WhereScape Red

Incremental Loading with High-Water Marks

Document Version 1.1.1, 2021/10/11, used on Wherescape Red 6.1.*, Snowflake Target, SQL Server Red Metadata

Author: Christopher Siegfried

Modified Loading with a High-Water Mark

Loading data with a High-Water Mark (HWM) allows incremental loading when there is no Change Data Capture feature available. A column, usually an identity column or datetime column, is selected as the HWM. After a load the maximum value of that column is stored. The next time that same data source is loaded, the stored value of the HWM column is used as a filter to load the most recent data.

The solution below was presented by Alex Nicolae in the WhereScape EMEA User Group on August 26th, 2021. The majority of the script logic is copied from the scripts shown in the video while the setup directions (which follow the user guide below) have been constructed using infoVia best practices.

User Guide

Set the Load Table HWM Column
After completing the setup process open the properties for a load table and select "Extended Properties" in the left window. Locate the HWM_COLUMN extended property. Set the value to the desired **Column Name** (as opposed to the *Source* Column Name) from the load table.
To reduce issues while using a timestamp HWM the column in the load table should use a timestamp data type and any values inserted into the column should be cast to a generally recognized format (e.g. YYYY-MM-DD HH:MM:SS)
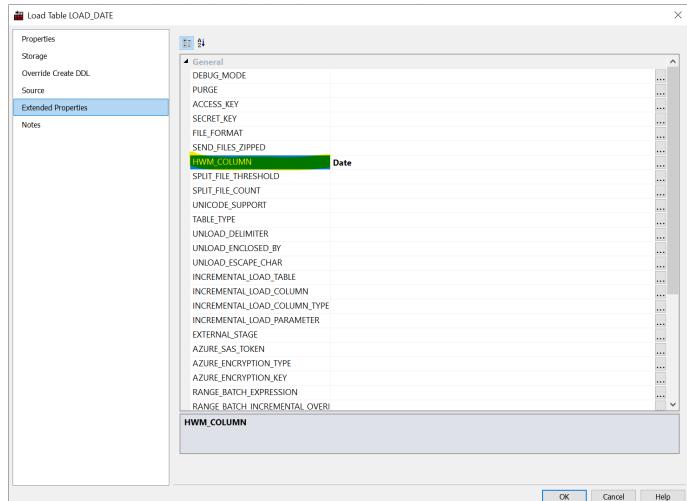
Set the Load Table Post-Load Script
Open the properties of a load table and set the Post Load Type to "Script". Select "Rebuild" next to the post-load script to set the template.

Post Load Type:	Script	
Post Load Script:	POST_LOAD_DATE	Edit Rebuild Regenerate

Template: wsl_hwm_snowflake_psscript_postload
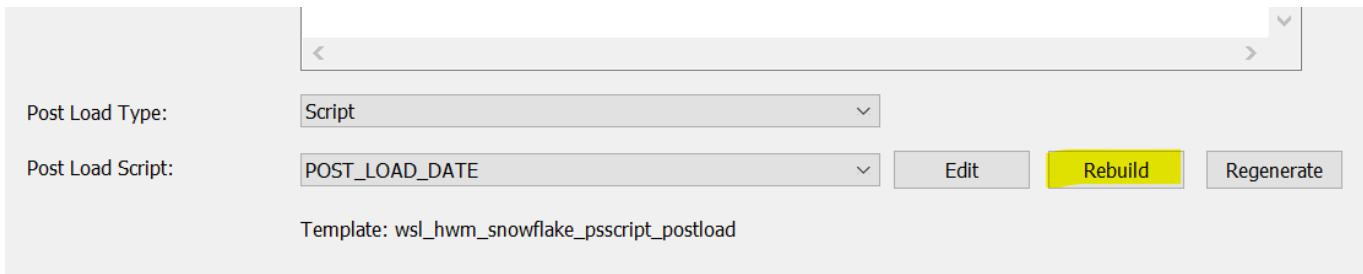
When the rebuild option window opens set the processing option "Template" to the template created during setup and click okay.

