

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Sudah banyak penelitian yang berkaitan dengan data warehouse diantaranya:

Penelitian milik (Paramita, 2016) yang berjudul “Perancangan *Data Warehouse* pada Perpustakaan Yayasan Lentera Insan”. Dalam penelitian ini, Aulia mengembangkan suatu data warehouse pada perpustakaan yang sebelumnya masih menggunakan teknik manual untuk menghubungkan tiap-tiap modul yang terpisah. Kebutuhannya disesuaikan dengan kebutuhan manajemen perpustakaan, dan hasilnya data menjadi lebih terstruktur dan terintegrasi sehingga memudahkan pihak perpustakaan dalam mengambil keputusan.

Selanjutnya penelitian berjudul “Implementasi Data Warehouse dan Data Mining untuk Penentuan Rencana Strategis Penjualan Batik (Studi Kasus Batik Mahkota Laweyan)” oleh (Al Irsyadi, 2014). Dalam penelitiannya, Fatah menggunakan metode ETL dengan model Snowflake Schema. Dan menggunakan monitoring Cube Browser. Informasi yang ditampilkan menggunakan proses drill down. Hasil dari penelitian ini data warehouse Batik Mahkota Laweyan memiliki empat table dimensi, empat table sub dimensi dan satu table Fakta.

Adapun penelitian oleh (Henderi & Winarko, 2015) dengan judul “Rancangan Sistem Data Warehouse Evaluasi Kinerja”. Penelitian ini melalui tahap menentukan subjek, studi literature, observasi proses bisnis, menyusun KPI, menentukan representasi KPI dan menentukan data. Hasil yang diperoleh ialah rancangan fisik dan model data.

Menurut (Wilanda, 2016) penelitiannya yang berjudul “Pembangunan Data Warehouse pada PT.PUPUK ISKANDAR MUDA”. Metode yang digunakan pada penelitian ini adalah pemodelan skema Fact constellations dan perancangan OLAP (OnLine Analytical Processing), dan untuk pengolahan dan analisis datanya dengan menggunakan teknik Roll-Up dan Drill-Down. Tujuan dari penelitian ini adalah untuk membantu perusahaan dalam mengintegrasikan data dan menganalisa penualan dan produksi barang. Hasil dari penelitian menunjukkan bahwa metode pemodelan skema

Fact constellations dan perancangan OLAP akurat diterapkan untuk mengintegrasikan dan menganalisa data penjualan dan produksi perusahaan.

Penelitian selanjutnya dilakukan oleh (Dharayani, Laksitowening, & Yanuarfiani, 2015). Dengan judul penelitian “Implementasi ETL (Extract, Transform, Load) Pangkalan Data Perguruan Tinggi dengan Menggunakan State Space Problem”. Penelitian ini mengimplementasikan ETL dengan menggunakan alur kerja state space problem. State Space Problem digunakan untuk menggambarkan alur proses ETL dan mencari urutan aktivitas pada proses ETL. Dari hasil pengujian ETL dilakukan perubahan urutan aktivitas dengan menggunakan transisi graf dan didapatkan hasil yang lebih optimal.

Penelitian yang lainnya oleh (Linda, 2015) dengan judul “Pengisian Borang Akreditasi Standar 3 Menggunakan Model Data Warehouse”. Penelitian ini dilakukan untuk pengisian boring akreditasi pada formulir standar 3 tentang data mahasiswa dan lulusan serta mendapatkan informasi berkaitan dengan jumlah calon mahasiswa baru, jumlah mahasiswa baru pada fakultas, jumlah lulusan yang sekaligus dikelompokkan berdasar jenis indeks prestasi kumulatif serta persentasinya. Penelitian ini menggunakan metode Online Analytical Processing dan metode Business Life Cycle.

Penelitian selanjutnya berjudul “Perancangan Data Warehouse Sistem Informasi Eksekutif dengan Metode Kimball (Kasus pada Universitas Muhammadiyah Magelang)” oleh (Asroni, 2014). Penelitian ini menggunakan Nine Step Methodology (metode 9 langkah) dari Kimball. Penggunaan metode ini dikarenakan metode ini memiliki tahap perancangan yang dimulai dari kebutuhan, sehingga sistem mampu menyajikan informasi dengan struktur lengkap dan sesuai dengan kebutuhan pengambil keputusan. Hasil dari penelitian ini yaitu Data Warehouse yang sudah digunakan untuk membangun suatu bagian Sistem Informasi Eksekutif yaitu bidang Akademik (SIE-Akademik), Sistem ini mampu menyajikan data yang terintegrasi dan dapat ditampilkan sesuai dimensi factual yang diinginkan berupa ringkasan IP mahasiswa selama 7 tahun.

