

# EDW Extractor User Guide

A Guide for Using the EDW Extractor v2.0 – *LIMITED to  
Extractor Installation & EDW Connection*

## Contents

|  |    |
|--|----|
| Document Control.....                              | 3  |
| Associated References .....                        | 4  |
| 1 Overview .....                                   | 5  |
| 2 Getting Started – Extractor Installation .....   | 5  |
| 2.1 Pre-requisites .....                           | 5  |
| 2.2 Download & Unzip Extractor v2 Executable ..... | 5  |
| 2.3 Extractor v2 Folder Structure.....             | 6  |
| 2.3.1 Build\Extractor Folder.....                  | 6  |
| 2.3.2 Dist\Extractor Folder.....                   | 6  |
| 2.4 Running the Extractor .....                    | 8  |
| 3 Troubleshooting.....                             | 10 |
| 3.1 Unable to connect to EDW.....                  | 10 |
| 3.1.1 Configuring ODBC Driver for EDW .....        | 10 |
| 3.1 Slow to Extract Data from EDW .....            | 14 |

## Document Control

|                            |   |
|----------------------------|---|
| <b>Document Title</b>      | EDW Extractor User Guide: LIMITED to EDW Access Testing |
| <b>Document Status</b>     | Draft   |
| <b>Document Owner</b>      | ABM Costing Team  |
| <b>Version</b>             | 0.2   |
| <b>File Name /Location</b> |   |

### Version History

| Version | Date       | Author                                | Changes   |
|---------|------------|---------------------------------------|---|
| 0.1     | 12.10.2023 | Lai-Mun Balnave                       | Initial Draft   |
| 0.2     | 22.02.2024 | Ranjit Sukumaran<br>& Jeanette Friend | Addition of troubleshooting section. Revision prior to release. |
|         |            |                                       |   |

## Distribution List

| Name                    | Title                        | Organisation / Team |
|-------------------------|------------------------------|---------------------|
| ABM Costing Team        |                              | ABM                 |
| Tony Hutton             | A/c A&I Team Manager         | ABM                 |
| Ranjit Sukumaran        | A&I Team Developer & Analyst | ABM                 |
| Emma Watson             |                              | CCLHD               |
| Sarah Wang              |                              | FWLHD               |
| Sandy Bull              |                              | HNELHD              |
| Chloe Laina             |                              | ISLHD               |
| Gamage Perera           |                              | Justice Health      |
| Thanaa Mosa             |                              | MLHD                |
| Tony Albuquerque        |                              | MNCLHD              |
| Susan Davies            |                              | NBMLHD              |
| Louise Savrda           |                              | NNSWLHD             |
| Sandra Thompson         |                              | NSLHD               |
| Micheline Maddaford     |                              | SCHN                |
| Andrew McDonell         |                              | SESLHD              |
| Tze-Yen Rossiter        |                              | SLHD                |
| Sharon McFarlane        |                              | SNSWLHD             |
| Anne-Marie Jimenez      |                              | SVHN                |
| Raymond Lal             |                              | SWSLHD              |
| Karen Storey/Steve Shea |                              | WNSWLHD             |
| Jennifer Killen         |                              | WSLHD               |

## Associated References

| Title   | Location  |
|---|---|
| NSW Health Cost Accounting Guidelines - Vol 3a Technical Overview | <a href="#">CAG Vol3a Technical Overview 14 Sept 2022.pdf (adobe.com)</a> |
| ABM Costing Requirements V0.7                                     | <a href="#">ABM - Costing Requirements v0.7 Work In Progress.xlsx</a>     |

## 1 Overview

The purpose of this version of the user guide is to document the installation instructions for the Costing Extractor V2, test that a successful connection to EDW can be achieved and that the Extractor process can run, resulting in a set of output files. The Costing Extractor v2 has been redeveloped in Python and is capable of connecting to either HIE or EDW as the data source, however for the purposes of testing, only a connection to EDW is to be used.

This document will be expanded to include details on how to use the Extractor in due course.

## 2 Getting Started – Extractor Installation

### 2.1 Pre-requisites

Before you download and run the Extractor v2 application, please ensure that the following pre-requisites are met.

#### 1. System/Workstation Requirements

- RAM: Minimum 4 GB, preferably 8 GB
- Processor: Minimum 1GHz
- OS: 64-bit Windows 10 OS
- Disk space: Up to 250Gb (Upper limit based on size of output datasets, which is determined by volume of data extracted and size of LHD)
- ODBC 64bit Driver
- SMTP for receipt of AMHCC flat file (LHDs/SHNs only) – Note, it is expected that the AMHCC data extraction may be available from EDW directly. This will be confirmed.

