Learn How to Implement SCD in Talend to Capture Data Changes

Business Overview

Slowly Changing Dimensions is one of the most important concepts in a Data Warehouse. The goal of a data warehouse is to analyze data from a historical perspective; as a result, we cannot simply overwrite the data and must rely on special techniques to preserve the history when it comes to the analytical and volume components. The implementation is based on Slowly Changing Dimensions in Data Warehouse.
For example- The employee dimension may hold attributes such as name, date of joining, hour rate, monthly salary gross, and phone number. The employee's details may change over time (e.g.changing hour rate, changing monthly salary gross). These changes can be accommodated by a slowly changing dimension

This is the second project in the Talend series. The first project covers the installation of the Talend tool, an overview of Talend, and running an ETL Talend job.
In this project, we are going to create an ETL pipeline in Talend Open Studio to capture data changes using the Slowly Changing Dimension technique. We are also going to schedule the task using Task Scheduler in order to trigger the event after a certain period of time. This project will help you learn all the basics of SCD, its features, and different types of SCD along with their implementation with proper use cases.

Tech Stack

➔
Tool: Talend Open Studios.
➔
Database: MySQL, Postgres.

Talend:

Talend is an open-source software integration platform that allows you to build ETL pipelines across multiple platforms. Talend Open Studio is available free of charge and easy for anyone to download and use. It provides software solutions for data preparation, data quality, data integration, application integration, data management, and big data. Talend offers separate products for each of these solutions.

Key Takeaways

- Understanding the Extract-Transform-Load process.
- Understanding the importance of Talend.
- Creating Jobs in Talend.
- Importing an existing job in Talend.
- Exporting a job in Talend.
- Creating a connection with the MySQL database.
- Creating a connection with the Postgres database.
- Creating pipelines in Talend.

- Load Source data to Postgres using Talend Job.
- Understanding the need to capture data change.
- Understanding the concept of Slowly Changing Dimension(SCD).
- Understanding the different types of SCD.
- Understanding the various features of SCD.
- Difference between Version and Active features of SCD.
- Implementation of different types of SCD with the proper use case.
- Run a job in Talend.
- Build a job in Talend.
- Schedule a Talend job using Task Scheduler.

Architecture diagram: