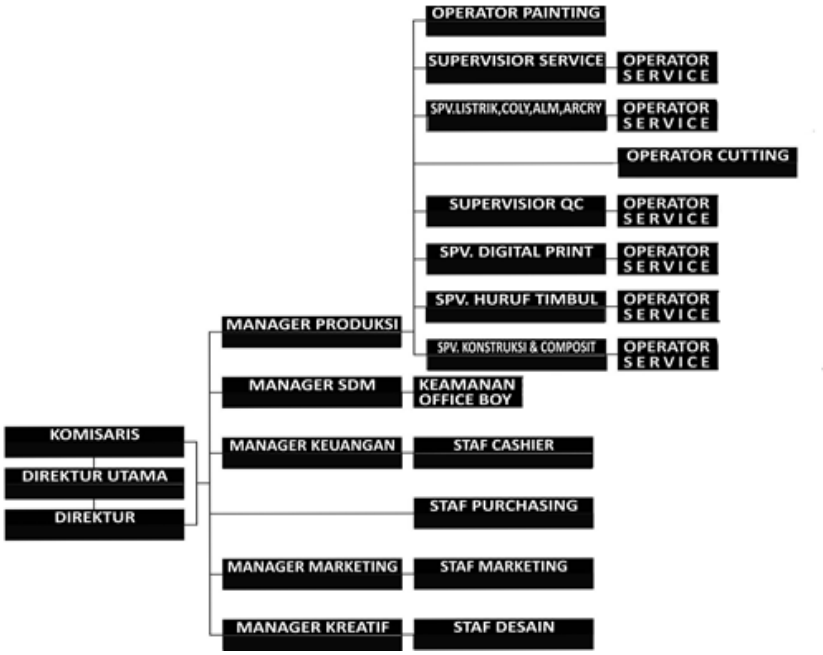
BAB I
TINJAUAN PUSTAKA

II.1 Profil Instansi

PT. Rajawali Neon adalah Perusahaan yang bergerak di bidang reklame yang didirikan tahun 1986 di Bandung, Media promosi yang ditawarkan meliputi: *Billboard, Neon Box, Neon Sign, Panel Neon, Letter Sign, Etsa, Digital Printing, Running Text*, dan jenis reklame lainnya. Konsumen yang memesan ke PT. Rajawali Neon terdiri dari berbagai lembaga pendidikan, partai politik, dan perusahaan yang membutuhkan kegiatan promosi pengiklanan. PT. Rajawali Neon berlokasi di jalan Sudirman No.727 Bandung.

II.2.7 Struktur Organisasi Instansi

Struktur organisasi adalah pola hubungan antara bagian-bagian dari instansi atau menggambarkan dengan jelas pemisahan kegiatan pekerjaan antara bagian yang satu dengan bagian yang lain dalam suatu instansi. Gambar II-1 Berikut merupakan struktur organisasi yang ada di PT. Rajawali Neon.



Gambar I-1 Struktur Organisasi PT. Rajawali Neon

Adapun deskripsi kerja dari struktur organisasi pada PT. Rajawali Neon adalah:

1. Manager Produksi
   a. Mengatur pengadaan bahan produksi.
   b. Bertanggungjawab atas hasil produksi
2. Operator Painting
   a. Bertanggungjawab terhadap proses *painting* reklame.
   b. Mengatur banyaknya jenis warna yang dibutuhkan
3. Supervisor Service
   a. Bertanggungjawab terhadap keseluruhan alat-alat dan mesin produksi
   b. Melakukan pemeriksaan secara teratur terhadap alat-alat dan mesin produksi.
   c. Melakukan perbaikan terhadap alat-alat dan mesin produksi yang rusak.
   d. Mengatur pengadaan alat-alat dan mesin produksi yang baru.
4. Supervisor Listrik, Coly, Alm, Arcry
   a. Bertanggunjawab terhadap kecukupan yang dibutuhkan untuk kebutuhan alat-alat dan mesin produksi
   b. Bertanggungjawab terhadap *coly, alm*, dan *arcry*
5. Supervisor Digital Print
   a. Mencetak reklame yang dicetak pada mesin *printer*.
   b. Meneima laporan hasil desain reklame untuk menentukan jenis cetak.
6. Manager Sumber Daya Manusia
   a. Mempertahankan kontribusi setiap divisi pada tingkat yang sesuai dengan kebutuhan perusahaan.
   b. Melakukan pelatihan keahlian sumber daya manusia.
7. Manager Keuangan
   a. Membuat laporan keuangan hasil produksi.
   b. Mencetak *invoice* hasil pemesanan.

8. Manager Marketing
    a. Mengatur hasil survey terhadap pemesanan.
    b. Melakukan analisa terhadap laporan penawaran dan pemesanan.
    c. Membina sumber daya manusia yang ada di lingkungan divisi marketing.
    d. Menentukan penerimaan hasil penawaran konsumen.
    e. Membuat laporan pemesanan secara keseluruhan.
9. Staff Marketing
    a. Melakukan penawaran reklame terhadap konsumen.
    b. Menangani pemesanan reklame terhadap konsumen.
    c. Mencatat hasil penawaran reklame.
    d. Meminta informasi konsumen yang memesan reklame.
    e. Melakukan survey terhadap proses penawaran.
    f. Membuat dokumentasi pemesanan dan penawaran.
10. Manager Kreatif
    a. Merancang reklame sesuai kebutuhan dan keinginan konsumen.
    b. Melakukan dokumentasi hasil desain.
    c. Menentukan bahan yang akan digunakan untuk kebutuhan produksi.
11. Staff Desain
    a. Melakukan perancangan reklame sesuai kebutuhan dan keinginan konsumen.
    b. Menerima hasil survey terhadap reklame yang akan dibuat.

