

Informatica® PowerExchange for Vertica 10.5

# User Guide for PowerCenter

Informatica PowerExchange for Vertica User Guide for PowerCenter 10.5 March 2021

#### © Copyright Informatica LLC 2014, 2021

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

Informatica, the Informatica logo, PowerCenter, and PowerExchange are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at https://www.informatica.com/trademarks.html. Other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties, including without limitation: Copyright DataDirect Technologies. All rights reserved. Copyright © Sun Microsystems. All rights reserved. Copyright © RSA Security Inc. All Rights Reserved. Copyright © Ordinal Technology Corp. All rights reserved. Copyright © Anadacht c.v. All rights reserved. Copyright Genivia, Inc. All rights reserved. Copyright Inc. All rights reserved. Copyright © Meta Integration Technology, Inc. All rights reserved. Copyright © Intalio. All rights reserved. Copyright © ComponentSource. All rights reserved. Copyright © Adobe Systems Incorporated. All rights reserved. Copyright © BataArt, Inc. All rights reserved. Copyright © ComponentSource. All rights reserved. Copyright © Microsoft Corporation. All rights reserved. Copyright © Glyph & Cog, LLC. All rights reserved. Copyright © Thinkmap, Inc. All rights reserved. Copyright © ComponentSource. All rights reserved. Copyright © Information Builders, Inc. All rights reserved. Copyright © OSS Nokalva, Inc. All rights reserved. Copyright Edifecs, Inc. All rights reserved. Copyright © Information Builders, Inc. All rights reserved. Copyright © International Organization for Standardization 1986. All rights reserved. Copyright © embedines Corporation. All rights reserved. Copyright © International Business Machines Corporation. All rights reserved. Copyright © Deserved. Copyright © International Business Machines Corporation. All rights reserved. Copyright © International Business Machines Corporation. All rights reserved. Copyright © Deserved. Copyright © International Business Machines Corporation. All rights reserved. Copyright © Deserved. Copyright © Exerved. Copyright © Exer

This product includes software developed by the Apache Software Foundation (http://www.apache.org/), and/or other software which is licensed under various versions of the Apache License (the "License"). You may obtain a copy of these Licenses at http://www.apache.org/licenses/. Unless required by applicable law or agreed to in writing, software distributed under these Licenses is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the Licenses for the specific language governing permissions and limitations under the Licenses.

This product includes software which was developed by Mozilla (http://www.mozilla.org/), software copyright The JBoss Group, LLC, all rights reserved; software copyright © 1999-2006 by Bruno Lowagie and Paulo Soares and other software which is licensed under various versions of the GNU Lesser General Public License Agreement, which may be found at http:// www.gnu.org/licenses/lgpl.html. The materials are provided free of charge by Informatica, "as-is", without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose.

The product includes ACE(TM) and TAO(TM) software copyrighted by Douglas C. Schmidt and his research group at Washington University, University of California, Irvine, and Vanderbilt University, Copyright (©) 1993-2006, all rights reserved.

This product includes software developed by the OpenSSL Project for use in the OpenSSL Toolkit (copyright The OpenSSL Project. All Rights Reserved) and redistribution of this software is subject to terms available at http://www.openssl.org and http://www.openssl.org/source/license.html.

This product includes Curl software which is Copyright 1996-2013, Daniel Stenberg, <a href="mailto:daniel@haxx.se">daniel@haxx.se</a>. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http://curl.haxx.se/docs/copyright.html. Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

The product includes software copyright 2001-2005 (®) MetaStuff, Ltd. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http://www.dom4j.org/ license.html.

The product includes software copyright © 2004-2007, The Dojo Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http://dojotoolkit.org/license.

This product includes ICU software which is copyright International Business Machines Corporation and others. All rights reserved. Permissions and limitations regarding this software are subject to terms available at http://source.icu-project.org/repos/icu/icu/trunk/license.html.

This product includes software copyright © 1996-2006 Per Bothner. All rights reserved. Your right to use such materials is set forth in the license which may be found at http://www.gnu.org/software/kawa/Software-License.html.

This product includes OSSP UUID software which is Copyright © 2002 Ralf S. Engelschall, Copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland. Permissions and limitations regarding this software are subject to terms available at http://www.opensource.org/licenses/mit-license.php.

