

Data Integration Approaches for AI-Ready Data Architecture: A Comparative Review of Relational Databases, Data Warehouses, and Knowledge Graphs

Kodrat Mahatma^{1*)}, Mamok Andri Senubekti²⁾, Binastya Anggara Sekti³⁾

¹⁾²⁾Manajemen Informatika, Universitas Teknologi Digital

³⁾Sistem Informasi, Universitas Esa Unggul

^{*)}Correspondence author: kodrat.mahatma@gmail.com, DKI Jakarta, Indonesia

DOI: <https://doi.org/10.37012/jtik.v12i1.3447>

Abstract

The rapid development of Artificial Intelligence (AI)-based systems has increased demand for integrated data architectures that support operational, analytical, and semantic data management. Organizations currently utilize relational databases, Data Warehouses, and Knowledge Graphs to support data integration in modern AI environments. However, previous studies generally discuss these approaches separately and rarely provide an integrated perspective regarding their roles within AI-ready architectures. This article conducts a comparative literature review of Relational Databases, Data Warehouses, and Knowledge Graphs as data integration approaches for AI-based systems. The study applies a conceptual literature review approach by analyzing scientific publications related to data integration, data warehouses, the semantic web, and knowledge graphs. The findings indicate that Relational Databases mainly support operational layers, Data Warehouses support analytical layers, while Knowledge Graphs provide semantic representation and contextual reasoning capabilities for modern AI systems. This article proposes a three-layer data integration perspective consisting of operational, analytical, and semantic layers as a conceptual framework for AI-ready data architectures.

Keywords: Artificial Intelligence, Data Integration, Data Warehouse, Knowledge Graph, Relational Database

Abstrak

Perkembangan sistem berbasis Artificial Intelligence (AI) meningkatkan kebutuhan organisasi terhadap integrasi data yang mampu mendukung pengelolaan data operasional, analitik, dan semantik secara terpadu. Berbagai organisasi saat ini memanfaatkan Relational Database, Data Warehouse, dan Knowledge Graph untuk mendukung integrasi data dalam lingkungan AI modern. Namun, sebagian besar kajian sebelumnya masih membahas ketiga pendekatan tersebut secara terpisah dan belum memberikan perspektif integratif mengenai posisi dan kontribusinya dalam arsitektur data siap AI. Artikel ini bertujuan melakukan kajian literatur komparatif terhadap Relational Database, Data Warehouse, dan Knowledge Graph sebagai pendekatan integrasi data untuk sistem berbasis AI. Metode penelitian yang digunakan adalah kajian literatur konseptual dengan pendekatan comparative review terhadap publikasi ilmiah terkait data integration, data warehouse, semantic web, dan knowledge graph. Hasil kajian menunjukkan bahwa Relational Database berperan pada lapisan operasional, Data Warehouse pada lapisan analitik, sedangkan Knowledge Graph mendukung lapisan semantik untuk AI modern. Artikel ini mengusulkan perspektif tiga lapisan integrasi data sebagai kerangka konseptual untuk membangun AI-ready data architecture.

Kata Kunci: Artificial Intelligence, Data Integration, Data Warehouse, Knowledge Graph, Relational Database

PENDAHULUAN

Perkembangan *Artificial Intelligence* (AI) meningkatkan kebutuhan organisasi terhadap pengelolaan dan integrasi data yang semakin kompleks. Sistem berbasis AI modern tidak hanya membutuhkan data dalam jumlah besar, tetapi juga membutuhkan data yang konsisten, terintegrasi, dan memiliki konteks semantik yang memadai untuk mendukung machine learning, analytics, reasoning, dan pengambilan keputusan berbasis AI (Russell & Norvig, 2021).

Permasalahan integrasi data telah menjadi salah satu isu utama dalam pengembangan sistem informasi modern. Lenzerini (2002) menjelaskan bahwa integrasi data bertujuan menyatukan data dari berbagai sumber heterogen agar dapat digunakan secara terpadu. Selain itu, Halevy et al., (2006) menekankan bahwa organisasi modern menghadapi peningkatan volume, variasi, dan kompleksitas data yang terus berkembang.

Dalam sejarah pengelolaan data organisasi, *relational database* menjadi pendekatan dominan untuk mendukung transaksi dan pengelolaan data operasional sejak diperkenalkan oleh Codd (1970). Pendekatan ini unggul dalam menjaga konsistensi data dan integritas transaksi. Seiring berkembangnya kebutuhan analitik, muncul pendekatan *data warehouse* yang berfokus pada integrasi data historis dan analytical intelligence (Chaudhuri & Dayal, 1997; Inmon, 2005).

