ANALISIS ETL (EXTRACT, TRANSFORM, LOAD) PADA REAL TIME DATA WAREHOUSE ETL (EXTRACT, TRANSFORM, LOAD) ANALYSIS IN REAL TIME DATA WAREHOUSE

Yudha Saputra Syam[1], Dhinta Darmantoro[2], Kusuma Ayu Laksitowening[3]

[1]Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Data Warehouse semakin banyak diterapkan oleh banyak perusahaan besar. Hal tersebut dipacu dengan semakin berkembangnya kebutuhan suatu perusahaan dalam mendapatkan informasi yang berkaitan dengan proses bisnisnya. Data Warehouse yang berisi rekaman data yang dimiliki perusahaan, di era E-Bussiness sekarang ini haruslah selalu menyajikan data yang up-to-date. Pada umumnya proses Data Warehouse akan direfresh per satuan waktu (per 2 hari, per 24 jam atau bahkan per jam). Namun akankah masih efektif cara tersebut dengan melihat teknologi informasi yang semakin berkembang, dimana perubahan data dapat terjadi kapanpun dan dimanapun, bahkan dalam hitungan detik data transaksi dapat mengalami perubahan hingga puluhan kali atau lebih. Oleh karena itu, untuk menjaga kevalidan data yang ada pada Data Warehouse maka diterapkanlah Real Time Data Warehouse.

Real Time Data Warehouse memiliki perbedaan yang cukup signifikan pada proses ETLnya. Dimana proses ETL dilakukan tiap kali ada perubahan data pada sumber datanya. Sehingga keakuratan data yang tersimpan dalam warehouse lebih terjamin dan proses ETL tidak memakan waktu yang lama karena tidak semua data sumber dilibatkan, hanya yang mengalami perubahan saja pada waktu tertentu.

Setelah proses implementasi dan pengujian Real Time Data Warehouse dilakukan, maka dapat dianalisa faktor-faktor yang menjadi kelebihan dan keunggulan sistem ini dibandingkan konvensional Data Warehouse. Analisa diutamakan dari segi keakuratan data, bentuk proses ETL dan penyajian informasi yang up-to-date.

Kata Kunci : Real Time Data Warehouse, ETL, keakuratan data.

Abstract

Many big enterprise use Data Warehouse as their storage. It happens because every enterprise needs to get business process's information. Nowadays, Data Warehouse which contains historical data of enterprise should provide up-to-date data. Generally, Data Warehouse's process will be refresh periodically (twice a day, every 24 hours or every hours.). But it will not too effective, because technology grows faster, data can be change everytime even in seconds. Because of that, to prevent validation of data on Data Warehouse we can use real-time datawarehouse.

Real-time datawarehouse has significant differences on ETL process where ETL process will be done when change of data happen on data source so the data which is stored in warehouse will be guaranteed. ETL process doen't take a long time to do because not all data involved.

After implementation and testing of real-time Data Warehouse done, we can analyze every factors that make real-time datawarehousemore excellent than konvesional Data Warehouse. Analysis of real-time Data Warehouse will be more spesific in data accuracy, ETL process dan providing up-to date information

Keywords : Real Time Data Warehouse, ETL, data accuracy

1. Pendahuluan

1.1. Latar Belakang Masalah

Dengan semakin berkembangnya kebutuhan suatu perusahaan dalam mendapatkan informasi, maka kebutuhan akan penerapan *Data Warehouse* semakin meningkat pula. *Data Warehouse*, pada tingkatan managerial, berfungsi sebagai sarana dalam proses pengambilan keputusan. Pengambilan keputusan didasari pada kondisi internal maupun external perusahaan yang semuanya terintegrasi dalam *Data Warehouse* tersebut.

Segala informasi yang ada didalam *Data Warehouse* merupakan seluruh data history yang dimiliki perusahaan, seperti data transaksi mengenai suatu proses bisnisnya. Di era E-Bussiness sekarang ini, seringlah terjadi perubahan yang sangat cepat pada data transaksi tersebut, karena setiap orang dapat bertransaksi secara bersamaan melalui media internet kapan dan dimana saja. Keadaan tersebut didukung oleh kemampuan proses ETL yang pada umumnya diterapkan pada pembuatan *Data Warehouse* membutuhkan waktu yang tidak sebentar dan membutuhkan *storage* yang besar. Suatu saat ketika proses ETL tersebut berlangsung, data source atau data transaksi telah mengalami perubahan dikarenakan dari telah berkembangnya E-Bussiness tadi, sehingga jelaslah informasi yang terbentuk dan tersimpan dalam *Data Warehouse* sudah tidak valid. Sehingga apakah keakuratan data didalamnya masih dapat terjaga, selanjutnya apakah masih berguna semua history data yang ada didalam *Data Warehouse* jika nantinya akan dipakai dalam proses pengambilan keputusan.

