

MATERI UTS DATA WAREHOUSE

A. DATA WAREHOUSE

Pengertian Data Warehouse

Data Warehouse merupakan database yang saling bereaksi yang dapat digunakan untuk query dan analisis, bersifat orientasi subjek, terintegrasi, time-variant, tidak berubah yang digunakan untuk membantu para pengambil keputusan.

Karakteristik Data Warehouse

1. Berorientasi pada subyek.

Data warehouse berorientasi subject artinya data warehouse didesain untuk menganalisa data berdasarkan subject-subject tertentu dalam organisasi, bukan pada proses atau fungsi aplikasi tertentu. Data warehouse diorganisasikan disekitar subjek-subjek utama dari perusahaan(customers,products dan sales) dan tidak diorganisasikan pada area-area aplikasi utama (customer invoicing,stock control dan product sales). Hal ini dikarenakan kebutuhan dari data warehouse untuk menyimpan data-data yang bersifat sebagai penunjang suatu keputusan, dari pada aplikasi yang berorientasi terhadap data.

2. Terintegrasi

Data warehouse dapat menyimpan data-data yang berasal dari sumber-sumber yang terpisah ke dalam suatu format yang konsisten dan saling terintegrasi satu sama lain. Data tidak dapat dipecah-pecah karena data yang ada merupakan suatu kesatuan yang menunjang keseluruhan konsep *data warehouse* itu sendiri.

Syarat integrasi sumber data dapat dipenuhi dengan cara konsisten dalam penamaan variabel, ukuran variabel, dan atribut fisik dari data. Misalnya untuk tabel pelanggan yang berasal dari sumber data I, format untuk jenis kelamin menggunakan format F dan M (Female dan Male). Sedangkan untuk tabel pelanggan yang berasal dari sumber data II menggunakan format P dan L (Perempuan dan Laki-Laki). Maka data tersebut sebelum masuk ke dalam *data warehouse* harus dilakukan integrasi terlebih dahulu. Misalnya memilih format dari salah satu sumber data atau menggunakan format yang baru, misalnya 0 untuk perempuan/female dan 1 untuk laki-laki/male.

3. Rentang waktu

Data yang berada di dalam *data warehouse* dapat dikatakan akurat dan valid pada titik waktu tertentu atau dalam interval waktu tertentu (hanya pada saat proses ETL/*update*). Setiap data yang dimasukkan ke *data warehouse* pasti memiliki dimensi waktu. Dimensi waktu ini akan dipergunakan sebagai pembanding dalam perhitungan untuk menghasilkan laporan yang

diinginkan. Selain itu, dengan menggunakan dimensi waktu, pembuat keputusan dapat mengenal kecenderungan dan pola dari suatu data.

4. Non - Volatile

Data pada *data warehouse* tidak di-*update* secara *real time*, tetapi di-*refresh* dari sistem operasional secara reguler. Data yang baru selalu ditambahkan sebagai tambahan bagi database itu sendiri. *Database* tersebut secara berkesinambungan mengambil data baru ini, dan kemudian disatukan dengan data sebelumnya. Data yang berada dalam *data warehouse* bersifat *read-only*, dan tidak berubah setiap saat sehingga hanya terdapat dua kegiatan dalam *data warehouse*, yaitu mengambil data (*loading data*), dan akses data (mengakses *data warehouse*, seperti melakukan *query* untuk menampilkan laporan yang dibutuhkan, tidak ada kegiatan *updating data*).

Keuntungan Data Warehouse

1. Data diorganisir dengan baik untuk *query* analisis dan sebagai bahan untuk pemrosesan transaksi.
2. Perbedaan diantara struktur data yang heterogen pada beberapa sumber yang terpisah dapat diatasi.
3. Aturan untuk transformasi data diterapkan untuk memvalidasi dan mengkonsolidasi data apabila data dipindahkan dari database OLTP ke data warehouse.
4. Masalah keamanan dan kinerja bisa dipecahkan tanpa perlu mengubah sistem produksi.

Keuntungan Strategis Data Warehouse bagi Perusahaan/Organisasi

1. Kemampuan untuk mengakses data yang besar
2. Kemampuan untuk memiliki data yang konsisten
3. Kemampuan kinerja analisa yang cepat
4. Mengetahui adanya hasil yang berulang-ulang
5. Menemukan adanya celah pada business knowledge atau business process.
6. Mengurangi biaya administrasi
7. Memberi wewenang pada semua anggota dari perusahaan dengan menyediakan kepada mereka informasi yang dibutuhkan agar kinerja bisa lebih efektif.

Tiga Jenis Dasar Sistem Data Warehouse

1. Functional Data Warehouse (*Data Warehouse Fungsional*)

Kata operasional disini merupakan database yang diperoleh dari kegiatan sehari-hari. Data warehouse dibuat lebih dari satu dan dikelompokkan berdasar fungsi-fungsi yang ada di dalam perusahaan seperti fungsi keuangan(*financial*),marketing,personalia dan lain-lain.

