

Rancangan Arsitektur Data Warehouse pada Industri Lingkungan & Sustainability pada perusahaan ElSusEnv



Disusun Oleh:

Muhammad Farhan	121450044
Mayada	121450145
Khoirul Anam	122450039
Alvia Asrinda Br Ginting	122450077
Syalaisha Andina Putrinsyah	122450121

**PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
2025**

BAGIAN I

Kebutuhan yang diperlukan oleh misi (ElSusEnv)

Data Warehouse digunakan oleh berbagai industri untuk mendukung analisis bisnis, baik yang berkaitan dengan karyawan maupun konsumen. Hasil analisis ini sangat penting karena dapat membantu para pemangku kebijakan dalam menentukan arah dan keberlangsungan perusahaan, sehingga informasi yang dihasilkan bersifat strategis. Data operasional yang dikumpulkan setiap hari umumnya belum mampu menyediakan informasi yang mendalam dan menyeluruh sesuai kebutuhan analisis. ElSusEnv sebuah perusahaan penyedia energi listrik yang berkomitmen terhadap prinsip keberlanjutan dan ramah lingkungan, berencana membangun arsitektur data warehouse yang mendukung visi dan komitmen tersebut. Dalam proses pengembangan arsitektur data warehouse, diperlukan penentuan pendekatan yang tepat. Pada perusahaan ElSusEnv, telah ditetapkan bahwa pendekatan yang digunakan adalah pendekatan Data-Driven. Pemilihan pendekatan ini didasarkan pada ketersediaan data dan hasil identifikasi awal yang telah dilakukan menggunakan diagram Entity-Relationship (E-R) pada laporan misi sebelumnya. Pendekatan Data-Driven memungkinkan pengidentifikasian semua skema multidimensi secara langsung dari basis data operasional. Terdapat tiga langkah utama dalam spesifikasi pendekatan Data-Driven, yaitu: identifikasi sistem sumber, proses derivasi, dan spesifikasi dokumentasi persyaratan.



Gambar 1. Tahapan dalam keperluan Spesifikasi pendekatan *Data-Driven approach*

I.1 Identifikasi Sumber data

Langkah ini bertujuan untuk mengidentifikasi sistem operasional yang dapat menjadi sumber data bagi data warehouse, dengan fokus awal pada sumber internal. Proses ini melibatkan analisis dokumentasi sistem, namun bila dokumentasi tidak tersedia atau usang, teknik *reverse engineering* dapat digunakan. Selain itu, kualitas dan karakteristik setiap sumber data seperti keandalan, ketersediaan, dan frekuensi pembaruan perlu dievaluasi. Pertemuan dengan penyedia data dilakukan untuk mendokumentasikan informasi tersebut, sehingga pada akhirnya dapat ditentukan sumber terbaik untuk setiap elemen data yang dibutuhkan.

I.2 Proses Derivasi

Identifikasi elemen multidimensi dalam data warehouse seperti fakta, ukuran, dimensi, dan hierarki dapat dilakukan melalui berbagai teknik yang bergantung pada model representasi data operasional (relasional atau entitas-relasi). Fakta dan ukuran diperoleh dari data yang sering

diperbarui, sedangkan dimensi biasanya berasal dari elemen yang statis. Proses ini bisa dilakukan secara otomatis, semi otomatis, atau manual, dengan keterlibatan ahli sistem atau pemahaman yang kuat tentang domain bisnis menjadi kunci untuk akurasi identifikasi, terutama dalam pendekatan manual.

I.3 Spesifikasi syarat dokumen

Seperti pada pendekatan *Business driven*, fase spesifikasi kebutuhan juga harus didokumentasikan. Dokumentasi tersebut harus menjelaskan elemen-elemen dari sistem sumber yang dapat dianggap sebagai fakta, ukuran, dimensi, dan hierarki. Informasi ini akan dimasukkan ke dalam metadata teknis. Selain itu, pada tahap ini sebaiknya melibatkan seorang ahli domain yang dapat membantu dalam mendefinisikan terminologi bisnis untuk elemen-elemen tersebut, serta memberikan penjelasan, misalnya apakah suatu ukuran bersifat aditif, semi-aditif, atau tidak aditif.

