

Schéma base Climato

1. But

Ce document décrit la structure de la base de données Climato. La base de données est stockée dans une base PostgreSQL avec l'extension TimeScaleDB.

2. Historique des mises à jour

- 2.0 : 21-04-2023 : re-écriture complète
- 3.0 : 6/1/2024 : re-écriture pour donner suite au changement de structure nécessaire pour accueillir les données de Météo France

3. postes : table des stations

Actuellement seuls les champs suivants sont utilisés par le serveur :

- **id** : Id de la station, référencée comme poste_id dans les autres tables
- **meteor** : code de la station au sein du réseau meteor.OI, Météo France
- **delta_timezone** : Nombre d'heures entre heure TU et heure locale (+4 pour La Réunion)
- **data_source** : 0: WeeWX/Meteor-OI, 1: Météo France
- **altitude/lat/long**: Mis à jour lors de la création automatique des stations Météo France, non utilisé par le serveur
- **other_code** : Contient le code Insee de la stations Météo France
- **stop_date** : Ne pas traiter les données venant après cette date, dans le chargement des JSON, ou de mise à jour à partir des backups
- **load_dump** : chargement des données à partir des dumps WeeWX
- **load_json** : booléen permettant d'activer ou non le chargement des fichiers Json.
 - o Si True -> Les backups ne sont plus chargés, et la mise à jour est faite exclusivement à partir des fichiers JSON
 - o Si False -> La base de données est mise à jour journalièrement à partir des mises à jour des dumps venant de la station. Les fichiers JSON sont ignorés, mais sont déplacés dans le répertoire « archive »
- **last_obs_date** : datetime UTC de l'obs la plus récente. Est maintenue à jour via un trigger sur la table obs. Est utilisé lors du chargement à partir des fichiers csv, à partir de l'API Météo France, à partir des dumps WeeWx.
- **last_obs_id** : id de l'obs la plus récente
- **last_extremes_date** : date de l'extrêmes la plus récente. Est maintenue à jour via un trigger sur la table extrêmes. Est utilisé seulement lors des importations à partir des dumps
- **last_extremes_id** : id de l'extrêmes la plus récente

4. mesures : table des mesures

**** DOIT ETRE REVU SUIVANT LES MESURES PRIS EN COMPTE ****

Cette table sert au paramétrage du serveur, et éventuellement pour récupérer des propriétés des mesures au niveau de l'interface. Voici les champs de la table :

- **id** : Id de la mesure, référencé comme mesure_id dans les autres tables
- **name** : nom de la mesure, seulement pour l'interface utilisateur

- **json_input** : nom de la colonne dans la table « obs » stockant la valeur de la mesure. Ce nom est aussi le nom de la clé utilisé dans le Json de mise à jour des mesures venant des stations
- **json_input_bis** : Si le Json ne contient pas la clé « json_input », ce synonyme peut être utilisé (Certaines mesures ont des noms différents entre la clé Json et le nom de la colonne, pour des raisons historiques)
- **archive_col** : nom de la colonne dans la table archive de WeeWX (utilisé par le programme de migration seulement)
- **archive_table** : Forçage du nom de la table pour retrouver les extrêmes dans la base de WeeWX. Par défaut, quand cette colonne est nulle, le nom du champ de WeeWX est préfixé avec « archive_day_ » pour retrouver le nom de la table de WeeWX stockant les valeurs extrêmes journaliers
- **is_avg** : booléen qui indique si cette mesure doit être agrégée en calculant les moyennes, ou par en faisant la somme des valeurs. Ce champ est utilisé par le programme utilisateur pour utiliser la bonne méthode d'agrégation.
- **is_wind** : indique s'il faut aller chercher la direction du vent dans la table des extrêmes de WeeWX
- **allow_zero** : non utilisé actuellement.
- **is_hourly** : indique que la mesure est mise à jour qu'une fois par heure, donc qu'une agrégation par minutes ne fait pas de sens. Non utilisé actuellement

5. obs : table des observations

Contient toutes les mesures élémentaires remontées par la station, et/ou issu de la base WeeWX de la station. Colonnes de la table :

- **id** : Id de l'observation, référencé comme obs_id dans les autres tables
- **date_local** : date et heure locale de la mesure
- **date_utc** : date et heure UTC de la mesure
- **poste_id** : id du poste concerné par la mesure
- **mesure_id** : id de la mesure
- **duration** : durée de la période d'intégration de la station (LOOP). Est en général de 5mn pour les données venant de WeeWx, 6mn pour les données venant de l'API Météo France, 60 mn pour les données horaires venant de météo France.
- **value** : valeur de la mesure (type float)
- **qa_value** : Indice de qualité de la mesure :
 - o **0** => Non renseigné
 - o **1** => Validé (par météo France, ou operateur Meteor.OI)
 - o **9** => Invalide

Les mesures invalides sont automatiquement supprimées des vues matérialisées (materialized views)

6. x_min: table des valeurs minimales

La table des valeurs minimales (records minimaux) garde par mesure, les valeurs extrêmes minimales relevées. Les minimaux peuvent venir de :

- o Mesure d'observation, lorsqu'aucune mesure extrême n'est fournie
- o Valeur extrême fournie (dans ce cas obs_id est null)

