

# Rapport BigData

William Arnault

January 2021

## 1 Question d'infrastructure

a) On a :

- $1\text{ligne} = 1\text{tweet} \Rightarrow 5Ko$ ,
- 504M de tweet par jour,
- Notre cluster contient 24 ordinateurs, chacun dispose d' 1To de mémoire donc notre cluster a 24To de mémoires.

$$504M * 5Ko = 2.52To$$

$$2.52Go * 3 = 7.56To$$

Par jour on a donc 7.56To de données.

$$24To \div 7.56To \approx 3\text{jours}$$

On peut donc stocker 3 jours de données sur notre cluster.

b) On a :

$$24To \approx 3\text{jours}$$

$$1\text{block} = 128Mo$$

Donc :

$$24To \div 128Mo = 187500$$

il nous faudra donc 187500 blocks

$$187500/24 = 7812.5 \approx 7812$$

Sur chaque machine on utilisera donc 7812 block en moyenne.

c) On sait qu'un jour de données revient a 7.56To

$$(5 * 365) * 7.56 = 13797To$$

il faudra donc 13797 machine pour stocker 5ans de données.