#### 2. Datawarehouse Access

- User access to EDW LRS FLAT schemas (Production environments). This is the State LRS (MOH LRS).
- Access to EDW is granted via the Data Warehouse Unit (DWU). For appropriate request forms refer: [EDW \(nsw.gov.au\)](https://www.nsw.gov.au/edw) and note that patient level data access is required. Please note that the lead time for access is usually a minimum of 2 weeks.

#### 3. Costing Collaborative Space Access

- This is the ABM secured network space designated for LHD Costing Teams and from where the Extractor v2 executable will need to be downloaded from. Access to this space, if you do not already have access, is via request to Kylie Hawkins, Manager, Clinical Cost Data Collections and Standards ([kylie.hawkins2@health.nsw.gov.au](mailto:kylie.hawkins2@health.nsw.gov.au))

### 2.2 Download & Unzip Extractor v2 Executable

1. Go to the Costing Collaborative Space and download the Extractor v2 zip file located here → [EDW Extractor Testing](#)

You will also find a copy of this User Guide document and the Checklist in this folder.





2. Download the zip file to the location you wish to run it from and unzip the contents to your local machine. For example, to run the extractor from D:\ drive, download the costing\_extractor zip file to D: and unzip. The Extractor v2 folder, sub-folders and key files will be installed to D:\costing\_extractor\.. The high level folder is: “costing\_extractor”

**Notes/Tips:**

- The Extractor v2 zip file you are downloading is approximately 250MB in size, so it might take time to download to your local machine.

## 2.3 Extractor v2 Folder Structure

After unzipping the Extractor, the following high-level folders under `..\costing_extractor\` will be created.

- ✓  `build`
  - >  `extractor`
- ✓  `dist`
  - >  `extractor`

### 2.3.1 Build\Extractor Folder

The “`..\costing_extractor\build\extractor`” folder contains the results of analysis and additional logs. You can ignore the contents of this folder unless you are trying to debug issues.

### 2.3.2 Dist\Extractor Folder

The “`..\costing_extractor\dist\extractor`” folder contains a number of subfolders and files. This includes the Extractor application, bundled as an executable file, together with any associated libraries and binary .dll files.

Everything necessary to run the Extractor application will be under this folder. There are key files which should not be moved or modified. These are:

1. Extractor Configuration file, located in the path `..\costing_extractor\dist\extractor\config.ini`  
This file contains details of the previous run so that user does not have to run the whole extraction in future for the same input parameters.
2. Extractor logo for the initial splash screen when the user runs the application, located in the path `..\costing_extractor\dist\extractor\hssg_logo.jpg`

The following sections detail the key subfolders and files under the `..\costing_extractor\dist\extractor` folder that are used in Extractor processing.

#### 2.3.2.1 Costing Sub-Folder

The costing subdirectory will be where input files are located. The path is:

`..\costing_extractor\dist\extractor\Costing`

This subfolder is where files provided to you by ABM Costing team should be located and where your specific LHD input files should be located. The following screenshot illustrates the types of files that are in this sub-folder.