Tabel 1. Elemen Multidimensional dengan pendekatan *Data-Driven approach*

Fakta	Ukuran	Dimensi dan Kardinalitas	Hierarki
Energy_Production	Total_Energy_Produced, Operational_Hours, Efficiency, Energy_Loss, CO ₂ _Emission, Maintenance_Cost	Time (1:n), Site (1:n), Energy_Type(1:n), Equipment(1:n), Weather (1:n)	Time(Day→Month→Quarter → Year), Site(Equipment_Site→Facility→Region→Country), Energy_Type(Energy_Subtype→Energy_Type→Energy_Category), Equipment(Component→Equipment→Equipment_Type→Manufacturer), Weather(Weather_Reading→Weather_Condition→Season)
Equipment_Performance	Downtime_Hours, Cost_Maintenance, Fuel_Consumption	Time(1:n), Site(1:n), Equipment(1:n), Fuel_Type(1:n)	Time(Day→Month→Quarter → Year) Site(Equipment_Site→Facility→Region → Country) Fuel_Type(fuel_category)
Operational_Efficiency	Actual_Energy_Output, Power	Time(1:n), Site(1:n), Equipment	Time(Day→Month→Quarter

ency	lanned_Energy_Output,E nergy_Output_Variance, Actual_Cost,Planned_Co st,Cost_Variance,Budget _Utilization,Efficiency_ Ratio	ergy_Type(1:n),Cost_C ategory(1:n),Project_T ype(1:n), Status(1:n), Budget (1:n)	er → Year) Site(Project_Site→Facility → Region → Country) Energy_Type(Energy_Subt ype→Energy_Type→Ener gy_Category) Cost_Category(Cost_Item →Cost_Type→Cost_Categ ory) Project_Type(Project_ID→ Project_Type→Strategic_O bjective) Status(Status_Detail→Statu s → Phase) Budget(Budget_Line→Bud get_Code→Budget_Catego ry)
Environmental	Emission_Amount,Resou rce_Consumed,Waste_ Generated,Waste_Recycl ed,Compliance_Status,Pe nalty_Amount,Emission _Intensity,Water_Usage_ Efficiency	Time(1:n),Site(1:n),Em ission_Type(1:n),Resou rce_Type(1:n),Waste_T ype(1:n),Regulation_C ategory(1:n)	Time(Day → Month → Quarter → Year) Site(Measurement_Locatio n → Facility → Region → Country) Emission_Type(Emission_ Source →Emission_Type→ Emission_Category) Resource_Type(Resource_I tem→Resource_Type→Res ource_Category) Waste_Type(Waste_Item→ Waste_Type→Waste_Categ ory) Regulation_Category(Regu lation_Code→Regulation_ Type→Regulation_Categor y)
Regulatory_Compl iance	Inspection_Count,Violati on_Count,Resolved_Viol	Time(1:n), Site(1:n), Regulation(1:n),	Time(Day→Month→Quart er → Year)

	ations, Penalty_Amount, Compliance_Score, Days_to_Resolve, Follow_Up_Actions	Inspection_Result(1:n)	Site(Inspection_Location → Facility → Region → Country) Regulation(Clause → Regulation_Code → Regulation_Type → Regulatory_Agency) Inspection_Result(Findin → Result_Type → Compliance_Status)
--	--	------------------------	--

BAGIAN II

Skema Konseptual Multidimensi

Granularitas Faktual

Setiap tabel fakta harus didefinisikan dengan jelas sampai pada tingkat pengukuran terkecil yang bermakna bagi proses bisnis. Artinya, setiap baris dalam tabel mencerminkan satu kejadian terperinci misalnya, jumlah energi yang dihasilkan per hari atau per shift untuk masing-masing unit pembangkit, metrik kinerja tiap perangkat per periode, efisiensi operasional per unit produksi pada waktu tertentu, catatan emisi atau kualitas lingkungan per lokasi dan waktu, serta hasil inspeksi kepatuhan per entitas pada saat inspeksi. Penentuan granularitas terendah ini memastikan tidak terjadi campur aduk level agregasi dan memudahkan analisis lintas periode atau entitas.

Kunci Dimensi – Surrogate dan Alternate Key

Di setiap tabel dimensi digunakan *surrogate key* sebagai primary key biasanya bilangan bulat berurutan untuk menjaga kestabilan relasi dengan fakta meski kunci di sistem sumber berubah. Semua *join* antara fakta dan dimensi sebaiknya mengacu pada *surrogate key* ini, bukan kunci bisnis asli. Di samping itu, setiap dimensi juga memuat *alternate key* (business key) untuk merekam identitas entitas di sumber data, sehingga memudahkan pelacakan asal muatan dan penjajaran dengan sistem sumber.

