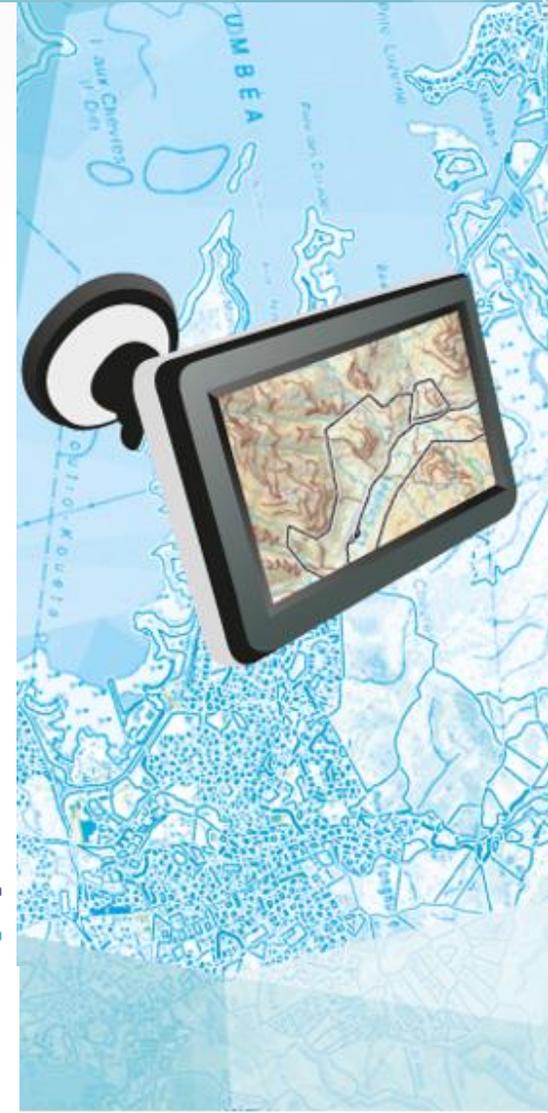


Entrepôt de donnée

Datawarehouse de la province Sud
(nom de code: DRHOUSE)



GISday
2022

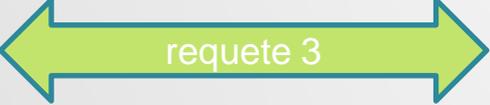


Sommaire

- Usage de la donnée jusqu'à présent
- Réflexions sur le processus décisionnel
- Architecture choisie
- Exemple décisionnel sur le trafic SIG du fond de carte

Usage de la donnée actuellement

Base de donnée



Application

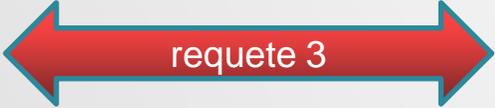


Décisionnel

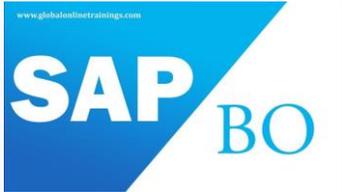


Usage de la donnée actuellement

Base de donnée



Applications



Rapport BI



Usage de la donnée actuellement

Avantages

- Accès en temps réel

Inconvénients

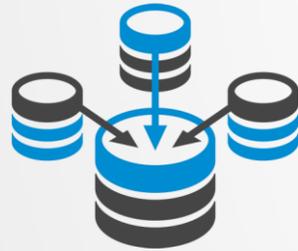
- Lenteur
- Stabilité (changement de modèle)
- Ralentir les applications métier
- Temporalité gérée uniquement au niveau du métier
- La gestion des droits
- Pas de description de la donnée