> dist > extractor > Costing

| Name                    | Date modified       | Type                  | Size     |
|-------------------------|---------------------|-----------------------|----------|
| AMHCC_Extract           | 22/02/2023 12:01 PM | Microsoft Excel C...  | 170 KB   |
| AMHCC_ExtractBlank      | 23/02/2022 2:28 PM  | Microsoft Excel C...  | 59 KB    |
| Class_Descriptions      | 7/07/2020 7:43 AM   | Microsoft Excel C...  | 279 KB   |
| CriticalCareGroup       | 5/07/2022 12:29 PM  | Microsoft Excel C...  | 42 KB    |
| DRGStandardWeights      | 3/09/2020 7:00 PM   | Microsoft Excel C...  | 191 KB   |
| DrgWeightMaker          | 3/09/2020 7:00 PM   | Microsoft Excel C...  | 20 KB    |
| EDRoleDelin             | 22/09/2021 9:22 AM  | Microsoft Excel C...  | 7 KB     |
| FinclassToFinType       | 3/03/2022 2:22 PM   | Microsoft Excel C...  | 2 KB     |
| ICU_RoleDelin           | 22/09/2021 9:22 AM  | Microsoft Excel C...  | 1 KB     |
| MDC                     | 3/09/2020 7:00 PM   | Microsoft Excel C...  | 8 KB     |
| PLA_AMHCC               | 3/09/2020 7:00 PM   | Microsoft Excel C...  | 4 KB     |
| PLA_Mapping_00          | 3/09/2020 7:00 PM   | Microsoft Excel C...  | 70 KB    |
| PLA_Role_Table          | 28/07/2022 1:31 PM  | Microsoft Excel C...  | 11 KB    |
| ReconciliationGroup     | 3/09/2020 7:00 PM   | Microsoft Excel C...  | 1 KB     |
| RoundDetails            | 21/08/2023 6:13 PM  | Microsoft Excel C...  | 1 KB     |
| SNAP_CareTypeToProdType | 3/09/2020 7:00 PM   | Microsoft Excel C...  | 1 KB     |
| SNAP_CaseTypeToProdType | 3/09/2020 7:00 PM   | Microsoft Excel C...  | 1 KB     |
| SNAP_CostingExtract     | 16/02/2023 8:14 AM  | Microsoft Excel 97... | 1,655 KB |
| SNAP_NWAU               | 23/02/2022 2:37 PM  | Microsoft Excel C...  | 1 KB     |
| SNAP_NWAU               | 16/02/2023 8:16 AM  | Microsoft Excel 97... | 659 KB   |
| SNAPRec                 | 23/02/2022 2:49 PM  | Microsoft Excel C...  | 1 KB     |
| SNAPRec                 | 16/02/2023 8:18 AM  | Microsoft Excel 97... | 24 KB    |
| SNAPStandardWeights     | 3/09/2020 6:00 PM   | Microsoft Excel C...  | 13 KB    |
| SpecialtyPortalMapping  | 24/06/2023 12:02 PM | Microsoft Excel C...  | 99 KB    |
| SpecialtyPortal         | 22/09/2021 9:22 AM  | Text Document         | 2 KB     |
| tbl_ExcludedEncounters  | 16/02/2023 2:01 PM  | Microsoft Excel C...  | 3 KB     |
| tbl_ExcludedEncounters  | 3/09/2020 7:00 PM   | Text Document         | 1 KB     |
| Tbl_PPM_transfer_AMO    | 24/06/2023 11:23 AM | Microsoft Excel C...  | 191 KB   |

If you need to change the parameter file (RoundDetails.csv) or update other input files, go to this sub-folder.

### Notes/Tips:

- Some of the files in the Costing sub-folder are only to be modified by the ABM Costing Team.

#### 2.3.2.2 ExtractorDB Subfolder

This folder contains the Extractor database where the EDW extracted data as loaded into, in addition to any flat files for post processing and transformation. There may be other staging files in this folder which should not be directly modified.

The path is `..\ costing_extractor\dist\extractor\ExtractorDB`

#### 2.3.2.3 Output Subfolder

The Output folder is where the output files generated by the Extractor will be located and this folder will be populated once the Extractor v2 application is run to completion.

The path is `..\ costing_extractor\dist\extractor\Output`

### Notes/Tips:

- Each successive Extractor run will overwrite the output files located in this folder. If you wish to retain the files from previous runs, please copy those output files to a different location.

#### 2.3.2.4 Event Log File

The event log will be generated each time the Extractor is run and records details of the processing undertaken. It is used for debugging purposes in the event the data extraction and transformation process fails. The content of this file will look similar to the following screenshot.

```

30-Jan-2024 10:37:59 - STARTING COSTING EXTRACTOR.
30-Jan-2024 10:37:59 - Import modules completed.
30-Jan-2024 10:37:59 - Created Costing, ExtractorDB, Output, temp_transform folders, if they did not exist before.
30-Jan-2024 10:37:59 - List of 64 bit DSN = {'ABM CASEMIX PROD': 'ODBC Driver 18 for SQL Server', 'CNQUATIPHA': 'ODBC Driver 18 for SQL Server', 'HIE_64bit': 'Sybase IQ', 'HIE_32bit': 'SQL Anywhere 11', 'LRS_MOH': 'SQL Server', 'LRS_MOH
30-Jan-2024 10:37:59 - Display splash screen.
30-Jan-2024 10:38:04 - The variance between the SNAP Costing Extract File and the SNAP Report is = 0
30-Jan-2024 10:38:05 - Checking if mandatory files exist

```

The event log is found under the Extractor subfolder. The path and filename is:

..\costing\_extractor\dist\extractor\python\_costing\_extractor\_log.txt

### Notes/Tips:

- The event log will be overwritten each time the Extractor v2 is run. If you wish to retain the event log from a previous run, please copy it to a different location prior to re-running the application.

