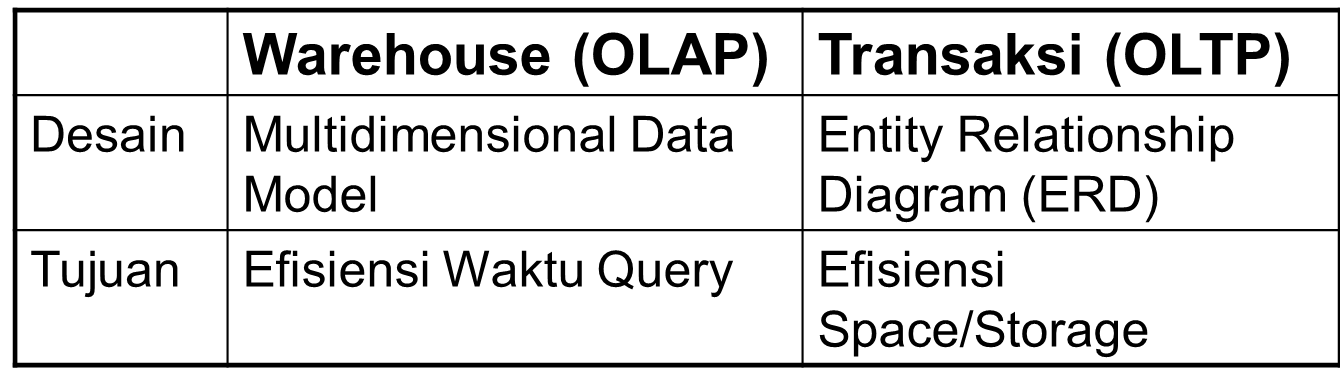
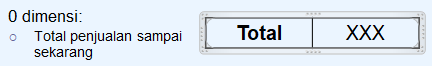
Rangkuman Data Warehouse

Bab I – Data Warehousing

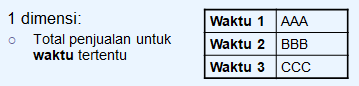
* **Data warehouse** merupakan database yang berisi data dari beberapa sistem operasional yang terintegrasi, teragregrasi dan terstruktur sehingga dapat digunakan untuk mendukung analisa dan proses pengambilan keputusan dalam bisnis,
* **Data warehouse** merupakan relasional database yang didesain untuk melakukan query dan analisa, bukannya untuk melakukan proses transaksi. Data warehouse biasanya berisi data histori yang didapatkan dari data transaksi.
* Selain relasional database, environment data warehouse juga mendukurng fitur *extraction, transportation, transformation & loading* (ETL), dan *Online Analytical Processing* (OLAP) serta kemampuan data mining, client analysis tools.
* **Mengapa data warehouse dibutuhkan ?**
  + Diperlukan adanya detail histori transaksi untuk keperluan analisa
  + Diperlukannya akses data untuk data histori lama
  + Membantu mengefisienkan penyimpanan dengan menyesuaikan level detail yang memang benar-benar dibutuhkan
  + Diperlukannya analisa data meskipun ada reorganisasi data
* **Data warehouse memiliki beberapa karakteristik, yaitu**
  + **Subject Oriented**, maksudnya, data warehouse dapat digunakan untuk menganalisa dan mendefinisikan informasi berdasarkan subject yang ada pada data warehouse tersebut, ***contohnya*** adalah ketika sebuah data warehouse digunakan untuk mencari bulan apa yang memiliki tingkat sales tertinggi di tahun 2017, dalam hal ini nsubjectnya adalah tingkat sales tertinggi
  + **Integrated,** maksudnya data warehouse dapat mendapatkan data dari banyak sumber, sehingga harus dilakukan penyesuaian terhadap masalah-masalah yang muncul (penamaan field yang berbeda, inkonsistensi pada unit pengukuran, dll)
  + **Nonvolatile,** maksudnya sekali data masuk ke data warehouse, data tidak akan dapat /boleh berubah.
  + **Time Variant/Multidimensional,** data warehouse terfokus pada perubahan waktu yang terjadi, hal ini disebabkan karena untuk menganalisa suatu informasi, data yang dibutuhkan berjumlah sangat besar dan harus meliputi varian waktu yang sudah lalu.
* **Desain data warehouse**
  + Data warehouse didesain untuk bisa melakukan query secara cepat, sehingga data warehouse terpisah dari database operasional
  + Terdapat perbedaan antara desain datawarehouse (OLAP) dan data transaksi (OLTP) antara lain,



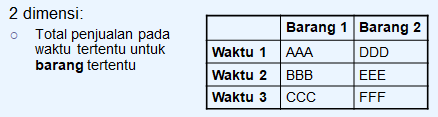
* + Multidimensional modeling merupakan modeling untuk data yang ada pada data warehouse, ciri-cirinya adalah
    - Subject oriented, yaitu melihat data dari berbagai persepektif
    - Berisikan data yang
      * Telah tervalidasi
      * Historikal
      * Terintegrasi
      * Mudah diakses
    - Direpresentasikan dalam bentuk **Data Cube**
  + **Data Cube** adalah representasi kumpulan data dalam multi-dimensi, msekipun bernama cube, namun data cube dapat direpresentasikan dalam N-dimensi
    - **0 Dimensi**



