

Atelier **A**DEMEC



Présentation par **Ségoène Albouy**
ingénieure d'études à l'Observatoire de Paris

—
Supports de l'atelier :

<https://cutt.ly/ke8Cp3U>

Plan de l'atelier

C'est quoi ?

Présentation de la suite elastic

À quoi ça sert ?

Dans quels buts et quels contextes utiliser ces technologies

Comment ça marche ?

Quelques mots sur le fonctionnement d'elasticsearch

Pourquoi kibana ?

Choisir son outil de visualisations de données

Point vocabulaire

Les termes à connaître

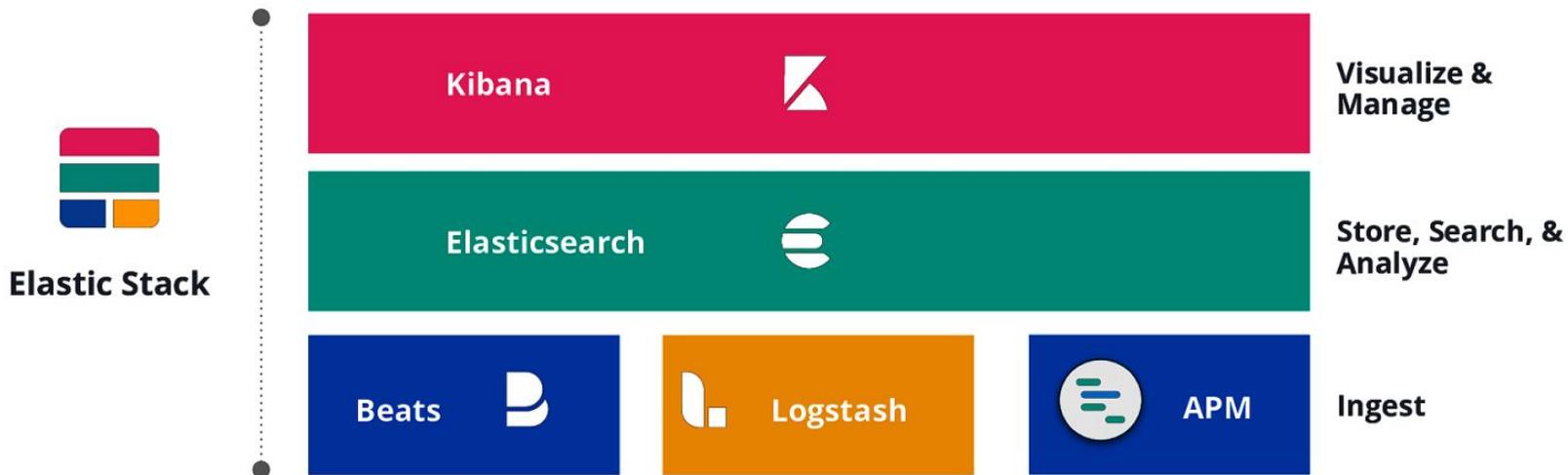
Travaux pratiques !

—
C'est quoi ?

**Présentation de la suite
elastic.**

C'est quoi ?

Elastic stack ?



Elasticsearch & kibana : *un duo de choc*

Elasticsearch

- Stockage des données
- Modalités de recherche
- Capacités analytiques

Kibana

- Fenêtre sur elastic stack
- Visualisation de données
- Accès à X-Pack et extensions avec *plugins*

—
À quoi ça sert ?

Dans quels buts et quels contextes utiliser ces technologies.

Recherche & analyse

Recherche

- **Zoom** dans un *dataset* pour trouver l'information désirée
- Parcourir de grand corpus de données textuelles
- Chercher parmi des données structurées

Analyse

- **Mise en perspective** d'un *dataset* pour contextualiser
- Calcul de statistiques sur les données
- Modalités d'agrégation avancées

Les cas d'usages

Analyse de logs

Ex : comprendre comment une application est utilisée, identifier certaines erreurs dans le code

Moteur de recherche

Ex : Rechercher parmi des articles et posts d'un forum ou les articles d'une boutique en ligne

Lien avec base de données relationnelle

Ex : Instance d'elasticsearch sur une base existante pour accélérer le requêtage de grand nombre de données

