

ANALISIS PERFORMANSI COLUMN-STORES DAN ROW-STORES PADA DBMS

Fatar¹, Kiki Maulana², Kusuma Ayu Laksitowening³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Performansi DBMS merupakan salah satu hal yang dijadikan pertimbangan dalam pemilihan sebuah DBMS. Untuk menjawab tantangan akan performansi tersebut, munculah sebuah paradigma arsitektur database baru yang disebut dengan column-stores database. Dengan paradigma baru ini, diharapkan permasalahan yang terkait dengan performansi DBMS dapat diselesaikan terutama dalam hal pengaksesan data.

Sekarang ini, terdapat dua cara untuk memetakan relasional tabel database ke dalam media penyimpanan yaitu menyimpan tabel row-by-row atau menyimpannya column-by-column. Kebanyakan DBMS konvensional mengimplementasikan arsitektur row-stores yang menyimpan tabel row-by-row karena sering digunakan di transactional application. Bagaimanapun terdapat sekumpulan aplikasi dimana arsitektur row-stores tidak berjalan efisien. Aplikasi ini biasanya lebih mengarah ke analytical application dimana query atau akses data menjadi operasi utama dalam rangka membuat perencanaan dan keputusan.

Melalui karakteristiknya yang menyimpan data per atribut di dalam media penyimpanan menjadikan arsitektur column-stores lebih efektif dan efisien dalam pengambilan data karena hanya atribut yang dibutuhkan saja yang akan diakses.

Kata Kunci : column-stores database, row-stores database, performansi, DBMS

Abstract

DBMS performance is one of many thing that have to be considered in choosing a DBMS. To fullfill the performance challenge, a new architecture paradigm which is called column-stores database arise. With this new paradigm, hopefully, the problem of DBMS performance especially in acces data will be solve.

There are two obvious ways to map a relational database table onto a storage interface: store the table row-by-row, or store the table column-by-column. Almost all conventional DBMS implement row-stores architecture because it often use on the most common application especially in transactional application. However, there are a set of emerging applications for database systems for which the row-by-row layout performs poorly. These applications are more analytical in nature, whose goal is to read through the data to gain new insight and use it to drive planning and decision making.

With it characteristic storing data by column rather than by row in storage so that column-stores can be more effective and efficient in query perform.

Keywords : column-stores database, row-stores database, performance, DBMS

1 Pendahuluan

1.1 Latar belakang

Selama dekade terakhir ini, hampir semua sektor telah sangat bergantung pada teknologi informasi. Jumlah data yang disimpan dalam sistem informasi tersebut secara kontinyu terus membesar. Seiring dengan semakin besarnya jumlah data munculah kebutuhan akan pengaksesan data yang cepat. Pengaksesan data yang cepat akan sangat berguna dalam rangka analisis, pengambilan keputusan, dan hal-hal penting lainnya. Hal-hal tersebut sangat penting bagi suatu organisasi agar dapat bertahan di tengah-tengah persaingan yang semakin ketat dewasa ini. Untuk mendukung sistem informasi di atas data yang semakin besar jumlahnya, *database management system* (DBMS) merupakan jawaban yang paling tepat dalam menjawab persoalan kecepatan akses data.

Untuk menyimpan data yang besar jumlahnya di DBMS dibutuhkan tabel dengan jumlah kolom dan baris yang besar. Baris merepresentasikan jumlah data, sedangkan kolom merepresentasikan detail informasi dari sebuah data. Pada saat *retrieve*, data akan dibaca dari tabel, dan pada kebanyakan proses *retrieve* tidak semua kolom yang ada dibutuhkan. Walaupun tidak semua kolom diambil akan tetapi pada pemrosesannya semua kolom akan dibaca (*full table scan*). Hal tersebut tentunya tidak efektif sehingga dibutuhkan sebuah cara untuk mengatasi persoalan tersebut. Sebuah cara diperkenalkan yaitu dengan memodifikasi arsitektur penyimpanannya melalui mekanisme *column-stores*.

Kebanyakan DBMS konvensional mengimplementasikan *row-stores storage system* dalam penyimpanan datanya, dimana semua atribut dari sebuah baris ditempatkan secara berdampingan dalam media penyimpanan data, sedangkan pada arsitektur *column-stores* yang akan disimpan adalah nilai dari tiap atribut, dan terpisah tiap atributnya. Dewasa ini, dengan semakin berkembang dan besarnya kebutuhan akan data dalam aplikasi-aplikasi besar, maka dibutuhkan pula tingkat pengaksesan yang cepat untuk pengambilan keputusan dan analisis. Dengan arsitektur *column-stores*, DBMS hanya perlu membaca nilai dari kolom yang dibutuhkan untuk memproses query yang diberikan, sehingga dapat menghindari membawa atribut-atribut yang *irrelevant* kedalam memori untuk diproses, seperti yang jamak terjadi pada arsitektur *row-stores* DBMS. Hal tersebut tentunya akan meningkatkan performansi dari DBMS tersebut. Dengan meningkatnya performansi tentu kecepatan akses akan data-data yang besar meningkat pula.

