Data Warehousing Concept Using ETL Process for SCD Type-2

K.Srikanth¹, N.V.E.S.Murthy², J.Anitha³
¹(Computer Science and Systems Engineering, Andhra University, India)
²(Computer Science and Systems Engineering, Andhra University, India)
³(Computer Science and Systems Engineering, Andhra University, India)

Abstract: SCD type 2 will store the entire history in the dimension table. In SCD type 2 effective date, the dimension table will have Start_Date and End_Date as the fields. If the End_Date is Null, then it indicates the current row. Know more about SCDs at Slowly Changing Dimensions Concepts. The new incoming record (changed/modified data set) replaces the existing old record in target. We will see how to implement the SCD Type 2 Effective Date in informatica. If there are retrospective changes made to the contents of the dimension, or if new attributes are added to the dimension which have different effective dates from those already defined, then this can result in the existing transactions needing to be updated to reflect the new situation. As an example consider the Employee dimension.

Keywords: ETL; Metadata; Mapping; Transformation.

I. INTRODUCTION

The beauty of this approach is it will maintain two versions, you will find two records the older version and the current version. In other words it maintains history. The thing to be noticed here is if there is any update in the salary of any employee then the history of that employee is displayed with the current date as the start date and the previous date as the end date. As in case of any SCD Type 2 implementation[1], here we need to first find out the set of SCD2 records which qualify for either INSERT or INSERT/UPDATE. Based on this approach, a typical mapping will contain expression, router and update strategy transformations but will not contain any lookup transformation.

Again we can implement Type 2 in following methods
1. Versioning
2. Effective Dates
3. By setting Current Flag values/Record Indicators.
4. We will divide the steps to implement the SCD type 2 Effective Date mapping into four parts.

II. SCD TYPE 2 EFFECTIVE DATE IMPLEMENTATION

Here we will see the basic set up and mapping flow require for SCD type 2 Effective Date. The steps involved are

Implementation:
Source:
Table 1: Oracle SQL Query On EMP Table

- Create the source and dimension tables in the database using Table 1.
- Open the mapping designer tool, source analyzer and either create or import the source definition.
- Go to the Warehouse designer or Target designer and import the target definition[2].
- Go to the mapping designer tab and create new mapping.
- Drag the source into the mapping[7].
- Go to the toolbar, Transformation and then Create.
- Select the lookup Transformation, Figure 1. enter a name and click on create. You will get a window as shown in the below image.

![Figure 1: Creating Lookup Transformation ports logic](image)

- Edit the lookup transformation, go to the ports tab and remove unnecessary ports. Just keep only Empkey, EMPNO and location ports in the lookup transformation. Create a new port (new_flag, update_flag) in the lookup transformation[3]. This new port needs to be connected to the output port of the Expression transformation.
- Go to the conditions tab of the lookup transformation and enter the condition as EMPNO= EMPNO1.
- Go to the properties tab of the LKP transformation and enter the below query in Lookup SQL Override[1]. Alternatively you can generate the SQL query by connecting the database in the Lookup SQL Override expression editor and then add the WHERE clause.

```sql
SELECT * FROM EMP;
```

14 rows selected.

![Figure 2: Creating Expression Transformation ports logic](image)

You can add ports to expression transformation either by selecting and dragging ports from other transformations or by opening the expression transformation and create ports manually[4]. Figure 2. We can add the port new_flag and update_flag using string datatype. In expression transformation implement the employee key either true or false.
1. If(ISNULL(EMPKEY), 'T', 'F'))
2. If( NOT ISNULL(EMPKEY) AND (SAL=SAL1), 'T', 'F'))

II. SCD TYPE 2 EFFECTIVE DATE IMPLEMENTATION
In this part, we will identify the new records and insert them into the target with Begin Date as the current date. The steps involved are:
- Go the properties tab of filter transformation and enter the filter condition as New_Flag=T and Update_Flag=T[5].
- Edit Router Transformation select groups port writing the Group Filter Condition in the inset and update flags, Figure 3.

Figure 3: Creating Router Transformation Groups logic
Insert : New_Flag='T' Update: Update_Flag='T'

III. SCD TYPE 2 EFFECTIVE DATE IMPLEMENTATION
In this part, we will identify the changed records and insert them into the target with Begin Date as the current date. Figure 4. The steps involved are:
- Now connect the ports of expression transformation (Nextval, Start_Date) to the Target definition ports (Emp_Keyword, End_Date)[6]. The part of the mapping flow is shown in the below image.
- Now drag the target definition into the mapping and connect the appropriate ports of update strategy transformation to the target definition.
- Drag and connect the NextVal port of sequence generator to the Expression transformation. In the expression transformation create a new output port (Start_Date and End_Date) and assign value SYSDATE to it.

Figure 4: Creating Expression Transformation ports logic
Start_date: Sysdate
IV. SCD TYPE 2 EFFECTIVE DATE IMPLEMENTATION

In this part, we will update the changed records in the dimension table with End Date as current date.

- Go to the ports tab of expression transformation and create a new output port (Start_Date and End_Date with date/time data type). Assign a value SYSDATE to this port[5].
- Now create an update strategy transformation and drag the ports of the expression transformation into it. Go to the properties tab and enter the update strategy expression as DD_UPDATE.
- Drag the target definition into the mapping and connect the appropriate ports of update strategy to it. Figure 5. The complete mapping image is shown below.

The complete Slowly Changing Dimension Mapping Design flow, Figure 6. This flow will provide completion information of SCD-Type-2 source data how to load target, maintain the data processing.
V. RESULTS

Table 2: Oracle SQL Query On EMP Table Target Data

Once load the target data after write oracle queries in insert and update the values Using connect the employee table. Table 2, Table 3. Below oracle table insert, update display the new type 2 complete updated data.

VI. CONCLUSIONS

Extraction-Transformation-Loading (ETL) tools are pieces of software responsible for the extraction of data from several sources. In this paper, we have focused on the problem A Type One change updates only the attribute, doesn't insert new records, and affects no keys. It is easy to implement but does not maintain any history of prior attribute values. Slowly Changing Dimensions (SCDs) are dimensions that have data that changes slowly, rather than changing on a time-based, regular schedule. In SCD type 2 effective date, the dimension table will have Start_Date and End_Date as the fields. If the End_Date is Null, then it indicates the current row. Know more about SCDs at Slowly Changing Dimensions Concepts. The new incoming record
(changed/modified data set) replaces the existing old record in target. Comprehensive ETL criteria were identified, testing procedures were developed, and this work was applied to commercial ETL tools. The study covered all major aspects of ETL usage and can be used to effectively compare and evaluate various ETL tools. We can implementation on SCD TYPE-2 based on SCD TYPE-1 and new fields like Versioning, Effective Dates, By setting Current Flag values/Record Indicators.

REFERENCES

Journal Papers:

Books:

Theses:

Proceedings Papers: