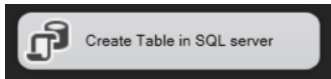


עבודה 2 - SSIS

לצורך הרצת המטלה, צריך להפעיל את packages לפי הסדר בדו"ח הבא.

Package – tables SQL

בחלק זה יצרנו את כל הטבלאות הנדרשות בSQL אליהן נזין את הנתונים מטבלאות csv שעתידות להתקבל באופן יומי. לטובת משימה זו השתמשנו בSQL task execute ויצרנו את הטבלאות הנדרשות באמצעות SQL.



נציין כי יש להריץ ראשית את package הזה על מנת ליצור את הטבלאות ורק לאחר מכן את ה package שנקרא execute.

Package – execute

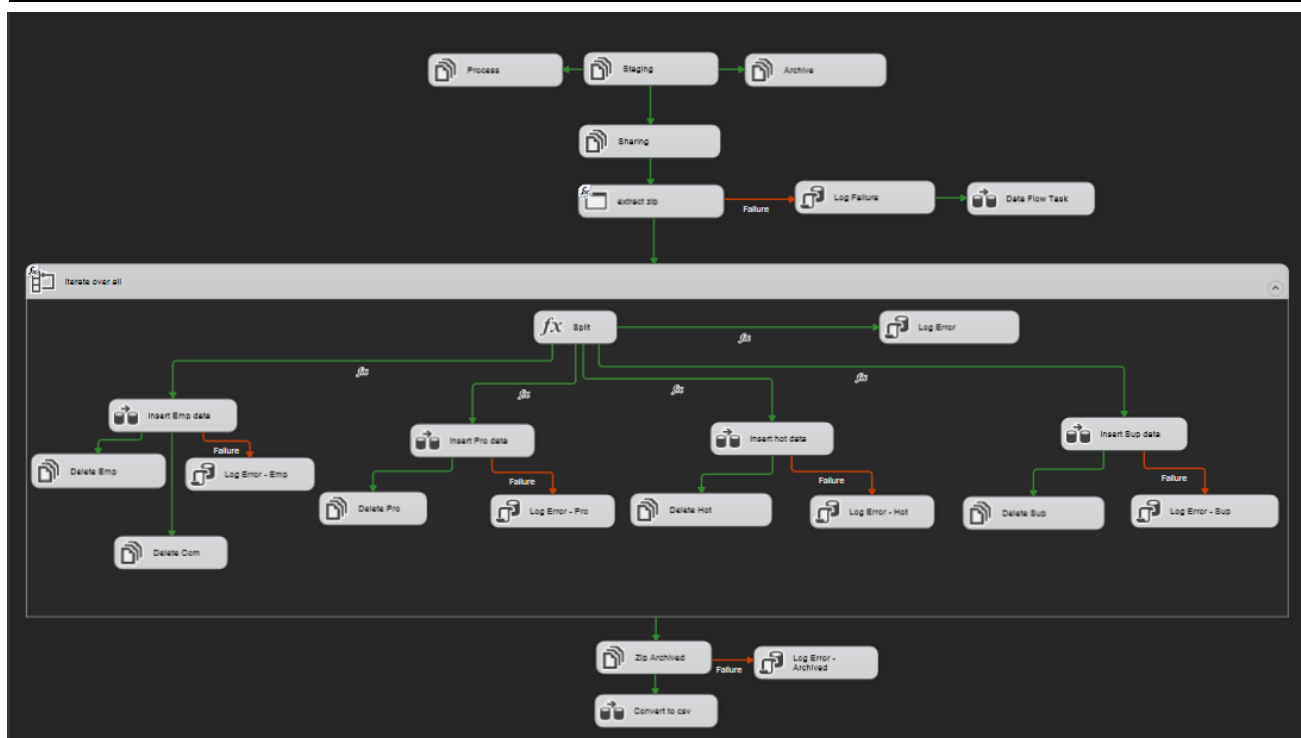
הגדרת משתנים

Name	Scope	Data type	Value	Expression
ArchivedFolder_path	excute	String	C:\Staging\ArchivedFolder	
Company_path	excute	String	C:\Staging\ProcessFolder\Company20230528.txt	@[User::ProcessFolder_path] + "\\Company" + @[Us...
Current_Date	excute	String	20230528	(DT_STR, 4, 1252) DATEPART("yy", GETDATE()) + RI...
Employee_path	excute	String	C:\Staging\ProcessFolder\Employee20230528.csv	@[User::ProcessFolder_path] + "\\Employee" + @[Us...
Errors_path	excute	String	C:\Staging\Errors.csv	@[User::Staging_path] + "\\Errors" + ".csv"
File_Cur	excute	String		
Hotel_path	excute	String	C:\Staging\ProcessFolder\Hotel20230528.csv	@[User::ProcessFolder_path] + "\\Hotel" + @[User::C...
ProcessFolder_path	excute	String	C:\Staging\ProcessFolder	
Product_path	excute	String	C:\Staging\ProcessFolder\Product20230528.csv	@[User::ProcessFolder_path] + "\\Product" + @[User::...
SharingFolder_path	excute	String	C:\Staging\SharingFolder	
Staging_path	excute	String	C:\Staging	
Supplier_path	excute	String	C:\Staging\ProcessFolder\Supplier20230528.csv	@[User::ProcessFolder_path] + "\\Supplier" + @[User::...
