

SAE 1.04 - Création d'une base de données

Compétence 4 :

Concevoir et mettre en place une base de données à partir d'un cahier des charges client

Apprentissages critiques :

AC 1 : Mettre à jour et interroger une base de données relationnelle (en requêtes directes ou à travers une application)

AC 3 : Concevoir une base de données relationnelle à partir d'un cahier des charges

Problématique professionnelle :

L'objectif est de montrer que dans le cadre d'un nouveau besoin il est nécessaire d'étudier l'existant, de le maîtriser et ensuite de proposer des extensions à la solution existante.

Déroulement :

Le travail doit se faire en binôme.

- 4h S1.04 en S14 du 11/12/2023 au 15/12/2023
- 2h S1.04 en S15 du 18/12/2023 au 22/12/2023
- 6h S1.04 en S16 du 08/01/2024 au 12/01/2024

Productions et évaluations :

Vous devez rendre un rapport clairement rédigé contenant l'ensemble des productions demandées dans le sujet sous forme d'un unique fichier au format PDF (le nom du fichier doit respecter cette nomenclature : 'G1A_NOM1_PRENOM1_NOM2_PRENOM2.pdf').

Votre rapport, d'une dizaine de pages au minimum, doit contenir une page de garde, un sommaire, une introduction et une conclusion. Il doit être exhaustif et décrire votre réflexion. L'ensemble de votre code doit être présent.

En plus d'une note de rapport vous serez évalués à l'oral pendant une séance de TP R1.05 (10 minutes par groupe). Cette évaluation aura lieu la semaine 17.

Vous devrez présenter votre travail sur machine en montrant les outils, la base utilisée, les requêtes, les graphiques. Si vous le souhaitez, vous pouvez vous appuyer sur des slides de présentation mais cela n'est pas obligatoire.

Vous devez également compléter individuellement la fiche portfolio de cette SAE. Bien entendu, vous pouvez reprendre les conclusions de votre rapport pour alimenter cette fiche. Vous trouverez cette fiche dans le partage réseau (pensez à nommer correctement ce fichier avec votre nom).

Vous devez déposer votre **rapport PDF** ainsi que votre **fiche portfolio (1 seul page au format PDF)** dans le « Casier-enseignants » :

« /Pédagogie/pedago-iut/INFO/Casiers-enseignants/MERILLOU_Nicolas/ »

La date limite de rendu est fixée au dimanche 14 janvier 2024 à minuit

Mail de contact : nicolas.merillou@unilim.fr

Sujet :

Vous devez utiliser les connaissances acquises dans la ressource R1.05 (Introduction aux bases de données et SQL) pour réaliser une étude s'appuyant sur une base de données. Le travail principal consiste à récupérer des données brutes au format CSV puis utiliser le langage SQL pour les analyser et les stocker dans une base de données que vous concevrez. Enfin vous devrez créer des requêtes SQL afin de produire des analyses sous forme graphique.

La source de données que vous devez utiliser contient une liste de jeux vidéo ainsi que leur nombre de ventes par secteur géographique et par année. Il s'agit du fichier suivant : « vgsales.csv ».

Pour en savoir plus sur cet ensemble de données, vous pouvez consulter le lien suivant :

<https://www.kaggle.com/gregorut/videogamesales>

Le format de ce fichier CSV est décrit comme suit :

- une ligne d'entête
- une virgule en séparateur de colonnes
- un point pour la virgule des nombres décimaux
- un guillemet pour les chaînes de caractères ayant des virgules

Le travail se déroule en 3 phases.

Phase 1 :

Dans cette phase vous devez analyser les données afin de bien comprendre leur structure et leurs particularités. L'objectif final est de construire un modèle de donnée en phase 2, permettant de mieux structurer l'information. Cette analyse peut se faire visuellement dans un premier temps puis en utilisant le SQL.

Pour cela vous pouvez insérer les données brutes sous forme de chaînes de caractères dans une unique table créée à cet effet. Cette table reprendra les colonnes du fichier CSV. Cette opération peut être effectuée grâce à l'assistant d'importation & exportation de données de Management Studio vu en TP. Par défaut, lors de la lecture d'un fichier plat, cet outil ne prend en compte que des colonnes de 50 caractères maximum ; à vous de trouver les bonnes options. Si vous ne souhaitez pas utiliser cet assistant, vous pouvez également générer des « insert » depuis le fichier CSV avec la méthode de votre choix.

