Install Guide - Setup Wizard - Microsoft Fabric

WhereScape Enablement Pack for Microsoft Fabric-RED 10.2+

This is a guide to installing the WhereScape Enablement Pack for Microsoft Fabric in WhereScape RED10

Table of Contents

Prerequisites For PostgreSQL Metadata

Before you begin the following prerequisites must be met:

- Create Database and ODBC DSN :
  - Supported* version of PostgreSQL (PostgreSQL 12 or higher)
    - A database to house the RED Metadata Repository.
    - A database for the Range Table DB (Optional)
    - A database to house scheduler (Optional)
- Software Installations
  - WhereScape RED10 with valid license key entered and EULA accepted
  - WhereScape Enablement Pack for target database version RED10
- Windows Powershell (64 bit) version 4 or higher
  - To check Windows Powershell Version:
    - Run below command in Windows Powershell

```
Get-Host|Select-Object Version
```

    - Run below command in Command Prompt

```
powershell $psversiontable
```

- Run the following command using PowerShell
  - The security protocol TLS 1.0 and 1.1 used by PowerShell to communicate with PowerShell gallery has deprecated and TLS 1.2 has been made mandatory

```
[Net.ServicePointManager]::SecurityProtocol = [Net.ServicePointManager]::
SecurityProtocol -bor [Net.SecurityProtocolType]::Tls12
Register-PSRepository -Default -Verbose
Set-PSRepository -Name "PSGallery" -InstallationPolicy Trusted
```

  - Progress bar placeholder info line

```
Install-Module -Name PoshProgressBar -SkipPublisherCheck -Force
```

* : *RED supports the following versions for the metadata repository: PostgreSQL 12 or higher*

Prerequisites For Microsoft Fabric

Before you begin the following prerequisites must be met:

- For **Windows**
  - Install the latest Azure-CLI package - Latest MSI of the Azure CLI (64-bit)
- For **Linux**
  - Install latest Azure CLI package using command:

```
curl -sL https://aka.ms/InstallAzureCLIDeb | bash
```

1. Open the Command prompt :

```
az login
```

   - Login using an Azure account having a Microsoft Fabric subscription

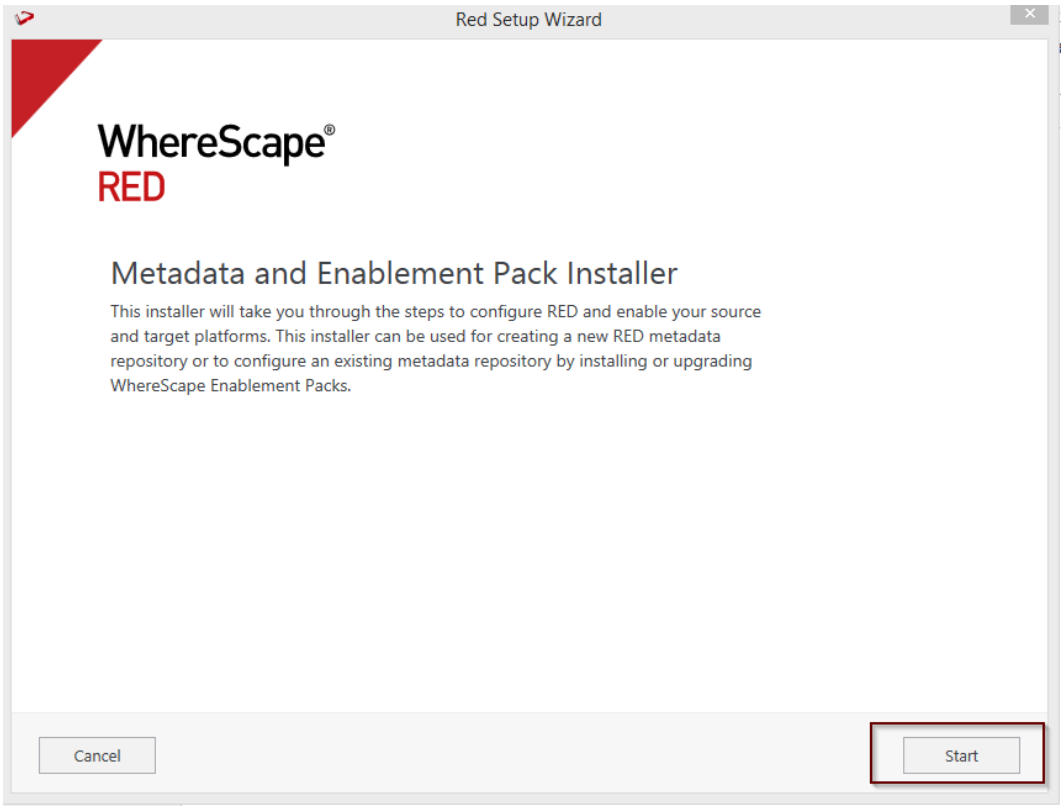
1. Create Database and ODBC DSN :
   - Microsoft Fabric
     - At least one schema available to use as a RED Data Warehouse Target
2. Python 3.11 or higher
   - Select "Add Python 3.11 to PATH" from installation Window
   - Pip Manager Install with command :

```
python -m pip install --upgrade pip
```

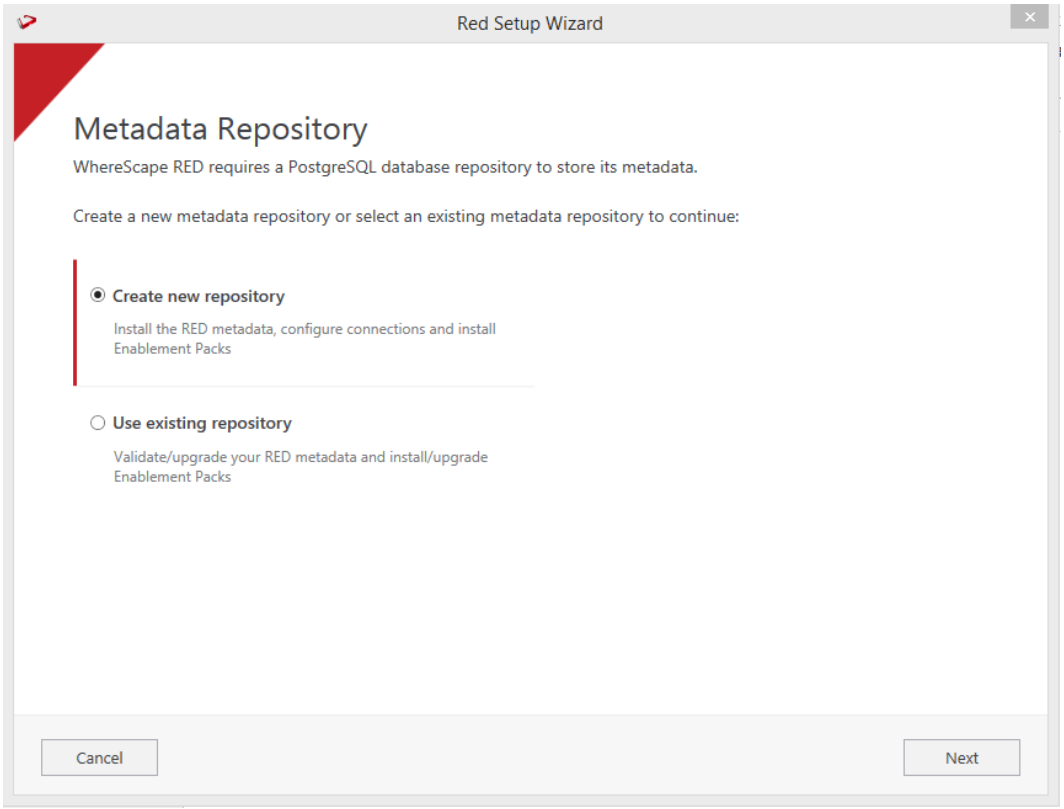
* We recommend using case-insensitive (CI) collations during the creation of your Fabric Warehouse

