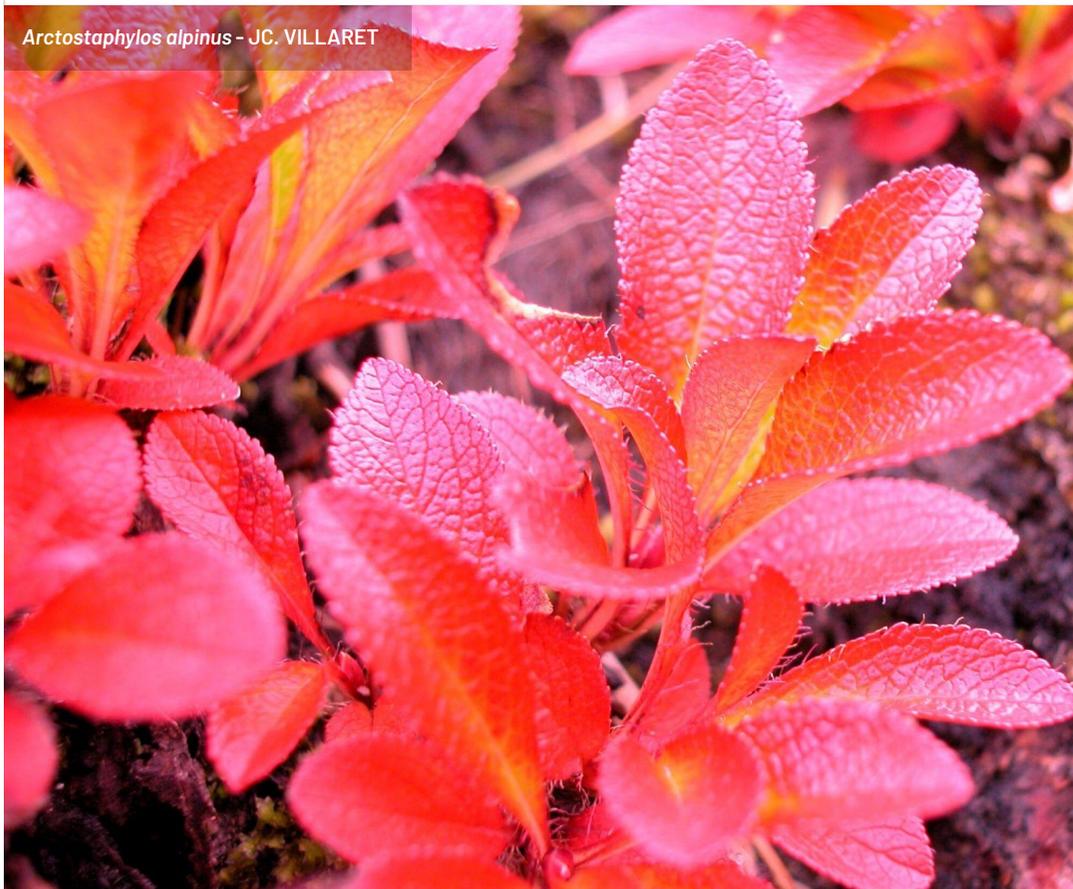


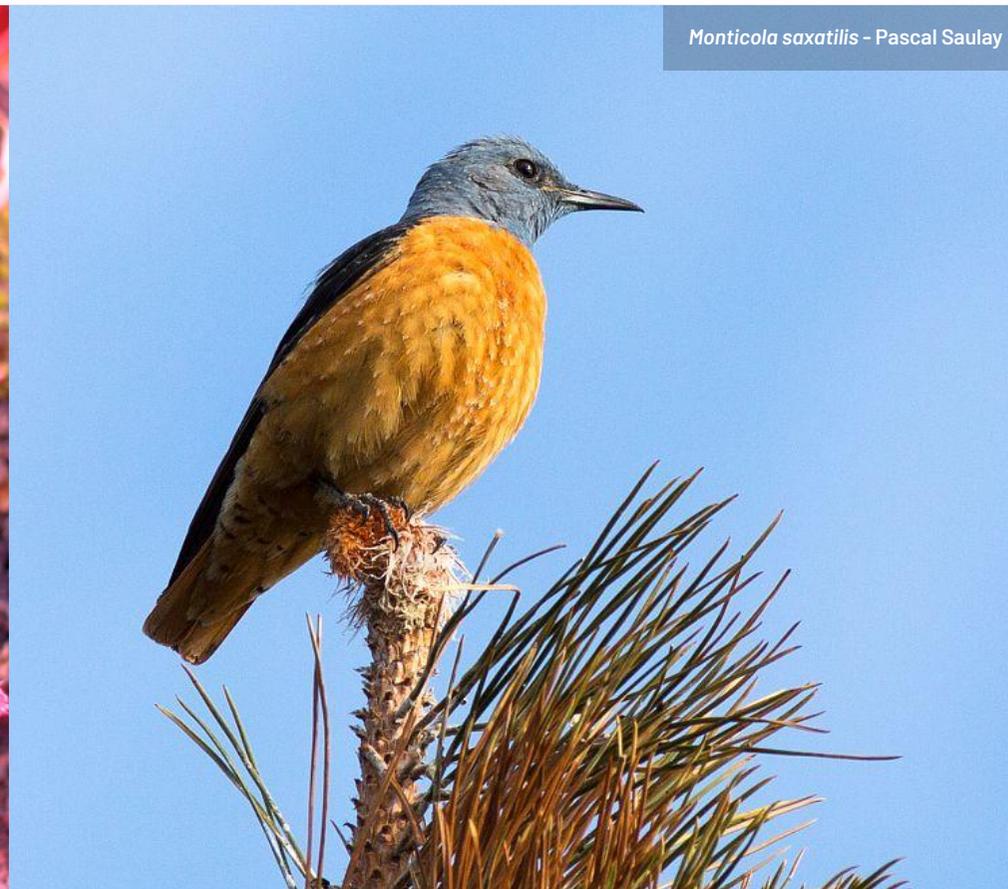


La Région
Auvergne-Rhône-Alpes

Arctostaphylos alpinus - JC. VILLARET



Monticola saxatilis - Pascal Saulay



Atelier

Performances dans GeoNature

Rencontres nationales GeoNature
26 & 27 juin 2024



Parc national
des Ecrins

Sommaire

1



Améliorer les performances de GeoNature

2



Un cas d'étude, le module de Synthèse

3



Analyse des performances de GeoNature

Améliorer les performances de GeoNature

Source de ralentissements

Solutions matérielles

Exemple d'architecture serveur

Solutions logicielles

Bonnes pratiques de développement



Grand Géocoucou, ou Road Runner

Source de ralentissements

- Limites matériels
 - Les performances de PostgreSQL dépendent fortement de la vitesse de lecture sur le stockage [1, 2]
- Concurrence de plusieurs services web sur la machine hôte
- Configuration du serveur HTTP et du SGBD [3]
- Requêtes SQL non optimisées (complexe, mauvais index...)
- PostgreSQL parallélise peu l'exécution d'une requête [4]

[1] <https://www.2ndquadrant.com/en/blog/tables-and-indexes-vs-hdd-and-ssd/>

[2] <https://public.dalibo.com/exports/formation/manuels/formations/perf1/perf1.handout.html>

