

Avancement SAE

S2.01 Conception et implémentation d'une base de données

Ibrahim BENKHERFELLAH Axel COULET

Université Sorbonne Paris Nord
BUT1 SD Semestre 2

June 1, 2025

Table des matières

- 1 Introduction
- 2 Exploration et compréhension des données
- 3 Modélisation de la base de données
- 4 Script de création et peuplement SQL
- 5 Interrogation et visualisation des données
- 6 Conclusion

Objectif du projet

- Comprendre et modéliser le commerce des technologies à faible émission de carbone.
- Concevoir une base de données normalisée à partir de deux jeux de données CSV.
- Interroger et visualiser les données pour produire des analyses pertinentes.

Objectif du projet

- Comprendre et modéliser le commerce des technologies à faible émission de carbone.
- Concevoir une base de données normalisée à partir de deux jeux de données CSV.
- Interroger et visualiser les données pour produire des analyses pertinentes.

Objectif du projet

- Comprendre et modéliser le commerce des technologies à faible émission de carbone.
- Concevoir une base de données normalisée à partir de deux jeux de données CSV.
- Interroger et visualiser les données pour produire des analyses pertinentes.

- Fichiers CSV étudiés :
 - `Trade_in_Low_Carbon_Technology_Products.csv`
 - `Bilateral_Trade_in_Low_Carbon_Technology_Products.csv`
- Exploration initiale avec Python (pandas) :
`.head()`, `.info()`, etc.
- Identification de colonnes clés : `CTS_Code`, `Indicator`, `Trade_Flow`, etc.
- Difficultés rencontrées :
 - Colonnes de dates (comme `Year`) au mauvais format ou non reconnues comme `Date`.
 - Colonnes "`F1994`" à "`F2023`" en format long à dépivoter (via `UNPIVOT` de SQL).

- Fichiers CSV étudiés :
 - `Trade_in_Low_Carbon_Technology_Products.csv`
 - `Bilateral_Trade_in_Low_Carbon_Technology_Products.csv`
- Exploration initiale avec Python (pandas) :
`.head()`, `.info()`, etc.
- Identification de colonnes clés : `CTS_Code`, `Indicator`, `Trade_Flow`, etc.
- Difficultés rencontrées :
 - Colonnes de dates (comme `Year`) au mauvais format ou non reconnues comme `Date`.
 - Colonnes "`F1994`" à "`F2023`" en format long à dépivoter (via `UNPIVOT` de SQL).

- Fichiers CSV étudiés :
 - `Trade_in_Low_Carbon_Technology_Products.csv`
 - `Bilateral_Trade_in_Low_Carbon_Technology_Products.csv`
- Exploration initiale avec Python (pandas) :
`.head()`, `.info()`, etc.
- Identification de colonnes clés : `CTS_Code`, `Indicator`, `Trade_Flow`, etc.
- Difficultés rencontrées :
 - Colonnes de dates (comme `Year`) au mauvais format ou non reconnues comme `Date`.
 - Colonnes "`F1994`" à "`F2023`" en format long à dépivoter (via `UNPIVOT` de SQL).

- Fichiers CSV étudiés :
 - `Trade_in_Low_Carbon_Technology_Products.csv`
 - `Bilateral_Trade_in_Low_Carbon_Technology_Products.csv`
- Exploration initiale avec Python (pandas) :
`.head()`, `.info()`, etc.
- Identification de colonnes clés : `CTS_Code`, `Indicator`, `Trade_Flow`, etc.
- Difficultés rencontrées :
 - Colonnes de dates (comme `Year`) au mauvais format ou non reconnues comme `Date`.
 - Colonnes "`F1994`" à "`F2023`" en format long à dépivoter (via `UNPIVOT` de SQL).

- Données structurées en colonnes d'années : F1994 à F2023.
- Présence de colonnes de description redondantes (CTS_Code, CTS_Name, CTS_Full_Descriptor).
- Première étape : lecture manuelle pour repérage des redondances et structures tabulaires.

- Données structurées en colonnes d'années : F1994 à F2023.
- Présence de colonnes de description redondantes (CTS_Code, CTS_Name, CTS_Full_Descriptor).
- Première étape : lecture manuelle pour repérage des redondances et structures tabulaires.

- Données structurées en colonnes d'années : F1994 à F2023.
- Présence de colonnes de description redondantes (CTS_Code, CTS_Name, CTS_Full_Descriptor).
- Première étape : lecture manuelle pour repérage des redondances et structures tabulaires.

