NAMA : RAHMA FITRI ASRIANI

KELAS : A

NIM : 11551201988

1. Pengertian Data Warehouses

Data warehouse adalah basis data yang saling berhubungan dan dapat diambil suatu informasi yang dapat di olah, serta bersifat orientasi subjek, terintegrasi, time variant, tidak berubah yang juga dapat digunakan untuk membantu pengambil keputusan.

Penerapan Data Warehouse di Perusahaan :

* Retailing dan sales, memprediksi penjualan, mencegah pencuri dan kecurangan, dan menentukan level inventori yang benar dan mendistribusikan jadwal sepanjang outlet. Sebagai contoh, penjual seperti AAFES menggunakan Fraud Watch dari Triversity untuk melawan penipuan oleh pekerja di 1400 toko.
* Banking. Level prediksi untuk peminjaman yang buruk dan kecurangan dalam penggunaan kartu kredit, prediksi pengeluaran kartu kredit oleh pelanggan yang baru, dan menentukan jenis mana dari pelanggan yang memberikan respon yang baik untuk menawarkan peminjaman.
* Manufacturing dan production, memprediksi kesalahan mesin, dan menemukan faktor kunci yang dapat mengoptimasi kapasitas produksi.
* Insurance. Memprediksi jumlah tagihan dan ongkon kesehatan, mengklasifikasikan elemen yang paling penting yang mempengaruhi jaminan kesehatan dan memoprediksi pelanggan mana yang akan membeli polis asuransi baru.
* Policework. Menemukan pola kejahatan, lokasi, dan perilaku tindak kejahatan; mengidentifikasi atribut yang berguna untuk menyelesaikan masalah kriminal.
* Healthcare. Korelasi demografis dari pasien yang memiliki penyakit yang parah, dan memperluas wawasan yang lebih baik tentang bagaimana mengidentifikasi dan gejala penyakit dan penyebabnya.
* Marketing. Mengklasifikasikan demografis pelanggan yang dapat digunakan untuk memprediksi pelanggan mana yang akan merespon untuk pembelian produk tertentu.
* Analisa Perusahaan dan Manajemen Resiko
* Perencanaan Keuangan dan Evaluasi Aset ,data Mining dapat membantu untuk melakukan analisis dan prediksi cash flow serta melakukan contingent claim analysis untuk mengevaluasi aset. Selain itu juga dapat menggunakannya untuk analisis trend.
* Perencanaan Sumber Daya (Resource Planning) ,dengan melihat informasi ringkas (summary) serta pola pembelanjaan dan pemasukan dari masing-masing resource, dapat memanfaatkannya untuk melakukan resource planning.
* Persaingan (Competition) Sekarang ini banyak perusahaan yang berupaya untuk dapat melakukan competitive intelligence. Data Mining dapat membantu untuk memonitor pesaing-pesaing dan melihat market direction mereka.
* Dapat melakukan pengelompokan customer dan memberikan variasi harga/layanan/bonus untuk masing-masing grup.
* Menyusun strategi penetapan harga di pasar yang sangat kompetitif. Hal ini diterapkan oleh perusahaan minyak REPSOL di Spanyol dalam menetapkan harga jual gas di pasaran.

Contoh Perusahaan Yang Menerapkan Data Warehouse :