Réflexions sur le processus décisionnel



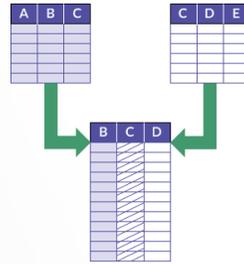
PRODUCTION

- Application métier
- Qgis
- Saisie de donnée



CENTRALISATION

- Métadonnée obligatoire
- Sécurité d'accès



DERIVATION

- Enrichir
- Anonymiser
- Métadonnée obligatoire
- Sécurité d'accès



SPECIALISATION

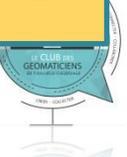
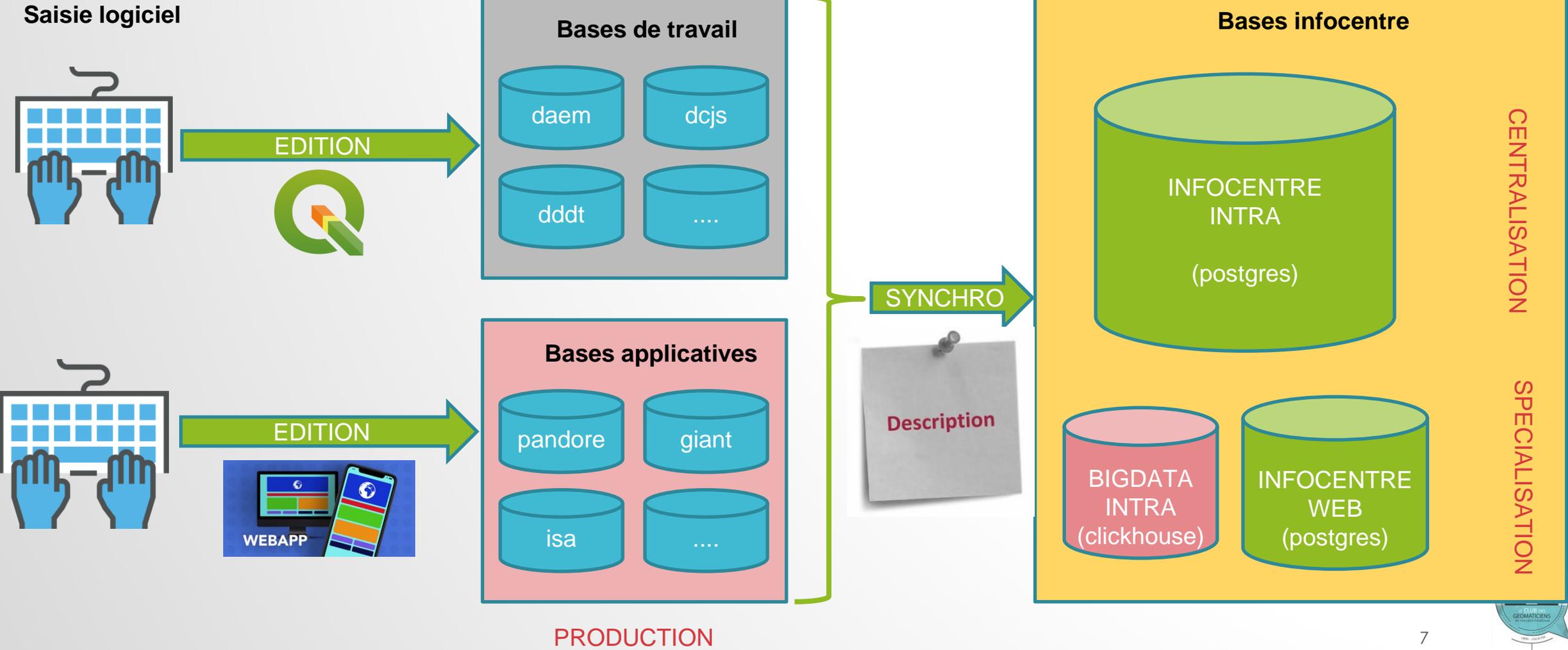
- Diffusion (WEB)
- Performance



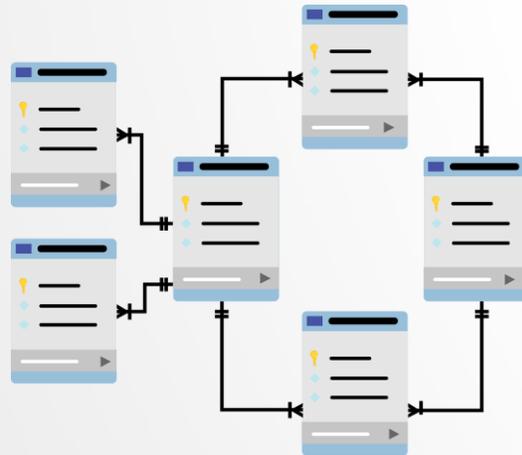
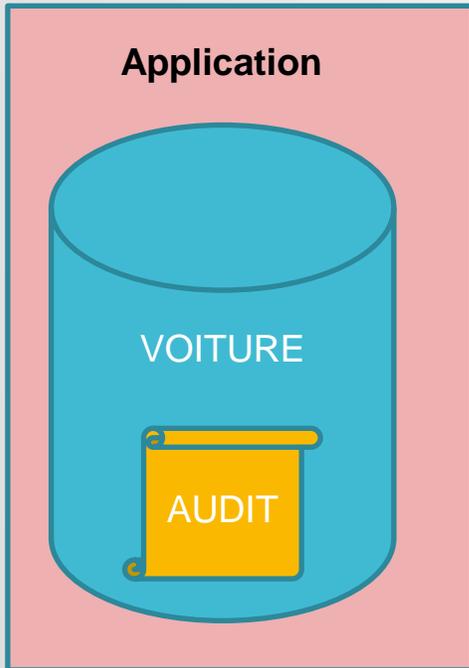
VISUALISATION

- BO
- Metabase
- Qgis
- Excel

De l'édition à l'infocentre



Temporalité: journal des modifications



- Ajoute
- Modifie
- Supprime

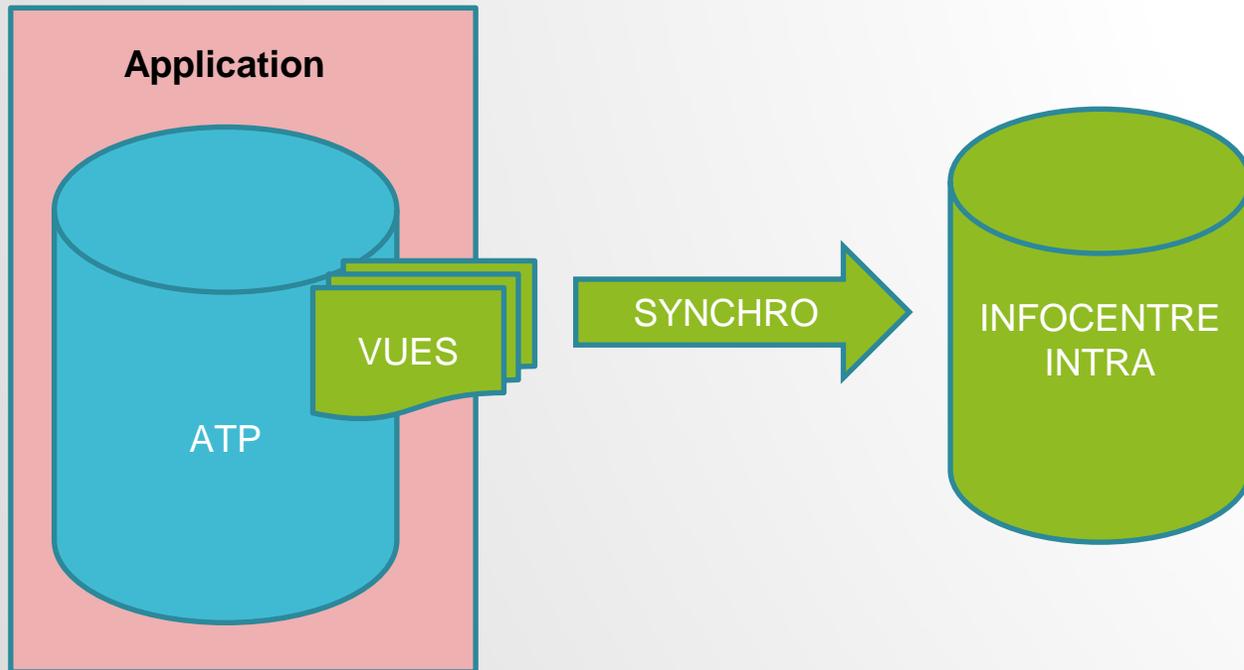
Table d'audit (journal)

Date Time	Action	Table	Donnée ligne
2022-11-16 10:00:00	CREATE	vehicule	{json}
2022-11-16 10:15:00	UPDATE	marque	{json}
2022-11-16 10:30:00	UPDATE	marque	{json}
2022-11-16 11:00:00	DELETE	marque	{json}

Le chef de projet peut configurer

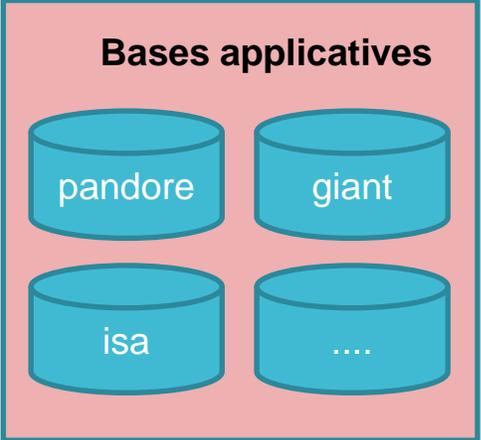
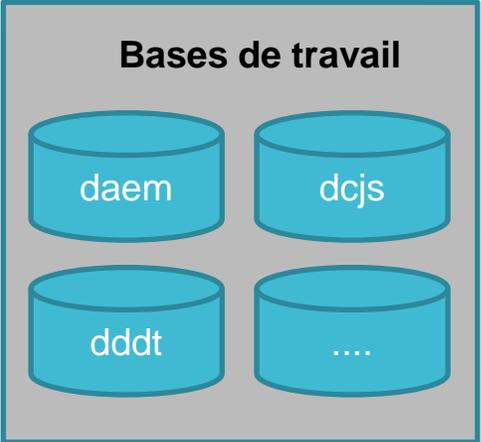
- les tables à auditer
- les propriétés à exclure (RGPD)

Sélection des données à synchroniser

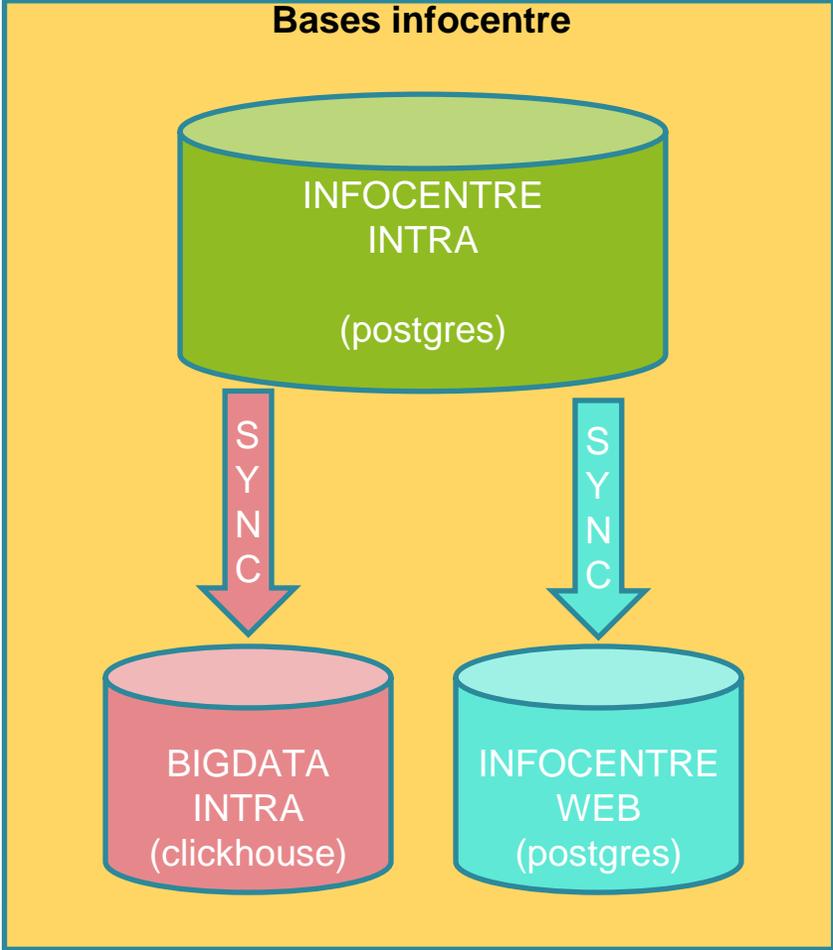


- Vue (table virtuelle par requête)
 - Contrat applicatif avec l'infocentre
 - Modèle à plat (pas de relationnel)
 - Les développeurs sont alertés si la vue est cassée
 - Cache la complexité de l'application
 - 1 vue = 1 métadonnée
- Drhouse contrôle les modèles avant synchro

Les types de synchronisations



- ISO identique
- Incrémentale (date)
- Modèle en étoile



phase spécialisation (Les différentes bases de données)

Pouvoir utiliser une base de donnée dans ce qu'elle sait faire de mieux

Besoin	Base de données
analyse statistique (fonctionnalités)	postgres (extensions diverses)
analyse spatiale (fonctionnalités)	postgres / postgis
analyse textuelle (recherche)	solr / elastic search
analyse de graphes	neo4j
analyse de logs (big data) > millions / milliards	clickhouse

