

BAB II

LANDASAN TEORI

2.1 Database

Database adalah sebuah tempat penyimpanan yang dapat digunakan secara bersamaan oleh beberapa departemen dan pengguna. *Database* memiliki data yang saling terintegrasi yang merupakan sumberdaya bersama di dalam perusahaan.

Database memiliki entitas, atribut, dan relasi. Entitas adalah objek yang berbeda yang terdapat pada sebuah database (orang, tempat, benda, konsep, atau peristiwa). Atribut adalah properti yang menjelaskan beberapa aspek dari sebuah objek yang ingin direkam. Relasi adalah hubungan antara satu entitas dengan entitas lainnya.

2.2 Data

Data adalah fakta mentah atau deskripsi dasar dari benda, peristiwa, aktivitas dan transaksi yang didapatkan, direkam, disimpan diklasifikasi tetapi belum terorganisir untuk dapat menyampaikan suatu arti yang spesifik (Turban & Rainer, 2009).

2.3 Informasi

Informasi adalah suatu koleksi fakta yang didapat dari data yang telah terorganisir dengan beberapa cara sehingga dapat memberikan suatu arti yang dapat dipahami oleh penerima (Turban& rainer, 2009).

2.4 Visualisasi Data

Visualisasi data adalah sebuah teknik penyajian data kedalam bentuk visual sehingga dapat memudahkan pengguna untuk memahami konsep dari data yang ditampilkan.

Penyajian informasi dalam bentuk visual akan mempermudah pengguna untuk memahami informasi yang diberikan, sehingga akan mempermudah proses pengambilan keputusan di dalam sebuah perusahaan.

2.4.1 Grafik Garis

Grafik garis adalah visualisasi data kedalam bentuk garis, yang diperoleh dari beberapa ruas garis yang menghubungkan titik-titik pada batang bilangan. Grafik garis memiliki dua garis sumbu yaitu horizontal dan vertikal.

Garis horizontal digunakan untuk menempatkan bilangan-bilangan yang bersifat tetap, seperti tahun dan ukuran-ukuran. Garis vertikal digunakan untuk menempatkan bilangan-bilangan yang sifatnya berubah-ubah, seperti harga, biaya, dan jumlah.

2.4.2 Grafik Lingkaran

Grafik lingkaran adalah visualisasi data kedalam bentuk lingkaran. Grafik lingkaran terdiri dari juring-juring (sektor) yang sesuai dengan nilai data tersebut. Pada diagram lingkaran bagian-bagian dari keseluruhan data dinyatakan dalam bentuk persen.

2.4.3 Grafik Batang

Grafik batang adalah visualisasi data kedalam bentuk balok (batang). Grafik ini berbentuk persegi panjang yang dilengkapi dengan skala tertentu, sesuai dengan data yang dimiliki. Pada grafik batang, batang dapat disusun dalam bentuk tegak atau mendatar.

2.5 *Dashboard*

Dashboard adalah sebuah koleksi data informasi yang digunakan untuk melakukan evaluasi performa dan memastikan tercapainya tujuan dari sebuah operasi (Ronald R, 2010).

Dashboard adalah tampilan visual dari informasi yang paling penting yang diperlukan untuk memahami dan mengelola satu atau lebih bidang organisasi yang diatur pada satu layar komputer sehingga dapat dipantau dengan mudah (Stephen Few, 2006).

Dari pengertian di atas, maka sebuah *dashboard* harus dapat mendukung proses *monitoring* secara visual seperti:

1. Melihat gambaran besar informasi.
2. Fokus pada item tertentu yang memiliki informasi penting.
3. Dapat melakukan drill ke informasi tambahan yang diperlukan untuk proses pengambilan keputusan.

2.5.1 Membuat *Dashboard* Yang Efektif

Komunikasi visual akan efektif bila diselaraskan dengan cara manusia melihat dan berfikir. Dengan kata lain, agar dapat bekerja secara efektif *dashboard* harus dapat dengan mudah dipahami dan dimengerti oleh pengguna. (Henry Dafid Thoreau, 2006) Hal yang penting pada proses preancangan desain adalah kesederhanaan desain.

Ketika merancang *dashboard* informasi yang ditampilkan hanya yang benar-benar dibutuhkan, menyederhanakan informasi tanpa harus menurunkan makna dari informasi tersebut, dan menampilkannya menggunakan mekanisme tampilan visual yang dapat dengan mudah dibaca dan dipahami.

