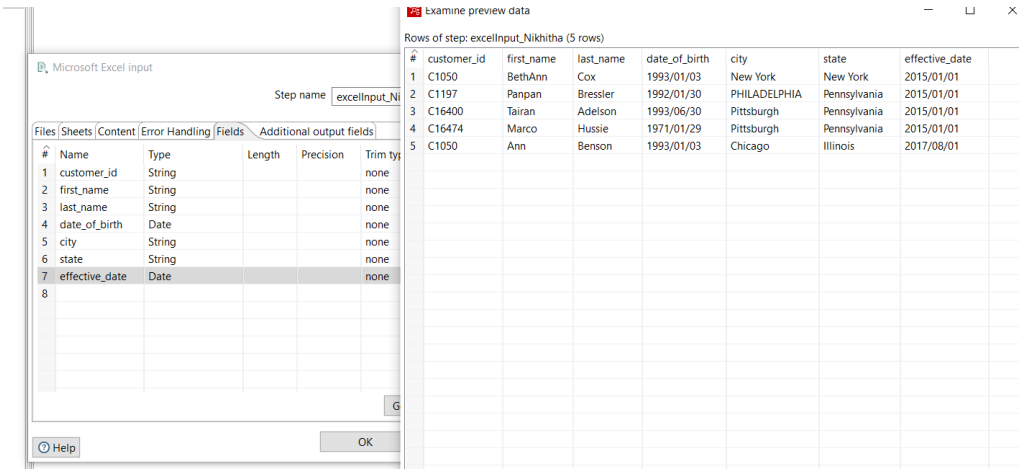
ITMD 526 – Data warehousing

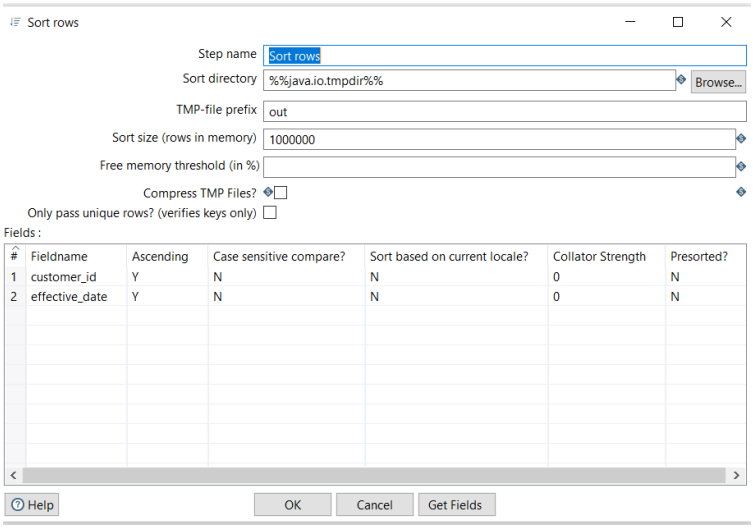
Assignment 3

Nikhitha Kamath – A20473286

Importing Excel File



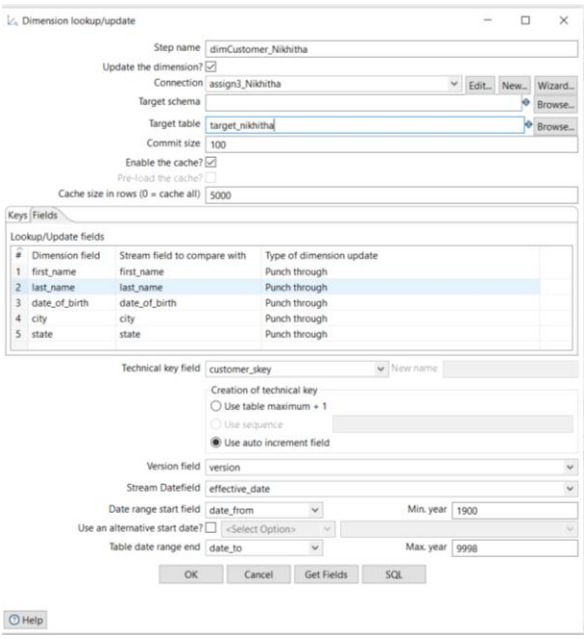
Sorting rows by **customer_id** and then by **effective_date**



