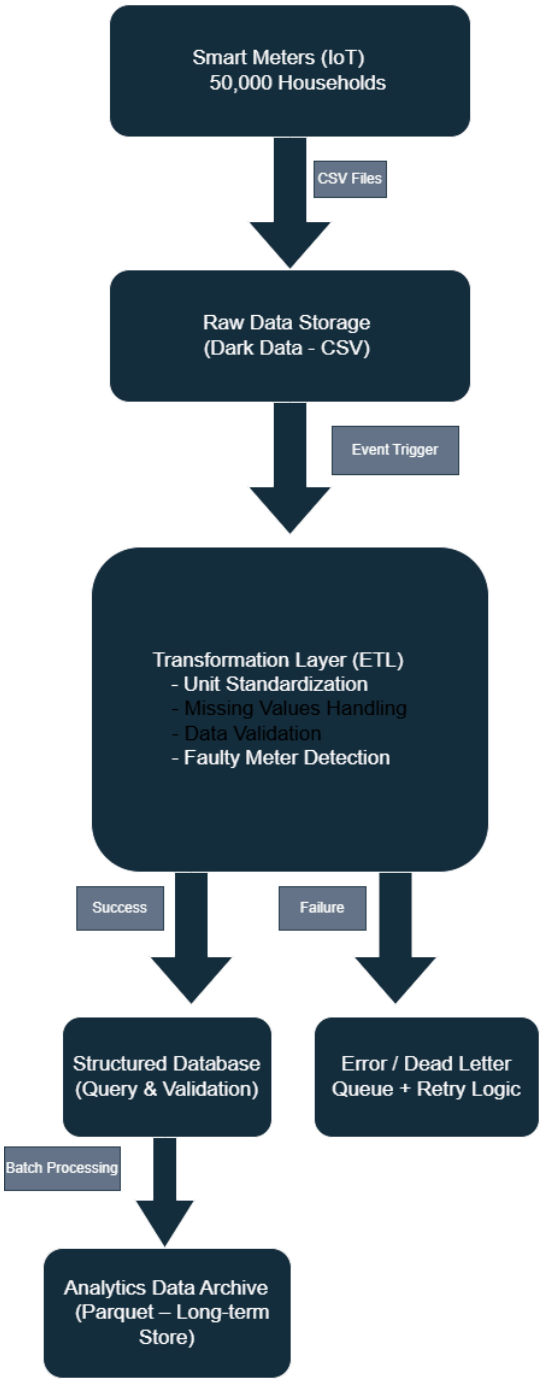
Task A: ETL Architecture Diagram (System Design)



Task B: Transformation Logic & Business Rules

This section defines the transformation rules applied during the ETL process to ensure data quality, consistency, and analytics readiness.

Rule 1: Energy Unit Standardization

To maintain consistency across all smart-meter readings:

- If `energy_unit = "W"`, the energy value is divided by 1000 and converted to `kW`.
- If `energy_unit = "kW"`, the value is kept unchanged.

All energy consumption values are stored using a unified measurement unit.

Rule 2: Missing Values Handling

Missing readings may occur due to temporary connectivity issues.

- If `energy_reading` is NULL:
    - The record is flagged as incomplete.
    - The record is excluded from peak consumption analysis.
    - Metadata is stored for potential future interpolation or recovery.

This approach prevents inaccurate analytics while preserving data traceability.

Rule 3: Data Validation

Each record is validated to ensure data integrity:

- Energy consumption values less than zero are rejected.
- Records with invalid or duplicated timestamps are flagged.
- Extreme outliers that significantly deviate from normal consumption patterns are identified for further review.

Invalid or suspicious records do not participate in analytical calculations.

Rule 4: Faulty Smart Meter Detection

Basic anomaly detection logic is applied to identify potentially faulty meters:

- If a smart meter reports zero consumption for an unusually long period.
- Or if it reports identical readings continuously over time.

Such meters are marked as potentially faulty for further inspection.

Task C: Single Record Lifecycle

This section describes how a single smart-meter record moves through the ETL pipeline from ingestion to long-term storage.

Step 1: Upload to Raw Storage

A smart meter sends a new energy reading in CSV format.
The file is stored without modification in the raw data storage layer, where it is considered unprocessed data.

Step 2: Triggering the Transformation Process

The arrival of a new CSV file automatically triggers the ETL pipeline through an event-driven mechanism.

Step 3: Data Cleaning and Validation

Within the transformation layer, the record undergoes:

- Energy unit standardization to kilowatts (kW).
- Detection and flagging of missing values.
- Validation of timestamps and energy consumption values.
- Identification of potential faulty meter behavior.

Step 4: Storage in Structured Database

If the record passes all validation checks, it is stored in a structured relational database.
This storage layer supports querying, auditing, and dashboard generation.

Step 5: Parquet Conversion and Archival

Validated records are processed in batches and converted into Parquet format.
The Parquet files are archived for long-term storage and optimized analytical processing, including future forecasting tasks.

Step 6: Success and Failure Handling

- Successful records continue through the pipeline and are archived.
- Failed records trigger automatic retry mechanisms.
- Records that repeatedly fail processing are moved to an error storage area and logged for investigation.