

BAB II

TINJAUAN PUSTAKA

2.1 Data dan Informasi

2.1.1 Pengertian Data

Menurut Kadir (1998), data adalah fakta mengenai objek, orang, dan lain-lain. Data dinyatakan dengan nilai (angka, deretan karakter, atau simbol).

Menurut Hariyanto (2004), data adalah fakta tentang sesuatu di dunia nyata yang dapat direkam dan disimpan pada media komputer. Data dapat diperluas dengan adanya basis data. Dengan demikian, pengertian data dapat diperluas menjadi fakta, teks, grafik, suara, serta video yang bermanfaat di lingkup pengguna.

Menurut artikel yang tertulis pada website merdeka.com, data merupakan suatu kumpulan yang terdiri dari fakta-fakta untuk memberikan gambaran yang luas terkait dengan suatu keadaan. Seseorang yang akan mengambil sebuah kebijakan atau keputusan umumnya akan menggunakan data sebagai bahan pertimbangan. Melalui data seseorang dapat menganalisis, menggambarkan, atau menjelaskan suatu keadaan. Di samping itu, data dikumpulkan melalui cara-cara tertentu kemudian diolah menjadi suatu informasi yang jelas dan mudah dipahami setiap orang. Data memiliki jenis dan fungsinya masing-masing, hal ini sesuai dengan kebutuhan pengelola data. Secara umum, data banyak digunakan untuk suatu penelitian tertentu. Namun, seiring berjalannya waktu, data dibutuhkan untuk memenuhi berbagai keperluan di berbagai bidang. Tentunya hal ini bertujuan untuk memberikan informasi yang jelas dan benar setelah data tersebut diolah lebih lanjut.

2.1.2 Pengertian Informasi

McFadden, dan kawan-kawan mendefinisikan informasi sebagai data yang telah diproses sedemikian rupa sehingga meningkatkan pengetahuan seseorang yang menggunakan data tersebut. Shannon dan Weaver, dua orang insinyur listrik melakukan pendekatan secara matematis untuk mendefinisikan informasi (Kroenke). Menurut mereka, informasi adalah jumlah ketidakpastian yang dikurangi ketika sebuah pesan diterima. Artinya, dengan adanya informasi, tingkat

kepastian menjadi meningkat. Menurut Davis, informasi adalah data yang telah diolah menjadi sebuah bentuk yang berarti bagi penerimanya dan bermanfaat dalam pengambilan keputusan saat ini atau saat mendatang (Kadir, 2014).

2.2 Pengertian Sistem Informasi

Sistem informasi mencakup sejumlah komponen (manusia, komputer, teknologi informasi dan prosedur kerja), ada sesuatu yang diproses (data menjadi informasi) dan dimaksudkan untuk mencapai suatu sasaran atau tujuan (Kadir, 2014).

2.3 Basis Data

2.3.1 Pengertian Basis Data

Menurut Rosa dan Shalahuddin (2018), sistem basis data adalah sistem terkomputerisasi yang tujuan utamanya adalah memelihara data yang sudah diolah atau informasi dan membuat informasi tersedia saat dibutuhkan. Pada intinya basis data adalah media untuk menyimpan data agar dapat diakses dengan mudah dan cepat. Sistem informasi tidak dapat dipisahkan dengan kebutuhan akan basis data apapun bentuknya, baik berupa *file* teks ataupun DBMS (*Database Management System*).

2.3.2 DBMS (*Database Management System*)

Menurut Rosa (2018), DBMS (*Database Management System*) atau dalam bahasa Indonesia sering disebut sebagai sistem manajemen basis data adalah suatu sistem aplikasi yang digunakan untuk menyimpan, mengelola, dan menampilkan data. Suatu sistem aplikasi disebut DBMS jika memenuhi persyaratan minimal sebagai berikut:

- a. Menyediakan fasilitas untuk mengelola akses data.
- b. Mampu menangani integritas data.
- c. Mampu menangani akses data.
- d. Mampu menangani backup data.