[3] <https://wiki-sinp.cbn-alpin.fr/serveurs/installation/web-srv/nginx>

[4] <https://makina-corpus.com/devops/paralleler-des-requetes-avec-postgresql>

Solutions matérielles

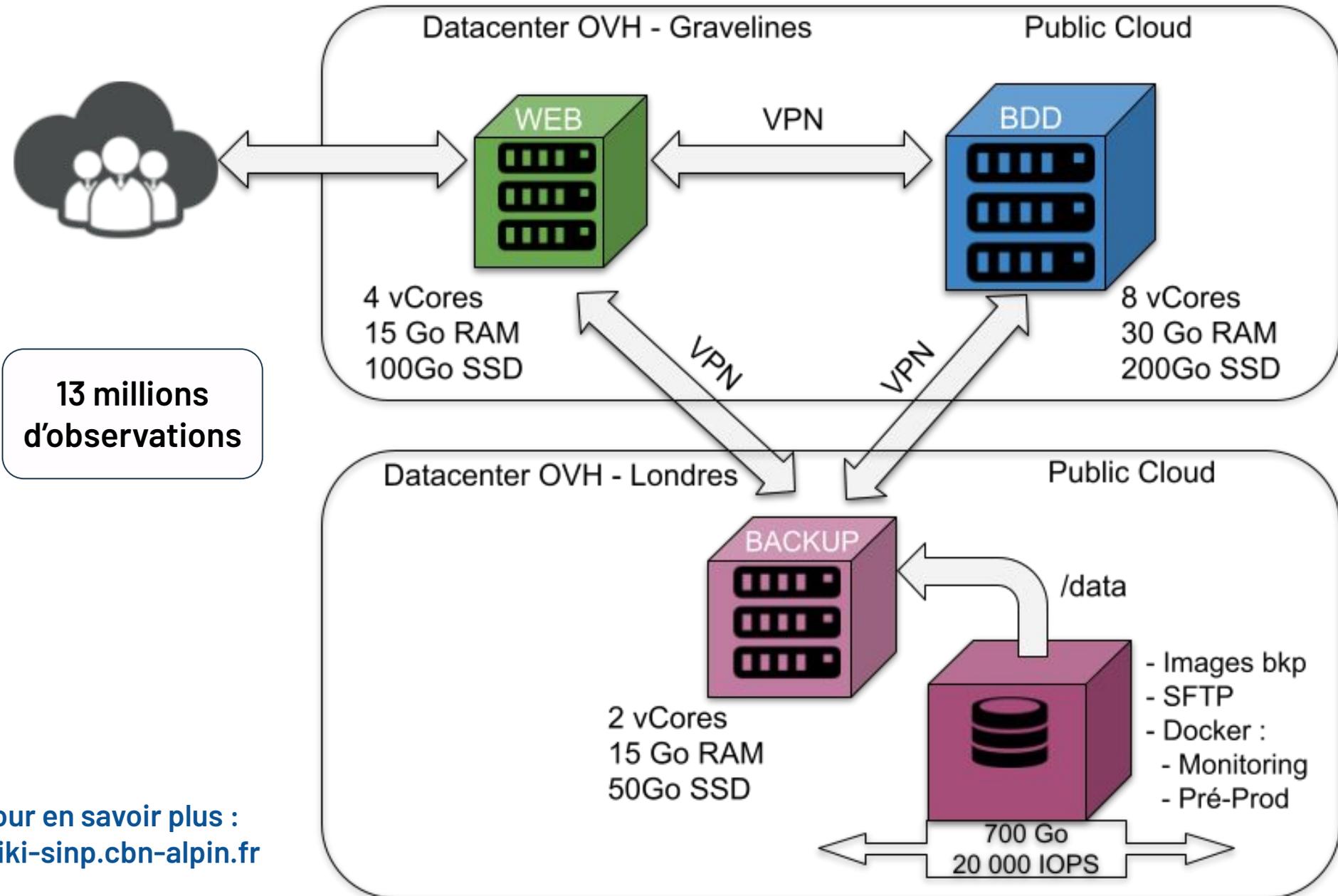
Hardware

Solution hardware => rapide mais coût financier important

- Opter pour des stockages rapides :
 - SSD SATA
 - ou mieux SSD NVME
- Isoler la base de données Postgresql sur une machine dédiée
- Augmenter la RAM de la machine hébergeant Postgresql

Architecture SINP PACA

Hardware



Pour en savoir plus :
wiki-sinp.cbn-alpin.fr

Solutions logicielles

Software

Optimiser la configuration de votre serveur

- Optimiser la configuration de PostgreSQL
(PgTune : <https://pgtune.leopard.in.ua>)
- Remplacer Apache par Nginx :
 - Nginx + performant, si beaucoup de requêtes
 - Apache si multi administrateurs (htaccess)
 - Ressources [1, 2]
 - Configuration pour GeoNature [3, 4]

The screenshot shows the PGTune web interface. On the left, under 'Parameters of your system', there are input fields for: DB version (16), OS Type (Linux), DB Type (Web application), Total Memory (RAM) (4 GB), Number of CPUs (8), Number of Connections (25), and Data Storage (SSD storage). A 'Generate' button is at the bottom of this section. On the right, there is a 'postgresql.conf' tab and an 'ALTER SYSTEM' button. Below this, a text area displays the generated configuration parameters, such as: # DB Version: 16, # OS Type: linux, # DB Type: web, # Total Memory (RAM): 4 GB, # CPUs num: 8, # Connections num: 25, # Data Storage: ssd, max_connections = 25, shared_buffers = 1GB, effective_cache_size = 3GB, maintenance_work_mem = 256MB, checkpoint_completion_target = 0.9, wal_buffers = 16MB, default_statistics_target = 100, random_page_cost = 1.1, effective_io_concurrency = 200, work_mem = 10485kB, huge_pages = off, min_wal_size = 1GB, max_wal_size = 4GB, max_worker_processes = 8, max_parallel_workers_per_gather = 4, max_parallel_workers = 8, max_parallel_maintenance_workers = 4. A 'Copy configuration' button is at the bottom right.

[1] <https://www.youtube.com/watch?v=9nyY-psbMs>

[2] <https://medium.com/@soulaimaneyh/nginx-internal-architecture-b94b013bc365>

[3] <https://wiki-sinp.cbn-alpin.fr/serveurs/installation/web-srv/nginx>

[4] <https://github.com/cbn-alpin/sinp-aura-srv/blob/main/web-srv/etc/nginx/sites-available/geonature.conf>

Bonnes pratiques de développement

- Transfert de données via l'API : minimum et constant
 - Pagination des web services
 - `paginate(select, *, page=None, per_page=None, max_per_page=None, error_out=True, count=True)` (flok-sqlalchemy)
- Construction des requêtes
 - Réduire le nombre de jointures dans les requêtes SQL
 - Favoriser l'utilisation d'INDEX
- Utilisation d'outils de diagnostic d'exécution des requêtes par PostgreSQL à l'aide de EXPLAIN
(<https://docs.postgresql.fr/15/sql-explain.html>)
- Analyse des besoins utilisateurs !
 - Pré-calculer et stocker les résultats de requêtes courantes (utilisation de vues matérialisées)

