

Laporan Analisis Kebutuhan Bisnis dan Teknis Perancangan Data Warehouse pada Industri Lingkungan & Sustainability



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

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

1. Pendahuluan

1.1. Tujuan Proyek

Pembangunan Data Warehouse pada PT. ElSusEnv bertujuan meningkatkan kinerja perusahaan dengan menciptakan sistem data terpusat dan terintegrasi. Proyek ini dirancang untuk memberdayakan manajemen dalam mengambil keputusan strategis sebagai upaya dalam meningkatkan efisiensi operasional dan mengoptimalkan alokasi sumber daya, serta mendorong keunggulan kompetitif melalui analisis data yang komprehensif dan akurat.

1.2. Ruang Lingkup

Proyek ini mencakup proses pengumpulan, integrasi, penyimpanan, dan penyajian data dari berbagai domain operasional utama perusahaan listrik berbasis energi terbarukan. Fokus utama proyek ini terdapat pada lima domain strategis, yaitu produksi listrik berdasarkan kondisi cuaca dan lingkungan, evaluasi kinerja pembangkit, analisis layanan dan kepuasan pelanggan, monitoring biaya operasional dan efisiensi proyek, serta kepatuhan terhadap regulasi teknis dan standar keselamatan. Data mentah dari setiap domain akan diproses melalui pipeline ETL dan disimpan dalam struktur data yang terintegrasi untuk mendukung analitik dan pelaporan secara menyeluruh.

1.3. Latar Belakang

Industri lingkungan dan sustainability saat ini menjadi salah satu sektor yang sangat strategis dan berkembang pesat di tengah tantangan global perubahan iklim dan kebutuhan transisi menuju energi bersih. Perusahaan penghasil listrik yang khususnya bergerak di bidang energi terbarukan, dihadapkan pada tekanan untuk tidak hanya menghasilkan energi yang ramah lingkungan, tetapi juga memastikan efisiensi operasional, kepatuhan regulasi, serta pelaporan dampak lingkungan yang transparan dan akurat. Di sisi lain, keberhasilan dalam industri ini sangat bergantung pada kemampuan untuk memanfaatkan data secara optimal dari berbagai sumber yang sangat beragam, mulai dari data teknis produksi, kondisi cuaca, hingga informasi finansial dan kepuasan pelanggan.

Di tingkat operasional, perusahaan ini mengelola aset dan fasilitas yang tersebar secara geografis dengan berbagai kondisi lingkungan yang dinamis, sehingga membutuhkan monitoring produksi listrik yang akurat serta evaluasi kinerja berdasarkan faktor-faktor lingkungan dan teknis. Selain itu, kebutuhan untuk memenuhi standar keselamatan dan regulasi teknis yang ketat serta pelaporan aspek ESG (Environmental, Social, and Governance) membuat pengelolaan data menjadi sangat kompleks. Data yang saat ini tersebar dalam berbagai sistem dan format yang tidak terintegrasi menyebabkan kesulitan dalam memperoleh informasi yang holistik dan real-time, sehingga berpotensi menghambat pengambilan keputusan yang cepat dan tepat.

Untuk mengatasi tantangan tersebut, perusahaan membutuhkan pembangunan sistem Data Warehouse yang mampu mengintegrasikan seluruh data dari berbagai sumber internal maupun eksternal dalam satu platform terpadu. Pendekatan ini memungkinkan pengolahan data secara menyeluruh dan konsisten, mendukung pengembangan dashboard analitik, laporan keberlanjutan, serta analisis kinerja teknis dan finansial secara real-time. Dengan demikian, perusahaan dapat meningkatkan efisiensi operasional, memastikan kepatuhan regulasi, dan memperkuat posisi sebagai pelaku industri energi terbarukan yang berorientasi pada keberlanjutan dan berbasis data.

2. Profil Perusahaan

2.1. Deskripsi Perusahaan

PT. ElSusEnv adalah perusahaan yang bergerak di bidang pengembangan dan pengoperasian pembangkit listrik dengan memanfaatkan sumber energi yang dapat diperbaharui secara alami dan berkelanjutan, seperti energi matahari, angin, air, panas bumi, biomassa, dan bioenergi sebagai bagian dari kontribusinya terhadap transisi energi bersih nasional. Perusahaan berfokus pada penyediaan listrik yang ramah lingkungan, efisien, dan andal, dengan berbagai proyek pembangkit listrik yang tersebar secara geografis. Dalam menjalankan operasionalnya, PT. ElSusEnv juga berkomitmen terhadap aspek Environmental, Social, and Governance (ESG). Komitmen ini diwujudkan melalui tiga pilar utama: penggunaan sumber energi yang ramah lingkungan, kemitraan dengan pemasok yang menerapkan praktik keberlanjutan, serta penerapan tata kelola perusahaan yang transparan dan akuntabel.

Namun, dalam operasionalnya, PT. ElSusEnv menghadapi tantangan besar dalam pengelolaan data yang masih tersebar di berbagai sistem yang terpisah (siloed), seperti SCADA, sensor IoT, ERP, laporan inspeksi teknis, dan data cuaca eksternal. Ketidakterpaduan format, keterbatasan akses antar sistem, serta dominasi proses manual menghambat integrasi data, memperlambat analisis, dan menyulitkan pelaporan kinerja teknis, efisiensi proyek, serta kepatuhan regulasi secara menyeluruh. Untuk mengatasi hal tersebut, perusahaan membutuhkan pembangunan sistem Data Warehouse sebagai platform analitik terpadu yang mampu mengintegrasikan seluruh data operasional dan strategis.

2.2. Stakeholder

Keterlibatan aktif dari berbagai pemangku kepentingan sangat penting dalam mendorong efektivitas dan efisiensi operasional perusahaan. Berikut adalah daftar pemangku kepentingan utama beserta tujuan bisnis spesifik yang menjadi fokus perhatian mereka.

Tabel 1. Stakeholder dan tujuan bisnis

Stakeholder	Tujuan Bisnis
CEO	Memantau dan menilai kinerja keseluruhan perusahaan untuk pengambilan keputusan strategis yang tepat.
CFO	Mengendalikan biaya operasional dan memantau efisiensi proyek untuk memastikan keberlanjutan finansial perusahaan.
Departemen Manajemen Pelanggan	Meningkatkan mutu layanan dan kepuasan pelanggan melalui pemantauan gangguan listrik dan analisis profil pelanggan.
Departemen K3	Memastikan kepatuhan perusahaan terhadap regulasi teknis dan standar keselamatan kerja.

3. Metodologi

3.1. Ringkasan Proyek

Proyek perancangan Data Warehouse perusahaan ElSusEnv dirancang dengan pendekatan data-driven, yang bertujuan untuk mengoptimalkan pengambilan keputusan berdasarkan data yang dihasilkan langsung dari operasional perusahaan. Data yang dikumpulkan dari berbagai sumber, seperti sistem SCADA, sensor IoT, laporan produksi, dan data eksternal tentang kondisi cuaca, akan diolah secara sistematis untuk memberikan wawasan yang lebih mendalam. Desain menggunakan Three-Tier Architecture memastikan pemisahan yang jelas antara penyimpanan data, logika aplikasi, dan antarmuka pengguna. Dengan star schema pada lapisan data, sistem ini akan memfasilitasi analisis yang lebih efisien, mendukung perencanaan dan operasional yang lebih baik, serta mempercepat respons terhadap dinamika pasar dan operasional energi terbarukan.

3.2. Metode Proyek

Proyek perancangan Data Warehouse perusahaan ElSusEnv akan diimplementasikan dengan menggunakan pendekatan Waterfall, yang mencakup tahapan-tahapan yang terstruktur dan berurutan. Proyek ini akan berfokus pada pendekatan bottom-up, di mana Data Mart dibangun terlebih dahulu untuk masing-masing departemen sebelum akhirnya diintegrasikan ke dalam Data Warehouse yang lebih besar. Dengan pendekatan ini, analisis dan laporan yang lebih terfokus dan relevan dapat segera dilakukan untuk masing-masing domain operasional, sementara integrasi keseluruhan akan memungkinkan analisis lintas divisi dan mendalam. Pada tahap akhir, sistem akan melalui pengujian untuk memastikan integrasi dan aliran data berjalan dengan baik, serta memenuhi standar operasional yang ditetapkan.

