Jurnal Sistem Informasi (JSI), VOL. 1, NO. 1, April 2009, *Halaman 47-54*
ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index

ANALISIS OPTIMASI QUERY PADA DATA MINING

Ermatita
Jurusan Sistem Informasi
Fakultas Ilmu Komputer, Universitas Sriwijaya
E-mail: Ermatita@ilkom.unsri.ac.id

Abstrak

Data mining is currently being used by organizations or companies to dig up information from the data that the data has been collected in a time long enough and in large numbers. To process a large amount of data in need of a strategy that can make processing data in an effective and efficient. Query optimization is a strategy or the way that can be used in the initialization and data processing in order to get the results of queries that is effective and efficient. Query optimization approach to rule-based and cost-based that can generate the required information quickly and accurately.

Keywords: data mining, query optimization, Query Optimizer, information, rule-based, cost-based

I. PENDAHULUAN

1.1 Latar Belakang

Kemajuan zaman dan teknologi saat ini telah merubah pola dalam perusahaan atau pun organisasi-organisasi. Data-data yang besar dan telah terkumpul dalam waktu yang lama dapat di olah menjadi sumber informasi yang dapat membantu menganalisis eksistensi dari sebuah perusahaan atau organisasi. Analisa otomatis dari data yang berjumlah besar atau kompleks yang terbentuk sebagai data mining, dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya saat ini telah banyak di lakukan .

Pengolahan data tersebut dapat dilakukan dengan mengakses data yang terdapat dalam database. Pengaksesan data tersebut dilakukan dengan melakukan query-query pada basisdata- basisdata dengan database manajemen sistem.

Pengaksesan data pada database perlu memperhatikan ketepatgunaan implementasi dari data itu sendiri serta waktu prosesnya. Ada banyak cara yang dapat dilakukan oleh database manajemen sistem dalam memproses dan menghasilkan jawaban sebuah query. Semua cara pada akhirnya akan menghasilkan jawaban (output) yang sama tetapi pasti mempunyai harga yang berbeda-beda, seperti misalnya total waktu yang diperlukan untuk menjalankan sebuah query.

Optimisasi query mencoba memberikan suatu pemecahan untuk menangani masalah tersebut dengan cara menggabungkan sejumlah besar teknik-teknik dan strategi, yang meliputi transformasi-transformasi logika dari query-query untuk mengoptimisasi jalan akses dan penyimpanan data pada sistem file terutama pada data-data yang besar dan lama

ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index

tersimpan. Setelah ditransformasikan, sebuah query harus dipetakan ke dalam sebuah urut-urutan operasi untuk menghasilkan data-data yang diminta. Hal ini dapat membantu suatu organisasi dalam mengolah informasi dari data mining yang terkumpul dengan pengaksesan data yang optimal.

Beberapa penelititan di bidang optimasi query telah dilakukan oleh peneliti-peneliti. Antara lain Jarke (1984) telah melakukan penelitian optimasi query dalam sistem basis data. Dia mengatakan bahwa query optimasi dalam sistem data base sangat penting di evaluasi untuk menghasilkan hasil yang optimal dalam melakukan query.

Zheng, melakukan penelititan ptimasi query dalam dabase grid. Dia mengembangkan sebuah query optimizer dalam DartGrid II dan pendekatan optimasi query heuristic, dinamis dan paralel dalam database grid

1.2 Tujuan

Tujuan yang ingin dicapai dalam tulisan ini adalah :

- Mempelajari dan memahami teknik yang digunakan oleh sebuah Database Manajemen Sistem (DBMS) untuk memproses dan mengoptimisasi sebuah query.
- Mempelajari dan memahami operasi-operasi dasar yang digunakan untuk mengeksekusi query.
- Menganalisis proses query optimasi pada datamining .

