

1. Historique

- v0.50 : premier draft
- v0.51 : clean des termes « mesure instantanée » et « donnée élémentaire »
- v0.52 : duration -> interval, data logger -> système d'acquisition, record -> extrême
- v0.55 : Révision majeure chapitres 5,6,7,8. Nouveaux chapitre 9/10
 - v0.56 : Intégration des fonctions de filtres/contrôle, et nouveaux champs de suivi qualité. Interval -> duration
- v0.57 : Ajout d'agrégation sur soilMinsoilMinTime. Suppression de l'agrégation de barometerMax et barometerMaxTime. Suppression de rainRate car rainRateMax est utilisée. Ajout d'une annexe C, avec les questions non répondues
- v0.58 : Ajout windChillMax, windChillMaxTime, barometerMax, barometerMaxTime, rainRateAvg, rainSumDuration, Annexe C,D
- v0.59 : Annexe A devient : Json ou colonnes (a completer)
- v0.60 : Ajout remarques Hubert dans Annexe E

2. TOC

1. Historique	1
2. TOC	1
3. Données de la VP2	2
a. Capteurs	2
b. Système d'acquisition de données	4
4. WeeWX	4
a. Période \$current	5
b. Période \$hour	5
c. Période \$day	5
d. Période \$month	5
e. Période \$year	5
f. Période \$alltime	5
5. Stratégie d'agrégation des données dans BD Climato	5
a. WeeWX	6
b. Agrégation seule, sans mesures	6
6. Gestion des extrêmes Mensuels/Annuels/Globaux	6
a. Extrêmes dans les données élémentaires ou pas	6
b. Gestion des extrêmes à l'insertion (importation)	6
c. Stratégie de régénération d'extrême en cas de suppression	7
7. Spécification du fichier JSON d'importation	8
a. Field format	8

BD Climato, Le voyage des données!

b.	header	8
c.	current	8
d.	agregation	10
8.	Schéma Base	12
a.	Poste	12
b.	Type Exclusion	13
c.	Tables observation	13
d.	Aggregation	13
9.	Structure du code	14
a.	Object Poste	14
b.	Objet TypeInstrument	14
10.	Algorithme importation	15
a.	Fonction ImportMeasures	15
b.	Fonction DeleteMeasure	15
ANNEXE A :	JSON ou Colonnes	16
ANNEXE B :	Date implementation	17
ANNEXE C :	Intégration des partenaires	18
ANNEXE D :	Modification des données (UI)	19
ANNEXE E :	Questions non réglées	20

3. Données de la VP2

Comme la plupart des stations du réseau MeteoR.OI sont des VP2, il est important de comprendre ce que la VP2 envoie comme information

a. Capteurs

	FONCTION	Standard	Capteur nécessaire	Plage
Pression barométrique	Pression barométrique (Élévation : -460 à +3650 m)	<input checked="" type="checkbox"/>	Inclut dans les consoles n°6312 ou 6316	880 à 1080 hPa, 660 à 810 mm
	Tendance barométrique (3 heures)	<input checked="" type="checkbox"/>		5 positions de la flèche : Montée (rapide ou lente), constante, ou descente (rapide ou lente)
ET	Évapotranspiration		Radiation solaire, thermo/hygromètre, ou	Jour : 999,9 mm; mois & année 1999,9

BD Climato, Le voyage des données!

	FONCTION	Standard	Capteur nécessaire	Plage
			ISS	mm
Humidité & point de rosée	Humidité intérieure	<input checked="" type="checkbox"/>	Inclut dans les consoles n°6312 ou 6316	10 à 90 %
	Humidité extérieure	<input checked="" type="checkbox"/>	Capteur température/humidité ou ISS	0 à 100 %
	Point de rosée (général)	<input checked="" type="checkbox"/>	Température/humidité ou ISS	-76° à +54°C
	Point rosée/gel élevé en humidité	<input checked="" type="checkbox"/>	Température/humidité ou ISS	-76° à +54°C
	Extra humidité		Température/humidité ou ISS	0 à 100 %
Humectation du feuillage	Humectation		Humectomètre n°6420 et station auxiliaire n°6345OV	0 à 15
Précipitation	Journalière et précipitation d'orage	<input checked="" type="checkbox"/>	Pluviomètre	0 à 999.9 mm
	Précipitations mensuelles et annuelles	<input checked="" type="checkbox"/>	Pluviomètre	0 à 19,999 mm
	débit de la pluie	<input checked="" type="checkbox"/>	Pluviomètre	2540 mm/h
Humidité du sol	Humidité du sol		sonde Watermark Humidité de sol n°6440 et station auxiliaire n°6345	0 to 200 cb
Rayonnement solaire	Puissance / énergie solaire		Pyranomètre capteur de rayonnement global n°6450	0 à 1800 W/m2
Température	Température intérieure	<input checked="" type="checkbox"/>	Inclut dans les consoles n°6312 ou 6316	
	Température extérieure	<input checked="" type="checkbox"/>	Capteur de température, sonde de température, capteur thermo/hygromètre, ou ISS	-40°C à + 60°C

