

## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Penelitian ini menggunakan beberapa sumber pustaka yang berhubungan dengan kasus atau latar belakang masalah yang akan diteliti. Penulis mendapat literatur dari penelitian terdahulu, sehingga dapat dijadikan acuan untuk dipelajari dan dipahami diantaranya, pada tahun 2022, Jayanta dkk pernah melakukan penelitian perancangan data warehouse dan dashboard di PT. Jaya Teknik. Dalam perancangan tersebut dilakukan dengan tujuan merancang dan menganalisis Data Warehouse yang dibutuhkan dalam penyediaan informasi pada PT. Jaya Teknik. Dalam perancangan tersebut peneliti memanfaatkan implementasi proses data dengan konsep *Extract Transform Load* (ETL) dengan memanfaatkan perangkat lunak *open source* yakni *Pentaho*. Perancangan sistem yang berhasil dilaksanakan pada penelitian tersebut menghasilkan kesimpulan yakni perancangan dan visualisasi dashboard berdasarkan data warehouse tersebut dapat membantu menganalisis peluang bisnis dan mengevaluasi pekerjaan sebelumnya. (Jayanta, dkk, 2022).

Penelitian kedua yang menjadi referensi dalam pembangunan sistem monitoring ini adalah penelitian yang dilakukan oleh Apriandi dan Fatoni. Penelitian tersebut dilakukan pada tahun 2022 dengan mengelola data penjualan obat kemudian memvisualisasinya dalam bentuk dashboard. Proses pengolahan data menggunakan konsep ETL dengan memanfaatkan tools *Pentaho*, sedangkan visualisasinya menggunakan aplikasi *Tableau*.

Berdasarkan hasil yang diperoleh pada penelitian ini, Apriliandi dan Fatoni menyimpulkan bahwa dashboard yang dihasilkan dari proses ETL yang dibangun dapat menyediakan informasi secara cepat dan mudah dipahami dalam pelaporan, serta dapat membantu penganalisisan masalah serta pengambilan kesimpulan. (Apriliandi & Fatoni, 2022) .

Kemudian, penelitian mengenai proses ETL juga pernah dilakukan oleh Praselia dan Kurniawan pada tahun 2021. Dalam penelitian tersebut, peneliti mencoba untuk menganalisis data yang dikumpulkan terkait produk, segment penjualan, transaksi dan sebagainya. Data tersebutlah yang diolah dengan konsep ETL. Proses yang dilakukan setelah terkumpulnya data yang dibutuhkan adalah transformasi data, dan kemudian hasil proses disimpan dalam database *MySQL* dan file *Microsoft Excel*. Pada akhir penelitian ini disimpulkan oleh penyusun bahwa pengolahan data yang dilakukan dengan proses ETL yang memanfaatkan tools Pentaho ini menghasilkan data yang lebih konsisten dan akurat. Penyusun menilai bahwa informasi yang dihasilkan juga lebih berguna bagi perusahaan (Praselia & Kurniawan, 2021).

Literatur anomali data pada penelitian ini diperoleh pada penelitian yang dilakukan oleh Setiawan dan Wibowo pada tahun 2023 dengan judul “Data Mining Implementation For Detection Of Anomalies In Network Traffic Packets Using Outlier Detection Approach”, dimana pada penelitian tersebut dilakukan untuk mendeteksi anomali data atau *outlier* pada data paket lalu lintas jaringan dengan metode memisahkan entitas *outlier* pada data set, kemudian dilakukan validasi dengan memeriksa informasi ahli. Dengan adanya

deteksi ini dapat dikategorikan data yang memiliki anomali dan diberikan label “Error” maupun “Warning” untuk dapat dilakukan investigasi lebih lanjut. (Setiawan & Wibowo, 2023).

Penelitian kelima yang menjadi literatur anomali data adalah penelitian yang dilakukan oleh Fadlilah,dkk tentang Deteksi Anomali Data Akademik Menggunakan DBSCAN Outlier Detection. Anomali data yang dimaksud dalam penelitian ini adalah ketika klaster memiliki data yang sedikit jika dibandingkan dengan klaster yang memiliki data banyak atau klaster yang dianggap normal. Penggunaan DBSCAN pada penelitian tersebut dianggap mampu mengidentifikasi anomali data secara baik (Fadlilah dkk, 2022).

