

BAB II

LANDASAN TEORI

2.1 Kajian Teori

2.1.1. Data Warehouse

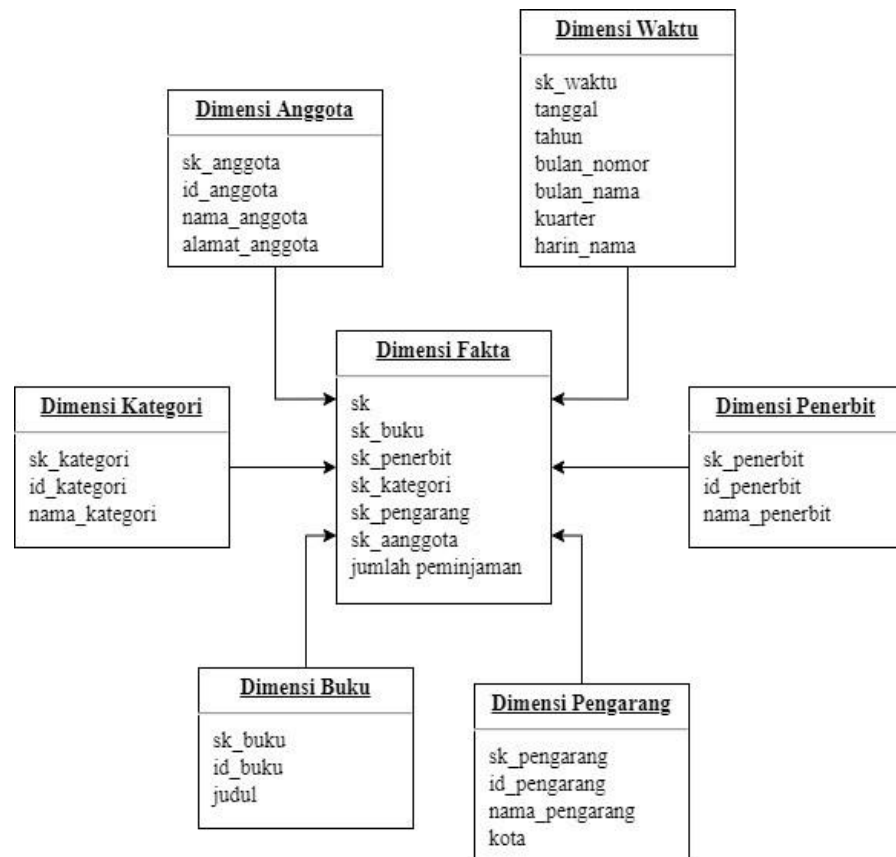
Data warehouse didefinisikan sebagai sekumpulan sebuah data yang saling terintegrasi dan bersifat objek oriented yang dibangun untuk menghasilkan sebuah informasi yang diperlukan sebagai pengambil keputusan. *Data warehouse* didefinisikan sebagai data yang terintegrasi, bersifat *subjek-oriented, non voliated*, dan *time varian* (Easterita et al., 2020). data yang tersimpan berupa data sejarah atau (*history data*) yang digunakan untuk menganalisa data dan sebagai proses pengambil keputusan. Sebuah basis data dapat digolongkan menjadi *data warehouse* apabila memiliki sebuah karakteristik sebagai berikut :

- a. Berorientasi objek, data diorganisasikan oleh objek secara terperinci, contohnya berdasarkan nama, alamat, waktu dan lain-lain.
- b. Terintegrasi, mencakup dari semua data aplikasi operasional organisasi dan data yang dibentuk secara konsisten.
- c. *time-variant*, data tidak menyediakan status saat ini, data disimpan untuk jangka Panjang missal periode lima atau sepuluh tahun bahkan lebih dan digunakan sebagai tren, perbandingan dan sebuah peramalan.
- d. *Nonvolatile* data yang dimasukkan ke dalam *data warehouse* adalah data yang hanya bisa dilihat atau *read-only*, tidak dapat dirubah.

