

# **LAPORAN MISI KEDUA 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**

## 1. Ringkasan Kebutuhan dari Misi

Berdasarkan studi kasus perusahaan RubicoNergi (RBN), diperlukan suatu gudang data yang mampu mengintegrasikan data konsumsi listrik, data spasial, hingga data pengguna sehingga digunakan suatu *business requirements specification based on data-driven*.

### 1.1. Identifikasi Sistem Sumber

Sistem ini mengintegrasikan berbagai sumber data untuk mendukung operasional dan analisis, yakni:

- Data Konsumsi Energi diperoleh dari dataset konsumsi listrik per menit (Kaggle) atau sistem meter pelanggan.
- Data Geospasial (informasi lokasi pelanggan, jaringan distribusi, dan wilayah administratif) yang bersumber dari sistem GIS.
- Data Pelanggan diambil dari sistem CRM/*Billing*.
- Data Gangguan berasal dari sistem pelaporan gangguan (*trouble ticket system*).
- Data Tagihan dari sistem tagihan listrik dan pembayaran.
- Data petugas Lapangan yang mencakup operasional dan pengawasan teknisi di lapangan.

### 1.2. Penerapan Proses Derivasi

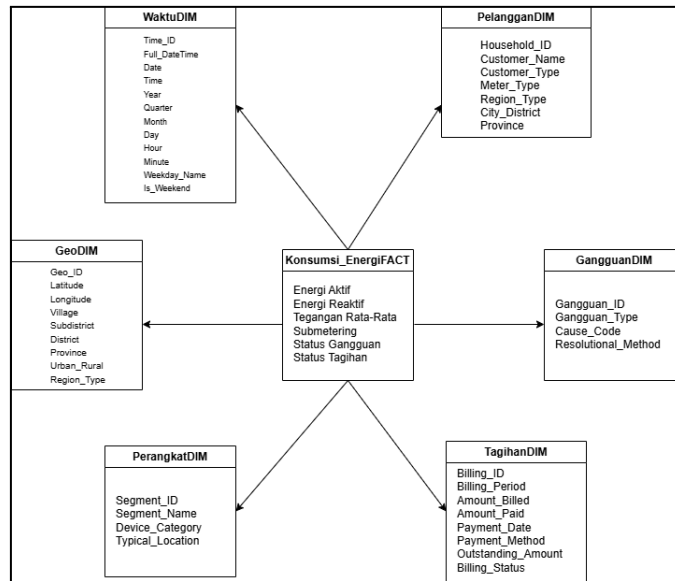
Fakta utama yang digunakan adalah Fakta Konsumsi Energi. Proses derivasi melibatkan beberapa komponen kunci, antara lain: Energi Aktif (kWh) yang diperoleh dari konversi *Global\_active\_power* ke energi per hari, Energi Reaktif yang bersumber dari *Global\_reactive\_power*, serta Tegangan (Voltage) yang dihitung berdasarkan rerata harian. Selain itu, terdapat Sub-metering 1-3 yang merepresentasikan area konsumsi spesifik dalam rumah tangga. Sistem ini juga mengintegrasikan Status Gangguan dari data jumlah gangguan dan waktu pemulihan, serta Status Tagihan yang mencakup kategori lunas, tertunda, atau menunggak. Dengan demikian, proses derivasi ini memungkinkan analisis yang lebih mendalam terhadap konsumsi energi dan faktor-faktor terkait.

### 1.3. Ringkasan Hasil Proses Derivasi

Tabel 1. Ringkasan Proses Derivasi

Fakta	Measures	Dimension & Cardinalities	Hierarchies & Levels
Konsumsi Energi	Energi Aktif, Energi Reaktif, Tegangan Rata-Rata, Sub-metering, Status Gangguan, Status Tagihan	WaktuDIM (1:n)	Menit → jam → hari → bulan → kuartal → tahun
		PelangganDIM (1:n)	Pelanggan → kab/kota → provinsi → region type
		GeoDIM (1:n)	Koordinat → desa → kecamatan → kab/kota → provinsi
		PerangkatDIM (1:n)	Submeter → fungsi/ruang
		GangguanDIM (1:n)	Jenis gangguan → penyebab → penanganan
		TagihanDIM (1:n)	Periode bulanan → status pembayaran per pelanggan

## 2. Skema Konseptual Multidimensi



Gambar 1. *Star Schema*

*Star Schema* pada gambar 1 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.

## 3. Penjelasan Setiap Komponen

### 3.1. 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.

### 3.2. Dimensi PelangganDIM

Dimensi ini, Household\_ID digunakan untuk mengidentifikasi setiap pelanggan. Atribut-attribut 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.

### 3.3. 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.

#### **3.4. 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.

#### **3.5. 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.

#### **3.6. 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.

### **4. Justifikasi Desain Konseptual**

#### **4.1. 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.

#### **4.2. Dimensi Waktu**

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

#### **4.3. Dimensi Pelanggan**

Dimensi pelanggan mendukung pembuatan kebijakan yang strategis dan tertarget.

#### **4.4. 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.

#### **4.5. Dimensi Jaringan**

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

#### 4.6. Dimensi Gangguan

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

#### 4.7. Dimensi Tagihan

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

### 5. 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:

#### 5.1. Kesesuaian dengan Tabel Fakta

Tabel 2. 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

#### 5.2. Kesesuaian dengan Tabel Dimensi

Tabel 3. 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