Visualiser du CSV

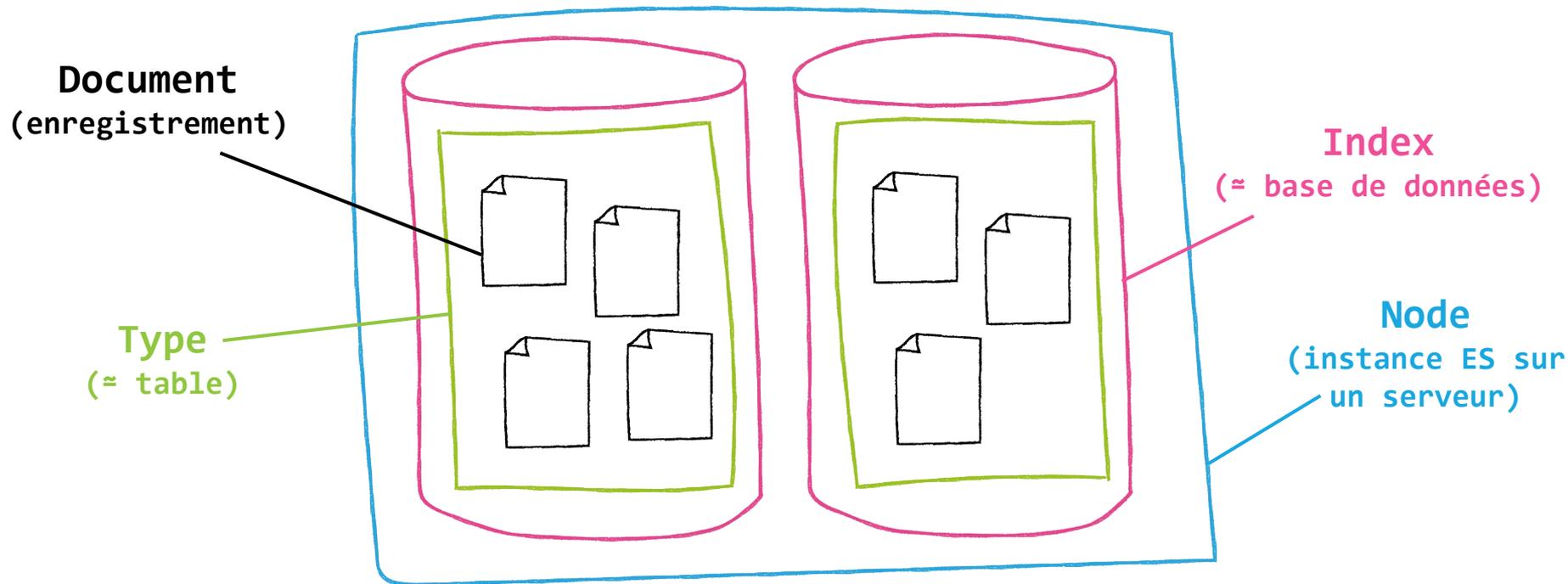
—

Comment ça marche ?

Le fonctionnement d'elasticsearch.

Comment ça marche ?

Structuration des données



Comment ça marche ?

La base de données

Auteur

id	nom	prénom
1	Zola	Émile
2	Balzac (de)	Honoré
3	Flaubert	Gustave
4	Maupassant (de)	Guy
5	Hugo	Victor

Oeuvre

id	titre	date
1	Germinal	1885
2	Le père Goriot	1842
3	Madame Bovary	1857
4	Bel ami	1885
5	L'Assomoir	1876

```
{
  "auteur": [
    {
      "id": "1",
      "prénom": "Émile",
      "nom": "Zola"
      "oeuvres": [
        {
          "id": "5",
          "titre": "L'Assomoir",
          "date": 1876
        },{
          "id": "1",
          "titre": "Germinal",
          "date": 1885
        }
      ]
    },{
      "id": "2",
      "prénom": "Honoré",
      "nom": "Balzac (de)"
      "oeuvres": [
        {
          "id": "2",
          "titre": "Le Père Goriot",
          "date": 1842
        }
      ]
    }
  ]
}
```

```
{
  "oeuvre": [
    {
      "id": "1",
      "titre": "Germinal",
      "date": 1885,
      "auteur": {
        "id": "1",
        "prénom": "Émile",
        "nom": "Zola"
      }
    },{
      "id": "2",
      "titre": "L'Assomoir",
      "date": 1876,
      "auteur": {
        "id": "1",
        "prénom": "Émile",
        "nom": "Zola"
      }
    }
  ]
}
```

