

## BAB II

### LANDASAN TEORI

#### 2.1 Penelitian Terdahulu

Beberapa penelitian terkait yang akan menjadi referensi dari penelitian ini adalah sebagai berikut:

No	Judul Artikel	Penulis	Jurnal	Pembahasan
1	An Overview of ETL Techniques, Tools, Processes and Evaluations in Data Warehousing [8]	Bilal Khan1, Saifullah Jan, Wahab Khan, and Muhammad Imran Chughtai	Journal on Big Data, Vol.6, 2024	<p>Masalah: Pentingnya melakukan proses ETL yang baik dan benar serta eksplorasi terhadap <i>tools</i> dan teknik yang tepat untuk mendukung proses ETL.</p> <p>Metode: Pendekatan menggunakan ETL (<i>Extract, Transform and Load</i>)</p> <p>Hasil: Proses ETL memiliki peran yang penting dalam membangun fondasi Data Warehouse.</p> <p>Kendala utama dalam menjalankan proses ini meliputi menjaga kualitas data, menangani</p>

No	Judul Artikel	Penulis	Jurnal	Pembahasan
				<p>kompleksitas keanekaragaman sumber data, serta tantangan dalam tahap proses ETL itu sendiri. Penekanan dilakukan pada perlunya penerapan proses ETL yang tepat untuk memastikan kualitas data dan mendukung manajemen data yang efektif.</p>
2	On-line Analytical Processing (OLAP) Operation for Outpatient Healthcare [5]	Talib M. J. Al Taleb, Sami Hasan, Yaqoob Yousif Mahd	Iraqi Journal of Science, 2021	<p>Masalah: Melakukan implementasi teknik analisis OLAP dalam mengolah data layanan kesehatan rawat jalan melalui integrasi data dari berbagai sumber menggunakan <i>Data Warehouse</i>.</p> <p>Metode: Pendekatan menggunakan <i>data warehousing, star schema</i>, dan OLAP</p> <p>Hasil: Menggunakan teknik analisa OLAP</p>

No	Judul Artikel	Penulis	Jurnal	Pembahasan
				untuk layanan kesehatan rawat jalan sebagai alat <i>business intelligence</i> dan mendukung keputusan dalam mengekstrak informasi secara cepat dan tepat dengan cara menarik data dari Data Warehouse.
3	An Overview of Data Warehouse and OLAP Technology [9]	Kirti Pathania	International Journal of Advanced Research in IT and Engineering, Vol.11, No.5, 2022	Masalah: Kebutuhan akan teknologi database yang mendukung <i>decision support system</i> dibandingkan dengan aplikasi transaksi tradisional, dengan fokus pada tantangan dalam <i>data warehousing</i> dan OLAP. Metode: Data warehousing dan OLAP. Hasil: OLAP memungkinkan pengguna untuk mengakses, melihat, dan melakukan analisis pada

No	Judul Artikel	Penulis	Jurnal	Pembahasan
				data dengan cepat dan efisien. Dengan analisis ini, pengguna dapat memahami data dari perspektif macro hingga micro melalui cube yang dibuat dengan sumber data terintegrasi pada data warehouse. Ini mendukung pengambilan keputusan yang lebih cepat dan mendalam berdasarkan analisis multidimensi.
4	OLAP analysis of user energy consumption based on multitemporal distribution characteristics [10]	Bin Liu1, Zhengyu Yang, Jiaqing Wu and Jie Gu, Ilko Kovacic, Christoph G. Schuetz, Bernd Neumayr, Michael Schrefl	Journal of Physics: Conference Series, 2022	Masalah: Penggunaan OLTP ( <i>Online Transaction Processing</i> ) dianggap kurang relevan dengan kebutuhan analisis, maka dibutuhkan pendekatan baru menggunakan OLAP. Metode: Pendekatan OLAP dengan berbagai metode analisis multidimensi seperti

No	Judul Artikel	Penulis	Jurnal	Pembahasan
				<p><i>slicing, dicing, drilling, dan rotating.</i></p> <p>Hasil: Hasil menunjukkan bahwa OLAP memungkinkan analisis data multidimensi dari berbagai perspektif dengan fleksibilitas tinggi. Berbagai metode dalam OLAP terbukti membantu pengambilan keputusan dengan memberikan wawasan yang lebih dalam kepada pengambil keputusan</p>
5	Data Warehouse Design with ETL Method (Extract, Transform, And Load) for Company Information Centre [11]	Wulan Stau Fana, Rini Soviab, Randy Permana, Md Ataul Islam	International Journal of Artificial Intelligence Research, Vol.5, No.2, 2021	<p>Masalah: Membuat <i>data warehouse</i> menggunakan skema <i>snowflake</i> untuk membantu perusahaan dalam pengambilan keputusan yang efektif</p> <p>Metode: Pendekatan <i>data warehouse</i> menggunakan skema</p>

No	Judul Artikel	Penulis	Jurnal	Pembahasan
				<p><i>snowflake</i></p> <p>Hasil: Penggunaan data warehouse dengan skema snowflake membantu dalam mengidentifikasi dan mengintegrasikan data dari berbagai sumber, mendukung transformasi data, serta menyediakan laporan untuk mempermudah pengambilan keputusan perusahaan.</p>
6	Improving the Quality of Service: ETL Implementation on Data Warehouse at Pharmacy Industry [12]	Fardhila Zahra Dwi Wardhani, Jansen Wiratama	Jurnal Tekno Kompak, Vol. 18, No.1, 2024	<p>Masalah: Pemantauan dan analisis data pada ketersediaan obat-obatan pada layanan farmasi dilakukan secara tradisional, sehingga pengambilan keputusan pada farmasi menjadi kurang optimal</p> <p>Metode: Pendekatan menggunakan <i>Data warehouse</i> dan OLAP</p>

No	Judul Artikel	Penulis	Jurnal	Pembahasan
				<p>Hasil: Penelitian menunjukkan bahwa penggunaan <i>star schema</i> pada <i>data warehouse</i> dan teknik analisis OLAP berhasil membuat data menjadi lebih terstruktur dan informasi menjadi lebih detail. Dengan ini, farmasi menjadi lebih mudah dalam melihat data dan pengambilan keputusan.</p>
7	Data Warehouse Modeling Using Online Analytical Approach [13]	Galih Budianto	Jurnal Ilmiah Informatikadan Ilmu Komputer, Vol.1, No.1, 2022	<p>Masalah: Persaingan antar perusahaan mendorong perlunya pemanfaatan teknologi untuk mengolah data secara cepat dan tepat. Tantangannya adalah perancangan sistem dan teknologi berbasis OLAP untuk mengumpulkan, memanipulasi, dan menyimpan data</p>