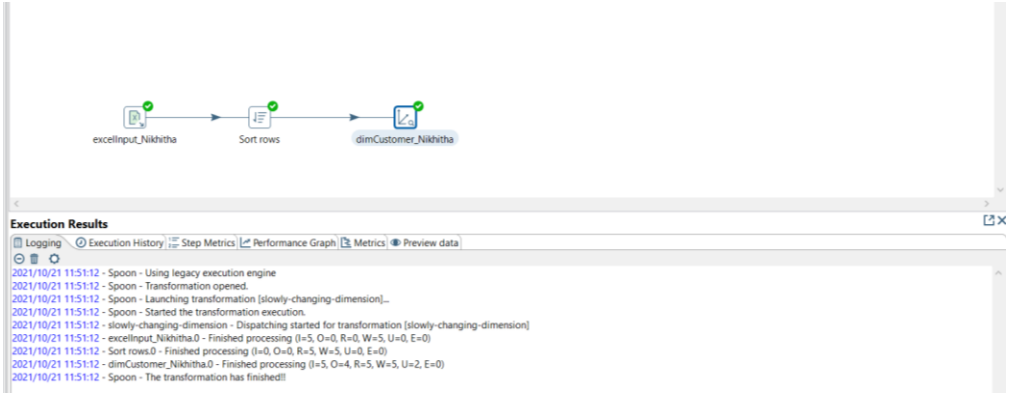
Dimension Lookup

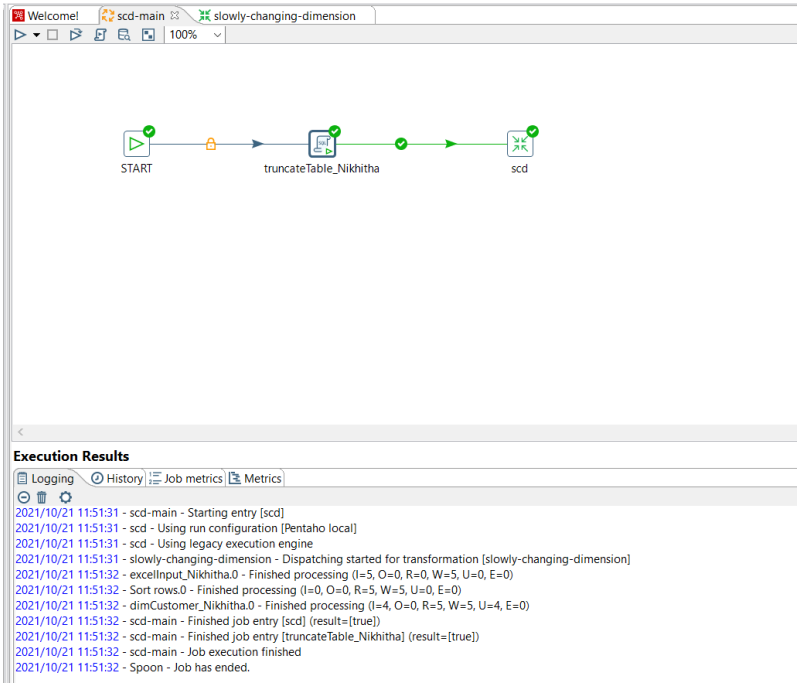
1. Connect to a database
2. Give a target table name
3. ‘Punch Through’ Values. This overwrites the most recent data. This is SCD Type 1

    - first_name: punch through; this is a scd type 1 variable.
    - last_name: punch through; this is a scd type 1 variable.
    - date_of_birth: punch through; this is a scd type 1 variable.
    - city: punch through; this is a scd type 1 variable.
    - State: punch through; this is a scd type 1 variable.



