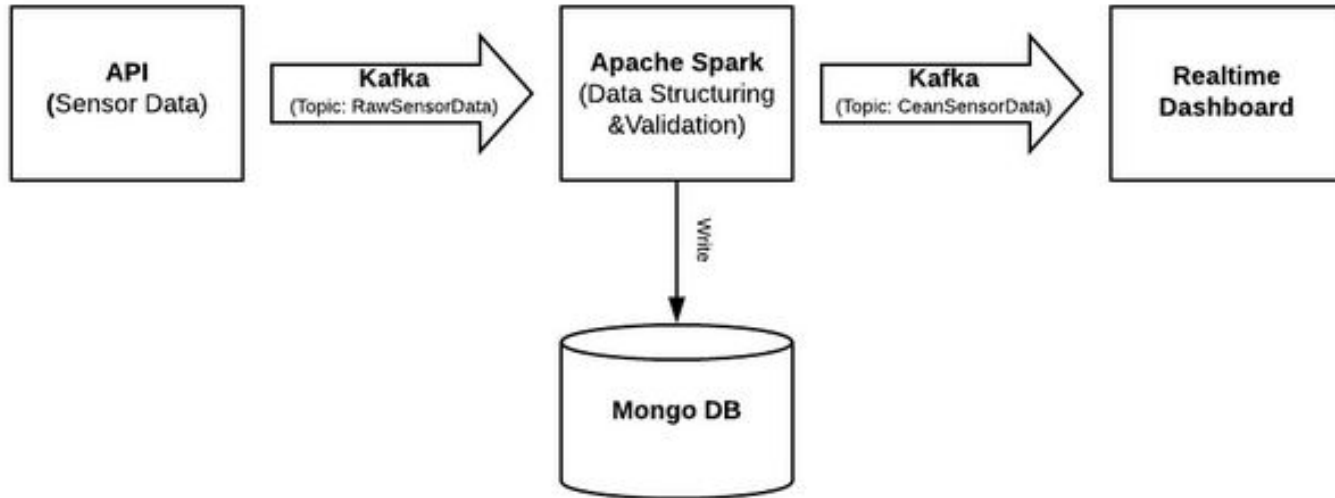


Rating series and movies

Leandro Rusotti, Adamo Canonne et Nadia Nejad-Sattary

Projet initial - Capteurs dans les rivières



L'intégration des données

Téléchargement des données sous format CSV :

<https://github.com/sharkdp/imdb-ratings/blob/master/ratings.csv>

Transfère de données avec SCP (Secure Copy Protocol)

```
scp -P 2222 -r /project maria_dev@127.0.0.1:/home/maria_dev
```

Copie des fichiers dans le cluster HDFS

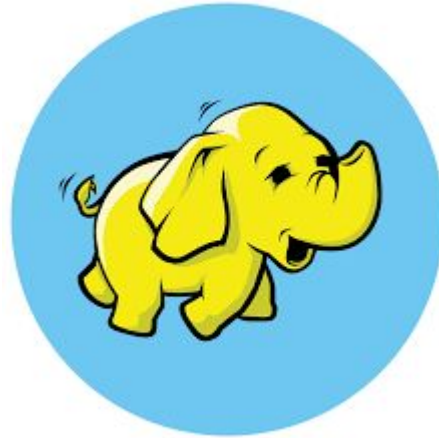
```
hdfs dfs -put ratings_v2 /user/maria_dev/project
```

L'utilisation des outils Cloudera

Spark Serveur - Traitement des données

HDFS - stockage des données

Python



Modifications

Afficher les ratings de films ou séries selon le choix de l'utilisateur

Utilisateur

- Choisi entre film et série
- Choisi le genre
- Choisi le nombre de ratings qu'il veut afficher

Amélioration possible



Problèmes rencontrés sur le nouveau projet

- RAM
- CSV
- User Input en python

Résultat

Notre GitHub : https://github.com/nadiasat/technologies_emergentes.git