No	Judul Artikel	Penulis	Jurnal	Pembahasan
				<p>multidimensi secara efisien.</p> <p>Metode: Perancangan sistem berbasis OLAP</p> <p>Hasil: Sistem yang dikembangkan berhasil mengimplementasikan OLAP untuk membantu dalam pengambilan keputusan. Tercatat respon query tidak lebih dari 10 detik untuk menampilkan grafik.</p> <p>Dengan adanya prototype ini dapat membantu perusahaan dalam pengambilan keputusan.</p>
8	Improve HR Decision-Making Based On Data Mart and OLAP [14]	Alaa Khalaf Hamoud et al	Journal of Physics: Conference Series, 2020	<p>Masalah: Bagaimana cara melakukan implementasi <i>data mart</i> untuk mengolah dan menganalisis data karyawan selama 15 tahun terakhir, untuk menentukan kriteria yang tepat bagi</p>



No	Judul Artikel	Penulis	Jurnal	Pembahasan
				<p>karyawan baru. Juga menghasilkan laporan OLAP yang sesuai dan efektif dalam pengambilan keputusan.</p> <p>Metode: Data Mart dan OLAP</p> <p>Hasil: Hasil penelitian menunjukkan bahwa implementasi OLAP pada data mart divisi HR dapat membantu perusahaan dalam pengambilan keputusan, meningkatkan standar, juga manajemen biaya.</p>
9	Permodelan Data Warehouse untuk Penjualan Ban Menggunakan Online Analytical Processing (OLAP) [15]	Eko Saputra	Teknologipintar.org, Vol.9, 2023	<p>Masalah: Tidak ada tempat penyimpanan data terstruktur untuk menyimpan data penjualan, dan pembuatan <i>report</i> masih dilakukan secara manual menggunakan excel file.</p> <p>Metode: <i>Data warehouse, nine step</i></p>

No	Judul Artikel	Penulis	Jurnal	Pembahasan
				<p><i>methodology</i>, dan OLAP</p> <p>Hasil: Pembuatan data warehouse dengan <i>star schema</i> menghasilkan sebuah tempat penyimpanan data yang lebih terstruktur, maka proses <i>reporting</i> dengan bantuan OLAP juga dapat dilakukan dengan lebih cepat.</p>
10	Analysis the Performance of Data Warehouse Architecture Types [16]	Karwn Jameel Merceedi et al	Journal of Soft Computing and Data Mining, Vol.3, No.1, 2022	<p>Masalah: Bagaimana cara mengelola dan mengoptimalkan <i>data warehouse</i> untuk mendukung perusahaan dalam mengambil keputusan. Fokus utama ada pada arsitektur <i>data warehouse</i> yang diharapkan bisa mengelola data historis secara efisien.</p> <p>Metode: <i>Data warehouse</i>, ETL</p>

No	Judul Artikel	Penulis	Jurnal	Pembahasan
				Hasil: Penelitian ini mengidentifikasi area penting dalam <i>data warehouseing</i> , seperti pembersihan data, optimisasi, pengelolaan, juga pemrosesan <i>query</i> yang efektif. Dengan menggunakan arsitektur <i>data warehouse</i> yang tepat, maka pengelolaan data akan lebih efisien.

Berdasarkan beberapa penelitian terdahulu yang sudah dijelaskan, terdapat beberapa kesimpulan yang dapat diambil, yaitu penggunaan *data warehouse* dan OLAP merupakan sebuah kombinasi yang efektif dalam melakukan analisis pada data yang berukuran besar. Proses ETL (*Extract, Transform, and Load*) merupakan proses yang sangat penting untuk membangun sebuah *data warehouse*, juga merupakan proses yang krusial [8], [12]. Proses ETL tidak hanya berfungsi untuk mengekstraksi data dari berbagai sumber, tetapi juga memastikan bahwa data telah diubah dan dimuat dengan kualitas yang baik ke dalam data warehouse. Proses ini menjadi fondasi penting dalam penelitian ini, mengingat kualitas data yang baik akan berdampak langsung pada akurasi analisis yang dilakukan [11], [12]. Penelitian ini memanfaatkan proses ETL untuk mengintegrasikan data penjualan dari berbagai sumber, membersihkan data yang tidak konsisten, serta mengelompokkan data berdasarkan dimensi-dimensi yang relevan.

Selain itu, struktur data yang digunakan dalam data warehouse juga menjadi aspek krusial. Seperti yang dijelaskan dalam penelitian [11], skema snowflake dianggap sebagai salah satu pendekatan yang efektif untuk menyusun data dalam data warehouse. Skema ini memungkinkan normalisasi tabel dimensi, yang berarti bahwa tabel dimensi dipecah menjadi tabel-tabel yang lebih kecil dan lebih spesifik. Dengan pendekatan ini, data menjadi lebih terorganisir, terperinci, dan mudah dikelola, terutama untuk analisis yang membutuhkan ketelitian tinggi. Penelitian ini menggunakan skema snowflake untuk menyusun data penjualan berdasarkan atribut-atribut penting. Dengan struktur ini, data dapat diakses dengan lebih efisien, dan analisis multidimensi menjadi lebih akurat.

Setelah data warehouse memiliki struktur yang baik, proses analisis dilakukan menggunakan teknik OLAP. Penelitian terdahulu, seperti yang diungkapkan oleh [5], [9], [14], [15], menunjukkan bahwa OLAP adalah alat yang sangat bermanfaat untuk menganalisis data dari berbagai dimensi secara mendalam. Fitur OLAP, seperti *slicing*, *dicing*, *drill-down*, dan *roll-up*, memungkinkan pengguna untuk menjelajahi data dengan fleksibel [11], [13]. Misalnya, dengan fitur *slicing*, pengguna dapat memfilter data berdasarkan dimensi tertentu, seperti waktu atau kategori produk, sedangkan fitur *drill-down* memungkinkan eksplorasi data yang lebih rinci, seperti melihat penjualan per subkategori produk atau lokasi spesifik.

