

Projet en groupe

1. Objectifs

L'objectif principal du projet est la mise en oeuvre d'une base de données NoSQL sur un jeu de données de votre choix. Un maximum de traitements devront être effectués via la base de données.

2. Données

Le choix du jeu de données est laissé à l'initiative de chaque groupe. Il peut y en avoir plusieurs. Le nombre total de lignes devra être de quelques milliers au moins.

Quelques idées :

- données de la SNCF : <https://data.sncf.com/explore/?sort=modified>
- données publiques : <https://www.data.gouv.fr/fr/>
- projet personnel ou professionnel (si données non confidentielles)
- autres :
<https://www.kaggle.com/datasets>
<https://data.world/datasets/open-data>
<https://www.forbes.com/sites/bernardmarr/2016/02/12/big-data-35-brilliant-and-free-d-ata-sources-for-2016/>
<https://github.com/awesomedata/awesome-public-datasets>
...

3. Attendus

3.1. Requêtes

Pour l'intérêt de l'exercice, on essaiera de mettre en oeuvre des requêtes complexes de type :

- expression régulière ('LIKE' en SQL)
- agrégation (somme, moyenne etc...)

3.2. Restitution

Une présentation de 20 minutes (+ ~5 minutes de questions) par groupe vendredi matin. Tout le monde doit parler ! Il faut donc répartir les rôles de manière la plus équilibrée possible.

Le support de présentation peut être basé sur:

- une présentation PowerPoint (15 slides maximum pour 20 minutes)
- une interface Web

- une démo en direct du fonctionnement de la base de données
- Excel pour faire des graphiques
- tout outil de votre choix qui apporte un plus...
- exemples : <https://public.tableau.com/s/gallery>

L'exercice est volontairement très libre : allez à l'essentiel pour répondre au travail demandé, vous pourrez toujours ajouter d'autres éléments s'il vous reste du temps !

4. Evaluation

Les critères d'évaluation sont :

% note	Description
50	<p>Mise en place d'une base de données NoSQL :</p> <ul style="list-style-type: none"> - (optionnel) bonus : installation de la base de données (si pas déjà installée pendant les séances) - (optionnel) bonus : déploiement sur un serveur cloud (OVH, AWS...) - justification du choix de la base de données vis-à-vis des données - insertion des données dans la base - complexité des traitements effectués (requêtes) - (optionnel) bonus : analyse du comportement de la base de données en cas de défaillance d'un noeud
20	<p>Interprétation des données grâce aux traitements faits par l'intermédiaire de la base de données</p> <p>Interface graphique/web pour l'analyse des données</p>
20	<p>Qualité du code/des requêtes (fournir un lien vers un repo public GitHub, Bitbucket ou autre)</p> <p>L'utilisation d'un client du langage de votre choix (Java, Python...) est fortement appréciée</p>
10	<p>Qualité de la présentation orale et du travail d'équipe</p> <p>(optionnel) bonus : simulation temps réel pendant la présentation (pour élections américaines par ex.)</p>