

Projet : Data Cleaning de parties Clash Royal

Villot Yliane, Alhousseine bah

PLAN



- 1-ANALYSE DU DATASET
- 2-IMPLÉMENTATION DU DATASET
- 3-ANALYSE DU RÉSULTAT
- 4-CONCLUSION

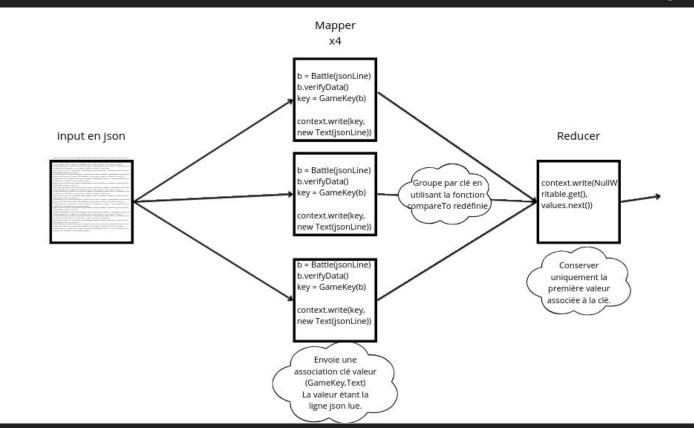
Analyse du dataset



```
"date": "2024-09-23T16:04:46Z",
"game": "pathOfLegend",
"mode": "Ranked1v1_NewArena",
"round":0,
"type": "pathOfLegend",
"winner":1,
"players":[
      "utag": "#U82CQ9C8Q",
      "ctag": "#QYPVC8RG".
      "trophies":5498.
      "exp":32,
      "league":1,
      "bestleague":2,
      "deck": "00010512213c5b5c".
      "evo":"".
      "tower": "6e",
      "strength":10.75,
      "crown":0,
      "elixir":12.41,
      "touch":1,
      "score":0
   { E
      "utag": "#8QRCGQJC",
      "trophies":7109,
      "exp":43,
      "league":1,
      "bestleague":5.
      "deck": "080c111416235b66",
      "evo":"08",
      "tower": "70",
      "strength": 11.1875,
      "crown":1,
      "elixir":2.74,
      "touch":1,
      "score":0
```

Implémentation





ANALYSE DU RÉSULTAT



12 Mapper au total

Entrées Map: 2,781,989 enregistrements ont été lus par les mappeurs.

Filtrage

Sorties Map : 2,765,874 enregistrements ont été émis par les mappeurs.

Perte de 16,115 enregistrements (2,781,989 -2,765,874) due aux filtrages





Entrées Map : 2,781,989 enregistrem ent

Filtrage

mappeurs
émettent
2,765,874
enregistrem
ents.

perte
16,115
enregistr

ements

ont été

filtrés

Entrées
Reduce: Les
réducteurs
reçoivent
2,381,199
enregistrement
s.

Reduce

Groupeme nt reduce: 2,184,191 groupes ont été créés

Sorties
Reduce les
réducteurs
émettent
2,184,191
enregistrement
s.



Combien de reducer ?

Une ligne de json ~= 523 octets. On a 2,184,191 parties en sortie

523 * 2184191 = 1142331893 octets Soit 1142Mo

1142 / 128 ~= 9.

Il faudrait donc en théorie environs 9 reducer pour avoir de bonnes performances.