

Interview - Codility

Spark - Question 1

The data file auth.csv is the aadhaar authentication file dataset.

I wanted all the SA(Service Agency) from the auth.csv file but with the below condition.

For all the AUA greater than "650000" fetch the SA which are only numeric and res_state_name is not Delhi.

Write plain scala program to achieve it. Do not use spark.

The problem statement is freezed and any query on the problem statement should be proceeded with suitable assumptions.

Please state your assumptions in the code that if you have assumed any.

The problem statement is simple, but the evaluation will be on the following criteria:

[1] Object Oriented Design - Design using traits and classes.

[2] High modular code. Usage of Design Patterns and Principles.

[2] Use of best practices in scala. - Custom exception handling, Either, Option, etc..,

[3] Bonus: Usage of functional design patterns : monads, monoids, semigroup, functors.

[4] Extra Bonus: Usage of cats/scalaz libraries. Using REST/ Slick is all fascinating.

Spark - Question 2

The data file auth.csv is the aadhaar authentication file dataset.

I wanted all the SA(Service Agency) from the auth.csv file but with the below condition.

For all the AUA greater than "650000" fetch the SA which are only numeric and res_state_name is not Delhi.

The query is same as Question 1 but write the spark code to solve it.

The problem statement is freezed and any query on the problem statement should be proceeded with suitable assumptions.

Please state your assumptions in the code that if you have assumed any.

You can either choose RDD or Dataset based approach but the code should be modular and well designed.