WhereScape Enablement Pack for Google BigQuery

This is a guide to installing the WhereScape Enablement Pack for Google BigQuery for WhereScape RED 8.6.6.+

Table of Contents

Prerequisites

Before you begin the following prerequisites must be met:

1. A supported* version of SQL Server or Azure SQL with
    - Database created to house the RED Metadata Repository with 'SELECT','INSERT','UPDATE','EXECUTE' permissions
    - (optionally) a database for the Range Table DB
    - ODBC DSN's created for these DB's

2. A BigQuery environment with
    - At least one schema available to use as a RED Data Warehouse Target

3. BigQuery software installed
    - BigQuery ODBC driver (64-bit) Simba ODBC driver 2.05.00.1001 or higher.
    - Google Cloud SDK for utilities GSUTIL and BQ installed and on the Windows system path
        - Download and Install Google Cloud SDK Installer from https://cloud.google.com/sdk/docs/quickstart
        - *Note:* To use the Google Cloud SDK and its associated tools (GSUTIL and BQ), authentication, initialization, and project configuration are required.
          Please refer to the provided instructions for completing this process from https://cloud.google.com/sdk/docs/install-sdk
        - *Important*: Google Cloud SDK and its associated tools (GSUTIL and BQ) must be authenticated again after installation. For reauthentication please refer this process from https://cloud.google.com/sdk/gcloud/reference/auth/application-default/login
          or run below command as admin in Command Prompt and follow prompted instructions.
          A complete system restart is recommended after this process.

          ```
          gcloud auth application-default login
          ```

    - ODBC DSN created to connect to your RED Data Warehouse on BigQuery
    - Google BigQuery Catalog Name
    - At least one storage bucket needs to created on Google Cloud Storage.
        - Permission Required : Upload, Read and Delete File

4. WhereScape RED version 8.6.6.0 or higher
    - Must be pre-installed with valid license key entered and EULA accepted

5. WhereScape Enablement Pack for BigQuery version 8.6.6.0 or higher
    - Downloaded and unpacked to a local temp folder

6. Python 3.8 or higher
    - Download python installer from https://www.python.org/downloads/
    - Select "Add Python 3.8 to PATH" from installation Window
    - PIP Manager
    - From Command Prompt (Run As Administrator) run below command

---

PIP Manager Install

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

---

* : RED supports the following versions for the metadata repository: MS SQL SERVER 2012 to 2019 and Azure SQL DB

Enablement Pack Setup Scripts

The Enablement Pack Install process is entirely driven by scripts. The below table outlines these scripts, their purpose and if "Run as Administrator" is required.

#	Enablement Pack Setup Scripts	Script Purpose	Run as Admin	Intended Application
1	Setup_Enablement_Pack.ps1	Setup and configure a **new** RED Metadata Repository for BigQuery	Yes	New and Existing installations
2	install_WslPowershell_Modules.bat	Installs or updates WslPowershell Modules on this machine	Yes	New and Existing installations
3	install_WslPython_Modules.bat	Installs or updates WslPython Modules and required Python libs on this machine	Yes	New and Existing installations
4	import_python_templates.ps1	Imports or updates the Python Templates (in BETA) to a RED Metadata Repository. Also includes any Script Imports	no*	Existing installations
5	set_default_templates.ps1	Applies the RED Connection defaults in a RED Metadata Repository for Python or Powershell templates	no*	Existing installations

Note

- *Some systems executing Windows Powershell scripts are disabled by default, see troubleshooting for workarounds*
- *install_WslPowershell_Modules.bat is required for creation of pilot range table as scripts required for pilot table i.e create,pause and restart are currently in powershell only*

Each Powershell script in the list above provides some help at the command line, this can be output by passing the "-help" parameter to the script.

