

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 *Data Warehouse***

Pengertian data warehouse menurut Inmon (2002, p31), “ *a data warehouse is a subject oriented, nonvolatile, time variant collection of data in support of management’s decisions* ” atau dapat diartikan “ *data warehouse* adalah koleksi data yang mempunyai sifat berorientasi subjek, terintegrasi , tidak mengalami perubahan dan mempunyai variasi waktu yang digunakan untuk mendukung proses pengambilan keputusan manajemen”.

Menurut Post (2002, p548) *data warehouse* adalah spesialisasi basis data yang dioptimasi untuk memenuhi permintaan mana jemen, data diekstrak dari sistem *online transaction processing* (OLTP), kemudian dibersihkan dan dioptimasisasi untuk pencarian dan analisis. Jadi dapat disimpulkan bahwa *data warehouse* adalah kumpulan data yang telah diringkas dan terintegrasi dari data operasional maupun data *external*, yang memiliki karakteristik *subject-oriented* , *integrated*, *nonvolatile* dan *time variant* yang berguna dalam pengambilan keputusan.

##### **2.1.1 Tujuan Perancangan *Data Warehouse***

*Data warehouse* yang digunakan selama ini memberikan kemudahan dan keuntungan karena *data warehouse* biasanya digunakan untuk melakukan empat tugas yang berbeda. Menurut Williams (1998, p533), keempat tugas *data warehouse* tersebut adalah sebagai berikut :

## 1. Pembuatan Laporan

Pembuatan laporan merupakan salah satu kegunaan *data warehouse* yang paling umum. Dengan menggunakan *query* sederhana dalam *data warehouse*, dapat dihasilkan informasi per tahun, per semester, per bulan, dan bahkan per hari.

## 2. *On-Line Analytical Processing (OLAP)*

*Data warehouse* digunakan dalam melakukan analisis bisnis untuk mengetahui kecenderungan pasar dan faktor-faktor penyebabnya, karena dengan adanya *data warehouse*, semua informasi baik detail maupun hasil *summary* yang dibutuhkan dalam proses analisa mudah didapat. Dalam hal ini *data warehouse* merupakan *tools* handal untuk analisa data yang kompleks. *OLAP* mendayagunakan konsep data multidimensi dan memungkinkan pemakai untuk menganalisa data sampai mendetail, tanpa mengetikkan satu pun perintah *SQL*. Hal ini dimungkinkan karena pada konsep data multidimensi, data berupa fakta yang sama bisa dilihat dengan menggunakan dimensi yang berbeda. Fasilitas lain yang ada pada *tools* perangkat lunak *OLAP* adalah *drill-down* dan *roll-up*. *Drill-Down* adalah kemampuan untuk melihat detail dari suatu informasi yang ditampilkan sedangkan *roll-up* adalah kebalikan dari *drill-down*.

## 3. Data mining

Data mining adalah proses untuk mencari informasi dan pengetahuan baru dengan cara menggali (mining) data yang berjumlah banyak pada *data warehouse*, dengan menggunakan kecerdasan buatan (*Artificial Intelligence*), statistik, dan

matematika. Data mining merupakan teknologi yang diharapkan bisa menjembatani komunikasi antara data dan pemakainya.



#### 4. Proses Informasi Eksekutif

*Data warehouse* digunakan untuk mencapai ringkasan informasi yang penting dengan tujuan membuat keputusan bisnis, tanpa harus menjelajahi keseluruhan data. Dengan menggunakan *data warehouse*, segala laporan telah diringkas dan dapat pula diketahui rinciannya secara lengkap. Hal ini akan mempermudah proses pengambilan keputusan. Informasi dan data pada laporan *data warehouse* menjadi target informatif bagi *user*, dimana *user* disini adalah pihak *eksekutif*.

## 2.2 Karakteristik *Data Warehouse*

### 1. Berorientasi Subjek

*Data warehouse* terorganisasi di seputar subjek kunci (atau entitas-entitas peringkat tinggi) dalam perusahaan, *Data warehouse* adalah tempat penyimpanan berdasarkan subyek bukan berdasarkan aplikasi.