- **id** : Id de la valeur mini
- **date_local** : date et heure locale de la mesure mini
- **poste_id** : id du poste concerné par la mesure mini
- **mesure_id** : id de la mesure mini
- **obs_id**: id the l'observation associée si le minimal provient d'une valeur mesurée
- **min**: Valeur minimale
- **min_time**: datetime de la valeur minimale
- **qa_min** : indice de qualité de la valeur minimale

Seules les valeurs minimales avec un indice de qualité valide/non spécifié sont pris en compte dans les vues matérialisées. L'indice de qualité d'un minimal provenant d'une mesure, est maintenu à jour avec l'indice de qualité de l'observation (si la mesure est invalidée, le minimal associé est invalidé)

7. x_max: table des valeurs maximales

La table des valeurs maximales (records maximaux) gardé par mesure, les valeurs extrêmes maximales relevées. Les maximaux peuvent venir de :

- Mesure d'observation, lorsqu'aucune mesure extrême n'est fournie
- Valeur extrême fournie (dans ce cas obs_id est null)

- **id** : Id de la valeur mini
- **date_local** : date et heure locale de la mesure mini
- **poste_id** : id du poste concerné par la mesure mini
- **mesure_id** : id de la mesure mini
- **obs_id**: id the l'observation associée si le minimal provient d'une valeur mesurée
- **max**: Valeur maximale
- **max_time**: datetime de la valeur maximale
- **max_dir** : direction du vent de la valeur maximale, pour les mesures de type vent
- **qa_max** : indice de qualité de la valeur maximale

Mêmes remarques que pour la table x_min

8. incidents : table des incidents

Le programme serveur stocke dans cette table les incidents rencontrés. Les incidents sont classés par

- **id** : Id de l'incident
- **dat** : date et heure de l'incident
- **source** : information sur la fonction ayant généré cet incident
- **level** : exception, error, info
- **reason** : message de l'incident
- **details** : Json contenant des informations complémentaires à cet incident. En cas d'exception pendant l'exécution, le stack est stocké dans le json.
- **active** : booléen pour indiquer si l'incident a été vu et traité (donc active -> false)

9. annotations : table des annotations

Cette table est utilisée par Grafana pour stocker des annotations sur les mesures relevées, comme des événements météorologiques spéciaux
Colonnes :

- **time** : date et heure du début de l'annotation
- **timeend** : date et heure de la fin de l'annotation
- **text** : description de l'annotation
- **tags** : tag de l'annotation

10. obs_hour/obs_day/obs_month : vue des observations

Chacune de ces vues contient une agrégation des mesures de la table obs. Les mesures invalides sont exclues de ces vues. Elle est automatiquement tenue à jour par TimeScaleDB. Elle contient les colonnes suivantes :

- **id** : Id de la ligne de la vue
- **date_local**: début de la période en date locale de la mesure agrégée
- **poste_id**: id du poste concerné
- **mesure_id** : id de la mesure concernée
- **is_avg** : Est-ce une mesure moyennée, sinon sommée
- **duration** : somme des 'durations' utilisées pour calculer cette agrégation
- **value** : Valeur de la mesure agrégée (par somme ou moyenne, suivant la mesure)

Les vues sont calculées les unes par-dessus les autres. Il est très fortement conseillé d'utiliser la vue correspondant au niveau d'agrégation souhaité, pour limiter les ressources du serveur, et améliorer le temps de réponse. Une vue par années peut être ajoutée si besoin est.

11. x_min_day/x_min_month : vue des records minimaux

Chacune de ces vues contient une agrégation des mesures de la table x_min. Les mesures invalides sont exclues de ces vues. Elle est automatiquement tenue à jour par TimeScaleDB. Elle contient les colonnes suivantes :

- **id** : Id de la ligne de la vue
- **date_local**: début de la période en date locale de la mesure agrégée
- **poste_id**: id du poste concerné
- **mesure_id** : id de la mesure concernée
- **min** : valeur minimale de la période d'agrégation
- **min_time** : datetime du minimal relevé

Les vues sont calculées les unes par-dessus les autres. Il est très fortement conseillé d'utiliser la vue correspondant au niveau d'agrégation souhaité, pour limiter les ressources du serveur, et améliorer le temps de réponse. Une vue par années peut être ajoutée si besoin est.

12. x_max_day/x_max_month : vue des records maximaux

Chacune de ces vues contient une agrégation des mesures de la table x_max. Les mesures invalides sont exclues de ces vues. Elle est automatiquement tenue à jour par TimeScaleDB. Elle contient les colonnes suivantes :

- **id** : Id de la ligne de la vue
- **date_local**: début de la période en date locale de la mesure agrégée
- **poste_id**: id du poste concerné
- **mesure_id** : id de la mesure concernée
- **max** : valeur maximale de la période d'agrégation
- **max_time** : datetime du maximal relevé

Les vues sont calculées les unes par-dessus les autres. Il est très fortement conseillé d'utiliser la vue correspondant au niveau d'agrégation souhaité, pour limiter les ressources du serveur, et améliorer le temps de réponse. Une vue par années peut être ajoutée si besoin est.