3.3. Tools dan Teknologi

Ada beberapa *tools* atau alat yang digunakan pada tugas ketiga ini yaitu :

- *Microsoft SQL Server*

- *SQL Server Integration Services (SSIS)*
- *SQL Server Analysis Services (SSAS)*
- *SQL Server Reporting Services (SSRS)*

Microsoft SQL Server yang berfungsi sebagai sistem manajemen basis data relasional (RDBMS). *SQL Server* digunakan untuk menyimpan, mengelola, dan mengambil data secara paralel dengan bahasa pemrograman *SQL* (Structured Query Language). Komponen utama *SQL Server* yang digunakan adalah Database Engine untuk pemrosesan kueri dan manajemen basis data, serta Data Management System (SSMS) sebagai antarmuka pengguna grafis (GUI) dengan manajemen basis data.

SQL Server Integration Services (SSIS) sebagai alat integrasi, transform, load (ETL). *SSIS* digunakan untuk mengambil data dari berbagai sumber seperti lembar kerja Excel, CSV, atau basis data lainnya, lalu membersihkannya dan memuatnya ke gudang data. Proses ETL terdiri dari Aliran Kontrol dan Aliran Data untuk mengatur dan memindahkan data dari sumber ke tujuan secara efisien.

SQL Server Analysis Services (SSAS) digunakan dalam pengembangan model dan tabel berbasis OLAP (Online Analytical Processing) yang memfasilitasi analisis dan penelitian data. Pendekatan ini memungkinkan terciptanya kubus, hierarki, dan metrik yang membantu meningkatkan kinerja bisnis. Dalam studi ini, contoh tersebut digunakan karena cocok untuk alat analisis seperti Power BI.

SQL Server Reporting Services (SSRS) digunakan sebagai alat untuk membuat dan menampilkan laporan. *SSRS* menyediakan beberapa komponen yang berguna seperti Report Builder, Report Builder, dan Report Builder yang memungkinkan pengguna bisnis membuat laporan secara teknis dan visual. Laporan disusun dalam format bahasa definisi laporan (RDL) berbasis XML.

Dalam pendekatan modern, studi ini juga mempertimbangkan komputasi awan untuk optimasi sinaps dan optimasi sumber daya. Kedua teknologi ini digunakan sebagai bagian dari integrasi data dalam lingkungan cloud dan mendukung konsep arsitektur data modern seperti Data Lakehouse dan Medallion Architecture, yang mengatur data ke dalam kategori perunggu (data mentah), perak (data bersih), dan emas (data siap analisis).

4. Perancangan Data Warehouse

4.1. Analisis Kebutuhan Bisnis

PT. ElSusEnv merupakan perusahaan yang bergerak di bidang Pembangkit Listrik energi terbarukan, yang menghadapi tantangan besar dalam mengelola dan menganalisis data operasional serta lingkungan secara efektif. Sifat produksi energi terbarukan yang sangat bergantung pada faktor eksternal seperti cuaca dan kondisi lingkungan membuat fluktuasi output menjadi hal yang umum. Selain itu, aset-aset pembangkit listrik yang tersebar secara geografis menambah kompleksitas dalam proses pemantauan dan pengambilan keputusan. Perusahaan juga menerapkan ESG dalam setiap proses perusahaan. Sumber data yang dimiliki sangat beragam, mulai dari sistem SCADA, sistem CMM, sensor IoT, data cuaca, hingga sistem ERP yang membuat integrasi data menjadi tantangan tersendiri. Untuk itu, diperlukan sistem data warehouse yang mampu mengintegrasikan dan mengelola data secara terpadu guna mendukung kebutuhan analitik dan pelaporan. Berikut adalah kebutuhan analitik yang telah diidentifikasi:

1. Analisis produksi listrik : Menilai produksi listrik harian, bulanan, dan tahunan per lokasi PLTS berdasarkan kondisi cuaca dan lingkungan.
 - a. Produksi listrik per hari/bulan/tahun per jenis pembangkit listrik energi terbarukan.
 - b. Produksi tertinggi dan terendah per jenis pembangkit listrik energi terbarukan dalam periode tertentu.
 - c. Korelasi produksi listrik dengan parameter cuaca.
2. Evaluasi kinerja pembangkit listrik: Analisis performa teknis pembangkit untuk mengidentifikasi potensi perbaikan operasional.
 - a. Capacity factor (CF) dan Capacity Utilization factor (CUF) per Pembangkit listrik di setiap periode waktu tertentu.
 - b. Tingkat ketersediaan dan keandalan (availability & reliability) tiap pembangkit.
3. Analisis Layanan dan Kepuasan Pelanggan: evaluasi kualitas teknis layanan listrik (berdasarkan gangguan) dengan analisis karakteristik dan perilaku pelanggan untuk meningkatkan kepuasan dan efektivitas pelayanan.
 - a. Distribusi konsumsi listrik berdasarkan segmen pelanggan (rumah tangga, industri, usaha, umum (*public*)).
 - b. Pelanggan dengan konsumsi tertinggi dan tingkat pertumbuhan konsumsi tertinggi.
 - c. Pola konsumsi energi per hari/bulan dan tren pertumbuhan konsumsi per segmen pelanggan.
 - d. Nilai SAIDI dan SAIFI rata-rata per wilayah/segmen pelanggan.
 - e. Wilayah dan segmen pelanggan dengan gangguan tertinggi (frekuensi dan durasi padam).
4. Monitoring Biaya Operasional dan Efisiensi Proyek: Pemantauan biaya aktual dan efisiensi ekonomi proyek.
 - a. Biaya operasional aktual per pembangkit listrik dan per proyek per bulan.
 - b. Proyek dengan biaya per MWh tertinggi dan terendah.

- c. Analisis ROI tiap proyek pembangkit.
5. Evaluasi Kepatuhan Regulasi Teknis dan Keselamatan: Monitoring penerapan standar keselamatan dan kepatuhan terhadap regulasi teknis dan standar keselamatan.
 - a. Tingkat kepatuhan terhadap standar teknis dan inspeksi berkala.
 - b. Tren peningkatan jumlah tenaga teknik bersertifikat sebagai indikator penguatan kapasitas SDM dan pemenuhan regulasi.

4.2. Desain Konseptual

Desain konseptual untuk data warehouse ini didasarkan pada analisis kebutuhan analitik yang telah dilakukan sebelumnya, di mana elemen-elemen dari sistem sumber diidentifikasi dan diklasifikasikan sesuai dengan kebutuhan laporan dan analisis yang spesifik. Elemen-elemen ini mencakup fakta, ukuran, dimensi, dan hierarki yang akan dimasukkan ke dalam metadata teknis. Tabel fakta dan dimensi akan diklasifikasikan sesuai dengan hierarki dan level yang dibutuhkan, serta akan melibatkan ahli domain untuk memastikan definisi terminologi bisnis yang tepat, termasuk apakah ukuran tersebut bersifat aditif, semi-aditif, atau tidak aditif. Struktur ini akan dirancang dalam skema bintang pada tahap desain logikal dan fisik. Struktur tersebut memungkinkan pengolahan data yang efisien dan memudahkan akses bagi pengguna non-teknis.

Tabel 2. Elemen Multidimensional dengan pendekatan Data-Driven

Fakta	Ukuran	Dimensi dan Kardinalitas	Hierarki
Energy_Production	Total_production, Total_Installed_capacity, Unit_Power_Plant	Time (1:n), Site (1:n), Energy_Type(1:n), Equipment(1:n), Weather (1:n)	Time(Day→Month→Quarter → Year) Site(Site_name→Region →Country) Weather(Temperature→Weather_Condition→Season)
Equipment_Performance	Downtime_Hours, Cost_Maintenance, Capacity_factor_Percentage, Capacity_Utilization_factor_Percentage	Time(1:n), Site(1:n), Equipment(1:n), Energy_Type(1:n)	Time(Day→Month→Quarter → Year) Site(Equipment_Site→Region → Country)
Customer_Service	Interrupted_Duration, SAIDI_Index, SAIFI_Index, Energy_Consumption	Time(1:n), Site(1:n), Equipment(1:n), Customer(1:n)	Time(Day→Month→Quarter → Year) Site(Project_Site→Region → Country) Customer(Customer_Type →Segment_name)