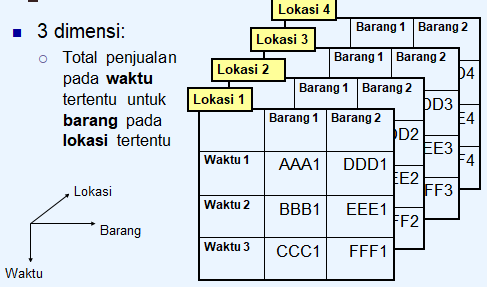
* + - **1 Dimensi**



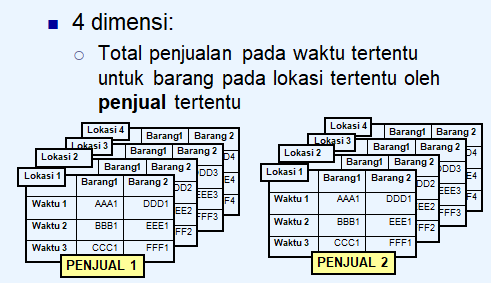
* + - **2 Dimensi**



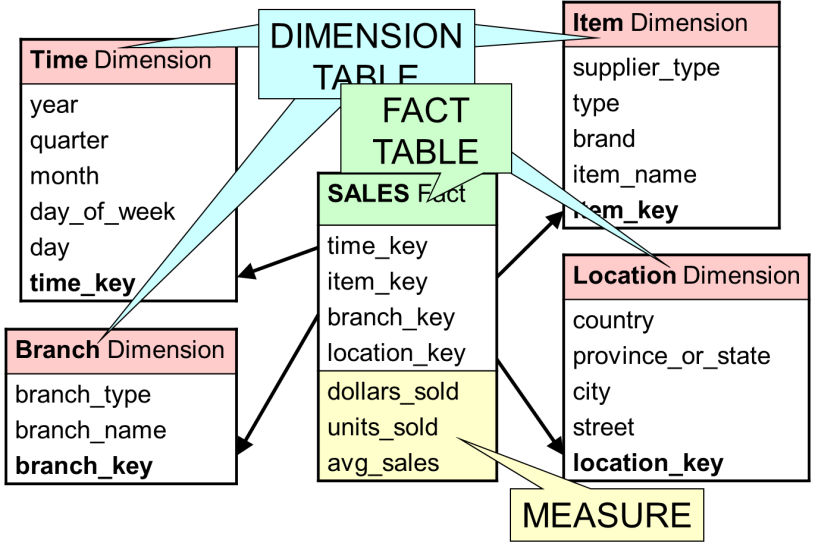
* + - **3 Dimensi**



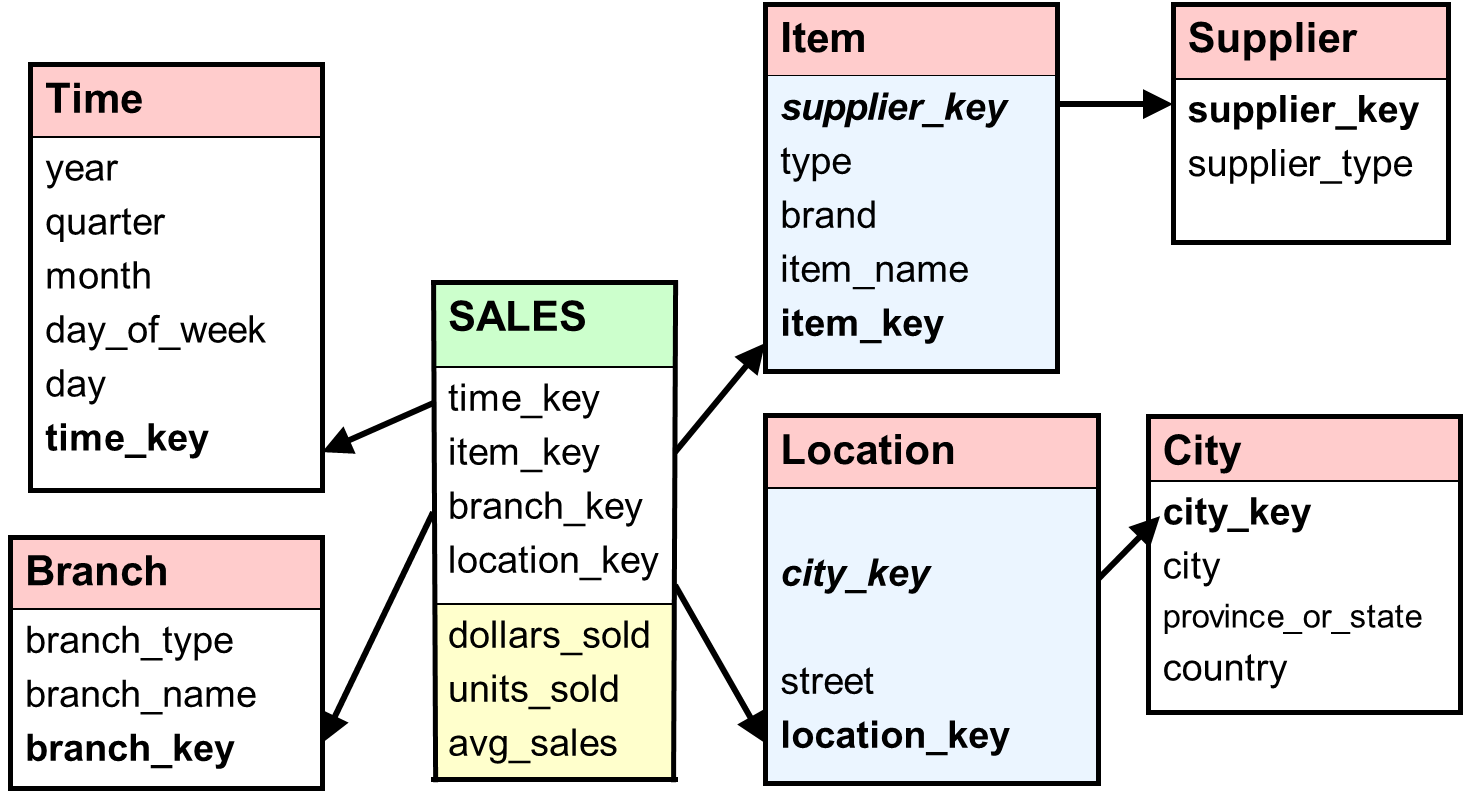
* + - 4 Dimensi



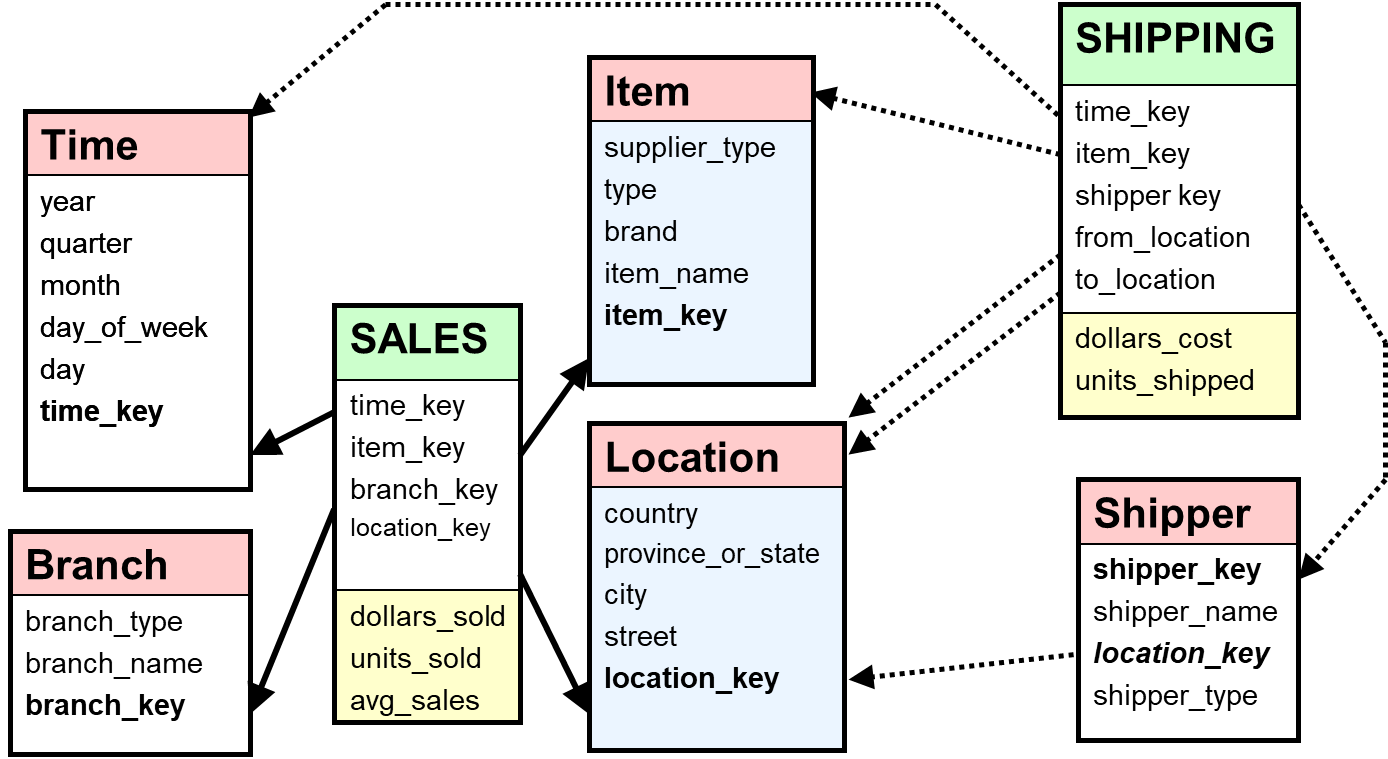
* + Dengan ini tabel dalam model data multidimensi dibagi menjadi dua macam, yaitu
    - Tabel fakta, menyimpan measure
    - Tabel dimensi, menyimpan atribut sumbu
  + Model dalam data multidimensi dapat direpresentasikan menjadi tiga skema, yaitu
    - Star schema
    - Snow flake schema
    - Fact constellation schema
  + **Skema star**
    - Pendekatan dimensional menggunakan skema star diorganisir menjadi **tabel fakta** dan **tabel dimensi**
    - Dimensi dalam star schema digunakan untuk melakukan analisa dan mengorganisasi data
    - Desain logika ditampilkan dalam bentuk representasi fisik yang akan mengoptimalkan performansi dan manajemen