- e. Ringkas, jika diperlukan, data operasional dimasukkan kedalam ringkasan. *Data warehouse* biasanya tidak dinormalisasikan sehingga masih terjadi redundansi atau duplikat data.
- f. Sumber, data yang tersedia dalam *data warehouse* merupakan data *internal* dan *external*.
- g. Metadata, metadata mengacu pada data tentang data yang menguraikan struktur dan beberapa arti tentang data, dengan demikian mendukung penggunaan yang efektif atau tidak efektif dari data.

2.1.2. Model Star Schema

Dalam *Data Warehouse* terdapat beberapa *schema*, diantaranya ada *star schema*, *snowflake schema*. *Star schema* merupakan sebuah pemodelan multidimensi yang paling sederhana yang bentuknya seperti bintang (Nurmalasari et al., 2019). Terdapat satu fakta (*fact tabel*) di pusat bintang dengan beberapa tabel dimensi (*dimensional tabels*) yang saling berelasi. Semua tabel dimensi berhubungan dengan tabel fakta. Tabel fakta memiliki beberapa key yang merupakan kunci indeks yang berhubungan dengan tabel dimensi. *Snowflake schema* atau sering disebut sekema bola salju merupakan sebuah pecahan dari skema bintang dengan tambahan beberapa tabel dimensi yang tidak berhubungan langsung dengan tabel fakta. tabel dimensi tersebut berhubungan dengan tabel dimensi yang lainnya.



Gambar 2.1 Contoh *Star Schema*

2.1.3. Metode Kimball

Metode *kimball* merupakan metode yang telah dibangun untuk perancangan *Data Warehouse* metode ini diciptakan oleh Ralph kimball pada tahun 1980 (Suni & Ridwan, 2018). Pemodelaan *data warehouse* diemnsional yang digunakan berdasarkan *kimballgroup.com* dalam pembuatan dimensional menggunakan tahap 4 langkah dalam pemrosesan dimensional. Pada metode 4 langkah ini untuk langkah pertama yaitu pemilihan proses bisnis, lalu melakukan deklarasi *grain*, identifikasi dimensi, indentifikasi fakta (Ferianto et al., 2021).

2.1.4. Pentaho Data Integration

Pentaho merupakan sebuah perusahaan yang menyediakan produk produk *source* untuk keperluan *data warehouse* dan *Bussines intelligence*. *Pentaho data integration* adalah bagian dari *Pentaho studio* yang digunakan untuk persiapan dalam pembentukan *Data Warehouse* dalam pemrosesan *Extraction , Load dan Transformation* (ETL) menggunakan pendekatan berbasis meta-data. Ketel merupakan *acronym* dari “*kettle E.T.T.L Environment*”, menyediakan fitur desain yang intuitif,grafis, *drag and drop design* dengan basis standart yang dapat diskalakan.dalam proses ETL dengan Pentaho dapat membaca berbagai sumber data yang ada dengan menggunakan fitur desain yang di sediakan yang telah disesuaikan berdasarkan kebutuhan dalam pembentukan *Data Warehouse*. (Müller & Keller, 2014)

2.1.5. XAMPP

Xampp adalah sebuah perangkat lunak yang memiliki sinkatan dari X (Empat Sistem Operasi Apapun). Xampp Merupakan perangkat lunak bebas yang mendukung beberapa operasi, dan kompilasi dari phpp, *Perl*. Xampp adalah tool yang digunakan untuk menyediakan beberapa perangkat lunak dalam satu paket. Dalam Xampp sudah terdapat Apche (Web Server), MySQL (Database), PHP (server side scripting), *Perl*, FTP server, PhpMyAdmin, dan berbagai Pustaka lainnya. Kepanjangan dari xampp yaitu :

X : Program ini dapat dijalankan di berbagai operasi seperti windows, linux, Mac OS, dan solaris.

A : Apache merupakan aplikasi web server. Apache bertugas menghasilkan halaman web yang benar kepada user berdasarkan kode PHP yang telah dituliskan oleh web atau user.

M : MySQL, merupakan aplikasi data server. SQL merupakan Bahasa terstruktur yang digunakan untuk mengolah database.

P : PHP merupakan Bahasa pemrograman web, dimana user dapat menggunakan bahasa pemrograman ini untuk membuat web yang bersifat server-side scripting.

P : perl , yaitu merupakan bahasa pemrograman untuk segala keperluan, dan dikembangkan kali oleh Larry Wall di mesin *Unix*.

