

Data WareHouse

Projet



Réalisé par :

Elghamch Zakaria (WISD)

Kamil Hamza (MASD)

Encadré par :

Pr. Guénaël Cabanes



I. La migration : Problème posé

La migration est un véritable problème auquel sont confrontés les pays.

La réponse à la question du motif de l'émigration peut être évidente à première vue,

mais avec l'exacerbation des données et la multiplicité des facteurs et aspects qui peuvent être utilisés pour étudier ce phénomène, nous recherchons désormais une étude scientifique pour obtenir une réponse claire.

La question posée est de bien comprendre ce phénomène à travers les données disponibles et de trouver les facteurs les plus contributifs à l'augmentation de ce phénomène.

Nous souhaitons mettre en place une solution logicielle permettant d'intégrer les données pertinentes et de pouvoir les interroger efficacement sous des angles divers.

II. Données disponibles :

Nous avons une base de données qui contient des données sur l'immigration (de l'année 2019), lorsque vous ouvrez ce fichier, vous verrez une description des données disponibles.

Vous trouverez trois tableaux, chacun contenant des données de migration spécifiques :

Migration de pays, migration de compétences, Migration de l'industrie.

Et un quatrième fichier (csv) contient *la section de chaque industrie*.

- ✓ Voir le dossier : [csv_files](#)
pour voir la base de données de notre sujet.

III. Étude des besoins :

Notre objectif principal dans ce projet, comme nous l'avons mentionné, est de trouver les facteurs contribuant à l'exacerbation de ce phénomène pour une bonne compréhension de la situation.

Sur une étude des besoins de notre problème et sur la base des données disponibles, Nous proposons les requêtes cibles suivantes :

- ✓ Quels sont les facteurs qui expliquent la migration ?

1. Économiques :

Requête cible :

Nombre d'immigrés
/ GNP
/ CAPITAL

2. Géographiques :

Requête cible :

Nombre d'immigrés
/ continent
/ region

3. Culturels :

Requête cible :

Nombre d'immigrés
/ language

4. Sociologiques :

Requête cible :

```
Nombre d'immigrés  
/ population  
/ surfacearea  
/ indepyear
```

5. Politiques :

Requête cible :

```
Nombre d'immigrés  
/ gouvernement form
```

✓ *Quelles catégories de compétences et les compétences migrent le plus ?*

Requête cible :

```
Nombre d'immigrés  
/ category_skill  
/ skillname
```

✓ *Quels secteurs de l'industrie et les industries migrent le plus ?*

Requête cible :

```
Nombre d'immigrés  
/ isic_section  
/ industry_name
```

IV. Étude des données :

1. Synthèse des données :

Les quatre fichiers csv dans le dossier [csv_files](#), chacun contenant des données de migration spécifiques :

<i>Nom du fichier (csv)</i>	<i>Contenu</i>
<i>country_migration</i>	<i>Le nombre d'immigrants entre les pays</i>
<i>industry_migration</i>	<i>Nombre d'immigrants industriels de chaque pays</i>
<i>skills_migration</i>	<i>Nombre d'immigrants par compétence par pays</i>
<i>linkedin_to_isic_rev_4_industry_mapping_0</i>	<i>Sections de chaque industrie</i>

D'autre part, on a deux tables sur une base de données, nommée :

Country : une table contient toutes les informations sur tous les pays du monde

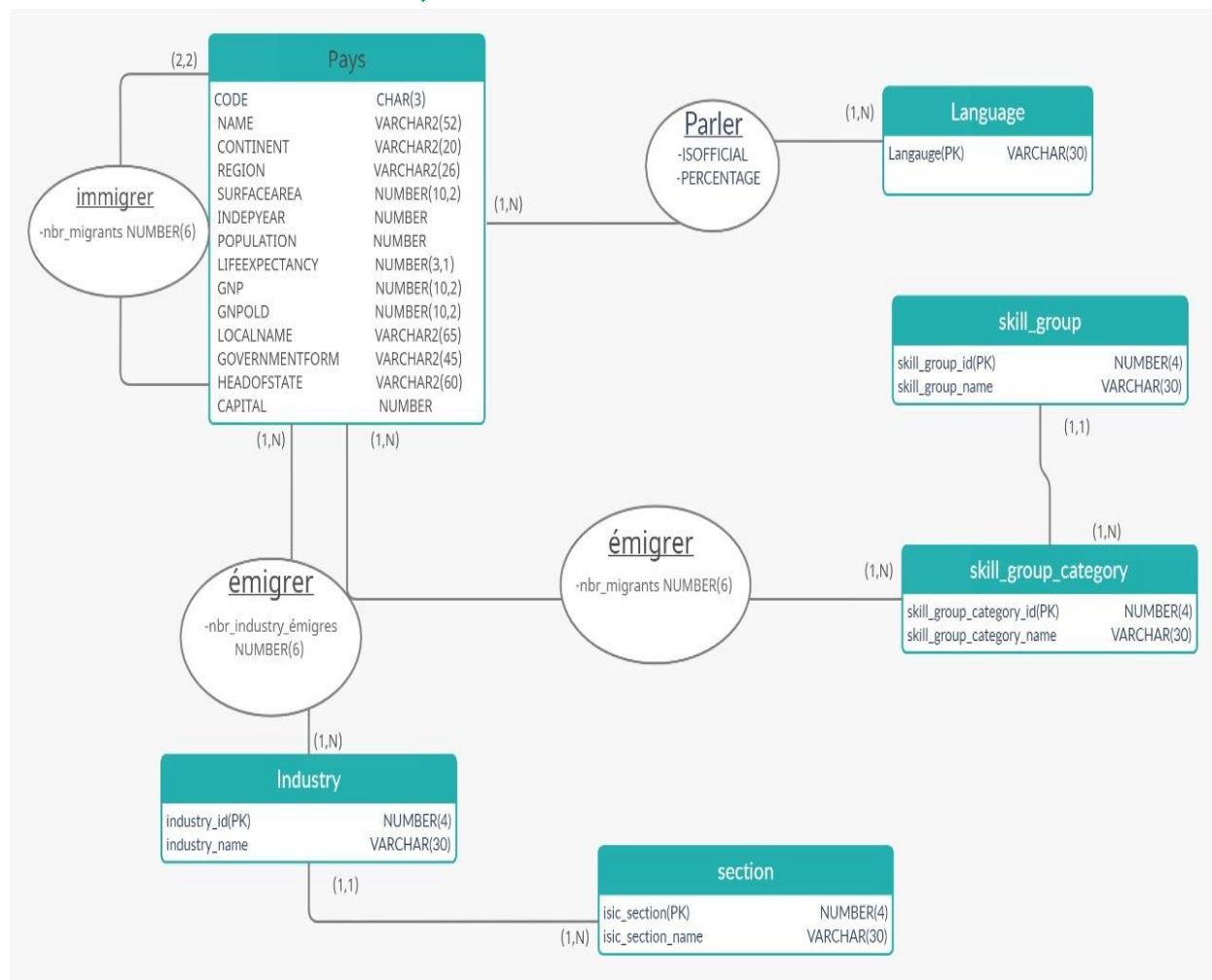
Countrylanguage : une table contient les langues communes dans chaque pays avec le pourcentage d'utilisation de cette langue.

