

Knowledge Graph Integration to Improve Reliability and Quality of Maintenance Data in the Oil and Gas Processing Industry

Adnan Buyung Nasution^{1*}, Ahir Yugo Nugroho², Deny Adhar³, Risman⁴,
Ramadani⁵

¹Sistem informasi, Universitas Borobudur

²Sistem informasi, universitas dharmawangsa

^{3,4,5}Sistem informasi, Universitas Muhammadiyah Asahan

Corresponding Author: Adnan Buyung Nasution

Adnan.buyung01@gmail.com

ARTICLE INFO

Keywords : Knowledge Graph, Data Reliability, Data Governance, Refinery Maintenance, Data Integration

Received : 20 May 2026

Revised : 23 June 2026

Accepted: 25 July 2026

©2026 Nasution, Nugroho, Adhar, Risman, Ramadani: This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

Data quality and reliability are fundamental challenges in asset maintenance management in the oil and gas processing industry, where equipment data is scattered across dozens of heterogeneous operational tables, including SAP notifications, work orders, pipe inspections, corrosion monitoring results, and metering data. This data fragmentation makes it difficult to trace relationships between entities, identify anomalies, and perform consistent cross-validation. This study proposes a Knowledge Graph (KG) integration approach as a data representation layer that connects equipment entities as central nodes with child nodes from monitoring tables through typed relationships. Implementation is carried out through the development of a Graph Manager component based on Flask and D3.js that synchronizes data from a PostgreSQL relational database to a Neo4j graph database on a scheduled basis. Evaluation is carried out on three aspects of data reliability, namely referential consistency, attribute completeness, and timeliness of data updates, before and after the implementation of the KG. Test results in a production environment show significant improvements in equipment tag inconsistency detection, decreased equipment history search time, and increased semantic search accuracy when the KG is combined with embedding vector (pgvector)-based search. This study concludes that the graph approach provides a real contribution to data governance and operational reliability of maintenance systems.

Integrasi Knowledge Graph untuk Peningkatan Reliabilitas dan Kualitas Data Pemeliharaan pada Industri Pengolahan Minyak dan Gas

Adnan Buyung Nasution^{1*}, Ahir Yugo Nugroho², Deny Adhar³, Risman⁴, Ramadani⁵

¹Sistem informasi, Universitas Borobudur

²Sistem informasi, universitas dharmawangsa

^{3,4,5}Sistem informasi, Universitas Muhammadiyah Asahan

Corresponding Author: Adnan Buyung Nasution

Adnan.buyung01@gmail.com

ARTICLE INFO

Kata Kunci: Knowledge Graph, Reliabilitas Data, Data Governance, Pemeliharaan Kilang, Integrasi Data

Received : 20 Mei 2026

Revised : 23 Juni 2026

Accepted: 25 Juli 2026

©2026 Nasution, Nugroho, Adhar, Risman, Ramadani: This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



ABSTRAK

Kualitas dan keandalan data merupakan tantangan mendasar dalam pengelolaan aset pemeliharaan pada industri pengolahan minyak dan gas, di mana data peralatan tersebar pada puluhan tabel operasional yang heterogen, meliputi notifikasi SAP, work order, inspeksi pipa, hasil pemantauan korosi, hingga data metering. Fragmentasi data tersebut menyebabkan kesulitan dalam menelusuri keterkaitan antar entitas, mengidentifikasi anomali, serta melakukan validasi silang yang konsisten. Penelitian ini mengusulkan pendekatan integrasi Knowledge Graph (KG) sebagai lapisan representasi data yang menghubungkan entitas peralatan sebagai node pusat dengan node turunan dari tabel-tabel pemantauan melalui relasi bertipe. Implementasi dilakukan melalui pembangunan komponen Graph Manager berbasis Flask dan D3.js yang melakukan sinkronisasi data dari basis data relasional PostgreSQL ke basis data graf Neo4j secara terjadwal. Evaluasi dilakukan terhadap tiga aspek reliabilitas data, yaitu konsistensi referensial, kelengkapan atribut, dan ketepatan waktu pembaruan data, sebelum dan setelah penerapan KG. Hasil pengujian pada lingkungan produksi menunjukkan peningkatan signifikan pada deteksi inkonsistensi tag peralatan, penurunan waktu penelusuran riwayat peralatan, serta peningkatan akurasi pencarian semantik ketika KG dikombinasikan dengan pencarian berbasis embedding vektor (pgvector). Penelitian ini menyimpulkan bahwa pendekatan graf memberikan kontribusi nyata terhadap tata kelola data (data governance) dan keandalan operasional sistem pemeliharaan.

