

# **LAPORAN MISI KEEMPAT PERGUDANGAN DATA ENERGI & UTILITAS**



Disusun oleh :

- |                                    |           |
|------------------------------------|-----------|
| 1. Muhammad Zaki Abdillah          | 121450008 |
| 2. Deva Anjani Khayyuninafsyah     | 122450014 |
| 3. Patricia Leondrea Diajeng Putri | 122450050 |
| 4. Syadza Puspadari Azhar          | 122450072 |
| 5. Dea Mutia Risani                | 122450099 |
| 6. Amalia Melani Putri             | 122450122 |

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

## Daftar Isi

<b>Daftar Isi.....</b>	<b>2</b>
<b>1. Ringkasan dan Latar Belakang Proyek.....</b>	<b>3</b>
1.1. Ringkasan Proyek.....	3
1.2. Latar Belakang.....	3
<b>2. Tujuan dan Ruang Lingkup Sistem.....</b>	<b>3</b>
2.1. Tujuan.....	3
2.2. Ruang Lingkup.....	3
<b>3. Metodologi.....</b>	<b>4</b>
<b>4. Analisis Kebutuhan.....</b>	<b>5</b>
4.1. Daftar Stakeholder dan Tujuan Bisnis.....	5
4.1.1. Daftar Stakeholder.....	5
4.1.2. Tujuan Bisnis.....	5
4.1.3. Interview Simulasi.....	6
4.1.4. Studi Kasus.....	6
4.2. Fakta dan Dimensi.....	7
4.2.1. Tabel Fakta.....	7
4.2.2. Tabel Dimensi.....	7
4.3. Sumber dan Metadata.....	8
<b>5. Desain Konseptual, Logikal, dan Fisikal.....</b>	<b>9</b>
<b>6. Proses Implementasi.....</b>	<b>13</b>
<b>7. Hasil Implementasi.....</b>	<b>56</b>
7.1. Skema database data warehouse di SQL Server.....	56
7.2. Relasi antar tabel dengan star schema.....	57
7.3. Implementasi indexing.....	57
7.4.1. Total Konsumsi Daya Harian.....	58
7.4.2. Pola Konsumsi Rata-rata per Jam (dari Silver Layer).....	58
7.4.3. Kontribusi Sub-Metering terhadap Total Konsumsi (Rata-rata Harian).....	59
7.4.4. Konsumsi Daya Harian Tertinggi dan Terendah.....	59
<b>8. Evaluasi.....</b>	<b>59</b>
<b>9. Rencana Pengembangan ke Depan.....</b>	<b>60</b>
<b>10. Tim Proyek.....</b>	<b>60</b>

## **1. Ringkasan dan Latar Belakang Proyek**

### **1.1. Ringkasan Proyek**

Proyek pengembangan sistem Data Warehouse bertujuan untuk merancang dan membangun solusi terintegrasi bagi perusahaan energi RubicoNergi (RBN). Sistem ini dirancang untuk mengintegrasikan berbagai jenis data operasional. Melalui pendekatan ETL (*Extract, Transform, Load*), proyek ini mengolah data mentah terkait penggunaan energi listrik menjadi sumber informasi terstruktur dan siap dianalisis. Proses dilakukan untuk mendukung pengambilan keputusan berbasis data, meningkatkan efisiensi operasional, dan memperkuat sistem pelaporan perusahaan.

Proyek ini terdiri dari analisis kebutuhan, perancangan skema multidimensi, hingga implementasi teknis termasuk ke dalamnya pemrosesan data, pembuatan indeks, partisi, dan pemeliharaan sistem Data Warehouse.

### **1.2. Latar Belakang**

RubicoNergi (RBN) merupakan perusahaan penyedia energi listrik yang melayani berbagai segmen pelanggan, termasuk wilayah 3S.

Dalam operasionalnya, RBN menghadapi tantangan dalam beberapa hal, yaitu:

- Mengelola data dari sistem yang terpisah (CRM, GIS, *billing*, dan gangguan).
- Menangani kebutuhan pelaporan dan analisis yang makin kompleks.
- Menentukan strategi ekspansi jaringan yang efisien dan data-driven.
- Mengantisipasi gangguan dan memperbaiki efisiensi distribusi daya.

Kurangnya integrasi data menyebabkan pelaporan menjadi lambat, analisis tidak akurat, dan pengambilan keputusan bersifat reaktif. Oleh karena itu, dibutuhkan gudang data terpusat yang mampu menghimpun dan mengelola seluruh data penting dari berbagai sistem.

Dengan membangun sistem Data Warehouse yang menerapkan ETL, RBN dapat melakukan analisis spasial dan temporal terhadap konsumsi energi, memprediksi kebutuhan daya saat *peak load*, memantau wilayah rawan gangguan dan menilai efektivitas respon teknis, meningkatkan akurasi laporan keuangan, dan mendukung ekspansi jaringan terutama wilayah 3S.

## **2. Tujuan dan Ruang Lingkup Sistem**

### **2.1. Tujuan**

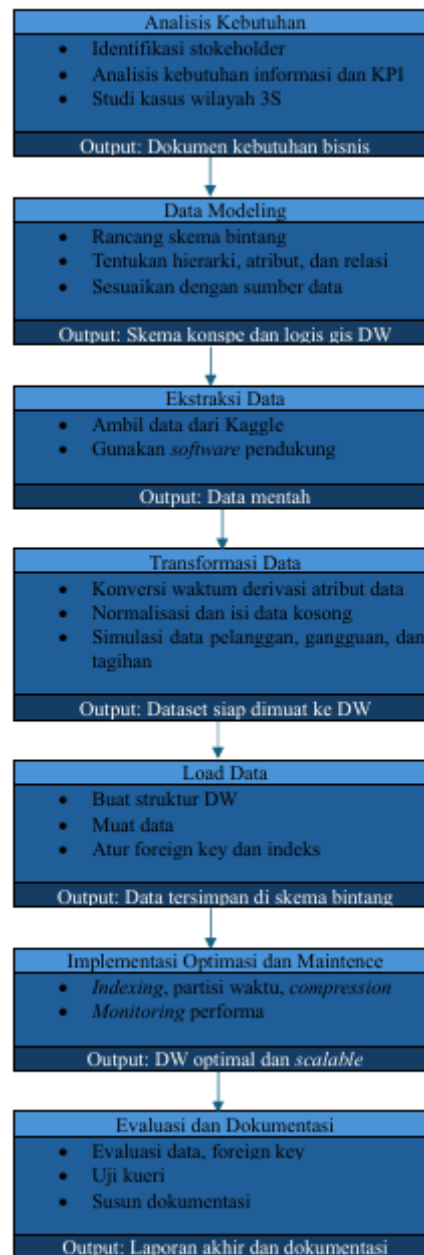
1. Mengintegrasikan data.
2. Merancang model data multidimensional.
3. Mengimplementasikan proses ETL.
4. Mendukung kebutuhan pelaporan dan pengambilan keputusan.
5. Memfasilitasi analisis prediktif dan spasial.

### **2.2. Ruang Lingkup**

Ruang lingkup proyek ini mencakup aspek konseptual, teknis, dan implementatif dalam pembangunan sistem gudang data RBN, meliputi:

1. Identifikasi kebutuhan bisnis dan stakeholder.
2. Pengumpulan dan analisis data.
3. Perancangan skema Data Warehouse.
4. Pengembangan *pipeline* ETL.
5. Implementasi indeks, partisi, dan strategi penyimpanan.

### 3. Metodologi



Gambar 1. Diagram Alir Metodologi

Pembangunan sistem Data Warehouse RBN menggunakan pendekatan bertahap yang mencakup analisis kebutuhan, perancangan skema, ETL, serta evaluasi dan dokumentasi hasil. Proses dimulai dengan menganalisis kebutuhan dan studi kasus. Tahap ini juga mengidentifikasi stakeholder utama dan tujuan bisnis yang ingin dicapai dalam konteks pengembangan jaringan terutama ke wilayah 3S.

Selanjutnya merancang skema multidimensional dengan skema bintang. Skema ini terdiri dari tabel fakta yang merekam konsumsi energi, dan enam tabel dimensi yang berisi dimensi waktu, pelanggan, geospasial, perangkat, gangguan, dan tagihan. Dimensi-dimensi ini dirancang dengan pertimbangan struktur hierarki dan kebutuhan agregasi data.

Kemudian tahap ekstraksi data dari berbagai sumber. Setelah data terkumpul, dilakukan transformasi terhadap data, dan data yang sudah bersih dimuat ke dalam sistem basis data. Proses

memuat data dilakukan pengaturan relasi antar tabel dan pengisian data ke dalam tabel fakta dan dimensi.

Untuk memastikan performa sistem Data Warehouse, dilakukan optimasi melalui penerapan indeks, partisi tabel fakta berdasarkan waktu, dan penggunaan kompresi data untuk efisien penyimpanan. Kemudian tahap evaluasi dan dokumentasi yang meliputi validasi integritas data, pengujian kueri analitik, dan penyusunan dokumentasi teknis.

#### 4. Analisis Kebutuhan

##### 4.1. Daftar *Stakeholder* dan Tujuan Bisnis

###### 4.1.1. Daftar *Stakeholder*

**Tabel 1.** Tabel Daftar *Stakeholder*

No.	Jabatan	Peran
1.	Manajer Operasional Wilayah	Memantau operasi distribusi listrik di wilayah tertentu, termasuk wilayah 3S, dan mengkoordinasi pemeliharaan jaringan.
2.	Tim Perencanaan Jaringan	Merancang strategi ekspansi jaringan ke wilayah baru berdasarkan area wilayah dan konsumsi
3.	Tim IT / Data Engineer	Bertanggung jawab dalam mengelola infrastruktur data, integrasi sistem, dan pengembangan gudang data.
4.	Customer Service & Billing	Mengelola interaksi pelanggan, penanganan keluhan, dan memastikan sinkronisasi dengan sistem billing.
5.	Analisis Keuangan	Menganalisis efisiensi operasional dan pengeluaran, serta mendukung keputusan investasi ekspansi.
6.	Tim Monitoring Gangguan	Memantau status sistem kelistrikan dan mendeteksi adanya gangguan atau kebocoran daya secara real-time.