	FONCTION	Standard	Capteur nécessaire	Plage
	Température extra		Capteur ou sonde de température n°6470. et station auxiliaire n°6345 et 6372	-40° à +60°C
Température apparente	Index de chaleur	<input checked="" type="checkbox"/>	Capteur Thermo/hygromètre ou ISS	-40° à +57°C
	Index d'humidité - soleil - vent		Capteur de radiation solaire, anémomètre & thermo/hygromètre ou ISS	-68° à +64°C
Heure et date	Heure	<input checked="" type="checkbox"/>	Inclut dans les consoles n°6312 ou 6316	24 heures
	Date	<input checked="" type="checkbox"/>		Jour/mois
Radiation UV	Index UV		Capteur rayonnement UV n°6490	0 à 16 index
	Dose UV		Capteur rayonnement UV n°6490	0 à 199 MEDs
Vent	Direction du vent	<input checked="" type="checkbox"/>	Girouette	0° à 360°
	Rose des vents	<input checked="" type="checkbox"/>	Girouette	division en 16 points
	Vitesse du vent (grandes coupelles)	<input checked="" type="checkbox"/>	Anémomètre	3 à 193 km/h ; 1 à 54 m/s
	Vitesse du vent (petites coupelles)		Anémomètre	5 à 282 km/h ; 1.5 à 79 m/s
	Refroidissement dû au vent (Wind chill)	<input checked="" type="checkbox"/>	Anémomètre et sonde de température, capteur de température ou thermo/hygromètre	-84° à +54°C

b. Système d'acquisition de données

La VP2 envoie des mesures instantanées toutes les deux secondes. Chaque envoi contient des mesures partielles, une fois la température, une autre fois la pression, etc... Le système d'acquisition doit regrouper ces informations partielles pour avoir une vision de tous les capteurs. Cela se faisant sur une certaine période, certaines valeurs doivent être agrégées à partir des mesures instantanées.

Chaque système d'acquisition peut avoir une stratégie différente dans les détails, mais ont globalement un fonctionnement assez semblable. Les mesures instantanées sont rarement

BD Climato, Le voyage des données!

archivées. Seule la version agrégée est stockée et est utilisée par les interfaces graphiques du système d'acquisition. Ce sont ces données agrégées que le système d'acquisition va stocker, ou distribuer. Nous appellerons ces données les **données élémentaires**, pour ne pas les confondre avec les agrégations faites dans la BD Climato.

A partir des mesures instantanées le système d'acquisition détecte les Min/Max, et garde en mémoire ces valeurs extrêmes, ainsi que l'heure ou cet extrême a été détecté. D'autres informations, suivant les instruments, comme la direction du vent au moment de l'extrême sont aussi gardées en mémoire.

4. WeeWX

WeeWX reçoit les mesures instantanées, appelé « LOOP » (correspond aux mesures instantanées). Un « LOOP » va contenir les informations de température, le suivant sur la pluie, le suivant la température, puis le vent... « LOOP » ne contient que des informations partielles sur un domaine donné. WeeWX analyse ces données et les sauve de deux façons :

- **Données élémentaires**, regroupement sur une période de 5mn, stockées dans la table « archive »
- **Valeurs Extrêmes** ainsi que des informations complémentaires (heure de l'extrême, direction du vent au moment de l'extrême relevé pour le vent...) dans des tables « records » spécifiques suivant différents niveaux de regroupement (daily, month,...).
Les extrêmes sont archivés par jour seulement.

