

חברת תרגילים למודול SSIS

בנית DataWareHouse

ותהליכי טעינת נתונים - ETL

התקנה:

יש להתקין את בסיס הנתונים [Worldwide Importers Database](#) מתוך הקישור המצורף :

או לחפש את קובץ bak של ה DB ב- github

קישורים נוספים נדרשים :

[SQL Server 2017 Installation](#)

[SQL Server Data Tools](#)

[SQL Server Management Studio](#)

1. בארו את המושגים הבאים:

a. Data warehouse

b. Operational DB

c. ETL

d. סוגי סכמות ב – DWH והמאפיינים שלהם

e. מהו dimension table?

f. מהו fact table?

g. מה ההבדל בין surrogate key לעומת business key

h. מה ההבדל בין variable לעומת parameter?

EXTRACT & LOAD

2. צרו package חדש בשם : LoadData.
יש לטעון את הקובץ products.csv לטבלת products בבסיס הנתונים Northwind,
עקבו אחר ההוראות הבאות:
(לפני תחילת התהליך יש:
1. להוסיף ב DB עמודת MAAM בטבלה מסוג money
2. לבטל את ה identity דרך הממשק.
2. בצעו המרה של הקלט למבנה המתאים בטבלה הקיימת ב DB.
3. עבור מוצרים שהקוד קיים בטבלה - יש לבצע עדכון מחיר, ומחיר כולל מע"מ,
עבור מוצרים חדשים יש לבצע הכנסה כולל לעמודה ה maam.
- הכניסו את התהליך ל container אחד.
3. צרו package חדש בשם: SplitData
טענו שוב את הקובץ products.csv. הפעם יש לפצל את הטעינה לשתי טבלאות שונות -
מוצרים המיועדים למכירה, ומוצרים שאינם מיועדים למכירה.
הדרכה: הוסיפו סקריפט ליצירת שתי טבלאות תואמות עבור prod_toSale, prod_NotToSale, בצעו split של הקליטה לשתי טבלאות שונות.
הכניסו את התהליך ל container אחד.
4. ענו על השאלות הבאות :
- a. מתי נשתמש ב - derived column
 - b. מתי נשתמש ב - data conversion
 - c. How many Control Flow in a package?
 - d. באילו component נעשה שימוש כאשר ה- source column מסוג של nvarchar כאשר ה- destination הוא varchar?
 - e. מהם 3 סוגי ה CONTAINERS הקיימים – מתי נשתמש בכל אחד מהם, הציגו דוגמא מעשית עבור כל אחד מהם.
 - f. עבור LookUp Component - בארו את שלושת הסוגים של cache Mode, מתי נשתמש בכל אחד מהם.
 - g. בארו את המושגים הבאים ומשמעותם:
 - i. dtsx
 - ii. SCD
 - iii. Task
 - iv. Container
 - h. מה היחס בין : Solution , Project , Package
 - i. Logging
- פרטו את סוגי הלוגים שניתן להפיק מתהליך SSIS, בגין כל סוג לוג - כתבו האם מדובר בקובץ טקסט או קובץ בינארי.

שם הלוג	קובץ טקסט או בינארי	מנוהל ע"י מערכת ההפעלה או SQL	ניתן לתשאל את הלוג ע"י SQL
SSIS Log Provider for SQL Server	בינארי	SQL	ניתן

5. צרו package חדש בשם: TRANSFORM_AND_LOAD
4. המטרה : לבצע ייצוא של טבלת EMPLOYEES למס' קבצי אקסל שונים.
5. שלבים בביצוע :
- יש לבצע את התיקונים הבאים – **derived column**
- ```
--when INT & Null Then -1
--when Date & Null Then '1900-01-01'
--when String & Null Then "Unknown String"
```
- יש להוסיף את העמודות: seniority, age, סך מכירות לעובד.
  - יש להוריד עמודות המכילות שדה שאינו מכיל טקסט, מספר, תאריך (NOTES, PHOTO).
  - להוסיף עמודות ע"פ משתני ה SYSTEM
- @System::CreatorComputerName
  - @System::PackageName
5. ייצוא לקובץ אקסל/CSV (לעיתים חסר דרייבר לייצא לאקסל).
6. עבור טבלת People, בבסיס נתונים WWI - הציגו דוח Profile:
- אילו עמודות מקבלות NULL
  - אילו עמודות יש בהם ערכים UNIQUE
  - עמודות עם מעל 20% ערכי NULL
  - עמודות שיכולות להיות או שכבר משמשות כ- PK
7. Loop Files
- בצעו ייצוא של הנתונים הבאים : שם מלא של העובד, מס הזמנות , שווי הזמנות. יש להוסיף LOOP CONTAINER אשר עובר על ספרייה ובה קבצי אקסל המכילים שמות עמודות בלבד ( ללא DATA), ומייצא את נתוני הטבלה לתוך כל אחד מהקבצים הללו. הסבר: בתחילת התהליך צרו ספרייה ובה 5 קבצי אקסל עם שמות עמודות בלבד. התהליך שלנו יבצע ייצוא בלולאה של הנתונים מה- DB לתוך כל אחד מהקבצים.
8. Loop Tables
- בצעו ריצה בלולאה על כל טבלאות בבסיס נתונים מסוים לבחירתכם והפיקו קובץ ובו העמודות הבאות : שם הטבלה, מס שורות בטבלה, מס העמודות באותה טבלה(שימו לב לטבלת sys.column