Custom Type	Script	
Script Connection	Runtime Connection for Scripts	
Template	wsl_hwm_snowflake_psscript_postload	
	wsl_hwm_snowflake_psscript_postload	
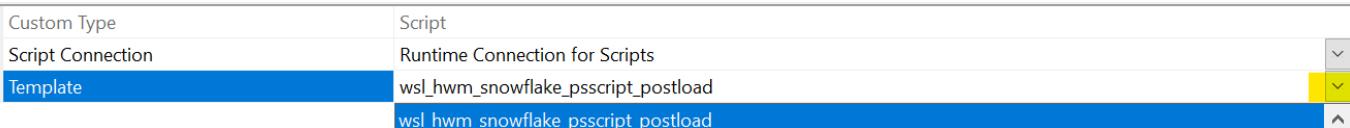
Select " Set the Post Load Script template to wsl_snowflake_psscript_postload_hwm (see creation directions in setup, below) and generate the code. The template can be set via the "Rebuild" option.

Generate HWM Parameters for load tables

Run the stored procedure created in setup to create a parameter for any load table which has a value in the HWM_COLUMN extended property. Load scripts for affected tables must be regenerated after completing this step.

Load Tables

Load tables as desired. The HWM value for a table can be manually adjusted by changing it's parameter value (Tools > Parameters). The name of a load table's HWM parameter will be HWM_<load table name>

Setup

Create Load Parameter Stored Procedure

Running this stored procedure will generate WhereScape Red parameters for each load table that has a value set for the HWM_COLUMN extended property. It also creates a WHERE statement on the load table to filter out values lower than the recorded HWM parameter value.

1. Right click the "Procedure" heading in the object pane and select "New Object", then enter a name for the procedure.
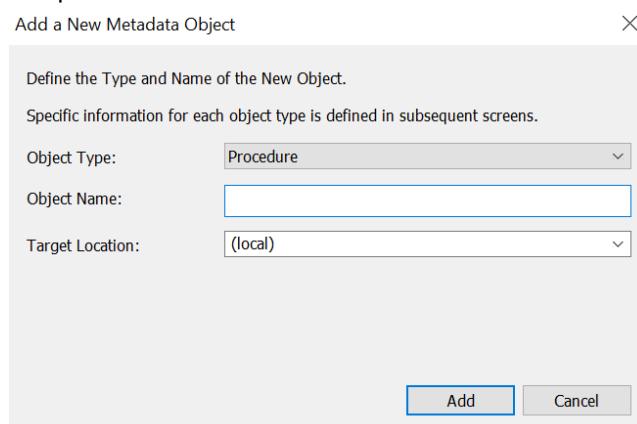
2. Click "OK" on the properties window that will appear
3. Right click the newly created object and select edit
4. An empty procedure file will open. Select Tools > Create Procedure Outline from the menu
5. Immediately following the comment block starting with "Insert custom code here..." add the below code

```
DECLARE @param VARCHAR(255);
DECLARE lt CURSOR FOR

SELECT 'HWM_' + lt_table_name
FROM ws_load_tab lt
    INNER JOIN ws_ext_prop_value ep
        ON lt.lt_obj_key = ep.epv_obj_key
    INNER JOIN ws_ext_prop_def ed
        ON ep.epv_def_key = ed.epd_key
            AND ed.epd_variable_name = 'HWM_COLUMN'
    LEFT JOIN dss_parameter param
        ON 'HWM_COLUMN' + lt.lt_table_name = param.dss_parameter_name
WHERE param.dss_parameter_name IS NULL;

OPEN lt;
FETCH NEXT FROM lt INTO @param;

WHILE @@FETCH_STATUS = 0
BEGIN
    EXEC WsParameterWrite @param, '', 'High water mark';
    FETCH NEXT FROM lt INTO @param;
END
CLOSE lt;
DEALLOCATE lt;

--Add WHERE clause to filter to load table
UPDATE lt
SET lt_where_clause = 'WHERE ' + lc.lc_src_column + ' > ''$PHWM_' + lt_table_name+
'$'' '
FROM ws_load_tab lt
INNER JOIN ws_ext_prop_value ep
    ON lt.lt_obj_key = ep.epv_obj_key
INNER JOIN ws_ext_prop_def ed
    ON ep.epv_def_key = ed.epd_key
    AND ed.epd_variable_name = 'HWM_COLUMN'
LEFT JOIN ws_load_col lc
    ON lc.lc_col_name = ep.epv_value
WHERE (ep.epv_value is not null and ep.epv_value <> '')
    and (lt.lt_where_clause is null OR CONVERT(VARCHAR(MAX), lt.lt_where_clause)
='');
```

6. Save and compile the code through the editor toolbar or by using the right-click menu on the
   procedure

Create the Post-Load Script Template
This script will run after a load has finished and record the MAX value of the HWM column in the load table's
HWM parameter.