Penelitian lainnya oleh (Fadilah, Winarno, & Amborowati, 2016). Penelitian yang berjudul “Perancangan Data Warehouse untuk Sistem Akademik STMIK Kediri”.

Metode pada penelitian ini adalah metode penelitian tindakan menggunakan action research diantaranya tahap Observasi, Reflect, Plan dan Act. Tujuan menggunakan metode ini untuk merencanakan subyek data, merancang skema, bintang, merancang data warehouse kemudian diimplementasikan dengan cara Ekstrak, Transform. Hasil dari penelitian ini adalah Rancangan Data Warehouse akademik yang sesuai dengan STMIK Kediri dan menghasilkan laporan berupa jumlah mahasiswa tiap angkatan, jenis kelamin, dan per kota asal mahasiswa.

Penelitian lain yang berkaitan dengan data warehouse adalah “Perancangan dan Implementasi Data Warehouse untuk mendukung Sistem Akademik (Studi Kasus pada STKIP Muhammadiyah Kotabumi)”. Penelitian ini dilakukan oleh (Khotimah & Sriyanto, 2016). Penelitian ini menggunakan metode Kimball “nine-step methodology”, untuk perancangan data warehouse, menggunakan metode bottom-up. Tujuan dari perancangan ini adalah melayani informasi strategis seperti jumlah total Mahasiswa baru per batch, status mahasiswa dan jumlah hasil berdasarkan PGA dan predikat. Hasil dari perancangan ini adalah pemodelan skema bintang yang menjadikan data terorganisir, memudahkan dan menyederhanakan dalam proses query analisis untuk penyajian data sebagai pendukung proses evaluasi dan pelaporan.

Penelitian selanjutnya oleh (Wayanwiswani, 2013) yang berjudul “Penerapan Hybrid Slowly Change Dimension untuk Nearly Realtime Data Warehouse”. Dalam penelitiannya, Ni Wayan menilai bahwa dalam melakukan perancangan data warehouse menggunakan pemodelan dimensi Hybrid Slowly Change Dimension dapat menangkap setiap perubahan data yang dapat mempengaruhi dimensi yang terjadi pada database sumber, karena Ni Wayan juga mengembangkan teknik ini menggunakan metode change dan capture. Sehingga penelitian ini mempunyai hasil terbentuknya desain SCD yang diimplementasikan dalam table-tabel form sistem thesis dan sistem disertasi yang berfungsi mencatat perubahan yang terjadi dalam sebuah kolom baru pada table, nilai kolom lama yang mengalami perubahan akan digunakan untuk menampung field awal sebelum perubahan dan kolom baru akan diisi nilai dari perubahan yang dilakukan.

Penelitian terakhir tentang data warehouse oleh (Yosua Adi Kharisma, Ramos Somya, 2017) yang berjudul “Perancangan Data Warehouse pada Toko Wisata Rasa Surabaya menggunakan Skema Galaxy”. Toko Wisata Rasa menggunakan aplikasi Point of Sales (POS) untuk mencatat transaksi penjualan sehari-hari. Pada penelitian ini Data Warehouse dibangun dengan menggunakan skema Galaxy yang diterapkan dengan basis data MySQL yang bersifat *free*. Selain itu juga dikembangkan aplikasi *Online Analytical Processing* (OLAP) berbasis web dengan memanfaatkan *library Highcharts* untuk menampilkan grafik. Aplikasi OLAP dikembangkan menggunakan metode Prototyping. Hasil pengujian ini menunjukkan bahwa *Data Warehouse* yang terdiri dari dua table fakta bermanfaat untuk menampung semua data transaksi selama ini dan aplikasi OLAP yang dibuat juga mampu menyajikan informasi yang baik melalui grafik yang telah diolah menggunakan *library Highcharts*.

2.2 Pengertian Data

Data terdiri dari banyak informasi maupun keterangan yang didapat melalui pengamatan dan pencarian ke sumber-sumber. Data yang didapat merupakan suatu anggapan atau fakta yang belum diolah lebih jauh. Apabila data yang ada sudah diolah menggunakan penelitian maupun percobaan sebelumnya, maka data yang ada akan menjadi lebih kompleks seperti database, informasi, maupun solusi untuk memecahkan masalah tertentu.

