
Vertica Knowledge Base Article

Vertica QuickStart for Informatica PowerCenter

Document Release Date: 1/30/2023

Legal Notices

Warranty

The only warranties for Open Text Corporation. All rights reserved. products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. OpenText shall not be liable for technical or editorial errors or omissions contained herein.

The information contained herein is subject to change without notice.

Restricted Rights Legend

Confidential computer software. Valid license from OpenText required for possession, use or copying. Consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

Copyright Notice

© Copyright 2015-2023 Open Text Corporation. All rights reserved.

Trademark Notices

Adobe™ is a trademark of Adobe Systems Incorporated.

Microsoft® and Windows® are U.S. registered trademarks of Microsoft Corporation.

UNIX® is a registered trademark of The Open Group.

This product includes an interface of the 'zlib' general purpose compression library, which is Copyright © 1995-2002 Jean-loup Gailly and Mark Adler.

Contents

Vertica QuickStart for Informatica PowerCenter	4
About the Vertica QuickStarts	4
VHist ETL Overview	4
Requirements	4
Install the Software	5
Install Informatica PowerCenter	5
Install the Vertica Database Server	5
Install the Client Driver	5
Install the QuickStart Application	6
Create a Connection to Vertica	6
Create the Data Warehouse	6
Import the .xml File	7
Execute the Workflow	8
Import Data Warehouse Objects	9
Import the seq_Batch_Id File	10
Import the wf_stg_vhist File	11
Import the wf_star_vhist File	12
Loading Data into the VHist Data Warehouse	14
Execute the Workflow	14
Validate the ETL	14
Schedule Incremental Loads	15
Find More Informatio	16

Vertica QuickStart for Informatica PowerCenter

The Vertica QuickStart for Informatica PowerCenter is a sample ETL application powered by the Vertica Analytic Database. The QuickStart extracts data from Vertica system tables and loads it into a data warehouse called VHist (Vertica History).

Details about the ETL processing and the source and target data sets are provided in the companion document, [Vertica VHist ETL Overview](#).

You can download the Vertica QuickStart for Informatica PowerCenter from the following location:

<https://www.vertica.com/quickstart/vertica-quickstart-for-informatica/>

About the Vertica QuickStarts

The Vertica QuickStarts are free, sample applications created using front-end products from Vertica technology partners. For an overview, watch this short [video](#).

The QuickStarts are posted for download on the [Vertica QuickStart Examples](#) page.

Note The Vertica QuickStarts are freely available for demonstration and educational purposes. They are not governed by any license or support agreements and are not suitable for deployment in production environments.

VHist ETL Overview

VHist ETL occurs in two steps:

1. Data is extracted from [system tables](#) in the V_CATALOG and V_MONITOR schemas and loaded into a staging schema called VHIST_STAGE. Only minimal transformation occurs during this step.
2. Data is extracted from VHIST_STAGE, transformed, and loaded into the VHist star schema.

For details, see [Vertica VHist ETL Overview](#).

Requirements

The Vertica QuickStart for Informatica PowerCenter requires a Vertica database server with the VHist schemas, a Vertica client, and Informatica PowerCenter 10 or later.

The QuickStart was created using Informatica PowerCenter 10 and Vertica 7.2.x.

Vertica also has a plug-in for Informatica PowerCenter. This plug-in makes integrating the two products easier. For more information, see [Vertica Plug-In for Informatica](#) in the Vertica documentation.

Install the Software

To install the software that is required for running the QuickStart, follow these steps:

- [Install Informatica PowerCenter](#)
- [Install the Vertica Database Server](#)
- [Install the Client Driver](#)
- [Install the QuickStart Application](#)

Install Informatica PowerCenter

Informatica is a data integration tool that promotes automation, reuse, and agility. To install a free version of Informatica PowerCenter, navigate to the [Informatica Marketplace](#) and click **Download Now**.

