



**UNIVERSITI PUTRA MALAYSIA**

**PRESTASI KAEDAH PENILAIAN PERTANYAAN DALAM  
PANGKALAN DATA TERAGIH**

**MOHD SHAHIDAN BIN ABDULLAH**

**FSAS 1997 4**

**PRESTASI KAEDAH PENILAIAN PERTANYAAN DALAM  
PANGKALAN DATA TERAGIH**

Oleh

**MOHD SHAHIDAN BIN ABDULLAH**

Tesis ini dikemukakan sebagai memenuhi keperluan bagi  
mendapatkan Ijazah Master Sains di Fakulti Sains dan  
Pengajian Alam Sekitar, Universiti Putra Malaysia

November 1997



**Dengan Nama Allah Yang Maha Pemurah Lagi Maha Mengasihani**



## PENGHARGAAN

Penulis ingin merakamkan setinggi-tinggi penghargaan kepada Jawatankuasa Penyeliaan yang dipengerusikan oleh Dr. Ali bin Mamat di atas bimbingan, tunjuk ajar dan nasihat serta galakan yang amat berguna sepanjang kajian ini dijalankan.

Penulis juga ingin merakamkan berbanyak terima kasih kepada ahli-ahli Jawatankuasa Penyeliaan lain yang terdiri daripada Dr. Md. Yazid bin Mohd Saman dan Dr. Hajjah Fatimah binti Ahmad di atas bimbingan dan nasihat mereka berdua.

Penulis merakamkan berbanyak terima kasih kepada Kerajaan Malaysia yang membiayai sepenuhnya semua perbelanjaan sepanjang penyelidikan ini dijalankan. Penulis juga ingin merakamkan sekalung kasih untuk keluarga tersayang di atas dorongan, pengorbanan dan kesabaran yang mereka berikan sepanjang kajian ini, terutamanya kepada ibu Hajjah Chik bt Jaafar, isteri Roslina bt Mohd Yussof, anak-anak Mohd Khairul Ikhwan, Mohd Khairul Anwar, Mohd Hamiz dan Nur Ayuni. Tidak ketinggalan juga buat rakan-rakan seperjuangan Khairuddin dan Zurni.

Akhir sekali, penulis mengucapkan ribuan terima kasih kepada semua yang terlibat secara langsung atau tidak dalam penghasilan tesis ini dan berdoa moga-moga Allah sahaja yang akan membalasnya dan dirahmati Allah di dunia dan akhirat, Amin.



## KANDUNGAN

### Muka Surat

PENGHARGAAN.....	ii
SENARAI JADUAL.....	vi
SENARAI RAJAH.....	vii
SENARAI NAMA SINGKATAN.....	x
ABSTRAK.....	xi
ABSTRACT.....	xiii

### BAB

I	PENDAHULUAN.....	1
	Pengenalan .....	1
	Isu-isu Penyelidikan Di Dalam PDT .....	7
	Latar Belakang Masalah.....	12
	Objektif Kajian.....	15
	Skop Kajian.....	16
	Manfaat Daripada Kajian.....	16
	Organisasi Tesis.....	17
II	SOROTAN LITERATUR.....	18
	Pengenalan.....	18
	Teknologi Pangkalan Data.....	18
	Seni bina SPPD.....	19
	Komponen-komponen SPPD.....	21
	Rangkaian Komputer.....	22
	Seni bina Rangkaian.....	27
	Pangkalan Data Teragih.....	30
	Pemprosesan Selari .....	35
	Kajian Lepas Yang Berhubungan .....	37
	Kesimpulan.....	39
III	PENGOPTIMUMAN PERTANYAAN PANGKALAN DATA TERAGIH (PDT).....	41
	Pendahuluan.....	41
	Pemprosesan Pertanyaan.....	42
	Pemecahan Pertanyaan.....	44
	Penormalan.....	44
	Analisa.....	46
	Membuang Kes Bertindih.....	50
	Menjana Semula Pertanyaan Aljabar.....	52
	Penempatan Data Pertanyaan Pada PDT.....	56