2.1.6. Mysql

Menurut Ramadhan dan Mukhayar (2020) dalam jurnal Penggunaan *Database Mysql* dengan Interface *PhpMyAdmin* sebagai Pengontrolan *SmartHome* Berbasis *Raspberry Pi*, *Mysql* adalah salah satu jenis *database* server yang cukup terkenal. *Mysql* menggunakan Bahasa SQL untuk *license Expection* data dan ada juga versi komersialnya. SQL (*structure query language*) merupakan salah satu *syntax* Bahasa yang digunakan untuk memasukkan, mengubah, menghapus, memilih dan melindungi data SQL bukanlah *database* aplikasi, tetapi digunakan untuk melakukan perintah perubahan database berupa pengguna SQL. Karena *Mysql* adalah server *multithreaded* maka akan memproses permintaan layanan.

2.1.7. Database

Basis data Sebagian dari data yang terstruktur dan berhubungan maka mudah untuk disimpan, dimanipulasi serta dipanggil oleh pengguna. Data mendeskripsikan sebuah domain (ranah) tertentu membuat pengguna mudah mendapatkan jawaban atas pertanyaan yang diajukan ke sistem basis data tersebut. Basis data sendiri dapat diartikan sebagai koleksi atau kumpulan data yang terorganisir sedemikian rupa sehingga data mudah untuk disimpan, manipulasi data (diperbarui, ditambah, dikelola dengan perhitungan tertentu) dan dihapus. manfaat basis data memiliki kecepatan dan kemudahan, pemanfaatan *database* memungkinkan kita untuk dapat menyimpan data dan melakukan perubahan terhadap data dan juga menampilkan-nya. Ruang penyimpanan yang efisien (spasial) yang menghilangkan jumlah pengulangan data, baik dengan menerapkan enkripsi atau dengan menciptakan hubungan dan bentuk kelompok data yang saling berhubungan. Akurasi (presisi) penggunaan encoding atau pembentukan hubungan data serta aturan tipe data, domain data, keunikan data, dll, dapat ditegakkan secara ketat dalam database. Seiring dengan tersedianya data dari waktu ke waktu, semakin banyak ruang penyimpanan yang dibutuhkan, baik dalam jumlah maupun jenisnya. Tidak semua data selalu dibutuhkan, jadi Anda bisa memilih untuk memiliki data utama atau master, data transaksional, data historis hingga data kadaluarsa. Kelengkapan (*completeness*) untuk memenuhi kebutuhan fungsional kelengkapan data semakin berkembang, yaitu dengan menambah record-record data dan elakukan perubahan struktur data dalam basis data, dalam bentuk penambahan objek baru berupa tabel atau dengan penambahan file-file baru. Keamanan (*security*) mencegah pengaksesan

oleh orang yang tidak berwenang, penggunaan *database* dikelola oleh sistem aplikasi yang mendukung lingkungan multiuser (Novendri, 2019).

2.1.8. ETL (Extract Transform Load)

ETL merupakan tahapan pertama yang akan dilakukan untuk mengolah data, yaitu menerima data yang berasal dari berbagai sumber data yang terstruktur maupun tidak terstruktur atau bisa disebut juga berupa data *heterogen*. menjadikannya sebuah data yang konsistensi (Yulianto, 2019). Data yang telah di olah dalam proses ETL akan menjadi lebih terstruktur dan konsistensi.

ETL adalah proses yang sangat penting dalam *data warehouse* dan ETL data dari operasi dapat dimasukkan ke dalam *data warehouse*. Anda juga dapat menggunakan ETL untuk mengintegrasikan data ke dalam sistem yang ada. Tujuan ETL adalah untuk mengumpulkan, memfilter, memproses, menggabungkan data yang relevan dari berbagai sumber, dan menyimpannya di gudang data. Hasil ETL memenuhi kriteria pergudangan data seperti data historis, konsolidasi, agregat, dan statis, menghasilkan data dengan struktur yang dirancang untuk proses analisis (Mulyana, 2014). *Data warehouse* sebagai sebuah sistem yang mengekstrak, mengisi dan mengirim data sumber ke dalam penyimpanan dimensional, menyesuaikan data yang di butuhkan oleh dimensi yang akan digunakan sebagai pendukung dan mengimplementasikan query dan analisis dalam mengambil sebuah keputusan (Yulianto, 2019).

