

# RAPPORT PROJET OpenFoodFacts

**Binôme:**

**AFONOUVI Kékéli**

**THIAW Steicy**

**Professeur:**

**Bousse Marc**

**2022/2023**

## Modélisations

Pour aborder ce projet nous avons tout d'abord créé un diagramme de classes en fonction des informations que nous avons et de notre compréhension du sujet.

Pour cela nous avons réalisé deux cubes:

- **Cube Produit** : Contenant toutes les informations relatives aux produits de la BD avec les dimensions usage "Date Création", "Barcode" et "Créateur". (Voir figure 2)
- **Cube Version produit** : Concernant les informations relatives aux différentes versions de produits avec les dimensions usage "Date Création", "Date modification", "Barcode", "Nutrition", "Pnns" et "Correcteur". (Voir figure 1)

Ensuite nous avons traduit ces diagrammes de classe en schéma relationnel avec deux tables des faits **BD\_faits\_Version produits** et **BD\_faits\_Produits**. Ces deux tables sont reliées à des tables dimensionnelles en fonction de notre modélisation précédente.

**NB:** Nous avons décidé de dégrader les dimensions "Data création" et "Date modification" pour limiter le nombre de jointures et créer moins de tables

(Voir figure 3)

## Transformations Kettle de remplissage des tables

### Remplissage des tables dimensionnelles

Grâce à l'outil Pentaho Data Integration (Kettle) nous avons rempli nos 4 tables dimensionnelles (**Insertion nutrition**, **Insertion barcode**, **Insertion pnns**, **Insertion contributeur**) dans une même transformation (**Extraction\_tables\_dimensionnelles**) en suivant les étapes données en cours.

**Remarque** : Concernant la table **Insertion nutrition**, nous avons utilisé un code sql pour transformer la colonne **nutrition\_score\_fr** en **nutriscore (A, B, C, D...)**

(Voir fichier KTR)

### Remplissage des tables des faits

#### Table Produit

Nous avons commencé par remplir notre table des faits Produit à partir de la table OFF\_version\_produit se trouvant dans la base\_bousse. Cela s'est fait dans une transformation différente de celle des tables dimensionnelles (**transformation\_FaitProduit**)

**Étape 1** : nous avons extrait de la base\_bousse les différents champs qui nous intéressent.

**Étape 2** : nous avons fait un lien, grâce à l'étape 'Recherche valeur dans base de données' entre la table précédente et la table BD\_barcode se trouvant dans notre base de donnée pour récupérer les identifiants.

Nous avons répété l'étape 2 pour récupérer les autres identifiants

**Étape 3 :** nous avons sélectionné les champs qui nous intéressent avec '**Altération structure flux**' et nous avons inséré dans notre base de donnée

**(Voir fichier KTR)**

### **Table Version Produit**

Nous avons rempli notre table des faits Versions Produit à partir de la table OFF\_version\_produit se trouvant dans la base\_bousse. Cela s'est fait dans une transformation différente des transformations précédentes (**transformation\_FaitVersionProduit**)

**Étape 1 :** nous avons extrait de la base\_bousse les différents champs qui nous intéressent.

**Étape 2 :** dans un soucis d'harmonisation nous avons remplacé les valeur 'NULL' de pnns2 par 'unknown' grâce à l'étape '**Remplacer valeur nulle**'

**Étape 3 :** nous avons fait un lien, grâce à l'étape '**Recherche valeur dans base de données**' entre la table précédente et la table BD\_barcode se trouvant dans notre base de donnée pour récupérer les identifiants.

Nous avons répété l'étape 2 pour récupérer les autres identifiants

**Étape 4:** nous avons sélectionné les champs qui nous intéressent avec '**Altération structure flux**' et nous avons inséré dans notre base de donnée

**(Voir fichier KTR)**

### **Création d'une tâche (JOB)**

Pour remplir la base de données correctement nous avons d'abord besoin d'exécuter la transformation des tables dimensionnelles avant celle des tables de faits , de ce fait nous avons créé un Job afin d'exécuter toutes les transformations de façon séquentielle.

