

Spark Part:

We are expecting you to develop a spark application to store files of the players according to their continent, please follow the following guidelines:

1. Use the provided data sample to store a new file that contains the player name and the continent he is coming from.
2. Using parallelism is a requirement, We are expecting one file per continent as an output.
3. Optimize your solution to expect that the data set we are providing is about a 100 GB of data.
4. After a while, we will have an updated document that has the latest salaries of the players, you are expected to build your spark application to support taking these updates and merge the new values of these players salaries into our warehouse. The new file will only have the players with the updated salaries only.
5. Output format should be csv.

Hive Part:

Build hive table/s that helps you the best to answer the following questions, we are expecting you to write down the queries that answers them:

- 1- Which top 3 countries that achieve the highest income through their players?
- 2- List The club that includes the most valuable players, The top 5 clubs that spends highest salaries
- 3- Which of Europe or America - on continent level - has the best FIFA players?