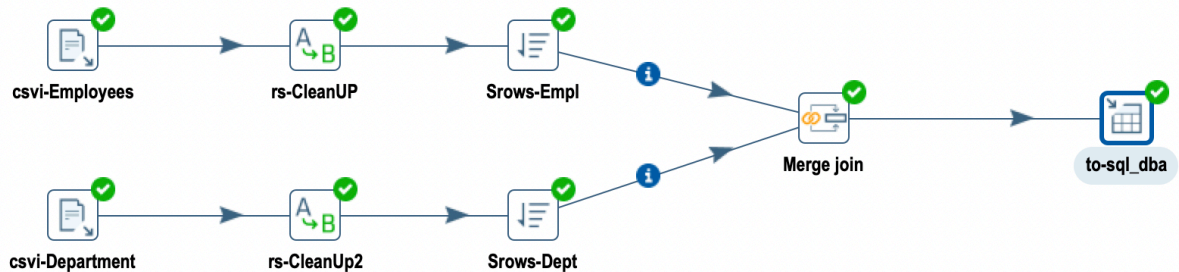


ETL PROJECT_2

Pentaho Data Integration



Pentaho Transformations

Pentaho Data Integration(PDI) provides the Extract, Transform, and Load (ETL) capabilities that facilitate the process of capturing, cleansing, and storing data using a uniform and consistent format that is accessible and relevant to end users and IoT technologies.

This is a 5 steps ETL process, this time it consist of 2 tables where cleansing happen before the tables are joined then loaded to sql Database.

1. **Extraction.** Getting the tables from local to PDI to be able to explore, trim, check the data type and can also preview the imported data.
2. **Data Cleansing.** The PDI "Replace in String" will clean all kind of symbols, replaces strings, add or remove, adjust a date type and many more...etc.
3. **Sort Rows.** The step will adjust the two tables by there primary keys.
4. **Merge Join.** This is simply the joining.
5. **Table Output.** Finally, it's the creation of the landing table in the database. Now the data is ready for further analysis.