Index inversé

- 1:** Ce n'est pas n'importe quel balai Harry, c'est un Nimbus 2000 !
- 2:** Tu es un sorcier Harry.
- 3:** On aurait pu se faire tuer, ou pire : se faire renvoyer !

terme	freq.	doc.
ce	2	1
est	2	1
harry	2	1,2
ne	2	1
un	2	1,2
balai	1	1
nimbus	1	1
renvoyer	1	3
sorcier	1	2

Systeme distribue : les *shards*

Horizontal scalability

- ↳ **Division des index en *shards***
- ↳ **Répartition des *shards* sur les serveurs**
- Prise en charge de grandes quantités de données
- Répartition de la puissance de calcul
- Tolérance aux pannes et résilience du système

Comment ça marche ?

Communiquer avec elasticsearch

REST API

Protocole HTTP
Requêtes JSON
Très compatible

Client API

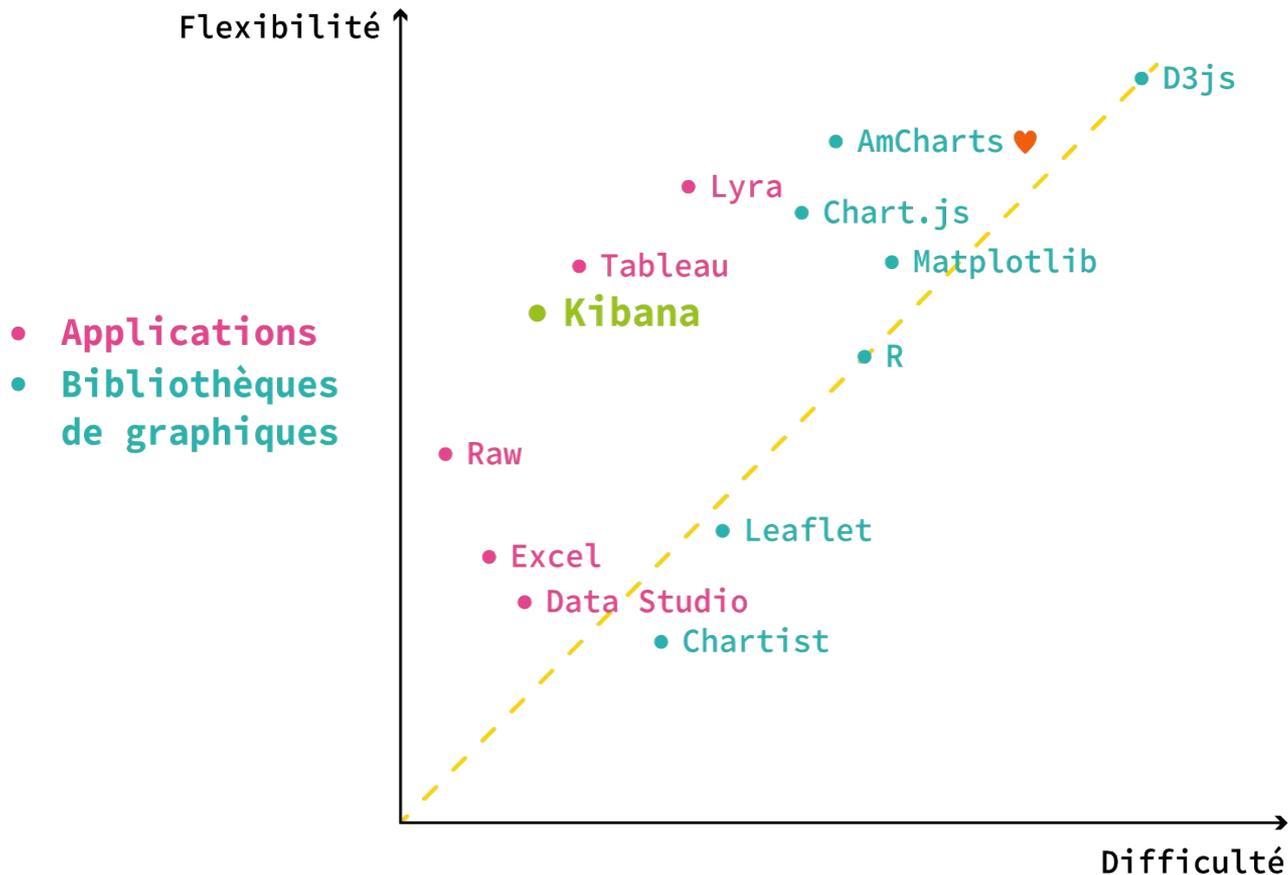
Nbx langages
Req. simplifiées
Abstraction

Application

Interface graphique
Pas de code
nécessaire

—
Pourquoi kibana ?

