

# Funzionamento INGESTION & ETL ENEL X-FS

La fase di Ingestion consiste in tabelle esterne che puntano a tabelle su HBase aggiornate dall'applicativo.

Le tabelle su HBase sono 3 anagrafiche ( tfruta:svaProducts,tfruta:subscribers e ubiq:anonymization ) e 1 di transazioni ( tfruta:transactions )

Per le tabelle anagrafiche , così come per l'ETL di Tfruta, ho implementato una logica di ETL in overwrite; mentre, per la tabella delle transazioni (partizionata) una logica incrementale.

Per ottenere le strutture (DDL) delle tabelle in produzione è sufficiente una show create table su Hive. Tuttavia, nello zip allegato le ho inserite, insieme agli script degli ETL

L'ETL è schedulato su Nificonfig con nome "enelpayprod" alle 03:00 con quest'ordine (I primi 3 girano in parallelo):

- 1) ETL\_ENELPAY\_TRANSACTIONS.sql
- 1) ETL\_ENELPAY\_SVAPRODUCTS.sql
- 1) ETL\_ENELPAY\_ANOM.sql
- 2) ETL\_ENELPAY\_SUBSCRIBERS.sql

Per fare il passaggio da una EXTERNAL TABLE Hive che punta a una su HBase a una tabella su Hive, bisogna:

- 1) Creare la tabella Hive (Non external formato-PARQUET ) che ospiterà i dati aggiornati , per esempio : a\_tra,a\_sub,a\_svapro ( per DDL vedi script allegato )
- 2) ETL che carica i dati dalla EXTERNAL ( x es e\_tra, e\_sub ) alla PARQUET di Hive ( per gli ETL vedi gli script in allegato )

In questo link [RELEASE NOTE\\_BI\\_02-07-2020\\_\\_Rilascio Aggiunta CodeSharer Subscriber ENELPAY](#) trovi una release in cui ci hanno chiesto di aggiungere un campo alla tabella dei subscribers ( e\_sub )

Per aggiungere un campo, bisogna:

- 1) inserire il campo nella ddl della tabella e\_
- 2) inserire il campo nei parametri "WITH SERDEPROPERTIES" hbase.columns.mapping della tabella e\_
- 3) inserire il nuovo campo nella ddl della tabella a\_
- 4) inserire il nuovo campo nei rispettivi ETL

Per il punto 2, è necessario sapere la column family su HBase, alcune cose da sapere:  
se il campo è float/intero/booleano su HBase , bisogna aggiungere #b nei parametri, ad esempio:

voglio aggiungere campo supplierTax (che è un float )appartenente alla colum family billingInfo, devo usare questa sintassi:

billingInfo:supplierTax\_f#b

Se invece il campo è string o di qualsiasi altro tipo, basta mettere: attr: NomeCampoSuHbase



DDL&ETL ENELPAY.zip

