



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB II

LANDASAN TEORI

2.1 Data

Data adalah wujud beberapa nilai dari atribut tertentu dalam catatan spesifik (Hall, 2011).

Data adalah fakta atau observasi mentah, yang umumnya mengenai fenomena fisik atau transaksi bisnis. Secara lebih spesifik, data adalah objek pengukuran atribut-atribut dari suatu entitas. (O'Brien & Marakas, 2010).

Menurut penulis, data adalah fakta mentah mengenai suatu transaksi yang merupakan objek pengukur dari suatu entitas.

2.2 Data Warehouse

2.2.1 Definisi Data Warehouse

Menurut Vaisman dan Zimanyi (2014, p. 53), "*A data warehouse is a particular database targeted toward decision support. It takes data from various operational databases and other data sources and transforms it into new structures that fit better for the task of performing business analysis*". Sebuah *data warehouse* adalah sebuah *database* khusus yang ditujukan untuk pendukung pengambilan keputusan. Data diambil dari *database* operasional yang bermacam-macam dan dari sumber data lainnya kemudian diubah menjadi struktur baru yang lebih cocok untuk melakukan analisis bisnis.

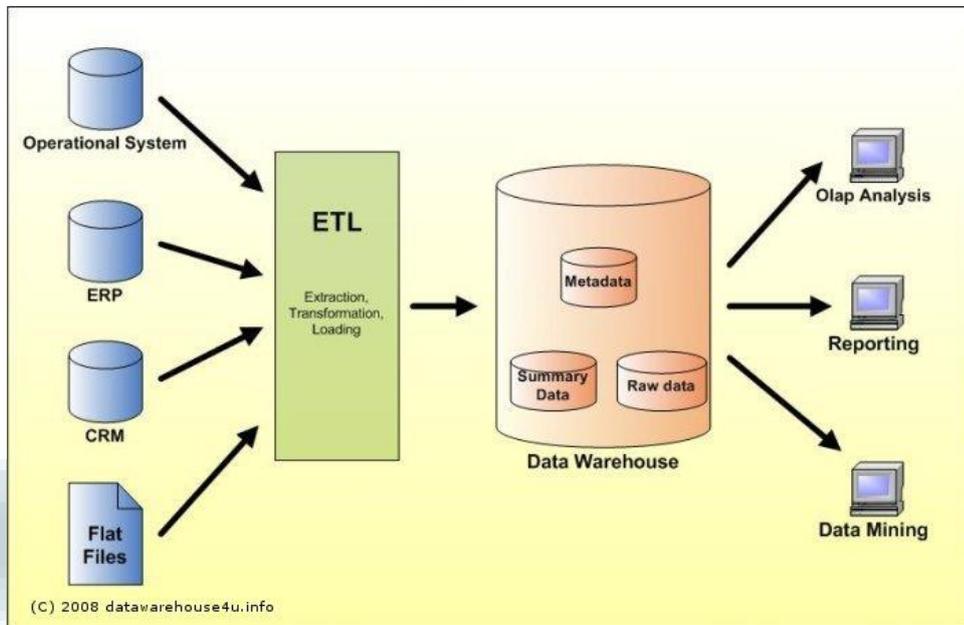
(Ponniah, 2010) dalam studi terbarunya mengemukakan bahwa data *warehouse* adalah informasi yang:

- Menyediakan data yang terintegrasi dari perusahaan
- Membuat informasi saat ini dan data histori dapat dengan mudah tersedia untuk digunakan sebagai strategi pembuatan keputusan.
- Membuat transaksi pendukung keputusan menjadi memungkinkan tanpa menghalangi sistem operasional
- Memberikan data organisasi secara konsisten.
- Menyajikan sumber informasi strategi yang fleksibel dan interaktif.

Menurut Kimball dan Ross (2013, p. 232), "*The data warehouse is the foundation that supports the panoramic 360-degree view of your customers*". *Data warehouse* adalah dasar atau pondasi yang mendukung pandangan secara 360 derajat dari sisi pelanggan.

Menurut penulis, *data warehouse* adalah sebuah tempat penampungan yang mengumpulkan data dari semua sumber *database* yang ada sehingga mempermudah para pembuat keputusan untuk melihat informasi-informasi yang ada pada perusahaan.

2.2.2 Arsitektur Data Warehouse



Gambar 2.1 Typical architecture of a data warehouse

2.2.2.1 Operational Data

Sumber data dari *data warehouse* diambil dari :

- Data operasional *mainframe* diadakan di hirarki generasi pertama dan jaringan *databases*.
- Departemen data diadakan di sistem kepemilikan *file* seperti VSAM, RMS, dan *relational DBMSs* seperti Informix dan Oracle.
- Data pribadi diadakan pada *workstation* dan *server* pribadi.
- Sistem eksternal seperti Internet, *databases* komersial, atau *databases* yang diasosiasikan dengan pelanggan atau penyuplai organisasi.

2.2.2.2 Operational Data Source

Operational Data Store (ODS) adalah sebuah data penyimpanan dari data operasional yang terintegrasi yang digunakan untuk analisis. ODS biasanya dibuat ketika sistem operasi legasi ditemukan tidak mampu memenuhi *requirement* laporan. ODS juga bisa memasok data yang telah dibersihkan dan yang di ekstrak dari sumber sistem.

2.2.2.3 ETL Manager

ETL manager melakukan semua operasi yang terasosiasi dengan ETL dari data ke dalam *warehouse*. Datanya dapat diekstrak secara langsung dari sumber data atau, lebih umumnya, dari penyimpanan data operasional.

2.2.2.4 Warehouse Manager

Manajer *warehouse* melakukan semua operasi yang terasosiasi dengan manajemen dari data di *warehouse*. Operasi yang dilakukan oleh *warehouse manager* adalah sebagai berikut :

- Analisis data untuk memastikan konsistensi;
- Transformasi dan penggabungan sumber data dari penyimpanan sementara ke tabel data *warehouse*;
- Pembuatan indeks-indeks dan *views* pada dasar tabel-tabel;
- Generasi dari denormalisasi (bila diperlukan);

- Generasi dari agregasi;
- *Backing up* dan *archiving data*.

2.2.2.5 Query Manager

Manajer *query* melakukan semua operasi yang berkaitan dengan manajemen dari *query* pengguna. Kompleksitas dari manajer *query* ditentukan oleh fasilitas yang disediakan *end-user access tools* dan database. Operasi yang dilakukan oleh komponen ini termasuk didalamnya *query* langsung pada tabel yang sesuai dan menjadwalkan eksekusi dari *query*.

2.2.2.6 Detailed Data

Area dari *data warehouse* dimana setiap detail data disimpan di skema database. Pada kebanyakan kasus, data tidak disimpan secara online tetapi sengaja dibuat *available* dengan mengagregasikan data lebih *detail* ke level selanjutnya. Namun secara teratur data secara *detail* di tambahkan ke *warehouse* untuk mendukung data yang teragregasi

2.2.2.7 Lightly and highly summarized data

Area pada bagian *data warehouse* ini menyimpan semua ringkasan data yang bersifat *lightly* dan *highly* yang dihasilkan oleh manajer *warehouse*. Bagian area *warehouse* ini bersifat sementara, karena akan berubah secara terus-menerus agar dapat menanggapi perubahan profil *query*.

Tujuan dari ringkasan informasi adalah untuk mempercepat performa dari *queries*. Walaupun akan adanya penambahan biaya operasional yang berasosiasi dengan ringkasan data untuk mengimbangi dengan menghilangkan syarat-syarat yang ditujukan untuk performa berkelanjutan (contohnya seperti *sorting* dan *grouping*) dalam menjawab *user queries*. Ringkasan data tersebut akan di-*update* ketika data baru masuk ke dalam *warehouse*.

