Use Case Name : Enriched_Bom Creation

Use Case Definition:

The objective of this use case is to create a comprehensive table that provides material, supplier, manufacturer, and category details. This table illustrates the entire flow of materials, from raw materials to finished goods.

Data Sources

1. **BOM Data**: Contains information about finished goods (FG) and raw materials.
2. **Supplier_Volume**: Details about suppliers for raw materials.
3. **SML Table**: Manufacturer details for raw materials.
4. **Spend Table**: Category information for raw materials.

Steps to Implement Use Case

Step 1: read source files

Read all source files and create raw delta tables and Create pipeline in azure datafactory.

Step 2: Join BOM and Supplier_Volume

- Perform a join between the `Supplier_Volume` table and the `BOM` table to fetch supplier details for the raw_materials.
- Conditions before joining Supplier_Volume:
    - Include only records where `last_3_yr_vol` is greater than `0`.
    - Exclude records where `suplr_id` is blank.

Step 3: Add Manufacturer Details (Two-Level Join with SML Table)

1. **Level-1 Join**:
    a. Join the result dataset from Step 1 with the `SML` table to add manufacturing details for raw_materials.
    b. Conditions to apply on SML for Level 1 Join:

i. Include only records where `last_used_date` is within or later than the last three years.
ii. Include only records where `status` is active (blank).

2. **Level-2 Join**:
    a. For records unmatched in Level-1, perform a second join with the SML table.
    b. Conditions to apply on SML for Level-2 Join:
        i. Include only records where `status` is inactive (X).
3. **Add Join-Level Column**:
    a. Introduce a new static column in the output to specify the join level:
        i. Records matched at Level-1 → Display **"level-1 join"**.
        ii. Records matched at Level-2 → Display **"level-2 join"**.
        iii. Records unmatched in both levels → Display **"No match"**.

Step 4: Add Category Information

- Join the result dataset from Step 2 with the Spend table to add category details for the raw_materials.

Step 5: Add visualization using matplotlib or plotly

- Create a pie chart visualization for material category distribution (percentwise)
- Create a bar chart visualization for percentage of nulls across every column in final output table.
- Or use databricks dashboard feature to create sample dashboard for category distribution (percentwise) and total number of materials information

Output

The final output table includes the following enriched data:

A) Final enrich bom table contains:

- Finished Material details
- Raw Material details
- Supplier details

- Manufacturer details (with join level)
- Category information

b) adf pipeline execution details which conatins :

- Pipeline Run id
- Job name (databricks notebook name)
- Pipeline trigger time
- Total execution time
- Status of pipeline (success or fail)

Expected Results

- Comprehensive and enriched table covering all material details.
- Accurate categorization, supplier and manufacturer information for raw materials.
- Proper handling of unmatched records with appropriate labeling.
- Execution of entire pipeline through adf
- Adf pipeline audit log table which will be helpful to monitor pipeline execution details.