Keuntungan dari bentuk data warehouse seperti ini adalah, sistem mudah dibangun dengan biaya relatif murah sedangkan kerugiannya adalah resiko kehilangan konsistensi data dan terbatasnya kemampuan dalam pengumpulan data bagi pengguna.

2. Centralized Datawarehouse (*Data Warehouse Terpusat*)

Bentuk ini terlihat seperti bentuk data warehouse fungsional, namun terlebih dahulu sumber data dikumpulkan dalam satu tempat terpusat, kemudian data disebar ke dalam fungsinya masing-masing, sesuai kebutuhan perusahaan. Data warehouse terpusat ini, biasa digunakan oleh perusahaan yang belum memiliki jaringan eksternal.

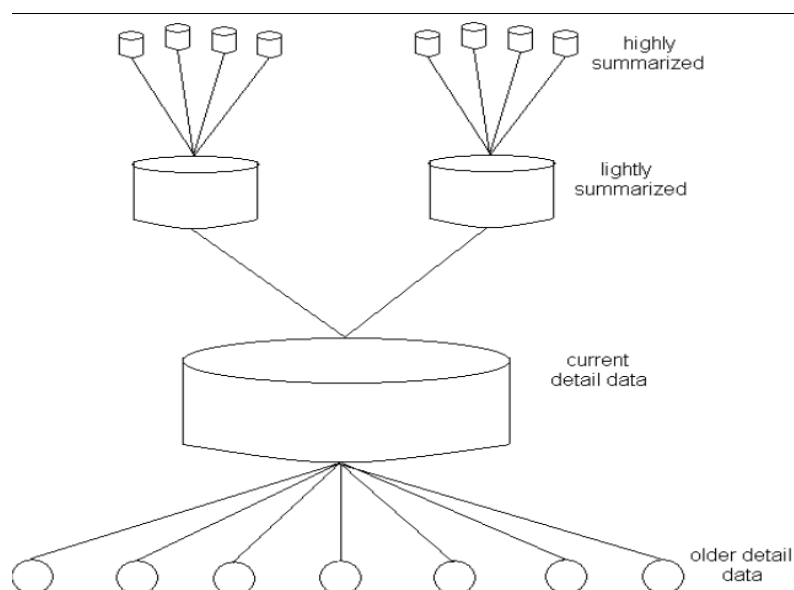
Keuntungan dari bentuk ini adalah data benar-benar terpadu karena konsistensinya yang tinggi sedang kerugiannya adalah biaya yang mahal serta memerlukan waktu yang cukup lama untuk membangunnya.

3. Distributed Data Warehouse (*Data Warehouse terdistribusi*)

Pada data warehouse terdistribusi ini, digunakan gateway yang berfungsi sebagai jembatan penghubung antara data warehouse dengan workstation yang menggunakan sistem beraneka ragam. Dengan sistem terdistribusi seperti ini memungkinkan perusahaan dapat mengakses sumber data yang berada diluar lokasi perusahaan(eksternal).

Keuntungannya adalah data tetap konsisten karena sebelum data digunakan data terlebih dahulu di sesuaikan atau mengalami proses sinkronisasi. Sedangkan kerugiannya adalah lebih kompleks untuk diterapkan karena sistem operasi dikelola secara terpisah juga biaya nya yang paling mahal dibandingkan dengan dua bentuk data warehouse lainnya.

Struktur Data Warehouse



1. *Current detail data*

Current detail data merupakan data detil yang aktif saat ini, mencerminkan keadaan yang sedang berjalan dan merupakan level terendah dalam *data warehouse*. Didalam area ini warehouse menyimpan seluruh detail data yang terdapat pada skema basis data. Jumlah data sangat besar sehingga memerlukan *storage* yang besar pula dan dapat diakses secara cepat. Dampak negatif yang ditimbulkan adalah kerumitan untuk mengatur data menjadi meningkat dan biaya yang diperlukan menjadi mahal.

2. *Older detail data*

Data ini merupakan data historis dari *current detail data*, dapat berupa hasil cadangan atau archive data yang disimpan dalam *storage* terpisah. Karena bersifat back-up(cadangan), maka biasanya data disimpan dalam *storage* alternatif seperti tape-desk.

Data ini biasanya memiliki tingkat frekuensi akses yang rendah. Penyusunan file atau directory dari data ini di susun berdasarkan umur dari data yang bertujuan mempermudah untuk pencarian atau pengaksesan kembali.

3. *Lightly summarized data*

Data ini merupakan ringkasan atau rangkuman dari *current detail data*. Data ini dirangkum berdasar periode atau dimensi lainnya sesuai dengan kebutuhan.

Ringkasan dari *current detail data* belum bersifat total summary. Data-data ini memiliki detil tingkatan yang lebih tinggi dan mendukung kebutuhan warehouse pada tingkat departemen. Tingkatan data ini di sebut juga dengan data mart. Akses terhadap data jenis ini banyak digunakan untuk view suatu kondisi yang sedang atau sudah berjalan.

4. *Highly summarized data*

Data ini merupakan tingkat lanjutan dari *Lightly summarized data*, merupakan hasil ringkasan yang bersifat totalitas, dapat di akses misal untuk melakukan analisis perbandingan data berdasarkan urutan waktu tertentu dan analisis menggunakan data multidimensi.

5. *Metadata*

Metadata bukan merupakan data hasil kegiatan seperti keempat jenis data diatas. Menurut Poe, metadata adalah 'data tentang data' dan menyediakan informasi tentang struktur data dan hubungan antara struktur data di dalam atau antara *storage*(tempat penyimpanan data).

Metadata berisikan data yang menyimpan proses perpindahan data meliputi database structure, contents, detail data dan summary data, matrices, versioning, aging criteria, versioning,

transformation criteria. Metadata khusus dan memegang peranan yang sangat penting dalam data warehouse. Metadata sendiri mengandung struktur data, algoritma, dan mapping.

B. OLAP DAN OLTP

1. OLAP

OLAP adalah singkatan dari *Online Analytical Processing* yang merupakan suatu proses atau juga bisa disebut dengan sistem yang dirancang atau digunakan untuk melakukan permintaan terhadap data dalam bentuk yang kompleks dan menganalisa data yang bervolume besar, maka dari itu OLAP seringkali disebut analisis data multidimensi. OLAP bekerja dengan data dalam bentuk multidimensi yang biasanya berbentuk 3 dimensi yang disebut kubus (cube).

OLAP biasanya digunakan untuk pengambilan suatu keputusan, misalkan sebagai contoh para top level di sebuah minimarket ingin melihat data transaksi perbulan, per 6 bulan atau bahkan per tahun untuk pengambilan keputusan barang apa saja yang paling laku dijual serta barang apa saja yang tidak laku di pasar, pada OLAP dikenal dengan istilah drill-down, slicing dan dicing

Terdapat beberapa kategori dari OLAP, diantaranya :

- Multi deminsional OLAP (MOLAP atau MD-OLAP)
- Relational OLAP (ROLAP), atau disebut juga Multi relational OLAP
- Managed Query Environment (MQE), atau disebut juga Hybrid OLAP (HOLAP)

Keuntungan dari OLAP, yaitu :

- Penyimpanan pengawasan organisasi melalui integritas data korporasi.
- Meningkatkan hasil dan keuntungan secara potensial dengan mengizinkan organisasi untuk merespon permintaan pasar lebih cepat.
- Meningkatkan produktifitas pemakai akhir bisnis, pengembang IT, dan keseluruhan organisasi. Pengawasan yang lebih dan akses tepat waktu terhadap informasi strategis dapat membuat pengambilan keputusan lebih efektif.
- Mengurangi “backlog” pengembangan aplikasi bagi staf IT dengan membuat pemakai akhir dapat merubah schema dan membangun model sendiri.
- Sebagai aplikasi OLAP tergantung pada data warehouse dan sistemOLTP untuk memperbaharui sumber tingkatan data mereka.
- Mengurangi aktifitas query dan lalu lintas jaringan pada system OLAP atau pada data warehouse.

Sedangkan kekurangan OLAP, yaitu :

- Tidak terdapat berfungsi untuk melakukan operasi pivot. Saat ini untuk melakukan operasi pivot dengan cara mengubah axis (baris dan kolom) dimensinya.

- Aplikasi OLAP belum memiliki fungsi updating yang berguna pada saat terdapat data hotspot terbaru.

2. OLTP

OLTP (*Online Transaction Processing*) merupakan suatu sistem yang memproses suatu transaksi secara langsung seperti insert, update, delete, melalui komputer yang terhubung dalam jaringan. OLTP berorientasi pada proses yang memproses suatu transaksi secara langsung melalui komputer yang terhubung dalam jaringan. OLTP sangat optimal untuk updating data.

OLTP ini sering kita jumpai disekitar kita seperti toko atau supermarket seperti misalnya kasir pada sebuah super market yang menggunakan mesin dalam proses transaksinya. OLTP mempunyai karakteristik beberapa user dapat creating, updating, retrieving untuk setiap record data, lagi pula OLTP sangat optimal untuk updating data.

Keuntungan dari OLTP, diantaranya :

- Kesederhanaan dan efisiensi untuk bisnis.
- Mengurangi jejak makalah sehingga lebih cepat, lebih akurat dalam perkiraan pendapatan dan beban.