2.2.2.8 Archive and Backup

Area ini menyimpan semua ringkasan *detail* data untuk proses *archiving* dan *backup*. Walaupun ringkasan data dihasilkan dari data yang detail, mungkin akan diperlukan untuk *back up* data ringkasan *online* jika data ini disimpan diluar periode penyimpanan data yang *detail*.

2.2.2.9 Metadata

Metada digunakan untuk berbagai tujuan seperti :

- Proses ekstrak dan *loading*: *metadata* digunakan untuk memetakan sumber-sumber data untuk pandangan umum dari data di *warehouse*.;
- Proses manajemen *warehouse*: *metadata* digunakan untuk mengotomatiskan produksi dari *summary tables*.

- Sebagai bagian dari proses manajemen *query*: *metadata* digunakan untuk mengarahkan *query* ke sumber data yang sesuai.

Struktur *metadata* berbeda diantara masing-masing proses, karena tujuannya yang berbeda. Ini menandakan bahwa beberapa Salinan dari *metadata* mendeskripsikan *item* data yang sama yang ada pada *data warehouse*. *End user access tools* menggunakan *metada* untuk memahami bagaimana cara membangun sebuah *query*.

2.2.2.10 End-User Access Tools

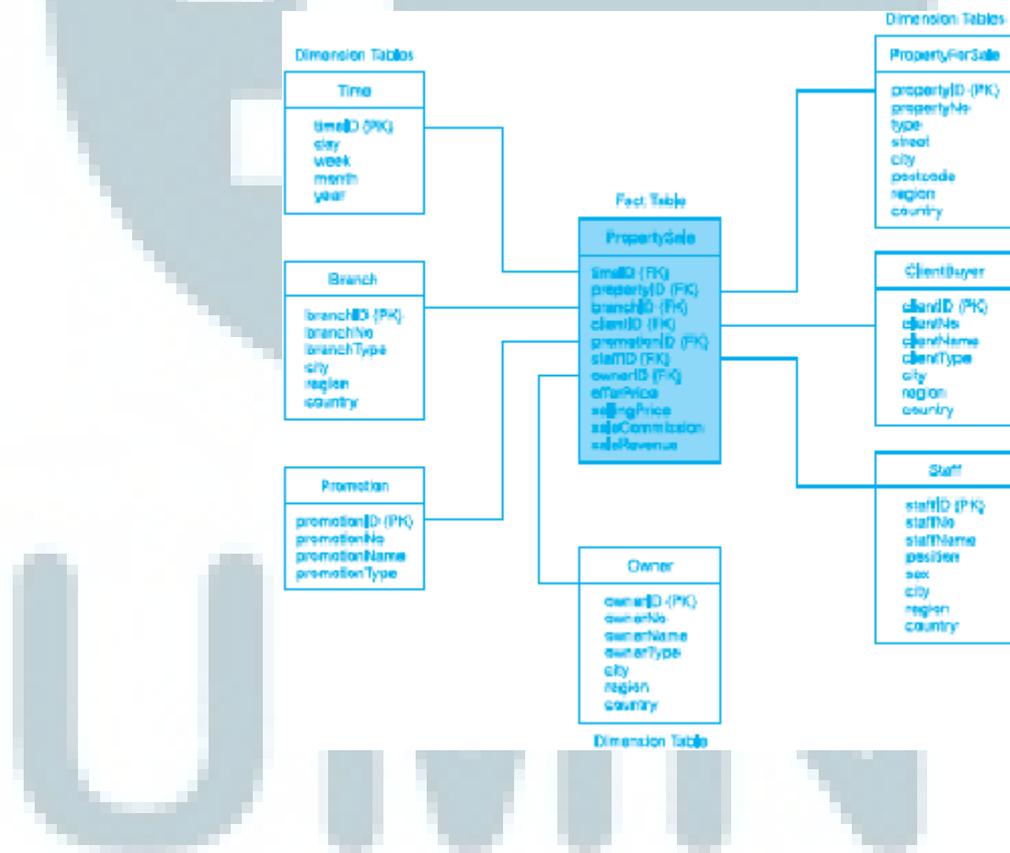
Tujuan utama dari *data warehousing* adalah untuk mendukung pembuat keputusan. Pengguna ini berinteraksi dengan *warehouse* menggunakan *end-user access tools*. *The data warehouse* harus bisa mendukung *ad hoc* dan analisis rutin secara efisien. Performa tinggi didapat dengan *preplanning* persyaratan untuk *joins*, *summations*, dan laporan periodic oleh *end-users*. *End-user access tools* dikategorikan menjadi 4 grup utama, yaitu :

