

Analisa Pemanfaatan Datawarehouse di Depnakertrans

Anto Pribadi¹, Dwi Erika Pitasari², I Wayan S. Wicaksana³, DjJuarsa MD⁴
^{1,2}Univ.Gunadarma & Depnakertrans, ³Univ. Gunadarma, ⁴Depnakertrans
^{1,2,4}{anto, erika, djuharsa}@nakertrans.go.d,
³iwayan@staff.gunadarma.ac.id

Ringkasan

Saat ini dikalangan dunia Teknologi Informasi telah dikenal teknologi datawarehouse, datawarehouse merupakan kumpulan data yang berasal dari berbagai sumber, dengan memiliki historikal data di atas lima tahun. Dengan feature yang dimiliki oleh datawarehouse, maka Depnakertrans bisa menggunakan teknologi ini untuk berbagai analisis memanfaatkan data mining. Faktor penting yang akan dilihat adalah bagaimana menganalisis pemanfaatan datawarehouse pada lingkungan Depnakertrans.

1 Pendahuluan

Data dan Informasi merupakan salah satu aset yang penting bagi kelancaran kegiatan pelaksanaan tugas pemerintahan baik menyangkut kegiatan pelayanan masyarakat dan kegiatan penyajian informasi bagi pengambilan keputusan, utamanya adalah dalam hal mendukung masalah pengangguran dan kemiskinan, Perencanaan Tenaga Kerja, Perencanaan Pembangunan Transmigrasi, Perumusan Ketenagakerjaan dan Ketransmigrasian, Peningkatan Kualitas dan Produktivitas Tenaga Kerja, Perlindungan TKI-LN, Pembinaan Hubungan Industrial, Pengawasan Ketenagakerjaan, Pengembangan Kawasan/Lokasi Transmigrasi, Pengawasan TKA dan lain-lain.

Dalam rangka mendukung kebutuhan data dan informasi di lingkungan Depnakertrans saat ini telah ada beberapa sistem informasi, namun demikian sistem yang satu dengan yang lainnya belum merupakan satu kesatuan sehingga belum dapat menyediakan kebutuhan data dan informasi secara cepat dan terpadu. Keadaan sistem informasi yang

tersebar dan platform yang berbeda - beda mengakibatkan duplikasi data, pengkodean yang tidak sama sehingga muncul data yang berbeda - beda padahal data tersebut adalah data yang sama. Saat ini telah dikenal teknologi Datawarehouse (gudang data). Datawarehouse merupakan kumpulan data yang berasal dari sumber - sumber berbeda, tetapi data yang dihasilkan tidak menimbulkan duplikasi dan pengkodean yang berbeda sumber.

Disisi lain, Depnakertrans memiliki data yang terdistribusi dan otonomi dengan waktu histori yang cukup panjang. Salah satu kendala utama adalah pada waktu akan digunakan, banyak data yang tidak bisa segera dikumpulkan ataupun di akses. Maka perlu dipikirkan bagaimana jika menerapkan datawarehouse untuk mengatasi dua hal utama ini.

Berdasarkan teori ini maka datawarehouse dapat diimplementasikan untuk menangani masalah duplikasi data dan pengkodean yang tidak sama di lingkungan Depnakertrans.

Dalam paper ini akan dibahas analisis pemanfaatan datawarehouse pada subject area Angkatan Kerja pada Depnakertrans. Sengaja pada langkah awal akan dilihat pada sebuah model kecil yang kemudian akan dibawa ke area yang lebih luas di Depnakertrans.

Sistematika penulisan penelitian ini terbagi atas beberapa bagian, dimana bagian awal menjelaskan latar belakang permasalahan. Bagian berikutnya akan melihat ke state of the art dari datawarehouse. Pada bagian ke-tiga akan membahas pendekatan analisis yang dipilih untuk diterapkan pada kasus datawarehouse Depnakertrans untuk bidang Angkatan Kerja.

