



13

## Parallel Database

by: Ahmad Syauqi Ahsan

# Latar Belakang

- Parallel Database Management System adalah DBMS yang diimplementasikan pada parallel computer yang mana terdiri dari sejumlah node (prosesor dan memory) yang saling terhubung menggunakan jaringan super cepat.
- Sistem ini berusaha sebaik-baiknya untuk memanfaatkan arsitektur mikroprosesor paling modern menggunakan solusi yang berorientasi software untuk pengelolaan data.

# Tujuan

3

- Permasalahan DBMS konvensional:
  - ▣ Waktu akses disk yang tinggi
  - ▣ Database dengan ukuran yang sangat besar tidak dapat didukung menggunakan satu sistem
- Parallel Database Management System adalah satu-satunya solusi yang mungkin untuk meningkatkan bandwidth I/O melalui parallelism dan untuk menyimpan database dalam jumlah yang sangat besar didalam satu sistem.

# Keuntungan Parallel DBMS

4

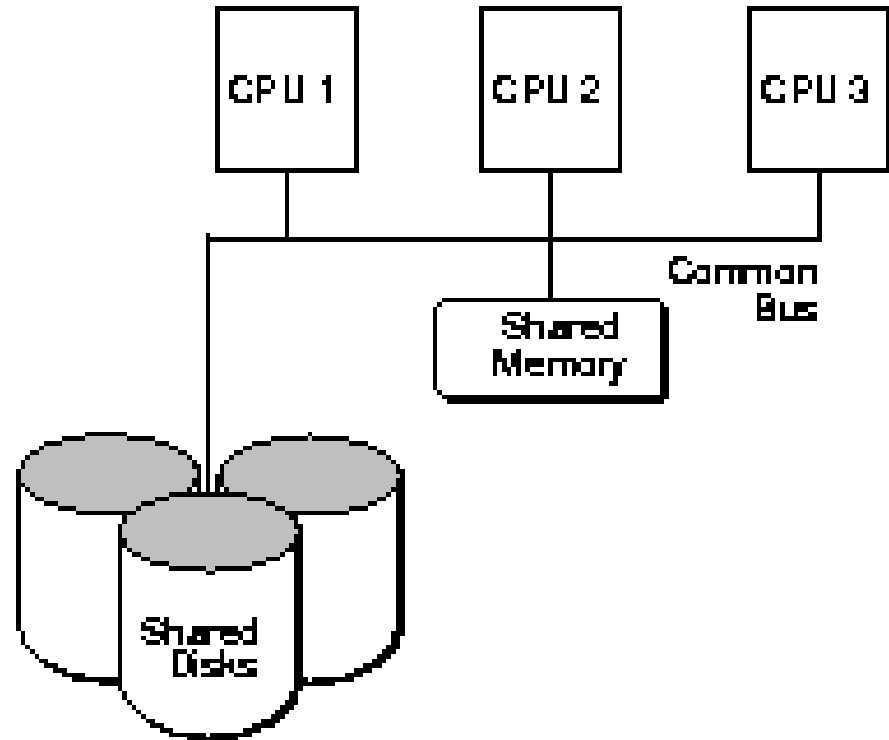
- Performa tinggi → meningkatkan throughput (intra-query parallelism) dan mengurangi waktu respon (intra-query parallelism).
- High Availability → menggunakan replikasi data
- Extensibility → dapat diperbesar kapasitasnya serta ditingkatkan kecepatannya secara linear.

# Arsitektur Parallel DBMS

5

## Shared Memory:

- Keuntungan:
  - ▣ Simplicity
  - ▣ Load Balancing
- Permasalahan:
  - ▣ Harga
  - ▣ Limited Extensibility
  - ▣ Low Availability