Pada Deskripsi kerja PT. Rajawali Neon dapat disimpulkan bahwa yang melakukan proses *analitycal* terhadap data pemesanan dan penawaran adalah manager marketing. Dimana manager marketing sudah memahami terhadap proses *analitycal* terhadap perangkat lunak data mart setelah diberi pemahaman oleh seorang *analyst*.

II.2 Landasan Teori

Sub bab ini berisi teori-teori pendukung yang digunakan dalam proses analisis dan implementasi dalam tugas akhir ini.

II.2.1 Pengertian Data

Data dan Informasi mempunyai keterkaitan antara satu dengan yang lainnya. Data adalah fakta – fakta atau pesan – pesan yang belum dievaluasi atau bahan mentah dari informasi yang melalui pengolahan tertentu dibentuk menjadi suatu informasi. Data merupakan kenyataan yang menggambarkan suatu kejadian – kejadian nyata.

" *Menurut Steven Alter, data merupakan fakta,gambar atau suara yang mungkin atau tidak berhubungan atau berguna bagi tugas tertentu* " [3].

" *Menurut McLeod, data terdiri dari fakta-fakta dan angka yang secara relatif tidak berarti bagi pemakai. Sedangkan informasi adalah data yang sudah diproses atau data yang memiliki arti "* [3].

Data merupakan suatu bentuk keterangan-keterangan yang belum diolah atau dimanipulasi sehingga belum begitu berarti bagi sebagian pemakai. Sedangkan informasi merupakan data yang sudah di olah sehingga memiliki arti.

II.2.2 Basis Data

Basis data terdiri dari dua kata, basis dan data. Basis dapat diartikan kurang lebih sebagai markas, gudang atau tempat berkumpul [3]. Sedangkan data adalah repesentasi fakta dunia nyata yang mewakili suatu objek seperti manusia (pegawai, siswa, pembeli dan lain – lain), barang dan sebagainya yang direkam dalam bentuk angka, huruf, simbol, teks, gambar, bunyi atau kombinasi lainnya. Basis data sendiri dapat didefinisikan dalam sejumlah sudut pandang seperti [3].

2. Himpunan kelompok data (arsip) saling berhubungan yang diorganisasikan sedemikian rupa agar kelak dapat dimanfaatkan kembalidengan cepat dan mudah.
3. Kumpulan file / table / arsip saling berhubungan yang disimpan dalam media penyimpanan elektronis.

II.2.3 Data Warehouse

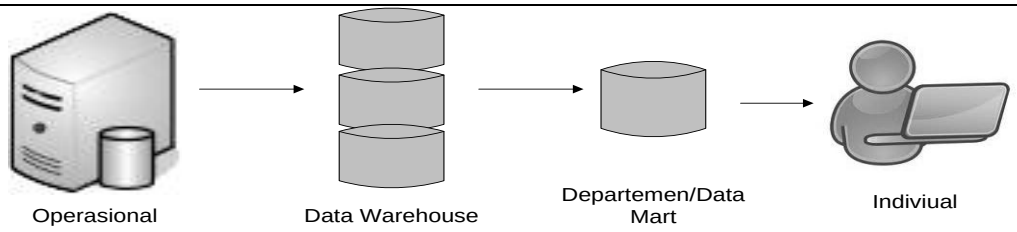
Data warehouse merupakan kumpulan informasi logikal, yang dikumpulkan dari berbagai operasional *database* yang mendukung aktivitas para bisnis analis dalam pekerjaan mengambil keputusan. Kedengaran cukup sederhana pada awalnya, tetapi *data warehouse* menggambarkan hal yang mendasar perbedaan cara pandang tentang mengorganisasi dan mengatur informasi dalam suatu organisasi.

Data warehouse adalah multidimensional dalam relasional model *database*, informasi ditentukan dalam suatu rangkaian dari tabel yang berdimensi dua. Tidak begitu halnya dengan *data warehouse*, banyak *data warehouse* adalah multidimensional, artinya bahwa *data warehouse* terdiri dari *layer-layer*, kolom dan baris. *Layer-layer* dalam suatu *data warehouse* menunjukkan informasi yang disampaikan pada dimensi yang berbeda.*Data warehouse* merupakan kombinasi dari *database* yang berbeda-beda, *datawarehouse* mengkombinasikan informasi dengan meringkas (*summarizing*) dan mengelompokkan (*aggregation)*. Pada saat mengekstrak informasi dari berbagai operasional *database* untuk membuat suatu *data warehouse*, pengumpulan informasi yang diperlukan untuk pembuat keputusan. Informasi yang diperlukan ini didefinisikan oleh persetujuan para pengguna sesuai dengan informasi yang mereka butuhkan dalam pengambilan keputusan. Jadi suatu *data warehouse* hanya berisi informasi yang relevan dengan kebutuhan *user* untuk mendukung pengambilan keputusan.

> *"Menurut W.H. Inmon dan Richard D.H., data warehouse adalah koleksi data yang mempunyai sifat berorientasi subjek,terintegrasi,time-variant, dan bersifat tetap dari koleksi data dalam mendukung proses pengambilan keputusan management"* [4].

Jadi, data warehouse merupakan metode dalam perancangan basis data, yang menunjang DSS (*Decission Support System*) dan EIS (*Executive Information System*). Dari definisi-definisi yang dijelaskan tadi, dapat disimpulkan data warehouse adalah basis data yang saling bereaksi yang dapat digunakan untuk *query* dan analisisis, bersifat orientasi subjek, terintegrasi, *timevariant*, tidak berubah yang digunakan untuk membantu para pendukung keputusan. Ada empat

tingkat data dalam lingkungan arsitek *database* yang di tunjukkan pada Gambar I-2. Tingkat Operasional, tingkat atom (atau data warehouse), tingkat departemen (atau data mart), dan tingkat individu. Tingkatan data yang berbeda ini adalah dasar dari arsitektur yang lebih besar yang bernama *corporate information factory* (CIF).