Zip	excute	String	C:\Staging\SharingFolder\data20230528.zip	@[User::SharingFolder_path] + "\\data" + @[User::Cu...
Zip_exe	excute	String	C:\Program Files\7-Zip\7z.exe	

יצרנו כמה סוגים של משתנים אשר יסייעו לנו להפוך את הפתרון לדינאמי ויעיל עבור המשתמש

- 1. משתנה התאריך** – יצרנו משתנה ששומר את התאריך של אותו היום באופן דינאמי ובמשתנה זה השתמשנו לטובת הגדרת שאר המשתנים עליהם נפרט בהמשך.
- 2. יצירת התיקיות** – יצרנו את התיקיות הנדרשות כך Foldername_path כאשר Foldername הוא שם התיקיה והערך הוא הנתבי לתיקיה זו במחשב. כך ששינוי הניתוב לא יפגע בתקינות התוכנית (Sharing, Process, Archived). בנוסף שמרנו ערך עבור תיקיית האב Staging, כך שגם שינוי של תיקיית האב לא תפגע בתקינות הקוד ותאפשר דינאמיות מרבית.
- 3. נתיב הקבצים** – יצרנו את נתיב הקבצים הנדרשים כך ששם הקובץ יתעדכן באופן דינאמי בהתאם לתאריך. למשל עבור קבצי ה Employee היומיים זה יראה כך - @[User::ProcessFolder_path] + "\\Employee" + @[User::Current_Date] + ".csv". ביצענו זאת עבור כל קבצי ה csv הנדרשים וכן עבור קובץ ה txt בהתאמה.
- 4. משתנה File_Cur** – משתנה אליו נשמור בכל איטרציה ב for each loop container את המסמך הנוכחי עליו נעבוד.
- 5. Zip_exe** – משתנה אליו נשמור את קובץ ה exe שבאמצעותו נחלץ את תיקיית הזיפ.
- 6. Zip** – ניתוב שבו שמור קובץ הזיפ היומי, אותו נחלץ ועליו נבצע את התהליך הנדרש.

המבנה הכללי של הפתרון שיצרנו



יצירת התיקיות



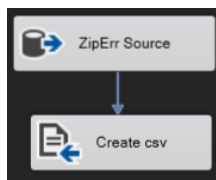
באמצעות File System Task יצרנו את התיקיות הנדרשות. ראשית תיקיית Staging ולאחר מכן את התיקיות Process, Sharing, Archived.

כמו שציינו קודם השתמשנו במשתנים עבור הניתוב אליו בו נפתח את כל אחת מהתיקיות כך שניתן לשנות זאת בלי לפגוע בתהליך.