✓ Voir le fichier [world.sql](#)

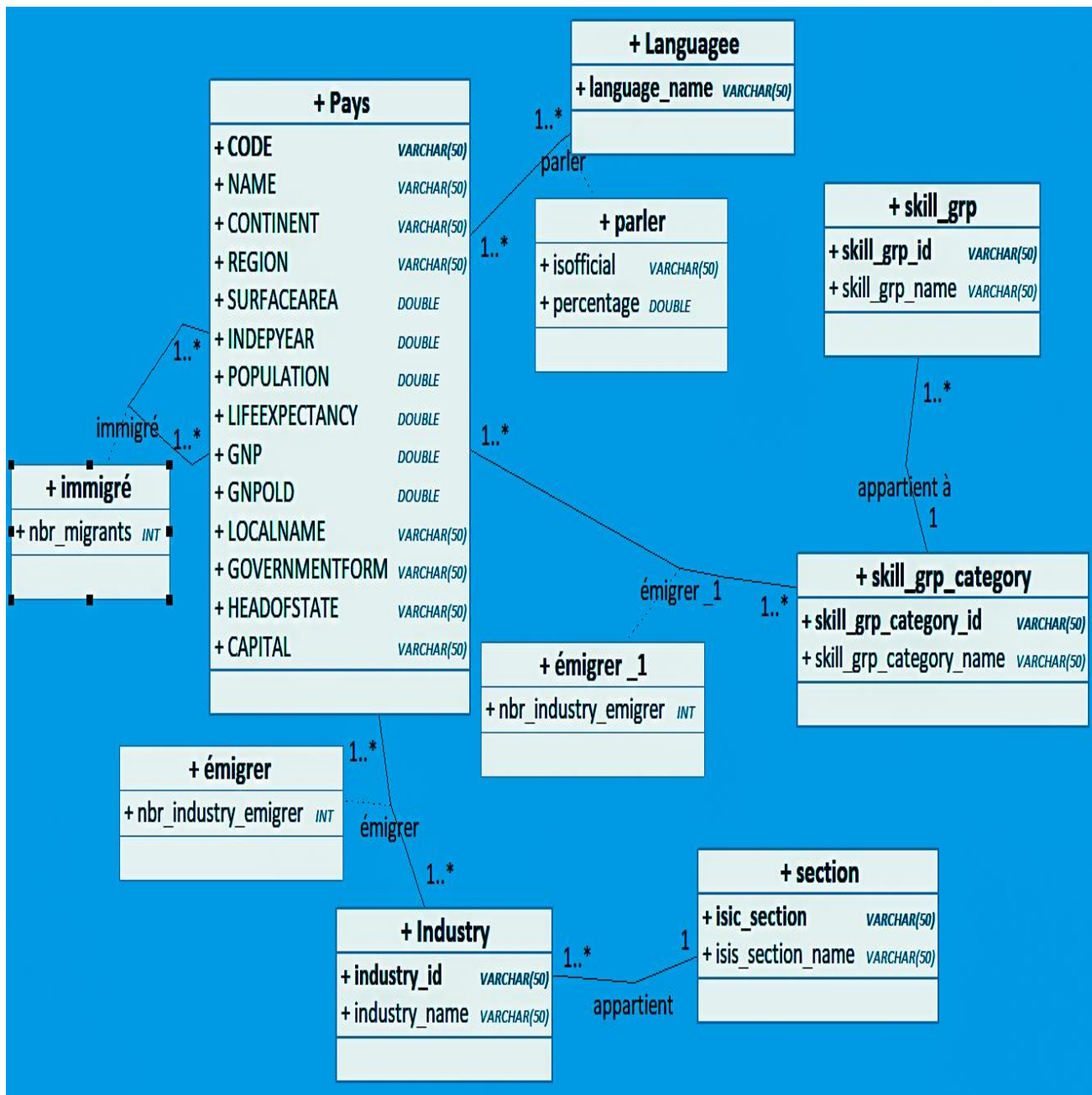
Après avoir étudié les données à notre disposition, nous proposons d'expliquer les données comme suit :

✓ Voir le fichier [synthèse.text](#)