Cas d'étude : le module de Synthèse

Sources de ralentissements

Solutions envisagées

Résultats des tests : tuiles MVT

Résultats des tests : SQL

Pour en savoir plus :
wiki-sinp.cbn-alpin.fr

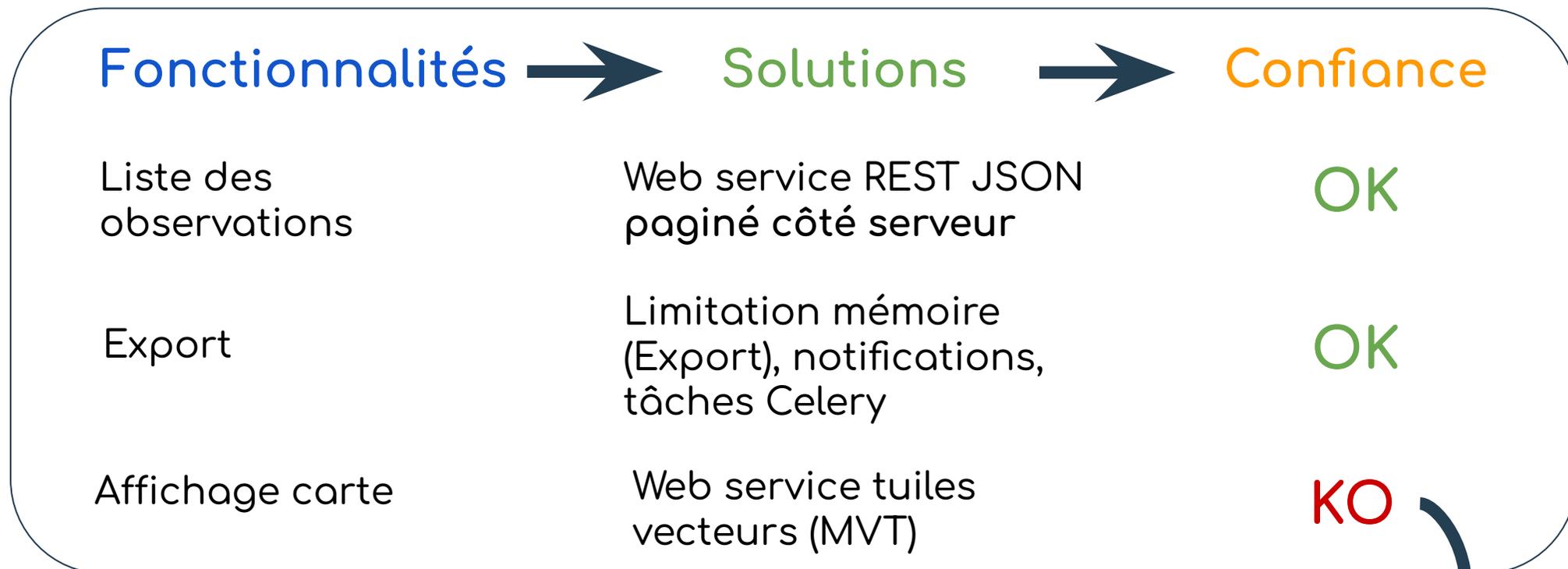


Source de ralentissements

- Recherche de la Synthèse (SINP AURA : 29 millions d'observations)
 - Web service **non paginé** (=> limite des 50 000 observations)
 - **Un seul web service** (GeoJson) pour récupérer géométries et informations
- Utilisation de **vues intermédiaires** (`v_synthese_for_...`) avec de **nombreuses jointures**
- Insertion de données dans la Synthèse :
 - => un trigger pour `cor_area_synthese` (378 millions de lignes)
 - => `st_intersects()` avec géométries de `l_areas`
 - => Si beaucoup de géométries dans `l_areas`
 - => **coût élevé !**

Solutions envisagées

Création de web services spécifiques : carte, liste, export



Amélioration des performances en :

- parallélisant l'accès aux données
- diminuant le volume de données par appel de web service
- agrégeant les données à petite échelle

Résultats tests : tuiles MVT

Avantages

- création de tuile très facile : fonctions Postgis
- parallélisation des requêtes en base : via appels de web services distincts
- nouvelles fonctionnalités : relief, 2.5D, rotation, thèmes, I18N...

Inconvénients

- performance accrue à certain niveau de zoom seulement
- mauvaise intégration à Leaflet (plugin) ou changement de framework (Maplibre)
- nécessite requêtes très rapides pour garantir fluidité de l'interface

Continuer à utiliser **Leaflet** et le **GeoJson**

=> **amélioration des requêtes SQL !**

Résultats tests : SQL

Pour améliorer les performances, utiliser :

- géométrie avec SRID 4326 pour les intersections
- opérateur `&&` plutôt que `st_intersects`
- index `SP-GIST` avec géométrie ne se chevauchant pas, sinon `GIST` plutôt que `BRIN`
 - Pour `synthese` ⇒ `SP-GIST`, pour `l_areas` ⇒ `GIST`
- table de relation (`cor_area_synthese`) plutôt que recherche spatiale
 - ajout d'une colonne `area_type_code` sur `cor_area_synthese`
- `st_intersects()` sur géométries complexes simplifiées via `st_subdivide()`
- table/VM de cache stockant les résultats de calcul (`COUNT`)
- requêtes s'exécutant uniquement sur une seule table bien indexée sans jointure

Requêtes du module Synthèse sur
une seule vue matérialisée bien indexée !