- Terperinci
- Hari ke hari
- Nilai saat ini
- Probabilitas tinggi dalam akses
- Berorientasi aplikasi

- Paling granular
- Time variant
- Terintegrasi
- Berorientasi subjek
- Beberapa ringkasan

- Terbatas
- Beberapa derived beberapa primitif
- Tipikal departemen
  - Accounts
  - Pemasaran
  - Teknisi
  - Aktuarial
  - manufaktur

- temporer
- ad hoc
- heuristik
- tidak berulang
- PC,berbasis workstation

Gambar I-2 Tingkatan Arsitektur *Database*

II.2.4 Data Mart

Data Mart merupakan suatu bagian dari *data warehouse* yang mendukung pembuatan laporan dan analisis data pada bagian operasional departemen suatu perusahaan. *Data mart* berisi informasi yang relevan bagi user yang ingin mengambil keputusan, sebuah perusahaan yang telah menginvestasikan baik uang dan waktunya dalam sebuah *bisnis operational* yang telah mempunyai dan menyimpan sebuah data dari pelanggan, penjualan dan inventori. Data tersebut merupakan *asset* dari suatu perusahaan yang perlu di *manage* untuk keperluan *kompetitif* perusahaan. Banyak perusahaan yang telah menerapkan *DSS* (*Decision Support system)* yang telah ada di suatu *data mart*, sebuah *data mart* terintegrasi

secara operasional dan kumpulan data histori untuk aplikasi pengambilan keputusan.

Ada empat tugas yang bisa dilakukan dengan adanya *data mart*, keempat tugas tersebut yaitu [4]:

a. Pembuatan laporan

Pembuatan laporan merupakan salah satu kegunaan *data mart* yang paling umum dilakukan. Dengan menggunakan *query* sederhana didapatkan laporan per hari, per bulan, per tahun atau jangka waktu kapanpun yang diinginkan untuk menunjang proses pendukung keputusan.

b. *On-Line Analytical Processing* (OLAP)

Dengan adanya *data mart*, semua informasi baik detail maupun hasil *summary* yang dibutuhkan dalam proses analisa mudah didapat. *OLAP* mendayagunakan konsep data multidimensi dan memungkinkan para*user* menganalisa data sampai mendetail, tanpa mengetikkan satupun perintah *SQL*. Hal ini dimungkinkan karena pada konsep multidimensi, maka data yang berupa fakta yang sama bisa dilihat dengan menggunakan fungsi yang berbeda.

c. *Data Mining*

Data mining merupakan proses untuk menggali (*mining)* pengetahuan dan informasi baru dari data yang berjumlah banyak pada *data mart*, dengan menggunakan kecerdasan buatan (*Artificial Intelegence*), statistik dan matematika. *Data mining* merupakan teknologi yang diharapkan dapat menjembatani komunikasi antara data dan pelakunya.

d. Proses informasi *executive*

Data mart dapat membuat ringkasan informasi yang penting dengan tujuan membuat keputusan bisnis, tanpa harus menjelajahi keseluruhan data. Dengan menggunakan *data mart* segala laporan telah diringkas dan dapat pula mengetahui segala rinciannya secara lengkap, sehingga mempermudah proses pengambilan keputusan. Informasi dan data pada laporan *data mart* menjadi target informatif bagi pengguna.

II.2.5 Karakteristik Sistem

Karakteristik *data warehouse* menurut Inmon yaitu [4]:

a. *Subject Oriented* (Berorientasi Subjek)

Data warehouse berorientasi subjek artinya *data warehouse* di desain untuk menganalisa data berdasarkan subjek-subjek 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 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. Jadi dengan kata lain, data yang disimpan adalah berorientasi kepada subjek bukan terhadap proses.

b. *Integrated* (Terintegrasi)

Data Warehouse dapat menyimpan data-data yang berasal dari sumber-sumber yang terpisah kedalam suatu format yang konsisten dan saling terintegrasi satu dengan lainnya. Dengan demikian data tidak bisa dipecah-pecah karena data yang ada merupakan suatu kesatuan yang menunjang keseluruhan konsep *data warehouse* itu sendiri. Syarat integrasi sumber data dapat dipenuhi dengan berbagai cara sepeti konsisten dalam penamaan

variable, konsisten dalam ukuran *variable*, konsisten dalam struktur pengkodean dan konsisten dalam atribut fisik dari data.

c. *Time Variant* (Rentang Waktu)

Seluruh data pada *data warehouse* dapat dikatakan akurat atau valid pada rentang waktu tertentu. Untuk melihat interval waktu yang digunakan dalam mengukur keakuratan suatu *data warehouse*, kita dapat menggunakan cara antara lain :

1. Cara yang paling sederhana adalah menyajikan *data warehouse* pada rentang waktu tertentu, misalnya antara 5 sampai 10 tahun ke depan.
2. Cara yang kedua, dengan menggunakan variasi atau perbedaan waktu yang disajikan dalam *data warehouse* baik *implicit* maupun *explicit* secara *explicit* dengan unsur waktu dalam hari, minggu, dan bulan. Secara *implicit* misalnya pada saat data tersebut diduplikasi pada setiap akhir bulan, atau per tiga bulan. Unsur waktu akan tetap ada secara *implicit* didalam data tersebut.
3. Cara yang ketiga, variasi waktu yang disajikan *data warehouse* melalui serangkaian *snapshot* yang panjang. *Snapshot* merupakan tampilan dari sebagian data tertentu sesuai keinginan pemakai dari keseluruhan data yang ada bersifat *read-only*.

d. Non Volatile

Karakteristik keempat dari *data warehouse* adalah *non-volatile*, maksudnya data pada *data warehouse* tidak di-*update* secara *real time* tetapi di *refresh* dari sistem operasional secara reguler. Data yang baru selalu ditambahkan sebagai suplemen bagi *database* itu sendiri dari pada sebagai sebuah perubahan. *Database* tersebut secara *continue* menyerap data baru ini, kemudian secara *incremental* disatukan dengan data sebelumnya.