Dalam penelitian ini, OLAP digunakan untuk menganalisis data penjualan tahun 2023 secara lebih mendalam dan terstruktur. Teknik ini memungkinkan visualisasi data dari berbagai perspektif, seperti analisis tren penjualan berdasarkan waktu, perbandingan performa antar kategori produk, dan identifikasi lokasi penjualan dengan kontribusi tertinggi. Penelitian terdahulu, seperti yang dijelaskan oleh [10], [15], [16], juga menunjukkan bahwa OLAP mampu memberikan wawasan yang mendalam bagi pengambilan keputusan bisnis. Oleh karena itu, penelitian ini mengadopsi teknik

OLAP untuk menghasilkan visualisasi berupa tiga dashboard analitik yang dirancang untuk membantu perusahaan memahami data penjualannya secara lebih komprehensif.

Dengan mengintegrasikan proses ETL yang baik, struktur data warehouse berbasis skema *snowflake*, dan analisis OLAP, penelitian ini diharapkan dapat memberikan kontribusi nyata dalam meningkatkan kemampuan analisis data penjualan di Matahari Department Store. Dukungan dari penelitian terdahulu ini semakin memperkuat argumen bahwa pendekatan yang digunakan dalam penelitian ini relevan dan efektif dalam mendukung kebutuhan bisnis yang kompleks.

## 2.2 Tinjauan Teori

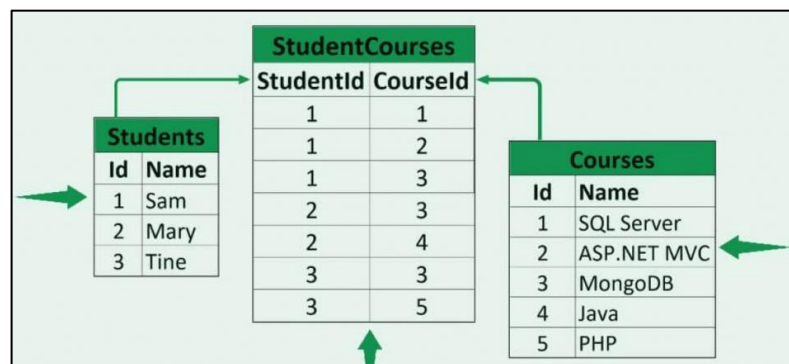
### 2.2.1 Database Management System

Database merupakan sebuah kumpulan informasi dari sebuah organisasi yang disimpan dengan baik dan rapi agar dapat dilakukan pemahaman dan mendapatkan informasi yang lebih mendetail [17]. Database berfungsi sebagai tempat penyimpanan data akan mempermudah sebuah organisasi dalam mengolah data menjadi lebih nyaman diakses, mudah untuk diolah, dan dikelola dengan efisien. Dengan adanya database, proses pencarian informasi akan lebih mudah dan cepat.

Tujuan dibuatnya sebuah database adalah untuk menjadi tempat pengimanan data, dan melakukan pengambilan, penghapusan, juga pengeditan menjadi lebih cepat dan nyaman [17]. Proses ini memungkinkan organisasi untuk melakukan manajemen data dengan lebih terorganisir, mengurangi resiko kehilangan data, dan memelihara data dalam jangka waktu yang lama. Dengan penggunaan database, maka perusahaan atau organisasi akan menyimpan data dengan lebih konsisten [18].

Sedangkan sebuah DBMS atau *Database Management System* merupakan sebuah perangkat lunak (*software*) yang digunakan untuk mengolah *database* dengan lebih efisien [19]. Tidak hanya berfungsi untuk menyimpan data,

DBMS juga dapat mengatur akses data, memperbaharui informasi secara langsung, juga beberapa fungsi pemeliharaan data [18]. Saat ini, DBMS juga dikenal sebagai RDBMS atau *Relational Database Management System*, yaitu melakukan penyajian data didalam tabel tabel yang saling terhubung [20], seperti contoh pada gambar 2.1.



Gambar 2. 1 Contoh Relational Database Management System [21]

### 2.2.2 Business Intelligence

Bisnis analitik atau *Business Intelligence* merupakan proses pengumpulan, mengelola, juga menganalisa data bisnis untuk menghasilkan wawasan yang dapat membantu perusahaan dalam mengambil keputusan bisnis [22]. BI mencakup beberapa proses, yaitu: *decision support system*, *query and reporting*, *OLAP (Online Analytical Processing)*, *statistical analysis*, *forecasting*, juga *data mining* [22].

Beberapa keuntungan dari penggunaan BI adalah kemampuannya dalam mengelola data secara otomatis, memberikan hasil yang lebih cepat dan meningkatkan akurasi dalam pengambilan keputusan, peningkatan kinerja operasional, juga kemampuannya untuk memprediksi pasar [23]. Dengan menggunakan BI, kita dapat menganalisa data mentah, dan *mengolah* data menjadi data yang siap untuk disajikan [23].

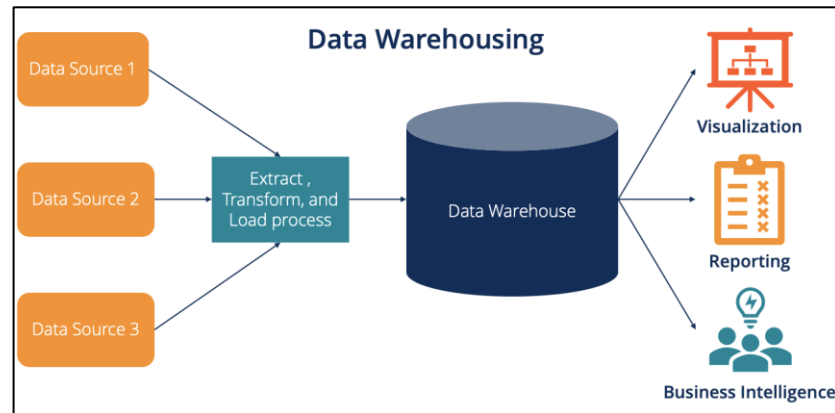
Terdapat beberapa tahapan implementasi BI, dimulai dari melakukan pengumpulan data dari berbagai sumber, data merupakan data mentah yang akan diolah. Kemudian menyiapkan penyimpanan data yang terstruktur untuk menyimpan data mentah. Data yang sudah disimpan kemudian akan di transformasikan dan diproses agar dapat dianalisa. Beberapa analisa yang dapat dilakukan adalah OLAP, *data mining*, juga visualisasi data sesuai dengan keperluan perusahaan. Hasil data yang sudah diolah akan digunakan untuk mendukung strategi perusahaan [22], [23], [24].

