

מטלה 3 - SSIS

לצורך הרצת המטלה שלנו, צריך להריץ את החבילות לפי סדר הופעתן בדו"ח זה.

תחילה, הגדרנו params עבור כל החבילות:

Name	Data type	Value
cities	String	C:\Data\Cities.csv
cities_file	String	Cities.csv
data	String	C:\Data
DB	String	DB_SSIS
employee_file	String	Employees.csv
employees	String	C:\Data\Employees.csv
path_to_download_files	String	C:\Users\MG\Desktop\DS\Semester 6\SSIS\HW3\
SERVER	String	DESKTOP-T5AASC7\SQLEXPRESS

Cities - יכיל את הנתיב לקובץ הערים תחת הניתוב המתאים לתיקיה data כפי שהתבקשנו.

Employees - יכיל את הנתיב לקובץ העובדים תחת הניתוב המתאים לתיקיה data כפי שהתבקשנו.

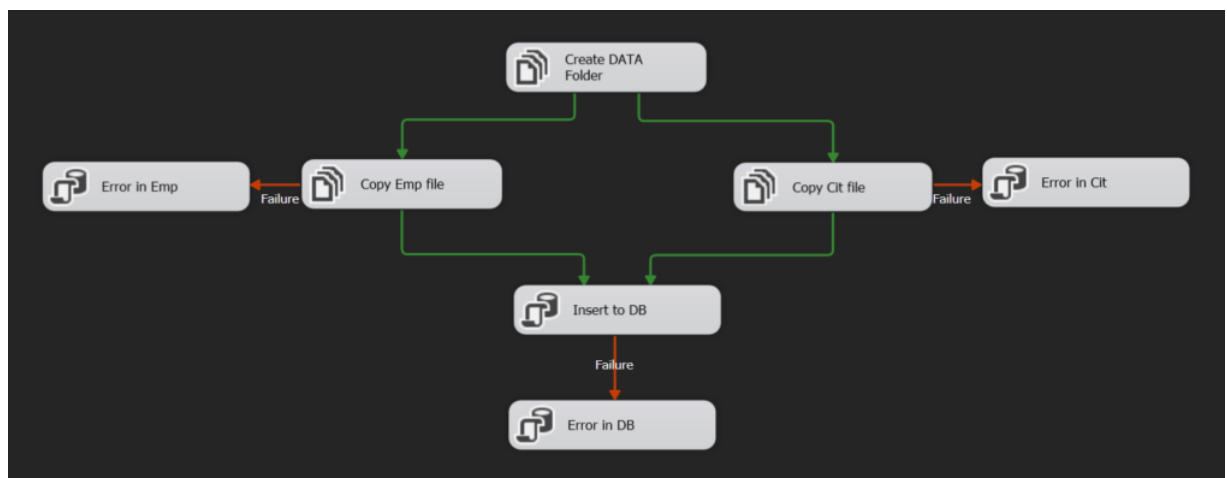
Path_to_download_files - יכיל את הנתיב במחשב שבו שמורים הקבצים שנתבקשנו להוריד.

Cities_file, employees_file - יכילו את שמות קבצי csv שנצטרך להעתיק לתיקיה data.

DB - מכיל את שם הdb.

SERVER - מכיל את שם servern במחשב.