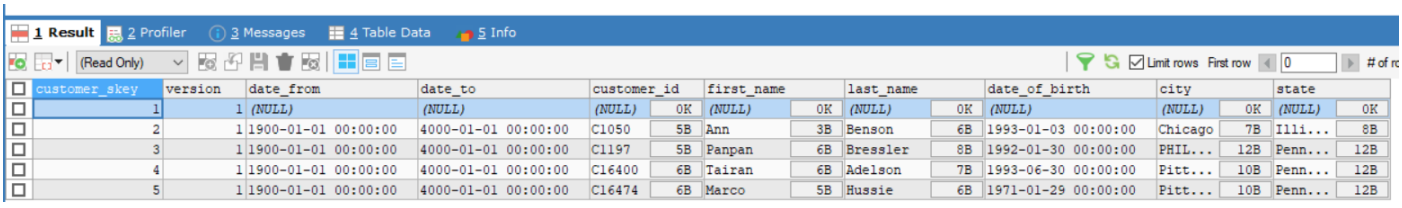
Successful Execution of Transformation and Job



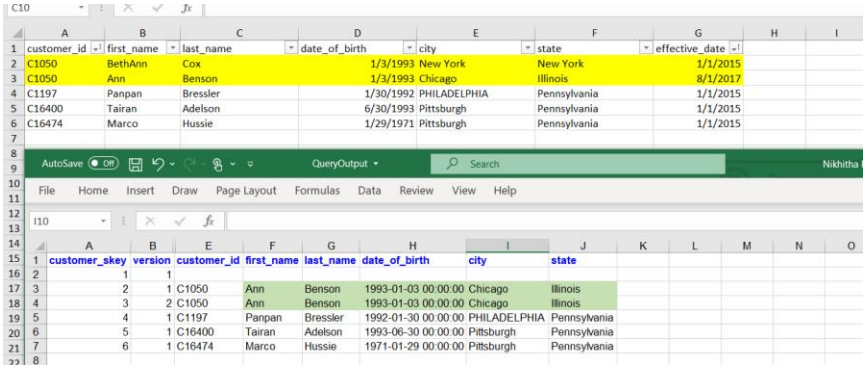


SCD Type 1 Output

All columns were punch through SCD type 1 variables, therefore the details were overwritten.



customer_skey	version	date_from	date_to	customer_id	first_name	last_name	date_of_birth	city	state
1	1	*(NULL)* 0K	*(NULL)* 0K	*(NULL)* 0K	*(NULL)* 0K	*(NULL)* 0K	*(NULL)* 0K	*(NULL)* 0K	*(NULL)* 0K
2	1	1900-01-01 00:00:00	4000-01-01 00:00:00	C1050 5B	Ann 3B	Benson 6B	1993-01-03 00:00:00	Chicago 7B	Illi... 8B
3	1	1900-01-01 00:00:00	4000-01-01 00:00:00	C1197 5B	Panpan 6B	Bressler 8B	1992-01-30 00:00:00	PHIL... 12B	Penn... 12B
4	1	1900-01-01 00:00:00	4000-01-01 00:00:00	C16400 6B	Tairan 6B	Adelson 7B	1993-06-30 00:00:00	Pitt... 10B	Penn... 12B
5	1	1900-01-01 00:00:00	4000-01-01 00:00:00	C16474 6B	Marco 5B	Hussie 6B	1971-01-29 00:00:00	Pitt... 10B	Penn... 12B