2. Le modèle relationnel normalisé :

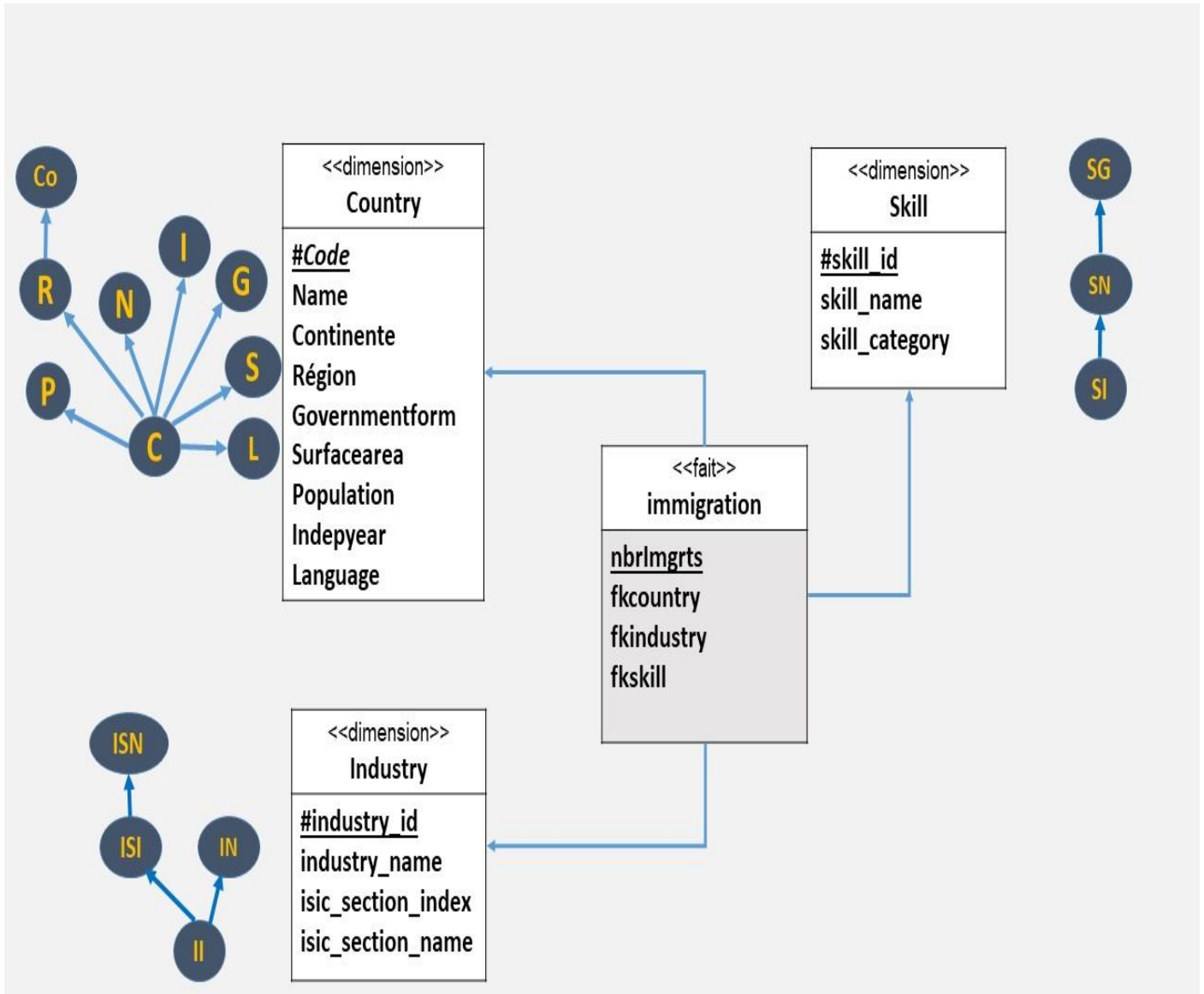


3. Le modèle conceptuel en **UML** correspondant au modèle relationnel normalisé :



V. Modélisation dimensionnelle :

À partir de l'étude des données et des besoins, on propose une **représentation logique finale** de notre **modèle dimensionnel en étoile** pour un **DataWarehouse** permettant de traiter la question générale qui a été posée :



✓ Description des tables de dimensions :

Attribut	Type	Commentaire	Qualité	Utilité
<i>#CODE</i>	CHAR(2)	PRIMARY KEY	2	2
<i>NAME</i>	VARCHAR2(52)	nom du pays	2	2
<i>CONTINENT</i>	VARCHAR2(20)	La continent du pays	2	2
<i>REGION</i>	VARCHAR2(26)	La région	2	2
<i>SURFACEAREA</i>	NUMBER(10,2)	La surface du pays	1	2
<i>INDEPYEAR</i>	NUMBER	L'année de l'indépendance	1	2
<i>POPULATION</i>	NUMBER	Population	1	2
<i>GNP</i>	NUMBER(10,2)	Produit national brut du pays	2	2
<i>GOVERNMENTFORM</i>	VARCHAR2(45)	Système de gouvernement du pays	2	2
<i>CAPITAL</i>	NUMBER	la somme d'investissements du pays	2	2
<i>LANGUAGE</i>	VARCHAR2(30)	La langue dominante	2	2

Description de la table de dimension *country*

Attribut	Type	Commentaire	Qualité	Utilité
<i>#SKILL_GROUP_ID</i>	NUMBER(5)	PRIMARY KEY	2	2
<i>SKILL_GROUP_CATEGORY</i>	VARCHAR2(100)	nom du catégorie de la compétence	2	2
<i>SKILL_GROUP_NAME</i>	VARCHAR2(60)	nom de la compétence	2	2

Description de la table de dimension *skill*

Attribut	Type	Commentaire	Qualité	Utilité
<i>#INDUSTRY_ID</i>	NUMBER(5)	PRIMARY KEY	2	2
<i>INDUSTRY_ID</i>	VARCHAR2(100)	nom du industrie	2	2
<i>SIC_SECTION_NAME</i>	VARCHAR2(100)	nom du section de industrie	2	2

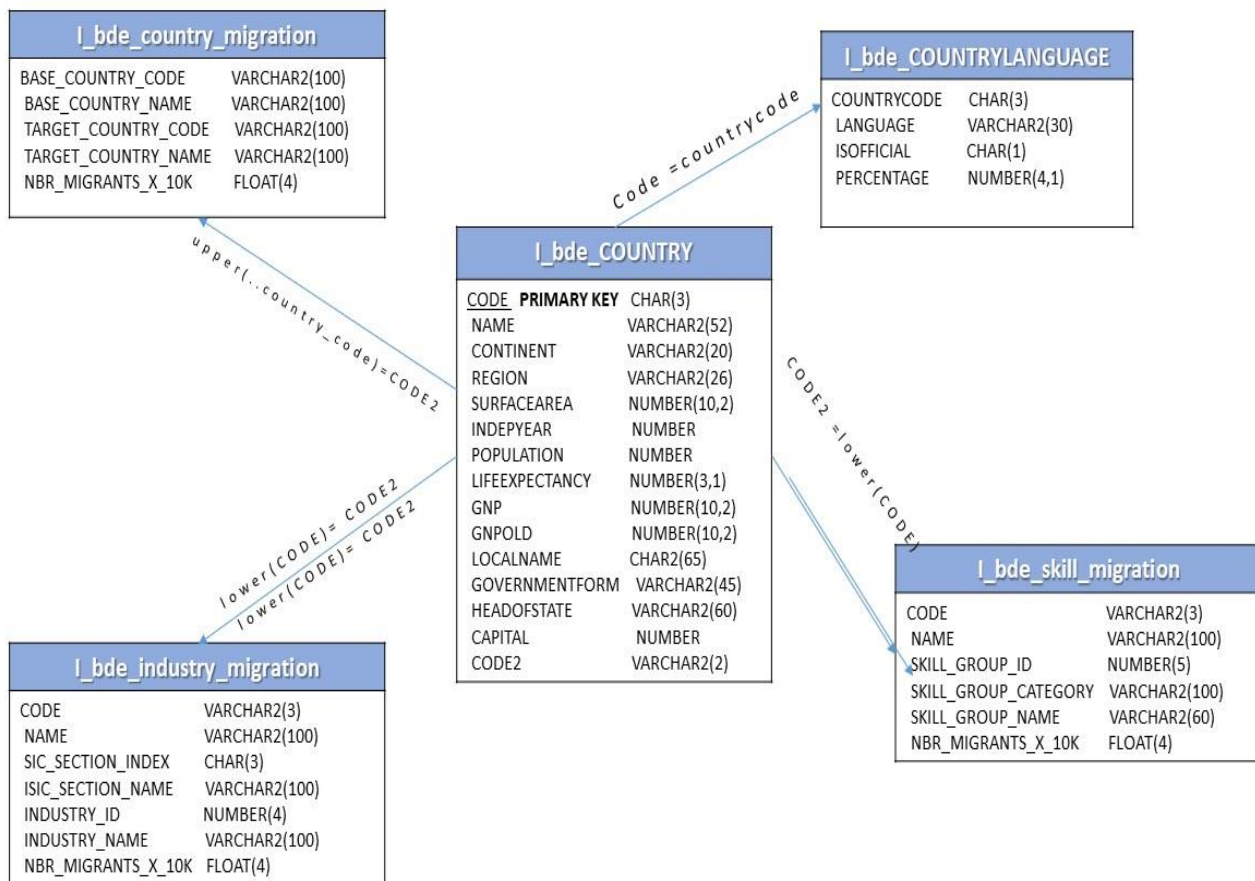
Description de la table de dimension *industry*