Outils d'analyse des performances de GeoNature

Outils de développement

Montée en charge

Suivi des erreurs

Analyse requêtes SQL

Surveillance système

Et vous ?



Outils dans GeoNature

pytest-benchmark

- Depuis la 2.14.1, ajout de tests de performances sur les routes avec de long temps de calculs
 - Utilisation de *pytest* et de l'extension *pytest-benchmark*

```
----- benchmark 'gn_meta': 2 tests -----
Name (time in ms)      Min          Max          Mean          StdDev          Median          IQR          Outliers          OPS          Rounds  Iterations
-----
test_list_datasets      9.9540 (1.0)  15.2306 (1.0)  11.1078 (1.0)  1.2337 (3.42)  10.6634 (1.0)  1.5744 (2.47)  8;1  90.0267 (1.0)  48  1
test_list_acquisition_frameworks  88.1332 (8.85)  89.0526 (5.85)  88.5526 (7.97)  0.3605 (1.0)  88.4947 (8.30)  0.6371 (1.0)  2;0  11.2927 (0.13)  7  1
-----

----- benchmark 'occhab': 6 tests -----
Name (time in ms)      Min          Max          Mean          StdDev          Median          IQR          Outliers          OPS          Rounds  Iterations
-----
test_export_all_habitats_geojson  3.5089 (1.0)  9.6823 (1.71)  3.8072 (1.01)  0.6479 (1.59)  3.5760 (1.00)  0.1514 (1.0)  19;50  262.6591 (0.99)  256
test_export_all_habitats_shapefile  3.5117 (1.00)  6.0609 (1.07)  3.8516 (1.03)  0.4501 (1.10)  3.6284 (1.02)  0.5820 (3.85)  37;8  259.6330 (0.98)  217
test_export_all_habitats_csv      3.5138 (1.00)  5.6726 (1.0)  3.7569 (1.0)  0.4081 (1.0)  3.5624 (1.0)  0.2225 (1.47)  32;33  266.1763 (1.0)  195
test_get_station                  8.0415 (2.29)  9.6859 (1.71)  8.5245 (2.27)  0.4615 (1.13)  8.4751 (2.38)  0.6817 (4.50)  4;0  117.3093 (0.44)  21