Step-By-Step Guide

Setup and configure a new RED Metadata Repository for BigQuery

Run Powershell as Administrator:

```
Script 1 > Powershell -ExecutionPolicy Bypass -File .\Setup_Enablement_Pack.ps1
```

Install or Update WhereScape Powershell Modules

Run Script As Administrator

```
Script 2 > install_WslPowershell_Modules.bat
```

Install or Update WhereScape Python Modules

Run Script As Administrator

```
Script 3 > install_WslPython_Modules.bat
```

There are two steps in this script:

1. Installs the WhereScape WslPython modules to C:\Program Data\WhereScape\Modules\
2. PIP to download/update required Python libraries - for offline install please see the required library list for Python in the Troubleshooting section.

Install or Update WhereScape Python Templates (For Existing Installations)

Run Script as Administrator

```
Script 1 > install_WslPython_Modules.bat
Script 2 > import_python_templates.ps1
Script 3 > set_default_templates.ps1
```

Note : Skip this step for new installations.

Set Connection defaults for a Template Set (For Existing Installations)

```
Script 1 > set_default_templates.ps1
```

Choose "Python" when prompted

Note : To get help for parameter values run the script for help. For Example Powershell.exe -ExecutionPolicy Bypass -File .\Setup_Enablement_Pack. ps1 -help

Note: The upgrade will not overwrite the existing Data Type Mappings, UI configs. The user either needs to manually delete them or use startAtstep for continuing the script execution.

Important Upgrade Notes

If RED repository exists, it will prompt to upgrade the repository.
This enablement pack will overwrite any existing Source Enablement Pack UI Configs:

Amazon S3	Load From Amazon S3
Azure Data Lake Storage Gen2	Load From Azure Data Lake Storage Gen2
Google Cloud	Load From Google Cloud
REST API	Load From REST API

To ensure existing Source Enablement Pack connections and associated Load Tables continue to browse and load:

Go into UI Configuration Maintenance in RED prior to installing this Enablement Pack and rename the affected UI Configurations. While the updated Load Template will work with previous Source Enablement Pack's we recommend moving these previous versions of Load Tables to newly created Parser based connections following this install. The earlier versions of the Source Enablement Pack will be deprecated following this release.

Post Install Steps - Optional

If you used the script 'Setup_Enablement_Pack.ps1' then the following optional post install steps are available

Configure Connections

There were two connections added that will optionally require your attention:

1. Connection: 'BigQuery' - This connection was setup as per parameters provided in the installation script 1
   a. open it's properties and check extended properties tab, set it up for **LOCATION**. Ignore if already set during installation script prompt
2. Connection: 'Database Source System' - this connection was setup as an example source connection,
   - open it's properties and set it up for a source DB in your environment
   - or you can remove it if not required

Enable Script Launcher Toolbar

There are a number of stand-alone scripts which provide some features, 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 'View->Toolbars->Script Launcher'

Service Authentication

A User DSN must be created for both the repository user and each individual user, as these references utilize a single, one-time use key.

In the event that a User DSN cannot be created for the repository user, a System DSN can be used as an alternative

