

DATASTREAM ELASTICSEARCH

Mehdi BOISSAT-BRON

leboncoin



Qui suis-je ?

Mehdi BOISSAT-BRON

Développeur Backend
Feature Team Import



mbbo.org



linkedin.com/in/mbbo128



twitter.com/mbbo128



github.com/mbbo128



leboncoin, leader de l'économie collaborative

Vous le savez peut-être mais leboncoin.fr c'est :

Le **6ème** site visité de France
39 millions d'annonces en ligne
32 millions d'utilisateurs uniques par mois
800 000 nouvelles annonces par jour
10 milliards de pages vues par mois

C'est aussi un groupe :

~ **1500** Collaborateurs
~ **600** Personnes Product & Tech
~ **60** Feature teams
[Blog tech](#)
[Page welcome to the jungle](#)

Sur Elasticsearch :

14 clusters en prod 2 pour l'infrastructure 1 pour la sécurité, et les autres dédiés aux feature team selon les besoins
1K/2K requêtes par seconde et **18 millions** de documents pour le plus gros cluster ES dédié

leboncoin
GROUPE

VIDE·DRESSING

à vendre
à louer

locasun

Groupe Argus

PAYCAR

PILGO

Agriaffaires

MachineryZone

SOMMAIRE

01. INTRODUCTION

SLIDE 5

02. FONCTIONNEMENT

SLIDE 8

03. DEMO POSTMAN

SLIDE 11

04. STACK MANAGEMENT

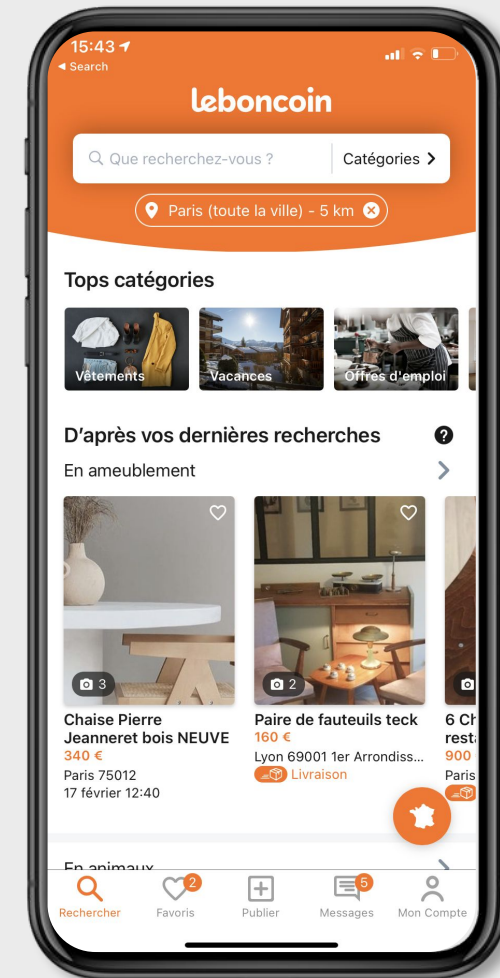
SLIDE 14

05. CONFIGURATION

SLIDE 16

06. CONCLUSION

SLIDE 18



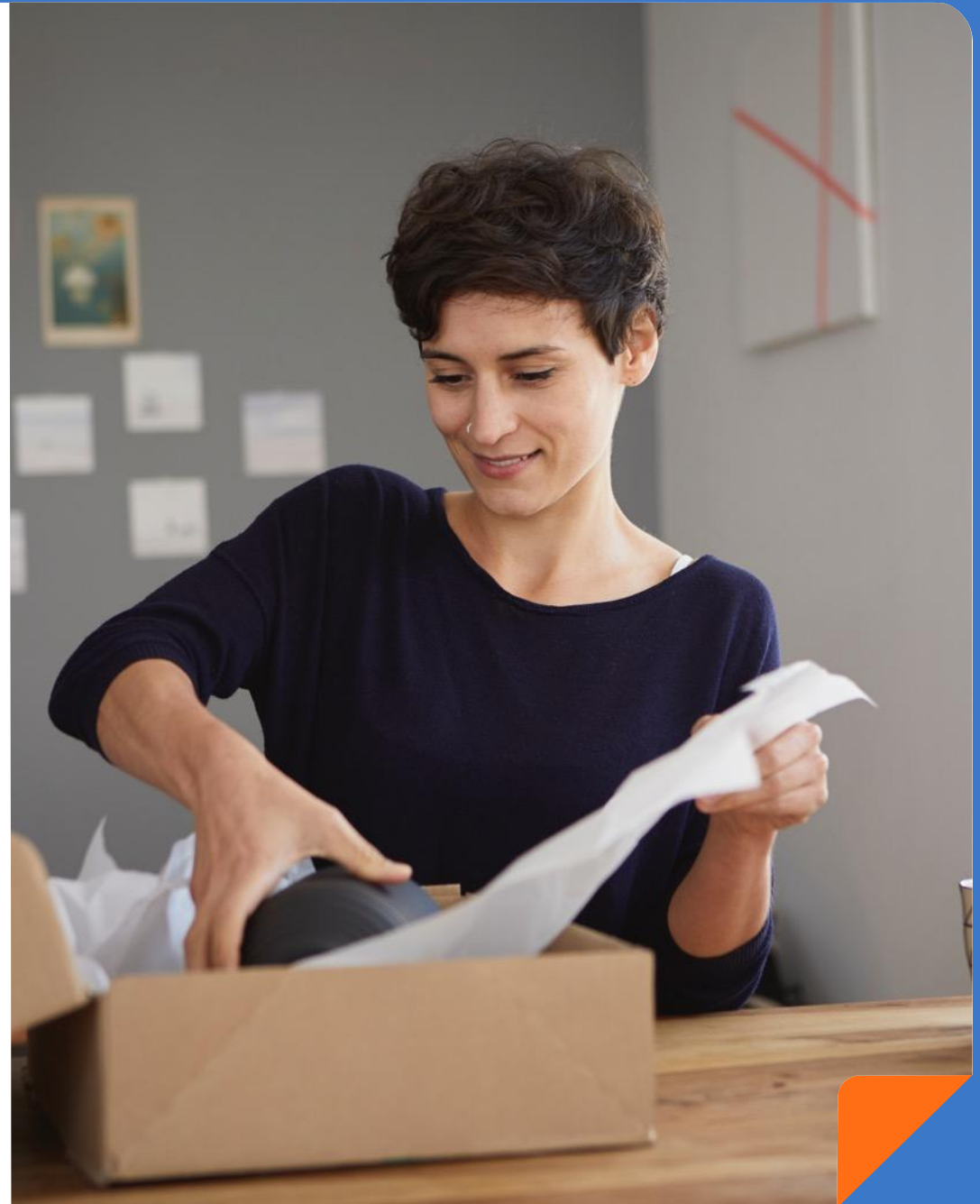
01

INTRODUCTION

INTRODUCTION

Problématique

Les clients diffusant sur leboncoin en utilisant les API ne disposent pas forcément de retours concernant ce qu'il s'est déroulé pendant leur dernière tentative de diffusions



INTRODUCTION

Diffusion en direct

Vue sur les 7 derniers jours

[Voir mes annonces](#)

