

Laporan Perancangan Data Warehouse pada Industri Lingkungan & Sustainability



Disusun Oleh:

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

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

1. Profil Industri & Masalah Bisnis :

Sebagai CEO perusahaan penghasil listrik berbasis energi terbarukan, saya bertanggung jawab mengoptimalkan operasi bisnis yang berorientasi pada keuntungan dan keberlanjutan. Perusahaan ini terdiri dari divisi kunci seperti Sumber Daya Manusia (SDM), Produksi, dan Finansial, yang masing-masing menghasilkan data heterogen seperti kinerja karyawan HR analytics, efisiensi mesin produksi IoT sensor data, dan laporan keuangan triwulan. Untuk menjawab kebutuhan spesifik tiap departemen, kami mengadopsi pendekatan bottom-up dengan membangun Data Mart terpisah: Data Mart Produksi menganalisis inefisiensi rantai pasok, Data Mart SDM mengevaluasi produktivitas karyawan, dan Data Mart Pemasaran merancang kampanye pelanggan berbasis perilaku konsumen. Data Mart ini dirancang dengan model dimensional skema bintang untuk kemudahan akses pengguna non-teknis, lalu diintegrasikan secara bertahap ke dalam Data Warehouse guna analisis lintas divisi, seperti mengidentifikasi korelasi antara pelatihan karyawan dan efisiensi produksi. Melalui pendekatan ini, perusahaan berhasil mengurangi biaya produksi serta meningkatkan inovasi karyawan. Efektivitas kampanye pelanggan juga meningkat, sembari membangun fondasi data terpadu untuk strategi jangka panjang tanpa mengorbankan kecepatan solusi masalah spesifik.

2. Daftar Stakeholder & Tujuan Bisnis (tabel):

Tabel 1. Stakeholder dan tujuan bisnis

Daftar Stakeholder	Tujuan Bisnis
CEO	Memastikan efisiensi operasional dan keberlanjutan perusahaan melalui keputusan berbasis data
Divisi Produksi	Meningkatkan efisiensi pembangkit energi terbarukan dan mengurangi downtime melalui analisis data sensor
Divisi Sumber Daya Manusia	Meningkatkan produktivitas dan retensi karyawan melalui pemantauan performa dan efektivitas pelatihan
Divisi Pemasaran	Merancang kampanye yang lebih efektif dan tepat sasaran melalui analisis perilaku pelanggan
Divisi Keuangan	Mengoptimalkan investasi dan operasional dengan menilai efisiensi biaya dan ROI pembangkit energi
Tim Data Analyst	Menyediakan laporan dan dashboard lintas divisi untuk mendukung pengambilan keputusan strategis
Regulator	Memantau dan menilai kepatuhan perusahaan terhadap kebijakan energi bersih dan keberlanjutan

investor	Mengevaluasi profitabilitas dan kelayakan proyek energi terbarukan untuk pengambilan keputusan investasi
Pelanggan	Mendapatkan layanan energi bersih yang andal dengan partisipasi dalam program keberlanjutan

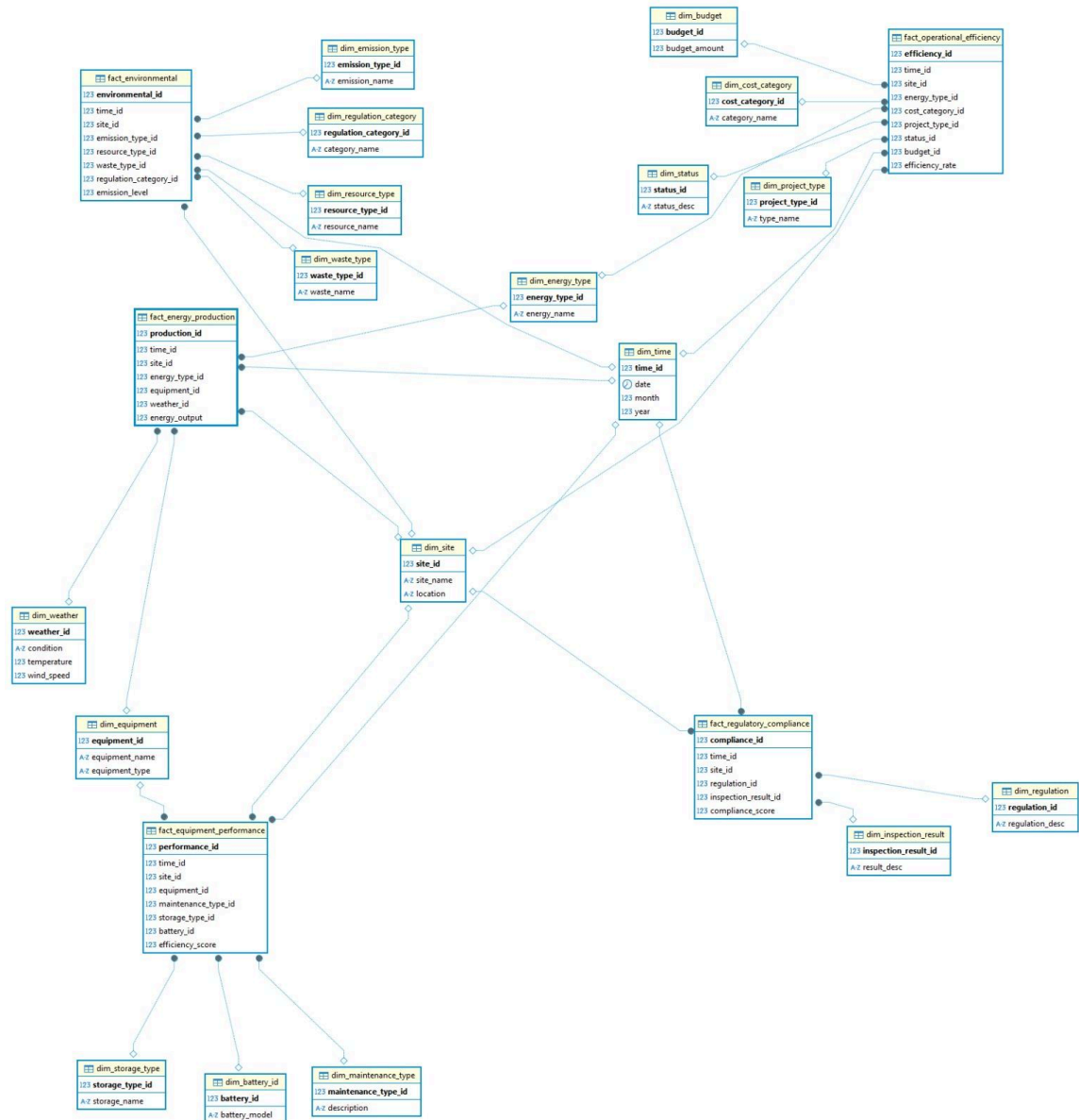
3. Fakta & Dimensi (diagram + penjelasan):

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

Tabel 2. Fakta dan dimensi

Kebutuhan	Fakta	Dimensi
Analisis produksi, konsumsi dan penyimpanan energi	Energy_Production	Time, Site, Energy_Type, Equipment, Weather
Evaluasi efisiensi peralatan, downtime, output energi, dan penyimpanan energi	Equipment_Performance	Time, Site, Equipment, Maintenance_Type, Storage_Type, Battery_ID
Monitoring biaya operasional dan efisiensi proyek peningkatan keberlanjutan	Operational_Efficiency	Time, Site, Energy_Type, Cost_Category, Project_Type, Status, Budget
Audit emisi karbon, penggunaan air, dan limbah dari operasional	Environmental	Time, Site, Emission_Type, Resource_Type, Waste_Type, Regulation_Category
Evaluasi kepatuhan terhadap regulasi lingkungan dan standar energi hijau	Regulatory_Compliance	Time, Site, Regulation, Inspection_Result

Dari hasil identifikasi fakta dan dimensi, selanjutnya dirancang desain konseptual dalam bentuk diagram ERD untuk memberikan gambaran hubungan antar entitas dalam skema data warehouse yang akan dibangun.



Gambar 1. Diagram ERD

Terdapat lima tabel fakta yang merepresentasikan berbagai aktivitas penting perusahaan, yaitu fact_energy_production, fact_equipment_performance, fact_operational_efficiency, fact_environmental, dan fact_regulatory_compliance. masing-masing mewakili aspek penting operasional. Setiap tabel fakta terhubung dengan sejumlah tabel dimensi melalui foreign key, yang memungkinkan analisis data dari berbagai sudut pandang. Misalnya, tabel fact_energy_production mencatat data kuantitatif terkait produksi energi, dan dianalisis berdasarkan waktu (dim_time), lokasi (dim_site), jenis energi (dim_energy_type), peralatan yang digunakan (dim_equipement), serta kondisi cuaca saat itu (dim_weather). Demikian pula, tabel fact_equipment_performance merekam performa

alat berdasarkan dimensi seperti jenis perawatan, tipe penyimpanan, dan baterai yang digunakan.

Hubungan serupa juga terlihat pada fact_operational_efficiency, yang menyimpan data efisiensi proyek berdasarkan jenis energi, kategori biaya, tipe proyek, status, serta alokasi anggaran. Tabel fact_environmental berfungsi untuk mencatat emisi, penggunaan sumber daya, dan limbah, serta keterkaitannya dengan regulasi lingkungan. Terakhir, fact_regulatory_compliance digunakan untuk memonitor kepatuhan terhadap regulasi melalui data inspeksi.

4. Sumber Data & Metadata (deskripsi + contoh data):

No	Sumber Data	Deskripsi	Frekuensi Update	Contoh Data	Link
1	Open Power System Data (OPSD)	Data teknis pembangkit energi terbarukan di Eropa.	Triwulanan	Kapasitas: 150 MW; Tgl: 2020-01-15; Teknologi: PV; Kode: DE1; Tegangan: 110 kV.	Link
2	Kaggle – HR Analytics Dataset	Riwayat kerja dan performa karyawan.	Triwulanan (internal SDM)	Kepuasan: 0.85; Proyek: 5; Promosi: 1; Jam: 160/bulan.	Link
3	Kaggle – Customer Personality Analysis	Preferensi dan demografi pelanggan.	Mingguan	Pendapatan: \$50,000; Respons: 3; Usia: 35; Belanja: 60.	Link
4	U.S. Energy Information Administration (EIA)	Laporan keuangan perusahaan energi global.	Triwulanan	Pendapatan: \$1,000,000; Laba: \$200,000; Arus Kas: \$150,000.	Link
5	Kaggle – Machine Sensor Data	Pembacaan sensor mesin untuk maintenance prediktif.	Harian (batch/streaming)	Waktu: 2024-04-15; ID: 101; Sensor1: 0.5; Sensor2: 75 psi; Kegagalan: Overheating.	Link

Semua data akan diproses melalui tahapan ETL menggunakan Python (Pandas, PyArrow) dan diotomasi dengan Apache Airflow. Pembersihan data mencakup imputasi (mean/median/mode), normalisasi (unit dan waktu), serta deteksi outlier (Z-score dan IQR). Seluruh data dikonversi ke format Parquet dan disusun dalam skema dimensional untuk mendukung integrasi dengan Power BI dan Tableau.