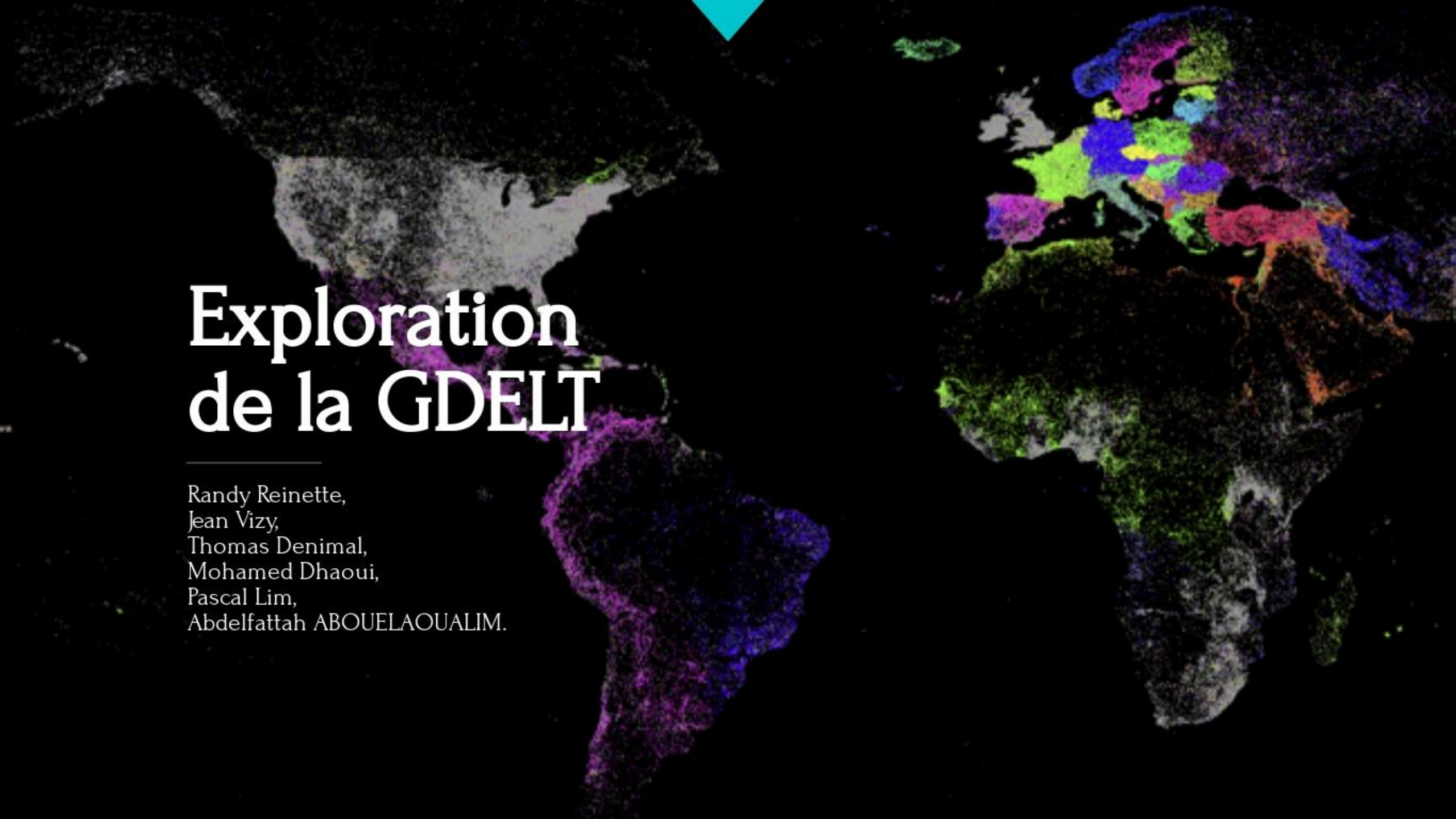
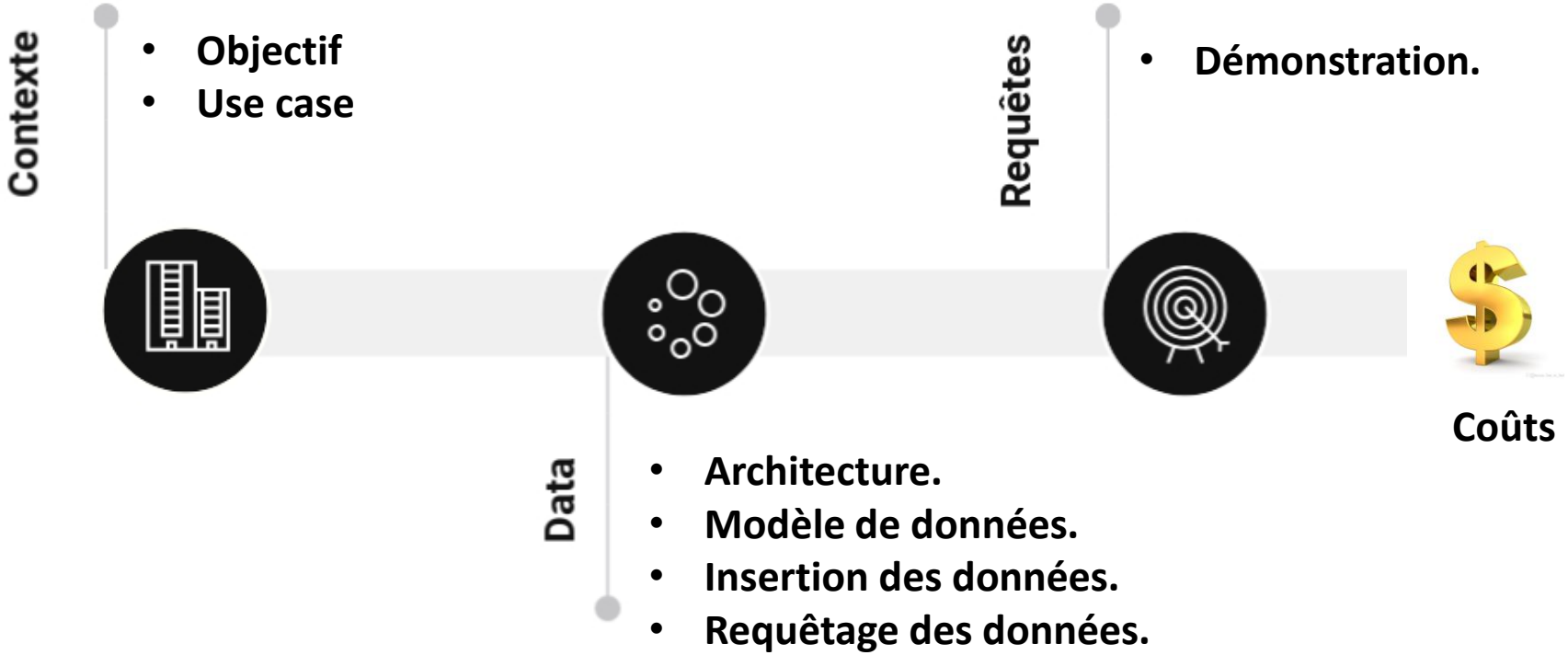