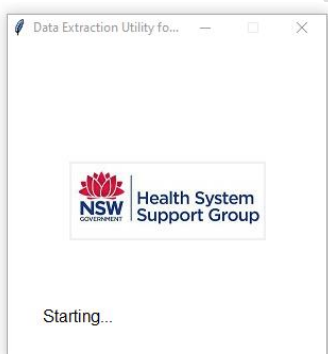
## 2.4 Running the Extractor

After completing the download and any changes to the input files in Costing\ subfolder, the application itself can be run.

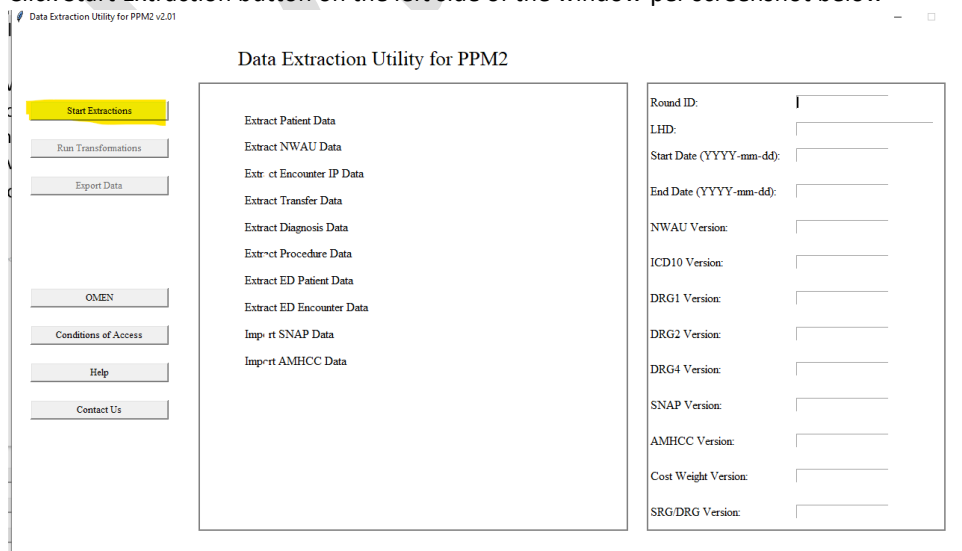
- Go to the folder where the application is located. The path is ..\costing\_extractor\dist\extractor
- Double click the extractor application icon shown in the screenshot below.

|           |                    |             |           |
|-----------|--------------------|-------------|-----------|
| config    | 20/08/2023 3:20 PM | INI File    | 1 KB      |
| extractor | 21/08/2023 6:08 PM | Application | 17,952 KB |
| extractor | 21/08/2023 4:35 PM | PY File     | 1,346 KB  |
| hssg_logo | 5/05/2023 11:31 AM | JPG File    | 8 KB      |

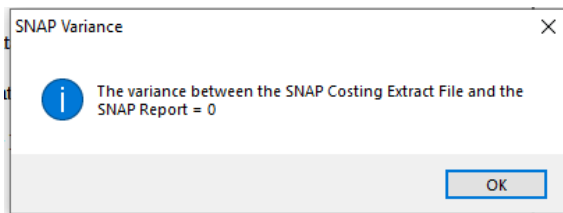
- The splash screen will appear.



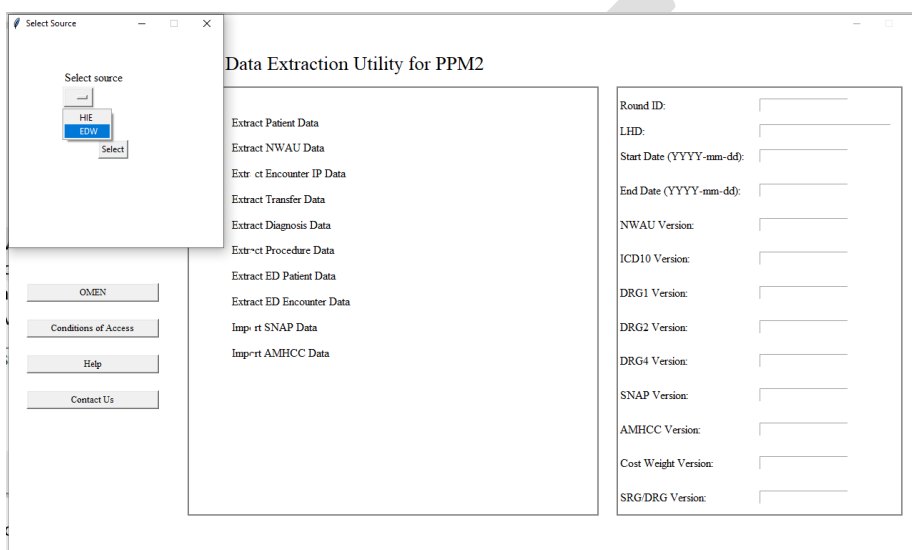
- Click Start Extraction button on the left side of the window per screenshot below



