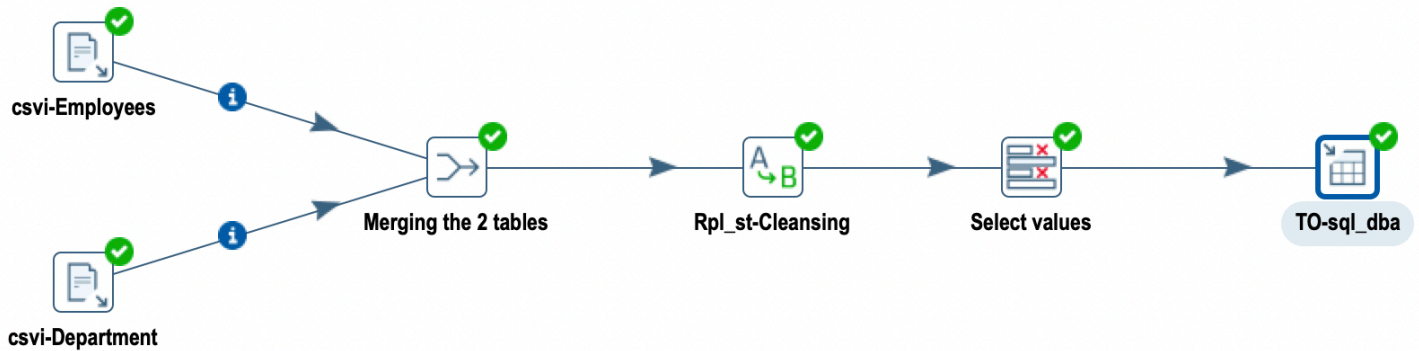


# ETL PROJECT\_3

## 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 Transformation, this time it consist of 2 tables where cleansing happen after 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. **Merge Join.** This is simply the joining.
3. **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.
4. **Select Values.** Select Values will finished it up last data preparation to correct all data type, add or remove any unwanted rows, correct date type...etc before loading to a sql database or data warehouse.
5. **Table Output.** Finally, it's the creation of the landing table in the database. Now the data is ready for further analysis.

By running the transformation, a green tick will appear if everything is correct as pictured above.