2.6 *Business Intelligence (BI)*

Business Intelligence adalah sekumpulan model matematika dan metodologi analisa yang secara sistematis dapat menghasilkan sebuah data yang dapat menghasilkan informasi dan pengetahuan yang berguna untuk mendukung proses pengambilan keputusan (vercellis, 2003).

(Turban, dkk, 2007) *Business Intelligence (BI)* terbagi kedalam lima jenis kategori yaitu:

1. *Enterprise Reporting*, digunakan untuk menghasilkan laporan-laporan statis yang dapat didistribusikan kepada banyak orang. Jenis laporan ini sangat sesuai untuk membuat laporan operasional dan *dashboard*.
2. *Ad Hoc Query and Analysis*, digunakan untuk memberikan akses kepada pengguna agar dapat melakukan *query* pada *database*, dan melakukan proses penggalian informasi sampai pada tingkat paling dasar dari sebuah informasi transaksional.
3. *Cube Analysis*, digunakan untuk menyediakan layanan analisis *OLTP multidimensional*, laporan hanya ditujukan kepada kalangan terbatas.
4. *Statistical Analysis and Data Mining*, digunakan untuk melakukan analisis prediksi atau menentukan korelasi diantara dua matrik.
5. *Delivery Report and Alert*, digunakan untuk melakukan pengiriman laporan secara lengkap atau memberikan peringatan kepada seluruh pengguna.

2.7 Online Transaction Processing (OLTP)

OLTP adalah sebuah sistem yang dirancang untuk dapat melakukan pemrosesan transaksi yang secara umum dapat melakukan perubahan kecil pada data operasional sebuah organisasi, yang bertujuan untuk menangani operasi sehari-hari (Connolly dan Begg, 2005).

2.8 Data Warehouse

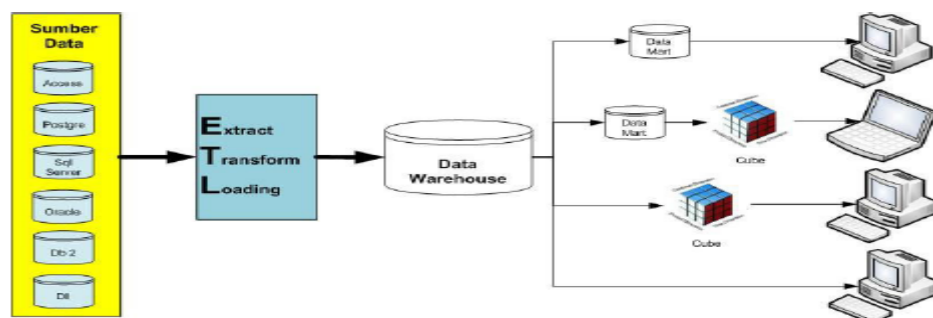
Data warehouse adalah sekumpulan *database* yang berasal dari berbagai sumber yang saling terintegrasi, berorientasi subjek yang dirancang untuk mendukung DSS, data warehouse berisi data perusahaan (Inmon, 2002).

Menurut (kimball, ross, 2002) *Data warehouse* adalah sebuah sarana yang dapat meningkatkan kualitas pengambilan keputusan, *data warehouse* berisi data yang dapat mendukung proses DSS.

Menurut (O'Brien, James A, 2007), *Data warehouse* merupakan sumber utama data dari *database* lain yang telah dibersihkan, diubah, dan di kategorikan untuk analisis bisnis dan mendukung proses DSS.

Dari keterangan di atas dapat disimpulkan bahwa *data warehouse* adalah sebuah wadah yang dapat menampung data-data yang dibutuhkan untuk melakukan analisis suatu kondisi dalam perusahaan yang akan menghasilkan sebuah informasi yang bermanfaat yang dapat mendukung proses DSS. Data yang digunakan dalam data warehouse dapat berasal dari data internal dan eksternal.