VI. Mise en place de la zone d'extraction :



Dans cette partie, on va charger une base de données contenant toutes les informations sur les types d'immigration (comme mentionné plus haut dans la section II *données disponibles* et dans la partie IV *études des données*) téléchargé depuis

Voilà le schéma relationnel des tables relationnelle de base de données Oracle :



Nous avons créé des tables externes pour chaque fichier csv enregistrer dans le dossier [csv_files](#) :

- Nous avons créé une vue [I_bde_country](#) pour la table de la base Oracle "country".
- Nous avons créé une vue [I_bde_countrylanguage](#) pour la table de la base Oracle "countrylanguage".
- Nous avons créé une table externe [I_bde_skill_migration](#) pour le fichier csv "skills_migration.csv".
- Nous avons créé une table externe [I_bde_industry_migration](#) pour le fichier csv "industry_migration.csv".
- Nous avons créé une table externe [I_bde_skill_migration](#) pour le fichier csv "skills_migration.csv".
- Nous avons créé une table externe [I_bde_country_migration](#) pour le fichier csv "country_migration.csv".

✓ *Voir le dossier [sql_files](#) (les fichiers : [I_bde_country](#), [I_bde_industry_migration](#), ...)*

! *Exécutez simplement le fichier [I_bde](#) pour créer la zone d'extraction.*

VII. Mise en place de la zone de transformation :

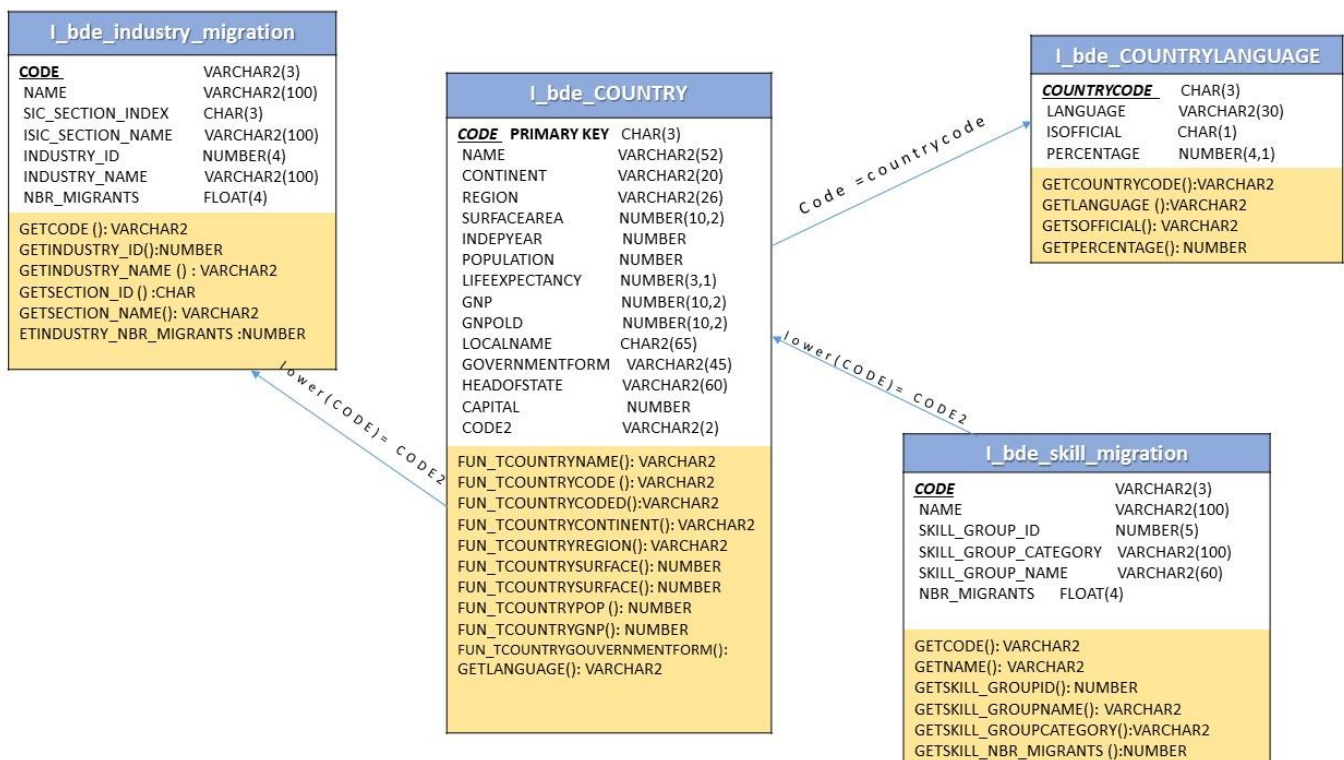
T

*Dans cette zone (de transformation), on va créer des tables objet-relationnelle (SQL3) qui ont des **fonctions** qui permet de transformer les données pour les mettre dans notre Data Warehouse.*

✓ Voir le dossier *sql_files* (les fichiers : *I_bdt_country*, *I_bdt_industry_migration*, ...)