Package 1 - Create Tables

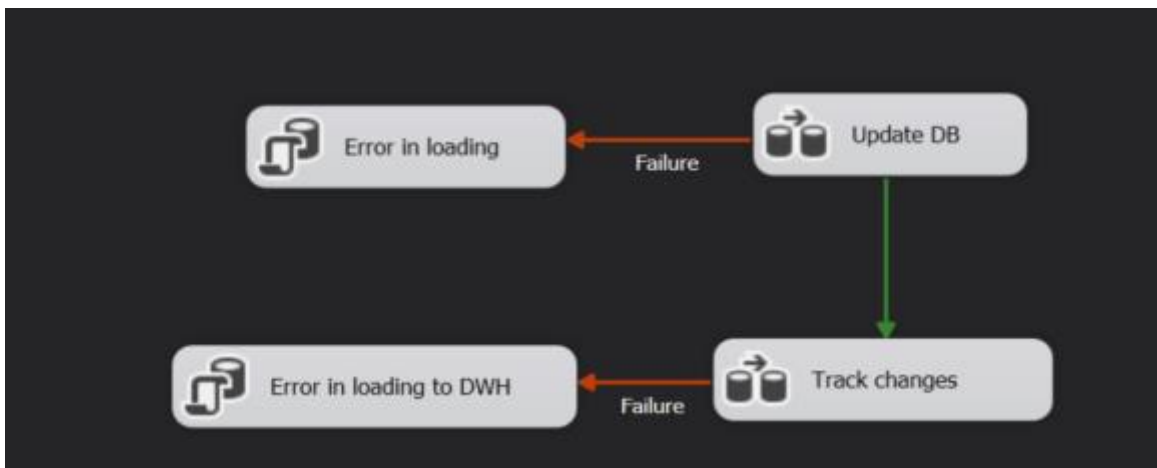


בחבילה זו:

- יצרנו את התיקייה DATA בהתאם להוראות, כלומר בנתיב: C:\Data
- העתקנו את הקבצים מהמקום בו הם שמורים במחשב אל הניתוב שהתבקשנו בעבודה - C:\Data.
- יצרנו טבלאות בdb כך שהעמודות בטבלאות תואמות לעמודות הרלוונטיות במקרה הצורך. את יצירת הטבלאות עשינו רק אם הטבלה לא קיימת עדיין בdb.
- בכל שלב ייצאנו את השגיאות לטבלת errors במידה והיו. לדוגמה אם לא נמצא קובץ הערים בתיקיה, תתווסף לטבלת השגיאות רשומה עם הפירוט הבא:

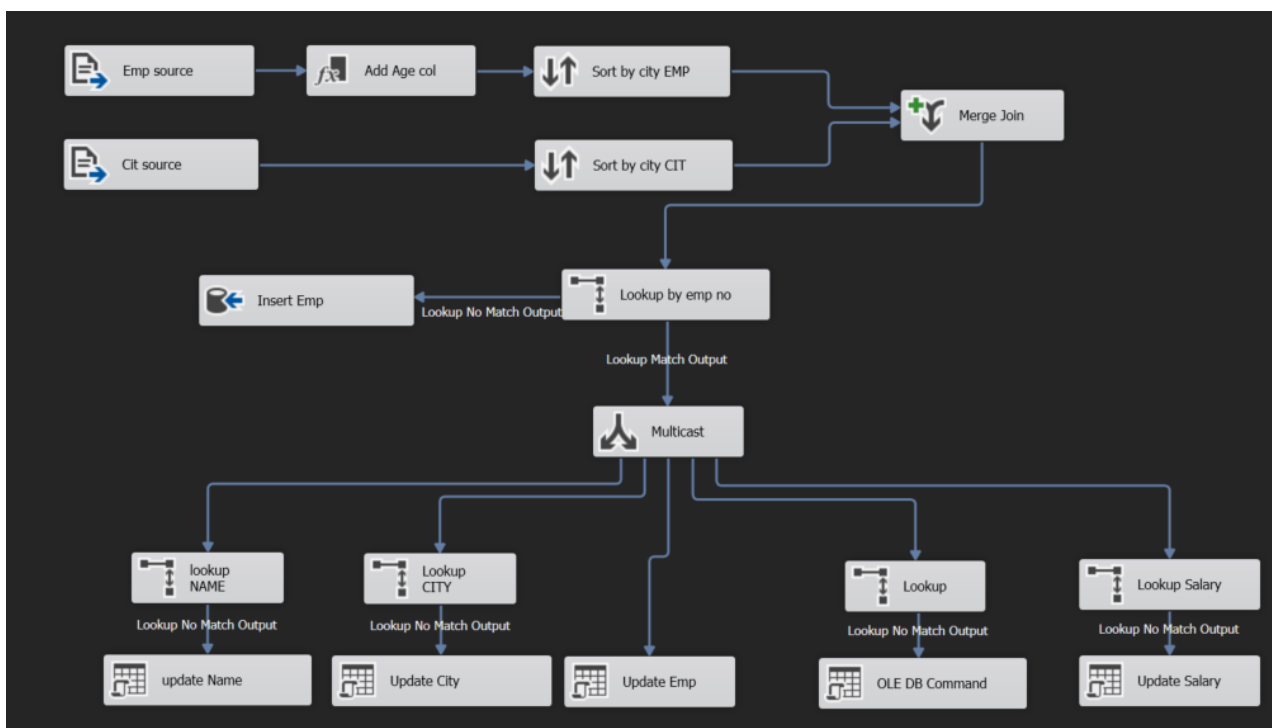
4	Cities.csv	File not found	2023-06-13
---	------------	----------------	------------

