ETL – Telecom Data Processing with SSIS

1. Introduction

This project demonstrates the design and implementation of an ETL pipeline for a telecom

company. The company generates CSV files containing subscriber transaction data every

5 minutes. The task was to load these files into a SQL Server database using SQL Server

Integration Services (SSIS) applies business rules, validates data quality, and stores rejected

records separately.

2. Requirements

- Load CSV files generated every 5 minutes.
- Apply validation rules (IMSI, IMEI, CELL, LAC, EVENT_TS).
- Reject invalid records and store them in a separate table.
- Log the original file name for rejected records.
- Archive processed CSV files in a separate folder.

3. Solution Architecture

The ETL process was implemented in SSIS and consists of two main components:

Control Flow: Uses a Foreach Loop Container to iterate over incoming CSV files. A Data Flow Task processes each file, and once completed, it is moved to an archive folder

using a File System Task.

Data Flow: Reads the CSV data, applies validation rules, performs transformations, and

loads valid records into the Transaction table while redirecting invalid rows to the Failed

Transactions table.

4. Implementation Details

• Flat File Source: Reads telecom event data from CSV files.

• Lookup Transformation: Joins IMSI with reference table to fetch subscriber_id (assigns -99999 if null).

• Derived Column: Creates TAC (first 8 chars of IMEI) and SNR (last 6 chars). If IMEI is null/too short, assigns -99999.

• Conditional Split/Error Output: Separates valid and invalid records based on rules.

• OLE DB Destination (Valid Data): Loads valid records into the Successful Transaction table.

• OLE DB Destination (Error Output): Stores rejected records in the Failed Transaction table, along with the file name.

• File System Task: Moves processed CSV file into archive folder.

5. Data Validation Rules

• ID: Transaction_id (as-is).

• IMSI: must not be null, else reject.

• IMSI: Lookup subscriber_id, assign -99999 if not found.

• IMEI: TAC & SNR If null/length < 14 => -99999

• CELL: must not be null, else reject.

• LAC: must not be null, else reject.

• EVENT_TYPE: As-is.

• EVENT_TS: must be a valid datetime, else reject.

6. Testing & Results

The ETL process was tested with sample telecom CSV data. Valid records were correctly inserted into the Transaction table, while invalid rows (e.g., missing IMSI or malformed timestamps) were redirected to the Failed Transaction table. Processed CSV files were successfully moved to the archive folder.

7. Conclusion

The ETL pipeline ensures automation, data quality, and robust error handling for telecom

transaction data. It reduces manual intervention and guarantees reliable storage of

subscriber records for analysis and reporting.

8. Future Enhancements

- Automate job execution with SQL Server Agent.
- Add data quality reporting and monitoring.
- Extend pipeline to integrate with cloud storage platforms.
- Implement incremental data loading strategies for scalability.