Reconcilation application

Prudential

PLAI Datalake and Azure DWH Reconciliation Application

Technical Specification Document

Contents

[1. Overview 3](#_Toc50122226)

[2. Code Structure and Configuration 3](#_Toc50122227)

[2.1. Code structure 3](#_Toc50122228)

[2.2. Data sources configuration 3](#_Toc50122229)

[2.2.1. Source 1 – Greenplum (postgres) 3](#_Toc50122230)

[2.2.2. Source 2 – Datalake Dev (postgres) 3](#_Toc50122231)

[2.2.3. Target 1 – Local Sql Server (sql server) 3](#_Toc50122232)

[2.2.4. Target 2 – Azure Data Warehouse (sql server) 4](#_Toc50122233)

[2.3. Source/Target Table Configuration 4](#_Toc50122234)

[2.4. Schema Configuration 4](#_Toc50122235)

[3. Reconciliation Criteria 4](#_Toc50122236)

[3.1. Row Count 4](#_Toc50122237)

[3.1.1. Source data path 4](#_Toc50122238)

[3.1.2. Column Specification 4](#_Toc50122239)

[3.2. Investment Operation (DTF) 5](#_Toc50122240)

[3.2.1. Source data path 5](#_Toc50122241)

[3.2.2. Column Specification 5](#_Toc50122242)

[3.2.3. Extraction batch time 5](#_Toc50122243)

[3.2.4. Source detail 5](#_Toc50122244)

[4. Reconciliation Report 6](#_Toc50122245)

[4.1. Fields 6](#_Toc50122246)

[4.1.1. id (unique identifier) 6](#_Toc50122247)

[4.1.2. criterion 6](#_Toc50122248)

[4.1.3. outcome 6](#_Toc50122249)

[4.1.4. source\_db 6](#_Toc50122250)

[4.1.5. target\_db 6](#_Toc50122251)

[4.1.6. source\_table 6](#_Toc50122252)

[4.1.7. source\_predicate 6](#_Toc50122253)

[4.1.8. target\_table 6](#_Toc50122254)

[4.1.9. source\_column 6](#_Toc50122255)

[4.1.10. target\_column 6](#_Toc50122256)

[4.1.11. source\_result 6](#_Toc50122257)

[4.1.12. target\_result 6](#_Toc50122258)

[4.1.13. check\_time 6](#_Toc50122259)

[4.2. Sample 6](#_Toc50122260)

[5. Customization 6](#_Toc50122261)

[5.1. Change existing data source 6](#_Toc50122262)

[5.2. Adding new data source 6](#_Toc50122263)

[5.2.1. Adding parameters in application.properties 6](#_Toc50122264)

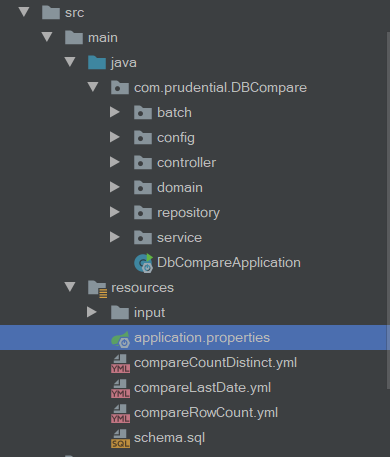
[5.3. Adding new criterion 6](#_Toc50122265)

# Overview

The reconciliation framework is a Java application built on Spring JPA Data framework. Its intention is to reconcile the data loaded onto Azure Data Warehouse with its corresponding sources e.g. datalake dev, Greenplum etc.

# Code Structure and Configuration

## Code structure



## Data sources configuration

Data sources are configured in application.properties. Below are the code snippets with passwords and user names masked for each data source.