Pada perkembangan berikutnya, muncul kebutuhan terhadap representasi data yang lebih fleksibel dan semantik untuk mendukung *interoperability*, *explainability*, dan *contextual intelligence* pada sistem berbasis AI. Dalam konteks ini, *knowledge graph* menjadi pendekatan yang semakin banyak digunakan karena mampu merepresentasikan entitas dan relasi secara eksplisit dalam bentuk graph semantik (Bizer et al., 2009; Hogan et al., 2022).

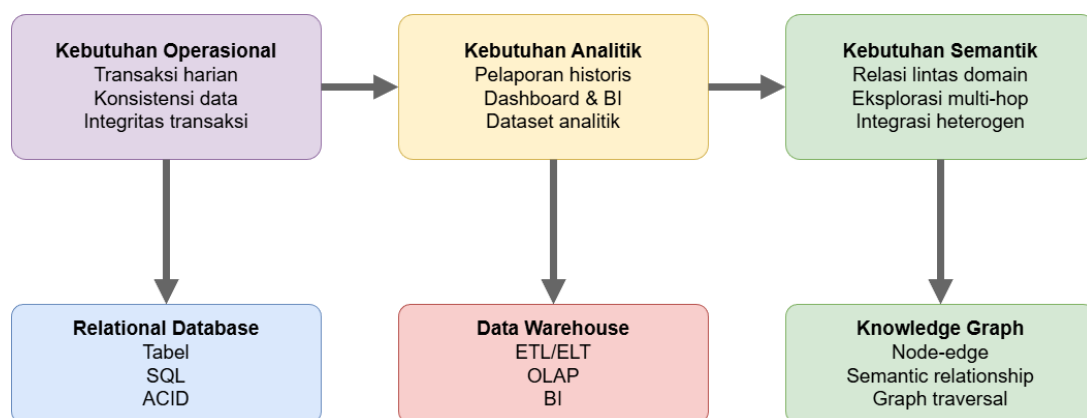
Kebutuhan tersebut juga berkaitan dengan berkembangnya model data berbasis graf dan *explainable AI*. *Graph database* memberikan cara yang lebih alami untuk merepresentasikan relasi kompleks antar entitas, terutama ketika data memiliki struktur hubungan yang padat dan heterogen (Angles & Gutierrez, 2008). Di sisi lain, sistem AI yang digunakan dalam pengambilan keputusan organisasi juga perlu dapat dijelaskan agar

hasilnya tidak hanya akurat, tetapi juga dapat ditelusuri dan dipertanggungjawabkan (Adadi & Berrada, 2018).

Meskipun berbagai penelitian telah membahas relational database, data warehouse, dan knowledge graph, sebagian besar kajian masih membahas pendekatan tersebut secara terpisah. Belum banyak penelitian yang mengkaji ketiga pendekatan tersebut secara komparatif dalam konteks integrasi data untuk mendukung sistem berbasis AI.

Artikel ini mengusulkan perspektif tiga lapisan integrasi data untuk sistem berbasis AI, yaitu lapisan operasional, lapisan analitik, dan lapisan semantik. Dalam perspektif ini, relational database diposisikan sebagai fondasi pengelolaan data operasional, data warehouse sebagai lapisan integrasi analitik, dan knowledge graph sebagai lapisan representasi semantik.

Tujuan penelitian ini adalah melakukan kajian literatur komparatif terhadap relational database, data warehouse, dan knowledge graph sebagai pendekatan integrasi data untuk sistem berbasis AI. Pemetaan umum hubungan antara kebutuhan operasional, analitik, dan semantik terhadap pendekatan integrasi data yang sesuai ditunjukkan pada Gambar 1.



Gambar 1. Pemetaan Kebutuhan Intergrasi Data terhadap Pendekatan Relational Database, Data Warehouse, dan Knowledge Graph

METODE PENELITIAN

Penelitian ini menggunakan metode kajian literatur konseptual (*conceptual literature review*) dengan pendekatan *comparative review*. Pendekatan ini digunakan untuk

membandingkan karakteristik, peran, keunggulan, dan keterbatasan relational database, data warehouse, dan knowledge graph dalam konteks integrasi data untuk sistem berbasis AI.

Sumber literatur diperoleh dari jurnal ilmiah, buku akademik, conference proceedings, dan artikel teknis yang membahas *data integration*, *relational database*, *data warehouse*, *semantic web*, *knowledge graph*, dan *AI-ready data architecture*. Literatur utama yang digunakan mencakup karya klasik dan modern seperti (Bizer et al., 2009; Chaudhuri & Dayal, 1997; Codd, 1970; Hogan et al., 2022; Inmon, 2005; Ji et al., 2022; Kimball & Ross, 2013; Lenzerini, 2002; Nickel et al., 2016).