test_list_stations                12.5866 (3.59)  17.5694 (3.10)  14.1588 (3.77)  1.1220 (2.75)  14.3941 (4.04)  1.8858 (12.46)  20;0  70.6275 (0.27)  53
test_list_stations_restricted     15.9495 (4.55)  131.6148 (23.20)  20.0906 (5.35)  18.8780 (46.26)  16.6401 (4.67)  1.2055 (7.96)  1;3  49.7744 (0.19)  37
-----

----- benchmark 'occtax': 2 tests -----
Name (time in ms)      Min          Max          Mean          StdDev          Median          IQR          Outliers          OPS          Rounds  Iterations
-----
test_list_relevés_restricted     19.0176 (1.0)  137.0526 (4.43)  30.5188 (1.03)  35.3517 (67.56)  19.6464 (1.0)  0.8446 (1.31)  1;2  32.7667 (0.98)  11  1
test_list_relevés_unrestricted   29.1133 (1.53)  30.9189 (1.0)  29.7602 (1.0)  0.5233 (1.0)  29.5907 (1.51)  0.6436 (1.0)  4;0  33.6019 (1.0)  13  1
-----

----- benchmark 'ref_geo': 2 tests -----
Name (time in ms)      Min          Max          Mean          StdDev          Median          IQR          Outliers          OPS          Rounds  Iterations
-----
test_get_areas_without_geom     91.7221 (1.0)  97.7884 (1.0)  94.2566 (1.0)  1.8153 (1.0)  94.6034 (1.0)  2.3623 (1.0)  4;0  10.6093 (1.0)  11  1
test_get_areas_with_geom       104.5232 (1.14)  118.9449 (1.22)  108.4920 (1.15)  5.8981 (3.25)  106.3268 (1.12)  3.9748 (1.68)  1;1  9.2173 (0.87)  5  1
-----

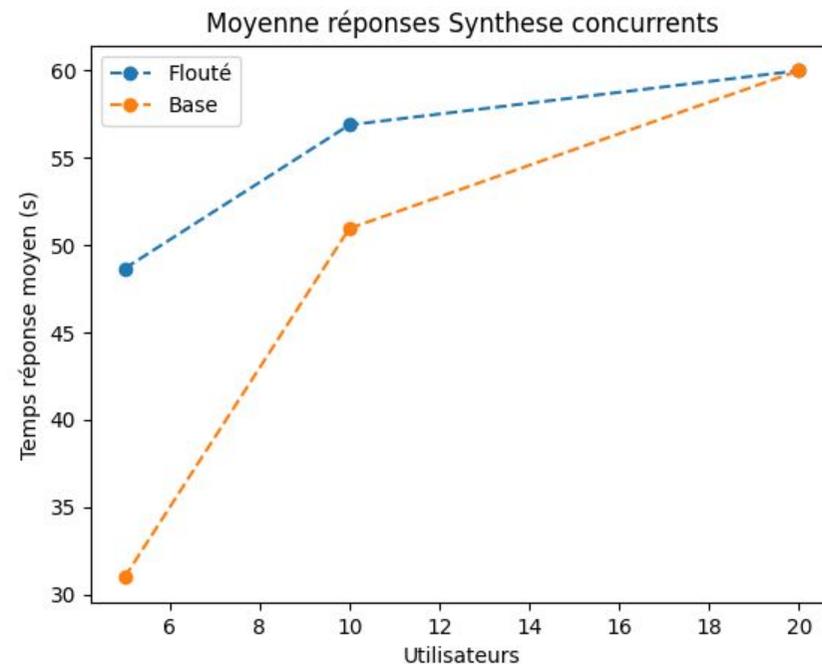
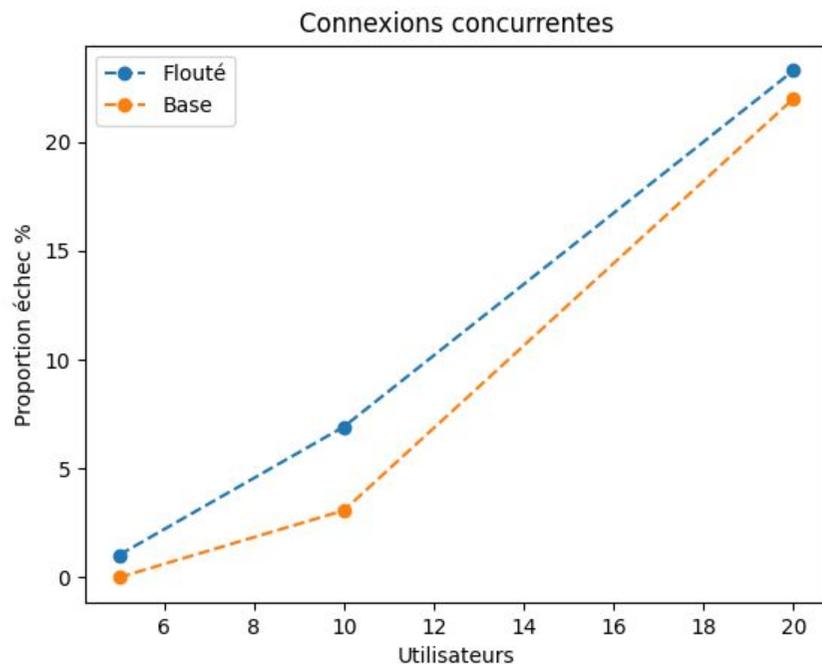
----- benchmark 'synthese': 38 tests -----
Name (time in us)      Min          Max          Mean          StdDev          Median          IQR          Outliers          OPS          Rounds  Iterations
-----
OPS
Rounds  Iterations
```