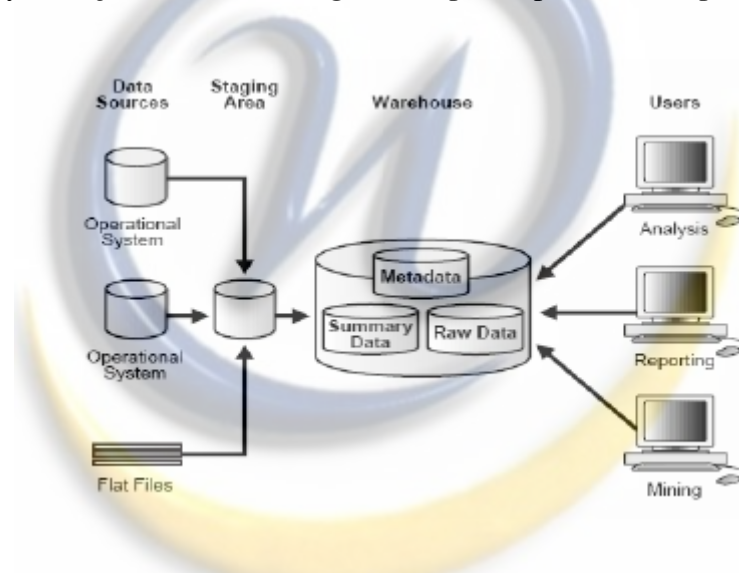
2.8.1 Arsitektur Data Warehouse



Gambar 2.1 Arsitektur Data Warehouse

Arsitektur *data warehouse* Pada gambar 2.1 terdiri dari sumber data, *ETL*, *data mart* dan *cube*. Sumber data adalah data operasional yang disimpan dalam *database*, yang selanjutnya akan dilakukan proses *ETL* dan diintegrasikan kedalam *data warehouse*. Sedangkan *data mart* dan *cube* berisi data-data yang mendukung fungsi bisnis.

Data mart adalah *sub-set* dari *data warehouse* yang umumnya terdiri dari sebuah subjek tunggal (inmon, 2005). Jadi, *data mart* merupakan serangkaian data yang hanya menjelaskan suatu fungsi dari operasi pada sebuah perusahaan.



Gambar 2.2 Arsitektur *Data Warehouse* dengan *Staging Area* (Paul Lane, 2002)

Pada gambar 2.2 terlihat proses pengolahan data operasional sebelum dimasukkan ke dalam *data warehouse* melalui *staging area* terlebih dahulu. *Staging area* digunakan untuk memudahkan dalam melakukan integrasi dan pembersihan data sehingga dapat menghasilkan data yang berkualitas. Karena di dalam *staging area* terdapat proses untuk melakukan penggabungan data, pembersihan, dan standarisasi data.

2.8.2 Pemodelan Dimensional

(Connolly dan Begg, 2002) Pemodelan dimensional adalah sebuah teknik perancangan logika yang bertujuan untuk merepresentasikan data ke dalam sebuah standar bentuk intuiting yang dapat diakses dengan performa yang tinggi. Setiap model dimensional terdiri dari tabel fakta dan sekumpulan tabel yang lebih kecil yang disebut tabel dimensi.

2.8.3 Skema Bintang (*Star Schema*)

Struktur skema bintang adalah suatu struktur yang dapat dengan mudah dipahami dan digunakan oleh pengguna. Dalam skema bintang terdapat dua jenis tabel, yaitu tabel fakta dan tabel dimensi yang berisi referensi data.

Tabel fakta pada skema bintang dihasilkan dari *event-event* yang terjadi dimasa lampau, fakta yang paling berguna dalam tabel fakta adalah perhitungan *numeric*, atau fakta yang terjadi dari setiap *record*. Sedangkan tabel dimensi mengandung informasi tekstual yang deskriptif.

2.8.4 Perancangan *Data Warehouse*

(Connolly dan Begg, 2002) Terdapat sembilan tahap metodologi dalam membangun *data warehouse* yang dikenal dengan *nine-step methodology* yang terdiri dari:

1. Memilih Proses (*Choosing The Process*)
2. Memilih Sumber (*Choosing The Grain*)

3. Mengidentifikasi dan Penyesuaian Dimensi (*Identifying and Conforming The Dimensions*)
4. Memilih Fakta (*Choosing The Fact*)
5. Menyimpan Perhitungan Awal dalam Tabel Fakta (*Storing Pre-Calculation in The Fact Table*)
6. Melihat Kembali Table Dimensi (*Rounding Out The Dimension Tables*)
7. Memilih Durasi *Database* (*Choosing The Duration of Database*)
8. Menelusuri Perubahan dari Dimensi secara Perlahan (*Tracking Slowly Changing Dimension*)
9. Memutuskan Prioritas *Query* dan Tipe *Query* (*Deciding The Query Priorities and The Query Models*)

2.8.5 Data Mart

(Inmon, 2005) yang dimaksud dengan *data mart* adalah struktur data per departement yang berasal dari *data warehouse* dilakukan denormalisasi berdasarkan kebutuhan informasi dari setiap departement.