Install the Vertica Database Server

If you do not have Vertica, you can download the Community Edition free of charge:

1. Navigate to [Vertica Community Edition](#).
2. Log in or click **Register Now** to create an account
3. Follow the on-screen instructions to download and install the Vertica Community Edition.

Install the Client Driver

Informatica PowerCenter uses an ODBC Data Source Name (DSN) to connect to Vertica. To create the DSN, download the Vertica client package and install the ODBC driver, then configure the DSN in the ODBC Driver Manager.

To download the Vertica client package:

1. Navigate to the [Vertica Client Drivers](#) page.
2. Download the version of the Vertica client package that is compatible with the architecture of your operating system and Vertica server version. The Vertica client package includes client components and all the drivers.

Note Starting in Vertica 8.1.1, the ODBC drivers are both forward and backwards compatible. For details, see [Client Driver and Server Version Compatibility](#) in the Vertica documentation.

To install the ODBC driver and create the DSN, follow the instructions in the Vertica documentation:

- To install the ODBC driver, see [Client Drivers](#).
- To create the DSN, see [Creating an ODBC Data Source Name \(DSN\)](#).

You can choose to install the complete Vertica client package or just the ODBC driver.

Install the QuickStart Application

1. Navigate to vertica.com/quickstart.
2. Select **Vertica QuickStart for Informatica PowerCenter**.
3. Log in or create an account.
4. Click **Download**.
5. Save the compressed file, `VHIST_ETL_INFORMATICA.zip`, on your machine.
6. Extract the contents of the file to any local directory. You will see the following files:
 - `seq_batch_id.xml`: A reusable sequence generator transformation.
 - `wf_vhist_ddl.xml`: A workflow to drop and create the required database tables for implementing VHist. This file is run only once to create the required tables and is not a part of the run schedule.
 - `wf_stg_vhist.xml`: A workflow to populate staging tables from system tables. This workflow is designed to trigger the next phase of VHist and is critical for the job scheduler.
 - `wf_star_vhist.xml`: A workflow to populate dimension and fact tables.
 - `Param_stg_vhist.txt`: A parameter file for the `wf_stg_vhist` workflow. You can edit this file if you want to customize connection names.
 - `Param_star_vhist.txt`: A parameter file for the `wf_star_vhist` workflow. You can edit this file if you want to customize connection names.
 - `Param_ddl_vhist.txt`: A parameter file for the `wf_vhist_ddl` workflow. You can edit this file if you want to customize connection names.

Create a Connection to Vertica

1. Open the Informatica PowerCenter Workflow Manager.
2. From the top menu, select **Connections > Relational > New > Vertica > OK**.
3. Enter the following parameters:
 - Connection name: The default connection name is `vhist_stage` for both the source and target. You can change the connection name but must update it in all parameter files.
 - Database user name and password
 - Connection string: DSN value
 - Code page: UTF-8 encoding of Unicode
4. Click **OK**.

Create the Data Warehouse

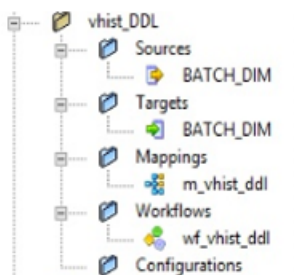
To create the data warehouse, you must import `wf_vhist_ddl.xml` into the Repository Manager Client and execute the workflow in the Workflow Manager.

