# Lab Spark SQL

L’objectif de ce lab est de se familiariser avec l’api spark sql.

la dataset cible(IFA-World-Cup) comprend : 44 341 résultats de matchs internationaux de football, depuis le tout premier match officiel en 1872 jusqu'en 2022 . Ces matchs couvrent un large éventail de compétitions, de la Coupe du Monde de la FIFA à la FIFI Wild Cup, en passant par les matchs amicaux. Il s'agit exclusivement de matchs internationaux masculins ; les données n'incluent pas les Jeux Olympiques ni les matchs opposant au moins une équipe nationale B, une équipe U23 ou une sélection de championnat.

A partir du dataset des **matchs internationaux de football (1872–2022) on veut analyser et répondre aux questions suivantes :**

# Découverte & requêtes simples

1. Combien de matchs sont présents dans le dataset ?
2. Quelle est la **première** et la **dernière année** couverte par les données ?
3. Lister les **10 tournois les plus fréquents**.
4. Combien de matchs ont été joués sur terrain neutre ?
5. Lister les **10 pays** ayant accueilli le plus de matchs.
6. Combien de matchs ont terminé sur un **score nul** ?
7. Afficher les matchs où le score total (home + away) est supérieur à 6.

# Agrégations & statistiques

8. Nombre total de matchs joués par chaque équipe (domicile + extérieur).
9. Top 10 des équipes ayant marqué le plus de buts (toutes compétitions confondues).
10. Moyenne de buts par match par décennie.
11. Nombre de matchs joués par tournoi et par année.
12. Classement des équipes ayant remporté le plus de matchs à domicile.
13. Nombre de victoires, défaites et nuls pour chaque équipe.
14. Score moyen des matchs joués sur terrain neutre vs non neutre.
15. Top 5 des matchs avec l’écart de score le plus élevé.

# Requêtes analytiques (fenêtres & logique avancée)

16. Calculer le **goal average** (buts marqués − buts encaissés) par équipe.

17. Classement des équipes par nombre de victoires par année (ranking avec `ROW_NUMBER`).
18. Évolution du nombre de matchs par décennie.
19. Identifier les équipes invaincues sur une année donnée.
20. Calculer la **longue série de victoires consécutives** par équipe.
21. Pour chaque tournoi, déterminer l'équipe la plus victorieuse.
22. Comparer les performances à domicile vs à l'extérieur pour chaque équipe.