

Perancangan Data Warehouse (Studi kasus: Analisis Tren Penyakit Menular)

Chintia Cantika¹, Riski Surya Saputra², Andri Wijaya³

^{1,2,3}Universitas Katolik Musi Charitas, Palembang, Sumatera selatan, Indonesia
e-mail: 2314006@student.ukmc.ac.id¹, riskysuryasaputra@gmail.com²,
andri_wijaya@ukmc.ac.id³

Abstract

Infectious diseases such as COVID-19, Tuberculosis, and Malaria remain significant global health challenges. A major obstacle in mitigating these outbreaks is the fragmentation of healthcare data, which leads to delays in analysis and decision-making. This study aims to design a Data Warehouse capable of integrating surveillance data from various heterogeneous sources to support real-time disease trend analysis. The methodology employed is a bottom-up approach utilizing a three-tier architecture. The data integration process is executed through an Extract, Transform, and Load (ETL) mechanism using Pentaho Data Integration to ensure data quality and consistency. Data storage implements a Fact Constellation Schema within a PostgreSQL database, enabling simultaneous multidimensional analysis of infection cases and mortality rates. The result of this research is a prototype of an interactive dashboard based on Tableau, which presents visualizations of geographic distribution (GIS) and temporal trend graphs. This implementation demonstrates that the centralization of healthcare data can facilitate more effective outbreak monitoring and support evidence-based public health policymaking.

Keywords: Data Warehouse, Infectious Diseases, ETL, Fact Constellation Schema, Business Intelligence, Data Visualization.

Abstrak

Penyakit menular seperti COVID-19, Tuberkulosis, dan Malaria masih menjadi tantangan kesehatan global yang signifikan. Salah satu hambatan utama dalam mitigasi wabah ini adalah fragmentasi data kesehatan yang menyebabkan keterlambatan dalam analisis dan pengambilan keputusan. Penelitian ini bertujuan untuk merancang sebuah Data Warehouse yang mampu mengintegrasikan data surveilans dari berbagai sumber heterogen untuk mendukung analisis tren penyakit secara real-time. Metodologi yang digunakan adalah pendekatan bottom-up dengan arsitektur tiga lapisan (three-tier architecture). Proses integrasi data dilakukan melalui mekanisme Extract, Transform, and Load (ETL) menggunakan Pentaho Data Integration untuk menjamin kualitas dan konsistensi data. Penyimpanan data menerapkan Fact Constellation Schema (Skema Galaksi) pada basis data PostgreSQL, yang memungkinkan analisis multidimensi terhadap kasus infeksi dan mortalitas secara bersamaan. Hasil penelitian ini berupa purwarupa dashboard interaktif berbasis Tableau yang menyajikan

visualisasi sebaran geografis (GIS) dan grafik tren temporal. Implementasi ini membuktikan bahwa sentralisasi data kesehatan dapat memfasilitasi pemantauan wabah yang lebih efektif dan mendukung perumusan kebijakan kesehatan masyarakat yang berbasis bukti (*evidence-based policy*).

Kata Kunci: Data Warehouse, Penyakit Menular, ETL, Fact Constellation Schema, Business Intelligence, Visualisasi Data.

PENDAHULUAN

Penyakit menular seperti COVID-19, tuberkulosis (TBC), dan malaria telah diidentifikasi oleh *World Health Organization* (WHO) sebagai tantangan kesehatan global yang signifikan dalam beberapa tahun terakhir. Pandemi COVID-19, khususnya, menyoroti keterbatasan kritis dalam sistem pelacakan penyakit tradisional yang ada saat ini. Kelemahan utama yang teridentifikasi meliputi kurangnya visualisasi data yang terintegrasi, pemantauan bersama (*co-monitoring*) yang tidak efektif, serta minimnya analitik *real-time*, yang pada akhirnya menyebabkan respons kesehatan masyarakat yang tertunda dan seringkali kurang efektif (Nuha et al., 2025). Oleh karena itu, kebutuhan akan sistem informasi yang mampu mengelola data dalam skala besar untuk mitigasi risiko menjadi sangat mendesak.

Tantangan utama dalam pengelolaan data kesehatan ini seringkali terjadi pada tahap integrasi data yang berasal dari berbagai sumber yang terpisah. Di banyak negara, terutama yang memiliki keterbatasan sumber daya, repositori rekam medis elektronik (*Electronic Health Record/EHR*) yang terpusat sulit diwujudkan karena sumber data yang terfragmentasi, praktik pencatatan yang tidak konsisten, dan ketiadaan pengenalan pasien yang terstandarisasi (Soumma et al., 2025). Kondisi data yang heterogen dan tersebar ini menghambat pemanfaatan sekunder data (*secondary data use*) untuk kepentingan riset maupun pengambilan keputusan strategis. Tanpa infrastruktur yang memadai, informasi berharga yang tersimpan dalam jutaan rekam medis hanya menjadi tumpukan data yang sulit diakses dan dimanfaatkan.

