tatilbudur.com .Net Hiring Challenge

Subject: Challenge - Large data integration task

Background

For your project, you need to get your customer’s past sales data to the project database. This data is stored in a central location and can be accessed over network as a stream with related credentials. Because this data is too detailed and too large for your needs, only a summary of those records is needed on the project you are participating.

Definition

- You are required to read and summarize incoming streaming sales data with millions of records and write it to a relational database.
- Data stream is either coming from network or a file. Approximate data size is >200GB.
- Data is needed to be summarized by the week of the year, Store, Company, Brand and Product. In the resulting data set we need to see total sales volume and total price.
- Summarized data will be inserted to the database.
- Assume reading data and writing to DB is already implemented.
- Please just focus on the summarizing task and write the most efficient code in terms of CPU,memory and speed.
- Your code will be run on a Linux container on cloud and has 4 vCPU, 8GB RAM 64GB SSD storage.
- We are also concerned about code readability and maintainability.
- You can add any thoughts as comments inside your code block.
- Use provided data classes.
- Week number value should be in year combined with week of year format. For example, for the first week of 2021 it should be integer value of “202101”.
- First week of the year starts with first Monday of the year.

Rules

- Test your code on attached sample project.
- Evaluation will be made using the same sample project.
- Write your code in “*private static IEnumerable<SummarizedSalesData> Challenge1Solution(IEnumerable<SalesData> items)*” method only.
- Your solution method must return summarized values to save them to the database.
- Do not alter any other code, because it is not tested on unchanged sample project.
- If possible do not use any variables other than local variables.
- If needed you can use Console.Writeline for logging purposes.

You can assume incoming data is correct, no need to validate id, name or any other field

GOOD LUCK!

tatilbudur.com

techcareer.net