Sedangkan kekurangan dari OLTP adalah :

- Seperti halnya sistem pengolahan informasi, keamanan dan keandalan adalah suatu pertimbangan, bila organisasi memilih untuk mengandalkan OLTP, operasi dapat sangat mempengaruhi jika sistem transaksi atau database karena tidak tersedia.
- Data yang rusak, kegagalan sistem, atau masalah ketersediaan jaringan.
- Selain itu, seperti banyak solusi modern teknologi informasi online, beberapa sistem membutuhkan pemeliharaan offline yang selanjutnya mempengaruhi pada analisa biaya dan manfaat.

3. Perbedaan OLAP dengan OLTP

Perbedaan antara OLAP dengan OLTP adalah :

a. User (Pengguna)

Dalam OLTP, penggunanya adalah IT Professional sedangkan OLAP penggunanya adalah Knowledge worker maksudnya penggunanya adalah seorang yang bertindak dalam subyek tertentu, atau petinggi dalam suatu perusahaan.

b. Data

Dalam OLTP datanya adalah hari ini, update setiap saat sedangkan OLAP datanya adalah sekarang dan hari ini yang berguna untuk melakukan analisis ke depan.

c. Function (Fungsi)

OLTP digunakan sehari-hari untuk proses bisnis seperti toko atau swalayan, sedangkan OLAP digunakan untuk pengambilan keputusan.

d. Design DB (Desain Database)

Desain dalam OLTP bersifat Entity Relational atau databasenya dinormalisasi dulu sebelum digunakan. Untuk OLAP desain databasenya di de-normalisasi.

e. Ukuran Database

Ukuran database untuk OLTP sekitar MB-GB, sedangkan OLAP bisa sampai GB-TB

f. Penggunaan

OLTP digunakan setiap saat, sedangkan OLAP digunakan seperlunya saja.

g. Access (Akses)

OLTP aksesnya bisa write, read dan lain-lain. Sedangkan OLAP sering dibaca karena digunakan untuk analisa.

h. Unit Pekerjaan

OLTP pekerjaannya hanya sederhana misalnya transaksi dalam swalayan. Untuk OLAP query untuk menampilkan data sangat kompleks

i. Jumlah rekaman yang di akses

OLTP sekitar ratusan sampai ribuan, tapi jika OLAP data yang diakses bisa sampai jutaan bahkan milyaran.

j. Jumlah Pengguna

Untuk OLTP penggunaanya adalah puluhan, tapi kalau OLAP penggunaanya bisa sampai ratusan bahkan ribuan

C. BUSINESS INTELLIGENCE

Definisi BI

Business Intelligence (BI) merupakan sekumpulan alat dan teknik yang memungkinkan suatu perusahaan dapat mengubah data bisnis menjadi sebuah informasi yang aktual dan akurat untuk proses pengambilan keputusan, agar dapat digunakan oleh orang-orang yang tepat dengan tujuan paling tepat. *Business Intelligence* ini bekerja dengan cara melakukan ekstraksi data-data operasional perusahaan dan mengumpulkannya dalam sebuah *Data Warehouse*. Selama proses ekstraksi juga dapat dilakukan transformasi dengan menerapkan berbagai formula, agregasi, maupun validasi sehingga didapat data yang sesuai dengan kepentingan analisis bisnis. Kemudian data pada *Data Warehouse* diproses menggunakan berbagai analisis statistik dalam proses *Data Mining*, sehingga didapat berbagai kecenderungan atau pattern dari data. Hasil dari penyederhanaan ini disajikan kepada *end user* yang biasanya merupakan pengambil keputusan bisnis pada suatu perusahaan.

Artificial Intelligence

Artificial Intelligence (AI) adalah suatu bentuk kecerdasan buatan yang diciptakan guna meniru suatu kepintaran ilmiah yang biasanya dimasukkan dalam mesin komputer untuk menyelesaikan suatu masalah. Artificial Intelligence ini merupakan sistem yang mampu memberikan keputusan secara langsung bagi pengguna.

Perbedaan AI dengan BI

Secara umum perbedaan antara *Artificial Intelligence* (AI) dengan Business Intelligence (BI) yaitu, Artificial Intelligence merupakan penentu sebuah keputusan, sedangkan *Business Intelligence* (BI) merupakan pembantu pemberian keputusan.

Teknologi BI

Terdapat beberapa kombinasi teknologi dari *Business Intelligence* yang digunakan untuk mendapatkan hasil yang diinginkan, diantaranya adalah DW (*Data Warehouse*), DM (*Data Mart*), OLAP (*Online Analytical Processing*), DM (*Data Mining*), dan ditambahkan dengan sebuah *Open Data* yang bersifat optional.

