

# RAPPORT DU PROJET : MODELISATION ET CONCEPTION DE BASES DE DONNEES MULTIDIMENSIONNELLES POUR L'ANALYSE DES DONNEES CHEZ NETFLIX

CFD 2024

**Mohamed Falilou Fall**

EMAIL: MFF.FALILOU.FALL@GMAIL.COM +221-77-914-15-89

## Table des matières

1- INTRODUCTION.....	3
2- CONSTRUCTION DE LA BASE DE DONNÉES OPÉRATIONNELLE NETFLIX .....	4
2-1 Le script SQL de création de la base de données netflix.....	4
2-1-1 Création de la table pays (pays).....	4
2-1-2 Création de la table ville(ville) .....	4
2-1-3 Création de la table langue(langue).....	4
2-1-4 Création de la table catégorie film(categorie_film) .....	4
2-1-5 Création de la table client(client).....	5
2-1-6 Création de la table acteur (acteur) .....	5
2-1-7 Création de la table location(location).....	5
2-2 Le diagramme ER (Entité-Relation) de la base de données netflix.....	6
3- DÉFINITION DES AXES D'ANALYSE .....	7
3-1 Identification des dimensions analytiques pertinentes .....	7
3-2 Définition des indicateurs clés de performance (KPI).....	8
3-3 Scénarios d'analyse stratégique .....	9
4- MODÉLISATION ET IMPLÉMENTATION D'UN ENTREPÔT DE DONNÉES (DWH) DÉDIÉ "DWHNETFLIX.....	10
4-1 Modèle en étoile (star schema) .....	10
4-1 Conception logique et physique de l'entrepôt de données sous PostgreSQL avec DBEAVER ....	11
4-1-1 la table de dimension client .....	11
4-1-2 la table de dimension géographie.....	11
4-1-3 la table de dimension temps .....	11
4-1-4 la table de dimension contenu .....	11
4-1-5 la table de faits location .....	12
4-2 ETL (Extraction, Transformation, Loading) : chargement des données brutes .....	12
4-2-1 Insertion dans dim_temps.....	12
4-2-2 Insertion dans faits_location .....	12
4-2-3 Insertion dans dim_géographie .....	12
4-2-4 Insertion dans dim_client.....	13
4-2-5 Insertion dans dim_temps.....	13
5- CONCEPTION DU CUBE OLAP AVEC PENTAHO SCHEMA WORKBENCH .....	14
5-1 Définition des dimensions et mesures dans le cube.....	14
5-2 Optimisation du schéma pour les requêtes multidimensionnelles .....	15
6- RÉALISATION DE REQUÊTES MDX POUR L'ANALYSE STRATÉGIQUE .....	16
6-1 Analyse des revenus mensuels par type de contenu au Sénégal .....	16

6-2 Top 10 des films les plus loués sur la période d'une année au Senegal .....	16
6-3 <b>Taux de rétention</b> : Pourcentage d'utilisateurs qui restent abonnés sur une période d'une année par rapport au total des utilisateurs du début de la période (1 <sup>er</sup> janvier) au Senegal .....	17
6-4 Comparaison des locations entre le semestre 1 et le semestre 2 au Senegal .....	18
6-5 Montant total des locations au Senegal.....	19
6-6 Impact des langues sur les locations au Senegal .....	19
6-7 Analyse des pics de location par jour de la semaine au Senegal.....	19
6-8 Evolution de la location au fil de l'année (les 12 mois) au Senegal .....	20
6-9 Segmentation des utilisateurs selon leurs comportements de location au Senegal .....	20
6-10 Prévision de la demande basée sur les tendances historiques au Senegal.....	21
7- PROBLEMES RENCONTRES .....	22

## 1- INTRODUCTION

Dans un contexte où les données jouent un rôle central dans les prises de décision stratégiques, ce projet académique vise à concevoir et optimiser une base de données multidimensionnelle dédiée à l'analyse de la performance de la plateforme Netflix. L'objectif est de fournir à l'entreprise un entrepôt de données (Data Warehouse) qui permettra de centraliser et exploiter efficacement les données transactionnelles, comportementales, et géographiques afin de mieux comprendre les préférences des utilisateurs et d'améliorer la stratégie commerciale.