2 State of the art Datawarehouse

Datawarehouse(gudang data) adalah repository utama dari kumpulan historical data. Yang terdiri dari data raw yang dapat dimanfaatkan untuk manajemen decision Support System [4] . Menurut Bill Inmon [3] data warehouse harus mengikuti terms sebagai berikut;

1. subject - oriented data dalam database terorganisasi sehingga semua elemen data berelasi dengan objek yang dihubungkan bersama - sama.
2. Time Variant - perubahan data didalam database direkam sehingga laporan yang ditampilkan bisa memperlihatkan perubahan yang terjadi berdasarkan perubahan waktu.
3. Non Volatile - data di dalam database tidak pernah dituliskan lebih dari satu kali atau di-delete. Sekali data committed, data bersifat statis,read only, tetapi bisa dimunculkan untuk laporan yang akan datang.
4. Integrated - database mengandung data dari semua operasional aplikasi yang terorganisasi sehingga data konsisten

Melihat dari karakteristik datawarehouse di atas, maka manfaat yang dapat dipetik adalah :

1. Menganalisa Data sehingga bisa dibuat query yang kompleks dan analisisnya seperti data mining.
2. Menganalisa informasi tanpa menurunkan operasional sistem.

Metadata adalah data tentang data, khususnya ada datawarehouse maka ini adalah merupakan informasi mengenai datawarehouse dan data yang diisi ke dalam datawarehouse. Metadata dibagi dua bagian yaitu:

1. Teknis data. Teknis data berisi penjelasan tentang operational database dan penjelasan dari data warehouse Dari kedua penjelasan itu, atau skema, operasi pergerakan data dapat diimplementasikan. Data ini membantu administrasi data warehouse menjaga datanya dan mengetahui dari mana semua data berasal.

2. Data bisnis. Data bisnis membantu pemakai mencari informasi dalam data warehouse tanpa mengetahui implementasi databasenya. The business data memberikan informasi kepada pemakai:

- Pada saat data dipindahkan ke dalam warehouse (*how current it is*)
- Dari mana data berasal. (*which operational database*).
- Memungkinkan pengguna mengetahui reliabilitas dari data.

Pemanfaatan dari datawarehouse adalah dengan menggunakan pendekatan dari Data Mining, dimana definisi dari Data Mining adalah

1. Mencari informasi yang berharga di dalam suatu data yang berjumlah besar.
2. Eksplorasi dan analisa secara otomatis atau semiotomatis dari suatu kuantitas data yang besar yang bertugas untuk mencari pola dan aturan yang berarti

Melihat dari manfaat tersebut, maka tugas Data Mining secara umum dapat dibagi menjadi dua metode yaitu:

1. Metoda prediksi Menggunakan beberapa variable untuk memperkirakan suatu nilai yang tidak diketahui dari variable yang lain.
2. Metoda deskripsi Mencari suatu pola yang dapat ditafsirkan manusia sehingga data dapat digambarkan atau diuraikan.
3. Jenis pendekatan untuk mencapai dua metode tadi adalah :

- Classification [Predictive]
- Clustering [Descriptive]
- Association Rule Discovery [Descriptive]
- Regression [Predictive]
- Deviation Detection [Predictive]

Untuk menganalisis apakah sebuah datawarehouse sesuai atau tidak untuk sebuah organisasi perlu dilakukan pertimbangan beberapa faktor sebagai berikut [5, 2]:

- Organisasi, sejauh mana sebuah organisasi dalam menjalankan kegiatannya sangat bersandar akan data terutama data histori yang digunakan.
- Kebutuhan, apakah memang diperlukan model datawarehouse yang bersifat tersentral atau terdistribusi dengan model *tightly couple* ataupun *loosely couple*.
- Bahan dasar, apakah memang cukup tersedia data transaksi elektronika dasar beserta juga standarisasi untuk menggabungkan data-data dari berbagai sumber.
- Teknologi, pilihan teknologi sangat beragam, ini akan sangat bergantung kepada unjuk kerja yang ingin dicapai, dana, waktu yang tersedia.
- Unjuk kerja, seberapa jauh unjuk kerja sebuah datawarehouse yang diharapkan. Khusus untuk unjuk kerja dapat dilihat faktor utama [1] sebagai berikut :
 - kapasitas : melihat kepada DBMS makan kemampuan kapasitas penyimpanan dan kecepatan akses perlu diperhitungkan dengan matang, sesuai dengan data yang dimiliki oleh sebuah organisasi dan perkembangan ke depan.
 - unjuk kerja *loading* dan *indexing* : seberapa cepat DBMS untuk melakukan penarikan dan index data dasar pada sistem produksi atau alat untuk pembuatan datawarehouse.
 - *operational integrity, reliability and manageability*: bagaimana kemampuan teknologi untuk memberikan solusi pada sudut pandang operasional, apakah mampu untuk kondisi 24x7, seberapa sulit untuk melakukan online maupun mengakses datanya.
 - *client/server connectivity* : seberapa jauh mendukung terutama untuk produk di pasar termasuk untuk interface, middle-ware dan SQL dialek.
 - *query processing performance* : seberapa baik DBMS menghandel sebuah query, termasuk memindai tabel, pencarian dengan kondisi, pencarian dari berbagai tabel.