Analisis dilakukan melalui beberapa tahapan, yaitu: 1. Identifikasi konsep utama integrasi data. 2. Klasifikasi karakteristik masing-masing pendekatan. 3. Analisis peran pendekatan dalam sistem berbasis AI. 4. Perbandingan keunggulan dan keterbatasan. 5. Penyusunan perspektif konseptual tiga lapisan integrasi data. Tahapan tersebut digunakan untuk memastikan bahwa perbandingan tidak hanya bersifat deskriptif, tetapi juga menghasilkan kerangka konseptual yang relevan bagi pengembangan arsitektur data siap AI.

Dimensi analisis yang digunakan dalam penelitian ini meliputi struktur data, fokus utama, fleksibilitas skema, kemampuan integrasi, dukungan terhadap analitik, representasi semantik, interoperabilitas (*interoperability*), skalabilitas (*scalability*), keterjelasan AI (*explainability*), kecerdasan kontekstual (*contextual intelligence*), dan relevansi terhadap AI modern. Dimensi analisis tersebut digunakan untuk membandingkan relational database, data warehouse, dan knowledge graph secara lebih seimbang, sebagaimana dirangkum pada Tabel 1.

Tabel 1. Dimensi Analisis Perbandingan Pendekatan Integrasi Data

Dimensi	RDB	DW	KG	Relevansi AI
Struktur data	Relasional	Multidimensi	Graph	Menentukan fleksibilitas pengolahan data
Fokus utama	Operasional	Analitik	Semantik	Menentukan peran dalam arsitektur AI
Fleksibilitas skema	Rendah–sedang	Sedang	Tinggi	Penting untuk data heterogen

Dimensi	RDB	DW	KG	Relevansi AI
Integrasi data	Terbatas	Multi sumber	Semantik	Mendukung interoperability
Dukungan analitik	Dasar	Tinggi	Graph analytics	Mendukung machine learning
Representasi semantik	Rendah	Terbatas	Tinggi	Mendukung contextual understanding
Interoperabilitas	Sedang	Sedang	Tinggi	Mendukung integrasi lintas sistem
Skalabilitas/ Scalability	Tinggi untuk transaksi	Tinggi untuk analitik	Tinggi untuk relasi kompleks	Mendukung AI skala besar
Keterjelasan/ Explainability	Rendah	Sedang	Tinggi	Mendukung AI yang dapat dijelaskan
Kecerdasan kontekstual	Rendah	Sedang	Tinggi	Mendukung reasoning AI
Relevansi terhadap AI modern	Sedang	Tinggi	Sangat tinggi	Menentukan kontribusi terhadap AI-ready architecture

HASIL DAN PEMBAHASAN

Relational Database dalam Integrasi Data

Relational Database merupakan salah satu pendekatan paling dominan dalam pengelolaan data organisasi sejak diperkenalkan oleh Codd (1970). Model relasional menggunakan struktur tabel yang terdiri atas baris dan kolom untuk menyimpan data secara terstruktur. Pendekatan ini dirancang untuk mendukung konsistensi data, integritas transaksi, dan efisiensi pengelolaan data operasional.

Dalam konteks integrasi data, relational database banyak digunakan sebagai sumber utama data operasional organisasi. Sistem Enterprise Resource Planning (ERP), Customer Relationship Management (CRM). Keunggulan utama pendekatan ini terletak pada kemampuan menjaga konsistensi transaksi melalui prinsip ACID (Atomicity, Consistency, Isolation, Durability).

Selain itu, relational database juga mendukung penggunaan Structured Query Language (SQL) yang telah menjadi standar dalam pengelolaan dan manipulasi data terstruktur. Kemampuan ini menjadikan relational database tetap relevan dalam berbagai lingkungan organisasi modern.

Namun demikian, relational database memiliki beberapa keterbatasan dalam konteks AI modern. Struktur skema yang relatif kaku menyebabkan proses integrasi data lintas domain menjadi lebih kompleks, terutama ketika organisasi harus mengelola data semi-terstruktur, data tidak terstruktur, atau relasi yang sangat kompleks. Selain itu, pendekatan relasional juga kurang optimal dalam merepresentasikan hubungan semantik dan contextual relationships yang semakin dibutuhkan dalam sistem berbasis AI.

Dibandingkan dengan pendekatan lain, Relational Database lebih berorientasi pada efisiensi transaksi operasional dibandingkan dengan representasi pengetahuan atau reasoning semantik. Oleh karena itu, pendekatan ini lebih tepat diposisikan sebagai lapisan operasional dalam arsitektur data siap AI.

