IMPLEMENTATION GUIDELINES

01/05/23

# AIM

This guide shows how to use the Pentaho ETL to bring the [Malawi Synthetic data](https://github.com/Inspire-Mubas/Malawi-IDSR-COVID-19-Synthetic-DataSet/tree/main/v2.0)  into an OMOP database in Postgres.

**Prerequisites**: [POSTGRES](https://www.enterprisedb.com/downloads/postgres-postgresql-downloads), pgadmin, [java](https://www.java.com/en/download/) installed, [Pentaho data integration](https://www.hitachivantara.com/en-us/products/dataops-software/data-integration-analytics/pentaho-community-edition.html) (PDI Base install) using Spoon batch file

# Environment preparation

* Download the [transformations](https://drive.google.com/drive/folders/1rqxrfAzAdqFosDhQMvf7dI661U7jKlb5?usp=sharing) save under Working folder
* Download the [MW CSV files](https://drive.google.com/drive/u/1/folders/1FhCfP5X-aGvymgzIZloOvr0yEj05-re-) save under Working folder

# DATABASE PREPARATION

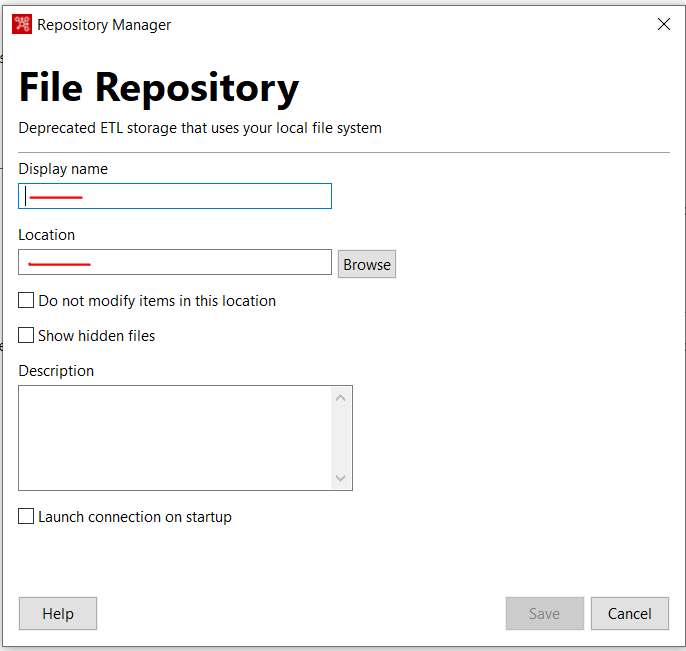
1. Create a database in Postgres pgAdmin4 and give it name MW\_IDSR [name does not matter]
2. Using Postgres pgAdmin4, open the Query Tool and create schemas using this [SQL statement](https://drive.google.com/file/d/1JZ9q4Nfg8DNJXyMemE51kdjT9R_bqoxK/view?usp=sharing).
3. Import the downloaded Malawi CSV file into the **synthetic\_v1.mw\_idsr\_synthetic\_v1** table using the Import functionality:

| i. Select the **header**, leave the **Delimiter** to ‘comma’ and remove the other options like **Quote** |  |
| --- | --- |
|  |  |