Ce projet intègre une approche complète, allant de la construction de la base de données opérationnelle à la modélisation avancée pour des analyses multidimensionnelles. En mettant en place un modèle d'entrepôt de données (modèle en étoile), nous visons à faciliter l'accès aux informations clés sur les transactions, utilisateurs, contenus, et périodes. Ces données permettront une analyse approfondie des tendances de consommation et des performances, tant au niveau local (Sénégal) qu'au sein de l'espace UEMOA et au-delà.

## 2- CONSTRUCTION DE LA BASE DE DONNÉES OPÉRATIONNELLE NETFLIX

### 2-1 Le script SQL de création de la base de données netflix

#### 2-1-1 Création de la table pays (pays)

```
CREATE TABLE if not exists pays
(
  id_pays INTEGER PRIMARY KEY,
  code_pays VARCHAR(4) NOT NULL,
  pays TEXT not null
);
```

#### 2-1-2 Création de la table ville(ville)

```
create table if not exists ville
(
  id_ville INTEGER PRIMARY key,
  code_ville VARCHAR(10),
  ville text not null,
  id_pays INTEGER,
  FOREIGN KEY (id_pays) REFERENCES pays(id_pays)
);
```

#### 2-1-3 Création de la table langue(langue)

```
create table if not exists langue
(
  id_langue INTEGER primary key ,
  code_langue VARCHAR(3),
  libelle CHAR(20)
);
```

#### 2-1-4 Création de la table catégorie film(categorie\_film)

```
create table if not exists categorie_film
(
  id_categorie integer primary key ,
  code_categorie varchar(2),
  libelle text not null
);
```

#### 2-1-5 Création de la table client(client)

```
create table if not exists client
(
  id_client integer primary key ,
  code_client varchar(10),
  prenom varchar(100),
  nom varchar(100),
  email varchar(100),
  id_ville integer,
  FOREIGN KEY (id_ville) REFERENCES ville(id_ville)
);
```

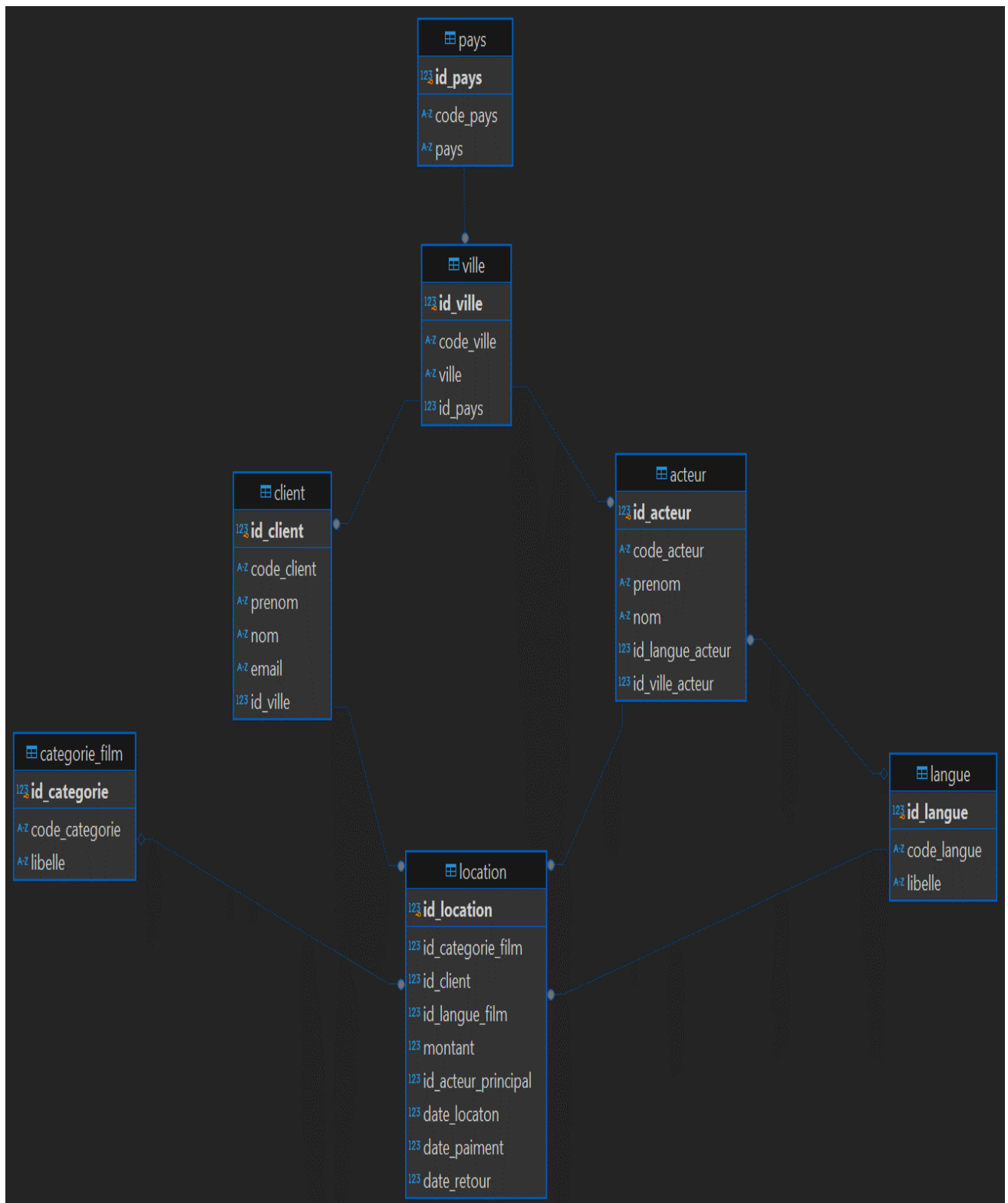
#### 2-1-6 Création de la table acteur (acteur)

```
create table if not exists acteur
(
  id_acteur integer primary key ,
  code_acteur varchar(5),
  prenom varchar(100),
  nom varchar(100),
  id_langue_acteur integer ,
  id_ville_acteur integer ,
  FOREIGN KEY (id_langue_acteur) REFERENCES langue(id_langue),
  FOREIGN KEY (id_ville_acteur) REFERENCES ville(id_ville)
);
```

