**Question préliminaire:**

1) Nous avons besoin d’un grand espace de stockage et nous savons pas précisément ce qui va être stocké. Ducoup un data lake fera l’affaire. On peut donc considérer un AZURE DATA LAKE STORAGE 2 de microsoft.

2) On cherche à avoir le rapport disponible le plus vite possible pour intervenir le plus tôt. On a une contrainte de temps et on doit analyser en continu et non à la fin de la journée, donc un stream de données est nécessaire.

3) Ils ont engagé seulement une équipe de data scientist qui en principe sont spécialisés dans l’écriture d’algorithme et de statistiques et aucun data engineer / architect qui auraient pu s’occuper de l’architecture.

4) Il manque une donnée correspondant à la date du rapport.

Grâce à cette donnée, on pourra prédire à l’avance les zones qui seront susceptibles d’avoir un peacescore moyen élevé selon la date et davantage de peacewatchers pourront être déployés.

PROJET ARCHI:

Data Lake -> AZURE DATA LAKE STORAGE 2

Rapport petit -> STREAM -> KAFKA

Langage -> SCALA ou PYTHON