* Telekomunikasi  
  Sebuah perusahaan telekomunikasi menerapkan data warehouse dan data mining untuk melihat dari jutaan transaksi yang masuk, transaksi mana sajakah yang masih harus ditangani secara manual (dilayani oleh orang), Tujuannya tidak lain adalah untuk menambah layanan otomatis khusus untuk transaksi-transaksi yang masih dilayani secara manual. Dengan demikian jumlah operator penerima transaksi manual tetap bisa ditekan minimal.
* Keuangan  
  Financial Crime Enforcement Network di Amerika Serikat baru-baru ini menggunakan menggunakan data minimng untuk menambang trilyunan daru berbagai subjek seperti property, rekening bank dan transaksi keuangan lainnya untuk mendeteksi transaksi-transaksi keuangan yang mencurigakan seperti money laundry. Mereka menyatakan bahwa hal tersebut akan susah dilakukan jika menggunakan analisis standar.
* Asuransi  
  Australian Health Insurance Commision menggunakan data mining untuk mengidentifikasi layanan kesehatan yang sebenarnya tidak perlu tetapi tetap dilakukan oleh peserta asuransi. Hasilnya? mereka berhasil menghemat satu juta dollar pertahun.
* Olahraga  
  IBM Advanced Scout menggunakan data mining untuk menganalisis statistik permainan NBA (jumlah shots blocked, assists dan fouls) dalam rangka mencapai keunggulan bersaing (competitive advantage) untuk tim New York Knick dan Miami Heat.
* Astronomi  
  Jet Propulsion Laboratory (JPL) di Pasadena, California dan Palomar Observartory berhasil menemukan 22 quasar dengan bantuan data mining. Hal ini merupakan salah satu kesuksesan penerapan data mining di bidang astronomi dan ilmu ruang angkasa.
* Internet Web Surf-Aid  
  IBM Surf-Aid menggunakan algoritma data mining untuk mendata akses halaman web khususnya yang berkaitan dengan pemasaran guna melihat prilaku dan minat costumer serta melihat keefektifan pemasaran melalui web.

1. Karakteristik Datawarehouse

* Subject oriented

Database yang semua informasi yang tersimpan di kelompokkan berdasarkan subyek tertentu misalnya : pelanggan, gudang, pasar, dsb. Data yang disusun menurut subyek berisi hanya informasi yang penting bagi pemprosesan decision support. Semua Informasi tersebut disimpan dalam suatu sistem *data warehouse*.Data-data di setiap subyek dirangkum ke dalam dimensi, misalnya : periode waktu, produk, wilayah, dsb, sehingga dapat memberikan nilai sejarah untuk bahan analisa.

* Integrated

Jika data terletak pada berbagai aplikasi yang terpisah dalam suatu lingkungan operasional, encoding data sering tidak seragam sehinggga bila data dipindahkan ke data warehouse maka coding akan diasumsikan sama seperti lazimnya. Data di dalam warehouse distruktur berdasarkan pada model korporasi secara keseluruhan melewati batas fungsional dari kebiasaan yang ada. Ini mencakup standar penamaan, sistem pengukuran dan perulangan

* Time variant

Data warehouse adalah tempat untuk storing data selama 5 sampai 10 tahun atau lebih, data digunakan untuk perbandingan atau perkiraan dan data ini tidak dapat diperbaharui. Data di dalam data warehouse mempunyai karakter khusus berupa time-series dalam bentuk data historical Data terdiri dari suatu seri dari “pemotretan” keadaan yang ditandai dengan waktu dan nilai data pada saat “pemotretan” tersebut. Ini dapat digunakan untuk melakukan trend analysis dari data tersebut.

* Non-volatile

Data tidak dapat diperbaharui atau dirubah tetapi hanya dapat ditambah dan dilihat. Data warehouse tidak diupdate secara terus menerus (seperti insert, delete dan change) seperti dalam sebuah sistem OLTP. Data di dalam data warehouse secara periodik di upload dalam jangka waktu yang sama.

1. Kelebihan dan kekurangan penggunaan data warehouse

Kelebihan Data warehouse :