Benchmark pour le bigdata

Requête spatiale en agrégeant des points (30 millions de lignes)

pc portable
i7-8750H CPU @ 2.20GHz (6 cores)
memoire de 16 Go
SSD

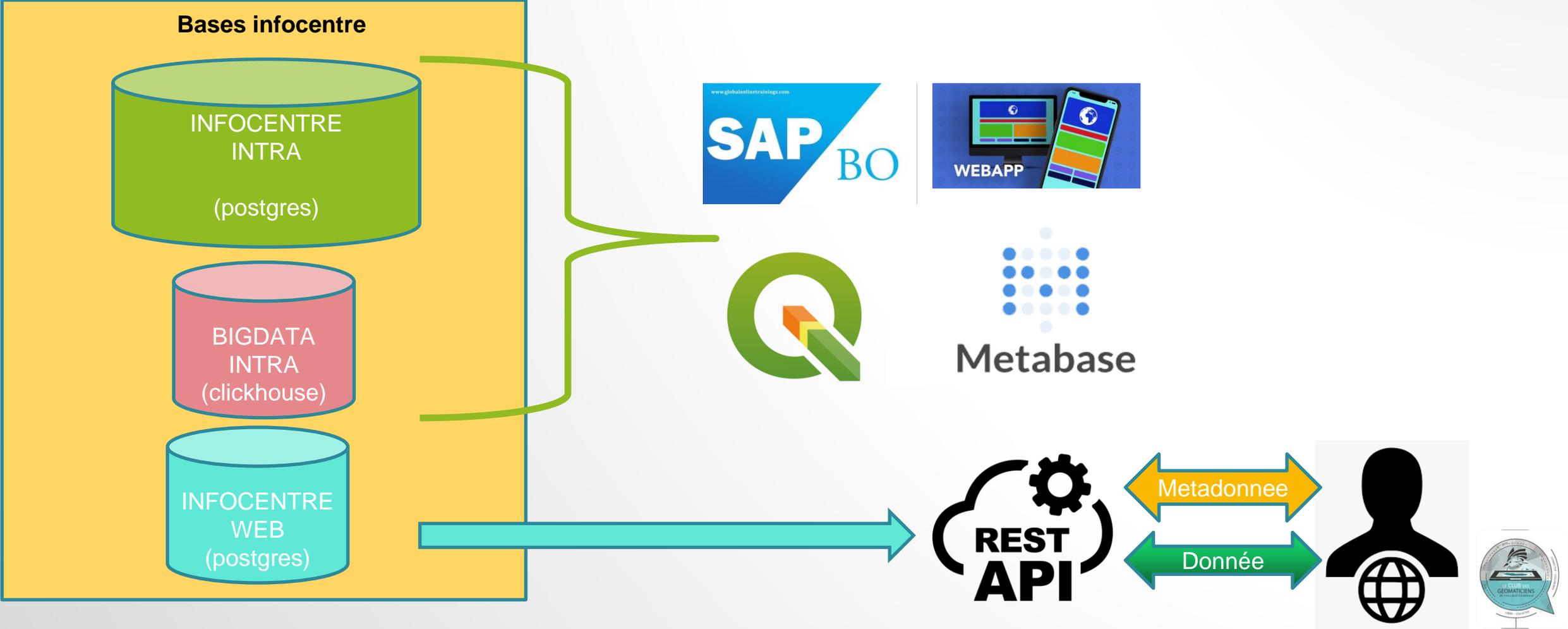
Base de donnée	Temps	Résultat exact
Postgresql (v13)	45 secondes	oui
ElasticSearch (v7)	11 secondes	non, résultat approximatif
Clickhouse (v18)	300 millisecondes 60 millisecondes (optimisation)	oui

Pour le BIGDATA, **clickhouse** est largement vainqueur

750 fois plus rapide que **postgres**

180 fois plus rapide que **elasticSearch** avec un résultat juste

Usage des données (BI / partage)



A la recherche d'une Solution BI Tableaux de bord

- Stage de 2 mois d'un étudiant en BI master 1 er année

Les critères

- Facilité d'utilisation / Ergonomie
- La richesse des composants (carte)
- La gestion de la sécurité et la possibilité de scripter
- La performance
- Le prix