1. Data Warehouse
2. OLAP
3. Data Mining

Data mining merupakan sebuah atau serangkaian proses untuk menggali nilai tambah atau pencarian secara otomatis yang berupa informasi berguna dalam tempat penyimpanan data yang besar. Terdapat beberapa istilah lain dari data mining yaitu Knowledge Discovery (mining) in Database (KDD), Knowledge Extraction, Business Intelligence, dan merupakan alat yang penting untuk memanipulasi data dalam penyajian informasi sesuai kebutuhan pengguna dengan tujuan untuk membantu dalam analisis koleksi pengamatan perilaku. Teknik data mining ini biasanya digunakan untuk menjawab pertanyaan dalam dunia bisnis, seperti “Barang apakah yang sering di beli oleh pelanggan?”, “Siapakah pelanggan yang sering berbelanja?”, dan lain sebagainya.

4. Open Data

Open data atau data terbuka merupakan informasi atau data yang tersedia yang bebas digunakan, digunakan kembali (reusable) dan didistribusikan kembali (redistributed) oleh siapa saja. Open data ini dimaksudkan untuk mewujudkan komitmen yang konkret dari kalangan pemerintah dalam peningkatan partisipasi masyarakat, pemberantasan korupsi dan pemanfaatan teknologi-teknologi baru yang bertujuan untuk mewujudkan tata kelola pemerintahan yang lebih terbuka, efektif dan akuntabel, yang disebut birokrasi terbuka atau open government. Tentunya, semakin banyak data yang dibuka, semakin besar pula potensi pemanfaatan dari data tersebut oleh masyarakat. Keuntungan lain dari

open data ini dapat berupa, peningkatan layanan publik, peningkatan kepercayaan publik, dan dapat meningkatkan partisipasi masyarakat.

Tiga Arsitektur DW

Terdapat tiga jenis arsitektur Data Warehouse, yaitu Arsitektur Central, Arsitektur Federal, dan Arsitektur Tiered.

a. Arsitektur Central merupakan arsitektur yang menghimpun semua data dari client yang berbeda menjadi terpusat.

b Arsitektur Federal

Arsitektu Federated merupakan arsitektur yang dimana data disimpan pada tempat penyimpanan yang berbeda-beda, masing-masing client memiliki database yang berbeda.

c. Arsitektur Tiered

Arsitektur Tiered merupakan arsitektur yang dimana data tersebar pada satu gudang data atau gudang data yang bertingkat. Data hanya akan dikumpulkan atau dikurangi bergiliran melalui tingkatan-tingkatan tersebut.

D. BIG DATA

Big Data

Apa itu Big Data ? Jika diterjemahkan secara mentah-mentah maka Big Data berarti suatu data yang berkapasitas besar. Big Data dapat didefinisikan suatu teknologi yang dapat melakukan pengolahan, penyimpanan dan analisis data yang sangat kompleks dalam beragam bentuk/format (Variety), berukuran besar (Volume) dan penambahan data yang sangat cepat (Velocity) yang kemudian akan dianalisa atau diolah lagi untuk keperluan tertentu seperti membuat keputusan (decision making), prediksi, dan lainnya. Pada perusahaan Big Data memiliki manfaat untuk mengolah data yang mereka miliki maupun data eksternal untuk membantu dalam mengambil keputusan-keputusan strategis. Contoh dari Big Data dapat berupa data yang berukuran hingga petabytes atau lebih, seperti milyaran hingga triliunan catatan personal seseorang yang semuanya berasal dari sumber berbeda seperti web, customer service, social media, dan sebagainya, bahkan data-data tersebut dapat berupa video, gambar, logs, binary, dan lainnya.

Big Data dan Data Warehouse

Dalam mengolah data pada Data Warehouse dan Big Data diperlukan infrastruktur yang mumpuni seperti dengan menggunakan supercomputer atau yang sering disebut sebagai high-level computing. Supercomputer bukan seperti komputer-komputer pada umumnya. Supercomputer adalah mesin komputasi yang memiliki kinerja tingkat tinggi dan memiliki kecepatan pemrosesan data yang

sangat cepat. Supercomputer memiliki berbagai macam aplikasi, seperti, scientific calculations, modeling simulations, and rendering large 3D graphics. Supercomputer biasanya memiliki ukuran yang lebih besar dari komputer biasa dan membutuhkan daya/power yang lebih besar dari komputer biasa. Bahkan, sebuah Supercomputer dapat terdiri dari serangkaian komputer yang membutuhkan space ruangan yang besar. Untuk menggunakan infrastruktur yang telah disediakan, maka diperlukan sebuah perangkat lunak yang mampu menangani pemrosesan Data Warehouse dan Big Data. Salah satu perangkat tersebut adalah Apache Hadoop yang merupakan sebuah framework perangkat lunak open-source.

