

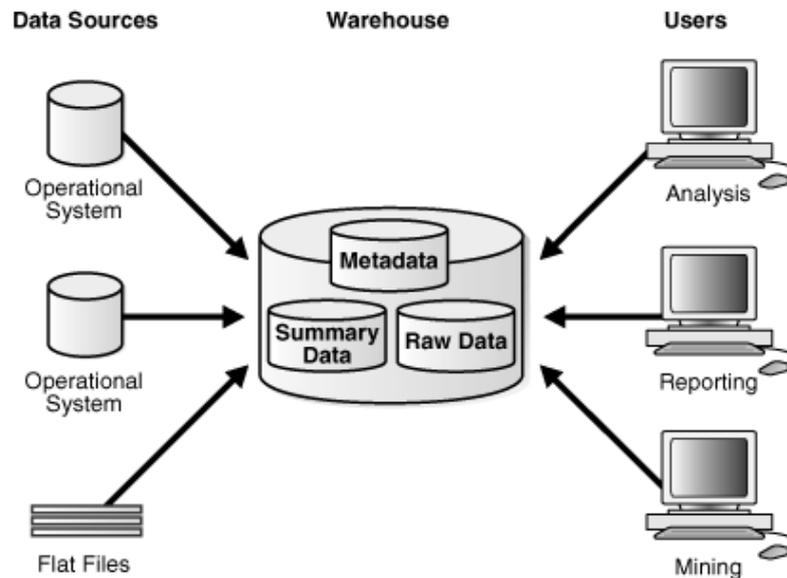
## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Data Warehouse**

Data Warehouse merupakan sebuah basis data relasional yang dirancang khusus untuk keperluan analisis dan eksplorasi data. Selain itu, fungsi utamanya adalah untuk menyimpan data historis yang berasal dari berbagai sumber informasi. Data Warehouse juga memisahkan antara proses analisis dengan proses transaksi. Hal ini bertujuan untuk menggabungkan data dari beberapa sumber [6]. Banyak perusahaan membutuhkan data warehouse untuk melakukan integrasi data, integrasi data dilakukan dengan menggabungkan data dari berbagai sistem atau aplikasi yang berbeda kemudian menyimpannya dalam satu penyimpanan yang besar. Hasil dari integrasi data dapat diolah sesuai dengan kebutuhan perusahaan [7].

Data warehouse adalah serangkaian data yang terfokus pada subjek tertentu, terintegrasi dengan baik, bersifat stabil, dan mengalami perubahan seiring waktu yang bertujuan untuk mendukung proses pengambilan keputusan. Berikut adalah beberapa karakteristik yang dapat diidentifikasi pada data warehouse [4]:



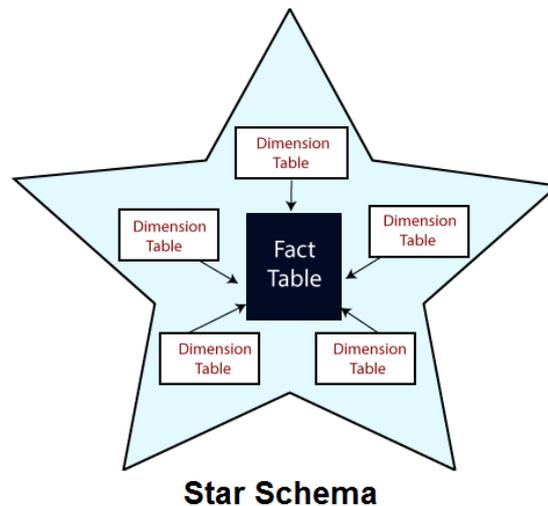
**Gambar 2.1** Karakteristik Data Warehouse