חילוף קובץ הזיפ



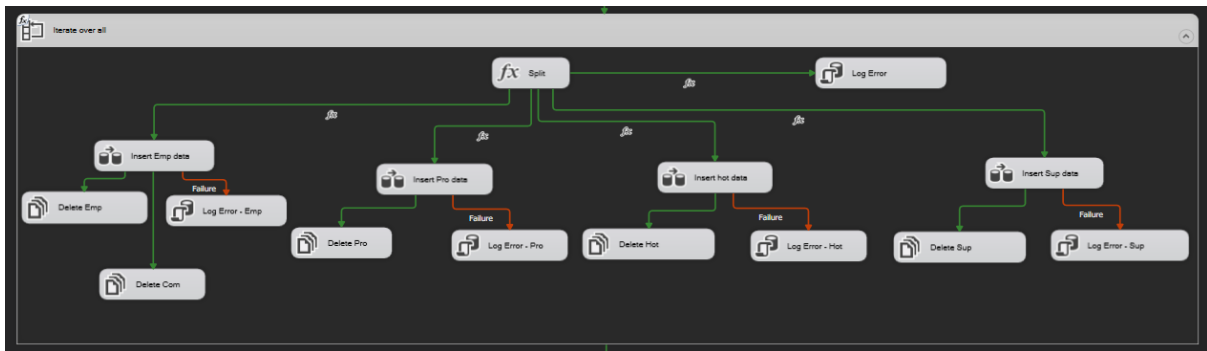
לטובת חילוף קובץ הזיפ השתמשנו בתוכנת 7-Zip\7z.exe. באמצעות Execute process task ביצענו את החילוף של התיקיה מהזיפ.



לאורך העבודה השתמשנו בExecute SQL task על מנת להתמודד עם שגיאות שעשויות לצוף לאורך התוכנית. למשל במקרה זה אי הצלחה בחילוף הזיפ/ קובץ הזיפ לא קיים בתיקיה. לאחר מכן השתמשנו בData flow על מנת להמיר את הנתונים הנמצאים בטבלת Errors בשרת הSQL לקובץ CSV.

בתוך Data flow השתמשנו בOLEDB source ובFlat file destination.

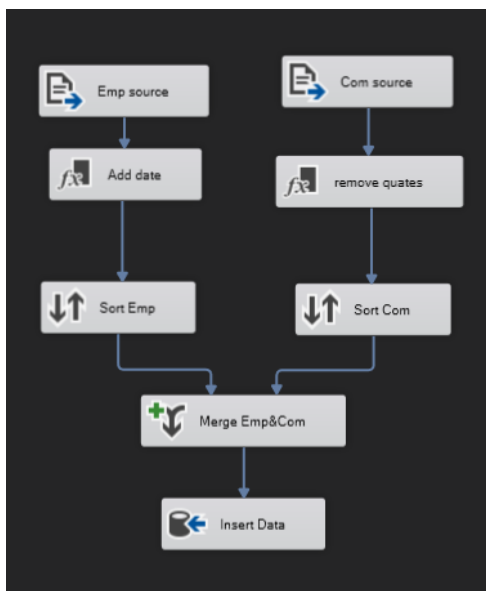
יצירת loop container for each



- נציין כי עבור כל connection manager יצרנו משתנה דינאמי של connection string באמצעות שימוש ב-variables וב-Expressions כך שהחיבורים ישמרו בצורה דינאמית ויאפשרו גמישות מרבית בין הימים ובין תיקיות.

containern יעבור על כל קובץ ויבצע את הפעולות הבאות:

1. באמצעות Expression Task ו-Precedence Constraint הקונטיינר ייגש לקובץ קובץ יבדוק אם הוא בשם הנכון והתאריך הנכון ובמידה וכן ימשיך לשאר התוכנה אחרת יכניס שגיאה לטבלת השגיאות.
2. עבור כל אחת מהטבלאות ביצענו את התהליך הבא ב-Data flow:



- א. גישה לטבלת csv
 - ב. הוספת עמודת תאריך
 - ג. מחיקת גרשיים במידת הצורך
 - ד. ביצוע merge ו-sort במידת הצורך
 - ה. כתיבה לשרת ה-SQL לטבלה המתאימה
3. מחיקת הקובץ מתיקית Process באמצעות File System Task לאחר ההכנה לשרת ה-SQL.
 4. כתיבת שגיאה בטבלת השגיאות במידת הצורך.
- להלן דוגמא עבור ההכנסה של הנתונים לטבלת Employee כפי שקורה ב-Date flow בשם 'Insert Emp data'.
- נציין כי פעלנו בצורה דומה בהתאם להוראות עבור שאר בטבלאות.

מחיקת הגרשיים בוצעה עבור כל הטבלאות באופן הבא:
נוסיף כי ב-derived column שכתוב 'Add date' מתבצע גם מחיקת הגרשיים.

