1. **Pengertian Datawarehouse**

Data warehouse sebagai gabungan dari beberapa sistem yang terintegrasi dan didesain untuk menyediakan informasi yang dibutuhkan untuk pengambilan keputusan

1. Fact table adalah table yang berisikan nilai2 numerik. Nilai numerik ini bisa berupa fakta dan foreign key. Nilai2 numerik tersebut biasanya disebut dengan measure.

Dimension table adalah tabel2 yang berisi data grouping dari data pengukuran yang dilakukan.

1. **Kapan top down dan up down dilakukan?**

* **Pendekatan Top Down (Top Down Approach)**

Pendekatan ini dilakukan dengan membuat perancangan data warehouse terlebih dahulu secara keseluruhan. Dalam pembuatannya, ditentukan apa yang menjadi sumber data dan bagaimana cara untuk memasukkannya ke dalam data warehouse yang telah dirancang terlebih dahulu. Pendekatan ini menyatakan bahwa pembangunan desain arsitektur data warehouse adalah hal pertama yang dilakukan dalam membangun suatu data warehouse.

* **Pendekatan Bottom Up (Bottom Up Approach)**

Berbeda dengan pendekatan Top Down, pendekatan ini menyatakan bahwa data warehouse berawal dari kumpulan data mart yang telah dibangun terlebih dahulu untuk masing-masing departemen. Data mart yang telah dibangun kemudian digabungkan untuk membangun suatu data warehouse yang utuh.

1. **Perbedaan Dependent Data Mart dan Independent Data Mart (IDM).**

Perbedaan dari kedua arsitektur tersebut hanya terletak pada ketergantungannya terhadap data warehouse. Pada Dependent Data Mart, data yang diperoleh sangat tergantung pada data warehouse terpusat sedangkan pada IDM, data yang digunakan terpisah dari data warehouse terpusat dan bersifat independent (berdiri sendiri). Perbandingan antara data warehouse dan data mart dapat dilihat pada tabel berikut ini.

1. **Karekteristik yang dimiliki oleh sebuah data warehouse, yaitu :**

* Orientasi Subyek.

Suatu data warehouse didesain dan dibangun secara khusus dari database transaksional berdasarkan keperluan perusahaan, semisal data warehouse untuk kastemer, sales, dan lain-lain. Hanya data yang benar-benar diperlukan yang dimasukkan kedalam database.

* Integrasi data.

Untuk mendapatkan informasi yang diinginkan, data yang ada dalam data warehouse akan dibangun dari beberapa macam sistem yang ada diperusahaan. Kemudian data tersebut ditransformasi dan diload. Karena itu ketika disimpan ke dalam data warehouse data tersebut akan diintegrasikan sehingga hanya ada satu cara dan atribut dengan format dan unit yang sama.

* Nonvolatile

Pada database transaksional, operasi yang dilakukan adalah operasi update (insert, delete, dan update). Sedangkan dalam data warehouse metode ini tidak digunakan. Data disimpan ke dalam data warehouse pada periode waktu tertentu setelah dilakukan beberapa perhitungan (calculation) dan rangkuman (summary).

* Setiap unit data akan relevan dengan waktu tertentu

Setiap data yang dimasukkan ke data warehouse pasti memiliki dimensi waktu. Dimensi waktu ini dipergunakan sebagai pembanding dalam perhitungan untuk menghasilkan laporan yang diinginkan. Selain itu, dengan menggunakan dimensi waktu pembuat keputusan dapat mengenal kecenderungan (trend) dan pola dari suatu data.

1. **Kapan oltp diperlukan utk transaksi ?**

Data Warehouse terbagi kedalam 2 sistem yaitu OLTP (menangani transaksi) dan OLAP (bagian analisis data). Secara umum dapat dijelaskan bahwa sistem OLTP menyediakan sumber data untuk Data Warehouse, kemudian sistem OLAP akan melakukan analisis terhadap data tersebut.

* OLTP

OLTP (Online Transactional Processing) merupakan sekumpulan fungsi yang bekerja secara bersama-sama dalam mengelola, mengumpulkan, menyimpan, memproses serta mendistribusikan informasi. OLTP (On-line Transaction Processing) memiliki karakteristik dengan jumlah data yang besar namun transaksi yang dilakukan cukup sederhana seperti insert,update, dan delete.

* OLAP

OLAP (Online Analytical Proccessing) merupakan suatu proses yang digunakan untuk melakukan permintaan terhadap data dalam bentuk yang kompleks dan menganalisa data yang bervolume besar. OLAP merupakan teknologi yang memproses data di dalam database dalam struktur multidimensi, menyediakan jawaban yang cepat untuk query dan analisis yang kompleks. Data multidimensi adalah data yang dapat dimodelkan sebagai atribut dimensi dan atribut ukuran.