**Choisir son outil de
visualisation de données.**



Adapté ou pas ?

Adapté

- Visualiser dynamiquement sur un dataset évolutif
- Explorer ses données dans une visée heuristique
- Mettre en rapport des visualisations

Pas adapté

- Créer des visualisations très spécifiques
- Exporter ses visualisations au format image
- Gestion des utilisateurs multiples

—

Point vocabulaire

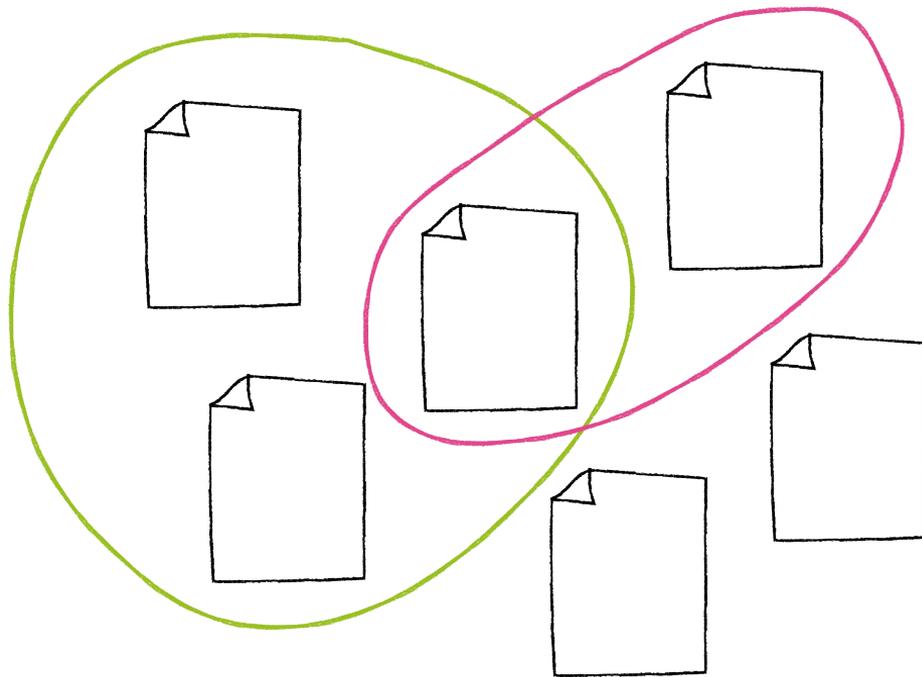
Les termes à maîtriser pour
bien utiliser kibana.