1.3 Ruang Lingkup

Ruang lingkup yang membatasi permasalahan yang akan dibahas pada tulisan ini adalah :

- Proses optimisasi query pada sebuah datamining
- Pembahasan teknik-teknik yang digunakan pada proses optimisasi query dibatasi hanya pada dua teknik utamanya, yaitu *Heuristic Optimization* dan *Cost based optimization*.
- menganalisis optimisasi query yang dapat diterapkan pada datamining

II. BAHASAN UTAMA

2.1 Data Mining

Ada beberapa definisi dari data mining yang dikenal di buku-buku teks data mining. Diantaranya adalah :

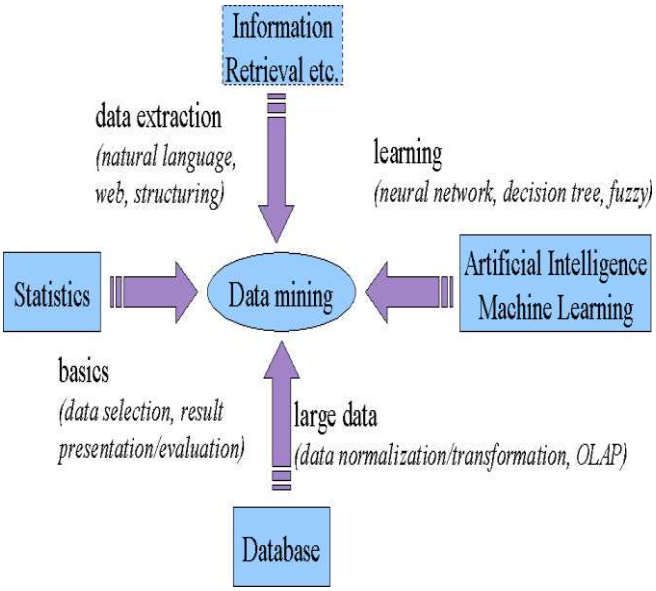
- Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.
- Data mining adalah analisa otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya

ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index

Menarik untuk diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Dari definisi-definisi itu, dapat dilihat ada beberapa faktor yang mendefinisikan data mining :

- data mining adalah proses otomatis terhadap data yang dikumpulkan di masa lalu
- objek dari data mining adalah data yang berjumlah besar atau kompleks
- tujuan dari data mining adalah menemukan hubungan-hubungan atau pola-pola yang mungkin memberikan indikasi yang bermanfaat

Data mining merupakan ilmu yang merupakan gabungan dari beberapa ilmu. Gambar 1 menunjukkan bahwa data mining memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistic, database dan juga information retrieval.



Gambar 1 Perspektif Datamining dari beberapa disiplin ilmu

Beberapa teknik yang sering disebut-sebut dalam literatur data mining seperti classification, neural network, genetic algorithm dll. sudah lama dikenal di dunia kecerdasan buatan. Statistik memberikan kontribusi pada data mining

Data mining digunakan untuk melakukan *information discovery* yang informasinya lebih ditujukan untuk seorang *Data Analyst* dan *Business Analyst* (dengan ditambah visualisasi tentunya). D*ata mining adalah* bidang ilmu yang diperkaya oleh beberapa ilmu, seperti: *information science* (ilmu informasi), *high performance computing*, visualisasi, *machine learning*, statistik, *neural networks* (jaringan syaraf tiruan), pemodelan matematika, *information retrieval* dan *information extraction* serta pengenalan pola. Bahkan pengolahan citra (*image processing*) juga digunakan dalam rangka melakukan *data mining* terhadap data *image*/*spatial*.

ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index

2.2 Optimasi Query

Teknik optimasi dapat dilakukan dengan beberapa cara. Terdapat 2 pendekatan optimasi yang umum digunakan sebagaimana diungkapkan oleh Chanowich (2001), yakni:

a. Heuristik atau *rule-based*
   Teknik ini mengaplikasikan aturan heuristik untuk mempercepat proses *query*. Optimasi jenis ini mentransformasikan query dengan sejumlah aturan yang akan meningkatkan kinerja eksekusi, yakni:
   - melakukan operasi *selection* di awal untuk mereduksi jumlah baris
   - melakukan operasi projection di awal untuk mereduksi jumlah atribut
   - mengkonversikan query dengan banyak join menjadi query dengan banyak subquery
   - melakukan operasi selection dan join yang paling kecil keluarannya sebelum operasi lain

b. *Cost-based*
   Teknik ini mengoptimasikan cost yang dipergunakan dari beberapa alternatif untuk kemudian dipilih salah satu yang menjadi cost terendah. Teknik ini mengoptimalkan urutan join terbalik yang dimungkinkan pada relasi-relasi r1 $\rightarrow$ r2 $\rightarrow$ ... rn. Teknik ini dipergunakan untuk mendapatkan pohon left-deep join yang akan menghasilkan sebuah relasi sebenarnya pada node sebelah kanan yang bukan hasil dari sebuah intermediate join.

