

BAB II

TINJAUAN PUSTAKA

2.1 Studi Sebelumnya

Penelitian yang dilakukan Arsyad [2] pada tahun 2021 membahas integrasi data menggunakan metode ETL (*Extract, Transform, Load*) dengan pendekatan *incremental loading* dalam SSIS (*SQL Server Integration Services*). Pendekatan ini meningkatkan efisiensi sinkronisasi data dengan hanya mengekstraksi perubahan terbaru dari sistem sumber. Hasil penelitian menunjukkan bahwa SSIS dapat mengelola data dari berbagai sumber, namun terdapat kendala seperti variasi struktur data, ketiadaan atribut penanda waktu (*timestamp*), dan absennya indeks pada tabel sumber yang menghambat waktu pemuatan. SSIS menyediakan konektor untuk berbagai *platform* sebagai solusi.

Penelitian yang dilakukan Fanani dkk[3], pada tahun 2025 membahas rekonstruksi arsitektur *database* untuk meningkatkan performa proses *load data* menggunakan metode Pendekatan Rekonstruksi Arsitektur (*Partitioning, Penggunaan Indeks, ETL Pipeline*). Dengan menerapkan teknik *partitioning, indeks, dan pipeline* ETL yang lebih efisien, penelitian ini berhasil meningkatkan kecepatan *load data* hingga 51% dibandingkan arsitektur sebelumnya. Hasil penelitian menunjukkan bahwa metode ini juga mengurangi *downtime* sistem, meningkatkan efisiensi sumber daya melalui *cloud computing*, serta meningkatkan skalabilitas dan kualitas data dengan validasi *real-time* dan *automated data cleansing*.

Penelitian yang dilakukan oleh Kushwaha[4] pada tahun 2022 berfokus pada pembangunan dan optimalisasi *pipeline ETL* untuk memprediksi tingkat *tune-in* dan *churn rate* pengguna. Dalam penelitian ini, proses ETL dikembangkan menggunakan *Spark* dan *Databricks* sebagai alat utama. Hasil penelitian menunjukkan bahwa optimalisasi *pipeline ETL* mampu mempercepat proses pemuatan dan penyegaran data di *data warehouse* secara efisien, sehingga mendukung kebutuhan analisis dan prediksi dalam skala besar.

Penelitian yang dilakukan Ali, dkk[5] pada tahun 2024 membahas integrasi Python dengan Microsoft Excel menggunakan metode *System Development Life Cycle* (SDLC) dan pendekatan *Extract, Transform, Load* (ETL). Proses ETL digunakan untuk mengekstrak data dari berbagai sumber, mentransformasikan data melalui pembersihan dan pengolahan otomatis, serta memuatnya kembali ke Excel untuk analisis yang lebih efisien. Dengan pendekatan ini, pengguna dapat memanfaatkan kemampuan Python dalam lingkungan Excel tanpa memerlukan pengodean. Hasil penelitian ini menunjukkan peningkatan efisiensi dalam menangani data besar dan kompleks.

Penelitian yang dilakukan Nafisa, dkk[6] pada tahun 2022 membahas perancangan *dashboard Business Intelligence* (BI) dengan metode ADDIE menggunakan Microsoft Power BI untuk mendukung pengambilan keputusan terkait UMK di Indonesia. *Dashboard* ini membantu *investor* dan pemerintah dalam merumuskan strategi peningkatan daya saing UMK.

Penelitian yang dilakukan Steven, dkk[7] pada tahun 2021 membahas perancangan sistem *Business Intelligence* (BI) untuk mengelola data penjualan di PT. Suryaplas Intitama menggunakan metode kimball nine step (*Extract Transform, Load* (ETL)) dan *Business Intelligence Roadmap*. Proses ETL dilakukan dengan *Pentaho Data Integration* (PDI), sedangkan visualisasi data menggunakan Microsoft Power BI. Hasil penelitian menunjukkan bahwa *dashboard* yang dikembangkan membantu *stakeholder* dalam menganalisis transaksi penjualan dan mendukung keputusan berbasis data.