Exploration de la GDELT

Randy Reinette,
Jean Vizy,
Thomas Denimal,
Mohamed Dhaoui,
Pascal Lim,
Abdelfattah ABOUELAOUALIM.

Sommaire


Contexte
• Objectif
• Use case
Requêtes
• Démonstration.
Coûts
Data
• Architecture.
• Modèle de données.
• Insertion des données.
• Requêtage des données.

Contexte

Objectif & Use case

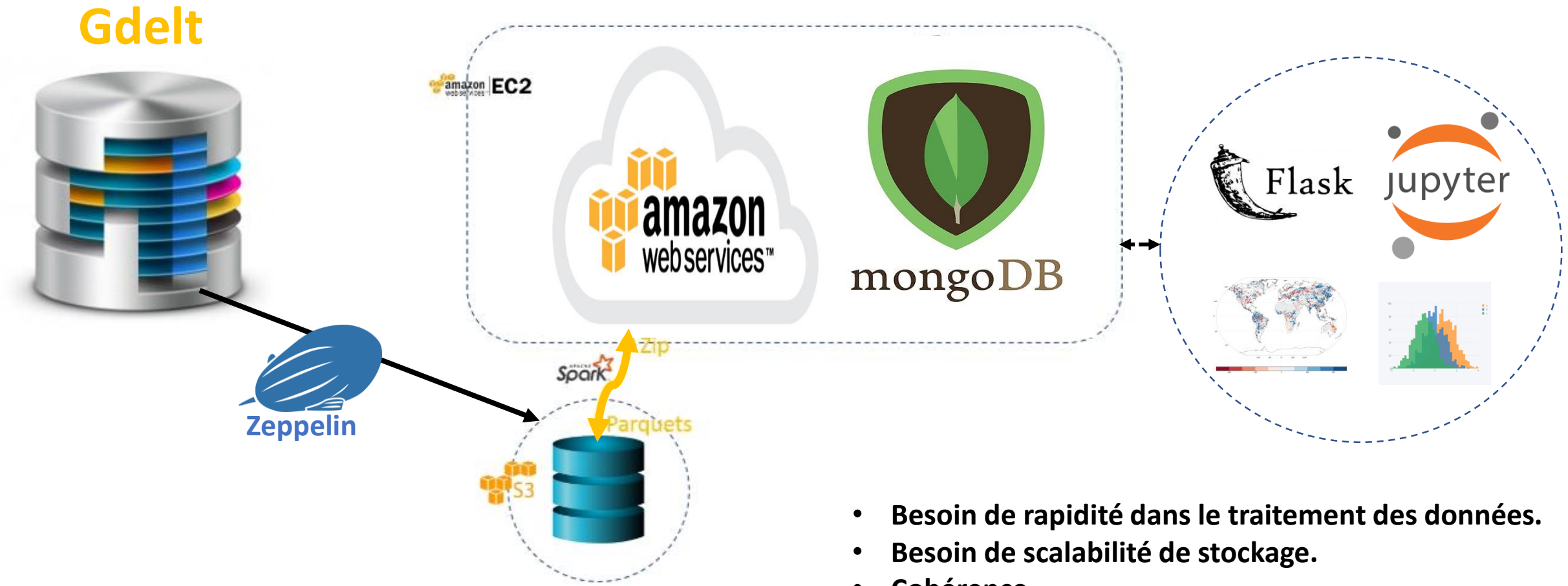
Objectif :

- concevoir un système pour analyser l'évolution des relations entre les pays,
- En s'appuyant sur le ton des mentions dans les médias de chaque pays,
- A partir des données de la base GDELT

Use case :

- Proposer un système de stockage distribué, résilient et performant sur AWS pour les données de GDELT
- Capable de traiter un an de données
- Tolérant aux pannes
- Pour un budget max. de 300€