1. Data warehouse yang berfokus pada subjek tertentu menggambarkan pendekatan desain data warehouse yang terpusat pada analisis sesuai dengan kebutuhan dalam suatu organisasi.
2. Data warehouse yang terintegrasi mampu menyimpan data dari berbagai sumber sistem aplikasi ke dalam format yang konsisten dan saling terhubung satu sama lain.
3. Data warehouse memiliki rentang waktu yang memperlihatkan akurasi atau validitas data pada periode tertentu. Variasi tersebut dinyatakan melalui periode data baik secara tersirat maupun terungkap, yang mana data mencerminkan serangkaian gambaran singkat.
4. Data warehouse mengalami perubahan seiring waktu, hal ini menandakan bahwa data tidak mengalami pembaruan secara langsung, melainkan diperbarui secara berkala dari sistem operasional. Basis data secara rutin mengambil data baru yang kemudian diintegrasikan secara bertahap dengan data yang sudah ada..

Berdasarkan definisi diatas dapat disimpulkan bahwa data warehouse merupakan struktur penyimpanan basis data yang memuat informasi historis yang diperoleh dari berbagai sumber data

## 2.2 Star Schema

Star schema merupakan salah satu jenis struktur skema (*schema*) dalam pemodelan data dengan tabel fakta di tengah dan dikelilingi oleh tabel dimensi yang berisi referensi data, setiap tabel dimensi hanya berhubungan dengan tabel fakta. [4].



**Gambar 2.2** Star Schema

### 1. Fact Table

Tabel fakta, atau yang sering disebut sebagai *fact table*, adalah struktur tabel yang biasanya berisi data yang dapat diukur, seperti penjualan, pendapatan, kuantitas, dan sejenisnya. *Fact table* juga berperan sebagai himpunan kunci asing yang berasal dari kunci utama yang ada di setiap tabel dimensi. Selain itu, tabel fakta juga mencakup data historis.

### 2. Dimension Table

Tabel dimensi merupakan struktur tabel yang memuat informasi rinci yang menjelaskan kunci asing yang ada di dalam tabel fakta. Atribut yang ada di dalam tabel dimensi berfungsi untuk melakukan penyaringan, pengelompokan, atau analisis data dalam tabel fakta.

## 2.3 Pentaho Data Integration

Pentaho Data Integration adalah sebuah tools dalam domain bisnis yang berfungsi untuk melaksanakan proses Ekstraksi, Transformasi, dan Pengisian

(ETL). Pentaho juga memiliki kemampuan untuk melakukan migrasi data, melakukan pembersihan data, serta melakukan pengisian data dari file ke database atau sebaliknya, bahkan dalam jumlah data yang sangat besar. Pentaho memiliki keunggulan yaitu *Graphical user Interface (GUI)* yang *friendly* serta fitur drag and drop yang mempermudah pengguna. Elemen kunci dalam Pentaho Data Integration (PDI) adalah Transformasi dan Job. Transformasi merujuk pada rangkaian petunjuk yang disusun untuk mengubah masukan menjadi hasil yang diharapkan, dan mengikuti pola masukan-proses-keluaran. Sedangkan job adalah kumpulan intruksi untuk mengeksekusi transformation. Pentaho memiliki tiga komponen penting yaitu Spoon, Pan dan Kitchen [8] .

1. Spoon

Pentaho Data Integration (PDI) merupakan perangkat lunak dalam domain bisnis yang digunakan untuk melaksanakan proses Ekstraksi, Transformasi, dan Pengisian atau *Loading (ETL)*. Spoon adalah antarmuka grafis dalam Pentaho Data Integration (PDI). Ini adalah lingkungan pengembangan yang memungkinkan pengguna untuk membuat dan mengelola alur kerja (*jobs*) dan transformasi data (*transformations*). Spoon memungkinkan pengguna untuk mendesain dan mengatur langkah-langkah yang diperlukan untuk memperoleh, mentransformasi, dan memuat informasi. Dalam spoon, pengguna dapat membangun dan menguji alur kerja dan transformasi sebelum menjalankannya.

2. Pan

Pan adalah program yang digunakan untuk menjalankan transformasi data (*transformations*) yang telah dibuat di spoon. Dengan kitchen, transformasi data akan berjalan secara otomatis. Ini berguna saat ingin mengotomatisasi proses transformasi data tanpa harus melakukannya secara manual.

3. Kitchen

Kitchen adalah program yang digunakan untuk menjalankan alur kerja (*jobs*) yang telah dibuat di spoon. Pan dapat menjadwalkan dan

menjalankan alur kerja ETL secara otomatis tanpa harus bergantung pada antarmuka grafis. Ini sangat berguna untuk menjalankan proses ETL secara terjadwal di latar

## **2.4 ETL**

ETL adalah bagian dari menyiapkan data yang melibatkan proses pengambilan informasi, integrasi data, serta menyimpan data ke dalam database. Dalam proses ETL, informasi dari beberapa sumber secara bergilir akan diekstrak dan digabungkan ke dalam database. Penetapan lingkup ETL dilakukan dengan menganalisa setiap tabel tujuan baik dimensi maupun fakta, yang sebaiknya dilakukan pada tahap awal pembangunan arsitektur proses ETL. Berikut adalah proses ETL[7] :

### **1. Extract**

Pengambilan data atau ekstraksi melibatkan proses seleksi dan penarikan informasi dari berbagai sumber data. Dalam konteks ekstraksi data dari database operasional, transfer data dari sumber asli ke dalam data warehouse menjadi langkah yang krusial. Proses ini penting karena besar volume data yang akan dipindahkan dari sumber asli (sistem legacy) ke dalam data warehouse. Sumber informasi yang berakhir di dalam data warehouse bisa berasal dari internal organisasi (sistem legacy) atau eksternal.

### **2. Transform**

Transformasi merujuk pada rangkaian aturan atau fungsi yang diterapkan pada data yang diekstraksi. Hal ini menentukan cara data akan dipergunakan untuk analisis, yang bisa mencakup berbagai proses seperti encoding data, merging data, splitting data, melakukan perhitungan, dan pembuatan surrogate key. Hasil dari proses transformasi ini adalah data yang terstruktur dengan baik dan konsisten dengan data yang telah disimpan di dalam data warehouse. Data yang telah melalui tahap transformasi ini kemudian siap untuk digunakan dalam analisis oleh pengguna data warehouse.

### 3. Loading

Tahap Load adalah langkah di mana data dimasukkan ke dalam tujuan akhir, yang seringkali merupakan data warehouse. Waktu yang diperlukan untuk tahap ini bervariasi sesuai dengan kebutuhan organisasi. Beberapa data warehouse dapat mengisi semua informasi yang ada secara kumulatif setiap minggu, sementara yang lain mungkin hanya menambahkan data baru dalam format historis ke dalamnya, baik secara keseluruhan atau hanya pada bagian tertentu dari data warehouse yang sama..

## 2.5 OLAP

OLAP merupakan teknik dari suatu tools yang berfungsi untuk mengevaluasi data yang tersimpan di dalam gudang data (data warehouse), dengan tujuan menghasilkan laporan sesuai dengan kebutuhan pengguna. Hal ini memungkinkan pengguna untuk memperoleh keuntungan melalui akses secara langsung, responsif, dan tetap [4]. Dalam konteks ini, OLAP memungkinkan konversi data transaksional dari OLTP ke dalam format data analitis, yang kemudian disimpan dalam tabel fakta. Proses OLAP berfokus pada data historis yang memiliki histori waktu dan mendukung proses analitis, sementara data yang berkaitan dengan transaksi diolah menggunakan *On Line Transactional Processing* yang mendukung operasi dan transaksi. [6]. Dalam penelitian ini, OLAP dilakukan dari dataset yang telah dibersihkan. Dataset tersebut dianalisis untuk menentukan dimensi dan melakukan proses slicing pada data multi-dimensi. Berikut adalah fitur-fitur pemanfaatan OLAP yang relevan dalam konteks ini.[4]:

#### 1. Consolidation (Roll-up)

Roll up adalah fitur dalam teknologi OLAP yang berlawanan dengan drill down. Fitur ini memungkinkan pengguna untuk melihat ringkasan atau agregasi data dari tingkat yang lebih rendah ke tingkat yang lebih tinggi dalam hierarki data.