Data Warehouse dalam Integrasi Data

Data Warehouse berkembang sebagai solusi untuk mendukung kebutuhan analitik organisasi yang tidak dapat dipenuhi secara optimal oleh sistem transaksi operasional. Chaudhuri & Dayal (1997) menjelaskan bahwa Data Warehouse dirancang untuk mengintegrasikan data historis dari berbagai sumber guna mendukung proses analitik dan pengambilan keputusan.

Menurut Inmon (2005), data warehouse memiliki karakteristik *subject-oriented*, *integrated*, *time-variant*, dan *non-volatile* sebagai fondasi bagi pengambilan keputusan manajemen. Sementara itu, Kimball & Ross (2013) menekankan pentingnya *dimensional modeling* untuk mendukung analitik bisnis dan *Business Intelligence*. Dengan demikian, data warehouse tidak hanya berfungsi sebagai tempat penyimpanan data historis, tetapi juga sebagai lapisan analitik yang membantu organisasi memahami pola, tren, dan kinerja bisnis.

Dalam konteks integrasi data, data warehouse berfungsi sebagai pusat konsolidasi data dari berbagai sistem operasional melalui proses *Extract, Transform, Load* (ETL). Pendekatan ini memungkinkan organisasi membangun repositori data historis yang terintegrasi untuk mendukung *dashboard*, *reporting*, *Business Intelligence*, dan *machine*

learning. Dalam sistem berbasis AI, data warehouse juga dapat berperan dalam penyediaan data analitik, persiapan dataset, dan pelaporan berbasis data.

Meskipun demikian, data warehouse masih memiliki keterbatasan dalam merepresentasikan relasi semantik dan hubungan kontekstual yang kompleks. Dibandingkan dengan Relational Database yang berfokus pada transaksi operasional, Data Warehouse lebih unggul dalam integrasi historis dan analitik. Namun, kemampuan representasi semantik dan penalaran kontekstualnya masih lebih terbatas dibandingkan dengan Knowledge Graph.

Knowledge Graph dalam Integrasi Data

Knowledge Graph merupakan pendekatan representasi data berbasis graf yang digunakan untuk memodelkan entitas dan hubungan antarentitas secara eksplisit. Pendekatan ini berkembang dari konsep *Semantic Web* dan *Linked Data*, yang menekankan keterhubungan data, makna, dan interoperabilitas antar sumber data (Bizer et al., 2009). Berbeda dengan model relasional yang menekankan struktur tabel, knowledge graph merepresentasikan data dalam bentuk simpul (*node*) dan relasi (*edge*), sehingga relasi antar-entitas dapat menjadi bagian utama dari proses analisis.

Dalam konteks integrasi data, knowledge graph memiliki keunggulan dalam menghubungkan data heterogen dari berbagai domain. Pendekatan ini mendukung *semantic interoperability*, penalaran kontekstual, keterjelasan AI (*explainability*), serta pemanfaatan data untuk aplikasi seperti *semantic search*, *recommendation system*, *question answering*, dan *Retrieval-Augmented Generation* (RAG). Ji et al. (2022) menjelaskan bahwa knowledge graph dapat meningkatkan kemampuan representasi, akuisisi, dan pemanfaatan pengetahuan dalam sistem AI.

Dari perspektif model data, penggunaan graf memperluas cara organisasi memandang integrasi data. Jika pendekatan relasional menekankan tabel dan kunci, pendekatan graf menekankan keterhubungan antar-entitas sebagai objek analisis utama (Angles & Gutierrez, 2008). Hal ini relevan untuk sistem berbasis AI karena banyak kebutuhan modern, seperti pencarian semantik, rekomendasi, dan analisis keterkaitan, membutuhkan eksplorasi relasi secara eksplisit.

Namun demikian, implementasi knowledge graph juga memiliki tantangan, seperti kebutuhan standardisasi semantik, pembangunan ontologi, kualitas data, dan tata kelola relasi antar-entitas. Karena itu, Knowledge Graph tidak selalu menggantikan Relational Database atau Data Warehouse, tetapi lebih tepat diposisikan sebagai lapisan semantik yang melengkapi lapisan operasional dan analitik dalam arsitektur data siap AI.

Contoh Perangkat Lunak dan Platform

Berbagai pendekatan integrasi data didukung oleh platform yang berbeda. Berbagai pendekatan integrasi data didukung oleh perangkat lunak dan platform yang berbeda. Relational Database umumnya menggunakan platform seperti MySQL, PostgreSQL, Oracle Database, dan Microsoft SQL Server. Data Warehouse banyak memanfaatkan platform seperti Snowflake, Amazon Redshift, Google BigQuery, dan Microsoft Azure Synapse Analytics. Sementara itu, Knowledge Graph banyak menggunakan Neo4j, GraphDB, Stardog, dan Amazon Neptune. Contoh platform yang merepresentasikan masing-masing pendekatan integrasi data ditampilkan pada Tabel 2.