### 2. Terintegrasi

Data yang tersimpan dalam *data warehouse* didefinisikan menggunakan konversi penamaan yang konsisten, format-format, struktur terkodekan, serta karakteristik-karakteristik yang berhubungan, Sumber data yang ada dalam *data warehouse* tidak hanya berasal dari *database* operasional (*internal source*) tetapi juga berasal dari data diluar sistem (*external source*). Data pada sumber berbeda dapat di-*encode* dengan cara yang berbeda.

### 3. Memiliki Dimensi Waktu

Data yang tersimpan dalam *data warehouse* mengandung dimensi waktu yang mungkin digunakan sebagai rekaman bisnis untuk tiap waktu tertentu, *Data warehouse* menyimpan sejarah (*historical data*). Bandingkan dengan kebutuhan sistem operasional yang hampir semuanya adalah data mutakhir, waktu merupakan tipe atau bagian data yang sangat penting didalam *data warehouse*.

### 4. Non Volatile

Data yang tersimpan dalam *data warehouse* diambil dari *system* operasional yang sedang berjalan, tetapi tidak dapat diperbaharui (*di-update*) oleh pengguna (bersifat 'hanya-baca), Sekali masuk kedalam *data warehouse*, data-data, terutama data tipe transaksi, tidak akan pernah di *update* atau dihapus (*delete*) Terlihat, bahwa keempat karakteristik ini saling terkait kesemuanya harus diimplementasikan agar suatu *data warehouse* bisa efektif memiliki data untuk mendukung pengambilan-keputusan. Dan, implementasi keempat karakteristik ini membutuhkan struktur data dari *data warehouse* yang berbeda dengan database sistem operasional.

## 2.3 Struktur Data Warehouse

Sebuah *data warehouse* memiliki beberapa struktur, seperti :

### 1. Physical Data warehouse

Tempat dimana semua data untuk *data warehouse* disimpan bersama metadata dan proses logis untuk *scrubbing* (menghapus), *organizing* (mengatur), *packaging* (mengumpulkan) dan proses dari detail data.

## 2. Logical Data warehouse

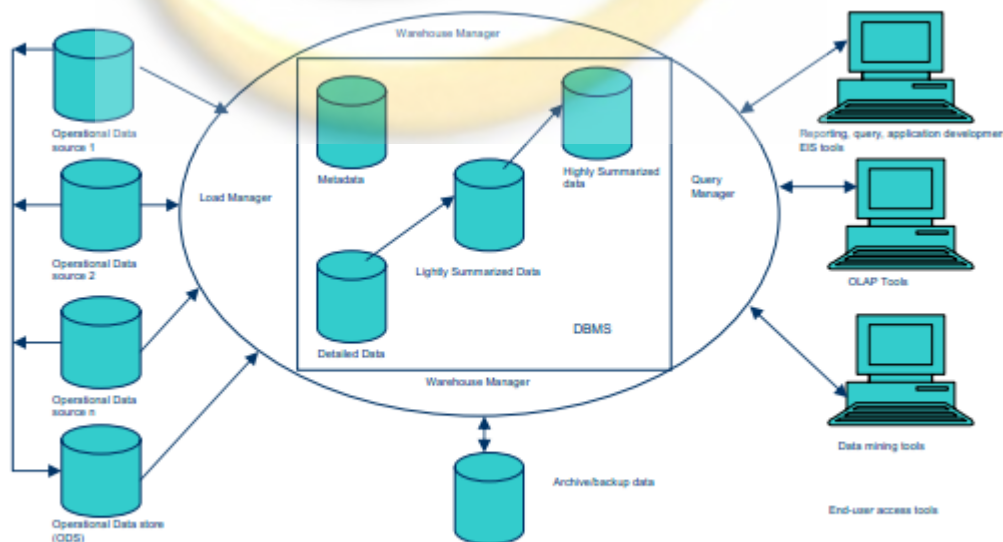
Berisikan metadata termasuk *enterprise rules* dan proses logis untuk *scrubbing*

(menghapus), *organizing* (mengatur), *packaging* (mengumpulkan) dan proses data. Tetapi tidak berisikan data yang aktual. Disamping itu juga berisikan informasi yang diperlukan untuk mengakses data dimana saja.

## 3. Data Mart

*Data mart* adalah suatu bagian dari *data warehouse* yang dapat mendukung pembuatan laporan dan analisa data pada suatu unit, bagian atau operasi pada perusahaan. Sebagai bagian dari proses pengembangan *data warehouse* yang selalu berulang, sebuah perusahaan perlu membangun sebuah rangkaian *physical data mart* dan menghubungkannya melalui *enterprise-wide logical data warehouse* atau dimasukkan dari *single physical data warehouse*.