#### 2-1-7 Création de la table location(location)

```
create table if not exists location
(
  id_location integer primary key,
  id_categorie_film integer ,
  id_client integer ,
  id_langue_film integer,
  montant float,
  id_acteur_principal integer ,
  date_locaton integer,
  date_paiement integer,
  date_retour integer ,
  FOREIGN KEY (id_location) REFERENCES location(id_location),
  FOREIGN KEY (id_categorie_film) REFERENCES categorie_film(id_categorie),
  FOREIGN KEY (id_client) REFERENCES client(id_client),
  FOREIGN KEY (id_langue_film) REFERENCES langue(id_langue),
  FOREIGN KEY (id_acteur_principal) REFERENCES acteur(id_acteur)
);
```

## 2-2 Le diagramme ER (Entité-Relation) de la base de données netflix



### 3- DÉFINITION DES AXES D'ANALYSE

#### 3-1 Identification des dimensions analytiques pertinentes

Utilisateurs

Contenus

Periodes

Transactions(paiement)

Geographie



### 3-2 Définition des indicateurs clés de performance (KPI)



## Scenario 1

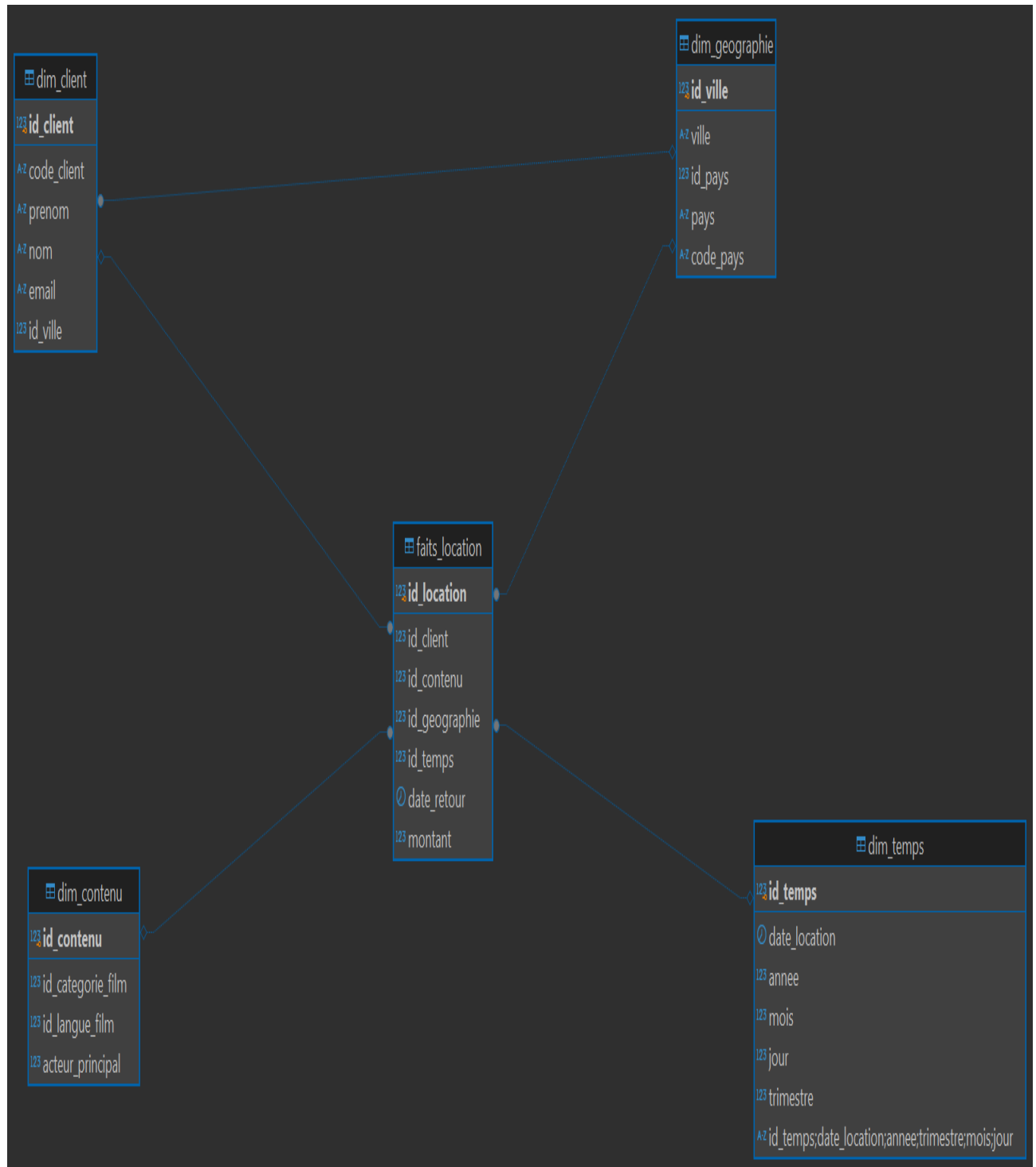
- suivi des tendances de consommation