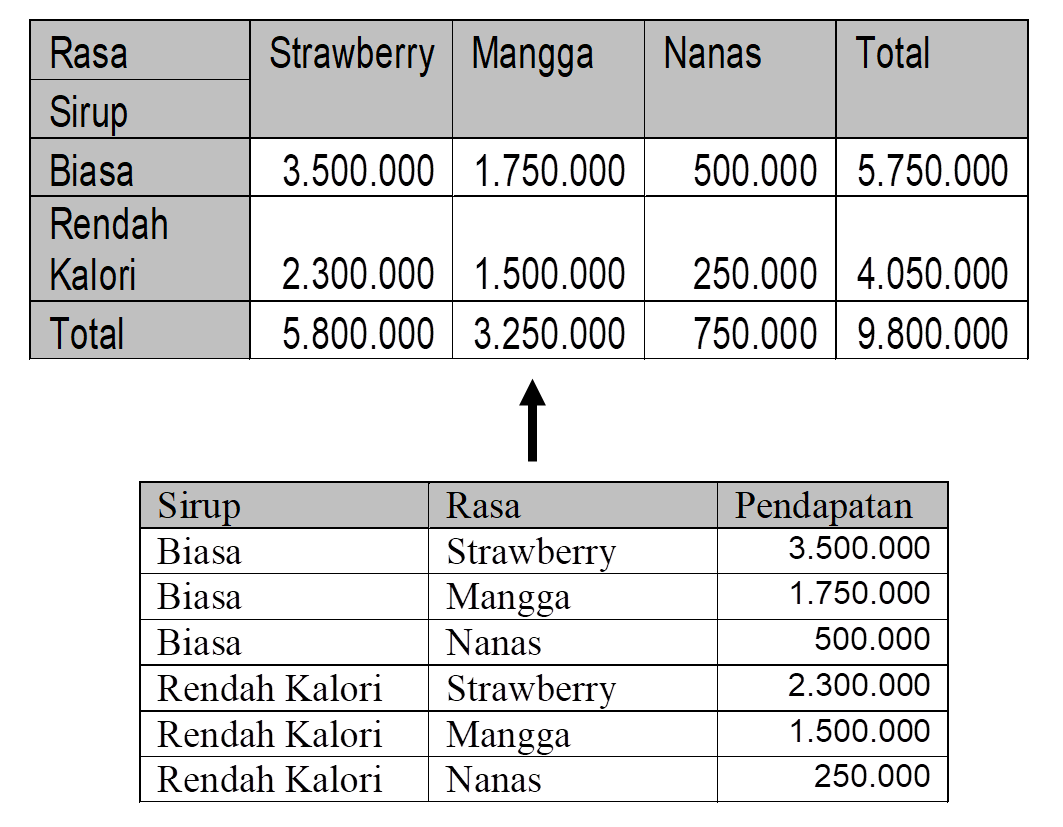
* + **Snowflake Schema**
    - Pendekatan dimensional menggunakan snowflake diorganisir menjadi tabel fakta dan tabel dimensi, namun pada tabel dimensi lebih di breakdown lagi menjadi tabel yang berbeda, namun hanya berhubungan dengan tabel dimensi saja



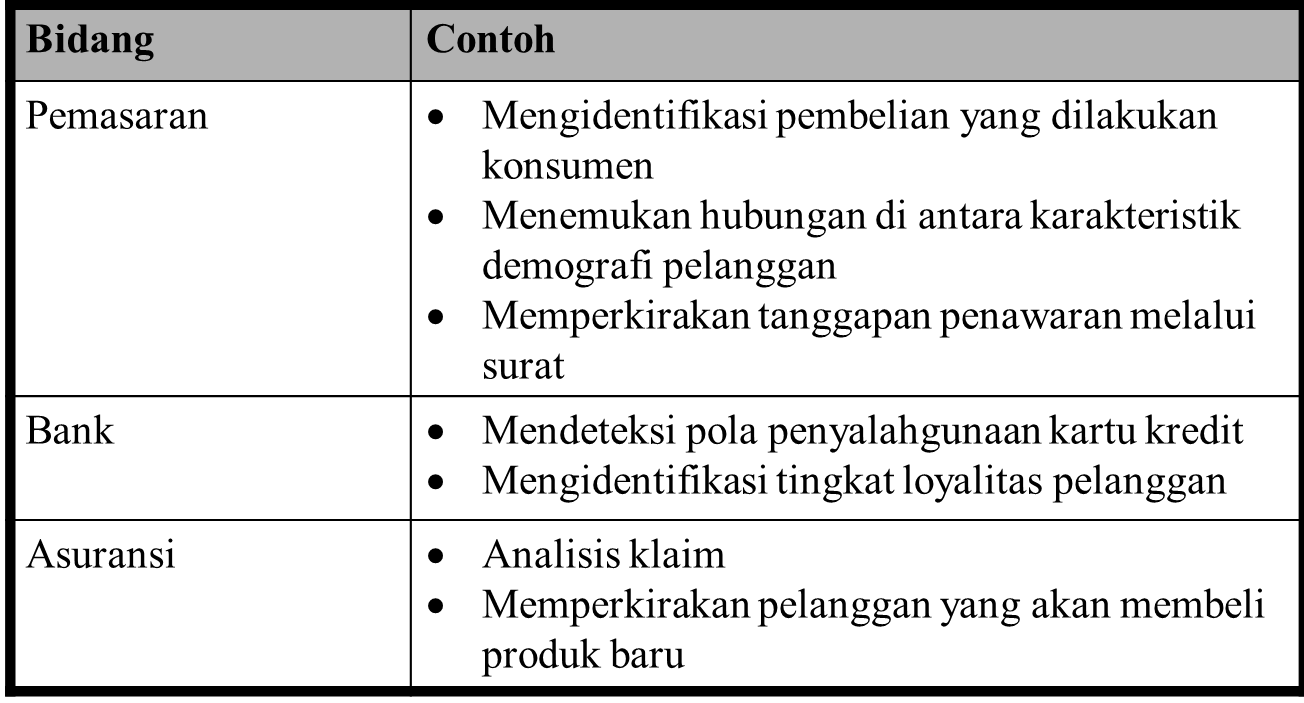
* + **Fact Constellation**
    - Pendekatannya desain fact constellation merupakan gabungan dari star schema dan snowflake schema dimana **suatu dimensi** **dapat digunakan lebih dari satu tabel fakta**