II.2.6 Model Multidimensional (*Multidimensional Modelling*)

Model multidimensional adalah teknik untuk memvisualisasi model data sebagai suatu kumpulan dari ukuran yang dideskripsikan dengan aspek-aspek bisnis (Chuck Ballard, Dirk Herreman, Don Schau, Ronda Bell, Eunseang Kim, Ann Valencic; IBM Corp., 1998). Hal ini khususnya sangat berguna untuk meringkas dan menyusun data dan memperlihatkan data untuk mendukung para analis data. *Multi Dimensional modeling* fokus pada data numerik, seperti harga, jumlah, berat, keseimbangan dan kejadian-kejadian.

Multi Dimensional modeling mempunyai beberapa konsep :

a. *Fact*

Fact adalah suatu koleksi dari relasi data-data *items*, terdiri dari ukuran-ukuran dan konteks data. Setiap *fact* biasanya merepresentasikan sebuah bisnis *item*, suatu transaksi bisnis, atau sebuah kejadian yang dapat digunakan dalam analisis bisnis atau proses bisnis. Dalam *data mart*, *fact* di implementasikan dalam tabel dasar dimana semudah data *numeric* dan disimpan [3].

b. Dimensions

Dimensions adalah suatu koleksi dari anggota atau unit-unit data dengan tipe yang sama. Dalam sebuah diagram, suatu dimensi biasanya direpresentasikan dengan suatu *axis.* Dalam dimensional model, semua data menunjukan *fact table* yang diasosiasikan dengan satu dan hanya satu member setiap *multiple dimensions.* Jadi dimensi menunjukan latar belakang kontekstual dari *fact.* Banyak proses analisis yang digunakan untuk menghitung (*quatify*) dampak dari dimensi pada *fact* [3].

Dimensi adalah parameter dari apa yang ingin lakukan dalam *On-Line Analytical Processing (OLAP)* [6]. Sebagai contoh, dalam suatu *database* untuk menganalisa semua penjualan dari produk, dimensi berikut ini harus ada waktu, lokasi, pembeli, penjual, dan scenario seperti kejadian, pembiayaan atau angka estimasi.Dimensi biasanya juga dapat

dipetakan bukan angka numerik, *entity* yang bersifat informatif seperti merek atau karyawan.

Dimension member : suatu dimensi berisi banyak anggota-anggota. Suatu anggota dimensi adalah nama pembeda atau *indetifier* yang digunakan untuk membedakan posisi suatu data item. Sebagai contoh, semua bulan, empat bulanan dan tahunan membuat dimensi waktu. Dan semua kota, wilayah dan negara merupakan dimensi geografi [5].

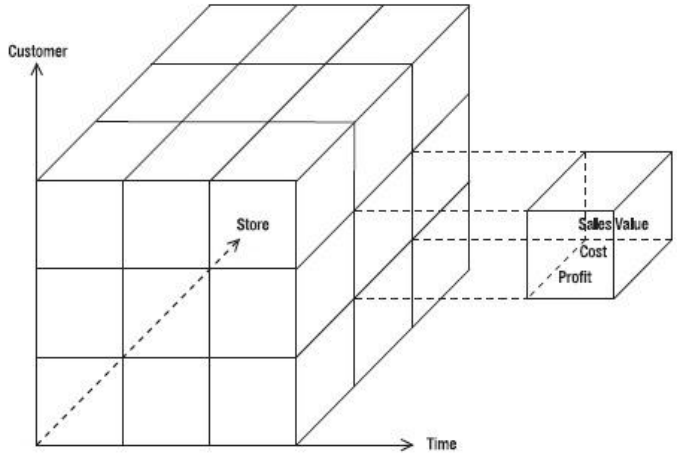
Dimension Hierarchies : dapat mengatur anggota dari suatu dimensi kedalam satu atau banyak hirarki. Setiap hirarki dapat juga mempunyai banyak level hirarki. Setiap member dari dimensi tidak boleh dilokasikan kedalam satu struktur hirarki [6].

c. *Measures*

Suatu *measures* (ukuran) adalah suatu besaran (angka numerik) atribut dari sebuah *fact*, yang menunjukan *performance* atau *behavior* (tingkah laku) dari bisnis secara relatif pada suatu dimensi [5].Angka atau nomor yang ditunjukan disebut dengan *variable*. Sebagai contoh ukuran dari penjualan dalam bentuk uang, besarnya penjualan, jumlah pengadaan, biaya pengadaan, banyaknya transaksi dan lainnya. Suatu ukuran dijelaskan dengan kombinasi dari member dari suatu dimensi dan diletakkan dalam *fact.*

II.2.7 Visualisasi Multidimensional Model

Cara yang paling populer dalam memvisualisasikan suatu model dimensional adalah dengan menggambarkan sebuah *cube* data. Contohnya dapat dilihat pada gambar II.3 yang dapat menggambarkan tiga model dimensional menggunakan sebuah kubus. Biasanya suatu dimensional model terdiri dari lebih tiga dimensi dan digambarkan sebagai suatu *hypercube*, akan tetapi *hypercube* sulit untuk divisualisasikan, jadi sebuah kubus lebih biasa digunakan [6]. Gambar berikut merupakan *measurement* adalah volume dari pemesanan konsumen, yang mana dijelaskan dengan kombinasi dari tiga dimensi yaitu lokasi, konsumen dan waktu.

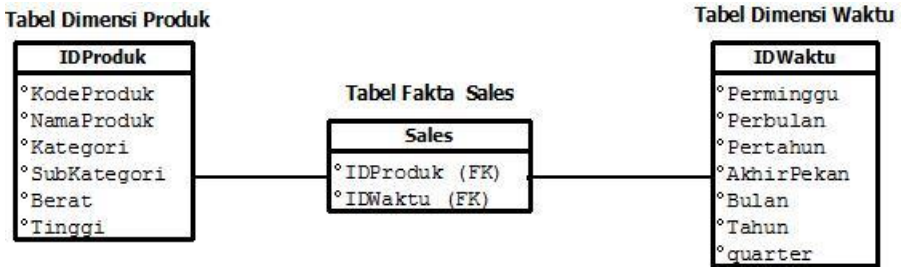


Gambar I-3 Visualisasi Multidimensional Model dengan kubus

Terdapat tiga konsep yang dapat memodelkan lebih jelas dalam penentuan *fact table,* dimensi dan *measure*, yaitu :

a. Skema *star*

Skema *star* adalah model dasar dari pemodelan multidimensi yang dapat di lihat pada gambar II-4. Model star memiliki satu tabel induk yang dinamakan *table fact* dan kumpulan dari tabel-tabel kecil yang disebut tabel dimensi, disusun dalam pola-pola melingkar mengelilingi *table fact* [4].

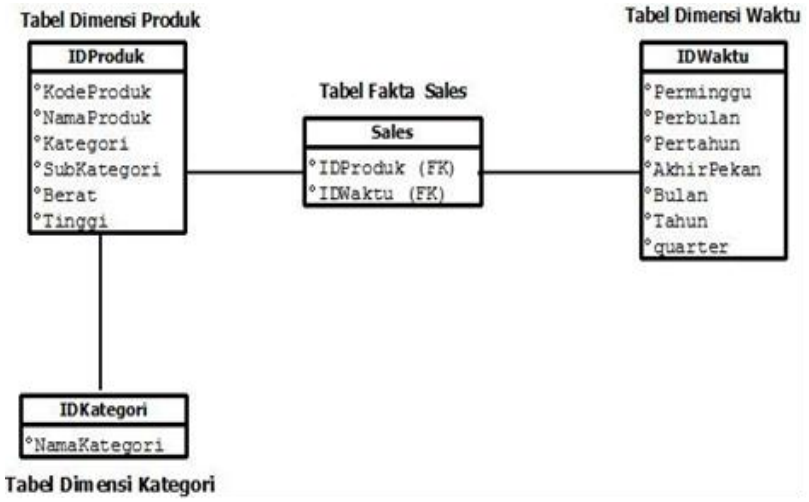


Gambar I-4 Skema *Star*

b. Model *Snowflake*

Selain skema *star* dalam menampilkan desain sebuah *data mart* juga dapat menggunakan model *Snowflake* sebagai contoh dapat dilihat pada **Error! Reference source not found.** Dalam sebuah model *snowflake* terdapat sebuah *fact table* yang dikelilingi oleh beberapa *dimension table.* Namun *dimension table* itu sendiri

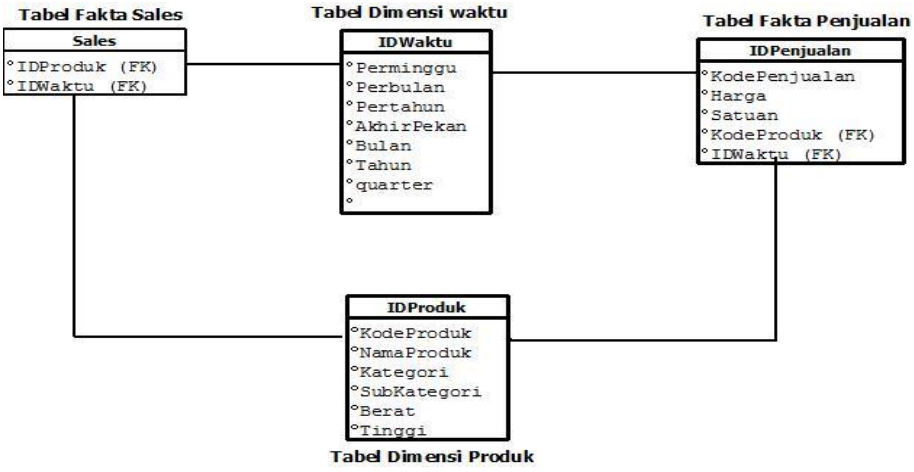
dapat seolah-olah berupa sebuah *fact table* lain yang juga memiliki dimensinya sendiri [4].



Gambar I-5 Gambar *Snowflake*

c. Skema *Constellation.*

Beberapa *fact tables* berbagi tabel dimensi, sebagai contoh dapat di lihat pada gambar II-6. skema *constellation* ditampilkan sebagai koleksi dari kumpulan skema bintang yang sering disebut sebagai skema *galaxy* [4].



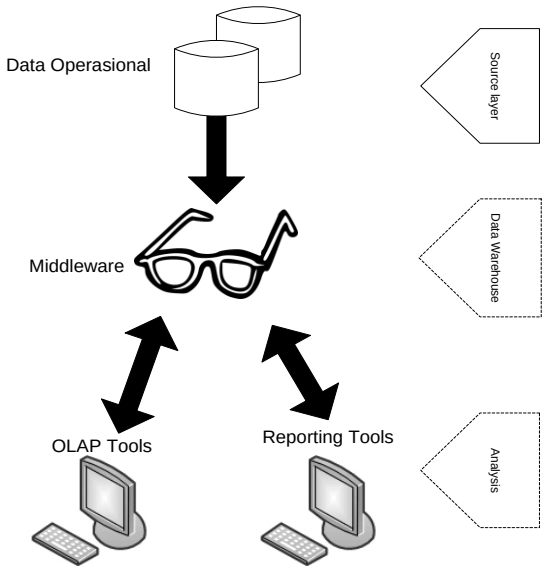
Gambar I-6 Skema *Constellation*

II.2.8 Arsitektur Sistem

Arsitektur *Data Warehouse* atau *Data mart* menurut Matteo Golfarelli dan Stefano Rizzi dalam bukunya yang berjudul *Data Warehouse Design: Modern Principles and Methodologies,* mengelompokkan arsitektur *data ware house* menjadi 3 kelompok, terdiri dari *single layer architechture, two layer architecture,* dan *three layer architecture*.