#### 2. Drill-down

Drill down adalah fitur penting dalam teknologi OLAP yang memungkinkan pengguna untuk menjelajahi data dengan lebih mendalam dengan menurunkan level detail dari informasi yang disajikan.

### 3. Slice and dice

Slicing adalah proses memilih satu atau lebih dimensi tertentu dari kubus data dan mengambil subset data yang relevan. Ini memungkinkan pengguna untuk memfokuskan analisis pada area tertentu dari data tanpa perlu memperhatikan dimensi lainnya. Di sisi lain, teknik dice menghasilkan sub-kubus dengan memilih dua atau lebih dimensi dari kubus data yang ada.

## 2.6 Visualisasi

Visualisasi adalah proses mewakili data atau informasi dalam bentuk grafis, seperti grafik, diagram, peta, atau gambar lainnya. Tujuan utama dari visualisasi adalah untuk mengkomunikasikan informasi dengan cara yang mudah dipahami dan menarik secara visual. Metode representasi grafis yang digunakan untuk menggambarkan data atau informasi dapat diterapkan di banyak bidang, termasuk bisnis, ilmu pengetahuan, teknologi, arsitektur, dan seni. Berbagai contoh penggambaran data meliputi bar, grafik, heatmap, pie chart dan lainnya. Penggunaan visualisasi dapat diterapkan untuk menggambarkan berbagai macam jenis data, termasuk data numerik, kualitatif, dan abstrak. Fokus utama dari visualisasi adalah untuk mengklarifikasi pemahaman atas informasi yang rumit atau sulit dipahami melalui representasi visual yang mudah dimengerti. Dengan demikian, visualisasi bermanfaat dalam proses pengambilan keputusan, penjelasan konsep, serta presentasi informasi dengan cara yang menarik dan interaktif.. [9]

Proses awal dalam melakukan penggambaran data adalah mengunduh data asli, kemudian dilanjutkan dengan proses pembersihan data dan pra-pemrosesan untuk menyiapkan data tagar dapat diproses lebih lanjut. Setelah data dipersiapkan, pengguna dapat memilih jenis penggambaran data yang sesuai dengan karakteristik data yang dimilikinya. Untuk mempermudah analisis

data pengguna dapat menambahkan fitur filter dan parameter yang sesuai, yang memungkinkan untuk menyaring dan mengatur data dengan lebih baik. Selain itu, pengguna juga dapat membuat penggambaran data untuk menyajikan hasil analisis secara interaktif. Visualisasi memiliki beberapa tujuan, antara lain:

1. Mengkomunikasikan Informasi dengan Jelas: Visualisasi data membantu menyajikan informasi secara visual, membuatnya lebih mudah dipahami oleh audiens. Ini membantu mengkomunikasikan temuan, tren, dan pola secara efektif..
2. Mengidentifikasi Pola dan Tren: Dengan visualisasi data, pola dan tren yang mungkin tidak terlihat dalam data mentah dapat dengan mudah diidentifikasi. Ini memungkinkan pengambilan keputusan berdasarkan pemahaman yang lebih dalam tentang data.
3. Mendeteksi Anomali: Visualisasi data dapat membantu dalam mendeteksi anomali atau data yang tidak biasa. Ketika data dipresentasikan secara visual, anomali tersebut mungkin lebih mudah terlihat, memungkinkan tindakan yang cepat dan tepat.
4. Meningkatkan Pengambilan Keputusan: Visualisasi data membantu dalam pengambilan keputusan yang lebih baik dan lebih cepat. Dengan menyajikan data secara visual, pemangku kepentingan dapat melihat hubungan antara variabel, memahami dampak keputusan, dan merencanakan tindakan yang sesuai.
5. Peningkatan Penggunaan: Penggambaran data yang sesuai dan informatif dapat meningkatkan partisipasi pengguna. Grafik, diagram, dan visualisasi interaktif dapat membuat pengguna lebih tertarik untuk menjelajahi dan memahami data.
6. Mengungkap Cerita dari Data: Visualisasi data memungkinkan untuk menceritakan cerita dari data dengan lebih efektif. Dengan mengatur data secara visual dalam bentuk naratif, informasi kompleks dapat disampaikan dengan cara yang mudah dipahami.
7. Memfasilitasi Kolaborasi: Visualisasi data memfasilitasi kolaborasi antara berbagai pemangku kepentingan. Dengan melihat data secara visual,