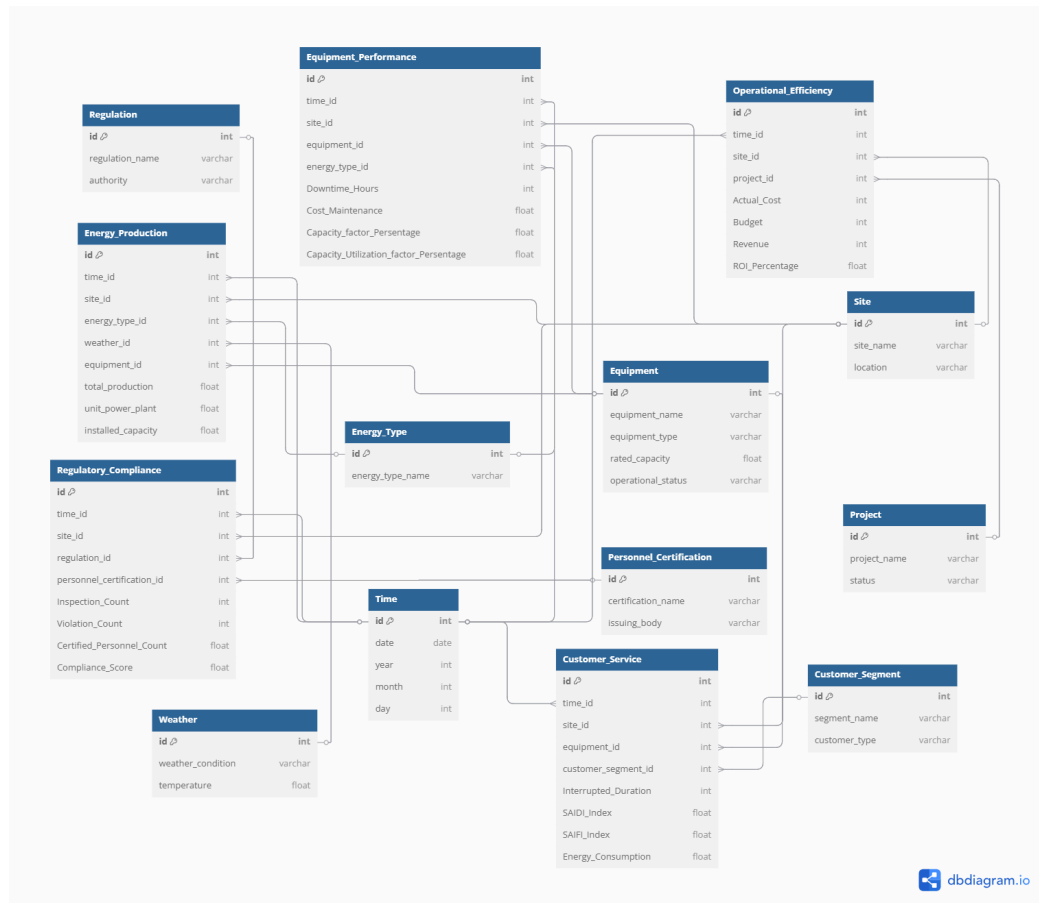
Operational_Efficiency	Actual_Cost, Budget, Revenue, ROI_Percentage	Time(1:n), Site(1:n), project (1:n)	Time(Day → Month → Quarter → Year) Site(Project_Site → Region → Country) Project(Project_name → Project_Category → Department)
Regulatory_Compliance	Inspection_Count, Violation_Count, Certified_Personnel_Count, Compliance_Score	Time(1:n), Site(1:n), Regulation(1:n), Personnel_Certification (1:n)	Time(Day → Month → Quarter → Year) Site(Inspection_Location → Facility → Region → Country) Regulation(enforcement_agency → Regulation_name → compliance_level → penalty_risk_level)

Berdasarkan masalah dan tujuan bisnis, maka dapat diidentifikasi sejumlah kebutuhan bisnis yang dirincikan pada tabel berikut.

Tabel 3. Fakta dan dimensi sesuai kebutuhan

Kebutuhan	Fakta	Dimensi
Analisis produksi listrik per jenis pembangkit listrik	Energy_Production	Time, Site, Energy_Type, Weather, Equipment
Evaluasi kinerja pembangkit listrik	Equipment_Performance	Time, Site, Equipment, Energy_Type
Analisis layanan dan kepuasan pelanggan	Customer_Service	Time, Site, Customer_Segment
Monitoring biaya operasional dan efisiensi proyek	Operational_Efficiency	Time, Site, Project
Evaluasi kepatuhan regulasi teknis dan keselamatan	Regulatory_Compliance	Time, Site, Regulation,

Setelah mengidentifikasi tabel fakta dan dimensi yang dibutuhkan, Seluruh komponen digambarkan dalam bentuk Entity Relationship Diagram (ERD), sebagai berikut.

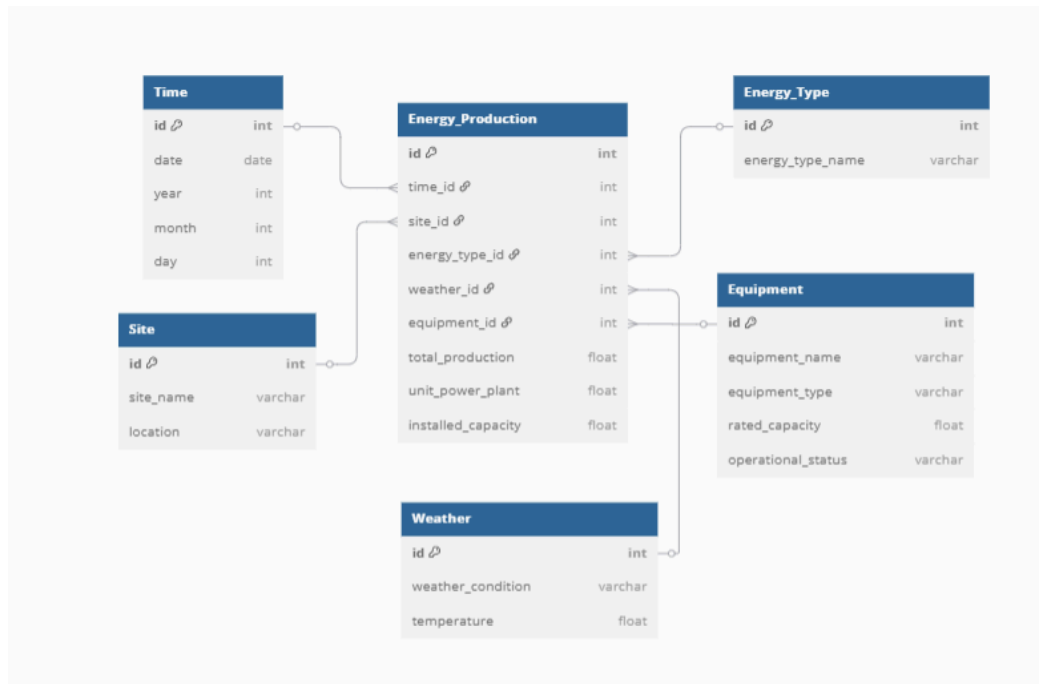


Gambar 1. ERD

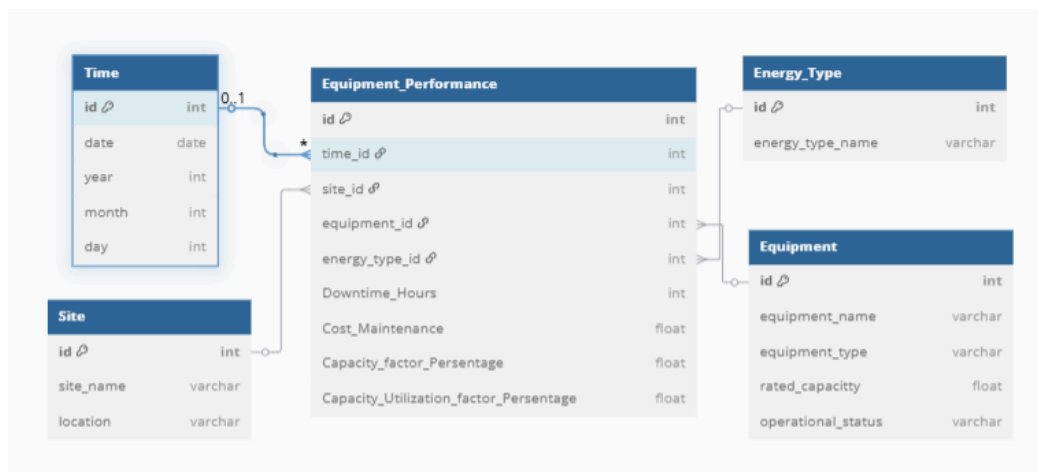
ERD RD ini akan menggambarkan hubungan antara tabel fakta dan dimensi, serta struktur data yang dibutuhkan untuk mendukung analisis dan laporan yang efisien. Dalam ERD, kita akan memetakan tabel fakta sebagai pusat dari diagram, yang dihubungkan dengan tabel dimensi melalui foreign keys. Hal ini memungkinkan untuk pemetaan data secara lebih terorganisir dan mendukung analisis lintas dimensi.

4.3. Desain Logikal dan Fisik

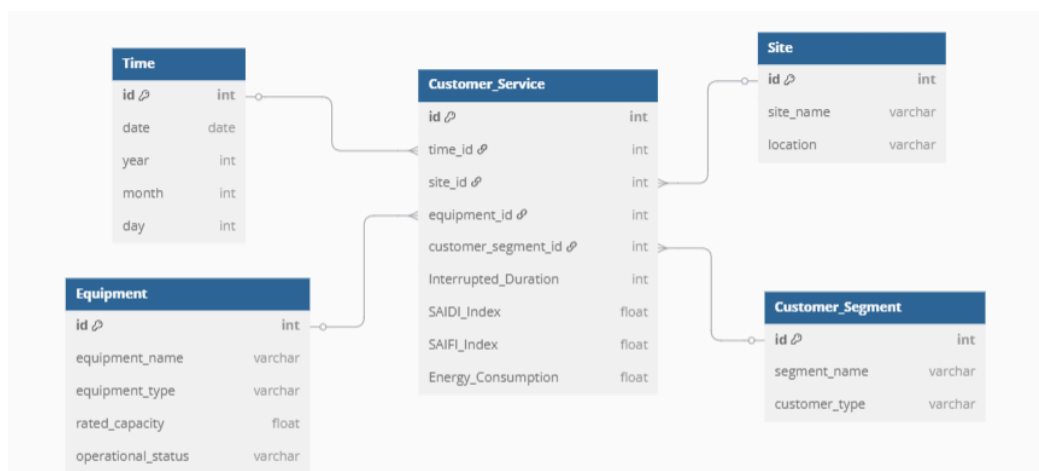
Desain logikal dan fisik merupakan tahap lanjutan dalam pembangunan data warehouse yang berfokus pada perancangan struktur data secara sistematis dan terintegrasi. Pada tahap ini, Seluruh komponen kebutuhan analitik yang diidentifikasi sebelumnya dirancang dalam kerangka skema bintang (star schema) untuk memastikan performa yang optimal dalam proses pengambilan keputusan.



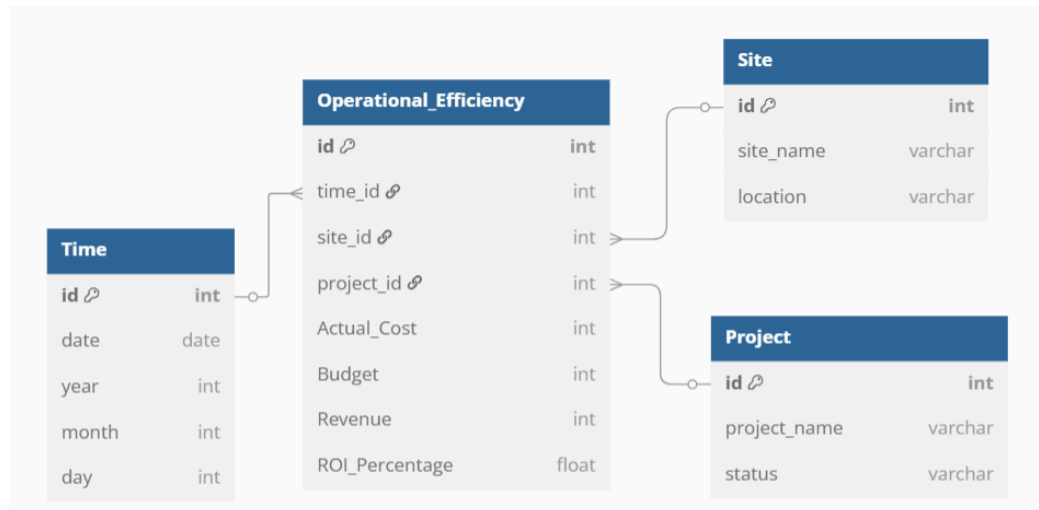
Gambar 2. Star Schema Produksi



Gambar 3. Star Schema Kinerja Peralatan



Gambar 4. Star Schema Layanan



Gambar 5. Star Schema Operasional



Gambar 6. Star Schema Regulasi

Tabel 4. Detail tabel fakta dan dimensi

Fakta	Measure	Dimensi	Penjelasan
Energy_Production	1. Total_Production : total energi listrik yang dihasilkan	Time: date	Menunjukkan kapan aktivitas produksi terjadi.
		Site: Site_name,	Lokasi Pembangkit Listrik .

	(Kwh atau MWh)	Locatio (Region)	
	2. Unit_power_plant: Jumlah unit pembangkitt listrik	Energy_Type: Energy_type_name	Jenis pembangkit listrik energi terbarukan contoh: PLTS (surya) , PLTBM (biomassa), dll
	3. Installed_capacity: kapasitas terpasang dari unit pembangkit tersebut (dalam MW atau kW).	Weather: Weather_condition , Temparetaur	Kondisi saat produksi
		Equipment: Equipment_name, Equipment_type, Rated_capacity, Operational_status	Detail peralatan (jenis turbin, panel surya, inverter, dsb)
Equipment_Performance	1. Down_time_hours : total waktu peralatan tidak beroperasi	Time: date	Waktu pengamatan atau pelaporan kinerja
		Site: Site_name, Location	Lokasi pembangkit tempat peralatan dipasang
	2. Cost_maintenance : biaya yang dikeluarkan untuk perawatan peralatan	Equipment: Equipment_name, Equipment_type, Rated_capacity, Operational_status	Jenis dan identitas peralatan (misalnya: turbin T-01, inverter I-05)
	3. Capacity_factor_persentase: rasio antara produksi aktual terhadap kapasitas maksimum		
	4. Capacity_utilization_factor_persentase: tingkat pemanfaatan kapasitas pembangkit	Energy_Type: Energy_type_name	Jenis pembangkit listrik energi terbarukan tempat peralatan dipasang contoh: PLTS (surya) , PLTBM (biomassa), dll
Customer_Service	1. Interrupted_duration: total durasi gangguan listrik yang dialami pelanggan (jam atau menit).	Time: date	waktu kejadian (gangguan: pemadaman) atau periode pelaporan.
		Site: Site_name, Location	wilayah layanan pelanggan
	2. SAIDI_index: rerata total gangguan per pelanggan (jam/pelanggan).	Customer_Segment : Customer_name, Customer_type	Jenis pelanggan contohnya: Rumah tangga, Industri, usaha, umum/public)
	3. SAIFI_index: rerata frekuensi gangguan per		

	<p>pelanggan.</p> <p>4. Energy_Consumption: total energi yang dikonsumsi oleh segmen pelanggan tertentu dalam periode (kWh).</p>		
Operational_Efficiency	<p>1. Actual_cost: biaya aktual yang dikeluarkan untuk proyek</p> <p>2. Budget: anggaran yang direncanakan untuk proyek</p> <p>3. Revenue: Pendapatan yang dihasilkan dari proyek dalam periode tertentu.</p> <p>4. ROI_percentage: Mengukur profitabilitas proyek.</p>	Time: date	Menunjukkan waktu/ periode pelaporan
		Site: Site_name, Location	Lokasi proyek
		Project: Project_name, status	Proyek yang dikerjakan
Regulatory_Compliance	<p>1. Inspection_count: Jumlah inspeksi yang dilakukan</p> <p>2. Violation_count: Jumlah pelanggaran terhadap regulasi yang ditemukan.</p> <p>3. Certificated_personnel_count: jumlah karyawan bersertifikasi</p> <p>4. Compliance_score : skor atau indeks kepatuhan terhadap regulasi</p>	Time: date	Tanggal atau periode inspeksi
		Site: Site_name, Location	Lokasi pembangkit atau fasilitas yang diawasi.
		Regulation: Regulation_name, Authority	Regulasi yang diterapkan

Membangun sistem data warehouse merupakan langkah strategis bagi perusahaan penyedia listrik berbasis energi terbarukan dalam mengintegrasikan data operasional yang tersebar di berbagai unit, seperti produksi energi, performa peralatan, layanan pelanggan, efisiensi proyek, dan kepatuhan regulasi. Dengan struktur yang terdiri dari tabel fakta dan dimensi, perusahaan dapat menyatukan informasi penting untuk dianalisis secara historis maupun

real-time. Implementasi skrip SQL dalam proses pembuatan data warehouse untuk merepresentasikan struktur data tersebut. Pendefinisian relasi antar tabel, serta penyusunan query analitik yang disesuaikan dengan kebutuhan analisis yang telah diidentifikasi, seperti monitoring produksi energi, evaluasi efisiensi operasional, pengukuran dampak lingkungan, dan audit kepatuhan terhadap regulasi.