Plus d'info (anglais) sur : <http://www.weewx.com/docs/customizing.htm>

Avec WeeWX les mesures instantanées ne sont pas accessibles (sauf programmation spécifique comme on verra).

Un « template » est utilisé pour générer un fichier JSON tous les 5mn, qui est envoyé vers le serveur Climato. Le « template » utilisé par WeeWX est défini dans (cela va changer dans le futur):

`/srv/weewx/root/skins/Bootstrap/json/daily.json.tmpl`

Le fichier JSON généré est stocké dans le raspberry dans `/srv/weewx/root/public_html/json`, avant envoi vers le serveur BD Climato.

Les extrêmes sont archivés par jour, puis agrégés par mois/année/global. On verra l'importance de cela en cas de modification d'une mesure.

Les « Templates » utilisent les variables de weeWX, préfixées par une période.

a. Période \$current

Pour faire simple et concis, \$current représente le dernier enregistrement archive (la donnée élémentaire). Par exemple \$current.rainSum correspond à la somme des pluies relevée sur la fréquence de génération des archives (5mn).

Pour plus de détail : <http://www.weewx.com/docs/customizing.htm>, recherche sur le texte : « time period \$current »

BD Climato, Le voyage des données!

b. Période \$hour

Données agrégées par heure. Ces données viennent de la base de données, de la table archive. Il est possible de décaler le \$hour, de prendre l'heure précédente... Par défaut, les extrêmes par heure sont simulés en prenant les min/max à partir des archives (agrégation de 5mn), car non stockés dans la base de données de WeeWX.

c. Période \$day

Données agrégées et extrême par jour. Ces données viennent de la table archive_day. Le extrêmes sont les extrêmes relevés.

d. Période \$month

Données agrégées et extrêmes par mois. Les extrêmes sont calculés à partir des extrêmes du jour du mois concerné.

e. Période \$year

Données agrégées et extrêmes par année. Les extrêmes sont calculés à partir des extrêmes du jour de l'année concernée.

f. Période \$alltime

Données agrégées et extrêmes globaux. Les extrêmes sont calculés en prenant les min/max de toutes les valeurs extrêmes journalières archivées.

5. Stratégie d'agrégation des données dans BD Climato

En général les agrégations vont être calculées par la fonction d'importation.

Seule la future donnée « durée d'ensoleillement », et ses agrégations seront récupérées de WeeWX, car l'agrégation nécessite un calcul spécifique pour son agrégation. Cela sera à approfondir quand on ajoutera le plug-in de supportant cette mesure calculée par WeeWX.

a. WeeWX

Le fichier JSON généré toutes les 5 mn (en fait tous les « duration » de temps) va contenir les données élémentaires. Il faut noter que cette période de 5mn pourra être différent pour certaines stations, ainsi qu'en cas de resynchronisation d'une base WeeWX régénérée à partir des données externes (Par exemple WUunderground ne garde que les données agrégées toutes les 15mn environ, ou pour une reprise de données manquantes, ou historique).

En cas de reprise d'historique, ou de correction de données erronées déjà importées, seules les informations \$current seront présentes dans le fichier JSON.

Pour avoir une fonction d'importation générale qui soit capable d'importer et de corriger des données erronées, il est préférable de calculer les données agrégées à partir des données \$current, sans utiliser les données agrégées par weeWX, sauf cas particulier. Les agrégations pour l'heure-jour-mois-année courant(e) représente une agrégation partielle, car la période est en cours, et non terminée. L'agrégation devient complète seulement quand la période est terminée (si toutes les données sont présentes). Certaines agrégations journalières, ont un décalage dans le temps.

BD Climato, Le voyage des données!

Pour une agrégation de type somme (cumul de pluie...) il suffit d'ajouter/soustraire la valeur en cas d'insertion/suppression de données mesure. Pour les sommes journalières, les sommes pourront partir de 0hTU, 0h local, ou d'une autre heure.

Pour une agrégation de type moyenne il sera nécessaire de garder la somme des valeurs, ainsi que la somme des durées, en plus de la moyenne calculée (qui pourra être une colonne calculée dans la BD).

b. Agrégation seule, sans mesures

Il y a une station qui a des données historiques agrégées par mois. Aussi certaines informations de météo France nous seront fournies agrégées par heure. Dans ces deux cas, il n'y aura pas de section « Current » dans le fichier JSON

Dans ce cas l'importation devra importer directement la valeur agrégée dans la table correspondant à la période d'agrégation (heure, jour...). Les agrégations supérieures seront automatiquement calculées de la même façon que dans le cas général.