berbagai tim dan departemen dapat bekerja sama untuk memahami masalah yang kompleks dan mencari solusi bersama.

8. Mendorong Inovasi: Visualisasi data dapat menginspirasi inovasi dan pemikiran kreatif. Dengan melihat data dalam berbagai bentuk visual, individu atau tim dapat menemukan pola atau hubungan yang tidak terduga, membuka pintu untuk solusi baru atau pendekatan yang inovatif.
9. Meningkatkan Efisiensi dan Produktivitas: Dengan menyajikan informasi secara visual, visualisasi data dapat meningkatkan efisiensi dan produktivitas. Ini membantu pengguna untuk dengan cepat memahami informasi kunci tanpa harus menyortir melalui data mentah yang kompleks.
10. Membantu Perencanaan dan Strategi: Visualisasi data dapat membantu dalam perencanaan strategis dan pengambilan keputusan jangka panjang. Dengan memahami tren dan pola dari data historis, organisasi dapat membuat strategi yang lebih baik untuk masa depan.

## **2.7 Tableau**

Tableau adalah sebuah aplikasi penggambaran data yang menghasilkan tampilan data yang dinamis dan informatif. Tableau juga memiliki kemampuan untuk mengelola dan menggambarkan data dengan baik, menganalisis data dalam jumlah besar. Penggunaan Tableau umumnya difokuskan pada analisis data yang mendukung proses pengambilan keputusan di berbagai perusahaan.

Tableau adalah alat yang berguna untuk menciptakan visualisasi data dalam format dashboard, memungkinkan data untuk menjadi lebih interaktif dan mudah dianalisis. Proses visualisasi ini melibatkan pengubahan dari data transaksional secara berkala, mengubahnya dari format tabel menjadi grafik atau pemetaan geografis, sehingga memudahkan pembacaan data[10].

Tableau adalah salah satu perangkat lunak Business Intelligence (BI) yang digunakan untuk menciptakan visualisasi data yang lebih interaktif, mudah dipahami, dan mudah dianalisis. Proses visualisasi ini melibatkan mengubah data tabular yang kaku menjadi berbagai bentuk grafik, diagram, peta geografis,

dan lainnya. Tujuannya adalah untuk memvisualisasikan perubahan dan perbedaan dalam data yang telah diolah agar lebih mudah dipahami [11].

Tableau adalah software yang dipakai untuk menganalisis dan menyajikan sekumpulan data dalam bentuk visual yang mendukung proses pengambilan keputusan. Suite perangkat lunak Tableau mencakup Tableau Desktop, Tableau Prep, Tableau Online, dan Tableau Public. Penerapan Tableau ini bertujuan untuk melakukan analisis data yang mendukung pengambilan keputusan. [5]

Tools visualisasi data ini sangat populer karena relatif mudah digunakan dan sangat kuat. Tableau terintegrasi dengan ratusan sumber untuk mengimpor data dan membuat berbagai visualisasi data. Selain itu tools ini juga dapat memvisualisasikan data dari bagan ke peta. Tableau juga tersedia dalam versi publik. Ini berarti siapa pun dapat mengakses dan melihat visualisasi data. Ada juga versi Tableau Desktop yang melindungi visualisasi data dengan kata sandi. Versi tableau desktop ini sangat cocok untuk bisnis dan perusahaan.

