

Q. What is the concept of Slowly changing dimensions (SCD)?

A. SCD is a task that includes storing and managing of both current and historical data over a span of time in a Data Warehouse. It is a crucial ETL task for tracking the history of dimensional records.

Q. What are its types?

A. SCD has three types which can be defined, deployed, and loaded using a Warehouse builder (is a ETL tool produced that offers a graphical environment to build, manage and maintain data integration processes in business intelligence systems). Three types are:

1) Overwriting:

In this the new data overwrites the old existing data. Hence the existing data is lost as it isn't stored anywhere else. This one is a default type as don't need to perform any specific actions to achieve this one.

2) Creating Another Dimension Record:

This one stores the entire history of data. When a change is made to a chosen value the current records is closed. A new record is created with the changed values and this new records becomes our current record. Each record contain the effective and expiration time to identify the time period between which the record was active.

3) Creating A Current Value Field:

This store two versions of values for a certain level attributes. Each record stores the previous value and the current value of a selected attribute. When a change is made to the data the current records becomes the old value and new data becomes the current value.

4) Type 4:

In this group of attributes in a dimension changes rapidly so mini-dimension forms. Its also known as rapidly changing monster dimension.

The frequently changing attributes will be removed from the main dimension and added to mini-dimension.

Q. Explain with examples (Scd1, Scd2, Scd3, Scd4)

A. Let's consider a warehouse of apple store records. If a store changes its Landline number then

SCD1 Example:

Overwriting the new data of Landline is type1. This is the simplest and easiest SCD.

SCD2 Example:

Storing the new data in another record and adding another column for every record version. It stores all the data with a version number. The maximum number in version column will have the latest value.

SCD3 Example:

It stores records in pairs. So it will store old Landline number as current number and the new landline number as new number and will replace old landline with new when a new number is added.

SCD4 Example:

It creates a history table containing current and previous values. After Landline change we insert new record in source table.