6. Gestion des extrêmes Mensuels/Annuels/Globaux

Les extrêmes sont différents des min/max des données élémentaires.

a. Extrêmes dans les données élémentaires ou pas

WeeWX ne sait pas fournir les extrêmes dans les données élémentaires, mais sait les fournir par jour. Il est prévu dans une version ultérieure de pouvoir fournir cette information, en développant un plug-in pour WeeWX. Certaines stations ne savent pas fournir de valeurs extrêmes. La difficulté arrive quand une observation est supprimée (pour être modifiée à la suite d'une erreur). Dans ce cas il peut être nécessaire d'ajuster les valeurs extrêmes.

b. Gestion des extrêmes à l'insertion (importation)

Les valeurs extrêmes peuvent être fournies de façons suivantes :

Cas 1 : Avec les données élémentaires, avec le futur plugin WeeWX

Cas 2 : Agrégés par heure/jour/mois (par jour pour WeeWX sans plug-in). Ce niveau d'agrégation est appelé « agrégation minimale de l'extrême »

Cas 3 : Non fournies, ce qui sera le de nouvelles données corrigeant une observation supprimée

Règle générale : « Si on a la donnée extrême on l'utilise, sinon on la calcule à partir des Min/Max des informations élémentaires. »

Dans tous les cas, lors de l'importation de données, si un Min/Max est au-dessous/au-dessus d'un extrême, il faut le considérer comme nouvel extrême. Cela peut arriver à la suite d'une suppression d'une donnée contenant un extrême, suivi d'une insertion d'une nouvelle mesure corrigée. (Je peux détailler au besoin ☺)

c. Stratégie de régénération d'extrême en cas de suppression

On a besoin de régénérer un extrême, **seulement** dans le cas d'une suppression d'une observation couvrant une période pendant laquelle un extrême a été

BD Climato, Le voyage des données!

établi.

Il faut établir dans quel cas de gestion des extrêmes on se trouve.

i. Cas 1 : Extrêmes dans les observations

Il est facile de retrouver les valeurs extrêmes d'une période, en parcourant la période de niveau inférieur, ou les observations pour les extrêmes par heures.

ii. Cas 2, suivant le niveau d'agrégation cible

1. Extrême agrégé plus petit que « l'agrégation minimale de l'extrême »

On doit retrouver le Min/Max à partir des observations, ce qui va donner à une perte d'information. (Par exemple pour régénérer un extrême par heure avec WeeWX actuel, il est nécessaire de retrouver les min/max à partir des observations.

2. Extrême agrégé égal ou supérieur à « l'agrégation minimale de l'extrême »

Il faut repartir des extrêmes de la sous période correspondant. Par exemple avec WeeWX actuel, pour recalculer les extrêmes mensuels, il faut repartir des extrêmes journaliers.

iii. Cas 3 : Pas d'extrêmes

Quel que soit le niveau, il faut retrouver le Min/Max à partir des observations

iv. Conclusion

Deux algorithmes sont nécessaires :

- recalcul de l'agrégation à partir d'une agrégation inférieure
- régénération des agrégations sur une période donnée à partir des observation.

Le plus délicat étant de savoir dans quel cas on se trouve. Pour une 1ere version, il serait possible de stocker cette information comme un champs de la table poste

7. Spécification du fichier JSON d'importation

En rouge les nouveaux champs

a. Field format

Toutes les valeurs sont stockées dans une 'string'. Format UTF-8.

Les dates sont en UTC, sous la forme : « yyyy-mm-dd hh:mm:ss »

Les coordonnées géographiques sont au format : « +/-dd :mmm,ss»

BD Climato, Le voyage des données!

b. header

title	Météo en direct
poste_id	** Nouveau **
location	\$station.location
dat	\$current.dateTime .New name
lat	\$station.latitude[0]° \$station.latitude[1]' \$station.latitude[2]
lon	\$station.longitude[0]° \$station.longitude[1]' \$station.longitude[2]
alt	\$station.altitude
alt_unit	«meter»
hardware	\$station.hardware
uptime	\$station.uptime
serverUptime	\$station.os_uptime
weewxVersion	\$station.version
duration	Durée de la période d'agrégation de la donnée élémentaire

c. current

Ces données correspondent aux données élémentaires, quand il y en a. Si cette section est manquante, le programme d'importation ne devra tenir compte que des données agrégées.