Installation Through Setup Wizard

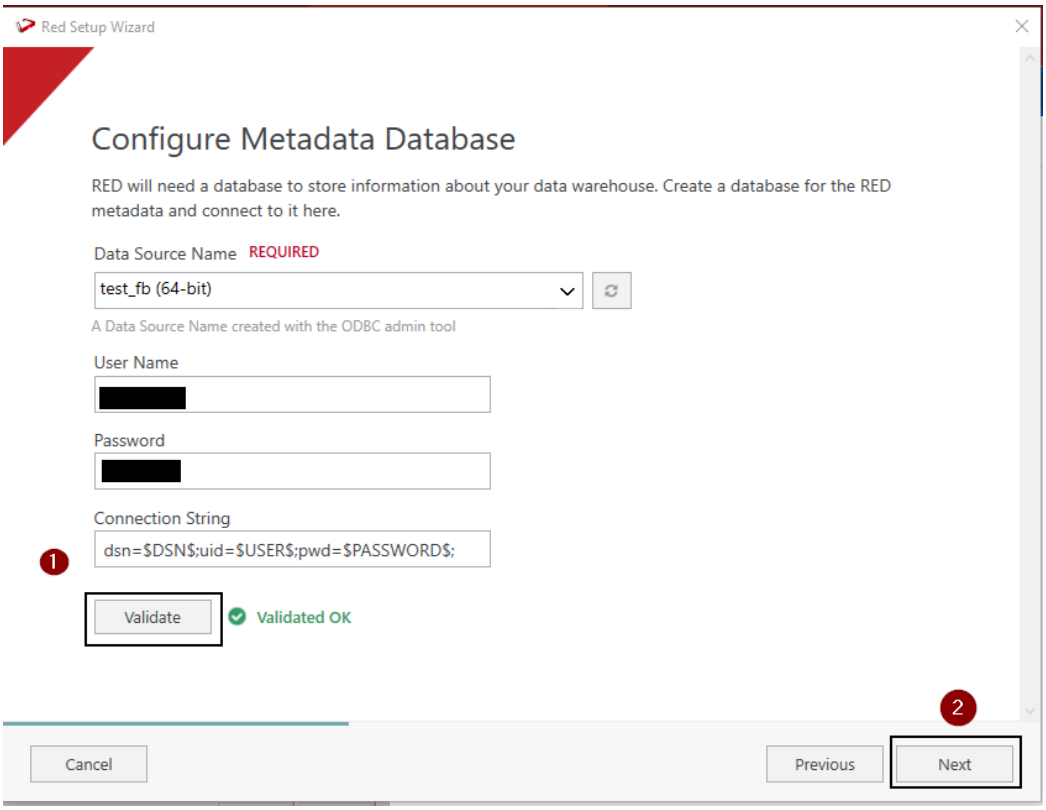
Run Setup Wizard as administrator



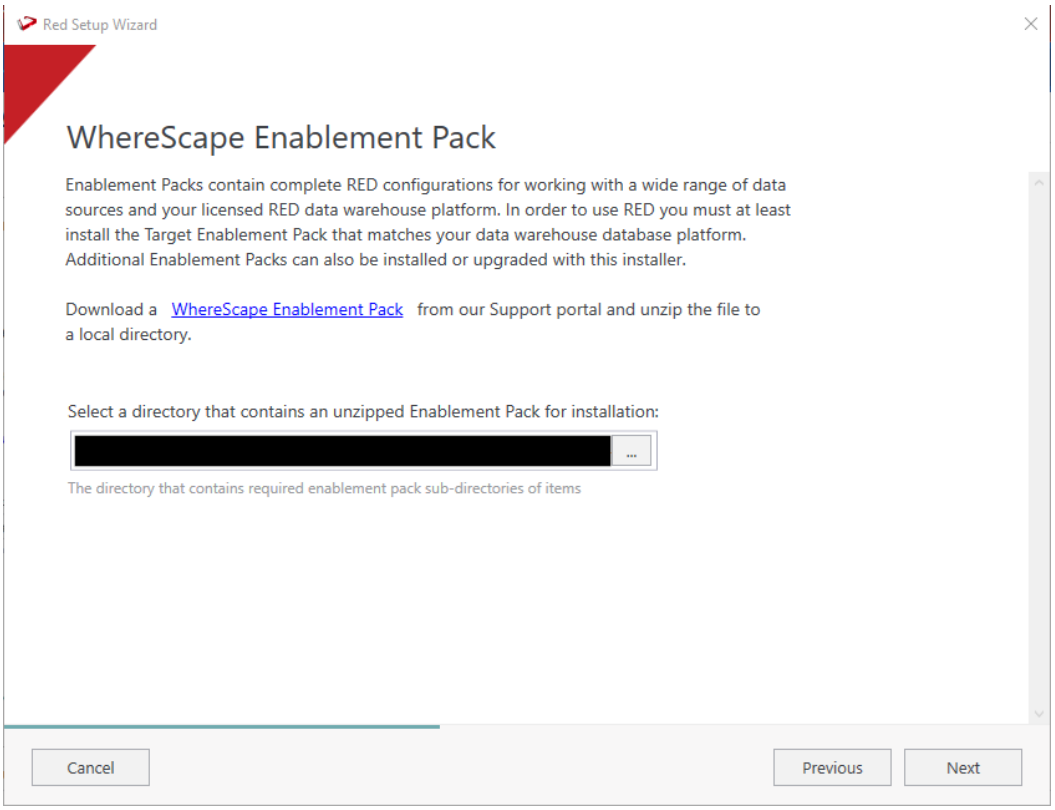
Create new repository or upgrade already existing repository.



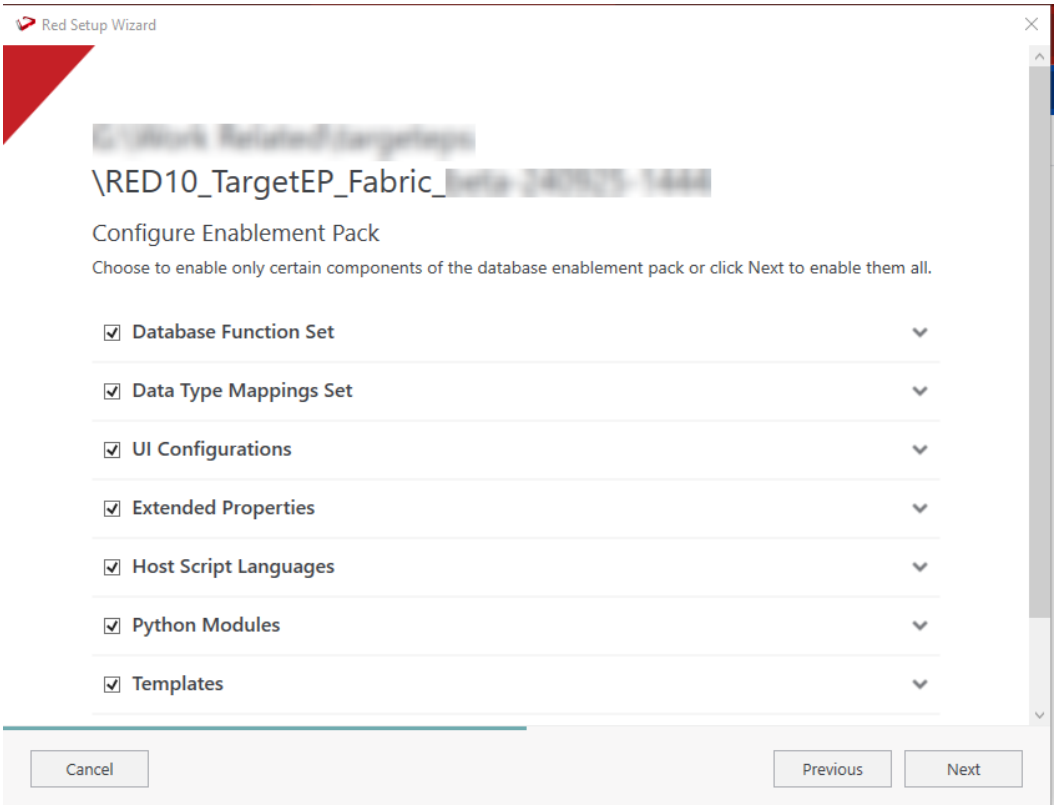
Select the created ODBC DSN, input login details with connection string and then select "Validate". Press Next