1.2 Perumusan masalah

Berdasarkan latar belakang tersebut, maka permasalahan yang akan dibahas adalah sebagai berikut :

1. Bagaimana tingkat perbandingan performansi antara arsitektur *column-stores* database dengan *row-stores* database.
2. Faktor-faktor yang menyebabkan performansi DBMS yang satu lebih baik daripada DBMS yang lain.

3. Adakah situasi dimana *column-stores* database atau *row-stores* database efisien atau tidak efisien untuk digunakan.

Untuk menghindari meluasnya materi pembahasan Tugas Akhir ini, maka Penulis membatasi permasalahan sebagai berikut :

1. Menggunakan data akademik IT Telkom.
2. Pemodelannya menggunakan relasional model.
3. Yang diamati hanyalah performansi dengan parameter *throughput* dan *response time*.
4. Implementasi *column-stores* database menggunakan MonetDB 5.
5. Implementasi *row-stores* database menggunakan Oracle 10g.

1.3 Tujuan

Tujuan yang ingin dicapai dalam Tugas Akhir ini adalah :

1. Membangun perangkat lunak yang dapat digunakan untuk menguji performansi kedua DBMS sehingga bisa dijadikan sebagai sarana menarik kesimpulan.
2. Mengukur tingkat performansi antara kedua DBMS pada skenario-skenario yang diambil.
3. Menganalisis faktor-faktor yang menyebabkan salah satu DBMS lebih unggul dari sisi performansi dibanding yang lain.
4. Mencari situasi dimana *column-stores* database atau *row-stores* database efisien atau tidak efisien untuk digunakan.

1.4 Metodologi penyelesaian masalah

Berikut ini adalah metodologi penyelesaian masalah yang dipergunakan dalam menyelesaikan Tugas Akhir ini:

1.4.1 Study Literatur

Pada tahap ini yang dilakukan adalah mencari referensi yang dapat digunakan sebagai acuan dalam pembangunan basis data dalam DBMS MonetDB 5 atau dalam DBMS Oracle 10g, meliputi DDL, DML maupun query yang akan diterapkan pada skenario pengujian. Selain itu referensi lainnya juga digunakan untuk membantu pembuatan aplikasi dalam Java.

1.4.2 Analisis dan Desain

Tahapan ini adalah tahapan yang meliputi analisis dan perancangan terhadap basis data yang akan dibangun, merancang kebutuhan dan mendesain perangkat lunak yang akan digunakan untuk menguji performansi DBMS.

1.4.3 Implementasi Basis Data dan Perangkat Lunak

Tahap ini adalah tahap pembangunan basis data baik di tradisional DBMS maupun di *column-stores* DBMS. Selain membangun basis data, pada tahap ini dibangun pula perangkat lunak yang akan digunakan sebagai alat untuk membandingkan performansi dari kedua DBMS. Pada tahap ini DBMS yang digunakan untuk *row-stores* adalah Oracle 10g, untuk *column-stores* menggunakan MonetDB 5, sedangkan perangkat lunaknya menggunakan bahasa pemrograman Java.

1.4.4 Testing

Pada tahap ini akan dilakukan pengujian terhadap perangkat lunak yang telah dibangun, apakah sudah bekerja dengan benar. Pengujian dilakukan dengan memberikan statement SQL yang sama ke dua DBMS tersebut.

1.4.5 Analisis Hasil

Output dari perangkat lunak berupa *throughput* dan *response time*, baik dari konvensional DBMS maupun *column-stores* DBMS akan dibandingkan, sehingga akan diketahui DBMS mana yang lebih baik dari segi performansi. Setelah didapat DBMS mana yang lebih baik, kemudian akan dianalisis apa faktor-faktor apa yang menyebabkan salah satu DBMS lebih baik dari yang lainnya.

1.4.6 Pengambilan Kesimpulan dan Penyusunan Laporan

Pada tahap terakhir ini diambil kesimpulan dari hasil analisis yang telah dilakukan pada tahap sebelumnya untuk kemudian disusun laporan terhadap analisis yang telah dilakukan.

1.5 Sistematika Penulisan

Tugas akhir ini disusun dengan sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Berisi pemaparan mengenai latar belakang masalah, tujuan, perumusan masalah, batasan masalah, metodologi tugas akhir, dan sistematika penulisan.

BAB II LANDASAN TEORI

Berisi uraian mengenai landasan teori yang akan digunakan, meliputi teori tentang konsep *column-stores* dan *row-stores* database dan teori-teori pendukung yang berkaitan dengan penelitian tugas akhir ini.

BAB III DESAIN DAN PERANCANGAN SISTEM

Berisi tentang *use case*, *sequence diagram*, *class diagram* dan *entity relationship diagram* dari aplikasi dan database.

BAB IV ANALISIS DAN EVALUASI SISTEM

Berisi hasil analisa dari penelitian dan pengujian yang dilakukan.