Si une clef n'est pas fournie, la valeur est considérée comme nulle

Si l'intervalle (duration) est petit, les valeurs peuvent être des valeurs agrégées sur l'intervalle, ou instantanées si cela n'est pas possible. Cependant si l'intervalle est grand (une heure ou plus), les valeurs doivent représenter une valeur agrégée sur la période

Clef JSON	Commentaire
outTemp	
outTempMax	Seulement si extrêmes avec les observations
outTempMaxTime	Seulement si extrêmes avec les observations
outTempMin	Seulement si extrêmes avec les observations
outTempMinTime	Seulement si extrêmes avec les observations
inTemp	<u>Donnée non météo</u>
windchill	
heatIndex	
dewpoint	
humidity	
humidityMax	Seulement si extrêmes avec les observations
humidityMaxTime	Seulement si extrêmes avec les observations
humidityMin	Seulement si extrêmes avec les observations
humidityMinTime	Seulement si extrêmes avec les observations
inHumidity	Name change for consistency <u>données non météo</u>
barometer	Sea-level pressure
barometerMax	Seulement si extrêmes avec les observations <u>on garde ?</u>

Deleted: Value

Deleted: Value

BD Climato, Le voyage des données!

barometerMaxTime	Seulement si extrêmes avec les observations <u>idem</u>
barometerMin	Seulement si extrêmes avec les observations
barometerMinTime	Seulement si extrêmes avec les observations
pressure	Station level pressure
windSpeed	
windDir	
windGust	Seulement si extrêmes avec les observations
windGustTime	si non fourni mais windGust fourni : prend le milieu de la periode
windGustDir	Seulement si extrêmes avec les observations
rainRateMax	Seulement si extrêmes avec les observations
rainRateMaxTime	Seulement si extrêmes avec les observations
rainSum	
rainSum1h	Sur une heure <u>avant</u> l'heure de l'observation
rainSum3h	Sur 3 heures <u>avant</u> l'heure de l'observation
rainSum6h	Sur 6 heures <u>avant</u> l'heure de l'observation
rainSum12h	Sur 12 heures <u>avant</u> l'heure de l'observation
rainSum24h	Sur 24 heures <u>avant</u> l'heure de l'observation
soilTemp	
soilTempMin	Seulement si extrêmes avec les observations
soilTempMinTime	Seulement si extrêmes avec les observations
UV	<u>c'est dose moyenne UV ou c'est indice UV ? Dans la table archive UV semble être l'indice.</u>
UV indice	<u>idem</u>
ETP	Changement de nom <u>TRES IMPORTANT</u> . En fait dans la table archive l'ETP n'est disponible que toutes les heures. C'est donc TRES probablement l'ETP horaire, ce qui paraît logique car on ne calcule pas l'ETP toutes les 5 minutes.
solarRadiation	<u>Rayonnement global moyen sur l'intervalle en W/m² c'est à dire en J/s/cm², on peut considérer que c'est de l'instantané. Pour avoir la valeur cumulée il faut multiplier par la durée de l'intervalle en secondes et ramener en W/cm² pour diminuer la taille du nombre obtenu. A voir si là on garde la valeur « instantanée » ou le cumul car pour la suite des agrégations le cumul est plus pertinent.</u>
InsolationDuration	A définir... avec les agrégations nécessaires...

Deleted: à partir de

Deleted: à partir de

Deleted: à partir de

Deleted: à partir de

Deleted: à partir de

La table observation

d. agregation

Le niveau de l'agrégation est indiqué dans la clef « agregation ». Les valeurs pour un seul niveau peuvent être données, les agrégations de niveau supérieur étant calculé lors de l'importation

Si les données extrêmes ne sont pas fournies dans la section « current », elles peuvent être données ici, cela correspond au « niveau d'agrégation minimale » expliqué auparavant.

BD Climato, Le voyage des données!

