Lab 4: Part 1 Slowing Changing Dimension

SalesEmployeeDimention Table

|  |  |
| --- | --- |
| Field | Description |
| SalesEmployeeSSN | This is the employee id number and is the same as employee social security number. Since the employee SSN is unique to each employee, it does not change, and can be represented in integer format, the employee SSN is used in the transactional database as employee id number, and is retained in the dimension table as dimension table primary key. |
| SalesEmployeeName | This is the First Name, Last Name of the employee. Over course of employment, there are cases where either employee’s first name or last name or both have changed. |
| SalesEmployeePosition | This is the position title of the sales employee and can take different values such as sales associate, senior sales associate, account executive etc. Over course of employment as a sales person, an employee may hold different position titles over time. |
| SalesDistrict | The sales territory is organized into a hierarchy of regions and districts. The district and region to which a sales employee is assigned may change over time. Each sales district belongs to a region. However, the sales territory can be re-organized and the assignments of districts to regions can also vary over time. |
| SalesRegion | A sales region consists of sales districts. The assignment of sales districts to sales regions can change when there are reorganizations. Such reorganizations are infrequent. |

1. What are the pros/cons of modeling the changes to employee dimension as type 1, 2 or 3 SCD. What is the best way? Can it be implemented using the current dimension design? What changes would need to be made to the dimension table design?

Type1: overwrite the old values with new values

Pros: A. Easy to implement B. no need to add additional rows or columns;

Cons: Lose all the historical values;

Yes, we can use the current dimension design to implement type 1 SCD.

No changes need to be made to the design.

Type2: Add new rows when there are new changes to any of the column values; add additional three columns: Start Date/ Effective Date, End Date/Expiration Date, Indicator(current or expired);

Pros: All the historical values will be kept; No need for future tables structure changes;

Cons:Plenty of records referring to the same product/customer/entity;

No, we need to add additional 3 columns: Start Date/ Effective Date, End Date/Expiration Date, Indicator(current or expired);

Type3: Add a new columns each time we have a new changes; it is usefull for predictable changes

Pros:Easy to retrieve one record with all the changes/historical values; Historical values will be stored;

Cons: No time stamp associated with the change; any change from any entity will result in a new column;

Yes, we can still use the current table design but once there is a new change the table design need to be updated;

1. Assume that the organization has decided to model all the changes to employee dimension as type 2 SCD. At the next quarterly update of the data warehouse, the ETL analyst informs you that EmployeeNo 123456789 has had both a change in SalesDistrict and change in SalesEmployeePosition. Should these changes be added to the dimension table as one new row that captures both changes, or two new rows with one row for Sales Position Change and one row for district change?

It depends on the time/date of the updates. If in the type 2 SCD, we keep the effective date and end date columns, and both of the changes are made on the same date, tes, we can create one row for those changes;

If the changes are made on different date, sometime different time on the same date, we need to create 2 rows for 2 changes;

3. Certain departments, such as HR, expect frequent changes to records within their systems. For example, an employee’s designation can change multiple times a year. In these cases, companies generally want to be able to maintain the history of data while still being able to quickly query it based on the most current attribute value.

In DWB, you could choose to deal with this situation by setting up the Job Title as an SCD6 field. Given the original table below, design a type 6 dimension table for employee dimension.

|  |  |  |  |
| --- | --- | --- | --- |
| Employee ID | Employee Code | Employee Name | Job Title |
| 30032 | 303 | Frodo Baggins | Sales Manager |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Employee ID | Employee Code | Employee Name | Historic Job Title | Current Job Title | Start Date | End Date | Current Indicator |
| 30032 | 303 | Frodo Baggins | Sales Manager | Sales Manager | 2010-01-01 | 9999-12-31 | Yes |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Employee ID | Employee Code | Employee Name | Historic Job Title | Current Job Title | Start Date | End Date | Current Indicator |
| 30032 | 303 | Frodo Baggins | Sales Manager | Program Manager | 2010-01-01 | 2021-10-20 | No |
| 23940 | 303 | Frodo Baggins | District Manager | Program Manager | 2021-10-21 | 2024-12-21 | No |
| 23967 | 303 | Frodo Baggins | Program Manager | Program Manager | 2025-12-21 | 9999-12-31 | Yes |