PENDAHULUAN

Industri pengolahan minyak dan gas mengoperasikan ribuan unit peralatan yang masing-masing memiliki riwayat pemeliharaan, status integritas, serta data operasional yang dicatat pada berbagai sistem informasi yang berbeda. Pada konteks operasional Unit Pengolahan (Refinery Unit), data tersebut umumnya tersimpan dalam sistem ERP seperti SAP, basis data internal untuk pemantauan kondisi alat seperti Inspection Checklist Unit (ICU), Above Ground Tank (ATG), serta sistem pelaporan bad actor dan Basic Operational Condition (BOC). Setiap sistem ini dirancang untuk kebutuhan fungsional tertentu, sehingga keterkaitan antar data sering hanya direpresentasikan secara implisit melalui kesamaan nomor tag peralatan.

Pendekatan basis data relasional konvensional cukup memadai untuk operasi transaksional, namun mengalami keterbatasan ketika diperlukan analisis lintas tabel yang kompleks, misalnya menelusuri seluruh riwayat anomali, perbaikan, dan inspeksi suatu peralatan dalam satu pandangan terpadu. Permasalahan ini semakin terasa ketika jumlah tabel yang saling terkait mencapai puluhan, dengan struktur relasi yang tidak selalu dinormalisasi secara konsisten.

Knowledge Graph menawarkan paradigma representasi data berbasis entitas dan relasi yang secara alami merepresentasikan struktur jaringan keterkaitan antar objek. Dengan menempatkan entitas peralatan sebagai node pusat dan menghubungkannya dengan node-node dari tabel pemantauan melalui relasi bertipe (typed relationships), pengguna dapat melakukan penelusuran multi-hop yang sebelumnya memerlukan banyak operasi JOIN pada basis data relasional.

Penelitian ini bertujuan untuk: (1) merancang arsitektur integrasi Knowledge Graph yang menyinkronkan data dari basis data relasional operasional ke basis data graf; (2) mengevaluasi dampak penerapan KG terhadap tiga dimensi reliabilitas data, yaitu konsistensi referensial, kelengkapan atribut, dan ketepatan waktu; serta (3) mengukur peningkatan efisiensi penelusuran data dan akurasi pencarian semantik dibandingkan dengan pendekatan basis data relasional murni.

TINJAUAN PUSTAKA

Reliabilitas Data dan Data Governance

Reliabilitas data dalam konteks tata kelola data mencakup dimensi akurasi, konsistensi, kelengkapan, ketepatan waktu, dan keterhubungan (linkage) antar entitas data. Kerangka kerja data governance menempatkan katalog data, lineage, serta kontrol kualitas sebagai komponen inti untuk menjamin bahwa data yang digunakan dalam pengambilan keputusan operasional dapat dipercaya. Pada lingkungan industri proses, kegagalan dalam menjaga konsistensi data peralatan dapat berdampak langsung pada keakuratan analisis keandalan dan perencanaan pemeliharaan.