Tabel 2. Contoh Perangkat Lunak dan Platform

Pendekatan	Contoh Platform	Karakteristik	Dukungan AI
Relational Database	MySQL, PostgreSQL	Transaksi operasional	Data source
Data Warehouse	BigQuery, Snowflake	Historical analytics	ML analytics
Knowledge Graph	Neo4j, GraphDB	Semantic reasoning	Explainable AI

Hasil Komparasi Utama

Hasil kajian menunjukkan bahwa masing-masing pendekatan memiliki fokus dan karakteristik berbeda dalam mendukung integrasi data untuk sistem berbasis AI. Relational Database unggul dalam pengelolaan transaksi operasional dan konsistensi data. Data Warehouse unggul dalam analitik historis dan konsolidasi data lintas sistem. Sementara itu, Knowledge Graph unggul dalam representasi semantik, contextual reasoning, dan explainability.

Temuan ini menunjukkan bahwa tidak ada satu pendekatan tunggal yang mampu memenuhi seluruh kebutuhan integrasi data AI modern. Sebaliknya, organisasi perlu memanfaatkan kombinasi pendekatan yang sesuai dengan kebutuhan operasional, analitik,

dan semantik. Perbandingan karakteristik utama antara Relational Database, Data Warehouse, dan Knowledge Graph disajikan pada Tabel 3.

Tabel 3. Perbandingan Karakteristik Relational Database, Data Warehouse, dan Knowledge Graph

Karakteristik	RDB	DW	KG
Fokus utama	Transaksi	Analitik	Semantik
Struktur	Tabel	Multidimensi	Node-edge
Fleksibilitas	Rendah	Sedang	Tinggi
Historical analytics	Rendah	Tinggi	Sedang
Semantic reasoning	Tidak ada	Terbatas	Tinggi
Explainability	Rendah	Sedang	Tinggi
AI readiness	Sedang	Tinggi	Sangat tinggi
Use case	ERP/CRM	BI/analytics	RAG/recommendation

Sebagaimana terlihat pada Tabel 3, Relational Database lebih kuat pada transaksi operasional, Data Warehouse pada analitik historis, sedangkan Knowledge Graph pada fleksibilitas semantik dan explainability

Kaitan Integrasi Data dengan Explainable AI dan Data-Centric AI

Dalam sistem berbasis AI, integrasi data tidak hanya berperan sebagai proses teknis untuk menyatukan sumber data, tetapi juga sebagai fondasi bagi keterjelasan, pengendalian kualitas, dan akuntabilitas model. Adadi & Berrada (2018) menekankan bahwa *explainable AI* diperlukan agar proses dan hasil model dapat dipahami oleh pengguna, pengembang, dan pengambil keputusan. Dalam konteks ini, knowledge graph memiliki posisi penting karena relasi antar entitas dapat direpresentasikan secara eksplisit sehingga alur keterkaitan data lebih mudah ditelusuri.

Selain aspek keterjelasan, kesiapan data juga menjadi faktor kunci keberhasilan AI. Sambasivan et al. (2021) menunjukkan bahwa banyak kegagalan sistem AI berawal dari permasalahan data seperti bias, inkonsistensi, kurangnya dokumentasi, dan lemahnya pekerjaan kurasi data. Schelter et al. (2018) juga menyoroti pentingnya pengelolaan dan monitoring kualitas data dalam pipeline machine learning agar perubahan data tidak menurunkan performa model.

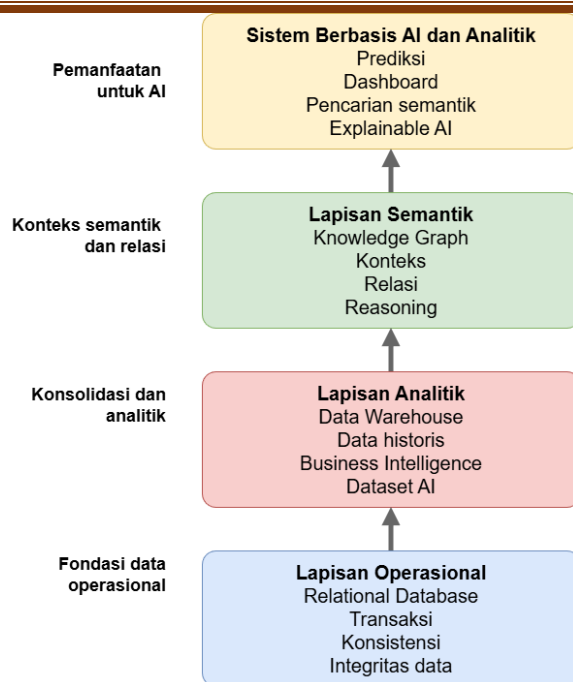
Dari sisi integrasi lintas sumber, perbedaan skema dan struktur data masih menjadi tantangan mendasar. Rahm & Bernstein (2001) menjelaskan bahwa *schema matching* merupakan salah satu persoalan utama dalam integrasi data karena sumber data yang berbeda sering menggunakan struktur, terminologi, dan tingkat granularitas yang tidak sama. Oleh karena itu, Data Warehouse dan Knowledge Graph perlu dilihat sebagai mekanisme pelengkap: Data Warehouse mendukung konsolidasi dan historisasi data, sedangkan Knowledge Graph memperkuat pemetaan semantik dan eksplorasi relasi.