- Conversion des colonnes FXXXX en format court (année, valeur).
- Normalisation des années (F1994 à F2023) :
 - Chaque colonne FXXXX devient une ligne avec une valeur d'année.
 - Transformation effectuée via `UNION ALL` dans SQL pour correspondre à la logique de `stack()` en Python.
- Problèmes rencontrés :
 - Ambiguïté sur le pays origine.
 - Duplication de lignes si filtrage incorrect.

- Conversion des colonnes FXXXX en format court (année, valeur).
- Normalisation des années (F1994 à F2023) :
 - Chaque colonne FXXXX devient une ligne avec une valeur d'année.
 - Transformation effectuée via `UNION ALL` dans SQL pour correspondre à la logique de `stack()` en Python.
- Problèmes rencontrés :
 - Ambiguïté sur le pays origine.
 - Duplication de lignes si filtrage incorrect.

- Conversion des colonnes FXXXX en format court (année, valeur).
- Normalisation des années (F1994 à F2023) :
 - Chaque colonne FXXXX devient une ligne avec une valeur d'année.
 - Transformation effectuée via `UNION ALL` dans SQL pour correspondre à la logique de `stack()` en Python.
- Problèmes rencontrés :
 - Ambiguïté sur le pays origine.
 - Duplication de lignes si filtrage incorrect.

Transformation des données : format large vers format long

Format large (wide) :

Chaque année est une colonne

...	Scale	F1994	F1995	F1996
...	Units	190885.0	2433262.0	17679060.0
...	Units	0.0004146	0.0053298	0.0347023
...	Units	668268200.0	612865500.0	493126000.0



Format long (tidy) :

Chaque ligne correspond à une observation unique (année, valeur).

...	Scale	Year	Value
...	Units	F1994	190885.0
...	Units	F1994	0.0004146
...	Units	F1994	668268200.0

- Création d'une table `Bilateral_Trade` pour les données bilatérales :
 - Relation réflexive entre deux pays (deux clés étrangères vers `country`).
- Table `Trade` pour les données nationales.
- Lien systématique aux dimensions : pays, indicateur, catégorie, année.

- Création d'une table `Bilateral_Trade` pour les données bilatérales :
 - Relation réflexive entre deux pays (deux clés étrangères vers `country`).
- Table `Trade` pour les données nationales.
- Lien systématique aux dimensions : pays, indicateur, catégorie, année.

- Création d'une table `Bilateral_Trade` pour les données bilatérales :
 - Relation réflexive entre deux pays (deux clés étrangères vers `country`).
- Table `Trade` pour les données nationales.
- Lien systématique aux dimensions : pays, indicateur, catégorie, année.

Schéma Entité-Association

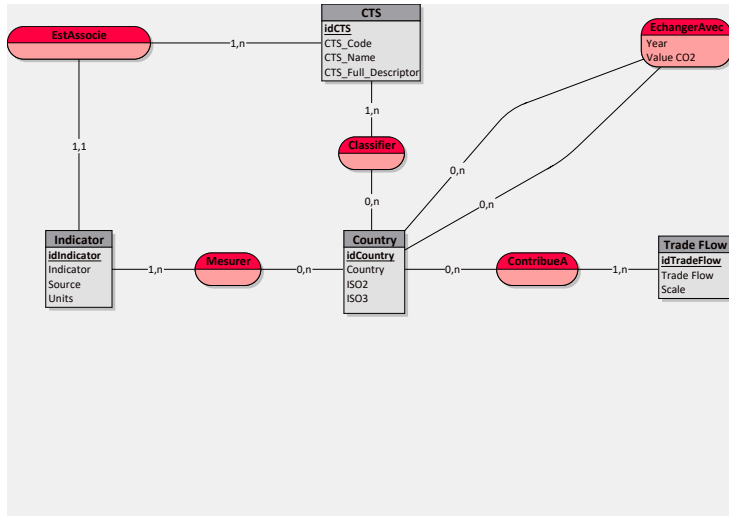


Figure: Schéma Entité-Association Initial

Schéma Entité-Association

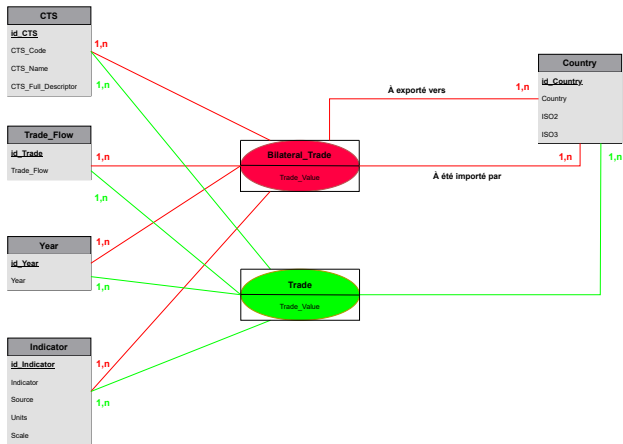


Figure: Schéma Entité-Association Final

Modèle conceptuel (EA)