Import the .xml File

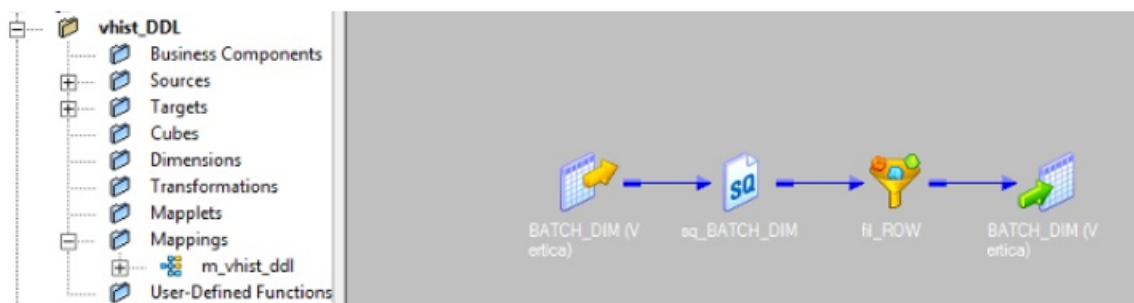
1. Open the Repository Manager Client.
2. Connect to the repository where you want to import objects.
3. Select the repository.
4. Select **Folder > Create**.
5. Name the folder vhist_DDL and click **OK**.
6. Select the vhist_DDL folder you created.
7. Select **Repository > Import Objects**.
8. Click **Browse** and select the wf_vhist_ddl.xml file.
9. Click **Next**.
10. Click **Add All**.
11. Click **Next**. Verify that both the destination folder and the folder in the .xml file are the same.
12. Select **Next > New Rule**.
13. Under **Select Resolution**, select **Replace**.
14. Click **Next**.
15. Under **Status**, verify that **Resolved** is displayed. If yes, click **Import**.
16. Click **Done**.
17. Right-click the vhist_DDL folder and click **Refresh** to view imported objects and their dependent child objects.

After a successful import, you should see:

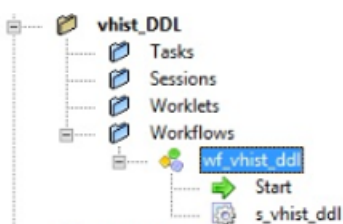
REPOSITORY MANAGER:



DESIGNER:



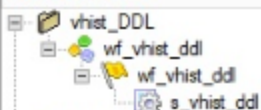
WORKFLOW MANAGER:



Execute the Workflow

1. Open the Workflow Manager and connect to the Repository where you imported the objects.
2. Right-click the vhist_DDL folder and click **Open**.
3. Double-click **Workflows** and select the **wf_vhist_ddl.xml** workflow.
4. Place the **Param_ddl_vhist.txt** parameter file into the local machine folder C:\Informatica\10.0.0\server\infa_shared\BWParam.
5. Under the **Workflow** tab, click **Edit**.
6. In the **General** tab, check the box for your integration service.
7. Click **OK**.
8. Click the **Workflow** tab and select **Start Workflow**.

In the workflow monitor, verify that the workflow successfully runs.

Workflow Run	Start Time	Completion Time	Status
	8/16/2016 12:22:03 PM	8/16/2016 12:22:13 PM	Succeeded
	8/16/2016 12:22:03 PM	8/16/2016 12:22:09 PM	Succeeded

At the database level, you should see the following tables:

- vhist_stage.batch_dim
- vhist_stage.VHIST_STAGE_LOAD_LOG
- vhist_stage.nodes
- vhist_stage.users
- vhist_stage.system_resource_usage
- vhist_stage.user_sessions
- vhist_stage.QUERY_PROFILES
- vhist_stage.LOAD_STREAMS
- vhist_stage.DISK_STORAGE
- vhist_stage.RESOURCE_REJECTIONS
- vhist_stage.PROJECTION_STORAGE
- vhist_stage.PROJECTION_USAGE
- vhist.batch_dim
- vhist.VHIST_LOAD_LOG
- vhist.time_dim
- vhist.date_dim
- vhist.nodes_dim
- vhist.users_dim
- vhist.sru_fact
- vhist.user_sessions_fact
- vhist.QUERY_PROFILES_FACT
- vhist.LOAD_STREAMS_FACT
- vhist.DISK_STORAGE_FACT
- vhist.RESOURCE_REJECTIONS_FACT
- vhist.PROJECTION_STORAGE_FACT
- vhist.PROJECTION_USAGE_FACT