Query DDL(Skema Bintang atau Zone Gold)

```
-- Dimensi Waktu
CREATE TABLE dim_time (
    time_id INT PRIMARY KEY,
    date DATE
);

-- Dimensi Lokasi
CREATE TABLE dim_site (
    site_id INT PRIMARY KEY,
    site_name VARCHAR(100),
    location VARCHAR(255)
);

-- Dimensi Tipe Energi
CREATE TABLE dim_energy_type (
    energy_type_id INT PRIMARY KEY,
    energy_type_name VARCHAR(100)
);

-- Dimensi Cuaca
CREATE TABLE dim_weather (
    weather_id INT PRIMARY KEY,
    weather_condition VARCHAR(100),
    temperature DECIMAL(5,2)
);

-- Dimensi Peralatan
CREATE TABLE dim_equipment (
    equipment_id INT PRIMARY KEY,
    equipment_name VARCHAR(100),
    equipment_type VARCHAR(100),
    rated_capacity DECIMAL(10,2),
    operational_status VARCHAR(50)
);

-- Dimensi Customer
CREATE TABLE dim_customer (
    customer_id INT PRIMARY KEY,
    customer_segment VARCHAR(50),
    customer_name VARCHAR(100),
    customer_type VARCHAR(50)
);

-- Dimensi Proyek
CREATE TABLE dim_project (
    project_id INT PRIMARY KEY,
    project_name VARCHAR(100),
    status VARCHAR(50)
```

```

);

-- Dimensi Regulasi
CREATE TABLE dim_regulation (
    regulation_id INT PRIMARY KEY,
    regulation_name VARCHAR(100),
    authority VARCHAR(100)
);

-- Fakta Produksi Energi
CREATE TABLE fact_energy_production (
    fact_id INT PRIMARY KEY,
    time_id INT,
    site_id INT,
    energy_type_id INT,
    FOREIGN KEY (time_id) REFERENCES dim_time(time_id),
    FOREIGN KEY (site_id) REFERENCES dim_site(site_id),
    FOREIGN KEY (energy_type_id) REFERENCES dim_energy_type(energy_type_id)
);

-- Fakta Kinerja Peralatan
CREATE TABLE fact_equipment_performance (
    fact_id INT PRIMARY KEY,
    time_id INT,
    site_id INT,
    equipment_id INT,
    energy_type_id INT,
    FOREIGN KEY (time_id) REFERENCES dim_time(time_id),
    FOREIGN KEY (site_id) REFERENCES dim_site(site_id),
    FOREIGN KEY (equipment_id) REFERENCES dim_equipment(equipment_id),
    FOREIGN KEY (energy_type_id) REFERENCES dim_energy_type(energy_type_id)
);

-- Fakta Layanan Pelanggan
CREATE TABLE fact_customer_service (
    fact_id INT PRIMARY KEY,
    time_id INT,
    site_id INT,
    customer_id INT,
    FOREIGN KEY (time_id) REFERENCES dim_time(time_id),
    FOREIGN KEY (site_id) REFERENCES dim_site(site_id),
    FOREIGN KEY (customer_id) REFERENCES dim_customer(customer_id)
);

-- Fakta Efisiensi Operasional
CREATE TABLE fact_operational_efficiency (
    fact_id INT PRIMARY KEY,
    time_id INT,
    site_id INT,
    project_id INT,
    FOREIGN KEY (time_id) REFERENCES dim_time(time_id),
    FOREIGN KEY (site_id) REFERENCES dim_site(site_id),
    FOREIGN KEY (project_id) REFERENCES dim_project(project_id)
);

-- Fakta Kepatuhan Regulasi
CREATE TABLE fact_regulatory_compliance (
    fact_id INT PRIMARY KEY,
    time_id INT,

```

```

site_id INT,
regulation_id INT,
FOREIGN KEY (time_id) REFERENCES dim_time(time_id),
FOREIGN KEY (site_id) REFERENCES dim_site(site_id),
FOREIGN KEY (regulation_id) REFERENCES dim_regulation(regulation_id)
);

```

```

UPDATE fact_energy_production f
SET total_production = sub.total_production
FROM (
    SELECT
        f.fact_id,
        SUM(log.production_kwh) AS total_production
    FROM
        fact_energy_production f
    JOIN dim_equipment e ON e.energy_type_id = f.energy_type_id
    JOIN production_log log ON log.equipment_id = e.equipment_id AND log.time_id =
f.time_id
    GROUP BY f.fact_id
) sub
WHERE f.fact_id = sub.fact_id;

UPDATE fact_equipment_performance f
SET
    down_time_hours = sub.down_time_hours,
    cost_maintenance = sub.cost_maintenance,
    capacity_factor_percentage = sub.capacity_factor_percentage,
    capacity_utilization_factor_percentage = sub.capacity_utilization_factor_percentage
FROM (
    SELECT
        f.fact_id,
        SUM(log.down_time_hours) AS down_time_hours,
        SUM(log.maintenance_cost) AS cost_maintenance,
        AVG(log.actual_production / log.max_capacity * 100) AS capacity_factor_percentage,
        AVG(log.used_capacity / log.total_capacity * 100) AS
capacity_utilization_factor_percentage
    FROM
        fact_equipment_performance f
    JOIN log_fact_equipment_performance log ON log.time_id = f.time_id
    GROUP BY f.fact_id
) sub
WHERE f.fact_id = sub.fact_id;

UPDATE fact_equipment_performance f
SET
    down_time_hours = sub.down_time_hours,
    cost_maintenance = sub.cost_maintenance,
    capacity_factor_percentage = sub.capacity_factor_percentage,
    capacity_utilization_factor_percentage = sub.capacity_utilization_factor_percentage
FROM (
    SELECT
        log.time_id,
        e.equipment_id,
        SUM(log.downtime_hours) AS down_time_hours,
        SUM(log.maintenance_cost) AS cost_maintenance,
        AVG(log.actual_output_kwh / NULLIF(e.rated_capacity * log.operating_hours, 0)) *

```



```

100 AS capacity_factor_percentage,
    AVG(log.utilization / NULLIF(e.rated_capacity, 0)) * 100 AS
capacity_utilization_factor_percentage
FROM equipment_log log
JOIN dim_equipment e ON log.equipment_id = e.equipment_id
GROUP BY log.time_id, e.equipment_id
) sub
WHERE f.time_id = sub.time_id
AND f.equipment_id = sub.equipment_id;

UPDATE fact_customer_service f
SET
    interrupted_duration = sub.interrupted_duration,
    saidi_index = sub.saidi_index,
    saifi_index = sub.saifi_index,
    energy_consumption = sub.energy_consumption
FROM (
    SELECT
        log.time_id,
        c.customer_id,
        SUM(log.interruption_duration_minutes)/60.0 AS interrupted_duration,
        AVG(log.saidi_index) AS saidi_index,
        AVG(log.saifi_index) AS saifi_index,
        SUM(log.energy_consumed_kwh) AS energy_consumption
    FROM customer_log log
    JOIN dim_customer c ON log.customer_id = c.customer_id
    GROUP BY log.time_id, c.customer_id
) sub
WHERE f.time_id = sub.time_id
AND f.customer_id = sub.customer_id;

UPDATE fact_operational_efficiency f
SET
    actual_cost = sub.actual_cost,
    budget = sub.budget,
    revenue = sub.revenue,
    roi_percentage = sub.roi_percentage
FROM (
    SELECT
        log.time_id,
        p.project_id,
        SUM(log.actual_cost) AS actual_cost,
        SUM(log.budget) AS budget,
        SUM(log.revenue) AS revenue,
        (SUM(log.revenue) - SUM(log.actual_cost)) / NULLIF(SUM(log.actual_cost), 0) * 100
    AS roi_percentage
    FROM project_log log
    JOIN dim_project p ON log.project_id = p.project_id
    GROUP BY log.time_id, p.project_id
) sub
WHERE f.time_id = sub.time_id
AND f.project_id = sub.project_id;

UPDATE fact_regulatory_compliance f
SET
    inspection_count = sub.inspection_count,
    violation_count = sub.violation_count,
    certificated_personnel_count = sub.certificated_personnel_count,
    compliance_score = sub.compliance_score

