**Question préliminaire:**

1) Besoin d’un grand volume de stockage -> data lake

2) Rapport disponible le plus vite possible. On a une contrainte de temps et on doit analyser en continu, 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