	Pengoptimuman Pertanyaan Sejagat.....	61
	Model Kos.....	62
	Statistik PDT.....	64
	Saiz Hubungan Sementara Yang Dihasilkan.....	66
	Turutan Cantum Di Dalam Fragmen.....	69
	Cantum Berbanding Dengan Semicantum.....	74
	Pengoptimuman Pertanyaan Setempat.....	76
	Kesimpulan.....	76
IV	METODOLOGI.....	78
	Pengenalan .....	78
	Kepentingan Pengoptimuman Pertanyaan.....	79
	Kaedah-kaedah Pengoptimuman Pertanyaan Teragih.....	81
	Kaedah <i>Hill-Climbing</i> .....	81
	Kaedah SDD-1 .....	85
	Kaedah Sistem R* .....	90
	Kaedah AHY (Apers-Henver-Yao) .....	94
	Pelaksanaan Kajian .....	96
	Pangkalan Data .....	96
	Pertanyaan.....	101
	Pengambilan Data.....	103
	Kesimpulan .....	108
V	HASIL DAN PERBINCANGAN.....	109
	Pengenalan .....	109
	Analisis Perbandingan .....	110
	Pengukuran Kos Keseluruhan Bagi Pertanyaan Ringkas....	111
	Pengukuran Kos Keseluruhan Bagi Pertanyaan Kompleks..	113
	Perbandingan Masa Gerak Balas Bagi Pertanyaan Ringkas Menggunakan Satu Pemproses Dan Multipemproses.....	115
	Perbandingan Masa Gerak Balas Bagi Pertanyaan Kompleks Menggunakan Satu Pemproses Dan Multipemproses.....	119
VI	KESIMPULAN DAN CADANGAN.....	123
	Kesimpulan.....	123
	Cadangan Penyelidikan Di Masa Hadapan.....	126

BIBLIOGRAFI .....	129
LAMPIRAN	
A    Algoritma Kaedah-Kaedah Pertanyaan.....	135
B    Kod Pseudo Aturcara-Aturcara .....	138
C    Senarai Aturcara-Aturcara Sumber.....	151
BIODATA.....	177



## SENARAI JADUAL

Jadual		Muka Surat
1	Jadual Perbandingan Kaedah-kaedah Pertanyaan.....	13
2	Bilangan Rekod Dan Kos Mengikut Hubungan .....	83
3	Pengiraan Faktor Pilihan Dan Kos Keseluruhan.....	88
4	Saiz Hubungan Baru Selepas Pengoptimuman Pertanyaan..	90
5	Hubungan Pekerja (nop, nama, tugas).....	98
6	Hubungan Tempoh (nop, noj, khid, jangka_masa).....	99
7	Hubungan Projek (noj, namaj, budget).....	99
8	Hubungan Gaji (khid, gaji).....	100
9	Hubungan Elaun (nop, elaun).....	100
10	Jadual Senarai Nama dan Fungsi Aturcara-Aturcara.....	105
11	Pecahan Bilangan Rekod Mengikut Hubungan.....	110
12	Hasil Kajian Kos Keseluruhan Bagi Pertanyaan Ringkas (P12).....	112
13	Hasil Kajian Kos Keseluruhan Bagi Pertanyaan Kompleks (P13).....	114
14	Hasil Kajian Masa Gerak Balas Pertanyaan Ringkas (P12) Menggunakan Satu Pemproses.....	116
15	Hasil Kajian Masa Gerak Balas Pertanyaan Ringkas (P12) Menggunakan Multipemproses.....	118
16	Hasil Kajian Masa Gerak Balas Pertanyaan Kompleks (P13) Menggunakan Satu Pemproses.....	120
17	Hasil Kajian Masa Gerak Balas Pertanyaan Kompleks (P13) Menggunakan Multipemproses	122