Architecture

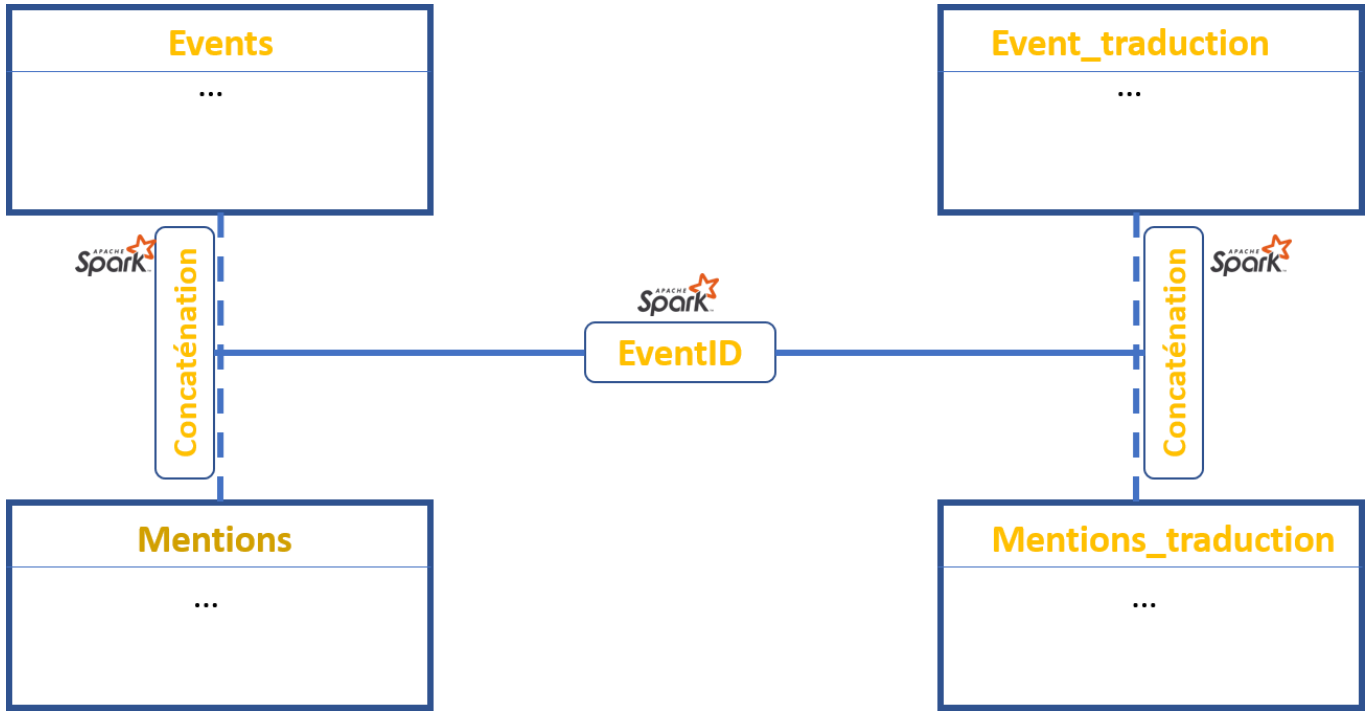


- **Besoin de rapidité dans le traitement des données.**
- **Besoin de scalabilité de stockage.**
- **Cohérence.**
- **Disponibilité tolérance au panne .**

Data

Exploration des données:

- 4 tables + Métadata
- 150 GB sur Mongo.
