

## Assignment 3

## SCD Type 1:

It overwrites the data. If we are having 2 data for single customer (Same Customer\_ID) then it overwrites with the new data and original data is removed.

First two records of C999 has been removed and it did overwrite with the current new values.

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
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
C1050	Ann	Benson	1/3/1993	Chicago	Illinois	8/1/2017
C999	Madhu	Jagdale	9/20/1993	Warsaw	Indiana	1/9/2018
C999	Madhuli	Thorat	9/20/1993	Chicago	Illinois	1/23/2021
C999	Madhuli	Thorat	10/17/1990	Chicago	Illinois	10/19/2021

customer_id	first_name	last_name	date_of_birth	city	state	version	date_from	date_to
C1050	Ann	Benson	1993-01-03	Chicago	Illinois	1	1900-01-01	9999-01-01 00:00:00
C1197	Panpan	Bressler	1992-01-30	PHILADELPHIA	Pennsylvania	1	1900-01-01	9999-01-01 00:00:00
C16400	Tairan	Adelson	1993-06-30	Pittsburgh	Pennsylvania	1	1900-01-01	9999-01-01 00:00:00
C16474	Marco	Hussie	1971-01-29	Pittsburgh	Pennsylvania	1	1900-01-01	9999-01-01 00:00:00
C999	Madhuli	Thorat	1993-09-20	Chicago	Illinois	1	1900-01-01	9999-01-01 00:00:00

## SCD Type 2:

It keeps the original data and add more rows with all the new changes, so that you can know what the value of any particular field was before. I have added one more column for the version which tells us for how many times data has been changed, it increments by 1.

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
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
C1050	Ann	Benson	1/3/1993	Chicago	Illinois	8/1/2017
C999	Madhu	Jagdale	9/20/1993	Warsaw	Indiana	1/9/2018
C999	Madhuli	Thorat	9/20/1993	Chicago	Illinois	1/23/2021

customer_key	customer_id	first_name	last_name	date_of_birth	city	state	version	date_from	date_to
2	C1050	BethAnn	Cox	1993-01-03	New York	New York	1	1900-01-01	2017-06-01 00:00:00
3	C1050	Ann	Benson	1993-01-03	Chicago	Illinoi	2	2017-06-01	9999-01-01 00:00:00
4	C1197	Panpan	Bressler	1992-01-30	PHILADELPHIA	Pennsylvania	1	1900-01-01	9999-01-01 00:00:00
5	C16400	Tairan	Adelson	1993-06-30	Pittsburgh	Pennsylvania	1	1900-01-01	9999-01-01 00:00:00
6	C16474	Marco	Hussie	1971-01-29	Pittsburgh	Pennsylvania	1	1900-01-01	9999-01-01 00:00:00
7	C999	Madhu	Jagdale	1993-09-20	Warsaw	Indiana	1	1900-01-01	2021-01-23 00:00:00
8	C999	Madhuli	Thorat	1993-09-20	Chicago	Illinois	2	2021-01-23	9999-01-01 00:00:00

## Mixed Type Update:

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
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
C1050	Ann	Benson	1/3/1993	Chicago	Illinois	8/1/2017
C999	Madhu	Jagdale	9/20/1993	Warsaw	Indiana	1/9/2018
C999	Madhuli	Thorat	9/20/1993	Chicago	Illinois	1/23/2021
C999	Madhuli	Thorat	10/17/1990	Chicago	Illinois	10/19/2021

In the above diagram I have used Type1 and Type 2 SCD. By mixing these two types data can be updated. I have changed date of birth of customer ID C999 and it updated date of birth value for each record of C999 customer.

customer_key	customer_id	first_name	last_name	date_of_birth	city	state	version	date_from	date_to
4	C1197	Panpan	Bressler	1992-01-30	PHILADELPHIA	Pennsylvania	1	1900-01-01	9999-01-01 00:00:00
5	C16400	Tairan	Adelson	1993-06-30	Pittsburgh	Pennsylvania	1	1900-01-01	9999-01-01 00:00:00
6	C16474	Marco	Hussie	1971-01-29	Pittsburgh	Pennsylvania	1	1900-01-01	9999-01-01 00:00:00
7	C999	Madhu	Thorat	1990-10-17	Chicago	Illinois	1	1900-01-01	2021-01-23 00:00:00
8	C999	Madhuli	Thorat	1990-10-17	Chicago	Illinois	2	2021-01-23	9999-01-01 00:00:00