1. Right click the "Template" heading in the object pane and select "New Object" using "wsl_snowflake_psscript_postload_hwm" for the name
2. Set Target DB to "Common"
3. Set Type to PowerShell (64 bit) Script
4. Select Okay
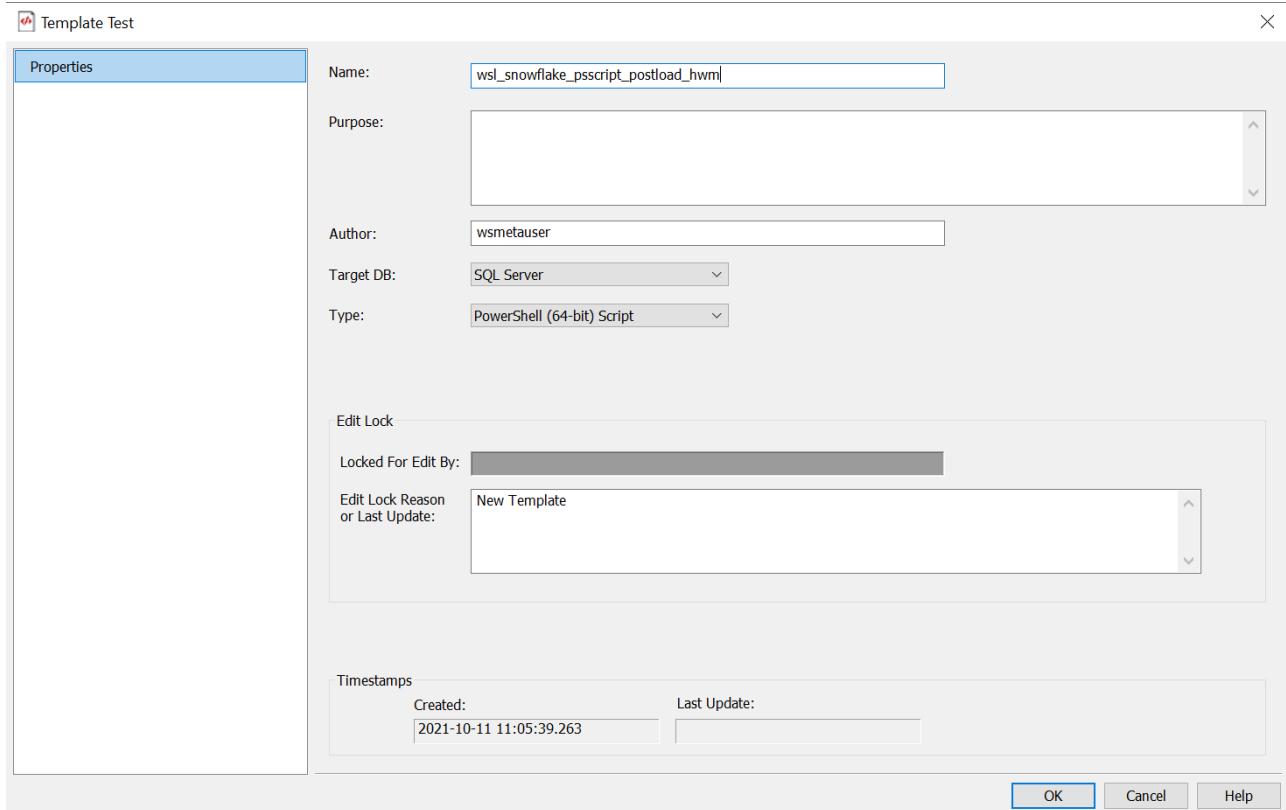
5. Right click the newly created object and select edit
6. An empty template file will open. Paste in the below code.

```
Import-module -Name WslPowershellCommon -DisableNameChecking
Import-module -Name WslPowershellSnowflake -DisableNameChecking
$tgtConn = New-Object System.Data.Odbc.OdbcConnection
Hide-Window

# Main call
try {
    $Col = Get-ExtendedProperty -PropertyName "HWM_COLUMN" `
-TableName ${env:WSL_LOAD_TABLE}
    $ParamName = "HWM_${env:WSL_LOAD_TABLE}"
    $ParamValue = (WsParameterRead "$ParamName")[0]

    $SQL = @"
SELECT
    CASE WHEN COUNT(*) > 0 THEN MAX($Col)
    WHEN '$ParamValue' IS NULL THEN ''
    ELSE '$ParamValue' END max_value FROM ${env:WSL_LOAD_FULLNAME};
"@
    $Result = Run-RedSQL -sql $sql -dsn ${env:WSL_TGT_DSN} `
-uid ${env:WSL_TGT_USER} -pwd ${env:WSL_TGT_PWD} -odbcConn $tgtConn
```

```
    $max_value = $Result.max_value
    WsParameterWrite -ParameterName "$ParamName" -ParameterValue `
"$max_value" -ParameterComment "Updated by post load script"
    $status = 1
    $return_Msg = "Updated parameter"
}
catch {
    $status = 2
    $return_Msg = "Failure $_.Exception.Message"
}

$status
$return_Msg
```