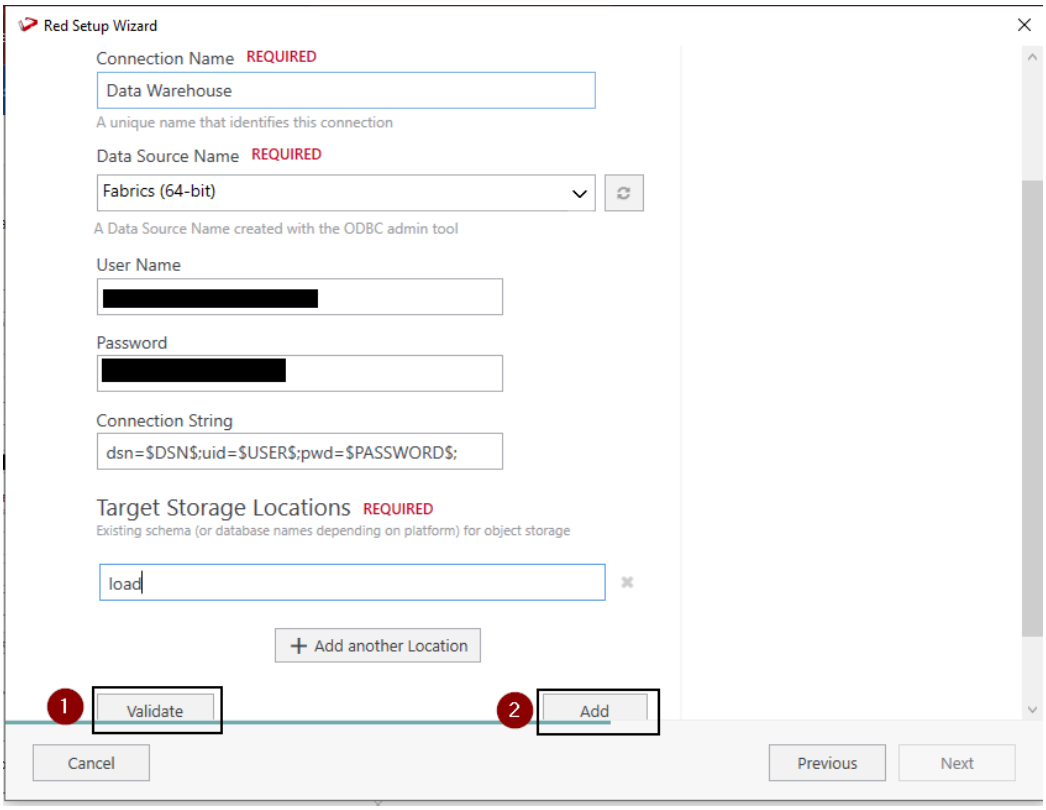
Select the directory that contains unzipped Enablement Pack for installation. Press Next



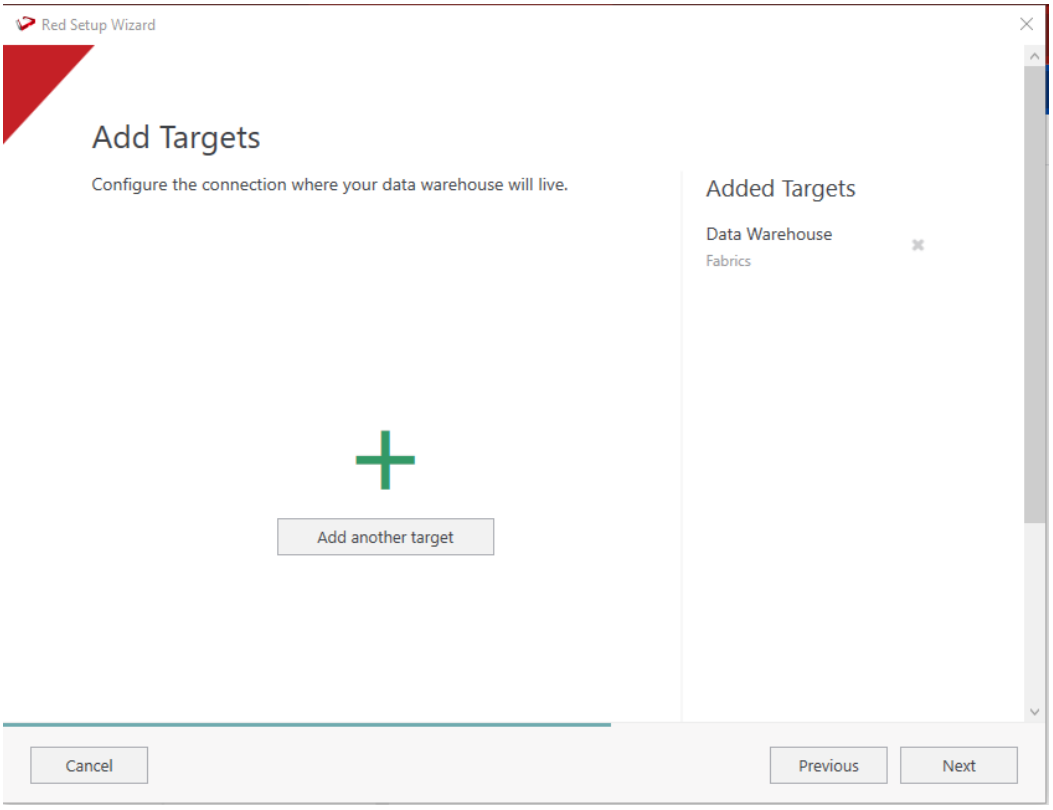
Using the check boxed list, include or exclude the components that are to be installed. Press Next



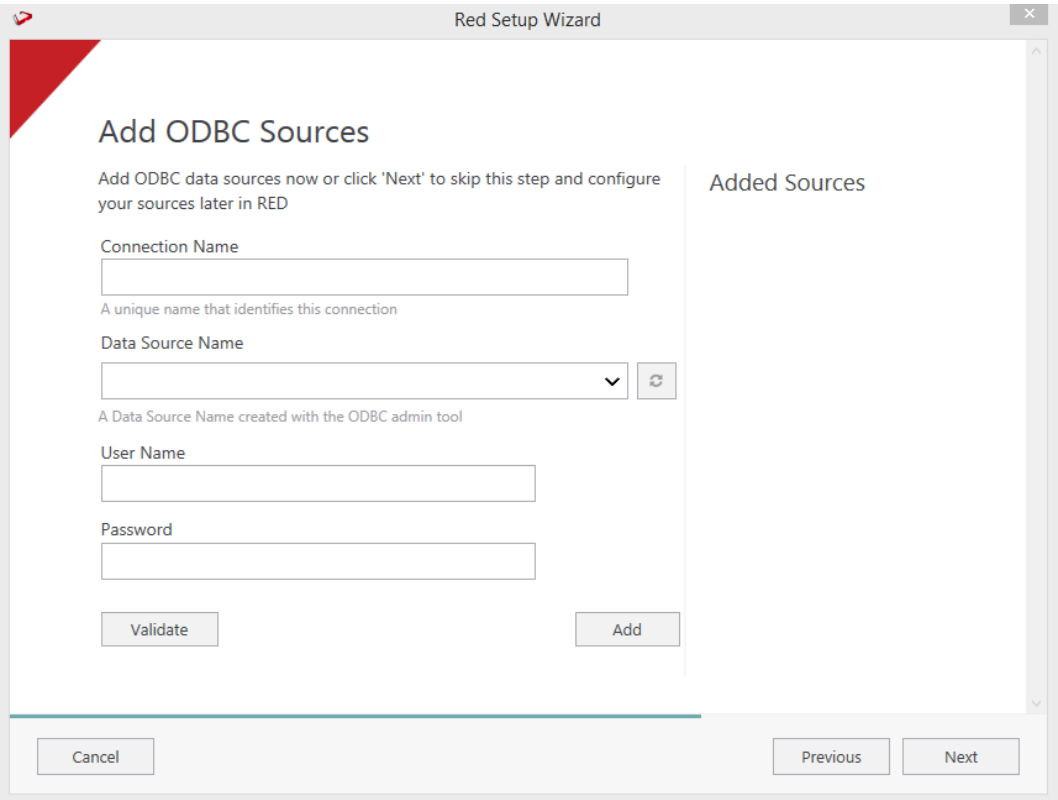
Configure a target connection (example, Data Warehouse) and its target locations.Validate and press ADD.



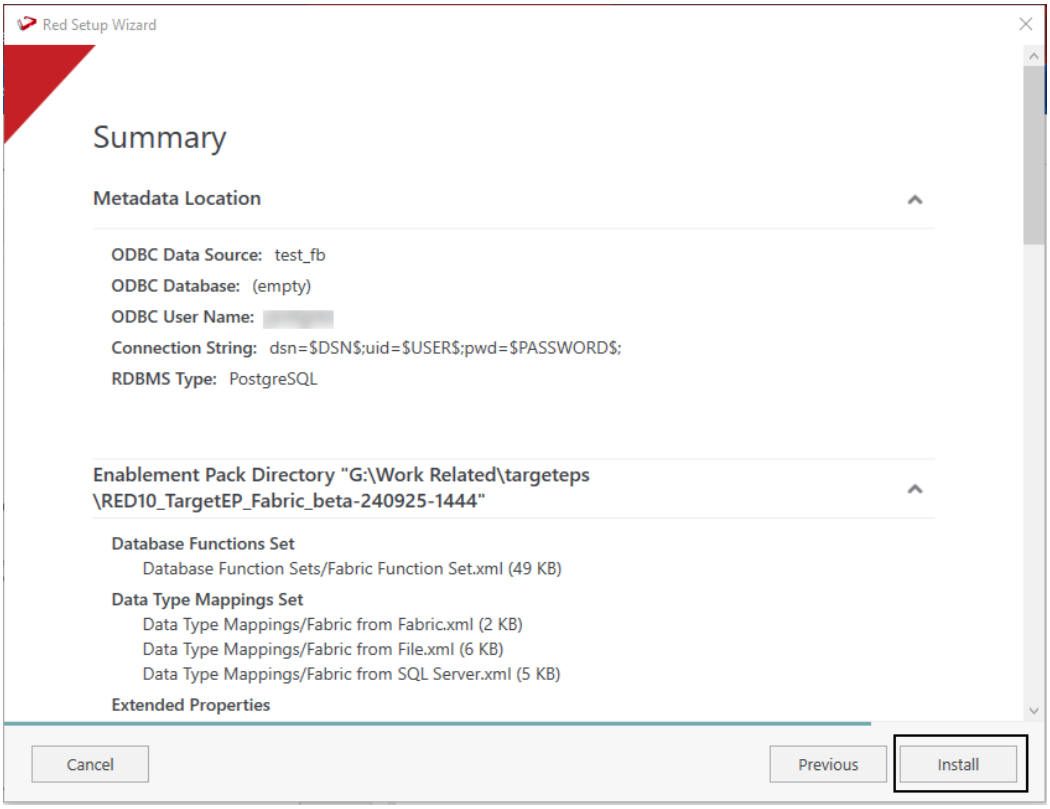
When done, press ADD and then Press Next to advance.



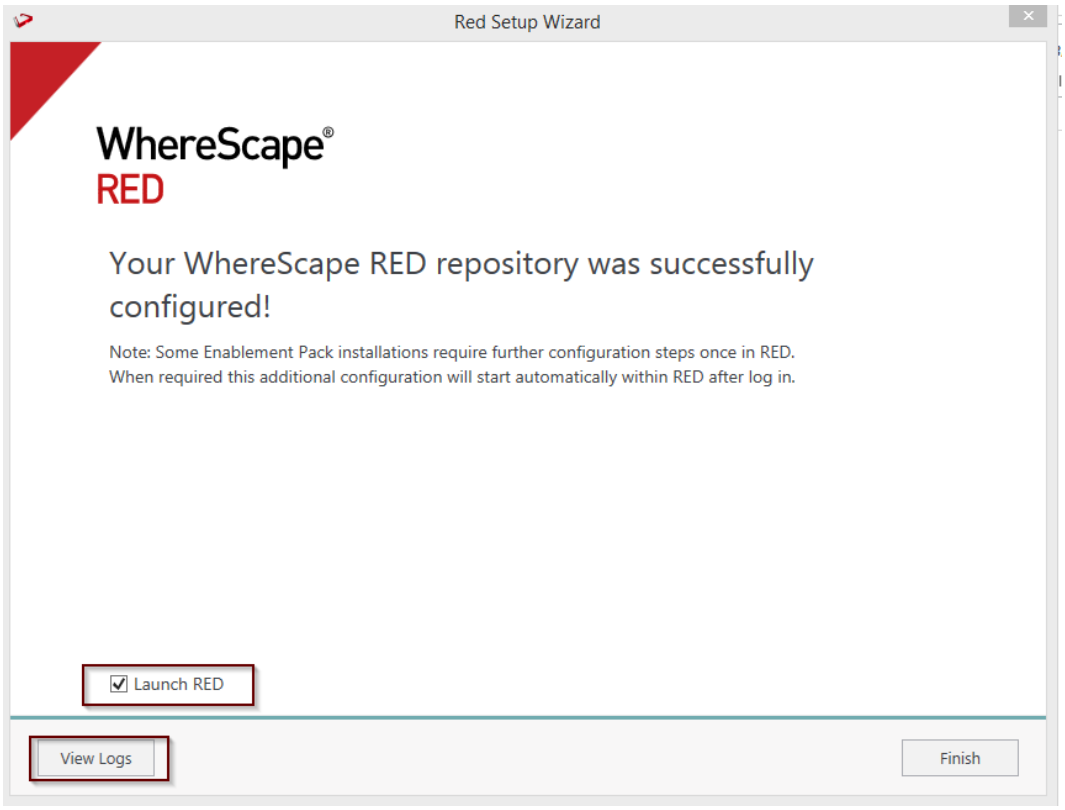
Configure a data source connection (optional) and its target locations. Validate and press ADD. Press Next to advance.



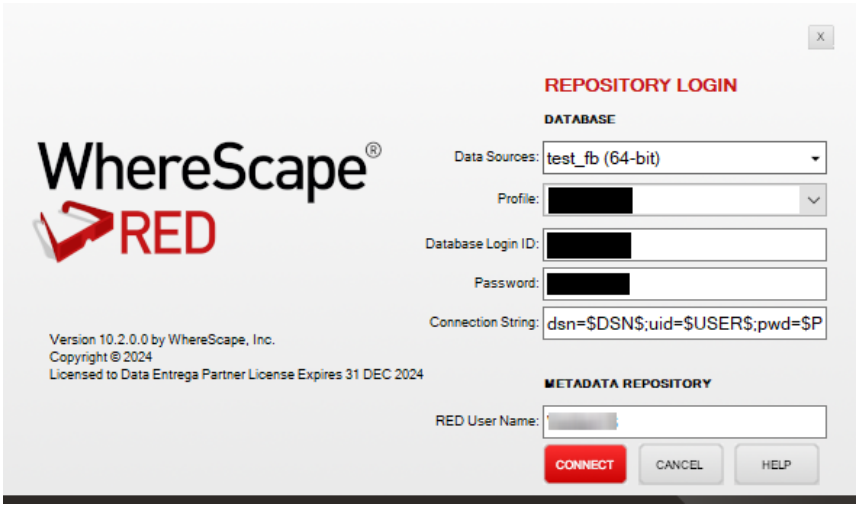
Review the installation summary and press Install