Package 2 – Insert and Update DB



נביט data flow של כל אחד מהם בנפרד ונפרט על התוכן, גם כאן הוספנו תיעוד של שגיאות במקרה והמשימה לא צלחה.

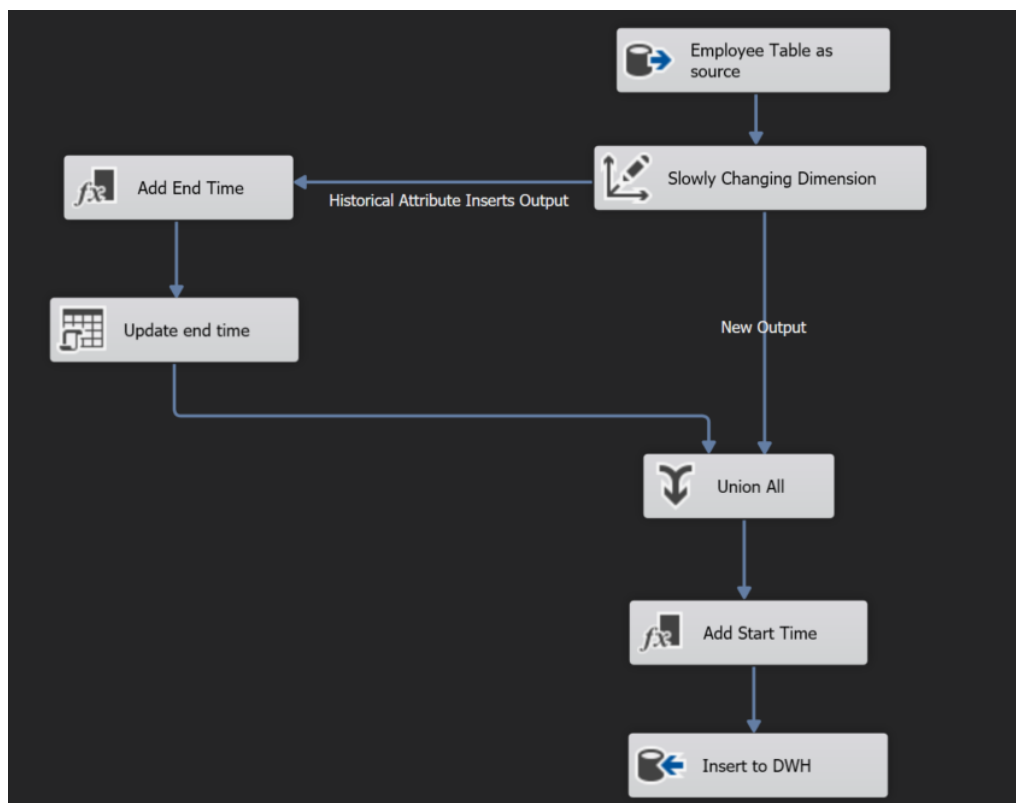
:Update DB



1. הגדרנו את המקור של הנתונים שאיתם נעבוד, במקרה הזה אלו שני קבצי CSV.
2. לנתוני העובדים הוספנו עמודה - Age, את החישוב שלה ביצענו כך: $\text{YEAR}(\text{GETDATE}()) - \text{YEAR}(\text{DOB})$.
3. לצורך מיזוג בין 2 הטבלאות, מיינו את שתיהן לפי עמודת ה-city, ולאחר מכן ביצענו מיזוג מסוג inner join כך שעבור כל עובד נוספה שם העיר בה הוא גר, והורדנו את העמודה שמכילה את מס' העיר. את המיזוג ביצענו לפי מס' העיר.
4. ביצענו lookup לפי מס' העובד בטבלת Employees, נשים לב כי מס' העובד זה מס' ייחודי עבור כל עובד ולכן התייחסנו אליו כמפתח.
 - א. עבור עובד שלא קיים בטבלה, נוסיף אותו בשלמותו לטבלה.
 - ב. עבור עובד שנמצאה התאמה למס' שלו בטבלה, נבצע lookup עבור כל ערך שיש לעובד בטבלה שעשוי להשתנות, ובמידה ולא נמצא התאמה - נעדכן את טבלת Employees בערך החדש.

ג. עבור כל שינוי שבוצע אצל עובד הקיים בטבלת העובדים, נוסף רשומה שמתעדת את השינוי בטבלת Log.