To update the 'Services' settings if jobs are not running through the scheduler, please follow the steps shown in the attached screenshots:
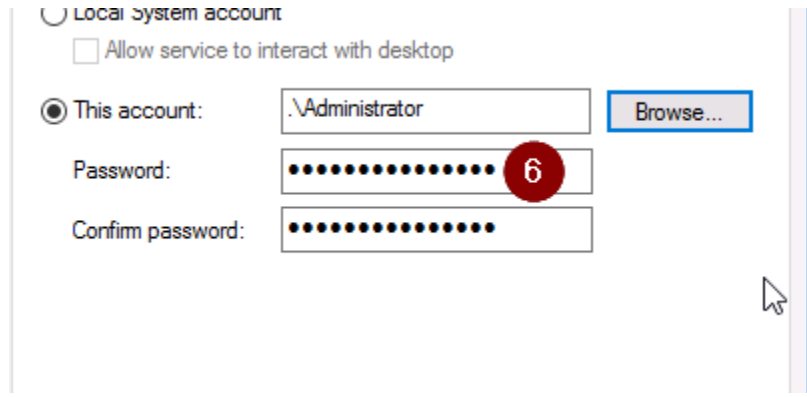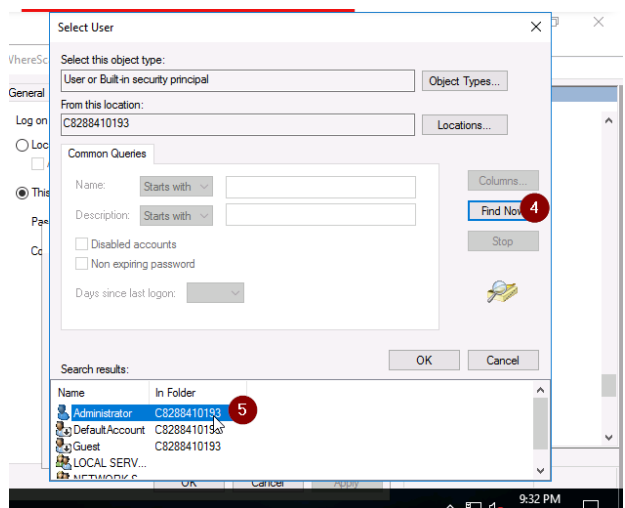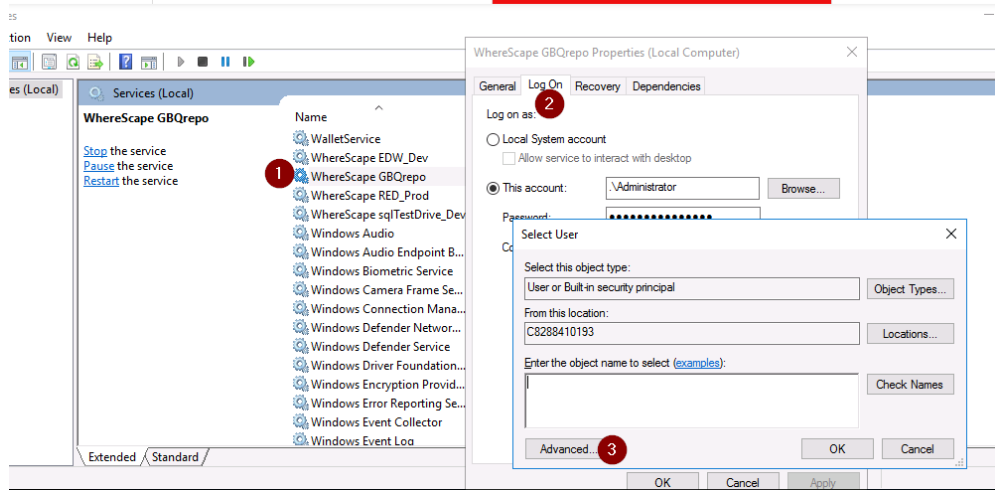
Source Enablement Pack Support

Source Pack Name	Supported By Google BigQuery	Supported Features	Prerequisites/Permissions Required Google Big Query
Google Cloud Storage	Yes	Download to local and load	None
Google Drive	No	N/A	None
Azure Data Lake Storage Gen2	Yes	Download to local and load	None
Amazon S3	Yes	Download to local and load	None

Windows Parser	1. CSV 2. Excel 3. JSON 4. XML 5. AVRO 6. ORC 7. PARQUET	Load Template, Source Properties will have option to select parser type to load the files.	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 on the internet 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>

Restarting failed scripts

Some of the setup scripts will track each step and output the step number when there is a failure. To restart from the failed step (or to skip the step) provide the parameter "-startAtStep <step number>" to the script.

Example:

Powershell -ExecutionPolicy Bypass -File .\<script_file_name.ps1> -startAtStep 123

Tip: to avoid having to provide all the parameters again you can copy the full command line with parameters from the first "INFO" message from the beginning of the console output.

Python requirements for offline install

Additionally to the base Python installation being required, the WhereScape Python Template set also requires certain additional Python libraries. The install scripts uses the PIP (package manager) to download these libraries, however for offline installs you will need to install the required libraries yourself.

Required Python libraries/add-ons:

- pywin32-ctypes
- python-tds
- pywin32
- glob2
- gzip-reader
- regex
- pyodbc

If a valid RED installation can not be found

If you have Red 8.5.1.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.