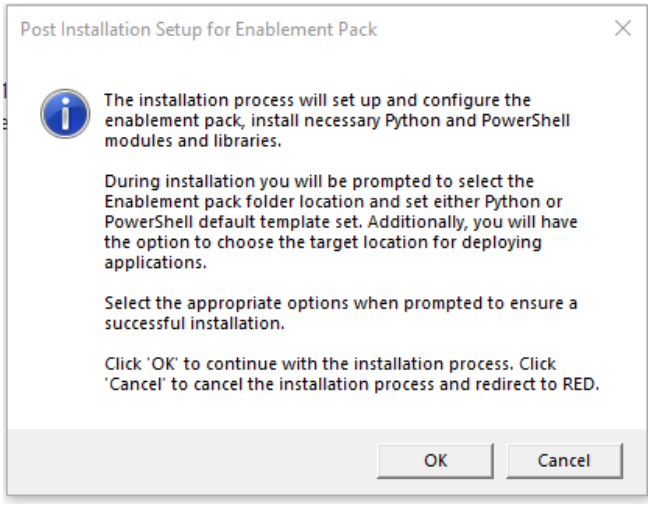
Clicking on the View Logs will take to the installation log. Click on Finish once the installation is completed successfully.



Login to WhereScape RED.

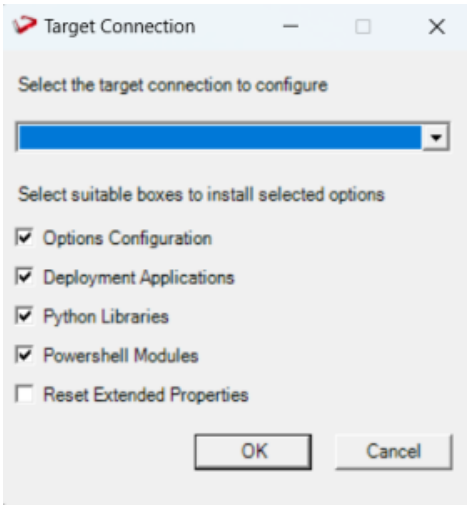


Note: There is a post-install script that will run at the first login to RED10 to complete the post setup wizard installation process.
You will be directed to below PowerShell window which will give brief explanation about post installation process.

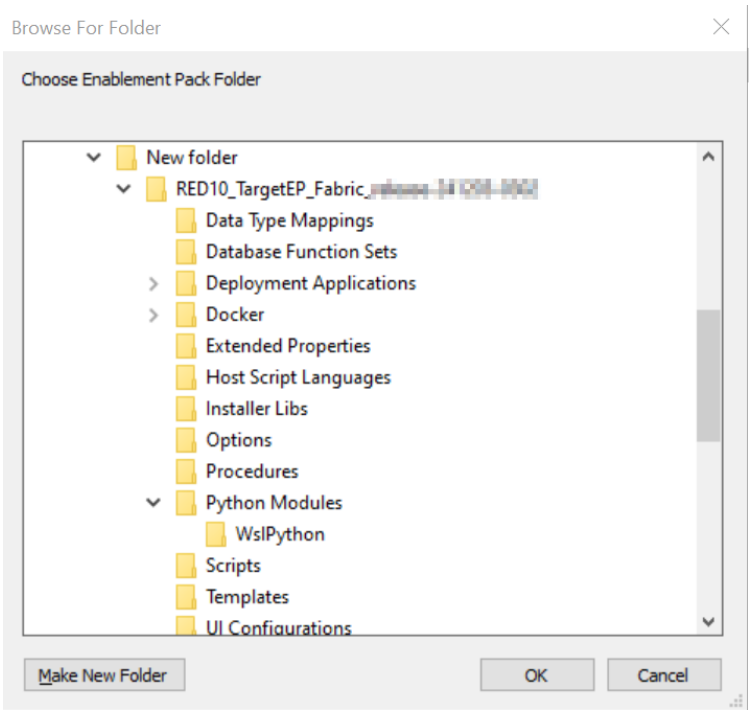


Press OK to start the post installation. If pressed Cancel installation will stop and user will be directed to RED.

The user will be directed to the window below, where they have to select the target connection to be configured. Additionally, by deselecting the provided options, the user can choose not to install a particular option. "Reset Extended Properties" is deselected by default.

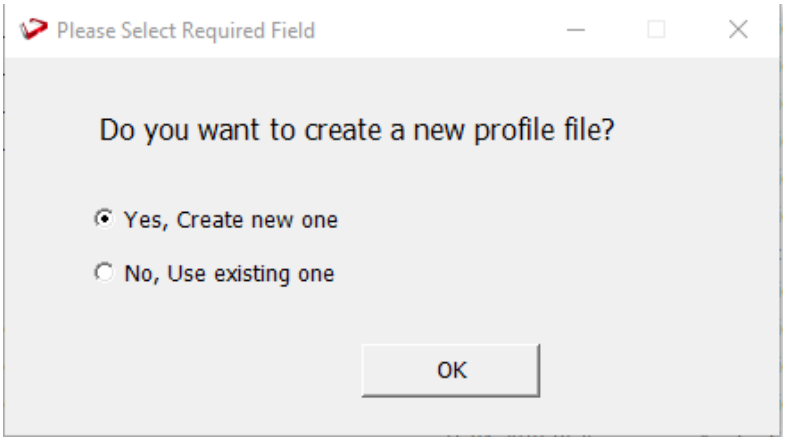


You will be directed to below PowerShell window. Provide the directory that contains unzipped Enablement Pack.



Press OK

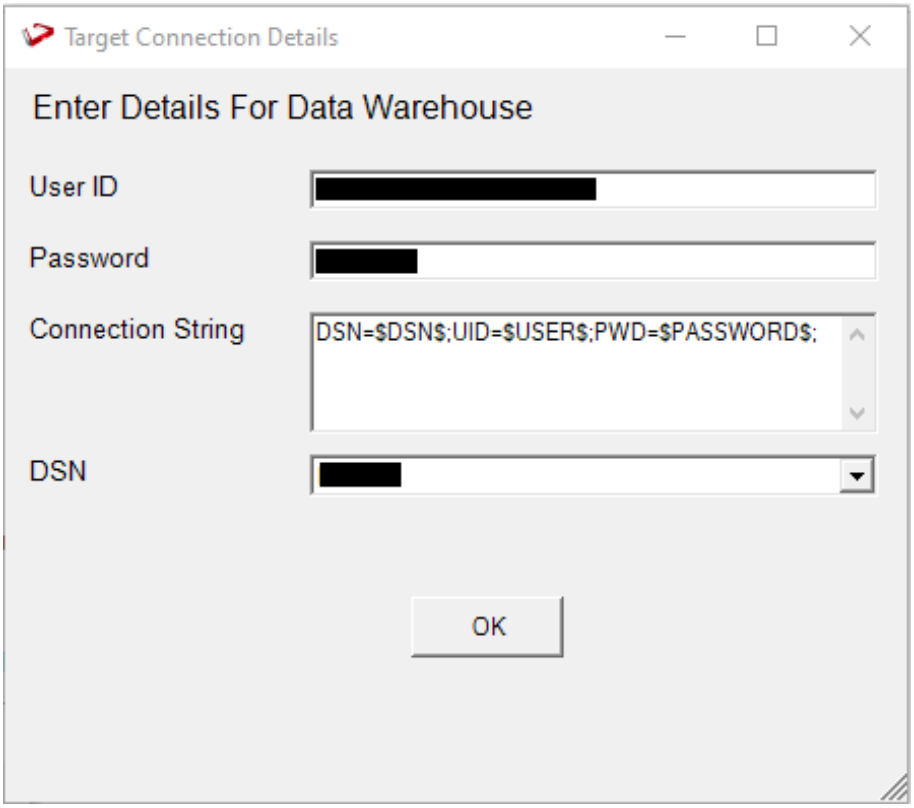
The user will be directed to the window below, where they have to select the Create new profile or use existing one option.



