

## Data Migration

Con una frecuencia inicial en principio mensual, se generará la migración de los datos de contactabilidad (teléfonos y mails) presentes en la base Active hacia las tablas de Datos Adicionales del Moroso, particularmente, DAM\_CONTACTABILIDAD.

Para esto, se seguirán los siguientes pasos (similares a los de cualquier enriquecimiento por proveedor externo):

## Data Cleaning

1. Se correrá una query para traer todos los datos de contactabilidad de Active (teléfonos y mails) de todos los campos donde tengamos posibles inserciones.
2. Se ejecuta el notebook Data Cleaning - Mails. Si tuviésemos duplicados, pasarían al notebook de Duplicados correspondiente.
  - a. Entre otras cosas, el notebook revisa si la string tiene un @, y además si existen o no algunos caracteres especiales (\$, -, /).
  - b. Se genera la limpieza de los números de documento.
  - c. Adicionalmente, se revisan algunos casos particulares de ocurrencia de mails que no son verdaderos (ejemplo: no@no.com).
3. Se ejecuta el notebook Data Cleaning - Teléfonos. Si tuviésemos duplicados, pasarían al notebook de duplicados correspondientes.
  - a. Se generan ciertas condiciones relacionadas con la longitud de los caracteres
    - i. En particular, que tenga al menos 2 caracteres diferentes. Por lo que nos quedamos con aquellos que tengan una longitud mayor a 1.
  - b. Se genera la limpieza de los números de documento.
  - c. Se revisa si existen o no algunos caracteres especiales en la string (\$, -, /) para eliminarlos.
4. Se exportan los datos en formato .csv.

## Data Insert

Los datos deberán ser insertados dentro de la tabla DAM\_CONTACTABILIDAD, que se encuentra en la database Active (Active.dbo.DAM\_CONTACTABILIDAD).

Como la cantidad de records es alta (en el orden de millones), se deberá generar una inserción por batch, en un horario que no interfiera con las operaciones (entre 8PM y 8AM aproximadamente). Para esto se deberá generar un notebook que divida el .csv en batches (pueden ser dataframes) de 100k records cada uno aproximadamente (validar con IT). Se deberá notificar al equipo de IT sobre la inserción, para ver en qué franja horaria puede realizarse.