! Exécutez simplement le fichier *I_bdt* pour créer la zone de transformation.

Voilà le schéma relationnel des tables objets relationnelle de base de données Oracle :

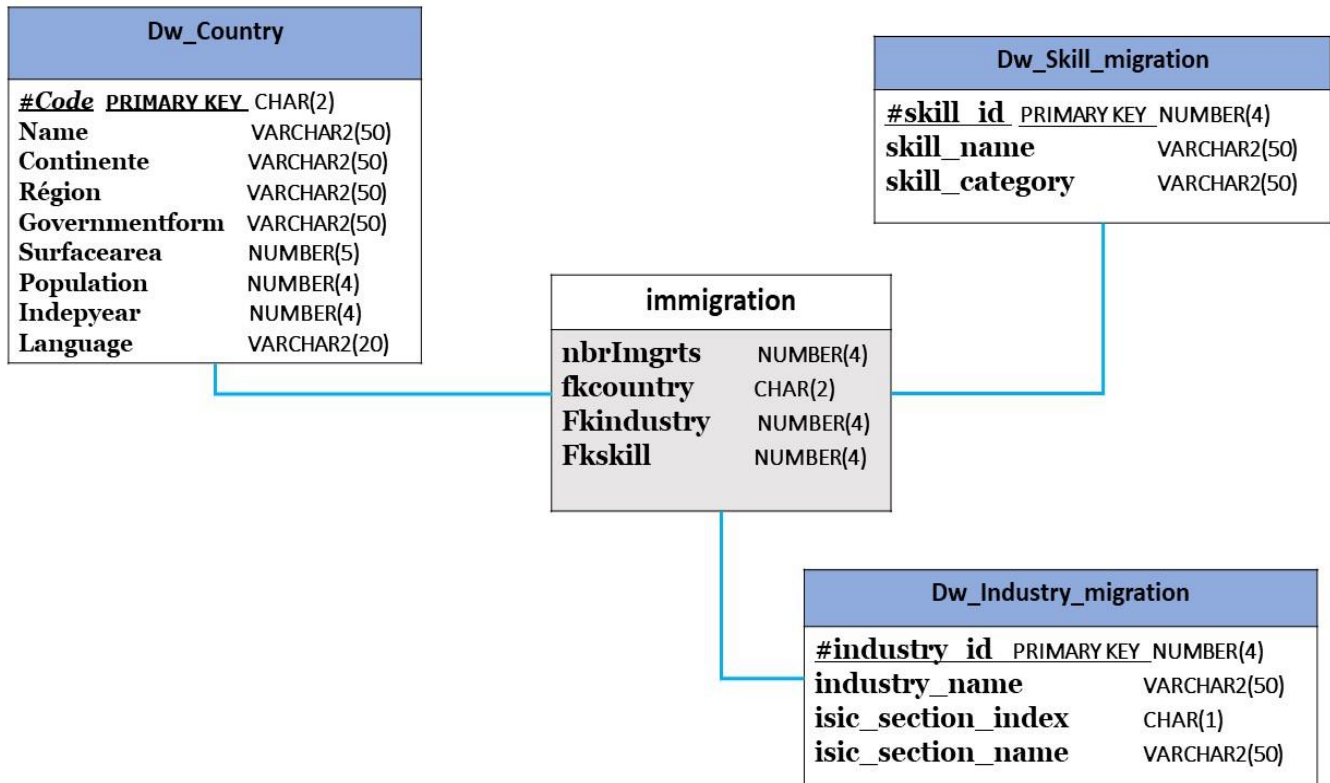


VIII. Mise en place de la zone d'exploitation :

L

L'objectif est maintenant de créer notre DataWarehouse.

Voilà le schéma *DataWarehouse* :



La partie « *Load* » consiste à mettre les données dans chaque table de dimension.

✓ Voir le fichier *I_DW* dans le dossier *sql_files*.

IX. Reporting :



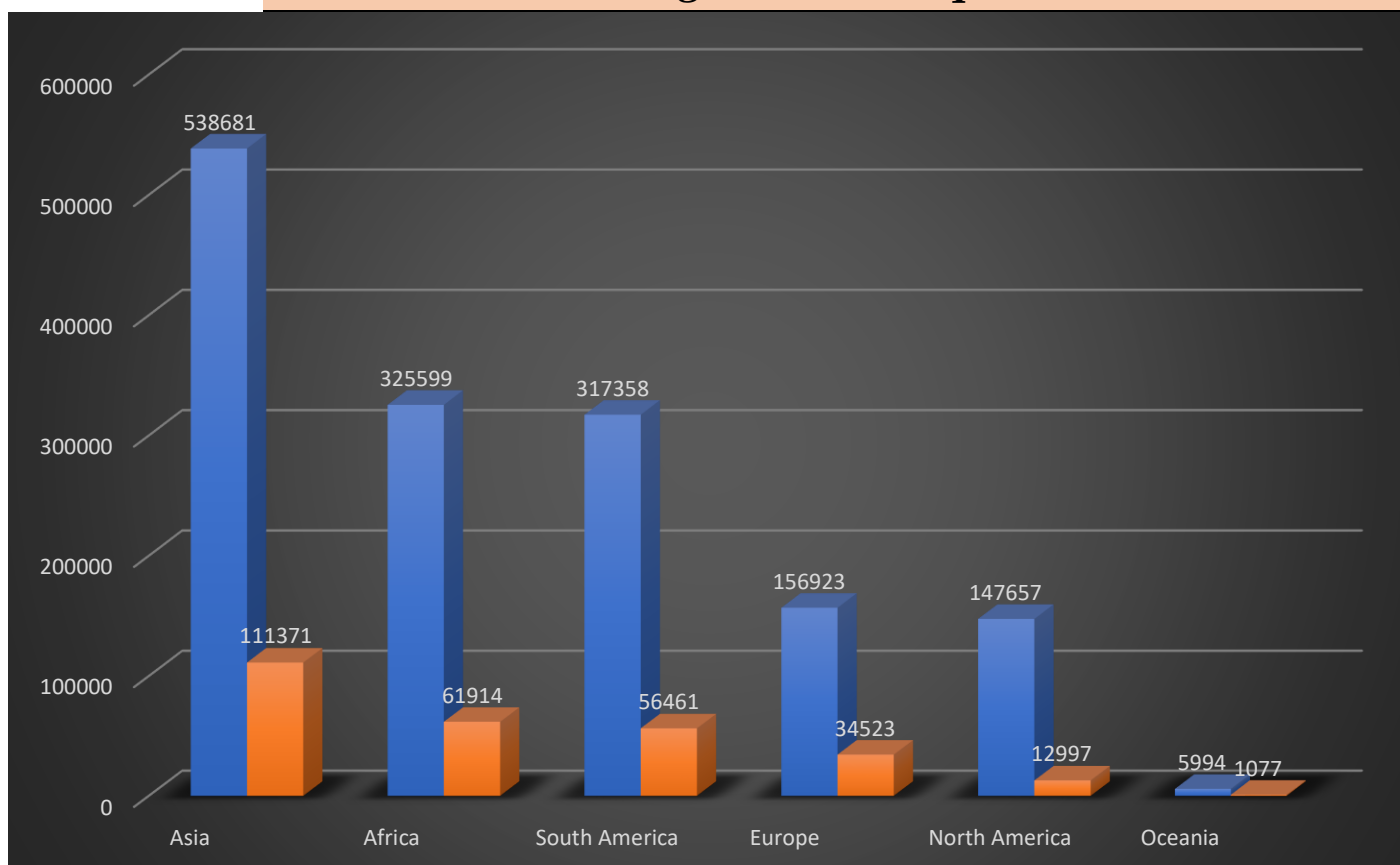
*Dans cette section, nous présenterons quelques **statistiques** qui répondent aux questions posées au début.*

