

# COMMENT ANALYSER VOTRE E-RÉPUTATION SUR LE WEB?

(CAS DES CANDIDATS PRÉSIDENTIELLES 2017 EN FRANCE)

Présenté par : Ahmed KISMA

Le : 03/04/2017



## Mon Profil

**Nom** : Ahmed KISMA

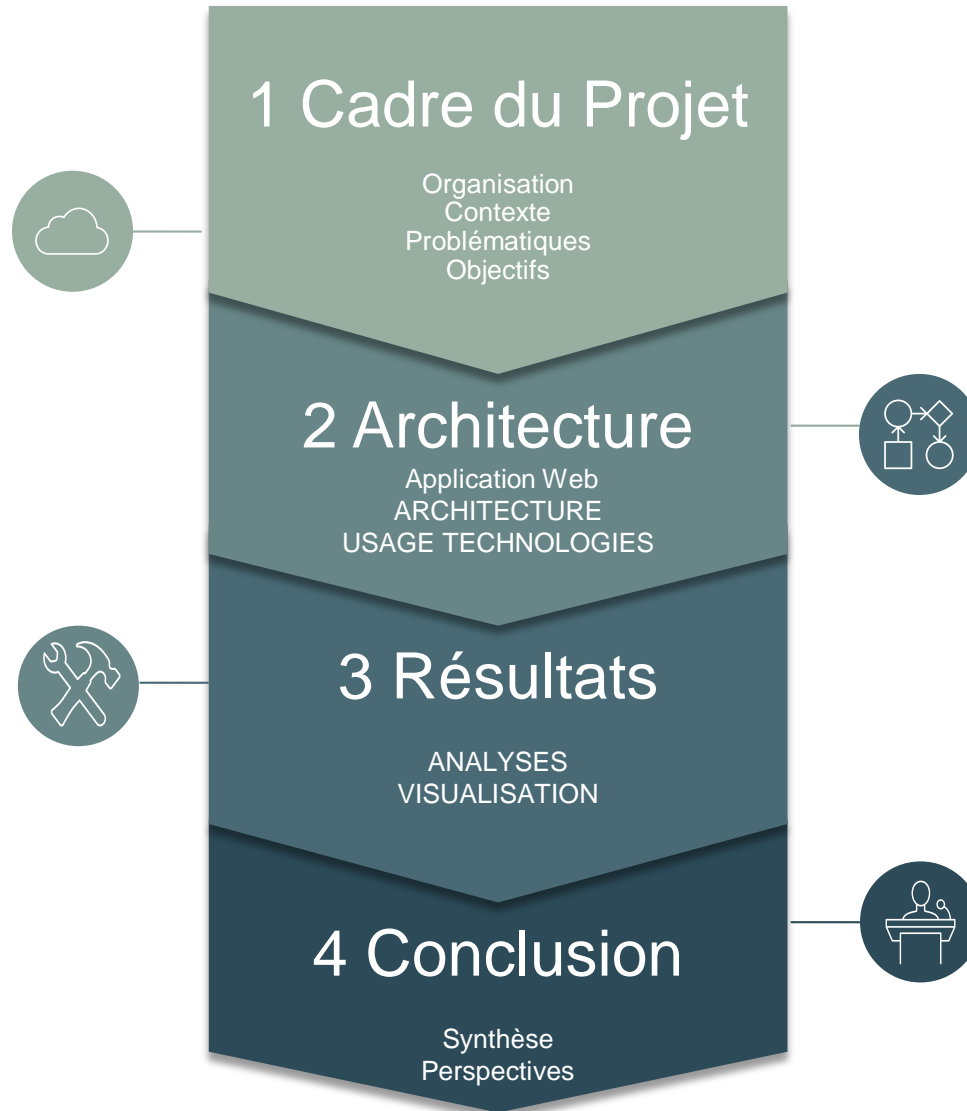
**Formation** : Ingénieur Système d'information et Analyse décisionnelle

**Expériences pro** : Développements et BI

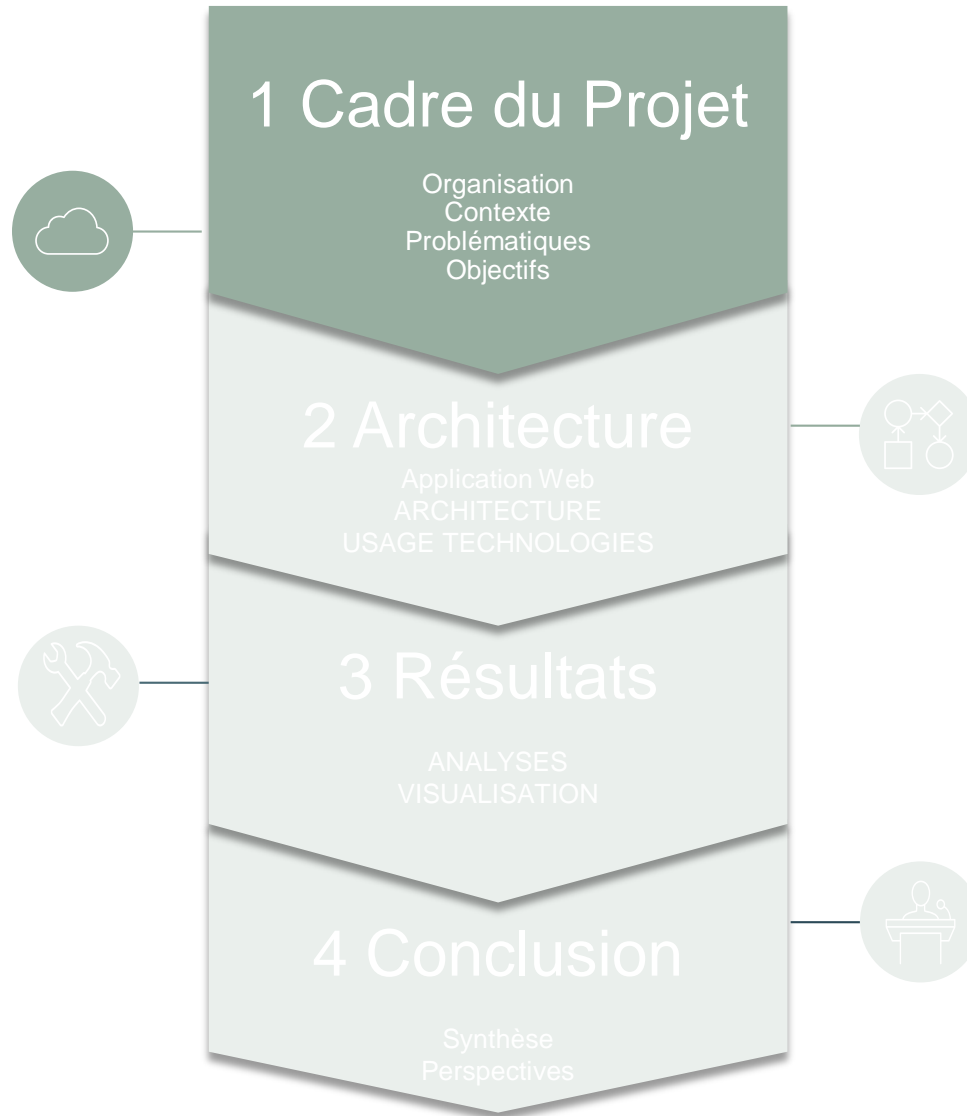
**Perspectives** : Data Analyste en BIG DATA



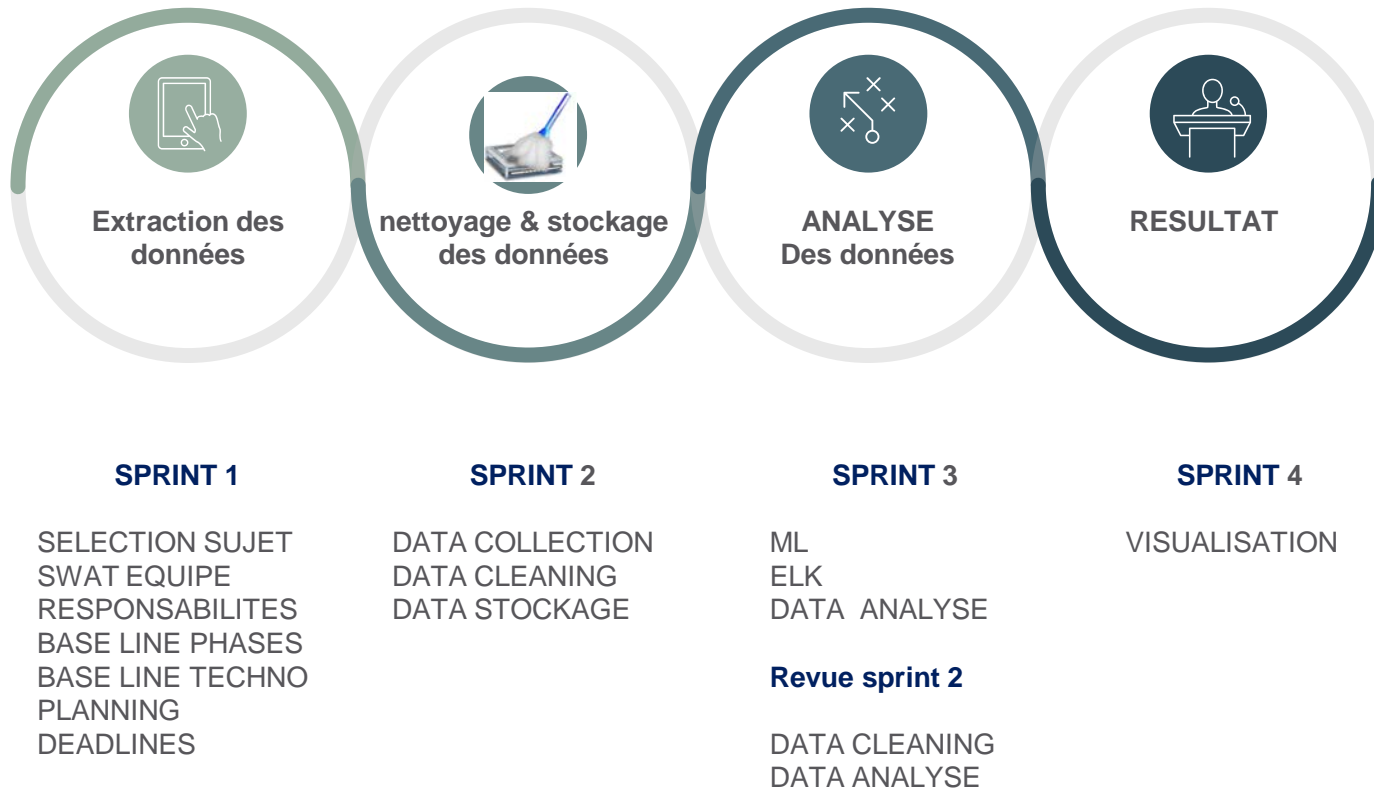
# Plan



# Plan



## Cadre du Projet : Organisation



\*Daily meeting: 15 min/jour



## Cadre du Projet : Contexte



## Cadre du Projet : Contexte (pourquoi?)

....

**59% des internautes passent plus de 12 heures par semaine sur internet, et 23% surfent sur leur smartphone.**

**WEB**  
**Internet**



....

**74% des internautes apprécient quand l'entreprise s'engage sur les réseaux sociaux.**



## Cadre du Projet : Contexte (Chiffres)



....

**1,8 milliard d'euros,  
c'est l'estimation du marché  
mondial de l'e-réputation**

....

