

Slowly Changing Dimensions (SCD)

Slowly Changing Dimensions (SCD) is a data warehouse concept used to manage changes in dimensional data over time. It is a method of managing data changes over time, particularly in huge databases. It allows us to keep track of how data changes while also maintaining the information's correctness and integrity.

Types of Slowly Changing Dimensions (SCD)

1. Type-1 (Overwrite the changes)

Type 1 maintains only the current data without preserving any historical information. When a change occurs, the existing data is overwritten with the updated values. This approach is useful when historical data is not significant or when maintaining history is not required. However, it does not retain any historical context.

PK	Name	City	Date Effective
001	Sadaf	Islamabad	28-05-2015
002	Aamir	multan	14-04-2010
003	qamar	Lahore	29-09-2014
004	zain	Gujrat	27-08-2011
005	muneeb	karchi	28-08-2022

Example, Consider a dimension table with attributes such as ID, Name, City and Date Effective. In Type 1, If any individual decides to modify their current location, the dimension table's existing city entry gets replaced with the new city, resulting in the loss of the previous address.

2. Type-2 (History will be added as new row)

Type 2 :By generating new entries for each update while retaining the initial values, historical changes are effectively monitored. The dimension table is expanded with additional rows, comprising a **surrogate key**, start date, end date, and modified attributes. This approach ensures a comprehensive chronicle of alterations while preserving referential integrity.

SK	PK	Name	City	From	To	Status Flag
45322	002	SADAF	HAFIZABAD	14-APR-10	06-09-19	ARCHIVED
34256	002	SADAF	ISLAMABAD	07-SEPT-19	∞	ACTIVE
12128	004	SULTAN	LAHORE	27-MAR-11	27-10-19	ARCHIVED

23486	004	SULTAN	MULTAN	28-SEPT-19	∞	ACTIVE

Example: In SCD Type 2, if Sadaf and Sultan changes their location, a new row is added to the table with a new surrogate key, start date, and end date. The old row remains unchanged, preserving the previous address and associating it with the appropriate time frame.

3. Type-3 (History will be added as a new column)

SCD Type 3 keeps track of limited historical changes by adding new columns to the dimension table. It maintains the current value in the original attribute and adds separate columns to capture the previous values. It sacrifices full historical tracking for simplicity and efficiency.

SK	PK	NAME	PREV_CITY	CURR_CITY	FROM_DATE
45322	002	SADAF	HAFIZABAD	ISLAMABAD	07-03-19
45322	002	SADAF	ISLAMABAD	LAHORE	10-01-20
12321	004	AAMIR	HAFIZABAD	KARACHI	11-09-20

Example: Extending the dimension table, in Type 3, when Sadaf or Aamir change their locations the exiting column is split into two separate columns like current and Previous city. Previous city is containing value of previous data and current city column contain data of new location.

Type-4 (A new dimension will be added)

Type 4, also known as a separate history table, stores historical changes in a separate table linked to the main dimension table. It maintains a one-to-many relationship between the main table and the history table, allowing for efficient storage and retrieval of historical data.

PK	Name	City	Date Effective
001	Sadaf	Islamabad	28-05-2015
002	Aamir	multan	14-04-2010
003	qamar	Lahore	29-09-2014
004	zain	Gujrat	27-08-2011
005	muneeb	karchi	28-08-2022

History Table (New Dimension) *

SK	PK	NAME	PREV_CITY	CURR_CITY	FROM_DATE
45322	002	Sadaf	hafizabad	ISLAMABAD	07-SEPT-19
12128	004	Aamir	muktan	MARDAN	28-SEPT-19

Example: Using the dimension table, in Type 4, a separate history table is created. Whenever a change occurs, a new row is inserted into the history table with the original attribute values and a timestamp. The main dimension table only contains the current data.