✖ 2 références en erreur

✔ 343 annonces en ligne

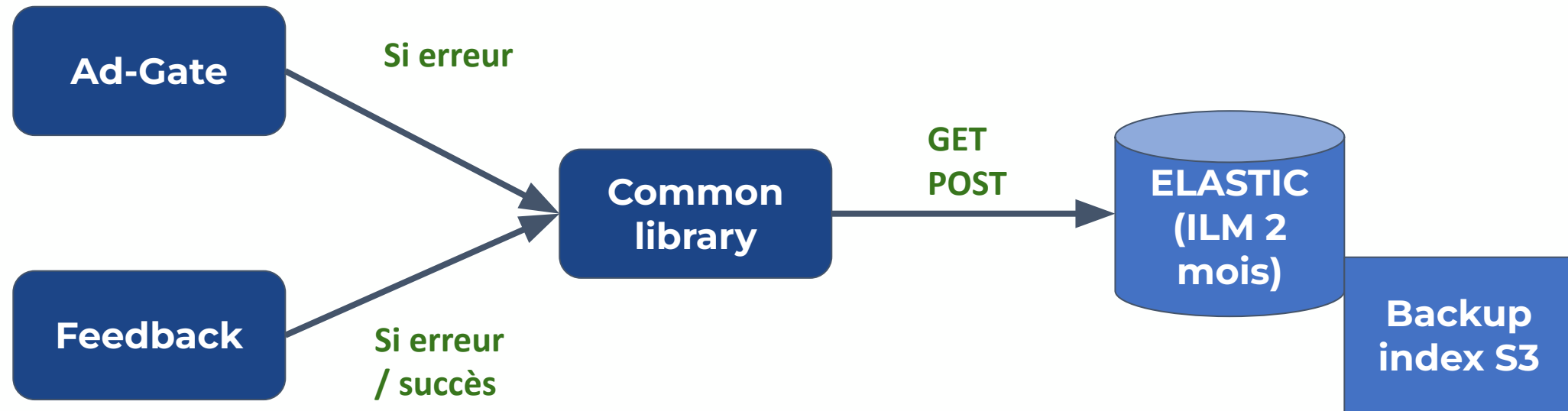
Références en erreur	Message d'erreur	Explication	
9438_Terrain_a_batir_1106m2 Maison a batir 1106 m2 avec un superbe terrain constructible. Affaire a saisir rapid...	La référence de l'annonce doit être renseignée sans dépasser 30 caractères (règle du portail)	La référence de l'annonce doit être renseignée sans dépasser 30 caractères (règle du portail)	10 fev.
9438_Terrain_a_batir_1106m2 Maison a batir 1106 m2 avec un superbe terrain constructible. Affaire a saisir rapid...	La référence de l'annonce doit être renseignée sans dépasser 30 caractères (règle du portail)	La référence de l'annonce doit être renseignée sans dépasser 30 caractères (règle du portail)	10 fev.
9438_Terrain_a_batir_1106m2 Maison a batir 1106 m2 avec un superbe terrain constructible. Affaire a saisir rapid...	La référence de l'annonce doit être renseignée sans dépasser 30 caractères (règle du portail)	La référence de l'annonce doit être renseignée sans dépasser 30 caractères (règle du portail)	9 fev.
9438_Terrain_a_batir_1106m2 Maison a batir 1106 m2 avec un superbe terrain constructible. Affaire a saisir rapid...	La référence de l'annonce doit être renseignée sans dépasser 30 caractères (règle du portail)	La référence de l'annonce doit être renseignée sans dépasser 30 caractères (règle du portail)	2 fev.
9438_Terrain_a_batir_1106m2 Maison a batir 1106 m2 avec un superbe terrain constructible. Affaire a saisir rapid...	La référence de l'annonce doit être renseignée sans dépasser 30 caractères (règle du portail)	La référence de l'annonce doit être renseignée sans dépasser 30 caractères (règle du portail)	19 jan.

02

FONCTIONNEMENT

FONCTIONNEMENT

Schéma



FONCTIONNEMENT

Solutions

Solution 1

Un index par jour suffix date

broadcast-monitoring-01-05
broadcast-monitoring-02-05
broadcast-monitoring-03-05
...



Gestion dans le code

Solution 2

Rollover automatique

broadcast-monitoring-00001
broadcast-monitoring-00002
broadcast-monitoring-00003
...



Plus recommandé

Solution 3

Datastream

broadcast-monitoring
.ds-broadcast-monitoring-[date]-000001
.ds-broadcast-monitoring-[date]-000002
.ds-broadcast-monitoring-[date]-000003
...