Processing (Pemrosesan Data) dengan Apache Hadoop

Salah Satu Tools yang digunakan oleh big data adalah The Apache Hadoop Software Library , Hadoop merupakan Framework yang memungkinkan untuk melakukan pemrosesan yang terdistribusi dari Data yang besar diantara kluster komputer menggunakan pemrograman yang simple. Framework ini di design agar meningkatkan kemampuan menyimpan dari server tunggal ke ribuan mesin komputer yang masing-masing menyediakan komputasi dan penyimpanan lokal.

- **Analisa Data dengan OLAP**

OLAP (Online Analytical Processing) merupakan metode pendekatan untuk menyajikan jawaban dari permintaan proses analisis yang bersifat dimensional secara cepat, yaitu desain dari aplikasi dan teknologi yang dapat mengoleksi, menyimpan, memanipulasi suatu data multidimensi untuk tujuan analisis. Database yg dikonfigurasi untuk pelayanan OLAP model data multidimensi, bisa digunakan untuk analisis kompleks dan kueri khusus. konsep OLAP adalah Analisis Cepat dari Informasi Multidimensi yang dapat di sharing (FASMI).

- **Transaksi Data dengan OLTP**

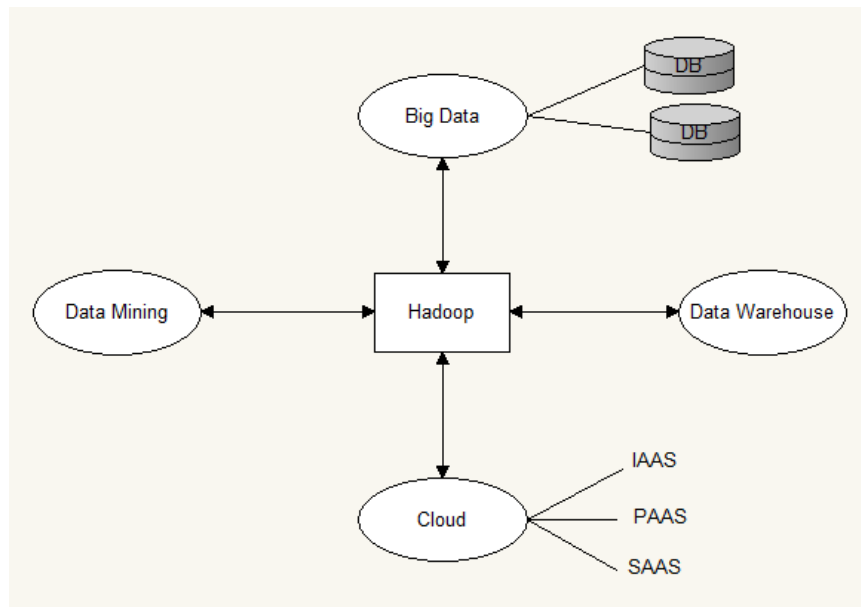
OLTP (Online Transaction Processing) merupakan suatu sistem yang memproses suatu transaksi secara langsung seperti insert, update, delete, melalui komputer yang terhubung dalam jaringan. OLTP berorientasi pada proses yang memproses suatu transaksi secara langsung melalui komputer yang terhubung dalam jaringan. OLTP sangat optimal untuk updating data.

- **ETL (Extract, Transform, Loading)**

ETL (Extract Transform Loading) merupakan fase pemrosesan data dari sumber data masuk ke dalam data warehouse. Tujuan ETL adalah mengumpulkan, menyaring, mengolah dan menggabungkan data yang relevan dari berbagai sumber untuk disimpan ke dalam data warehouse. ETL juga dapat digunakan untuk mengintegrasikan data. ETL merupakan proses yang harus dilalui dalam pembentukan *data warehouse*

Implementasi Big Data

Dalam implementasinya, Big Data berkaitan dengan Apache Hadoop dan No SQL sebagai tools pendukung nya serta Vendor sebagai penyedia layanan Big Data baik berupa hardware maupun software. Hadoop merupakan project opensource dibawah Apache. Hadoop merupakan implementasi opensource project yang dapat memecahkan 3 karakteristik masalah pada Bigdata: Volume, Velocity, dan Variety. Berikut merupakan gambar dari bagan Arsitektur dari implementasi tersebut.

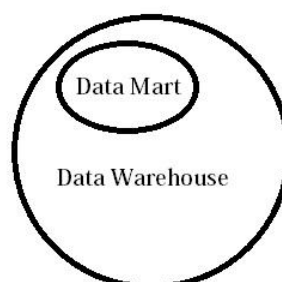


Dalam proses Hadoop yang digambarkan pada gambar diatas yang mengolah data yang sangat besar secara terdistribusi dan berjalan di atas cluster yang terdiri dari beberapa komputer yang saling terhubung, Hadoop berjalan pada karnel Linux dan menggunakan Oprasi Sistem Linux. Hadoop memiliki kaitan dengan Big Data yang terdiri dari database-database, Hadoop juga memiliki kaitan dengan Data Maining dan Data Warehouse yang berada pada proses Analytic, serta Cloud dengan ketiga layanannya yaitu IAAS, PAAS, dan SAAS.

E. DATA MART

Definisi Data Mart

Data mart adalah suatu bagian pada data warehouse yang mendukung pembuatan laporan dan analisa data pada suatu unit, bagian atau operasi pada suatu perusahaan. Jika digambarkan maka akan tampak seperti gambar diagram berikut.



Dalam beberapa implementasi data warehouse, data mart adalah miniature dari data warehouse. Data mart sering digunakan untuk memberikan informasi kepada segmen fungsional organisasi. Contoh umum dari data mart adalah untuk departemen penjualan, departemen persediaan dan pengiriman, departemen keuangan, manajemen tingkat atas, dan seterusnya. Data mart juga dapat digunakan untuk gudang data segmen data untuk mencerminkan bisnis secara geografis terletak di mana masing-masing daerah relatif otonom. Sebagai contoh, sebuah organisasi layanan yang besar mungkin memperlakukan pusat operasi regional sebagai unit usaha perorangan, masing-masing dengan data mart sendiri yang memberikan kontribusi untuk gudang data master. Contoh produk data mart yaitu SmartMart (IBM), Visual Warehouse (IBM), PowerMart (Informatica), dan lain sebagainya.

Keuntungan dan Kerugian Data Mart

Keuntungan :

Data mart dapat meningkatkan waktu respon pengguna akhir, karena berisi data mentah yang memungkinkan sistem komputer untuk fokus pada satu tugas, sehingga meningkatkan kinerja. Berbeda dengan sistem OLTP, data mart juga dapat menyimpan data historis yang memungkinkan pengguna untuk menganalisis kecenderungan data. Selain itu, data mart tidak begitu mahal dan kompleks sebagai data gudang untuk setup dan melaksanakan karena masalah teknis tidak begitu sulit untuk diselesaikan.

Kerugian :

Data mart memiliki nilai yang terbatas karena data mart tidak dapat melihat organisasi secara keseluruhan dan pelaporan dan analisis potensi terbatas.

Perbedaan Data Mart dengan Data Warehouse

Data warehouse merupakan gabungan dari beberapa data mart dan levelnya berada pada perusahaan atau organisasi. Sedangkan data mart merupakan bagian dari datawarehouse dan berada level departemen pada perusahaan atau organisasi tersebut. Data mart menangani sebuah business proses, misalkan penjualan, maka hanya proses penjualan saja yang ditangani pada data mart, Perbandingan antara data warehouse dan data mart dapat dilihat pada tabel berikut ini.

Tabel 1. Perbedaan Data Mart dengan Data Warehouse

	Data Warehouse	Data Mart
Lingkupan (Scope)	Perusahaan/Enterprise	Departemen
Subjek (Subjects)	Multiple	Single
Sumber data (Data Sources)	Banyak	Sedikit
Ukuran data	100 GB sampai > 1 TB	<100GB

Waktu implementasi	Berbulan-bulan bahkan sampai bertahun-tahun	Beberapa tahun
--------------------	---	----------------

ETL dan ELT

ETL (Extraction, Transformation, Loading) merupakan proses yang harus dilalui dalam pembentukan data warehouse, yang dimana pada proses ini yaitu semua data akan di extract terlebih dahulu kemudian data akan disaring dan data yang akan diambil yaitu data yang penting, pada proses transformation data akan diubah sesuai format yang disetujui dan data di load ke data warehouse, kemudian data akan diolah selanjutnya akan di teruskan ke user. Untuk lebih jelasnya berikut merupakan penjelasan dari tiap proses.

a. Extrsact (Ekstraksi Data)

Ekstraksi data adalah proses dimana data diambil atau diekstrak dari berbagai sistem operasional, baik menggunakan query, atau aplikasi ETL.

b. Transformation (Transformasi Data)

Transformasi adalah proses dimana data mentah (raw data) hasil ekstraksi disaring dan diubah sesuai dengan kaidah bisnis yang berlaku