**2,7 milliards de personnes  
peuvent voir ce qui est publié  
sur Internet.**





## Cadre du Projet : Contexte (chiffres)

»»»»

**83% des entreprises auront  
une crise d'e-réputation d'ici  
5 ans.**

»»»»

**40% des entreprises considèrent  
l'e-réputation comme  
le risque numéro 1.**

»»»»

**80% de ce qui se dit au sujet  
d'une marque sur Internet  
ne provient pas de la marque.**



## Cadre du Projet : Contexte (chiffres)

....

**41% des mécontentements s'expriment d'ailleurs à égalité sur Facebook et Twitter.**



....

**11% seulement s'expriment sur un blog et 7% sur Youtube.**



## Cadre du Projet : Contexte (l'importance)

La propagande virtuelle et la manipulation d'opinion sont les armes de guerre d'aujourd'hui

D'où l'importance :

- De **gérer** la vitalisation de l'entreprise
- De **contrôler et maîtriser** son identité virtuelle (action /réaction)
- De **savoir** ce que les internautes pensent et disent de l'entreprise



## Cadre du Projet : Contexte

### L'e-réputation des politiciens est un enjeu essentiel du marketing politique

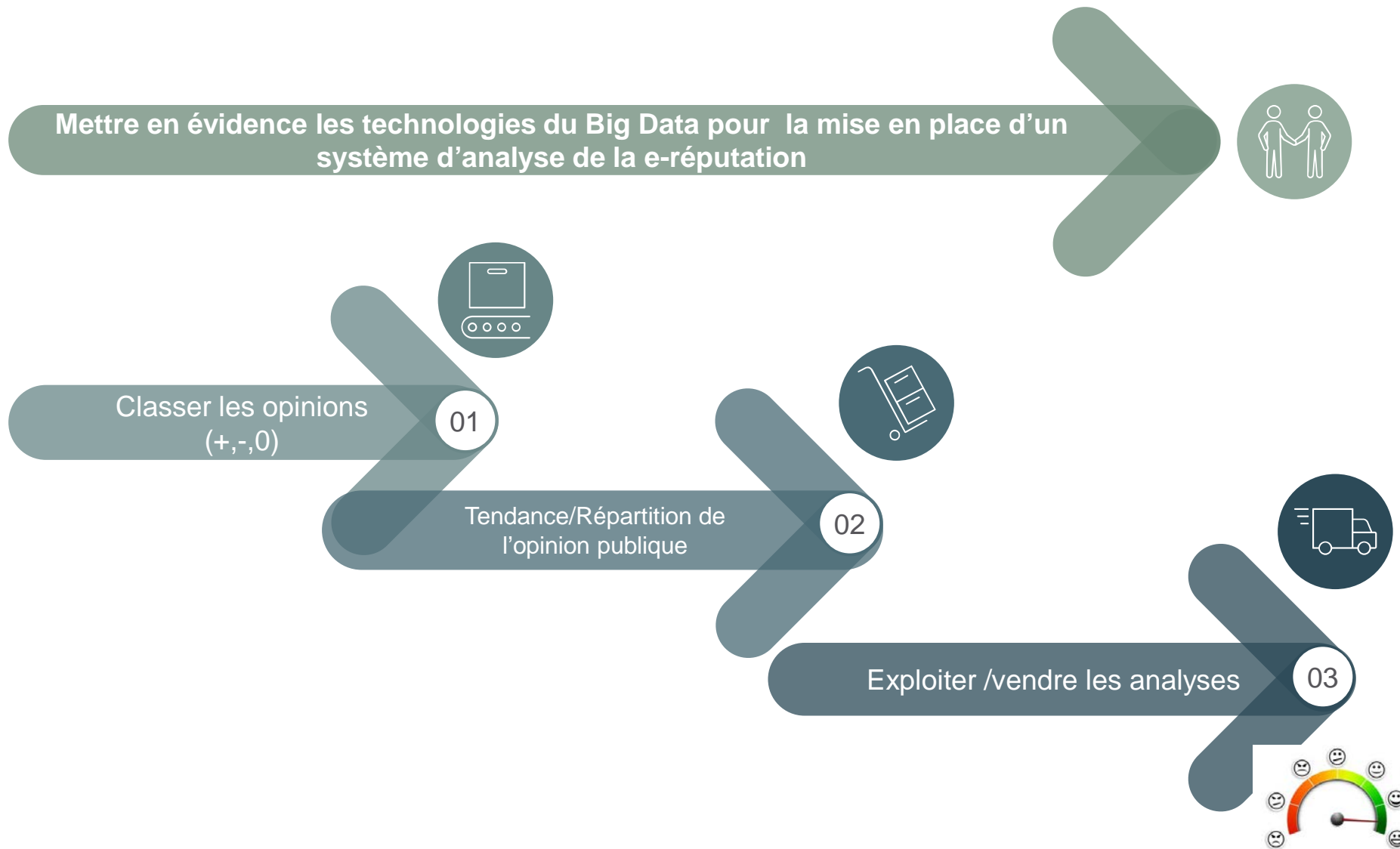
- ❖ Contrôle de leur image via les médias
- ❖ Adaptation de leur discours en fonction des cibles et des événements.



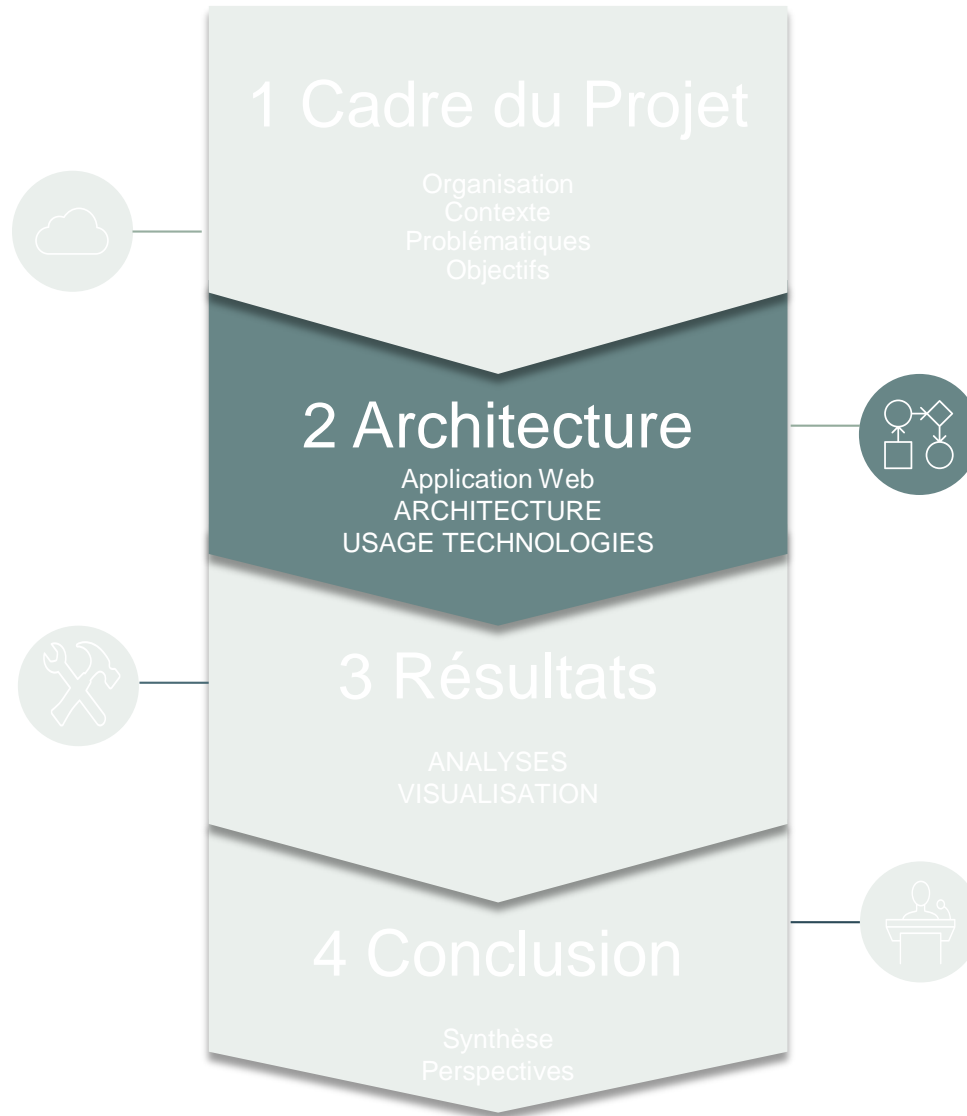
*La présence sur les réseaux sociaux, un enjeu stratégique pour les politiciens*



## Cadre du Projet : Objectifs



# Plan

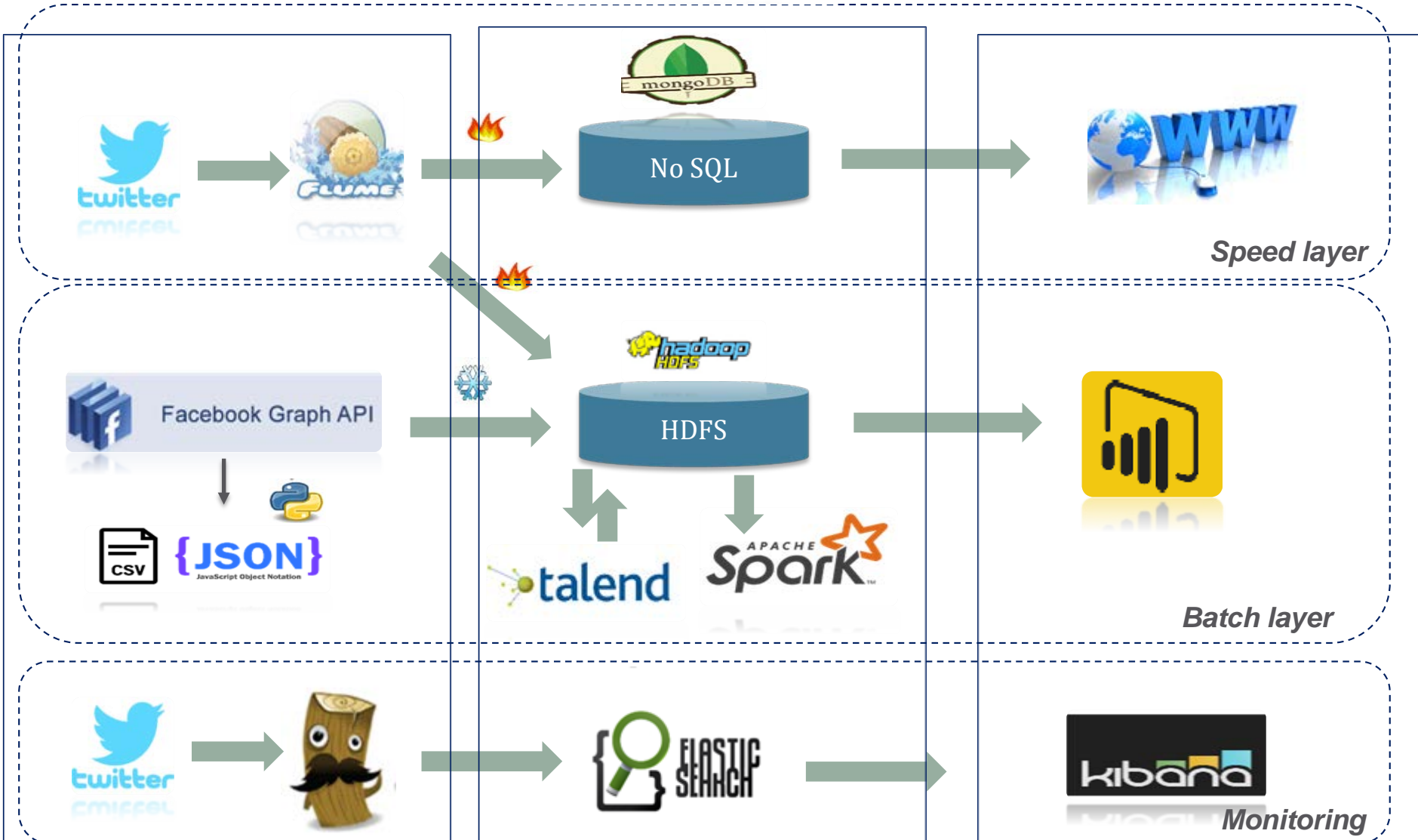


# Architecture Big Data

## DATA COLLECTION

## STOCKAGE

## VISUALISATION



## 2. ARCHITECTURE - COLLECTE DE DONNEES

### Streaming



**{JSON}**  
JavaScript Object Notation

Limitation 2400 tweet  
par jour (30 MB de texte)



