

# [DEFECT] Dimension Deduplication

---

## Problem

As a data engineer, I want to fix the data quality problem caused by a faulty ETL process which led to redundant records in the table `dim_dep_agreement`.

## Example

sk	agrmnt_id	actual_from_dt	actual_to_dt	client_id	product_id	interest_rate
1	101	2015-01-01	2015-02-20	20	305	3.5%
2	101	2015-02-21	2015-05-17	20	345	4%
3	101	2015-05-18	2015-07-05	20	345	4%
4	101	2015-07-06	2015-08-22	20	539	6%
5	101	2015-08-23	9999-12-31	20	345	4%
6	102	2016-01-01	2016-06-30	25	333	3.7%
7	102	2016-07-01	2016-07-25	25	333	3.7%
8	102	2016-07-26	2016-09-15	25	333	3.7%
9	102	2016-09-16	9999-12-31	25	560	5.9%
10	103	2011-05-22	9999-12-31	30	560	2%

In the above sample, a redundancy can be seen within the records 2, 3 and 6, 7, 8.

## Expected Behaviour

New row in the table is created if at least one out of three business attributes (`client_id`, `product_id`, `interest_rate`) changed for a given agreement (`agrmnt_id`).

## Proposed Solution

Prepare a SQL script to create new table `dim_dep_agreement_compacted` with redundant records "collapsed". For example, instead of the rows 2-3, there should be a single row with the period from 2015-02-21 (`actual_from_dt`) to 2015-07-05 (`actual_to_dt`).

## Acceptance Criteria

- The SQL query is designed as a single statement, i.e. without using intermediate/temporary tables, updates or deletes.
- The table `dim_dep_agreement_compacted` should have "smooth history" for every agreement (`agrmnt_id`), i.e. no gaps or intersections for the validity intervals (from `actual_from_dt` to `actual_to_dt`).
- The query logic is described and documented.