```

```

FROM (
  SELECT
    log.time_id,
    r.regulation_id,
    COUNT(log.inspection_id) AS inspection_count,
    SUM(log.violation_count) AS violation_count,
    COUNT(log.certified_employee_id) AS certificated_personnel_count,
    AVG(log.compliance_score) AS compliance_score
  FROM regulation_log log
  JOIN dim_regulation r ON log.regulation_id = r.regulation_id
  GROUP BY log.time_id, r.regulation_id
) sub
WHERE f.time_id = sub.time_id
AND f.regulation_id = sub.regulation_id;

```

Query Analitik

Berikut query analitik yang disesuaikan dengan kebutuhan analisis yang telah diidentifikasi sebelumnya:

1. Monitoring produksi energi

```

SELECT
  et.energy_type_name,
  DATE_TRUNC('month', t.date) AS month,
  SUM(f.total_production) AS total_energy_produced
FROM fact_energy_production f
JOIN dim_time t ON f.time_id = t.time_id
JOIN dim_energy_type et ON f.energy_type_id = et.energy_type_id
GROUP BY et.energy_type_name, DATE_TRUNC('month', t.date)
ORDER BY et.energy_type_name, month;

```

2. Evaluasi kinerja pembangkit listrik

```

SELECT
  s.site_name,
  AVG(f.capacity_factor_percentage) AS avg_capacity_factor,
  AVG(f.capacity_utilization_factor_percentage) AS avg_utilization,
  SUM(f.down_time_hours) AS total_downtime_hours
FROM fact_equipment_performance f
JOIN dim_site s ON f.site_id = s.site_id
GROUP BY s.site_name
ORDER BY avg_utilization DESC;

```

3. Analisis layanan dan kepuasan pelanggan

```
SELECT
    cs.customer_segment,
    AVG(f.saidi_index) AS avg_saidi,
    AVG(f.saifi_index) AS avg_saifi,
    AVG(f.interrupted_duration) AS avg_interruption_duration
FROM fact_customer_service f
JOIN dim_customer cs ON f.customer_id = cs.customer_id
GROUP BY cs.customer_segment
ORDER BY avg_interruption_duration DESC;
```

4. Monitoring biaya operasional dan efisiensi proyek

```
SELECT
    p.project_name,
    SUM(f.actual_cost) AS total_actual_cost,
    SUM(f.budget) AS total_budget,
    SUM(f.revenue) AS total_revenue,
    AVG(f.roi_percentage) AS avg_roi_percentage
FROM fact_operational_efficiency f
JOIN dim_project p ON f.project_id = p.project_id
GROUP BY p.project_name
ORDER BY avg_roi_percentage DESC;
```

5. Evaluasi kepatuhan regulasi teknis dan keselamatan

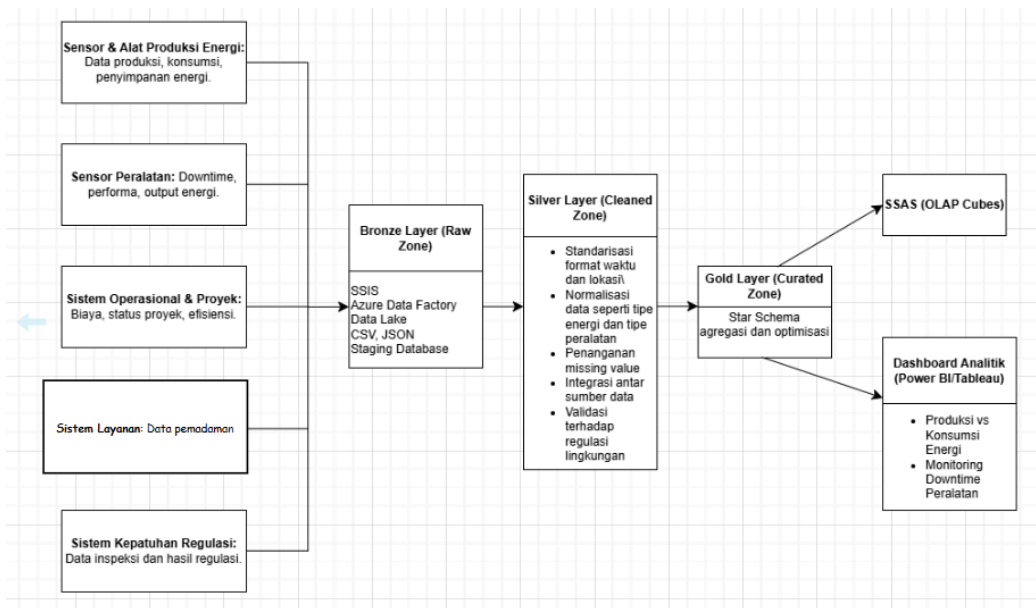
```
SELECT
    r.regulation_name,
    COUNT(f.inspection_count) AS total_inspections,
    SUM(f.violation_count) AS total_violations,
    AVG(f.compliance_score) AS avg_compliance_score
FROM fact_regulatory_compliance f
JOIN dim_regulation r ON f.regulation_id = r.regulation_id
GROUP BY r.regulation_name
ORDER BY avg_compliance_score ASC;
```

4.4. Sumber Data

Data pada tugas kali ini diperoleh dari "Statistik Ketenagalistrikan 2022" menjelaskan sumber data yang digunakan untuk penyusunan laporan statistik ketenagalistrikan. Data utama berasal dari PT PLN (Persero) dan Direktorat Jenderal Ketenagalistrikan sebagai acuan pembuatan data dummy ini mencakup informasi mengenai penyediaan tenaga listrik, perusahaan tenaga listrik, dan pemanfaatan energi. Selain itu, data pendukung juga diperoleh dari Direktorat Jenderal Minyak dan Gas Bumi, Direktorat Jenderal Mineral dan Batubara, serta

Direktorat Jenderal Energi Baru Terbarukan dan Konservasi Energi. Hal ini menunjukkan adanya kolaborasi antar berbagai unit di lingkungan Kementerian Energi dan Sumber Daya Mineral untuk menghasilkan informasi yang komprehensif dan akurat.

4.5. Aliran Data



Gambar 7. Alur Aliran Data

Pada diagram alur aliran data menunjukkan bagaimana data yang dikumpulkan dari berbagai sumber diproses dan dibuat menjadi informasi yang dapat digunakan oleh stakeholder perusahaan energi terbarukan. Proses dimulai dengan pengumpulan data mentah dari beberapa sistem utama. Ini termasuk sensor dan alat produksi energi (data tentang produksi, konsumsi, dan penyimpanan energi), sensor peralatan (data tentang downtime, kinerja, dan output energi), sistem operasional dan proyek (data tentang biaya dan efisiensi proyek), dan sistem kepatuhan regulasi (data tentang hasil inspeksi dan kepatuhan regulasi).

Pertama, data disimpan dalam Lapisan Bronze (Raw Zone) menggunakan platform seperti SSIS, Data Factory, Azure, atau Data Lake dalam format CSV atau JSON, atau melalui staging database. Kemudian, Lapisan Silver (Cleaned Zone) menangani standarisasi format waktu dan lokasi, normalisasi tipe data, integrasi antar sumber data, dan validasi peraturan lingkungan yang berlaku.

Data yang telah disiapkan dan dibersihkan kemudian dikirim ke Lapisan Emas (Zona Kuadrat). Pada tahap ini, data disusun dalam skema bintang dan dilakukan agregasi dan optimisasi untuk membuatnya mudah diakses untuk

analisis lebih lanjut. Output Gold Layer didistribusikan ke dua jalur: pertama, ke OLAP Cubes (SSAS) untuk analisis multidimensi oleh pihak berwenang seperti CEO dan divisi keuangan; dan kedua, ke dashboard analitik, yang dapat divisualisasikan dan digunakan untuk membuat keputusan tentang produksi versus konsumsi energi, serta untuk memantau kinerja peralatan oleh tim data analyst, divisi produksi, dan pihak terkait lainnya.