Solution retenue

03

DEMO POSTMAN

DEMO POSTMAN

Index lifecycle management

Définition d'une policy qui se nomme **meetup-policy** se trouve les règles pour déclencher un rollover d'index ainsi que la durée de vie de l'index jusqu'à sa suppression



DEMO POSTMAN

Index template

Le data stream utilise un index template nommé **meetup-template**

L'index template peut contenir un **component template** pour définir le mapping de nos index.



04

STACK MANAGEMENT

STACK MANAGEMENT

Staging

Le data stream utilise un **index template** nommé broadcast-monitoring

L'index template lui-même utilise un **component templates** nommé morpheus-mapping qui contient le mapping de nos index.

Tous ces éléments sont visibles dans la section index management du kibana.

05

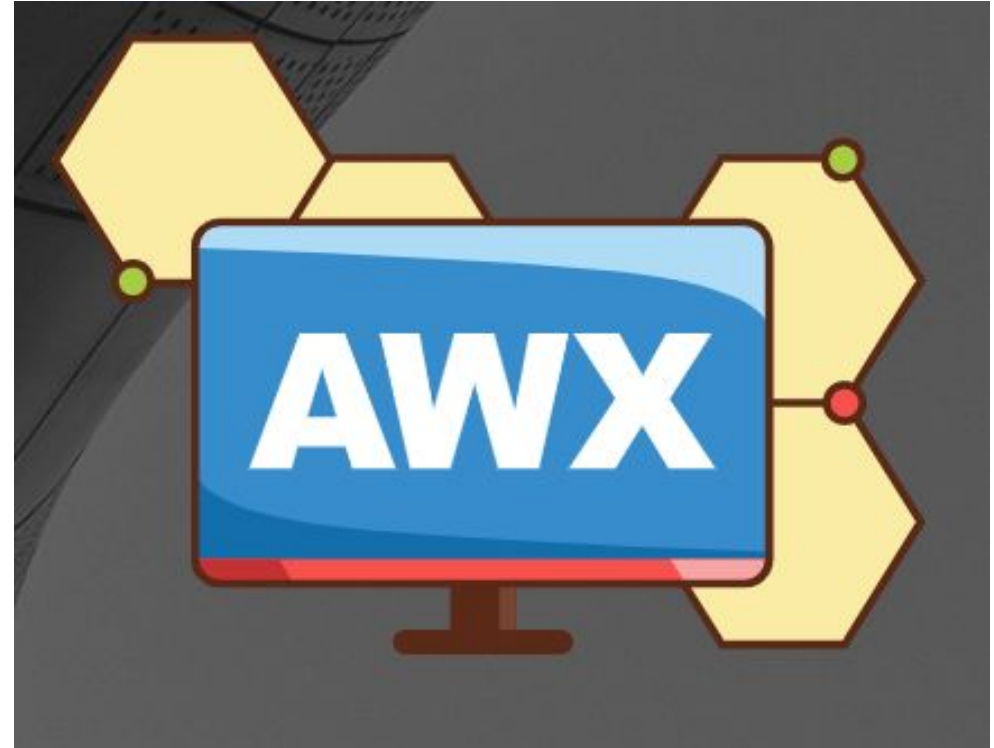
CONFIGURATION

CONFIGURATION

Le configuration sur kibana est versionnée et se fait via un projet ansible

Voici un exemple de la configuration de staging

La conf est load dans **AWX**



06

CONCLUSION

CONCLUSION

Bénéfices :

- Abstraction concernant la création des index
- Scalabilité pour le stockage des données
- Possibilité de définir une durée de vie pour les index
- Simplification au niveau de l'archivage

Documentations :

elastic.co/guide/en/elasticsearch/reference/master/data-streams.html

elastic.co/guide/en/elasticsearch/reference/current/set-up-a-data-stream.html

Sources :

github.com/mbbo128/datastream

Merci

Avez-vous des questions ?