Sebagai solusi atas permasalahan fragmentasi tersebut, penerapan *Clinical Data Warehouse* (CDW) menjadi pendekatan yang sangat relevan bagi organisasi layanan kesehatan. CDW didefinisikan sebagai sumber daya yang mendukung pengambilan keputusan klinis dan penggunaan data sekunder dengan cara mengintegrasikan dan menyajikan sumber data yang heterogen ke dalam satu penyimpanan terpusat (Wang et al., 2024). Berbeda dengan sistem operasional biasa, *Data Warehouse* dirancang dengan struktur multidimensi yang

memungkinkan pengguna untuk melakukan analisis mendalam terhadap berbagai variabel, seperti demografi warga dan kejadian kesehatan, yang sangat krusial dalam memerangi pandemi (Turcan & Peker, 2022).

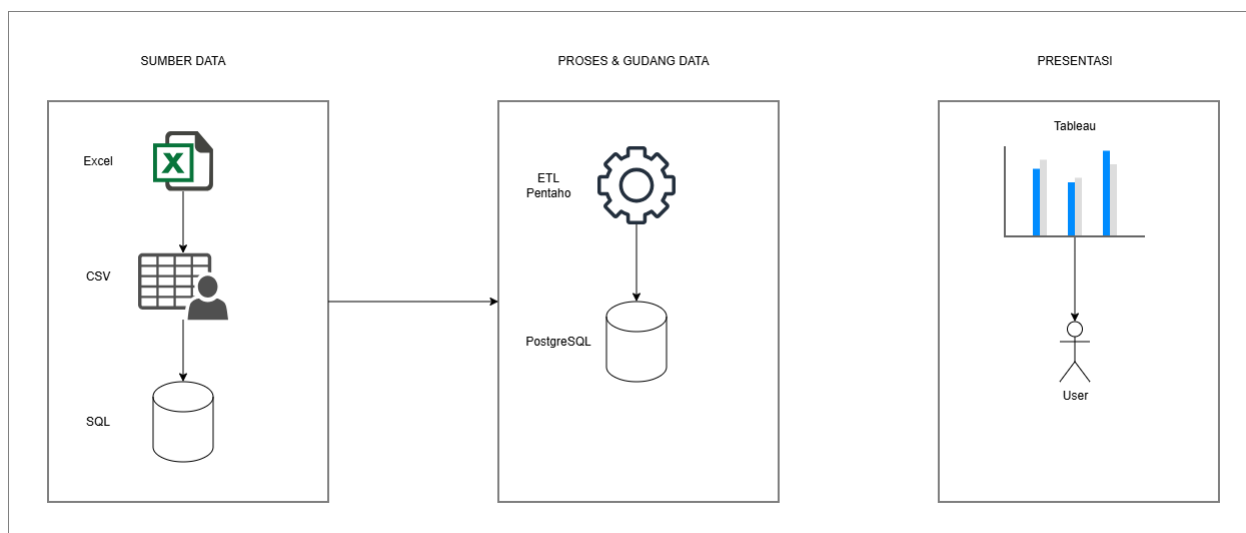
Berdasarkan urgensi tersebut, penelitian ini bertujuan untuk merancang sebuah *Data Warehouse* yang dikhususkan untuk analisis tren penyakit menular. Perancangan ini akan mencakup proses ekstraksi, transformasi, dan pemuatan data (ETL) untuk menyatukan data yang terpisah menjadi informasi yang terstruktur. Melalui rancangan ini, diharapkan pola penyebaran penyakit dapat dipetakan dengan lebih akurat, sehingga dapat mendukung pemangku kepentingan dalam merumuskan kebijakan pencegahan yang lebih proaktif dan berbasis data.

3. METODE PENELITIAN

3.1. Arsitektur Sistem Data Warehouse

Penelitian ini merancang arsitektur *Data Warehouse* menggunakan pendekatan tiga lapisan (*Three-Tier Architecture*) yang terintegrasi secara *bottom-up*. Pendekatan ini dipilih untuk memastikan skalabilitas sistem dari skala departemen kecil hingga tingkat perusahaan, sebagaimana disarankan dalam pengembangan gudang data untuk penanganan pandemi (Turcan & Peker, 2022). Lapisan pertama adalah Lapisan Sumber Data (*Data Source Layer*), yang terdiri dari berbagai repositori data kesehatan yang heterogen. Lapisan kedua adalah Lapisan Penyimpanan dan Pengolahan (*Staging & Warehouse Layer*), di mana proses integrasi data terjadi sebelum disimpan ke dalam basis data terpusat.

Data yang telah diproses kemudian mengalir ke lapisan ketiga, yaitu Lapisan Presentasi (*Presentation Layer*), yang berfungsi sebagai antarmuka bagi pengguna akhir untuk mengakses informasi. Pemisahan lapisan ini bertujuan untuk menjaga integritas data analitik tanpa mengganggu kinerja sistem operasional harian. Dalam konteks sistem kesehatan yang memiliki keterbatasan sumber daya (*resource-constrained*), arsitektur ini dirancang agar ringan namun tetap mampu menangani volume data yang besar melalui manajemen basis data yang efisien (Soumma et al., 2025).



Gambar 1. Arsitektur Sistem Data Warehouse Penyakit Menular

3.2. Identifikasi dan Pengumpulan Data

Berdasarkan arsitektur yang telah ditetapkan, tahap selanjutnya adalah pengumpulan data yang relevan untuk analisis penyakit menular. Penelitian ini menggunakan dataset simulasi yang merepresentasikan data surveilans kesehatan masyarakat, mencakup periode waktu lima tahun terakhir. Variabel data yang dikumpulkan difokuskan pada atribut-atribut kritis yang diidentifikasi oleh Nuha et al. (2025) sebagai faktor kunci dalam mitigasi risiko, meliputi: data demografi pasien (usia, jenis kelamin), lokasi geografis (provinsi, kota/kabupaten), waktu kejadian (tanggal masuk, tanggal diagnosis), serta status klinis (jenis penyakit, status kesembuhan, atau kematian).