This product includes software developed by Boost (http://www.boost.org/) or under the Boost software license. Permissions and limitations regarding this software are subject to terms available at http://www.boost.org/LICENSE\_1\_0.txt.

This product includes software copyright <sup>®</sup> 1997-2007 University of Cambridge. Permissions and limitations regarding this software are subject to terms available at http://www.pcre.org/license.txt.

This product includes software copyright © 2007 The Eclipse Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at http://www.eclipse.org/org/documents/epl-v10.php and at http://www.eclipse.org/org/documents/edl-v10.php.

http://www.schneier.com/blowfish.html; http://www.jmock.org/license.html; http://xsom.java.net; http://benalman.com/about/license/; https://github.com/CreateJS/EaseJJS/blob/master/src/easeJjs/display/Bitmap.js; http://www.h2database.com/html/license.html#summary; http://jsoncpp.sourceforge.net/LICENSE; http://jdbc.postgresql.org/license.html; http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto; https://github.com/rantav/hector/blob/master/LICENSE; http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html; http://jibx.sourceforge.net/jibx-license.html; https://github.com/lokato/libgeohash/blob/master/LICENSE; https://github.com/libsodium/blob/master/LICENSE; http://one-jar.sourceforge.net/index.php?page=documents&file=license; https://github.com/EsotericSoftware/kryo/blob/master/License.txt; http://www.scalalang.org/license.html; https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt; http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html; https://github.com/documentcloud/underscore-contrib/blob/master/LICENSE, and https://github.com/apache/hbase/blob/master/LICENSE.txt.

This product includes software licensed under the Academic Free License (http://www.opensource.org/licenses/afl-3.0.php), the Common Development and Distribution License (http://www.opensource.org/licenses/cddl1.php) the Common Public License (http://www.opensource.org/licenses/cpl1.0.php), the Sun Binary Code License Agreement Supplemental License Terms, the BSD (http://www.opensource.org/licenses/bsd-license.php), the new BSD License (http://www.opensource.org/licenses/bsd-license.php), the Artistic License (http://www.opensource.org/licenses/artistic-license-1.0) and the Initial Developer's Public License Version 1.0 (http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/).

This product includes software copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. All rights reserved. Permissions and limitations regarding this software are subject to terms available at http://xstream.codehaus.org/license.html. This product includes software developed by the Indiana University Extreme! Lab. For further information please visit http://www.extreme.indiana.edu/.

This product includes software Copyright (c) 2013 Frank Balluffi and Markus Moeller. All rights reserved. Permissions and limitations regarding this software are subject to terms of the MIT license.

See patents at https://www.informatica.com/legal/patents.html.

DISCLAIMER: Informatica LLC provides this documentation "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of noninfringement, merchantability, or use for a particular purpose. Informatica LLC does not warrant that this software or documentation is error free. The information provided in this software or documentation may include technical inaccuracies or typographical errors. The information in this software and documentation is subject to change at any time without notice.

#### **NOTICES**

This Informatica product (the "Software") includes certain drivers (the "DataDirect Drivers") from DataDirect Technologies, an operating company of Progress Software Corporation ("DataDirect") which are subject to the following terms and conditions:

- 1. THE DATADIRECT DRIVERS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY. FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.
- 2. IN NO EVENT WILL DATADIRECT OR ITS THIRD PARTY SUPPLIERS BE LIABLE TO THE END-USER CUSTOMER FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF THE USE OF THE ODBC DRIVERS, WHETHER OR NOT INFORMED OF THE POSSIBILITIES OF DAMAGES IN ADVANCE. THESE LIMITATIONS APPLY TO ALL CAUSES OF ACTION, INCLUDING, WITHOUT LIMITATION, BREACH OF CONTRACT, BREACH OF WARRANTY, NEGLIGENCE, STRICT LIABILITY, MISREPRESENTATION AND OTHER TORTS.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa\_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2021-04-16