**{JSON}**  
JavaScript Object Notation

Limitation 2400 tweet  
par jour (30 MB de texte)

### Scraping



**Tweepy**

**{JSON}**  
JavaScript Object Notation

5 ans, 70 MB  
de texte



**URLLIB2  
& BS4**



6 ans, 100 MB  
de texte





### ➤ Réseaux Sociaux

twitter

1 660 000 tweets



1 280 000 commentaires

### ➤ Data Scraping (size) / TEXT



≈ **1Giga**



## Nettoyage et structuration des données



### - SUPPRIMER

les espaces  
les doublons.  
les lignes vides  
les lignes dupliquées



### - Corriger les erreurs

### - Gérer les virgules

*Des données propres et fiables garantissent de meilleures décisions*



## Remplacer les émoticônes avec les mots

*Extrait d'un fichier CSV scrapé depuis Facebook avant Nettoyage*



Remplacer

Conten



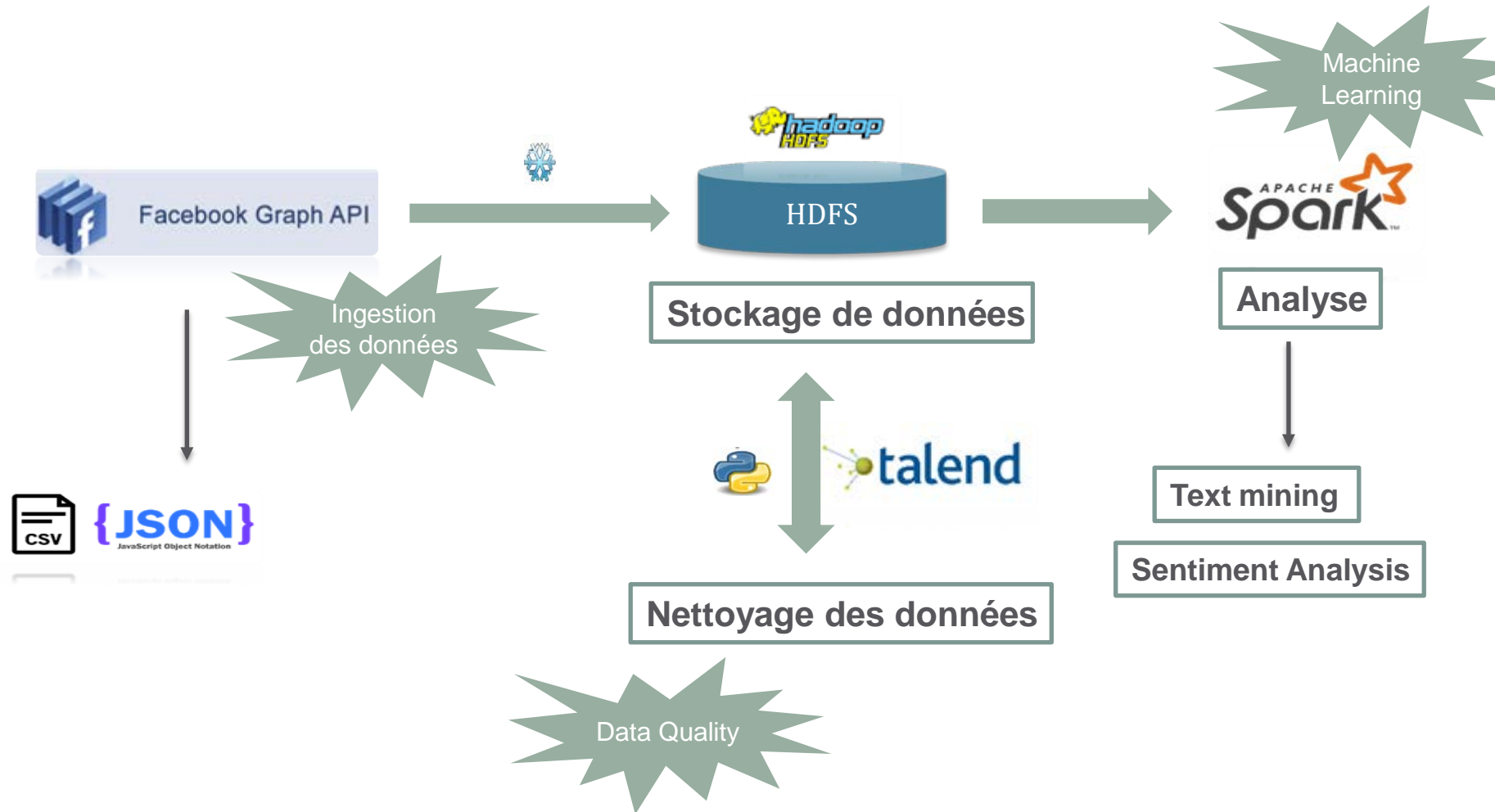
Remplacer

Amour

2264,,	AIDONS LES TPE ET PME IL Y AURA PLUS DE BOULOUT,	Gilbert An
2264,,	"avec ds abeilles, une ruche",	Jacques Faelchlin,2017-03-03 11:10:
2264,,	"Hamon écrit son texte tout seul, lui au moins, contrairement au	
2264,10155083893312264_1933771066909380,	JLM ne lit même pas s	
2264,,	Je vote Hamon ♥♥♥♥,	Mari Lyn,2017-03-03 11:10:20,7,0,0
2264,,	Votez Hamon,	Fayçal Zouaoui,2017-03-03 11:10:56,4,0,0
2264,,	Bon courage,	Rahim Lagarois,2017-03-03 11:10:58,0,70,70
2264,,	Allez tous Avec B.Hamon,	Maherzia Champiré,2017-03-03 11:11:0
2264,,	Au top bon programme 😊😊😊😊😊😊😊😊,	Annie Fleury Ps,2017-
2264,,	Bravo et merci !,	Stephane Tirtiaux,2017-03-03 11:11:10,4,30,20
2264,,	On va gagner !,	Mari Lyn,2017-03-03 11:11:12,5,0,0
2264,,	En zone rurale c'est important les espaces de coworking. Ça pern	
2264..	Hamon est bien..	Gustave Parking.2017-03-03 11:11:14.0.0.20



## Architecture Big Data : Batch Layer



## ► Framework Spark:

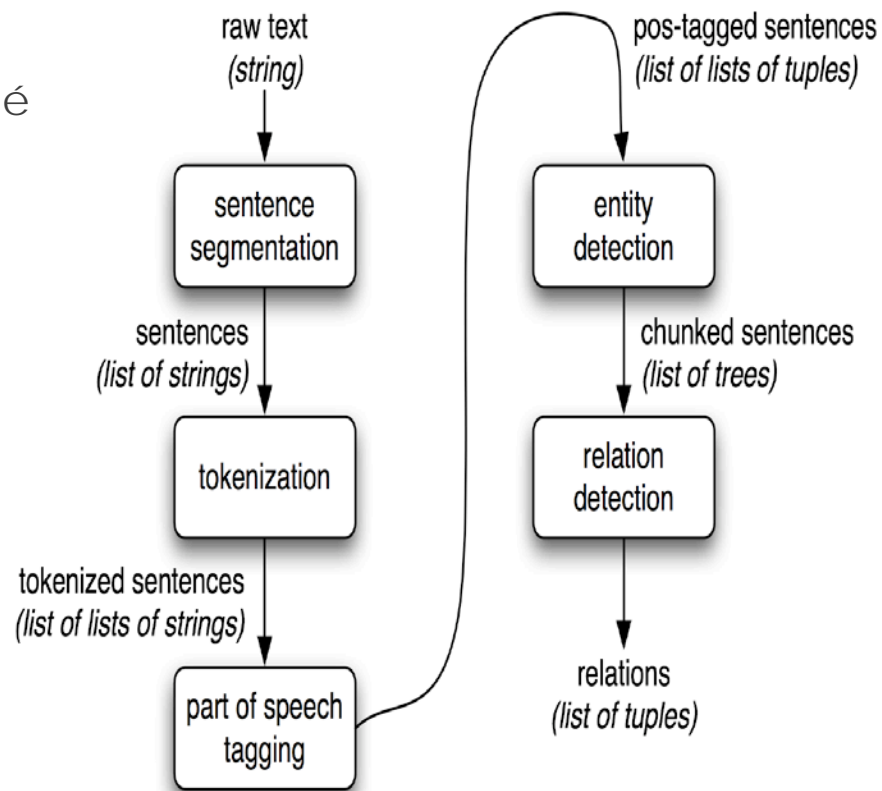
- ❖ Framework dédié au calcul distribué
- ❖ Réalisation de calculs sur des RDD

## ► NLTK

- ❖ Librairie pour manipuler du texte

## ► Objectifs:

- i. Recherche des termes les plus cités : Thématiques/mots clé
- i. Avoir en résultat en nuage des mots

