set schema 'colleges2';

-----

-- PREREQUIS : Ce script ne fonctionne que si vous avez défini un répertoire de travail (Working directory)

-- par défaut dans lequel se trouve votre script et les données sous forme de fichiers CSV

-- \_region - infos dans fr-en-etablissements-ep

/\*WbImport -file=./fr-en-etablissements-ep.csv

-header=true

-delimiter=';'

-quoteChar='"'

-table=\_region

-schema=colleges2

-mode=insert,update

-filecolumns=$wb\_skip$,$wb\_skip$,$wb\_skip$,$wb\_skip$,$wb\_skip$,$wb\_skip$,$wb\_skip$,$wb\_skip$,libelle\_region,$wb\_skip$,$wb\_skip$,$wb\_skip$,$wb\_skip$,$wb\_skip$,$wb\_skip$,$wb\_skip$,$wb\_skip$,$wb\_skip$,code\_region

-importColumns=libelle\_region,code\_region

-keyColumns=code\_region; \*/

/\* Cette approche est loin d'être optimale et va requérir de charger plusieurs fois le même fichier CSV pour différentes tables.

Importing file '/Users/delhay/vrac/git/enseignement/but\_info/but\_info\_bdd/bdd\_s2/sae\_204/colleges-ips/enonce/soluce/./fr-en-etablissements-ep.csv' into table \_region

19 row(s) inserted

8430 row(s) updated

Execution time: 3m 11s

Il vaut donc mieux charger une bonne fois pour toute tous les fichiers CSV dans de grosses tables temporaires puis faire des requêtes. \*/

-- Données générales pour l'année 2022-2023

drop table if exists \_temp\_fr\_en\_etablissements\_ep cascade;

create table \_temp\_fr\_en\_etablissements\_ep(

uai char(8) primary key,

ep\_2022\_2023 varchar(7), -- ep pour l'année 2022-2023 : REP, REP+, HORS REP. Pas utilisé a priori

nom\_etablissement varchar(108),

type\_etablissement varchar(39), -- libelle\_nature

statut\_public\_prive varchar(18), -- secteur

libelle\_academie varchar(16), -- lib\_academie

libelle\_departement varchar(23), -- nom\_departement

nom\_commune varchar(45), -- nom\_de\_la\_commune

libelle\_region varchar(34), -- libelle\_region

uai\_tete\_de\_reseau char(8), -- pas utilisé

qp\_a\_proximite\_o\_n char(16), -- 'Dans QP' ou NULL

qp\_a\_proximite char(8), -- code\_quartier\_prioritaire

nom\_du\_qp varchar(85), -- nom\_quartier\_prioritaire

nombre\_d\_eleves integer, -- effectifs

code\_postal char(5), -- code\_postal

code\_commune char(5), -- code\_insee\_de\_la\_commune

code\_departement char(3), -- code\_du\_departement

code\_academie char(2), -- code\_academie

code\_region char(2), -- code\_region

libelle\_nature varchar(39), -- libelle\_nature

code\_nature char(3), -- code\_nature

latitude numeric,

longitude numeric

);

delete from \_temp\_fr\_en\_etablissements\_ep;

WbImport -file=./fr-en-etablissements-ep.csv

-header=true

-delimiter=';'

-quoteChar='"'

-table=\_temp\_fr\_en\_etablissements\_ep

-schema=colleges2

-mode=insert, update -- car il y a des doublons (toute la ligne)

-filecolumns=uai,ep\_2022\_2023,nom\_etablissement,type\_etablissement,statut\_public\_prive,libelle\_academie,libelle\_departement,nom\_commune,libelle\_region,uai\_tete\_de\_reseau,qp\_a\_proximite\_o\_n,qp\_a\_proximite,nom\_du\_qp,nombre\_d\_eleves,code\_postal,code\_commune,code\_departement,code\_academie,code\_region,libelle\_nature,code\_nature,$wb\_skip$,latitude,longitude

