

# Visualisation interactive unifiée d'une variété de sources d'information

Un cas d'usage de Grafana

Jonathan Schaeffer <jonathan.schaeffer@univ-grenoble-alpes.fr>

rdatadev - 14 décembre 2021

## Contents

<b>1</b>	<b>Un mot de contexte</b>	<b>2</b>
<b>2</b>	<b>La problématique à résoudre</b>	<b>2</b>
2.1	Visualiser l'activité de production . . . . .	2
2.1.1	<b>Note</b> : ce n'est pas de la supervision à proprement parler . . . . .	2
2.2	Fournir un tableau de bord de pilotage . . . . .	2
2.3	Une diversité de sources d'information . . . . .	2
<b>3</b>	<b>Solution proposée</b>	<b>3</b>
3.1	Bonnes propriétés . . . . .	3
3.1.1	Permet de connecter une grande variété de sources de données . . . . .	3
3.1.2	Permet de dessiner une grande diversité de graphiques ( <i>panels</i> ) et de tableaux de bords ( <i>dashboards</i> ) . . . . .	4
3.2	Implémentation d'une solution . . . . .	4
<b>4</b>	<b>Démonstration</b>	<b>4</b>
4.1	Vues administrateur système . . . . .	4
4.1.1	Voir les ressources utilisées en regard de l'activité des webservices ATTACH . . . . .	4
4.1.2	Activité d'intégration des données ATTACH . . . . .	4
4.2	Vue responsable scientifique . . . . .	6
4.2.1	Distribution de données ATTACH . . . . .	6
4.2.2	Détails des volumes hébergés ATTACH . . . . .	6
<b>5</b>	<b>Conclusions</b>	<b>6</b>
5.1	Points d'attention . . . . .	6
5.2	Évolutions . . . . .	7



## 1 Un mot de contexte

L'entrepôt de données national Résif-DC réceptionne les données sismologiques brutes et validées et offre les services de distribution à une communauté mondiale d'utilisateurs.

Quelques ordres de grandeur :

- **20000** canaux décrits en métadonnées
- **585** canaux actifs
- **80To** de données sous forme de séries temporelles
- **80 millions** de requêtes aux webservice par an
- plus de **80To** distribués en 2021

## 2 La problématique à résoudre

### 2.1 Visualiser l'activité de production

- identifier l'intensité de l'activité (le fond, les pics)
- faire le lien avec des métriques systèmes de base ou plus évoluées
  - CPU/RAM/load
  - trafic réseau
  - charge sur le SGBD
  - le nombre de requêtes aux services d'accès à la donnée ou d'intégration

Avoir une vision instantanée et dans le temps.

#### 2.1.1 Note : ce n'est pas de la supervision à proprement parler

On n'a pas besoin de définir des seuils, des système d'alerte, etc.

### 2.2 Fournir un tableau de bord de pilotage

Montrer des statistiques demandées par la responsable scientifique. Ce sont des statistiques d'activité métier permettant d'indiquer:

- le nombre de requêtes aux données sismologiques
- le détail des livraisons de données
- des détails sur les données hébergées, ...

### 2.3 Une diversité de sources d'information

- influxdb pour les métriques systèmes
- elasticsearch pour les logs des requêtes http
- postgresql pour les statistiques métier calculées par des programmes d'analyse de logs et d'agrégation.