4.6. Arsitektur Sistem

a. 1. Arsitektur Logikal

Arsitektur logikal Data Warehouse ElSusEnv dibangun berdasarkan skema konseptual yang telah dikembangkan dan diimplementasikan pada Misi Pertama dan Kedua. Struktur ini terdiri dari lima *fact table* utama: produksi energi, performa alat, efisiensi operasional, dampak lingkungan, dan kepatuhan regulasi. Kelima fakta tersebut dirancang dengan skema *star schema* yang terpisah secara fungsional namun terintegrasi secara semantik melalui *conformed dimensions*, yaitu waktu, lokasi, peralatan, dan kebijakan regulatif.

Setiap *fact table* beroperasi dengan granularitas harian dan terhubung langsung ke dimensi kontekstual yang menyimpan atribut relevan untuk pelaporan dan analitik keberlanjutan. Dimensi waktu mencakup atribut standar serta pelabelan kalender keberlanjutan berdasarkan musim dan tahun fiskal. Dimensi lokasi memuat informasi struktur fasilitas dan wilayah administratif yang telah diselaraskan pada tahap transformasi data. Dimensi alat dan regulasi terhubung ke beberapa fakta dengan tujuan memungkinkan kueri lintas proses seperti evaluasi performa alat terhadap hasil inspeksi atau nilai kepatuhan terhadap emisi.

Pendekatan *Slowly Changing Dimension Type 2* diterapkan pada atribut-atribut yang mengalami perubahan periodik dan berdampak pada analitik longitudinal. Implementasi SCD2 digunakan secara eksplisit pada status alat, klasifikasi proyek efisiensi, serta cakupan kebijakan regulasi lingkungan. Seluruh dimensi dibangun berdasarkan data yang tersedia pada sumber eksternal seperti Kaggle dan EIA, serta data internal yang telah melalui proses validasi di Misi Kedua.

Hubungan antar entitas dibentuk tidak hanya untuk memfasilitasi pelaporan, tetapi juga untuk memungkinkan evaluasi kausalitas dalam konteks lingkungan. Produksi energi yang tercatat dalam *fact_energy_production* dapat

dianalisis bersama dengan tingkat downtime pada *fact_equipment_performance* dan nilai kepatuhan dari *fact_regulatory_compliance* menggunakan dimensi waktu dan lokasi yang konsisten. Dengan arsitektur ini, sistem mampu menjawab kebutuhan analitik multidimensi, lintas domain, dan historis yang diperlukan oleh EISusEnv dalam mendukung operasional berbasis data dan prinsip keberlanjutan.

b. Arsitektur Fisik

Desain fisik Data Warehouse EISusEnv difokuskan pada keandalan pemrosesan data skala besar, efisiensi penyimpanan, dan dukungan terhadap kebutuhan analitik harian maupun historis. Seluruh tabel disimpan dalam format *Apache Parquet* untuk memanfaatkan struktur berbasis kolom dan kemampuan kompresi tinggi. Penyimpanan dilakukan dalam file yang terpartisi berdasarkan *time_id* dan *site_id*, sesuai dengan struktur kueri utama yang telah diidentifikasi sejak tahap eksplorasi pada Misi Kedua.

Data dimuat secara harian melalui pipeline *extract-transform-load* yang dibangun menggunakan Python dan dijalankan terjadwal menggunakan Apache Airflow. Proses ini mencakup konversi satuan energi ke dalam MWh, suhu dalam Celsius, serta penyesuaian format waktu ke dalam struktur dimensi waktu terstandar. Pembersihan data dan validasi referensial juga dilakukan terhadap seluruh dimensi menggunakan kunci bisnis dan *lookup mapping* yang telah dibangun sebelumnya.

Indeks dibentuk secara terfokus pada kolom yang memiliki tingkat selektivitas tinggi dalam kueri pelaporan dan agregasi. Kolom seperti *equipment_id*, *emission_type_id*, dan *regulation_category_id* diprioritaskan karena terbukti digunakan secara berulang pada skenario analitik utama yang dievaluasi selama Misi Kedua. Strategi pembentukan indeks ini dilakukan adaptif berdasarkan analisis *query plan* dari hasil uji performa.

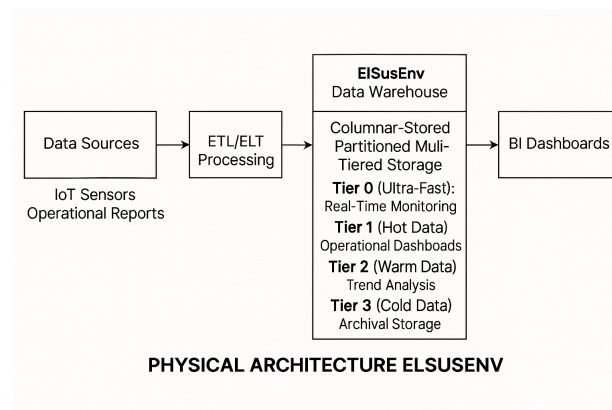
Struktur penyimpanan fisik diorganisasikan dalam tiga *tier* untuk mengelola data berdasarkan usia dan kebutuhan akses.

Tier 1 mencakup data aktif tiga bulan terakhir dan digunakan untuk dashboard harian seperti performa produksi energi, emisi harian, serta pemantauan alat. Data ini disimpan pada media latensi rendah untuk mendukung visualisasi interaktif melalui Power BI.

Tier 2 menyimpan data historis satu hingga dua tahun terakhir yang digunakan untuk evaluasi tren produksi energi tahunan, efektivitas proyek efisiensi, dan pelaporan triwulanan. Penyimpanan menggunakan media campuran dengan optimalisasi throughput dan penghematan ruang.

Tier 3 digunakan untuk menyimpan data historis lebih dari dua tahun yang mencakup hasil inspeksi terdahulu, pelaporan tahunan SDG, dan arsip proyek. Seluruh data pada tier ini disimpan dalam sistem *object storage* berbasis cloud dan tetap terhubung melalui metadata katalog yang memungkinkan akses historis saat dibutuhkan.

Dengan struktur partisi, format kolom, strategi indeks, dan tier penyimpanan ini, arsitektur fisik Data Warehouse EISusEnv mampu menangani kebutuhan integrasi lintas domain, analitik keberlanjutan, serta pelaporan kepatuhan secara terstruktur, efisien, dan dapat diskalakan.



Gambar 8. Arsitektur Fisik Gudang Data EISusEnv Berbasis Penyimpanan Berlapis (Tiered Storage)

5. Implementasi dan Hasil

5.1. Proses ETL

Proses ETL menggunakan alat SSIS yang disediakan oleh SQL Server Management Studio. Sebelum masuk ke dalam proses ETL, terlebih dahulu membentuk database di Database Engine sebagai alat penyimpanan data serta menjalankan *query* SQL seperti membentuk database lalu membentuk tabel fakta dan tabel dimensi. Proses ETL dalam aliran data terletak dalam lapisan *Bronze raw* dengan menggunakan SSIS yang ada di SQL Server Management System (SSMS).

5.1.1 Extract (Tahap Bronze Zone)

Tahap pertama yakni proses Extract data, dimana sumber data berformat csv akan dimasukkan ke dalam SQL server, menyesuaikan dengan tabel yang sudah dibentuk sebelumnya. Menggunakan fungsi dari *import flat file* dapat memasukkan data dari sumber ke dalam tabel di SSMS. Penamaan tabel tidak jauh berbeda dari nama sumber data nya.

5.1.2 Transform (Tahap Silver Zone)

Pada proses *Transform* ini akan dilakukan beberapa agregasi seperti Standarisasi waktu dan lokasi, Normalisasi tipe energi dan peralatan, Penanganan missing value.

```
SELECT
    equipment_id,
    site_id,
    CONVERT(DATE, Time_Date, 120) AS standardized_date,
    UPPER(LTRIM(RTRIM(site_name))) AS site_name,
    UPPER(LTRIM(RTRIM(Site_Location))) AS region
INTO stg_equipment_performance_silver
FROM bronze.equipment_performance;
```