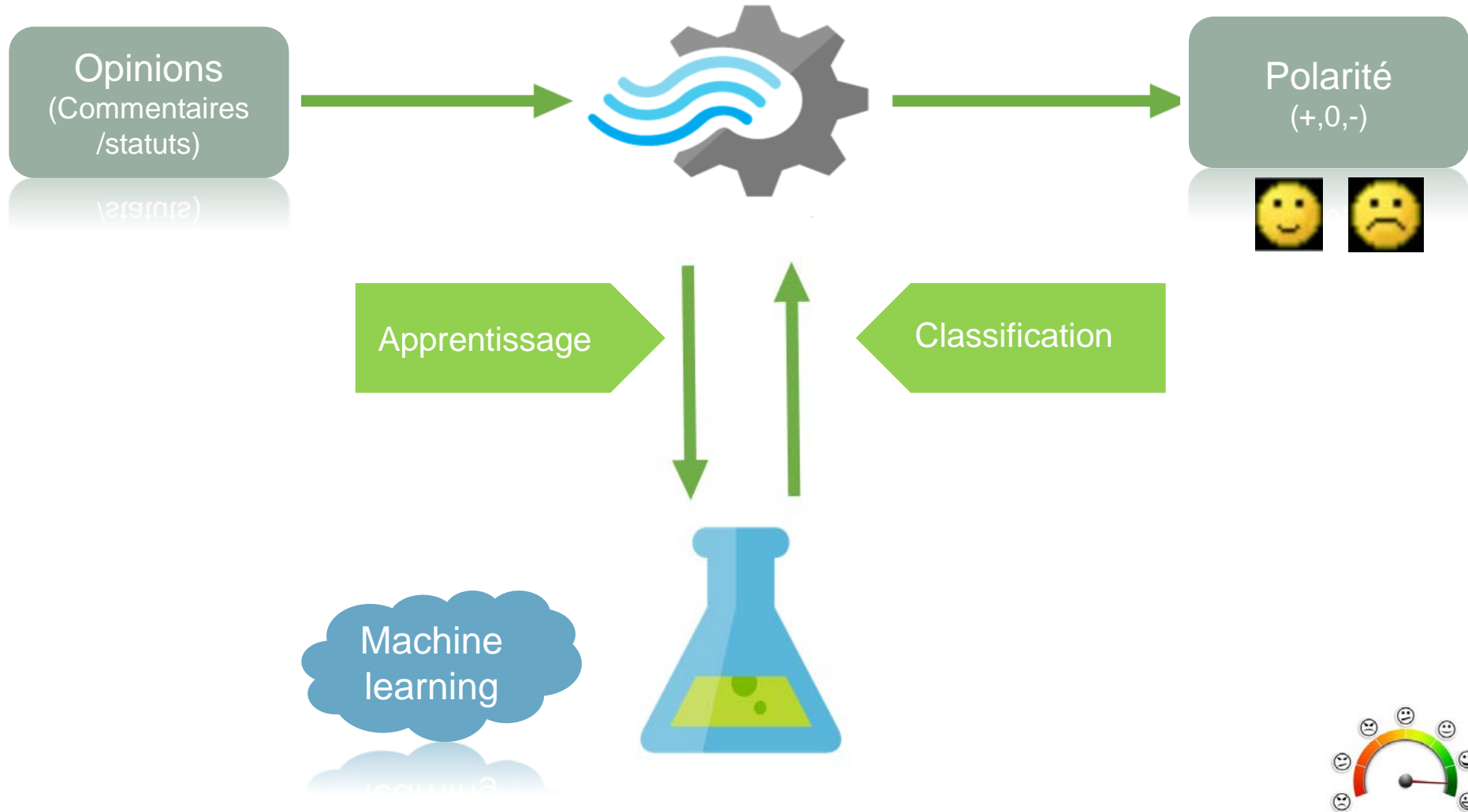


## Résultats

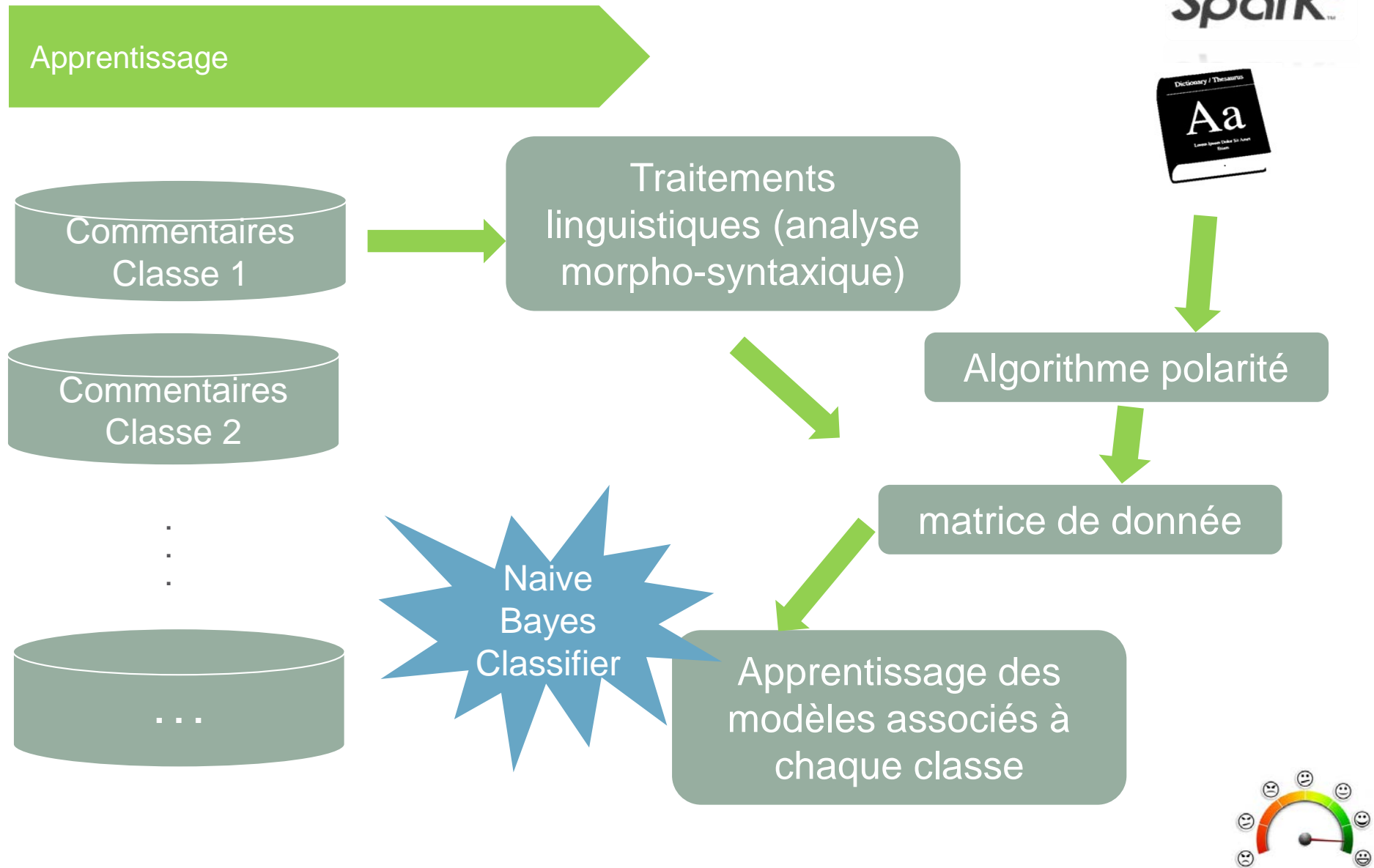
- Les thèmes les plus cités  
(nuage de mots + courbe de distributions)

Nuage de mots pour les commentaires de François Fillon sur Facebook (2016)