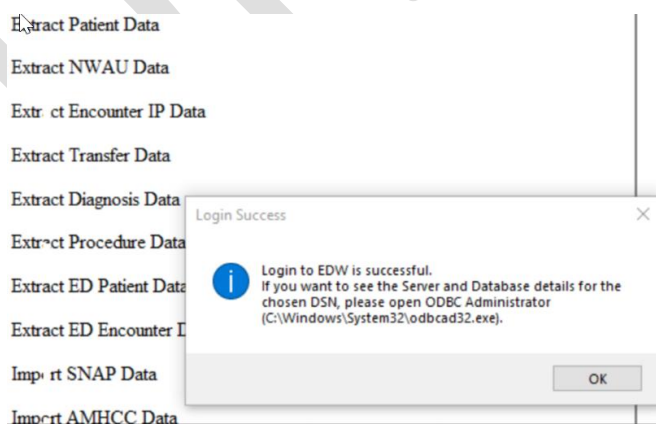
- When you click extraction, the program will check that all files are present and that the SNAP record count captured in SNAPRec agrees with the records count in the two SNAP extract files. If the SNAP input file is not present, the Extractor will continue without it, and similarly if the AMHCC file is not available, processing will also continue. Just select OK per screenshot below.



- You will then be prompted to select the database source. Select EDW.



- You should get a message to confirm that you have successfully connected to EDW per the screenshot below. If you do not receive this message, refer to 3 Troubleshooting

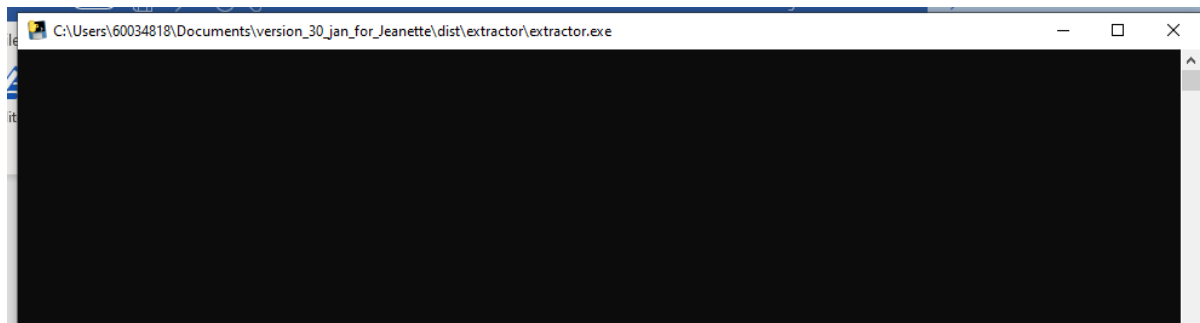


## Notes/Tips

**Important:** A black console window will open along with the costing extractor application. The black console window will show errors that have not been captured by the log file. If the application errors and the console

window show the error message, please take a screenshot of the error message and send the screenshot along with the log file to the support team.

The following screenshot illustrates the console window.



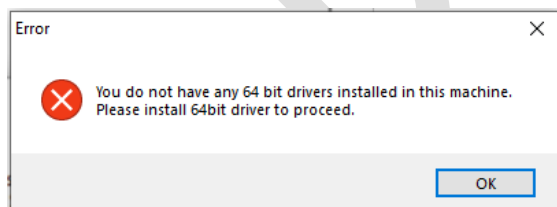
### 3 Troubleshooting

The following section describes common issues and suggestions to resolve them.

#### 3.1 Unable to connect to EDW

Reasons why you may be unable to connect to EDW include not having access to the environment required or not having the ODBC driver correctly set up. Please be aware that access granted to pre-production environments does not automatically include access to EDW production.