Knowledge Graph dalam Manajemen Aset

Knowledge Graph merepresentasikan domain pengetahuan sebagai kumpulan node (entitas) dan edge (relasi) yang dapat diberi label dan properti. Dalam manajemen aset industri, representasi ini memungkinkan pemodelan hierarki peralatan, hubungan fungsional antar komponen, serta keterkaitan dengan kejadian operasional seperti notifikasi kerusakan dan order pemeliharaan. Basis data graf seperti Neo4j menyediakan bahasa kueri deklaratif (Cypher) yang mempermudah penulisan kueri penelusuran multi-hop dibandingkan SQL pada basis data relasional.

Sinkronisasi Data Relasional ke Graf

Pendekatan extract-transform-load (ETL) digunakan untuk memindahkan data dari basis data sumber ke representasi graf. Proses ini perlu memperhatikan idempotensi operasi sinkronisasi, sehingga eksekusi berulang tidak menghasilkan duplikasi node atau relasi. Strategi umum yang digunakan adalah operasi MERGE pada Cypher, yang melakukan pembuatan node baru hanya jika belum terdapat node dengan identitas yang sama, serta pembaruan properti pada node yang telah ada.

Pencarian Semantik Berbasis Embedding

Selain representasi graf, pencarian semantik berbasis embedding vektor menggunakan model representasi teks memungkinkan pencarian dokumen atau data berdasarkan kemiripan makna, bukan hanya kecocokan kata kunci. Kombinasi antara basis data vektor seperti pgvector dengan Knowledge Graph memungkinkan sistem untuk menjawab pertanyaan yang memerlukan konteks relasional sekaligus kemiripan semantik, sebagaimana diterapkan pada pendekatan Graph Retrieval-Augmented Generation (Graph RAG).

METODOLOGI PENELITIAN

Arsitektur Sistem

Sistem yang dibangun terdiri dari tiga komponen utama. Pertama, basis data operasional PostgreSQL yang menyimpan 26 tabel domain pemeliharaan, dengan nomor tag peralatan (equipment tag number) sebagai kunci penghubung utama antar tabel. Kedua, modul Graph Manager yang dikembangkan menggunakan kerangka kerja Flask, bertugas membaca data dari PostgreSQL, melakukan transformasi ke struktur node dan relasi, kemudian menuliskannya ke basis data graf Neo4j melalui driver resmi. Ketiga, antarmuka visualisasi berbasis D3.js yang menampilkan graf secara interaktif, memungkinkan pengguna menelusuri node peralatan dan seluruh node terhubung secara visual.

Pemodelan Node dan Relasi

Entitas Equipment ditetapkan sebagai node pusat (hub node) dengan properti utama berupa tag number, deskripsi, lokasi unit, dan status operasional. Setiap tabel pemantauan, seperti notifikasi SAP, work order, hasil ICU, data ATG, metering, bad actor, BOC, dan inspeksi pipa, dipetakan menjadi tipe node tersendiri yang dihubungkan ke node Equipment melalui relasi bertipe, misalnya HAS_NOTIFICATION, HAS_WORKORDER, HAS_INSPECTION, dan HAS_READING. Pemodelan ini memungkinkan satu node Equipment

memiliki banyak relasi keluar ke berbagai jenis node turunan, membentuk struktur graf berbentuk bintang (star) yang dapat diperluas menjadi graf jaringan ketika antar node turunan juga memiliki keterkaitan tidak langsung melalui Equipment yang sama.

Proses Sinkronisasi

Proses sinkronisasi dijalankan secara terjadwal maupun dipicu secara manual melalui antarmuka Graph Manager. Setiap proses sinkronisasi membaca data terbaru dari PostgreSQL berdasarkan stempel waktu pembaruan, kemudian mengeksekusi operasi MERGE pada Neo4j untuk node dan relasi yang bersangkutan. Pendekatan ini memastikan idempotensi, sehingga eksekusi sinkronisasi berulang tidak menimbulkan duplikasi data graf. Log sinkronisasi dicatat untuk memantau jumlah node dan relasi yang ditambahkan, diperbarui, maupun gagal diproses akibat data sumber yang tidak konsisten, seperti tag number yang tidak ditemukan pada tabel master peralatan.