-importColumns=uai,ep\_2022\_2023,nom\_etablissement,type\_etablissement,statut\_public\_prive,libelle\_academie,libelle\_departement,nom\_commune,libelle\_region,uai\_tete\_de\_reseau,qp\_a\_proximite\_o\_n,qp\_a\_proximite,nom\_du\_qp,nombre\_d\_eleves,code\_postal,code\_commune,code\_departement,code\_academie,code\_region,libelle\_nature,code\_nature,latitude,longitude

-keyColumns=uai;

-- Données sur les effectifs de 2019-2020 à 2022-2023

drop table if exists \_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv;

create table \_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv(

rentree\_scolaire char(9) not null,

region\_academique varchar(26) not null,

academie varchar(16) not null,

departement varchar(21) not null,

commune varchar(32) not null,

numero\_college char(8) not null,

denomination\_principale varchar(30) not null,

patronyme varchar(30),

secteur varchar(6),

rep numeric(1),

rep\_plus numeric(1),

nombre\_eleves\_total numeric not null default 0, -- si champ vide alors 0

nombre\_eleves\_segpa numeric not null default 0,

nombre\_eleves\_ulis numeric not null default 0,

\_6eme\_total numeric not null default 0,

\_6eme\_segpa numeric not null default 0,

\_6eme\_ulis numeric not null default 0,

\_6eme\_filles numeric not null default 0,

\_6emes\_garcons numeric not null default 0,

\_6eme\_lv1\_allemand numeric not null default 0,

\_6eme\_lv1\_anglais numeric not null default 0,

\_6eme\_lv1\_espagnol numeric not null default 0,

\_6eme\_lv1\_autres numeric not null default 0,

\_6eme\_lv2\_allemand numeric not null default 0,

\_6eme\_lv2\_anglais numeric not null default 0,

\_6eme\_lv2\_espagnol numeric not null default 0,

\_6eme\_lv2\_italien numeric not null default 0,

\_6eme\_lv2\_autres numeric not null default 0,

\_5eme\_total numeric not null default 0,

\_5eme\_segpa numeric not null default 0,

\_5eme\_ulis numeric not null default 0,

\_5eme\_filles numeric not null default 0,

\_5emes\_garcons numeric not null default 0,

\_5eme\_lv1\_allemand numeric not null default 0,

\_5eme\_lv1\_anglais numeric not null default 0,

\_5eme\_lv1\_espagnol numeric not null default 0,

\_5eme\_lv1\_autres numeric not null default 0,

\_5eme\_lv2\_allemand numeric not null default 0,

\_5eme\_lv2\_anglais numeric not null default 0,

\_5eme\_lv2\_espagnol numeric not null default 0,

\_5eme\_lv2\_italien numeric not null default 0,

\_5eme\_lv2\_autres numeric not null default 0,

\_4eme\_total numeric not null default 0,

\_4eme\_segpa numeric not null default 0,

\_4eme\_ulis numeric not null default 0,

\_4eme\_filles numeric not null default 0,

\_4emes\_garcons numeric not null default 0,

\_4eme\_lv1\_allemand numeric not null default 0,

\_4eme\_lv1\_anglais numeric not null default 0,

\_4eme\_lv1\_espagnol numeric not null default 0,

\_4eme\_lv1\_autres numeric not null default 0,

\_4eme\_lv2\_allemand numeric not null default 0,

\_4eme\_lv2\_anglais numeric not null default 0,

\_4eme\_lv2\_espagnol numeric not null default 0,

\_4eme\_lv2\_italien numeric not null default 0,

\_4eme\_lv2\_autres numeric not null default 0,

\_3eme\_total numeric not null default 0,

\_3eme\_segpa numeric not null default 0,

\_3eme\_ulis numeric not null default 0,

\_3eme\_filles numeric not null default 0,

\_3emes\_garcons numeric not null default 0,

\_3eme\_lv1\_allemand numeric not null default 0,

\_3eme\_lv1\_anglais numeric not null default 0,

\_3eme\_lv1\_espagnol numeric not null default 0,

\_3eme\_lv1\_autres numeric not null default 0,

\_3eme\_lv2\_allemand numeric not null default 0,

\_3eme\_lv2\_anglais numeric not null default 0,

\_3eme\_lv2\_espagnol numeric not null default 0,

\_3eme\_lv2\_italien numeric not null default 0,

\_3eme\_lv2\_autres numeric not null default 0,

code\_postal char(5) not null,

constraint \_temp\_effectif\_pk primary key (rentree\_scolaire, numero\_college)