Aggregations

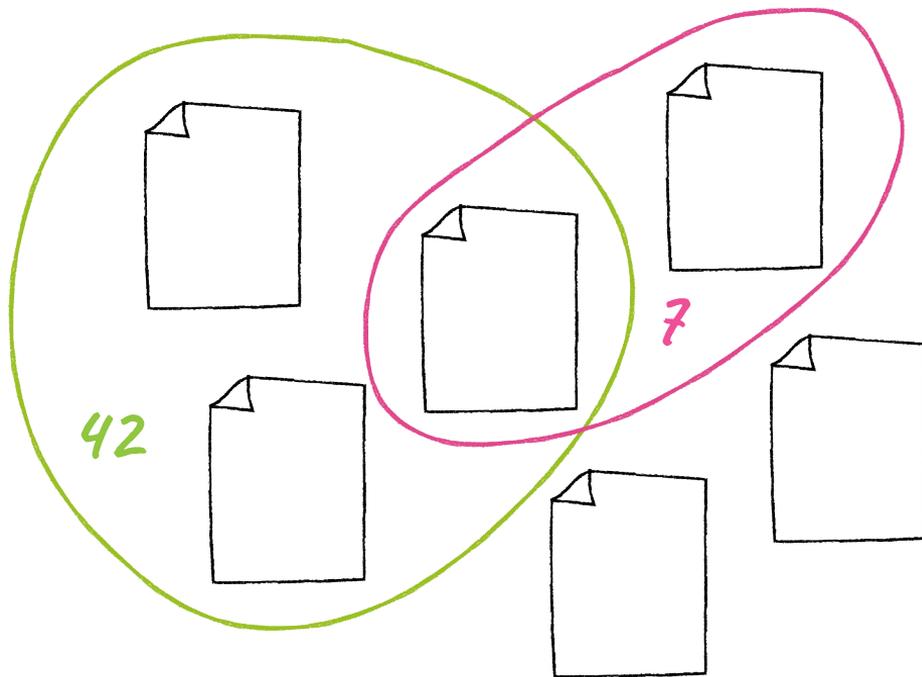
Buckets

Metrics

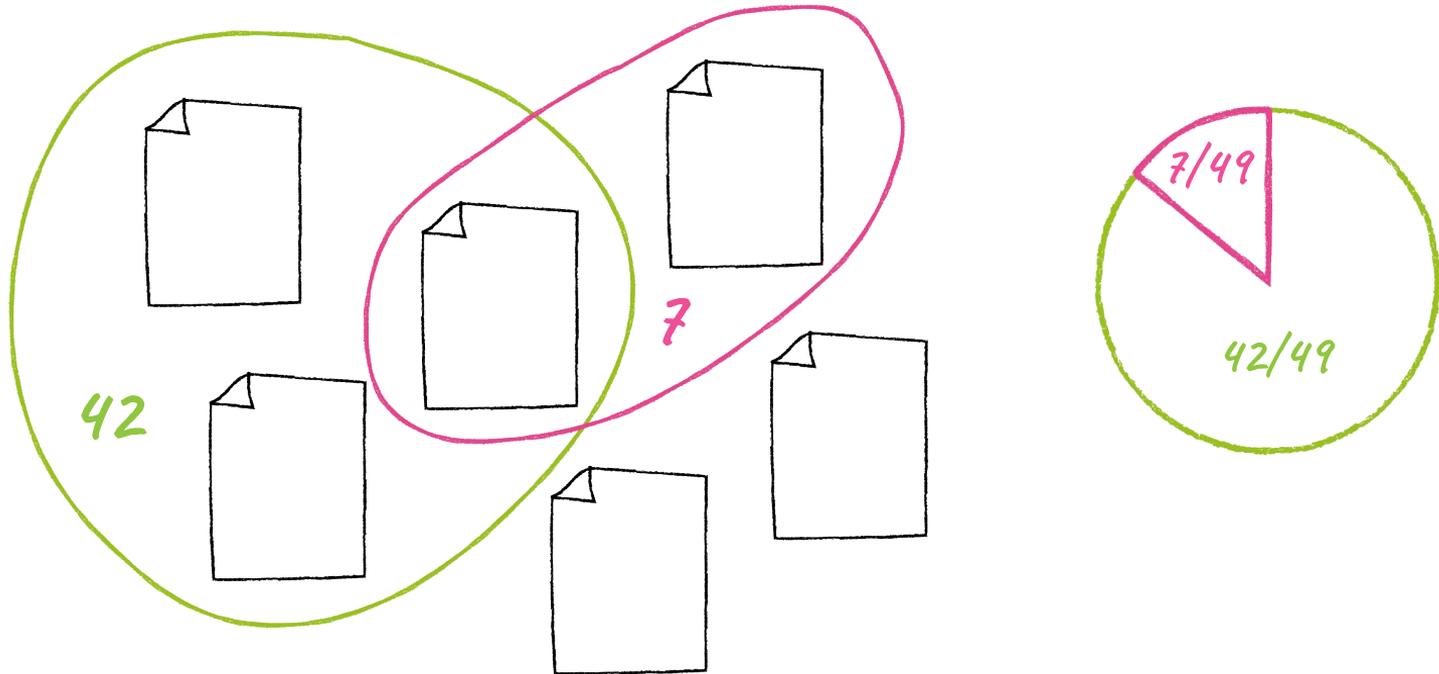
Bucket aggregation



Metrics aggregation



Application dans une visualisation



Types d'agrégation

Buckets

(principaux type)

- *Terms*
- *Histogram / Date histogram*
- *Range / Date range*
- *Filter*
- *Geo hash*

Metrics

(principaux type)

- *Count*
- *Sum*
- *Average*
- *Min / Max*
- *Unique count*

—
Des questions ?



C'est parti !