## Scenario 2

- segmentation des clients

## 4- MODÉLISATION ET IMPLÉMENTATION D'UN ENTREPÔT DE DONNÉES (DWH) DÉDIÉ "DWHNETFLIX"

### 4-1 Modèle en étoile (star schema)



## 4-1 Conception logique et physique de l'entrepôt de données sous PostgreSQL avec DBEAVER

### 4-1-1 la table de dimension client

```
CREATE TABLE IF NOT EXISTS dim_client (  
    id_client INTEGER PRIMARY KEY,  
    code_client VARCHAR(10) NOT NULL,  
    prenom VARCHAR(100),  
    nom VARCHAR(100),  
    email VARCHAR(100),  
    id_ville INTEGER,  
    FOREIGN KEY (id_ville) REFERENCES dim_geographie(id_ville)  
);
```

### 4-1-2 la table de dimension geographie

```
CREATE TABLE IF NOT EXISTS dim_geographie (  
    id_ville INTEGER PRIMARY KEY,  
    ville VARCHAR(100),  
    id_pays INTEGER,  
    pays VARCHAR(100),  
    code_pays VARCHAR(4)  
);
```

### 4-1-3 la table de dimension temps

```
CREATE TABLE IF NOT EXISTS dim_temps (  
    id_temps INTEGER PRIMARY KEY,  
    date_location DATE,  
    annee INTEGER,  
    mois INTEGER,  
    jour INTEGER,  
    trimestre INTEGER  
);
```