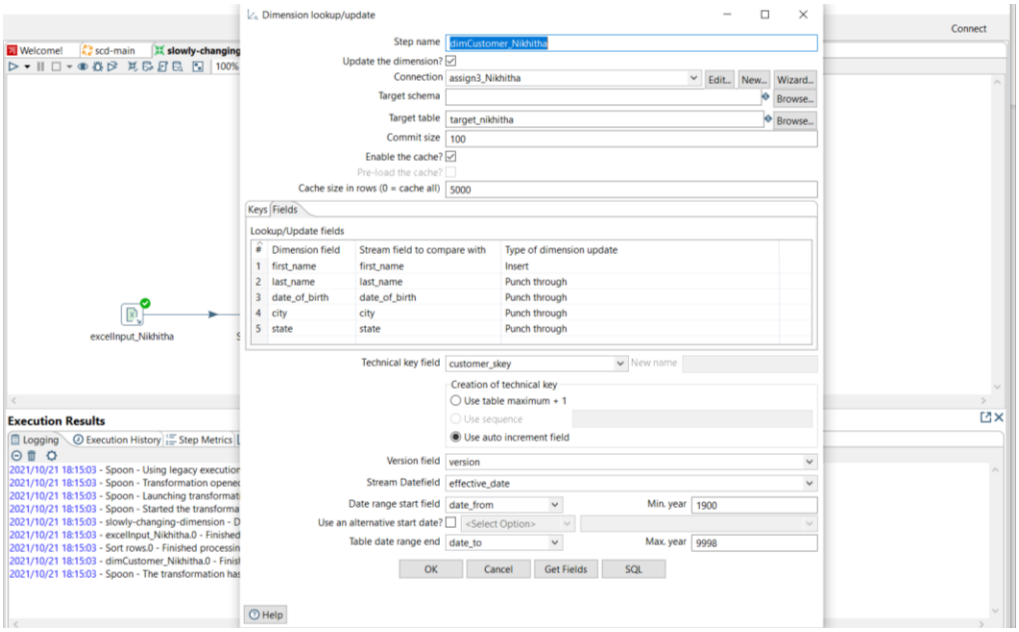
customer_id	first_name	last_name	date_of_birth	city	state	effective_date
C1050	BethAnn	Cox	1/3/1993	New York	New York	1/1/2015
C1050	Ann	Benson	1/3/1993	Chicago	Illinois	8/1/2017
C1197	Panpan	Bressler	1/30/1992	PHILADELPHIA	Pennsylvania	1/1/2015
C16400	Tairan	Adelson	6/30/1993	Pittsburgh	Pennsylvania	1/1/2015
C16474	Marco	Hussie	1/29/1971	Pittsburgh	Pennsylvania	1/1/2015

customer_skey	version	customer_id	first_name	last_name	date_of_birth	city	state
1	1						
2	1	C1050	Ann	Benson	1993-01-03 00:00:00	Chicago	Illinois
3	2	C1050	Ann	Benson	1993-01-03 00:00:00	Chicago	Illinois
4	1	C1197	Panpan	Bressler	1992-01-30 00:00:00	PHILADELPHIA	Pennsylvania
5	1	C16400	Tairan	Adelson	1993-06-30 00:00:00	Pittsburgh	Pennsylvania
6	1	C16474	Marco	Hussie	1971-01-29 00:00:00	Pittsburgh	Pennsylvania

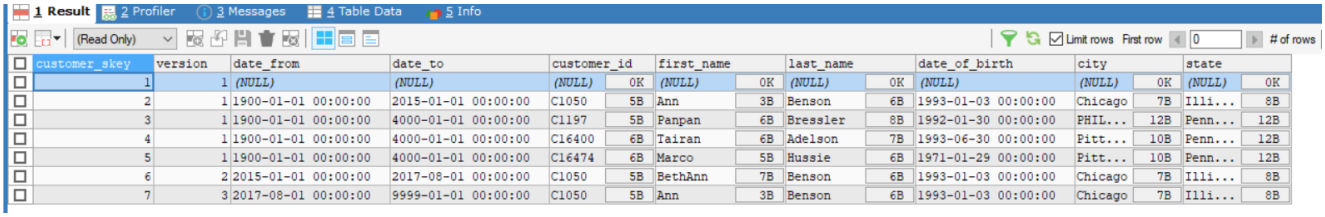
SCD Type 2

1. Connect to a database
2. Give a target table name
3. 'Insert' Values. Adds a new row. This is SCD Type 2

    - first_name: insert; this is a scd type 2 variable.
    - last_name: punch through; this is a scd type 1 variable.
    - date_of_birth: punch through; this is a scd type 1 variable.
    - city: punch through; this is a scd type 1 variable.
    - State: punch through; this is a scd type 1 variable.