9. **מצורף DEMO :** ייצאו את טבלת ההזמנות ל CSV תוך חלוקה לקבצים שונים לפי שנת המכירה של כל הזמנה . 1996.csv , 1997.csv , 1998.csv

- פתרון ליצירה אוטומטית של הקבצים בספריה

הדרכה: יש לבצע פקודה לשליפת כל השנים שיש מהם מכירות, להכניס לתוך record set, לעבור בלולאת ADO על כל השורות, ועבור כל שורה - לייצר קובץ עם השם המתאים, להעתיק אליו נתונים המתאימים.

10. Loop

צרו תוכנית המקבלת כפרמטר מספר ומציגה את העצרת שלו. יש לייצר פרמטר קלט, להעתיק אותו למשתנה, ולרוץ בלולאה לחישוב העצרת, בסיום יש להציג חלון הודעה ובו תוצאת החישוב.

11. ייבוא של קובץ PRODUCTS תוך שימוש ב LOOKUP ע"מ להבדיל בין UPDATE ו INSERT  
בדקו את התהליך עבור כל אחד מהמקרים - UPDATE לשורה קיימת או INSERT לשורה חדשה. **מצורף DEMO :**

10. להוסיף עמודות ע"פ משתני ה SYSTEM (הודגם בכיתה – יש לבחור מתוך רשימת המשתנים של SYSTEM)

1. [System::CreatorComputerName]@

2. [System::PackageName]@

12. MAIL & LOG :

יש לבצע תהליך להצגת הפלט של הפרוצדורה sp\_spaceuse עבור כל אחת מהטבלאות בבסיס נתונים WideWorldImporters .  
יש להציג event viewer את האירועים הבאים : תחילת התוכנית, וסיום התוכנית.  
במידה והתוכנית נכשלה יש לשלוח מייל לאופרטור המתאים.

תיאור שלבי יצירת ה DTS :

a. צרו package חדש בשם LOOP\_TABLES.

b. הריצו את פקודת sp\_spaceused עם פרמטר של שם טבלה ובחנו את הפלט

פתרון : sp\_spaceused employees

c. יש לייצר טבלה ב – DB בהתאם לפלט של הפרוצדורה sp\_spaceused  
פתרון:

Create Table SpaceUsed

(

TableName NVARCHAR(200),  
TotalRows NVARCHAR(200),  
Reserved NVARCHAR(200),  
TableData NVARCHAR(200),  
Index\_Size NVARCHAR(200),  
Unused NVARCHAR(200))

- d. יש לנקות את הטבלה בתחילת כל תהליך.
- e. יש לעבור foreachloop על כל הטבלאות ב DB
- f. להריץ עליהם פקודת SQL שלעיל, ולהכניס את הפלט לתוך הטבלה המתאימה.
- מצורף Solution**
- g. יש לשלוח מייל דרך script task

13. Loop Files + שימוש בפרמטר  
הוסיפו package עבור מעבר על כל הקבצים בספריה נתונה כפרמטר, עבור כל קובץ יש לבצע את הפעולות הבאות.
- a. טעינה שלו לתוך ה DB - כל הקבצים במבנה זהה.
- b. העתקת הקובץ לספריית archive
- c. הוספת תאריך\_שעה\_דקה\_שניות לשם הקובץ ע"י שימוש ב expression.
- מצורף Solution**

14. Deployment  
לאחר deploy של ה PACKAGE , בצעו את הפעולות הבאות :
- a. הריצו את ה package מתוך ה SSMS
- b. פתחו דוח EXECUTION לניטור התהליך.
- c. שנו את ערך הפרמטר של שם ה DB עבור התהליך LOOP\_TABLES.
15. צרו JOB המתוזמן לריצה בכל רגע שה CPU ב IDLE - ירידה מתחת ל 5% למשך 10 דקות לטעינת הנתונים מ OP – ל DW.
16. עבור הקובץ CourseGrades.xlsx הנתון , יש לפעול בהתאם לרשום בגיליון הוראות.
17. יש לייצר ADMIN solution - העשרה.