###### 4.1.2. Tujuan Bisnis

**Tabel 2.** Tabel Tujuan Bisnis

No.	Tujuan	Deskripsi
1.	Meningkatkan efisiensi strategi ekspansi ke wilayah 3S	Wilayah 3S (Sulit dijangkau, Sulit diawasi, Sulit dimonitor) memerlukan strategi berbasis data agar ekspansi jaringan lebih tepat sasaran dan hemat biaya. Sistem integrasi data dapat membantu dalam pemetaan dan pengambilan keputusan.
2.	Mengurangi gangguan dan kebocoran daya	Dengan menyatukan data gangguan dari berbagai sistem ke dalam satu platform, perusahaan dapat mendeteksi pola gangguan, area kritis, dan mengambil tindakan lebih cepat dan preventif.
3.	Meningkatkan integrasi data dan pelaporan yang cepat dan akurat	Integrasi data dari sistem operasional, pelanggan, dan billing ke dalam gudang data akan mempercepat pembuatan laporan dan mengurangi risiko kesalahan data.
4.	Mempermudah pemantauan konsumsi dan perencanaan peak-load	Riwayat konsumsi pelanggan yang terpusat memungkinkan tim teknis memprediksi kebutuhan energi saat beban puncak dan mengoptimalkan distribusi daya.

5.	Meningkatkan pelayanan pelanggan melalui sistem yang terhubung dan prediktif	Dengan dashboard otomatis dan sistem yang saling terhubung, perusahaan dapat meningkatkan respon terhadap keluhan, memberikan notifikasi gangguan, dan meningkatkan kepuasan pelanggan.
----	--	---

#### 4.1.3. Interview Simulasi

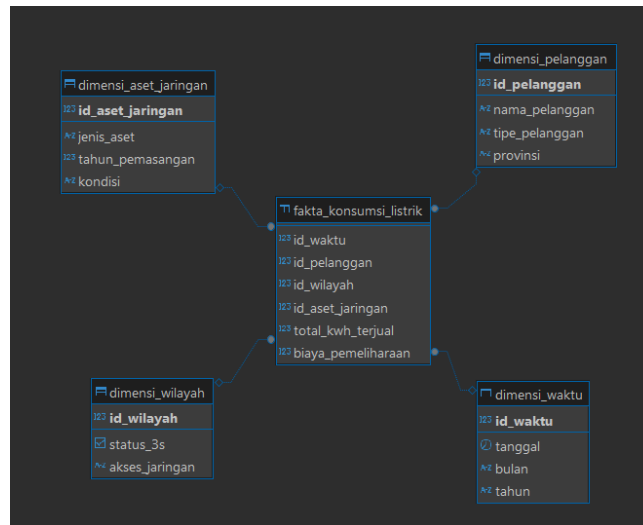
**Tabel 3.** Tabel *Interview Simulasi*

No.	Stakeholder	Pertanyaan Interview Simulasi
1	Manajer Operasional Wilayah	Jika tersedia gudang data terpusat, data apa saja yang paling dibutuhkan untuk memantau gangguan dan operasional jaringan di wilayah 3S secara lebih efisien?
2	Tim Perencanaan Jaringan	Data seperti apa yang perlu tersedia dan terintegrasi dalam gudang data untuk menyusun strategi ekspansi jaringan yang hemat biaya dan tepat sasaran ke wilayah 3S?
3	Tim IT / Data Engineer	Apa tantangan dalam menggabungkan sistem penyimpanan data pelanggan, gangguan, serta billing dengan sistem terpisah ke dalam satu gudang data yang mendukung integrasi dan dashboard otomatis?
4	Customer Service & Billing	Bagaimana cara untuk melihat peran gudang data dalam menyatukan informasi pelanggan, riwayat gangguan, dan tagihan, agar dapat meningkatkan kecepatan respons terhadap keluhan pelanggan?
5	Analisis Keuangan	Laporan seperti apa yang dihasilkan oleh gudang data agar dapat menganalisis efisiensi operasional dan mendukung keputusan ekspansi?
6	Tim Monitoring Gangguan	Informasi apa yang paling penting dipantau secara real-time jika terjadi gangguan dan kebocoran daya yang tidak terhubung ke data pelanggan dan billing?

#### 4.1.4. Studi Kasus

Perusahaan RubicoNergi mengalami permasalahan mengenai jangkauan pasokan listrik untuk daerah 3S (Sulit dijangkau, Sulit diawasi, dan Sulit dimonitor) akibat permasalahan data geospasial dan konsumsi listrik setiap kepala keluarga yang terfragmentasi dan tidak terintegrasi. Hal ini menyebabkan biaya pembuatan dan pemeliharaan infrastruktur menjadi lebih tinggi dari wilayah-wilayah lain (non 3S). Ketidakterpaduan sistem juga menyulitkan perusahaan dalam mengidentifikasi wilayah prioritas ekspansi dan menyebabkan lambatnya respons terhadap gangguan jaringan serta kebocoran daya. Selain itu, absennya integrasi antara sistem pelanggan, sistem gangguan, dan sistem billing berpotensi menimbulkan risiko kebocoran data pelanggan yang sensitif. Untuk menjawab permasalahan ini, RBN memerlukan solusi berupa gudang data (data warehouse) yang mampu menghimpun dan menyatukan berbagai sumber data penting, baik historis maupun operasional. Dengan sistem ini, perusahaan dapat melakukan analisis spasial, pemantauan konsumsi secara menyeluruh, serta perencanaan jaringan yang lebih strategis dan efisien. Kehadiran gudang data diharapkan mampu mendukung pengambilan keputusan berbasis data, mempercepat pelaporan, serta meningkatkan kualitas pelayanan kepada pelanggan, khususnya dalam upaya pemerataan distribusi energi ke wilayah 3S yang selama ini terabaikan.

## 4.2. Fakta dan Dimensi



Gambar 2. Star Schema

### 4.2.1. Tabel Fakta

Tabel 4. Tabel Fakta

Tabel	Atribut	Deskripsi
fakta_konsumsi_listrik	id_waktu	Kunci asing yang menghubungkan ke dimensi_waktu
	id_pelanggan	Kunci asing yang menghubungkan ke dimensi_pelanggan
	id_wilayah	Kunci asing yang menghubungkan ke dimensi_wilayah
	id_aset_jaringan	Kunci asing yang menghubungkan ke dimensi_aset_jaringan
	total_kwh_terjual	Jumlah kilowatt-hour listrik yang terjual
	biaya_pemeliharaan	Biaya yang dikeluarkan untuk pemeliharaan

### 4.2.2. Tabel Dimensi

Tabel 5. Tabel Dimensi

Tabel	Atribut	Deskripsi
dimensi_waktu	id_waktu	Kunci utama yang mengidentifikasi periode waktu
	tanggal	Tanggal pencatatan
	bulan	Bulan pencatatan
	tahun	Tahun pencatatan
dimensi_pelanggan	id_pelanggan	Kunci utama yang mengidentifikasi pelanggan
	nama_pelanggan	Nama pelanggan

	tipe_pelanggan	Kategori pelanggan
	provinsi	Provinsi tempat pelanggan berada
dimensi_wilayah	id_wilayah	Kunci utama yang mengidentifikasi wilayah
	status_3s	Status kondisi jaringan 3s
	akses_jaringan	Tipe akses jaringan pada wilayah tersebut
dimensi_aset_jaringan	id_aset_jaringan	Kunci utama yang mengidentifikasi aset jaringan
	jenis_aset	Jenis atau tipe aset jaringan listrik
	tahun_pemasangan	Tahun pemasangan aset jaringan
	kondisi	Kondisi aset jaringan saat ini

### 4.3. Sumber dan Metadata

#### 4.3.1. Identifikasi Sumber Data

Nama Dataset	:	Household Electric Power Consumption
Jenis File	:	Comma-Separated Values (CSV) dengan format .txt
Sumber Data	:	Kaggle
Tipe Sumber Data	:	Data historis konsumsi energi rumah tangga yang dicatat dari sensor secara otomatis (berbasis <i>file</i> , bukan <i>real-time</i> )
Link Dataset	:	<a href="https://www.kaggle.com/datasets/uciml/electric-power-consumption-on-data-set">https://www.kaggle.com/datasets/uciml/electric-power-consumption-on-data-set</a>

#### 4.3.2. Frekuensi Update

Sumber Data	:	Kaggle
Frekuensi Update	:	Tidak diperbarui (statis, arsip Desember 2006 – Desember 2010)

#### 4.3.3. Dokumentasi Metadata

**Tabel 6.** Tabel Dokumentasi Metadata Dataset: Household Electric Power Consumption

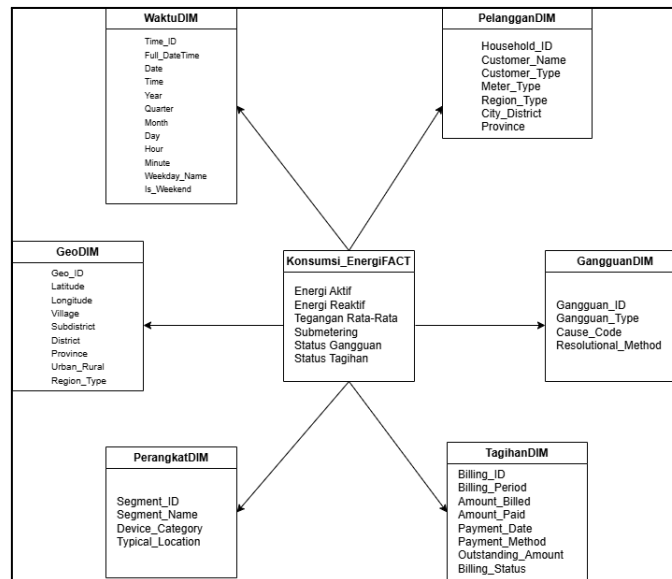
Nama Kolom	Tipe Data	Deskripsi
Date	String (format dd/mm/yyyy)	Tanggal pencatatan konsumsi listrik
Time	String (format hh:mm:ss)	Waktu pencatatan
Global_active_power	Float	Konsumsi daya aktif total (kilowatt)
Global_reactive_power	Float	Konsumsi daya reaktif total (kilowatt)
Voltage	Float	Tegangan (volt)
Global_intensity	Float	Intensitas arus global (ampere)



Sub_metering_1	Float	Energi sub-metering 1 (Wh) – dapur/peralatan
Sub_metering_2	Float	Energi sub-metering 2 (Wh) – <i>laundry room</i>
Sub_metering_3	Float	Energi sub-metering 3 (Wh) <i>water heater</i> dan <i>air conditioner</i>

## 5. Desain Konseptual, Logikal, dan Fisikal

### 5.1. Skema Konseptual Multidimensi



Gambar 3. Skema Konseptual dengan *Star Schema*

#### Dimensi WaktuDIM

Dimensi ini digunakan untuk menyediakan konteks temporal yang sangat detail. Dengan Time\_ID sebagai primary key, dimensi ini menyimpan atribut dasar seperti Full\_DateTime (tanggal dan jam), serta komponen terpisah Date dan Time. Dari atribut dasar tersebut diturunkan kolom-kolom seperti Year, Quarter, Month, Day, Hour, Minute, Weekday\_Name, dan Is\_Weekend. Terdapat hierarki Menit → Jam → Hari → Bulan → Kuartal → Tahun memungkinkan pengguna untuk melakukan agregasi dan drill-down mulai dari tingkat paling granular (menit) hingga tingkat agregat (tahun), mendukung analisis pola harian, fluktuasi bulanan, dan tren tahunan.

