



Assignment

Week9: Apache Spark - General Purpose
Cluster Computing Framework

IMPORTANT

Self-assessment enables students to develop:

1. A sense of responsibility for their own learning and the ability & desire to continue learning,
2. Self-knowledge & capacity to assess their own performance critically & accurately, and
3. An understanding of how to apply their knowledge and abilities in different contexts.

All assignments are for self assessment only. Solutions will be released on every subsequent week. Once the solution is out, evaluate yourself.

No discussions/queries are allowed on assignment questions on slack channel.

Note: *You can raise your doubts once the solution is released.*

Problem 1 - Recently asked in HCL

Given an Input data set with name, age and city

if age > 18 add a new column that's populated with 'Y' else 'N'

Please solve this using apache spark



Problem 2 :

Given the input file where columns are stationId, timeOfTheReading, readingType, temperatureRecorded, and few other columns...

We need to find the minimum temperature for each station id.

Please do it using apache spark.





5 Star Google Rated
Big Data Course

LEARN FROM THE EXPERT



9108179578

Call for more details