   Data hasil query lebih dari kurang dari atau sama dengan 5165 record, penggunaan query dengan cross product membutuhkan waktu yang lebih kecil dibanding dengan metode subset query. Sebaliknya untuk data lebih besar dari 5165 record, penggunaan subset query jauh lebih cepat disbanding dengan menggunakan cross product.

   Desain aplikasi saja tidak cukup untuk meningkatkan unjuk kerja harus didukung dengan optimasi dari perintah SQL yang digunakan pada aplikasi tersebut. Dalam mendesain database, seringkali lokasi fisik data tidak menjadi perhatian penting. Karena hanya desain logik saja yang diperhatikan. Padahal untuk menampilkan hasil query dibutuhkan pencarian yang melibatkan struktur fisik penyimpanan data. Inti dari optimasi query adalah meminimalkan "jalur" pencarian untuk menemukan data yang disimpan dalam lokasi fisik. Index pada database digunakan untuk meningkatkan kecepatan akses data. Pada saat query dijalankan, index mencari data dan menentukan nilai ROWID yang membantu menemukan lokasi data secara fisik di disk. Akan tetapi penggunaan index yang tidak tepat, tidak akan meningkatkan unjuk kerja dalam hal ini kecepatan akses data.

   Proses akses data akan lebih cepat jika data terletak pada block tabel yang berdekatan daripada harus mencari di beberapa datafile yang terletak pada block yang berbeda.

ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index

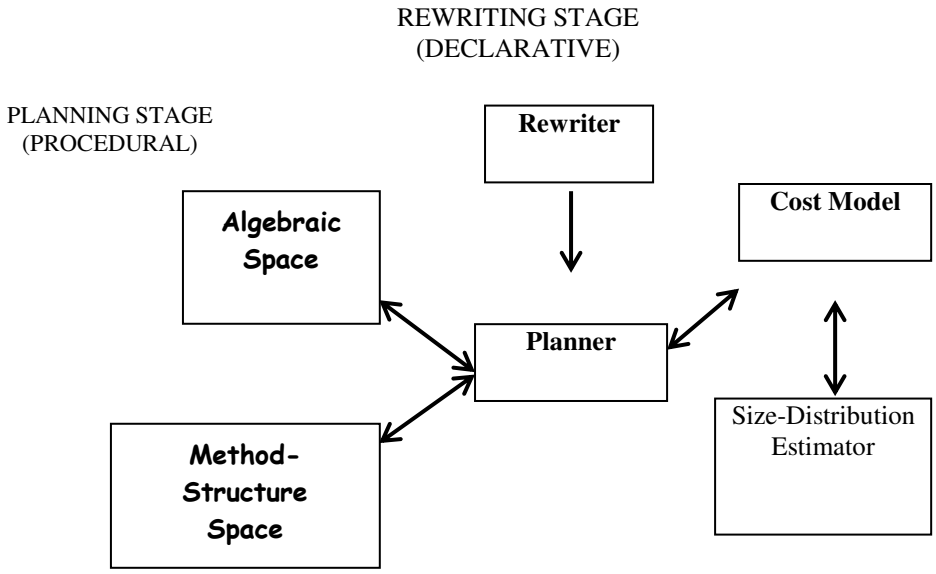
2.3 Query Optimizer

Query optimizer adalah bagian dari DBMS yang berfungsi mengoptimasi query.

Proses yang biasanya terjadi dalam optimizer adalah optimizer memeriksa semua ekspresi-ekspresi aljabar yang sama yang diberikan query dan memilih salah satunya yang memiliki harga taksiran paling rendah. Tugas dari optimizer adalah untuk mentransformasikan inisial ekspresi query ke dalam sebuah rencana evaluasi yang menghasilkan record yang sama.