Keterkaitan antara knowledge graph dan AI juga dapat dilihat dari relational machine learning. Nickel et al. (2016) menjelaskan bahwa knowledge graph dapat menjadi basis bagi pembelajaran relasional karena data tidak hanya dipandang sebagai tabel fitur, tetapi sebagai jaringan relasi antar-entitas. Dengan demikian, perspektif tiga lapisan yang diusulkan dalam artikel ini tidak hanya memetakan teknologi integrasi data, tetapi juga menjelaskan bagaimana data operasional, analitik, dan semantik dapat mendukung AI yang lebih dapat dijelaskan dan lebih siap digunakan secara organisasi.

Perspektif Tiga Lapisan Integrasi Data untuk Sistem Berbasis AI

Berdasarkan hasil kajian, artikel ini mengusulkan perspektif tiga lapisan integrasi data untuk mendukung sistem berbasis AI. Lapisan pertama adalah lapisan operasional yang didukung oleh relational database. Lapisan kedua adalah lapisan analitik yang didukung oleh Data Warehouse. Lapisan ketiga adalah lapisan semantik yang didukung oleh Knowledge Graph.

Perspektif ini menunjukkan bahwa Relational Database, Data Warehouse, dan Knowledge Graph tidak bersifat saling menggantikan, melainkan saling melengkapi dalam membangun *AI-ready data architecture*. Perspektif tiga lapisan integrasi data yang diusulkan dalam artikel ini divisualisasikan pada Gambar 2.



Gambar 2. Perspektif Tiga Lapisan Integrasi Data untuk Sistem Berbasis AI

Isu Kualitas, Tata Kelola, dan Kesiapan Data untuk AI

Implementasi sistem berbasis AI tidak hanya bergantung pada teknologi integrasi data, tetapi juga sangat dipengaruhi oleh kualitas data, tata kelola data, dan kesiapan organisasi dalam mengelola data secara konsisten. Banyak permasalahan pada sistem AI modern berawal dari persoalan data seperti bias, inkonsistensi, kurangnya dokumentasi, dan lemahnya kurasi data (Sambasivan et al., 2021).

Dalam konteks pipeline machine learning, pemantauan kualitas data juga penting agar perubahan data tidak menurunkan performa model (Schelter et al., 2018). Oleh karena itu, integrasi data untuk sistem berbasis AI perlu memperhatikan beberapa isu utama, seperti inkonsistensi data (*data inconsistency*), heterogenitas skema (*schema heterogeneity*), ambiguitas semantik (*semantic ambiguity*), keterlacakan data (*data lineage*), interoperabilitas (*interoperability*), skalabilitas (*scalability*), bias data (*data bias*), dan silo data (*data silos*).

Heterogenitas skema berkaitan langsung dengan persoalan *schema matching*, yaitu proses mencocokkan struktur atau elemen data dari berbagai sumber yang berbeda. Isu ini telah lama dibahas dalam literatur integrasi data karena menjadi salah satu hambatan utama dalam penggabungan data lintas sistem (Rahm & Bernstein, 2001). Sementara itu,

ambiguitas semantik dan keterlacakan data menjadi penting karena sistem berbasis AI tidak hanya membutuhkan data yang tersedia, tetapi juga data yang dapat ditafsirkan, ditelusuri, dan dipertanggungjawabkan.

Knowledge Graph memberikan peluang untuk meningkatkan interoperabilitas semantik, tetapi implementasinya tetap membutuhkan tata kelola data yang baik. Beberapa isu utama yang memengaruhi kesiapan data untuk sistem berbasis AI dirangkum pada Tabel 4.