2.1.9. OLAP

OLAP (*online analytical processing*) adalah seperangkat alat yang mendukung proses menganalisis dan membandingkan data dalam basis data, suatu pendekatan yang memberikan jawaban cepat dan beragam untuk kueri analitik multidimensi. (Linda, 2015). Struktur operasional yang paling utama pada OLAP merupakan struktur data *cube* (kubus). *Cube* dalam OLAP adalah struktur data yang multidimensi (*actual/virtual*) yang dapat melakukan analisis data secara cepat. Susunan data pada kubus memiliki fungsi untuk mengatasi keterbatasan *database* relational. Sederhananya, OLAP adalah teknik untuk meminta data dalam format yang kompleks dan menganalisis data dalam jumlah besar. Untuk alasan ini, OLAP disebut analisis data multidimensi. (RMS, 2019)

2.1.10. Power BI

Power BI adalah kombinasi layanan, aplikasi, dan perangkat lunak konektor. Power BI adalah alat pelaporan dan visualisasi data dan rujukan untuk menganalisis data sebagai pengambil keputusan proses bisnis.. Power BI dapat membaca data dari berbagai sumber yang sulit di pahami dan di buat menjadi sebuah laporan yang dengan mudah untuk di baca dan di analisa (Edhya, 2021).

2.2 Kajian Penelitian Terdahulu

Penelitian sebelumnya digunakan untuk referensi atau perbandingan dengan penelitian penulis dan dapat digunakan untuk menghindari kesamaan antar penelitian.

Penelitian yang dilakukan oleh Risky Ferianto, Arie Nugroho, dan Teguh Andriyanto tahun 2021 dengan judul “*Data warehouse* Pengolahan data Penjualan Studi Kasus UD HF Bersaudara” mengatakan bahwa *Data Warehouse*

Digunakan untuk menyimpan data transaksi penjualan, namun sistem POS tidak tertata dengan baik dan dibuat secara otomatis oleh sistem. Untuk memudahkan pengguna melihat data yang dibutuhkan, penulis membuat database OLAP yang sama persis dengan data yang akan digunakan. Dihasilkan oleh aplikasi kasir. (Ferianto et al., 2021). Perbedaan penelitian ini dengan penelitian yang dilakukan penulis adalah tempat penelitian, pada penelitian penulis melakukan penelitian untuk TBM data buku sedangkan tempat penelitian pada jurnal terdahulu adalah pada data data penjualan.

Penelitian lain juga dilakukan oleh Rio Irawan tahun 2021 dengan judul “Pemodelan *Data warehouse* Perpustakaan Fakultas Tarbiyah dan Ilmu Keguruan (FTIK) Institut Agama Islam Negeri (IAIN) Palangka Raya” pada perpustakaan Institut Agama Islam Negeri (IAIN) Palangka Raya, yaitu Penelitian ini menghasilkan pemodelan *data warehouse* dengan *snowflake schema* yang dilakukan dengan proses *extract, loading, transformation* (ETL). *Data warehouse* yang telah dirancang menghasilkan sebuah informasi yang berguna untuk kepentingan pengolah perpustakaan FTIK IAIN Palangka Raya dalam mendukung setiap manajerial yang dibutuhkan nantinya (Irawan, 2021). Perbedaan Penelitian ini dengan penelitian yang dilakukan penulis adalah penelitian terdahulu menggunakan metode *kimball 9* langkah sedang pada penulis metode yang digunakan menggunakan metode *kimball 4* langkah dan pada penelitian penulis menggunakan *Star schema* .