Pada penelitian diatas menjadi literatur dalam pembangunan sistem monitoring anomali data yang menggabungkan metode *Extract, Transform, dan Load* dengan *Apache NIFI* dalam pemrosesan data anomali yang kemudian akan divisualisasikan pada sebuah *Grafana Dashboard*.

Tabel 2.1 merupakan tabel tinjauan pustaka yang dibuat untuk mendefinisikan masing masing penelitian yang telah dilakukan sebelumnya, yang menjadi literatur dalam penelitian ini :

**Table 2.1 Tinjauan Pustaka**

Penulis	Tahun	Overview	Perangkat
Jayanta dkk	2022	Perancangan data warehouse menggunakan metode ETL untuk membantu menganalisis peluang bisnis.	<i>Pentaho, Qlik Sense</i>
Apriliandi & Fatoni	2022	Pengolahan data menggunakan metode ETL dan memvisualisasikannya.	<i>Pentaho, Power BI</i>
Prasetia & Kurniawan	2021	Pengolahan data penjualan dengan metode ETL menghasilkan data yang lebih konsisten.	<i>Pentaho, Microsoft Excel, MySql</i>

Setiawan & Wibowo	2023	Pendeteksian anomali data menggunakan metode pemisahan data yang <i>outlier</i>	-
Fadlillah, Chrisnanto, & Ningsih	2022	Pendeteksian outlier data yang dianggap sebagai anomali dengan menggunakan DBSCAN	<i>DBSCAN</i>
Yang akan diusulkan : Tri Tuter Prawesti	2023	Sistem monitoring yang memanfaatkan metode ETL dan Visualisasi Dashboard untuk mendeteksi anomali data berdasarkan ketersediaan data.	<i>Apache Nifi, Python, Grafana Dashboard, MySql</i>

## 2.2 Landasan Teori

### 2.2.1 Anomali Data

Anomali data merupakan sebuah penyimpangan atau ketidaknormalan yang signifikan yang terdapat dalam sebuah kumpulan data (Kamil, 2023). Anomali data merupakan pola yang menyimpang dari aliran data yang diharapkan ( Tianyuan dan Zhao, 2023 ). Deteksi anomali pada sebuah kumpulan data ini dilakukan dengan tujuan mengidentifikasi data yang memiliki kemungkinan terjadinya *missing* data, runtun waktu yang tidak lengkap serta normalisasi data (Munir dkk, 2021). Penentuan anomali data biasanya mengacu pada pola yang telah ditetapkan, dengan kata lain anomali data berarti data tidak dapat memenuhi data yang dianggap normal ( Tianyuan dan Zhao, 2023 ).

Deteksi anomali ini perlu dilakukan untuk mendukung fungsi data yang akan diolah menjadi sebuah informasi sehingga dapat membantu pengambilan keputusan. Apabila data yang tersedia tidak lengkap atau *invalid* maka informasi yang dihasilkan juga akan berpengaruh. Selain itu anomali data ini juga dapat memberikan indikasi bahwa proses sebelumnya terjadi gangguan teknis atau peluang yang mungkin belum terdeteksi ( Kamil, 2023). Manfaat

lain dari deteksi anomali data adalah penyelesaian masalah serta pengalokasian sistem dapat lebih maksimal, sehingga dapat menghemat biaya ( Amazon, 2023).

Sistem monitoring yang dibangun dalam penelitian ini akan menjadi salah satu sarana pendeteksian anomali data yang mungkin terjadi. Anomali yang dimaksudkan adalah ketersediaan data pada source, sehingga dapat membantu analisis validasi atau kelengkapan data.

### **2.2.2 *Extract, Transfer, Load***

Dalam pengolahan data sering sekali terdapat istilah *Extract, Transfer Load* atau biasa disebut sebagai ETL. ETL dapat didefinisikan sebagai kumpulan proses dalam mempersiapkan data *operational source* ( Jayanta dkk, 2022). Sedangkan menurut Saraswati dan Martini pada tahun 2020, ETL merupakan proses yang bertujuan untuk mengambil dan memproses data dari suatu sumber kepada sumber baru. Dan menurut Oslan dan Kristanto pada tahun 2019 mengatakan bahwa ETL merupakan fungsi integrasi data yang melibatkan penggalian data dari sumber luar dan mengubahnya sesuai dengan kebutuhan bisnis, dan akhirnya diloading pada sebuah gudang data. Sehingga apabila disimpulkan konsep Extract, Transfer, dan Load ini merupakan sebuah proses yang dibangun untuk menggali atau mengambil data dari suatu sumber dan diolah sesuai dengan kebutuhan, kemudian disimpan pada suatu penyimpanan yang dipersiapkan sebagai data *operational source*. Sumber data dari proses

ETL sendiri dapat dari berbagai macam seperti flat file, Excel, bahkan sosial media (Oslan & Kristanto, 2019).

### **2.2.3 Visualisasi Data**

Salah satu cara efektif untuk menyajikan data yang detail menjadi sebuah informasi yang dapat lebih mudah diterima adalah dengan abstraksi data menjadi informasi visual (Syaripul & Bachtiar, 2016). Visualisasi data merupakan penyajian data sehingga dapat lebih menarik dan lebih mudah untuk dipahami, sehingga dapat membantu mempercepat dalam pengambilan keputusan (Irmayani, 2021). Visualisasi data ini diperlukan untuk memberikan pemahaman yang lebih baik bagi penerima data terutama ketika data yang disajikan dalam jumlah besar ( Irmayani, 2021). Visualisasi data dapat dilakukan dengan menyajikan data dalam bentuk grafik, diagram, chart, heatmaps, dan sebagainya.

### **2.2.3 Apache Nifi**

*Apache Nifi* merupakan sebuah platform transfer teknologi yang dikembangkan oleh *National Security Agent (NSA)* pada tahun 2014 yang didasarkan pada perangkat lunak *Niagara Files* (nifi.apache.org , 2024). Apache Nifi merupakan perangkat lunak open source yang dapat digunakan untuk otomasisasi dan pemantauan pergerakan data ( Cakir, A.dkk, 2022). Selain itu Apache Nifi merupakan salah satu dari beragam perangkat lunak

yang dapat digunakan untuk pemrosesan data dengan konsep Extract, Load, dan Transform.

Menurut dokumentasi pada situs resmi Apache Nifi, terdapat 5 konsep utama diantaranya :

#### 1. *Flowfile*

*Flowfile* atau aliran data merupakan objek yang bergerak melalui sistem. Pada aliran data tersebut, *Apache Nifi* membaca *attribute* maupun *key / value* yang dipasangkan.

#### 2. *Flowfile Processor*

*Flowfile processor* merupakan prosesor yang dapat melakukan *extract* data, *routing* data, transformasi data, dan *loading* data pada sistem. Pada *Apache Nifi*, *flowfile processor* tersedia beragam jenis mulai dari *convert* data *csv* menjadi format yang lain, *execute script*, *listen udp* dan sebagainya. *Flowfile processor* membaca konfigurasi serta atribut yang tersedia pada *flowfile*.

#### 3. *Connection*

*Connection* pada *Apache Nifi* ini merupakan jalur yang menjadi penghubung antar *processor* pada sistem. Pada *connection* inilah antrean atau *queue* data yang telah diproses atau yang akan diproses.

#### 4. *Flowfile Controller*

*Flowfile controller* ini merupakan *controller* yang dapat melakukan *maintenance* terhadap koneksi, *threads*, dan alokasi penyimpanan yang digunakan oleh semua proses.

## 5. *Process group*

*Process group* merupakan kumpulan proses pengolahan data yang dibangun dalam *Apache Nifi*, sehingga pemrosesan data dapat dikelompokkan sesuai kebutuhan. Antar *process group* tetap dapat terhubung dengan memanfaatkan *input* dan *output* port.

### 2.2.4 *Grafana Monitoring Dashboard*

*Grafana Monitoring Dashboard* merupakan salah satu perangkat lunak *open source* yang menyediakan visualisasi, *alert on*, dan eksplorasi berdasarkan *query* yang digunakan. Visualisasi data yang mungkin dilakukan pada *Grafana* ini dapat berupa grafik, time series, table, pie chart, dan sebagainya. *Grafana* mendukung banyak sumber data yang dapat digunakan. *Grafana* memungkinkan pengguna untuk menggunakan *plugin framework* sumber data yang beragam mulai dari *SQL/ NoSQL database*, *Prometheus*, *AWS Cloud*, *Google Cloud*, dan sebagainya. *Grafana* juga menyediakan fitur *alert* berdasarkan visualisasi yang ditampilkan. *Alert* yang telah dipasang pada konfigurasi dashboard, dapat dikirimkan melalui berbagai macam aplikasi seperti *Telegram* dan email.