### 2.4 Arsitektur Data Warehouse



Gambar 2.1 Arsitektur Data Warehouse (Connolly, 2005, p1157)

Menurut Connolly dan Begg (2002, p1052), komponen utama *data warehouse* , antara lain :

#### 1. Data Operasional

Data operasional adalah data yang digunakan untuk mendukung proses bisnis sehari-hari.

#### 2. *Operational Data Store (ODS)*

Operational data store adalah tempat penyimpanan data operasional yang bersifat current dan terintegrasi yang digunakan untuk analisis. Atau dengan kata lain, *ODS* mendukung proses transaksi operasional maupun proses analisis. Dengan adanya *ODS* maka pembangunan data warehouse menjadi lebih mudah karena *ODS* dapat menyediakan data yang telah diekstrak dari sumber dan telah dibersihkan sehingga proses pengintegrasian dan restrukturisasi data untuk data warehouse menjadi lebih sederhana.

#### 3. *Load Manager*

Disebut juga komponen *front end* menangani semua operasi yang berhubungan dengan fungsi *extract data* (mengambil data) dan fungsi *loading data* (menaruh data) ke dalam *data warehouse*.

#### 4. *Warehouse Manager*

*Warehouse manager* menangani semua operasi yang berhubungan dengan management data dalam *data warehouse* . Operasi-operasi yang dijalankan oleh *warehouse manager* mencakup :

- a. Analisis data untuk menjaga konsistensi data.

- b. Melakukan transformasi dan penggabungan sumber data dari tempat penyimpanan sementara ke dalam tabel-tabel *data warehouse*.
- c. Melakukan denormalisasi.
- d. Melakukan agregasi.
- e. Menyimpan (*archive*) dan *back-up data*.

#### 5. *Query Manager*

*Query manager* (disebut juga komponen *backend*) menangani semua operasi yang berhubungan dengan management permintaan *user* (*user queries*). Operasi yang dijalankan oleh *query manager* meliputi kegiatan mengarahkan permintaan ke tabel-tabel data yang tepat dan melakukan penjadwalan eksekusi terhadap permintaan.

#### 6. *Detailed Data*

Dalam *data warehouse*, area ini adalah tempat penyimpanan semua *detailed data* dalam skema basis data. *Detailed data* dibagi menjadi 2, yaitu *current detail data* (tempat penyimpanan semua *detailed data* yang bersifat *current*) dan *old detailed data* (tempat penyimpanan semua *detailed data* yang bersifat *old*).

#### 7. *Lightly and Highly Summarized Data*

Area ini adalah tempat penyimpanan sementara *data predefined* isi yang ringkas secara *light* dan *high* (*predefined lightly and highly summarized*) yang dihasilkan oleh *warehouse manager*. Tujuan dari ringkasan informasi ini adalah untuk mempercepat tanggapan terhadap permintaan *user*. Ringkasan data di-*update* secara berkala seiring dengan bertambahnya data dalam *data warehouse*.



## 8. *Archive / Backup Data*

Dalam *data warehouse*, area ini digunakan untuk menyimpan *detailed data* dan data yang telah diringkas. Tujuannya adalah untuk penyimpanan (*archiving*) dan *backup*. Data kemudian *ditransfer* ke media penyimpanan seperti *magnetic tape* atau *optical disk*.

## 9. Metadata

Digunakan untuk menyimpan semua definisi metadata (keterangan tentang data) yang digunakan dalam seluruh proses *warehouse*.

Metadata digunakan untuk berbagai tujuan, antara lain :

- a. proses *extracting* dan *loading*
- b. proses *warehouse management*
- c. sebagian proses *query management*

## 10. *End-User Access Toolss*

*End-user access toolss* adalah *tools* yang memanfaatkan kegunaan dari *data warehouse*. Kegunaan *data warehouse* tersebut, antara lain untuk pembuatan laporan, *OLAP*, *data mining* dan proses informasi eksekutif.