Note: For fresh installation RED will create profile file with same name as DSN, which the user can use or choose to create new profile file.

Press Ok.

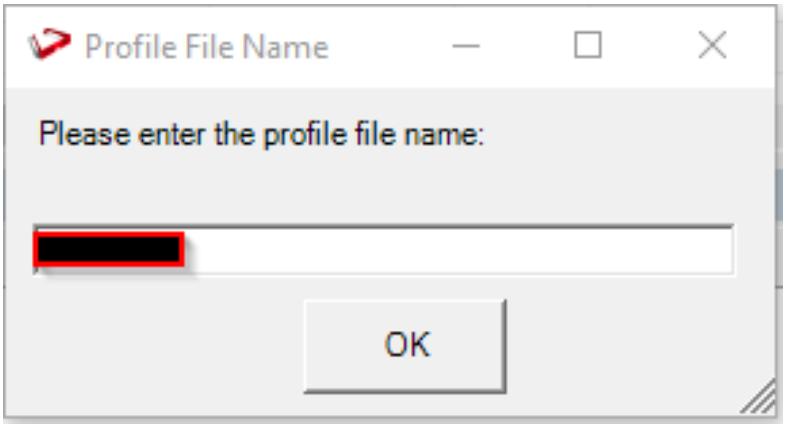
For "Yes, Create new one" option , user will be directed to the window below.



Note: User can use default connection string or input new one.

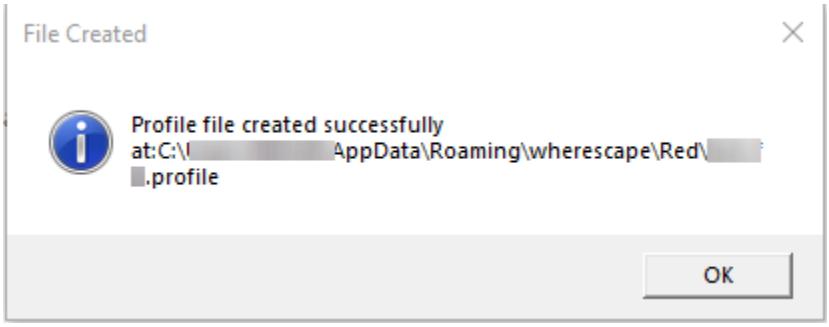
Press OK

The user will be directed to the window below, where user can add profile name.



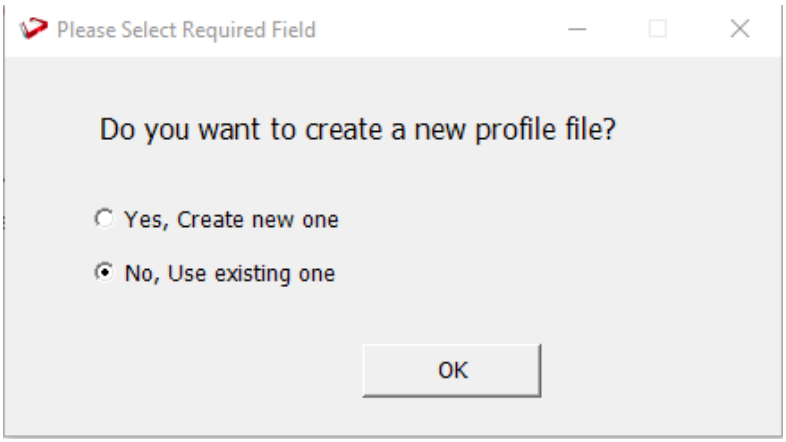
Press Ok.

The below pop up will come to confirm the user that profile is created at that location



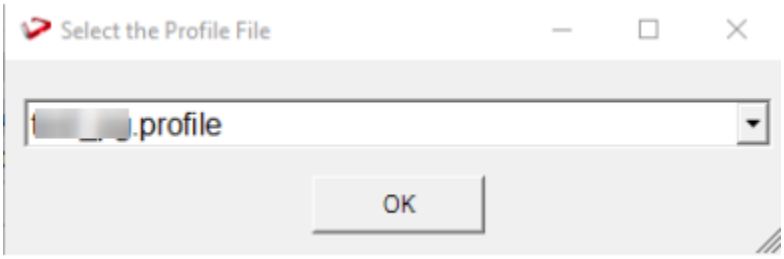
Press OK.

If the user choose "No, Use existing one" option.



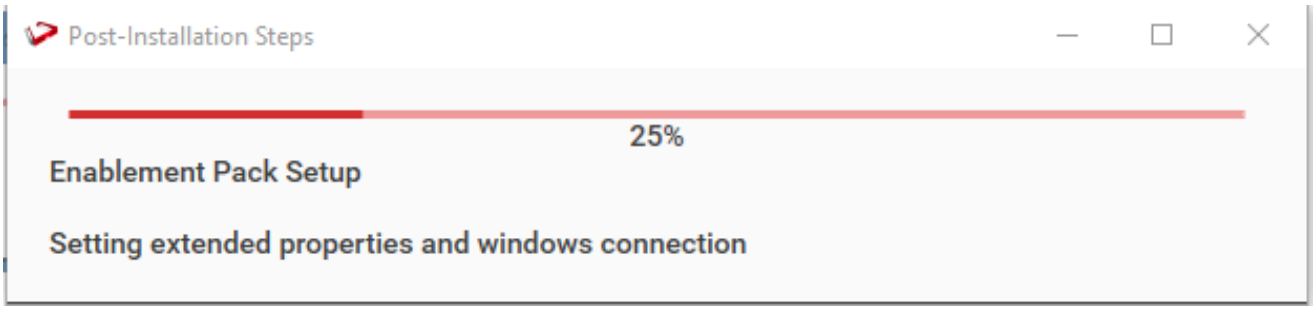
Press OK

The user will be directed to the window below ,where user can select the exiting profile file.

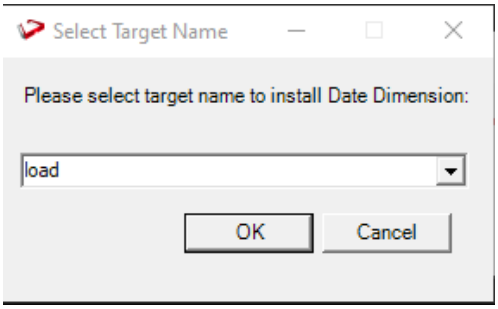


Press OK.

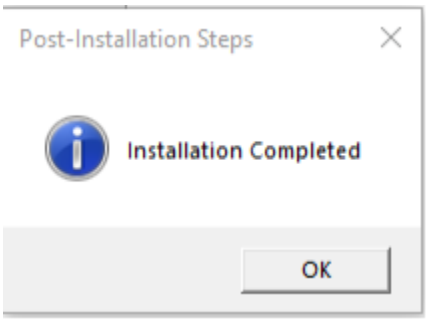
The progress bar will show the post installation progress.



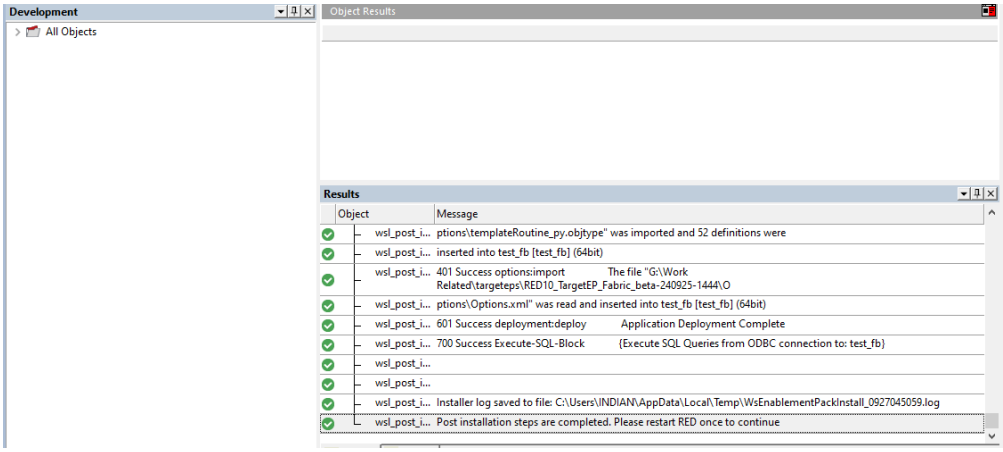
User will have to choose the schema for the target setting that were provided. One pop up will come for setting default target schema for Date Dimension.