****

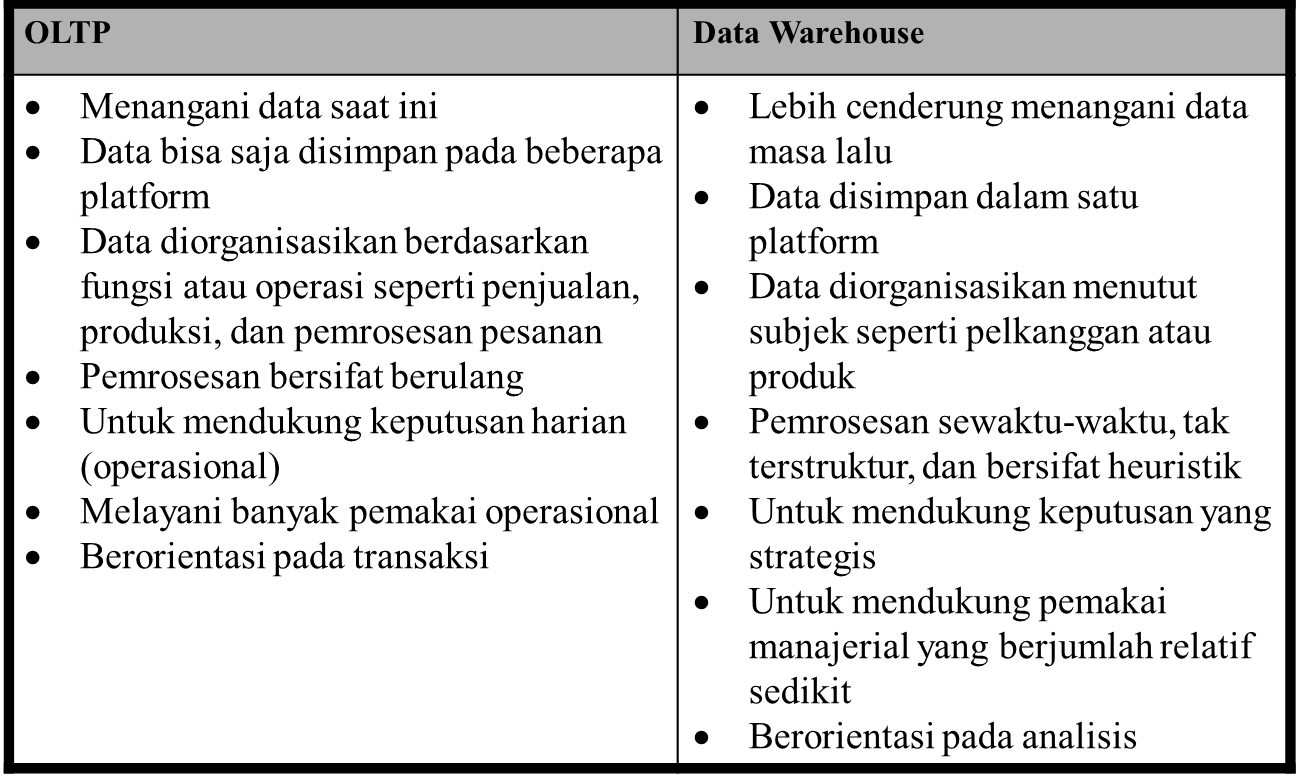
* + Ada beberapa pertimbangan dalam mendesain data warehouse, yaitu
    - Desain yang dirancang khusus untuk pengelolaan
      * Harus mudah di **backup** secara teratur
      * Diperhatikan pada saat proses **loading** new data
      * Memperhatikan pada saat proses **aggregarting** new data
      * Memperhatikan ketika melakukan aktifitas **pemeliharaan** data
    - Desain yang dirancang khusus untuk performa
      * Mengimplementasikan **indeks**
  + Data warehouse di desain sebagai suatu sistem yang terpisah pisah, hal ini dikarenakan karena sistem operasional biasa tidak memiliki data histori (data time series, atau data yang bergantung pada dimensi waktu) sehingga tidak dapat dianalisa.
    - Maka dari itu, dalam data warehouse, ekstraksi data dari sistem operasional dilakukan terlebih dahulu lalu dikombinasikan dan di transformasikan dalam format yang seragam dan di load ke dalam database
* **Data Mart** adalah subyek spesifik atau aplikasi spesifik dari data warehouse yang berisi data untuk satu kali transaksi, misalnya untuk penjualan, atau suplai. Data mart memiliki beberapa karakteristik, yaitu
  + Bersifat dependent atau independent bergantung pada sumber informasinya, jika sumber informasi bersifat dependent adalah berupa data warehouse itu sendiri
  + Sumber informasi bersifat independent jika tidak terdapat data warehouse, sebagai gantinya, sumber informasi didapatkan dari ekstraksi data da ri sistem operasional
  + Data mart fokus hanya pada kebutuhan pemakai yang terkait dalam sebuah departemen atau fungsi bisnis
  + Biasanya tidak mengandung data operasional yang rinci seperti data warehouse
  + Hanya mengandung sedikit informasi dibandingkan dengan data warehouse, sehingga lebih mudah dipahami dan dinavigasi
* **Definisi alternatif Data Mart,** bagian dari data warehouse yang mendukung kebutuhan pada tingkat departemen atau fungsi bisnis tertentu dakan perusahaan
* **OLAP (Online Analytical Processing),** jenis pemrosesan yang memanipulasi dan menganalisa data bervolume besar dari berbagai perspektif (multidimensi). Juga seringkali disebut *analisis data multidimensi*
  + Data muldimensi sendiri adalah data yang dapat dimodelkan sebagai atribut dimensi & atribut ukuran
  + Contoh atribut dimensi adalah nama, barang, warna barang, dll
  + Contoh atribut ukuran adalah jumlah barang, jumlah penjualan, dll
  + OLAP memiliki beberapa kemampuan, diantara lain
    - **Konsolidasi**, atau **rollup** melibatkan pengelompokan data. Contohnya adalah
      * Kantor-kantor cabang dapat dikelompokkan menurut kota atau provinsi
      * Transaksi penjualan dapat dikelompokkan menurut tahun, triwulan, semester, quarter, bulan, dsb
    - **Drill-down,** merupakan kebalikan dari konsolidasi, yaitu memungkinkan suatu data yang ringkas dijabarkan menjadi data yang lebih detail (breakdown data)
    - **Sliceing & Dicing**, atau **Pivoting** memungkinkan untuk melihat data dari berbagai sudut pandang