**(Voir fichier KTR)**

## **Code MDX des requêtes**

**Requête 1 : répartition des produits par contributeur («créateur») et année de création (colonnes)**

```
select
NON EMPTY
Hierarchize(Union({[Date Création].[Tous années]}, [Date Création].[Tous années].Children)) ON COLUMNS,
NON EMPTY
Hierarchize(Union({[Créateur].[Tous créateurs]}, [Créateur].[Tous créateurs].Children)) ON ROWS
from [Produits]
where ([Measures].[Nombre produits])
```

**Résultat :** (Voir figure 4)

Pour chaque créateur visible en ligne nous pouvons voir le nombre de produits qu'il a créé par an , ainsi que le nombre de produits créé toutes années confondues.

**Requête 2 : nombre de versions par mois (lignes) et pnns1 (colonnes) des produits créés en 2017.**

```
select  
NON EMPTY  
{[Pnns].[pnns1].Members} ON COLUMNS,  
NON EMPTY  
{[Date Modification].[Année_Modification].Members} ON ROWS  
from [Versions Produits]  
WHERE ( [Measures].[Nombre Versions Produits],[Année_Creation].[2017])
```

**Résultat: (Voir figure 5)**

D'après notre compréhension de la requête , il s'agit d'avoir le nombre de versions de produit par mois indépendamment de l'année mais nous n'arrivons pas à concaténer par les mois indifféremment des années. Nous avons donc décidé de regrouper les mois concernant une année de modification sous l'année en question.

## **Autres requêtes choisies**

**Requête 3 : nombre de versions faites par le contributeurs openfoodfact en 2017 par pnns1 (en lignes)**

```
select [Measures].[Nombre Versions Produits] ON COLUMNS,  
[Pnns].[Pnns1].Members ON ROWS  
FROM [Versions Produits]  
WHERE([Correcteur].[openfoodfacts-contributors])
```

**Résultat: (Voir figure 6)**

Cette requête est intéressante car elle nous permet de concentrer nos recherche sur un contributeur en particulier et de connaitre son impacte sur la base open food facts

**Requête 4 : répartition du nombre de versions de produits par pnns(lignes) et par nutri score(colonnes) en 2018**

```
select [Nutrition].[Nutri Score].Members ON COLUMNS,  
[Pnns].[Pnns1].Members ON ROWS  
FROM [Versions Produits]  
WHERE([Date Modification].[2018])
```

Cette requête est intéressante car elle nous permet , pour une date donnée (2018) de visualiser l'évolution du nutriscore de chaque produit.