);

-- Reformatage des années en années scolaires.

drop function if exists ftg\_rentree4\_rentree9();

create or replace function ftg\_rentree4\_rentree9() returns trigger as $$

declare

annee\_int integer;

begin

annee\_int := new.rentree\_scolaire::integer;

new.rentree\_scolaire := to\_char(annee\_int,'FM9999') || '-' || to\_char(annee\_int+1,'FM9999');

return new;

end;

$$ language 'plpgsql';

-- Trigger se déclenchant sur ajout de données dans \_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv

drop trigger if exists tg\_rentree4\_rentree9 on \_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv;

create trigger tg\_rentree4\_rentree9

before insert or update of rentree\_scolaire

on \_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv

for each row

execute procedure ftg\_rentree4\_rentree9();

-- rentree\_scolaire;region\_academique;academie;departement;commune;numero\_college;denomination\_principale;patronyme;secteur;rep;rep\_plus;nombre\_eleves\_total;nombre\_eleves\_total\_hors\_segpa\_hors\_ulis;nombre\_eleves\_segpa;nombre\_eleves\_ulis;6eme\_total;6eme\_hors\_segpa\_hors\_ulis;6eme\_segpa;6eme\_ulis;6eme\_filles;6emes\_garcons;6eme\_lv1\_allemand;6eme\_lv1\_anglais;6eme\_lv1\_espagnol;6eme\_lv1\_autres;6eme\_lv2\_allemand;6eme\_lv2\_anglais;6eme\_lv2\_espagnol;6eme\_lv2\_italien;6eme\_lv2\_autres;5eme\_total;5eme\_hors\_segpa\_hors\_ulis;5eme\_segpa;5eme\_ulis;5eme\_filles;5emes\_garcons;5eme\_lv1\_allemand;5eme\_lv1\_anglais;5eme\_lv1\_espagnol;5eme\_lv1\_autres;5eme\_lv2\_allemand;5eme\_lv2\_anglais;5eme\_lv2\_espagnol;5eme\_lv2\_italien;5eme\_lv2\_autres;4eme\_total;4eme\_hors\_segpa\_hors\_ulis;4eme\_segpa;4eme\_ulis;4eme\_filles;4emes\_garcons;4eme\_lv1\_allemand;4eme\_lv1\_anglais;4eme\_lv1\_espagnol;4eme\_lv1\_autres;4eme\_lv2\_allemand;4eme\_lv2\_anglais;4eme\_lv2\_espagnol;4eme\_lv2\_italien;4eme\_lv2\_autres;3eme\_total;3eme\_hors\_segpa\_hors\_ulis;3eme\_segpa;3eme\_ulis;3eme\_filles;3emes\_garcons;3eme\_lv1\_allemand;3eme\_lv1\_anglais;3eme\_lv1\_espagnol;3eme\_lv1\_autres;3eme\_lv2\_allemand;3eme\_lv2\_anglais;3eme\_lv2\_espagnol;3eme\_lv2\_italien;3eme\_lv2\_autres;code\_postal

delete from \_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv;

WbImport -file=./fr-en-college-effectifs-niveau-sexe-lv.csv

-header=true

-delimiter=';'

-quoteChar='"'

-table=\_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv

-schema=colleges2

-mode=insert, update -- car il y a des doublons (toute la ligne)