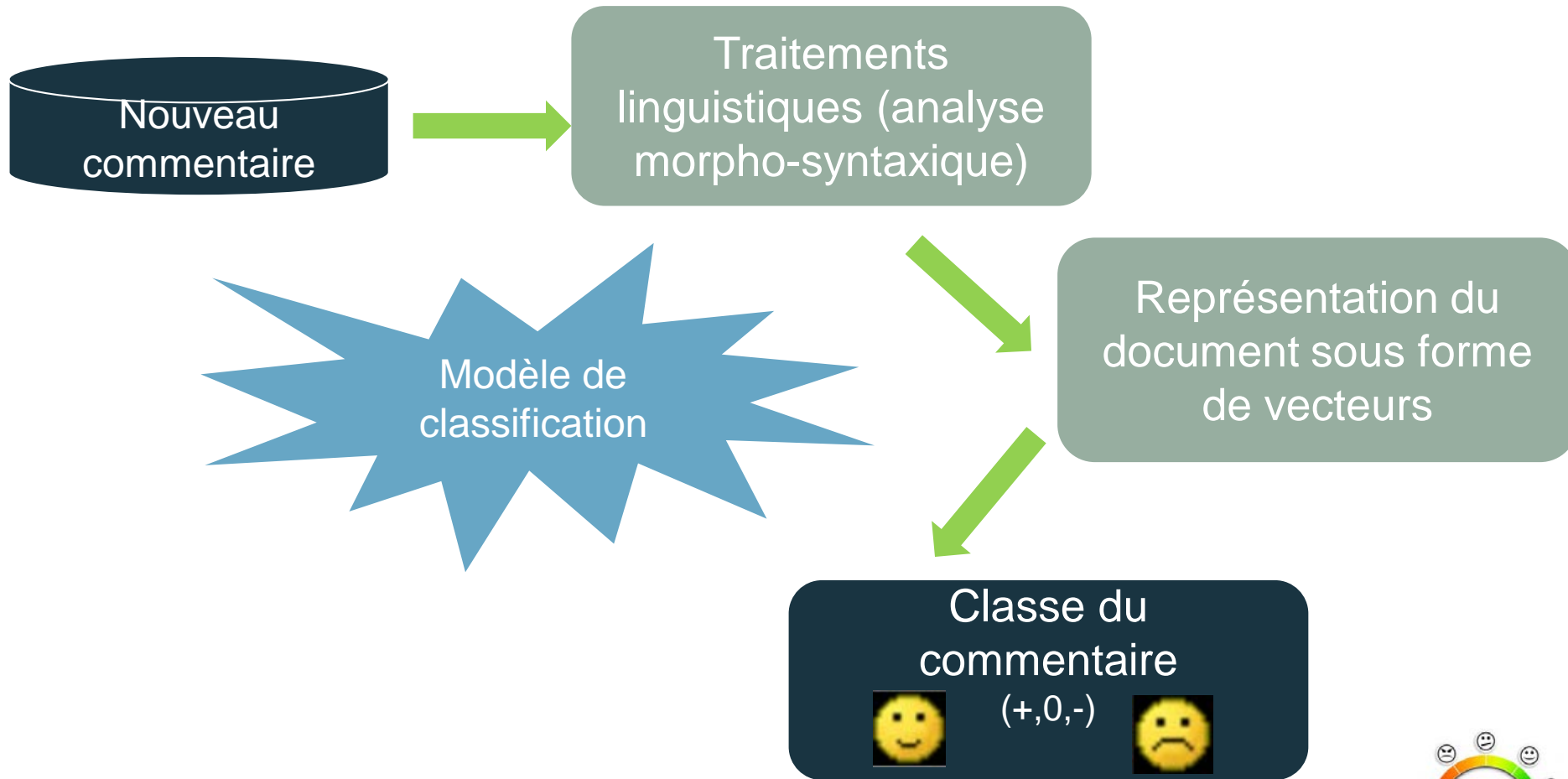


## Analyse des données : Analyse des sentiments

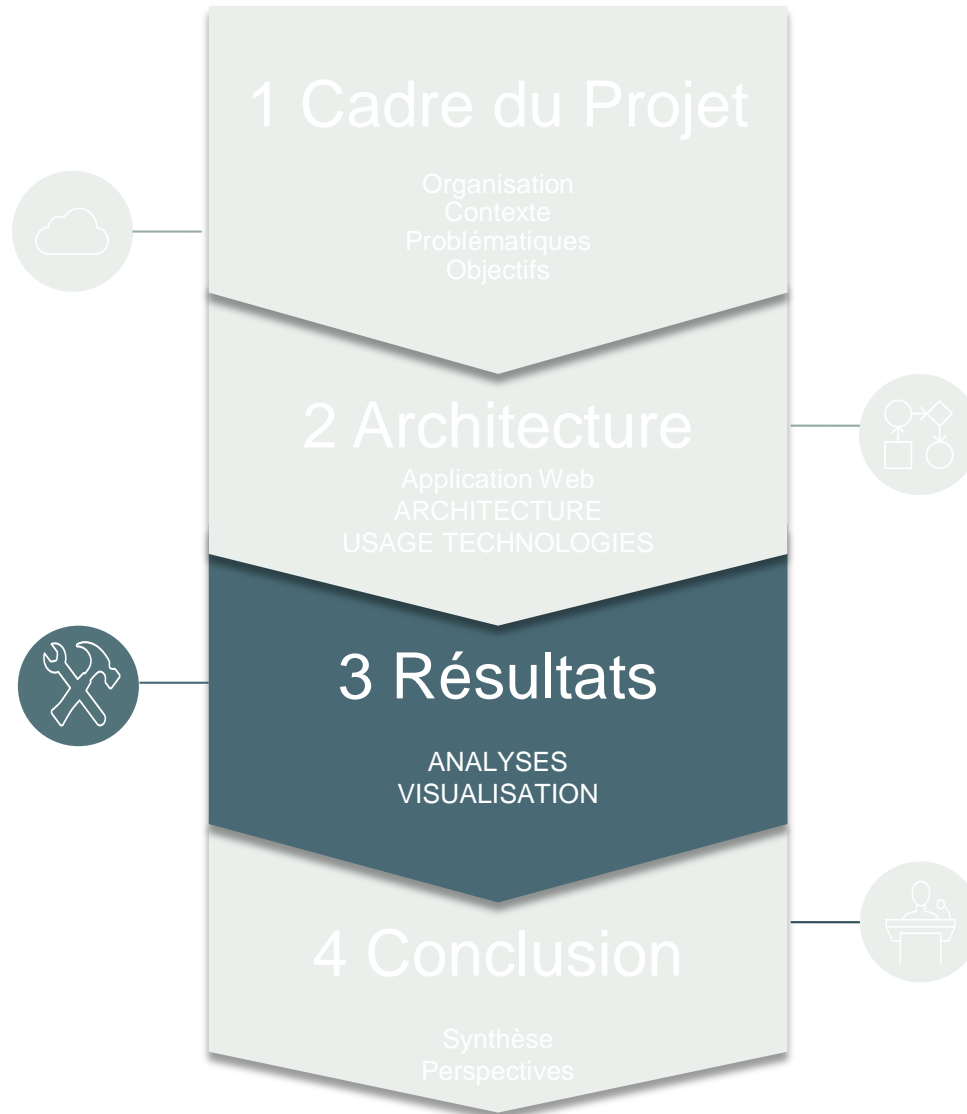




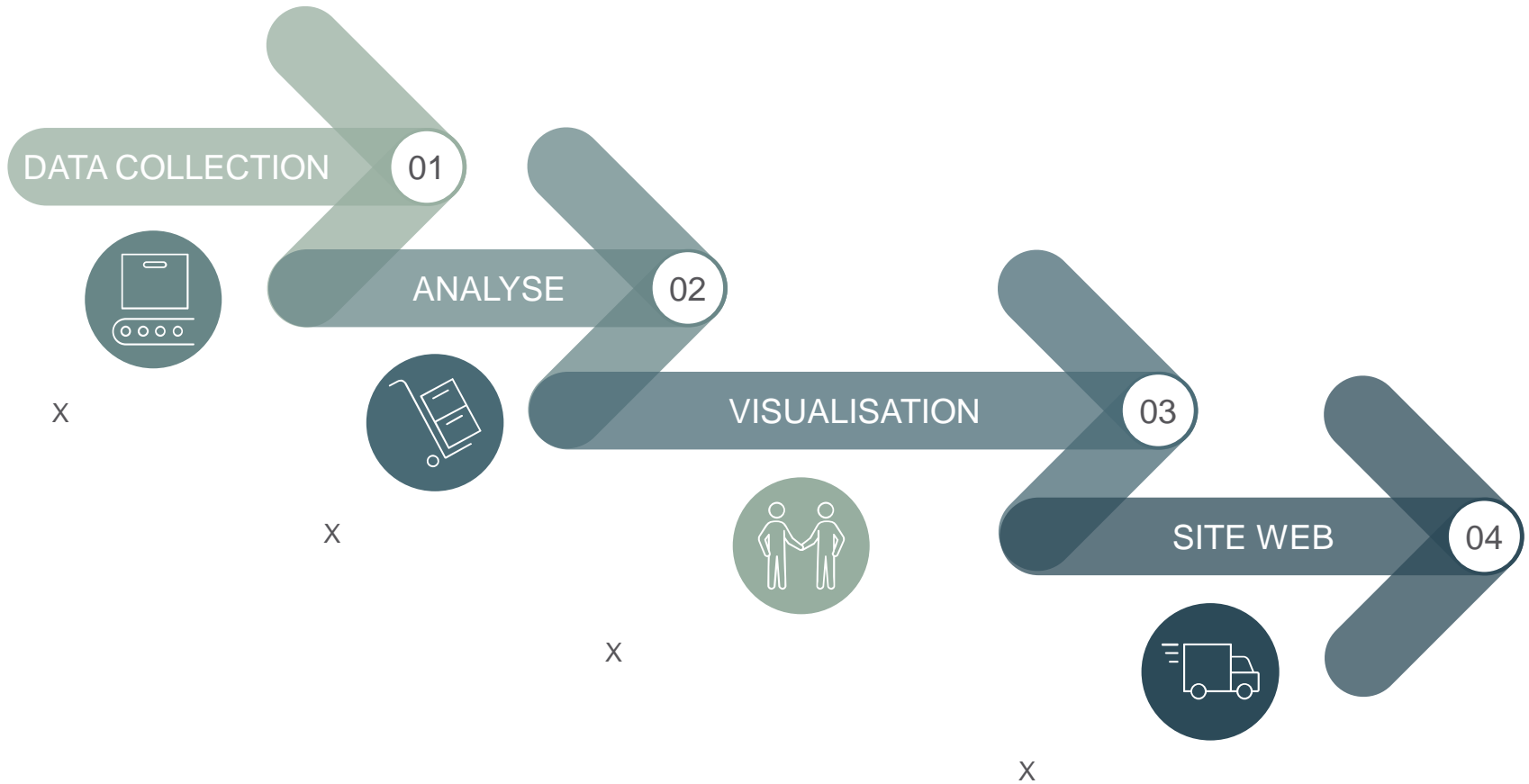
Classification



# Plan

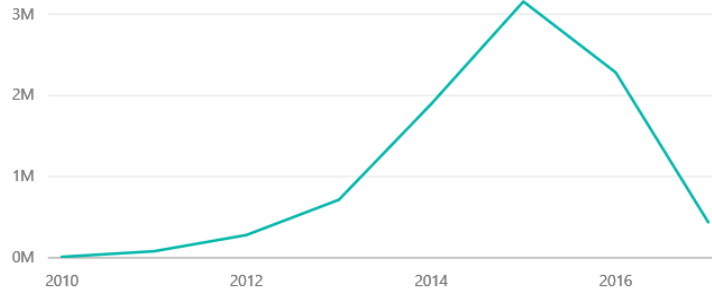


### 3. RESULTATS



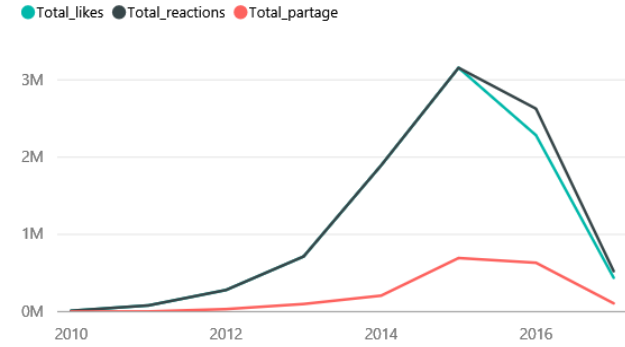
# Résultats : Nbre like & émotions

Total\_likes par Année



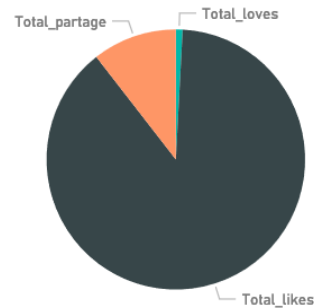
Nombre de Like Marine Le Pen

Total\_likes, Total\_reactions et Total\_partage par Année



Total de réaction selon les like et les partages des

Total\_loves, Total\_likes, Total\_wows, Total\_sads, Total\_hahas, Total\_a...

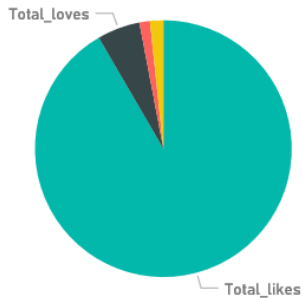


Marine LE PEN



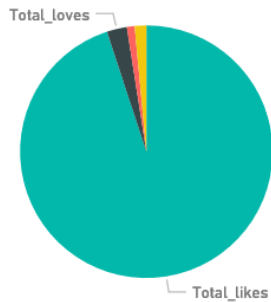
# Résultats : nbre like & Emotions

Total\_likes, Total\_loves, Total\_hahas, Total\_angrys, Total\_sads et ...



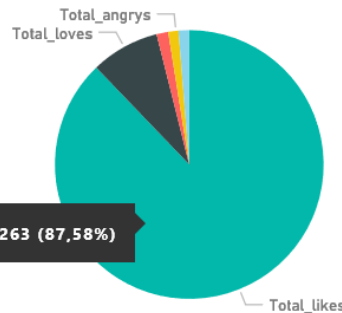
François Fillon

Total\_likes, Total\_loves, Total\_hahas, Total\_angrys, Tot...



Mélenchon

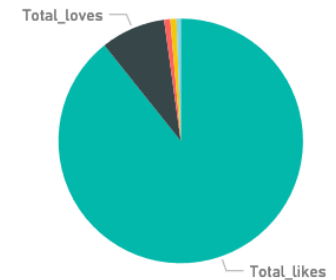
Total\_likes, Total\_loves, Total\_hahas, Total\_angrys, Total\_sads et...



Emmanuel Macron

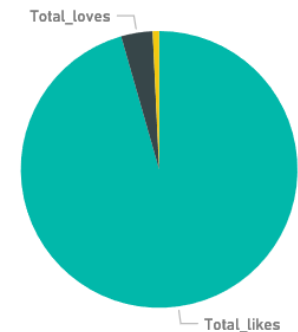


Total\_likes, Total\_loves, Total\_hahas, Total\_angrys...



Benoit Hamon

Total\_likes, Total\_loves, Total\_hahas, Total\_angrys, Total\_s...



Marine LE PEN



# Résultats : Réaction en émotions sur facebook

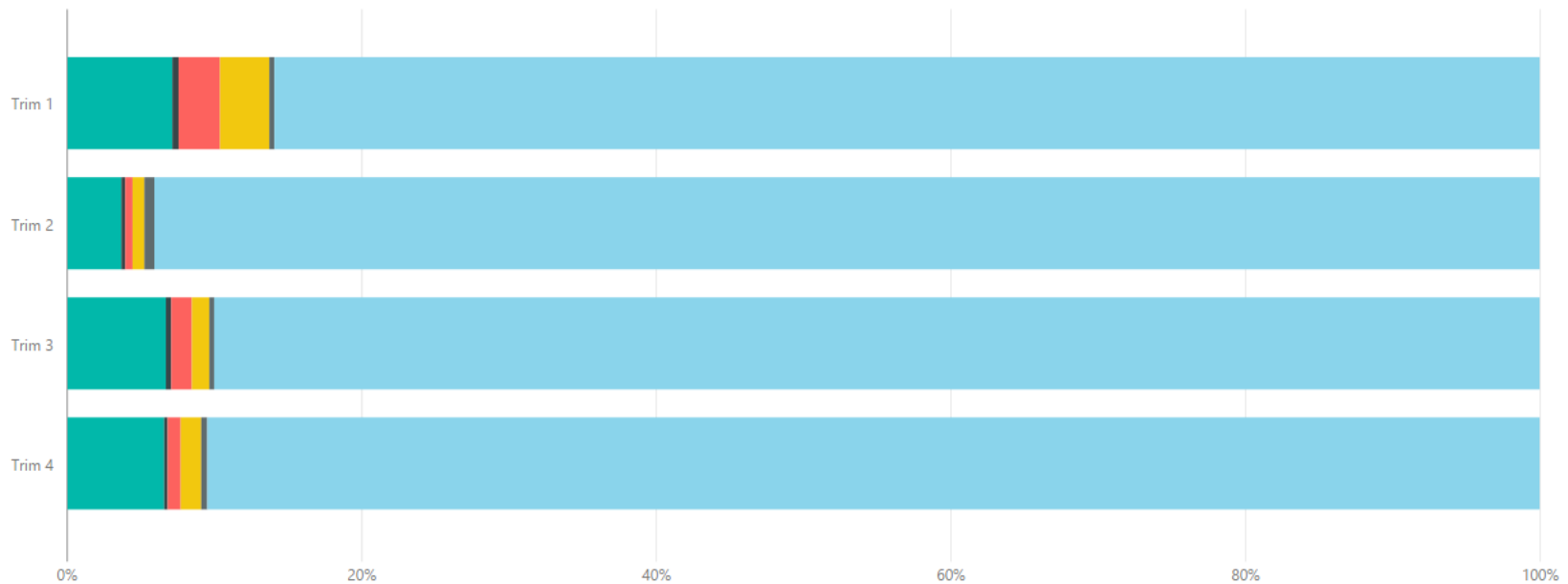
## Les réactions menseulle des interneutes sur les statuts postés sur Facebook Par Fillon