Berdasarkan pengertian dari beberapa sumber diatas dapat digaris bawahi bahwa tableau merupakan tools yang berguna untuk melakukan analisis dan visualisasi data. Keunggulan utamanya terletak pada antarmuka yang mudah digunakan dan kemampuan untuk mengintegrasikan berbagai sumber data. Dengan tableau, pengguna dapat dengan cepat mengubah data menjadi visualisasi yang mendukung pengambilan keputusan,. Tableau juga memiliki beberapa versi ada Tableau Desktop, Tableau Prep, Tableau Online, dan Tableau Public sehingga populer dan mudah digunakan dalam analisis data.

## **2.8 Penelitian Terdahulu**

Penelitian yang merujuk pada [12] oleh I Putu Susila Handika, Putu Praba Santika pada tahun 2020 dengan judul “Implementasi Data Warehouse dan Business Intelligence Untuk Data Penjualan Pada PT. ABC” menggabungkan business intilience dengan visualisasi data untuk mendapatkan informasi yang cepat dan akurat. Metode yang digunakan adalah nine steps kimball dengan

mengintegrasikan data transaksi melalui proses ETL sehingga menghasilkan data yang relevan sesuai dengan kebutuhan

Penelitian dengan judul “Analisa Kategori Barang Dengan Penjualan Terbanyak dalam Jangka Waktu 3 Bulan Menggunakan Data Warehouse” oleh I Putu Agus Eka Pratama dan Rey Bernard pada tahun 2021 memanfaatkan data transaksional pada bulan juli hingga september 2020 yang digitalisasikan melalui proses ETL berbasis Pentaho Data Integration. Hasil pengujian menunjukkan bahwa beras merupakan kategori barang dengan penjualan terbanyak pada kurun waktu 3 bulan [8].

Penelitian yang dilakukan oleh I Kadek Anom Sukawirasa, I Gede Acintia Udayana, I Made Yoga Mahendra, Gde Deva Dimastawan Saputra, Ida Bagus Made Mahendra dengan judul “Implementasi Data warehouse Dan Penerapannya Pada PHI-Minimart dengan Menggunakan Tools Pentaho dan Power BI” sumber data yang digunakan adalah data warehouse penjualan, proses ETL dilakukan menggunakan alat bernama Pentaho. Model yang digunakan adalah starschema dengan dimensi cabang, produk, karyawan, dan waktu yang terhubung dengan fact table penjualan. Sumber data yang digunakan adalah data Warehouse penjualan. Proses ETL (Extraction, Transformation, Loading) menggunakan alat Pentaho. Dibuatlah starschema dengan dimensi cabang, produk, karyawan, dan waktu yang terhubung dengan tabel fakta penjualan.. Data Warehouse yang dihasilkan dianalisis menggunakan OLAP, pembuatan kubus OLAP menggunakan schema workbench, dan terakhir pembuatan dashboard. Hasil dari penelitian ini menampilkan dashboard penjualan, yang berisi barang yang terjual dan sub total penjualan tiap cabang [13].

Penelitian yang merujuk pada [4] Oleh Rizki Riksazany dan Mewati Ayub dengan judul “Explorasi Data Warehouse Penjualan dengan Tableau” pada tahun 2019 menggunakan teknologi OLAP yang dibantu dengan software tableau. Implementasi OLAP menggunakan perangkat lunak Tableau, dapat dilakukan dengan pemisahan data dari berbagai jenis menjadi informasi dalam

bentuk tabel dimensi dan tabel ukuran (measures). Hasil dari penelitian ini berupa laporan perbandingan pada tahun 2015 hingga 2018.