## SENARAI RAJAH

Rajah		Muka Surat
1	Konsep Pangkalan Data Teragih .....	3
2	Hubungan Antara Isu-isu Penyelidikan.....	11
3	Rajah Tiga Paras Seni bina ANSI-SPARC .....	20
4	Rangkaian Konsepsi Nod-nod Dan PSE.....	24
5	Topologi Cincin .....	25
6	Topologi Pokok .....	25
7	Topologi Sambungan Sepenuhnya .....	26
8	Topologi Bintang .....	26
9	Topologi Tak Tersusun.....	27
10	Rajah Seni bina Tujuh-lapisan ISO/OSI .....	28
11	Rajah Seni bina SPPDT Homogen.....	32
12	Rajah Seni bina Skema SPPDT Homogen.....	34
13	Fragmentasi Hubungan R.....	35
14	Konfigurasi Sistem Multipemproses Dengan Ingatan Utama Berkongsi.....	36
15	Konfigurasi Sistem Multipemproses Dengan Ingatan Utama Tersendiri .....	37
16	Empat Aras Pemprosesan Pertanyaan Teragih.....	43
17	Graf Perkaitan Bagi Ungkapan Kalkulus P3.....	48
18	Graf Perkaitan Kalkulus P4 Yang Mempunyai Ralat.....	50
19	Rajah Pokok Hubungan Aljabar Sebelum Penjanaan.....	54

20	Penulisan Semula Hubungan Pokok Aljabar.....	55
21	Pertanyaan Generik Pada Proses Pilihan.....	59
22	Pertanyaan Terturun Pada Proses Pilihan.....	59
23	Pertanyaan Generik Pada Proses Cantum.....	60
24	Pertanyaan Terturun Pada Proses Cantum.....	61
25	Perbandingan Kos Keseluruhan dan Masa Gerak Balas.....	64
26	Turutan Pemindahan Rekod Pada Operasi Cantum.....	69
27	Graf Perkaitan Operasi Cantum (Join) .....	70
28	Pendekatan Cantum.....	74
29	Pendekatan Semicantum .....	75
30	Graf Perkaitan Bagi Pertanyaan P11.....	92
31	Reka Bentuk Sistem PDT Yang Diwujudkan Untuk Kajian	97
32	Rajah Perkaitan Hubungan .....	101
33	Graf Bagi Pertanyaan Ringkas (P12) Bagi Kos Keseluruhan Lawan Bilangan Rekod.....	112
34	Graf Bagi Pertanyaan Kompleks (P13) Bagi Kos Keseluruhan Lawan Bilangan Rekod.....	114
35	Graf Bagi Pertanyaan Ringkas (P12) Bagi Masa Gerak Balas Lawan Bilangan Rekod Menggunakan Satu Pemproses.....	117
36	Graf Bagi Pertanyaan Ringkas (P12) Bagi Masa Gerak Balas Lawan Bilangan Rekod Menggunakan Multipemproses.....	118

37	Graf Bagi Pertanyaan Kompleks (P13) Bagi Masa Gerak Balas Lawan Bilangan Rekod Menggunakan Satu Pemproses.....	121
38	Graf Bagi Pertanyaan Kompleks (P13) Bagi Masa Gerak Balas Lawan Bilangan Rekod Menggunakan Multipemproses.....	122