257260

Nombre de commentaires

Total\_loves, Total\_wows, Total\_hahas, Total\_angrys, Total\_sads et Total\_likes par Trimestre

● Total\_loves ● Total\_wows ● Total\_hahas ● Total\_angrys ● Total\_sads ● Total\_likes



# Résultats : Réaction en émotions sur facebook

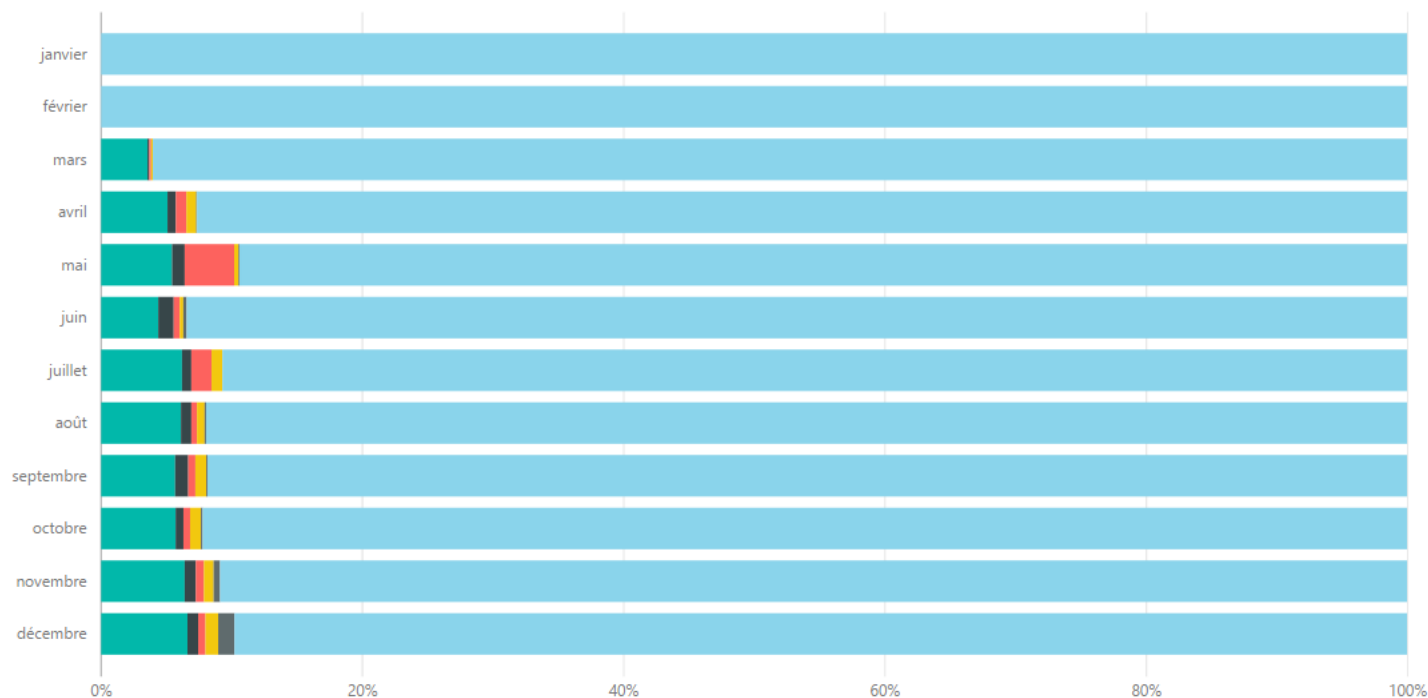
## Les réactions menseulle des interneutes sur les statuts posté sur Facebook Par Emmanuel Macron

Total\_loves, Total\_hahas, Total\_angrys, Total\_wows, Total\_sads et Total\_likes par Mois

18608

Nombre Commentaires

● Total\_loves ● Total\_hahas ● Total\_angrys ● Total\_wows ● Total\_sads ● Total\_likes



année

□ 2014

□ 2015

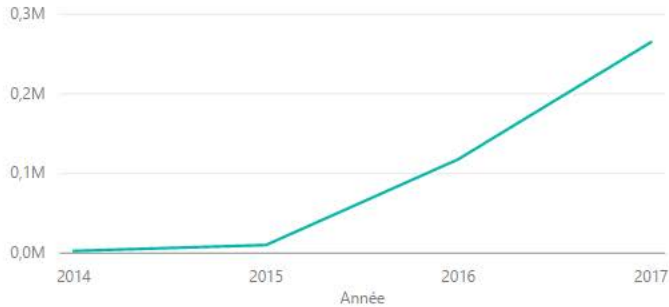
■ 2016

□ 2017



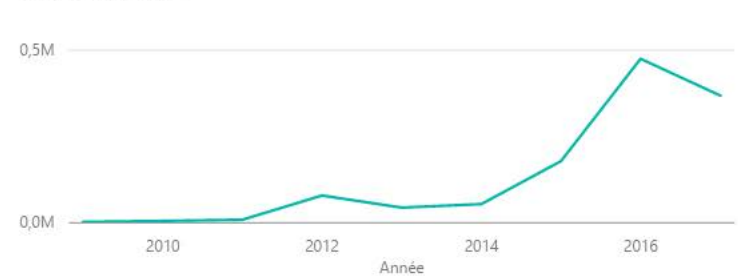
# Résultats : Nombre de Like sur Facebook