- Type Entités : Country, Indicator, CTS, Trade_Flow, Year
- Type Associations :
 - Bilateral_Trade (réflexive sur Country)
 - Trade (relation simple avec agrégation)
- Justification des cardinalités et des relations

Modèle conceptuel (EA)

- Type Entités : Country, Indicator, CTS, Trade_Flow, Year
- Type Associations :
 - Bilateral_Trade (réflexive sur Country)
 - Trade (relation simple avec agrégation)
- Justification des cardinalités et des relations

- Type Entités : Country, Indicator, CTS, Trade_Flow, Year
- Type Associations :
 - Bilateral_Trade (réflexive sur Country)
 - Trade (relation simple avec agrégation)
- Justification des cardinalités et des relations

- Schéma relationnel résultant de la modélisation EA :
 - Country(idCountry, Country, ISO2, ISO3)
 - CTS(idCTS, CTS_Code, CTS_Name, CTS_Full_Descriptor)
 - Trade_Flow(id_Trade_Flow, Trade_Flow, Scale)
 - Indicator(idIndicator, Indicator, Source, Units)
 - Year(id_Year, Year)
 - Bilateral_Trade(id_country, id_counterpart, id_indicator, id_cts, id_Trade_Flow, id_year, trade_value)
 - Trade(id_country, id_indicator, id_cts, id_Trade_Flow, id_year, trade_value)

Forme normale : 1NF

- Toutes les valeurs sont atomiques : aucun champ multivalué ou composé.
- Transformation des colonnes F1994 à F2023 en une colonne `id_year` via pivot SQL.
- Tables relationnelles sans redondance horizontale.
- **Conclusion** : le schéma respecte la **première forme normale (1NF)**.

- Le schéma est en 1NF.
- Aucune dépendance fonctionnelle partielle dans les tables à clés composites.
- Exemple : les descripteurs CTS ont été extraits dans une table distincte, reliée par CTS_Code.
- **Conclusion** : toutes les dépendances fonctionnelles concernent la clé entière → **2NF validée**.

- Le schéma est en 2NF.
- Suppression des dépendances transitives.
- Exemple : l'unité d'un indicateur dépend uniquement de l'indicateur, pas d'un pays ou autre entité.
- **Conclusion** : toutes les colonnes non-clés dépendent uniquement de la clé primaire → **schéma en 3NF**.

- Toutes les dépendances fonctionnelles ont un antécédent qui est une super-clé.
- Exemple : dans `Bilateral_Trade`, seule la combinaison complète des identifiants détermine `trade_value`.
- Il n'existe pas de dépendance fonctionnelle violant cette condition.
- **Conclusion** : le schéma respecte la forme normale de **Boyce-Codd (BCNF)**.

Création de la table Country

```
CREATE TABLE Country(  
    id_Country SMALLSERIAL PRIMARY KEY,  
    Country VARCHAR(50),  
    ISO2 CHAR(2),  
    ISO3 CHAR(3)  
);
```

Création de la table Indicator

```
CREATE TABLE Indicator(  
    id_Indicator SMALLSERIAL PRIMARY KEY,  
    Indicator VARCHAR(80),  
    Source VARCHAR,  
    Units VARCHAR(50),  
    Scale VARCHAR(5)  
);
```

Création de la table CTS

```
CREATE TABLE CTS(  
    id_CTS SMALLSERIAL PRIMARY KEY,  
    CTS_Code VARCHAR(6),  
    CTS_Name VARCHAR(100),  
    CTS_Full_Descriptor VARCHAR(150)  
);
```


Création de la table Trade_Flow

```
CREATE TABLE Trade_Flow(  
    id_Trade_Flow SMALLSERIAL PRIMARY KEY,  
    Trade_Flow VARCHAR(20)  
);
```

Création de la table Year

```
CREATE TABLE Year(  
    id_Year SMALLSERIAL PRIMARY KEY,  
    Year DATE  
);
```