DBMS sudah mulai berkembang sekitar tahun 1960an. Sekitar tahun 1970an mulai berkembang teknologi RDBMS (*Relational DBMS*) yaitu DBMS yang berbasis relasional model. Secara sederhana relasional model dapat dipahami sebagai suatu model yang memandang data sebagai sekumpulan tabel yang saling terkait. Hampir semua DBMS mengadopsi SQL sebagai bahasa untuk mengelola

data pada DBMS. Karena pentingnya data pada suatu perusahaan, maka hampir sebagian besar perusahaan memanfaatkan DBMS dalam mengelola data yang mereka miliki.

2.3.3 SQL (*Structured Query Language*)

Menurut Rosa dan Shalahuddin (2018), SQL (*Structured Query Language*) adalah bahasa yang digunakan untuk mengelola data pada RDBMS. SQL awalnya dikembangkan berdasarkan teori aljabar relasional dan kalkulus. Meskipun SQL diadopsi dan diacu sebagai bahasa standar oleh hampir sebagian besar RDBMS yang beredar saat ini, tetapi tidak semua standar yang tercantum dalam SQL diimplementasikan oleh seluruh RDBMS tersebut. Sehingga terkadang ada perbedaan perilaku (hasil yang ditampilkan) oleh DBMS yang berbeda padahal *query* yang dimasukkan sama.

2.4 *Data Warehouse*

2.4.1 *Pengertian Data Warehouse*

Menurut Purwanti (2017), *Data Warehouse* (DW) adalah suatu *database* khusus yang digunakan sebagai “gudang data” atau data yang telah terkonsolidasi dari sumber-sumber data berbagai sistem informasi yang ada pada suatu organisasi/perusahaan. Misalkan jika kita memiliki sistem pembelian (*purchasing*), sumber daya manusia (*human resources*), penjualan (*sales*), inventori, dan lain-lain. Maka *data warehouse* adalah hasil konsolidasi atau penggabungan data-data dari berbagai sistem OLTP tersebut dalam satu kesatuan sehingga kaitan antara satu dengan yang lainnya menjadi jelas dan kuat. Dengan adanya konsolidasi tersebut, diharapkan banyak informasi berharga yang bisa didapatkan sehingga menciptakan nilai strategis yang tinggi dalam kaitan pengambilan keputusan.

Struktur *data warehouse* tentunya perlu dioptimalkan untuk berbagai kepentingan sistem pelaporan terutama yang bersifat analisa menyeluruh dan dengan kaitan perkembangan OLAP yang semakin penting, *data warehouse* saat ini lebih banyak mengadopsi pemodelan yang cocok untuk dikonsumsi sistem OLAP dengan baik.

2.4.2 *Tahap Perancangan Data Warehouse*

Berikut ini adalah sembilan langkah untuk merancang suatu *data warehouse* yang dicetuskan oleh Kimball (2002), yaitu:

a. Pemilihan Proses

Hal yang pertama dilakukan adalah menganalisa dan menentukan subjek permasalahan yang dihadapi, lalu mengidentifikasi proses bisnis yang berhubungan dengan permasalahan tersebut. Pemilihan proses dilakukan untuk memperjelas batasan data warehouse yang akan dibuat.

b. Menentukan *Grain*

Grain merupakan calon fakta yang dapat dianalisis. Pemilihan grain dilakukan untuk memutuskan apa yang dipresentasikan *record* dari tabel fakta. Pada proses kedua ini dipeilihlah data dari calon fakta dengan memutuskan *record* apa yang akan dipresentasikan pada tabel fakta.

c. Mengidentifikasi dan Penyesuaian Dimensi

Dalam tahap ini dilakukan identifikasi dimensi untuk setiap tabel fakta yang ada. Misal, setiap data pelanggan pada tabel dimensi pembeli dilengkapi dengan *id_pelanggan*, *no_pelanggan*, *tipe_pelanggan*, *tempat_tinggal*, dan lain sebagainya.

d. Memilih Fakta

Pada tahap ini dipilih fakta-fakta yang akan mengisi disetiap tabel fakta, dimana fakta yang dipilih harus sesuai dengan *grain* yang telah ditentukan. Biasanya penyesuaian dimensi dan grain ini ditampilkan dalam bentuk matriks.