Aussi dans certains cas spécifiques décrits auparavant, le fichier JSON peut ne contenir que des données d'agrégation, sans avoir de données observations (« Current »).

agregation	« heure », « jour », « mois », « annee » ou « global »
	Type Temp
outTempAvg	Calculé : outTemp/outTempInterval
outTempSum	<u>Ne peut on pas s'en passer puisqu'on a Avg et Interval ?</u>
outTempInterval	<u>c'est le temps (minutes ou secondes) pendant lequel la donnée est disponible ou c'est le nombre de données élémentaires disponibles ?</u>
outTempMax	
outTempMaxTime	
outTempMin	
outTempMinTime	
heatIndexMax	
heatIndexMaxTime	
windChillMax	<u>Windchill est toujours \leq outTemp, donc le max a t'il du sens ?</u>
windChillMaxTime	
windChillMin	
windChillMinTime	
Humidity	?? make sense ?? -> add sum and sum interval....
humidityMax	
humidityMaxTime	
humidityMin	
humidityMinTime	
dewPointMax	
dewPointMaxTime	
dewPointMin	
dewPointMinTime	
inTempMax	Donnée non météo
inTempMaxTime	-
inTempMin	-
inTempMinTime	-
inHumidityMax	-
inHumidityMaxTime	-
inHumidityMin	-
inHumidityMinTime	-
soilMin	
soilMinTime	
	Type Pression
barometerAvg	Calculé : barometerSum/barometerInterval
barometerSum	<u>Voir outTemp</u>
barometerInterval	<u>Voir outTemp</u>

BD Climato, Le voyage des données!

barometerMax	
barometerMaxTime	
barometerMin	
barometerMinTime	
	Type Rain
rainSum	
rainRainAvg	Calculé à partir de (rainSum * 60) / rainSumDuration <u>à mon sens pas pertinent</u>
rainSumDuration	<u>Nuance avec Interval ? Même question que outTemp</u>
rainRateMax	
rainRateMaxTime	
	Type Wind
windAvg	Calculé : windSum/windInterval
windSum	
windInterval	<u>Voir outTemp</u>
windMax	
windMaxDir	
windMaxTime	
windRms	?? <u>à mon sens pas pertinent</u>
windVecAvg	?? <u>idem</u>
windVecDir	?? <u>idem</u>
	Type Solar
UVMax	
UVMaxTime	
UVMin	?? make sense ?? <u>sur l'heure pourquoi pas mais au-delà toujours à 0.</u>
UVMinTime	-
radiationMax	<u>Radiation c'est un cumul à partir de l'heure... à discuter</u>
radiationMaxTime	
radiationMin	?? make sense ?? <u>idem UVmin</u>
radiationMinTime	-
etpAvg	????? <u>etp se cumule comme la pluie. Compte tenu remarque plus haut on ne dispose pas d'ETp en-dessous de l'heure. Je vois plus etpSum ou etp tout court. A partir du mois pourquoi pas des min/max quotidien.</u>
etpMax	
etpMaxTime	
etpMin	
etpMinTime	
	Type Divers
rxAvg	Calcule : rxSum / rxInterval <u>donnée non météo</u>
rxSum	-
rxInterval	-

BD Climato, Le voyage des données!

rxMax	-
rxMaxTime	-
rxMin	?? -
rxMinTime	-
voltAvg	Calculé : voltSum/voltInterval -
volSum	-
voldInterval	-
VoltMax	-
VoltMaxTime	-
VoltMin	-
VoltMinTime	-

8. Schéma Base

a. Poste

Deux parties, ce qui est vraiment nécessaire, et ce qui ne sera pas utilisé pour le moment

id		Serial, pk (no interne a bd climato)
meteor		Code ou nom meteor
meteoFR		Code meteoFR, si envoyé à météo France
title		Nom usuel de la station ?? (n'est-ce pas meteor ??)
type		Type station (vp2,...)
cas_gestion_extreme		1,2,3 (cas de gestion de l'extreme)
agg_min_extreme		Niveau d'agrégation minimum des extrêmes envoyé dans Json dans le cas 2
		Données non fondamentales...
owner		Nom du responsable de la station
email		Email
phone		Tel (+xxx yyyyyyyyyyy)
address		Texte libre pour adresse
zip		
city		
country		
latitude		Stocke sous le format : ????
longitude		
start		Date integration dans le reseau
end		Default = date(2100,12,31)
comment		Long texte libre

b. Type Exclusion

id	int	Serial, pk
poste_id	smallint	
type_data	varchar(20)	temp, wind, rain....
start_x	date	
end_x	date	Default = date(2100,12,31)
model_value	jsonb	Json, with {columnName : « type »} qui sera utilisé dans les valeurs par défaut.
value	jsonb	Json avec {columnName : « valeur par défaut »}