Une fois cette table créée et alimentée, l'analyse approfondie des données à l'aide de requêtes SQL peut commencer :

- Vérification des champs vides
- Vérification des types de chaque colonne (vous pouvez utiliser la commande `CONVERT`)
- Vérification de la redondance d'informations
- ...

Production attendue :

- Description de la méthode utilisée pour obtenir la table demandée
- Analyse détaillée des données colonne par colonne

Vous devez inclure l'ensemble des requêtes SQL utilisées ainsi que les photos d'écrans qui vous semblent nécessaires

Phase 2 :

Dans cette phase vous devez concevoir un modèle de données puis créer les tables et contraintes choisies. Vous alimenterez ensuite vos nouvelles tables en utilisant les données de la table de la phase 1.

Pour transférer ces données, vous pouvez utiliser les commandes « INSERT INTO MA_TABLE SELECT ... » vues en cours ainsi que la fonction « ROW_NUMBER ».

La fonction ROW_NUMBER permet de numéroter le résultat trié d'un SELECT ; plus d'information ici : <https://learn.microsoft.com/fr-fr/sql/t-sql/functions/row-number-transact-sql>

Exemple de syntaxe :

```
INSERT INTO GENRE_LIST
SELECT ROW_NUMBER() OVER(ORDER BY Genre) AS id, Genre FROM
(SELECT DISTINCT Genre FROM VGSALES) LISTE_DES_GENRES
```

Production attendue :

- Schéma de la base et justification des choix
- Script de création de la base et des contraintes
- Requêtes SQL permettant d'alimenter les tables

Phase 3 :

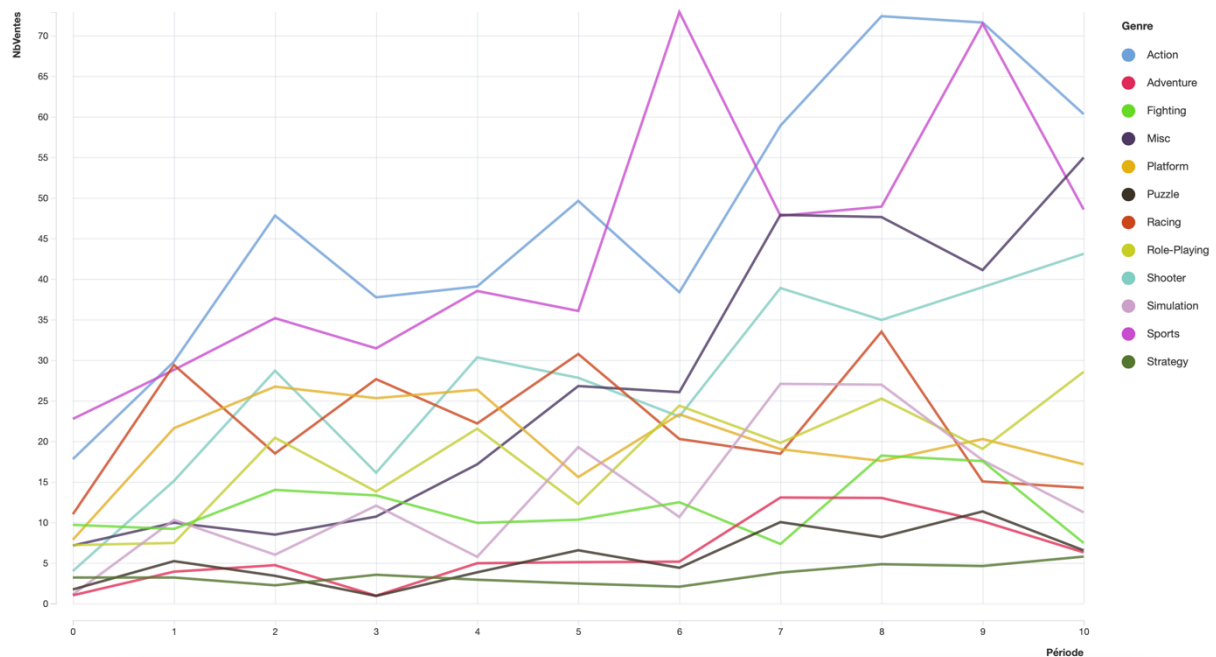
Dans cette phase vous devez écrire et exécuter des requêtes SQL sur votre modèle de données et produire des résultats sous la forme de graphiques. Ces graphiques seront obtenus grâce à l'application SQLPad disponible sur les ordinateurs de l'IUT. Cette application permet de publier sur le web des graphiques à partir de données obtenues avec des requêtes SQL (<https://getsqlpad.com>).

