dan pengaksesannya harus melalui interface yang sama.

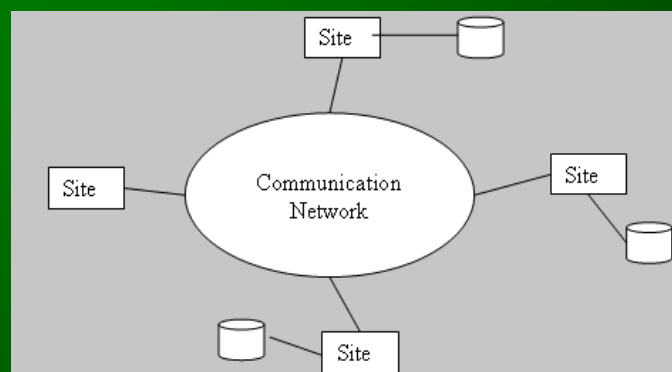
APA ITU BASIS DATA TERDISTRIBUSI? (lanj)

- Distribusi data yang terpisah secara fisik akan menimbulkan masalah yang tidak dijumpai ketika database ada di komputer yang sama.
→ Inilah yang membedakan antara DDBS dengan Multiprocessor.
- DDBS bukan merupakan sistem dimana meskipun ada jaringan komputer, tapi databasenya hanya terletak disatu node dari jaringan tsb.



APA ITU BASIS DATA TERDISTRIBUSI? (lanj)

- Jaringan Komputer atau Kolesi File, tidak cukup untuk membentuk DDBS!!!.



Sistem homogen dan heterogen

- Sistem yang homogen berarti sistem database terdistribusi yang menghubungkan DBMS yang bertipe sama, kompatibel, struktur dan deskripsi datanya sama
- Sistem yang menghubungkan DBMS yang tidak kompatibel adalah sistem yang heterogen

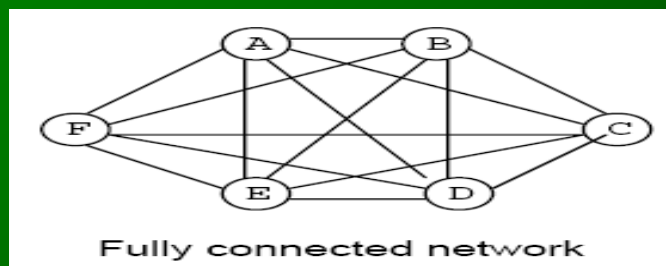
Sistem database terdistribusi

- Pada sistem database terdistribusi, data disimpan pada beberapa komputer. Komputer dalam sebuah sistem terdistribusi berbeda dalam ukuran dan fungsi. Prosesor mungkin mencakup mikrokomputer workstation, minikomputer atau sistem komputer besar. Prosesor-prosesor ini dapat dikaitkan dengan sejumlah nama yang berbeda seperti: side, node, komputer, dan sebagainya.
- Sistem database terdistribusi terdiri dari kumpulan tempat, masing-masing turut serta dalam pengeksekusian transaksi yang dapat mengakses data pada sebuah tempat atau berbeda tempat.
- Perbedaan utama antara centralized dan sistem database terdistribusi adalah pada sistem terpusat data berada pada satu lokasi sedangkan pada database terdistribusi data berada pada beberapa lokasi.

Struktur dari database terdistribusi

- Sistem database terdistribusi terdiri dari kumpulan tempat, masing-masing memelihara sistem database lokalnya, masing-masing tempat dapat memproses transaksi lokal. Beberapa konfigurasi dari sistem database terdistribusi:
 - Fully connected Network
 - Partially connected Network
 - Tree connected Network
 - Star network
 - Ring network

Fully connected Network

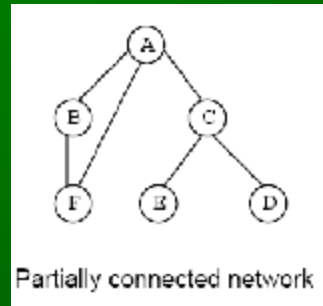


Fully Connected network :

Keuntungan : kalau salah satu node rusak, yang lainnya masih dapat berjalan (tetapi biaya mahal).