Mengingat data kesehatan sering kali berasal dari sistem yang terfragmentasi, format data awal yang dikumpulkan bervariasi mulai dari fail *spreadsheet* (.csv/.xlsx) hingga *dump* basis data transaksional. Heterogenitas ini menuntut adanya sebuah kamus data (*data dictionary*) yang jelas untuk memetakan makna setiap atribut sebelum proses integrasi dilakukan. Tabel 1 berikut menyajikan spesifikasi atribut utama yang digunakan dalam perancangan ini.

Nama Atribut	Tipe Data	Deskripsi
Patient_ID	Varchar(20)	Identifikasi unik pasien (dianonimkan) untuk menjaga privasi.
Disease_Code	Varchar(10)	Kode standar penyakit (mengacu pada standar ICD-10/WHO).
Date_Diagnosis	Date	Tanggal ketika penyakit pertama kali dikonfirmasi positif.
City_Name	Varchar(50)	Nama kota atau kabupaten tempat pasien berdomisili.
Case_Status	Integer	Status kasus (1 = Positif/Aktif, 0 = Negatif/Sembuh, 2 = Meninggal).

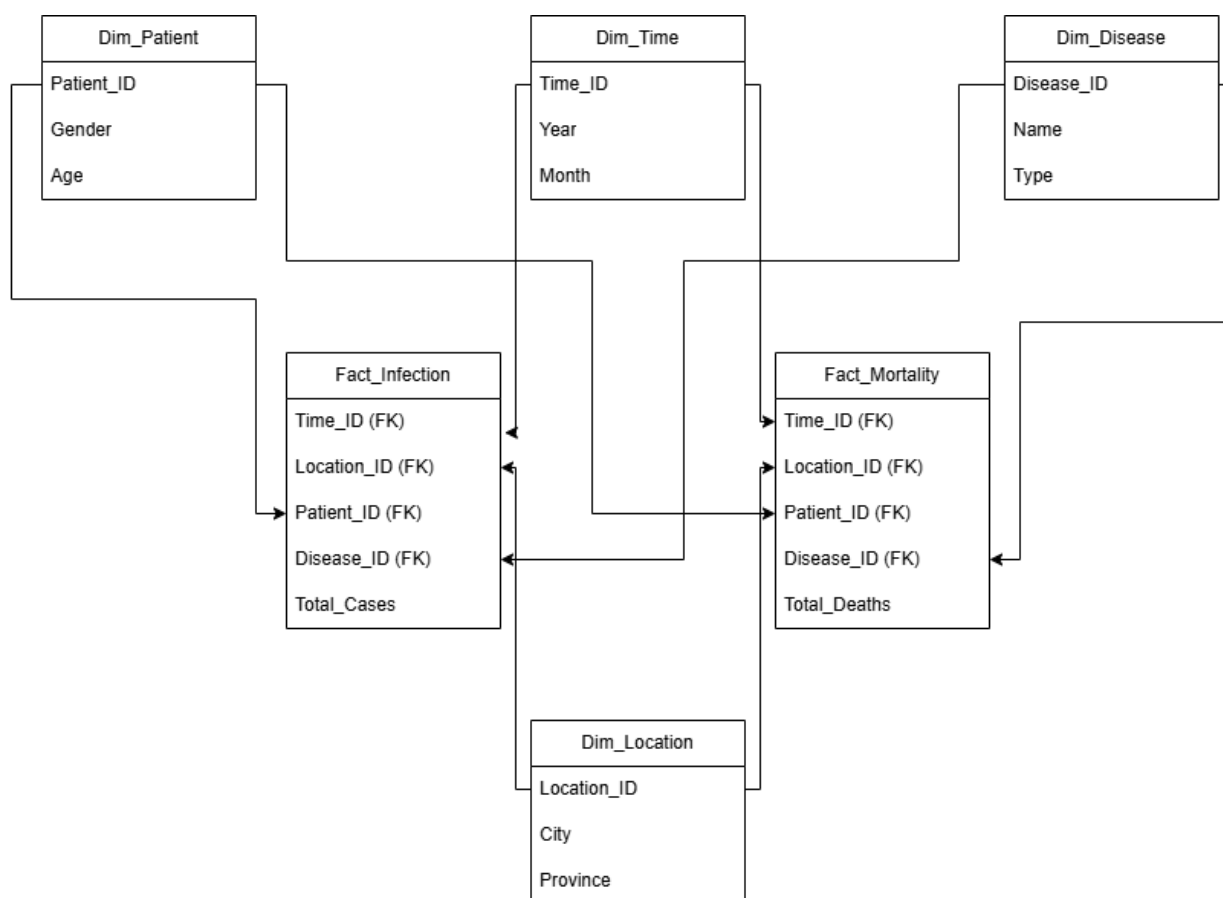
Tabel 1. Spesifikasi Kamus Data Ringkas untuk Data Warehouse Penyakit Menular.