Total\_likes par Année



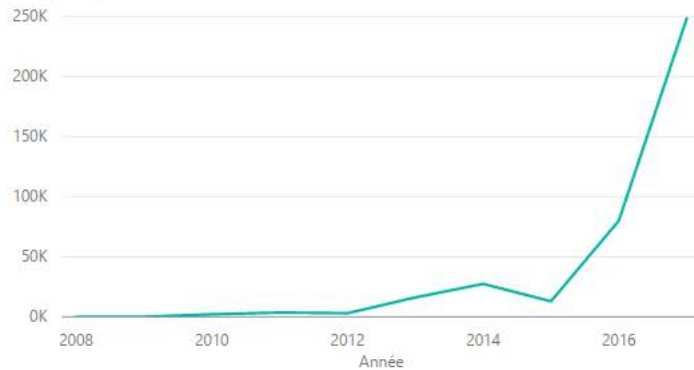
Nombre de Like pour Emmanuel Macron

Total\_likes par Année



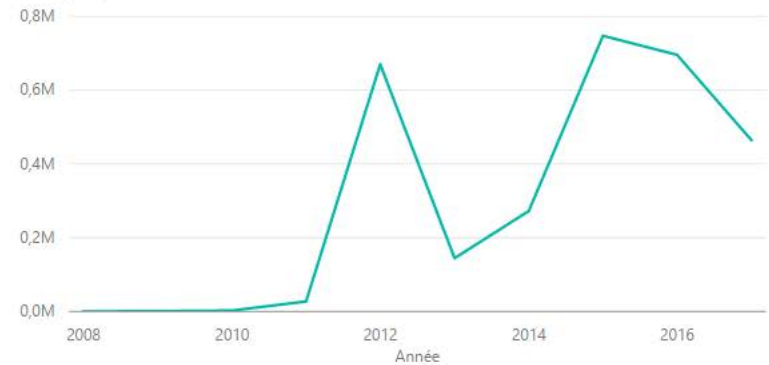
Nombre de Like François Fillon

Total\_likes par Année



Nombre de Like pour Benoit

Total\_likes par Année



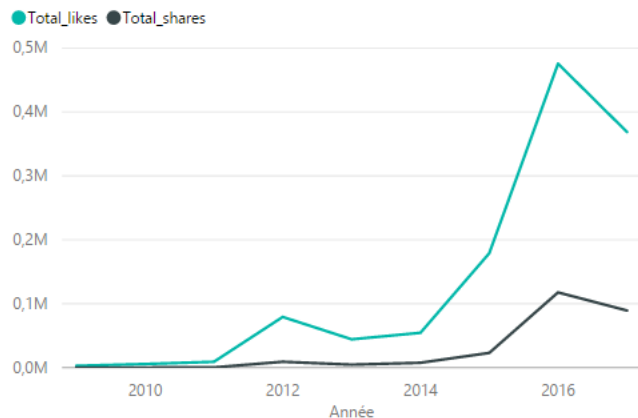
Nombre de Like Mélençon



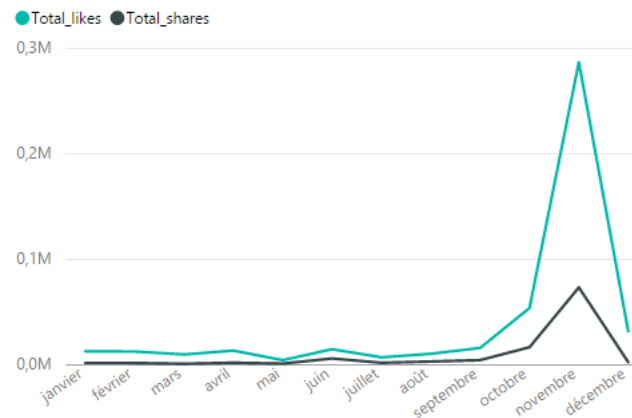


# Résultats : Nombre de partage & like sur Facebook

Total\_likes et Total\_shares par Année



Total\_likes et Total\_shares par Mois

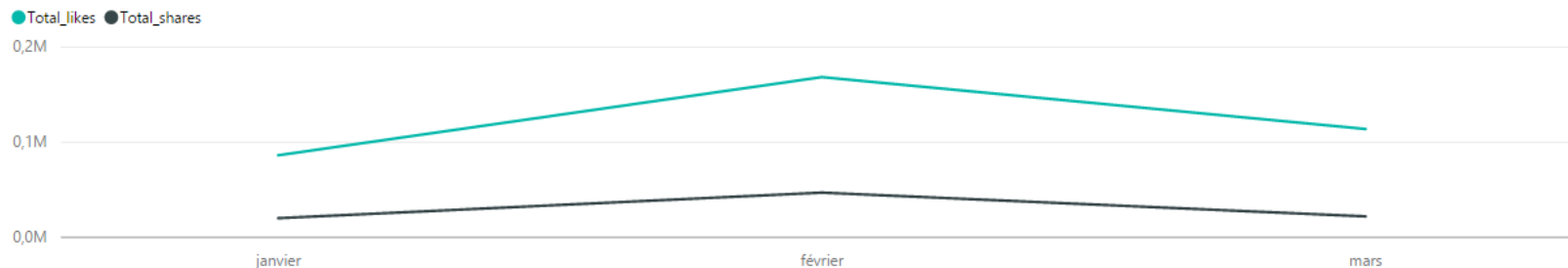


Total des Like et partages de statuts de Fillon par année

Total des Like et partages de statuts de Fillon en 2016

François Fillon

Total\_likes et Total\_shares par Mois



Total des Like et partages de statuts de Fillon par mois en 2017



## Résultats : Mesure de E-réputation (polarité)

NEUTRE

NEGATIF

POSITIF



Mesure de l'e-réputation



# Résultats : Qualification d'un commentaire

## Mesure de l'e-réputation par commentaire pour Hamon

Nom de l'auteur

Profil

Adrien Bruno Broult

☐ Adrien Bar

☐ Adrien Barton Carton

☐ Adrien Beauquier

☐ Adrien Boedec

☐ Adrien Boissiere

☐ Adrien Boudet

☐ Adrien Brière

☒ Adrien Bruno Broult

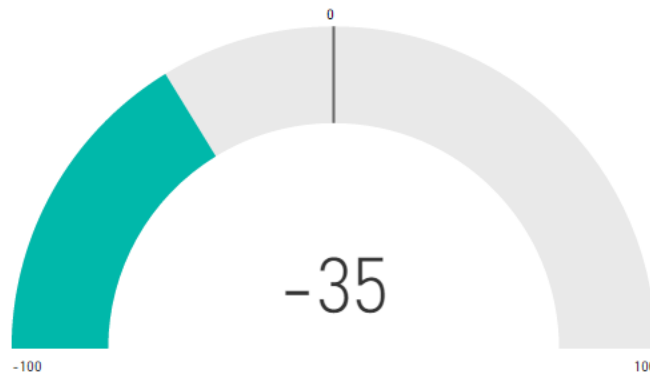
☐ Adrien Bureau

### Commentaire

Message

C'est donc vous dont on parle sur vie de carabin, qui avez demandé de faire nettoyer un hôpital et cacher des malades pour vos vœux? :p

polarity, min, max et origine



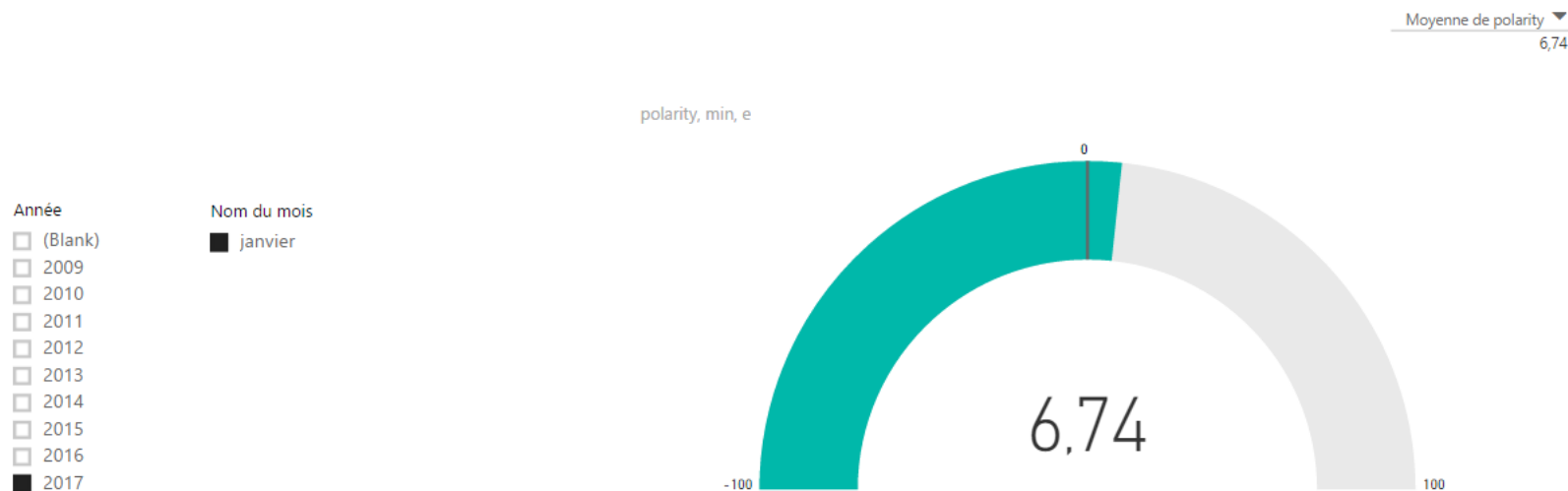
**NEGATIF**

Premier Reputation



# Résultats : E-réputation pour Hamon par mois

Vision d'ensemble sur la réputation de Benoît Hamon sur le Facebook



Polarité (Positif, Négatif, Neutre)

La Sélection de la période

Date de Publication

11/02/2009 23/01/2017



# VISUALISATION DASHBOARD KIBANA



Données sur 23 % de personnes qui partagent leurs localisation sur twitter



# Résultats : Publication sur le WEB



➤ Statistiques

➤ Résultats e-réputation



Nouvel utilisateur

Utilisateur existant

ACCÉDEZ GRATUITEMENT  
À VOTRE TABLEAU DE BORD E-RÉPUTATION

☐ Entreprise ☒ Individu

 Prénom \*

 Nom \*

 Adresse E-mail \*

 Mot de passe \*

 Téléphone

**S'INSCRIRE GRATUITEMENT**

En vous inscrivant, vous approuvez les [Conditions Générales d'Utilisation](#).

\* champ obligatoire

“ Maîtriser notre e-réputation,  
c'est améliorer notre image de marque. ”



Découvrez l'image  
que vous donnez sur internet :

☐ ENTREPRISE ☒ INDIVIDU

Prénom

Nom

**MESUREZ VOTRE E-RÉPUTATION**

[Politique de confidentialité](#)

François Fillon **FR**

IDENTIFICATION DE VOS RÉSEAUX SOCIAUX

IMAGE	RÉSEAU SOCIAL	IDENTIFICATION
	<b>youtube</b> François Fillon - YouTube <a href="https://www.youtube.com/channel/UCp1R4BFjKw34PFUc3GDLkw">https://www.youtube.com/channel/UCp1R4BFjKw34PFUc3GDLkw</a>	Non modifiable
	<b>Twitter</b> François Fillon (@FrancoisFillon)   Twitter <a href="https://twitter.com/FrancoisFillon">https://twitter.com/FrancoisFillon</a>	Twitter
Terminer l'identification des réseaux sociaux		



# Statut de l'analyse de e-réputation / USER



En Attente de Validation



En cours d'analyse



MESUREZ



SURVEILLEZ



AMÉLIOREZ





# Tableau de bord e-réputation / USER



Bienvenue François Fillon,

Construisons ensemble votre tableau de bord e-réputation

☒ Analyser votre e-réputation personnelle : (Ex: Steve Jobs)

François Fillon

FR

☐ Analyser l'e-réputation de votre entreprise : (Ex: Apple)

Nom de l'entreprise

FR

+ Ajouter un autre mot-clé (Ex: Iphone, Ipad, Samsung, ...)

COMMENCER



**Bonjour FILLON, François**  
**Tableau de bord pour Lundi 27/03/2017**

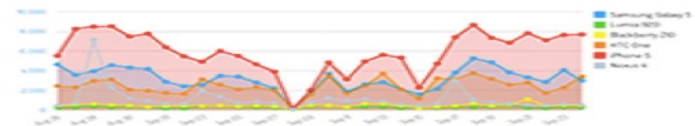
Gmail Images

## OVERVIEW

between Sep 25, 2013 and Sep 24, 2013

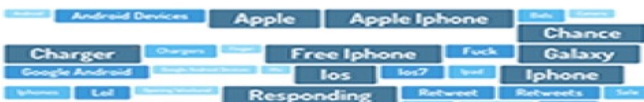
	VOLUME	TREND	SENTIMENT	RANKING	MENTIONS PER RANK		
					HIGH	MEDIUM	LOW
IPHONE 5	176585	↑ 10%	Positive	3	High	Medium	Low
SAMSUNG GALAXY S4	94893	↓ 15%	Neutral	4	Medium	High	Low
HTC ONE	72005	↑ 10%	Positive	5	Medium	Medium	High
NEXUS 4	34306	↑ 10%	Positive	6	Medium	Medium	High
BLACKBERRY Z10	14676	↓ 10%	Neutral	7	Medium	Medium	High
LUMIA 920	9408	↓ 10%	Neutral	8	Medium	Medium	High

## TREND



## KEY CONCEPTS

about iPhone 5 between Sep 25, 2013 and Sep 26, 2013



## SENTIMENT

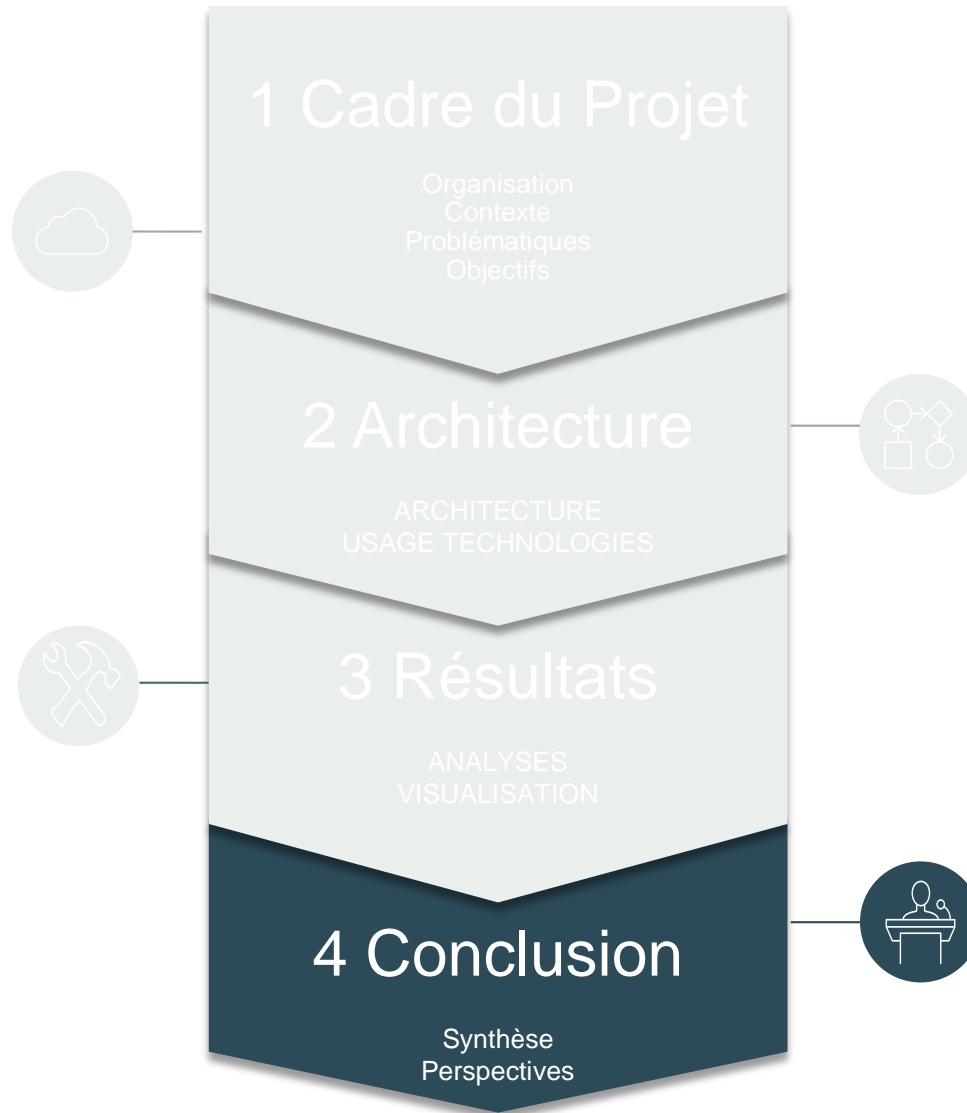
about iPhone 5 between Sep 25, 2013 and Sep 26, 2013