- 40 GB sur S3



Modèle de données

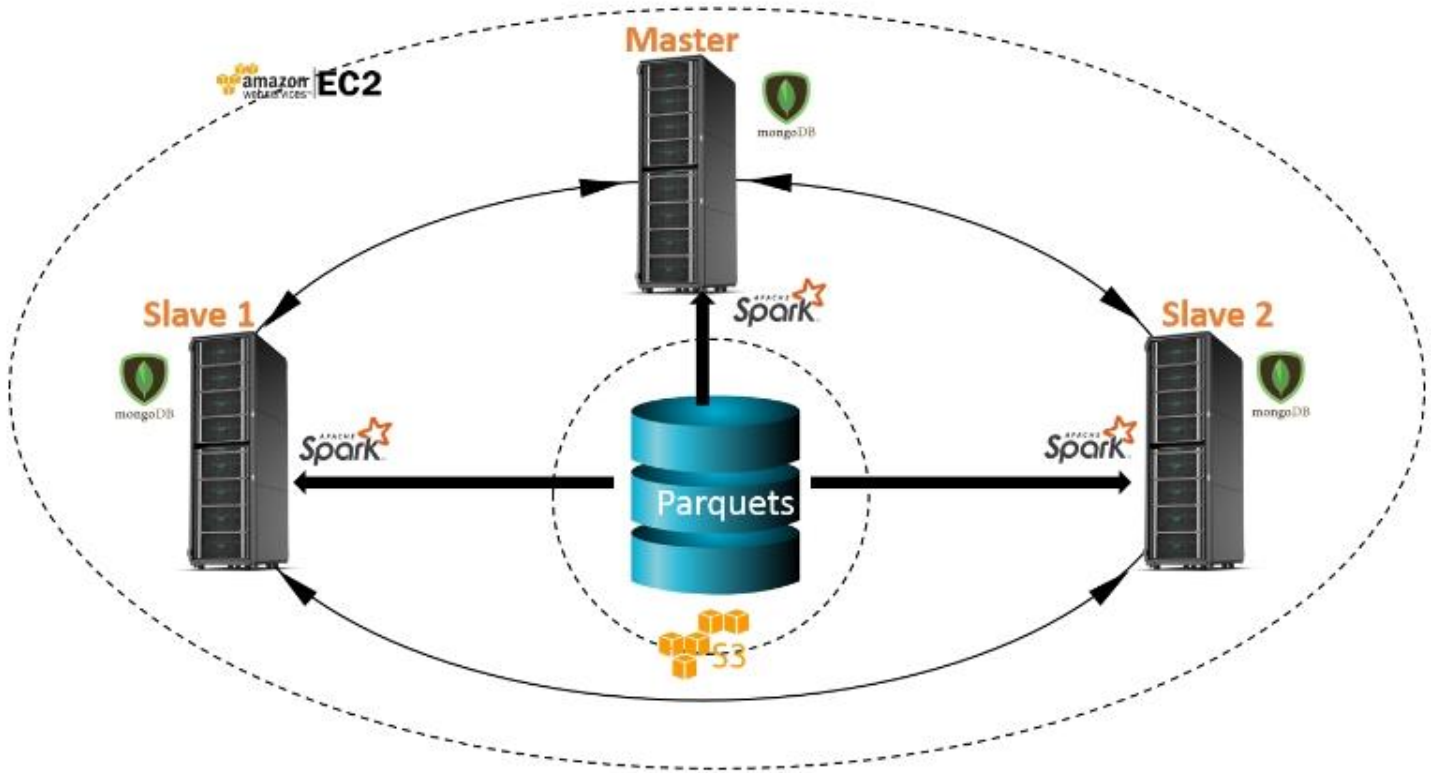
Choix du système du stockage:

- **MongoDB:**
  - Fonctionnalité et accès aux données.
  - Scalabilité et performance.
  - Souplesse d'évolution de l'architecture.
  - Tolérance aux pannes.
  - Sharding - Multi-indexing.