2.2.1 Pengertian Database

Database adalah kumpulan data yang terdiri dari arsip, table, file yang disimpan dalam sistem computer yang diolah dengan mudah untuk mendapatkan maupun menghasilkan informasi bagi pengguna.

2.2.2 Pengertian Data Warehouse

Data Warehouse terdiri dari data yang diintegrasikan yang sudah disimpan dalam jangka waktu tertentu dan digunakan untuk menambahkan informasi. data warehouse merupakan koleksi lengkap / tetap (*non-volatile*) yang digunakan untuk mendukung proses pengambilan keputusan, yang berorientasi pada subyek, terintegrasikan dan memiliki dimensi waktu (Inmon, 2002). Beliau juga menyatakan bahwa data warehouse merupakan dasar dari proses pengambilan keputusan.

2.2.3 Karakteristik Data Warehouse

Data Warehouse menurut (Thomas Connolly, 2016) memiliki beberapa karakteristik, diantaranya yaitu:

1. *Subjected-oriented*, ialah data warehouse yang didesain berdasar subject tertentu dari organisasi, yang berdasar pada area aplikasi utama, bukan berdasarkan fungsi maupun proses aplikasi. Data warehouse dikelompokkan pada sekitar subjek utama dari para pelanggan, produk dan penjualan dan juga berdasar dengan aplikasi pembutan faktur pelanggan, pengendalian persediaan juga penjualan produk. Hal ini dikarenakan data yang ada pada data warehouse adalah data yang digunakan untuk mengambil keputusan bukan data yang berorientasi terhadap aplikasi.
2. *Integrated*
Data Warehouse dapat menyimpan data yang berasal dari sumber-sumber data yang terpisah dan dari aplikasi yang berbeda, kedalam format yang konsisten dan dapat terintegrasi satu sama lain. Sehingga data tidak akan terpecah dan konsisten untuk terintegrasi dari data ke user.
3. *Time-variant*
Data warehouse menyimpan data histori atau hanya menyimpan data yang akurat dan valid pada rentang waktu tertentu.
4. *Non-Volatile*
Data yang juga terdapat pada data warehouse tidak di-upgrade secara real time tetapi di perbaru dari sistem operasional secara umum. Data terbaru juga selalu ditambah ke database itu sendiri. Database itu secara bertahap akan menerima dan menyerap data baru, lalu akan disatukan dengan data sebelumnya.

2.3 Pengertian OLAP dan OLTP

2.3.1 OLTP (Online Transaction Processing)

OLTP adalah sebuah proses data yang berfokus pada tugas yang berorientasi dalam transaksi sehari-hari (*insert, update, delete*) dimana datanya adalah data yang terbaru dan spesifik. Penekanan utama dalam sistem OLTP ini diletakkan pada proses

query yang cepat, menjaga integritas didalam lingkungan multi-akses dan efektivitas diukur dengan jumlah transaksi per detik.