- *Reporting and query tools*;
- *Tools* pengembangan aplikasi;
- *Tools OLAP*;
- *Tools data mining*.

2.2.3 Desain Data Warehouse

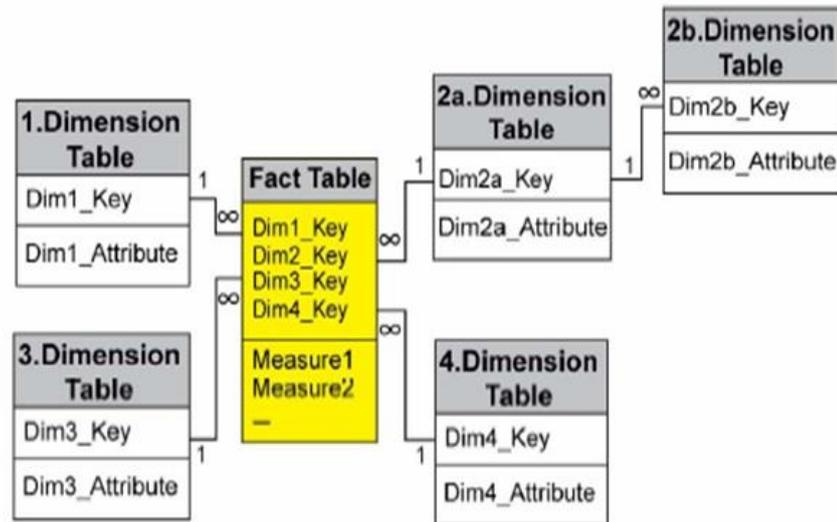
2.2.1 Star Schema

Star Schema adalah model data berbentuk dimensi yang memiliki *fact table* di pusatnya dan dikelilingi oleh tabel dimensi denormalisasi. *Star schema* mengeksploitasi karakteristik dari *factual data* seperti fakta-fakta yang di-generate oleh *events* yang muncul di masa lalu, dan mungkin berubah, terlepas dari bagaimana cara mereka menganalisis.



Gambar 2.2 Star Schema

2.2.2 Starflake Schema

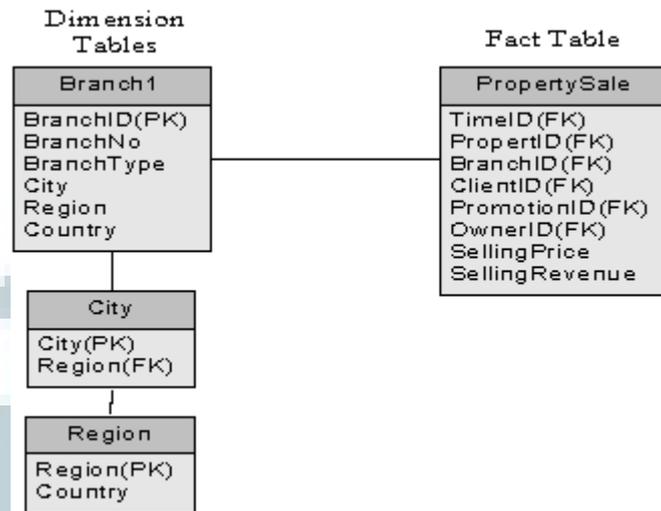


Gambar 2.3 Star Schema

Starflake Schema adalah model data dimensi yang memiliki *fact table* di bagian pusatnya atau tengah dan dikelilingi oleh tabel dimensi normalisasi.