Tabel 4. Isu Data yang Relevan dalam Arsitektur Data Siap AI

Isu Data	Dampak terhadap AI	Penanganan
Inkonsistensi data	Prediksi tidak stabil	Tata kelola data
Heterogenitas skema	Sulit integrasi lintas sistem	Pencocokan skema dan pemetaan semantik
Ambiguitas semantik	Salah tafsir makna data	Ontologi dan standardisasi semantik
Keterlacakan data	Sulit audit dan penelusuran keputusan	Manajemen metadata
Interoperabilitas	Integrasi sistem sulit	Standar data dan Knowledge Graph
Skalabilitas	Pemrosesan data AI terhambat	Data Warehouse dan arsitektur terdistribusi
Bias data	Model AI bias	Pemantauan kualitas data
Silo data	Fragmentasi data	Arsitektur hybrid

Rekomendasi Berdasarkan Kebutuhan Organisasi

Pemilihan pendekatan integrasi data perlu disesuaikan dengan kebutuhan organisasi. Organisasi yang berfokus pada transaksi operasional membutuhkan relational database sebagai fondasi utama. Organisasi yang berfokus pada Business Intelligence dan historical analytics memerlukan Data Warehouse. Sementara itu, organisasi yang membutuhkan semantic integration, explainability, dan contextual intelligence dapat mempertimbangkan penggunaan Knowledge Graph.

Dalam praktiknya, organisasi modern umumnya membutuhkan kombinasi ketiga pendekatan tersebut untuk membangun *AI-ready data architecture*. Rekomendasi pemilihan pendekatan integrasi data berdasarkan kebutuhan organisasi disajikan pada Tabel 5.

Tabel 5. Rekomendasi Pendekatan Integrasi Data Berdasarkan Kebutuhan Organisasi

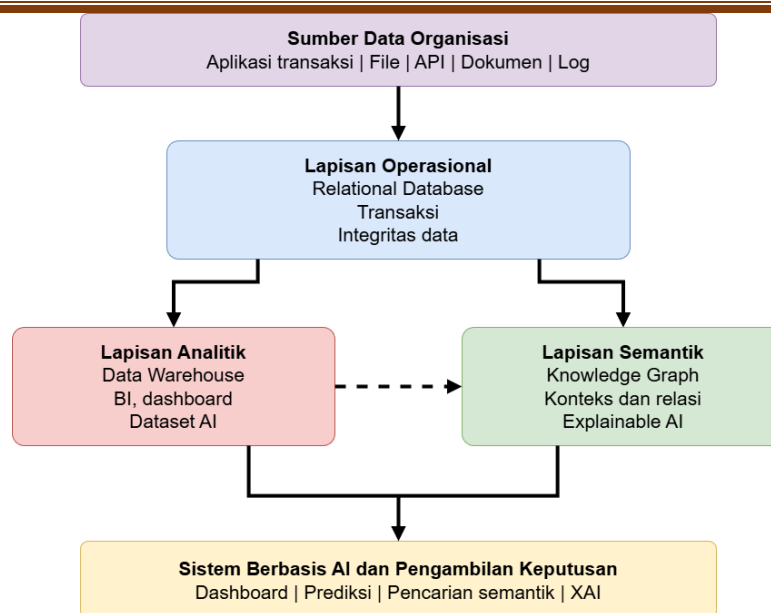
Kebutuhan	Pendekatan	Dukungan AI
Transaksi operasional	Relational Database	Operational AI
Business Intelligence	Data Warehouse	ML analytics
Semantic integration	Knowledge Graph	Explainable AI
RAG dan LLM	KG + DW	Generative AI
Enterprise AI platform	Hybrid architecture	Full AI ecosystem

Implikasi Arsitektur Hybrid

Hasil kajian menunjukkan bahwa arsitektur hybrid menjadi pendekatan yang semakin relevan dalam implementasi sistem berbasis AI modern. Pendekatan hybrid memungkinkan organisasi mengombinasikan keunggulan Relational Database, Data Warehouse, dan Knowledge Graph dalam satu ekosistem data. Dalam arsitektur hybrid, Relational Database tetap digunakan untuk mendukung transaksi operasional, Data Warehouse digunakan untuk analitik historis dan konsolidasi data, sedangkan Knowledge Graph digunakan untuk representasi semantik dan contextual intelligence.

Pendekatan ini juga semakin relevan dalam era Generative AI dan Retrieval-Augmented Generation (RAG), di mana sistem AI membutuhkan kombinasi data operasional, data historis, dan pengetahuan kontekstual secara terpadu. Ilustrasi implementasi arsitektur hybrid yang menggabungkan lapisan operasional, analitik, dan semantik ditunjukkan pada Gambar 3.

Sebagaimana ditunjukkan pada Gambar 3, arsitektur hybrid memungkinkan organisasi menghubungkan sumber data operasional, data warehouse, dan knowledge graph untuk mendukung layanan AI dan pengambilan keputusan. Dengan pola ini, organisasi dapat mempertahankan fungsi transaksi, analitik historis, dan representasi semantik dalam satu ekosistem data yang saling melengkapi.