2.3.2 OLAP (Online Analytical Processing)

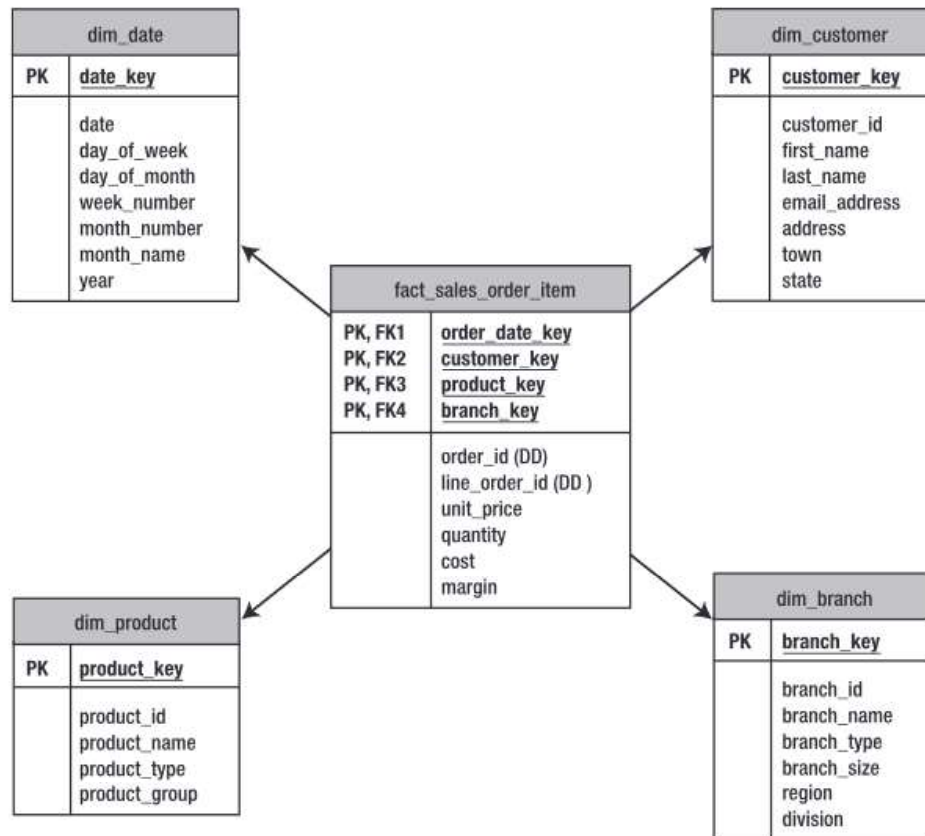
OLAP biasanya ditandai dengan volume transaksi yang relative rendah. Query biasanya sangat rumit dan terdapat agregasi. Untuk sistem OLAP waktu respon adalah sebuah ukuran efektivitas. Aplikasi OLAP banyak digunakan untuk teknik data mining. Pada databse OLAP ada agregasi, data historis, disimpan dalam skema multi-dimensional (biasanya berupa skema bintang).

2.4 Model Data Multidimensi

Model Data ini berupa kubus (*cube*) yang dapat menyimpan juga mengakses ddata dalam batas yang telah ditetapkan. Adapun definisi dari Model data ini terdiri dari table dimensi dan table fakta. Tabel dimensi berupa table yang berisi data rinci yang menjelaskan tentang foreign key yang ada pada table fakta. Sedangkan table fakta ialah table yang berisi kumpulan foreign key dari primary key yang ada pada masing-masing table dimensi. Pemodelan dimensional ini berupa teknik pemodelan data warehouse yang paling sering digunakan karena relasi dan table dapat memaksimalkan performa dalam mengambil keputusan pada database relasional. Adapun beberapa model skema yang paling sering digunakan yaitu:

1. Star schema

Star schema adalah data warehouse yang bentuknya paling sederhana karena bentuknya yang menyerupai bintang. Skema ini tersusun dari satu table fakta yang berada ditengah, lalu dikelilingi oleh table dimensi. Setiap table dimensi selalu berhubungan dengan table fakta. Dan satu table fakta harus mempunyai relasi dengan satu table dimensi. Karakteristik dari skema bintang ini ialah table dimensinya yang tidak dinormalisasikan Berikut contoh skema bintang:

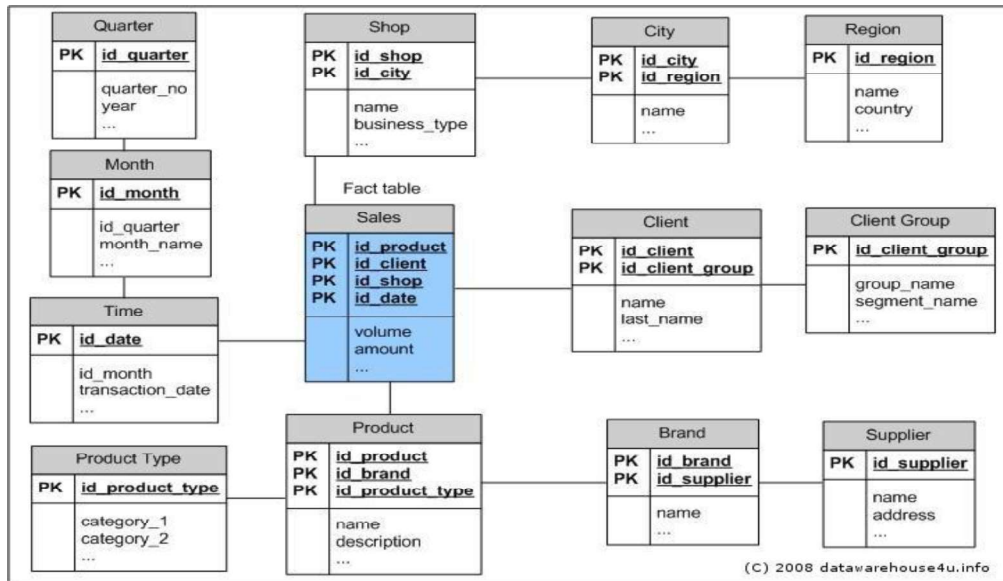


Gambar 2.1 *Star Schema* (Rainardi 2008)