* + Software yang dapat mengimplementasikan OLAP adalah
    - Express server (oracle)
    - Powerplay (cognos software)
    - Mentacube (informix)
    - HighGate Project (Sybase)
* **Data Mining**, adalah ekstraksi pola yang menarik dari sebuah data dalam jumlah besar. Dalam database, data mining digunakan untuk menemukan pola tersembunyi atau relasi-relasi pada database yang akan digunakan untuk memperkirakan perilaku di masa mendatang
  + Contoh penerapan data mining di berbagai bidang pekerjaan



* + Data mining terus berkembang, sehingga teknologi dari data mining juga akan semakin bertambah, antara lain adalah
    - Statistik
    - Penerapan algoritma neural network
    - Penerapan logika fuzzy
    - Penerapan algoritma genetika
    - Penerapan teknologi kecerdasan buatan lainnya
  + Pendekatan data mining juga dapat dilakukan melalui visualisasi data, pada pendekatan ini user akan dibantu untuk menemukan sendiri pola dari data berukuran besar dengan didasarkan pada visualisasi oleh data mining
* **Langkah-langkah membangun data warehouse**
  + Menentukan sasaran bisnis dalam pembentukan data warehouse
  + Mengidentifikasi data dari database operasional atau sumber lainnya yang diperlukan bagi data warehouse
  + Menentukan standarasisasi penamaan data dan makna pada item-item yang ada
  + Merancang desain database untuk data warehouse
    - Menentukan antara satu database atau lebih
    - Membuat database dalam oracle
    - Menentukan skema untuk database
    - Mengatur data file dan tablespace
    - Membuat tabel fakta dan tabel dimensi
    - Mendefinisikan constraint nya
    - Membuat index
    - Membuat partisi
    - Membuat View
    - Dan mempertimbangkan keamanannya
  + Membangun kebijakan dalam pengarsipan data lama, sehingga penyimpanan yang diperlukan efektif
  + Melakukan ekstraksi data operasional ke dalam database milik data warehouse
  + Melakukan testing sebelum rilis produksi
    - Tes waktu yang dibutuhkan untuk melakukan reload data
    - Membersihkan data dan transformasi
    - Tes waktu respon query
    - Tes data summary yang dibutuhkan
    - Tes waktu yang dibutuhkan untuk tugas pengelolaan
* **Fitur2 data warehousing dalam oracle**
  + Bitmap join indexes
  + Partisi
  + Proses ETL(Extract-Transform-Load)
    - Mekanisme untuk mengidentifikasi perubahan baris dalam database
    - Tabel eksternal
    - Perintah SQL baru, “merge”
    - Eksekusi paralel
  + Materialized views
* **Tantangan dalam membangun data warehouse**
  + Memerlukan tuning untuk mendapatkan performa yang lebih baik
  + Diperlukannya metadata
  + Volume data akan terus meningkat pesat
  + Ketersediaannya harus tinggi
  + User yang mengakses akan lebih banyak
  + Tipe aplikasi baru akan selalu muncul
* **Perbedaan Data Warehouse dan OLTP**



BAB II – Index

* Bitmap Index Datawarehouse
  + Bitmap index, digunakan pada kolom yang memiliki varian nilai yang sedikit, semisal jensi kelamin (Laki/Perempuan), status nikah (Lajang/Menikah), dll
  + Selain itu, cocok digunakan untuk kolom yang mengandung nilai NULL
  + Namun, bitmap index hanya cocok digunakan pada local partitioning index
  + Keuntungan
    - Mengurangi waktu respon untuk query data yang besar
    - Mengurangi alokasi penyimpanan untuk index dibandingkan index yang lain
    - Bekerja secara minimum di hardware (penggunaan CPU rendah, dan memori yang keil)
    - Maintenance dapat dilakukan secara efisien selama dilakukan proses DML dan loading data
* B-Tree Index Datawarehosue
  + Index berbentuk tree, dan analoginya seperti pada saat mencari topik pada sebuah buku melalui daftar isi
  + Cocok digunakan pada sebuah kom yang memiliki rentang nilai yang besar (nama, tanggal lahir, alamat, dll)
  + B-Tree Index dapat bekerja pada local maupun global partitioning index
* Index Compression
  + Index compression akan secara otomatis dilakukan oleh oracle pada saat pembuatan index Bitmap atau B-Tree
  + Tujuan dari index compression adalah untuk efisiensi storage, namun akan meningkatkan kinerja CPU
* Local Index vs Global Index
  + Pada oracle versi 8i atau sebelumnya, penggunaan global index tidak direkomendasikan dalam data warehouse
    - Hal ini dikarenakan, pada statement DDL dalam partisi (ALTER TABLE, DROP PARTITION, dll), dapat mengakibatkan semua index tidak valid lagi, sehingga harus membangun kembali index yang tentunya langkah ini tidak efisien
  + Pada oracle versi 10g, ada perubahan pada global index, sehingga global index dapat dimaintenance setelah menggunakan DDL. Sehingga validasi index lebih mudah dilakukan dan membuat global index lebih efektif untuk data warehouse
  + **NAMUN,** meskipun terdapat perubahan, local index lebih umum digunakan daripada global index.
* **TIPS PEMBUATAN INDEX DATA**
  + Buat index hanya pada baris record yang besar (>1000 record)
  + Buat index hanya pada kolom yang digunakan pada klausa SQL where
  + Contoh pada tabel **emp** pada hr schema
    - SELECT \* FROM EMP WHERE ENAME=’Robert’;
    - Yang harus di index adalah pada kolom ENAME
    - Karena ename adalah data dengan range yang lebar (nama) maka index yang digunakan adalah B-Tree Index
  + Contoh yang lain
    - SELECT \* FROM EMP EHERE ENAME=**’**Robert’ AND deptno=40;
    - Yang harus dibuat indexnya adalah **index gabungan** antara ename dan deptno
  + Contoh yang lain
    - SELECT \* FROM EMP WHERE ENAME=’Robert’ or deptno=40;
    - **Index harus dibuat terpisah** yaitu index kolom ename dan index kedua adalah index kolom deptno