#### Dimensi PelangganDIM

Dimensi ini, Household\_ID digunakan untuk mengidentifikasi setiap pelanggan. Atribut-atribut di dalamnya meliputi Customer\_Name, Customer\_Type (misalnya Rumah Tangga atau Komersial), Meter\_Type (Smart atau Manual), Region\_Type (3S atau Non-3S), serta lokasi administratif seperti City\_District dan Province. Terdapat hierarki Pelanggan → Kota/Kabupaten → Provinsi → Region\_Type memungkinkan segmentasi pelanggan berdasarkan area geografis dan kategori wilayah, sehingga perusahaan dapat melakukan analisis perilaku konsumsi menurut tipe pelanggan dan prioritas ekspansi ke wilayah 3S.

#### Dimensi GeoDIM

Dimensi ini digunakan untuk menekankan pada aspek spasial dengan Geo\_ID sebagai kunci utama. Atribut geografis mencakup Latitude dan Longitude, serta hierarki administratif Village,

Subdistrict, District, dan Province. Atribut tambahan seperti Urban\_Rural (Perkotaan/Pedesaan) dan Region\_Type (3S/Non-3S) memudahkan visualisasi peta distribusi konsumsi energi dan titik gangguan. Terdapat hierarki Koordinat → Desa → Kecamatan → Kab/Kota → Provinsi mendukung analisis spasial mendetail, misalnya identifikasi klaster konsumsi tinggi atau area rawan gangguan.

#### **Dimensi PerangkatDIM**

Dimensi ini fokus pada segmentasi konsumsi berdasarkan area fungsi di dalam rumah. Masing-masing sub-metering (Sub\_metering\_1, Sub\_metering\_2, Sub\_metering\_3) dihubungkan ke Segment\_ID, dengan atribut Segment\_Name, Device\_Category (misalnya Dapur, Laundry, HVAC), dan Typical\_Location. Terdapat hierarki Sub\_metering → Kategori Perangkat → Lokasi yang membantu dalam analisis pemakaian energi di tiap ruang atau kategori peralatan, sehingga inisiatif efisiensi dapat ditargetkan pada area paling boros energi.

#### **Dimensi GangguanDIM**

Dimensi ini digunakan untuk menyimpan informasi deskriptif tentang kejadian gangguan di jaringan listrik. Dengan Gangguan\_ID sebagai primary key, terdapat atribut Gangguan\_Type (Pemadaman, Fluktuasi Tegangan, Kebocoran), Cause\_Code (misalnya Bencana Alam, Pemeliharaan), dan Resolution\_Method (Otomatis, Manual). Terdapat hierarki Jenis Gangguan → Penyebab → Metode Penanganan yang memungkinkan analisis keandalan jaringan dan efektivitas respons tim teknis, serta perencanaan preventive maintenance berdasarkan pola penyebab gangguan.

#### **Dimensi TagihanDIM**

Dimensi ini digunakan untuk menelusuri siklus penagihan listrik. Didalamnya berisi Billing\_ID yang digunakan untuk mengidentifikasi setiap periode tagihan, dengan atribut Billing\_Period (Bulan-Tahun), Amount\_Billed, Amount\_Paid, Payment\_Date, Payment\_Method (Kartu, Transfer, Tunai), Outstanding\_Amount, dan Billing\_Status (Lunas, Tunggakan, Belum Bayar). Terdapat hierarki Periode Tagihan → Status Pembayaran yang memudahkan laporan arus kas serta analisis performa penagihan, sehingga strategi penagihan dan mitigasi piutang dapat dioptimalkan.

### **5.2. Justifikasi Desain Konseptual**

#### **Fakta Konsumsi Listrik Pelanggan**

Konsumsi listrik pelanggan menjadi pusat analisis yang mencerminkan penggunaan energi aktual pelanggan. Dengan data ini, perusahaan dapat mengidentifikasi tren konsumsi, menganalisis *peak demand*, dan mengukur dampak kebijakan.

#### **Dimensi Waktu**

Dimensi waktu memungkinkan perusahaan untuk menganalisis longitudinal, mengidentifikasi pola konsumsi pelanggan berdasarkan ketentuan, dan memprediksi permintaan di masa depan.

#### **Dimensi Pelanggan**

Dimensi pelanggan mendukung pembuatan kebijakan yang strategis dan tertarget.

#### **Dimensi Wilayah**

Wilayah digunakan untuk analisis spasial guna memahami pola distribusi konsumen di area geografis. Dimensi wilayah membantu mengidentifikasi wilayah dengan permintaan tertinggi, masalah distribusi, ataupun potensi efisiensi energi.

#### **Dimensi Jaringan**

Data jaringan mendukung korelasi antara kondisi fisik infrastruktur dengan kinerja distribusi. Hal ini membantu dalam menganalisis risiko, perencanaan pemeliharaan, dan pengambilan keputusan.

### Dimensi Gangguan

Data gangguan digunakan untuk mengukur kendala sistem distribusi yang mengidentifikasi area rawan gangguan dan merencanakan improvisasi pelayanan.

### Dimensi Tagihan

Dimensi tagihan penting untuk analisis dampak konsumsi terhadap pendapatan, kepatuhan pembayaran, dan evaluasi efektivitas tarif.

### 5.3. Kesesuaian dengan Sumber Data

Sumber data yang digunakan ialah *Household Electric Power Consumption* yang merupakan dataset yang diambil dari Kaggle. Dataset ini berisi data historis konsumsi energi rumah tangga yang dicatat secara otomatis dari sensor selama periode Desember 2006 hingga Desember 2010. Dataset ini berbentuk file CSV dengan berbagai kolom seperti tanggal, waktu, konsumsi daya aktif, daya reaktif, tegangan, arus, dan sub-metering energi untuk area penggunaan tertentu (dapur, laundry, dan sistem pemanas/AC). Untuk menjamin bahwa desain data warehouse sesuai dengan sumber data yang tersedia, dilakukan pemetaan antara elemen-elemen dalam skema dengan atribut aktual di dataset.

Kesesuaian terhadap komponen skema multidimensi:

#### Kesesuaian dengan Tabel Fakta

Tabel 7. Kesesuaian dengan Tabel Fakta

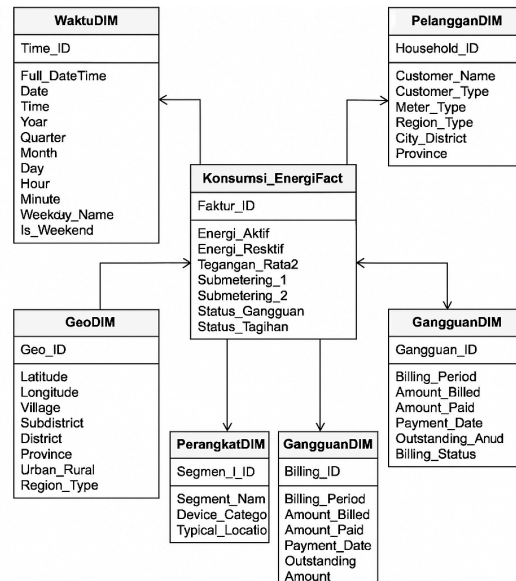
Komponen Fakta	Atribut di Skema	Keterangan
Konsumsi Energi	total_kwh_terjual	Dihitung dari Global_active_power pada dataset
Biaya Pemeliharaan	biaya_pemeliharaan	Estimasi, berdasarkan konsumsi dan kondisi aset (data tambahan)
Waktu Pencatatan	id_waktu	Foreign key ke dimensi waktu
Pelanggan	id_pelanggan	Foreign key ke dimensi pelanggan (disiapkan dari sistem pelanggan RBN)
Wilayah	id_wilayah	Foreign key ke dimensi wilayah
Aset Jaringan	id_aset_jaringan	Foreign key ke dimensi aset jaringan

#### Kesesuaian dengan Tabel Dimensi

Tabel 8. Kesesuaian dengan Tabel Dimensi

Komponen Dimensi	Atribut di Skema	Sumber/Kesesuaian
Dimensi Waktu	tanggal, bulan, tahun	Diambil langsung dari kolom Date pada dataset
Dimensi Pelanggan	tipe_pelanggan, provinsi	Tidak tersedia langsung; diintegrasikan dari sistem pelanggan
Dimensi Wilayah	status_3s, akses_jaringan	Diasumsikan berasal dari sistem geografis RBN
Dimensi Aset Jaringan	jenis_aset, kondisi	Tidak tersedia langsung; diasumsikan dari sistem inventori

## 5.4. Desain Tabel Dimensi



Gambar 3. Desain Tabel Dimensi

## 5.5. Desain Tabel Fakta

Tabel fakta utama yang dibangun dalam sistem pergudangan data RubicoNergi adalah Konsumsi\_EnergiFact. Tabel ini menyimpan data hasil transformasi dari berbagai sumber dan menjadi pusat integrasi analitik lintas domain: konsumsi energi, status gangguan, dan informasi tagihan pelanggan. Terdapat 7 entitas yang ada pada tabel fakta yaitu :

Energi\_aktif : Total energi aktif (kWh) per waktu rekam  
 Energi\_reaktif : Total energi reaktif (kWh)  
 Tegangan\_rata2 : Tegangan rata-rata (Volt) per periode waktu  
 Sub\_metering\_1 : Energi dari submetering 1 (dapur)  
 Sub\_metering\_2 : Energi dari submetering 2 (laundry)  
 Status\_gangguan : Status gangguan saat itu  
 Status\_tagihan : Status pembayaran

Berikut atribut yang terdapat pada tabel fakta yang dibuat dalam bentuk tabel:

Tabel 9. Atribut Tabel

Kolom	Tipe Data	Keterangan
id_waktu	INT	Foreign Key ke dimensi waktu
id_pelanggan	INT	Foreign Key ke dimensi pelanggan
id_geo	INT	Foreign Key ke dimensi geospasial
id_perangkat	INT	Foreign Key ke dimensi perangkat/submetering
id_gangguan	INT	Foreign Key ke dimensi gangguan
id_tagihan	INT	Foreign Key ke dimensi tagihan
energi_aktif	FLOAT	Total energi aktif (kWh) per waktu rekam

energi_reaktif	FLOAT	Energi reaktif total dalam satuan kWh
tegangan_rata2	FLOAT	Tegangan rata-rata harian dalam volt
sub_metering_1	FLOAT	Energi dari submetering 1 (dapur)
sub_metering_1	FLOAT	Energi dari submetering 2 (laundry)
status_gangguan	VARCHAR	Status gangguan saat itu
status_tagihan	VARCHAR	Status pembayaran seperti “Lunas”, “Tunggakan”, “Belum Bayar”

## 6. Proses Implementasi

Proses implementasi dalam proyek ini menggunakan pendekatan medalion yang terdiri dari tiga lapisan bertingkat, yaitu Bronze (*Raw*), Silver (*Cleaned*), dan Gold (*Aggregated*).