Création de la table Bilateral_Trade

```
CREATE TABLE Bilateral_Trade (  
    id_Country SMALLINT REFERENCES Country(id_Country),  
    id_Counterpart SMALLINT REFERENCES Country(id_Country),  
    id_Indicator SMALLINT REFERENCES Indicator(id_Indicator),  
    id_CTS SMALLINT REFERENCES CTS(id_CTS),  
    id_Trade_Flow SMALLINT REFERENCES Trade_Flow(...),  
    id_Year SMALLINT REFERENCES Year(id_Year),  
    trade_value DOUBLE PRECISION,  
    PRIMARY KEY(id_Country, id_Counterpart_Country,  
                id_Indicator, id_CTS, id_Trade_Flow, id_Year)  
);
```

Création de la table Trade

```
CREATE TABLE Trade (  
    id_Country SMALLINT REFERENCES Country(id_Country),  
    id_Indicator SMALLINT REFERENCES Indicator(id_Indicator),  
    id_CTS SMALLINT REFERENCES CTS(id_CTS),  
    id_Trade_Flow SMALLINT REFERENCES Trade_Flow(...),  
    id_Year SMALLINT REFERENCES Year(id_Year),  
    trade_value DOUBLE PRECISION,  
    PRIMARY KEY(id_Country, id_Indicator, id_CTS,  
    id_Trade_Flow, id_Year)  
);
```

Peuplement de la base

- Utilisation de COPY, INSERT INTO avec jointures.
- Alignement avec les traitements Python (Avec `.stack()`).
- Cas particuliers traités :
 - Échanges d'un pays avec lui-même
 - Lignes NULL / Trade Flow = Not Applicable
- Difficultés rencontrées :
 - Contraintes techniques : NULL, unicité, jointures complexes.
 - Problèmes de données : fusion sans doublons, nettoyage, volume élevé.

- Utilisation de COPY, INSERT INTO avec jointures.
- Alignement avec les traitements Python (Avec `.stack()`).
- Cas particuliers traités :
 - Échanges d'un pays avec lui-même
 - Lignes NULL / Trade Flow = Not Applicable
- Difficultés rencontrées :
 - Contraintes techniques : NULL, unicité, jointures complexes.
 - Problèmes de données : fusion sans doublons, nettoyage, volume élevé.

- Top 10 des pays importateurs de technologies en provenance de Chine (2021)
- Répartition mondiale des exportations de technologies en 2021 (%)
- Part annuelle des échanges bilatéraux vs nationaux dans le commerce mondial (1994 à 2023)
- Nombre de pays importateurs depuis la Chine (2021)
- Nombre de flux d'exportation de la Chine enregistrés (2021)
- Valeur totale des exportations de la Chine (2021)

- Utilisation de Metabase pour interroger la BD
- Utilisation de Tableau pour valider les valeurs issues du CSV
- Comparaison SQL / Python → vérification de l'intégrité
- Exemples visuels :

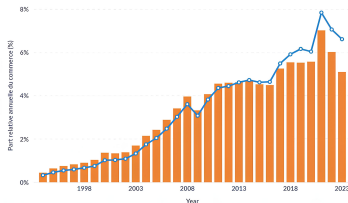
Tableau de Bord - Étude sur la Chine

2021 : l'année de la domination chinoise dans le commerce vert

Comment les flux d'exportation de technologies bas carbone se sont-ils structurés à l'échelle mondiale en 2021, et quel a été le rôle stratégique de la Chine dans cette dynamique commerciale ?

Part annuelle des échanges bilatéraux vs nationaux dans le commerce mondial (1994 à 2023)

● Flux Nationaux ● Flux (Bilatéraux)



399 Exports

Nombre de flux d'exportation de la Chine enregistrés (2021)

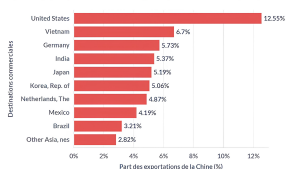
215 Pays

Nombre de pays importateurs depuis la Chine (2021)

258,7B \$

Valeur totale des exportations de la Chine (2021)

Top 10 des pays importateurs de technologies en provenance de Chine (2021)



Répartition mondiale des exportations de technologies en 2021 (%)

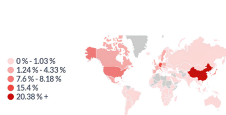


Figure: 2021 : L'année de la domination chinoise dans le commerce vert

- Base de données fonctionnelle et fidèle au CSV (dans nos rêves)
- Respect des étapes du processus de modélisation
- Difficultés sur l'import (pivot, valeurs nulles, sens du flux)
- Vérification croisée avec Python
- Améliorations possibles :

Merci pour votre attention

Questions ?