Derived Column Name	Derived Column	Expression	Data Type	Length
DateInsert	<add as new column>	GETDATE()	database timestamp ...	
Emp_No	Replace 'Emp_No'	REPLACE(Emp_No,"'", "")	string [DT_STR]	5
Name	Replace 'Name'	REPLACE(Name,"'", "")	string [DT_STR]	5
DOB	Replace 'DOB'	REPLACE(DOB,"'", "")	string [DT_STR]	5
Gender	Replace 'Gender'	REPLACE(Gender,"'", "")	string [DT_STR]	5
Salary	Replace 'Salary'	REPLACE(Salary,"'", "")	string [DT_STR]	5
GeographyID	Replace 'Geographyl...	REPLACE(GeographyID,"'", "")	string [DT_STR]	5

העברת התיקיה לארכיון



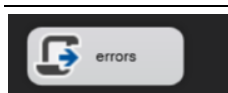
לאחר ביצוע התהליך על התיקיה היומית, נבצע העברה של תיקיית הזיפ
 Archived לSharing.
 במידה וישנה בעיה בהעברה נכתוב זאת בטבלת השגיאות.

המרה של טבלת השגורות בשרת הSQL לקובץ CSV



כפי שציינו קודם, כל השגיאות בתהליך נשמרו בטבלת Errors וכעת נבצע ייצור של הטבלה
 לקובץ CSV באמצעות Date flow.

הקפצת התראה למשתמש בעת שגיאה



שימוש ב - event handle תחת onError, יצרנו script task על מנת להקפיץ הודעה למסך
 המשתמש בעת שגיאה מעבר להוספת הרשומה בטבלת השגיאות.

תוצאות הרצה – תיקיה עם תאריך נכון

ניתן לראות כי כאשר הרצנו את התוכנה עם התאריכים הנכונים, המערכת עובדת בשלמותה והנתונים הוזנו
 בטבלאות בשרת הSQL.

טבלת Employee

	Emp_No	Name	DOB	Gender	Salary	LocationID	Company	City	State	Email	DateTime
1	100	Mahesh	12/1/1965	Male	30000	1	ABC Pvt. Ltd.	Noida	UP	abc.pqr@gmail.com	2023-05-29 23:00:19.833
2	105	Sunita	12/1/1958	Female	60000	1	ABC Pvt. Ltd.	Noida	UP	abc.pqr@gmail.com	2023-05-29 23:00:19.833
3	102	Anurag	12/21/1969	Male	38000	2	PQR Pvt. Ltd.	Dehradun	UK	pqr220@mycompany.com	2023-05-29 23:00:19.833
4	103	AJAY	12/1/1961	Male	40000	3	TTT Solutons Ltd.	Shimla	HP	info@ttsolutions.com	2023-05-29 23:00:19.833
5	104	Neha	12/1/1971	Female	19000	3	TTT Solutons Ltd.	Shimla	HP	info@ttsolutions.com	2023-05-29 23:00:19.833
6	107	Anil	12/29/1963	Male	45000	3	TTT Solutons Ltd.	Shimla	HP	info@ttsolutions.com	2023-05-29 23:00:19.833
7	101	Suresh	11/14/1963	Male	34000	4	AAA Solutons Ltd.	Bhopal	MP	info@aaasolutions.com	2023-05-29 23:00:19.833
8	106	Amar	12/12/1978	Male	45000	4	AAA Solutons Ltd.	Bhopal	MP	info@aaasolutions.com	2023-05-29 23:00:19.833