### Source 1 – Greenplum (postgres)

spring.datasource3.url=jdbc:postgresql://10.171.211.50:5433/PRUDB

spring.datasource3.jdbcUrl=${spring.datasource3.url}

spring.datasource3.username=\*\*\*

spring.datasource3.password=\*\*\*

spring.datasource3.driver-class-name = org.postgresql.Driver

### Source 2 – Datalake Dev (postgres)

spring.datasource.url=jdbc:postgresql://10.170.49.62:5433/prudb

spring.datasource.jdbcUrl=${spring.datasource.url}

spring.datasource.username=\*\*\*

spring.datasource.password=\*\*\*

spring.datasource.driver-class-name = org.postgresql.Driver

### Target 1 – Local Sql Server (sql server)

spring.datasource2.url=jdbc:sqlserver://10.170.49.196:1433

spring.datasource2.jdbcUrl=${spring.datasource2.url}

spring.datasource2.username=\*\*\*

spring.datasource2.password=\*\*\*\*

spring.datasource2.driver-class-name = com.microsoft.sqlserver.jdbc.SQLServerDriver

### Target 2 – Azure Data Warehouse (sql server)

spring.datasource4.url=jdbc:sqlserver://sdb-sgrass-dev-az1-sql001.database.windows.net:1433;databaseName=ifrs17\_dw

spring.datasource4.jdbcUrl=${spring.datasource4.url}

spring.datasource4.username=\*\*\*

spring.datasource4.password=\*\*\*

spring.datasource4.driver-class-name = com.microsoft.sqlserver.jdbc.SQLServerDriver

## Source/Target Table Configuration

Source and target tables are specified in the three .yml files: compareRowCount.yml, compareCountDistinct.yml and compareLastDate.yml. The business logic is described in [section 3](#_Reconciliation_Criteria).

## Schema Configuration

In a JPA-based app, we can choose to let Hibernate create the schema or use schema.sql. Here we chose the latter approach. DDL is specified in the schema.sql file placed under resources folder.

# Reconciliation Criteria

## Row Count

## Count distinct of a specific column

# Reconciliation Report

## Fields

The reconciliation report is named as reconciliation\_report under the ifrs17 schema in datalake dev and consists of the following fields:

### id (unique identifier)

### criterion

### outcome

### source\_db

### target\_db

### source\_table

### source\_predicate

### target\_table

### source\_column

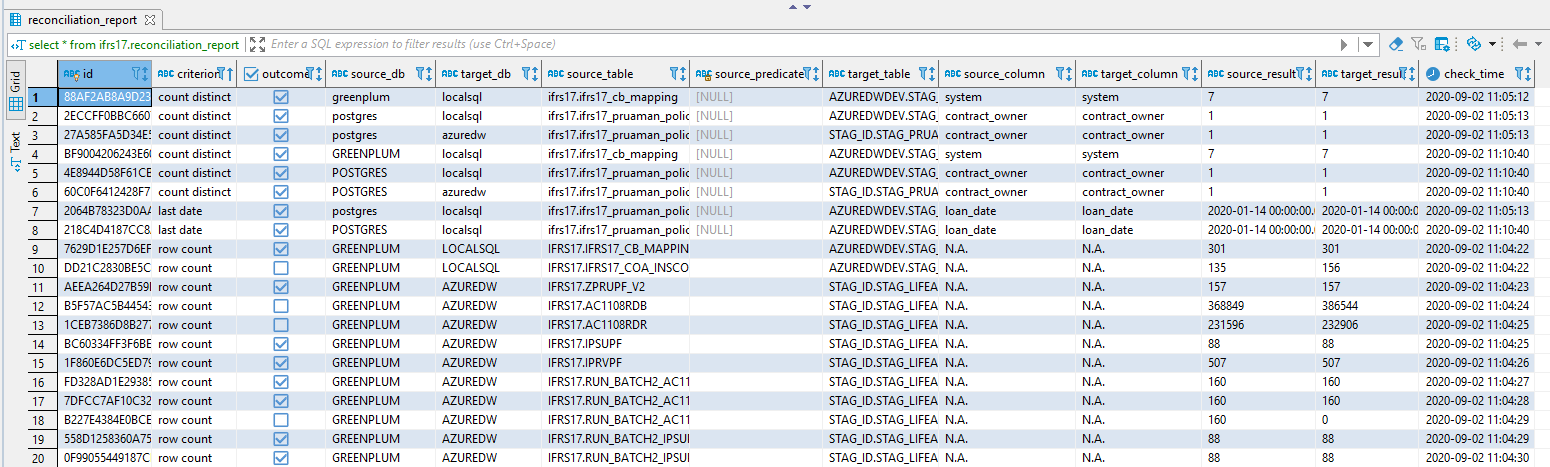
### target\_column

### source\_result

### target\_result

### check\_time

## Sample



# Customization

## Change existing data source

## Adding new data source

### Adding parameters in application.properties

## Adding new criterion