



Assignment

Week14: Apache Spark - Optimization
Part-2

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. Solutions will be released on every subsequent week. Once the solution is out, evaluate yourself.

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

Note: You can raise your doubts in the subsequent week once the solution is released

TRENDYTECH 9108179578

Problem 1:

Find total orders and total amount per status per day. The result should be sorted by order date in descending, order status in ascending and total amount in descending and total orders in ascending. Aggregation should be done using below methods. However, sorting can be done using a dataframe or RDD. Perform aggregation in each of the following ways :

==>{USING BROADCAST-JOIN} & {USING PERSIST(StorageLevel)}

(A). Just by using Data Frames API - here order_date should be YYYY-MM-DD format (convert these into a DATE_FORMAT(to_date)) So, that when we do sorting in future, the sorting happens on DATE NOT a String)

(B). Using Spark SQL - here order_date should be YYYY-MM-DD format (convert these into a DATE_FORMAT(to_date)) So, that when we do sorting in future, the sorting happens on DATE NOT a String)
using "HASH-AGGREGATION"

Problem 2:

Create jar for problem 1 spark project and run the jar on terminal

Problem 3:

Execute the jar created in problem 2 on cluster terminal with below configuration

- 1.) Disable dynamic allocation
- 2.) Mode -cluster
- 3.) Config your spark.ui
- 4.) Master-yarn
- 5.) Num-executer 6
- 6.) Executer-memory 3g
- 7.) Executer-cores 2

TRENDY TECH 9108179578



5 Star Google Rated
Big Data Course

LEARN FROM THE EXPERT



9108179578

Call for more details