



מעבדה לאיסוף וניהול נתונים

עבודה מספר 3 ב- SSIS

מטרת התרגיל

בתרגיל זה תפתחו תהליכי טעינה מ- OLTP ל- DWH במתודולוגיית Slowly changing dimension.

הנחיות הגשה

- יש להגיש את כל המטלות במועד על פי ההנחיות שתפורסמה ב Moodle.
- הגשה באיחור תגרוור הורדת ניקוד.
- את המטלה יש להגיש לתיבת ההגשה ב- Moodle. קובץ הפתרון יהיה קובץ מכוון מסוג zip ששמו מורכב ממספרי תעודת הזהות של המגישים, כאשר ביניהם ישנו קו תחתון. (לדוגמה: ID1_ID2.zip)
- קובץ ההגשה יכול לכלול את הקבצים הבאים:
- קבצי SSIS שיצרתם (שיוכלו לשמש להרצה לשם בדיקת הפתרון).
- קובץ Word שמכיל תרשימים מה- SSIS של ControlFlow ו- DataFlow.
- הוראות הרצה מיוחדות (רק במידש ונדרש) בקובץ Word.

משימות התרגיל

- צרו מתחת לכונן C: את התיקייה Data והעתיקו אליה את הקבצים Employees.csv ו- Cities.csv. המצורף לתרגיל. קובץ זה מכיל נתונים עדכניים של עובדים ממערכות OLTP.
- ההנחה היא שמערכות OLTP יוציאו פלט נתוני עובדים מעודכן לקובץ בשם Employees.csv לתיקיית C:\Data בתדירות כלשהיא.
- הגדירו בבסיס נתונים ב- SQL-Server טבלת עובדים בשם Employee שהמבנה שלה הוא כמו של הקובץ Employees.csv ותיטען מקובץ זה בתדירות של אחת לדקה.
- לאחר מכן, יש להוסיף/ לעדכן בטבלת Employees את העמודות הבאות:
 - Age (שדה חדש)** – גיל העובד (במספר שלם מעוגל כפלי מטה). יש לחשב לפי שדה תאריך הלידה (DOB).
 - City (שדה קיים)** – שם עיר המגורים של העובד. יש להשתמש בקוד העיר של העובד שנלקח מהקובץ Employees.csv ולהחליף אותו בשם העיר שילקח מהקובץ Cities.csv.
 - UpdateDateTime (שדה חדש)** – תאריך ושעת העדכון האחרון של כל רשומת עובד שחל בה עדכון בשדה אחד לפחות.
- בנוסף, הגדירו בבסיס הנתונים את טבלת DWH_Employees שתשמור נתונים היסטוריים של עובדים כמחסן נתונים ותכלול את כל המידע על כל עובד מטבלת Employee.
- קונפיגורציית Incremental תעבוד בשיטת עדכון נתונים בזמן אמת (אחת לדקה). כלומר, לאחר שטענתם את הקובץ Employee.csv לטבלה בשם Employee בתדירות של אחת לדקה, עליכם לבצע השוואה מול נתוני טבלת DWH_Employees ולזהות אילו עדכונים בוצעו לנתונים של אילו מהעובדים. בהתאם לשינויים הללו, שמרו במחסן הנתונים את כל העדכונים שבוצעו בתצורה של Slowly changing dimension מסוג 2. יש להוסיף לטבלה DWH_Employees את כל השדות והרשומות הנדרשים על מנת לתמוך בתצורה זו.



7. בעת עדכון טבלת DWH_Employees, עליכם לתמוך בפעולות הבאות בהתאם לזמני עדכון הרשומות בטבלת Employee:
- הוספת רשומת עובד חדש.
 - עדכון פרטי עובד קיים – אין לאפשר שינוי בשדה [ת"ז עובד] מכיוון שהוא שדה מפתח.
8. הנחת היסוד היא שמטבלת Employees לא נמחקים עובדים ולא נמחקות/מתווספות עמודות.
9. תעדו בטבלת Log את כל פרטי השינויים ההיסטוריים שבוצעו בנתוני טבלת העובדים המקורית. להלן מבנה הטבלה לדוגמה:

EmployeeID	FieldName	OldValue	NewValue	UpdateDateTime
12345	Salary	10,000	11,000	21/5/2022 23:56

10. תעדו בטבלת Errors את פרטי כל השגיאות, באיזה שלב בתהליך הן קרו ואת זמן התרחשותן.
11. בסוף ההרצה של כל אחת מהקונפיגורציות יש לבצע את הפעולות הבאות:
- יצאו את טבלת Employees לקובץ בשם Employees_OLTP.csv בתיקייה C:\Data
 - יצאו את טבלת DWH_Employees לקובץ בשם DWH_Employees.csv בתיקייה C:\Data
 - יצאו את טבלת Log לקובץ ששמו Log.csv בתיקייה C:\Data
 - יצאו את טבלת השגיאות Errors לקובץ ששמו Errors.csv בתיקייה C:\Data

הארות והבהרות:

- אין להשתמש בסקריפטים באף אחת מהמשימות !!
- יש לממש את כל התהליך שלעיל באמצעות Package (או יותר) של SSIS.
- שימו לב להשתמש בשמות משמעותיים במשימות (Tasks) ובאנוטציות על מנת להסביר תהליכים מורכבים ב-SSIS.

בהצלחה