(poniah, 2001) *data mart* adalah *data warehouse* kecil yang dirancang untuk tiap departement. Keuntungan dari *data mart* adalah biayanya yang murah dan waktu penerepannya relatif cepat.

(poniah, 2001) Adapun perbedaan antara *data mart* dan *data warehouse* adalah :

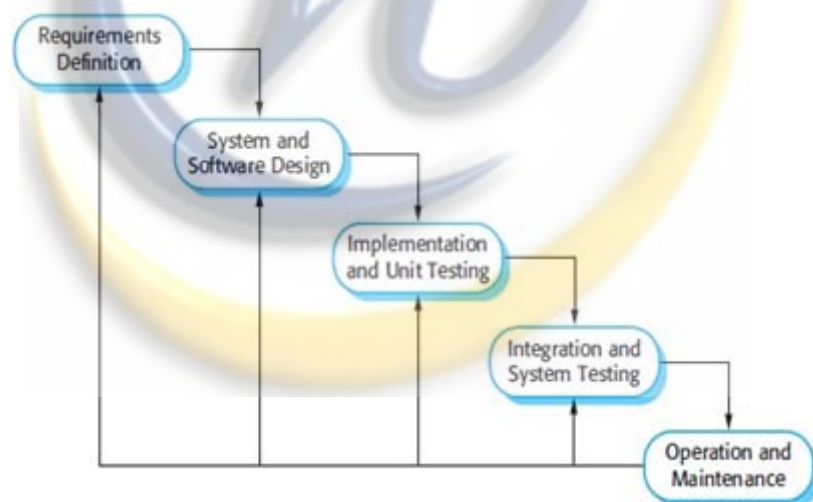
Tabel 2.1 perbedaan antara *data warehouse* dan *data mart*

Data Warehouse	Data Mart
Mencakup pada keseluruhan organisasi	Mencakup departement tertentu

Gabungan dari semua <i>data mart</i>	Sebuah proses bisnis
Data diterima dari <i>staging area</i>	<i>Star Join (fakta dan dimensi)</i>
Di <i>query</i> dari sumber tampilan	Teknologi maksimal untuk melakukan akses dan analisa data
Menampilkan informasi organisasi	Menampilkan informasi per departement

2.9 Waterfall Model

Menurut (I Sommerville, 2011) terdapat 5 tahapan pada *waterfall model*, yaitu *requirement analysis and definition*, *system and software design*, *implmentation and unit testing*, *integration and system testing*, dan *operation and maintenance*.



Gambar 2.3 Waterfall Model (I Sommerville, 2011)

1. *Requirement Analysis and Definition*

Merupakan tahapan penetapan fitur, kendala, serta tujuan sistem melalui proses wawancara yang dilakukan dengan pengguna sistem. Proses ini bertujuan untuk menentukan spesifikasi sistem yang akan dibangun.

2. *System and Software Design*

Tahapan ini merupakan tahapan pembentukan suatu arsitektur sistem berdasarkan persyaratan yang telah ditentukan pada proses sebelumnya. Tahapan ini berfungsi sebagai gambaran abstraksi dasar sistem perangkat lunak dan hubungan-hubungannya.

3. *Implementasi and System Testing*

Pada tahapan ini dilakukan proses implementasi desain yang telah dibuat pada tahapan sebelumnya kedalam bentuk program dengan menggunakan bahasa pemrograman. Setiap unit akan dilakukan pengujian untuk memastikan semua fungsi telah sesuai dengan yang diharapkan.

4. *Integration and System Testing*

Pada tahapan ini setiap unit program akan dilakukan pengintegrasian antara satu dengan yang lainnya, serta dilakukan proses ujicoba sebagai satu sistem yang utuh untuk memastikan sistem yang dibangun telah memenuhi persyaratan yang ada.

5. *Operation and Maintenance*

Tahapan ini merupakan tahapan terakhir dimana pada tahapan ini dilakukan proses penggunaan sistem. Pada tahapan ini dapat dilakukan perbaikan apabila ada *error* yang ditemukan.