a. *Single Layer Architechture*

Single Layer Architechture pada umumnya tidak sering digunakan dalam suatu kasus. Tujuannya adalah untuk meminimalkan jumlah data yang disimpan, untuk mencapai tujuan ini, ia bisa menghilangkan redudansi data. Gambar II-7 menunjukkan hanya lapisan fisik yang tersedia. ini berarti bahwa *data warehouse* di implementasikan sebagai pandangan multidimensi data operasional yang dibuat oleh *middleware* tertentu, atau lapisan pengolahan menengah. *Middleware* merupakan komponen perantara yang memungkinkan *client* dan (lapisan aplikasi dan sistem operasi) saling terhubung dan berkomunikasi satu sama lain



Gambar I-7 *Single Layer Architechture*

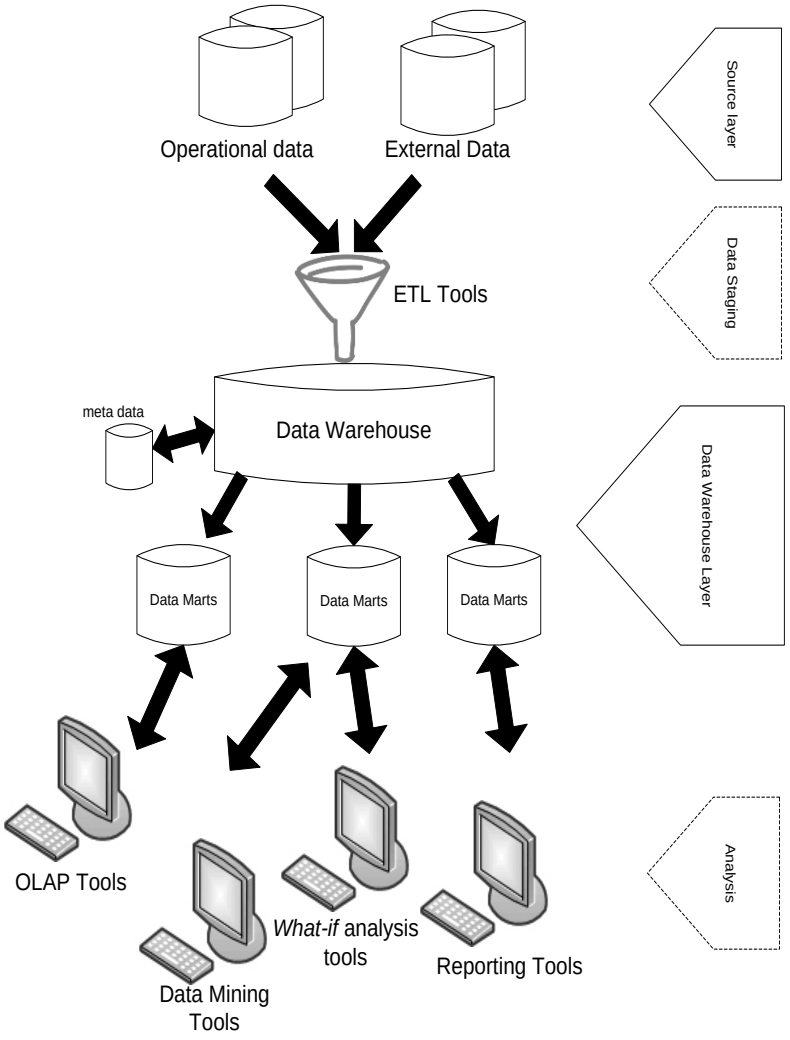
Kelemahan arsitektur ini terletak pada kegagalan untuk memenuhi persyaratan untuk pemisahan antara pengolahan analisis dan transaksional.

Query analisis yang disampaikan kepada data operasional setelah *middleware* menganalisis. Untuk alasan ini, pendekatan virtual untuk *data warehouse* dapat berhasil hanya jika analisis kebutuhan utama dibatasi dan volume data untuk tujuan analisis sangat besar.

b. *Two Layer Architechture*

Two layer architechture yang di tunjukan pada **Error! Reference source not found.** merupakan arsitektur dari *data warehouse* yang menggunaka dua layer atau lapisan sebagai proses pemisahan antara sumber yang tersedia *dengan data warehouse*. Arsitektur ini memiliki empat tahap aliran data, yaitu:

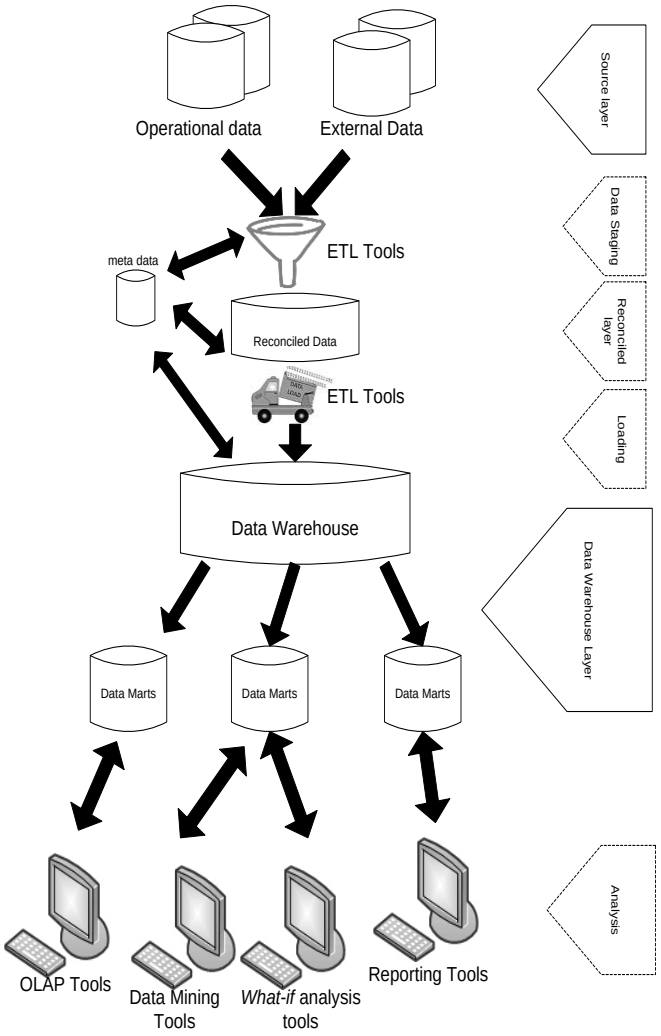
1. *Source Layer*, Sebuah sistem *data warehouse* menggunakan sumber data heterogen. Data yang awalnya disimpan ke *database* relasional perusahaan atau mungkin berasal dari sistem informasi perusahaan.
2. *Staging Data*, Data yang disimpan ke sumber harus diekstrak, kemudian dibersihkan untuk menghilangkan inkonsistensi, dan terintegrasi untuk menggabungkan sumber yang heterogen menjadi satu skema umum. Skema umum tersebut adalah Ekstraksi, Transformasi, dan *Loading* (ETL).
3. *Data warehouse layer*, pada lapisan ini data dapat diakses secara langsung, tetapi juga dapat digunakan sebagai sumber untuk membuat *data mart* dan dirancang untuk departemen perusahaan tertentu. *Meta data* sebagai *repository* yang menyimpan informasi pada sumber, *staging data*, dan sebagainya.
4. *Analysis layer*, data yang terintegrasi secara efisien dan fleksibel diakses untuk mencetak laporan, menganalisis informasi yang di butuhkan dan mensimulasikan hipotesis skenario bisnis yang membahas berbagai jenis pengambilan keputusan perusahaan.



Gambar I-8 *Two layer Architechture*

c. *Three Layer Architechture*

Dalam arsitektur ini, lapisan ketiga adalah lapisan penyesuaian data (*reconciled data*) . Lapisan ini terwujud dari data operasional yang diperoleh setelah mengintegrasikan dan membersihkan sumber data. Akibatnya, data tersebut adalah terintegrasi, konsisten, dan rinci. Gambar II-9 menunjukkan sebuah data warehouse yang tidak di dapat dari sumbernya secara langsung, tetapi diperoleh dari *reconciled data*.



Gambar I-9 *Three Layer Architechture*

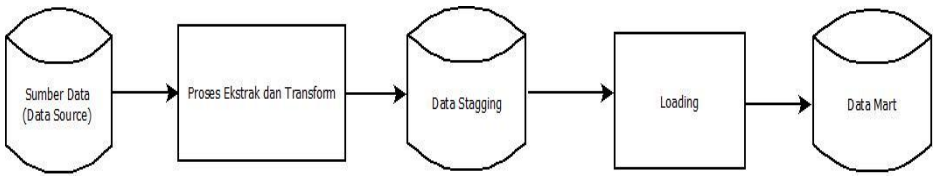
II.2.9 ETL (Extract, Transform, Loading)

Extraction merupakan proses untuk mengidentifikasi seluruh sumber data yang relevan dan mengambil data dari sumber data tersebut. Untuk lebih jelasnya dapat dilihat pada gambar II-10. Penentuan pendekatan yang digunakan pada proses ekstraksi sangat terkait dengan analisis bisnis proses, pendefinisian area subjek , serta desain logik/fisik *data mart* [2]. Proses ini meliputi penyaringan data yang akan digunakan dalam pembuatan *data warehouse*. Dapat langsung dimasukkan

langsung dalam *data warehouse* atau dimasukkan dalam tempat penampungan sementara terlebih dahulu. Pada hakikatnya bagian dari ekstraksi melibatkan penguraian dari data yang telah diekstrak, menghasilkan suatu pengecekan jika data bertemu dengan suatu struktur atau pola yang diharapkan. Jika bukan, data tersebut mungkin ditolak secara keseluruhan.

Proses yang ke dua adalah transformasi data yang telah diekstrak ke dalam format yang diperlukan. Hal ini perlu dilakukan mengingat data yang diambil berasal dari sumber yang berbeda yang kemungkinan memiliki standardisasi yang berbeda pula. Data dari beberapa sistem perlu ditransformasi ke dalam format umum yang disepakati dan digunakan dalam *data warehouse*. *Transformation* merupakan proses yang mempunyai peran dalam melakukan perubahan dan integrasi skema serta struktur yang berbeda - beda ke dalam skema dan struktur yang terdefinisi dalam *data mart* [6] .

Loading merupakan proses pemindahan data secara fisik dan sistem operasional ke dalam *data mart* [6] .Tahap *load* adalah men-load data ke dalam target akhir (*end target*), yang pada umumnya adalah *data warehouse*. Bergantung pada kebutuhan organisasi, proses ini bervariasi secara luas.