Slowly Changing Dimension (SCD)

Untuk atribut dimensi yang perlu melacak riwayat perubahan seperti alamat fasilitas, spesifikasi perangkat, atau kebijakan lingkungan diterapkan SCD Tipe 2. Setiap kali ada perubahan pada atribut penting, dibuatlah baris baru dengan *surrogate key* baru, sedangkan baris lama tetap disimpan sebagai jejak sejarah, lengkap dengan tanggal mulai dan tanggal berakhir keabsahan. Baris terkini ditandai dengan indikator *current*. Sebaliknya, untuk atribut yang tidak memerlukan

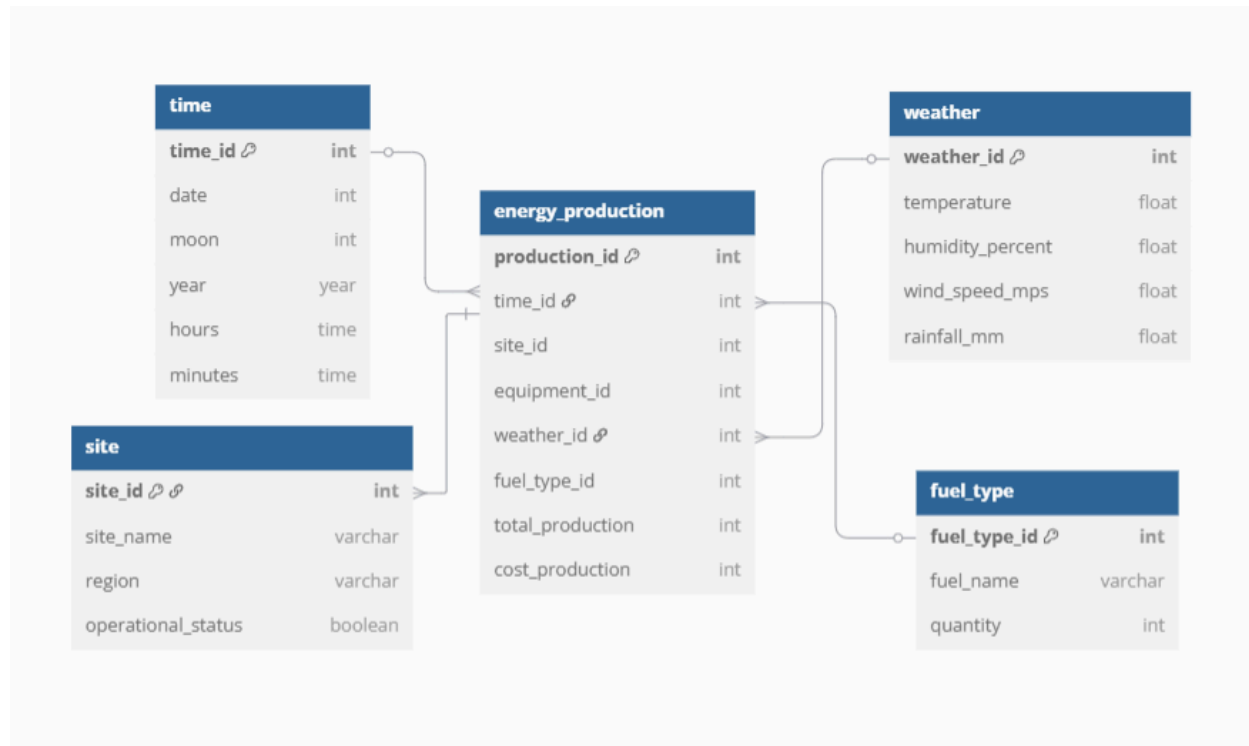
histori, cukup gunakan SCD Tipe 1 dengan menimpa nilai lama agar struktur tabel tetap sederhana.