:Track Changes

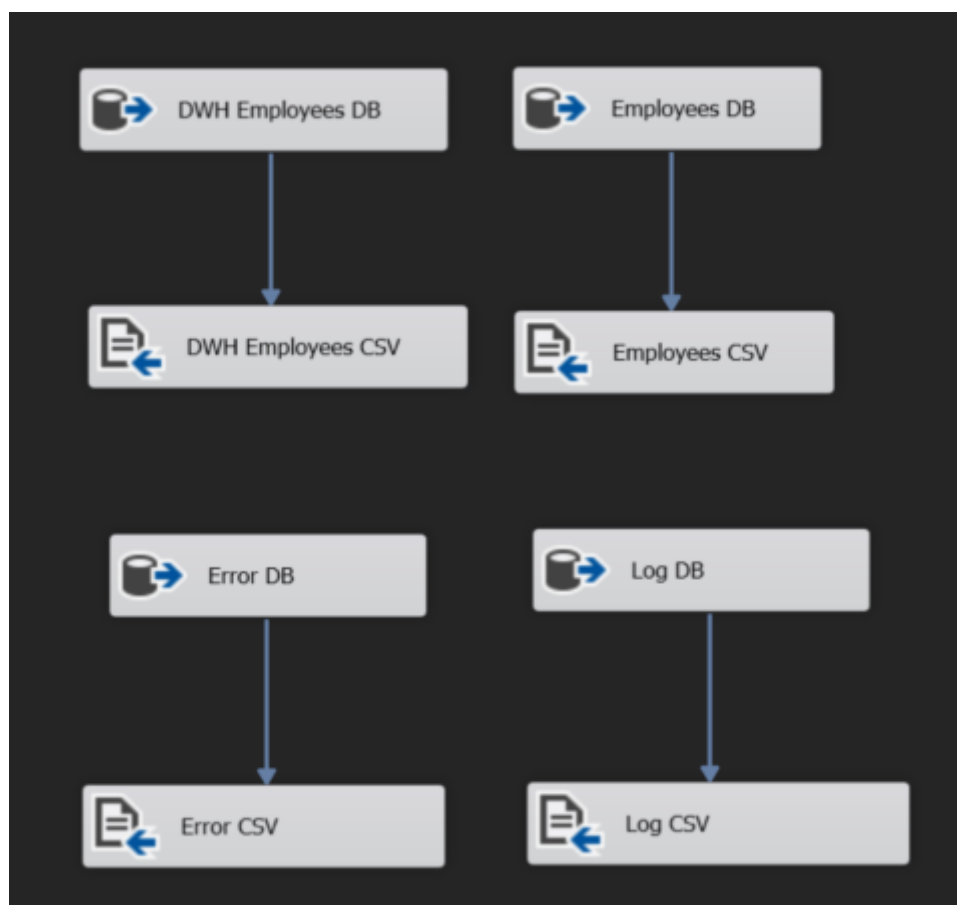


1. הגדרנו את הטבלה employees מהdb כטבלת המקור שלנו.
2. ביצענו בדיקה של SCD, כך שעבור כל רשומה בטבלת המקור מתבצעת השוואה עם הנתונים הקיימים בdwh וכאשר נמצאת רשומה שעברה עדכון וקיים בה ערך שונה מהאחרון שעודכן בdwh נפעל כך:
 - א. את הערך האחרון ששמרנו שלו אין תאריך סיום, כלומר קיים null במקום – נעדכן ונגדיר את זמן הסיום להיות הזמן הנוכחי שבו מתבצעת הבדיקה.
 - ב. ניצור רשומה חדשה שמכילה את הנתונים העדכניים כפי שהם מופיעים בטבלת המקור, עם זמן התחלה שהוא הזמן כרגע וזמן סיום לא מוגדר – null.
3. את השינויים שביצענו נכניס לתוך טבלת DWH Employees בdb שלנו.

Package 3 – Create Csv files from tables



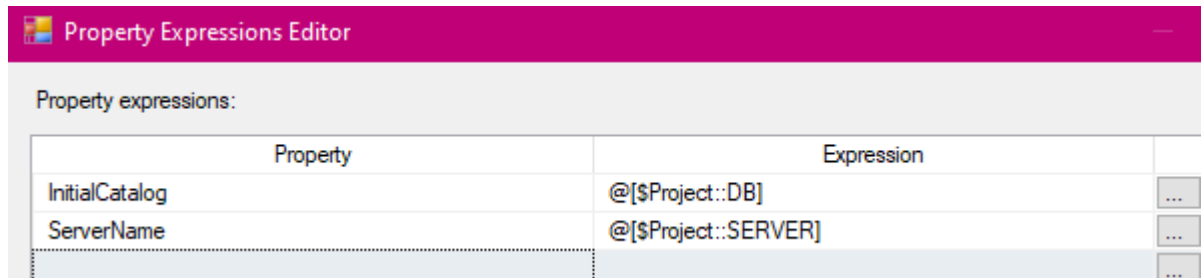
בחבילה זו אנחנו יוצרות קבצי csv אם אינם קיימים עדיין ביעד, ומזינות לתוכם את הטבלאות מהDB כפי שנתבקשנו בהוראות. נביט בdata flow:



נוכל לראות כי עבור כל טבלה שנתבקשנו לייצא, יצרנו גישה למקור בserver שלנו ולאחר מכן ייצאנו את התוכן אל קובץ csv בהתאם לנתוני הטבלה.

:Connection mangers

עבור החיבורים הרלוונטיים הוספנו expression שמשמש בפרמטרים שהגדרנו בתחילה המטלה, כך שהחיבור יהיה דינאמי ונוח לשימוש.