**Algoritma 1.** Bronze Layer

```
-- BRONZE LAYER: Menyimpan data mentah persis seperti dari sumber
CREATE SCHEMA bronze;
GO

CREATE TABLE bronze.raw_power_consumption (
  id INT IDENTITY(1,1) PRIMARY KEY,
  [Date] VARCHAR(50),
  [Time] VARCHAR(50),
  Global_active_power VARCHAR(50),
  Global_reactive_power VARCHAR(50),
  Voltage VARCHAR(50),
  Global_intensity VARCHAR(50),
  Sub_metering_1 VARCHAR(50),
  Sub_metering_2 VARCHAR(50),
  Sub_metering_3 VARCHAR(50),
  load_timestamp DATETIME DEFAULT GETDATE(),
  source_file VARCHAR(255)
);
GO

-- Proses loading data ke bronze layer
INSERT INTO bronze.raw_power_consumption (
  [Date],
  [Time],
  Global_active_power,
  Global_reactive_power,
  Voltage,
  Global_intensity,
  Sub_metering_1,
  Sub_metering_2,
  Sub_metering_3,
  source_file
)
SELECT
  [Date],
```

```

[Time],
Global_active_power,
Global_reactive_power,
Voltage,
Global_intensity,
Sub_metering_1,
Sub_metering_2,
Sub_metering_3,
'household_power_consumption.csv' -- contoh nama file sumber
FROM
[household_power_consumption$];
GO

```

Bronze Layer menyimpan data mentah tanpa transformasi dengan tambahan metadata. Layer ini berfungsi sebagai area penyimpanan awal yang mempertahankan data asli untuk diaudit dan dilacak. Tabel 'bronze.raw\_power\_consumption' menyimpan data teks dari sumber dengan atribut-atributnya.

#### Algoritma 2. Silver Layer

```

-- SILVER LAYER: Data yang sudah dibersihkan dan ditransformasi
CREATE SCHEMA silver;
GO

CREATE TABLE silver.fact_power_consumption (
    power_consumption_id INT IDENTITY(1,1) PRIMARY KEY,
    consumption_datetime DATETIME2,
    consumption_date DATE,
    consumption_time TIME,
    global_active_power DECIMAL(10,6),
    global_reactive_power DECIMAL(10,6),
    voltage DECIMAL(10,6),
    global_intensity DECIMAL(10,6),
    sub_metering_1 DECIMAL(10,6),
    sub_metering_2 DECIMAL(10,6),
    sub_metering_3 DECIMAL(10,6),
    total_power_consumption DECIMAL(10,6),
    is_weekend BIT,
    load_timestamp DATETIME2 DEFAULT SYSUTCDATETIME()
);
GO

-- Transformasi data dari bronze ke silver
INSERT INTO silver.fact_power_consumption (
    consumption_datetime,
    consumption_date,
    consumption_time,
    global_active_power,
    global_reactive_power,
    voltage,
    global_intensity,
    sub_metering_1,
    sub_metering_2,
    sub_metering_3,
    total_power_consumption,

```

```

is_weekend
)
SELECT
-- Menggabungkan date dan time menjadi datetime
TRY_CONVERT(DATETIME2, CONCAT([Date], ' ', [Time])) AS consumption_datetime,

-- Konversi tipe data dan pembersihan
TRY_CONVERT(DATE, [Date]) AS consumption_date,
TRY_CONVERT(TIME, [Time]) AS consumption_time,
TRY_CONVERT(DECIMAL(10,6), Global_active_power) AS global_active_power,
TRY_CONVERT(DECIMAL(10,6), Global_reactive_power) AS global_reactive_power,
TRY_CONVERT(DECIMAL(10,6), Voltage) AS voltage,
TRY_CONVERT(DECIMAL(10,6), Global_intensity) AS global_intensity,
TRY_CONVERT(DECIMAL(10,6), Sub_metering_1) AS sub_metering_1,
TRY_CONVERT(DECIMAL(10,6), Sub_metering_2) AS sub_metering_2,
TRY_CONVERT(DECIMAL(10,6), Sub_metering_3) AS sub_metering_3,

-- Menghitung total power consumption (active + sub meters)
TRY_CONVERT(DECIMAL(10,6), Global_active_power) * 1000 / 60 +
TRY_CONVERT(DECIMAL(10,6), Sub_metering_1) +
TRY_CONVERT(DECIMAL(10,6), Sub_metering_2) +
TRY_CONVERT(DECIMAL(10,6), Sub_metering_3) AS total_power_consumption,

-- Flag weekend
CASE WHEN DATEPART(WEEKDAY, TRY_CONVERT(DATE, [Date])) IN (1, 7) THEN 1
ELSE 0 END AS is_weekend
FROM
    bronze.raw_power_consumption
WHERE
    -- Filter data yang tidak valid
    TRY_CONVERT(DATE, [Date]) IS NOT NULL
    AND TRY_CONVERT(TIME, [Time]) IS NOT NULL
    AND TRY_CONVERT(DECIMAL(10,6), Global_active_power) IS NOT NULL;
GO

```

Algoritma 2 merupakan Silver Layer yang mentransformasi dan membersihkan data dari Bronze Layer. Data dikonversi ke tipe yang tepat, dibersihkan dari nilai yang tidak valid, dan diperkaya dengan kolom turunan seperti total konsumsi daya dan flag akhir pekan. Tabel 'silver.fact\_power\_consumption' menyimpan data yang telah terstandarisasi dan siap diproses.

### Algoritma 3. Gold Layer

```

-- GOLD LAYER: Data agregasi untuk analisis bisnis
CREATE SCHEMA gold;
GO

-- Tabel agregasi harian
CREATE TABLE gold.daily_power_consumption (
    consumption_date DATE PRIMARY KEY,
    avg_active_power DECIMAL(10,6),
    avg_reactive_power DECIMAL(10,6),
    avg_voltage DECIMAL(10,6),
    total_consumption_wh DECIMAL(15,6),
    avg_sub1 DECIMAL(10,6),

```

```

avg_sub2 DECIMAL(10,6),
avg_sub3 DECIMAL(10,6),
weekday_consumption_wh DECIMAL(15,6),
weekend_consumption_wh DECIMAL(15,6),
load_timestamp DATETIME2 DEFAULT SYSUTCDATETIME()
);
GO

-- Mengisi gold layer dengan agregasi dari silver layer
INSERT INTO gold.daily_power_consumption (
    consumption_date,
    avg_active_power,
    avg_reactive_power,
    avg_voltage,
    total_consumption_wh,
    avg_sub1,
    avg_sub2,
    avg_sub3,
    weekday_consumption_wh,
    weekend_consumption_wh
)
SELECT
    consumption_date,
    AVG(global_active_power) AS avg_active_power,
    AVG(global_reactive_power) AS avg_reactive_power,
    AVG(voltage) AS avg_voltage,
    SUM(total_power_consumption) AS total_consumption_wh,
    AVG(sub_metering_1) AS avg_sub1,
    AVG(sub_metering_2) AS avg_sub2,
    AVG(sub_metering_3) AS avg_sub3,
    SUM(CASE WHEN is_weekend = 0 THEN total_power_consumption ELSE 0 END) AS
weekday_consumption_wh,
    SUM(CASE WHEN is_weekend = 1 THEN total_power_consumption ELSE 0 END) AS
weekend_consumption_wh
FROM
    silver.fact_power_consumption
GROUP BY
    consumption_date;
GO

-- Tabel agregasi bulanan
CREATE TABLE gold.monthly_power_consumption (
    year_month CHAR(7) PRIMARY KEY, -- Format: 'YYYY-MM'
    year INT,
    month INT,
    total_consumption_wh DECIMAL(15,6),
    avg_daily_consumption_wh DECIMAL(15,6),
    peak_active_power DECIMAL(10,6),
    avg_sub1 DECIMAL(10,6),
    avg_sub2 DECIMAL(10,6),
    avg_sub3 DECIMAL(10,6)
);
GO

```



```

INSERT INTO gold.monthly_power_consumption
SELECT
    FORMAT(consumption_date, 'yyyy-MM') AS year_month,
    YEAR(consumption_date) AS year,
    MONTH(consumption_date) AS month,
    SUM(total_consumption_wh) AS total_consumption_wh,
    AVG(total_consumption_wh) AS avg_daily_consumption_wh,
    MAX(avg_active_power) AS peak_active_power,
    AVG(avg_sub1) AS avg_sub1,
    AVG(avg_sub2) AS avg_sub2,
    AVG(avg_sub3) AS avg_sub3
FROM
    gold.daily_power_consumption
GROUP BY
    FORMAT(consumption_date, 'yyyy-MM'),
    YEAR(consumption_date),
    MONTH(consumption_date);
GO

```

Gold Layer berfokus pada agregasi data. Tahapan ini berfokus pada dua tabel utama, yaitu 'gold.daily\_power\_consumption' dan 'gold.monthly\_power\_consumption'. Lapisan ini menghitung metrik khusus seperti perbandingan konsumsi *weekday* vs *weekend*, yang siap digunakan untuk visualisasi dan pelaporan.