Tabel 2. 1 Studi Sebelumnya

No	Peneliti	Tahun	Tujuan	Metode	Hasil
1	Arsyad[2]	2021	memecahkan masalah bisnis yang kompleks dengan menyalin atau mengunduh file, memuat gudang data, membersihkan dan menambang data, dan mengelola data SQL Server	<i>Extract Transform Load</i>	SSIS dapat mengelola data multi-platform secara efisien melalui <i>incremental load</i> , namun terkendala struktur data tidak seragam dan ketiadaan timestamp atau indeks di sumber data.
2	Fanani, dkk[3]	2025	merancang arsitektur aliran data untuk meningkatkan performa <i>load data</i>	Pendekatan Rekonstruksi Arsitektur (<i>Partitioning</i> , Penggunaan Indeks, ETL Pipeline)	Waktu <i>load data</i> meningkat 51%, downtime berkurang, skalabilitas meningkat, <i>error data</i> menurun melalui <i>cleansing</i> otomatis.
3	Kushwaha[4]	2022	Membangun dan mengoptimalkan <i>pipeline ETL</i> untuk prediksi <i>tune-in</i> dan <i>churn rate</i> pengguna	Pengembangan <i>ETL pipeline</i> menggunakan <i>Spark</i> dan <i>Databricks</i>	Proses ETL dapat dioptimalkan untuk memuat dan memperbarui konten <i>data warehouse</i> secara efisien guna keperluan analisis dan prediksi

4	Ali, dkk[5]	2024	menggabungkan bahasa Python dengan Microsoft Excel guna menyederhanakan proses analisis data	<i>System Development Life Cycle - SDLC</i>	Integrasi Python Excel memungkinkan non-programmer melakukan analisis dan otomatisasi data tanpa coding, meningkatkan kemudahan, performa, dan kompatibilitas
5	Nafiisa, dkk [6]	2022	Merancang <i>dashboard</i> berbasis <i>Business Intelligence</i> (BI) menggunakan Microsoft Power BI untuk membantu pengambilan keputusan terkait UMK di Indonesia.	<i>Analysis, Design, Development, Implementation, Evaluation (ADDIE)</i>	<i>Dashboard</i> visualisasi yang menampilkan data jumlah usaha di sektor <i>green economy</i> , seluruh sektor, jumlah pekerja, pendapatan, dan pengeluaran UMK
6	Steven, dkk[7]	2021	Merancang aplikasi <i>Business Intelligence</i> dengan <i>dashboard</i> visualisasi data untuk membantu memantau transaksi penjualan di PT. Suryaplas Intitama,	<i>kimball nine step (Extract Transform, Load (ETL)) dan Business Intelligence Roadmap</i>	<i>Dashboard</i> visualisasi data menggunakan Microsoft Power BI

2.2 Dasar Teori

2.2.1 Extract, Transform, Load (ETL)

ETL (*Extract, Transform, Load*) adalah proses dalam pengolahan data yang terdiri dari tiga tahap utama:

- *Extract*: Mengambil data dari berbagai sumber, seperti *database, file excel, API, atau data warehouse* lainnya[8].
- *Transform*: Membersihkan, mengubah, dan menggabungkan data agar sesuai dengan kebutuhan analisis dan penyimpanan[9][10].
- *Load*: Memasukkan data yang sudah diproses ke dalam sistem tujuan, seperti *database* terpusat atau *platform* visualisasi data [11][12][13].

ETL sering digunakan dalam sistem *Business Intelligence* (BI) dan *data warehousing* untuk memastikan bahwa data yang digunakan dalam analisis selalu terkini dan akurat[14][15][16][17].

2.2.2 Full Refresh

Full refresh merupakan metode pemuatan ulang seluruh data pada sistem target dengan data terbaru dari sumber. Dalam pendekatan ini, setiap baris dan kolom akan diambil kembali secara menyeluruh tanpa melakukan pengecekan perubahan data. Karena proses ini menghapus seluruh data lama sebelum memuat ulang data baru, metode ini juga dikenal sebagai *destructive load*[18].

2.2.3 Visualisasi Data dengan Power BI

Power BI adalah alat *Business Intelligence* (BI) yang memungkinkan pengguna untuk membuat *dashboard* interaktif dan laporan berbasis data[19]. Beberapa fitur utama Power BI meliputi:

- *Data Importing*: Kemampuan untuk mengambil data dari berbagai sumber, termasuk Excel, *database* SQL, dan API.
- *Data Transforming*: Menggunakan *Power Query* untuk membersihkan dan menggabungkan data.

- *Data Visualization* Menampilkan data dalam bentuk grafik, tabel, atau, indikator KPI untuk mendukung pengambilan keputusan[14][20].

Dengan integrasi Python dan Power BI, data yang telah diproses secara otomatis dapat di visualisasikan dalam bentuk *dashboard* yang intuitif, sehingga memudahkan pemantauan kinerja operasional di salah satu divisi PT XYZ.