-filecolumns=rentree\_scolaire,region\_academique,academie,departement,commune,numero\_college,denomination\_principale,patronyme,secteur,rep,rep\_plus,nombre\_eleves\_total,$wb\_skip$,nombre\_eleves\_segpa,nombre\_eleves\_ulis,\_6eme\_total,$wb\_skip$,\_6eme\_segpa,\_6eme\_ulis,\_6eme\_filles,\_6emes\_garcons,\_6eme\_lv1\_allemand,\_6eme\_lv1\_anglais,\_6eme\_lv1\_espagnol,\_6eme\_lv1\_autres,\_6eme\_lv2\_allemand,\_6eme\_lv2\_anglais,\_6eme\_lv2\_espagnol,\_6eme\_lv2\_italien,\_6eme\_lv2\_autres,\_5eme\_total,$wb\_skip$,\_5eme\_segpa,\_5eme\_ulis,\_5eme\_filles,\_5emes\_garcons,\_5eme\_lv1\_allemand,\_5eme\_lv1\_anglais,\_5eme\_lv1\_espagnol,\_5eme\_lv1\_autres,\_5eme\_lv2\_allemand,\_5eme\_lv2\_anglais,\_5eme\_lv2\_espagnol,\_5eme\_lv2\_italien,\_5eme\_lv2\_autres,\_4eme\_total,$wb\_skip$,\_4eme\_segpa,\_4eme\_ulis,\_4eme\_filles,\_4emes\_garcons,\_4eme\_lv1\_allemand,\_4eme\_lv1\_anglais,\_4eme\_lv1\_espagnol,\_4eme\_lv1\_autres,\_4eme\_lv2\_allemand,\_4eme\_lv2\_anglais,\_4eme\_lv2\_espagnol,\_4eme\_lv2\_italien,\_4eme\_lv2\_autres,\_3eme\_total,$wb\_skip$,\_3eme\_segpa,\_3eme\_ulis,\_3eme\_filles,\_3emes\_garcons,\_3eme\_lv1\_allemand,\_3eme\_lv1\_anglais,\_3eme\_lv1\_espagnol,\_3eme\_lv1\_autres,\_3eme\_lv2\_allemand,\_3eme\_lv2\_anglais,\_3eme\_lv2\_espagnol,\_3eme\_lv2\_italien,\_3eme\_lv2\_autres,code\_postal

-importColumns=rentree\_scolaire,region\_academique,academie,departement,commune,numero\_college,denomination\_principale,patronyme,secteur,rep,rep\_plus,nombre\_eleves\_total,nombre\_eleves\_total\_hors\_segpa\_hors\_ulis,nombre\_eleves\_segpa,nombre\_eleves\_ulis,\_6eme\_total,\_6eme\_hors\_segpa\_hors\_ulis,\_6eme\_segpa,\_6eme\_ulis,\_6eme\_filles,\_6emes\_garcons,\_6eme\_lv1\_allemand,\_6eme\_lv1\_anglais,\_6eme\_lv1\_espagnol,\_6eme\_lv1\_autres,\_6eme\_lv2\_allemand,\_6eme\_lv2\_anglais,\_6eme\_lv2\_espagnol,\_6eme\_lv2\_italien,\_6eme\_lv2\_autres,\_5eme\_total,\_5eme\_hors\_segpa\_hors\_ulis,\_5eme\_segpa,\_5eme\_ulis,\_5eme\_filles,\_5emes\_garcons,\_5eme\_lv1\_allemand,\_5eme\_lv1\_anglais,\_5eme\_lv1\_espagnol,\_5eme\_lv1\_autres,\_5eme\_lv2\_allemand,\_5eme\_lv2\_anglais,\_5eme\_lv2\_espagnol,\_5eme\_lv2\_italien,\_5eme\_lv2\_autres,\_4eme\_total,\_4eme\_hors\_segpa\_hors\_ulis,\_4eme\_segpa,\_4eme\_ulis,\_4eme\_filles,\_4emes\_garcons,\_4eme\_lv1\_allemand,\_4eme\_lv1\_anglais,\_4eme\_lv1\_espagnol,\_4eme\_lv1\_autres,\_4eme\_lv2\_allemand,\_4eme\_lv2\_anglais,\_4eme\_lv2\_espagnol,\_4eme\_lv2\_italien,\_4eme\_lv2\_autres,\_3eme\_total,\_3eme\_hors\_segpa\_hors\_ulis,\_3eme\_segpa,\_3eme\_ulis,\_3eme\_filles,\_3emes\_garcons,\_3eme\_lv1\_allemand,\_3eme\_lv1\_anglais,\_3eme\_lv1\_espagnol,\_3eme\_lv1\_autres,\_3eme\_lv2\_allemand,\_3eme\_lv2\_anglais,\_3eme\_lv2\_espagnol,\_3eme\_lv2\_italien,\_3eme\_lv2\_autres,code\_postal