Kerugian : control management tidak terjamin

Partially connected Network

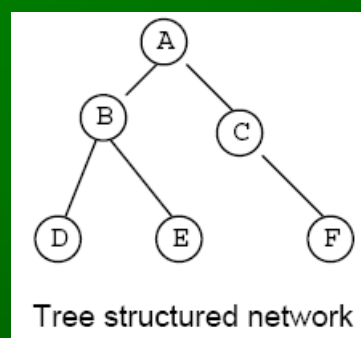


Partially connected network :

Keuntungan : reliability rendah, biaya dapat ditekan

Kerugian : control management tidak terjamin

Tree connected Network

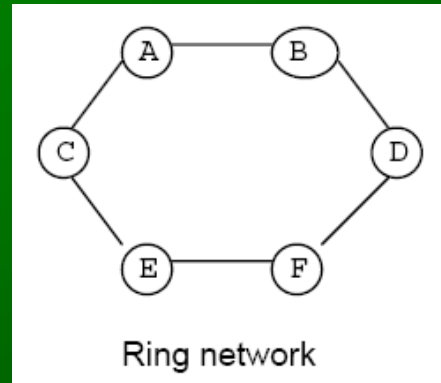


Tree structure network :

Keuntungan : bersifat sentral, control management lebih terjamin

Kerugian : kalau node pusat (A) rusak, semua akan rusak.

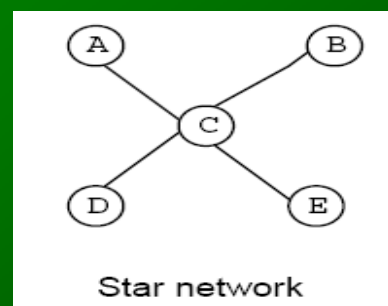
Cat : setiap proses dimulai dari bawah.



Ring Network (LAN) :

Keuntungan : rusak satu, yang lain masih berjalan

Kerugian : Control management kurang terjamin karena bersifat desentralisasi



Star Network (LAN) :

Keuntungan : - control management lebih terjamin, karena bersifat sentral
- reliability rendah

Kerugian : kalau pusat rusak, yang lainnya rusak

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 Sentral: 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

Teknik Memecah Basis Data

- Memecah basis data menjadi unit-unit logika yang disebut **fragmen-fragmen** yang bisa *diassign* untuk disimpan di berbagai site
- Penggunaan **replikasi data** yang memungkinkan data tertentu untuk disimpan di lebih dari satu site
- Proses **alokasi** fragmen –atau replika fragmen– untuk penyimpanan di berbagai site

Teknik Memecah Basis Data (2)

- Teknik-teknik di atas digunakan dalam proses perancangan basis data terdistribusi
- Informasi yang berhubungan dengan fragmentasi, alokasi dan replikasi data disimpan dalam sebuah **global directory** yang diakses oleh aplikasi DDBS

Fragmentasi Data

- Fragmentasi Horizontal
 - Membagi sebuah relasi secara horisontal dengan mengelompokkan baris-baris untuk membuat subset dari tuples
- Fragmentasi Vertical
 - Membagi sebuah relasi secara vertikal berdasarkan kolom
- Fragmentasi Hibrid

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 (2)

- 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

Replikasi dan Alokasi Data (3)

- Masing-masing fragmen –atau salinan dari sebuah fragmen- harus diassign ke site tertentu dalam sistem terdistribusi. Proses ini disebut **distribusi data** (atau **alokasi data**)

Keuntungan dan kerugian pada database terdistribusi

- Alasan mengapa diperlukan database terdistribusi, yaitu:
 - Shering of data (berbagi data)
 - Reliability dan availability
 - Mempercepat pemrosesan query
- Beberapa kerugian penerapan database terdistribusi, yaitu:
 - Biaya pengembangan perangkat lunak
 - Kemungkinan terjadinya bug dan adanya pengeluaran tambahan untuk pemrosesan

Keuntungan dari distribusi data

- Keuntungan utama dari sistem database terdistribusi adalah kemampuannya untuk berbagi dan mengakses data dalam bentuk yang efisien dan dapat dipercaya.