2.2.3 Snowflake Schema

Snowflake Schema adalah model data dimensi yang memiliki *fact table* di bagian pusatnya atau tengah dan dikelilingi oleh tabel dimensi normalisasi dan denormalisasi



Gambar 2.4 Snowflake Schema

2.2.4 Metodologi Data Warehouse

2.2.4.1 Top Down

Persyaratan dari berbagai pengguna digabungkan sebelum proses desain dimulai, dan skema unik dibangun. Selanjutnya dilakukan pemisahan *views* dari koresponden. Pendekatan ini susah untuk dilakukan dan harganya mahal karena *databases* yang besar dan *developer* yang tidak berpengalaman (Vaisman & Zimanyi, 2014).

2.2.4.2 Bottom Up

Skema terpisah yang dibangun untuk setiap grup pengguna dengan persyaratan berbeda, dan kemudian, selama fase penglihatan integrasi, skema-skema ini digabungkan untuk membentuk skema konseptual global bagi seluruh *database*. Pendekatan ini biasanya digunakan untuk *databases yang besar*. (Vaisman & Zimanyi, 2014)

2.2.5 Keuntungan *Datawarehouse*

Keuntungan-keuntungan dari *data warehouse* (Connolly & Begg, 2010).

2.2.5.1 Potensi investasi yang tinggi

Sebuah organisasi harus melakukan sejumlah besar dari sumber-sumber yang ada untuk memastikan implementasi *data warehouse* yang sukses dan biaya sangat bervariasi mulai dari £50,000 sampai £10 *million* karena berbagai macam solusi teknis yang tersedia.

2.2.5.2 Keuntungan Kompetitif

Keuntungan besar atas investasi untuk perusahaan-perusahaan yang telah berhasil mengimplementasikan *data warehouse* adalah bukti dari keuntungan besar yang menyertai teknologi ini. Keunggulan kompetitif diperoleh dengan memungkinkan para pembuat keputusan untuk mengakses data yang sebelumnya tidak tersedia, tidak diketahui, dan belum dimanfaatkan sebagai informasi, misalnya pelanggan, tren, dan permintaan.

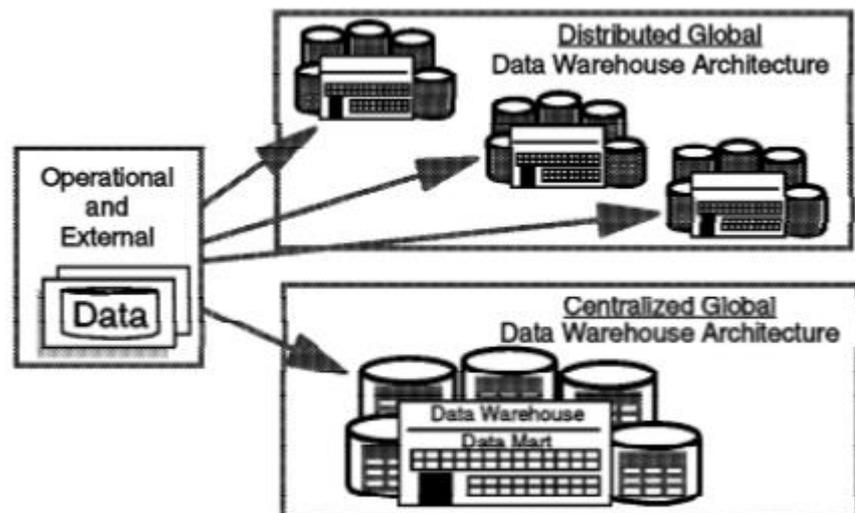
2.2.5.3 Menambah Produktifitas dari pembuat keputusan

Data warehousing meningkatkan produktifitas dari para pembuat keputusan dengan membuat integrasi data yang konsisten, berorientasi subjek, dan data historis. *Data warehouse* mengintegrasikan data dari beberapa sistem yang tidak kompatibel ke

sebuah bentuk yang menyediakan satu *view* yang konsisten dari organisasi. Dengan mengubah data menjadi informasi yang berguna, *data warehouse* memungkinkan para pembuat keputusan perusahaan untuk melakukan analisis yang lebih *substantive*, akurat, dan konsisten.