### 4-1-4 la table de dimension contenu

```
CREATE TABLE IF NOT EXISTS dim_contenu (  
    id_contenu INTEGER PRIMARY KEY,  
    id_categorie_film INTEGER,  
    id_langue_film INTEGER,  
    acteur_principal INTEGER,  
    FOREIGN KEY (id_categorie_film) REFERENCES categorie_film(id_categorie),  
    FOREIGN KEY (id_langue_film) REFERENCES langue(id_langue),  
    FOREIGN KEY (acteur_principal) REFERENCES acteur(id_acteur)  
);
```

#### 4-1-5 la table de faits location

```
CREATE TABLE IF NOT EXISTS faits_location (  
    id_location INTEGER PRIMARY KEY,  
    id_client INTEGER,  
    id_contenu INTEGER,  
    id_geographie INTEGER,  
    id_temps INTEGER,  
    date_retour INTEGER,  
    montant FLOAT,  
    FOREIGN KEY (id_client) REFERENCES dim_client(id_client),  
    FOREIGN KEY (id_contenu) REFERENCES dim_contenu(id_contenu),  
    FOREIGN KEY (id_geographie) REFERENCES dim_geographie(id_ville),  
    FOREIGN KEY (id_temps) REFERENCES dim_temps(id_temps)  
);
```

#### 4-2 ETL (Extraction, Transformation, Loading) : chargement des données brutes

##### 4-2-1 Insertion dans dim\_temps

```
INSERT INTO dim_temps (id_temps, date_location, annee, mois, jour, trimestre)  
SELECT  
    L.id_location AS id_temps,  
    TO_DATE(L.date_location::text, 'YYYYMMDD') AS date_location,  
    EXTRACT(YEAR FROM TO_DATE(L.date_location::text, 'YYYYMMDD')) AS annee,  
    EXTRACT(MONTH FROM TO_DATE(L.date_location::text, 'YYYYMMDD')) AS mois,  
    EXTRACT(DAY FROM TO_DATE(L.date_location::text, 'YYYYMMDD')) AS jour,  
    CEIL(EXTRACT(MONTH FROM TO_DATE(L.date_location::text, 'YYYYMMDD')) / 3.0) AS  
trimestre  
FROM netflix.public.location L;
```

##### 4-2-2 Insertion dans faits\_location

```
INSERT INTO faits_location (  
    id_location, id_client, id_contenu, id_geographie, id_temps, date_retour,  
montant  
)  
SELECT  
    L.id_location,  
    L.id_client,  
    L.id_location AS id_contenu,  
    c.id_ville AS id_geographie,  
    L.id_location AS id_temps,  
    CAST(L.date_retour AS INTEGER) AS date_retour,  
    L.montant  
FROM netflix.public.location L  
JOIN netflix.public.client c ON L.id_client = c.id_client
```

##### 4-2-3 Insertion dans dim\_geographie

```
INSERT INTO dim_geographie (id_ville, ville, id_pays, pays, code_pays)  
SELECT v.id_ville, v.ville, p.id_pays, p.pays, p.code_pays  
FROM netflix.public.ville v  
JOIN netflix.public.pays p ON v.id_pays = p.id_pays;
```

#### 4-2-4 Insertion dans dim\_client

```
INSERT INTO dim_client (id_client, code_client, prenom, nom, email, id_ville)
SELECT c.id_client, c.code_client, c.prenom, c.nom, c.email, c.id_ville
FROM netflix.public.client c;
```

#### 4-2-5 Insertion dans dim\_temps

```
INSERT INTO dim_temps (id_temps, date_location, annee, mois, jour, trimestre)
SELECT
    L.id_location AS id_temps,
    DATE(L.date_location) AS date_location,
    strftime('%Y', DATE(L.date_location)) AS annee,
    strftime('%m', DATE(L.date_location)) AS mois,
    strftime('%d', DATE(L.date_location)) AS jour,
    (CAST(strftime('%m', DATE(L.date_location)) AS INTEGER) + 2) / 3 AS trimestre
FROM netflix.public.location L;
```