Assuming your EDW access has been granted, if you are unable to successfully connect to EDW you may need to configure the ODBC driver, especially if you receive a message similar to the following:

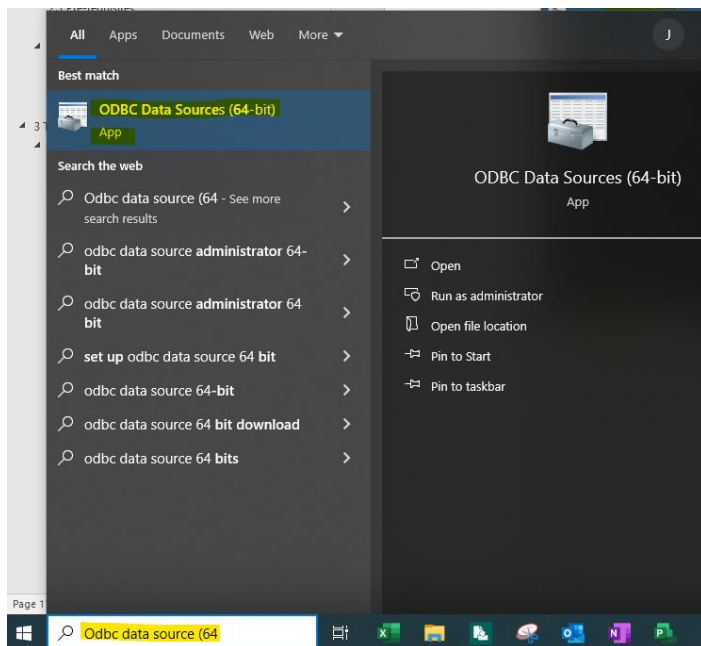


If this is the case Refer: 3.1.1 Configuring ODBC Driver for EDW

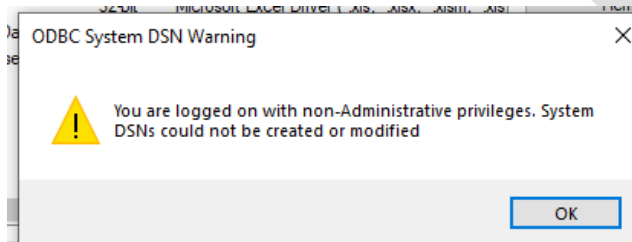
##### 3.1.1 Configuring ODBC Driver for EDW

You may need to configure the ODBC driver to be able to connect to EDW. To do so, please follow these instructions.

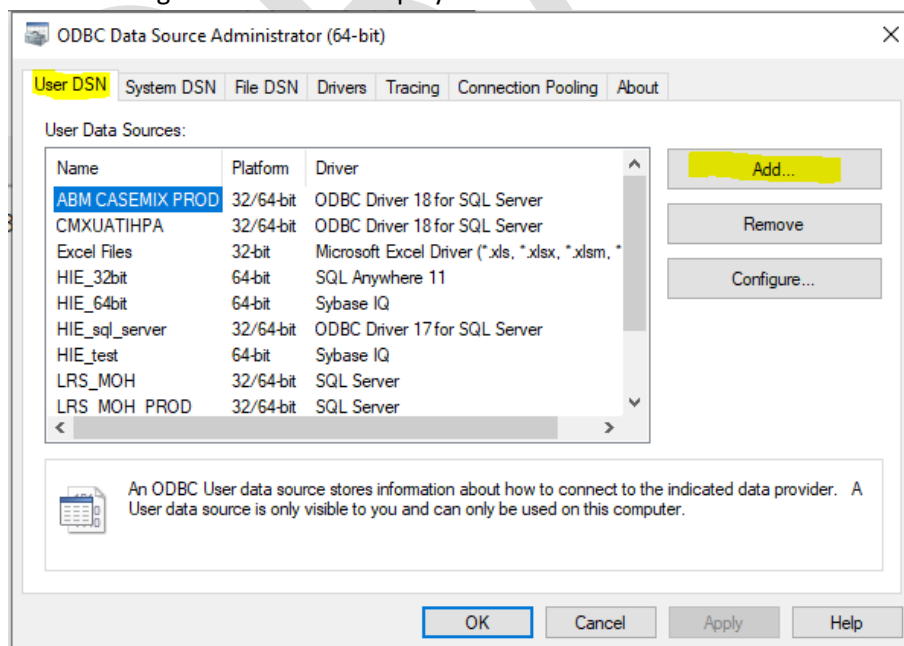
1. In windows search bar, type ODBC Data Sources (64-bit) per the screenshot below and select ODBC Data Sources (62-bit) App. If ODBC Data Sources (64-bit) is not available, search with "ODBC Administrator" and select that application.