BAB V PENUTUP

Berisi kesimpulan dan saran untuk pengembangan lebih lanjut terhadap hasil penelitian ini.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Dari berbagai macam uji coba dan analisis yang telah dilakukan maka dapat diambil kesimpulan bahwa :

1. *Column-stores* database memiliki kecepatan rata-rata 2 kali lebih baik dibanding *row-stores* pada operasi *update* dan pengaksesan data yang memiliki banyak baris akan tetapi hanya beberapa atribut dari data tersebut saja yang diambil karena memang membaca sebagian data lebih cepat dibanding harus melakukan pembacaan semua data.
2. *Row-stores* database memiliki performansi rata-rata 2 kali lebih baik pada operasi *insert*, dan *delete* karena pada operasi tersebut dibutuhkan pengaksesan semua atribut dimana pada arsitektur *row-stores* atribut data dapat ditemukan sekaligus pada sebuah blok karena letaknya yang berdampingan.
3. *Column-stores* memiliki performansi rata-rata lebih baik pada semua jenis query seperti *join*, *subquery* dan *kompleks query*, kecuali pada query yang membutuhkan pengambilan semua atribut(*select**).
4. *Column-stores* memiliki rata-rata *throughput* 2 kali lebih baik dibanding *row-stores* pada eksekusi query, sebaliknya pada eksekusi non-query *row-stores* memiliki rata-rata *throughput* 2 kali lebih baik dibanding *column-stores*.
5. Melalui arsitekturnya *column-stores* dapat meminimalisir I/O cost dengan hanya mentransfer blok-blok yang mengandung atribut-atribut yang dibutuhkan saja.

5.2 Saran

Setelah menyelesaikan Tugas Akhir ini, Penulis memiliki beberapa saran diantaranya :

1. Beberapa *column-stores* DBMS memiliki cara yang berbeda dalam menerapkan konsep *column-stores*-nya seperti penggunaan *virtual-oid* pada MonetDB atau penggunaan *B-tree Map* pada LucidDB, untuk selanjutnya disarankan untuk menganalisa performansi pada *column-stores* DBMS yang lain.
2. Banyak sistem informasi sekarang telah menerapkan sistem *online* dalam aplikasinya yang memungkinkan pada suatu waktu akan terdapat banyak proses pengaksesan sekaligus. Karena itu Penulis menyarankan pengujian dan analisis konsep arsitektur *column-stores* pada konkurensi data untuk melihat bagaimana tingkat performansinya.
3. Pada Tugas Akhir ini penulis melakukan analisis performansi menggunakan RDBMS, untuk selanjutnya disarankan menggunakan jenis DBMS yang lain seperti OODBMS atau ORDBMS.

6 Referensi

- [1] Oracle, 2005, "Oracle Database 10g: 2 Day DBA", Oracle
- [2] Oracle, 2005, "Oracle Database 10g: Administration Workshop Volume 1", Oracle
- [3] Nugroho, Adi, 2008, "Menjadi Administrator Basis Data Oracle 10g"
- [4] http://en.wikipedia.org/wiki/Column-oriented_DBMS, diakses pada tanggal 8 Agustus 2009
- [5] <http://en.wikipedia.org/wiki/MonetDB>, diakses pada tanggal 8 Agustus 2009
- [6] <http://monetdb.cwi.nl/projects/monetdb/SQL/>, diakses pada tanggal 8 Agustus 2009
- [7] Abadi, D., Madden, Samuel, Hachem, Nabil, 2008, "Column-Stores and Row-Stores: How Different Are They Really?"
- [8] Abadi, D., 2008, "Query Execution in Column-Oriented Database System", MIT PhD Dissertation
- [9] Vermeij, M., Quak, W., Kersten, M., Nes, N., "MonetDB, A Novel Spatial Column-Store DBMS"
- [10] Boncz, P., 2002, "Monet A Next Generation DBMS Kernel for Query Intensive Applications"
- [11] The MonetDB Development Team, 2008, "MonetDB/SQL Reference Manual"
- [12] Anastassia Ailamaki, 2000, "Architecture-Conscious Database Systems", A dissertation for Doctor of Philosophy, University of Wisconsin — Madison
- [13] Stonebraker, M., Abadi, D., 2005, "C-Store: A Column-Oriented DBMS"
- [14] Miguel C. Ferreira, 2005, "Compression and Query Execution within Column Oriented Databases", Massachusetts Institute of Technology
- [15] The MonetDB Development Team, 2008, "MonetDB Server Reference Manual"
- [16] Idreos, Manegold, Kersten, SIGMOD'09 "Self-Organizing Tuple Reconstruction in Column-Stores"
- [17] Abadi, Myers, DeWitt, and Madden. ICDE 2007. "Materialization Strategies in a Column-Oriented DBMS"
- [18] Silberschatz, Korth, Sudharsan, 2002, "Database System Concepts", Mc Graw Hill
- [19] Clement T. Yu, Weiyi Meng, "Principles of Database Query Processing for Advanced Applications"