A la recherche d'une Solution BI Tableaux de bord

Les produits du benchmark

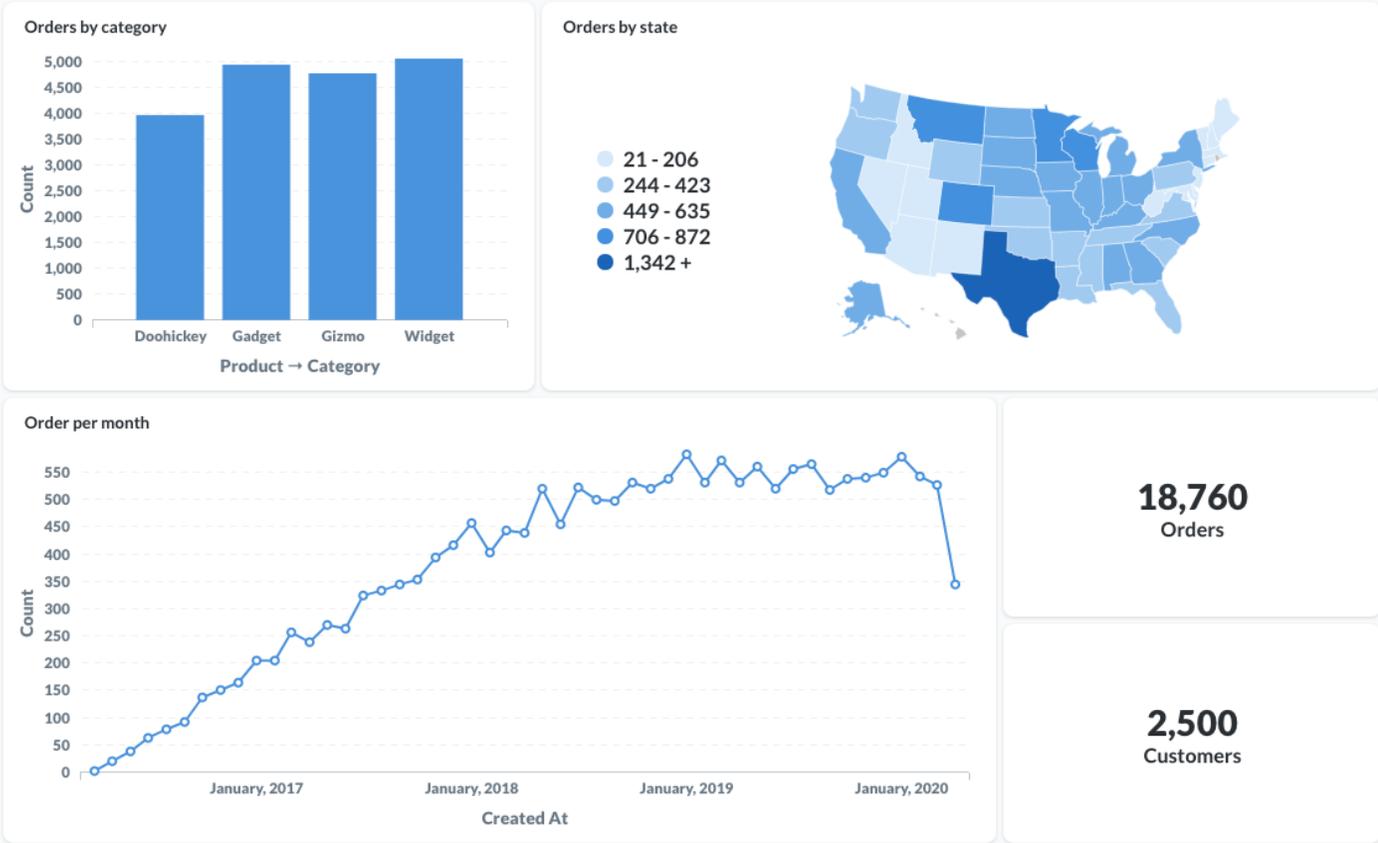
- Metabase
- PowerBI
- Tableau
- DataStudio
- Apache superset



- Langage ensembliste
 - Grammaire et vocabulaire simple (une 20 aine de mots à connaitre)
 - Interopérable (la plupart des logiciels supporte le SQL)
 - Extension SQL Postgis pour le spatial (base postgres)
-
- formations niveau débutant à l'IFAP cette année
 - formations débutant/avancé l'année prochaine

Demo metabase

Interactive dashboard



Navigation dans cartosud (geoportail de la province sud)

- Un agent souhaite utiliser notre portail pour
 - Voir précisément une zone (idéalement le plus récent)
 - Remonter le temps pour constater l'évolution



20 km

Echelle: 1 000 000

Coordonnées: Pas d'affichage



Dates (Fond de carte): 1976 à 2019

Sources: Province Sud / Nasa / province sud / Gouvernement de la Nouvelle-

Navigation dans cartosud (geoportail de la province sud)

- Fond de carte par défaut
 - Jusqu'au 1/2000 (Image Satellite de **0,5 m** de résolution de **mars 2021**)
 - En dessous du 1/2000 (Orthophoto de **0,2 m** de résolution de **septembre 2014**)
- Fond de carte Historique (demandé par l'utilisateur)
 - En dessous du 1/2000
 - Avant 2014
 - Ortho-photo de **0,25 m** de **septembre 2008**
 - Avant 2008
 - Ortho-photo de **0,5 m** de **septembre 2005**

- Quelles sont les zones à réactualiser en haute résolution par priorité?
(sachant que la DAEM n'a pas le budget pour tout acheter)
 - Projets avenir
 - **Trafic du fond de carte de l'année écoulée**

Intégration DRHOUSE

PHASE centralisation

DRHOUSE - Synchronisation des données

Administration

- DataSources
 - log
 - jdbc
 - Click house data src
 - Oracle data src
 - Postgres data src
- Référentiels
- Synchronisation
- Drhouse
 - Export des vues
 - Historique des modèles

Postgres data src (4)

Ajouter Ouvrir Rafraichir Supprimer Outils

Libelle	Commentaire	Mode	Annuaire	Web datasource
Ancienne base externe		Application	GUD	
externe_travail	Base de donnée contenant les ...	Bdd de Travail	GUD	
geodata	accès à la métadonnée du cat...	Application	GUD	
Infocentre (intra)	Infocentre principal à usage int...	InfoCentre	GUD	

Intégration DRHOUSE

Postgres data src (4) × geodata (jdbc:postgresql://pgsql1-prod.province-sud.prod:5432/geodata) ×

Enregistrer Rafrachir Supprimer Connexion réussie Synchroniser les droits Tout synchroniser Options admin

Info Connections error

Libellé: geodata Mode: Application Description: accès à la métadonnée du catalogue provincial et aux logs de geodata

Database: geodata Annuaire: GUD

Host: [REDACTED] Priorité de synchro en tant que source: 100

Port: [REDACTED] Ecriture si geocat:

Lecture seule:

Habilitation: [REDACTED]

Ajouter un schéma Rafrachir Synchroniser schém

Schéma	e...	a...
topology	OUI	NON
_drhouse_travail	OUI	OUI
tiger	OUI	NON
tiger_data	OUI	NON
geodata	OUI	OUI
repack	OUI	NON
_drhouse_audit	OUI	OUI
public	OUI	OUI

Ouvrir Rafrachir Ajouter une vue Supprimer l'objet Sync From Sync To

Table	_	e.	a.	cfe	Sync from	liste Sync to
catalogue_query	V...	O...		N...		
catalogue_query_m	V...	O...		N...		
databasechangelog	T...	O...		N...		
databasechangeloglock	T...	O...		N...		
drhouse_echelle_psud	V...	O...		N...		Infocentre (intra)/geo.ec...
drhouse_export_queries	V...	O...		N...		
drhouse_geocat_mot_cle	V...	O...		N...		Infocentre (intra)/psud_si...
drhouse geodata queries	V...	O...		N...		Infocentre (intra)/psud si...