## SENARAI NAMA SINGKATAN

<b>AHY</b>	Apers, Henver dan Yao
<b>ANSI</b>	American National Standard for Information system (Sistem Maklumat Piawaian Kebangsaan Amerika)
<b>ASCII</b>	American national Standard Code for Information Interchange (Kod Piawaian Kebangsaan Amerika untuk Saling Tukar Maklumat)
<b>DDBMS</b>	Distributed Database Management System (Sistem Pengurusan Pangkalan Data Teragih)
<b>DBMS</b>	Database Management System (Sistem Pengurusan Pangkalan Data)
<b>FP</b>	Faktor Pilihan
<b>ISO</b>	International Organisation for Standardization (Organisasi Piawaian Antarabangsa)
<b>ISO/OSI</b>	International Standard Organisation/Open System Interconnection reference model (Organisasi Piawaian Antarabangsa / Model Rujukan Saling Berhubungan Sistem Terbuka)
<b>LAN</b>	Local Area Network (Rangkaian Kawasan Setempat)
<b>PDT</b>	Pangkalan Data Teragih
<b>PSE</b>	Packets Switch Exchange (Pertukaran Suis Paket)
<b>SDD</b>	System for Distributed Databases (Sistem untuk Pangkalan Data Teragih)
<b>SPMPD</b>	Sistem Pengurusan Multi-Pangkalan Data
<b>SPPD</b>	Sistem Pengurusan Pangkalan Data
<b>SPPDT</b>	Sistem Pengurusan Pangkalan Data Teragih
<b>SQL</b>	Structured Query Language (Bahasa Pertanyaan Berstruktur)
<b>WAN</b>	Wide Area Network (Rangkaian Kawasan Luas)



Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk mendapatkan ijazah Master Sains.

## **PRESTASI KAEDAH PENILAIAN PERTANYAAN DALAM PANGKALAN DATA TERAGIH**

Oleh

**MOHD SHAHIDAN BIN ABDULLAH**

November 1997

Pengerusi : Dr. Ali bin Mamat

Fakulti : Sains dan Pengajian Alam Sekitar

Tesis ini membincangkan tentang kajian kaedah-kaedah penilaian pertanyaan yang dipilih di dalam pangkalan data teragih (PDT). PDT adalah gabungan dua konsep teknologi iaitu pangkalan data dan sistem rangkaian komunikasi. Pertanyaan (query) yang diberi kepada PDT adalah dalam bentuk ungkapan kalkulus yang akan diterima oleh pemproses pertanyaan (query processor) dan diterjemahkan melalui empat aras pemprosesan iaitu pemecahan pertanyaan, penempatan data, pengoptimuman pertanyaan dan pengoptimuman setempat. Empat kaedah penilaian pertanyaan yang dipilih di dalam kajian ini ialah *Hill-Climbing*, SDD-1, AHY dan Sistem R\*. Setiap kaedah mempunyai penyelesaian tersendiri yang melibatkan



operasi pilihan (selection), unjuran (projection), cantum (join) dan semicantum (semijoin).

Kajian dilakukan dengan menggunakan lima hubungan (relation) iaitu PEKERJA, PROJEK, TEMPOH, GAJI dan ELAUN serta satu fail katalog. Data ujian dimasukkan ke dalam setiap hubungan. Dua jenis pertanyaan digunakan iaitu pertanyaan ringkas dan pertanyaan kompleks. Setiap pertanyaan itu akan dioptimumkan mengikut empat kaedah penilaian yang dipilih dan bahasa pengaturcaraan C digunakan untuk menjana aturcara mengikut hasil pendekatan kaedah penilaian pertanyaan masing-masing. Prestasi setiap kaedah diukur berdasarkan kos keseluruhan (total cost) yang diukur dalam unit bait. Perbandingan prestasi setiap kaedah dari segi masa gerak balas menggunakan mesin multipemproses (multiprocessors) dan mesin satu pemproses juga dilakukan. Dengan itu perbezaan prestasi masa gerak balas dapat diukur pencapaiannya antara mesin satu pemproses dan mesin multipemproses.

Hasil daripada kajian ini memberi panduan tentang memilih kaedah penilaian pertanyaan untuk digunakan dalam pembangunan satu sistem pengurusan pangkalan data teragih.



Abstract of thesis submitted to the senate of Universiti Putra Malaysia in fulfilment of the requirements for the Degree of Master of Science.