דוגמאות להרצות ותוצאות:

תוצאות להרצה של הקבצים כפי שקיבלנו אותם:

טבלת Employee בSQL server:

	Emp_No	Name	DOB	Gender	Salary	Age	StartDate	EndDate	City	UpdateDateTime
1	100	Malesh	1965-12-01	Male	35000	58	2023-06-14 14:13:52.000	NULL	Mumbai	2023-06-14 14:13:51.470
2	101	Sureah	1963-11-14	Male	40000	60	2023-06-14 14:13:52.000	NULL	Mumbai	2023-06-14 14:13:51.470
3	102	Rajat	1969-12-21	Male	23000	54	2023-06-14 14:13:52.000	NULL	Delhi	2023-06-14 14:13:51.470
4	103	Kalpana	1961-12-01	Female	40000	62	2023-06-14 14:13:52.000	NULL	Bangalore	2023-06-14 14:13:51.470
5	104	Neha	1971-12-01	Female	19000	52	2023-06-14 14:13:52.000	NULL	Bangalore	2023-06-14 14:13:51.470

טבלת DWH_Employee בSQL server:

	Emp_No	Name	DOB	Gender	Salary	Age	City	UpdateDateTime
1	100	Malesh	1965-12-01	Male	35000	58	Mumbai	2023-06-14 14:13:51.470
2	101	Sureah	1963-11-14	Male	40000	60	Mumbai	2023-06-14 14:13:51.470
3	102	Rajat	1969-12-21	Male	23000	54	Delhi	2023-06-14 14:13:51.470
4	103	Kalpana	1961-12-01	Female	40000	62	Bangalore	2023-06-14 14:13:51.470
5	104	Neha	1971-12-01	Female	19000	52	Bangalore	2023-06-14 14:13:51.470