# **Table of Contents**

Preface
Informatica Resources
Informatica Network
Informatica Knowledge Base
Informatica Documentation
Informatica Product Availability Matrices
Informatica Velocity
Informatica Marketplace
Informatica Global Customer Support
Chapter 1: Introduction to PowerExchange for Vertica 8
PowerExchange for Vertica Overview
Introduction to Vertica
Chapter 2: PowerExchange for Vertica Configuration 10
PowerExchange for Vertica Configuration Overview
Prerequisites
Vertica Configuration File for Linux
Configuration Options
Sample Vertica Configuration File
Vertica Configuration for Windows
PowerExchange for Vertica Plug-in Registration
Upgrading PowerExchange for Vertica
Chapter 3: Vertica Sources and Targets 14
Vertica Sources and Targets Overview
Configuring ODBC Data Sources on Windows
Configuring ODBC Data Sources on Linux
Importing Vertica Source and Target Definitions
Chapter 4: Vertica Mappings
Vertica Mappings Overview
Vertica Mapping Example
Chapter 5: Vertica Sessions
Vertica Sessions Overview
PowerExchange for Vertica Connections
Vertica Relational Connection Properties
Vertica Bulk Connection Properties
Configuring a Vertica Connection

Copy Method	22
Exception File	22
Rejected File	23
Non-Data Nodes	24
Staging File and Compression	24
Session Configuration for Vertica Targets in Bulk Mode	24
Partitioning for Vertica Targets	26
Appendix A: Vertica Data Type Reference	:8
Data Type Reference Overview	28
Vertica and Transformation Data Types	28
Index 3	1

# Preface

Use the *Informatica® PowerExchange®* for *Vertica User Guide* to learn how to extract data from and load data to Vertica by using PowerCenter Client. Learn to create a Vertica connection, develop mappings, and run sessions in an Informatica domain.

# Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

#### Informatica Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit <a href="https://network.informatica.com">https://network.informatica.com</a>.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- · Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

# Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit <a href="https://search.informatica.com">https://search.informatica.com</a>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at <a href="https://ksearch.informatica.com">KB\_Feedback@informatica.com</a>.

#### Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit <a href="https://docs.informatica.com">https://docs.informatica.com</a>.

## Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at <a href="https://network.informatica.com/community/informatica-network/product-availability-matrices">https://network.informatica.com/community/informatica-network/product-availability-matrices</a>.

## Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at <a href="http://velocity.informatica.com">http://velocity.informatica.com</a>. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at <a href="mailto:ips@informatica.com">ips@informatica.com</a>.

# Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at https://marketplace.informatica.com.

## Informatica Global Customer Support

You can contact a Global Support Center by telephone or through the Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

https://www.informatica.com/services-and-training/customer-success-services/contact-us.html.

To find online support resources on the Informatica Network, visit <a href="https://network.informatica.com">https://network.informatica.com</a> and select the eSupport option.

## CHAPTER 1

# Introduction to PowerExchange for Vertica

This chapter includes the following topics:

- PowerExchange for Vertica Overview, 8
- Introduction to Vertica, 8

# PowerExchange for Vertica Overview

PowerExchange for Vertica provides connectivity between PowerCenter and Vertica.

You can use relational mode or bulk mode to process Vertica data.

In bulk mode, you can use PowerExchange for Vertica to write large volumes of data to a Vertica database. You can then run queries on the Vertica database to analyze the data. You use an ODBC data source to connect to the Vertica database. When you run a Vertica session in bulk mode, the PowerCenter Integration Service creates a pipe, runs the copy command, and writes the data to a pipe. The copy command reads the data from the pipe and writes the data to the Vertica target.

You can also use relational mode to read or write data. Use relational mode when you want to read large volumes of data from a Vertica source or when you want to update or delete data in a Vertica target.

#### Example

You work for a social media organization that wants to track user activity on the website. You can use a Vertica database to store the user online activity. You can then query the Vertica database and use it for analysis. For example, you can use PowerExchange for Vertica to find where your top users are located.

# Introduction to Vertica

The Vertica database is a distributed storage system that you can use to store and analyze terabyte to petabytes of data.

Unlike traditional databases, Vertica uses a column-oriented approach for storing data. It stores each column separately on the disk. When you run a query on a Vertica database, only the columns that are called in the query will be accessed. The columns that are not called in the query will be ignored. Therefore, you can

process queries very efficiently since the amount of data accessed is much smaller than a traditional row store.

Vertica also uses encoding, sorting, and compression while storing data.