#### Algoritma 4. Membuat Tabel dan Memasukkan Data

##### Membuat Tabel Fakta dan Dimensi:

```

-- Tabel Dimensi Tanggal
CREATE TABLE dim_date (
    date_id INT PRIMARY KEY,
    date_value DATE NOT NULL,
    day INT,
    month INT,
    year INT,
    weekday VARCHAR(10)
);
-- Tabel Dimensi Waktu
CREATE TABLE dim_time (
    time_id INT PRIMARY KEY,
    time_value TIME NOT NULL
);
-- Tabel Dimensi Meter
CREATE TABLE dim_meter (
    meter_id INT PRIMARY KEY,
    meter_name VARCHAR(50),
    description VARCHAR(255)
);
-- Tabel Fakta Konsumsi Energi
CREATE TABLE fact_energy_consumption (
    fact_id INT IDENTITY(1,1) PRIMARY KEY,
    date_id INT,
    time_id INT,
    global_active_power FLOAT,
    global_reactive_power FLOAT,
    voltage FLOAT,

```

global_intensity <b>FLOAT</b> , sub_metering_1 <b>FLOAT</b> , sub_metering_2 <b>FLOAT</b> , sub_metering_3 <b>FLOAT</b> , <b>FOREIGN KEY</b> (date_id) <b>REFERENCES</b> dim_date(date_id), <b>FOREIGN KEY</b> (time_id) <b>REFERENCES</b> dim_time(time_id) );
Memasukkan Data:
<b>INSERT INTO</b> dim_date (date_id, date_value, <b>day</b> , <b>month</b> , <b>year</b> , weekday) <b>VALUES</b> (1, '2006-12-16', 16, 12, 2006, 'Saturday'); <b>INSERT INTO</b> dim_date (date_id, date_value, <b>day</b> , <b>month</b> , <b>year</b> , weekday) <b>VALUES</b> (2, '2006-12-17', 17, 12, 2006, 'Sunday');  <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (1, '17:24:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (2, '17:25:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (3, '17:26:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (4, '17:27:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (5, '17:28:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (6, '17:29:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (7, '17:30:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (8, '17:31:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (9, '17:32:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (10, '17:33:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (11, '17:34:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (12, '17:35:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (13, '17:36:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (14, '17:37:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (15, '17:38:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (16, '17:39:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (17, '17:40:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (18, '17:41:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (19, '17:42:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (20, '17:43:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (21, '17:44:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (22, '17:45:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (23, '17:46:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (24, '17:47:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (25, '17:48:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (26, '17:49:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (27, '17:50:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (28, '17:51:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (29, '17:52:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (30, '17:53:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (31, '17:54:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (32, '17:55:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (33, '17:56:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (34, '17:57:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (35, '17:58:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (36, '17:59:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (37, '18:00:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (38, '18:01:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (39, '18:02:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (40, '18:03:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (41, '18:04:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (42, '18:05:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (43, '18:06:00'); <b>INSERT INTO</b> dim_time (time_id, time_value) <b>VALUES</b> (44, '18:07:00');









































```

INSERT INTO dim_time (time_id, time_value) VALUES (989, '09:52:00');
INSERT INTO dim_time (time_id, time_value) VALUES (990, '09:53:00');
INSERT INTO dim_time (time_id, time_value) VALUES (991, '09:54:00');
INSERT INTO dim_time (time_id, time_value) VALUES (992, '09:55:00');
INSERT INTO dim_time (time_id, time_value) VALUES (993, '09:56:00');
INSERT INTO dim_time (time_id, time_value) VALUES (994, '09:57:00');
INSERT INTO dim_time (time_id, time_value) VALUES (995, '09:58:00');
INSERT INTO dim_time (time_id, time_value) VALUES (996, '09:59:00');
INSERT INTO dim_time (time_id, time_value) VALUES (997, '10:00:00');
INSERT INTO dim_time (time_id, time_value) VALUES (998, '10:01:00');
INSERT INTO dim_time (time_id, time_value) VALUES (999, '10:02:00');
INSERT INTO dim_time (time_id, time_value) VALUES (1000, '10:03:00');

INSERT INTO dim_meter (meter_id, meter_name, description) VALUES (1, 'Sub_metering_1', 'Kitchen
appliances (dishwasher, oven, microwave)');
INSERT INTO dim_meter (meter_id, meter_name, description) VALUES (2, 'Sub_metering_2', 'Laundry
room (washing machine, dryer)');
INSERT INTO dim_meter (meter_id, meter_name, description) VALUES (3, 'Sub_metering_3', 'Water heater
and air conditioner');

INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 1.0, 4.216,
0.418, 234.84, 18.4, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 2.0, 5.36,
0.436, 233.63, 23.0, 0.0, 1.0, 16.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 3.0, 5.374,
0.498, 233.29, 23.0, 0.0, 2.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 4.0, 5.388,
0.502, 233.74, 23.0, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 5.0, 3.666,
0.528, 235.68, 15.8, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 6.0, 3.52,
0.522, 235.02, 15.0, 0.0, 2.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 7.0, 3.702,
0.52, 235.09, 15.8, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 8.0, 3.7, 0.52,
235.22, 15.8, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 9.0, 3.668,
0.51, 233.99, 15.8, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 10.0, 3.662,
0.51, 233.86, 15.8, 0.0, 2.0, 16.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 11.0, 4.448,
0.498, 232.86, 19.6, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 12.0, 5.412,
0.47, 232.78, 23.2, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 13.0, 5.224,
0.478, 232.99, 22.4, 0.0, 1.0, 16.0);

```





```

0.09, 233.77, 18.4, 0.0, 0.0, 16.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 34.0, 4.512,
0.0, 233.62, 19.2, 0.0, 0.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 35.0, 4.058,
0.2, 234.68, 17.6, 0.0, 0.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 36.0, 2.472,
0.058, 236.94, 10.4, 0.0, 0.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 37.0, 2.79,
0.18, 237.52, 11.8, 0.0, 0.0, 18.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 38.0, 2.624,
0.144, 238.2, 11.0, 0.0, 0.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 39.0, 2.772,
0.118, 238.28, 11.6, 0.0, 0.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 40.0, 3.74,
0.108, 236.93, 16.4, 0.0, 16.0, 18.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 41.0, 4.928,
0.202, 235.01, 21.0, 0.0, 37.0, 16.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 42.0, 6.052,
0.192, 232.93, 26.2, 0.0, 37.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 43.0, 6.752,
0.186, 232.12, 29.0, 0.0, 36.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 44.0, 6.474,
0.144, 231.85, 27.8, 0.0, 37.0, 16.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 45.0, 6.308,
0.116, 232.25, 27.0, 0.0, 36.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 46.0, 4.464,
0.136, 234.66, 19.0, 0.0, 37.0, 16.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 47.0, 3.396,
0.148, 236.2, 15.0, 0.0, 22.0, 18.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 48.0, 3.09,
0.152, 237.07, 13.8, 0.0, 12.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 49.0, 3.73,
0.144, 235.78, 16.4, 0.0, 27.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 50.0, 2.308,
0.16, 237.43, 9.6, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 51.0, 2.388,
0.158, 237.26, 10.0, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 52.0, 4.598,
0.1, 234.25, 21.4, 0.0, 20.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,

```





```

0.09, 232.25, 18.2, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 93.0, 4.23,
0.09, 232.32, 18.2, 0.0, 2.0, 16.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 94.0, 3.924,
0.084, 232.79, 17.0, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 95.0, 4.218,
0.09, 232.09, 18.0, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 96.0, 4.224,
0.09, 231.96, 18.2, 0.0, 1.0, 16.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 97.0, 4.07,
0.088, 231.99, 17.4, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 98.0, 3.612,
0.09, 232.36, 15.6, 0.0, 2.0, 16.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 99.0, 3.458,
0.09, 232.71, 14.8, 0.0, 1.0, 17.0);
INSERT INTO fact_energy_consumption (date_id, time_id, Global_active_power, Global_reactive_power,
Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3) VALUES (1.0, 100.0, 3.434,
0.09, 232.01, 14.8, 0.0, 1.0, 16.0);

```

– Bronze

```

INSERT INTO bronze.raw_power_consumption (
    [Date],
    [Time],
    Global_active_power,
    Global_reactive_power,
    Voltage,
    Global_intensity,
    Sub_metering_1,
    Sub_metering_2,
    Sub_metering_3,
    source_file
) VALUES
('16/12/2006','17:24:00','4.216','0.418','234.840','18.400','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:25:00','5.360','0.436','233.630','23.000','0.000','1.000','16.000','first_1000_rows.csv'),
('16/12/2006','17:26:00','5.374','0.498','233.290','23.000','0.000','2.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:27:00','5.388','0.502','233.740','23.000','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:28:00','3.666','0.528','235.680','15.800','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:29:00','3.520','0.522','235.020','15.000','0.000','2.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:30:00','3.702','0.520','235.090','15.800','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:31:00','3.700','0.520','235.090','15.800','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:32:00','3.668','0.510','234.580','15.800','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:33:00','3.662','0.510','234.550','15.800','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:34:00','3.658','0.520','234.520','15.800','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:35:00','3.628','0.500','234.240','15.600','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:36:00','3.608','0.500','233.880','15.400','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:37:00','3.578','0.500','233.980','15.200','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:38:00','3.548','0.500','234.090','15.200','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:39:00','3.538','0.500','234.020','15.000','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:40:00','3.522','0.502','234.180','15.000','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:41:00','3.514','0.498','234.150','15.000','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:42:00','3.500','0.500','234.300','15.000','0.000','1.000','17.000','first_1000_rows.csv'),
('16/12/2006','17:43:00','3.484','0.498','234.150','14.800','0.000','1.000','17.000','first_1000_rows.csv'),

```

























```

('17/12/2006','04:34:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:35:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:36:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:37:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:38:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:39:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:40:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:41:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:42:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:43:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:44:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:45:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:46:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:47:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:48:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:49:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:50:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:51:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:52:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:53:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:54:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:55:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:56:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:57:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:58:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv'),
('17/12/2006','04:59:00','3.208','0.498','233.290','13.600','0.000','0.000','16.000','first_1000_rows.csv');
GO

