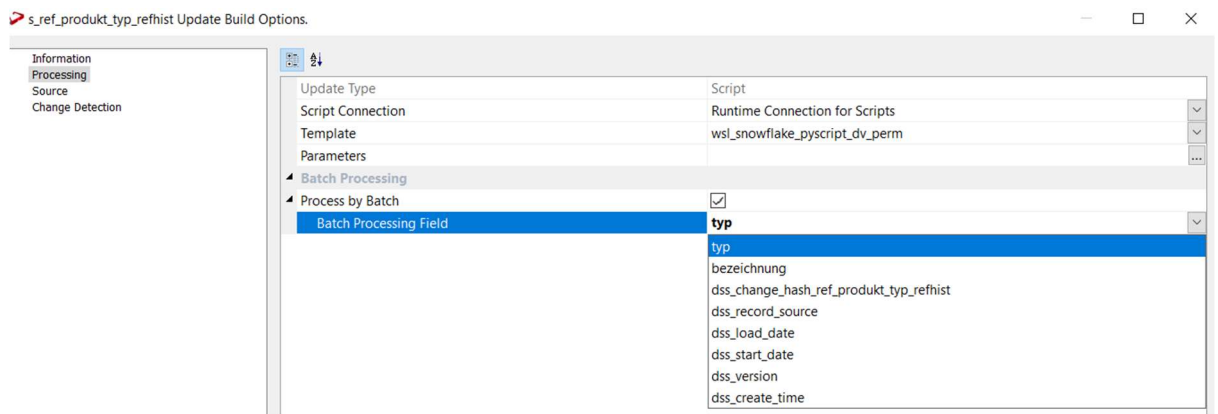


7. Duplicate Data

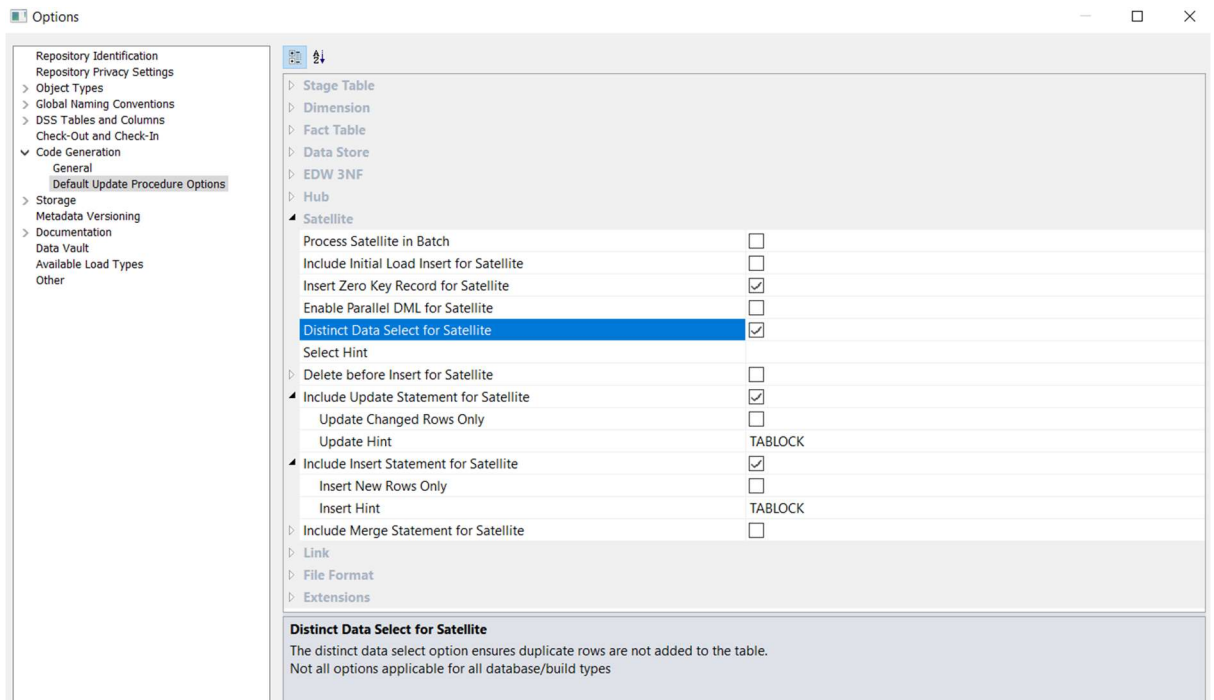
Option 1

Technical DV duplicate data - two identical rows arriving at the same batch timestamp

- a) Use a batch id different from the business key or add a version column using for example a SQL windowed function like row_number() to insert



- b) In WhereScape RED there is an option for each object to set a distinct property on an object.



Option 2

Example two rows with the same business key but different data arrive on the same batch. Meaning the source is erroneous.

Solutions:

1. Change the satellite as a multi-active satellite, and load the received data as it is.
2. Communicate with the team delivering the source data and let them fix it.
3. Apply soft rules in Business Vault to fix the data

Technically, one can use an error handling to create an error reporting table for the team responsible for cleansing the data.

Add a stage entity with a row_number column.

```
ROW_NUMBER() OVER (PARTITION BY <business_key>  
                    ORDER BY <business_key>)
```

Records with a value 1 from the stage go into the vault , and all other records go to an error reporting table.