**Full Detailed Analysis and Proposed Solution for ETL Workflow Design for Hierarchy Update for SP/ITCL**

**1. Introduction**

This document outlines a comprehensive analysis and proposes a solution for an ETL workflow designed to automate hierarchy updates for SP/ITCL data within the Universal Set table. The primary objective is to eliminate the need for manual intervention in reflecting hierarchy changes within historical data.

**2. Problem Statement**

The current process for managing hierarchy changes in SP/ITCL data suffers from the following shortcomings:

* **Manual Intervention:** When hierarchy changes occur in the RDR data source, manual updates are required to reflect these changes in the historical data stored within the GDR dataset. This approach is time-consuming, prone to errors, and leads to inconsistencies.
* **Outdated Snapshots:** Existing snapshots (SP Monthly, Weekly, ITCL Monthly, Weekly, Combined Monthly/Weekly) rely on data from the GDR, which may not reflect the latest hierarchy updates. This results in inaccurate information being presented in reports and data views that depend on these snapshots.

**3. Business Requirements**

To address the aforementioned issues, the proposed ETL workflow aims to achieve the following:

* **Automated Updates:** Implement an automated ETL workflow that triggers daily or upon completion of the daily SP/ITCL incremental job execution. This workflow will ensure the Universal Set table is updated with the latest hierarchy data fetched from RDR.
* **Fresh Snapshots:** Newly created monthly and weekly snapshots for SP/ITCL data will leverage the updated Universal Set table, reflecting the latest hierarchy information.
* **Preserved History:** Existing historical snapshots will remain unaltered to maintain a record of past data states. Users can decide how to utilize these snapshots in the future (e.g., for reference purposes only).
* **Universal Data Exposure:** The updated Universal data will be readily available to designated reports and data views, including:
  + SP Monthly
  + SP Weekly
  + ITCL Monthly
  + ITCL Weekly
  + SP and ITCL Combined Universal Set
  + SP and ITCL Combined Monthly
  + CRP
* **Incremental Table Consistency:** The behavior of the incremental table will remain unchanged. It will continue to rely on FG hierarchy data uploaded by the data loader for the past 90 days, ensuring consistency within the data governance UI export.

**4. Proposed Solution**

The proposed ETL workflow will encompass the following stages:

* **Data Acquisition**
  + A scheduler component will initiate the ETL job based on the defined trigger (daily or upon daily SP/ITCL incremental job completion).
  + The ETL job will establish a connection to the RDR data source and extract the latest hierarchy data.
* **Data Transformation**
  + The extracted data will undergo transformations to align with the structure of the Universal Set table.
  + Data cleansing techniques may be applied to rectify potential inconsistencies or errors within the RDR data.
* **Data Loading**
  + The transformed data will be loaded into the Universal Set table.
  + Update operations will target specific records using a combination of person ID and cost center code for accuracy.
* **Snapshot Creation**
  + Separate ETL processes will be established to generate monthly and weekly snapshots for SP/ITCL data.
  + These processes will extract data from the updated Universal Set table, ensuring the latest hierarchy information is captured.
  + The newly created snapshots will incorporate the updated hierarchy, while existing historical snapshots will remain unmodified.
* **Data Exposure**
  + The updated Universal data will be accessible to the designated reports and data views.
  + Reporting tools will be reconfigured to point to the updated Universal Set table as the source of hierarchy data.
* **Incremental Table Update**
  + The incremental table will continue to utilize data from FG uploads for the past 90 days, preserving its current behavior.

**5. Impact Analysis**

* **Existing Historical Snapshots:** A decision needs to be made regarding how to handle existing historical snapshots. Options include:
  + Leaving them unchanged for reference purposes.
  + Implementing a separate process to update them with the latest hierarchy data (may require additional resources and testing). This approach should be carefully evaluated to determine the cost-benefit trade-off.
* **Data Validation:** Strategies for validating the accuracy of the data fetched from RDR will be established. This may involve data quality checks, reconciliations with source systems, or implementing data profiling techniques.

**6. Next Steps**

* **Detailed Workflow Design:** Develop a comprehensive technical document outlining the specific steps involved in each stage of the ETL workflow. This document will include details on data extraction methods, transformation logic, data loading procedures, and snapshot generation processes.
* **Impact Analysis Report:** Prepare a detailed report outlining the potential impacts of the proposed solution on existing systems, processes, and users. This report will address how existing historical snapshots will be handled and the approach for data validation.