טבלת Employee_OLTP בקובץ csv:

UpdateDa	City	Age	Salary	Gender	DOB	Name	Emp_No
13:51.5	Mumbai	58	35000	Male	01/12/1965	Malesh	100
13:51.5	Mumbai	60	40000	Male	14/11/1963	Sureah	101
13:51.5	Delhi	54	23000	Male	21/12/1969	Rajat	102
13:51.5	Bangalore	62	40000	Female	01/12/1961	Kalpana	103
13:51.5	Bangalore	52	19000	Female	01/12/1971	Neha	104

טבלת DWH_Employee בSQL server:

EndDate	StartDate	UpdateDa	City	Age	Salary	Gender	DOB	Name	Emp_No
	14/06/2023 14:13	13:51.5	Mumbai	58	35000	Male	01/12/1965	Malesh	100
	14/06/2023 14:13	13:51.5	Mumbai	60	40000	Male	14/11/1963	Sureah	101
	14/06/2023 14:13	13:51.5	Delhi	54	23000	Male	21/12/1969	Rajat	102
	14/06/2023 14:13	13:51.5	Bangalore	62	40000	Female	01/12/1961	Kalpana	103
	14/06/2023 14:13	13:51.5	Bangalore	52	19000	Female	01/12/1971	Neha	104

כעת, שינינו ערך אחד בכל עמודה של נתונים בטבלת העובדים, וטענו אותה לdb שלנו.

טבלת העובדים לפני(המקורית):

Emp_No	Name	DOB	Gender	Salary	City
100	Malesh	01/12/1965	Male	35000	1
101	Sureah	14/11/1963	Male	40000	1
102	Rajat	21/12/1969	Male	23000	2
103	Kalpana	01/12/1961	Female	40000	3
104	Neha	01/12/1971	Female	19000	3

טבלת העובדים אחרי שינויים(השינויים מסומנים בצהוב):

Emp_No	Name	DOB	Gender	Salary	City
100	Maleshak	01/12/1965	Male	35000	1
101	Sureah	14/11/1963	Female	40000	1
102	Rajat	21/12/1969	Male	230000	2
103	Kalpana	01/12/1961	Female	40000	5
104	Neha	01/12/1971	Female	19000	3

לאחר השינוי:

טבלת Employee בSQL server:

	Emp_No	Name	DOB	Gender	Salary	Age	City	UpdateDateTime
1	100	Maleshak	1965-12-01	Male	35000	58	Mumbai	2023-06-14 14:18:22.490
2	101	Sureah	1963-11-14	Female	40000	60	Mumbai	2023-06-14 14:18:27.507
3	102	Rajat	1969-12-21	Male	230000	54	Delhi	2023-06-14 14:18:32.510
4	103	Kalpana	1961-12-01	Female	40000	62	Ahmedabad	2023-06-14 14:18:42.533
5	104	Neha	1971-12-01	Female	19000	52	Bangalore	2023-06-14 14:18:37.520

טבלת DWH_Employee בSQL server:

	Emp_No	Name	DOB	Gender	Salary	Age	StartDate	EndDate	City	UpdateDateTime
1	100	Malesh	1965-12-01	Male	35000	58	2023-06-14 14:13:52.000	2023-06-14 14:18:42.000	Mumbai	2023-06-14 14:13:51.470
2	101	Sureah	1963-11-14	Male	40000	60	2023-06-14 14:13:52.000	2023-06-14 14:18:42.000	Mumbai	2023-06-14 14:13:51.470
3	102	Rajat	1969-12-21	Male	23000	54	2023-06-14 14:13:52.000	2023-06-14 14:18:42.000	Delhi	2023-06-14 14:13:51.470
4	103	Kalpana	1961-12-01	Female	40000	62	2023-06-14 14:13:52.000	2023-06-14 14:18:42.000	Bangalore	2023-06-14 14:13:51.470
5	104	Neha	1971-12-01	Female	19000	52	2023-06-14 14:13:52.000	NULL	Bangalore	2023-06-14 14:13:51.470
6	100	Maleshak	1965-12-01	Male	35000	58	2023-06-14 14:18:42.000	NULL	Mumbai	2023-06-14 14:18:22.490
7	101	Sureah	1963-11-14	Female	40000	60	2023-06-14 14:18:42.000	NULL	Mumbai	2023-06-14 14:18:27.507
8	102	Rajat	1969-12-21	Male	230000	54	2023-06-14 14:18:42.000	NULL	Delhi	2023-06-14 14:18:32.510
9	103	Kalpana	1961-12-01	Female	40000	62	2023-06-14 14:18:42.000	NULL	Ahmedabad	2023-06-14 14:18:42.533

טבלת Log בSQL server:

	EmployeeID	FieldName	OldValue	NewValue	UpdateDateTime
1	102	Salary	23000	230000	2023-06-14 17:49:18.430
2	100	Name	Malesh	Maleshak	2023-06-14 17:49:18.443
3	103	City	Bangalore	Ahmedabad	2023-06-14 17:49:18.450
4	101	Gender	Male	Female	2023-06-14 17:49:18.457

כעת, ביצענו הוספת עובד לטבלת העובדים ששינינו קודם, עובד מס' 105:

City	Salary	Gender	DOB	Name	Emp_No
1	35000	Male	01/12/1965	Maleshak	100
1	40000	Female	14/11/1963	Sureah	101
2	230000	Male	21/12/1969	Rajat	102
5	40000	Female	01/12/1961	Kalpana	103
3	19000	Female	01/12/1971	Neha	104
10	6000	Female	02/12/1971	Nora	105

לאחר השינוי:

טבלת Employee בSQL server:

	Emp_No	Name	DOB	Gender	Salary	Age	City	UpdateDateTime
1	100	Maleshak	1965-12-01	Male	35000	58	Mumbai	2023-06-14 17:52:19.690
2	101	Sureah	1963-11-14	Female	40000	60	Mumbai	2023-06-14 17:52:24.700
3	102	Rajat	1969-12-21	Male	230000	54	Delhi	2023-06-14 17:49:32.953
4	103	Kalpana	1961-12-01	Female	40000	62	Ahmedabad	2023-06-14 17:49:42.983
5	104	Neha	1971-12-01	Female	19000	52	Bangalore	2023-06-14 17:49:37.957
6	105	Nora	1971-12-02	Female	6000	52	Jaipur	2023-06-14 17:52:14.613

טבלת DWH_Employee בSQL server:

	Emp_No	Name	DOB	Gender	Salary	Age	StartDate	EndDate	City	UpdateDateTime
1	100	Malesh	1965-12-01	Male	35000	58	2023-06-14 17:47:47.000	2023-06-14 17:49:43.000	Mumbai	2023-06-14 17:47:46.833
2	101	Sureah	1963-11-14	Male	40000	60	2023-06-14 17:47:47.000	2023-06-14 17:49:43.000	Mumbai	2023-06-14 17:47:46.833
3	102	Rajat	1969-12-21	Male	23000	54	2023-06-14 17:47:47.000	2023-06-14 17:49:43.000	Delhi	2023-06-14 17:47:46.833
4	103	Kalpana	1961-12-01	Female	40000	62	2023-06-14 17:47:47.000	2023-06-14 17:49:43.000	Bangalore	2023-06-14 17:47:46.833
5	104	Neha	1971-12-01	Female	19000	52	2023-06-14 17:47:47.000	NULL	Bangalore	2023-06-14 17:47:46.833
6	100	Maleshak	1965-12-01	Male	35000	58	2023-06-14 17:49:43.000	NULL	Mumbai	2023-06-14 17:49:22.930
7	101	Sureah	1963-11-14	Female	40000	60	2023-06-14 17:49:43.000	NULL	Mumbai	2023-06-14 17:49:27.940
8	102	Rajat	1969-12-21	Male	230000	54	2023-06-14 17:49:43.000	NULL	Delhi	2023-06-14 17:49:32.953
9	103	Kalpana	1961-12-01	Female	40000	62	2023-06-14 17:49:43.000	NULL	Ahmedabad	2023-06-14 17:49:42.983