BAB III – TUNING DATA WAREHOUSE

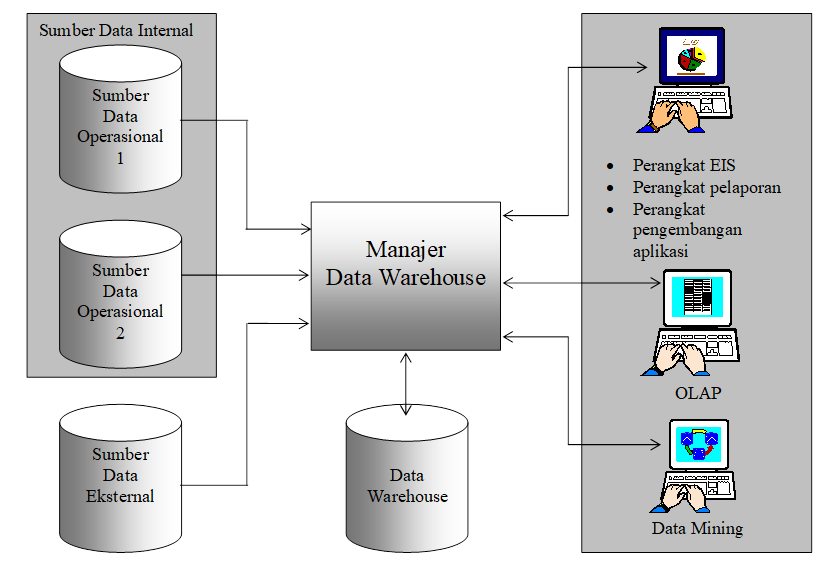
* Performa Input dan output dalam data warehouse harus selalu menjadi pertimbangan utama pada saat melakukan desain data warehouse.
* Beban kerja yang harus dihadapi oleh data warehouse antara lain
  + Input output yang intensif karena loading data berjumlah besar
  + Pada saat membuat index
  + Pada saat membuat materialized view
  + Dan pada saat melakukan query dengan volume data yang besar
* Pada tuning data warehouse, ada beberapa hal yang perlu diperhatikan, antara lain
  + Mengkonfigurasi I/O bandiwth
    - Mengkonfigurasikan storage dalam I/O bandwith lebih penting dibandingkan dengan menambah kapasitas storage nya
    - Yang dikonfigurasikan pada I/O bandwith antara lain
      * I/O throughput rates
      * RPM
      * Buffer cache
      * Write dan Read Speed
  + Membagi Storage
    - Dalam mengkonfigurasikan storage, secara umum untuk setiap object database yang ada disimpan dalam beberapa disk dengan akses channel yang berbeda beda
    - Dalam oracle, pembagian storage ini dapat dilakukan dengan membuat beberapa datafiles dalam database oracle
    - Datafiles tersebut nantinya dapat didistribusikan ke beberapa disk yang berbeda-beda
  + Menggunakan Storage Redundancy sebagai backup
    - Karena jumlah data pada data warehouse sangat masif, maka storage yang dibutuhkan akan sangat besar, sehingga sangat memungkinkan terjadinya kerusakan storage. Untuk mengatasi hal tersebut sebaiknya dibuat juga disk redundancy sebagai alat backup data pada datawarehouse
    - Setiap disk master (storage master) setidaknya memiliki satu disk backup (disk redudancy)
    - Teknologi yang sering digunakan adalah RAID untuk membuat disk redudancy
  + Test sistem I/O sebelum membuat database
    - Sekali file database dibuat, akan sangat sulit mengkonfigurasikannya kembali, maka harus dilakukan testing sistem I/O terlebih dahulu sebelum database dibuat.
    - Hal ini dilakukan untuk menghindari adanya pembangunan ulang database yang disebabkan adanya keharusan untuk me-rekonfigurasi ulang I/O sistemnya
  + Merencanakan desain untuk pertumbuhan data kedepannya
    - Dalam pembuatan data warehouse, wajib untuk designer datawarehouse untuk memiliki pandangan terhadap pertumbuhan data kedepannya, hal ini dilakukan untuk meningkatkan efisiensi dan durability dari suatu data warehouse

PERTANYAAN KISIKISIKULIAH

1. Jelaskan pengertian dari datawarehouse

**Data warehouse** merupakan relasional database yang didesain untuk melakukan query dan analisa, bukannya untuk melakukan proses transaksi. Data warehouse biasanya berisi data histori yang didapatkan dari data transaksi.

1. Jelaskan secara detail sistem datawarehouse dan gambarkan diagramnya
   * Data warehouse mengambil sumber data dari data operasional, dan sumber eksternal,
   * Data operasional dapat diperoleh dari database pelanggan ataupun produk dari dalam organisasi
   * Sumber eksternal diperoleh dari internet, database yang bersifat komersil dan lain lain
   * Berbagai data dari macam macam sumber tersebut lalu digabungkan dan diproses lebih lanjut oleh manajer data warehouse dan disimpan dalam database tersendiri
   * Selanjutnya perangkat lunak/aplikasi seperti OLAP, data mining atau, perangkat EIS, dll digunakan untuk menunjang akses data warehouse dan kepentingan analisa data pada datawarehouse



1. Jelaskan apa itu ETL, dan mengapa 80% proses dataawarehouse itu semuanya ETL

ETL (extraction/transormation/load) adalah salah satu fitur data warehouse, ETL memakan 80% data warehouse karena kebanyakan proses data warehouse adalah untuk

