# **Data Collection & Processing**

## **Task Summary**

Collect and structure a **time-series electricity consumption dataset** for AI-powered energy theft and leak detection .

* Dataset includes meter ID, timestamp (UTC), and cumulative energy consumption.
* Saved as a **CSV file** in /data/.

## **Expected Outputs**

## AI-Powered Energy Theft and Leak Detection System must :

## Identify abnormal energy or pipelines usage, alerting potential thefts or leaks, forecasting or baseline predicting consumptions, visualizing anomalies on dashboards, as well as insights into the data quality and the performance of AI models to support decision making.

* File: data/sample\_consumption\_5meters.csv
* 5 meters, 3 years, hourly readings (~657,000 rows)
* Timestamps in UTC, meter IDs anonymized

## **Processing Steps**

1. Validate schema: check for presence of all required fields
2. Remove duplicates and deal with missing readings.
3. Resample to hourly or daily intervals.
4. Output features: consumption delta, rolling statistics, anomaly scores.
5. Label data using known theft/leak events (if available).

## **Quality Checks**

* Completeness (>95% of expected readings present)
* Unique (meter\_id, timestamp) entries
* Valid value ranges (e.g., voltage, cumulative energy)
* Temporal consistency (cumulative readings should not decrease)