Figure. Extrait de la sortie de *pytest-benchmark* dans GeoNature

Test de montée en charge - Gatling

Travaux menés par Antoine Cecchi (PatriNat)

- Scénario : chargement des données dans la Synthèse sans filtre + floutage
- Configuration Serveur : CPU 2 coeurs + RAM 8Go



Utiliser Sentry : analyse performances, suivi erreurs

The screenshot displays the Sentry Performance dashboard for the organization 'Parc national de...'. The left sidebar contains navigation options: Issues, Projects, Performance, Alerts, Discover, Dashboards, Releases, User Feedback, Stats, and Settings. The main content area features three summary charts: 'User Misery' (0,314), 'Transactions Per Minute' (30,378), and 'Failure Rate'. A 'Project Thresholds' tooltip is visible, explaining that gauges performance using metrics like response time, per project, for the Apdex and User Misery Scores. Below the charts is a table of transaction performance metrics.

TRANSACTION	PROJECT	TPM	P50 ↓	P95	FAILURE RATE	APDEX	USERS	USER MISERY
geonature.core.gn_profiles.tasks.refresh_profiles	depobio2-prod-geonature	0.0006	1.06hr	1.06hr	0%	0	0	
gn_module_dashboard.tasks.refresh_vm	depobio2-prod-geonature	0.0006	14.31min	14.31min	0%	0	0	
gn_module_dashboard.tasks.refresh_vm	geonat-test-geonature	0.0006	7.03min	7.03min	0%	0	0	
gn_module_dashboard.tasks.refresh_vm	ginco2-centre-geonature	0.0006	4.10min	4.10min	0%	0	0	
gn_module_dashboard.tasks.refresh_vm	ginco2-occtax-geonature	0.0006	1.66min	1.66min	0%	0	0	
gn_module_dashboard.tasks.refresh_vm	depobio2-preprod-geonature	0.0006	53.11s	53.11s	0%	0	0	
gn_module_import.tasks.do_import_in_synthese	ginco-nationale-preprod-geonature	0.0013	42.17s	1.33min	0%	0.5	0	
geonature.core.gn_profiles.tasks.refresh_profiles	ginco2-guyane-geonature	0.0006	34.70s	34.70s	0%	0	0	
geonature.core.gn_profiles.tasks.refresh_profiles	ginco2-centre-geonature	0.0006	25.77s	25.77s	0%	0	0	
gn_synthese.get_autocomplete_taxons_synthese	ginco2-occtax-geonature	0.0006	20.09s	20.09s	0%	0	1	
geonature.core.gn_profiles.tasks.refresh_profiles	depobio2-preprod-geonature	0.0006	17.10s	17.10s	0%	0	0	
geonature.core.gn_profiles.tasks.refresh_profiles	geonat-test-geonature	0.0006	16.28s	16.28s	0%	0	0	
gn_module_dashboard.tasks.refresh_vm	ginco2-guyane-geonature	0.0006	13.30s	13.30s	0%	0	0	

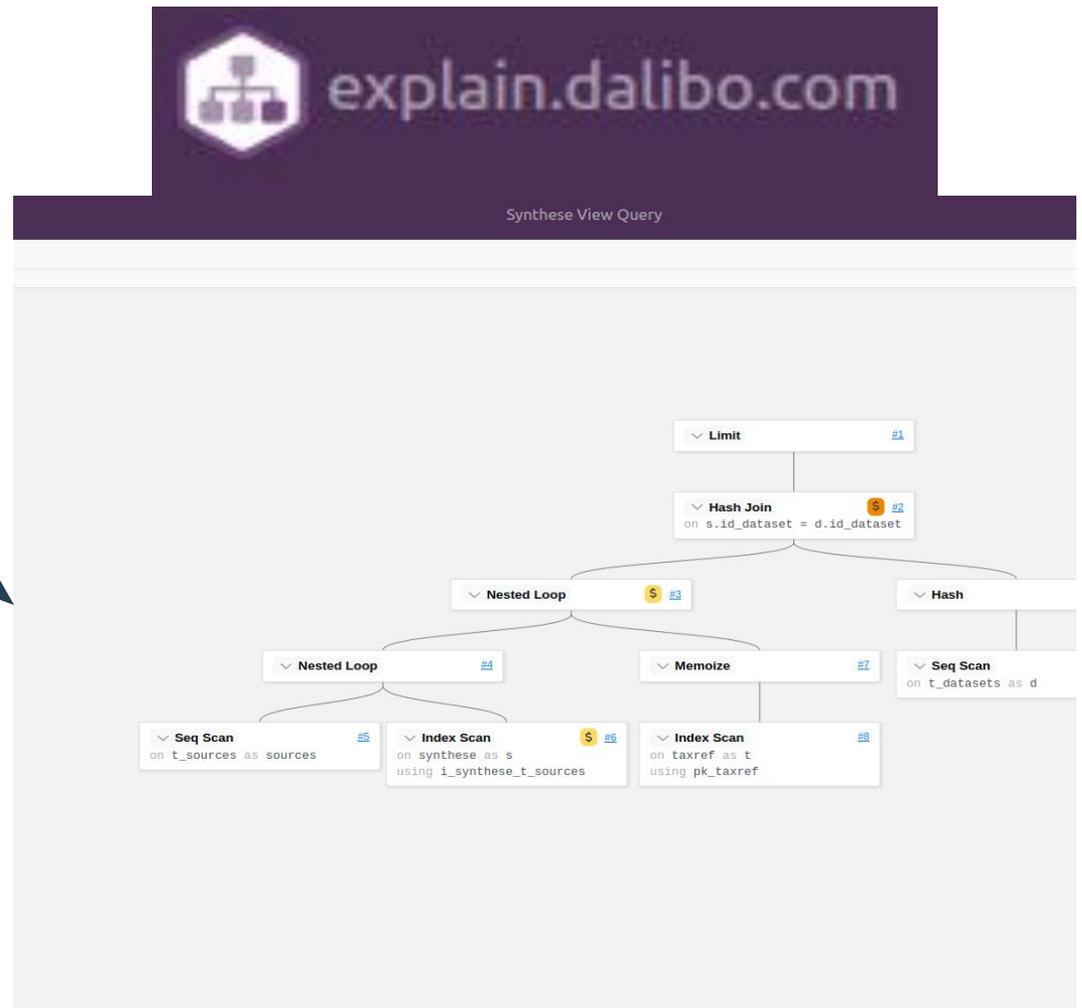
Alternative => GiltchTip

PEV : analyser requêtes SQL

REQUÊTE dans Postgresql

```
EXPLAIN select *  
from  
gn_synthese.v_synthese  
_for_web_app  
limit 100
```