Dans tout le reste nous symbolisons :

Nombres d'émigrent des compétences

Nombres d'émigrent d'industrie

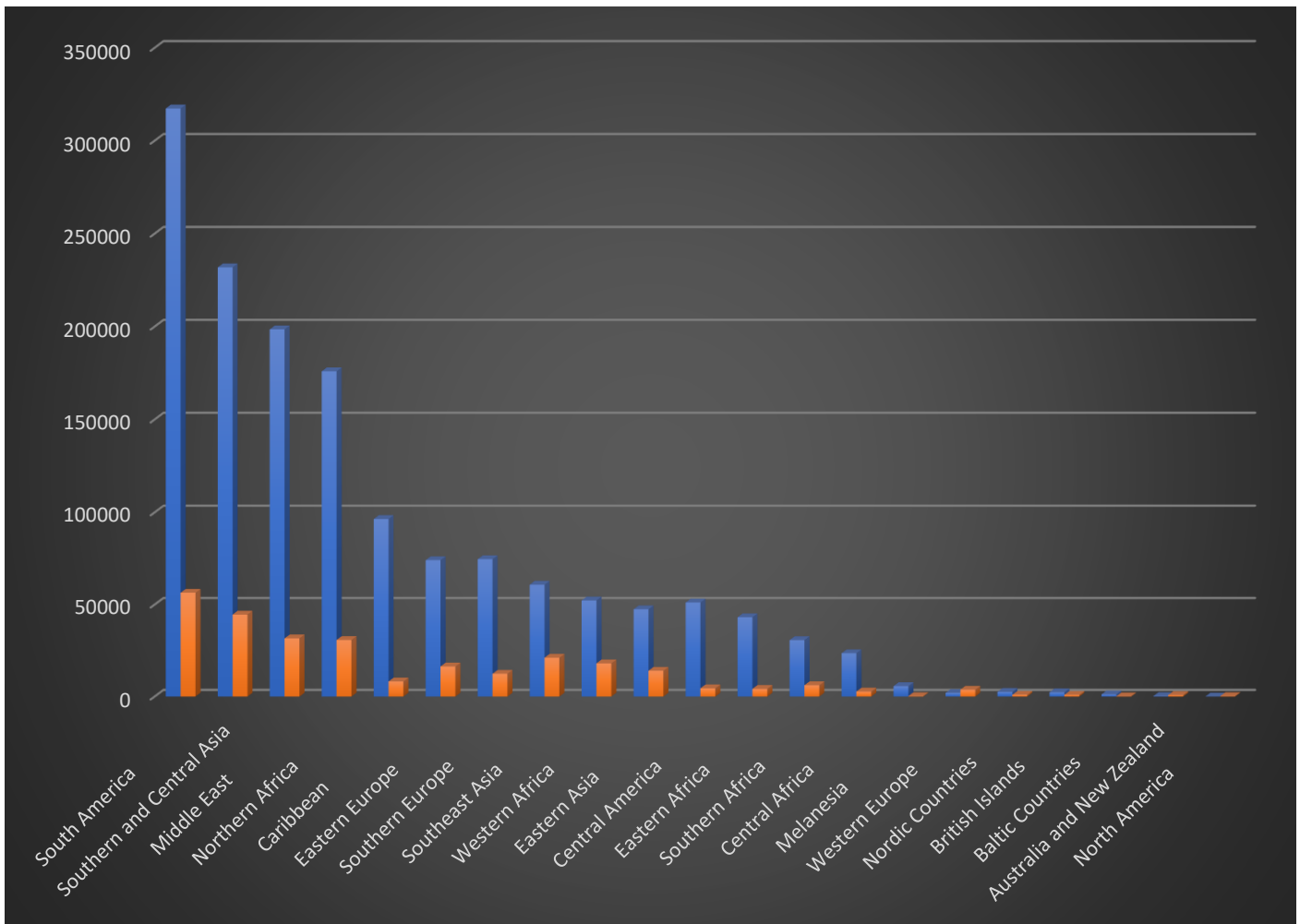
■ *Le nombre d'immigrants de chaque continent :*



❖ *Commentaires :*

- *L'Asie se classe au **premier rang** sur le continent où la plupart de ses habitants migrent, que ce soit en termes de compétences ou d'industrie, Et puis l'Afrique.*
- *Le continent **océanien** peut être considéré comme **invulnérable** au phénomène de migration*