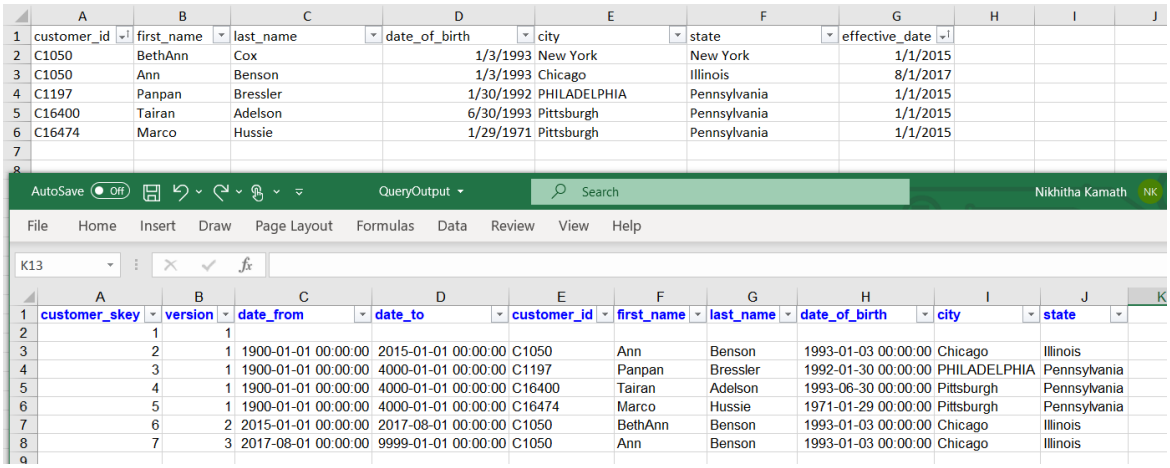


SCD Type 2 Output



customer_skey	version	date_from	date_to	customer_id	first_name	last_name	date_of_birth	city	state
1	1	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
2	1	1900-01-01 00:00:00	2015-01-01 00:00:00	C1050	Ann	Benson	1993-01-03 00:00:00	Chicago	Illi...
3	1	1900-01-01 00:00:00	4000-01-01 00:00:00	C1197	Panpan	Bressler	1992-01-30 00:00:00	PHIL...	Penn...
4	1	1900-01-01 00:00:00	4000-01-01 00:00:00	C16400	Tairan	Adelson	1993-06-30 00:00:00	Pitt...	Penn...
5	1	1900-01-01 00:00:00	4000-01-01 00:00:00	C16474	Marco	Hussie	1971-01-29 00:00:00	Pitt...	Penn...
6	2	2015-01-01 00:00:00	2017-08-01 00:00:00	C1050	BethAnn	Benson	1993-01-03 00:00:00	Chicago	Illi...
7	3	2017-08-01 00:00:00	9999-01-01 00:00:00	C1050	Ann	Benson	1993-01-03 00:00:00	Chicago	Illi...

First name is the type 2 variable, it inserts a new row with the first name changed to Ann. Rest of the columns were punch through SCD type 1 variables, therefore the details were overwritten.



customer_id	first_name	last_name	date_of_birth	city	state	effective_date
C1050	BethAnn	Cox	1/3/1993	New York	New York	1/1/2015
C1050	Ann	Benson	1/3/1993	Chicago	Illinois	8/1/2017
C1197	Panpan	Bressler	1/30/1992	PHILADELPHIA	Pennsylvania	1/1/2015
C16400	Tairan	Adelson	6/30/1993	Pittsburgh	Pennsylvania	1/1/2015
C16474	Marco	Hussie	1/29/1971	Pittsburgh	Pennsylvania	1/1/2015

customer_skey	version	date_from	date_to	customer_id	first_name	last_name	date_of_birth	city	state
1	1								
2	1	1900-01-01 00:00:00	2015-01-01 00:00:00	C1050	Ann	Benson	1993-01-03 00:00:00	Chicago	Illinois
3	1	1900-01-01 00:00:00	4000-01-01 00:00:00	C1197	Panpan	Bressler	1992-01-30 00:00:00	PHILADELPHIA	Pennsylvania
4	1	1900-01-01 00:00:00	4000-01-01 00:00:00	C16400	Tairan	Adelson	1993-06-30 00:00:00	Pittsburgh	Pennsylvania
5	1	1900-01-01 00:00:00	4000-01-01 00:00:00	C16474	Marco	Hussie	1971-01-29 00:00:00	Pittsburgh	Pennsylvania
6	2	2015-01-01 00:00:00	2017-08-01 00:00:00	C1050	BethAnn	Benson	1993-01-03 00:00:00	Chicago	Illinois
7	3	2017-08-01 00:00:00	9999-01-01 00:00:00	C1050	Ann	Benson	1993-01-03 00:00:00	Chicago	Illinois