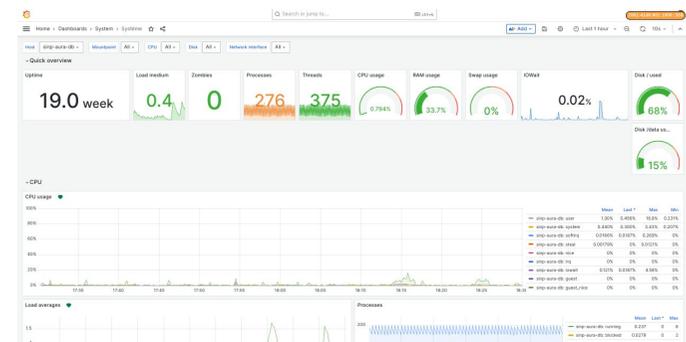
Résultat
du
EXPLAIN



Monitoring

Solution sous Docker :

- **Telegraf, InfluxDB, Grafana** => système
- **Portainer** => Docker
- **GiltchTip** => un Sentry allégé
- **HeathChecks** => exécution des Cron
- **GoAccess & Awstats** => logs du serveur web
- **Matomo** => stats dynamiques des sites web
- **PG Activity** => commande `top` pour Postgresql

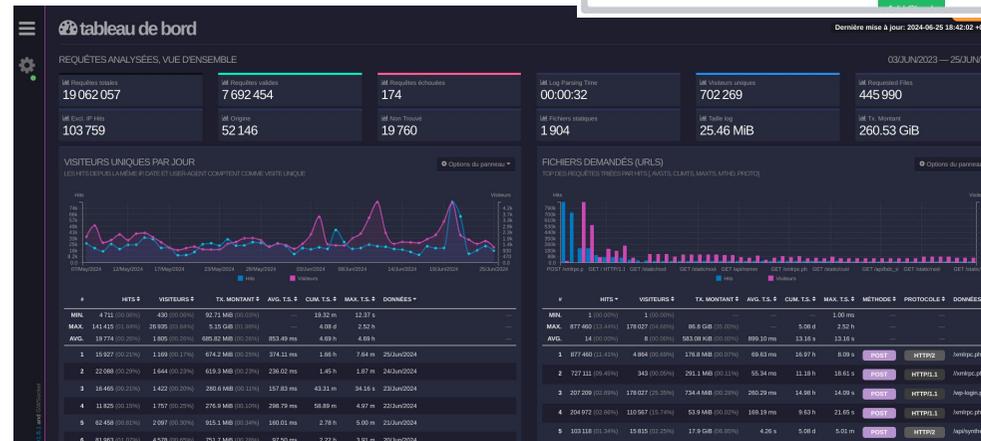


A screenshot of the Portainer interface. The top bar says 'Upgrade to Business Edition' and 'Environments'. The main content area is titled 'Home' and shows a list of Docker environments. Each environment entry includes a name (e.g., 'bhp-aura-simp'), a group, status, creation time, and resource usage (CPU, RAM, Volumes). There are 'Live connect' buttons for each environment.

A screenshot of a dashboard showing a list of checks. The table has columns: Name, UUID, Integrations, Period, and Last Ping. The data rows include checks like 'root backups', 'Prod as byz', 'DB Backups', 'Weekly Reports', and 'Clean Uploads'.

Name	UUID	Integrations	Period	Last Ping
root backups	a46958bc-341a-4044-b08e-c19f158320e9		12 hours	2 months, 1 week ago
Prod as byz	ff1342d3-00ea-44fa-b2f9-339f3980c0d0		5 minutes	5 minutes ago
DB Backups	7c324798-5d75-4fc4-a272-4380e184c899		1 day	25 minutes ago
Weekly Reports	f6d8d3b-aa7a-43a2-87d5-a7b6d67f4815		1 week	25 minutes ago
Clean Uploads	1bae1439-5d88-4ec1-a652-55a362131005		1 day	7 months, 2 weeks ago

A screenshot of the GiltchTip interface. It shows a sidebar with navigation options like 'Biodiv/AURA', 'Problems', 'Monteurs de disponibilité', 'Performance', 'Projets', 'Versions', 'Paramètres', 'Profil', 'Se déconnecter', and 'Soutenir GiltchTip'. The main area displays an 'AttributeError' event with details like 'NoneType' object has no attribute 'is_dict'. There are also sections for 'Tags' and 'Exception (most recent call first)'.



Et vous ?

Questions ouvertes

- **Quelle est votre infrastructure serveur pour votre GeoNature ?**
(avantages/inconvénients)
- **Quels outils utilisez-vous ?**
(monitoring, analyse de performance, statistiques, développement...)
- **Quelles solutions/techniques ont améliorés les performances de votre installation GeoNature ?**



MERCI DE VOTRE ATTENTION

Auteurs :

Jacques FIZE (PNE)

Jean-Pascal MILCENT (CBNA)