Kerugian dari distribusi data

- Biaya pengembangan perangkat lunak
- Memperbesar terjadinya kesalahan (bug), karena pengolahan terdistribusi ini memungkinkan adanya operasi paralel
- Meningkatkan biaya pemrosesan

KELEBIHAN DAN KELEMAHAN BASIS DATA TERDISTRIBUSI

- KELEBIHAN:
 - **Otonomi Lokal:** Karena data terdistribusi, kelompok user yang bisa menggunakan data tersebut dapat menyimpannya di site dimana dia bekerja, sehingga masing-masing site mempunyai kontrol lokal.
 - **Performansi Tinggi:** Karena data yang digunakan umumnya lebih dekat dengan user, maka performansi akses ke database dapat ditingkatkan.
 - **Kepercayaan:** Karena ada replikasi data di lebih dari satu tempat, maka kemungkinan crashnya satu node/site, atau hilangnya aksesibilitas karena kegagalan link komunikasi, tidak membuat data untuk tidak dapat diakses.
 - **Ekonomis:**
 - Dari sisi communication cost: Lebih ekonomis dengan membagi aplikasi dan menjalankannya di beberapa situs lokal, dari pada satu aplikasi dipaksa untuk akses ke database yang tersebar.
 - Dari sisi biaya normal: Lebih ekonomis membeli komputer kecil-kecil yang kemampuannya sama dengan satu komputer besar.
 - **Ekspansinya Mudah:** Dapat disesuaikan dengan mudah seiring dengan berkembangnya ukuran database. Paling sekitar penambahan pengolahan dan kemampuan penyimpanan di jaringan. Jelas, bahwa kemampuannya tidak akan bertambah secara linier, tapi paling tidak improvisasi dapat dimungkinkan.
 - **Shareability:** Untuk merespon trend organisasi ke arah distributed fashion

KELEBIHAN DAN KELEMAHAN BASIS DATA TERDISTRIBUSI

• KELEMAHAN:

- **Kurangnya Pengalaman**: Minimnya percobaan di lingkup kerja nyata.
- **Kompleksitas**: Menambah masalah baru dari masalah yang ada pada DBS.
- **Biaya**: Menambah biaya hardware (hardware u/ mekanisme komunikasi). Memang H/W sekarang murah shg biaya tdk terlalu signifikan, tapi biaya pengembangan teknologi/software dalam masalah teknis (distributed debuggers dll) sangat diperlukan. Selain itu adalah biaya untuk replikasi usaha/tenaga (manpower), yang dibutuhkan di beberapa site → Analisa trade-off antara keuntungan dari efisiensi/efektivitas penggunaan informasi dengan biaya penambahan personel.
- **Kontrol yang Terdistribusi**: Ini kelebihan yang sudah disebut, tapi sayangnya ini dapat menimbulkan masalah sinkronisasi dan koordinasi
- **Keamanan**: Sekuriti jaringan lebih kompleks dari sekuriti akses database yang tersentralisasi.
- **Migrasi yang susah**: Belum ada tool atau metodologi yang dapat membantu untuk konversi dari DBS ke DDBS.

FAKTOR PENGHAMBAT

• Kompleksitas tambahan dengan adanya lingkup kerja yang terdistribusi

- Replikasi data (semua atau sebagian) di jaringan, tapi tidak perlu di semua site. Yang esensial hanya database harus ada di tempat lebih dari satu. Duplikasi ini karena pertimbangan reliability dan efisiensi.



DDBS harus:

Memilih satu diantara site yang menyimpan data yang dibutuhkan.
Mengupdate setiap site yang menyimpan data yang sudah diupdate

- Apabila ada suatu kegagalan (site crash atau link komunikasi putus) selama masa update, sistem harus segera dapat mengupdate site yang tidak tergapai tadi sesegera mungkin (saat sistem tsb kembali normal)
- Sinkronisasi transaksi lebih dipertimbangkan lagi dibandingkan sistem yang tersentralisasi (DBS).