2.2.6 Bentuk-bentuk *datawarehouse*

Bentuk fisik dari data warehouse bisa terpusat dan terdistribusi di dalam perusahaan (Ballard, et al., 1998).



Gambar 2.5 Dua pendekatan arsitektur data warehouse (Ballard, et al., 1998)

2.2.6.1 Terpusat

Datawarehouse yang terpusat akan digunakan oleh seluruh organisasi yang berada dalam 1 lokasi dan dikelola oleh departemen sistem informasi.

2.2.6.2 Terdistribusi

Datawarehouse yang terdistribusi juga akan digunakan oleh seluruh organisasi, akan tetapi akan terdistribusi dari berbagai lokasi fisik di dalam organisasi dan ini dikelola oleh departemen sistem informasi.

2.2.7 Karakteristik

Inmon(1993), Karakteristik data dibagi menjadi 4 yaitu:

2.2.7.1 *Subject oriented*

Sebagai data warehouse mengorganisasi *major subject* dari perusahaan (contohnya seperti pelanggan, produk-produk, dan penjualan) daripada area *major application* (seperti pembuatan *invoice* pelanggan, pemantauan stok, dan penjualan produk). Ini merupakan refleksi dari kebutuhan untuk menyimpan data *decision-support* daripada data *application-oriented*.

2.2.7.2 *Integrated*

Karena sumber data yang datang bersamaan dari sistem aplikasi perusahaan besar yang berbeda-beda, maka sumber data seringkali tidak konsisten, contohnya, format yang berbeda-beda. Sumber data yang terintegrasi harus dibuat konsisten untuk memunculkan kebutuhan data *user* atau pengguna.

2.2.7.3 *Time Variant*

Karena data dalam *warehouse* hanya akurat dan *valid* pada saat-saat tertentu saja. *Time variant* dari *data warehouse* muncul

pada waktu dimana data itu disimpan, dan asosiasi waktu eksplisit dan implisit dengan semua data, pada kenyataannya data itu dimunculkan kembali sebagai *a series of snapshots*.

2.2.7.4 Non Volatile

Karena data tidak di-update secara *real time* tetapi di-*refresh* dari sistem operasional secara teratur. Data baru selalu ditambahkan sebagai *supplement* dan bukan diganti pada *database*. *Database* ini akan menyerap data baru secara berkelanjutan dan berintegrasi dengan data sebelumnya secara bertahap.

2.3 Database

Database adalah bagian kumpulan dari data yang terkait dengan logika yang terbagi dan, deskripsi dari data itu sendiri, didesain untuk memenuhi kebutuhan informasi dan untuk membantu aktifitas organisasi (Vaisman & Zimanyi, 2014).

Menurut Paulraj Ponniah (2010, p. 559), *database* adalah sebuah repository dimana koleksi data *enterprise* terintegrasi dan diminta yang disimpan untuk pemrosesan komputer dan pembagian data.

Menurut penulis, *database* adalah kumpulan data-data dari komputer yang dapat diolah dan dimanipulasi sesuai keinginan, untuk dijadikan sebagai informasi.