Import Data Warehouse Objects

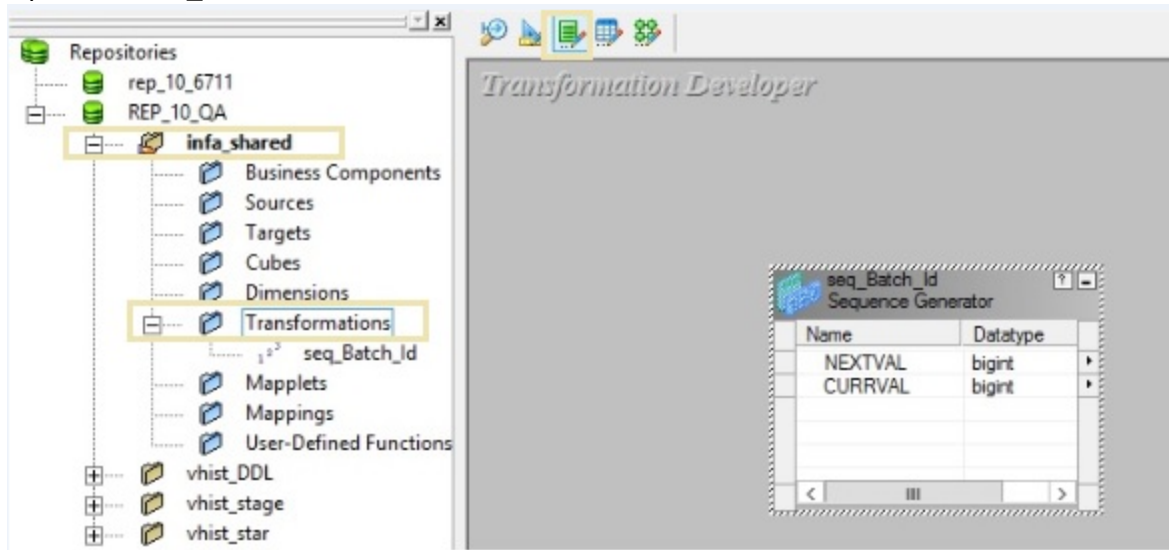
Before you import the wf_stg_vhist.xml file, you must import the seq_Batch_Id.xml file. The seq_Batch_Id.xml file imports all stage-related mappings and workflows.

Import the seq_Batch_Id File

1. Open the Repository Manager Client.
2. Connect to the Repository where you want to import objects.
3. Click the Repository.
4. Select **Folder > Create**.
5. Name the folder infa_shared.
6. Under **Options**, check **Allow Shortcut**.
7. Click **OK**.
8. Select the infa_shared folder you created.
9. Select **Repository > Import Objects**.
10. Click **Browse** and select the seq_Batch_Id.xml file.
11. Click **Next**.
12. Click **Add All**.
13. Click **Next**. Verify that both the destination folder and the folder in the .xml file are the same.
14. Select **Next > New Rule**.
15. Under **Select Resolution**, select **Replace** from the drop-down menu.
16. Click **Next**.
17. Under **Status**, verify that Resolved is displayed. If yes, click **Import**.
18. Click **Done**.
19. Right-click the infa_shared folder and click **Refresh** to view the imported object and its dependent child objects.

After you successfully import the object, follow these steps to verify the properties of the object in Designer:

1. Log in to the environment where the object was imported.
2. Open the `infa_shared` folder.



3. Navigate to the transformation developer.
4. Open the `seq_Batch_Id` shared object by double-clicking it.
5. Click **Properties**.
6. Change the current value field to 1 and click **Save**.

