

ETL - Telecom - Garage Education

[المصدر: مجموعة [الفيديوهات](#) المنشورة على قناة Garage Education بواسطة المهندس/ مصطفى علاء،
وبها تم التطبيق من خلال لغة Scala، الهدف من هذا المستند هو تكرار **فكرة** السلسلة ولكن باستخدام
[SQL Server Integration Services SSIS

إحدى شركات الاتصالات تواصلت معك بصفتك ETL Developer/Consultant للقيام بعملية تخزين البيانات
وفقاً لما يلي

1. الشركة لديها سيستم يقوم بحفظ ملف csv بشكل دوري كل 5 دقائق، هذا الملف يشتمل على
البيانات الأساسية للحركات المختلفة التي قام بها العملاء خلال فترة زمنية محددة.
2. قدمت الشركة لك الجدول التالي والذي يوضح البيانات المخزنة في ملف csv و

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. المعالجة المطلوبة إتمامها على هذه البيانات قبل تخزينها في قاعدة البيانات

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. الحركات المرفوضة rejected records، سيتم تخزينها في جدول مستقل، مع تسجيل اسم الملف الأصلي csv

5. خلال عملية تسجيل البيانات في قاعدة البيانات، هناك بعض البيانات الإضافية المطلوب تسجيلها لضمان جودة عملية تخزين البيانات

- عدد الحركات في ملف csv
- عدد الحركات التي تم تخزينها في قاعدة البيانات
- عدد الحركات المرفوضة لعدم تحقق الشروط المطلوبة
- ربط البيانات المخزنة في قاعدة البيانات مع الملف الأصلي csv

6. بعد إتمام عملية تخزين البيانات في قاعدة البيانات وفقا للشروط السابقة، انقل ملف csv الى فولدر آخر