2.4 Junk Dimension

Junk Dimension adalah sebuah pengelompokan dari indikator berkardinalitas rendah. Dengan menciptakan *junk dimension*, maka dapat

menghapus indikator dan *flags* dari table fakta dan menempatkannya pada sebuah kerangka dimensi yang berguna. (Kimball & Ross, 2013)

2.5 *Factless Fact Table*

Tabel *factless fact* dapat digunakan untuk menganalisis apa yang tidak terjadi. *Query* ini selalu memiliki 2 bagian yaitu sebuah *factless fact* yang berisi semua kemungkinan peristiwa yang mungkin akan terjadi dan yang telah terjadi sebelumnya. Ketika aktivitas dikurangi dari cakupan, maka menghasilkan rangkaian peristiwa yang tidak terjadi. (Kimball & Ross, 2013)

2.6 Data Mart

Menurut Connolly dan Begg (2010, p. 1214), *data mart* adalah sebuah *database* yang berisi *subset* dari data korporat untuk membantu kebutuhan analitikal dari unit bisnis tertentu (seperti departemen penjualan) atau untuk membantu pengguna yang berbagi *requirement* yang sama untuk menganalisis proses bisnis tertentu (seperti penjualan properti).

Menurut Paulraj Ponniah (2010, p. 559), *data mart* adalah koleksi dari data yang terkait dari sumber *internal* dan *external* yang berubah, terintegrasi, dan disimpan untuk tujuan penyediaan informasi strategi bagi pengguna tertentu dalam perusahaan.

Menurut penulis, *data mart* adalah *subset data warehouse* yang membantu analisis data perusahaan yang fokus pada departemen tertentu.

2.7 OLAP

OLAP (*Online Analytical Processing*). Mencakup spektrum yang luas dari analisis multidimensi kompleks melibatkan perhitungan yang rumit dan memerlukan waktu respon yang cepat (Ponniah, 2010).

. Menurut Vaisman dan Zimanyi (2014, p. 3) OLAP (*Online Analytical Processing*) adalah sistem yang memungkinkan pengguna untuk berinteraksi dengan *query* dan mengagregasikan data yang terdapat dalam *data warehouse* secara otomatis. OLAP dapat dibagi menjadi 3 pendekatan, yaitu:

2.7.1 *Relational OLAP (ROLAP)*

Pendekatan yang menyimpan data di *relational databases* dan mendukung *extensions* ke SQL dan memiliki metode akses khusus untuk mengimplementasi model data *multidimensional* dan operasi yang berkaitan secara efisien.

2.7.2 *Multidimensional OLAP (MOLAP)*

Pendekatan yang menyimpan data di struktur data *multidimensional* khusus (contohnya *array*) dan mengimplementasi operasi OLAP pada struktur data tersebut.

2.7.3 *Hybrid OLAP (HOLAP)*

Pendekatan yang menggabungkan pendekatan *relational OLAP* dan *multidimensional OLAP*).

2.8 OLTP

OLTP (*Online Transaction Processing*) proses-proses dalam aplikasi yang mengoleksi data *online* selama eksekusi transaksi bisnis. Pemrosesan *order* adalah OLTP. (Ponniah, 2010)

Online Transaction Processing adalah proses Transaksi dan kemampuan untuk mengontrol konkurensi serta teknik *recovery* yang menggaransi konsistensi data. (Vaisman & Zimanyi, 2014)

2.9 Business Intelligence

Business Intelligence (BI) adalah aplikasi yang biasa digunakan secara sinonim dengan informasi yang tersedia di sebuah perusahaan untuk membuat keputusan strategi (Ponniah, 2010).

Business Intelligence (BI) adalah serangkaian proses dan struktur data yang digunakan untuk menganalisis data dan informasi yang digunakan untuk mendukung strategi pembuatan keputusan. (Imhoff, Galemme, & Geiger, 2003)

2.10 Dashboard

Dashboard adalah mekanisme tampilan *visual* untuk mengaktifkan pengguna bisnis pada setiap *level* untuk menerima informasi yang mereka butuhkan untuk membuat keputusan yang lebih baik yang dapat meningkatkan performa. (Ponniah, 2010)

Dashboard adalah alat visualisasi yang paling populer di *business intelligence* yang dipakai untuk melihat data organisasi dan memanfaatkan model pengukuran *performance* yang berbeda untuk mengidentifikasi dan

mengimplementasi ukuran semua level pada organisasi (Vaisman & Zimanyi, 2014).