3 Pembahasan

Dari hasil tinjauan pustaka maka dapat dihasilkan arsitektur Enterprise Data Warehouse dan melihat kondisi di lapang yang terdiri dari:

- Sumber data / Data Sources dalah sumber data yang ada di lingkungan Departemen Tenaga Kerja dan Transmigrasi, baik internal maupun eksternal memungkinkan untuk di gabungkan karena telah tersedianya standard untuk pertukaran data walau belum secara keseluruhan. Permasalahan adalah lambatnya tersedia data beserta tingkat kelengkapan dan validitas data yang perlu sekali ditingkatkan.
- Process Extract, Transform dan Loading (ETL) ETL., adalah proses untuk mengambil dan memilih data dari sumber data yang akan diolah ke dalam data warehouse yang disesuaikan dengan kebutuhan bisnis. Ini sangat penting di Depnakertrans, karena adanya pelayanan masyarakat langsung terutama dalam kondisi-kondisi kritis seperti beberapa kasus Angkatan Kerja
- Menejemen Metadata Manajemen Metadata adalah alat pendukung meliputi informasi/keterangan mengenai data itu sendiri, baik bersifat teknis maupun bisnis. Sehingga pengguna informasi akan mudah memahami suatu istilah atau proses informasi
- Staging, staging adalah tempat penyimpan data yang bersifat sementara dimana datanya:Disimpan dalam satu database tertentu, Tidak menggunakan index untuk mempercepat proses transformasi , Tidak dapat diakses oleh user hanya khusus untuk developer , Data di staging akan digunakan untuk proses transformasi data ke dalam Datawarehouse.
- Enterprise Data Warehouse, EDW Departemen Tenaga Kerja dan Transmigrasi berisi data yang terintegrasi yang akan digunakan untuk keperluan pelaporan dan analisa. EDW berisi data yang bersifat atomic maupun summary, atau keduanya, sesuai kebutuhan bisnis Departemen Tenaga Kerja dan Transmigrasi. Dalam EDW data akan disimpan secara historical sesuai kebutuhan bisnis di Departemen

Tenaga Kerja dan Transmigrasi. Data tersebut akan disimpan secara terpusat dalam satu server.

- Operation Data Store (ODS) ODS adalah tempat penyimpanan data transaksi yang dirancang untuk kebutuhan EDW, dimana dalam EDW data dapat berubah sesuai perubahan data operasional sehingga kebutuhan laporan yang real time dapat diatasi dengan performance dan scalability yang cukup baik.
- Metadata Teknik Metadata teknis digunakan untuk mempermudah pengontrolan proses ETL, sehingga setiap terjadi perubahan proses baik dari sumber data, aturan bisnis, dan target datawarehouse-nya akan dapat diketahui seberapa besar dampak perubahan tersebut. Metadata Teknis digambarkan sebagai berikut:
 1. Panjang dan definisi Field di dalam datawarehouse
 2. Pemetaan field ke field antara sumber dan target data
 3. Waktu respon query
 4. Penggunaan query table-table agregat
 5. Waktu yang dibutuhkan untuk loading, update dan menyimpan data pada saat ke dalam atau ke luar datawarehouse
 6. Waktu yang dibutuhkan dan memverifikasi kesuksesan proses transfer batch file
- Metadata Bisnis Metadata bisnis adalah informasi atau definisi yang diinginkan oleh pengguna mengenai arti, kualitas dan timelines dari data tersebut, sehingga definisi dan bussiness rules dari informasi nantinya akan unik dan mudah dimengerti pengguna, baik pada saat pengembangan maupun pada saat dipublikasikan.
- Akses Data Pemanfaatan data warehouse nantinya akan digunakan untuk: Ad Hoc analisis, Fix Report/ Operational dan EIS / OLAP Pelaporan Akses data-nya akan melalui Portal