## CHAPTER 2

# PowerExchange for Vertica Configuration

This chapter includes the following topics:

- PowerExchange for Vertica Configuration Overview, 10
- Prerequisites, 11
- Vertica Configuration File for Linux, 11
- · Vertica Configuration for Windows, 12
- PowerExchange for Vertica Plug-in Registration, 12
- Upgrading PowerExchange for Vertica, 13

# PowerExchange for Vertica Configuration Overview

PowerExchange for Vertica is installed with PowerCenter.

You must configure PowerExchange for Vertica before you read data from or write data to a Vertica database. You can configure PowerExchange for Vertica on Windows and Linux operating systems.

To configure PowerExchange for Vertica, perform the following steps:

- 1. Complete the prerequisites.
- 2. Define the configuration options that the Vertica ODBC driver must use based on the operating system you use.
  - If you use Linux operating systems, create a configuration file named vertica.ini and define the Vertica configuration options in it. Define the VERTICAINI environment variable and set it to the path where you stored the vertica.ini configuration file.
  - If you use Windows operating systems, define the configuration options in the ODBC data source that
    you create to access the Vertica database.
- 3. If you use bulk mode to write data, register the PowerExchange for Vertica plug-in with the PowerCenter repository.

# **Prerequisites**

Before you configure PowerExchange for Vertica, perform the following prerequisite tasks:

- Install PowerCenter.
- 2. Install the Vertica database.
- Install the Vertica ODBC driver to import Vertica source and target definitions. You can download the Vertica ODBC driver from the following site: https://my.vertica.com/download/vertica/client-drivers/

# Vertica Configuration File for Linux

If you use Linux operating systems, the Vertica ODBC driver uses a Vertica configuration file to determine the configuration options that it must use while connecting to a Vertica database.

Use a text editor to create the Vertica configuration file and name the file as <code>vertica.ini</code>. In the <code>vertica.ini</code> file, define the configuration options that you want the Vertica ODBC driver to use to connect to a Vertica database.

After you configure the vertica.ini file, define an environment variable called VERTICAINI and set it to the path where you stored the vertica.ini configuration file.

## **Configuration Options**

The configuration options that you define in the vertica.ini file determine how the Vertica ODBC driver connects to the Vertica database.

#### DriverManagerEncoding

Defines the UTF encoding standard that the driver manager must use to connect to the Vertica database.

Set this option to UTF-16 for the Vertica ODBC driver.

#### **ODBCInstLib**

Defines the path where you stored the installer library file for the Vertica ODBC driver.

Set this option to the file path where you stored the libodbcinst.so file.

#### **ErrorMessagesPath**

Defines the path where the localized error message files for the Vertica ODBC driver are stored. These files are stored in the same directory as the Vertica ODBC driver files.

This option is required. If you do not set it and the ODBC driver encounters an error, the ODBC driver prints an error message stating that the ODBCMessages.xml file is missing.

#### LogLevel

Defines the minimum severity for the Vertica ODBC driver to log messages in the log file. These messages are specific to the connection errors that the Vertica ODBC driver encounters.

Set this option to one of the following values:

0. No logging

- 1. Fatal errors
- · 2. Errors
- · 3. Warnings
- 4. Informational messages
- 5. Debug messages
- · 6. Trace, all messages

For example, if you set the value to 4, the Vertica ODBC client driver logs all informational messages, warnings, errors, and fatal errors.

**Note:** The LogLevel value does not affect the tracing level session property that you specify in the PowerCenter session properties. The tracing level session property determines the amount of detail that the PowerCenter Integration Service writes in the session log.

#### LogPath

Defines the path where the Vertica ODBC driver must create the log file.

# Sample Vertica Configuration File

You can use the sample Vertica configuration file as a guideline to create the file.

[VerticaDriverName]
DriverManagerEncoding=UTF-16
ODBCInstLib=<Informatica install Dir>/ODBC7.1/lib/libodbcinst.so
ErrorMessagesPath=/opt/vertica/lib64
LogLevel=4
LogPath=/tmp

# Vertica Configuration for Windows

If you use Windows operating systems, define the configuration options in the ODBC data source that you create to access the Vertica database.

For information about Vertica configuration for Windows, see <u>"Configuring ODBC Data Sources on Windows" on page 14.</u>

# PowerExchange for Vertica Plug-in Registration

To write data to a Vertica database in bulk mode, you must register the PowerExchange for Vertica plug-in with the PowerCenter repository. You do not need to register the plug-in with the repository to read or write data in relational mode. When you upgrade PowerCenter from an earlier version and want to write data in bulk mode, you must update the PowerExchange for Vertica plug-in registration.