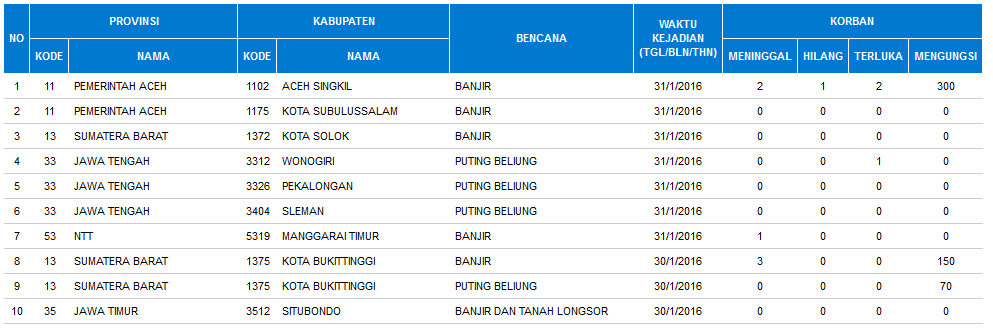
1. Data terorganisir dengan baik untuk query analisis dan sebagai bahan yang baik untuk proses transaksi.
2. Perbedaan struktur data yang banyak macamnya dari sumber yang berbeda dapat diatasi.
3. Memiliki aturan transformasi untuk memvalidasi dan menkonsolidasi data dari OLTP ke Data warehouse.
4. Masalah keamanan dan kinerja dapat dipecahkan tanpa perlu mengubah sistem produksi.

Kekurangan Data Warehouse :

1. Datawarehouse bukan merupakan lingkungan yang cocok untuk data yang tidak terstruktur.
2. Data perlu di ekstrak, diubah (ETL) dan di load ke datawarehouse sehingga membutuhkan tenggang waktu untuk memindahkannya.
3. Semakin lama dipelihara, semakin besar biaya untuk merawat sebuah datawarehouse.
4. Datawarehouse dapat menjadi ketinggalan dari data terbaru yang relatif cepat, karena data yang digunakan di datawarehouse tidak di update secara cepat. sehingga data yang ada tidak optimal.
5. Star dan Snowflake Schema

* Star Schema

Sebagai contoh kasus, kita akan mencoba membuat suatu star schema dari data http://dibi.bnpb.go.id/data-bencana

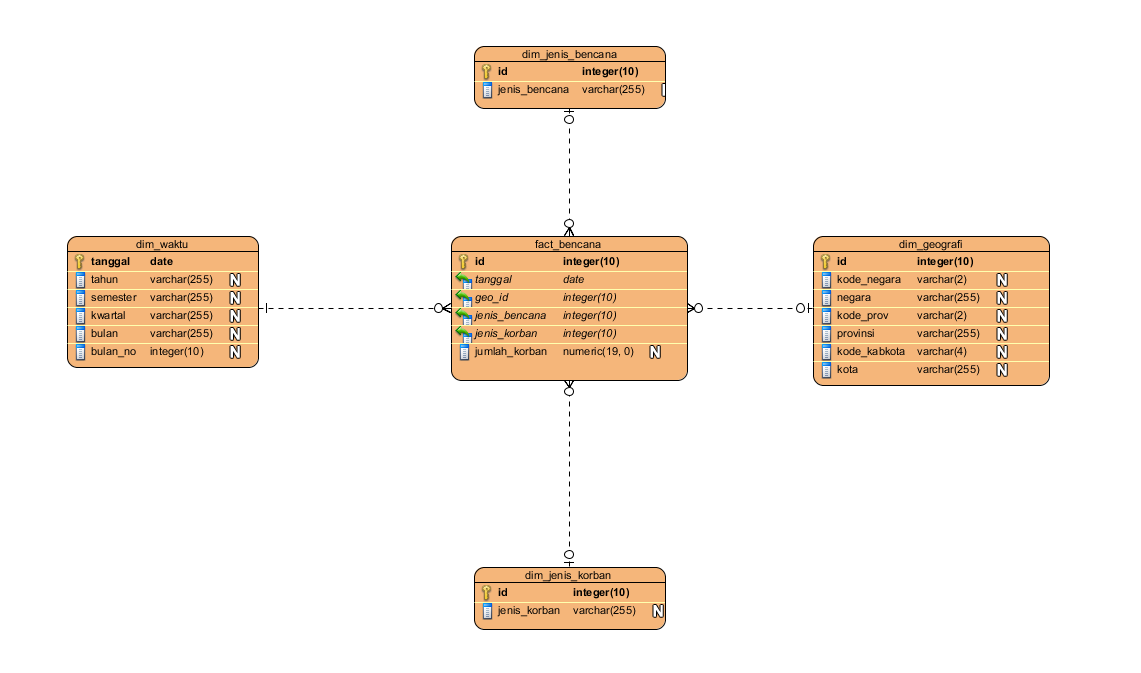


Kalau lihat data di atas, maka hanya ada satu measure yaitu "jumlah korban". Ada berapa dimensi? Paling tidak ada 4 dimensi, yaitu:

1. Dimensi waktu (waktu kejadian (tgl/bln/tahun)
2. Dimensi geografi (provinsi, kabupaten)
3. Dimensi bencana (banjir, puting beliung, dll)
4. Dimensi jenis korban (meninggal, hilang, dll)

Dimensi waktu adalah dimensi yang hampir selalu ada dalam setiap Business Intelligence, sehingga semua solusi BI biasanya memiliki pre-defined tabel dimensi ini. Di Tableau misalnya, begitu suatu kolom type nya adalah tanggal, maka otomatis dikenali sebagai dimensi waktu lengkap dengan atribut lainnya seperti bulan, kwartal, semester, tahun, dll. Dengan dimensi waktu yang predefined ini kita hampir tidak perlu membuat table dimensi sendiri, namun untuk kasus ini kita akan membuat table dimensi waktu sendiri.

Dengan dimensi dan measure di atas, maka star schema nya menjadi sbb



Selanjutnya, setelah kita mendapatkan Star Schema yang sesuai adalah memasukkan data yang ada ke masing-masing table ini.

Untuk table dimensi korban hanya ada 4 baris, yaitu:

|  |  |
| --- | --- |
| 1 | Meninggal |
| 2 | Hilang |
| 3 | Terluka |
| 4 | Mengungsi |

Untuk table dimensi bencana berisi

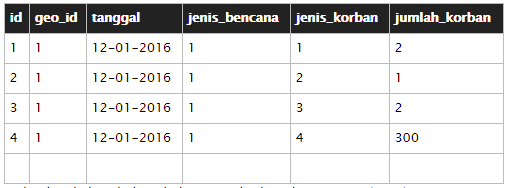
|  |  |
| --- | --- |
| 1 | Banjir |
| 2 | Angin Puting Beliung |
| 3 | Tanah Longsor |
| 4 | Kebakaran Hutan |
| 5 | Gempa Bumi |

Untuk dimensi geografi



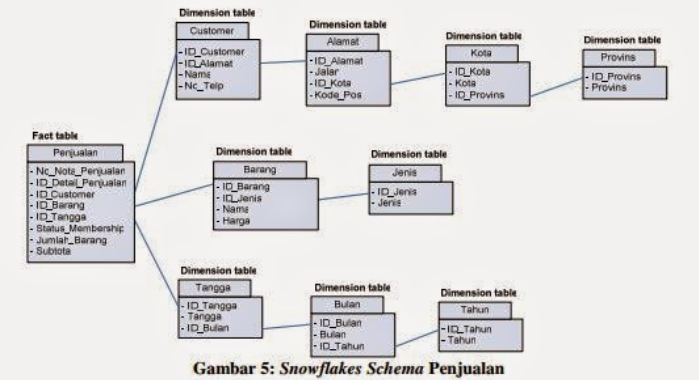
Kita bisa saja membuat listnya secara manual seperti di atas, atau kita bisa gunakan ETL untuk memasukkan data ini secara program. Kita akan bahas ETL di tulisan berikut.

Untuk table facta disusun dari kolom-kolom dimensi dan measure tablenya menjadi sbb.