Le model_value va être utilisé pour demander les valeurs par défaut à fournir, ou au moins dans une routine de validation de l'input. Toute donnée non spécifiée aura un défaut null.

c. Tables observation

poste_id	id du poste
dat	Date de l'observation
last_rec_dat	Date de dernière mise à jour
duration	Durée en secondes
qa_modifications	Nombre de mesures contenant des modifications par la fonction de filtrage ou trigger
qa_incidents	Nombre de mesures contenant des données rejetées par la fonction de contrôle qualité
qa_check_done	Agrégation vérifiée par le contrôle qualité
j	Jsonb contenant le json de l'observation

d. Aggregation

poste_id	id du poste
dat	Date du début de l'agrégation
last_rec_dat	Date de dernière mise à jour
duration	Durée en secondes
qa_modifications	Nombre d'agrégations inférieures contenant des modifications par la fonction de filtrage ou trigger
qa_incidents	Nombre d'agrégations inférieures contenant des données rejetées par la fonction de contrôle qualité
qa_check_done	Agrégation vérifiée par le contrôle qualité
j	Jsonb contenant le json de l'agrégation

9. Structure du code

a. Object Poste

i. Static (ça existe/python?)getInfo(posteld) : postInfo

➔ Retrouve les données postes + tableau des type invalidés et valeurs par défaut

BD Climato, Le voyage des données!

ii. `cas_gestion_extreme` : int

Retourne le `cas_gestion_extreme`. Pourra être implémenté dans le future en scannant les données observations

iii. `agg_min_extreme` : int ou string ??

Retourne l'agrégation mini de l'extrême. Pourra être implémenté dans le future en scannant les données observations

iv. `get_agregations(timestamptz)`: List json

Retourne un tableau de toutes les agrégations de ce poste pour la date/heure correspondant. Créé les agrégations manquantes

v. `get_observation(timestamptz)`: json, bool

Retourne un « json observation » avec l'observation demandée. Crée un record si n'existe pas. Retourne le Json, et un booléen pour dire si une observation existait ou pas à la date.

b. Objet `TypeInstrument`

Cette classe d'objet implémente le comportement du type objet. C'est lui qui va savoir insérer ou supprimer une observation reçue dans la table observation, tout en calculant les agrégations et les extrêmes.

i. Méthodes publiques

1. `process_observation`

Paramètres :

- Données élémentaires lue (fichier JSON traité)
- « json observation »
- Poste objet
- `Add_flag` : true pour ajouter, false pour soustraire

Met à jour le « json observation » à partir des données élémentaires. Les fonctions de filtrage en entrée et trigger comme décrit dans le document « `filtres et controles.doc` » seront implémentés ici.

2. `process_aggregates`

Paramètres :

- Données élémentaires lue (fichier JSON traité)
- Tableau des agrégations
 - [0] -> Heure
 - [1] -> Jour
 - [2] -> Mois
 - [3] -> Année
 - [4] -> Global
- Poste objet
- `Add_flag` : true pour ajouter, false pour soustraire

BD Climato, Le voyage des données!

Met à jour les agrégations et extrêmes à partir des données élémentaires. Aussi agrégations des données de qualité (qa_modifications, qa_incidents, qa_check_done)

ii. Méthodes privées

1. To be written...

10. Algorithme importation

a. Fonction ImportMeasures

- Lire fichier JSON (trié par poste) dans variable <measure>
- Begin Transaction
- Si postId différent -> Call poste.getInfo(posteId)
- json_obs, add_flag = poste.get_observation(<measure.dat>)
- json_aggs = poste.get_aggregations(<measure.dat>)
- if flag => call deleteMeasure(measure, json_obs, json_aggs)
- Pour chaque type_instrument
 - type_i.process_observation(measure, json_obs, poste, true)
 - type_i.process_aggregations(measure, json_aggs, poste, true)
- Update observation with json_obs
- Update all aggregations with json_aggs
- Commit transaction, and delete JSON file
- Si erreur -> Rollback Transaction + notify + passe au JSON suivant

b. Fonction DeleteMeasure

- Pour chaque type_instrument
 - type_i.process_observation(measure, json_obs, poste, false)
 - type_i.process_aggregations(measure, json_aggs, poste, false)
- return

BD Climato, Le voyage des données!