### **2.2.3 Data Warehouse**

#### **2.2.3.1 Pengertian Data Warehouse**

*Data Warehouse* atau biasa disingkat DW merupakan sebuah tempat penyimpanan yang dibuat secara khusus sebagai wadah data yang diambil dari berbagai sumber yang beragam, mulai dari sistem operasional, database transaksi, juga data eksternal dan berbagai department yang ada di sebuah organisasi atau perusahaan [25]. Data yang sudah dikumpulkan kemudian akan diolah dan disimpan secara terstruktur dan terintegrasi di dalam *Data Warehouse* [26].

Salah satu kegunaan dari pembuatan data warehouse adalah untuk memproses data dan menjadikannya sebagai dasar dalam pengambilan keputusan penting di perusahaan. Dengan mengolah data secara terpusat pada *Data Warehouse*, perusahaan atau organisasi dapat melakukan analisis data historis, melakukan identifikasi pola maupun tren, juga menyusun laporan yang lebih akurat. Gambar 2.2 di bawah ini merupakan salah satu contoh ilustrasi dari proses *Data Warehousing*.



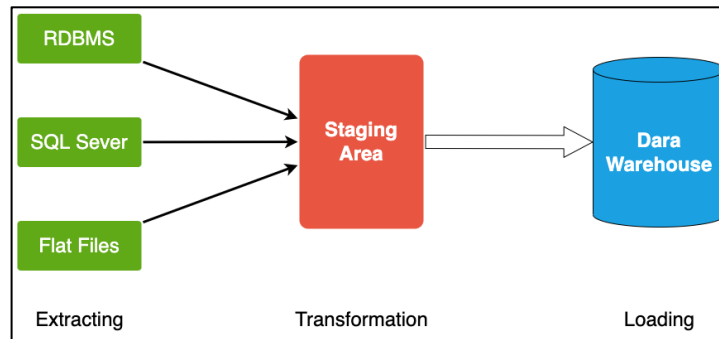
Gambar 2. 2 Ilustrasi Data Warehousing [27]

Dimulai dari mengambil data, kemudian melakukan proses ETL (*extract, transform, dan load*), kemudian memasukkan data kedalam *Data Warehouse*. Setelah data berhasil di bersihkan, maka kemudian dapat dilakukan visualisasi data, atau digunakan untuk reporting, atau untuk *business intelligence*.

### 2.2.3.2 Proses ETL

Dalam proses melakukan *data warehousing* terapat beberapa tahapan penting untuk mengumpulkan dan memproses data, yaitu proses ETL (*Extract, Transform, and Load*) [8]. Proses ini bertujuan untuk mempersiapkan data agar siap digunakan dan dianalisa untuk pengambilan keputusan bisnis sebuah perusahaan. Proses ETL terdiri dari tiga proses utama, seperti pada gambar 2.3, yaitu *Extracting, Transformation, dan Loading*.





Gambar 2. 3 Proses Extract, Transform, dan Load [28]

1. **Extract**, merupakan tahap pertama dalam proses ETL, yang melibatkan pengumpulan data dari berbagai sumber atau database yang ada di perusahaan [8]. Proses ekstraksi bertujuan untuk mengumpulkan semua data yang relevan dan dibutuhkan dalam satu sumber data terpusat. Hal ini memudahkan dalam melakukan analisis data, karena data yang tersebar di berbagai sistem atau departemen dapat digabungkan menjadi satu kesatuan. Setelah data dikumpulkan, data tersebut akan ditempatkan dalam sebuah area penyimpanan sementara yang disebut *Staging Area*. Di dalam *staging area*, data akan disiapkan terlebih dahulu sebelum dipindahkan ke dalam data warehouse, memungkinkan data untuk diseragamkan dan disiapkan untuk proses berikutnya.
2. **Transform**, setelah data terkumpul ke dalam *staging area*, langkah berikutnya adalah melakukan transformasi data agar sesuai dengan format yang dibutuhkan dan mengintegrasikan data dari berbagai sumber yang berbeda. Pada tahapan ini, proses pembersihan dan penyesuaian data dilakukan untuk memastikan data yang dimuat ke dalam data warehouse merupakan data yang sudah bersih, konsisten, dan siap di analisa [8]. Beberapa proses yang biasanya terjadi pada tahapan ini adalah sebagai berikut:
  - a. **Filtering**, proses ini merupakan proses pemilihan data yang sesuai dengan kriteria yang dibutuhkan. Sebagai contoh untuk analisis data

penjualan, data hanya akan dipilih berdasarkan bulan tertentu, seperti data penjualan hanya untuk bulan Juni.

- b. *Cleaning*, pada tahap ini, semua data yang tidak konsisten atau memiliki format yang berbeda-beda akan diseragamkan. Sebagai contoh, penulisan nama wilayah seperti 'Jakarta' akan disingkat menjadi 'JKT', atau 'Daerah Istimewa Yogyakarta' akan disingkat menjadi 'DIY', dan seterusnya. *Cleansing* dilakukan untuk menghindari data yang tidak konsisten dan mengganggu proses analisa.
  - c. *Joining*, data dari berbagai tabel yang saling berhubungan akan digabungkan untuk membentuk informasi yang lebih lengkap. Proses ini dilakukan untuk memastikan integrasi antar data yang berasal dari berbagai sumber yang berbeda-beda untuk membentuk satu kesatuan informasi.
  - d. *Sorting*, data akan diurutkan berdasarkan kriteria atau aspek tertentu, seperti tanggal, bulan, atau kategori lainnya yang relevan. Pengurutan data dilakukan untuk membuat hasil analisis menjadi lebih terstruktur dan mudah dihapami.
3. **Load**, setelah proses pengecekan, integrasi, dan transformasi data selesai, tahap berikutnya adalah memindahkan data ke dalam data warehouse [29]. Pada tahap ini, data yang telah disiapkan di dalam data warehouse siap untuk dianalisis lebih lanjut. Data warehouse berperan sebagai tempat penyimpanan pusat bagi data yang telah diproses, dan analisis dapat dilakukan melalui berbagai metode, salah satunya adalah dengan membuat visualisasi data. Proses pemuatan data memastikan bahwa data yang dimasukkan ke dalam data warehouse telah terstruktur dengan baik dan sesuai dengan kebutuhan analisis yang diinginkan,