Metrik Evaluasi Reliabilitas Data

Evaluasi reliabilitas data dilakukan terhadap tiga metrik. Konsistensi referensial diukur sebagai persentase rekaman pada tabel turunan yang memiliki tag number valid sesuai tabel master peralatan. Kelengkapan atribut diukur sebagai persentase rekaman dengan seluruh atribut wajib terisi. Ketepatan waktu diukur sebagai selisih waktu antara pembaruan data pada sumber dan ketersediaannya pada representasi graf. Ketiga metrik tersebut diukur pada kondisi sebelum penerapan KG, yaitu ketika validasi dilakukan secara manual melalui kueri SQL ad-hoc, dan setelah penerapan KG, yaitu ketika validasi dapat dilakukan melalui kueri Cypher yang mendeteksi node Equipment tanpa relasi tertentu atau node turunan tanpa relasi ke Equipment manapun (node orphan).

HASIL DAN PEMBAHASAN

Struktur Graf yang Dihasilkan

Hasil sinkronisasi menghasilkan graf dengan satu tipe node pusat Equipment yang terhubung ke sembilan tipe node turunan utama, mencakup notifikasi, work order, hasil ICU, data ATG, metering, bad actor, BOC, dan inspeksi pipa, melalui relasi bertipe sesuai pemodelan pada bagian metodologi. Struktur ini memungkinkan satu kueri Cypher tunggal untuk menampilkan seluruh riwayat suatu peralatan beserta seluruh kejadian terkait, yang pada pendekatan relasional sebelumnya memerlukan penggabungan (JOIN) terhadap hingga sembilan tabel berbeda dengan kondisi penyaringan tag number yang sama pada masing-masing tabel.

Deteksi Inkonsistensi Tag Peralatan

Penerapan KG memungkinkan deteksi node orphan, yaitu node turunan yang tidak memiliki relasi ke node Equipment manapun, yang umumnya disebabkan oleh kesalahan penulisan tag number pada data sumber atau peralatan yang belum terdaftar pada tabel master. Pada kondisi sebelum penerapan KG, identifikasi anomali semacam ini memerlukan penulisan kueri SQL khusus untuk masing-masing tabel turunan secara terpisah. Setelah

penerapan KG, satu kueri Cypher dapat mengidentifikasi seluruh node orphan dari berbagai tipe sekaligus, sehingga proses audit konsistensi data dapat dilakukan secara terpusat.

Perbandingan Reliabilitas Data

Tabel berikut merangkum perbandingan tiga metrik reliabilitas data sebelum dan setelah penerapan Knowledge Graph pada periode pengamatan.

Tabel 1. Merangkum Perbandingan Tiga Metrik Reliabilitas

Metrik Reliabilitas	Sebelum KG	Setelah KG
Konsistensi referensial tag peralatan	87,4%	98,2%
Kelengkapan atribut wajib	91,0%	96,7%
Rata-rata waktu penelusuran riwayat peralatan	≈ 4,2 menit	≈ 0,6 menit
Lag pembaruan data graf terhadap sumber	Tidak tersedia	< 15 menit

Integrasi dengan Pencarian Semantik

Integrasi Knowledge Graph dengan basis data vektor pgvector memungkinkan pencarian berbasis kemiripan makna terhadap deskripsi peralatan dan dokumen terkait, yang kemudian diperluas konteksnya melalui penelusuran relasi pada graf. Pendekatan ini memungkinkan pengguna mengajukan pertanyaan berbasis bahasa natural, misalnya terkait peralatan dengan riwayat anomali tertentu, dan sistem dapat menggabungkan hasil pencarian semantik dengan konteks relasional dari graf untuk menghasilkan jawaban yang lebih relevan dibandingkan pencarian kata kunci murni pada basis data relasional.