## 5- CONCEPTION DU CUBE OLAP AVEC PENTAHO SCHEMA WORKBENCH

### 5-1 Définition des dimensions et mesures dans le cube

#### Dimension Client

- id\_client
- code\_client
- prenom
- nom
- email

#### Dimension Contenu

- id\_contenu
- id\_categorie\_film
- id\_langue\_film
- acteur\_principal

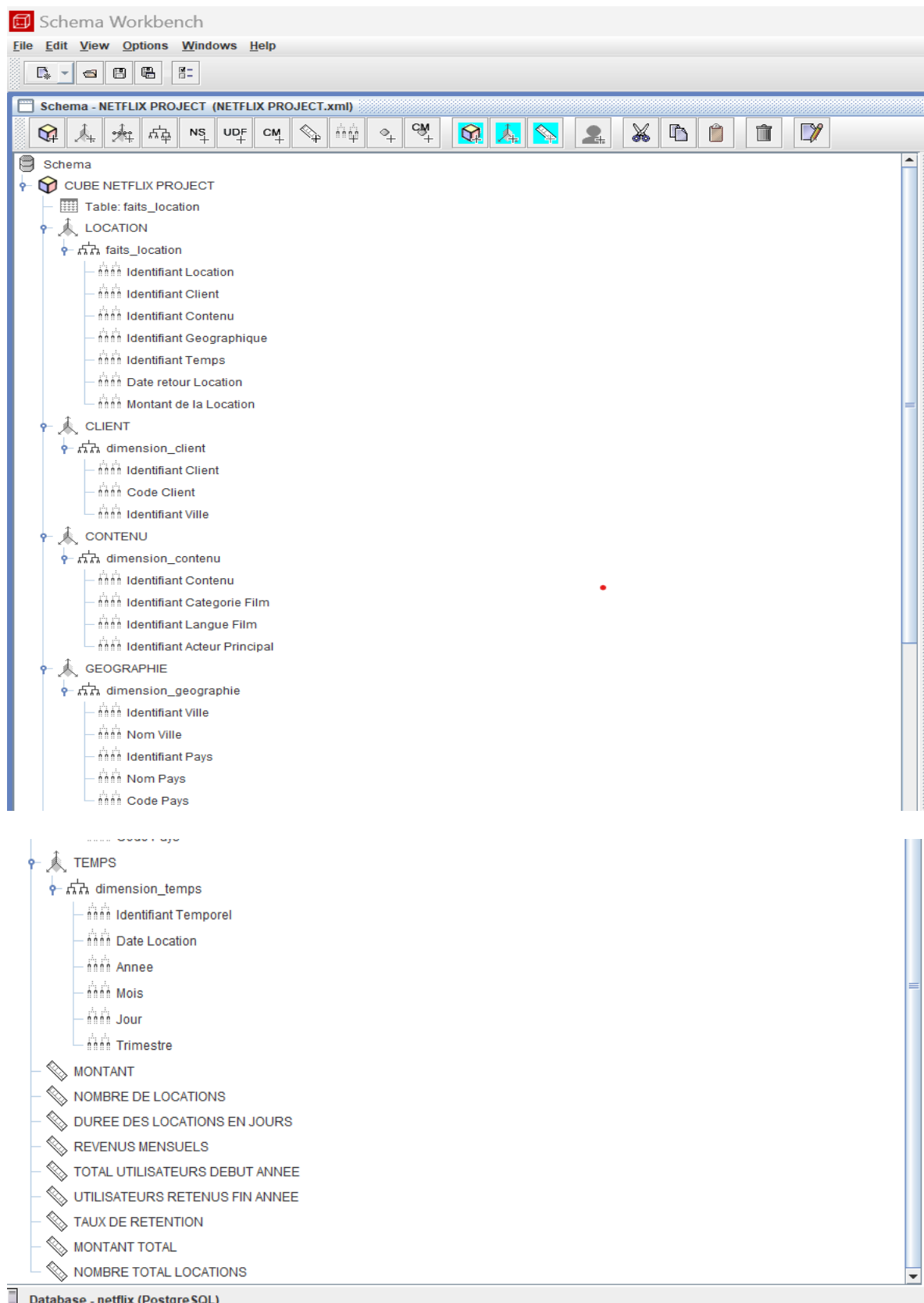
#### Dimension Temps

- id\_temps
- date\_location
- annee
- mois
- jour
- trimestre

#### Dimension Geographie

- id\_ville
- ville
- id\_pays
- pays
- code\_pays

## 5-2 Optimisation du schéma pour les requêtes multidimensionnelles





## 6- RÉALISATION DE REQUÊTES MDX POUR L'ANALYSE STRATÉGIQUE

### 6-1 Analyse des revenus mensuels par type de contenu au Senegal

```
WITH
  MEMBER [Measures].[REVENUS MENSUELS] AS
    SUM(
      [Location].[Date retour Location].CurrentMember.Children,
      [Measures].[MONTANT]
    )
SELECT
  {
    [TEMPS].[Mois].Members
  } ON ROWS,
  {
    [CONTENU].[Identifiant Categorie Film].Members
  } ON COLUMNS
FROM
  [CUBE_NETFLIX_PROJECT]
WHERE
  ([GEOGRAPHIE].[Nom Pays].&[Senegal])
```

### 6-2 Top 10 des films les plus loués sur la période d'une année au Senegal

```
WITH
  MEMBER [Measures].[NOMBRE DE LOCATIONS] AS
    SUM([LOCATION].[Identifiant Location].MEMBERS, [Measures].[NOMBRE DE LOCATIONS])

SELECT
  TOPCOUNT(
    [CONTENU].[Identifiant Contenu].MEMBERS,
    10,
    [Measures].[NOMBRE DE LOCATIONS]
  ) ON ROWS,

  [TEMPS].[Annee].&[2023] ON COLUMNS

FROM
  [CUBE_NETFLIX_PROJECT]
WHERE
  ([TEMPS].[Mois].MEMBERS, [GEOGRAPHIE].[Nom Pays].&[Sénégal])
```

**6-3 Taux de rétention** : Pourcentage d'utilisateurs qui restent abonnés sur une période d'une année par rapport au total des utilisateurs du début de la période (1<sup>er</sup> janvier) au Senegal

```