### 2.4.1 **Infrastruktur *Data Warehouse***

Infrastruktur *data warehouse* terdiri dari *software*, *hardware*, pelatihan-pelatihan dan komponen-komponen lainnya yang me mberikan dukungan yang dibutuhkan untuk mengimplementasikan arsitektur *data warehouse* Poe (1998, p43). Salah satu instrumen yang mempengaruhi keberhasilan pengembangan *data warehouse* adalah pengidentifikasian arsitektur mana yang terbaik dan infrastruktur yang dibutuhkan. Arsitektur dan infrastruktur sangat erat hubungannya. Arsitektur

yang sama mungkin akan membutuhkan infrastruktur yang berbeda, tergantung pada lingkungan perusahaan ataupun organisasi.

#### 2.4.2 Metodologi Perancangan *Data Warehouse*

Berdasarkan kutipan Connolly dan Begg (2002, p1083) metodologi yang dikemukakan oleh Kimball dalam membangun *data warehouse* ada 9 tahapan, dikenal dengan Nine-step Methodology.

##### 1. Memilih proses

Pilihlah subjek dari permasalahan yang sedang dihadapi, kemudian identifikasi proses bisnisnya. Data mart adalah bagian dari *data warehouse* yang pembuatan laporan dan analisis data pada suatu unit, bagian atau operasi pada perusahaan.

##### 2. Memilih grain

Tentukan tabel fakta dan identifikasi dimensi. Tabel fakta merupakan tabel yang mengandung angka dan *data history* dimana *key* yang dihasilkan sangat banyak karena merupakan kumpulan – kumpulan *foreign key* dan *primary key* yang ada pada masing – masing tabel dimensi yang berhubungan. Sedangkan tabel dimensi adalah tabel yang berisi kategori dengan ringkasan data detail yang dapat dilaporkan, seperti laporan keuntungan pada tabel fakta, sebagai dimensi waktu (perbulan, persemester, pertahun).

##### 3. Identifikasi dan penyesuaian dimensi

Identifikasi dimensi dalam detail yang secukupnya untuk mendeskripsikan sesuatu. Ketika tabel dimensi ada pada dua atau lebih data mart, maka tabel dimensi tersebut

harus mempunyai dimensi yang sama atau salah satu merupakan subset dari yang lainnya. Apabila suatu tabel dimensi digunakan lebih dari satu data mart, maka dimensinya harus disesuaikan.

#### 4. Memilih fakta

Tentukan fakta-fakta dari tabel fakta yang akan digunakan pada data mart.

Fakta – fakta tersebut harus numerik dan dapat ditambah.

#### 5. Menyimpan pre-kalkulasi pada tabel fakta

Setelah fakta-fakta dipilih maka lakukan pengkajian ulang untuk menentukan apakah ada fakta-fakta yang dapat diterapkan pre-kalkulasi (kalkulasi awal) dan lakukan penyimpanan pada tabel fakta.

#### 6. Melengkapi tabel dimensi

Dalam langkah ini, kita kembali pada *dimension table* dan menambahkan gambaran teks terhadap dimensi yang memungkinkan. Gambaran *teks* harus mudah digunakan dan dimengerti oleh *user*. Kegunaan suatu data mart ditentukan oleh lingkup dan atribut tabel dimensi.

#### 7. Memilih durasi dari *database*

Tentukan waktu dari pembatasan data yang diambil dan dipindahkan ke dalam tabel fakta. Seperti data perusahaan tiga tahun lalu atau lebih diambil dan dimasukkan dalam tabel fakta.

#### 8. Melacak perubahan dari dimensi secara perlahan

Amati perubahan dari dimensi pada *dimension table*. Ada tiga tipe dasar dari perubahan dimensi yang perlahan, yaitu :

- a. Perubahan atribut dimensi ditulis ulang (*over write*).
  - b. Perubahan atribut dimensi mengakibatkan pembuatan suatu dimensi baru.
  - c. Perubahan atribut dimensi mengakibatkan sebuah atribut alternatif dibuat, jadi antar atribut yang lama dan yang baru diakses secara bersama – sama.
9. Memutuskan prioritas dan *mode query*

Pertimbangkan pengaruh dari perancangan fisik, seperti keberadaan dari ringkasan (*summaries*) dan penjumlahan (*agregate*). Selain itu, masalah administrasi, *backup data*, *recovery data*, kinerja *indeks* dan keamanan juga merupakan faktor yang harus diperhatikan.