Dimensi Konform dan Denormalisasi

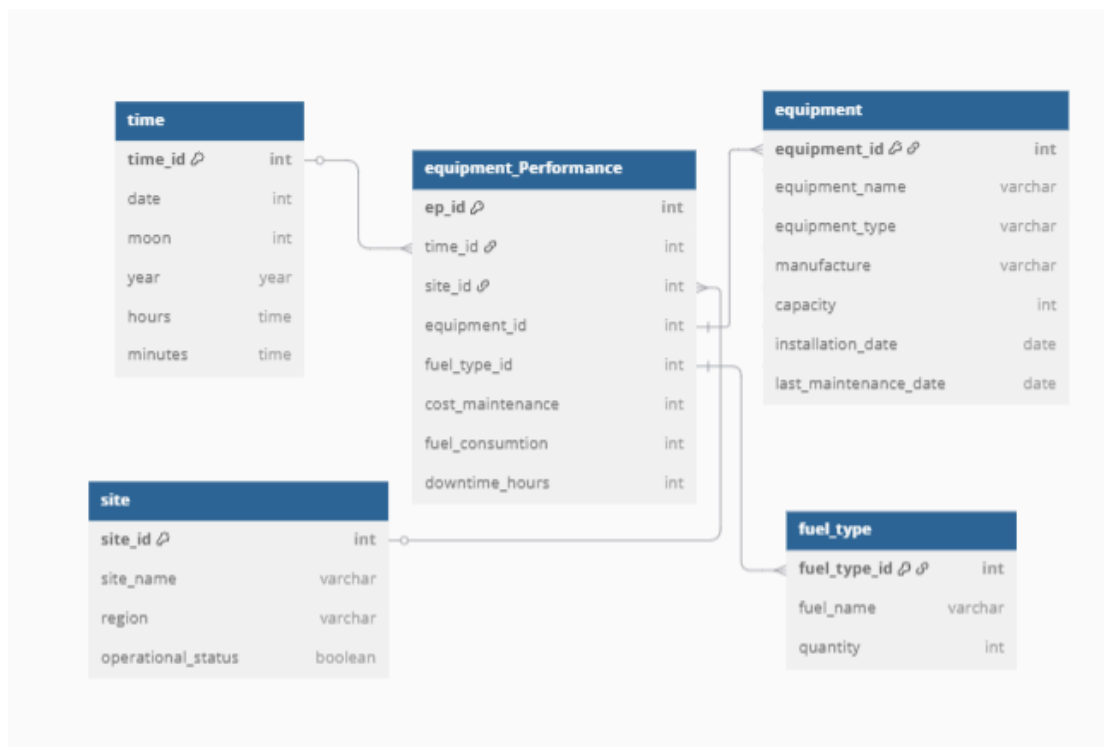
Dimensi yang dipakai bersama oleh beberapa tabel fakta sebaiknya dibuat *conformed*, artinya definisinya konsisten di seluruh skema. Contoh paling umum adalah dimensi waktu, yang harus satu untuk semua fakta agar laporan berkaitan periode sejalan. Demikian pula, jika entitas seperti lokasi atau perangkat dipakai di banyak fakta, buatlah satu dimensi khusus. Agar kinerja *query* optimal, struktur dimensi dibuat denormalisasi (flat/star schema) sehingga setiap fakta hanya perlu *join* langsung ke satu tabel dimensi. Normalisasi multilapis (*snowflake*) hanya dipilih bila benar-benar diperlukan untuk mengurangi redudansi yang signifikan.

Star Schema vs Snowflake & Hierarki Alami

Praktik terbaik merekomendasikan skema bintang karena jumlah *join* minimal dan kecepatan akses lebih baik. Dalam dimensi flat, sertakan atribut yang lengkap misalnya di dimensi geografi sajikan negara, provinsi, dan kota dalam satu tabel agar drill-down mudah dilakukan. Untuk hierarki alami seperti waktu (Tahun→Kuartal→Bulan→Hari), buatlah kolom terpisah dalam satu dimensi menggunakan *surrogate key* misalnya YYYYMMDD agar urutan data sesuai kalender. Hierarki lain, semisal organisasi (Divisi→Unit Kerja) atau lokasi (Negara→Provinsi→Kota), juga dapat diterapkan serupa untuk mendukung analitik mendalam tanpa melanggar prinsip desain dimensional. Dengan struktur ini, skema ElSusEnv menjadi lebih sederhana, cepat, dan konsisten untuk berbagai kebutuhan pelaporan dan analisis.