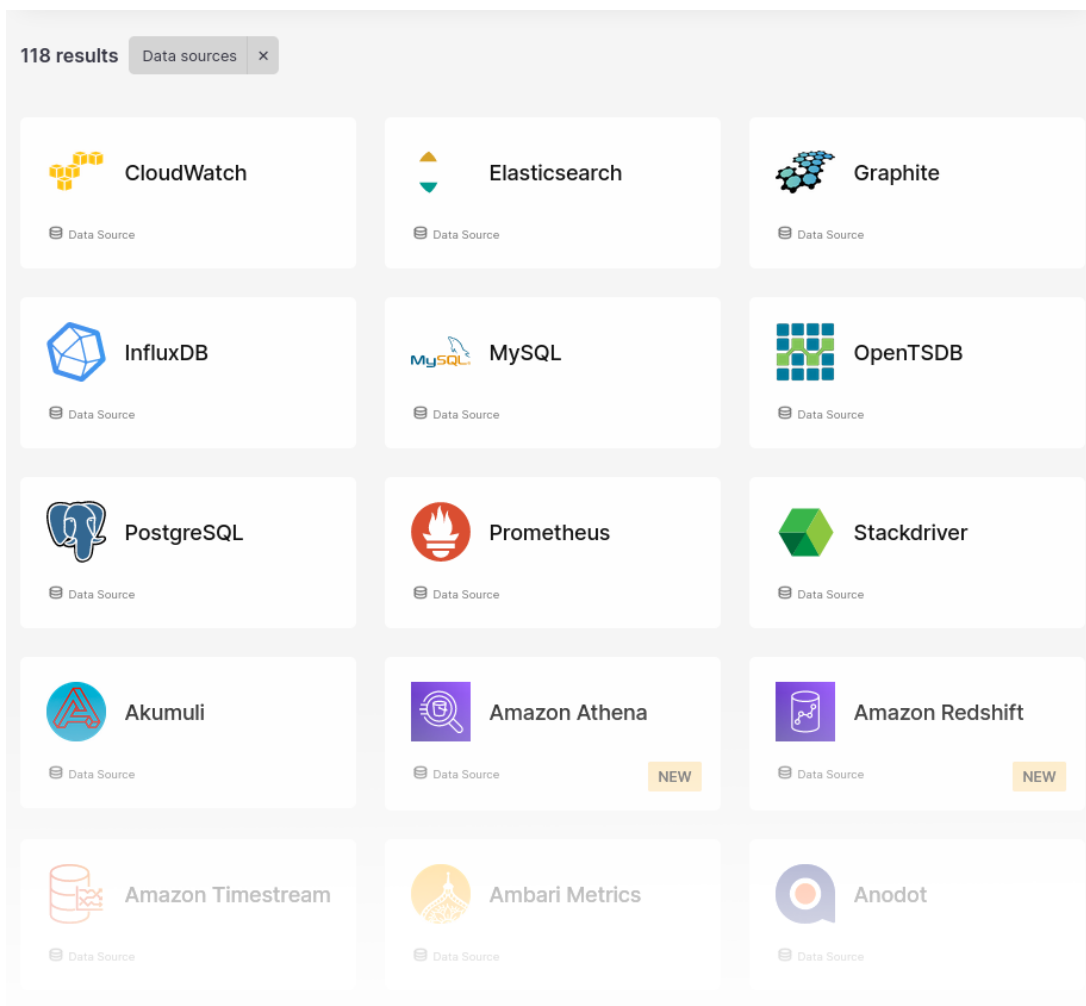
### 3 Solution proposée



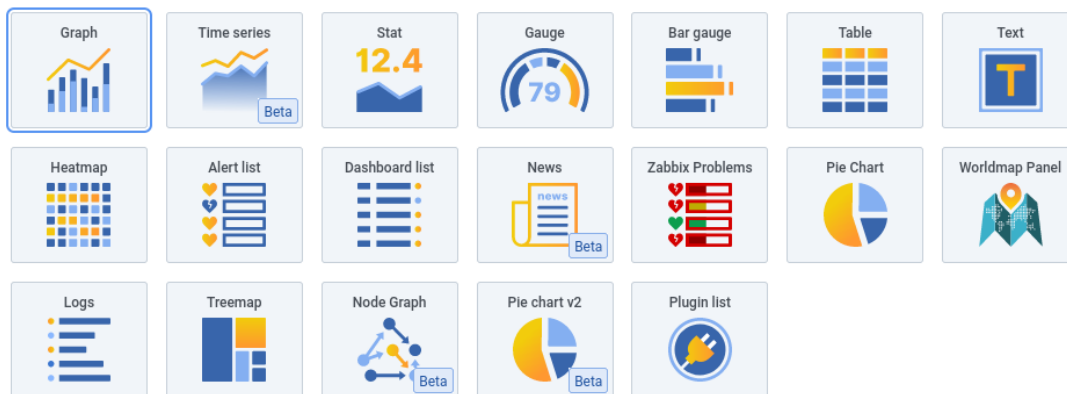
à la rescousse !

#### 3.1 Bonnes propriétés

##### 3.1.1 Permet de connecter une grande variété de sources de données



### 3.1.2 Permet de dessiner une grande diversité de graphiques (*panels*) et de tableaux de bords (*dashboards*)



## 3.2 Implémentation d'une solution

1. Configuration d'une *organisation* pour Résif-DC
2. Déclaration de plusieurs *datasources* dans l'organisation
  - postgresql, base de données des statistiques
  - influxdb pour les métriques système
  - elasticsearch pour l'activité des requêtes http instantannée
3. Conception de plusieurs *dashboards* correspondant chacun à un point de vue (administrateur d'infra, responsable scientifique)
  - les 3 *datasources* sont bien intégrés dans l'outil de conception des panels
  - mais rapidement, on a voulu affiner les requêtes : SQL a été d'une grande aide pour la partie postgres. Grafana permet d'écrire une requête tant que le format de retour est bien supporté par le type de visualisation.

## 4 Démonstration

### 4.1 Vues administrateur système

#### 4.1.1 Voir les ressources utilisées en regard de l'activité des webservices ATTACH

- datasource elasticsearch pour les webservices instantannés
- datasource influxdb pour les métriques systèmes

Dashboard de démonstration

#### 4.1.2 Activité d'intégration des données ATTACH

Dashboard pour visualiser différentes étapes d'intégration de la donnée et comment tous les process répartis se comportent.