**Tabel 2.1** Penelitian Terdahulu

| No | Penulis  | Judul   | Metode Penelitian  | Hasil penelitian  |
|----|--|---|--------------------|---|
| 1  | (I Putu Susila Handika & Putu Praba Santika, 2020) | Implementasi Data Warehouse dan Business Intelligence Untuk Data Penjualan Pada PT. ABC | nine steps kimball | <ol style="list-style-type: none"> <li>1. Sistem dapat menampilkan jumlah dan total penjualan berdasarkan waktu</li> <li>2. Sistem dapat menampilkan jumlah dan total penjualan berdasarkan produk</li> <li>3. Sistem dapat menampilkan jumlah dan total penjualan berdasarkan waktu dan produk</li> <li>4. Sistem dapat menampilkan jumlah dan total penjualan berdasarkan waktu, produk, dan store</li> </ol> |

|   |   |   |              |   |
|---|---|---|--------------|---|
|   |   |   |              | <p>5. Sistem dapat menampilkan jumlah dan total penjualan berdasarkan waktu, departemen, dan produk</p> <p>6. Sistem dapat menampilkan jumlah dan total penjualan berdasarkan waktu, region, dan store</p> <p>7. Sistem dapat menampilkan jumlah dan total penjualan berdasarkan waktu, region, store, departemen, dan produk</p> |
| 2 | (I Putu Agus Eka Pratama dan Rey Bernard, 2021) | Analisa Kategori Barang Dengan Penjualan Terbanyak dalam Jangka Waktu 3 Bulan | ETL dan OLAP | Hasil pengujian menunjukkan bahwa beras merupakan kategori barang dengan penjualan terbanyak  |

|   |   |  |              |   |
|---|---|--|--------------|---|
|   |   | Menggunakan Data Warehouse   |              | pada kurun waktu 3 bulan  |
| 3 | (I Kadek Anom Sukawirasa, I Gede Acintia Udayana, I Made Yoga Mahendra, Gde Deva Dimastawan Saputra, Ida Bagus Made Mahendra, 2021) | Implementasi Data warehouse Dan Penerapannya Pada PHI-Minimart dengan Menggunakan Tools Pentaho dan Power BI | ETL dan OLAP | Dashboard interaktif dengan menampilkan <ul style="list-style-type: none"> <li>a. Total Penjualan dari bulan dan nama cabang</li> <li>b. Total penjualan dari bulan dan nama produk</li> <li>c. Total penjualan dari bulan dan nama propinsi</li> </ul> |
| 4 | (Rizki Riksazany dan Mewati Ayub, 2019 )  | Explorasi Data Warehouse Penjualan dengan Tableau  | ETL dan OLAP | Hasil berupa reports <ul style="list-style-type: none"> <li>a. Perbandingan total penjualan produk pada tahun 2015 hingga 2018</li> <li>b. Perbandingan total penjualan produk berdasarkan beberapa kota pada tahun 2015 hingga 2018</li> </ul>         |

|  |  |  |  |   |
|--|--|--|--|---|
|  |  |  |  | <p>c. Perbandingan beberapa distributor yang melakukan transaksi penjualan pada beberapa kota pada tahun 2015 hingga 2018</p> <p>d. Perbandingan beberapa distributor yang melakukan transaksi penjualan di kota Jakarta pada tahun 2016 hingga 2017</p> <p>e. Perbandingan total penjualan berdasarkan type di beberapa distributor di kota Jakarta pada tahun</p> |
|--|--|--|--|---|

|  |  |  |  |  |
|--|--|--|--|--|
|  |  |  |  | 2016 hingga 2017<br>f. Perbandingan total penjualan berdasarkan produk pada type tertentu pada salah satu distributor di kota Jakarta berdasarkan bulan pada tahun 2017. |
|--|--|--|--|--|

Pada penelitian ini akan akan diimplementasikan teknologi OLAP dan visualisasi data untuk mengetahui tren pasar perusahaan dengan bantuan tools Tableau dan Pentaho data Integration.