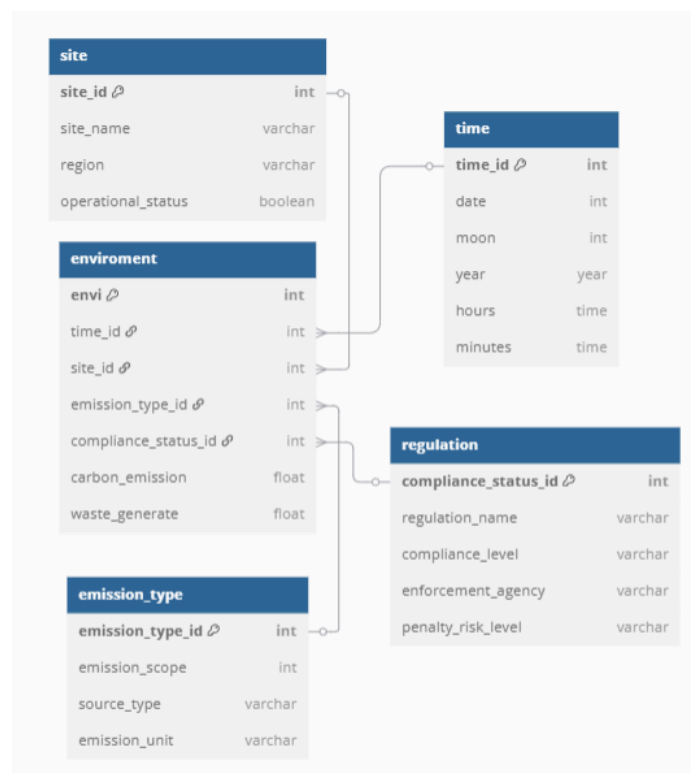
Gambar 2. Star Schema Produksi



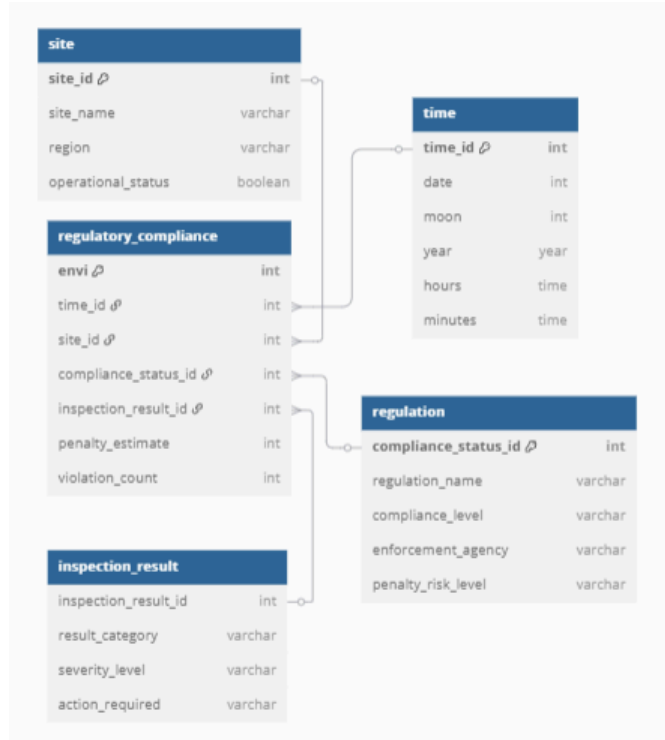
Gambar 3. Star Schema Kinerja peralatan



Gambar 4. Star Schema Operational



Gambar 5. Star Schema lingkungan



Gambar 6. Star Schema Kepatuhan

BAGIAN III

Penjelasan Tiap Komponen

3.1 Star Schema Produksi

3.1.1 Tabel Fakta

Menyimpan data utama hasil produksi energi.

production_id	Primary key tabel produksi.
time_id	Foreign key ke tabel waktu.
site_id	Foreign key ke tabel lokasi.
equipment_id	ID peralatan.
weather_id	Foreign key ke tabel cuaca.
fuel_type_id	Foreign key ke tabel bahan bakar.
total_production	Total energi yang dihasilkan.
cost_production	Biaya produksi energi.

3.1.2 Tabel Dimensi

a. Time

Dimensi waktu penting untuk analisis tren dan perbandingan antar periode.

time_id	Primary key.
date	Tanggal produksi.
moon	Bulan.
year	Tahun.
hours	Jam.
minutes	Menit.

b. Site

Berguna untuk analisis kinerja per lokasi dan per wilayah.

site_id	Primary key.
site_name	Nama lokasi.
region	Wilayah.
operational_status	Status operasional.

c. Weather

Berguna dalam analisis pengaruh kondisi lingkungan terhadap produksi.

weather_id	Primary key.
temperature	Suhu (°C).
humidity_percent	Kelembapan (%).
wind_speed_mps	Kecepatan angin (meter per detik).
rainfall_mm	Curah hujan (mm).

d. Fuel_type

Dapat digunakan untuk menghitung efisiensi atau emisi produksi.

fuel_type_id	Primary key.
fuel_name	Nama jenis bahan bakar.
quantity	Jumlah bahan bakar yang digunakan.