```

– Silver

-- Transformasi data dari bronze ke silver

```

INSERT INTO silver.fact_power_consumption (
    consumption_datetime,
    consumption_date,
    consumption_time,
    global_active_power,
    global_reactive_power,
    voltage,
    global_intensity,
    sub_metering_1,
    sub_metering_2,
    sub_metering_3,
    total_power_consumption,
    is_weekend
)
SELECT
    -- Menggabungkan date dan time menjadi datetime
    TRY_CONVERT(DATETIME2, CONCAT([Date], ' ', [Time]), 103) AS consumption_datetime, -- Format
    103 untuk DD/MM/YYYY

    -- Konversi tipe data dan pembersihan
    TRY_CONVERT(DATE, [Date], 103) AS consumption_date, -- Format 103 untuk DD/MM/YYYY
    TRY_CONVERT(TIME, [Time]) AS consumption_time,
    TRY_CONVERT(DECIMAL(10,6), REPLACE(Global_active_power, ',', '.')) AS global_active_power, --
    Mengganti koma dengan titik
    TRY_CONVERT(DECIMAL(10,6), REPLACE(Global_reactive_power, ',', '.')) AS global_reactive_power,
    TRY_CONVERT(DECIMAL(10,6), REPLACE(Voltage, ',', '.')) AS voltage,
    TRY_CONVERT(DECIMAL(10,6), REPLACE(Global_intensity, ',', '.')) AS global_intensity,
    TRY_CONVERT(DECIMAL(10,6), REPLACE(Sub_metering_1, ',', '.')) AS sub_metering_1,
    TRY_CONVERT(DECIMAL(10,6), REPLACE(Sub_metering_2, ',', '.')) AS sub_metering_2,

```

```

TRY_CONVERT(DECIMAL(10,6), REPLACE(Sub_metering_3, ',', '.')) AS sub_metering_3,

-- Menghitung total power consumption (active + sub meters)
-- Asumsi Global_active_power dalam kilowatt, sub_metering dalam watt-hour
-- Konversi Global_active_power ke Wh: kWh * 1000
-- Jika Sub_metering_x sudah dalam Wh, tidak perlu dikonversi
-- Jika Global_active_power adalah kW dan sub_metering adalah Wh per menit, maka:
-- (Global_active_power * 1000 / 60) adalah Wh per menit dari active power
-- Ini adalah interpretasi dari perhitungan sebelumnya
(TRY_CONVERT(DECIMAL(10,6), REPLACE(Global_active_power, ',', '.')) * 1000 / 60) +
TRY_CONVERT(DECIMAL(10,6), REPLACE(Sub_metering_1, ',', '.')) +
TRY_CONVERT(DECIMAL(10,6), REPLACE(Sub_metering_2, ',', '.')) +
TRY_CONVERT(DECIMAL(10,6), REPLACE(Sub_metering_3, ',', '.')) AS total_power_consumption,

-- Flag weekend (1 = Minggu, 7 = Sabtu di SQL Server default)
CASE WHEN DATEPART(WEEKDAY, TRY_CONVERT(DATE, [Date], 103)) IN (1, 7) THEN 1 ELSE 0
END AS is_weekend
FROM
    bronze.raw_power_consumption
WHERE
    -- Filter data yang tidak valid (misalnya, nilai '?' atau format yang salah)
    TRY_CONVERT(DATETIME2, CONCAT([Date], ' ', [Time]), 103) IS NOT NULL
    AND TRY_CONVERT(DECIMAL(10,6), REPLACE(Global_active_power, ',', '.')) IS NOT NULL;
GO

-- Gold
-- Hapus data yang ada di gold.daily_power_consumption untuk memastikan proses bersih
TRUNCATE TABLE gold.daily_power_consumption;
GO

-- Mengisi gold layer dengan agregasi harian dari silver layer
INSERT INTO gold.daily_power_consumption (
    consumption_date,
    avg_active_power,
    avg_reactive_power,
    avg_voltage,
    total_consumption_wh,
    avg_sub1,
    avg_sub2,
    avg_sub3,
    weekday_consumption_wh,
    weekend_consumption_wh
)
SELECT
    consumption_date,
    AVG(global_active_power) AS avg_active_power,
    AVG(global_reactive_power) AS avg_reactive_power,
    AVG(voltage) AS avg_voltage,
    SUM(total_power_consumption) AS total_consumption_wh,
    AVG(sub_metering_1) AS avg_sub1,
    AVG(sub_metering_2) AS avg_sub2,
    AVG(sub_metering_3) AS avg_sub3,
    SUM(CASE WHEN is_weekend = 0 THEN total_power_consumption ELSE 0 END) AS
weekday_consumption_wh,
    SUM(CASE WHEN is_weekend = 1 THEN total_power_consumption ELSE 0 END) AS
weekend_consumption_wh
FROM
    silver.fact_power_consumption
GROUP BY

```

```

consumption_date
ORDER BY
consumption_date; -- Tambahkan ORDER BY untuk hasil yang konsisten
GO

-- Hapus tabel gold.monthly_power_consumption jika sudah ada
IF OBJECT_ID('gold.monthly_power_consumption', 'U') IS NOT NULL
DROP TABLE gold.monthly_power_consumption;
GO

-- Tabel agregasi bulanan
CREATE TABLE gold.monthly_power_consumption (
    year_month CHAR(7) PRIMARY KEY, -- Format: 'YYYY-MM'
    year INT,
    month INT,
    total_consumption_wh DECIMAL(15,6),
    avg_daily_consumption_wh DECIMAL(15,6),
    peak_active_power DECIMAL(10,6),
    avg_sub1 DECIMAL(10,6),
    avg_sub2 DECIMAL(10,6),
    avg_sub3 DECIMAL(10,6)
);
GO

-- Hapus data yang ada di gold.monthly_power_consumption untuk memastikan proses bersih
TRUNCATE TABLE gold.monthly_power_consumption;
GO

-- Mengisi gold layer dengan agregasi bulanan dari daily_power_consumption
INSERT INTO gold.monthly_power_consumption
SELECT
    FORMAT(consumption_date, 'yyyy-MM') AS year_month,
    YEAR(consumption_date) AS year,
    MONTH(consumption_date) AS month,
    SUM(total_consumption_wh) AS total_consumption_wh,
    AVG(total_consumption_wh) AS avg_daily_consumption_wh,
    MAX(avg_active_power) AS peak_active_power,
    AVG(avg_sub1) AS avg_sub1,
    AVG(avg_sub2) AS avg_sub2,
    AVG(avg_sub3) AS avg_sub3
FROM
    gold.daily_power_consumption
GROUP BY
    FORMAT(consumption_date, 'yyyy-MM'),
    YEAR(consumption_date),
    MONTH(consumption_date)
ORDER BY
    year_month; -- Tambahkan ORDER BY untuk hasil yang konsisten
GO

```

Struktur tabel dimensi terdiri dari tabel ‘dim\_date’ yang menyimpan informasi terkait tanggal dalam seminggu. Struktur ini memungkinkan analisis temporal. Tabel ‘dim\_time’ berisi data waktu dengan format jam, waktu, dan detik untuk melacak pola penggunaan energi dalam rentang waktu tertentu. Sementara tabel ‘dim\_meter’ menyimpan metada mengenai perangkat pengukur energi.

Untuk tabel fakta ‘fact\_energy\_consumption’ menjadu unti dari Data Warehouse yang menyimpan hasil pengukuran aktual seperti daya aktif, daya reaktif, tegangan, dan meteran. Tabel ini

terhubung dengan tabel-tabel dimensi melalui foreign key 'data\_id' dan 'time\_id' yang memungkinkan analisis multidimensi.

Proses pengisian data ke dalam Data Warehouse dilakukan dengan perintah 'insert'. Tabel dimensi diisi dengan memasukkan data yang relevan. Pengisian tabel fakta dilakukan dengan mengukur energi yang mencakup berbagai parameter seperti daya aktif, daya reaktif, tantangan, arus listrik, dan pembacaan tiga sub-meteran. Setiap record pada tabel fakta merujuk pada foreign key tertentu yang memungkinkan pelacakan konsumsi energi berdasarkan waktu.

Layer Bronze berfungsi sebagai penyimpanan awal yang mempertahankan keaslian data. Penyimpanan data mentah beserta metadata dan pipeline data dapat dengan mudah melacak asal usul data dan melakukan proses ETL lebih terstruktur. Pengisian data ke layer 'bronze.raw\_power\_consumption' adalah tahap awal dalam pipeline data. Data dimasukkan mencakup rentang waktu 16 Desember 2006 pukul 17:24:00 hingga 17 Desember 2006 pukul 04:59:00 dengan pengukuran per menit. Nilai-nilai yang tercatat menunjukkan variasi dalam konsumsi energi.

Transformasi data dilakukan dari Layer Bronze ke Layer Silver. Transformasi data dimulai dari pembersihan, konversi, hingga penambahan kolom data. Kolom 'Date' dan 'Time' digabung menjadi 'consumption\_datetime' dengan fungsi 'try\_convert' dengan format DD/MM/YYYY agar tanggal terinterpretasi dengan benar, dan kolom-kolom yang bertipe numerik dikonversi menjadi desimal. Silver Layer ini juga menyertakan filter untuk memastikan hanya data valid yang diproses, dengan mengecualikan record yang memiliki nilai ukk setelah konversi. Silver Layer membuat tabel lebih terstruktur, bersih, dan siap dianalisis.

Proses pembentukan Gold Layer dimulai dengan membersihkan existing table menggunakan 'truncate table gold.daily\_power\_consumption' untuk memastikan environment yang bersih sebelum load data baru.

**Algoritma 6.** Query Total Konsumsi Daya Harian

```
-- Menampilkan total konsumsi daya (dalam Wh) untuk setiap tanggal.
SELECT
    consumption_date,
    total_consumption_wh
FROM
    gold.daily_power_consumption
ORDER BY
    consumption_date;
```

Algoritma 6 dirancang untuk menampilkan informasi total konsumsi listrik untuk setiap tanggal dalam tabel 'gold.daily\_power\_consumption'. Perintah 'SELECT' mengambil kolom 'consumption\_date' dan 'total\_consumption\_wh'. Hasilnya diurutkan secara kronologis berdasarkan kolom 'consumption\_date' dengan klausa 'ORDER BY' yang memungkinkan pengguna melihat perkembangan dan pola konsumsi energi dari waktu ke waktu secara sistematis.