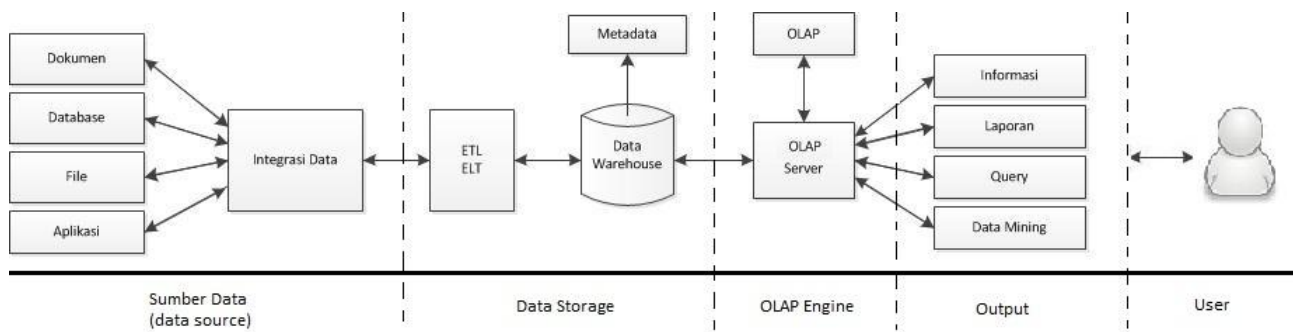
c. Loading (Pengisian Data)

Proses terakhir yang perlu dilakukan adalah proses pemuatan data yang didapatkan dari hasil transformasi ke dalam data warehouse.

ELT (Extracition, Loading, Transformation)

ELT (Extracition, Loading, Transformation) adalah variasi dari ETL (Extraction, Transformation, Loading), proses integrasi data di mana transformasi berlangsung pada server menengah sebelum dimuat ke sasaran. Sebaliknya, ELT memungkinkan data mentah yang akan dimuat langsung ke sasaran dan berubah di sana. Kemampuan ini sangat berguna untuk memproses set data yang besar diperlukan untuk intelijen bisnis (BI) dan analisis data yang besar. Sebenarnya, ELT hampir sama dengan ETL, hanya saja pada ELT setelah mengekstrak data dari sumber data, ELT meloading data ke dalam data warehouse kemudian barulah data tersebut di diolah sesuai kebutuhan.

Bagan



Berdasarkan dengan gambar diatas terdapat 4 bagian yaitu pada bagian pertama merupakan Sumber Data (Data Sources), bagian kedua merupakan penyimpanan data (Data Storage), bagian ketiga yaitu OLAP Engine dan pada bagian keempat yaitu Output yang merupakan hasil keluaran dari semua proses yang ada. Kemudian ke 4 bagian tersebut akan dimanfaatkan oleh pengguna (perorangan maupun organisasi) sesuai dengan kebutuhan-kebutuhan yang ada misalnya sebagai DSS (Decision Support System), Business Intelligence, Smart Economic, Big Data serta kebutuhan lainnya didalam mendukung proses bisnis.

Pada bagian pertama terlihat beberapa sumber data seperti document, database, file, aplikasi dan sebagainya merupakan hasil dari proses OLTP yaitu data yang tersimpan setelah melakukan transaksi data seperti insert, update, dan delete. Kemudian sumber-sumber data tersebut akan diintegrasikan dan akan diteruskan menggunakan proses ETL ataupun ELT yang terdapat pada bagian kedua, sesuai dengan kebutuhan untuk disimpan pada data warehouse dan setiap data pada data warehouse dilengkapi oleh metadata. Kegunaan dari metadata ini adalah untuk menyimpan data dari sebuah data yang dikumpulkan menjadi sebuah informasi mengenai data tersebut seperti panjang field, tipe field, tanggal, dan lain sebagainya. Kemudian pengguna akan memberikan perintah untuk menampilkan informasi, OLAP Engine yang terdapat pada bagian keempat akan melakukan analisis terhadap data yang telah tersimpan pada data warehouse. Kemudian, hasil olahan tersebutlah yang akan ditampilkan kepada pengguna sesuai dengan yang diinginkan oleh pengguna tersebut baik dalam format informasi, pelaporan, query, hasil data mining dan lain sebagainya.

Daftar Pustaka

- [1] <https://id.scribd.com/doc/227593819/PAPER-Data-Warehouse> [diakses tanggal 24 Oktober 2015].
- [2] <https://www.academia.edu/5354417/Makalah-data-warehouse> [diakses tanggal 24 Oktober 2015].
- [3] <http://edipuspayasa.blogspot.co.id/2015/09/olap-dan-oltp.html> [diakses tanggal 24 Oktober 2015].
- [4] <http://edipuspayasa.blogspot.co.id/2015/09/data-warehouse-dw-dan-business.html> [diakses tanggal 24 Oktober 2015].
- [5] <http://edipuspayasa.blogspot.co.id/2015/10/data-warehouse-dan-big-data.html> [diakses tanggal 24 Oktober 2015].
- [6] <http://edipuspayasa.blogspot.co.id/2015/10/data-mart-data-warehouse-etl-elt-dan.html> [diakses tanggal 24 Oktober 2015].