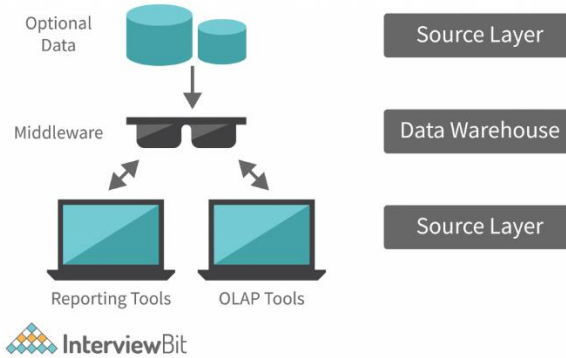
serta dapat menyediakan data yang tepat guna dalam mendukung pengambilan keputusan bisnis yang akurat.

### 2.2.3.3 Arsitektur Data Warehouse

Arsitektur dalam data warehouse merupakan sebuah kerangka desain yang menjelaskan bagaimana cara data dikumpulkan, disimpan, dan diakses oleh data warehouse. Arsitektur menjadi hal yang penting karena menentukan bagaimana cara data akan diolah dan bagaimana proses analisis data dapat dilakukan secara efisien [30]. Dalam implementasinya, terdapat tiga jenis arsitektur yang umum digunakan dalam pembuatan *data warehouse*, yaitu *single-tier architecture*, *two-tier architecture*, dan *three-tier architecture* [26]. Setiap arsitektur memiliki karakteristik, kelebihan, dan kekurangannya masing-masing sesuai dengan kebutuhan dan skala organisasi yang menggunakannya.

- a. *Single-tier architecture*, merupakan jenis arsitektur *data warehouse* yang dirancang untuk meminimalkan jumlah data yang akan disimpan. Penggunaan arsitektur ini dapat membantu mengurangi resiko adanya redundansi data, sehingga data yang disimpan lebih efisien dan tidak terduplikasi. Salah satu kekurangan dari jenis arsitektur *single-tier* ini adalah tidak memiliki pemisah antara proses analisis dan transaksi, hal ini menyebabkan data yang digunakan untuk analisis dan transaksi disimpan dalam satu tempat yang sama, seperti pada gambar 2.4. Kekurangan ini dapat menghambat kinerja sistem karena proses analisis dan transaksi berjalan dan dilakukan dengan sumber data yang sama. Tanpa pemisah antara analisis dan transaksi akan menghambat efisiensi dan kecepatan proses analisis.

### Single-Tier Data Warehouse Architecture

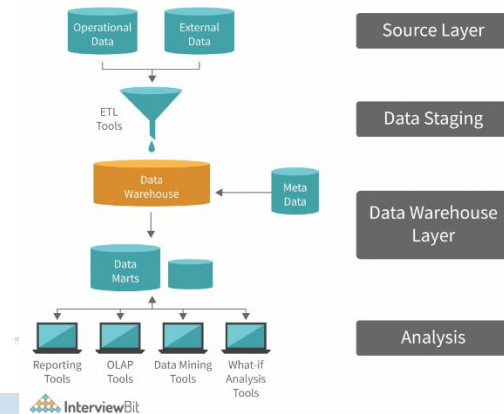


Gambar 2. 4 Contoh Single Tier Architecture Data Warehouse [31]

- b. *Two-tier architecture*, merupakan arsitektur yang lebih kompleks dibandingkan dengan model arsitektur *single tier* karena memiliki komponen tambahan yaitu *staging area*. *Staging area* berfungsi untuk tempat penyimpanan data secara sementara agar dapat menyesuaikan data dengan format yang ada sebelum dimasukkan kedalam *data warehouse*. Dengan adanya *staging area*, maka data akan lebih terstruktur dan terorganisir. Salah satu kekurangan dari *two-tier* adalah keterbatasan dalam hal jaringan dan kapasitas penyimpanan, akan menjadi hambatan jika volume data yang harus di proses sangat besar. Gambar 2.5 merupakan contoh dari arsitektur *two-tier*.

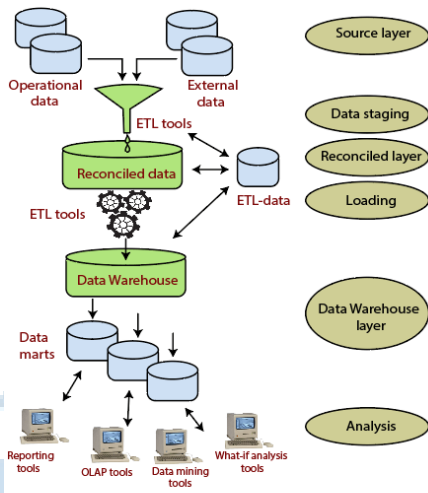
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## Two-Tier Data Warehouse Architecture



Gambar 2. 5 Contoh Two Tier Architecture Data Warehouse [31]