2.10 PHP (*PHP Hypertext Preprocessor*)

(A Saputra, 2011) yang dimaksud dengan PHP adalah suatu bahasa pemrograman yang berfungsi untuk membangun suatu website dinamis. PHP bersifat *open source* sehingga dapat dipakan secara Cuma-Cuma dan mampu

berjalan pada sistem operasi windows maupun linux. Adapun keunggulan yang dimiliki PHP yaitu:

1. Mudah untuk dipelajari
2. PHP dapat diaplikasikan keberbagai *Operasi Sistem* dan dapat diakses oleh banyak web browser
3. Bersifat *Open Source*
4. Memiliki tingkat akses yang cepat
5. Didukung oleh beberapa web server, seperti apache, IIS, lighttpd, xitami.
6. Dapat mendukung database, baik yang gratis maupun yang berbayar, seperti MySQL, PostgreSQL, mSQL, Informix, SQL server, Oracle.

2.11 Structure Query Language (SQL)

(connolly dan begg, 2005) yang dimaksud dengan SQL adalah bahasa yang dirancang dengan penggunaan relasi untuk merubah masukan menjadi keluaran sesuai dengan kebutuhan. Sebagai sebuah bahasa standar internasional SQL menetapkan 2 komponen pokok yaitu:

1. *Data Definition Language (DDL)* untuk mendefinisikan struktur pada basis serta untuk melakukan akses kontrol terhadap data.
2. *Data Manipulating Language (DML)* untuk mengembalikan dan memperbrui data.

2.12 Query

Query adalah kemampuan untuk menampilkan suatu data dari dataase dimana data diambil dari tabel-tabel yang ada di dalam database sesuai dengan kebutuhan. Bahasa *query* merupakan bahasa khusus yang digunakan untuk

melakukan *query* pada basis data. Contoh penggunaan bahasa *query* adalah *SELECT * FROM SERVICE WHERE KD_TIME=2*. *Query* tersebut meminta untuk menampilkan semua *record* yang terdapat pada tabel *service* dengan *KD_TIME=2*.

2.13 High Chart

High chart adalah sebuah *charting library* yang ditulis dengan menggunakan bahasa pemrograman *javascript*. Highchart menawarkan grafik yang interaktif yang dapat digunakan pada aplikasi berbasis web. Saat ini *highchart* mendukung grafik tipe *line*, *spline*, *area*, *arespline*, *column*, *bar*, *pie*, dan *scatter*. *Highcart* dapat berjalan di hampir semua browser.

2.14 Pentaho

Pentaho adalah sebuah nama perusahaan dan juga nama dari produk yang dimiliki oleh perusahaan tersebut. Pentaho merupakan sebuah perusahaan yang khusus membuat software untuk keperluan *data warehouse*(DWH) dan *business intelligence*(BI). Produk-produk yang ditawarkan dalam dua bentuk edisi, yaitu *Enterprise Edition* (EE) dan *Community Edition* (CE).

Perbedaan produk EE dan CE terletak pada:

1. Produk EE memiliki fitur-fitur yang tidak disediakan pada produk CE.
2. Untuk menggunakan produk EE diharuskan membayar biaya berlangganan tahunan dan akan mendapatkan dukungan layanan langsung dari pentaho corporation.

3. Produk CE bersifat gratis selama digunakan sesuai dengan perjanjian *lisensi* yang telah disediakan, produk CE tidak mendapatkan dukungan dari pentaho corporation.

2.14.1 Pentaho Data Integraion(PDI)

Pada tahun 2001, matt casters membuat sebuah perangkat ETL yang diberi nama kettle yang merupakan singkatan dari KDE ETTL Environment, karena pada awalnya perangkat ini ditujukan untuk dapat dijalankan di atas K Desktop Environment(KDE) linux. ETTL pada kettle merupakan singkatan dari Extraction, Transportation, Transformation, dan Loading.

Karena kettle memiliki beragam keunggulan dibanding perangkat ETL *open source* lainnya, pada april 2006, kode sumber kettle resmi diambil alih oleh pentaho. Oleh pentaho, kettle diberinama pentaho data integration (PDI).

2.14.2 Step

Step merupakan blok bangunan inti dari *transformation*. Masing-masing step memiliki fungsi dan tugas tertentu, misalkan ada *step* yang berfungsi membaca file excel, ada step yang berfungsi untuk memilih *field*, ada step yang berfungsi untuk melakukan proses *insert/ update* data kedalam database, dan lain sebagainya.

Kettle memiliki ratusan step siap pakai, dalam spoon setiap step diawali oleh ikon-ikon unik. Antara step yang satu dengan step yang lainnya dihubungkan menggunakan transformation hop.