Gambar 3. Ilustrasi Arsitektur Hybrid yang Menggabungkan Lapisan Operasional, Analitik, dan Semantik

KESIMPULAN DAN REKOMENDASI

Artikel ini telah melakukan kajian literatur komparatif terhadap Relational Database, Data Warehouse, dan Knowledge Graph sebagai pendekatan integrasi data untuk sistem berbasis AI. Hasil kajian menunjukkan bahwa ketiga pendekatan memiliki karakteristik, fungsi, dan kontribusi yang berbeda dalam mendukung pengelolaan data organisasi.

Relational Database berperan sebagai fondasi pengelolaan data operasional dan transaksi. Data Warehouse mendukung integrasi data historis dan analitik organisasi, sedangkan Knowledge Graph memberikan kemampuan representasi semantik, penalaran kontekstual, dan keterjelasan AI yang semakin penting dalam sistem AI modern.

Kontribusi utama artikel ini adalah usulan perspektif tiga lapisan integrasi data yang terdiri atas lapisan operasional, analitik, dan semantik sebagai kerangka konseptual untuk mendukung pembangunan arsitektur data siap AI. Perspektif ini menegaskan bahwa ketiga pendekatan tersebut tidak saling menggantikan, tetapi saling melengkapi sesuai dengan kebutuhan pengelolaan data yang berbeda.

Kajian ini juga menunjukkan bahwa pendekatan hybrid menjadi strategi yang paling relevan dalam implementasi sistem berbasis AI modern. Oleh karena itu, organisasi perlu mempertimbangkan kombinasi berbagai pendekatan integrasi data sesuai kebutuhan operasional, analitik, dan semantik.

Penelitian selanjutnya dapat mengembangkan evaluasi implementasi arsitektur hybrid pada lingkungan enterprise, termasuk integrasi dengan Generative AI, Large Language Model (LLM), dan Retrieval-Augmented Generation (RAG). Evaluasi tersebut dapat dilakukan melalui studi kasus empiris untuk menilai manfaat, keterbatasan, dan biaya implementasi masing-masing pendekatan secara lebih objektif.

REFERENSI

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Angles, R., & Gutierrez, C. (2008). Survey of graph database models. *ACM Computing Surveys*, 40(1), 1–39. <https://doi.org/10.1145/1322432.1322433>
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data: The story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. <https://doi.org/10.4018/jswis.2009081901>
- Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26(1), 65–74. <https://doi.org/10.1145/248603.248616>
- Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6), 377–387. <https://doi.org/10.1145/362384.362685>
- Halevy, A. Y., Rajaraman, A., & Ordille, J. J. (2006). Data integration: The teenage years. *Proceedings of the 32nd International Conference on Very Large Data Bases*, 9–16. <https://dl.acm.org/doi/10.5555/1182635.1164130>
- Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2022).

- Knowledge Graphs. *ACM Computing Surveys*, 54(4), 1–37. <https://doi.org/10.1145/3447772>
- Inmon, W. H. (2005). *Building the Data Warehouse* (4th ed.). Wiley. <https://www.wiley.com/en-es/Building+the+Data+Warehouse%2C+4th+Edition-p-9780764599446>
- Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2022). A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2), 494–514. IEEE Transactions on Neural Networks and Learning Systems. <https://doi.org/10.1109/TNNLS.2021.3070843>
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). Wiley. <https://www.wiley.com/en-jp/The%2BData%2BWarehouse%2BToolkit%3A%2BThe%2BDefinitive%2BGuide%2Bto%2BDimensional%2BModeling%2C%2B3rd%2BEdition-p-9781118530801>
- Lenzerini, M. (2002). Data integration: A theoretical perspective. *Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 233–246. <https://doi.org/10.1145/543613.543644>
- Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A Review of Relational Machine Learning for Knowledge Graphs. *Proceedings of the IEEE*, 104(1), 11–33. <https://doi.org/10.1109/JPROC.2015.2483592>
- Rahm, E., & Bernstein, P. A. (2001). A survey of approaches to automatic schema matching. *The VLDB Journal*, 10(4), 334–350. <https://doi.org/10.1007/s007780100057>
- Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson. <https://www.pearson.com/en-us/subject-catalog/p/artificial-intelligence-a-modern-approach/P200000003500/9780137505135>
- Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). “Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21*, 1–15. <https://doi.org/10.1145/3411764.3445518>

Schelter, S., Biessmann, F., Januschowski, T., Salinas, D., Seufert, S., & Szarvas, G. (2018).

On Challenges in Machine Learning Model Management. *IEEE Data Eng. Bull.*

<https://www.semanticscholar.org/paper/On-Challenges-in-Machine-Learning>

Model-Management-Schelter

Biessmann/46038cb1606041001be2f42b4e09f630d4b4147f