**PERFORMANCE OF QUERY EVALUATION METHODS IN  
DISTRIBUTED DATABASES**

By

**MOHD SHAHIDAN BIN ABDULLAH**

November 1997

Chairman : Dr. Ali bin Mamat

Faculty : Science and Environmental Studies

This thesis reports a study on the selected existing query evaluation methods in a distributed database system. A distributed database system is a combination of database and communication network system. A query processor receives queries in the form of relational calculus and then breaks it through four layers of processing. The four layers are query decomposition, data localization, query optimization and local optimization. Four distributed query evaluation methods selected in the research are Hill-Climbing, SDD-1, AHY and System R\*. Every method has its own solution involving selection, projection, join and semijoin.



In evaluating the queries, five relations and one catalogue file are used. These relations are PEKERJA, PROJEK, TEMPOH, GAJI and ELAUN. Two types of query that is simple and complex are considered in the experiment and each query is executed using the four methods mentioned before one at a time. The execution of the query is implemented in C language. The performance of each method is measured based on the total cost in byte. The performance in terms of response times of each method implemented on one processor and multiprocessor machines is also compared.

The result from this study will help users to choose appropriate query evaluation methods in developing a distributed database management system.





# **BAB I**

## **PENDAHULUAN**

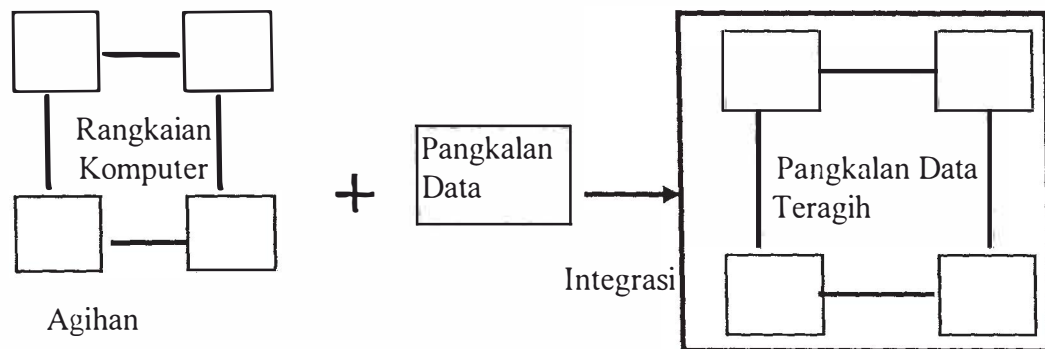
### **Pengenalan**

Pengurusan data adalah aktiviti paling utama pada sesebuah organisasi yang menggunakan komputer di dalam pengurusannya. Pada peringkat awal penggunaan komputer, kaedah lama (primitive) iaitu menggunakan sistem pengurusan fail digunakan untuk menguruskan data. Kaedah ini telah menimbulkan banyak kerumitan iaitu dari segi membangun, mengubahsuai dan menyelenggarakan sistem pengurusan data. Untuk menyelesaikan masalah-masalah itu pangkalan data diperkenalkan, dengan semua fail yang terdapat pada sesebuah organisasi disepadukan bagi membentuk satu pangkalan data. Pangkalan data yang dibentuk ini sangat penting bagi organisasi. Pangkalan data ini disimpan di dalam satu sistem komputer utama dan dipanggil pangkalan data terpusat. Berbagai-bagai aplikasi dibangunkan berasaskan kepada pangkalan data tersebut. Walau bagaimanapun keperluan kepada pengurusan maklumat yang lebih cekap dan berkesan oleh sesebuah organisasi telah mendorong pangkalan data terpusat yang terhad

kepada lokasi tertentu, dipecahkan kepada beberapa unit pangkalan data dan setiap satunya disimpan di dalam komputer berasingan. Ini membawa kepada konsep sistem pangkalan data teragih (Bell dan Grimson, 1992).

