

ETL – Telecom case study

One of the telecommunications companies contacted you as an ETL Developer/Consultant to perform data storage according to the following:

1. The company has a system that saves a CSV file periodically every 5 minutes. This file contains the basic data of various transactions performed by customers during a specified time period.
2. The company has provided you with the following table, which illustrates the data stored in the CSV file:

Column Name	Data Type	Length	Is Nullable	Sample
ID	Int		False	156
IMSI	String	9	False	310120265
IMEI	String	14	True	490154203237518
CELL	Int		False	123
LAC	Int		False	12
EVENT_TYPE	String	1	True	1
EVENT_TS	Timestamp		False	16/1/2020 7:45:43

3. The required processing needs to be performed on this data before storing it in the database

Column Name	Mapping Rules	Target Model
ID	As-is	Transaction_id
IMSI	As-is, reject the record if null	IMSI
IMSI	Join with IMSI reference and get subscriber id, add - 99999 if null	subscriber_id
IMEI	Substring first 8 chars, if null or size is less than 15 add - 99999	TAC
IMEI	Substring last 6 chars, if null or size is less than 15 add - 99999	SNR
IMEI	As-is	IMEI

CELL	As-is, reject the record if null	CELL
LAC	As-is, reject the record if null	LAC
EVENT_TYPE	As-is	EVENT_TYPE
EVENT_TS	Validate the timestamp format to be YYYY-MM-DD HH:MM:SS, reject the record if null	EVENT_TS

4. The rejected records will be stored in a separate table along with recording the name of the original CSV file.
5. During the data registration process in the database, there are additional data that needs to be recorded to ensure the quality of the data storage process:
 - a. The number of transactions in the CSV file.
 - b. The number of transactions that have been stored in the database.
 - c. The number of rejected transactions due to not meeting the required conditions.
 - d. Linking the stored data in the database with the original CSV file.
6. After completing the data storage process in the database according to the previous conditions, move the CSV file to another folder.