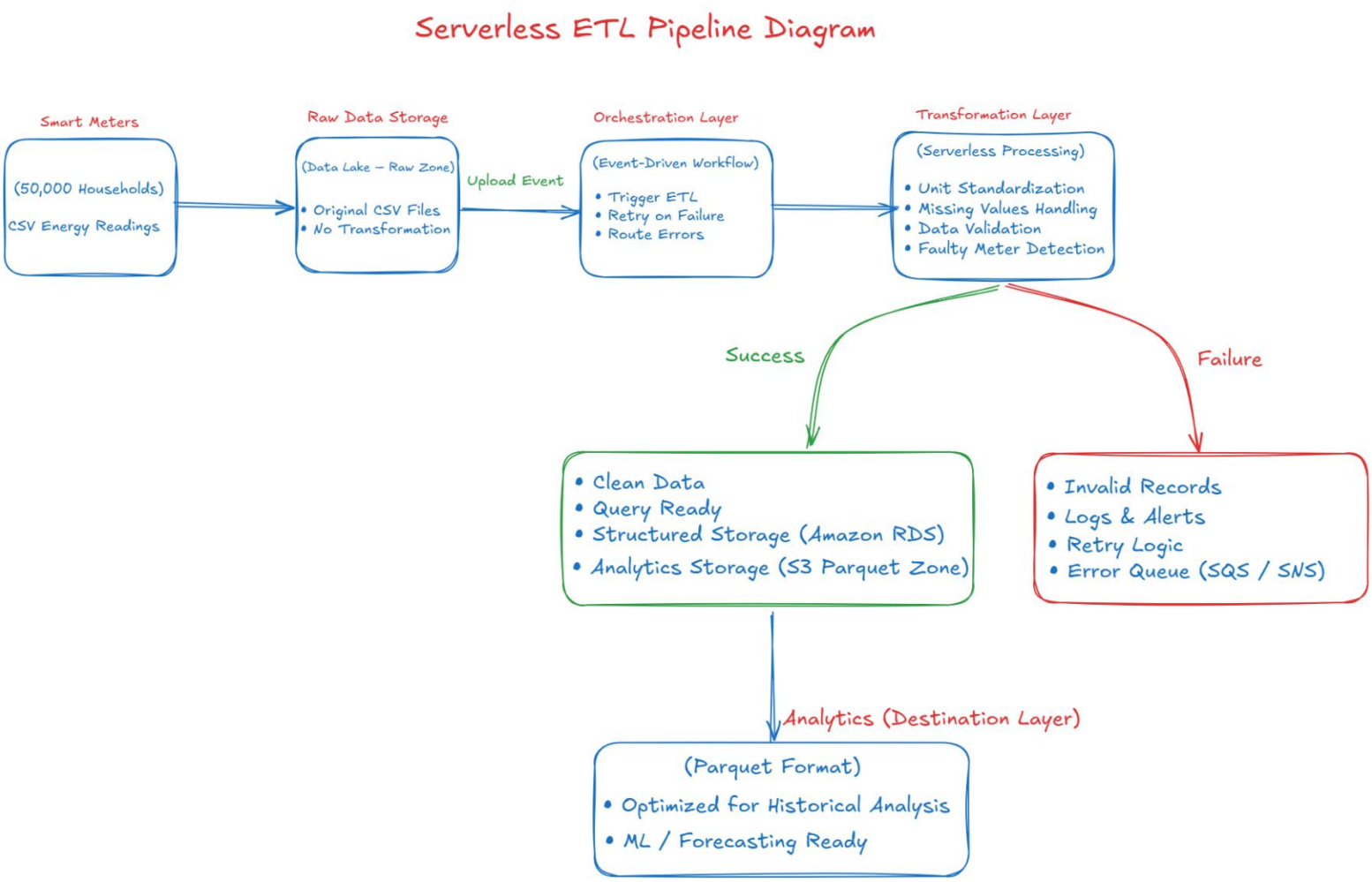
Task A : System Design



Task B : Transformation Logic

1. **Unit Standardization:**
   All energy readings are converted to kilowatts (kW). Values reported in watts (W) are divided by 1000 to ensure consistency.

2. **Missing Values:**
   Records with missing energy readings are flagged and excluded from peak consumption analysis but retained for monitoring purposes.

3. **Data Validation:**
   Records with invalid timestamps, negative values, or unrealistic consumption levels are rejected to maintain data quality.

4. **Duplicate Records:**
   Duplicate readings with the same meter ID and timestamp are resolved by keeping the most recent record.

5. **Faulty Meter Detection:**
   Meters reporting zero or near-zero consumption for extended periods are flagged as potentially faulty.

Task C: Single Record Lifecycle

single smart-meter record follows a clearly defined lifecycle within the pipeline:

1. **Raw Upload:**
   The record is uploaded in CSV format to raw storage without modification.
2. **Event Triggering:**
   The upload event triggers the orchestration workflow, initiating the transformation process.
3. **Transformation and Validation:**
   The record undergoes unit conversion, validation checks, and missing value evaluation. Fault detection logic is applied if necessary.
4. **Structured Storage:**
   If the record passes all validation rules, it is stored in a relational database in a structured format suitable for querying and analysis.
5. **Analytics Archival:**
   The record is later converted into Parquet format and archived in the analytics data lake for historical analysis and predictive modeling.
6. **Failure Handling:**
   If the record fails at any stage, retries are attempted automatically. Persistent failures result in the record being stored in an error repository with appropriate logs and alerts.