**Résultat: (Voir figure 7)**

# Annexes

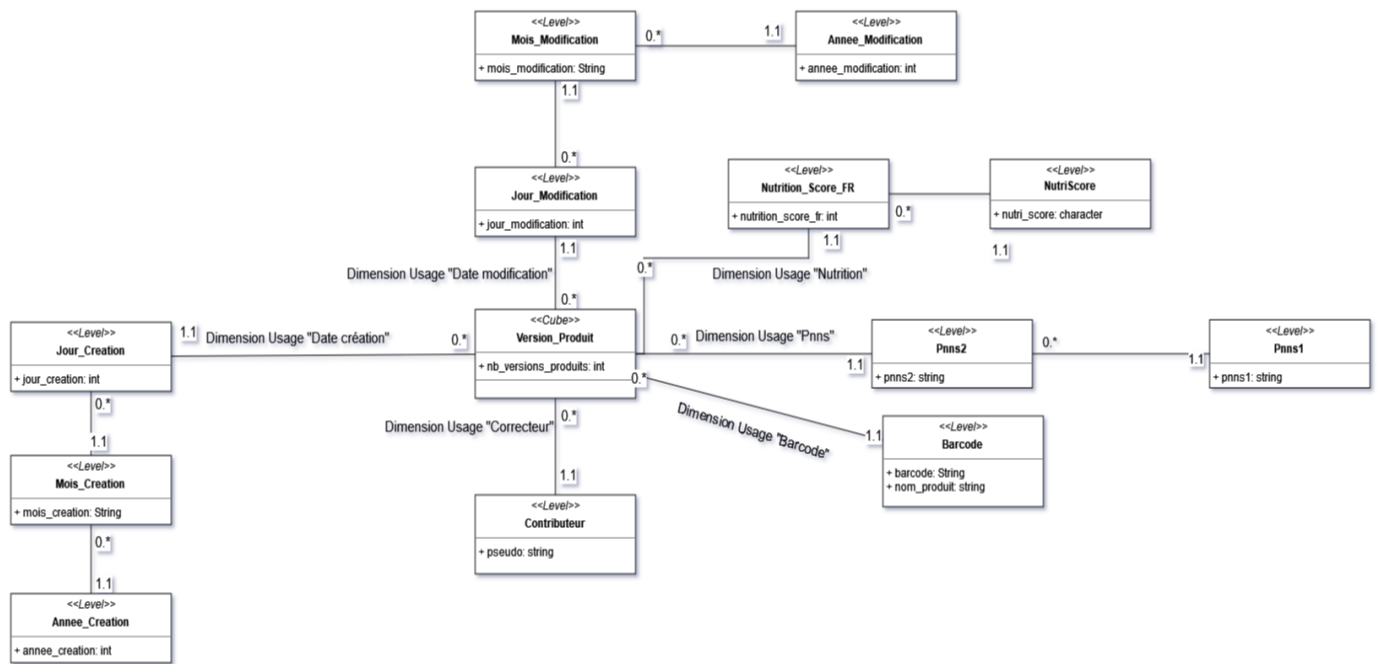


Figure 1: diagramme de classe VersionProduits

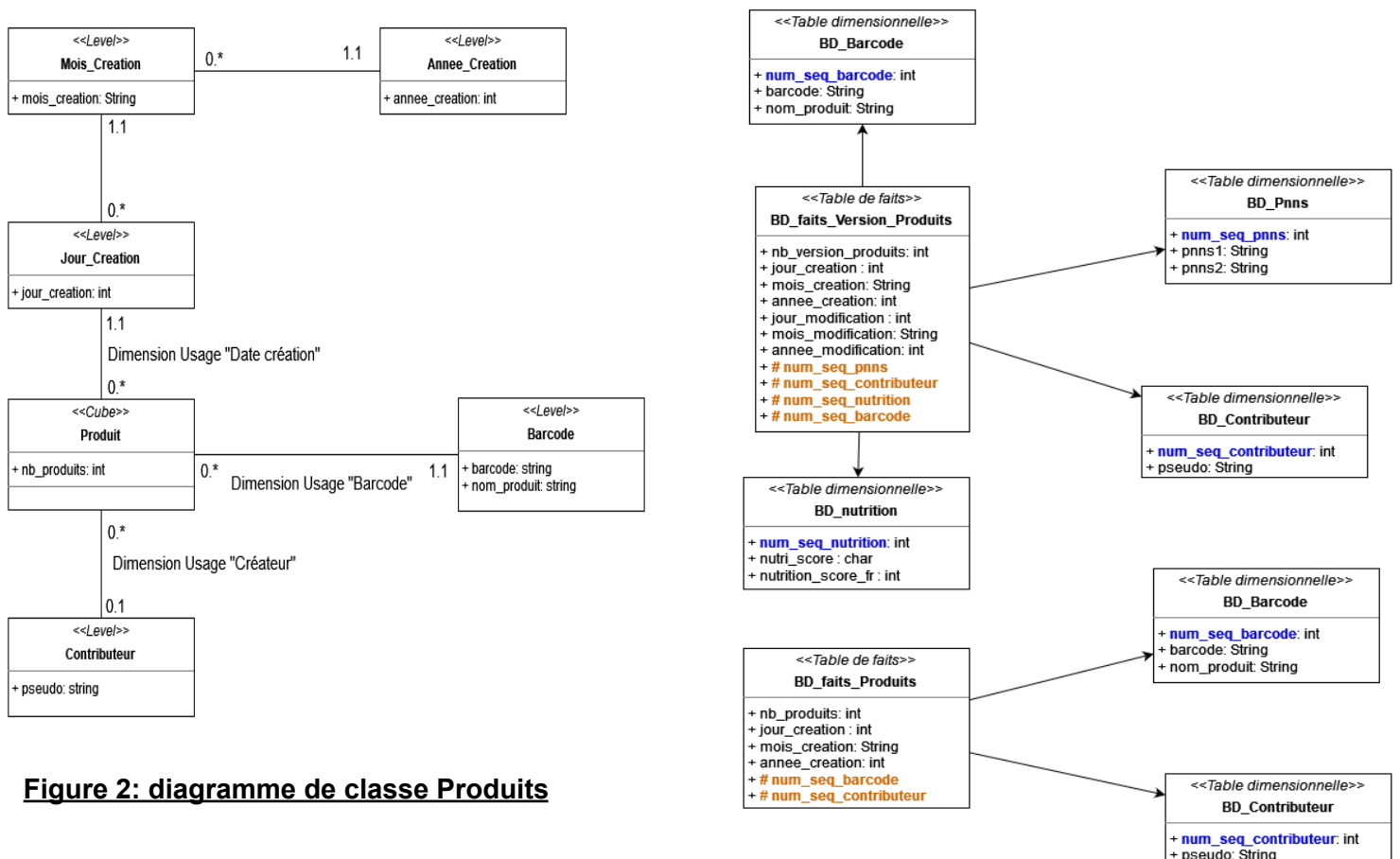


Figure 2: diagramme de classe Produits

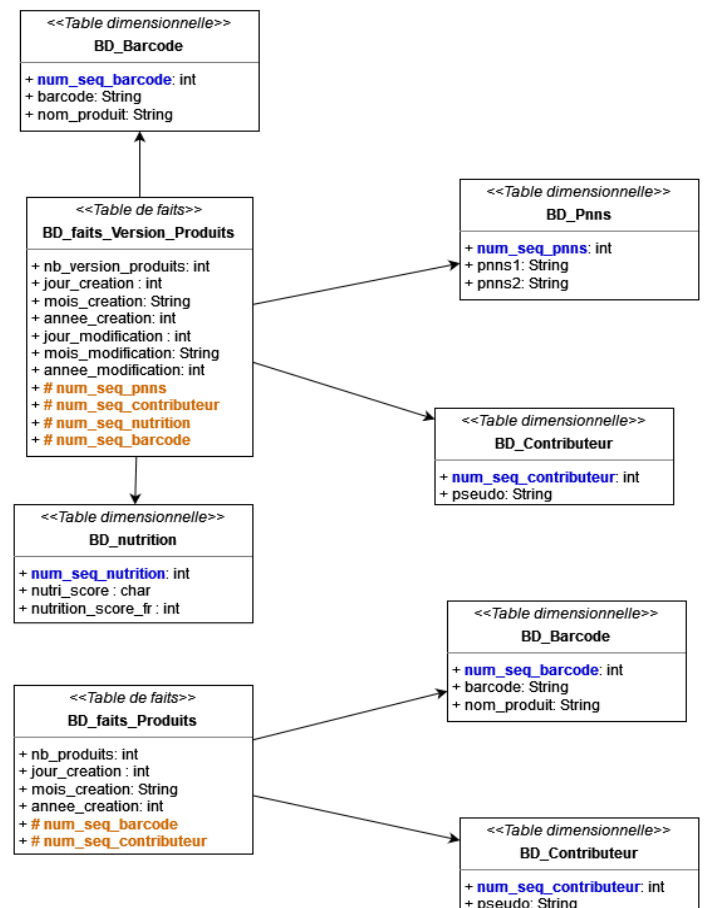


Figure 3: schéma relationnel

	Date Création						
Créateur	<input type="checkbox"/> Tous années	<input type="checkbox"/> 2013	<input type="checkbox"/> 2014	<input type="checkbox"/> 2015	<input type="checkbox"/> 2016	<input type="checkbox"/> 2017	<input type="checkbox"/> 2018
<input checked="" type="checkbox"/> Tous créateurs	404	3	3	36	22	139	201
agamitsudo	1				1		
aleene	4			1	3		
andre	2	2					
aristoi	1		1				
back94	1					1	
beniben	2					1	1
beretman73	1						1
bouf	1						1
claire-87	1						1
daffyl8	1						1