WITH
-- Nombre total d'utilisateurs actifs au debut de l'annee (1er janvier)
MEMBER [Measures].[TOTAL UTILISATEURS DEBUT ANNEE] AS
    Count(
        Filter(
            [CLIENT].[dimension_client].[Identifiant Client].Members,
            ([TEMPS].[dimension_temps].[Date Location].[1-Jan], [Measures].[NOMBRE DE LOCATIONS])
            > 0
        )
    )

-- Nombre d'utilisateurs encore actifs a la fin de l'annee (31 decembre)
MEMBER [Measures].[UTILISATEURS RETENUS FIN ANNEE] AS
    Count(
        Filter(
            [CLIENT].[dimension_client].[Identifiant Client].Members,
            ([TEMPS].[dimension_temps].[Date Location].[31-Dec], [Measures].[NOMBRE DE
LOCATIONS]) > 0
        )
    )

-- Calcul du Taux de Retention
MEMBER [Measures].[TAUX DE RETENTION] AS
    IIF(
        [Measures].[TOTAL UTILISATEURS DEBUT ANNEE] > 0,
        ([Measures].[ UTILISATEURS RETENUS FIN ANNEE] * 100.0) / [Measures].[ TOTAL
UTILISATEURS DEBUT ANNEE],
        NULL
    )

SELECT
    {[Measures].[TAUX DE RETENTION]} ON COLUMNS,
    {[GEOGRAPHIE].[dimension_geographie].[Nom Pays].&[Senegal]} ON ROWS
FROM [CUBE_NETFLIX_PROJECT]
WHERE ([TEMPS].[dimension_temps].[Annee].&[2024])
```

#### 6-4 Comparaison des locations entre le semestre 1 et le semestre 2 au Senegal

```
WITH
MEMBER [Temps].[Semestre].[S1] AS
    Aggregate(
        {[Temps].[Annee].CurrentMember, [Temps].[Mois].&[1]:[Temps].[Mois].&[6]}
    )
MEMBER [Temps].[Semestre].[S2] AS
    Aggregate(
        {[Temps].[Annee].CurrentMember, [Temps].[Mois].&[7]:[Temps].[Mois].&[12]}
    )

SELECT
    {[Temps].[Semestre].[S1], [Temps].[Semestre].[S2]} ON COLUMNS,
    {[GEOGRAPHIE].[Nom Pays].&[Senegal]} ON ROWS
FROM [CUBE_NETFLIX_PROJECT]
WHERE
    (Measures.[NOMBRE DE LOCATIONS])
```

