Manipulation intelligente des données : Dépendances Fonctionnelles

1. Objectif :

Développez les mécanismes (procédures et fonctions PL/SQL) qui permettent de vérifier les dépendances sémantiques (-DF- ; -EQ- ; -LE- ; -GE-) qui peuvent éventuellement exister entre les colonnes d'une table telles que

Algorithme simplifié pour vérifier la DF entre 2 colonnes :

- Eliminer les doubles et compter le nombre d'occurrences
  - Visualiser les données sans doublons (COL1, COL2)
  - Visualiser le comptage du nombre d'occurrence pour chaque valeur de COL1 Si le nombre d'occurrences maximal est supérieur à 1 Alors la DF n'est pas vérifiée !
- Tester la DF : COL1 -DF- COL2

2.Notre solution :

nous avons implémenté l'algorithme simplifié ci-dessous en 2 étapes :

- Une procédure MySql qui permet de déterminer les dépendances entre 2 colonnes : **VERIFDF** (tableName , DR_SemanticDependencies , LEFTCOL , RIGHTCOL , INOUT PERCENTAGE)
- Un script python qui parcourt toutes les colonnes d'une tables et déterminer toutes les dépendance entre les colonnes : **generateDependencies**(connection,table_name)

2.Exemple :

Soit la table suivante :

COL1	COL2	COL3	COL4	COL5
EPINAY-SUR-SEINE	FRANCE	1	1	EUROPE
PARIS	FRANCE	1	1	EUROPE
VILLETANEUSE	FRANCE	1	1	EUROPE
EPINAY-SUR-SEINE		1	1	EUROPE
EPINAY SUR SEINE	FRANCE	1	1	EUROPE
PARIS	FRANCE	1	1	EUROPE
PARIS		1	1	EUROPE
PARIS	FRANCE	1	1	EUROPE
MARSEILLE	FRANCE	1	1	EUROPE
ORLY-VILLE	FRANCE	1	1	EUROPE
MARSEILLE	FRANC	1	1	EUROPE
PARYS	FR	1	1	EUROPE
EPINAY-SUR-SEINE	FRANCE	1	1	EUROPE
PARIS	france	1	1	
Bruxelles	Belgique	1	1	europe
MOSCOU	RUSSIE	1	1	ASIE
ALGER	ALGERIE	1	1	AFRIQUE
RABAT	MAROC	1	1	afrique
TUNIS	TUNISIE	1	1	AFRIQUE

Les dépendances de cette table sont :

TableName	LEFTCOL	SEMANTICDEPENDENCY	RIGHTCOL	PERCENTAGE
TABLACONTROLER	col1	-DF-	col2	68
TABLACONTROLER	col1	-GE-	col3	100
TABLACONTROLER	col1	-GE-	col4	100
TABLACONTROLER	col1	-DF-	col5	89
TABLACONTROLER	col2	-DF-	col1	58
TABLACONTROLER	col2	-GE-	col3	89
TABLACONTROLER	col2	-GE-	col4	89
TABLACONTROLER	col2	-DF-	col5	89
TABLACONTROLER	col3	-DF-	col1	37
TABLACONTROLER	col3	-DF-	col2	53
TABLACONTROLER	col3	-EQ-	col4	100
TABLACONTROLER	col3	-DF-	col5	79
TABLACONTROLER	col4	-DF-	col1	37
TABLACONTROLER	col4	-DF-	col2	53
TABLACONTROLER	col4	-EQ-	col3	100
TABLACONTROLER	col4	-DF-	col5	79
TABLACONTROLER	col5	-DF-	col1	42
TABLACONTROLER	col5	-DF-	col2	58
TABLACONTROLER	col5	-GE-	col3	95
TABLACONTROLER	col5	-GE-	col4	95
TABLACONTROLER	col2,col3,col4,col5	-DF-	col1	44
TABLACONTROLER	col1,col3,col4,col5	-DF-	col2	58
TABLACONTROLER	col1,col2,col3,col4	-DF-	col5	84