

BAB II TINJAUAN PUSTAKA

2.1 Database

Database adalah kumpulan informasi yang disimpan di dalam komputer secara sistematis sehingga dapat diperiksa menggunakan suatu program komputer untuk memperoleh informasi dari basis data tersebut. *Database* adalah representasi kumpulan fakta yang saling berhubungan disimpan secara bersama sedemikian rupa dan tanpa pengulangan yang tidak perlu, untuk memenuhi berbagai kebutuhan. *Database* merupakan sekumpulan informasi yang saling berkaitan pada suatu subjek tertentu pada tujuan tertentu pula [1].

Konsep dasar dari basis data adalah kumpulan dari catatan-catatan, atau potongan dari pengetahuan. Sebuah basis data memiliki penjelasan terstruktur dari jenis fakta yang tersimpan di dalamnya: penjelasan ini disebut skema. Skema menggambarkan objek yang diwakili suatu basis data, dan hubungan di antara objek tersebut. Model yang umum digunakan sekarang adalah model relasional, yang menurut istilah layman mewakili semua informasi dalam bentuk tabel-tabel yang saling berhubungan di mana setiap tabel terdiri dari baris dan kolom.

2.2 Tracer study

Tracer study merupakan sebuah penelitian yang dilakukan oleh perguruan tinggi, sebagai bentuk pengawasan dalam hal pencarian kerja, situasi kerja, dan pemanfaatan pemerolehan kompetensi yang telah dihasilkan oleh perguruan tinggi [2]. Pengertian *tracer study* menurut [3] merupakan studi yang bertujuan untuk memperoleh informasi mengenai lulusan yang sudah bekerja dan belum bekerja. Selain itu *tracer study* juga bertujuan untuk mengetahui hasil pendidikan dalam bentuk penguasaan dan pemerolehan kompetensi lulusan yang diimplementasikan kedalam di dunia kerja serta transisi dari dunia pendidikan tinggi ke dunia usaha dan industri. Manfaat yang dapat diperoleh dari pelaksanaan *tracer study*, yaitu diantaranya :

1. Dapat melihat kepuasan *stakeholder*, sebagaimana terlihat dari hasil proses *learning experience* pada setiap alumni, yang nantinya akan dijadikan tolak ukur evaluasi kinerja perguruan tinggi

2. Dapat membangun hubungan antara alumni dengan mahasiswa,
3. Dapat mengetahui persebaran lulusan alumni perguruan tinggi.

2.3 Data warehouse

Menurut [4], *Data warehouse* merupakan suatu ringkasan dari kumpulan data-data logic berdasarkan dari *Database* operasional. Konsep dari *data warehouse* adalah suatu kombinasi teknologi yang memfasilitasi organisasi untuk mengelola dan memelihara data historis yang diperoleh dari sistem atau aplikasi operasional. *Data warehouse* juga menurut pendapat [5] adalah sekumpulan data yang sudah terintegrasi, berorientasikan subjek *database* yang dirancang untuk mendukung fungsi dari *Decision Support*. Fungsi dari pembuatan *data warehouse* sebagai media penyimpanan data untuk mendukung sistem pengambilan keputusan yang dibutuhkan oleh *top level* management dengan bantuan visualisasi data berupa dashboard yang berisi sebuah *report* yang fleksibel. *Data warehouse* juga dapat memfasilitasi pekerjaan utama intelijen bisnis - pelaporan dan analisis. Dalam *data warehouse* terdapat tempat penyimpanan data terpusat, dimana data dari basis data operasional dan sumber lainnya dapat diintegrasikan, dibersihkan dan diarsipkan untuk mendukung pengambilan keputusan. [6]

2.3.1 Karakteristik *Data warehouse*

Menurut [7] terdapat empat karakteristik dalam sebuah *data warehouse*, yaitu :

1) *Subject Oriented*

Suatu *data warehouse* dirancang untuk memenuhi kebutuhan analisis data berdasarkan subjek tertentu.