-keyColumns=rentree\_scolaire,numero\_college;

-- Données sur la valeur ajoutée

drop table if exists \_temp\_fr\_en\_indicateurs\_valeur\_ajoutee\_colleges;

create table \_temp\_fr\_en\_indicateurs\_valeur\_ajoutee\_colleges(

session char(9),

uai char(8),

nom\_de\_l\_etablissement varchar(87) not null,

commune varchar(32) not null,

departement varchar(21) not null,

academie varchar(16) not null,

secteur char(2) not null,

nb\_candidats\_g numeric default 0,

taux\_de\_reussite\_g numeric(5,1),

va\_du\_taux\_de\_reussite\_g numeric(5,1),

nb\_candidats\_p numeric default 0,

taux\_de\_reussite\_p numeric(5,1),

note\_a\_l\_ecrit\_g numeric(4,1),

va\_de\_la\_note\_g numeric(4,1),

note\_a\_l\_ecrit\_p numeric(4,1),

taux\_d\_acces\_6eme\_3eme numeric(5,1),

part\_presents\_3eme\_ordinaire\_total numeric(5,1),

part\_presents\_3eme\_ordinaire\_g numeric(5,1),

part\_presents\_3eme\_ordinaire\_p numeric(5,1),

part\_presents\_3eme\_segpa\_total numeric(5,1),

nb\_mentions\_ab\_g numeric default 0,

nb\_mentions\_b\_g numeric default 0,

nb\_mentions\_tb\_g numeric default 0,

constraint indicateurs\_valeur\_ajoutee\_pk primary key (uai, session)

);

-- reformatage des années de session en années scolaires

drop function if exists ftg\_insert\_session();

create or replace function ftg\_insert\_session() returns trigger as $$

declare

annee\_session integer;

begin

annee\_session := new.session::integer;

new.session := to\_char(annee\_session-1,'FM9999') || '-' || to\_char(annee\_session,'FM9999');

return new;

end;

$$ language 'plpgsql';

drop trigger if exists tg\_insert\_session on \_temp\_fr\_en\_indicateurs\_valeur\_ajoutee\_colleges;

create trigger tg\_insert\_session

before insert or update

on \_temp\_fr\_en\_indicateurs\_valeur\_ajoutee\_colleges

for each row

execute procedure ftg\_insert\_session();

-- session;uai;nom\_de\_l\_etablissement;commune;departement;academie;secteur;nb\_candidats\_g;taux\_de\_reussite\_g;va\_du\_taux\_de\_reussite\_g;nb\_candidats\_p;taux\_de\_reussite\_p;note\_a\_l\_ecrit\_g;va\_de\_la\_note\_g;note\_a\_l\_ecrit\_p;taux\_d\_acces\_6eme\_3eme;part\_presents\_3eme\_ordinaire\_total;part\_presents\_3eme\_ordinaire\_g;part\_presents\_3eme\_ordinaire\_p;part\_presents\_3eme\_segpa\_total;nb\_mentions\_ab\_g;nb\_mentions\_b\_g;nb\_mentions\_tb\_g;nb\_mentions\_global\_g

delete from \_temp\_fr\_en\_indicateurs\_valeur\_ajoutee\_colleges;

WbImport -file=./fr-en-indicateurs-valeur-ajoutee-colleges.csv

-header=true

-delimiter=';'

-quoteChar='"'

-table=\_temp\_fr\_en\_indicateurs\_valeur\_ajoutee\_colleges

-schema=colleges2

-mode=insert, update -- car il y a des doublons (toute la ligne)

