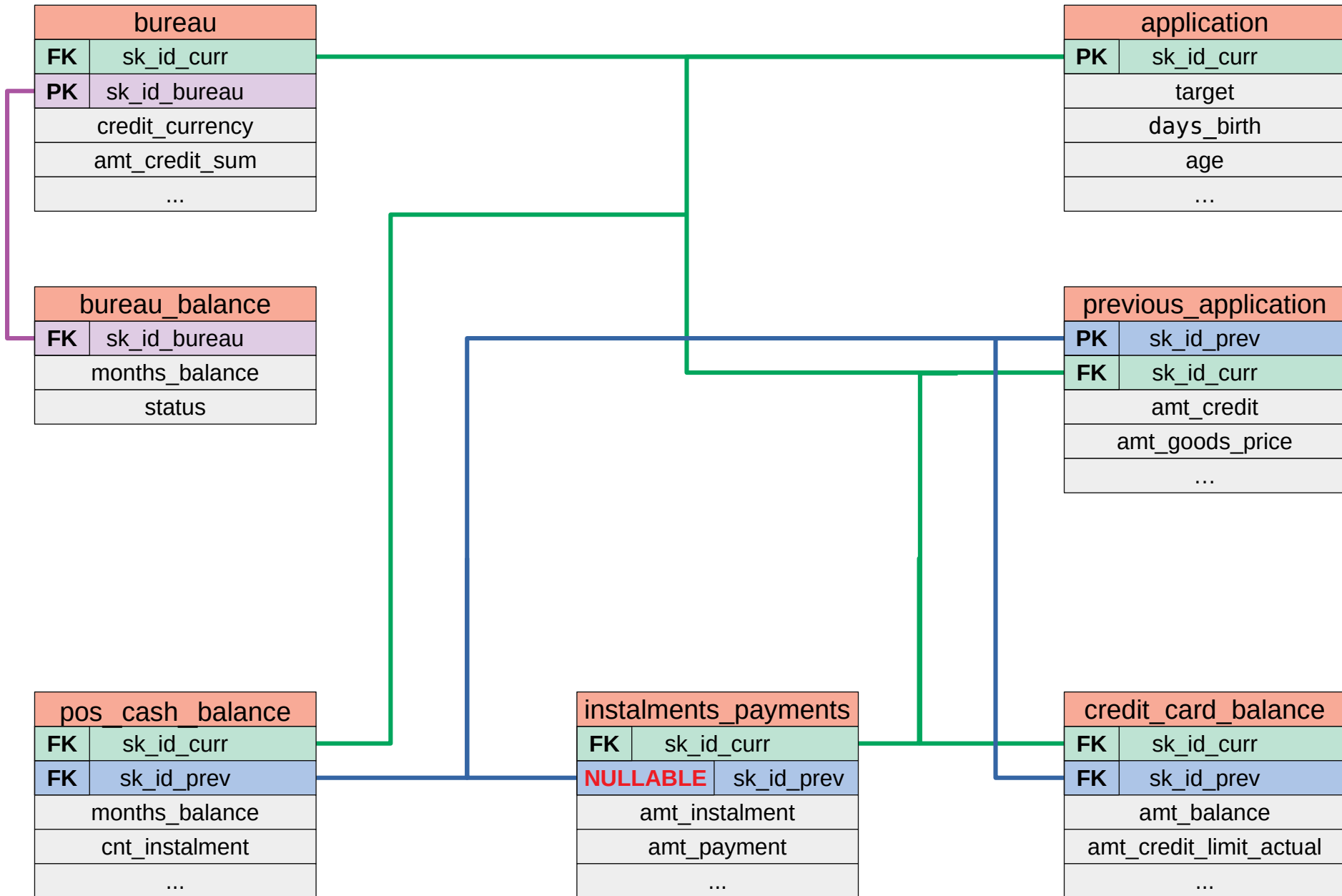


Initial schema of HomeCredit Competition DB



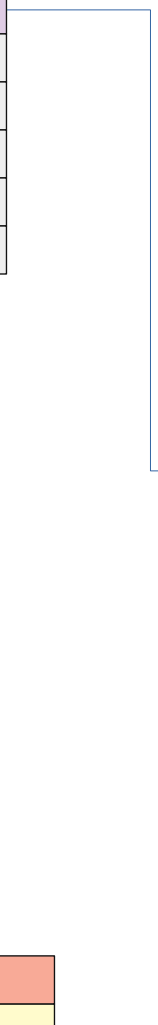
| calendar | |
|--------------|-----------------|
| PK | date_de_demande |
| year | |
| month | |
| day | |
| weekday | |
| week_of_year | |

| client | |
|----------------|-----------|
| PK | id_client |
| birthday | |
| gender | |
| education_type | |

| demande_de_credit | |
|-------------------|-----------------|
| PK | id_demande |
| FK | id_client |
| FK | date_de_demande |
| FK | type_de_credit |
| FK | type_de_achat |
| prix_de_achat | |
| montant_demande | |
| montant_credit | |
| type_accompagne | |
| status | |

| credit_types | |
|--------------|---------|
| PK | id_type |
| nom_de_type | |

| achat_types | |
|-------------|---------|
| PK | id_type |
| nom_de_type | |



| Table name | lines | columns |
|--------------------------|-----------|---------|
| application_test.csv | 48 741 | 122 |
| application_train.csv | 307 511 | 123 |
| previous_application.csv | 1 670 214 | 38 |

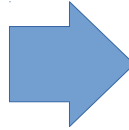


| Table name | lines | columns |
|-------------------|-----------|---------|
| calendar | 4255 | 6 |
| credit_types | 4 | 2 |
| demande_de_credit | 1 670 214 | 10 |
| achat_types | 28 | 2 |
| client | 356 255 | 4 |

Schéma d'Architecture Logique

Dossier avec des fichiers *.csv 

| Table name | lines | columns |
|--------------------------|-----------|---------|
| application_test.csv | 48 741 | 121 |
| application_train.csv | 307 511 | 122 |
| previous_application.csv | 1 670 214 | 37 |



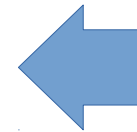
ETL



- lire le dossier avec les fichiers et insérer le nom des fichiers dans la table temporaire
- lire la première ligne des fichiers, divisée par des virgules et créer des tables avec colonnes du type texte
- lire les données des fichiers csv et remplir les tableaux



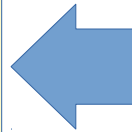
DB Prod



Extrait les données pour le projet



- Scripts SQL pour la création de tables finales basées sur les données brutes .
- Transformation de données
- Nettoyage les tables supplémentaires



Transformation finale de la base de données BI



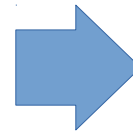
- Convertir les colonnes en un type de données approprié
- Ajouter des index et des contraintes



DB BI



| Table name | lines | columns |
|-------------------|-----------|---------|
| calendar | 4255 | 6 |
| credit_types | 4 | 2 |
| demande_de_credit | 1 670 214 | 10 |
| achat_types | 28 | 2 |
| client | 356 255 | 4 |



Requêtes SQL



Résultats



