

# Sujet de Programmation large échelle PLE

## 2023-2024

L'objectif de ce projet est de concevoir un système d'analyse de combinaison de cartes pour le jeu vidéo Clash Royale. Dans ce jeu chaque joueur possède un ensemble de cartes (96 maximum) et il doit en choisir 8 pour constituer un « deck » qui lui permettra d'affronter un adversaire qui lui aussi possède un deck de 8 cartes.

Nous avons enregistré près de 2.000.000 de parties de ce jeu dans notre master dataset, nous disposons pour chaque partie du deck de chaque joueur ainsi que des informations sur le niveau d'expérience des joueurs mais aussi sur le niveau moyen des cartes des decks.

L'objectif du projet est d'obtenir un système similaire à celui proposé ici: <https://royaleapi.com/decks/popular> pour aider les joueurs à choisir les meilleures combinaisons de cartes à jouer.

Le projet se décompose en 2 parties non indépendantes.

### Partie I (Map/Reduce Hadoop):

Dans cette partie, nous vous demandons pour chaque deck utilisé de calculer :

- Le nombre de victoires
- Le nombre d'utilisations du deck
- Le nombre de joueurs différents qui l'ont utilisé au moins une fois
- Le niveau de clan le plus élevé où le deck a obtenu une victoire.
- La différence moyenne de la force du deck lorsqu'il gagne.

Nous voulons calculer ces informations à plusieurs niveaux de granularité. Pour chaque semaine, pour chaque mois et pour l'ensemble des données.

Pour chaque niveau de granularité nous souhaitons obtenir les k plus grandes valeurs.

Implémentez votre solution en MapReduce/HBase et en nodejs express pour le serving layer.

Vous trouverez dans le fichier deck.js le code nécessaire pour obtenir à partir des codes des decks les cartes et les images des cartes du jeu clash royale.

### Partie II (Spark) :

Dans cette partie nous voulons générer pour chaque deck toutes les combinaisons possibles de 1, 2, 3, 4, 5, 6, 7 et 8 cartes. Cela représente 255 possibilités pour chaque deck de notre jeu de données.

Pour chacune de ces combinaisons nous voulons calculer les mêmes statistiques que celles calculées dans la partie I.

En utilisant le Sequencefile résultat de la partie I, proposez une solution en Spark pour généraliser les calculs de la partie I à tous les n-grams de cartes.

Dans votre interface web, permettez à l'utilisateur d'entrer un n-gram (n cartes) et d'obtenir comme dans <https://books.google.com/ngrams/> l'évolution de l'utilisation de ce ngram ainsi que l'évolution de ses statistiques.