2. *Snowflake Schema*

Snowflake schema ialah skema yang lebih kompleks dari skema bintang yang sama-sama digunakan untuk menyimpan data. Hal yang membedakan dari skema bintang adalah data yang ada pada table dimensi dilakukan normalisasi sebagian maupun keseluruhan data, gunanya untuk mengurangi duplikat data pada table.

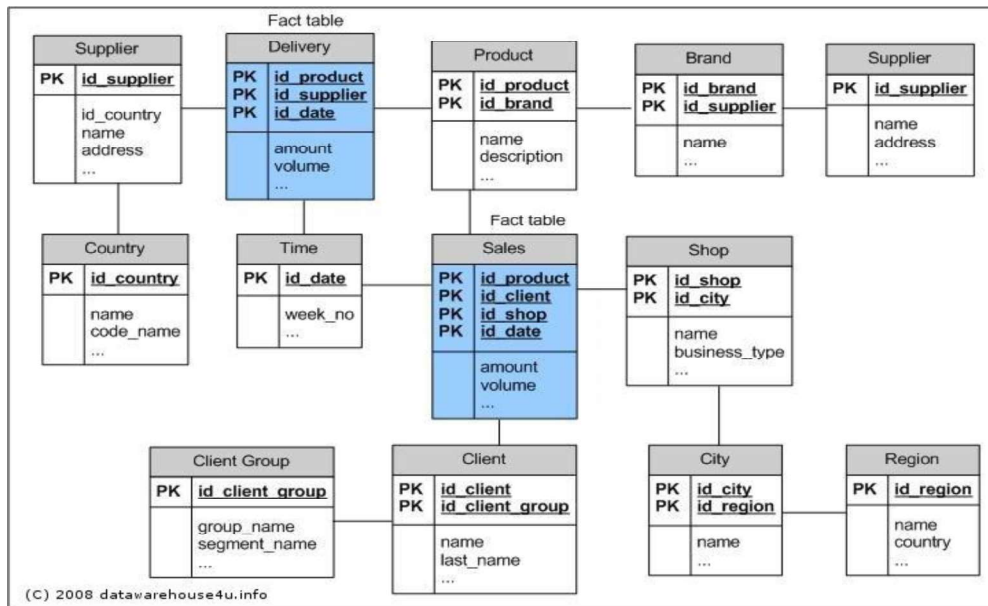
Berikut gambar dari *snowflake skema*:



Gambar 2.2 *Snowflake Schema* (Rainardi, 2008)

3. Constellation Schema

Constellation Schema ialah bentuk pengembangan dari *star schema*. Skema ini terdiri dari lebih dari satu table fakta yang saling berbagi satu atau lebih table dimensi. Keuntungan dari skema ini adalah dapat menghemat memori.



Gambar 2.3 *Constellation Schema* (Inmonna, n.d.)

2.5 ETL (Extraction, Transforming, and Loading)

Pengertian dari ETL adalah suatu proses dalam pembentukan data warehouse. Proses ETL memiliki tujuan yaitu untuk mengumpulkan, menyaring, mengelola dan menggabungkan data yang valid dari berbagai sumber yang berbeda-beda lalu untuk disimpan ke dalam sebuah data warehouse. Proses ETL ini terdiri dari Extracting, Transforming, dan Loading. Berikut penjelasan tiap proses:

1. *Extract*

Extract ialah proses pengambilan data yang diekstrak dari sistem operasional, Bisa menggunakan query atau aplikasi ETL itu sendiri. Proses ini dapat berlangsung berkali-kali guna mendapatkan data yang berkualitas.

2. *Transform*

Transform merupakan proses kedua setelah proses ekstrak. Data yang didapat dari hasil ekstrak akan di bersihkan dan diubah dari data asli ke bentuk data yang sesuai dengan standard data warehouse.

3. *Load*

Load merupakan proses terakhir yang kegunaannya untuk memasukkan data yang sudah ditransform ke dalam pangkalan data.

2.6 Pengertian Fact Tabel dan Dimension Tabel

1. Fact Tabel

Tabel Fakta ialah tabel yang berisi nilai-nilai numerik atau nilai-nilai yang dapat diukur (*measure*), seperti harga, jumlah barang dan lain-lain. Tabel fakta ini juga berisi kumpulan dari foreign key dari primary key yang ada pada masing-masing table dimensi. Terdapat data historis pada table.

2. Dimension table

Tabel dimensi adalah table yang berisikan data detail/rinci yang menjelaskan tentang foreign key yang ada pada table fakta. Atribut yang ada pada dimensi table dibuat secara hirarki untuk memudahkan proses query.

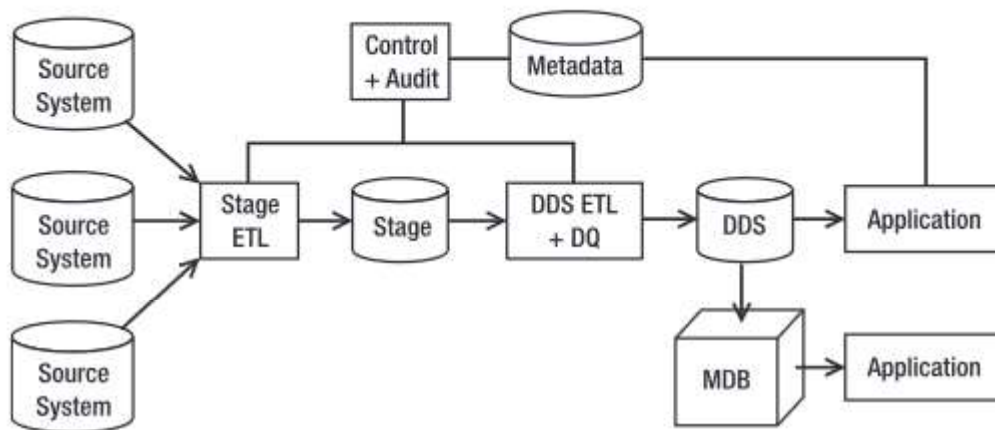
2.7 Arsitektur Data Warehouse

Data Warehouse atau Pangkalan data memiliki dua arsitektur yaitu, arsitektur data *flow* dan arsitektur sistem. Arsitektur data *flow* adalah tentang bagaimana data

store disimpan dalam data warehouse dan bagaimana aliran data dari sistem sumber pada user melewati data store ini. Ada beberapa jenis arsitektur data *flow* yaitu single DDS, NDS + DDS, ODS + DDS, dan federated data warehouse. Berikut penjelasan tiap arsitektur data flow:

2.7.1 Single DDS

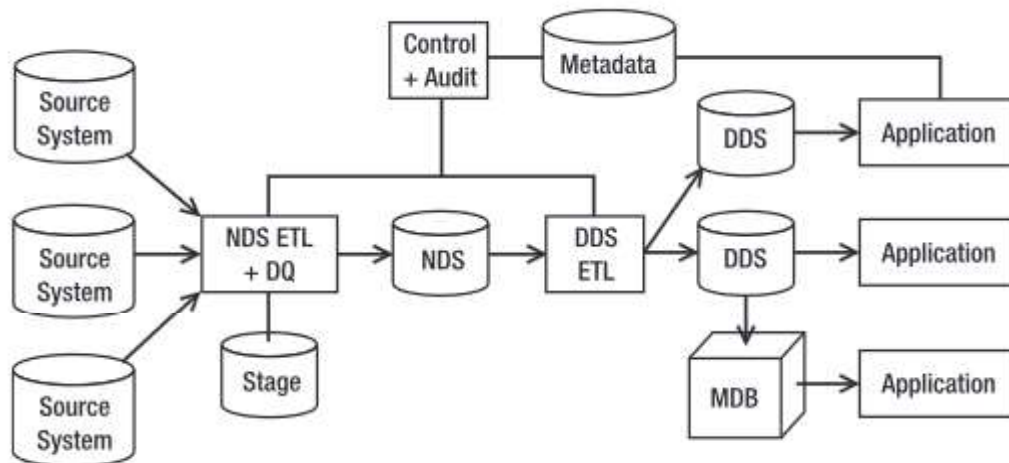
Pada arsitektur single DDS, hanya ada satu dimensional data store. Dimensional data store ini terdiri dari satu dimensional data mart. Dimensional data mart berisi kumpulan table fakta yang saling berhubungan dan table dimensi berhubungan berisi pengukuran. Sebuah paket ETL mengekstrak data dari beberapa sistem sumber dan menemukannya pada stage. Berikut adalah bentuk arsitektur Single DDS



Gambar 2.4 *Single DDS Architecture*(Rainardi, 2008)

2.7.2 NDS + DDS

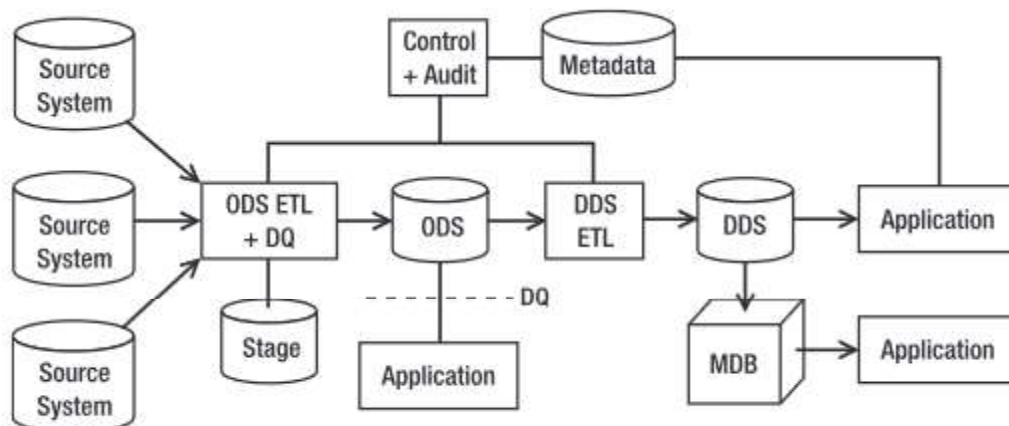
Arsitektur ini sama dengan arsitektur single DDS. Yang sedikit membedakannya ialah terdapat sistem normalized data di depan DDS. NDS memiliki kemampuan untuk *me-load* data ke beberapa DDS.



Gambar 2.5 *NDS + DDS Architecture* (Rainardi 2008)

2.7.3 ODS + DDS

Arsitektur ini juga mirip mirip dengan NDS + DDS, namun memiliki ODS pada NDS. Berbeda dengan NDS, yang berupa internal data store. Pada arsitektur ini DDS adalah master data store. ODS tidak memiliki master data histori, dan hanya berisi versi terbaru dari data master.

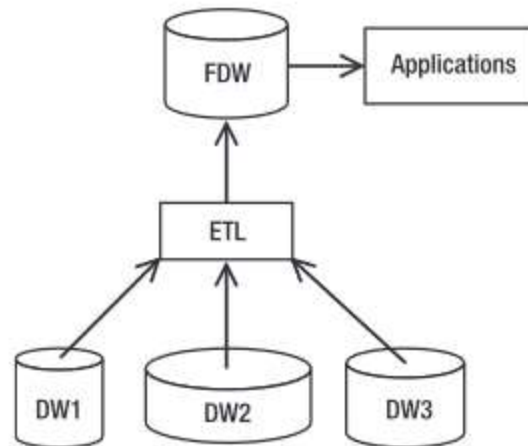


Gambar 2.6 *ODS + DDS Architecture* (Rainardi 2008)

2.7.4 Federated Data Warehouse (FDW)

FDW tersusun dari beberapa pangkalan data yang dipuncaknya terdapat lapisan pengambilan data. Sebuah FDW menerima data dari pangkalan data menggunakan

ETL dan me-load data ke dimensional data store yang baru. Karena adanya kegiatan penggabungan dan pembebasan, maka kita dapat memiliki tiga pangkalan data. Yaitu pangkalan data dimensi, normalisasi pangkalan data bentuk normal ketiga, dan yang terakhir adalah relasional data warehouse dengan table transaksi besar yang mereferensikan banyak table referensi.



Gambar 2.7 *Federated Data Warehouse Architecture* (Rainardi 2008)