2.11 Tools

Tools yang digunakan dalam pembuatan *data warehouse* ini adalah Pentaho *Business Analytics*. *Pentaho Business Analytics* adalah sebuah rangkaian produk *business intelligence* (Vaisman & Zimanyi, 2014). Yang terbagi menjadi 2 versi : *enterprise edition* yang bersifat komersial dan *community edition* yang bersifat *open source*. Komponen-komponen utama dari *pentaho business analytics* yaitu :

2.11.1 Pentaho Business Analytics Platform

Komponen ini bertindak sebagai *connection point* untuk semua komponen. Ini memungkinkan untuk solusi unifikasi (kompak) dan integrasi *end-to-user* data menjadi data visualisasi. Ini juga menyediakan satu set alat untuk pengembangan, *deployment*, manajemen dari aplikasi-aplikasi yang ada.

2.11.2 Pentaho Analysis Services

Pentaho analysis services dikenal juga sebagai *Mondrian*, yaitu sebuah *relational OLAP server*. *Mondrian* mendukung MDX (*multidimensional expression*) *query* dan XML, untuk analisis dan *OLAP4J* spesifikasi antarmuka. *Mondrian* membaca dari SQL dan sumber data lain dan mengagregasi data dalam sebuah *memory cache*.

2.11.3 Pentaho Data Mining

Menggunakan *Waikato environment* untuk pentahuan analisis (*weka*) untuk mencari pola data. *Weka* terdiri dari mesin algoritma untuk tugas *data mining*. Ini berisi fungsi dari *data processing*, *regression analysis*, *classification methods*, *cluster analysis*, dan visualisasi.

2.11.4 Pentaho Data Integration

Dikenal juga dengan nama *Kettle*, berisi sebuah *engine* integrasi data (ETL) dan *GUI* (*graphical user interface*) yang memungkinkan pengguna untuk men-define *data integration job* dan *transformations*. *Pentaho Data Integration* juga mendukung *deployment* dalam *node* tunggal dan *cloud* atau *cluster*.

2.11.5 Pentaho Data Integration

Pentaho report designer adalah sebuah penulis *report visual* yang dapat melakukan *query* dan memakai data dari banyak sumber. *Pentaho Report Designer* terdiri dari *core reporting engine*, dan mampu men-generate laporan dalam beberapa format dengan dasar *XML file*.

2.12 Penelitian Terdahulu

Metode yang digunakan dalam mendesain *data warehouse* ini adalah *Nine Steps Methodology* oleh Kimball. Metode ini dipilih peneliti karena pertimbangan peneliti terhadap kesuksesan hasil penelitian yang dilakukan oleh peneliti sebelumnya. Berikut ulasan singkatnya:

Penelitian oleh Andreas, Fransisca Theresia, Eileen Octavia Wibowo dari Universitas Bina Nusantara dengan skripsi yang berjudul ANALISIS DAN PERANCANGAN DATA WAREHOUSE KREDIT KENDARAAN BERMOTOR PADA PT. MANDALA MULTIFINANCE . Hasil dari penelitian ini menunjukkan pengambilan data yang lebih efektif untuk analisis pola perilaku konsumen pada tiap cabang perusahaan.

Penelitian oleh Victor Prasetya dari Universitas Multimedia Nusantara dengan skripsi berjudul ANALISIS DAN PERANCANGAN APLIKASI BUSINESS INTELLIGENCE PADA BAGIAN PEMASARAN UNIVERSITAS MULTIMEDI NUSANTARA. Hasil dari penelitian ini adalah *dashboard* yang interaktif yang mampu menjawab kebutuhan informasi *user* dengan baik.

Penelitian oleh John Kendar dari Universitas Multimedia Nusantara dengan skripsi berjudul ANALISIS DAN PERANCANGAN DATA WAREHOUSE GARUDA FREQUENT FLYER PADA PT. GARUDA INDONESIA (PERSERO) Tbk.

UMMN