טבלת Employee_OLTP בקובץ csv:

	UpdateDa	City	Age	Salary	Gender	DOB	Name	Emp_No
	52:19.7	Mumbai	58	35000	Male	01/12/1965	Maleshak	100
	52:24.7	Mumbai	60	40000	Female	14/11/1963	Sureah	101
	52:29.7	Delhi	54	230000	Male	21/12/1969	Rajat	102
	52:39.7	Ahmedaba	62	40000	Female	01/12/1961	Kalpana	103
	52:34.7	Bangalore	52	19000	Female	01/12/1971	Neha	104
	52:14.6	Jaipur	52	6000	Female	02/12/1971	Nora	105

טבלת DWH_Employee בSQL server:

EndDate	StartDate	UpdateDa	City	Age	Salary	Gender	DOB	Name	Emp_No
14/06/2023 17:49	14/06/2023 17:47	47:46.8	Mumbai	58	35000	Male	01/12/1965	Malesh	100
14/06/2023 17:49	14/06/2023 17:47	47:46.8	Mumbai	60	40000	Male	14/11/1963	Sureah	101
14/06/2023 17:49	14/06/2023 17:47	47:46.8	Delhi	54	23000	Male	21/12/1969	Rajat	102
14/06/2023 17:49	14/06/2023 17:47	47:46.8	Bangalore	62	40000	Female	01/12/1961	Kalpana	103
	14/06/2023 17:47	47:46.8	Bangalore	52	19000	Female	01/12/1971	Neha	104
	14/06/2023 17:49	49:22.9	Mumbai	58	35000	Male	01/12/1965	Maleshak	100
	14/06/2023 17:49	49:27.9	Mumbai	60	40000	Female	14/11/1963	Sureah	101
	14/06/2023 17:49	49:33.0	Delhi	54	230000	Male	21/12/1969	Rajat	102
	14/06/2023 17:49	49:43.0	Ahmedaba	62	40000	Female	01/12/1961	Kalpana	103
	14/06/2023 17:52	52:14.6	Jaipur	52	6000	Female	02/12/1971	Nora	105

טבלת Log לאחר הוספת עובד – אינה משתנה:

E	D	C	B	A
UpdateDa	NewValue	OldValue	FieldName	Employeee
49:18.4	230000	23000	Salary	102
49:18.4	Maleshak	Malesh	Name	100
49:18.4	Ahmedaba	Bangalore	City	103
49:18.5	Female	Male	Gender	101

כעת עשינו שינויים מרובים באותה הרשומה של עובד מסוים בטבלת העובדים וטענו אותה לdb שלנו(השינוי בצהוב):

City	Salary	Gender	DOB	Name	Emp_No
6	30000	Male	01/12/1965	Maleshak	100
1	40000	Female	14/11/1963	Sureah	101
2	230000	Male	21/12/1969	Rajat	102
5	40000	Female	01/12/1961	Kalpana	103
3	19000	Female	01/12/1971	Neha	104
10	6000	Female	02/12/1971	Nora	105

טבלת Log לאחר השינוי הכפול – ניתן לראות את שורות 5 ו6.

EmployeeID	FieldName	OldValue	NewValue	UpdateDateTime
102	Salary	23000	230000	2023-06-14 17:49:18.430
100	Name	Malesh	Maleshak	2023-06-14 17:49:18.443
103	City	Bangalore	Ahmedabad	2023-06-14 17:49:18.450
101	Gender	Male	Female	2023-06-14 17:49:18.457
100	Salary	35000	30000	2023-06-14 20:06:37.740
100	City	Mumbai	Chennai	2023-06-14 20:06:37.760

יצרנו בכוונה מצבים שמביאים לשגיאות, כך נראית טבלת Error לאחר מכן:

FileName	ErrorDescription	Date
csv file	File not found or could not be loaded	2023-06-14
Cities.csv	File not found	2023-06-14
Employees.csv	File not found	2023-06-14
Export CSV	Failed exporting DB to csv files	2023-06-14