# Arsitektur Parallel DBMS (2)

6

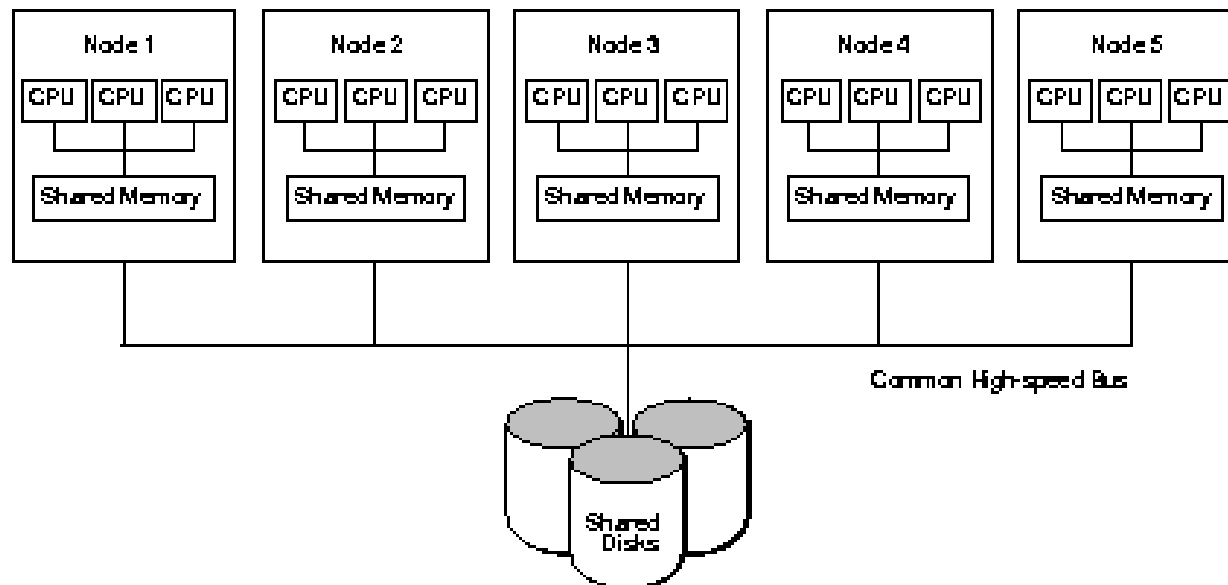
## Shared Disk

### □ Keuntungan:

- ▣ Biaya, Extensibility, Load Balancing, Availability

### □ Permasalahan:

- ▣ Higher Complexity, Potential Coherence problem



# Arsitektur Parallel DBMS (3)

7

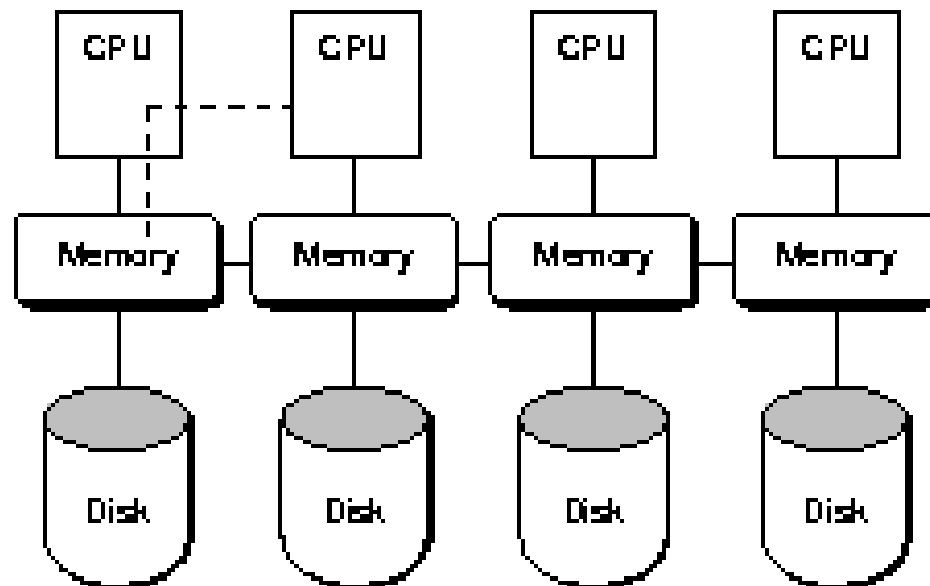
## Shared Nothing

### □ Keuntungan:

- ▣ Cost, Extensibility, Availability

### □ Permasalahan

- ▣ Complex, penambahan node baru membutuhkan pengaturan ulang database



# Teknik Parallel DBMS

- **Data Allocation:** Metode untuk menyimpan database kedalam beberapa disk yang berbeda guna untuk memastikan proses input-output secara parallel yang efisien.
- **3 Strategi Partitioning (Fragmentation):**
  - Round Robin: tuple ke  $l$  disimpan di partisi ke  $(l \bmod n)$  dimana  $n$  adalah jumlah partisi
  - Hashing: mengaplikasikan fungsi hash untuk beberapa atribut guna memberikan nomor partisi
  - Range Partitioning – Distribute tuples based
    - Range Partitioning: mendistribusikan tuple berdasarkan nilai dari beberapa atribut.



# Perbandingan Strategi Partitioning

9

	Round Robin	Hashing	Range
Sequential Scan	Best/good parallelism	Good	Good
Point Query	Difficult	Good for hash key	Good for range vector
Range Query	Difficult	Difficult	Good for range vector

# Penggunaan Data Fragmentation

10

- Memaksimalkan performa sistem
- Meminimalkan waktu respon (melalui intra-query parallelism)
- Memaksimalkan throughput (melalui inter-query parallelism)
- **Problems:** distribusi data yang tidak seimbang membuat partisi yang tidak seragam sehingga akan mengganggu (mengurangi manfaat) load balancing