Penelitian lain juga dilakukan oleh Sucipto, Sucipto, dan Arie Nugroho tahun 2020 dengan judul penelitiannya “Analisis *Data warehouse* Pada

Perpustakaan MAN X untuk Efisiensi Manajemen” pada penelitian ini menemukan sebuah masalah yaitu admin perpustakaan kesulitan dalam memperoleh informasi terkait buku apa yang sering dipinjam maupun yang jarang dipinjam, serta jumlah buku yang sedikit tidak sepadan dengan buku yang sering dipinjam. Hal ini dikarenakan pada sebuah kegiatan operasional seperti absensi buku, peminjaman buku, pengembalian buku masih dicatat dalam buku induk. Maka dari itu dibangunlah sebuah *data warehouse*. hasil dari penelitian ini menghasilkan sebuah aplikasi *business intelligence* berbasis web. Dalam aplikasi ini dapat menampilkan jumlah peminjaman buku berdasarkan nama anggota, pengarang, judul buku, penerbit, kategori dan tanggal peminjaman sehingga dapat digunakan untuk mendapatkan informasi terkait peminjaman buku pada perpustakaan MAN X (Sucipto et al., 2020) Perbedaan Penelitian ini dengan penelitian yang dilakukan penulis adalah tempat penelitian, penulis melakukan penelitian pada TBM, sedangkan pada penelitian terdahulu melakukan penelitian pada perpustakaan.

Penelitian yang telah dilakukan lainnya yaitu dengan judul “Pemetaan Data Siswa Menggunakan *Data Warehouse* Untuk Promosi di SMK Z” dalam penelitian ini dilakukan untuk melakukan pemodelan *data warehouse* supaya dapat membantu pihak sekolah dalam menganalisis data dengan mudah. Masalah yang dialami adalah data yang tersimpan dalam dua aplikasi yang berbeda sehingga pihak sekolah mengalami kesulitan dalam menganalisis data pendaftaran. Penelitian ini dilakukan karna mengingat pentingnya pengolahan data pada Lembaga Pendidikan agar proses pengolahan data dapat dilakukan

dengan mudah dan cepat (Khoirudin et al., 2020). Perbedaan penelitian ini dengan penelitian penulis adalah sumber data yang diperoleh.

Penelitian lain yang dilakukan oleh Syamsul Bakhri, dan Yamin Nuryamin tahun 2018 dengan judul “Rancangan *Data warehouse* untuk Penunjang Sistem Informasi Eksekutif Pada Yayasan UMMU’L QURO di Depok” pelaporan dianggap *statis* tidak fleksibel dalam menjelajahi berbagai informasi untuk melihat data dari berbagai dimensi. Dibangunnya sebuah rancangan *Data Warehouse* yang dapat diimplementasikan untuk menunjang keputusan pada proses penyusunan laporan di yayasn Ummu’l Quro menjadi lebih sederhana. Media presentasi yang digunakan pada yayasan adalah aplikasi berbsis web ,maka menjadikan kebutuhan laporan yang lebih fleksibel karena dapat diakses melalui internet. (Syamsul Bakhri, 2018) Perbedaan Penelitian ini dengan penelitian yang dilakukan penulis adalah penelitian terdahulu melakukan penelitian dengan pedekatan *top down*, sedangkan pada penelitian penulis menggunakan pendekatan kuantitatif deskriptif.

Penelitian yang telah dilakukan lagi selanjutnya yaitu dengan judul “Implementasi Data Warehouse Dan Penerapannya Pada Toko Magnifique Clothes Dengan Menggunakan Tools Pentaho” oleh yasman. Dalam penelitiannya membahas tentang implememntasi *data warehouse* pada toko Magnifique Clothes. Pemilik toko memerlukan sebuah informasi tentang penjualannya, pada toko ini hanya menyediakan informasi penjualan yang berasal dari *M.Access* yang ada di komputer toko tersebut dan tidak dapat menyajikan data dengan cepat. Untuk mengatasi masalah tersebut makan dibangunlah sebuah *data warehouse* supada

mendapatkan informasi yang lebih mudah dan cepat. Dalam penelitian ini untuk membangun sebuah *data warehouse* menggunakan *tools* pentaho. Hasil dalam penelitian mencakup data penjualan yang telah terjadi sejak tahun 2013-2017 berupa tampilan grafik ataupun dashboard penjualan.(Subuh & Yasman, 2019). Perbedaan penelitian ini dengan penelitian penulis adalah tempat penelitiannya , pada penelitian ini mengolah data penjualan dan penelitian yang sedang dilakukan penulis membahas tentang data buku dari TBM, serta sumber data yang diambil dari penelitian ini berasal dari M.Access pada penelitian penulis sumber data yang akan diolah berasal dari excel dan teks CSV.