e. Menyimpan Perhitungan Awal dalam Tabel Fakta

Sesuai dengan *grain* yang telah ditentukan sebelumnya yang merupakan calon-calon fakta, masing-masing fakta memiliki data yang dapat dihitung. Pada tahap inilah perhitungan tersebut dilakukan. Hal ini terjadi apabila fakta kehilangan *statement*.

f. Melihat Kembali Tabel Dimensi

Pada tahap ini ditambahkan keterangan selengkap-lengkapny pada tabel dimensi. Keterangan harus bersifat intuitif dan mudah dipahami oleh pengguna.

g. Memilih Durasi *Database*

Pada proses ini yang dilakukan adalah menentukan pembatasan waktu untuk data yang diambil dan dipindahkan ke dalam tabel fakta. penentuan durasi ini tergantung terhadap informasi perusahaan.

h. Menelusuri Perubahan dari Dimensi

Pada proses ini yang dilakukan adalah mengamati perubahan data dari tabel dimensi dengan cara:

1. Atribut yang telah dirubah terisi ulang.
2. Atribut yang telah berubah menimbulkan sebuah dimensi yang baru.
3. Atribut dimensi yang telah berubah menimbulkan alternatif sehingga nilai atribut yang lama dan yang baru dapat diakses secara bersama pada dimensi yang sama.

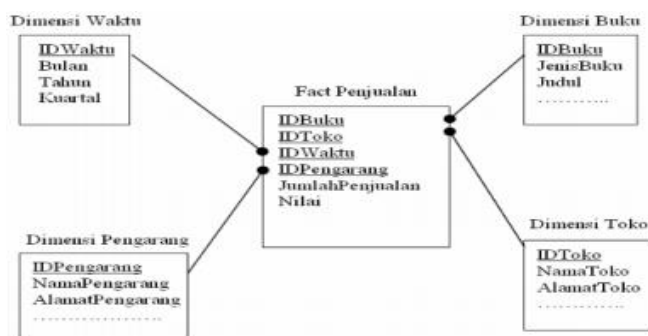
i. Memutuskan Prioritas *Query* dan Tipe *Query*

Tahap ini difokuskan pada perancangan fisik untuk *data warehouse*, seperti keberadaan dari ringkasan (*summaries*) dan penjumlahan (*aggregate*) serta masalah proses ETL (*extract, Transformation Loading*), backup, dan keamanan (*security*) yang memeberikan batasan kepada user merupakan faktor yang haruslah diperhatikan.

2.5 Pemodelan Data Multidimensional

2.5.1 Star Schema

Menurut Purwanti (2017), *star schema* adalah alat dimana pemodelan dimensional berisi sebuah tabel fakta pusat. Tabel fakta berisi atribut deskriptif yang digunakan untuk proses query dan foreign key untuk menghubungkan ke tabel dimensi. Atribut analisis keputusan terdiri dari ukuran performa, metrik operasional, ukuran agregat, dan semua metrik yang lain yang di perlukan untuk menganalisis performa organisasi. Tabel fakta menunjukkan apa yang didukung oleh *data warehouse* untuk analisis keputusan. Tabel dimensi mengelilingi tabel fakta pusat. Tabel dimensi berisi atribut yang menguraikan data yang dimasukkan dalam tabel fakta. Tabel dimensi menunjuk bagaimana data akan dianalisis.



Gambar 2.1 *Star Schema* Pada *Data Warehouse*

Tabel dimensi waktu adalah suatu tabel dimensi yang harus ada dalam sebuah *data warehouse* karena setiap data warehouse adalah time series. Waktu merupakan dimensi pertama yang harus diperhatikan dalam proses *sort order* dalam sesuatu database karena ketika hal itu pertama kali dilakukan, *loading* data secara berturut-turut dalam interval waktu tertentu akan masuk kedalam tempat yang sebenarnya dalam sebuah disk.