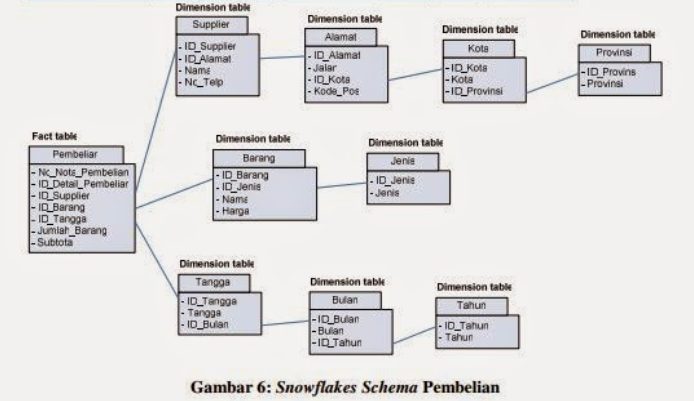


Perhatikan bahwa kolom-kolum jenis korban di transpose (pivot)

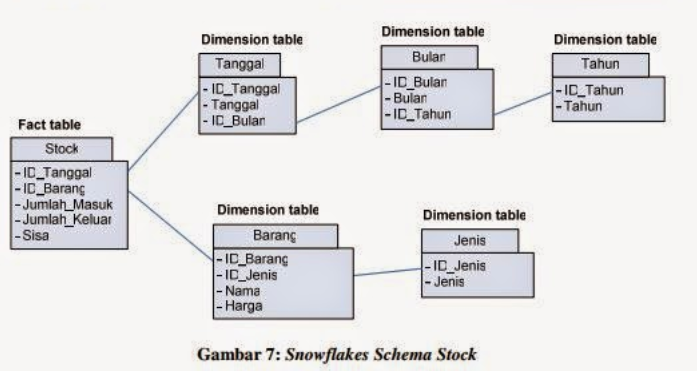
* Perancangan Snowflake Schema



Snowflakes schema merupakan pengembangan dari star schema, dimana setiap dimension table yang ada dipecah kembali menjadi bentuk yang lebih sederhana sesuai dengan kebutuhan pengambilan keputusan. Gambar 5 di atas merupakan gambar snowflakes schemauntuk proses penjualan. Star schema pembelian dapat dipecah kembali menjadi bentuk yang lebih detail pada dimension table-nya sehingga akan menghasilkan sebuah snowflake schema. Berikut adalah gambar dari snowflake schemauntuk proses pembelian :



Sedangkan gambar dibawah ini merupakan gambar snowflakes schemadari proses stock control:



Pada ketiga snowflakes schema diatas diketahui bahwa tiap-tiap dimension table mengalami normalisasi. Hal ini bertujuan untuk mengetahui hirarki dari tabel-tabel yang ada sehingga dapat mendukung pengambilan keputusan berdasarkan hirarki tersebut.

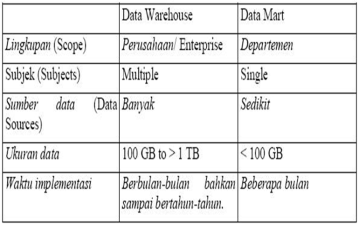
1. Perbedaan antara data warehouse dan data mart adalah sebagi berikut:

* Data warehouse merupakan gabungan dari beberapa data mart dan levelnya berada pada perusahaan atau organisasi.
* 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.

Data warehouse sendiri adalah database yang berisi data dari beberapa system operasional yang terintegrasi dan terstruktur sehingga dapat digunakan untuk mendukung analisa dan proses pengambilan keputusan dalam bisnis. Data warehouse didesain untuk kita bisa melakukan query secara cepat. Informasi diturunkan dari data lain, dilakukan rolling up untuk dijadikan ringkasan, dilakukan operasi drilling down untuk mendapatkan informasi lebih detail, atau melihat pola yang menarik atau melihat trend (kecenderungan).

Kemudian yang dimaksud data mart adalah lapisan akses dari data warehouse lingkungan yang digunakan untuk mendapatkan data keluar bagi pengguna. Data mart adalah subset dari data warehouse yang biasanya berorientasi pada lini bisnis tertentu atau tim. Dalam beberapa penyebaran, masing-masing departemen atau unit bisnis dianggap sebagai pemilik data mart yang termasuk semua perangkat lunak, perangkat keras dan data.

Untuk lebih jelasnya perbedaan data warehouse dan data mart adalah sebagai berikut:

[](https://aullyaarvianto.files.wordpress.com/2014/01/dat.png)