### 2.4.3 Konsep Pemodelan *Data Warehouse*

#### 1. *Entity Relationship Modelling (ER Modelling)*

Menurut Thomas Connolly dan Carolyn Begg *ERModelling* adalah sebuah pendekatan *top-down* untuk perancangan basis data yang mulai dengan mengidentifikasi data yang penting disebut *entites* dan *relationship* antar data harus direpresentasikan dalam model.

Menurut Ms. Alpa R. Patel dan Jayesh M. Patel model ER diwakili oleh diagram ER, yang menggunakan tiga simbol grafis dasar untuk konsep data: entitas, relasi, dan atribut.

#### - Entitas

Entitas didefinisikan sebagai orang, tempat, benda, atau peristiwa yang menarik bagi bisnis atau organisasi. Entitas merupakan kelas obyek, dimana merupakan hal-hal di dunia nyata yang dapat diamati dan digolongkan oleh sifat dan karakteristik

### - Relasi

Sebuah relasi dihubungkan dengan garis yang ditarik antara entitas. Ini menggambarkan interaksi struktural dan asosiasi di antara entitas dan model. Sebuah hubungan gramatikal ditunjuk oleh kata kerja, seperti memiliki, milik, dan memiliki. Hubungan antara dua entitas dapat didefinisikan dalam hal kardinalitas. Ini adalah jumlah maksimum contoh satu kesatuan yang berhubungan dengan satu kejadian di tabel lain dan sebaliknya. Kardinalitas yang mungkin adalah: satu-ke-satu(1:1), satu-ke-banyak (1:M), dan banyak-ke-banyak (M:M).

### - Atribut

Atribut menggambarkan karakteristik properti dari entitas. Untuk klarifikasi, penamaan atribut konvensi adalah hal yang sangat penting. Nama atribut harus unik dalam suatu entitas dan harus cukup jelas. Ketika sebuah *instance* tidak memiliki nilai atribut, kardinalitas minimum atribut adalah nol, yang berarti baik *nullable* atau opsional. Dalam pemodelan ER, jika kardinalitas maksimum atribut lebih dari 1, pemodel akan mencoba untuk menormalisasikan entitas dan akhirnya meningkatkan entitas lainnya. Oleh karena itu, biasanya kardinalitas maksimum atribut adalah 1.

## 2. Dimensionality Modelling

Menurut Thomas Connolly dan Carolyn Begg *dimensionality modeling* adalah teknik *logical design* yang bertujuan untuk menyajikan data standar, bentuk intuitif yang memungkinkan untuk mengakses *high performance*.

*Dimensionality modeling* menggunakan konsep dari *Entity Relationship (ER)* dengan beberapa batasan yang penting. Setiap Dimensional model terdiri dari satu atau lebih tabel yang memiliki banyak *Primary key (composite Primary*

*key*), yang disebut tabel dimensi (*dimensional table*). Setiap tabel dimensi memiliki satu buah (*non-composite*) *primary key* yang berhubungan dengan salah satu *primarykey* di tabel fakta. Karakteristik ini disebut skema bintang (*star schema*) atau *starjoin*.

## 2.5 Visualisasi Data

### A. Tabel

Tabel adalah kumpulan angka-angka yang disajikan dalam baris dan kolom menurut kategori-kategori tertentu sehingga dapat memudahkan dalam pembuatan analisis data. Tabel digunakan untuk menampilkan angka, tingkatan, proporsi, dan persentase kumulatif. Penyajian dengan tabel dapat memberikan angka-angka yang lebih teliti.

### B. Bar Chart

*Bar chart* biasanya digunakan untuk menampilkan data yang berkategori, data kontinu, dan data diskontinu. *Bar chart* dapat disajikan secara vertikal atau horizontal. Skala pengukurannya ialah nominal atau ordinal.

### C. Line Chart

*Line chart* ialah tipe grafik yang memvisualisasikan trend data dalam kurun waktu tertentu. *Line chart* biasanya menyajikan informasi dalam rangkaian titik data yang dihubungkan dengan segmen garis lurus.

#### ***D. Pie Chart***

*Pie chart* ialah tipe grafik yang memvisualisasikan data dalam bentuk frekuensi atau kategori. *Pie chart* ialah sebuah cara yang paling ilustratif untuk menampilkan kuantitas sebagai persentase dari total data. Total area dari sebuah *pie chart* merepresentasikan 100% dari kuantitas (jumlah dari nilai variabel pada seluruh kategori).