Untuk pencapaian komponen-komponen di atas maka beberapa langkah perlu dipersiapkan dengan baik. Langkah-langkah tersebut adalah :

- Pemilihan sumber data sebagai materi input. Pada Nakertrans data masih tersimpan dalam format kertas dan digital. Khusus untuk format kertas diperlukan suatu pendekatan khusus untuk proses digitalisasi. Untuk yang digital perlu dipertimbangkan masalah keragaman database dari berbagai sumber.
- Keragaman pada sistem database yang perlu diatasi dalam sebuah organisasi titik beratnya adalah pada masalah perbedaan *syntactic*. Contoh perbedaan syntactic adalah misalkan pemberian nama atribut yang berbeda seperti jenis kelamin dengan sex, pengkodeannya yang berbeda seperti wanita dengan perempuan, perbedaan nilai inisial, perbedaan tipe data misalkan text atau numerik.
- *Data cleaning* adalah suatu proses untuk mengurangi noise dari data, seperti data rangkap, data yang tidak konsisten misalkan ada manusia dengan tinggi 250cm.
- Teknologi yang digunakan pada pembangunan datawarehouse adalah sangat beragam. Maka pemilihan teknologi yang tepat merupakan hal yang sangat penting. Pemilihan dengan mempertimbangkan kebutuhan, ketersediaan SDM, dana, pengembangan dan pemeliharaan.
- Analisis data pada datawarehouse akan sangat mempengaruhi manfaat dari penerapan datawarehouse. Analisis ini akan memilih salah satu algoritma dari data mining. Data mining adalah merupakan faktor penting dalam pemanfaatan datawarehouse.
- Publikasi dan optimalisasi pemanfaatan dari datawarehouse dan datamining adalah penting agar dapat dimanfaatkan secara optimal untuk berbagai pihak yang berkepentingan dan berhak menggunakan.
- Keamanan merupakan isu standard pada sistem database. Terutama pada datawarehouse yang adalah merupakan repository penyimpanan besar yang secara teori menyimpan minimal data selama lima tahun. Maka perlindungan akan data yang disimpan adalah merupakan hal yang sangat penting untuk menghindari penyalahgunaan.

4 Kesimpulan

Dari Tinjauan Pustaka mengenai teori - teori tentang datawarehouse serta analisa sistem yang ada di lingkungan Depnakertrans, maka dapat diambil kesimpulan bahwa datawarehouse dapat diterapkan di lingkungan Depnakertrans dengan memperhatikan aliran dan keterbukaan data transaksi.

Datawarehouse di lingkungan Depnakertrans bisa dimanfaatkan untuk Ad Hoc analisis, Fix Report /Operational dan EIS / OLAP Pelaporan

Pustaka

- [1] Dave Browning and Joy Mundy. Data Warehouse Design Considerations. Technical article, Microsoft, [http://msdn2.microsoft.com/en-us/library/aa902672\(sql.80\).aspx#sql_dwdesign_topic3](http://msdn2.microsoft.com/en-us/library/aa902672(sql.80).aspx#sql_dwdesign_topic3), 2000.
- [2] Mark Demarest. A Data Warehouse Evaluation Model. <http://www.noumenal.com/marc/oracle7.html>, 1995. accessed Oct 20, 2007.
- [3] none. Bill Inmof father of Datawarehouse. http://en.wikipedia.org/wiki/Bill_Inmon. accessed Oct 24, 2007.
- [4] none. Data Warehouse. <http://en.wikipedia.org/wiki/Datawarehouse>. access Oct 24, 2007.
- [5] none. Business Intelligence Tutorial: Introduction to the Data Warehouse Cente. <ftp://ftp.software.ibm.com/ps/products/db2/info/vr8/pdf/letter/db2tue80.pdf>, 2002. accessed Oct 22, 2007.