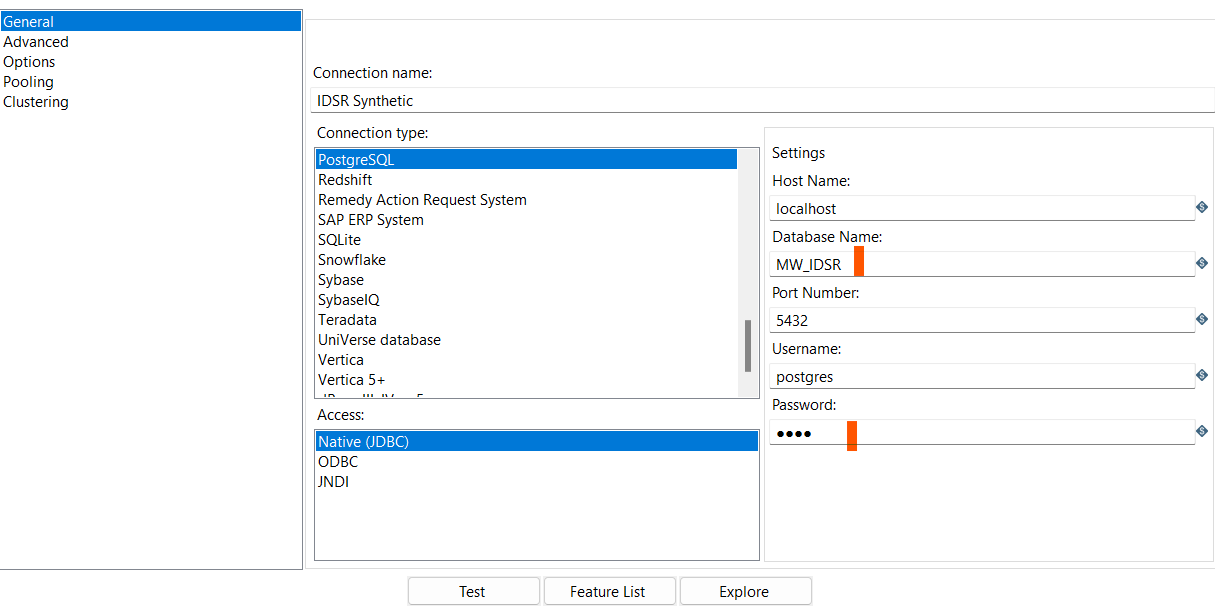
| ii. And finally click on the Columns tab to make sure they get selected automatically: |  |
| --- | --- |

# PENTAHO TASKS

1. Open Pentaho Spoon and navigate to Connect, there create a **File Repository** with any given name like **OMOP** and the location should be the **Pentaho v1** folder where the transformations are stored:



1. After creating the repository, select it and connect. Verify that the File Repository is connected by checking the top corner to see the name of the repository you created.
2. Within Pentaho, open **00 MASTER.kjb**
3. Edit a database connection **IDSR Synthetic** within the job by navigating to view -> 00 MASTER -> IDSR Synthetic. Edit the Database Name and Password :



*Do not change the database connection name “IDSR Synthetic”. By changing it, all other transformations will need to be modified to the new name.*

1. Test the database connection by clicking **Test** at the bottom of the dialog box.
2. Right click on the Database connection and **SHARE**
3. Run the **Pentaho files v1/00 MASTER.kjb** job.

**NOTE**: Incase there happens to be any error and the process that to be rerun: rerun the SQL statement, import the synthetic data and run the 00 Master again

# MIGRATING TO OMOP 5.4 FROM 6.0

1. Open **Pentaho files v1/OMOP 6.0 to 5.4/00 MASTER JOB.kjb** using Pentaho Spoon, make sure the database connection is okay and then run it. (Ensure that you are still connected to the file repository created earlier before running any jobs)
2. Refresh (DB\_MW\_OMOP) database and see if omop\_5\_4\_v1 schema has been created

# TRANSFORMATION STEPS

The modifications we made to transformation steps to accommodate Malawi IDSR data are already reflected inside the Pentaho. This [document](https://docs.google.com/document/d/1MJCF3uJwJ0fcrTPtQJi8kHIcdrg-f2y9/edit?usp=sharing&ouid=105967308948805982311&rtpof=true&sd=true) gives a high level explanation of the modifications and steps taken to come up with this Pentaho installation. A detailed explanation is contained in this [document](https://docs.google.com/document/d/1Gh9qVyaHMCmv5SnKNIQCLzpbgDd5fcbUqtyYvNOJScw/edit?usp=share_link).