

ISEN Nantes	CIPA 4 Projet Base de données	V 1.0
-------------	--	--------------

Projet BDD CIPA 4
« Gestion des installations photovoltaïques »
Cahier des charges
Octobre 2025

Introduction

Dans le cadre du cours de Base de Données, on vous demande de concevoir et d'interroger une base portant sur les installations photovoltaïques résidentielles (le jeu de données est réel). Les objectifs sont les suivants :

- construire un schéma relationnel en 3FN,
- nettoyer et homogénéiser les fichiers source, puis
- charger les données dans une base de données PostgreSQL, et ensuite
- répondre à un ensemble de requêtes d'analyse.

Les fichiers de base sont fournis, à savoir

- data.csv, qui reprend les installations, avec les informations suivantes : puissance crête, surface, marques, orientation, etc., et
- communes-france-2024-limite.csv, qui est le référentiel des communes, avec le code INSEE, la commune, le département, la région, et le code postal.

Exigences fonctionnelles

À partir des données CSV fournies, on vous demande

- de créer une base respectant les 3 premières formes normales (1FN, 2FN, 3FN),
- de nettoyer et d'homogénéiser les données (notamment les formats, les unités et les libellés),
- d'insérer les données dans une base de données PostgreSQL.

Requêtes d'interrogation des données

Créez les requêtes permettant de vérifier que votre BDD est valide et créez les requêtes répondant aux points suivants :

- 1) Afficher la puissance crête d'une installation selon la marque et le modèle des panneaux.
- 2) Afficher, afin de pouvoir les comparer, toutes les installations d'un même modèle de panneau (en affichant leurs caractéristiques détaillées : puissance, surface, orientation/pente, marque d'onduleur, et modèle d'onduleur).
- 3) Comparer 3 communes au choix : nombre d'installations, puissance totale et production moyenne par commune.
- 4) Déterminer, par marque d'onduleur, le nombre d'installations, la puissance moyenne, la puissance mini et maxi et la surface moyenne (arrondir à 2 décimales).
- 5) Déterminer, par orientation, le nombre d'installations et la puissance totale, ainsi que la répartition par pente (catégoriser les pentes par tranches à définir).
- 6) Déterminer les 20 installations à la puissance crête la plus élevée en affichant toutes leurs caractéristiques pertinentes (commune, département, marque panneaux/onduleur, surface,

orientation, pente).

- 7) Déterminer les 20 installations avec la densité (production/surface) la plus élevée.
- 8) Calculer le nombre d'installations dont la densité (puissance/surface) est inférieure à la moyenne globale.
- 9) Déterminer les 10 installations avec la densité (production/puissance) la plus faible.

Modification de la base de données

- 10) Ajoutez une colonne dans la base de données permettant d'enregistrer la date et l'heure de la dernière modification de l'enregistrement.

Questions complémentaires

- 11) On souhaiterait explorer le lien entre la surface des installations et leur puissance crête.
Proposez une ou plusieurs requêtes qui permettent d'explorer cette question et expliquez votre démarche (corrélations simples, regroupements par tranches de surface, etc.).
- 12) Choisissez 3 indicateurs pertinents et créez une vue par indicateur.
Exemples d'indicateurs :
 - puissance totale et moyenne par région/année ;
 - densité de puissance moyenne par commune ;
 - top N marques d'onduleurs par nombre d'installations.
- 13) (BONUS 1) Quel(s) autre(s) indicateur(s) proposeriez-vous ? Justifiez et écrivez la/les requête(s) correspondante(s) (ex. comparaison pente et pente optimum, orientation vs orientation optimum).
- 14) (BONUS 2) Choisissez 2 requêtes parmi toutes celles que vous avez mises au point pour le projet et essayez de les optimiser. Décrivez votre démarche et prouvez l'optimisation

Livrables

Les livrables à fournir sont les suivants :

- un planning initial et final des tâches accomplies (Gantt),
- un Modèle Conceptuel de Données initial, expliqué lors d'une présentation orale,
- un MLD issu du Modèle Conceptuel de Données précédent et au format textuel (non graphique) suivant : **Relation** (id, nom, prenom, #service)
- un Modèle Physique de Données sous forme de scripts SQL (création de tables, des clés, etc.),
- un rapport final, intégrant les points précédents et expliquant le processus de récupération des données, de nettoyage et d'insertion,
- les scripts d'insertion des données, un jeu de requêtes de tests, et les requêtes demandées avec systématiquement la requête et son résultat (ou au moins un extrait du résultat s'il est long),

- les explications concernant les questions complémentaires (ainsi que les vues accessibles dans la base de données),
- l'installation de la base de données sur le serveur mis à disposition

Livrables sur le Moodle :

- les documents précisés ci-dessus au format **PDF** (schémas, rapport, etc.)
- les fichiers de scripts au format **SQL**

Les fichiers devront comporter les noms et prénoms des membres du groupe !!!

Le rapport final devra être livré sur le Moodle pour le jeudi 16 octobre à 17h00.

Suggestion d'outils pour le Gantt :

- <https://www.onlinegantt.com/#/gantt>
- <https://www.ganttproject.biz>

Informations sur les données

Quelques informations sur les unités :

- Surface : Mètres carré
- Pente : degré
- Pente optimum : degré
- Orientation : degré par rapport au SUD (et pas au nord)
- Orientaton optimum : degré par rapport au SUD (et pas au nord)
- Production_pvgis : kWh annuel

Planning

Jour	Travail
Lundi 13/10	Début du projet
Mardi 14/10	Validation du MCD
Jeudi 16/10	Le rapport final devra être livré sur le Moodle
Vendredi 17/10	Présentation finale + remise du rapport final