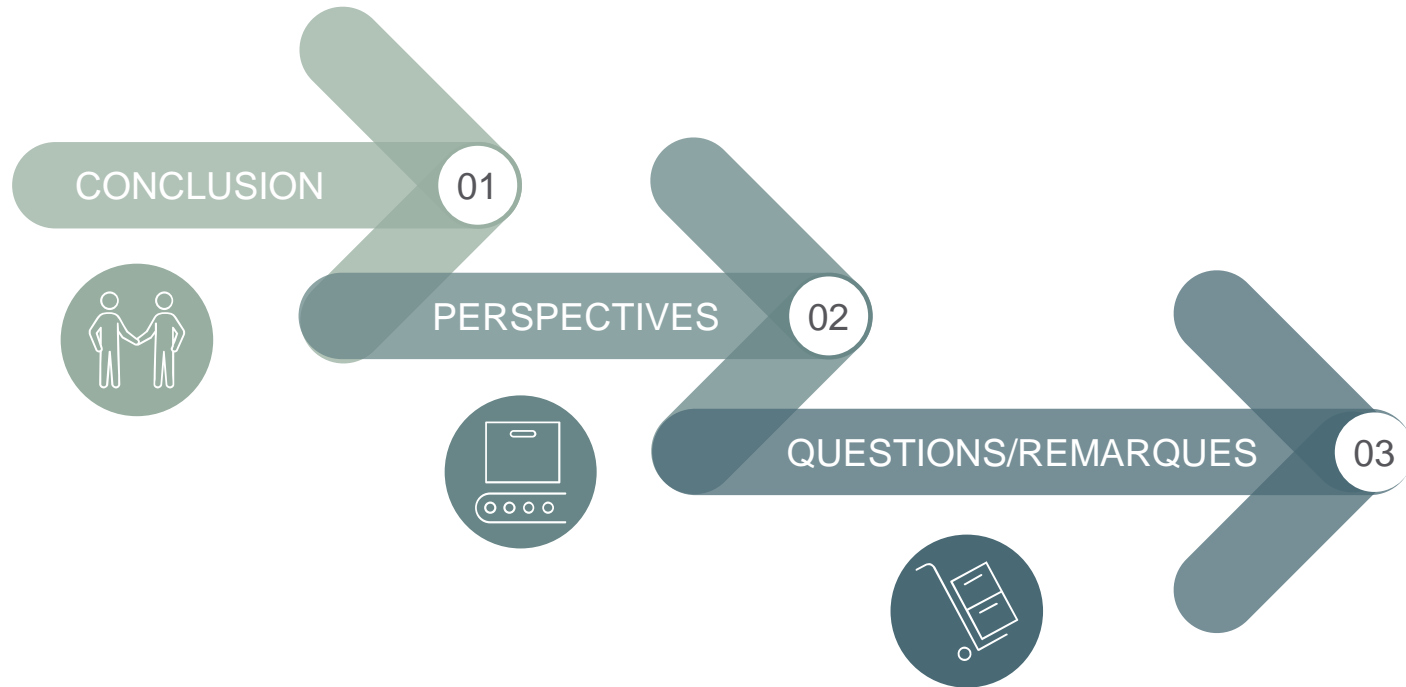
Note: Mentions without sentiment are not shown in this analysis.



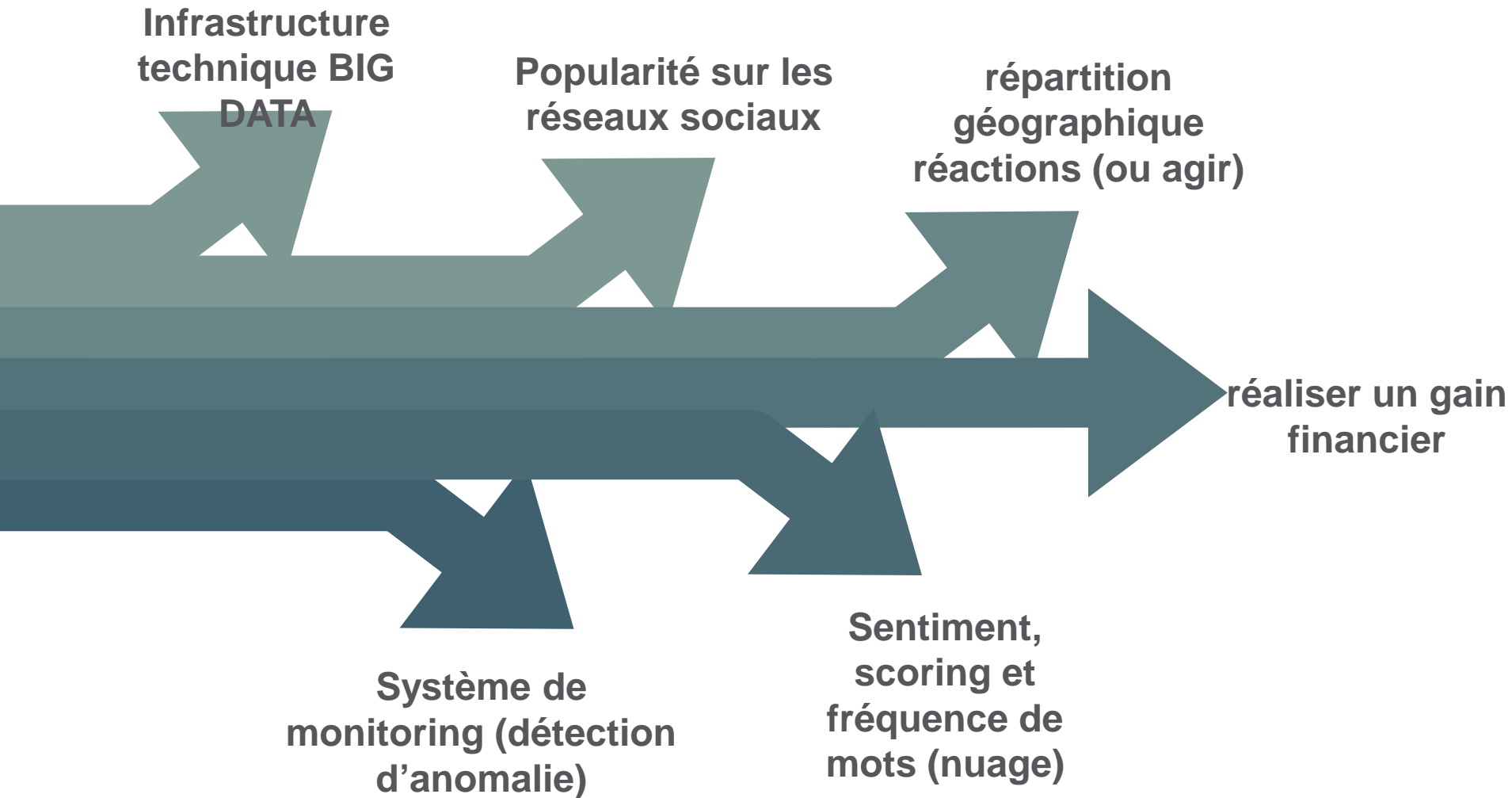
# Plan



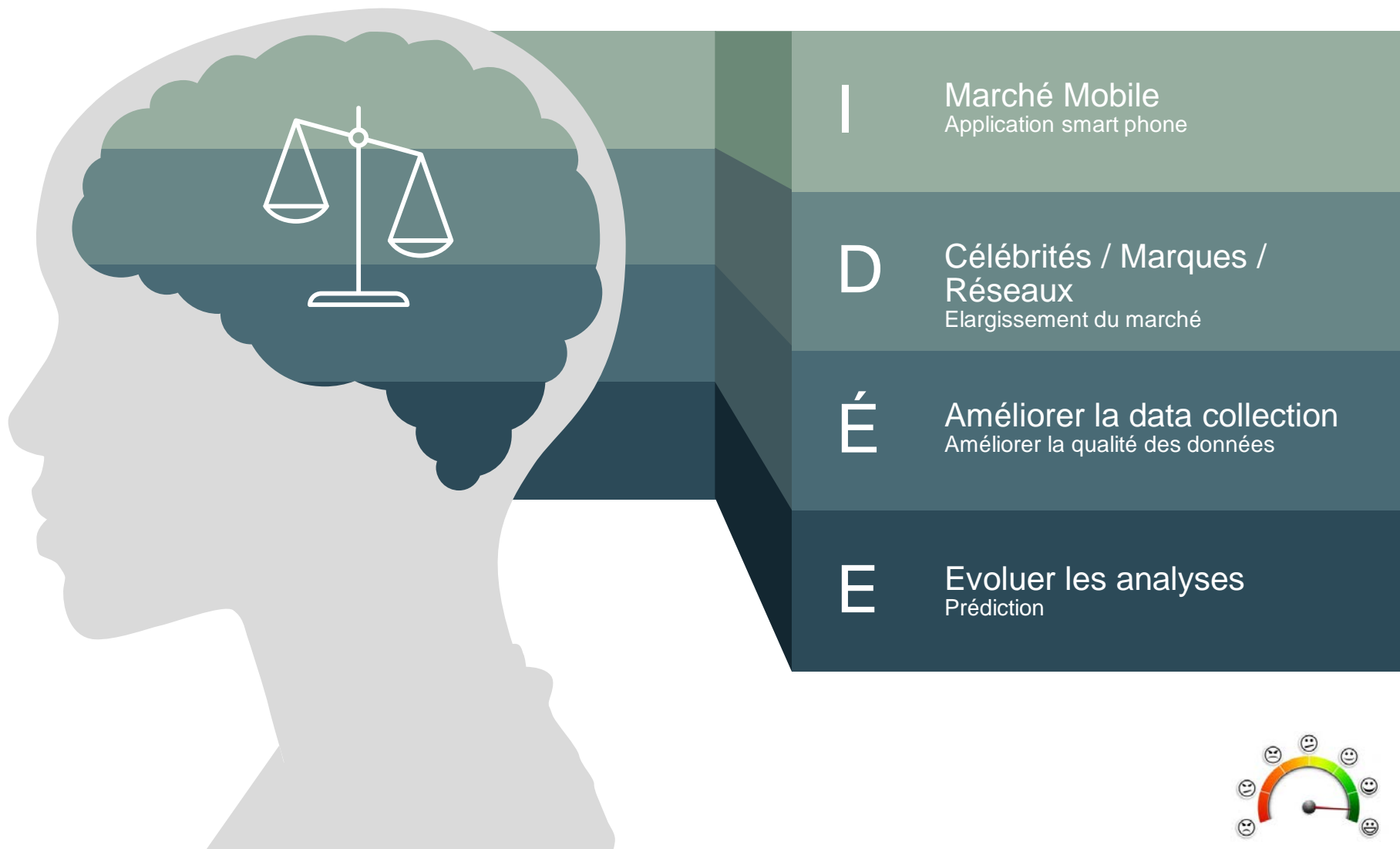
## 4. CONCLUSION & PRESPECTIVES



## 4. CONCLUSION



## 4. PERSPECTIVES



## 4. QUESTIONS



Merci à FITEC et les formateurs



Vos questions? Vos  
remarques ?



A votre disposition pour des précisions  
!



Intéressé par notre projet?  
Contactez nous sur [a.kisma16@gmail.com](mailto:a.kisma16@gmail.com)

