Instructions – Data Integration Engineer Technical Assessment

Introduction:

- As a Data Integration Engineer you are expected to have good hands-on coding skills. As part of the recruitment process, we want to give you the opportunity to demonstrate your skills
- This code review involves solving a problem which is outlined below
- It must be a scalable solution
- We are interested in the approach you follow for solving this problem.
- You can use any of these programming languages (Java/.NET/Python/Scala) to solve this problem.
- If you use online resources to assist with solving a requirement, please include the url (in a comment in the code).

Guidelines for Submission:

When writing your code, please be mindful of the following:
Your code should be free of bugs and solve the stated problem.
Your code should be easy to understand and maintain by other developers.
Your code should demonstrate common error handling scenarios.
Unit tests are expected (feel free to use your preferred unit test framework).
If you have made any assumptions in solving the problem, please mention those assumptions in comments in the code.
This is a confidential test. If you provide a repository with the code no mentions to Deloitte can be seen (repo name or README.md).

Task/Problem Statement:

- Write a program to fetch:
    - The passenger count, trip time (in minutes), distance travelled, temperature grouped by pickup start date (YYYY-MM-DD format).
    - Output file should be a CSV and can either be written to local or Azure blob storage.
- Code should be able to read the below input parameters:
    - Medallion
    - Pickup Start Date
    - Pickup End Date
- Not all input parameters are mandatory:
    - Medallion – Optional, and multiple values can be passed separated by ‘,’
    - Pickup Start Date – Mandatory. Input date format is YYYY-MM-DD
    - Pickup End Date – Mandatory. Input date format is YYYY-MM-DD
- Code should be production ready with setting up SQL server, inserting data, and the program should be able to extract the above-mentioned columns.

Files provided:

ny_cab_trips_data.sql – use this file to create table and insert data. Below is the table schema.

Column	Data Type	Sample Value
Medallion	text	D7D598CD99978BD012A87A76A7C891B7
hack_license	text	82F90D5EFE52FDFD2FDEC3EAD6D5771D
vendor_id	text	VTS
rate_code	int	1
store_and_fwd_flag	text	NULL
pickup_datetime	datetime	2013-12-01 00:13:00
dropoff_datetime	datetime	2013-12-01 00:31:00
passenger_count	int	1
trip_time_in_secs	int	1080
trip_distance	double	3.79
pickup_longitude	double	NULL
pickup_latitude	double	NULL
dropoff_longitude	double	NULL
dropoff_latitude	double	NULL

Weather.csv -- to get temperature for pickup start date. Below is the schema.

Column	Data Type	Sample Value
Datetime	text	12/01/2013
Temperature	float	42.6
Condition	text	Clear

Sample CSV output:

pickup_date, passenger_count, trip_time, distance_travelled, temperature

2013-12-01,569,689,126, 42.6

Please note that the header in the CSV file is not mandatory, and the above is just a sample output and not the actual data for the date mentioned.

Once you finish the assessment, please package all your files into a single .zip file and ensure your name is part of the file name. If you have used a GitHub repository, please send us the URL.