**Figure 4: Résultat requête 1**

Date Modification	Pnns									
	<input type="checkbox"/> Beverages	<input type="checkbox"/> Cereals and potatoes	<input type="checkbox"/> Composite foods	<input type="checkbox"/> Fat and sauces	<input type="checkbox"/> Fish Meat Eggs	<input type="checkbox"/> Fruits and vegetables	<input type="checkbox"/> Milk and dairy products	<input type="checkbox"/> Salty snacks	<input type="checkbox"/> Sugary snacks	<input type="checkbox"/> Unknown
<input checked="" type="checkbox"/> 2017	4	1	6	2	15	2	7	2	2	140
<input type="checkbox"/> April					2	1	1	1		3
<input type="checkbox"/> August					1		1		1	25
<input type="checkbox"/> December	1		2	1	3				2	27
<input type="checkbox"/> February			2		4				1	4
<input type="checkbox"/> January		1	2		2		1	1	1	3
<input type="checkbox"/> July	1		1		3					11
<input type="checkbox"/> June					3		1		1	12
<input type="checkbox"/> March	1				1					1
<input type="checkbox"/> May			2		1	1	1		1	13
<input type="checkbox"/> November				1	1					19
<input type="checkbox"/> October	1		3		3	2	5			31
<input type="checkbox"/> September	3				2	1	4		1	41
<input checked="" type="checkbox"/> 2018	3		6		11	3	7		6	50
<input type="checkbox"/> April					2	1	1	1		3
<input type="checkbox"/> August					1		1		1	25
<input type="checkbox"/> February			2		4				1	4
<input type="checkbox"/> January		1	2		2		1	1	1	3
<input type="checkbox"/> July	1		1		3					11
<input type="checkbox"/> June					3		1		1	12
<input type="checkbox"/> March	1				1					1
<input type="checkbox"/> May			2		1	1	1		1	13
<input type="checkbox"/> October	1		3		3	2	5			31
<input type="checkbox"/> September	3				2	1	4		1	41

Slicer: [Mesure=Nombre Versions Produits] [Année\_Creation=2017]

**Figure 5: Résultat requête 2**

	Mesures
Pnns	⊗ Nombre Versions Produits
<input type="checkbox"/> Alcoholic beverages	
<input type="checkbox"/> Beverages	26
<input type="checkbox"/> Cereals and potatoes	25
<input type="checkbox"/> Composite foods	34
<input type="checkbox"/> Fat and sauces	10
<input type="checkbox"/> Fish Meat Eggs	36
<input type="checkbox"/> Fruits and vegetables	25
<input type="checkbox"/> Milk and dairy products	44
<input type="checkbox"/> Salty snacks	13
<input type="checkbox"/> Sugary snacks	63
<input type="checkbox"/> Tripe dishes	
<input type="checkbox"/> unknown	76

Slicer: [Correcteur=openfoodfacts-contributors]

**Figure 6: Résultat requête 3**

	Nutrition					
Pnns	#null	A	B	C	D	E
<input type="checkbox"/> Alcoholic beverages						
<input type="checkbox"/> Beverages			2		1	
<input type="checkbox"/> Cereals and potatoes			4	4	2	
<input type="checkbox"/> Composite foods			4	6	1	
<input type="checkbox"/> Fat and sauces					4	6
<input type="checkbox"/> Fish Meat Eggs				3		7
<input type="checkbox"/> Fruits and vegetables			2	1		
<input type="checkbox"/> Milk and dairy products			12	2	4	
<input type="checkbox"/> Salty snacks					3	
<input type="checkbox"/> Sugary snacks					5	8
<input type="checkbox"/> Tripe dishes						
<input type="checkbox"/> unknown			1	2	2	

Slicer: [Année\_Modification=2018]

**Figure 7: Résultat requête 4**