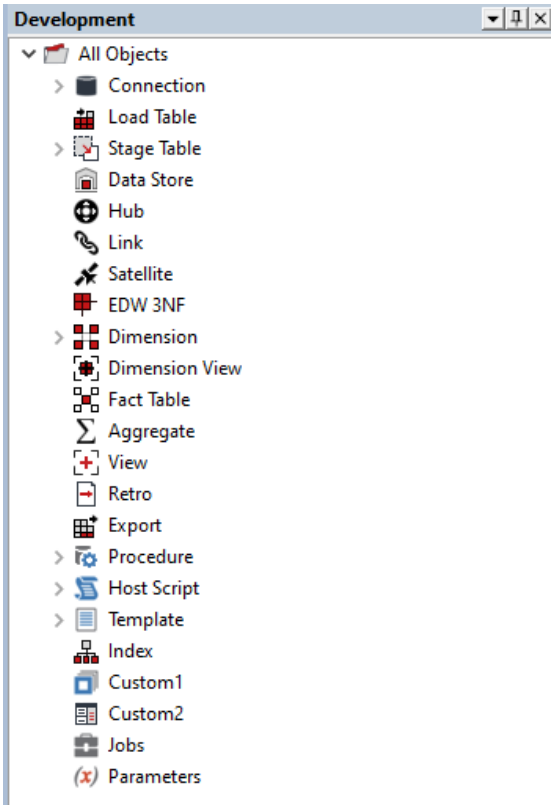
After selecting the target schema progress bar will show the progress for the installation and once it's completed, you will get the below pop up.



After pressing OK RED10 will open automatically.



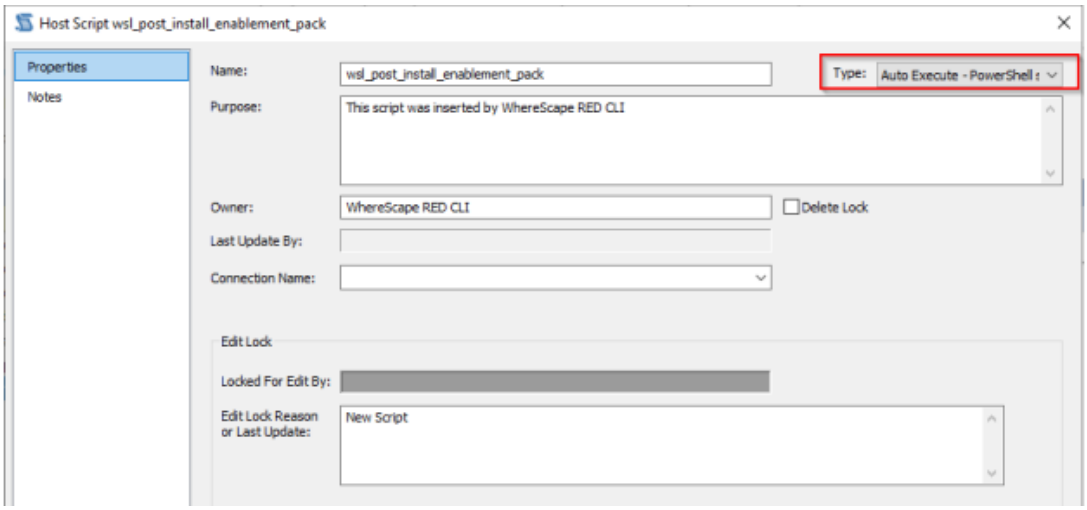
User will need to refresh the All Objects tree once.



Upgrade Of Existing Repository

For upgrade of existing repository

- From host script set script type of *wsl_post_install_enablement_pack* as Auto Execute - PowerShell Script



Guide for setting Fabric Data Factory Pipeline

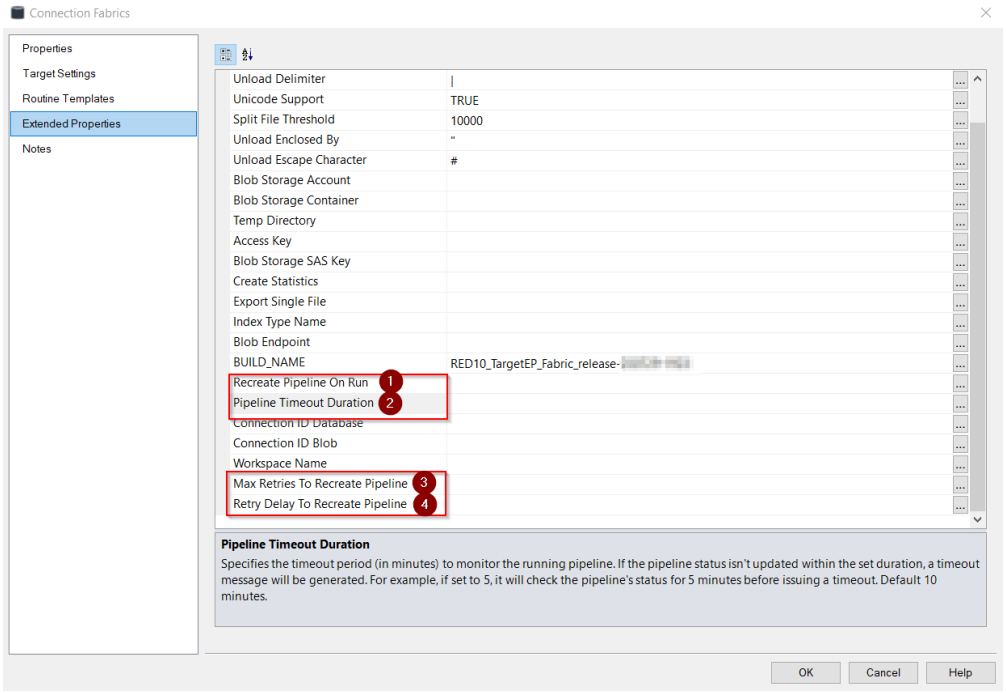
The EP contains 2 load templates:

1. wsl_fabric_pyscript_load will support load for Database, OneLake and traditional file parser.
2. wsl_fabric_pyscript_load_df template is for Fabric Data Factory

Fabric Data Factory Guide using load_df template:

In Fabric Data Factory, currently only OneLake sources are supported. When creating a connection to browse files, the Lakehouse name must be specified in the source connection. After browsing the connection and selecting a file, dragging the table will automatically populate the source properties. Four additional extended properties are specifically available for pipelines: “Recreate Pipeline on run”, “Pipeline Timeout Duration”, "Max Retries to Recreate Pipeline", "Retry Delay to Recreate Pipeline".

The first property, setting “Recreate Pipeline on run” to “True” ensures that a new pipeline is created each time the job is executed, with any existing pipeline with same name being deleted beforehand. If set to “False,” the existing pipeline remains intact, and only the execution is performed. However, if a pipeline fails due to a data type issue or any other property-related problem, the pipeline must be deleted and recreated, as the pipeline's JSON code cannot be modified when the “Recreate Pipeline on run” property is set to “False”. The default to this property is set to “True” in the script.
Whenever the Pipeline is created and data is loaded to the table, changing the Extended property to “False” will keep that existing pipeline as it is and act as a starting point for other operations which can be done.