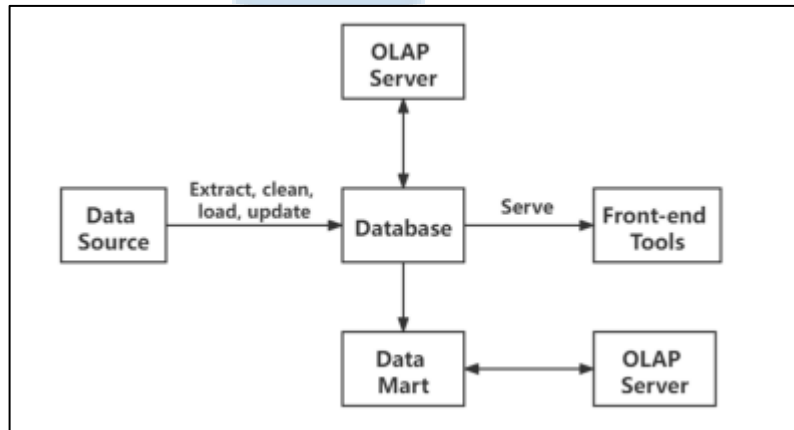
- c. *Three-tier architecture*, merupakan jenis arsitektur data warehouse yang paling umum digunakan. Terdiri dari tiga lapisan yang masing-masing memiliki perannya penting dalam memproses data. Lapisan pertama yaitu *bottom tier* merupakan lapisan yang berfungsi untuk membersihkan data yang akan dimasukkan ke dalam data warehouse. Pada lapisan ini juga terjadi transformasi data untuk menyeragamkan format data yang diterima agar lebih konsisten saat dianalisa. Lapisan kedua adalah *middle tier*, merupakan lapisan yang bertugas sebagai penghubung antara *end user* dengan database. Berfungsi sebagai alat komunikasi, sehingga memudahkan *end user* dalam mengakses data yang dibutuhkan. Sedangkan lapisan terakhir, atau lapisan ketiga, adalah *top tier*, terdiri dari berbagai *tools* dan API yang memungkinkan *end user* untuk mengakses data juga melakukan analisa data yang sudah disimpan di dalam data warehouse. Dengan adanya beberapa lapisan ini, membuat *three-tier architecture* menjadi lebih mudah dan efisien dalam membantu *end user* mendapatkan wawasan yang mendalam mengenai data yang ada.



Three-Tier Architecture for a data warehouse system

Gambar 2. 6 Three Tier Architecture Data Warehouse [31]

Data warehouse juga memiliki beberapa komponen penting pada arsitektur data nya, seperti pada gambar 2.7 di bawah ini terdapat beberapa komponen dasar yang penting untuk ada dalam data warehouse [10].



Gambar 2. 7 Data Warehouse System Architecture [26]

- a. *Data Source*, merupakan pondasi dari sebuah data warehouse dan tempat dimana semua sumber data. Terdiri dari dua jenis sumber data yaitu internal dan eksternal data yang akan disimpan dalam RDBMS (*Rational Database Management System*) [10].

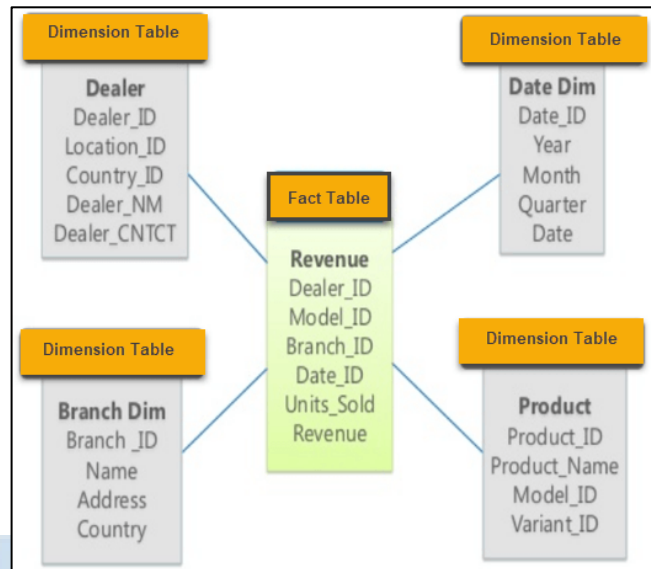
- b. *Database and Data Mart*, merupakan sebuah inti atau *core* dari sebuah data warehouse. Data dari berbagai sumber yang sudah dikumpulkan akan dilakukan proses ETL untuk kemudian dimasukkan kedalam database [10]. Data yang sudah diolah juga dapat dipecah berdasarkan kebutuhan user, seperti untuk departmen tertentu.
- c. OLAP Server, dimana data akan dilakukan integrasi dan dipecah berdasarkan kebutuhan analisa menggunakan dimensi tertentu. Data akan dipecah berdasarkan dimensi untuk membuat hasil analisa yang lebih mendalam berdasarkan sudut dan tingkatan tertentu [10], [32].
- d. Front-end Tools, merupakan tools untuk membuat analisa data yang sudah dilakukan OLAP. Beberapa tools tersebut adalah report tools, juga data analysis tools. Tools dapat dipilih sesuai dengan kebutuhan analisa oleh user. Umumnya, untuk tools analisis data akan menuju pada OLAP Server, sedangkan untuk report lebih berfokus pada data warehouse.

#### 2.2.3.4 Model Data Warehouse

Pada saat pembuatan data warehouse, terdapat beberapa model skema yang dapat digunakan atau dipilih, yaitu:

- a. Star Schema

Pada skema bintang atau *star schema*, terdapat tabel fakta dan tabel dimensi. Tabel fakta atau *fact table* merupakan tabel yang berisikan fakta dan *foreign key* yang merujuk pada tabel dimensi. Sedangkan tabel dimensi berisikan data dari tabel fakta yang lebih mendetail [33].



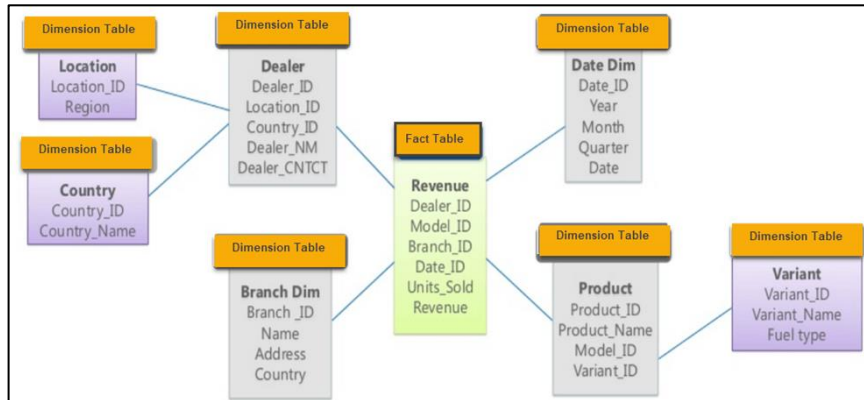
Gambar 2. 8 Contoh Star Schema [34]

Gambar 2.8 di atas merupakan contoh dari penerapan skema bintang. Pada fact table, terdapat beberapa fakta yang diambil dari tabel dimensi, seperti Dealer\_ID diambil dari tabel Dealer, begitu juga yang lainnya.

b. Snowflake Schema

Jika *star schema* artinya tabel fakta langsung berhubungan dengan tabel dimensi, maka berbeda dengan *snowflake schema*. *Snowflake schema* dimana tabel fakta tidak berhubungan langsung dengan tabel dimensi, artinya data akan dipecah menjadi tabel tambahan untuk mengurangi redundansi data [35].