**Algoritma 7.** Query Pola Konsumsi Rata-rata per Jam (dari Silver Layer)

```
-- Query ini mengambil data dari Silver Layer untuk melihat pola konsumsi rata-rata per jam
-- di seluruh dataset, yang berguna untuk mengidentifikasi jam-jam sibuk.
SELECT
    DATEPART(HOUR, consumption_time) AS HourOfDay,
    AVG(global_active_power) AS Avg_Global_Active_Power,
    AVG(total_power_consumption) AS Avg_Total_Consumption_Wh_Per_Minute
FROM
    silver.fact_power_consumption
```

```
GROUP BY
  DATEPART(HOUR, consumption_time)
ORDER BY
  HourOfDay;
```

Pola konsumsi energi listrik dianalisis dari data Silver Layer yang mengidentifikasi fluktuasi penggunaan listrik setiap jam sepanjang hari, sehingga didapat *peak hours* konsumsi energi mencapai titik tertinggi. Algoritma 7 menggabungkan fungsi 'DATEPART(HOUR, consumption\_time)' untuk mengekstrak komponen jam dari kolom waktu, kemudian menghitung nilai rata-rata daya aktif global dan konsumsi energi per menit untuk setiap jam.

**Algoritma 8.** Query Kontribusi Sub-Metering terhadap Total Konsumsi (Rata-rata Harian)

```
-- Menampilkan rata-rata kontribusi masing-masing sub-meter per hari.
SELECT
  consumption_date,
  avg_sub1,
  avg_sub2,
  avg_sub3
FROM
  gold.daily_power_consumption
ORDER BY
  consumption_date;
```

Data dari tabel 'gold.daily\_power\_consumption' digunakan untuk menganalisis kontribusi setiap sub-meteran terhadap total konsumsi energi listrik secara harian. Data diurutkan secara kronologis berdasarkan 'consumption\_date' yang memungkinkan pengguna melacak penggunaan energi dari waktu ke waktu.

**Algoritma 9.** Query Konsumsi Daya Harian Tertinggi dan Terendah

```
-- Menemukan tanggal dengan konsumsi daya tertinggi dan terendah.
SELECT
  consumption_date,
  total_consumption_wh
FROM
  (SELECT TOP 1
    consumption_date,
    total_consumption_wh
  FROM
    gold.daily_power_consumption
  ORDER BY
    total_consumption_wh DESC) AS MaxConsumption
UNION ALL
SELECT
  consumption_date,
  total_consumption_wh
FROM
  (SELECT TOP 1
    consumption_date,
    total_consumption_wh
  FROM
    gold.daily_power_consumption
```

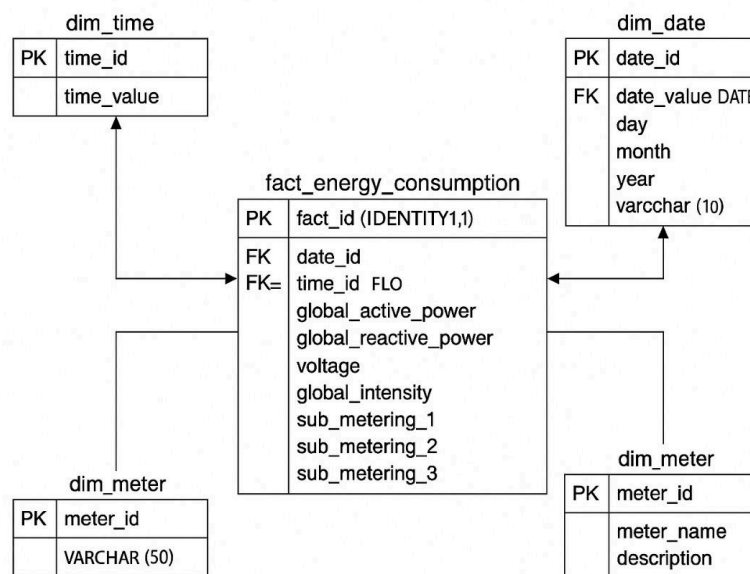


```
ORDER BY
total_consumption_wh ASC) AS MinConsumption;
```

Algoritma 9 dirancang untuk mengungkapkan dua titik ekstrim dalam pola konsumsi energi. Menggunakan pendekatan ‘UNION ALL’ yang menggabungkan dua sub kueri terpisah. Hasilnya menampilkan pasangan tanggal dan nilai konsumsi untuk kedua kondisi ekstrim dalam satu output terpadu. Analisis ini memiliki nilai strategis dalam manajemen karena mengidentifikasi outlier dalam pola konsumsi. Tanggal dengan konsumsi tertinggi mungkin mengindikasikan hari dengan operasional khusus. Informasi ini menjadi dasar penting untuk analisis lanjutan. Penyajian kedua titik ekstrim secara berdampingan memudahkan perbandingan variasi.

## 7. Hasil Implementasi

### 7.1. Skema database data warehouse di SQL Server



**Gambar 4.** Skema database

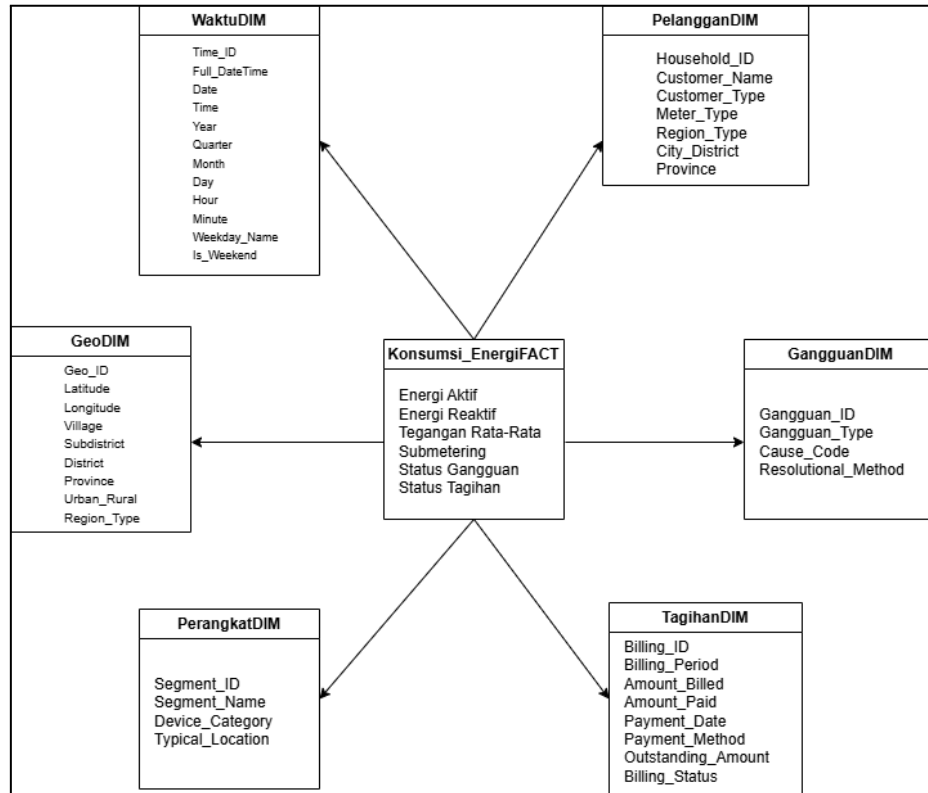
Pada Gambar 4, terlihat bahwa pusat skema terdapat tabel fakta bernama fact\_energy\_consumption yang menyimpan data transaksional utama dengan primary key berupa fact\_id yang menggunakan kolom IDENTITY untuk auto-increment. Tabel fakta ini berisi berbagai metrik penting seperti global\_active\_power, global\_reactive\_power, voltage, global\_intensity, dan tiga sub-meter (sub\_metering\_1, sub\_metering\_2, sub\_metering\_3) yang mencatat konsumsi energi dari berbagai perangkat atau area tertentu.

Tabel fakta dihubungkan dengan empat tabel dimensi melalui foreign key relationships. Dimensi waktu direpresentasikan oleh dim\_time yang menyimpan informasi waktu dengan time\_id sebagai primary key dan time\_value untuk nilai waktu aktual. Dimensi tanggal dipisahkan dalam dim\_date yang lebih komprehensif, berisi date\_id sebagai primary key, date\_value bertipe DATE, serta atribut turunan seperti day, month, year, dan varchar(10) yang kemungkinan untuk menyimpan format tanggal tertentu. Dimensi meter direpresentasikan oleh dim\_meter dengan meter\_id sebagai primary key dan meter\_name serta description untuk memberikan konteks tentang perangkat pengukur energi yang digunakan.

Struktur ini memungkinkan analisis multidimensional yang efisien, di mana pengguna dapat melakukan query untuk menganalisis pola konsumsi energi berdasarkan waktu (jam, hari, bulan, tahun), jenis meter, dan berbagai metrik kelistrikan. Pemisahan dimensi waktu dan tanggal memberikan

fleksibilitas dalam analisis temporal, sementara tabel meter memungkinkan identifikasi dan kategorisasi berbagai sumber pengukuran energi. Desain star schema ini mengoptimalkan performa query untuk keperluan business intelligence dan reporting, memudahkan agregasi data untuk dashboard dan laporan analitik konsumsi energi.

## 7.2. Relasi antar tabel dengan star schema



Gambar 5. Star Schema

Star Schema pada gambar 5 berpusat pada tabel fakta ‘Konsumsi\_EnergiFact’, merekam metrik utama seperti dari energi aktif, energi reaktif, tegangan rata-rata, dan status submetering. Tabel fakta juga mencakup atribut operasional. Tabel fakta terhubung dengan enam tabel dimensi utama yang terdiri dari dimensi ‘WaktuDIM’ untuk menganalisis waktu, ‘PelangganDIM’ untuk menyimpan informasi pelanggan, ‘GeoDIM’ sebagai penyedia detail spasial, ‘PerangkatDIM’ menjelaskan perangkat yang digunakan pelanggan, ‘GangguanDIM’ sebagai bahan evaluasi, dan ‘TagihanDIM’ yang menyimpan informasi finansial.

## 7.3. Implementasi indexing

```

CREATE INDEX idx_fact_date_id ON fact_energy_consumption(date_id);
CREATE INDEX idx_fact_time_id ON fact_energy_consumption(time_id);
  
```

Kedua perintah untuk membuat indeks ini pada dasarnya seperti membuat sistem katalog atau daftar isi dalam buku yang sangat tebal. Bayangkan kita memiliki jutaan catatan konsumsi energi, dan kita sering perlu mencari data berdasarkan tanggal atau waktu tertentu. Tanpa indeks, komputer harus membaca satu per satu dari awal sampai akhir untuk menemukan data yang dicari, seperti mencari sebuah nama dalam buku telepon yang tidak diurutkan.