3.2 Star Schema Kinerja Peralatan

3.2.1 Tabel Fakta

Menyimpan data numerik terkait performa setiap peralatan.

ep_id	Primary key dari record kinerja peralatan.
time_id	Foreign key ke dimensi waktu.
site_id	Foreign key ke dimensi lokasi/site.
equipment_id	Foreign key ke dimensi peralatan.
fuel_type_id	Foreign key ke dimensi bahan bakar.
cost_maintenance	Biaya pemeliharaan.
fuel_consumption	Konsumsi bahan bakar.
downtime_hours	Lama waktu alat tidak beroperasi.

3.2.2 Tabel Dimensi

a. Time

Dimensi waktu penting untuk analisis tren dan perbandingan antar periode.

time_id	Primary key.
date	Tanggal produksi.
moon	Bulan.
year	Tahun.
hours	Jam.
minutes	Menit.

b. Site

Berguna untuk analisis kinerja per lokasi dan per wilayah.

site_id	Primary key.
site_name	Nama lokasi.
region	Wilayah.
operational_status	Status operasional.

c. Equipment

Berguna untuk melihat apakah umur alat, jenis, atau produsen tertentu mempengaruhi performa.

equipment_id	Primary key.
--------------	--------------

equipment_name	Nama alat.
equipment_type	Jenis alat.
manufacture	Produsen.
capacity	Kapasitas alat.
installation_date	Tanggal pemasangan alat.
last_maintenance_date	Tanggal perawatan terakhir.

d. Fuel_type

Dapat digunakan untuk menghitung efisiensi atau emisi produksi.

fuel_type_id	Primary key.
fuel_name	Nama jenis bahan bakar.
quantity	Jumlah bahan bakar yang digunakan.

BAGIAN IV

Justifikasi Desain Konseptual

Pendekatan desain konseptual dengan *Data Driven approach* berarti bahwa inisiasi pembangunan Data Warehouse dilakukan berdasarkan analisis struktur dan isi dari sistem operasional yang sudah ada. Pendekatan ini cocok digunakan ketika dokumentasi sistem tersedia dan data historis dapat diakses dengan baik. Setelah analisis dilakukan, langkah penting berikutnya adalah mengidentifikasi tabel-tabel fakta yang relevan dengan kebutuhan analisis bisnis. Tabel fakta dipilih berdasarkan aktivitas yang sering terjadi dan memiliki nilai kuantitatif (measure) yang dapat digunakan untuk evaluasi kinerja. Selain itu, proses ini juga melibatkan identifikasi dimensi yang memungkinkan pembentukan skema multidimensi secara lengkap, seperti star schema. Pemilihan fakta dan dimensi dalam desain konseptual harus mempertimbangkan kualitas data, frekuensi pembaruan, serta keterkaitan antar entitas di sistem operasional. Dengan demikian, pendekatan ini tidak hanya bersifat teknis tetapi juga mendukung kebutuhan informasi strategis perusahaan berdasarkan realitas data yang tersedia.

BAGIAN V

Kesesuaian dengan sumber data

Arsitektur data warehouse ElSusEnv dirancang berdasarkan pendekatan dimensional modeling dengan penerapan skema bintang (star schema) dan sebagian snowflake. Desain ini memungkinkan pengorganisasian data yang efisien dan terstruktur, serta mendukung kebutuhan eksplorasi dan analisis

multidimensi. Setiap kebutuhan analitik utama telah dipetakan secara eksplisit terhadap struktur data warehouse melalui relasi antara tabel fakta dan tabel dimensi. Validitas pemetaan tersebut ditunjukkan dalam Tabel 5.1.