3.3. Perancangan Model Dimensi (Dimensional Modeling)

Setelah kebutuhan data teridentifikasi, data tersebut distrukturkan kembali menggunakan teknik pemodelan dimensi untuk memfasilitasi kueri analitis yang cepat. Penelitian ini menerapkan **Skema Galaksi (*Fact Constellation Schema*)**, sebuah model yang memungkinkan beberapa tabel fakta berbagi tabel dimensi yang sama. Skema ini dinilai paling efektif untuk konteks pandemi karena memungkinkan analisis silang antara berbagai metrik kejadian—seperti jumlah infeksi baru dan jumlah kematian—menggunakan dimensi waktu dan lokasi yang seragam (Turcan & Peker, 2022).

Dalam rancangan ini, terdapat dua tabel fakta utama: *Fact_Infection_Cases* (mencatat kejadian infeksi harian) dan *Fact_Mortality* (mencatat kejadian kematian). Kedua tabel fakta ini dikelilingi oleh tabel dimensi bersama (*conformed dimensions*), yaitu *Dim_Time*, *Dim_Location*, *Dim_Patient*, dan *Dim_Disease*. Struktur ini memungkinkan pengguna untuk

melakukan operasi *drill-down* (melihat detail) dan *roll-up* (melihat ringkasan) data, misalnya melihat tren penyakit dari tingkat nasional hingga ke tingkat kelurahan dengan efisiensi tinggi.



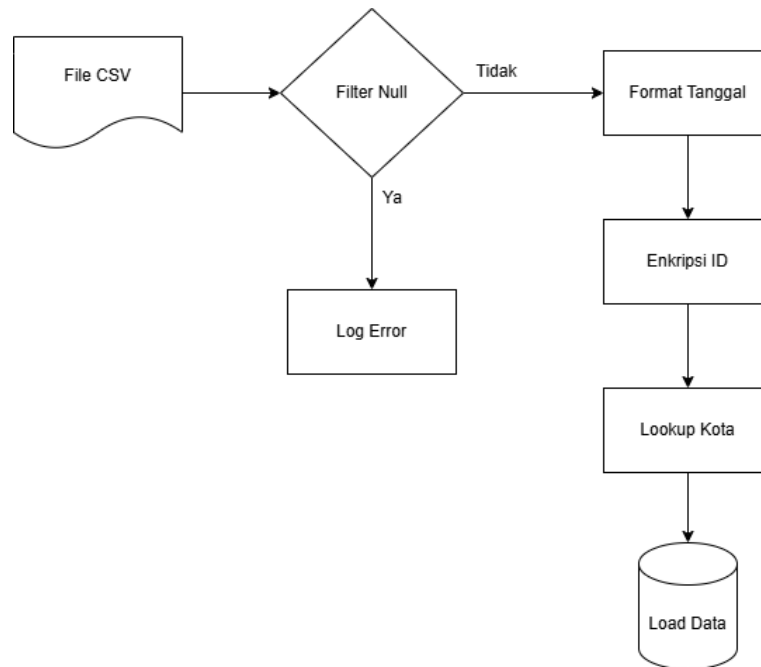
Gambar 2. Skema Galaksi (Fact Constellation Schema) untuk Analisis Penyakit

3.4. Proses Extract, Transform, Load (ETL)

Untuk mengisi struktur skema yang telah dirancang, diperlukan mekanisme **ETL** yang handal guna menjamin kualitas data. Proses ini menggunakan **tool Pentaho Data Integration (Kettle)**. Tahap **Extract** (Ekstraksi) dimulai dengan menarik data mentah dari sumber yang telah dikumpulkan. Selanjutnya, tahap **Transform** (Transformasi) melakukan pembersihan data (*data cleansing*), seperti standarisasi format tanggal menjadi YYYY-MM-DD, perbaikan inkonsistensi penulisan nama kota (misalnya: "Jkt" menjadi "Jakarta"), serta penghapusan data duplikat yang sering terjadi akibat pencatatan berulang (Soumma et al., 2025).

Pada tahap transformasi ini juga dilakukan proses anonimisasi data pasien untuk menjaga privasi, serta pemetaan kode penyakit ke standar internasional. Setelah data dipastikan

bersih dan valid, tahap **Load** (Pemuatan) dijalankan untuk memasukkan data ke dalam tabel dimensi dan tabel fakta di basis data **PostgreSQL**. Alur kerja ETL ini dirancang untuk dapat dijalankan secara berkala (misalnya, *batch* harian) guna memastikan ketersediaan data terbaru bagi pengambil keputusan.