-filecolumns=session,uai,nom\_de\_l\_etablissement,commune,departement,academie,secteur,nb\_candidats\_g,taux\_de\_reussite\_g,va\_du\_taux\_de\_reussite\_g,nb\_candidats\_p,taux\_de\_reussite\_p,note\_a\_l\_ecrit\_g,va\_de\_la\_note\_g,note\_a\_l\_ecrit\_p,taux\_d\_acces\_6eme\_3eme,part\_presents\_3eme\_ordinaire\_total,part\_presents\_3eme\_ordinaire\_g,part\_presents\_3eme\_ordinaire\_p,part\_presents\_3eme\_segpa\_total,nb\_mentions\_ab\_g,nb\_mentions\_b\_g,nb\_mentions\_tb\_g

-importColumns=session,uai,nom\_de\_l\_etablissement,commune,departement,academie,secteur,nb\_candidats\_g,taux\_de\_reussite\_g,va\_du\_taux\_de\_reussite\_g,nb\_candidats\_p,taux\_de\_reussite\_p,note\_a\_l\_ecrit\_g,va\_de\_la\_note\_g,note\_a\_l\_ecrit\_p,taux\_d\_acces\_6eme\_3eme,part\_presents\_3eme\_ordinaire\_total,part\_presents\_3eme\_ordinaire\_g,part\_presents\_3eme\_ordinaire\_p,part\_presents\_3eme\_segpa\_total,nb\_mentions\_ab\_g,nb\_mentions\_b\_g,nb\_mentions\_tb\_g

-keyColumns=session,uai;

-----------------------

-- Données sur IPS

drop table if exists \_temp\_fr\_en\_ips\_colleges\_ap2022;

create table \_temp\_fr\_en\_ips\_colleges\_ap2022(

rentree\_scolaire char(9),

academie varchar(16) not null,

code\_du\_departement char(3) not null,

departement varchar(21) not null,

uai char(8),

nom\_de\_l\_etablissment varchar(79) not null,

code\_insee\_de\_la\_commune varchar(5) not null,

nom\_de\_la\_commune varchar(35) not null,

secteur varchar(18) not null,

effectifs integer not null,

ips numeric(5,1) not null,

ecart\_type\_de\_l\_ips numeric(4,1) not null,

constraint ips\_colleges\_ap2022\_pk primary key(rentree\_scolaire,uai)

);

-- rentree\_scolaire;academie;code\_du\_departement;departement;uai;nom\_de\_l\_etablissment;code\_insee\_de\_la\_commune;nom\_de\_la\_commune;secteur;effectifs;ips;ecart\_type\_de\_l\_ips

delete from \_temp\_fr\_en\_ips\_colleges\_ap2022;

WbImport -file=./fr-en-ips-colleges-ap2022.csv

-header=true

-delimiter=';'

-quoteChar='"'

-table=\_temp\_fr\_en\_ips\_colleges\_ap2022

-schema=colleges2

-mode=insert, update -- car il y a peut-être des doublons (toute la ligne)

-filecolumns=rentree\_scolaire,academie,code\_du\_departement,departement,uai,nom\_de\_l\_etablissment,code\_insee\_de\_la\_commune,nom\_de\_la\_commune,secteur,effectifs,ips,ecart\_type\_de\_l\_ips

-importColumns=rentree\_scolaire,academie,code\_du\_departement,departement,uai,nom\_de\_l\_etablissment,code\_insee\_de\_la\_commune,nom\_de\_la\_commune,secteur,effectifs,ips,ecart\_type\_de\_l\_ips

-keyColumns=rentree\_scolaire,uai;

--------------------

--------------------

-- STATS SUR LES DONNEES : a décommenter pour obtenir le résultat

/\*

select count(distinct uai) as nb\_row\_temp\_fr\_en\_etablissements\_ep

from \_temp\_fr\_en\_etablissements\_ep;

-- 8500 établissements en tout

select count(distinct uai) as nb\_row\_colleges\_temp\_fr\_en\_etablissements\_ep

from \_temp\_fr\_en\_etablissements\_ep

where type\_etablissement = 'Collège';

-- 1414

select count(\*) as nb\_row\_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv

from \_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv;

-- 32982 (colleges,annee)

select count(distinct numero\_college) as nb\_row\_colleges\_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv

from \_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv;