Import the `wf_stg_vhist` File

1. Open the Repository Manager Client.
2. Connect to the Repository where you want to import objects.
3. Select the Repository.
4. Select **Folder > Create**.
5. Name the folder `vhist_stage`.
6. Click **OK**.
7. Select the `vhist_stage` folder you created.
8. Select **Repository > Import Objects**.
9. Click **Browse** and select the `wf_stg_vhist.xml` file.
10. Click **Next**.
11. Click **Add All**.
12. Click **Next**. Verify that both the destination folder and the folder in the .xml file are the same.
13. Select **Next > New Rule**.
14. Under **Select Resolution**, select **Replace** from the drop-down menu.
15. Click **Next**.

16. Under **Status**, verify that **Resolved** is displayed. If yes, click **Import**.
17. Click **Done**.
18. Right-click the `vhist_stage` folder and click **Refresh** to view the imported object and its dependent child objects.

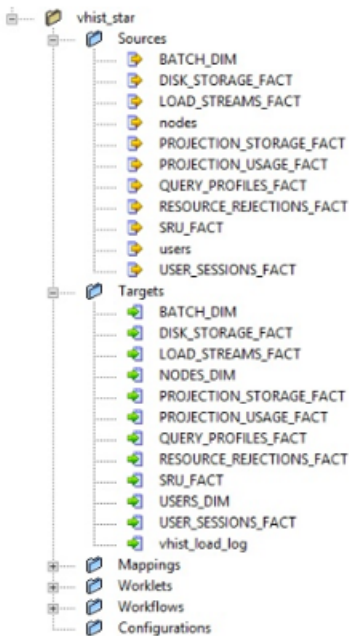
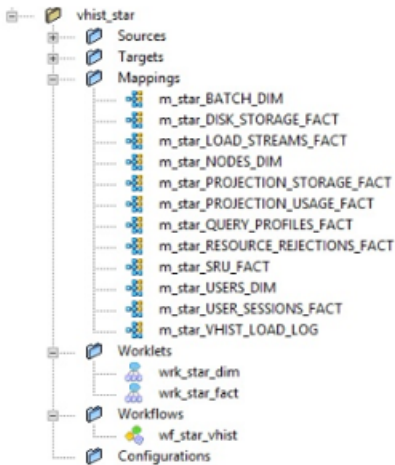
Import the wf_star_vhist File

To configure the star stage, fact tables, and dimension table population, you must import the `wf_stg_star.xml` file. This imports all the star-related mappings and workflows. You can load data from the source `V_CATALOG` to target both the stage and star tables.

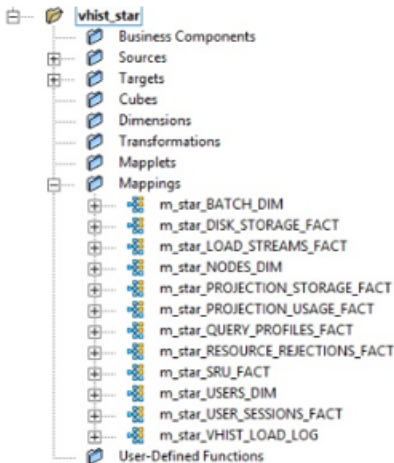
1. Open the Repository Manager Client.
2. Connect to the Repository where you want to import objects.
3. Click the Repository.
4. Select **Folder > Create**.
5. Name the folder `vhist_star`.
6. Click **OK**.
7. Select the `vhist_star` folder you created.
8. Select **Repository > Import Objects**.
9. Click **Browse** and select the `wf_star_vhist.xml` file.
10. Click **Next**.
11. Click **Add All**.
12. Click **Next**. Verify that both the destination folder and the folder in the `.xml` file are the same.
13. Select **Next > New Rule**.
14. Under **Select Resolution**, select **Replace** from the dropdown menu.
15. Click **Next**.
16. Under **Status**, verify that **Resolved** is displayed. If yes, click **Import**.
17. Click **Done**.
18. Right-click the `infa_shared` folder and click **Refresh** to view the imported object and its dependent child objects.