Lien du dashboard de démonstration



Figure 1: Dashboard pour croiser les requêtes aux webservices (en haut) et l'activité système



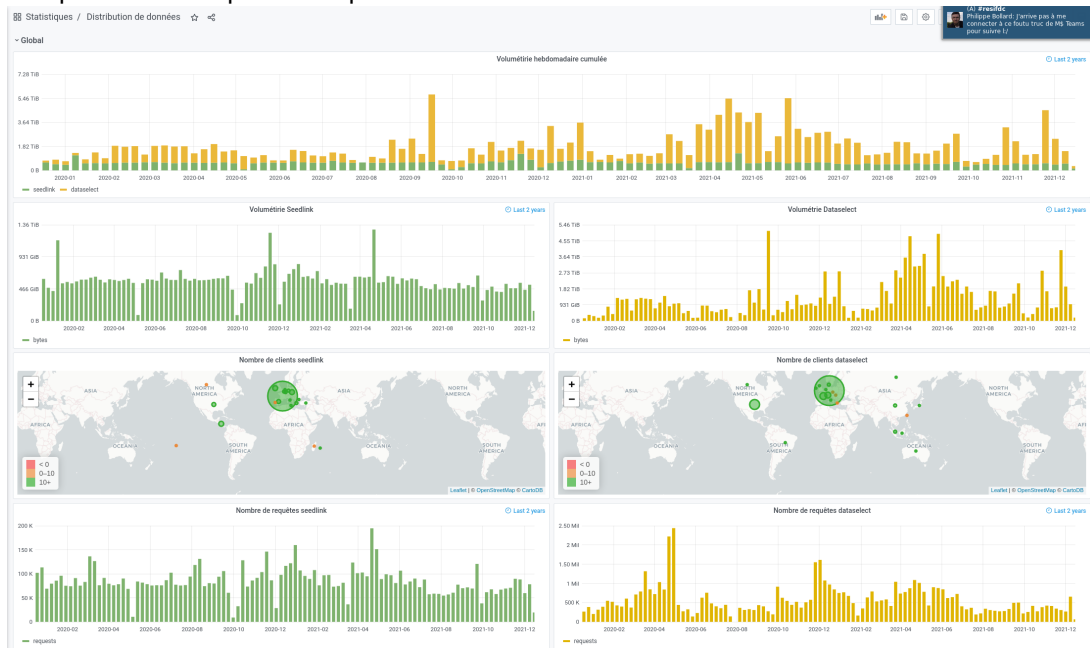
Figure 2: Dashboard pour croiser l'activité d'intégration et les ressources de calcul mobilisées

## 4.2 Vue responsable scientifique

### 4.2.1 Distribution de données

ATTACH

Pour le protocole temps réel et pour le webservice Dashboard de démonstration



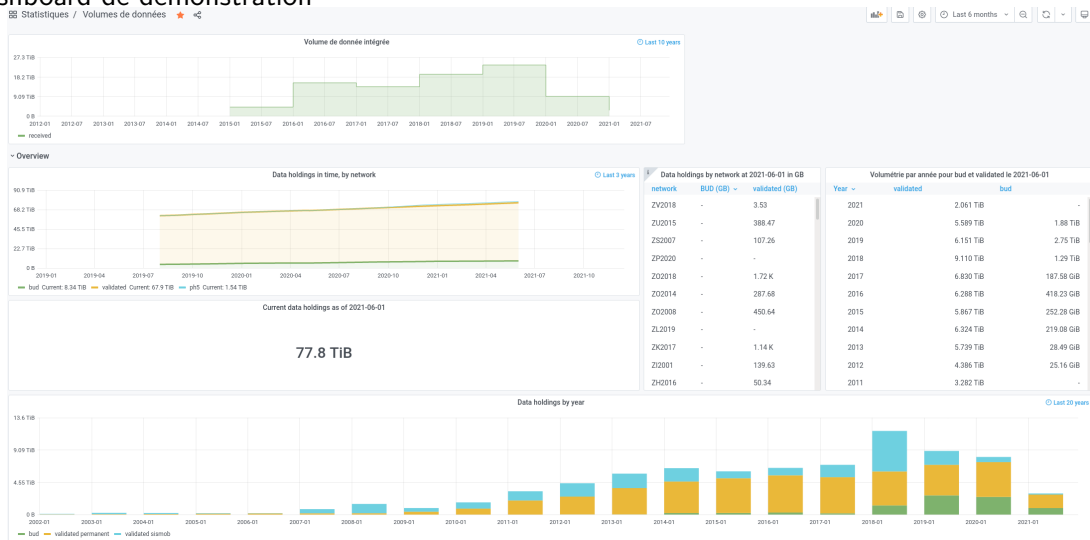
Dash-

board pour regarder la distribution des données aux utilisateurs

### 4.2.2 Détails des volumes hébergés

ATTACH

Dashboard de démonstration



## 5 Conclusions

L'outil de visualisation est très apprécié par les utilisateurs.

### 5.1 Points d'attention

- La conception de panels et de dashboards est assez fastidieuse. Mieux vaut bien identifier ce qu'on souhaite visualiser avant de se lancer.

- La maintenance des statistiques et de la visualisation doit être prise en compte

## 5.2 **Évolutions**

On n'est pas très loin de centraliser les logs et de les exposer dans certains dashboards