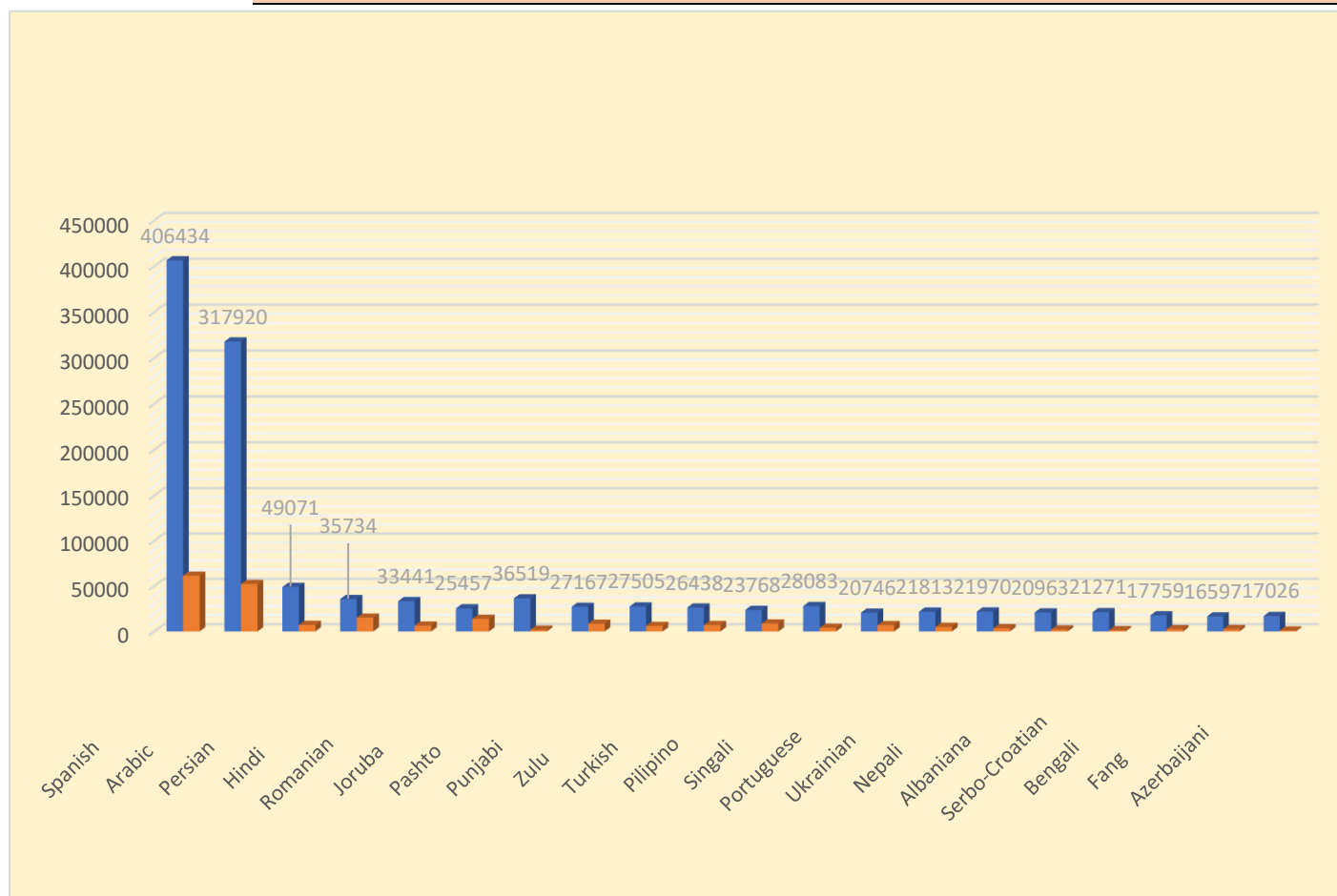
■ *Nombre d'immigrants de chaque région du monde :*



❖ *Commentaires :*

- *La migration des compétences dépasse **de loin** la migration de l'industrie à travers le monde.*
- *L'étude du nombre d'immigration diffère selon les continents et selon les régions, là où l'Asie était classée première, mais lorsque nous nous concentrons sur les régions, nous trouvons l'Amérique centrale, c'est le lieu où l'idée d'émigration est la plus répandue.*

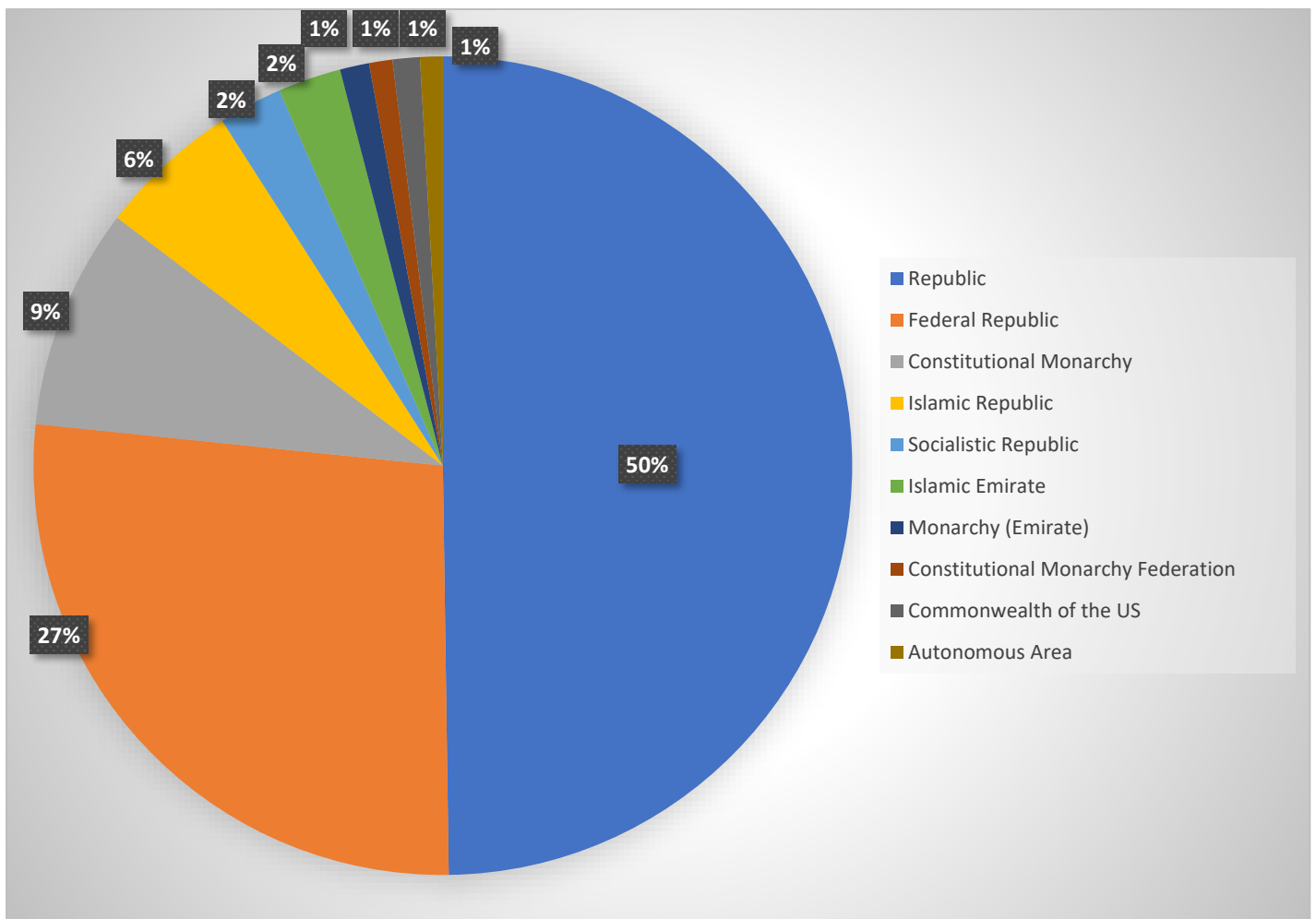
■ **Nombre d'immigrants selon la concentration linguistique :**