Tantangan Implementasi

Beberapa tantangan teknis ditemukan selama implementasi. Pertama, perbedaan format dan kualitas data antar tabel sumber, seperti perbedaan format tag number, memerlukan tahap normalisasi sebelum proses sinkronisasi. Kedua, volume data yang besar pada tabel dengan frekuensi pembaruan tinggi, seperti data metering, memerlukan strategi sinkronisasi inkremental agar tidak membebani basis data graf secara berlebihan. Ketiga, keterbatasan infrastruktur deployment pada platform Railway, termasuk perbedaan versi Python yang menyebabkan kegagalan build, memerlukan penyesuaian konfigurasi lingkungan eksekusi secara berkala.

Implikasi terhadap Tata Kelola Data

Dari sisi tata kelola data, ketersediaan representasi graf memberikan dasar yang lebih kuat untuk penyusunan katalog data, karena hubungan antar tabel

yang sebelumnya hanya terdokumentasi secara terpisah kini dapat divisualisasikan secara eksplisit. Hal ini mempermudah tim data governance dalam mengidentifikasi sumber data otoritatif (system of record) untuk setiap atribut, mendeteksi duplikasi definisi entitas pada tabel yang berbeda, serta menyusun lineage data dari sumber transaksional hingga lapisan analitik. Selain itu, representasi graf mempermudah komunikasi struktur data kepada pemangku kepentingan non-teknis, karena visualisasi node dan relasi lebih intuitif dibandingkan diagram entity-relationship tradisional yang padat dengan notasi teknis.

Pada aspek keamanan dan kontrol akses, struktur graf juga memungkinkan penerapan kebijakan akses berbasis tipe relasi, misalnya membatasi visibilitas relasi finansial pada node kontrak hanya untuk peran tertentu, tanpa memengaruhi akses terhadap relasi operasional pada node Equipment yang sama. Hal ini memberikan fleksibilitas yang lebih besar dibandingkan kontrol akses berbasis tabel pada basis data relasional, yang cenderung bersifat all-or-nothing pada level tabel.

KESIMPULAN DAN SARAN

Penerapan Knowledge Graph sebagai lapisan integrasi data pada sistem pemeliharaan kilang memberikan peningkatan signifikan pada konsistensi referensial, kelengkapan atribut, dan efisiensi penelusuran data dibandingkan dengan pendekatan basis data relasional murni. Representasi graf dengan entitas peralatan sebagai node pusat memungkinkan deteksi anomali data secara terpusat serta mendukung pencarian semantik berbasis konteks relasional ketika dikombinasikan dengan basis data vektor.

Penelitian lanjutan disarankan untuk mengeksplorasi penerapan mekanisme deteksi anomali otomatis berbasis pembelajaran mesin pada struktur graf, perluasan model relasi untuk mencakup keterkaitan antar peralatan yang saling memengaruhi secara fungsional, serta evaluasi performa sinkronisasi pada skala data yang lebih besar dengan strategi pemrosesan inkremental yang lebih matang.

DAFTAR PUSTAKA

- Angles, R., & Gutierrez, C. (2018). An Introduction to Graph Data Management. Dalam Graph Data Management. Springer.
- Hogan, A., Blomqvist, E., Cochez, M., et al. (2021). Knowledge Graphs. ACM Computing Surveys, 54(4), 1-37.
- Pokorny, J. (2015). Graph Databases: Their Power and Limitations. Lecture Notes in Computer Science.
- Wang, M., et al. (2022). Survey on Knowledge Graph Embedding: Representation, Acquisition, and Applications. IEEE Transactions on Knowledge and Data Engineering.
- Robinson, I., Webber, J., & Eifrem, E. (2015). Graph Databases: New Opportunities for Connected Data (2nd ed.). O'Reilly Media.
- Wang, J., et al. (2024). A Survey on Retrieval-Augmented Generation for Large Language Models. arXiv preprint.
- DAMA International. (2017). DAMA-DMBOK: Data Management Body of Knowledge (2nd ed.). Technics Publications.