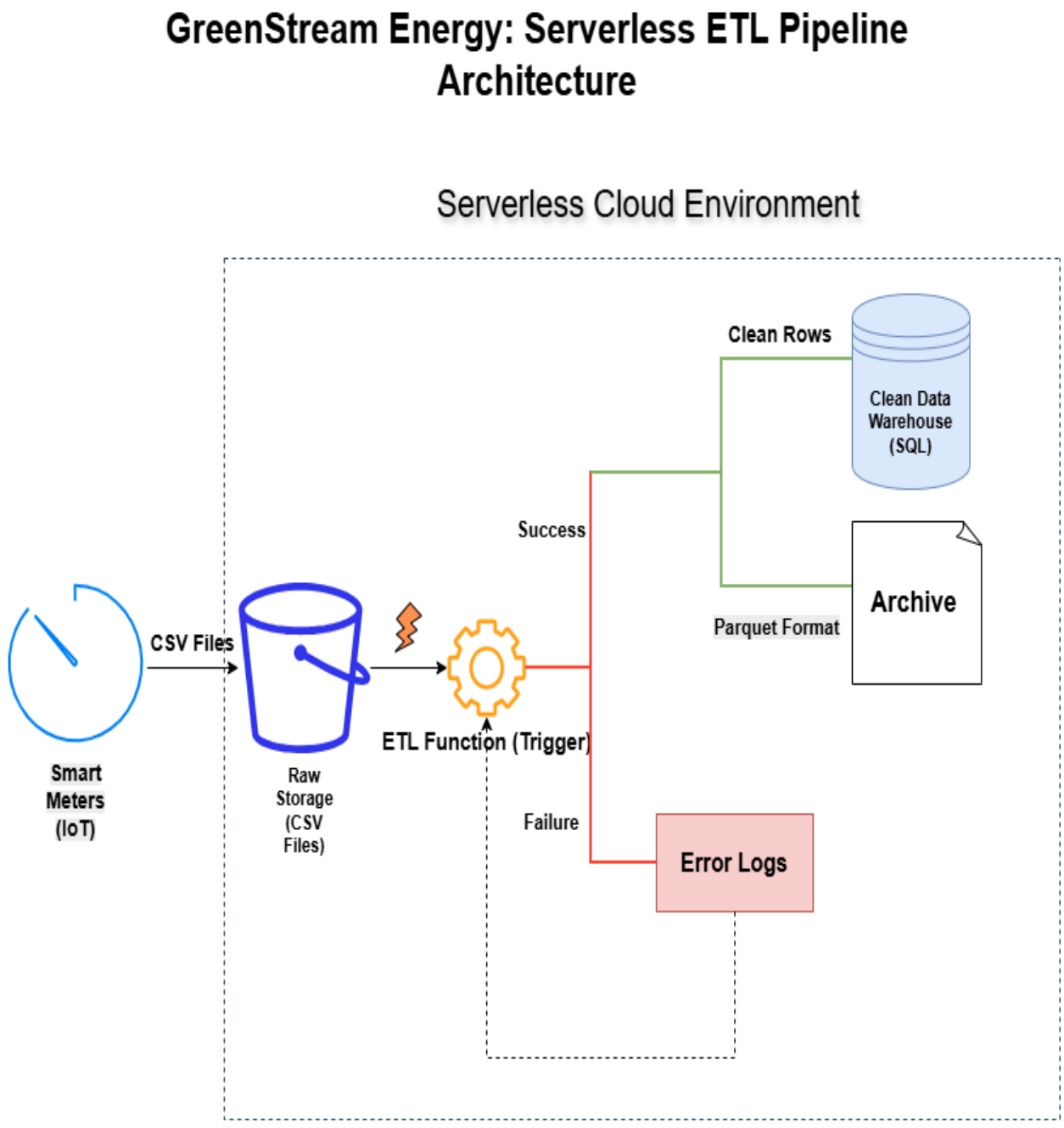
Green Stream Energy Data Pipeline Design

Task A: ETL Architecture Diagram:



Task B: Transformation Logic & Business Rules:

Rule Name	Business Logic Description
1. Unit Standardization	Check the unit column. **IF** unit is "W" (Watts), divide the reading value by 1000 and update unit to "kW". **ELSE** keep value as is.
2. Missing Values Handling	**IF** reading is NULL or empty, flag the record as "Incomplete" and exclude it from peak-usage aggregations (do not delete, just flag).
3. Faulty Meter Detection	**IF** a meter reports consumption = 0 for more than 24 consecutive hours, flag the MeterID as "Potential Faulty" for maintenance review.
4. Data Format Optimization	Convert the final cleaned data from row-based **CSV** to columnar **Parquet** format to optimize storage and query speed for historical analysis.

Task C: Single Record Lifecycle:

Step-by-Step Lifecycle of a Smart Meter Record:

1. **Ingestion:** A smart meter in a household sends a raw CSV file (containing reading: 1500 W) to the cloud **Raw Storage (S3)**.
2. **Triggering:** The arrival of this file automatically triggers the serverless **ETL Function** (Event-Driven Architecture).
3. **Validation & Transformation:** The function reads the record. It detects the unit is "W", so it applies **Rule #1** (divides by 1000) to convert it to 1.5 kW. It also checks for NULLs.

4. **Loading (Structured Storage):** The cleaned record (1.5 kW) is inserted into the **Data Warehouse** for immediate reporting.
5. **Archiving:** A copy of the clean data is converted to **Parquet format** and saved in the **Archive Storage** for future Machine Learning tasks.
6. **Error Handling:** If the record was corrupt or had invalid formats, it would have been routed to the **Error Logs** for manual inspection, bypassing the warehouse.