Dari permasalahan yang timbul diatas, muncullah suatu ide untuk menciptakan sebuah *Data Warehouse* yang selalu *Up-to-Date.* Sehingga segala informasi yang telah berubah di data sumber pada saat itu langsung dapat muncul pada *Data Warehouse* secara otomatis, sehingga siapapun yang mengakses *Data Warehouse* tersebut baik pimpinan perusahaan selaku pengambil keputusan, staf *call center* perusahaan, dan semua perangkat perusahaan yang berhubungan langsung dengan konsumen maupun konsumen dari perusahaan itu sendiri, akan selalu mendapatkan informasi yang akurat setiap saat. Ide atau konsep tersebut dikenal dengan *Real-Time Data Warehouse*. Yang menjadikan *Real Time Data Warehouse* berbeda dengan *Data Warehouse* biasa terletak pada bentuk proses ETL-nya. Sehingga pada Tugas Akhir ini bentuk ETL tersebut akan menjadi analisa utamanya.

1.2 Perumusan Masalah

Setelah mengetahui perlunya diterapkan suatu *Real Time Data Warehouse* untuk menjaga keakuratan data yang tersimpan pada *Data Warehouse* yang akan digunakan dalam proses pengambilan keputusan, maka masalah yang akan diangkat pada proposal tugas akhir ini antara lain :

1. Bagaimana mengimplementasikan *Real Time Data Warehouse*.
2. Bagaimana menerapkan mekanisme proses ETL yang mendukung dalam pembangunan *Real Time Data Warehouse*.

3. Bagaimana menganalisa faktor-faktor yang menjadi keunggulan dan kelemahan *Real Time Data Warehouse* dibandingkan *Data Warehouse* yang telah digunakan pada umumnya.

Berdasarkan rumusan masalah diatas, ada beberapa hal yang menjadi batasan masalah pada tugas akhir ini, yaitu :

1. Pengimplementasian *Real Time Data Warehouse* menggunakan dengan SQL Server 2005.
2. Reporting ditampilkan dalam bentuk aplikasi web menggunakan Report Builder yang telah terintegrasi dengan Microsoft SQL Server 2005.
3. Menggunakan data dummy yaitu mengenai beberapa proses bisnis yang dijalankan oleh perpustakaan di kampus IT Telkom, diantaranya : absensi pengunjung, peminjaman dan pengembalian buku.
4. Untuk *front application* maupun *reporting application* dibuat tidak memperhatikan unsur *GUI*.

1.3 Tujuan

Ada beberapa tujuan yang ingin dicapai pada pembuatan tugas akhir ini antara lain :

1. Mengimplementasikan *Real Time Data Warehouse* pada suatu studi kasus.
2. Menganalisa keunggulan dan kelemahan yang ada pada *Real Time Data Warehouse* dari beberapa aspek antara lain, bentuk proses ETL-nya, keakuratan data yang tersimpan, dan penyajian data cepat dan *up-to-date.*

1.4 Metodologi Penyelesaian Masalah

a. Studi Literatur :

Mencari referensi dari sumber-sumber lainnya sebagai acuan dan masukan dalam pembangunan *Real Time Data Warehouse.*

b. Pengumpulan Data :

Pengumpulan data dilakukan sejauh mungkin dengan mencari studi kasus pada suatu perusahaan, namun jika mendapat perusahaan yang tepat akan menggunakan data dummy.

c. Analisis dan Desain :

Pada tahap ini dilakukan analisa dan desain kebutuhan terhadap semua informasi yang akan dimasukkan kedalam *Real Time Data Warehouse*, sesuai data yang diambil dari studi kasus yang telah didapat dari tahapan sebelum ini, dan membuat aplikasi *reporting*-nya

d. Implementasi :

Tahapan ini melakukan pembangunan pada *Real Time Data Warehouse* menggunakan SQL Server 2005 sebagai DBMSnya. Disamping itu akan dibangun suatu aplikasi *Reporting* berfungsi sebagai interface dalam

mengambil keputusan sesuai proses bisnis yang telah ditentukan sebelumnya.

e. Testing :

Pengujian terhadap *Real Time Data Warehouse* dan aplikasi *Repoting*-nya yang telah dibangun.

f. Analisis hasil :

Pada tahap analisa difokuskan pada kelebihan dan kelemahan yang ditimbulkan oleh *Real Time Data Warehouse* terutama pada bentuk proses ETL-nya.

g. Pengambilan kesimpulan dan penyusunan laporan tugas akhir :

Pengambilan 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 latar belakang, perumusan masalah, batasan masalah, tujuan, metodologi penyelesaian masalah, dan sistematika penulisan.