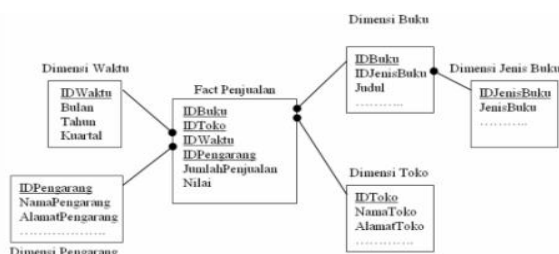
Data *designer* menyatakan secara eksplisit mengenai kebutuhan tabel dimensi waktu karena jika kunci waktu didalam tabel fakta adalah tipe data tanggal, beberapa *SQL query* mendapat kendala langsung dari kunci waktu pada tabel fakta dan menggunakan natural SQL untuk kendala pada bulan dan tahun. Karakteristik *Star Schema* adalah sebagai berikut :

- a. Pusat dari star disebut *fact table*
- b. *Fact table* mempunyai sebuah nilai *aggregate* dari data-data yang berasal dari tabel dimensi
- c. Setiap tabel dimensi berelasi langsung dengan *fact table*
- d. Tabel dimensi beisikan data tentang informasi atau waktu
- e. Relasi antara *fact table* dengan dimensi-dimensinya adalah 1-N (*one to many*)
- f. *Primary key* pada tabel dimensi akan menjadi *key* pada *fact table* atau dapat dikatakan bahwa fact table memiliki kombinasi *key* dari tabel dimensi tersebut.

Kelebihan model star adalah lebih simpel, mudah dipahami, hasil dari proses *query* juga relatif lebih cepat. Kekurangan model star adalah lebih boros dalam *space*.

2.5.2 *Snowflakes Schema*

Menurut Purwanti (2017), model *snowflake* merupakan perluasan dari star, yang sama-sama punya satu atau lebih dimensi. Pada *snowflake*, tabel yang berelasi pada *fact table* hanya tabel dimensi utama, sedangkan tabel yang lain



dihubungkan pada tabel dimensi utama. Model *snowflake* ini hampir sama seperti teknik normalisasi.

Gambar 2.2 *Snowflakes Schema Pada Data Warehouse*

Kelebihan model *snowflakes schema* adalah:

- a. Pemakaian *space* yang lebih sedikit
- b. *Update* dan *maintenance* yang lebih mudah.

Kekurangan model *snowflakes schema* adalah:

- a. Model lebih kompleks dan rumit
- b. Proses *query* lebih lambat
- c. *Performance* yang kurang bagus.

2.6 ETL (*Extract, Transform, Load*)

Menurut Purwanti (2017), ETL adalah sekumpulan proses untuk mengambil dan memproses data dari satu atau banyak sumber menjadi sumber baru, misalnya mengolah *database OLTP* menjadi *database OLAP*. Sumber data yang diolah ETL bisa dari berbagai sumber data, tidak hanya dari *database OLTP* saja, tetapi bisa juga dari *website, file teks, spreadsheet, database, email*, dan lain sebagainya. ETL terdiri dari 3 bagian utama, setiap bagian memiliki fungsi sesuai namanya.

- a. *Extract*, semua proses yang diperlukan untuk terhubung dengan beragam sumber data, dan membuat data tersebut tersedia bagi proses-proses selanjutnya; seperti contoh membaca *file* microsoft excel, mengambil data dari *database*, mengambil data kurs mata uang dari situs sebuah bank, dan lainnya.
- b. *Transform*, bagian ini mengacu pada fungsi apa saja yang berfungsi untuk mengubah data yang masuk menjadi data yang dikehendaki. Fungsi-fungsi dapat berupa pemindahan data, validasi data sesuai aturan yang ditetapkan, modifikasi isi, tipe atau struktur data, integrasi atau penggabungan data dari sumber-sumber lain, perhitungan, dan lain-lain. Seperti contoh mengubah tipe data dari string menjadi *date*, memeriksa apakah data nomor kartu kredit sesuai

format kartu kredit yang baku atau tidak, melakukan *lookup* untuk mendapatkan nilai tertentu berdasarkan kunci, dan lain-lain.