7. Save the script.

Create the HWM Extended Property

1. Select Tools > Extended Properties > Maintain Extended Properties
2. Select New
3. Set the new extended properties as shown below.
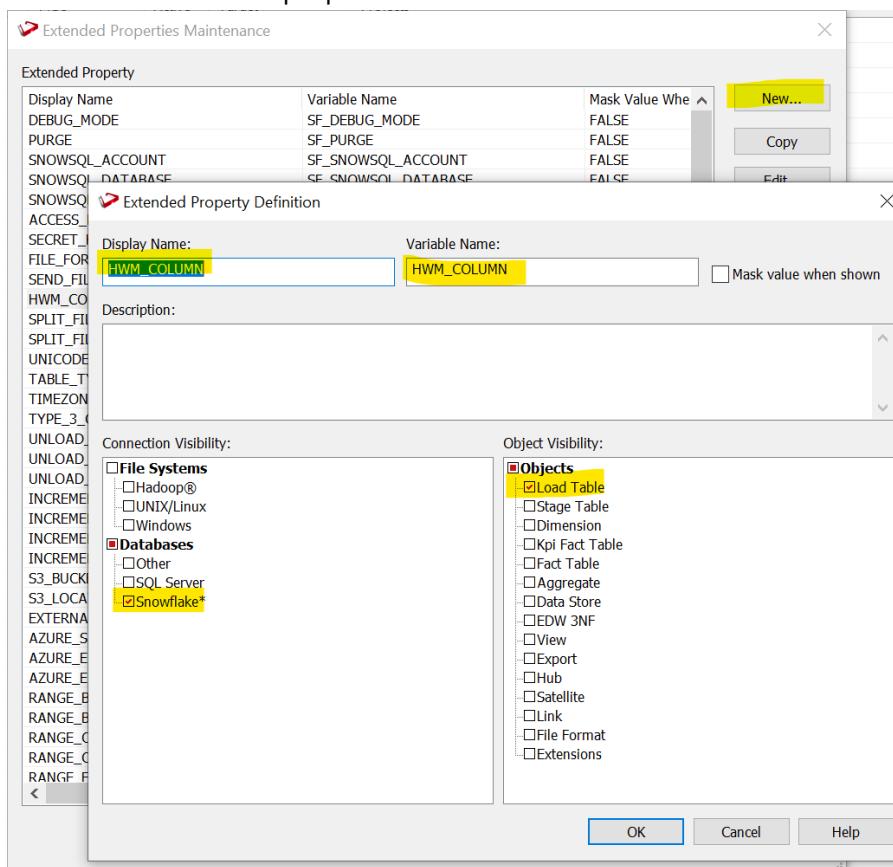
4. Select OK

Final Notes

The current version of the solution requires a moderately high amount of manual intervention as each load table must have its HWM_COLUMN value set as well as setting the post-load template. Future iterations will

aim to reduce the manual steps required and integrate WhereScape 3D.

Change Log:

1.1.2 -

Stored procedure corrected

Directions now default to use "Common" as the stored procedure type due to issues with "SQL Server" templates vanishing with some license types

1.1.1 -

Update document name

Reference Setup direction below User Guide

Add best practices for timestamp HWMs

Add images displaying setting template for post-load script

Add line breaks and backticks to post-load script

1.1 -

Adjust stored procedure to reference the load table source column name for WHERE clause Format procedure syntax