- (a) כתבו תהליך למעבר על כל בסיס הנתונים של המשתמש במערכת וביצוע גיבוי
- (b) כתבו תהליך למעבר על כל טבלאות ב DB הנשלח כפרמטר , וביצוע חישוב סטטיסטיקות + בניה מחודשת על כל האינדקסים.
- (c) הוסיפו תהליך למחיקת ה LOGS של המערכת אשר נוצרו לפני יותר מחודש ימים.

#### 18. משימה : ביצוע תהליך ETL ליצירת DW DB עבור Northwind

- יש לייצור סקריפט ליצירת DW DB .
  - יש לייצור את כל הטבלות ה DIM , FACT - סכמה בלבד. (ארכיטקטורה) .
  - יש לייצור טבלת DATE DIM , ולייצור פרוצדורה למילוי הטבלה .
  - יש לייצור תהליך SSIS - לטעינת נתונים ל DIM .
  - יש לייצור תהליך לטעינת נתונים ל Fact – .
  - יש לבצע את התיקונים הבאים – **derived column**
- when INT & Null Then -1
- when Date & Null Then '1900-01-01'
- when String & Null Then "Unknown String"

## Summary Project : Build WWI DW

1. יש לייצר DW עבור תהליך המכירה בלבד. מומלץ לפתח את הממדים הבאים :

Customer, city , Stock Item , Date , Employee

טבלת הFACT אמורה להראות כך :

| Column Name            | Data Type      | Allow Nulls                         |
|------------------------|----------------|-------------------------------------|
| [Sale Key]             | bigint         | <input type="checkbox"/>            |
| [City Key]             | int            | <input type="checkbox"/>            |
| [Customer Key]         | int            | <input type="checkbox"/>            |
| [Bill To Customer Key] | int            | <input type="checkbox"/>            |
| [Stock Item Key]       | int            | <input type="checkbox"/>            |
| [Invoice Date Key]     | date           | <input type="checkbox"/>            |
| [Delivery Date Key]    | date           | <input checked="" type="checkbox"/> |
| [Salesperson Key]      | int            | <input type="checkbox"/>            |
| [WWI Invoice ID]       | int            | <input type="checkbox"/>            |
| Description            | nvarchar(100)  | <input type="checkbox"/>            |
| Package                | nvarchar(50)   | <input type="checkbox"/>            |
| Quantity               | int            | <input type="checkbox"/>            |
| [Unit Price]           | decimal(18, 2) | <input type="checkbox"/>            |
| [Tax Rate]             | decimal(18, 3) | <input type="checkbox"/>            |
| [Total Excluding Tax]  | decimal(18, 2) | <input type="checkbox"/>            |
| [Tax Amount]           | decimal(18, 2) | <input type="checkbox"/>            |
| Profit                 | decimal(18, 2) | <input type="checkbox"/>            |
| [Total Including Tax]  | decimal(18, 2) | <input type="checkbox"/>            |
| [Total Dry Items]      | int            | <input type="checkbox"/>            |
| [Total Chiller Items]  | int            | <input type="checkbox"/>            |

2. יש לייצור DB : STAGE & MRR .

3. יש לייצור 3 סכמות:

a. Integration

b. Dimension

c. Fact

4. יש לייצור טבלה לניטור טעינות ל- DB בשם lineage – ובה שם טבלה, ותאריך טעינה

אחרון.(סקריפט מצורף).

5. יש לייצור פרוצדורה למילוי תאריכים (DIM DATES) החל ממועד המכירה הראשון בחברה

ועד לעוד שנתיים קדימה ברזולוציה של יום. יש לכתוב פרוצדורה שממלאת אתה טבלה

בתאריכים.

6. יש לכתוב תהליך לטעינת נתוני MRR , ותהליך לטעינת נתוני STG

7. יש לכתוב תהליך לטעינת נתוני DIM DW - יש לבצע זאת ע"י SCD

8. יש לכתוב תהליך לטעינת נתוני FACT DW

9. מבנה התהליך :