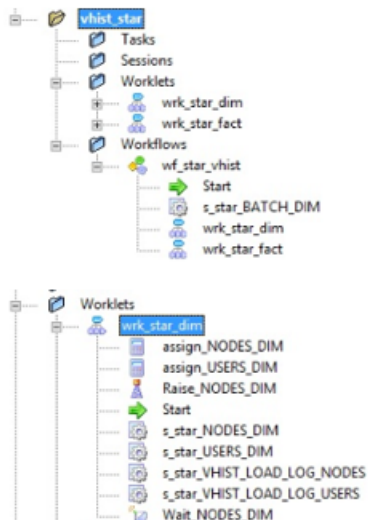
After a successful import, you should see the following:

REPOSITORY MANAGER:



DESIGNER:



WORKFLOW MANAGER:

Loading Data into the VHist Data Warehouse

The first time you run the jobs, you must populate the staging tables, fact tables, and dimension tables by following these steps.

Execute the Workflow

The first time you execute this workflow, it loads from source tables to stage tables and then load from stage tables to fact tables. Follow these steps to execute the workflow:

1. Open the Workflow Manager and connect to the Repository where you previously imported objects.
2. Right-click the **vhist_stage** folder and click **Open**.
3. Double-click **Workflows** and select the **wf_stg_vhist** workflow.
4. Place the Param_stg_vhist.txt parameter file into the local machine folder C:\Informatica\10.0.0\server\infa_shared\BWParam.
5. Under the **Workflow** tab, click **Edit**.
6. Under the **General** tab, check the box for your integration service.
7. Click **OK**.
8. Click the **Workflow** tab and select **Start Workflow**.
9. Upon successful completion, the next VHist job starts.

In the workflow monitor, you can verify if the workflow successfully runs.

Validate the ETL

The VHist ETL process records events in log tables that you can query to determine the success or failure of the data load.

To query the ETL log tables:

1. Connect to the target database using vsql or a client tool like DBVisualizer.
2. Run this query to validate the vhist_stage schema:

```
SELECT *
FROM vhist_stage.vhist_stage_load_log
WHERE batch_id =(SELECT max(batch_id)
FROM vhist_stage.vhist_stage_load_log)
```

3. Run this query to validate the vhist schema:

```
SELECT *
FROM vhist.vhist_load_log
WHERE batch_id =(SELECT MAX(batch_id)
FROM vhist.vhist_load_log);
```

Schedule Incremental Loads

Once the data warehouse has been created and populated, you can perform incremental loads to keep the warehouse up to date and start accumulating historical data. To refresh the data warehouse, schedule incremental loads to run at intervals.

You must schedule the **wf_stg_vhist** workflow on an incremental load based on your needs.

Take care when scheduling incremental loads to avoid placing undue demands on system resources or causing the data warehouse to grow too large. The amount of data stored in Vertica system tables depends on many factors, and the individual tables are not flushed at the same rate. Keep the following in mind:

- To avoid running incremental loads more often than is needed, try starting with daily loads then review the results in the log tables. If there are gaps in the results, decrease the interval between loads until you find an optimal balance.
- Repeated incremental loads increase the size of the data warehouse over time. The growth amount varies depending on system activity and frequency of collection.

Note The data that you load into VHist counts towards the limit specified in your Vertica license.

Tip If you are using the Community Edition, your license allows up to one terabyte of free storage. If you already have a licensed installation of Vertica, you can build and maintain the VHist warehouse using the Community Edition in a separate cluster.

Find More Informatio

- [Informatica](#)
- [Introduction to Using Informatica PowerCenter with Vertica](#)
- [Vertica Integration with Informatica: Connection Guide](#)
- [Vertica Integration with Informatica: Tips and Techniques](#)
- [Vertica VHist ETL Overview](#)
- [Vertica Community Edition](#)
- [Vertica User Community](#)
- [Vertica Documentation](#)