# SCD Type 2 Implementation Using Alter Row in Azure Data Factory

## Overview

SCD Type 2 ensures that changes in dimension data are captured as new records while preserving historical data. In Azure Data Factory (ADF), this can be implemented efficiently using the 'Alter Row' transformation within a Mapping Data Flow.

## Recommended Table Schema

Target SCD table structure should typically include:

|  |  |
| --- | --- |
| Column Name | Description |
| SurrogateKey | Unique identifier for each row version (e.g., auto-increment or GUID) |
| BusinessKey | Natural key (e.g., EmployeeID) to identify the entity |
| Attribute | Field(s) you want to track changes on (e.g., Name, Email) |
| StartDate | Date when this version became active |
| EndDate | Date when this version became inactive |
| IsCurrent | Boolean flag to mark the current active record |

## Steps to Implement Using Alter Row

1. 1. Create a Mapping Data Flow in Azure Data Factory.
2. 2. Add a Source transformation to bring in the new incoming data.
3. 3. Add a Lookup transformation to fetch existing current records (IsCurrent = true) from the target table.
4. 4. Use a Join (Inner or Left) to compare incoming data with existing records on the BusinessKey.
5. 5. Add a Derived Column to flag change detection (e.g., changed\_name = source.name != lookup.name).
6. 6. Add an Alter Row transformation with custom expressions:
7. - Mark rows as 'update' where change is detected (to expire existing record).
8. - Mark rows as 'insert' for the new version of the record.
9. 7. Add a Sink transformation configured to allow both 'insert' and 'update' operations.
10. 8. Set 'update' operation to set IsCurrent = false and EndDate = currentUTC().
11. 9. Set 'insert' operation to set IsCurrent = true and StartDate = currentUTC(), EndDate = null.
12. 10. Add the Data Flow to a pipeline and execute it.

## Best Practices

* Use `currentUTC()` to maintain consistency in date values across inserts and updates.
* Ensure surrogate key is auto-generated or uniquely created during insert.
* Use Conditional Split before Alter Row if you want to handle new, changed, and unchanged records separately.
* Validate your expressions in Alter Row for update/insert logic before execution.
* Filter Lookup transformation with `IsCurrent = true` to avoid updating inactive records.