Kemajuan pesat teknologi rangkaian komputer dan pangkalan data telah membolehkan wujudnya pangkalan data teragih (PDT). Tempat penyimpanan tidak terhad kepada satu tempat dan satu komputer sahaja, seperti mana yang berlaku pada pangkalan data terpusat, tetapi ianya boleh teragih lebih daripada satu tempat dan komputer. Komputer pula boleh terdiri dari jenis mesin yang sama (homogenous) atau jenis mesin yang berbeza (heterogenous) (Ceri dan Pelagatti, 1984). Teknologi pangkalan data teragih (Distributed Database) melibatkan gabungan dua komponen utama iaitu pangkalan data dan sistem rangkaian seperti ditunjukkan dalam Rajah 1 (Bell dan Grimson, 1992). Satu daripada ciri terpenting di dalam sistem teragih ini ialah setiap pangkalan data setempat (local database) adalah bersifat autonomi dan setiap pangkalan data itu dihubungkan dengan rangkaian komputer.

Satu unsur yang sangat penting di dalam penggunaan pangkalan data adalah pencapaian maklumat. Maklumat yang hendak dicari diungkapkan sebagai satu pertanyaan. Pemproses pertanyaan akan melaksanakan pertanyaan dan mengembalikan jawapan kepada pengguna. Dalam PDT apabila sesuatu pertanyaan diberikan oleh pengguna yang kebiasaannya diungkapkan dalam bentuk pertanyaan berasaskan hubungan kalkulus (relational calculus), pemproses pertanyaan akan



**Rajah 1: Konsep Pangkalan Data Teragih**

memecahkan pertanyaan tersebut kepada ungkapan hubungan aljabar (relational algebra) (Date, 1995). Selanjutnya ungkapan hubungan aljabar itu akan dilaksanakan. Oleh kerana PDT dipecahkan mengikut tapak-tapak, maka terdapat beberapa strategi pelaksanaan ungkapan hubungan aljabar. Masalahnya ialah bagaimana memilih strategi yang terbaik supaya pelaksanaan pertanyaan yang optimum dapat diperolehi. Semakin banyak pangkalan data terlibat, semakin banyak kemungkinan strategi pelaksanaan yang boleh dijanakan. Oleh yang demikian proses pengoptimuman pertanyaan (Sacco dan Yao, 1982) adalah sangat penting di dalam mendapatkan penyelesaian yang cekap dan berkesan. Beberapa kaedah pelaksanaan pertanyaan dalam PDT telah diperkenalkan seperti *Hill-Climbing* (Wong, 1977), Sistem R\* (Daniels dan Ng, 1982), SDD-1 (Bernstein dan Wong, 1981) dan AHY (Apers et al., 1983).

Kenapa perlu beralih kepada pangkalan data teragih? Jawapan kepada masalah ini adalah (Litwin, 1988):

- (i) Pada PDT, data diletakkan mengikut taburan organisasi. Dengan itu membenarkan kawalan dan pengendalian data kepunyaan sesuatu sub-organisasi dilakukan secara tempatan (local). Dengan kawalan secara tempatan ini meningkatkan kebolehpercayaan dan keberkesanan perolehan data.
- (ii) PDT sangat sesuai dilaksanakan apabila bahagian-bahagian di dalam organisasi yang sedia ada mempunyai pangkalan data masing-masing dan terdapat keperluan untuk memproses maklumat secara global. Untuk ini PDT direka bentuk secara bawah ke atas (bottom-up) kepada pangkalan-pangkalan data tempatan yang tersedia ada.
- (iii) Bagi organisasi yang terus berkembang dan menampilkan sub-organisasi baru, maka PDT sekaligus menyediakan kemudahan bagi pertumbuhan tersebut. Pertumbuhan PDT sendiri lebih mudah dikendalikan berbanding jenis terpusat. Pertambahan bilangan data, bilangan proses dan perubahan struktur fail dapat diubahsuai di dalam sistem teragih dengan lebih mudah, tanpa terikat dengan tapak-tapak (sites) lain.
- (iv) Di dalam suasana PDT, kebanyakan aplikasinya adalah bersifat tempatan. Dengan itu, penggunaannya mengurangkan kos komunikasi. Ini berbeza dengan pangkalan data terpusat di dalam rangkaian, di mana semua aplikasinya adalah berjauhan (remote) dan

data diletakkan pada satu tempat sahaja. Di dalam PDT, rekod dipecahkan mengikut nod-nod di dalam komputer berdasarkan:

- a) kadar penggunaan data/fail;
  - b) masa gerak balas diperlukan bagi sesuatu rekod;
  - c) kos storan dan komunikasi; dan
  - d) keperluan kebolehpercayaan.
- (v) Dengan wujudnya pemproses-pemproses yang berautonomi antara satu sama lain, maka gerak balas yang optimum diperolehi dengan adanya suasana proses secara selari antara pemproses-pemproses tersebut.
- (vi) Pertindihan atau salinan kedua data di tapak yang berbeza di dalam sistem, menyediakan ketersediaan (availability) dan kebolehpercayaan (reliability). Jika satu daripada nod mengalami kerosakan, data masih boleh diperolehi di tapak yang lain. Walau bagaimanapun replikasi data ini harus dilaksanakan secara berhati-hati, kerana konsep kemaskini dan kawalan pemulihan (recovery control) menjadi lebih kompleks apabila wujudnya data yang sama berada di dua atau lebih tapak yang berbeza (Edelstein, 1995).

Keupayaan pangkalan data adalah untuk menyatu dan menguruskan data-data yang berhubungan bagi sesuatu organisasi. Untuk pangkalan data jenis terpusat semua data diletakkan pada satu tapak dan mesin yang sama. Walau bagaimanapun struktur sesebuah organisasi itu sebenarnya dipecahkan kepada bahagian-bahagian

mengikut fungsi masing-masing. Kedudukan bahagian-bahagian ini kadang-kadang terletak jauh antara satu dengan lain. Pengalaman menunjukkan lebih 80% (Bell dan Grimson, 1992) capaian data sesuatu sub-organisasi kepada pangkalan data adalah data berkaitan dengannya sahaja.

Aliran kepada penggunaan pangkalan data teragih selain daripada desakan dari pengguna, juga disebabkan oleh perkembangan teknologi yang membolehkan kemudahan rangkaian komputer dilaksanakan. Kebaikan utama diperolehi dari pangkalan data teragih adalah (Rothnie dan Goodman, 1977):

- (i) Membenarkan evolusi sistem, perubahan keperluan pengguna dan membenarkan autonomi setempat;
- (ii) Menyediakan seni bina sistem yang ringkas dan lentur (flexible); dan
- (iii) Memberikan masa gerak balas yang baik.

Terdapat beberapa kelemahan yang dikenal pasti pada PDT berbanding dengan pangkalan data terpusat iaitu (Rothnie dan Goodman, 1977):

- (i) Kos penyediaan dan penyelenggaraan PDT. Sistem teragih memerlukan keperluan perkakasan tambahan terutamanya di dalam penyediaan keperluan rangkaian komunikasi. Di samping itu keperluan kepada kakitangan komputer juga bertambah kerana bilangan komputer yang ada lebih dari satu dan berada di lokasi yang berbeza.

- (ii) Kekompleksan (complexity) pengendalian PDT adalah lebih rumit berbanding dengan pangkalan data terpusat. Masalah-masalah seperti mereka bentuk PDT, kawalan serentak (concurrency control) dan pengoptimuman pertanyaan adalah antara masalah-masalah yang dikenal pasti di dalam persekitaran PDT.