Pada kode ini kita akan menyesuaikan atau melakukan standarisasi terhadap kolom waktu, site_name, dan site_location dengan menggunakan data yang sudah tersimpan di *bronze.equipment_performance* yang mana akan tersimpan dalam *stg_equipment_performance_silver*. Hasilnya akan seperti berikut:

	equipment_id	site_id	standardized_date	site_name	region
1	46	8	2024-07-03	LAHENDONG	DKI JAKARTA
2	89	23	2023-07-28	SORIK MARAPI	JAWA TIMUR
3	2	9	2023-10-05	ULUMBU	SULAWESI SELATAN
4	14	17	2025-05-02	SIDRAP	NTT
5	64	9	2024-07-13	ULUMBU	SULAWESI SELATAN
6	29	10	2024-05-23	MATALOKO	LAMPUNG
7	56	3	2024-10-08	SAGULING	JAWA TENGAH
8	99	3	2025-04-22	SAGULING	JAWA TENGAH
9	51	5	2023-07-19	DARAJAT	SULAWESI UTARA
10	9	17	2024-11-24	SIDRAP	NTT
11	71	14	2023-07-05	GRATI	JAWA TENGAH
12	9	1	2025-03-15	KARANGKATES	JAWA TIMUR
13	25	12	2024-10-25	SURALAYA	JAWA TIMUR
14	86	2	2025-04-02	CIRATA	JAWA BARAT
15	10	25	2023-09-13	MUARA LABOH	JAWA TENGAH
16	22	10	2025-02-20	MATALOKO	LAMPUNG
17	17	3	2024-08-28	SAGULING	JAWA TENGAH
18	43	2	2024-05-27	CIRATA	JAWA BARAT
19	34	13	2024-06-28	TANJUNG JAT...	JAWA BARAT
20	50	13	2023-08-10	TANJUNG JAT...	JAWA BARAT
21	10	5	2024-07-11	DARAJAT	SULAWESI UTARA
22	14	13	2023-11-12	TANJUNG JAT...	JAWA BARAT
23	6	1	2024-02-06	KARANGKATES	JAWA TIMUR
24	32	23	2024-08-29	SORIK MARAPI	JAWA TIMUR
25	73	5	2025-03-08	DARAJAT	SULAWESI UTARA
26	54	11	2024-05-11	PAITON	SUMATERA SELAT...
27	32	18	2024-03-25	JENEPONTO	BANTEN
28	60	9	2024-01-13	ULUMBU	SULAWESI SELATAN
29	72	21	2024-09-07	PATUHA	LAMPUNG
30	49	13	2024-06-15	TANJUNG JAT...	JAWA BARAT
31	29	18	2024-10-24	JENEPONTO	BANTEN
32	73	6	2023-05-29	WAYANG PLT...	NTT
33	88	10	2025-03-12	MATALOKO	LAMPUNG
34	96	24	2024-05-08	RANTAU DED	JAWA BARAT

Gambar 9. Standarisasi Waktu dan Lokasi

Terlihat dari kolom site_name dan region diterapkan UpperCase sehingga pada kolom tersebut semua nilai menggunakan huruf kapital. Dan juga pada kolom standarized_date sudah disesuaikan dengan format waktu ke YYYY-MM-DD. Pada tahap selanjutnya yakni kita akan

melakukan normalisasi pada tabel *equipment_performance* yang ada pada tahap *Bronze Zone*.

```
-- Tipe energi unik
SELECT DISTINCT
    energy_type_id,
    Energy_Type_Name
INTO dim_energy_type_silver
FROM bronze.equipment_performance
WHERE Energy_Type_Name IS NOT NULL;

-- Tipe peralatan unik
SELECT DISTINCT
    equipment_id,
    equipment_name,
    equipment_type
INTO dim_equipment_type_silver
FROM bronze.equipment_performance
WHERE equipment_name IS NOT NULL AND equipment_type IS
NOT NULL;
```

Hasil dari kode tersebut akan nampak seperti ini.

	energy_type_id	Energy_Type_Name
1	2	PLTB
2	1	PLTS
3	6	PLTG
4	4	PLTU
5	3	PLTA
6	5	PLTG

Gambar 10. Hasil Normalisasi data

Melakukan pembersihan terhadap data yang memiliki nilai kosong dan akan dibentuk ke dalam Silver Zone.

```
--Penanganan Missing Value
SELECT
    *,
    ISNULL(electricity_Production_kWh, 0) AS
fuel_consumption_cleaned
INTO stg_energy_production_silver
FROM bronze.energy_production;
```

5.1.3 Load (Tahap Gold Zone)

Dalam zona Gold dari arsitektur Medallion untuk kasus ElSusEnv, proses diarahkan pada pembentukan skema bintang (star schema) yang terdiri dari tabel fakta teragregasi dan tabel-tabel dimensi yang telah dibersihkan dan distandarisasi. Zona ini dirancang untuk mendukung analisis bisnis dan visualisasi data secara efisien melalui integrasi data historis yang telah disaring dan ditata rapi pada zona sebelumnya.

Pada tahap ini, kita membuat tabel fakta utama, yaitu `fact_energy_equipment_summary`, yang berisi ringkasan performa dan produksi energi berdasarkan penggabungan data dari dua sumber utama: `equipment performance` dan `energy production`. Tabel ini menyimpan metrik penting seperti total output energi, konsumsi bahan bakar, biaya pemeliharaan, serta waktu tidak beroperasi per kombinasi peralatan, lokasi, waktu, dan kondisi bahan bakar serta cuaca. Nilai-nilai ini dihitung dengan cara agregasi melalui fungsi `SUM` dari zona Silver yang sebelumnya telah dibersihkan dan disatukan.

Agar dapat dikaitkan ke berbagai dimensi analitis, dibentuk pula beberapa tabel dimensi, seperti `dim_equipment` untuk data peralatan, `dim_site` untuk lokasi dan wilayah, `dim_time` untuk kalender, `dim_fuel` untuk informasi bahan bakar dan kandungan energi, serta `dim_weather` untuk cuaca yang memengaruhi produksi energi. Setiap dimensi ini memiliki atribut-atribut penting yang memungkinkan pengguna akhir melakukan analisis berdasarkan kategori seperti jenis peralatan, wilayah geografis, jenis bahan bakar, atau rentang waktu tertentu.

Penggunaan foreign key pada tabel fakta menghubungkan data teragregasi ini dengan masing-masing tabel dimensi, membentuk struktur star schema yang ideal untuk pengolahan OLAP (Online Analytical Processing). Dengan struktur ini, sistem Business Intelligence atau visualisasi data seperti Power BI atau Tableau dapat dengan mudah mengakses, mengelompokkan, serta menelusuri data berdasarkan sudut pandang yang fleksibel. Secara keseluruhan, Gold Zone bertujuan untuk menyajikan data yang telah matang dan siap digunakan oleh pemangku kepentingan untuk pengambilan keputusan berbasis data, sekaligus menjaga integritas serta efisiensi dalam pemrosesan dan penyajian informasi.

5.2. Data

Dalam perancangan Data Warehouse untuk industri Lingkungan & Sustainability, data krusial yang akan dimanfaatkan bersumber dari lima domain utama yang saling terkait. Pertama, data energi produksi akan mencakup informasi detail mengenai kuantitas energi yang dihasilkan (misalnya, MWh), sumber energi (surya, angin, biomassa), lokasi fasilitas, serta parameter operasional terkait. Kedua, data efisiensi operasional akan menyediakan wawasan tentang konsumsi sumber daya seperti air dan listrik, metrik

penggunaan bahan bakar, serta data limbah dan emisi dari berbagai proses atau fasilitas, yang penting untuk mengidentifikasi area peningkatan. Selanjutnya, data kinerja peralatan akan memonitor kondisi dan performa aset kunci, termasuk data *uptime* dan *downtime*, frekuensi kegagalan, serta catatan pemeliharaan, guna memastikan operasional yang optimal dan prediktif. Keempat, data layanan pelanggan akan merekam interaksi dengan pelanggan terkait layanan lingkungan, termasuk jenis keluhan, waktu penyelesaian, dan tingkat kepuasan, yang esensial untuk meningkatkan kualitas layanan dan mengidentifikasi kebutuhan baru. Terakhir, data regulasi akan mencakup semua informasi kepatuhan terhadap standar dan peraturan lingkungan, termasuk batas ambang yang diizinkan, laporan audit, dan potensi pelanggaran, untuk memastikan operasi selalu sesuai dengan hukum yang berlaku. Kelima dataset ini, ketika diintegrasikan dalam Data Warehouse, akan menjadi fondasi bagi analisis komprehensif yang memungkinkan pengambilan keputusan strategis untuk keberlanjutan dan efisiensi di seluruh organisasi.

6. Tim Proyek

Nama	Peran	Kontribusi
Muhammad Farhan	DevOps Engineer (Lead Project)	Infrastructure setup & Deployment
Mayada	Analytics Engineer	Dashboard & visualization
Khoirul Anam	QA Engineer	Deployment Monitoring & Documentation
Alvia Asrinda Br Ginting	ETL Developer	ETL Development
Syalaisha Andina Putriansyah	Data Engineer	Architecture & System design

7. Lampiran