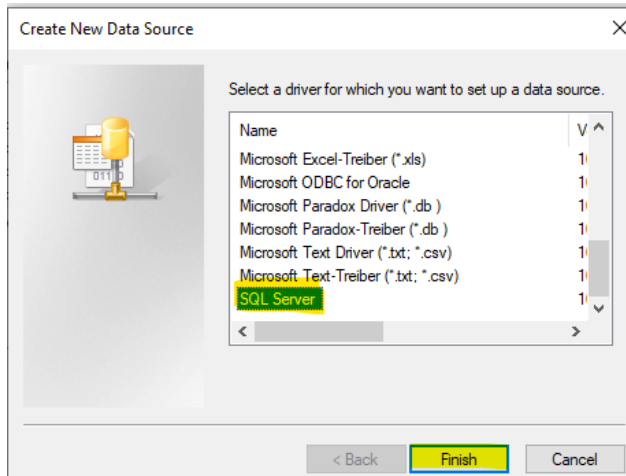
2. If you receive the following warning, just hit "Ok"



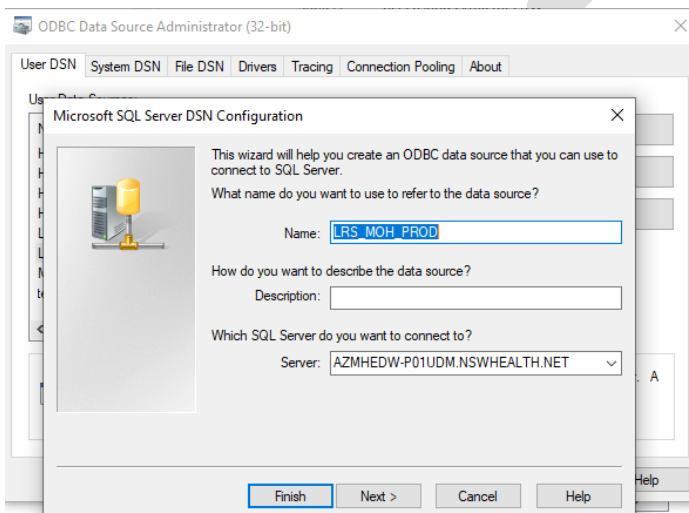
3. The following window will be displayed. Select "User DSN" tab and then the "Add" button.



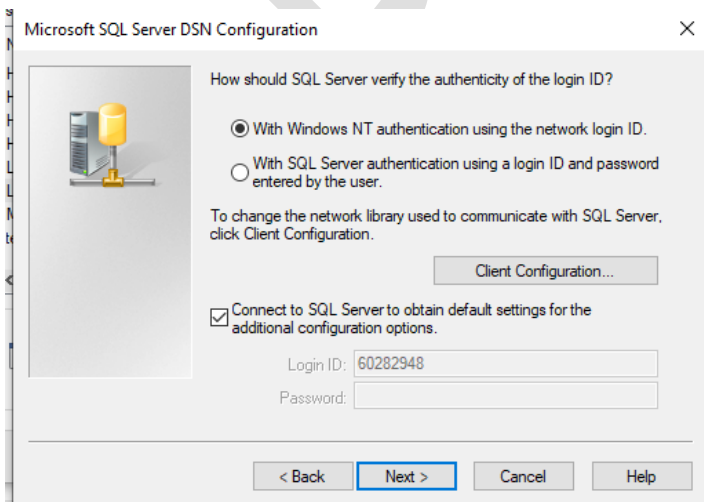
- Select SQL Server and then click on the “Finish” button



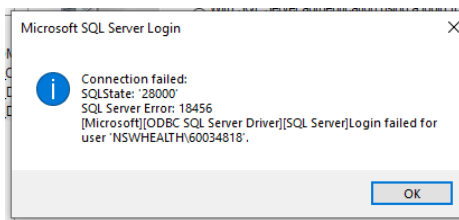
- The following screen will be displayed. Complete the details as shown below and click Next.