טבלת Hotel

	HotelID	HotelName	HotelCategory	City	Country	Address	Coordinates	DateTime
1	1	Good Night Hotel and Residence	4 Star	Toronto	Canada	1 ABC City Center	43.623409,-79.368683	2023-05-29 23:00:20.590
2	2	Pink Hotel Toronto	3 Star	Toronto	Canada	999 Kingsway Road	43.623409,-79.368683	2023-05-29 23:00:20.590
3	3	Tourist In	2 Star	Chicago	USA	ZZZ Broadway	41.881832,-87.623177	2023-05-29 23:00:20.590
4	4	City Center Hotel	4 Star	New York	USA	SSS Avenue	40.73061,-73.935242	2023-05-29 23:00:20.590
5	5	Venus In	3 Star	Adelaide	Australia	XXX Lake Avenue	-34.92123,138.599503	2023-05-29 23:00:20.590
6	6	Stoppage Inn Express & Suits	3 Star	Reno	USA	11 Knowledge Villa	39.530895,-119.814972	2023-05-29 23:00:20.590
7	7	Beach Hotel	4 Star	Miami	USA	12 BroadWay Drive	25.778135,-80.1791	2023-05-29 23:00:20.590
8	8	Traditional Hut Hotel	3 Star	Munnar	India	ZZZ Village Drive	10.089167,7.059723	2023-05-29 23:00:20.590
9	9	Big Ocean Hotel	4 Star	Miami	USA	AAA Beach Street	25.778135,-80.1791	2023-05-29 23:00:20.590
10	10	Entertainment Hotel	5 Star	Los Angeles	USA	BBB WestWood	34.052235,-118.243683	2023-05-29 23:00:20.590

טבלת Product

	ProductID	PName	SID	DateTime
1	1	Keyboard	1	2023-05-29 23:00:20.887
2	2	Monitor	2	2023-05-29 23:00:20.887
3	3	AC	2	2023-05-29 23:00:20.887
4	4	Bikes	3	2023-05-29 23:00:20.887
5	5	Bulb	3	2023-05-29 23:00:20.887
6	6	Processor	7	2023-05-29 23:00:20.887

טבלת Supplier

	SupplierID	CompanyName	DateTime
1	1	TVSGold	2023-05-29 23:00:21.153
2	2	LG	2023-05-29 23:00:21.153
3	3	Bajaj	2023-05-29 23:00:21.153
4	4	Maruti	2023-05-29 23:00:21.153
5	5	Orient	2023-05-29 23:00:21.153
6	6	Tata	2023-05-29 23:00:21.153

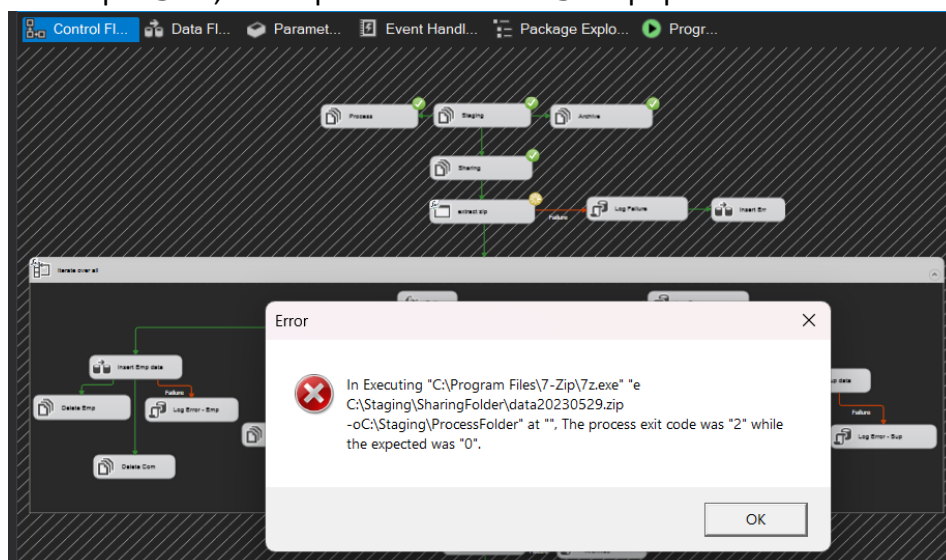
טבלת Errors

FileName	ErrorName	DateTime
----------	-----------	----------

ניתן לראות כי טבלאת השגיאות ריקה והתהליך רץ במלואו.

תוצאות הרצה – תיקיה עם תאריך שגוי

בעת הרצת התוכנית על קובץ זיפ שאינו מכיל את התאריך הנבדק, ראשית קופצת למשתמש הודעת שגיאה:



לאחר שנלחץ OK על הודעת השגיאה, השגיאה תהיה מתועדת בטבלת השגיאות בשרת הSQL:

	FileName	ErrorName	DateTime
1	C:\Staging\SharingFolder	The folder does not contain the Zip file.	2023-05-29 23:03:32.447

ובן מיוצאת לקובץ CSV:

FileName	ErrorName	DateTime
C:\Staging\SharingFolder	The folder does not contain the Zip file.	29/05/2023 23:03