- c. *Load*, semua proses yang diperlukan untuk mengisi data ke target, sebagai contoh Hasil dari proses sebelumnya disimpan ke dalam *file* Microsoft Excel, Hasil dari proses sebelumnya disimpan ke dalam *database* OLAP, dan lain-lain.

Aplikasi ETL sangat membantu sekali dalam merancang dan membentuk suatu *data warehouse*. Beberapa kemampuan yang harus dimiliki oleh ETL adalah sebagai berikut:

- a. Dapat membaca dan mengirim data dari dan ke berbagai sumber seperti *file teks*, *file Excel*, *table-table database relasional*, *web services*, dan lain sebagainya.
- b. Mampu melakukan manipulasi (transformasi) data dari yang sederhana sampai ke tahap rumit. Beberapa pengolahan data yang harus bisa dilakukan seperti melakukan normalisasi dan denormalisasi data, memecah kolom, melakukan perhitungan yang terkait tanggal dan jam, dan lain-lain.
- c. *Script* dapat dirancang dan diimplementasikan dengan tingkat produktivitas yang tinggi.
- d. Menghasilkan informasi meta data pada setiap perjalanan transformasi.
- e. Memiliki audit log yang baik.
- f. Dapat ditingkatkan performanya dengan *scale up* dan *scale out*.

Walaupun membangun solusi ETL dengan programming biasa seperti *coding* dengan bahasa Java atau .NET sangat dimungkinkan, namun tidak disarankan karena produktivitasnya yang cukup rendah dan berakibat ke pemborosan waktu dan biaya. Berbagai utilitas ETL telah tersedia di pasaran dan menunjukkan kinerja yang baik contohnya adalah Informatica, Data Transformation Service (DTS), Microsoft SQL Server Integration Service (SSIS), Pentaho Data Integration (Kettle), dan lain sebagainya.

2.7 OLAP (*Online Analytical Processing*)

Menurut Purwati (2017), *Online Analytical Processing* atau OLAP adalah sistem yang dirancang khusus untuk menghasilkan laporan analisa yang fleksibel,

kompleks dan dapat dihasilkan dalam waktu yang cepat. *Database* OLAP sangat berbeda dengan OLTP dari sisi beban kerja dimana OLAP dirancang dan difokuskan pada kecepatan pembacaan data sedangkan OLTP pada kecepatan perekaman dan perubahan data.

Umumnya *database* OLAP dihasilkan melalui suatu proses *batch* dan biasanya dilakukan dalam periode tertentu (tiap tengah malam, tiap minggu, dan sebagainya). *Database* OLAP biasanya juga telah merupakan *database* yang sudah diperkaya dari berbagai sumber data OLTP, dan biasanya merupakan suatu *data warehouse*. OLAP mengandung 2 tipe dasar yaitu *measure* dan *dimension*.

- a. *Measure* adalah data bilangan yang terukur, misalkan kuantitas (*quantity*), harga (*price*), nilai rata-rata (*average*) dari kelompok nilai tertentu, jumlah (*sum*) dari kelompok nilai tertentu, dan lain sebagainya.
- b. *Dimension* mengacu pada kategori yang digunakan untuk mengatur *measure*. Biasanya data dikelompokkan dalam bentuk bertingkat (*level*). Dimensi yang umumnya hampir selalu ada adalah dimensi waktu (*time dimension*). Pada dimensi ini hirarki yang disusun biasanya tahun, kuartal, bulan dan hari. Bisa juga jika diperlukan ada minggu di antara bulan dan hari.

Hampir semua *Database Management System* (DBMS) yang umum seperti Oracle Database, Microsoft SQL Server, IBM DBk, PostgreSQL, MySQL dan yang lainnya dapat digunakan untuk *database* OLTP maupun OLAP. Yang membedakan antara *database* OLTP dan OLAP adalah pada skema tabel yang dibentuk. Skema tabel *database* OLTP umumnya berbentuk *normalization*, sedangkan *database* OLAP menerapkan skema star atau *snowflake*. Pada OLTP bahasa *query* yang umum digunakan adalah *Structured Query Language* (SQL), sedangkan pada OLAP bahasa *query* yang digunakan adalah *Multidimensional Expressions* (MDX).