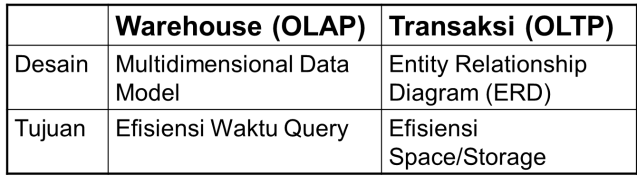
* + Menempatkan data, menulis program untuk ekstraksi
  + Memfilter dan membersihkan data
  + Mentransformasi data ke dalam skema standar
  + Load data ke dalam data warehouse

Dan harus diingat bahwa jumlah data yang ada pada data warehouse sangatlah banyak sehingga proses di atas membutuhkan waktu yang relatif besar dari keseluruhan proses datawarehouse

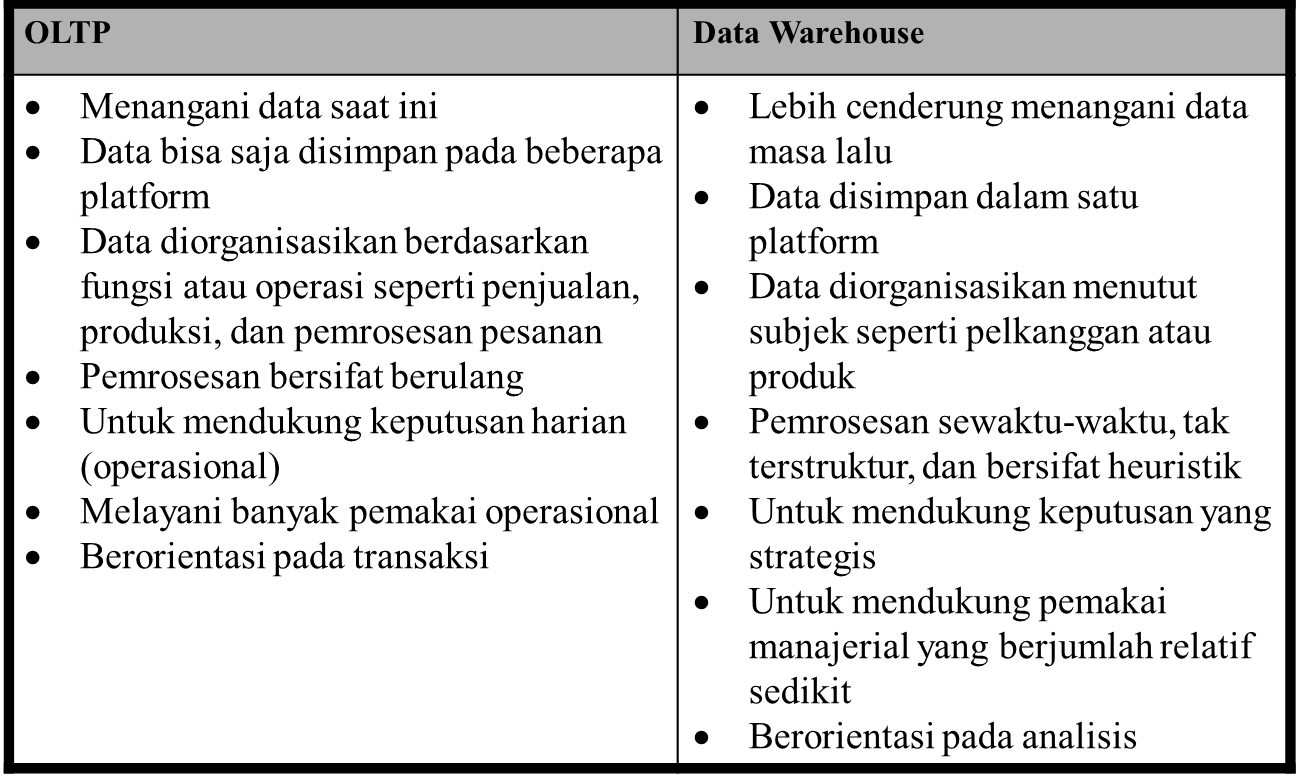
1. Jelaskan dan gambarkan skema yang ada pada datawarehouse

LIHAT PAGE 4

1. Jelaskan berbagai macam perbedaan dari OLAP dan OLTP



Atau



1. Jelaskan dengan detail langkah penggunaan data warehouse menggunakan oracle sampai pembuatan data di BISpreadSheet, jelaskan detail dan step-by-step
   1. Data warehouse di oracle
      * Membuat tablespace di oracle untuk tempat menyimpan data pada data warehousenya
      * Membuat tabel dimensi dan mendefinisikan setiap atributnya
      * Membuat Cube/tabel fakta dan mendefinisikan seluruh atributnya mulai dari primary key dan data measure nya
      * Membuat constraint antara tabel dimensi dan tabel fakta/cube
      * Mengisi data record untuk setiap tabel dimensi sebelum melakukan query
      * Membuat sequence dan indeks untuk meningkatkan performa data warehouse
   2. Langkah-langkah penggunaan BISpreadSheet
      * Install add-in BISPreadsheet pada MS.Excel
      * Klik add-in 🡪 oracleBI 🡪 New Query
      * Buat koneksi ke database pada tab connection editor, masukan description, hostname, port dan SID sesuai dengan instance database oracle yang ada lalu klik new
      * Pada tab OLAP connection, masukan nama koneksi yang telah dibuat tadi lalu masukan informasi username dan password lalu klik connect
      * Setelah muncul prompt Query Wizard klik next untuk membuat query baru
      * Pilih item apa saja dari database yang dibutuhkan untuk melakukan query
      * Pilih layout yang ingin digunakan untuk menampilkan data
      * Pilih detail yang akan ditampilkan berdasarkan dimensi yang telah ditentukan
      * Klik finish dan akan muncul interactable spreadhseet yang tabat di klik untuk melihat data yang ada