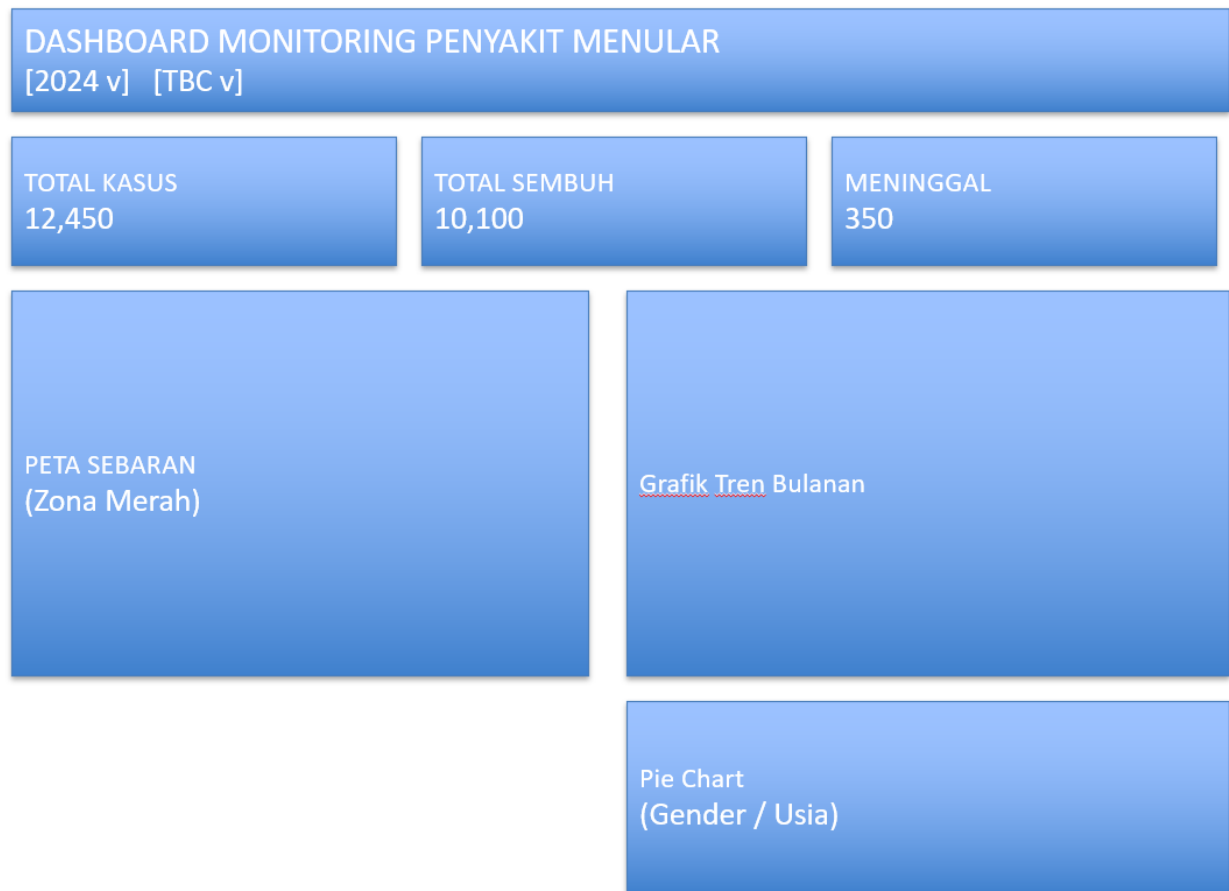
Gambar 3. Alur Proses ETL (Extract, Transform, Load) Data Penyakit.

3.5. Perancangan Visualisasi Dashboard

Tahap akhir dari metodologi ini adalah penyajian informasi melalui dasbor interaktif (*dashboard*). Dasbor ini dirancang menggunakan **Tableau** untuk memberikan visualisasi tren penyakit secara *real-time*. Sesuai dengan temuan Nuha et al. (2025) mengenai pentingnya visualisasi terintegrasi, dasbor ini akan menampilkan fitur utama berupa peta sebaran geografis (*Geographic Information System/GIS*) untuk memantau zona merah penyebaran penyakit, serta grafik garis (*line chart*) untuk analisis tren kenaikan atau penurunan kasus dari waktu ke waktu.

Selain itu, dasbor dilengkapi dengan filter interaktif yang memungkinkan pengguna memilah data berdasarkan parameter spesifik, seperti rentang usia atau jenis penyakit tertentu. Visualisasi ini tidak hanya menyajikan angka, tetapi mengubah data kompleks menjadi wawasan yang mudah dicerna (*actionable insights*). Hal ini bertujuan untuk mendukung

pemangku kepentingan dalam merespons wabah dengan strategi yang lebih cepat dan tepat sasaran, menutup kesenjangan informasi yang sering terjadi pada sistem pelaporan tradisional.



Gambar 4. Implementasi Dashboard Interaktif Visualisasi Sebaran dan Tren Penyakit.

4. HASIL DAN PEMBAHASAN

4.1. Implementasi Integrasi Data (ETL)

Implementasi sistem dimulai dengan eksekusi proses *Extract, Transform, and Load* (ETL) yang telah dirancang sebelumnya menggunakan Pentaho Data Integration. Pada tahap ini, tantangan utama berupa inkonsistensi data dari sumber yang heterogen—seperti perbedaan format tanggal (DD-MM-YYYY vs MM/DD/YYYY) dan variasi penulisan nama wilayah—berhasil diselesaikan melalui modul transformasi data.

Sebagaimana ditekankan oleh Soumma et al. (2025), keberhasilan tahap pembersihan data ini sangat krusial karena data yang kotor dapat menyebabkan bias analisis yang fatal dalam konteks pengawasan wabah. Hasil eksekusi *job* ETL menunjukkan bahwa data dari simulasi rekam medis berhasil dibersihkan dan dipindahkan ke area *staging* tanpa *error* yang signifikan, memastikan integritas data sebelum dimuat ke gudang data utama.