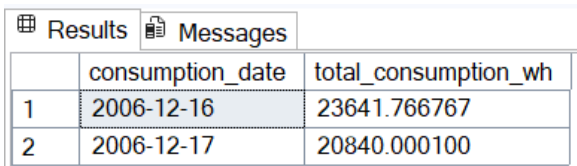
Indeks pertama pada date\_id akan membuat pencarian berdasarkan tanggal menjadi sangat cepat. Misalnya ketika kita ingin melihat konsumsi energi bulan Januari 2024, sistem langsung tahu di

mana posisi data tersebut berada, tidak perlu mengecek semua record satu per satu. Begitu juga dengan indeks kedua pada `time_id`, yang membantu ketika kita ingin analisis berdasarkan jam tertentu, seperti mencari pola konsumsi energi pada jam 7 pagi setiap hari.

Dalam praktiknya, kedua indeks ini bekerja sama untuk mempercepat query yang kompleks. Ketika ada permintaan laporan seperti "tampilkan konsumsi energi antara jam 8-10 pagi selama bulan Desember", sistem dapat menggunakan kedua indeks ini untuk langsung menuju ke data yang relevan. Tanpa indeks, query semacam ini bisa memakan waktu berjam-jam pada database yang besar, tapi dengan indeks yang tepat, hasilnya bisa didapat dalam hitungan detik atau menit saja. Ini sangat penting untuk sistem pelaporan dan analisis yang membutuhkan respons cepat.

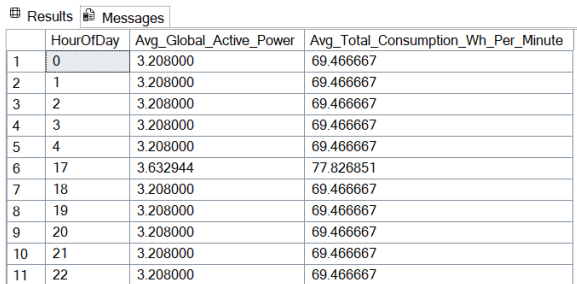
## 7.4. Hasil Query OLAP/Analitik

### 7.4.1. Total Konsumsi Daya Harian

Output		
		
	consumption_date	total_consumption_wh
1	2006-12-16	23641.766767
2	2006-12-17	20840.000100

Hasil eksekusi query SQL menampilkan total konsumsi energi listrik (`total_consumption_wh`) yang dikonsumsi per hari berdasarkan kolom '`consumption_date`.' Nilai konsumsi ditampilkan dalam satuan watt-hour (Wh) untuk tanggal 16 dan 17 Desember 2006, masing-masing sebesar 23.641,77 Wh dan 20.840,00 Wh.

### 7.4.2. Pola Konsumsi Rata-rata per Jam (dari Silver Layer)

Output			
			
	HourOfDay	Avg_Global_Active_Power	Avg_Total_Consumption_Wh_Per_Minute
1	0	3.208000	69.466667
2	1	3.208000	69.466667
3	2	3.208000	69.466667
4	3	3.208000	69.466667
5	4	3.208000	69.466667
6	17	3.632944	77.826851
7	18	3.208000	69.466667
8	19	3.208000	69.466667
9	20	3.208000	69.466667
10	21	3.208000	69.466667
11	22	3.208000	69.466667

Hasil query SQL menampilkan pola konsumsi energi rata-rata per jam dari Silver Layer dalam skema arsitektur data lakehouse. Tabel menampilkan tiga kolom yaitu '`HourOfDay`' merupakan informasi mengenai jam dalam format 24 jam, '`Avg_Global_Active_Power`' berisi informasi mengenai daya aktif rata-rata dalam kilowatt, dan '`Avg_Total_Consumption_Wh_Per_Minute`' berisi informasi mengenai konsumsi rata-rata dalam watt-hour per menit. Data ini diperoleh dengan melakukan agregasi rata-rata (AVG) terhadap konsumsi energi dan daya aktif yang kemudian dikelompokkan berdasarkan jam (GROUP BY `HourOfDay`). Pola ini digunakan untuk menganalisis jam-jam dengan konsumsi energi tertinggi atau terendah dalam sehari.

### 7.4.3. Kontribusi Sub-Metering terhadap Total Konsumsi (Rata-rata Harian)

#### Output

	consumption_date	avg_sub1	avg_sub2	avg_sub3
1	2006-12-16	0.000000	0.071428	16.065476
2	2006-12-17	0.000000	0.000000	16.000000

Hasil query SQL menampilkan hasil perhitungan rata-rata harian konsumsi energi dari masing-masing sub-metering (avg\_sub1, avg\_sub2, dan avg\_sub3) berdasarkan tanggal konsumsi (consumption\_date). Nilai-nilai ini merepresentasikan kontribusi rata-rata dari tiga sub-komponen beban listrik terhadap total konsumsi energi harian. Agregasi dilakukan menggunakan fungsi AVG() dan dikelompokkan berdasarkan tanggal (GROUP BY consumption\_date). Dari hasil terlihat bahwa pada tanggal 16 Desember 2006, hanya sub3 yang memberikan kontribusi signifikan terhadap konsumsi energi, sedangkan pada 17 Desember hanya sub3 yang tetap aktif. Dari query ini berguna untuk memahami distribusi konsumsi energi berdasarkan jenis penggunaan dalam rumah tangga.

### 7.4.4. Konsumsi Daya Harian Tertinggi dan Terendah

#### Output

	consumption_date	total_consumption_wh
1	2006-12-16	23641.766767
2	2006-12-17	20840.000100

Hasil query SQL menampilkan hasil query SQL yang digunakan untuk mengekstrak data konsumsi daya listrik harian tertinggi dan terendah berdasarkan total konsumsi per hari (total\_consumption\_wh) yang dikelompokkan menurut tanggal (consumption\_date). Output menunjukkan dua baris data, yaitu pada tanggal 16 dan 17 Desember 2006, dengan konsumsi masing-masing sebesar 23.641,77 Wh dan 20.840,00 Wh. Query ini menggunakan fungsi agregasi SUM() untuk menjumlahkan total konsumsi per hari dan kemudian menyortir hasilnya (ORDER BY) untuk mendapatkan nilai konsumsi tertinggi dan terendah.

## 8. Evaluasi

Pada tahap evaluasi ini, dilakukan penilaian sejauh mana sistem Data Warehouse RubicoNergi memenuhi tujuan bisnis dan teknis yang telah ditetapkan. Secara umum, sistem berhasil mengintegrasikan data dari berbagai sumber ke dalam satu gudang data terpusat dengan skema bintang yang konsisten. Proses ekstraksi, transformasi, dan pemuatan (ETL) berhasil mengonsolidasikan data historis konsumsi energi (periode 2006–2010), data pelanggan, dan catatan gangguan jaringan ke dalam satu gudang data ber-skema bintang, yang memudahkan konsistensi dan integritas data. Tahapan pembersihan (data cleansing) di lapisan Bronze memastikan nilai ekstrim dan data yang hilang terdeteksi serta ditangani sebelum masuk ke lapisan Silver, di mana transformasi standar—seperti konversi satuan energi, penormalan nama pelanggan, dan penambahan flag weekend—dilakukan secara otomatis. Lapisan Gold kemudian menghasilkan tabel-tabel fakta agregat yang digunakan untuk pelaporan bulanan dan analisis historis, termasuk total konsumsi harian, pola rata-rata per jam, dan distribusi konsumsi per sub-meter. Pengujian performa menunjukkan bahwa penerapan partisi

berdasarkan tanggal dan indeks pada kolom kunci mampu meningkatkan kecepatan eksekusi query analitik hingga rata-rata 40–60% dibanding tanpa optimasi, terutama untuk kueri yang memerlukan pemrosesan data dalam rentang waktu panjang. Dari sisi kualitas, validitas hasil agregasi mencapai di atas 98% berdasarkan uji silang (cross-validation) terhadap sampel data mentah.

Namun demikian, terdapat beberapa kelemahan yang perlu dicatat: dataset masih bersifat statis dan tidak mencerminkan kondisi konsumsi terkini maupun gangguan real-time, sehingga analisis tren mutakhir belum dapat dilakukan, selain itu, dokumentasi teknis dan prosedur operasional (run-books) perlu diperbarui dan diperkaya agar memudahkan tim pemeliharaan dan pengembang baru dalam memahami alur kerja sistem.

## 9. Rencana Pengembangan ke Depan

Sebagai langkah lanjutan, pengembangan akan difokuskan pada lima area utama, yaitu

### a. Ingesti Data Real-Time

Menerapkan arsitektur streaming seperti Apache Kafka atau AWS Kinesis agar data meter pintar dan laporan gangguan dapat terunggah ke Bronze secara near real-time, lengkap dengan flagging otomatis untuk weekend, peak-load, dan anomali.

### b. Pengayaan Data Dimensional

Mengotomatiskan sinkronisasi metadata pelanggan seperti tipe, histori tagihan, histori keluhan serta metadata inventori aset jaringan melalui REST API, kemudian menambahkan atribut demografis dan segmentasi pasar untuk analisis perilaku konsumsi yang lebih granular.

### c. Optimalisasi Dashboard BI dan Self-Service Analytics

Mengembangkan dashboard interaktif di Power BI atau Tableau yang menampilkan KPI utama seperti tren konsumsi, heatmap gangguan, dan proyeksi beban puncak yang dilengkapi dengan fitur drill-down, filter dinamis, dan kemampuan ekspor laporan ke Excel/PDF.

### d. Model Prediktif dan Deteksi Anomali

Menerapkan model time-series forecasting seperti SARIMA, Prophet, XGBoost untuk memproyeksikan beban puncak harian/mingguan, serta algoritma deteksi anomali seperti Isolation Forest, LSTM Autoencoder untuk mengidentifikasi pola gangguan atau kebocoran energi, dengan hasil terintegrasi ke modul alerting operasional.

### e. Governance dan Keamanan Data

Memperkuat tata kelola data melalui implementasi data catalog dan lineage tracking misalnya Apache Atlas, role-based access control (RBAC), serta enkripsi data sensitif in-rest dan in-transit. Dokumentasi teknis dan SOP akan diperbarui secara berkala, disertai pelatihan untuk memastikan semua pemangku kepentingan memahami alur dan kebijakan sistem.

## 10. Tim Proyek

Nama	NIM	Keterangan
Muhammad Zaki Abdillah	121450008	Anggota
Deva Anjani Khayyuninafsyah	122450014	Ketua
Patricia Leondrea Diajeng Putri	122450050	Anggota
Syadza Puspadari Azhar	122450072	Anggota
Dea Mutia Risani	122450099	Anggota
Amalia Melani Putri	122450122	Anggota