-- 8337

select count(distinct uai) from

(select uai

from \_temp\_fr\_en\_etablissements\_ep

where type\_etablissement = 'Collège'

intersect

select numero\_college as uai

from \_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv) as intersection;

-- 1414 (1414 si pas de restriction sur type\_etablissement)

-- A priori tous les colleges de fr\_en\_etablissements\_ep

\*/

------------------------------------------------------

---- DISTRIBUTION DES DONNEES DANS LES TABLES

------------------------------------------------------

-- \_region : 19 n-uplets

insert into \_region (code\_region, libelle\_region)

select distinct code\_region, libelle\_region

from \_temp\_fr\_en\_etablissements\_ep

where code\_region is not null;

-- \_departement : 98 n-uplets --> il en manque

insert into \_departement(code\_du\_departement, nom\_departement, code\_region)

select distinct code\_departement, libelle\_departement, code\_region

from \_temp\_fr\_en\_etablissements\_ep

where code\_departement is not null;

-- Mise en majuscules et suppression des accents

update \_departement

set nom\_departement = upper(nom\_departement);

update \_departement

set nom\_departement = replace(nom\_departement,'É','E');

update \_departement

set nom\_departement = replace(nom\_departement,'È','E');

update \_departement

set nom\_departement = replace(nom\_departement,'Ô','O');

-- \_commune :

-- Uniformisation des noms de commune

update \_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv

set commune = replace(commune, '-', ' ');

delete from \_commune;

insert into \_commune (code\_insee\_de\_la\_commune, nom\_de\_la\_commune, code\_du\_departement)

select distinct code\_postal as code\_commune, commune, code\_du\_departement

from \_temp\_fr\_en\_college\_effectif\_niveau\_sexe\_lv inner join \_departement

on departement = nom\_departement

order by code\_commune;

-- 4747 n-uplets

-- Difficulté sur la presentation des communes : prendre un référence

-- https://public.opendatasoft.com/explore/dataset/correspondance-code-cedex-code-insee/table/?flg=fr-fr

drop table if exists \_temp\_communes;

create table \_temp\_communes(

Code char(5) not null,

Libelle varchar(38),

Code\_INSEE char(5),

type\_code varchar(11) not null,

Nom\_commune varchar(45),

Nom\_departement varchar(23),

NOM\_EPCI varchar(75),

Nom\_region varchar(26)

);

delete from \_temp\_communes;

WbImport -file=./correspondance-code-cedex-code-insee.csv

-header=true

-delimiter=';'

--quoteChar='"'

-table=\_temp\_communes

-schema=colleges2

-mode=insert-- , update -- car il y a des doublons (toute la ligne)

-filecolumns=Code,Libelle,Code\_INSEE,type\_code,Nom\_commune,Nom\_departement,NOM\_EPCI,Nom\_region

-importColumns=Code,Libelle,Code\_INSEE,type\_code,Nom\_commune,Nom\_departement,NOM\_EPCI,Nom\_region

-keyColumns=code\_insee;

insert into \_commune (code\_insee\_de\_la\_commune, nom\_de\_la\_commune, code\_du\_departement)

select distinct coalesce(code\_insee, code) as code\_insee\_de\_la\_commune, commune, code\_du\_departement

from \_temp\_fr\_en\_indicateurs\_valeur\_ajoutee\_colleges va left outer join \_temp\_communes tc

on va.commune = tc.libelle

inner join \_departement d

on tc.nom\_departement = d.nom\_departement

where (code\_insee, commune) not in (select code\_insee\_de\_la\_commune, nom\_de\_la\_commune from \_commune)

and (code, commune) not in (select code\_insee\_de\_la\_commune, nom\_de\_la\_commune from \_commune);

insert into \_commune (code\_insee\_de\_la\_commune, nom\_de\_la\_commune, code\_du\_departement)

select distinct code\_commune, upper(nom\_commune), code\_departement

from \_temp\_fr\_en\_etablissements\_ep

where (code\_commune, upper(nom\_commune)) not in (select code\_insee\_de\_la\_commune, nom\_de\_la\_commune from \_commune) ;