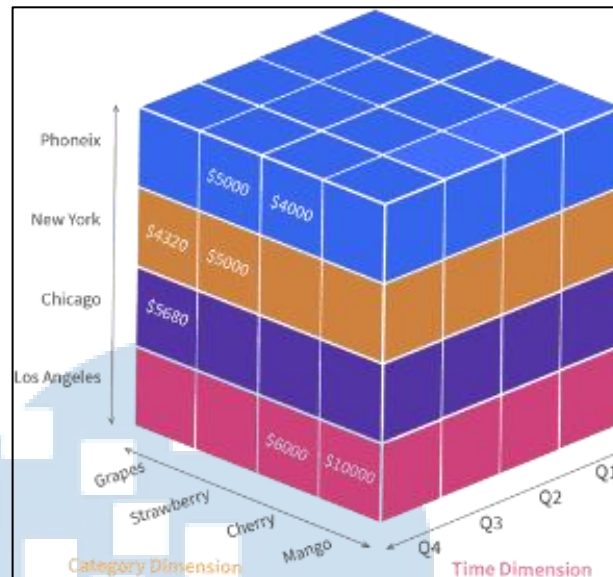
Gambar 2. 9 Contoh Snowflake Schema [36]

Gambar 2.9 di atas merupakan contoh dari *snowflake schema*, dimana sebelum dimasukkan ke tabel fakta, akan ada tabel tambahan untuk menormalisasi data.

## 2.2.4 OLAP (*Online Analytical Processing*)

### 2.2.4.1 Pengertian OLAP

OLAP atau *Online Analytical Processing* merupakan teknik pengolahan data yang digunakan untuk menganalisis informasi dari berbagai sudut pandang [7]. Teknik pengolahan data ini akan menyajikan hasil analisis dari sebuah Data Warehouse, tugasnya adalah untuk mengubah semua data yang ada di Data Warehouse menjadi sebuah data multidimensi atau kubus [10]. Perubahan data menjadi sebuah kubus multidimensional dilakukan untuk mempermudah dalam melakukan analisa. Dengan kubus multidimensional maka kita dapat melihat data berdasarkan sudut dan dalam tingkatan tertentu [10]. Gambar 2.10 merupakan contoh visualisasi dari kubus multidimensi.



Gambar 2. 10 Kubus Multidimensional OLAP [37]

#### 2.2.4.2 Karakteristik OLAP

Beberapa karakteristik umum dari teknik pengolahan OLAP adalah sebagai berikut:

1. Dapat melakukan analisis secara multidimensional, dengan menggunakan OLAP dapat melakukan analisis data berdasarkan dimensi yang berbeda-beda sesuai dengan kebutuhan, dan melakukan analisis berdasarkan perspektif yang berbeda-beda [10].
2. *Drill-Down dan Roll-Up*, OLAP juga memungkinkan analisa dilakukan dengan menelusuri data pada titik terendah (*drill-down*) atau melakukan analisa pada tingkat tertinggi (*roll-up*). Melakukan analisa dari yang paling mendetail sampai paling umum [38].
3. *Slice and Dice*, selain melakukan analisa di tingkat tertinggi dan terendah OLAP juga menyajikan analisa dalam bentuk *slice* dan *dice*. *Slice* maksudnya adalah melakukan analisa dengan memilih sub-set tertentu pada data, dan *dice* adalah melihat data dari kombinasi data tertentu [10].

### 2.2.4.3 Jenis OLAP

OLAP juga dibedakan berdasarkan bagaimana cara melakukan dan memproses data. Dibagi menjadi beberapa jenis [5], yaitu:

- a. MOLAP (Multidimensional OLAP), penyimpanan data dilakukan dengan cara kubus multidimensional [5]. Dengan menggunakan kubus multidimensional maka memungkinkan melakukan analisa data dengan lebih cepat dan terstruktur. Namun, ketika sumber data mengalami perubahan maka MOLAP harus dilakukan perubahan atau memasukkan data terbaru secara berkala.
- b. ROLAP (Relational OLAP), data akan disimpan dalam basis data relational, dan kubus multidimensi akan dihasilkan melalui kueri SQL [5]. Menggunakan tabel relational untuk menyimpan data dan mengolah data menggunakan kueri secara *real-time*, dan dapat menyimpan data dengan ukuran yang cukup besar .
- c. HOLAP (Hybrid OLAP), merupakan gabungan dari MOLAP dan ROLAP. Menggunakan tabel relational untuk penyimpanan data, berbentuk kubus multidimensi, dengan kueri yang cukup cepat [5]. Berikut tabel 2.1 untuk memperjelas perbedaan dari ketiga jenis OLAP [5] di atas.

Tabel 2. 1 Perbandingan Antara MOLAP, ROLAP, dan HOLAP

Feature	MOLAP	ROLAP	HOLAP
Data Storage	Cube	Relational Table	Relational Table
Query Performance	Fastest	Slowest	Fast
Storage Consumption	High	Low	Medium

### 2.2.5 User Acceptance Test

*User Acceptance Test* merupakan sebuah pengujian yang dilakukan bersama dengan *user* untuk memastikan bahwa aplikasi (dalam penelitian ini

adalah hasil visualisasi data) berhasil berjalan dengan baik dan memberikan *output* yang sesuai [39], [40]. Dengan menjalankan UAT, maka dapat diketahui bahwa hasil penelitian sudah memenuhi kebutuhan daripada *end user*, maka dari itu akan dilakukan UAT sebagai bagian dari tahapan *Evaluation*. Tabel UAT yang akan digunakan adalah sebagai berikut pada tabel 2.2.