Intégration DRHOUSE

DRHOUSE - Synchronisation des données

Administration

- DataSources
 - log
 - jdbc
 - Click house data src
 - Oracle data src
 - Postgres data src
- Référentiels
- Synchronisation
- Drhouse
 - Export des vues
 - Historique des modèles

Postgres data src (4) × geodata (jdbc:postgresql://pgsql1-prod.province-sud.prod:5432/geodata) × drhouse_geodata_queries ×

Enregistrer Rafrachir Synchroniser les droits Outils

Datasource

Datasource: geodata (jdbc:postgresql://[redacted])
Schema: geodata (geodata)
Objet: drhouse_geodata_queries
Type: Vue
Taille: [input]

Description

Metadonnee: Log des requêtes SIG de la p [Ouvrir]
Description: [input]
lock_description:

Paramètres de synchronisation en tant que cible

Mode synchro:
Période de synchro:
Mode d'erreur:
Date de dernière synchro:
Temps de dernière synchro:

Modèle (28) Indexes (0) Modèles liées Cibles à synchroniser (1) Source (provenance) Dataset (0) Debug

Synchroniser Archiver la cible Naviguer vers la cible Ajouter Ouvrir Rafrachir Supprimer Update Style Qgis Outils

Datasource sync to	Schema cible	Sync mode	Date last sync	Date bdd las... ↑	Nb...	Erreurs last sync	A	E.
Infocentre (intra) (jdbc:post...	psud_si_log	Ajout des changements (modèle ét...	10/11/2022 17:16:58	10/11/2022 17:16:58	0		<input checked="" type="checkbox"/>	N...



Métadonnée : Création d'une fiche

Log des requêtes SIG de la province sud

Enregistrer Rafranchir Supprimer Analyser Consulter l'analyse Dépublier Transformer en modèle Transformer en service vecteur Exporter

Général Diffusion Classification Situation Géographique Documents Liens URL Vecteur Historique Verrou

Table Url manquante

Titre: Log des requêtes SIG de la province sud

Référence Métier: 44e7f4b1-fdb7-4557-8b4f-a2ae8592745e

Description:
L'application provincial [geodata](#) fournit les services cartographiques.
[geodata](#) logue chaque requête en tenant du [context](#)
- l'application qui fait la demande des requêtes
- l'organisation de l'utilisateur
- la référence vers la métadonnée du service
- la résolution
Cette donnée est volumineuse et contient l'historique des requêtes avec une description détaillée

Fiabilité: Vérifiée

Donnée à caractère personnelle (RGPD) ?

Application provinciale: GEODATA

Contacts (1) Sources (1)

Ajouter Ouvrir Supprimer Outils

Nom	Fonction	Rôle	Organisme	Email	Téléphone
roman.de-scorraille	Responsable du bureau SIG	Responsable	DSIN	roman.de-scorraille@province-sud.nc	

Miniature



Métadonnée : Confidentialité

Log des requêtes SIG de la province sud

Enregistrer | Rafraîchir | Supprimer | Analyser | Consulter l'analyse | Dépublier | Transformer en modèle | Transformer en service vecteur | Exporter

Général | **Diffusion** | Classification | Situation Géographique | Documents | Liens URL | Vecteur | Historique | Verrou

Donnée

Règle de diffusion et sécurité

Portée de diffusion de la donnée: Réseau Extranet

Donnée téléchargeable:

Sécurité: Donnée sécurisée

Métadonnée

Groupes de diffusion (1)

Ajouter:

Groupe	Scope	Suppression
Agents psud		

Restictions

Récapitulatif des droits

Diffusion interne (psud.nc) (1)

Groupe	Nb user	Voir liste u...	Voir les do...
Agent (Affectation automatique)	1565		

Diffusion externe (province-sud.nc) (0)

Groupe	Nb user	Voir liste u...	Voir les do...
--------	---------	-----------------	----------------



Métadonnée : Classification

Log des requêtes SIG de la province sud

Enregistrer Rafrachir Supprimer Analyser Consulter l'analyse Dépublier Transformer en modèle Transformer en service vecteur Exporter

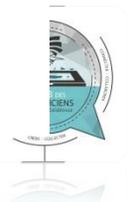
Général Diffusion **Classification** Situation Géographique Documents Liens URL Vecteur Historique Verrou

Mots clés (5) Thèmes (2)

Attacher mot clé Ajouter Ouvrir Attacher theme Ouvrir Rafrachir

Mot clé	Techni...
Logs	OUI
province sud	OUI
Sig	OUI
Statistiques	OUI
tuile	NON

Nom du thème	Description
Autre	other
Localisation	location



Intégration: Ajout d'une synchronisation

geodata / schéma geodata / modèle drhouse_geodata_queries -> Infocentre (intra) / schéma psud_si_log

Rafraichir Outils

Configuration Historique et erreurs Propriétés à exclure (0)

Sync from: drhouse_geodata_queries

Datasource cible

Datasource sync to: Infocentre (intra) (jdbc:postgresql://vm-059963.prod.province-sud.n

Schema cible: psud_si_log

Table cible: geodata_queries

Historique de dernière synchronisation

Date bdd last success sync: 14/11/2022 09:41:54

Date last sync: 14/11/2022 09:41:54

Nb d'entités synchronisées: 1535

Erreurs:

Durée: 8

Error:

Synchro en cours

Nb entite pending sync:

Date pending sync:

Algorithme de synchronisation

Algo: Ajout des changements (modèle étoile)

Période de synchro: Tous les 5 minutes

Mode erreur: insertion des lignes valides uniquement

Rafraichir vue mat aval:

Description: Ajoute uniquement les changements dans la cible
La cible contient l'historique de la source.
Il faut être vigilant à la période de synchro.
L'éclatement du modèle en étoile permet d'optimiser la volumétrie des données

Pré-requis:

- Avoir un id
- Avoir un champ date pour connaitre les données postérieures à la cible dans la source
- Avoir des attributs déclarés en dimensions pour le modèle en étoile
- Avoir des attributs déclarés en mesures pour le modèle en étoile

Enregistrer Supprimer Fermer

Intégration : Phase centralisation (postgresql)

Postgres data src (4) × geodata (jdbc:postgresql://pgsql1-prod.province-sud.prod:5432/geodata) × drhouse_geodata_queries × olap_geodata_queries ×

Enregistrer Rafrachir Synchroniser les droits Outils ▾

Datasource
Datasource: Infocentre (intra) (jdbc: [redacted])
Schema: psud_si_log (Infocentre (intra))
Objet: olap_geodata_queries
Type: Table
Taille: 38 GB

Description
Metadonnee: Log des requêtes SIG de la p [Ouvrir](#)
Description:
lock description:

Paramètres de synchronisation en tant que cible
Mode synchro: Ajout des changements (modè
Période de synchro: Tous les 5 minutes
Mode d'erreur: insertion des lignes valides uni
Date de dernière synchro: 2022-11-10T17:26:58+1100
Temps de dernière synchro: 1

Modèle (28) Indexes (1) Modèles liés Cibles à synchroniser (0) Source (provenance) Dataset (14) Debug

Ouvrir Rafrachir

Schema	Nom	Type	DDL_CONTENT	Metadonnee	Erreur	Relation
psud_si_log (Infocentre (intra))	olap_geodata_...	Vue	select ogeq.id , geocat_reference , data_title, ap...			DEPENDANCE



Intégration Phase spécialisation (clickhouse)

Enregistrer Rafrachir Synchroniser les droits Outils

Datasource

Datasource: infocentre (bigdata) (jdbc:clickhou:
 Schema: psud_si_log (infocentre (bigdata))
 Objet: olap_geodata_export_queries_eni
 Type: Table
 Taille: 3.51 GiB

Description

Metadonnee: Log enrichi des requêtes SIG **Ouvrir**
 Description: Geodata logue chaque requête pour en faire des statistique sur l usage des données dans le SIG. Chaque requête est enrichi par rapport à l infocentre pour l associer par exemple à une
 lock_description:

Paramètres de synchronisation en tant que cible

Mode synchro: Ajout des changements (même modèle)
 Période de synchro: Tous les 5 minutes
 Mode d'erreur: insertion des lignes valides uni
 Date de dernière synchro: 2022-11-10T17:27:38+1100
 Temps de dernière synchro: 9

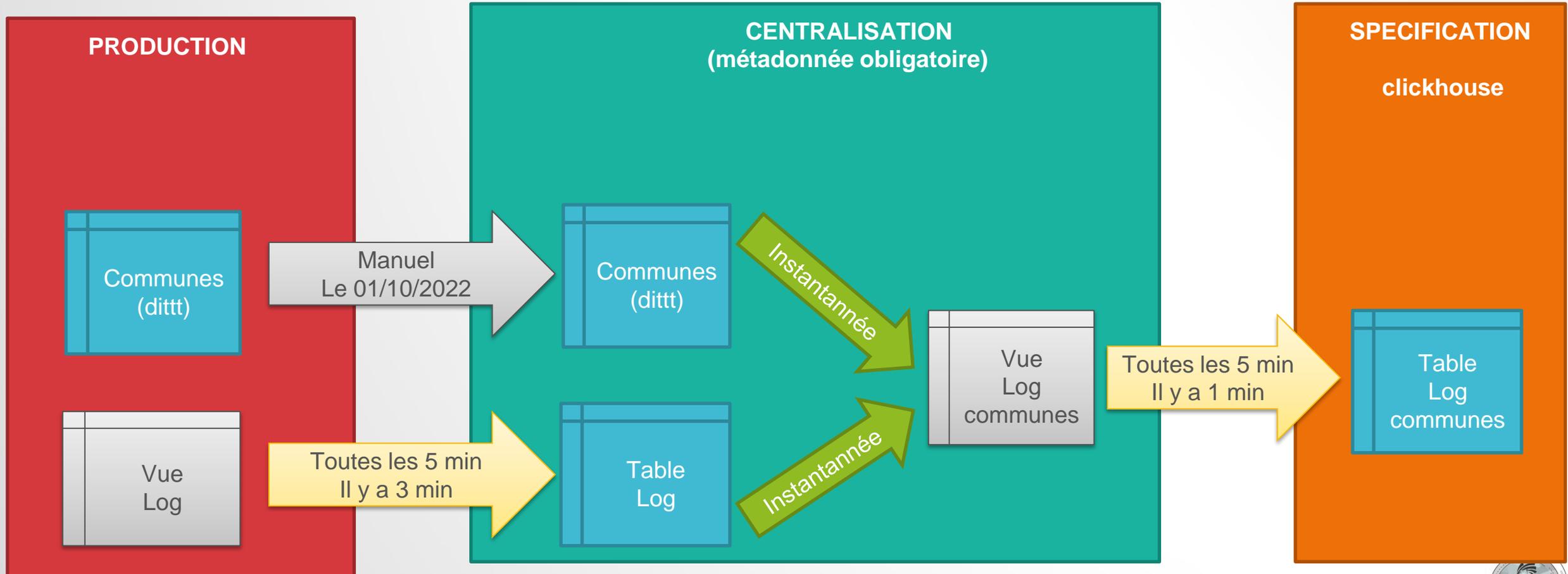
Modèle (27) Indexes (0) Modèles liés Cibles à synchroniser (0) Source (provenance) Dataset (0) Debug