-- Il manque des départements dans \_departement

insert into \_departement (code\_du\_departement, code\_region, nom\_departement)

values ('015','84','CANTAL'),

('032','76','GERS'),

('043','84','HAUTE-LOIRE'),

('048','76','LOZERE');

insert into \_commune (code\_insee\_de\_la\_commune, nom\_de\_la\_commune, code\_du\_departement)

select distinct code\_insee\_de\_la\_commune, nom\_de\_la\_commune, code\_du\_departement

from \_temp\_fr\_en\_ips\_colleges\_ap2022

where (code\_insee\_de\_la\_commune, nom\_de\_la\_commune) not in (select code\_insee\_de\_la\_commune, nom\_de\_la\_commune from \_commune) ;

----------------

--- \_academie

delete from \_academie;

insert into \_academie (code\_academie,lib\_academie)

select distinct code\_academie, libelle\_academie

from \_temp\_fr\_en\_etablissements\_ep

where code\_academie is not null;

-- Nettoyage des noms d'académies : majuscules et sans accents

update \_academie

set lib\_academie = upper(lib\_academie);

update \_academie

set lib\_academie = replace(lib\_academie,'É','E');

update \_academie

set lib\_academie = replace(lib\_academie,'È','E');

update \_academie

set lib\_academie = replace(lib\_academie,'Ô','O');

------------------------

-- Quartier prioritaire

delete from \_quartier\_prioritaire;

--- A compléter

INSERT INTO \_quartier\_prioritaire (code\_quartier\_prioritaire, nom\_quartier\_prioritaire)

SELECT DISTINCT qp\_a\_proximite, nom\_du\_qp

FROM \_temp\_fr\_en\_etablissements\_ep

WHERE (qp\_a\_proximite IS NOT NULL AND nom\_du\_qp IS NOT NULL)

ON CONFLICT (code\_quartier\_prioritaire) DO NOTHING;

delete from \_type;

--- A compléter

INSERT INTO \_type (code\_nature, libelle\_nature)

SELECT DISTINCT code\_nature, libelle\_nature

FROM \_temp\_fr\_en\_etablissements\_ep

WHERE (code\_nature, libelle\_nature) IS NOT NULL;

delete from \_etablissement;

--- A compléter

/\*INSERT INTO \_etablissement (uai,nom\_etablissement,secteur,code\_postal,lattitude,longitude,code\_insee\_de\_la\_commune,nom\_de\_la\_commune,code\_academie,code\_nature)

select distinct uai, nom\_etablissement,code\_postal,latitude,longitude,nom\_commune,code\_academie,code\_nature

from \_temp\_fr\_en\_etablissements\_ep

where (uai, nom\_etablissement,code\_postal,latitude,longitude,nom\_commune,code\_academie,code\_nature) is not null;\*/

-- Mise en majuscules et suppression des accents

update \_temp\_fr\_en\_etablissements\_ep

set nom\_commune = upper(nom\_commune);

update \_temp\_fr\_en\_etablissements\_ep

set nom\_commune = replace(nom\_commune,'É','E');

update \_temp\_fr\_en\_etablissements\_ep

set nom\_commune = replace(nom\_commune,'È','E');

update \_temp\_fr\_en\_etablissements\_ep

set nom\_commune = replace(nom\_commune,'Ô','O');

INSERT INTO \_etablissement (uai, nom\_etablissement, secteur, code\_postal, lattitude, longitude,code\_insee\_de\_la\_commune,nom\_de\_la\_commune, code\_academie, code\_nature)

SELECT DISTINCT

e.uai,

e.nom\_etablissement,

e.statut\_public\_prive,

e.code\_postal,

e.latitude,

e.longitude,

e.code\_commune,

e.nom\_commune,

e.code\_academie,

e.code\_nature

FROM

\_temp\_fr\_en\_etablissements\_ep e

JOIN

\_commune c ON e.code\_commune = c.code\_insee\_de\_la\_commune

WHERE

(e.uai, e.nom\_etablissement, e.code\_postal, e.latitude, e.longitude, e.code\_academie,e.nom\_commune, e.code\_nature) IS NOT NULL;

--- et tout le reste

commit;