❖ **Commentaires :**

- Les personnes qui parlent **espagnol** ou **arabe** sont plus touchées par l'idée d'immigration que les autres.
- **L'espagnol** est la seule langue européenne dont les locuteurs sont exposés à réfléchir de l'immigration, contrairement au reste des **autres langues européennes**, qui ne connaissent pas une proportion importante, et cela indique les différents lieux de vie des locuteurs de cette langue.

■ *Le taux des immigrants par système de gouvernement :*

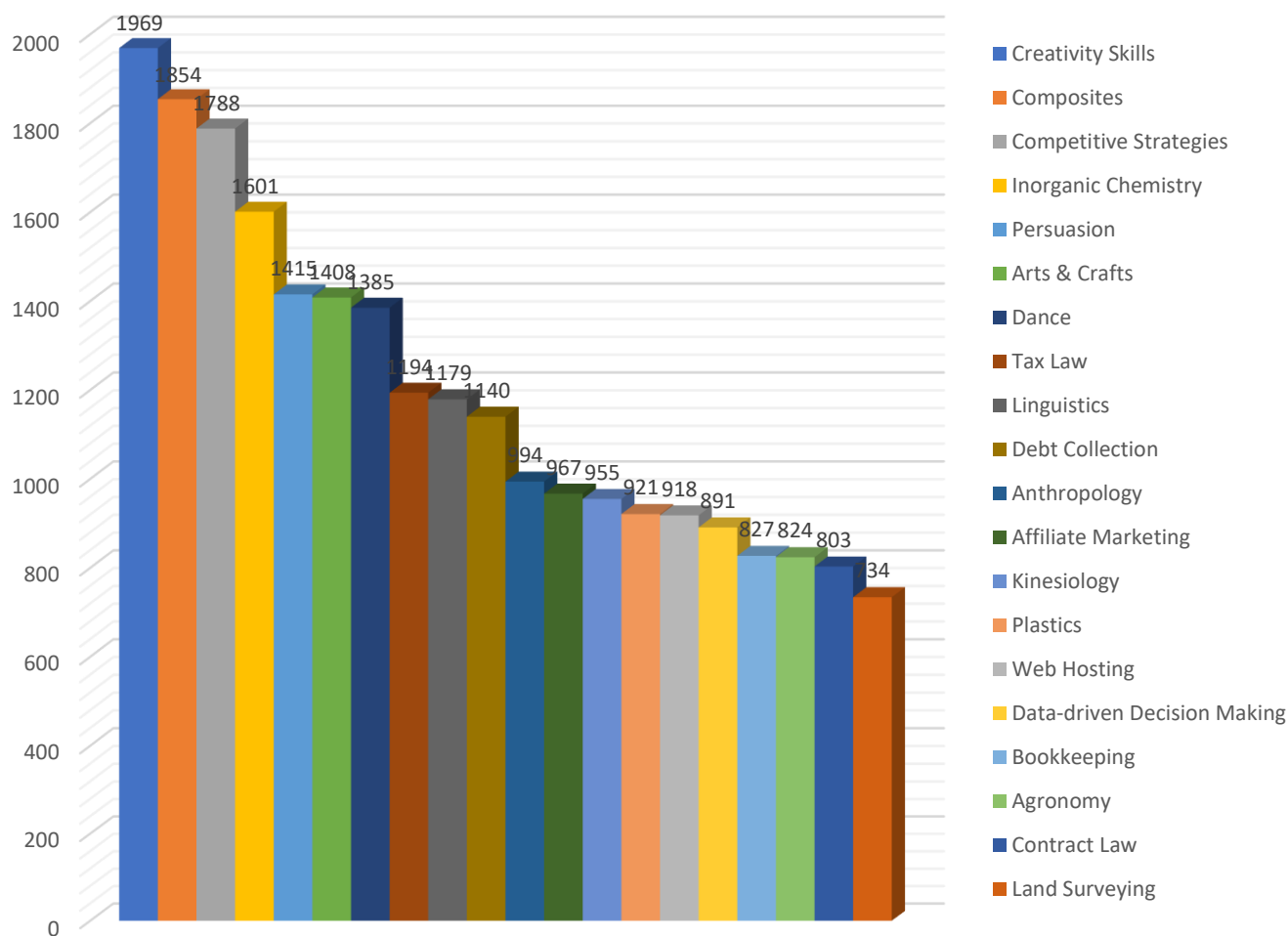


❖ *Commentaire :*

- La *moitié* des immigrants dans le monde viennent de pays *républicains*, et un *quart* des immigrants de pays *république fédérale*.

■ **Nombre de migration par catégorie de compétence :**

Les compétences qui migrent le plus



❖ **Commentaire :**

- *De nombreuses compétences quittent leur pays, et ce graphique montre que tous ces groupes migrent dans de larges proportions.*

X. Conclusion finale :

La migration est l'un des problèmes les plus mondiaux qui constitue une menace pour chaque pays, qu'il soit ciblé ou migrant.

Dans ce rapport, nous résumons ce que nous avons étudié concernant ce phénomène, Nous essayons de clarifier les dimensions les plus importantes qui peuvent être liées à l'immigration en tant que lieu de vie, système de vie et dimension linguistique.

Cette recherche visait une vision mathématique du phénomène migratoire à l'aide de moyens informationnels afin d'arriver à des explications pouvant être utiles à la prise de décision pour freiner l'immigration.

FIN