Keuntungan dari optimizer adalah dapat mengakses semua informasi statistik dari sebuah database. Selain itu optimizer juga dapat dengan mudah untuk melakukan optimisasi kembali apabila informasi statistik sebuah database berubah dan optimizer dapat menangani strategi yang berbeda-beda dalam jumlah besar yang tidak mungkin dilakukan oleh manusia.

Input dari optimizer adalah sebuah tree yang sudah mengalami proses parsing di dalam query parser. Tree tersebut biasanya disebut dengan *parse tree*. Sedangkan output dari optimizer adalah berupa rencana eksekusi (execution plan) yang siap untuk dikirimkan ke dalam query kode generator dan query processor untuk diproses untuk mendapatkan hasil akhir dari query tersebut.



Gambar 2.Arsitektur Umum Query Optimizer

Proses optimisasi query dapat dianggap mempunyai dua tingkatan. Dua tingkatan tersebut adalah : *rewriting* dan *planning*. Hanya ada satu modul pada tingkat pertama yaitu *Rewriter*, dimana semua modul-modul lainnya berada pada tingkat kedua. Tahap penulisan dapat disebut sebagai level *declarative*, sedangkan tahap perencanaan dapat juga disebut sebagai level *procedural*.

Fungsi-fungsi dari masing-masing modul pada gambar 3.2 akan dijelaskan secara lebih rinci :

ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index

Rewriter : Modul ini melakukan transformasi-transformasi untuk sebuah parse tree dari query yang diberikan dan menghasilkan query-query yang sama yang diharapkan lebih efisien.

Planner : Modul ini adalah modul utama yang menguji semua rencana-rencana eksekusi query yang dihasilkan pada tingkat sebelumnya dan memilih satu dari semua rencana yang termurah, yang akan digunakan untuk menghasilkan jawaban dari query yang asli. Planner menggunakan *search strategy*, yang memeriksa *space* (tempat) dari rencana-rencana eksekusi. Space ini ditentukan oleh dua modul lainnya dari optimizer, yaitu *Algebraic Space* dan *Method-Structure Space*. Untuk kebanyakan bagian, dua modul ini dan search strategy menentukan harga seperti *running time* dari optimizer itu sendiri yang seharusnya serendah mungkin. Rencana-rencana eksekusi yang diperiksa oleh planner dibandingkan berdasarkan perkiraan-perkiraan harganya dan dipilih yang perkiraan harganya.

2.4 FAKTOR LAIN YANG BERPENGARUH TERHADAP KECEPATAN AKSES DATA

Faktor lain yang berpengaruh terhadap kecepatan akses data, tidak hanya terletak pada optimasi perintah SQL, tapi terhadap hal-hal lain yang berpengaruh. Diantaranya adalah optimasi aplikasi dan penggunaan cluster dan index.

2.4.1 OPTIMASI APLIKASI

Pembuatan aplikasi, perlu memperhatikan apakah akses terhadap data sudah efisien. Efisien dalam hal penggunaan obyek yang mendukung kecepatan akses, seperti index atau cluster. Kemudian juga bagaimana cara database didesain. Apakah desain database sudah melakukan normalisasi data secara tepat. Kadangkala normalisasi sampai level yang kesekian, tidak menjamin suatu desain yang efisien. Untuk membuat desain yang lebih tepat, kadang setelah melakukan normalisasi perlu dilakukan denormalisasi. Misalnya tabel yang hubungannya one-to-one dan sering diakses bersama lebih baik disatukan dalam satu tabel.

2.4.2 CLUSTER DAN INDEX

Cluster adalah suatu segment yang menyimpan data dari tabel yang berbeda dalam suatu struktur fisik disk yang berdekatan. Konfigurasi ini bermanfaat untuk akses data dari beberapa tabel yang sering di-query. Penggunaan cluster secara tepat dilaksanakan setelah menganalisa tabel-tabel mana saja yang sering di-query secara bersamaan menggunaan perintah SQL join. Jika aplikasi sering melakukan query dengan menggunakan suatu kolom yang berada pada klausa WHERE, maka harus digunakan index yang melibatkan kolom tersebut. Penggunaan index yang tepat bergantung pada jenis nilai yang terdapat dalam kolom yang akan diindex. Dalam RDBMS Oracle, index B-Tree digunakan untuk kolom yang mengandung nilai yang cukup bervariasi, sedangkan untuk nilai yang tidak memiliki variasi cukup banyak, lebih baik menggunakan index bitmap.

ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index

2.5 BAHASAN

Optimisasi Query adalah suatu proses untuk menganalisa query untuk menentukan sumber-sumber apa saja yang digunakan oleh query tersebut dan apakah penggunaan dari sumber tersebut dapat dikurangi tanpa merubah output. Atau bisa juga dikatakan bahwa optimisasi query adalah sebuah prosedur untuk meningkatkan strategi evaluasi dari suatu query untuk membuat evaluasi tersebut menjadi lebih efektif. Optimisasi query mencakup beberapa teknik seperti transformasi query ke dalam bentuk logika yang sama, memilih jalan akses yang optimal dan mengoptimumkan penyimpanan data.

Optimisasi query merupakan bagian dasar dari sebuah sistem database dan juga merupakan suatu proses untuk menghasilkan akses yang efisien dari sebuah query di dalam sebuah database. Secara tidak langsung, sebuah rencana akses merupakan sebuah strategi yang nantinya akan dijalankan untuk sebuah query, untuk mendapatkan kembali operasi-operasi yang apabila dijalankan akan menghasilkan database record query. Ada tiga aspek dasar yang ditetapkan dan mempengaruhi optimisasi query, yaitu : *search space, cost model* dan *search strategy*.

Optimasi query perlu direncanakan dan dijalankan dengan baik. Baik itu terhadap database ataupun data mining. Basis data berperanan penting pada *data mining* karena *data mining* mengakses data yang ukurannya besar (bisa sampai terabyte) dan disini terlihat peran penting *database* terutama dalam optimisasi *query*-nya.

Pendahuluan perkembangan data mining yang pesat tidak dapat lepas dari perkembangan teknologi informasi yang memungkinkan data dalam jumlah besar terakumulasi. Sebagai contoh, toko swalayan merekam setiap penjualan barang dengan memakai alat POS(point of sales). Database data penjualan tsb. bisa mencapai beberapa GB setiap harinya untuk sebuah jaringan toko swalayan berskala nasional. Perkembangan internet juga punya andil cukup besar dalam akumulasi data. Investasi yang besar di bidang IT untuk mengumpulkan data berskala besar ini perlu dijustifikasi dengan dengan didapatnya nilai tambah dari kumpulan data ini.,

Dalam system basis data relasional sering terjadi dibutuhkan pengaksesan data terhadap beberapa tabel sekaligus. Ada beberapa cara dalam melakukan akses ini, namun perlu diperhatikan cara mana yang paling optimal sehingga diperoleh kecepatan akses tertinggi. Disini query optimizer akan dapat dimanfaatkan untuk mendapatkan hasil query yang optimal, yang efektif dan efisien. Dengan menganalisis proses query agar dapat optimal , maka optimasi query perlu dilakukan untuk mengolah data dalam data mining yang mengandung jumlah data yang besar, sehingga proses pencarian informasi dapat dilakukan dengan efektif dan efisien dan dapat membantu menganalisa informasi yang didapat dari data mining tersebut.

ISSN Print : 2085-1588
ISSN Online : 2355-4614
http://ejournal.unsri.ac.id/index.php/jsi/index

DAFTAR PUSTAKA

Chanowich, E. dan Sendelbach, E., 2001, "Query Optimization*" in CSE 498G – Advanced Database.*

Korth, H.F., dan Silberschatz, A., 1991, *Database System Concepts,* McGraw Hill, Singapura.

Setiawan, M.A., 2004, *Optimasi SQL Query untuk Informasi Retrievel pada Aplikasi Berbasis Web*, Proceedings Seminar Nasional Aplikasi Teknologi Informasi UII, Yogyakarta

Yanis,E Loannidis, 2004, Query Optimization, University of Wisconsin.

Jarke,M dan Koch,J, 1984, Query Optimization in Database System, Computing Survey, vol.16, No 2, Juni 1984

Zheng, X, Chen,H, Wu,Z and Mao,Y, Query Optimization in Database Grid, Grid Computing Lab, College of Computer Science, Zhejiang University, Hangzhou, 310027, China