#### 6-5 Montant total des locations au Senegal

```
SELECT
  {[Measures].[MONTANT]} ON COLUMNS,
  {[GEOGRAPHIE].[dimension_geographie].[Nom Pays].&[Senegal]} ON ROWS
FROM [CUBE_NETFLIX_PROJECT]
WHERE ([GEOGRAPHIE].[dimension_geographie].[Nom Pays].&[Senegal])
```

#### 6-6 Impact des langues sur les locations au Senegal

```
SELECT
  NON EMPTY
  {[Measures].[NOMBRE DE LOCATIONS], [Measures].[MONTANT]} ON COLUMNS,
  NON EMPTY
  (
    [CONTENU].[dimension_contenu].[Identifiant Langue Film].Members
  ) ON ROWS
FROM
  [CUBE_NETFLIX_PROJECT]
WHERE
  ([GEOGRAPHIE].[dimension_geographie].[Nom Pays].&[Senegal])
```

#### 6-7 Analyse des pics de location par jour de la semaine au Senegal

```
SELECT
  NON EMPTY
  [TEMPS].[Jour].Members ON COLUMNS,
  NON EMPTY
  {[Measures].[NOMBRE DE LOCATIONS]} ON ROWS
FROM
  [CUBE_NETFLIX_PROJECT]
WHERE
  ([GEOGRAPHIE].[Nom Pays].[Sénégal])
```

## 6-8 Evolution de la location au fil de l'année (les 12 mois) au Senegal

```
WITH
MEMBER [Measures].[MONTANT TOTAL] AS
SUM(
    [GEOGRAPHIE].[Nom Pays].[Senegal],
    [Measures].[MONTANT]
)
SELECT
    {[Measures].[MONTANT TOTAL]} ON COLUMNS,
    {[TEMPS].[Mois].Members} ON ROWS
FROM [CUBE_NETFLIX_PROJECT]
WHERE
    ([GEOGRAPHIE].[Nom Pays].[Senegal])
```

## 6-9 Segmentation des utilisateurs selon leurs comportements de location au Senegal

```
SELECT
    NON EMPTY
    {
        [Measures].[MONTANT],
        [Measures].[NOMBRE DE LOCATIONS]
    } ON COLUMNS,

    NON EMPTY
    (
        [CLIENT].[dimension_client].[Identifiant Client].Members *
        [GEOGRAPHIE].[dimension_geographie].[Nom Pays].[Senegal] *
        [TEMPS].[dimension_temps].[Annee].Members *
        [TEMPS].[dimension_temps].[Trimestre].Members
    ) ON ROWS

FROM [CUBE_NETFLIX_PROJECT]

WHERE
    (
        [GEOGRAPHIE].[dimension_geographie].[Nom Pays].[Senegal]
    )
```

## 6-10 Prévision de la demande basée sur les tendances historiques au Senegal

```
WITH
MEMBER [Measures].[MONTANT TOTAL] AS
SUM(
    [GEOGRAPHIE].[Nom Pays].&[Senegal],
    [Measures].[MONTANT]
)

MEMBER [Measures].[NOMBRE TOTAL LOCATIONS] AS
SUM(
    [GEOGRAPHIE].[Nom Pays].&[Senegal],
    [Measures].[NOMBRE DE LOCATIONS]
)

SELECT
{
    [Measures].[MONTANT TOTAL],
    [Measures].[Nombre Total Locations]
} ON COLUMNS,

NON EMPTY
(
    [TEMPS].[Annee].[Annee].Members *
    [TEMPS].[Mois].[Mois].Members
) ON ROWS

FROM [CUBE_NETFLIX_PROJECT]

WHERE ([GEOGRAPHIE].[Nom Pays].&[Senegal])
```

## 7- PROBLEMES RENCONTRES ET LIENS

### 7-1 Problème rencontré

Étant donné la configuration des données, il n'a pas été possible de mener une analyse à l'échelle de l'Espace économique de l'UEMOA. Le Sénégal étant le seul pays représenté dans l'échantillon, l'analyse a été restreinte à ce dernier.

### 7-2 Liens vers le DUMP et le Code XML

- [DUMP NETFLIX PROJECT MOHAMED FALILOU FALL](#)
- [CUBE NETFLIX PROJECT MOHAMED FALILOU FALL XML](#)