- **Cassandra:**
  - Moins de fonctionnalités par rapport à MongoDb.
  - Plusieurs fonctionnalité standard type SQL ne sont pas disponible.
  - Indexation moins riche que MongoDb.

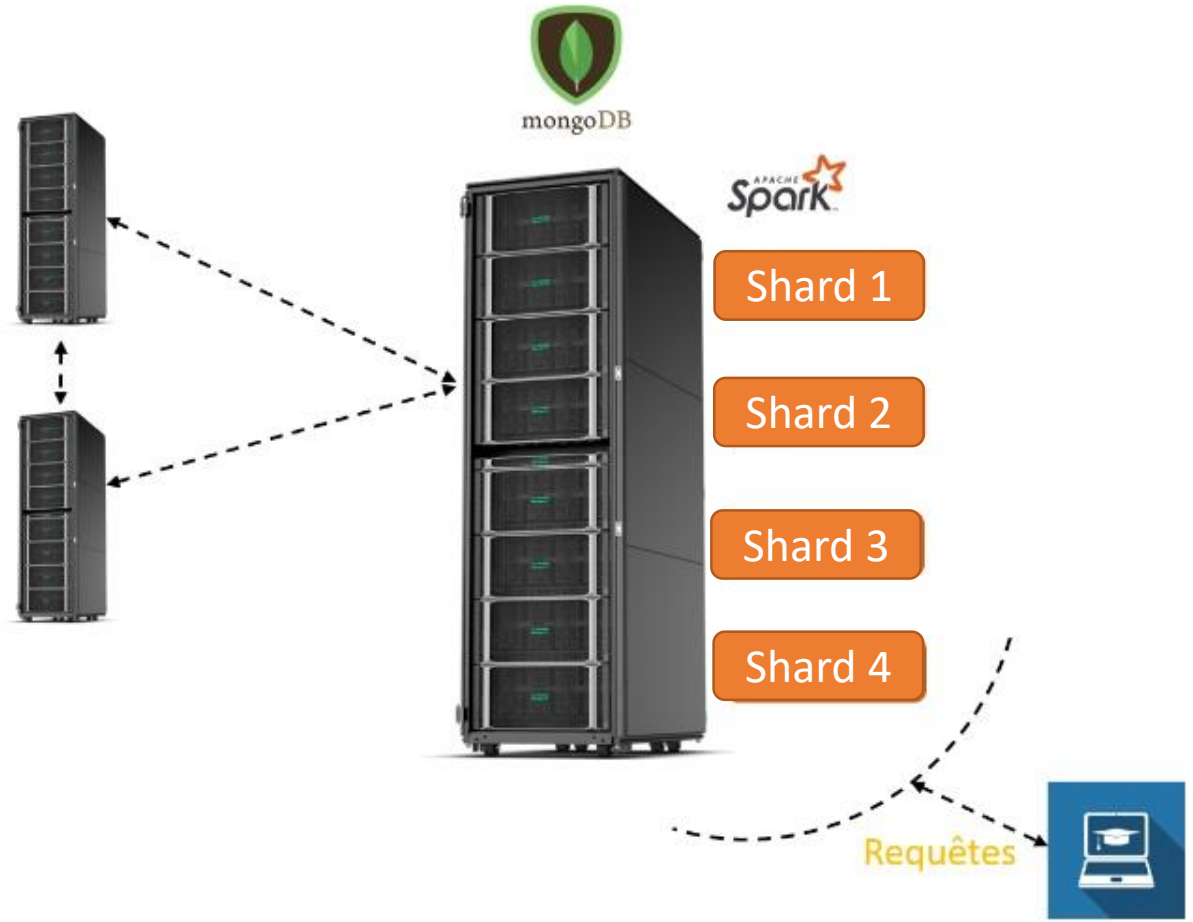
Architecture

Insertion



Architecture

Requêtage



Requêtes

Question 1

Le nombre d'articles/événements pour chaque (jour, pays de l'événement, langue de l'article).

Question 2

Pour un acteur(pays/organisation …) ⇒ afficher les événements qui y font référence.

Question 3

Les sujets (acteurs) qui ont eu le plus d'articles positifs/négatifs (mois, pays, langue de l'article).

Question 4

Acteurs/pays/organisation qui divisent le plus.

Coûts

Stockage/Computing	3 x t3.x2large + S3
Coût/jour	5 Dollars
Coût total	50 Dollars

?