2) *Integrated*

Suatu *data warehouse* harus mampu mengintegrasikan data dari berbagai sumber sistem operasional yang beragam. Oleh karena itu sebelum data dari berbagai sumber yang berbeda ini disimpan ke dalam *data warehouse* maka harus menghilangkan inkonsistensi data yang ada

3) *Non-volatile*

Data transaksi yang sudah dimasukkan ke dalam *data warehouse*, tidak akan pernah diupdate atau dihapus. Data yang tersimpan dalam *data warehouse* tidak dimaksudkan untuk aktifitas harian. Data dari sistem

operasional dipindahkan ke dalam *data warehouse* dalam interval waktu tertentu disesuaikan dengan kebutuhan bisnis. :

4) *Time Variant*

Data warehouse menyimpan sejarah (historical data). Waktu merupakan tipe atau bagian data yang sangat penting di dalam *data warehouse*. Di dalam *data warehouse* sering disimpan macam-macam waktu, seperti waktu terjadinya transaksi, waktu suatu transaksi dirubah atau waktu transaksi dibatalkan, kapan suatu transaksi bisa efektif, kapan suatu transaksi masuk ke komputer, dan kapan suatu transaksi masuk ke *data warehouse*. *Data warehouse* dikatakan bernilai bagus jika *data warehouse* menyimpan sejarah

2.4 Model Data

Pembuatan *data warehouse* didasarkan pada model data multidimensi. Model ini menampilkan data dalam bentuk kubus. Model data multidimensi terdiri dari dimensi dan fakta [8]. Skema basis data berisi kumpulan entitas dan hubungan antarentitas. Sebuah *data warehouse* memerlukan skema yang ringkas dan berorientasi subjek yang dapat digunakan dalam analisis data *on-line*. Tipe-tipe skema model data multidimensi adalah [9]

1) Skema Bintang

Skema bintang adalah skema *data warehouse* yang paling sederhana. Skema ini disebut skema bintang karena hubungan antar tabel dimensi dan tabel fakta menyerupai bintang, dimana satu tabel fakta dihubungkan dengan beberapa tabel dimensi. Titik tengah skema bintang adalah satu tabel fakta besar dan sudut-sudutnya adalah tabel-tabel dimensi. Keuntungan dari penggunaan skema bintang :

- a. Respon data lebih cepat daripada perancangan *database* operasional.
- b. Mempermudah dalam hal modifikasi atau pengembangan *data warehouse* yang terus menerus.
- c. *End-user* dapat menyesuaikan cara berpikir dan menggunakan data.
- d. Menyederhanakan pemahaman dan penelusuran metadata bagi pemakai dan pengembang.

2) Skema Snowflake

Skema snowflake adalah variasi dari skema bintang dimana beberapa tabel dimensi dinormalisasi. Sehingga dihasilkan beberapa tabel tambahan. Keuntungan yang didapat dengan menggunakan skema ini adalah penghematan memory, tapi waktu yang dibutuhkan untuk pemrosesan query menjadi lebih lama.

3) Skema Galaksi

Pada skema galaksi beberapa tabel fakta berbagi tabel dimensi. Keuntungan menggunakan skema ini adalah menghemat memori dan mengurangi kesalahan yang mungkin terjadi [10].

2.5 ETL (*Extraction, Transformation, Load*)

Menurut [11], proses ETL yang terdiri dari tiga proses yaitu *Extraction, Transformation, Load* merupakan proses pengambilan data dari sumber data dan mentransfer ke *data warehouse*. Pada proses tersebut, akan dilakukan proses pembersihan data agar mendapatkan data yang berkualitas.

Extract, transform, load (ETL) adalah proses yang sangat penting dalam *data warehouse*, data ETL dapat dimasukkan dari sistem operasi ke dalam *data warehouse*. Tujuan dari ETL adalah untuk mengumpulkan, menyaring, memproses, dan menggabungkan data dari berbagai sumber dan menyimpannya dalam sebuah gudang data.

Ekstraksi atau *extract* adalah proses memilih dan mengambil data dari satu atau lebih sumber dan membaca/mengakses data yang dipilih. Ini dapat menggunakan kueri atau implementasi ETL. disarankan agar pengguna menentukan persyaratan sumber data yang diperlukan sebelum melakukan ekstraksi yang lebih mudah.

Pada proses transform data yang telah diambil pada proses *extract* akan dibersihkan dan mengubah data dari bentuk asli menjadi bentuk yang sesuai dengan kebutuhan *data warehouse*. Kendala yang biasanya terjadi pada proses transform adalah sulitnya menggabungkan data dari beberapa sistem yang harus dibersihkan sehingga data bersifat konsisten.