ANNEXE A : JSON ou Colonnes

<https://docs.djangoproject.com/fr/3.1/ref/contrib/postgres/fields/>

ANNEXE B : Date implementation

**** a discuter plus tard ****

Table n2(id serial primary key, j jsonb, d timestampz)
datetime as :

- String in json
- Linux time in json
- Timestampz field

select * from n2;

```
id | j | d
---+---+---
1 | {"d": "2021-02-03 07:43:17.033231", "dl": 1612338197.033231} | 2021-02-03 07:43:17.033231+04
```

Get month from db :

select extract(month from timestampz(j->>'d')) as m_jsonts, extract(month from
to_timestamp((j->>'dl')::float)) as m_linuxtime, extract(month from d) as m_tz from n2;

```
m_jsonts | m_linuxtime | m_tz
-----+-----+-----
2 | 2 | 2
(1 row)
```

Get date in python :

...

Comparison :

Data type	size	get date type	+	-
Json : linux	4	to_timestamp((j->>'dl')::float)	Fast to compare Small	Convert to python
Json : string	20	timestampz(j->>'d'))	Easy to display	Convert to python Slow to compare Big, but compressed
timestampz	8	d	Easy to display convert to python	Not usable in json

BD Climato, Le voyage des données!

ANNEXE C : Intégration des partenaires

Si un partenaire nous fournit des données sous une forme différente (csv...), un module de transcodage des données vers notre format JSON sera développé. Une vérification de cohérence des données sera faite dans ce module.

ANNEXE D : Modification des données (UI)

Très probablement un programme de correction de données devra être disponible. Il faudra spécifier ce programme, mais l'idée générale est qu'il permette de :

- Supprimer une donnée ou une plage de données
- Corriger une donnée ou une plage de données en
 - Invalidant une mesure (ou un type, à réfléchir...)
 - ... autre cas ... voir les scenarios ...
- Modifier une observation (données élémentaires)

Ce programme générera un ou plusieurs fichiers json, qui seront traités par le module d'importation.

Aussi un programme de régénération des agrégations (sans perdre les valeurs extrêmes) devra être développé.

ANNEXE E : Questions non réglées

1. **Comment gere t'on les rain_sum1h/...24h ?**
⇒ **Dans les jsons horaire...**
2. Je ne suis pas convaincu que le rain_rate_mean a un intérêt à partir du moment où on a les rain_sum on peut recalculer. Pour étudier un épisode pluvieux on utilise principalement les cumuls glissants sur 1h, 3h, etc... pas trop les intensités horaires. Par contre les rain_rate_max sont vraiment intéressants car ils sont bien plus fins que ceux de MF qui sont lissés sur 6 minutes.
3. 3-pour le delta/trend qu'on trouve dans \$current, Micka tu confirmes que c'est sur les 3 dernières heures ?
4. 4-5 ok pour rajouter les min/max.
5. Micka pour les cumuls glissants je n'ai compris si tu préfères qu'on les calcule dans les tables où on prend ceux que tu as rajoutés dans \$current ? Et Nico tu penses que le temps de lecture/calcul risque d'être pénalisant ?
6. Je place ici des remarques générales sur les agrégations. Communément et notamment à MF on ne calcule pas de moyennes horaires pour les paramètres Tempe, Humidité, vitesse du vent, pression. On garde la valeur instantanée à l'heure ronde. Les moyennes quotidiennes sont calculées en prenant les 24 données horaires instantanées. Au pas de temps mensuel on fait la moyenne des moyennes journalières. D'autre part on fait la moyenne des TempeMax, TempeMin, HumiditéMax, HumiditéMin à partir des max/min quotidiens. A discuter donc de ce qu'on fait nous. On peut imaginer bien sûr de faire la moyenne des moyennes horaires et d'ailleurs le résultat ne devrait pas forcément être très différent.