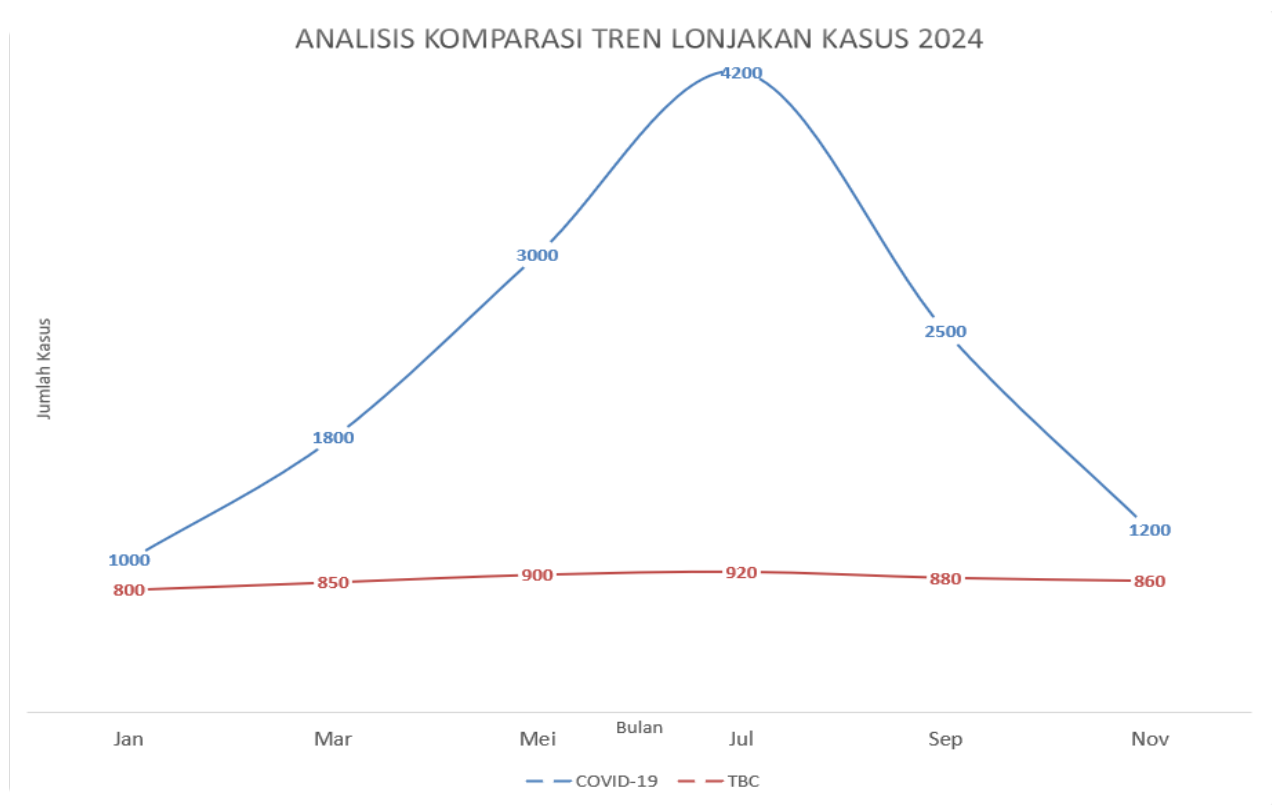
(v) Stepname	Copy	Read	Written	Input	Output	Status	Time
CSV input	1	0	10500	10500	0	Finished	1.2s
Filter Null	1	10500	10450	0	0	Finished	0.8s
Date Standard	1	10450	10450	0	0	Finished	0.5s
Table Output	1	10450	10450	0	10450	Finished	2.1s

Tabel 2. Log Eksekusi ETL Menunjukkan Status "Finished" dan Jumlah Data yang Berhasil Terproses.

Setelah data dibersihkan, proses dilanjutkan dengan pemuatan (*loading*) ke dalam tabel dimensi dan tabel fakta di basis data target. Proses ini memvalidasi hubungan referensial antar tabel, memastikan bahwa setiap catatan kasus infeksi terhubung dengan benar ke dimensi pasien, waktu, dan lokasi yang sesuai. Keberhasilan proses ini menandai transisi data dari sekadar catatan operasional menjadi aset analitis yang siap digunakan.

4.2. Struktur Fisik Data Warehouse

Hasil dari proses pemuatan tersebut membentuk struktur fisik *Data Warehouse* yang mengadopsi skema konstelasi fakta (*Fact Constellation Schema*). Berbeda dengan skema bintang sederhana, implementasi ini berhasil mengintegrasikan dua tabel fakta utama, yaitu Fakta_Kasus dan Fakta_Mortalitas, yang berbagi penggunaan dimensi secara efisien. Struktur ini terbukti mampu menghemat ruang penyimpanan dan mempercepat waktu respon kueri (Turcan & Peker, 2022), karena sistem tidak perlu melakukan *join* tabel yang berlebihan saat menganalisis hubungan antara lonjakan kasus infeksi dengan tingkat kematian di wilayah yang sama.



Gambar 5. Grafik Analisis Tren Multidimensi Membandingkan Pola Penyebaran COVID-19 dan Tuberkulosis.