Ouvrir Rafrachir Enregistrer coches Faire remonter la description du modèle Outils

O	Propriété	Type	Description	isId	isDi...	isMe...	isRo...	isTi...	f.	m	i.	e.
1	id	Nullable(String)		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	N...		<input type="checkbox"/>	O...
2	geocat_reference	Nullable(Int64)		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	N...		<input type="checkbox"/>	O...
3	data_title	Nullable(Int64)		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	N...		<input type="checkbox"/>	O...
4	appli_referer	Nullable(Int64)		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	N...		<input type="checkbox"/>	O...
5	geodata_code_src	Nullable(Int64)		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	N...		<input type="checkbox"/>	O...
6	direction	Nullable(Int64)		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	N...		<input type="checkbox"/>	O...
7	server_node	Nullable(Int64)		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	N...		<input type="checkbox"/>	O...

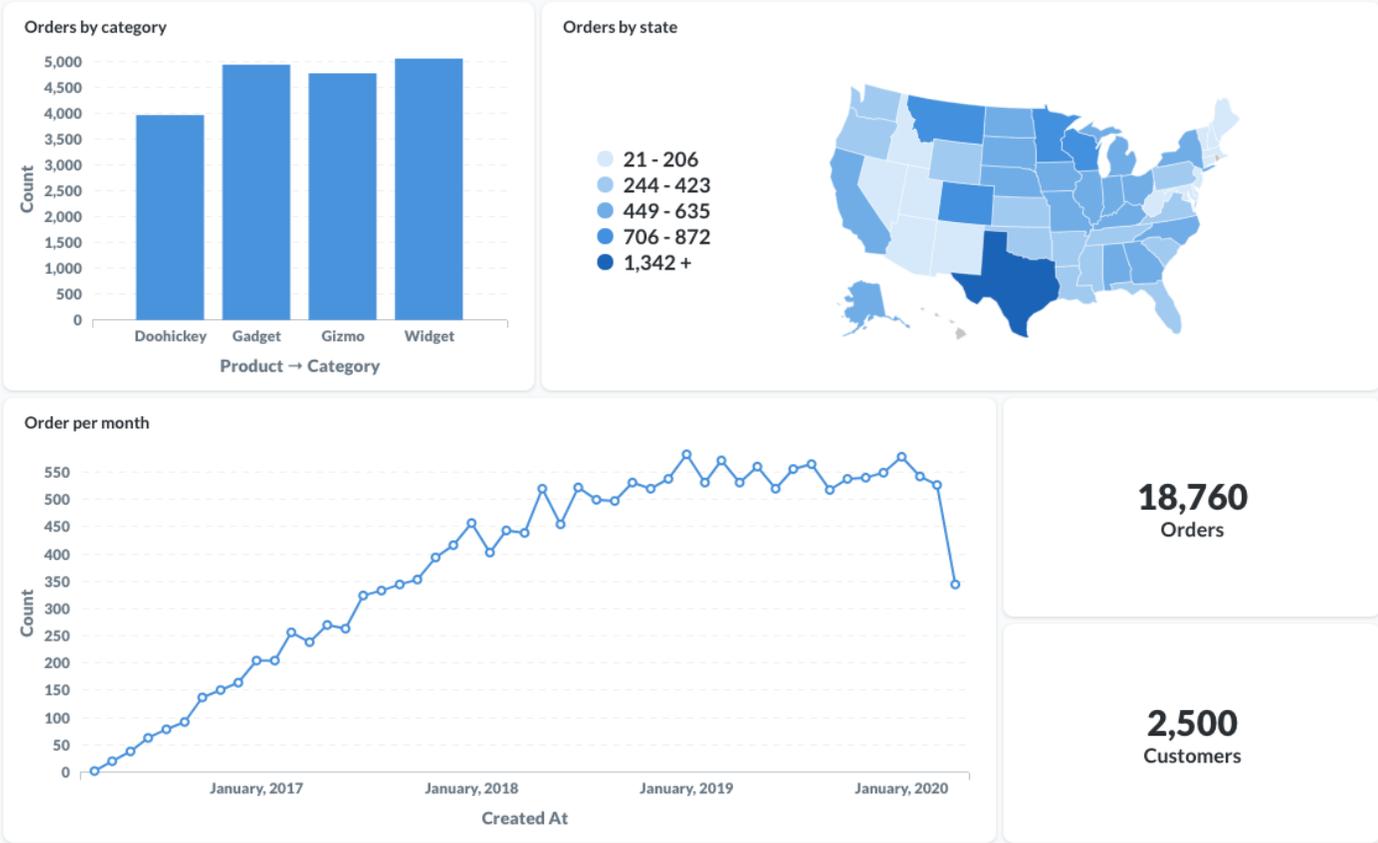


Graphe généalogie



Demo metabase

Interactive dashboard



METABASE Trafic de carte

The screenshot displays the Metabase web interface. At the top, there is a search bar with the text "Search..." and buttons for "+ New" and a settings gear. On the left, a sidebar contains navigation options: "Home", "BOOKMARKS" (with a dropdown arrow), "Recensement et Consommation de...", "COLLECTIONS" (with a dropdown arrow), "Our analytics", "Your personal collection", "si", and "DATA" (with a dropdown arrow). Below "DATA" is a "Browse data" link. The main content area features a greeting "Hey there, Roman" with a smiley face icon. Below this, a section titled "Pick up where you left off" contains several interactive cards: "Mosaïque dynamique Trafic au 0.01°", "statistique utilisation fond de carte psud", "Top 10 des mots clés utilisés", "requetes par application", "Concommodation service carto", and "Metabase tips". At the bottom of the interface is a decorative illustration of a lighthouse, a bridge, a sailboat, and a house.