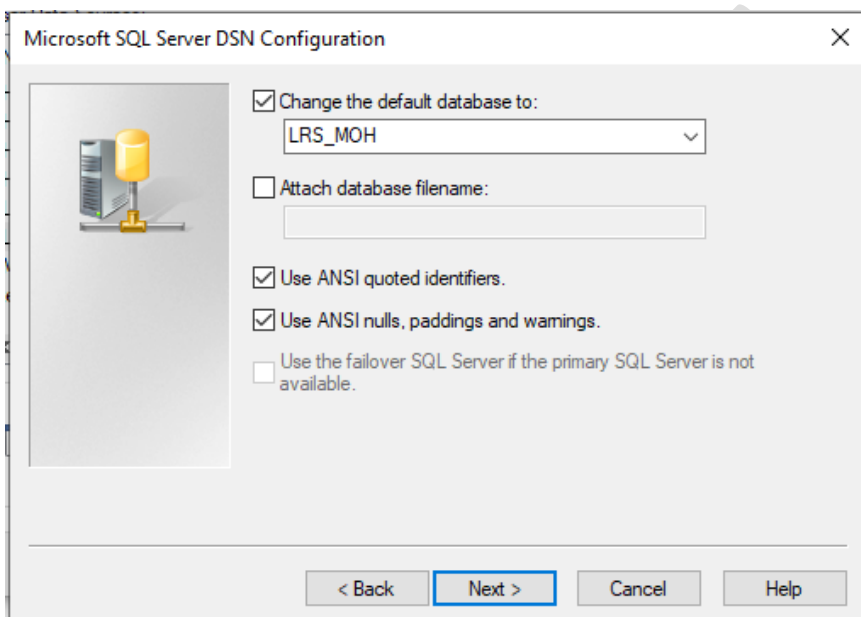
- The following screen will be displayed. Ensure the radio button “With Windows NT authentication using the network login ID” is selected, unless you have a different ID that you use to connect to EDW with. Select Next.



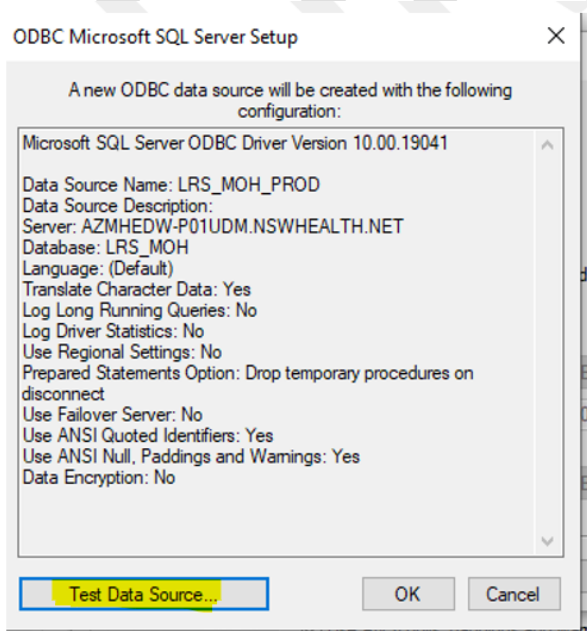
7. If you do not have access permission to EDW, the following message will be displayed. Hit OK and the ODBC setup will end. At this point you will need to resolve your EDW access via a SARA ticket or similar.



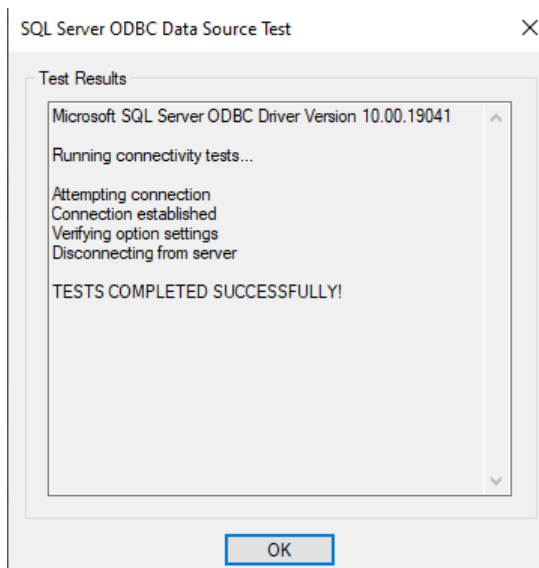
8. If you do have permission to access EDW, the following screen will be displayed. Rename the database to LRS\_MOH and select Next and then Finish.



9. To test the connection, select the "Click Test Data Source" button.



10. If connection to the EDW LRS is successful, you will receive the following message. Hit OK.



When you run the Extractor, you will need to select ODBC and select the data source name of the ODBC connection that you just configured i.e. LRS\_MOH\_PROD

### 3.1 Slow to Extract Data from EDW

Data extraction time from EDW is quite variable and it may take several hours to extract the required data for transformation purposes. Do not be surprised if this is the case. As EDW is a new solution, peak usage times and response times are still being analysed and performance may improve over time.