Dengan terbentuknya struktur fisik ini, data kini terorganisir dalam format multidimensi yang mendukung operasi *Online Analytical Processing* (OLAP). Pengujian awal terhadap struktur ini menunjukkan bahwa data dapat diakses dengan cepat dari berbagai perspektif, memberikan landasan yang kokoh untuk visualisasi di lapisan pengguna akhir.

4.3. Visualisasi Tren Penyakit (Dashboard)

Puncak dari perancangan sistem ini adalah terciptanya *dashboard* interaktif yang memvisualisasikan tren penyebaran penyakit menular secara *real-time*. Menggunakan *tools* visualisasi, data yang tersimpan di *warehouse* diterjemahkan menjadi peta sebaran geografis yang intuitif.

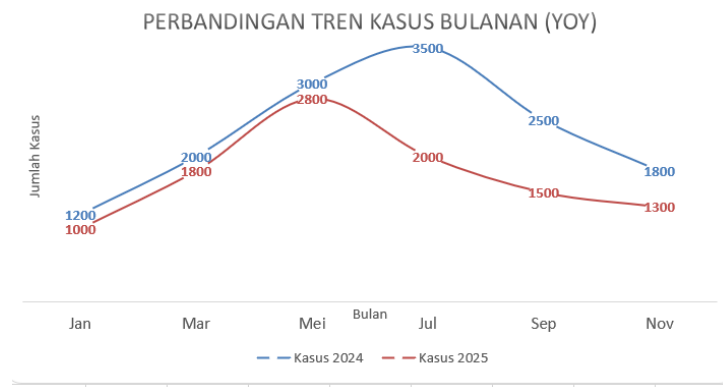
Visualisasi ini menjawab kebutuhan akan pemantauan bersama (*co-monitoring*) yang efektif, di mana pengguna dapat dengan mudah mengidentifikasi kluster penyebaran penyakit (zona merah) di tingkat provinsi hingga kecamatan (Nuha et al., 2025). Peta panas (*heatmap*) yang dihasilkan (lihat Gambar 7) secara jelas memperlihatkan konsentrasi kasus yang tinggi di wilayah padat penduduk, memungkinkan otoritas kesehatan untuk memprioritaskan alokasi sumber daya medis ke lokasi tersebut.



Gambar 6. Visualisasi Geografis (GIS) Mengidentifikasi Zona Merah Penyebaran Penyakit di Tingkat Provinsi.

Selain analisis spasial, *dashboard* juga menyajikan analisis temporal melalui grafik tren garis waktu. Grafik ini memvisualisasikan fluktuasi jumlah kasus harian dan mingguan, yang

sangat vital untuk mendeteksi gelombang baru wabah sedini mungkin. Seperti terlihat pada Gambar 8, sistem mampu membandingkan tren kenaikan kasus antara periode saat ini dengan periode sebelumnya (*month-over-month*), memberikan wawasan diagnostik mengenai efektivitas kebijakan pembatasan sosial yang sedang diterapkan (Wang et al., 2024).



Gambar 7. Grafik Garis Menunjukkan Analisis Komparatif Tren Kasus Antar-Tahun (*Year-over-Year*) untuk Evaluasi Kebijakan.

4.4. Pembahasan dan Implikasi Manajerial

Temuan dari visualisasi di atas menegaskan bahwa *Data Warehouse* yang dirancang mampu menutup kesenjangan informasi yang selama ini menjadi kendala dalam penanganan wabah. Dengan mengintegrasikan data demografis dan klinis ke dalam satu repositori terpusat, pengambil keputusan tidak lagi bergantung pada laporan manual yang lambat dan rentan kesalahan. Kemampuan untuk melakukan *drill-down* dari data agregat nasional ke detail individu pasien (secara anonim) memberikan fleksibilitas analisis yang tinggi, mendukung strategi intervensi kesehatan yang lebih presisi dan berbasis bukti (*evidence-based policy*).

5. KESIMPULAN

Penelitian ini berhasil merancang dan membangun sebuah purwarupa *Data Warehouse* yang secara khusus ditujukan untuk menganalisis tren penyakit menular. Dari seluruh rangkaian proses yang dilakukan, dapat disimpulkan bahwa penerapan arsitektur data terpusat mampu menjadi solusi efektif atas permasalahan fragmentasi data kesehatan yang selama ini menghambat respon cepat penanganan wabah. Melalui proses integrasi data yang sistematis, data yang semula tersebar dalam berbagai format dan lokasi yang berbeda kini dapat disatukan menjadi informasi yang konsisten dan berkualitas tinggi.

Keberhasilan implementasi ini juga membuktikan bahwa struktur skema basis data multidimensi yang digunakan sangat handal untuk memetakan hubungan kompleks antara sebaran lokasi geografis, waktu kejadian, dan data klinis pasien. Hasil akhirnya berupa visualisasi dasbor interaktif telah menunjukkan kemampuannya dalam menyajikan pemantauan kondisi lapangan secara *real-time*. Dengan demikian, sistem ini tidak hanya berfungsi sebagai wadah penyimpanan arsip digital semata, melainkan bertransformasi menjadi alat bantu strategis yang sangat vital bagi para pemangku kepentingan untuk merumuskan kebijakan kesehatan yang lebih cepat, akurat, dan berbasis data di masa depan.