a. שליפת המועד הטעינה האחרון והכנסה למשתנה

b. שליפת נתונים מהטבלה

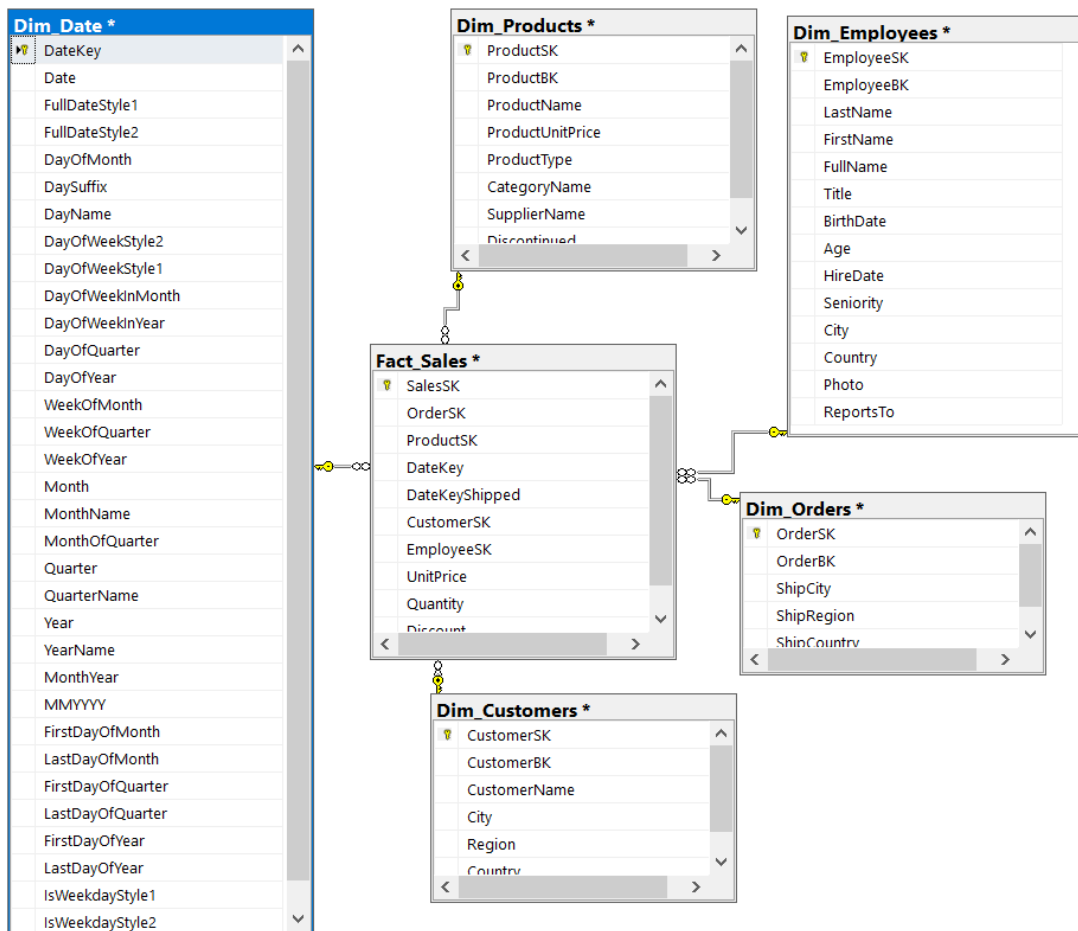
c. מניפולציה על הנתונים להשלמת ערכי NULL, והשלמת נתונים חסרים.

d. טעינה ל DW.

## **SSIS-Create Edate Data wareHouse**

1. Design and Implement a data warehouse for the "eDate" company based on the following requirements:
2. The management would like to measure sessions (count, min/max/average duration) according to the following dimensions:
  - a. Member Address (Continent \ Country \ City)
  - b. Member Gender and Sexual Preference
  - c. Member Age (Young \ Mid-Age \ Old)
  - d. Member Marital Status
  - e. Login Date (Year \ Quarter \ Month \ Day)
  - f. Session End Reason
3. The management would like to measure invitations (count, min/max/average time to response) according to the following dimensions:
  - a. Requesting Member Address (Continent \ Country \ City)
  - b. Requesting Member Gender and Sexual Preference
  - c. Requesting Member Age (Young \ Mid-Age \ Old)
  - d. Requesting Member Marital Status
  - e. Receiving Member Address (Continent \ Country \ City)
  - f. Receiving Member Gender and Sexual Preference
  - g. Receiving Member Age (Young \ Mid-Age \ Old)
  - h. Receiving Member Marital Status
  - i. Creation Date (Year \ Quarter \ Month \ Day)
  - j. Invitation Status
4. All the dimensions should change along with changes to the data source, so that each session or invitation is recorded with the dimension values at the time of creation (of the session or invitation).

# ERD for Northwind DatawareHouse



10. Which data flow source adapters can you use if you would like to read data from SQL Server?
  - A. flat file source
  - B. Raw File source
  - C. OLE DB source
  - D. ODBC source
  
11. 2. Which data flow destinations can you use if you would like to temporarily stage data to a file system? (Choose all that apply.)
  - A. OLE DB destination
  - B. Flat File destination
  - C. Raw File destination
  - D. Recordset destination
  
12. 3. Which statements are true regarding data source adapters? (Choose all that apply.)
  - A. You can change how source data is mapped to SSIS data types.
  - B. You can have only one data source adapter per data flow task.
  - C. You must always select all columns from the source adapter.
  - D. You can read data from an XML file by using SSIS.