Production attendue :

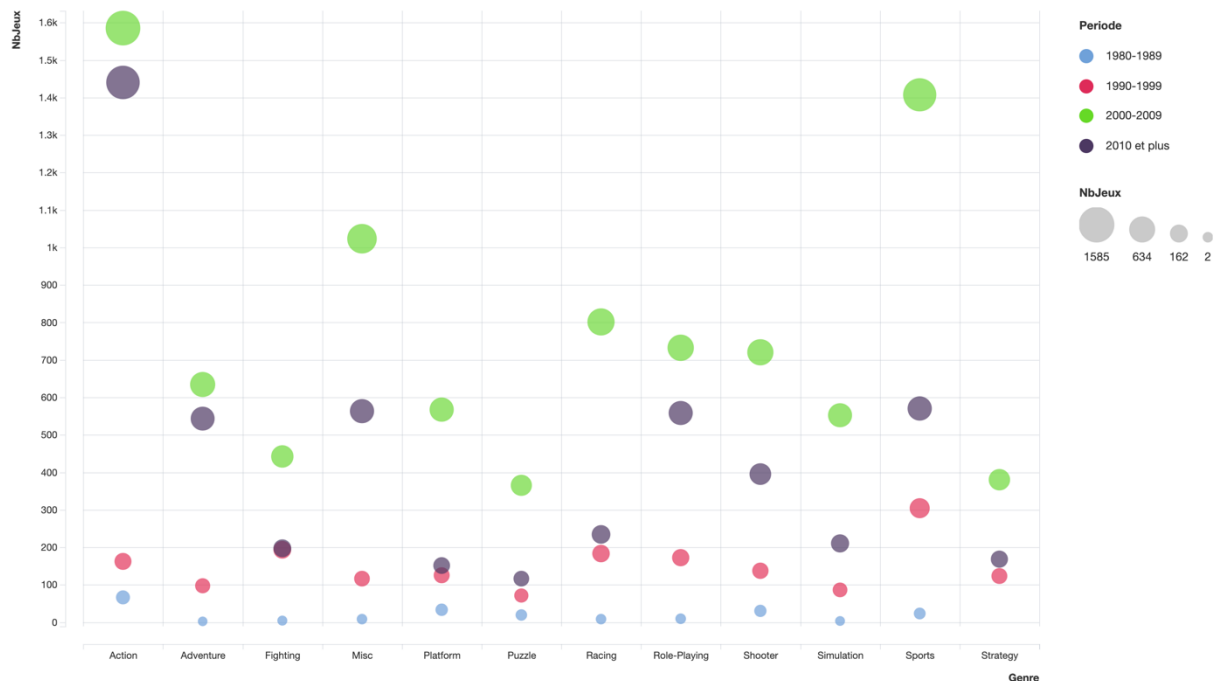
- L'ensemble des requêtes SQL et leurs résultats
- Vos graphiques SQLPad lorsqu'ils sont demandés

1. Affichez la liste des plateformes ayant plus de 20M de jeux vendus en Europe depuis 2010.
2. Affichez la liste des éditeurs de jeux PS4 dépassant les 2M de jeux vendus quelle que soit l'année, le résultat sera trié par nombre de jeux vendus du plus grand au plus petit.

- 3 Évolution du nombre de ventes de jeux par genre en Amérique du Nord sur la période 2000 à 2010. Pour cette requête vous construirez le graphique suivant :



- 4 Afficher le nombre de jeux édités en fonction du genre par tranche de 10 ans, soit 4 tranches : de 1980 à 1989, de 1990 à 1999, de 2000 à 2009, 2010 à maintenant. Pour cette requête vous construirez le graphique suivant :



- 5 Proposez au moins 2 requêtes supplémentaires d'analyse des données qui vous semblent intéressantes. Créez les graphiques adéquats correspondants.