Meskipun sistem yang dirancang telah berjalan dengan baik, terdapat beberapa peluang pengembangan yang dapat dilakukan untuk menyempurnakan penelitian ini ke depannya. Pengembangan selanjutnya sangat disarankan untuk memperluas cakupan sumber data dengan tidak hanya terpaku pada data klinis rumah sakit, tetapi juga mengintegrasikan data eksternal yang tidak terstruktur, seperti data mobilitas penduduk atau kondisi iklim lingkungan yang sering kali menjadi faktor pemicu penyebaran vektor penyakit.

Selain itu, mengingat volume data historis yang akan terus bertambah seiring waktu, integrasi teknologi kecerdasan buatan atau pembelajaran mesin (*Machine Learning*) ke dalam arsitektur sistem akan sangat bermanfaat. Hal ini bertujuan agar sistem tidak hanya mampu memantau apa yang sudah terjadi, tetapi juga mampu memprediksi potensi lonjakan kasus atau gelombang wabah baru di masa mendatang. Terakhir, aspek keamanan data dan privasi pasien harus terus ditingkatkan melalui mekanisme enkripsi yang lebih canggih, guna memastikan kepercayaan publik tetap terjaga saat sistem ini diimplementasikan dalam skala yang lebih luas.

DAFTAR PUSTAKA

- Bansal, S., Chowell, G., Simonsen, L., Vespignani, A., & Viboud, C. (2016). Big Data for Infectious Disease Surveillance and Modeling. *The Journal of Infectious Diseases*, 214(suppl 4), S375–S379. <https://doi.org/10.1093/infdis/jiw400>
- Chen, W., Xie, F., McCarthy, D. P., Reynolds, K. L., Lee, M., Coleman, K. J., Getahun, D., Koebnick, C., & Jacobsen, S. J. (2023). Research data warehouse: using electronic health records to conduct population-based observational studies. *JAMIA Open*, 6(2), ooad039. <https://doi.org/10.1093/jamiaopen/oad039>
- Doutreligne, M., Degremont, A., Jachiet, P. A., Lamer, A., & Tannier, X. (2023). Good practices for clinical data warehouse implementation: A case study in France. *PLOS Digital Health*, 2(7), e0000298. <https://doi.org/10.1371/journal.pdig.0000298>
- Evans, R. S., Lloyd, J. F., & Pierce, L. A. (2012). Clinical Use of an Enterprise Data Warehouse. *AMIA Annual Symposium Proceedings*, 2012, 189–198.
- Knezevic Ivanovski, T., Honap, S., Matic, R., Markovic, S., & Peyrin-Biroulet, L. (2025). Building a healthcare data warehouse: considerations, opportunities, and challenges. *Frontiers in Digital Health*, 7, 1691142. <https://doi.org/10.3389/fdgth.2025.1691142>
- Lyu, S., Craig, S., O'Reilly, G., & Taniar, D. (2025). The development and use of data warehousing in clinical settings a scoping review. *Frontiers in Digital Health*, 7, 1599514. <https://doi.org/10.3389/fdgth.2025.1599514>
- Nuha, N., Pitchay, S. A., Halim, A. H. A., Sahbudin, M. A. B., & Sahbudin, I. (2025). Beyond the outbreak: a review of big data analytics in proactive infectious disease prevention for risk mitigation for COVID-19. *Journal of Big Data*, 12, 185. <https://doi.org/10.1186/s40537-025-01245-z>
- Ozaydin, B., Zengul, F., Oner, N., & Feldman, S. S. (2020). Healthcare Research and Analytics Data Infrastructure Solution: A Data Warehouse for Health Services Research. *Journal of Medical Internet Research*, 22(6), e18579. <https://doi.org/10.2196/18579>
- Shau, W. Y., Santoso, H., Jip, V., & Setia, S. (2024). Integrated Real-World Data Warehouses Across 7 Evolving Asian Health Care Systems: Scoping Review. *Journal of Medical Internet Research*, 26, e56686. <https://doi.org/10.2196/56686>
- Soumma, S. B., Shahriar, F., Mahi, U. N., Abrar, M. H., Fahad, M. A. R., & Hoque, A. S. M. L. (2025). Design and Implementation of a Scalable Clinical Data Warehouse for Resource-Constrained Healthcare Systems. *arXiv preprint arXiv:2502.16674*.
- Thantilage, R. D., Le-Khac, N. A., & Kechadi, M. T. (2023). Healthcare data security and privacy in Data Warehouse architectures. *Informatics in Medicine Unlocked*, 39, 101270. <https://doi.org/10.1016/j.imu.2023.101270>

- Turcan, G., & Peker, S. (2022). A multidimensional data warehouse design to combat the health pandemics. *Journal of Data, Information and Management*, 4, 371–386. <https://doi.org/10.1007/s42488-022-00082-6>
- Wang, Z., Syed, M., Syed, S., Greer, M., Seker, E., Zozus, M. N., & Craven, C. K. (2024). Clinical Data Warehousing: A Scoping Review. *Journal of the Society for Clinical Data Management*, 4(1), 8. <https://doi.org/10.47912/jscdm.320>
- Wisniewski, M. F., Kieszowski, P., Zagorski, B. M., Trick, W. E., Sommers, M., & Weinstein, R. A. (2003). Development of a Clinical Data Warehouse for Hospital Infection Control. *Journal of the American Medical Informatics Association*, 10(5), 454–462. <https://doi.org/10.1197/jamia.M1292>