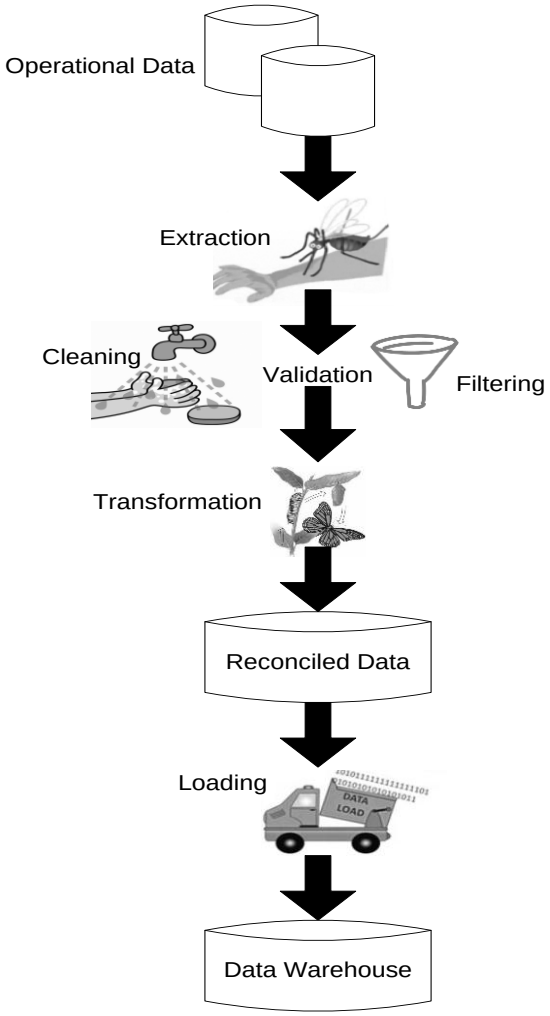


Gambar I-10 ETL Proses

II.2.10 Gambaran Proses ETL (*Data Staging*)

Pada arsitektur *data warehouse* yang umum biasanya menempatkan satu *server database* terpisah yang disebut *Staging* yang berfungsi untuk menangani proses *extract* dan *transform* ini sebelum dilakukan proses *load* ke tujuan akhir *data warehouse*.Menurut Matteo Golfarelli dan Stefano Rizzi dalam bukunya yang berjudul *Data Warehouse Design: Modern Principles and Methodologies*, proses *ETL* adalah proses ekstraksi, integrasi, dan pembersihan data dari data

operasional (*source layer*) sebelum kemudian dimasukkan ke dalam *data warehouse*. Seperti ditunjukkan pada gambar II-11 [4].



Gambar I-11 Data Staging

Di bawah ini adalah penjelasan dari tiap-tiap proses *Extract, Transform,* dan *Load* itu sendiri :

a. *extraction*. Pada proses ini, data operasional yang dibutuhkan pada *data warehouse* akan dibaca (*read*) untuk kemudian masuk dalam proses *cleaning*.
b. Pada poses *cleaning*, data operasional yang telah dibaca akan dilakukan peninjauan kembali pada proses *input* data. Misalnya terjadi redudansi dari data, *record* yang salah satu *field*-nya memiliki *null*

value akan dibersihkan.kemudian data yang sudah di bersihkan, akan di lakukan seleksi data yang memisahkan data yang berguna dan tidak berguna melalui proses *filtering*. Setelah itu proses pemeriksaan kebenaran atau *validating* suatu data yang dihasilkan dari penggabungan, dimana data tersebut telah melewati proses *cleaning* dan *filtering*.

c. Pada proses *transformation*, data yang sudah dibersihkan diubah dari format data opersional menjadi format *data warehouse*.

d. Pada proses *loading*, data yang sudah dibaca, dibersihkan, dan dirubah formatnya akan disimpan pada *data warehouse*. Ada dua teknik dalam proses *loading* ini. Pertama adalah *refresh*, teknik ini akan mengganti data lama di *data warehouse* dengan data yang baru jika diperlukan. Teknik kedua adalah *update*, teknik ini tidak mengganti data yang sudah ada di *data warehouse*, tapi terus menambahkan data yang baru ke dalam *data warehouse*.

II.2.11 OLAP (*On-Line Analytical Processing*)

OLAP (*On-Line Analytical Processing*) adalah salah satu cara untuk mengolah data yang ada pada sebuah *data warehouse*. *OLAP* memberikan jawaban terhadap *query analytic* untuk data yang bersifat multidimensional. *OLAP* menyediakan cara untuk menampilkan data multidimensional yang ada dalam *data mart* atau *data warehouse*, dengan *OLAP* dapat dibuat *cube* yang mengorganisasikan data dan membuat *summary* data untuk *query* yang efisien. Teknik dalam *OLAP* ini ada berbagai macam, di antaranya adalah [7]:

a. *Drill-down and roll-up*

Drill-down and roll-up adalah operasi untuk melihat data global atau detail di sepanjang level hirarki dimensi. *Roll-up* untuk melihat data secara *global* atau rangkuman (*summary*). *Drill-down* memandu pengguna untuk memperoleh data yang lebih detail. *Drill-down* ini biasa digunakan untuk menjawab pertanyaan atas suatu kasus tertentu.

Misalnya untuk menjawab pertanyaan ketika sebuah *summary* number (rata-rata atau jumlah) di bawah atau di atas harapan [7].

b. *Slice and dice*

Slice and dice adalah operasi untuk melihat data sebagai visualisasi dari kubus. Dengan *slice* dan *dice* pengguna dapat melihat data dari beberapa perspektif. Pengguna dapat mengekstrak bagian dari *data agregrated* dan dapat memeriksa dengan detail berdasarkan dimensi-dimensi yang diinginkan. *Data Agregrated* merupakan data pra-perhitungan (*precalculated*) dalam bentuk rangkuman data (data *summarized*) sehingga *query* pada kubus (*cube*) lebih cepat. *Slice* memotong kubus sehingga dapat memfokuskan pada perspektif yang spesifik (pada suatu dimensi). Sedangkan *dice* memberikan kemampuan untuk melihat pemilihan data pada dua dimensi atau lebih, yaitu dengan merotasi *cube* pada perspektif yang lain sehingga pengguna dapat melihat lebih spesifik terhadap data yang di analisis [7].

c. Pivotting

Pivotting merupakan suatu teknik untuk saling menukarkan dimensi data. Dengan melakukan *pivotting* maka dapat diamati suatu informasi atau data dari sudat pandang yang berbeda, sehingga diharapkan akan dapat memperjelas analisis yang dilakukan.