The second property, “Pipeline Timeout Duration,” is responsible for monitoring the pipeline once it is executed. By default, the timeout is set to 10 minutes in script. If the execution exceeds this duration, the pipeline will continue running in the background, but RED will notify the user that the pipeline is still in progress. At this point, the user can monitor the pipeline through the Fabric portal. It’s important to note that even if the timeout is set to 10 minutes, if the job completes in a shorter time—such as within a minute or two—RED will mark the job as successful and display the result in the result pane accordingly

The third property, "Max Retries to Recreate Pipeline", Specifies the maximum number of retry attempts allowed to recreate the pipeline in case of failure. If the pipeline recreation fails due to transient errors, the system will retry up to the defined number of times before giving up. This helps in improving reliability and handling temporary issues during pipeline recreation. Default value is 6.

The fourth property, "Retry Delay to Recreate Pipeline", this value is set to a positive number, the system will recreate the pipeline every time the run starts. The numeric value represents the delay (in seconds) before the pipeline is recreated to allow for any pending cleanup. This ensures that a fresh pipeline is used for execution, avoiding potential issues with stale configurations or corrupted state. Default value is 60 secs.

There can be issues with data types when loading data through Fabric Data Factory.In such cases, the Fabric pipeline will display an error, suggesting using varchar(8000) for the affected column (e.g., “”).The user should update the column’s data type to match the recommendation from Data Factory. After making the changes, the table should be recreated, the script regenerated,and the pipeline executed again with the extended property “Recreate Pipeline on run” set to “True” or left blank

Post Install Steps – Optional

If you used the script Setup Wizard for installation then the following optional post install steps are available.

Configure Connections

These connections added that will optionally require your attention:

1. Connection: Data Warehouse ('Fabrics')- This connection was setup as per parameters provided in Setup Wizard
    a. open Properties and click derive button for Database Host/Server and Database ID.
    b. open it's properties and check extended properties tab, set it up for **Blob Storage Account , Blob Storage Container** and **Blob Storage SAS Key**
2. Connection: 'Database Source System' - this connection was setup as an example source connection,
    a. open its properties and set it up for a source DB in your environment
    b. or you can remove it if not required

Enable Script Launcher Toolbar

There are a number of stand-alone scripts which provide some features such as "Ranged Loading", these scripts have been added to the Script Launcher menu but you will need to enable the menu toolbar item to see them.

To enable the Script Launcher menu in RED: Select menu item 'Home->Script Launcher'

Source Enablement Pack Support

Source Pack Name	Supported By Microsoft Fabric	Supported Features	Prerequisites/Permissions Required for Microsoft Fabric
Google Cloud Storage	Yes	Download to local and load	None
Azure Data Lake Storage Gen2	Yes	Download to local and load	None
Amazon S3	Yes	Download to local and load	None
Windows Parser	Yes	Load Template, Source Properties will have option to select parser type to load the files.	Refer to Windows Parser Guide
Azure One Lake	Yes	Download to local and load	Refer to Windows Parser Guide

Troubleshooting and Tips

Run As Administrator

Press the Windows Key on your keyboard and start typing cmd.exe, when the cmd.exe icon shows up in the search list right click it to bring up the context menu, select "Run As Administrator"

Now you have an admin prompt navigate to to the folder where you have unpacked your WhereScape Red Enablement Pack to using the 'cd' command:

```
C:\Windows\system32> cd <full path to the unpacked folder>
```

Run batch (.bat) scripts from the administrator prompt by simply typing the name at the prompt and hit enter, for example:

```
C:\temp\EnablementPack>install_WslPowershell_Modules.bat
```

Run Powershell (.ps1) scripts from the administrator prompt by typing the Powershell run script command, for example:

```
C:\temp\EnablementPack>Powershell -ExecutionPolicy Bypass -File .\Setup_Enablement_Pack.ps1
```

Notes: In the event you can not bypass the Powershell execution policy due to group policies you can instead try "-ExecutionPolicy RemoteSigned" which should allow unsigned local scripts.

Windows Powershell Script Execution

On some systems Windows Powershell script execution is disabled by default. There are a number of workarounds for this which can be found by searching the term "Powershell Execution Policy".

Here is the most common workaround which WhereScape suggests, which does not permanently change the execution rights:

Start a Windows CMD prompt as Administrator, change directory to your script directory and run the WhereScape Powershell scripts with this command:

- cmd:>Powershell -ExecutionPolicy Bypass -File .\<script_file_name.ps1>

Re-install Python Libraries

Press the Windows Key on your keyboard and start typing cmd.exe, when the cmd.exe icon shows up in the search list right click it to bring up the context menu, select "Run As Administrator"

Now you have an admin prompt navigate to to the folder where you have unpacked your WhereScape Red Enablement Pack to using the 'cd' command:

C:\Windows\system32> cd <full path to the unpacked folder>

Run batch (.bat) scripts from the administrator prompt by simply typing the name at the prompt and hit enter, for example:

C:\temp\EnablementPack>uninstall_WslPython_Modules.bat

For installation of Python libraries there are two methods

- Method 1

  Press the Windows Key on your keyboard and start typing cmd.exe, when the cmd.exe icon shows up in the search list right click it to bring up the context menu, select "Run As Administrator"

  Now you have an admin prompt navigate to to the folder where you have unpacked your WhereScape Red Enablement Pack to using the 'cd' command:

  C:\Windows\system32> cd <full path to the unpacked folder>

  Run batch (.bat) scripts from the administrator prompt by simply typing the name at the prompt and hit enter, for example:

  C:\temp\EnablementPack>install_WslPython_Modules.bat

- Method 2

  Press the Windows Key on your keyboard and start typing cmd.exe, when the cmd.exe icon shows up in the search list right click it to bring up the context menu, select "Run As Administrator"

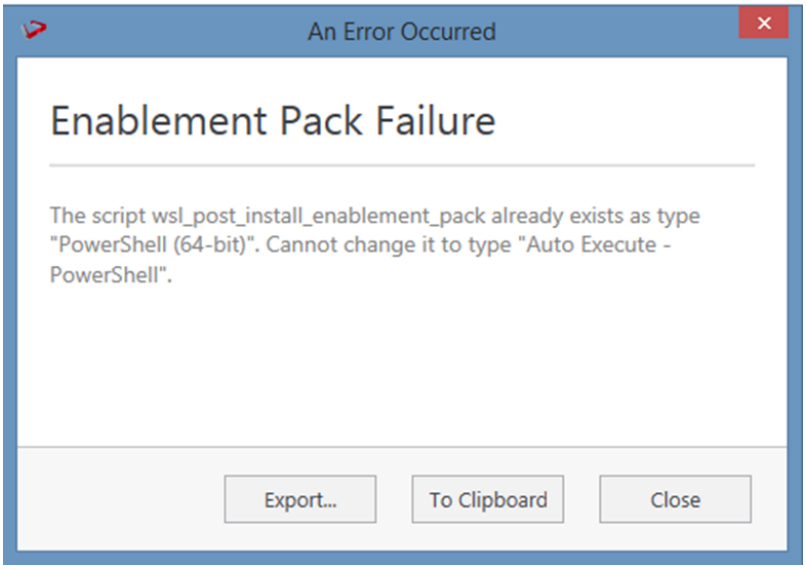
  Now you have an admin prompt navigate to to the folder where you have unpacked your WhereScape Red Enablement Pack to using the 'cd' command:

  C:\Windows\system32> cd <full path to the unpacked folder>

  Run the below command

  python -m pip install -r requirements.txt

For upgrade of existing repository



In upgrade of exiting repository if the user gets above error then it means the script type of *wsl_post_install_enablement_pack* is set to PowerShell(64-bit) change the script type to Auto Execute-PowerShell before upgrade or manually run the *wsl_post_install_enablement_pack* script from host script from RED after upgrade.

If a valid RED installation can not be found

If you have RED 10.x or higher installed but the script (Setup_Enablement_Pack.ps1) fails to find it on you system then you are most likely running PowerShell (x86) version which does not show installed 64 bit apps by default. Please open a 64 bit version of Powershell instead and re-run the script

Table name should be given in lowercase only

While loading the table, the table name should be given in lowercase, for example, load_tablename; otherwise, loaded data will now get displayed.