BAB II LANDASAN TEORI

Berisi teori-teori yang melandasi pembangunan sebuah Real Time *Data Warehouse* terutama pada tahap desain dan implementasi. Bab ini berisikan juga definisi mengenai *Data Warehouse* dan *real time Data Warehouse*, selain itu terdapat teknik-teknik pemecahan masalah dalam perancangan *ETL* pada *real time Data Warehouse.*

BAB III ANALISA DAN PERANCANGAN SISTEM

Berisi analisa perancangan sistem *real time Data Warehouse.*

.

BAB IV IMPLEMENTASI DAN ANALISIS SISTEM

Berisi gambaran bagaimana sistem diimplementasikan, diuji dan dianalisa kelebihan dan kekurangannya dari bentuk ETL yang telah dibuat dan dilihat dari tingkat ketersediaan data yang tepat dan akurat.

BAB V PENUTUP

Berisi kesimpulan dan saran dari analisa tugas akhir ini serta saran untuk pengembangan selanjutnya.

5. KESIMPULAN

5.1 Kesimpulan

Dari hasil implementasi dan analisis yang telah dilakukan maka dapat disimpulkan bahwa :

1. ETL pada Real Time Data Warehouse dapat diimplementasikan dengan menggunakan tambahan temporary table dan melakukan sistem penjadwalan. Dimana table tersebut digunakan untuk mendeteksi dan menangkap perubahan data atau transaksi yang terjadi pada data source.
2. Jumlah data yang diproses dalam ETL dipengaruhi oleh besarnya perubahan yang di-capture oleh temporary table.
3. Waktu ETL yang dibutuhkan jauh lebih cepat karena hanya memroses data yang mengalami perubahan.
4. Temporary table juga berpengaruh pada tingginya tingkat keakuratan data yang disajikan pada reporting tools. Sehingga memperkuat sifat fact table yang dinamis dan selalu bertambah setiap saat.
5. Dalam hal kebutuhan Real Time Data Warehouse cocok digunakan untuk proses bisnis yang mempunyai *lifecycle management* yang cepat.

5.2 Saran

Saran yang dapat diberikan untuk tugas akhir ini mengenai Real Time Data Warehouse antara lain :

1. Sebaiknya penjadwalan pada ETL dihapus dan menerapkan Push Method, yaitu bentuk lain dari implementasi *Real Time Data Warehouse*. Sehingga ketika ada perubahan data tidak perlu menunggu waktu ETL yang telah dijadwalkan.
2. Untuk studi kasus coba menggunakan data yang mempunyai lifcycle management lebih cepat lagi.

6. DAFTAR PUSTAKA

[1] Darmantoro,Dhinta,ST,MSCS,2006,*Data Warehousing*,IT Telkom.

[2] Kimpball, Ralph & Ross, Margy, 2002, *The Data Warehouse Toolkit*, second edition, Singapore, John Wiley & Sons, Inc.

[3] Kimpball, Ralph & friends,1998, *The Data Warehouse Lifecycle Toolkit*, Singapore, John Wiley & Sons, Inc.

[4] Kimball, Ralph & friends, 1996, *The Data Warehouse Toolkit : Practical Techniques for Building Dimensional Data Warehouse*, John Wiley & Sons, Inc.

[5] Nasution, Raudah (dkk). 2007. "Implementasi Data Warehouse pada Database Akademik STT TELKOM". *Prosiding Konferensi Nasional Sistem Informasi*. 251-256.

[6] Ankorion, Itamar,2005, *Change Data Capture-Efficient ETL for Real-Time BI*, dikutip 27 Februari, [online], Available : http://www.DMReview.com

[7] Cyran, Michael, 2003, *Oracle Warehouse Builder*, dikutip 27 Februari 2008, [online], Available: http://www.oracle.com/technology/products/warehouse/pdf/TWP_BIDW_OWB_Roadmap_1006.pdf

[8] DataMirror Corperation, 2002, *Considerations for Building a Real-time Data Warehouse*, dikutip 24 Februari 2008, [online], Available : http://www.wilshireconferences.com/EDF2002/ko-whitepaper.pdf

[9] Haisten, Michael, 1999, *Real-Time Datawarehouse Defined*, dikutip 24 Februari 2008, [online], Available : http://www.bettermanagement.com/library/default.aspx

[10] Haisten, Michael, 1999, *The next Stage in Data Warehouse Evolution, part 1*, dikutip 24 Februari 2008, [online], Available : http://www.dmreview.com

[11] Kimball, Ralph, 2005, *Real-time Data Warehouse Design Challenges*, dikutip 24 Februari 2008, [online], Available : http://www.ittoolbox.com

[12] Alam, M. Agus J, 2005, Pemrograman Transact-SQL pada SQL Server 2005, Jakarta : Gramedia