# Query Parallelism

11

- Inter query parallelism: eksekusi secara parallel dari beberapa query yang dihasilkan oleh beberapa transaksi yang berjalan secara bersamaan
- Intra query parallelism, terdapat dua macam:
  - Inter-operator: mengeksekusi beberapa operasi dari pohon query pada beberapa prosesor secara parallel
  - intra-operator: beberapa operator dieksekusi oleh beberapa prosesor, dimana masing-masing prosesor bekerja mengerjakan sebagian (subset) dari keseluruhan data

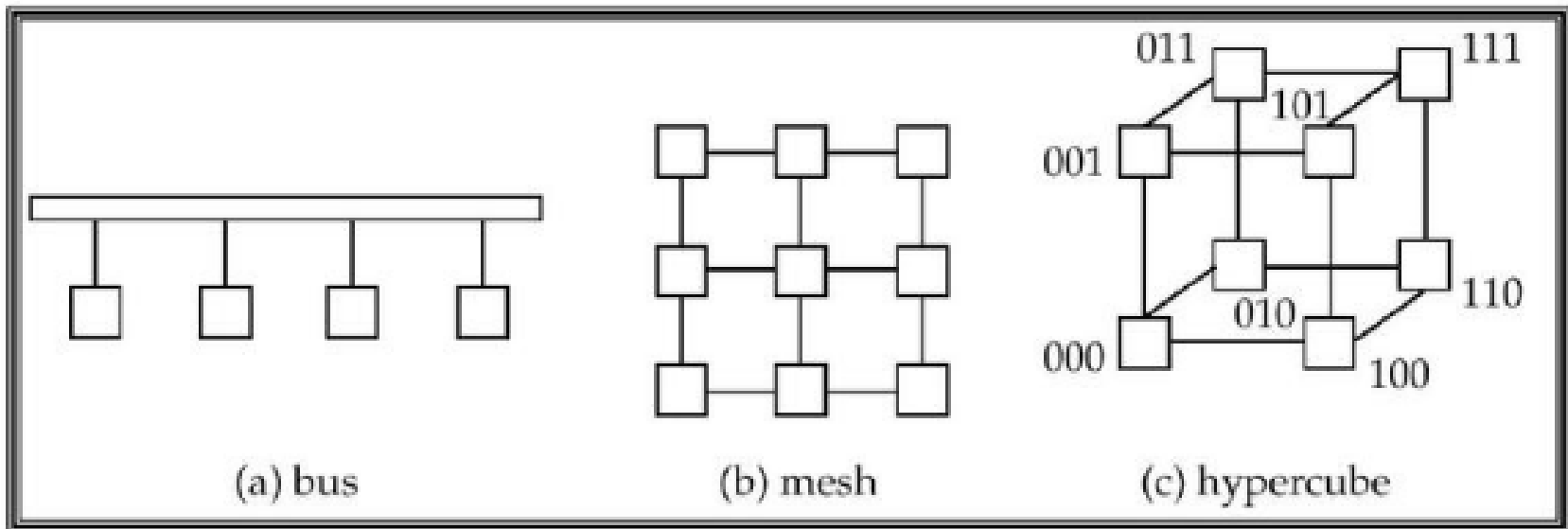
# Level Parallel

12

- **Coarse-grain Parallel:** mesin komputer dengan tipe ini terdiri dari sejumlah kecil processor dengan kemampuan tinggi
- **Massively Parallel** atau **Fine-grain Parallel:** mesin komputer dengan tipe ini memanfaatkan ribuan processor kecil

# Arsitektur Interkoneksi

13



# Isu-isu pada Parallel DBMS

14

- Partisi Data (Data Partitioning)
- Pemrosesan Query secara parallel (Parallel Query Processing)

# Tipe-tipe Query Database yang umum

15

- Sequential scan
- Point query
- Range query

# Perancangan Parallel System

16

- Beberapa isu dalam perancangan Parallel Systems:
- Dibutuhkan mekanisme untuk memasukkan data dari sumber eksternal secara parallel guna menangani masuknya data dalam jumlah yang sangat besar.
- Kelenturan untuk terjadi kegagalan pada processor atau disk
  - Kemungkinan terjadinya kegagalan (failure) beberapa disk atau processor pada parallel system lebih tinggi
  - System masih bisa beroperasi (mungkin dengan performa yang menurun) ketika terjadi kegagalan (failure)
  - Redundancy dapat dicapai dengan menyimpan tambahan copy untuk setiap data pada processor yang lain



# Perancangan Parallel System (2)

17

- Harus ada dukungan untuk pengelolaan data dan perubahan schema secara online.
  - Contoh: pembentukan index didalam database berukuran terabyte membutuhkan waktu beberapa jam atau hari, bahkan pada parallel system
    - Perlu untuk mengijinkan pemrosesan yang lain (insertions/deletions/updates) untuk tetap dapat dilakukan bahkan ketika index sedang dibentuk.
  - Ide dasar: pembentukan index melacak perubahan-perubahan dan
- Juga diperlukan dukungan untuk proses repartitioning secara online dan perubahan schema (dieksekusi secara bersamaan dengan pemrosesan lainnya)

18

# Tanya Jawab

Terima Kasih