Load adalah proses terakhir yang berfungsi untuk memasukkan data ke dalam target akhir, yaitu ke dalam *data warehouse*. Pada proses ini akan mengubah data kedalam bentuk *Dimensional Data Store* agar format data cocok untuk diterapkan pada proses analisis dan telah terintegrasi dengan beberapa sumber data.

Dalam ETL terdapat dua tabel yaitu tabel fakta dan tabel dimensi. Tabel fakta merupakan tabel yang mewakili subjek yang akan dianalisis dimana data dalam tabel berkaitan langsung dengan proses bisnis. Sedangkan tabel dimensi adalah tabel yang mewakili sumbu analisis yang digunakan untuk membatasi ruang lingkup kueri dan data yang ada dalam tabel dimensi jarang mengalami perubahan [12].

Sebagian besar data dalam sistem sumber sangat kompleks sehingga sangat sulit untuk mengidentifikasi data yang relevan. Tugas merancang dan membuat proses ekstraksi sangat memakan waktu. Data mentah dari sistem sumber biasanya dapat disimpan langsung di staging area dengan rekonstruksi minimal untuk menjaga keandalan data. Ada tiga metode ekstraksi yang umum digunakan yaitu *extract, transform and load*.



Gambar 2. 1 Skema Arsitektur ETL

2.6 OLAP (Online Analytical Processing)

OLAP (*Online Analytical Processing*) merupakan bentuk pemikiran bisnis yang merangkum hubungan antara pelaporan dan penggalian data. OLAP secara spesifik dapat diterapkan untuk memproses laporan bisnis terkait penjualan, pemasaran, penganggaran dan peramalan. Dengan bantuan OLAP, *query* pada *Database* dapat dilakukan dengan cepat, mudah dan efisien. Implementasi teknologi OLAP dapat menghasilkan laporan analisa data penjualan sebagai dasar pengambilan keputusan dan penentuan strategi perusahaan. Selain itu, implementasi *data warehouse* dan laporan dari proses OLAP, serta ditunjang dengan aplikasi dashboard mampu

menyuguhkan laporan dan status penjualan secara lebih sederhana sehingga memberikan pengguna kemudahan dalam membaca laporan guna mengambil suatu keputusan .

OLAP merupakan kumpulan aturan yang menyediakan sebuah kerangka dimensional untuk mendukung pengambilan keputusan. OLAP juga merupakan sebuah pendekatan secara cepat menyediakan jawaban-jawaban terhadap *query* analitik yang multi dimensi. OLAP adalah bagian dari kategori yang lebih global dari pemikiran bisnis, yang juga merangkum hubungan antara pelaporan dan penggalian data. Di dalam inti sebaran sistem OLAP merupakan konsep dari sebuah kubus OLAP (disebut juga sebagai kubus multidimensi atau *hypercube*) yang terdiri dari *numeric fact* yang disebut ukuran dan dikategorikan sebagai dimensi [13].

2.7 OLTP (Online Transaction Processing)

OLTP adalah *database* yang mengolah data atau mengakses data secara real time dan yang mendukung proses transaksi. Pada biasanya digunakan untuk aplikasi yang tergolong kritis. Menurut Rainer and Turban : OLTP (*Online Transaction Processing*) adalah pengelolaan data transaksi secara online langsung setelah transaksi itu terjadi. Menurut Connolly and Begg : OLTP adalah sebuah sistem yang dirancang untuk mengelola pemrosesan data transaksi tingkat tinggi dengan data yang membuat pengaruh kecil pada data operasional organisasi sehari-hari [14]

2.8 Pentaho Kettle

Menurut Phi-Integration.com, Pentaho merupakan sebuah perusahaan yang menawarkan produk business intelligence BI yang menyediakan data integrasi, pelayanan OLAP, pelaporan atau reporting, dashboarding, data mining dan kemampuan ETL atau extract transform load pada Kettle. Kettle merupakan aplikasi Extract, Transformation dan Load atau sering disingkat dengan ETL. Aplikasi Kettle merupakan salah satu bagian dari aplikasi BI Pentaho, yang dikenal dengan Pentaho Data Integration PDI.