



ETL Pipeline

Smart Meter System

NAVIGATION

Overview >

ETL Architecture

Business Rules

Record Lifecycle

ETL Pipeline Dashboard

Smart Meter Data Processing System - 50,000 Meters

Active Meters

50,000

Connected devices



Records/Hour

2.4M

Processing rate



Success Rate

99.7%

Validation pass



Alerts

23

Faulty meters detected



Pipeline Overview

Raw Data Ingestion

Active

Transform Processing

Running

Structured Storage

Synced

Parquet Archive

Updated

Edit with Lovable ×

System Status

- 50,000 Meters Active

Transformation Rules

- Unit Standardization (W to kW)
- Missing Value Handling
- Data Validation Rules
- Faulty Meter Detection

10 active business rules applied to every record

System Features

Task A

ETL Architecture Diagram
- Visual flow from raw data to analytics archive

Task B

Business Rules - Unit conversion, validation, and faulty meter detection

Task C

Record Lifecycle - Complete journey of a single smart meter reading



NAVIGATION

Overview

ETL Architecture >

Business Rules

Record Lifecycle

Task A: ETL Architecture Diagram

Conceptual System Design for Smart Meter Data Pipeline



Smart Meters (50,000)

IoT devices collecting real-time energy consumption data from residential and commercial properties



Raw Data Store

Raw CSV files stored exactly as received, preserving original data integrity



ETL Orchestrator

Serverless workflow triggered by file arrival events, coordinating transformation steps



Transform Step

Unit conversion (W to kW), missing data handling, validation checks, faulty meter detection



Structured Storage (DB)

Edit with Lovable



Clean tabular records ready for SQL queries, BI dashboards, and validation audits



Analytics Archive Layer

Parquet columnar format for compressed, fast analytics and long-term storage

❗ Error Handling Path

⟳ Retry Logic

Automatic retry up to 3 attempts on transformation failure

✖ Dead-Letter Storage

Failed records stored for manual review with error logs

Flow Legend

✓ Success Path

✖ Failure Path

⟳ Retry on Fail

Key Features

- Event-driven triggers

Edit with **Lovable**

System Status

● 50,000 Meters Active

- Serverless architecture
- Automatic scaling
- Data validation
- Error recovery



NAVIGATION

Overview

ETL Architecture

Business Rules >

Record Lifecycle

Task B: Transformation Logic & Business Rules

Complete set of rules for data processing and validation



Unit Standardization



Missing Values



Data Validation



Faulty Detection

Unit Standardization Rules



Rule 1 Convert Watts to Kilowatts

IF unit = "W"

→ value_kW = value_W / 1000, unit = 'kW'



Rule 2 Validate Numeric Fields

IF energy value cannot be parsed as number

→ Flag record as invalid_numeric

Missing Value Handling Rules



Rule 3 Missing Reading

IF reading_value = NULL

Flag as missing_reading, exclude from analysis



Rule 4 Short Gap Interpolation

IF missing data for 2 or less consecutive timestamps

→ Apply linear interpolation



Rule 5 Large Gaps (Wi-Fi Outage)

IF 3 or more missing timestamps in a row

→ Do not interpolate, mark as data_gap

Data Validation Rules

Edit with Lovable ×



Rule 6 Timestamp Validation

IF timestamp is out of order, duplicate, or out of range

→ Mark as invalid_timestamp



Rule 7 Range Validation

IF reading_kW < 0 OR reading_kW > technical_limit_max

→ Mark as out_of_range

⚠ Faulty Meter Detection Rules



Rule 8 Zero Consumption

IF meter reports 0 kW for 24+ consecutive hours

→ Flag as potential_faulty_meter



Rule 9 Unrealistic Spikes

IF consumption jumps more than 10x previous reading

→ Mark as anomalous_spike



Rule 10 Flatline Readings

IF exact same value for 100+ consecutive readings

→ Flag for inspection (flatline_behavior)

System Status

● 50,000 Meters Active



NAVIGATION

Overview

ETL Architecture

Business Rules

Record Lifecycle >

Task C: Single Record Lifecycle

Complete journey of one smart meter reading through the system

Sample Record

MeterID: 4812

Timestamp: 2025-01-22 14:00

Value: 450

Unit: W



Upload to Raw Storage

Step 1

The smart meter sends a reading that is stored exactly as received inside the Raw CSV Storage. No transformations occur at this stage.

Data preserved in original format for audit trail



Triggering the Transformation

Step 2

The moment the raw file lands, an event trigger notifies the ETL Orchestrator. The orchestrator starts a serverless transformation job that reads only the new file.

Event Trigger

Serverless Job

Incremental Read

Edit with Lovable ×



Data Cleaning & Validation

Step 3

3.1 Unit Standardization

Unit = W, Convert: $450W / 1000 = 0.45 \text{ kW}$

3.2 Missing Value Check

Value is present - No missing data flag

3.3 Validation Checks

Timestamp valid

Value not negative

Within range

No spike detected

3.4 Faulty Meter Rules

No faults detected - Record marked as valid



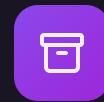
Storage in Structured Format

Step 4

The cleaned record is inserted into a structured table supporting SQL queries, BI dashboards, and validation audits.

Edit with **Lovable**

meter_id	timestamp	reading_kW	status	anomalies
4812	14:00	0.45	valid	none



Step 5

Archival in Parquet Format

At the end of the batch job, all validated records are converted into Parquet format for compression, fast columnar analytics, and long-term storage.

Compression

Columnar Format

Long-term Storage



Step 6

Success or Failure Handling

Success Path

Record logged as successfully processed
Pipeline continues to next record

Failure Path

Retry automatically up to 3 times
If still failing: Write to Dead-Letter Storage
Log error with meter_id, timestamp, error_type
Alert data engineering team

System Status

• 50,000 Meters Active

Edit with **Lovable**