1. **Kapan peristiwa ETL terjadi ?**

ETL adalah sekumpulan proses yang harus dilalui dalam pembentukan data warehouse. Tujuan ETL adalah mengumpulkan, menyaring, mengolah dan menggabungkan datadata yang relevan dari berbagai sumber untuk disimpan ke dalam data warehouse. Proses ETL sendiri terdiri dari extracting, transforming, loading, Berikut adalah penjelasan dari tiap proses :

* Extract

Extract adalah proses memilih dan mengambil data dari satu atau beberapa sumber dan membaca/mengakses data yang dipilih tersebut. Proses ini dapat menggunakan query, atau aplikasi ETL. Sebaiknya sebelum proses extract kita lakukan, akan lebih mudah jika user sudah mendefinisikan kebutuhan terhadap sumber data yang akan kita butuhkan.

* Transform

Pada Proses ini data yang telah diambil pada proses extract akan dibersihkan dan mengubah data dari bentuk asli menjadi bentuk yang sesuai dengan kebutuhan data warehouse. Kendala yang biasanya terjadi pada proses transform adalah sulitnya menggabungkan data dari beberapa sistem yang harus dibersihkan sehingga data bersifat konsisten.

* Load

Load adalah proses terakhir yang berfungsi untuk memasukkan data ke dalam target akhir, yaitu ke dalam data warehouse. Cara untuk memasukkan data adalah dengan menjalankan SQL script secara periodik.Pada proses ini akan mengubah data kedalam bentuk Dimensional Data Store agar format data cocok untuk diterapkan pada proses analisis dan telah terintegrasi dengan beberapa sumber data. Proses Load yang termasuk proses terakhir dalam ETL akan sampai ke berbagai macam output yang sesuai dengan skemanya, yaitu terdiri dari proses load-up data (lodupd), load-insert data (lodins), dan load bulk data (lodbld).

Data warehouse tidak mungkin ada tanpa adanya proses ETL karena Proses ETL merupakan suatu landasan dari sebuah data warehouse. Proses ETL ini sangat penting karena sangat berperan terhadap kualitas data dalam data warehouse, sehingga data warehouse nantinya dapat digunakan untuk keperluan business intelligence atau aktivitas analisis yang lain. Dikatakan Sebuah proses ETL berjalan dengan benar, jika pada proses itu melibatkan beberapa hal yaitu akan adanya proses mengekstraksi data dari sebuah sumber, mempertahankan kualitas data tersebut, menerapkan aturan-aturan standar, dan menyajikan data dalam berbagai bentuk, sehingga dapat digunakan dalam proses pengambilan keputusan

1. **Perbedaan Database dan Data warehouse**

Perbedaan database vs data warehouse secara detail dijelaskan berikut ini:

* Database dipergunakan dalam Transaksional Pengolahan Online (OLTP) dan dapat pula dipergunakan untuk kebutuhan data warehouse
* Data warehouse dipergunakan dalam Online Analytical Processing (OLAP) yang mampu membaca data historis bagi pengguna dalam proses pengambilan keputusan
* Tabel dan gabungan dalam database sangat kompleks sehingga dapat mengurangi data berlebihan serta menghemat ruang penyimpanan.
* Tabel dan gabungan dalam data warehouse tergolong sederhana untuk mengurangi waktu respon dan permintaan analitis.
* Database berorientasi pada pelanggan (customer oriented) dan dipergunakan dalam proses query serta transaksi, misalnya oleh kasir, pramuniaga, serta para profesional IT.
* Data warehouse berorientasi pada pasar (market oriented) dan dipergunakan dalam proses analisa data oleh orang-orang yang memiliki profesi knowlegde field, misalnya eksekutif, manager, serta para ahli analis data.
* Database mengatur dan mengolah data secara detil sehingga mudah dipergunakan dalam proses pengambilan keputusan.
* Data warehouse mengatur dan mengolah data-data historis dalam jumlah besar dan menyediakan beberapas fasilitas untuk keperluan menjumlah penjualan total serta menggabungkan penjualan pada tiap area, tiap bulan, dan seterusnya. Data warehouse menyimpan dan mengolah data menurut tingkat granularitas yang berbeda.
* Sistem database umumnya mengambil pemodelan data “entity relationship” serta desain yang berorientasi pada aplikasi.
* Sistem data warehouse umumnya mengambil pemodelan “snowflake” atau “star” serta desain yang berorientasi pada subyek.
* Database terfokus pada data real time dalam suatu organisasi atau instansi tanpa mengacu pada data-data historus maupun data-data dari organisasi lainnya.
* Data warehouse sering menjangkau berbagai tipe skema database dikarenakan proses evolusioner dari sebuah organisasi. Data warehouse juga berhubungan dengan berbagai informasi yang didapatkan dari beberapa organisasi lainnya, yang mana menggabungkan dari berbagai tempat penyimpanan data. Hal ini akan menyebabkan jumlah volume yang sangat besar sehingga data warehouse disimpan dalam beberapa tempat penyimpanan.
* Pola akses pada database terutama terdiri dari transaksi atomik yang pendek dan sistem seperti ini membutuhkan mekanisme “concurency control” serta “recovery”.
* Pola akses pada data warehouse sebagian besar adalah “read only”. Hal ini karena sebagian besar data yang tersimpan adalah data-data historis dan hanya sedikit data-data baru).