Tabel 5.1. Pemetaan Kebutuhan Analitik terhadap Sumber Data Warehouse

No	Kebutuhan Analitik	Tabel Fakta Terkait	Tabel Dimensi Terkait	Justifikasi Kesesuaian
1	Monitoring produksi energi per jenis, lokasi, dan waktu	fact_energy_production	dim_site, dim_time, dim_energy_type, dim_weather	Produksi energi direkam dalam atribut energy_output dan ditautkan dengan lokasi, jenis energi, serta parameter cuaca melalui relasi antar tabel.
2	Evaluasi efisiensi operasional dan biaya	fact_operational_efficiency	dim_project_type, dim_time, dim_cost_category, dim_status, dim_budget, dim_site	Nilai efisiensi dan biaya diidentifikasi melalui atribut efficiency_score, budget_id, dan cost_category_id, dengan dukungan dimensi proyek, status operasional, lokasi, dan waktu.
3	Evaluasi performa alat dan penyimpanan energi	fact_equipment_performance	dim_equipment, dim_battery, dim_storage_type, dim_maintenance_type, dim_time, dim_site	Struktur ini mendukung analisis performa teknis peralatan berdasarkan jenis penyimpanan energi, metode pemeliharaan, dan

				spesifikasi baterai yang digunakan.
4	Audit kepatuhan terhadap regulasi	fact_regulatory_compliance	dim_regulation, dim_inspection_result, dim_status, dim_regulation_category, dim_time, dim_site	Skor kepatuhan dan hasil inspeksi terhubung langsung ke sistem klasifikasi regulasi serta status operasional unit, memungkinkan evaluasi kepatuhan berbasis wilayah dan waktu.
5	Jejak emisi dan dampak lingkungan	fact_environmental	dim_emission_type, dim_waste_type, dim_resource_type, dim_time, dim_site, dim_regulation_category	Setiap emisi dan limbah diidentifikasi secara terstruktur melalui kategorisasi tipe emisi, limbah, dan sumber daya, serta terhubung dengan lokasi dan kategori regulasi terkait.
6	Tren biaya operasional jangka waktu tertentu	fact_operational_efficiency	dim_time, dim_cost_category, dim_budget	Kebutuhan analitik ini difasilitasi melalui pemetaan biaya terhadap waktu dan kategori pengeluaran yang direpresentasikan dalam struktur multidimensi.

7	Dampak kondisi cuaca terhadap produksi energi	fact_energy_production	dim_weather, dim_time, dim_site, dim_energy_type	Pengaruh suhu, kelembaban, dan kecepatan angin terhadap output energi dapat diestimasi secara longitudinal dan spasial melalui integrasi dimensi cuaca dan produksi.
8	Hubungan performa alat dan kepatuhan regulasi	fact_equipment_performance, fact_regulatory_compliance	dim_equipment, dim_maintenance_type, dim_time, dim_site, dim_regulation, dim_inspection_result	Korelasi antara performa teknis alat dan skor kepatuhan dapat diidentifikasi melalui foreign key yang seragam pada waktu dan lokasi dalam dua tabel fakta tersebut.

Struktur data warehouse dibangun dengan prinsip integritas referensial dan granularitas tinggi. Tabel fakta mencatat kejadian atomik yang terukur, sedangkan tabel dimensi menyediakan atribut deskriptif bersifat hierarkis dan klasifikatif. Pendekatan ini mendukung eksplorasi analitik dalam berbagai tingkat agregasi. Dimensi teknis seperti dim_equipment, dim_storage_type, dim_maintenance_type, dan dim_battery diintegrasikan untuk memungkinkan analisis presisi performa sistem energi dan pemodelan hubungan sebab-akibat pada skala mikro.

Dalam aspek regulasi, fact_regulatory_compliance mencakup atribut compliance_score yang terhubung dengan dimensi hasil inspeksi, jenis regulasi, dan kategori regulasi, memungkinkan analisis kepatuhan secara deskriptif, diagnostik, dan preskriptif. Dimensi waktu disusun dengan tingkat granularitas harian hingga tahunan, kompatibel untuk analisis musiman, prediksi deret waktu, dan estimasi berbasis waktu lainnya.

Skema star dan snowflake diterapkan secara selektif untuk menyeimbangkan efisiensi penyimpanan dan kecepatan query. Dimensi dengan struktur hierarkis, seperti dim_regulation dan dim_regulation_category, dinormalisasi tanpa mengurangi aksesibilitas data. Hubungan antar tabel fakta difasilitasi oleh foreign key konsisten seperti site_id dan time_id, sehingga memungkinkan analisis lintas-proses yang konsisten dan mendukung komputasi OLAP secara optimal.