Tabel 2. 2 Tabel User Acceptance Test

No	Questions	Expected Result	Pass/Failed
	Nama Menu		

### 2.3 Framework yang Digunakan

Terdapat beberapa framework yang bisa digunakan dan berkaitan dengan data warehousing. Penelitian ini menggunakan framework yang dirancang khusus dengan pendekatan data warehouse dan teknik analisis OLAP. Framework ini disesuaikan dengan kebutuhan penelitian untuk mempermudah proses pembuatan data warehouse yang mendukung analisis. Beberapa tahapan yang dilakukan, yaitu:

1. *Data Understanding*, pada tahapan ini akan dilakukan pemahaman pada setiap data penjualan milik perusahaan, dan memilih yang mana saja yang berkaitan dan relevan dengan penelitian ini.
2. *Arsitektur Data Warehouse*, melakukan pemilihan arsitektur *data warehouse* yang sesuai dengan penelitian, memikirkan beberapa aspek seperti sumber data, juga hasil yang diharapkan.
3. *Perancangan Data Warehouse*, merupakan tahapan untuk merancang *data warehouse*, dimulai dari membuat skema yang tepat, juga melakukan proses ETL.
4. *Visualisasi Data*, dengan menggunakan Power BI sebagai salah satu tools nya, maka akan dilakukan visualisasi data dari *data warehouse* yang sudah dibentuk.

5. *Evaluation*, pada tahapan ini akan dilakukan beberapa evaluasi yang berkaitan dengan penelitian, seperti melakukan *testing* bersama user IT, evaluasi penggunaan *data warehouse*, juga evaluasi implementasi OLAP pada *dashboard*.

Dengan *framework* ini, penelitian dapat lebih terarah dan berfokus pada pembuatan *data warehouse*, analisis OLAP, dan kebutuhan bisnis. Framework ini dirancang agar relevan, praktis, dan mampu menghasilkan data warehouse yang siap untuk dianalisis, sekaligus memberikan wawasan yang bermanfaat bagi pengambilan keputusan bisnis dari berbagai perspektif.

## 2.4 Tools yang Digunakan

### 2.4.1 MySQL

MySQL merupakan sebuah RDBMS atau *relational database management system* yang bersifat *open source*. MySQL sering digunakan untuk mengolah data dengan ukuran *small to medium size*, namun MySQL juga mampu untuk mengolah data dengan ukuran besar atau biasa disebut sebagai *big data* [41]. Kemampuannya untuk mengolah data kecil maupun besar menjadikan MySQL pilihan yang baik dan fleksibel untuk mengolah data penelitian.

Beberapa keunggulan dari penggunaan MySQL adalah kecepatannya dalam memproses kueri. Dirancang secara khusus untuk memberikan performa yang optimal dalam melakukan dan memproses kueri yang diberikan. Selain cepat, MySQL juga dapat mengatur penggunaan sumber daya dengan efisiensi, sehingga dapat menyimpan data dalam jumlah besar tanpa mengorbankan kinerja. MySQL menjadi pilihan yang tepat untuk mengolah berbagai jenis data, baik besar maupun kecil. Dibandingkan dengan RDBMS lainnya, MySQL dapat disebut sebagai *tools* yang lebih terkenal dan mudah untuk digunakan [41], [42].

### 2.4.2 Microsoft Power BI

Microsoft Power BI atau biasa dikenal sebagai Power BI merupakan sebuah *platform* analitik bisnis yang digunakan dan dirancang untuk

memudahkan pengguna dalam memahami dan menganalisa data yang berjumlah besar, mulai dari puluhan-ratusan-ribuan baris data [43]. Power BI juga dapat mengambil data dari berbagai sumber, baik yang berbasis *cloud* atau penyimpanan lokal. Kemudian, Power BI juga memungkinkan pengguna untuk mengubah data sesuai dengan format sebelum melakukan *load*. Dengan demikian, memungkinkan hasil analisa yang lebih efisien dikarenakan keseragaman data.

Beberapa fungsi dari penggunaan Power BI adalah [43], [44] :

- a. Melakukan koneksi dan meng-ekstrak data dari berbagai sumber, kemudian melakukan kompress pada data berukuran besar agar lebih mudah untuk di proses.
- b. Mendukung koneksi data, baik secara *cloud* atau koneksi ke data pada lokal komputer.
- c. Menghasilkan dashboard yang interaktif untuk mempermudah *end-user* memahami data berukuran besar dengan lebih mudah. Dengan tampilan yang mudah di mengerti dapat membantu dalam mempercepat pengambilan keputusan dan meningkatkan pemahaman pengguna terhadap tren dan pola dalam data.
- d. Hasil dari *reports* dan visualisasi data dapat dibagikan kepada user lain. Terdapat juga fitur kolaborasi yang memperbolehkan user melakukan *edit* dashboard secara bersamaan, dan menghasilkan hasil yang lebih efektif.
- e. *Reports* juga dapat diberikan keamanan data.

Secara keseluruhan, Microsoft Power BI menawarkan berbagai solusi Solusi untuk melakukan analisis data secara komprehensif dan intuitif, memudahkan pengguna untuk melakukan eksplorasi data, melakukan analisa data, juga memberikan visualisasi data secara sederhana untuk data yang besar.

### 2.4.3 MySQL Connector/ODBC 9.1

MySQL Connector/ODBC merupakan sebuah driver ODBC atau *open driver connectivity* yang dikembangkan oleh MySQL untuk menghubungkan aplikasi ODBC, salah satunya adalah Power BI [45]. Dengan menggunakan ODBC, maka akan ditetapkan standar yang dapat menghubungkan aplikasi dengan berbagai jenis database tanpa harus mengubah atau menyesuaikan konfigurasi aplikasi dengan database yang digunakan. Dengan menggunakan *connector*, maka aplikasi Power BI dapat mengakses, membaca, juga menulis data di dalam MySQL database.

Dari beberapa jenis *connector* yang ada, penelitian ini menggunakan MySQL Connector/ODBC 9.1 dikarenakan sifatnya yang *open source*, juga kemampuannya untuk menangani query besar dengan efisien. Pada penelitian ini akan menggunakan MySQL Connector/ODBC versi 9.1. Dengan versi terbaru ini, yaitu 9.1, terdapat beberapa kelebihan dibandingkan dengan versi sebelumnya. Dimana di versi 9.1 dapat membantu menangani data yang lebih kompleks, dan memberikan standarisasi yang lebih baik untuk melakukan integrasi dengan berbagai sistem, salah satunya aplikasi analitik Power BI.