### **Isu-isu Penyelidikan Di Dalam PDT**

Beberapa isu atau masalah teknikal telah dikenal pasti di dalam mengoptimumkan penggunaan pangkalan data teragih. Untuk menghasilkan Sistem Pengurusan Pangkalan Data Teragih (SPPDT) yang baik, penggunaan rangkaian di dalam pemindahan data dan penghantaran mesej perlu diminimumkan. Untuk mencapai tujuan ini beberapa masalah telah dikenal pasti di dalam bidang-bidang berikut (Ozsu dan Valduriez, 1991; Mohd Shahidan et al., 1996a):

#### **1. Reka Bentuk PDT**

Persoalan yang dikemukakan adalah bagaimana memperuntukkan pangkalan data dan aplikasi yang menggunakannya di antara tapak-tapak (sites) yang ada. Terdapat dua penyelesaian asas kepada persoalan tersebut iaitu pecahan data (partitioned) dan penyalinan (replicated). Mengikut pecahan adalah di mana pangkalan data dipecahkan kepada beberapa pecahan atau bahagian dan setiap pecahan itu diletakkan

pada tapak-tapak yang berbeza. Penyalinan pula terbahagi dua iaitu pertama penyalinan penuh (fully replicated), dengan keseluruhan data dimuatkan pada semua tapak-tapak, dan kedua adalah penyalinan separa dengan sesuatu pecahan data diletakkan pada lebih dari satu tapak (Ceri et al., 1983).

## **2. Pemprosesan Pertanyaan Teragih (Distributed Query Processing)**

Pemproses pertanyaan (query processor) adalah berperanan menerima pertanyaan (query) dan menukarkan pertanyaan tersebut kepada beberapa siri operasi pelaksanaan bagi mendapatkan hasil pertanyaan. Masalah yang dikenal pasti adalah menentukan strategi pelaksanaan yang terbaik di dalam persekitaran sistem teragih. Beberapa kaedah pelaksanaan telah diperkenalkan bagi mendapatkan laluan operasi yang terbaik seperti AHY dan SDD-1.

## **3. Pengurusan Katalog Teragih**

Katalog (directory) adalah tempat bagi menyimpan maklumat mengenai perihalan (description) dan lokasi berkaitan dengan butir-butir data di dalam pangkalan data. Masalah berkaitan dengan pengurusan katalog teragih adalah sama dengan masalah meletakkan fragmen-fragmen data di dalam rangkaian pangkalan data teragih. Kandungan katalog terbahagi kepada dua kemungkinan, pertama global iaitu mengandungi maklumat keseluruhan dan kedua tempatan iaitu maklumat yang ada hanya pada sesuatu tapak tertentu sahaja. Ianya boleh dipusatkan pada satu tapak atau dipecahkan kepada beberapa tapak.



#### **4. Kawalan Serentak Teragih (Distributed Concurrency Control)**

Kawalan serentak melibatkan pengasingkronian capaian data kepada PDT untuk membolehkan integriti pangkalan data dikekalkan. Di dalam persekitaran PDT kawalan ini lebih rumit berbanding pangkalan data terpusat kerana semua salinan pangkalan data pada tapak-tapak berbeza perlu dikawal keserentakannya supaya keadaan saling konsisten (mutual consistency) antara pangkalan-pangkalan data dapat dikekalkan. Antara penyelesaian yang dicadangkan (Orlowska dan Reid, 1995):

- (i) Pesimistik iaitu pengasingkronian pelaksanaan permintaan pengguna sebelum pelaksanaan bermula; dan
- (ii) Optimistik iaitu melaksanakan permintaan dan kemudian memeriksa sama ada pelaksanaan akan mengekalkan serentak pangkalan data.

Di dalam kedua-dua kaedah di atas terdapat dua pendekatan yang digunakan iaitu pertama penguncian (locking), di mana berdasarkan saling eksklusif (mutual exclusion) capaian kepada butir data. Kedua adalah turutan (timestamping), di mana transaksi dilaksanakan mengikut turutan tertentu.

#### **5. Pengurusan Kebuntuan Teragih (Distributed Deadlock Management)**

Masalah kebuntuan di dalam PDT adalah sama tabiinya dengan yang dihadapi oleh sistem pengoperasian. Persaingan antara pengguna-pengguna untuk mencapai data