2.8 Pentaho

Menurut Hidayati (2012), Pentaho merupakan *open source* BI (*Bisnis Intelijen*) sekaligus menyediakan *platform* dan program untuk membangun suatu aplikasi BI. Pentaho terdiri dari banyak program komputer yang bekerja secara bersamaan dan menyediakan solusi bisnis intelijen. Karena terdiri dari berbagai komponen,

dimana ada komponen level bawah, namun juga ada komponen yang menyediakan fungsionalitas level atas yang iasanya mengandalkan fungsionalitas yang ditawarkan oleh komponen level bawah, keseluruhan koleksi komponen dalam Pentaho dapat dilihat seperti stack, yang mana semakin tinggi levelnya maka akan semakin dekat dengan *end-user*.

2.9 Metode Pengumpulan Data

Menurut Rosa dan Shalahuddin (2018), hal pertama yang dilakukan dalam analisis sistem adalah melakukan pengumpulan data. Ada beberapa teknik pengumpulan data yang sering dilakukan, yaitu:

a. Teknik Wawancara

Pengumpulan data dengan menggunakan wawancara memiliki beberapa keuntungan, yaitu:

1. Lebih mudah dalam menggali bagian sistem mana yang dianggap baik dan bagian mana yang dianggap kurang baik.
2. Jika ada bagian tertentu yang perlu digali lebih dalam, maka dapat menanyakannya langsung kepada narasumber.
3. Dapat menggali kebutuhan user secara lebih bebas.
4. *User* dapat mengungkapkan kebutuhannya secara lebih bebas.

b. Teknik Observasi

Pengumpulan data dengan menggunakan observasi mempunyai keuntungan, yaitu:

1. Analisis dapat melihat langsung bagaimana sistem lama berjalan.
2. Mampu menghasilkan gambaran lebih baik jika dibanding dengan teknik lainnya.

c. Teknik Studi Pustaka

Studi pustaka dilakukan untuk memperoleh data dan informasi dengan membaca berbagai bahan penulisan, karangan ilmiah serta sumber-sumber lain mengenai permasalahan yang berhubungan dengan penulisan.

2.10 Penelitian Terkait

Hasil penelitian yang terdahulu terkait dengan penelitian yang dilakukan saat ini adalah sebagai berikut:

- a. Menurut Putu dan Gede (2018) dalam penelitiannya mengenai *data warehouse* sebagai prediksi penjualan produk pada Toko Mekarsari menyimpulkan bahwa dengan adanya *data warehouse* tersebut dapat membantu pemilik usaha dalam proses pengambilan keputusan barang yang akan dijual dan disimpan dengan stok lebih.
- b. Menurut Gusti dan Nyoman (2019) dalam penelitiannya mengenai *warehouse management system* Pada PT. Uniplastindo Interbuana Bali menyimpulkan bahwa penerapan *warehouse management system* dapat meningkatkan efisiensi biaya material *handling* serta dapat memberikan kemudahan bagi admin dalam mengumpulkan data karena dengan adanya sistem data dapat dikumpulkan dengan lebih mudah, cepat dan akurat.
- c. Menurut Khotimah dan Sriyanto (2016) dalam penelitiannya mengenai *data warehouse* untuk mendukung sistem akademik menyimpulkan bahwa *data warehouse* yang dibuat terkait dengan jumlah mahasiswa baru, status mahasiswa, dan lulusan dapat dijadikan dasar untuk menyajikan data yang terintegrasi sebagai data pendukung dalam proses evaluasi bidang akademik dan sebagai dasar proses pengambilan keputusan bagi manajemen.
- d. Menurut Rianto dan Hadis (2017) dalam penelitiannya mengenai *data warehouse* rumah sakit menyimpulkan bahwa dengan adanya *data warehouse* pasien, maka pengambilan keputusan untuk meningkatkan pelayanan menjadi lebih mudah.