A plug-in is an XML file that defines the functionality of PowerExchange for Vertica. The plug-in file for PowerExchange for Vertica is named VerticaConnector.xml. When you install PowerCenter, the installer copies the VerticaConnector.xml file to the following directory:

<Informatica Installation Directory>\server\bin\Plugin

To register the plug-in, the repository must be running in exclusive mode. Use the Administrator tool or the pmrep RegisterPlugin command to register the plug-in.

**Note:** If you do not have the Manage Services privilege to register the plug-in, contact the administrator who manages the PowerCenter Repository Service.

# Upgrading PowerExchange for Vertica

You can upgrade PowerExchange for Vertica to a later version to take advantage of new features. After you upgrade PowerExchange for Vertica, you must update the existing plug-in file registration to write data in bulk mode.

- 1. Upgrade PowerCenter.
- 2. Start the PowerCenter Repository Service in exclusive mode.
  - To change the PowerCenter Repository Service operating mode, you can use the Administrator tool or the *infacmd* UpdateRepositoryService command.
- 3. Run the pmrep connect command to connect to the PowerCenter Repository Service.
- 4. Run the pmrep upgradeMMDTablesForVertica command to upgrade PowerExchange for Vertica.
  - Enter the following command:
  - pmrep upgradeMMDTablesForVertica
- 5. In Informatica Administrator, click the **Plug-Ins** tab.
- 6. Click the Register Plug-in icon.
- Browse and select the Vertica plug-in file and select the option to update the existing plug-in file registration.
- 8. Enter your repository user name and password.
- 9. Click OK
  - The PowerCenter Repository Service updates the existing Vertica plug-in file registration.
- 10. Restart the PowerCenter Repository Service to run in normal mode.

# CHAPTER 3

# Vertica Sources and Targets

This chapter includes the following topics:

- Vertica Sources and Targets Overview, 14
- · Configuring ODBC Data Sources on Windows, 14
- Configuring ODBC Data Sources on Linux, 15
- Importing Vertica Source and Target Definitions, 15

# Vertica Sources and Targets Overview

Vertica source definitions and target definitions represent metadata for Vertica tables.

You can import Vertica source and target definitions. You can also create Vertica target definitions based on a source definition, or manually create target definitions.

Before you import a Vertica source definition or target definition, you must create an ODBC data source for the Vertica database.

# Configuring ODBC Data Sources on Windows

On Windows operating systems, use the ODBC Data Source Administrator to configure an ODBC data source for each Vertica database that you want to access. Use the Vertica ODBC driver to create a DSN connection to the Vertica database.

- 1. Start the ODBC Data Source Administrator.
  - On a Windows 32-bit system, open the Control Panel and click Administrative Tools. Then, doubleclick Data Sources (ODBC).
  - On a Windows 64-bit system, enter the following command at the command prompt: %windir%\SysWOW64\odbcad32.exe

The ODBC Data Source Administrator dialog box appears.

- 2. Click the System DSN tab, and then click Add.
  - The Create New Data Source dialog box appears.
- Select the Vertica ODBC driver from the list of available drivers and click Finish.
   An entry for the Vertica ODBC driver is added under the System Data Sources section.

- 4. On the System DSN tab, select the entry for the Vertica ODBC driver, and click Configure.
  - The HP Vertica ODBC DSN Configuration dialog box appears.
- 5. Enter a DSN name.
- 6. Enter the server name, port number, and the name of the Vertica database that you want to connect to.
- Enter the user name and password to connect to the Vertica database.
- 8. Click **Test connection** to verify if the connection is successful.
- 9. Click **OK** to save the changes.
- 10. Configure additional options in the ODBC Data Source Administrator as needed.

# Configuring ODBC Data Sources on Linux

On Linux operating systems, update the odbc.ini file entries to configure an ODBC data source for each Vertica database that you want to access.

- Locate the odbc.ini file in the following directory: <Informatica Installation Directory>
   \ODBC7.1\odbc.ini
- 2. In the odbc.ini file, add entries for all the Vertica databases that you want to connect to.

The following section shows a sample Vertica entry in the odbc.ini file:

```
[ODBC Data Sources]
vertica odbc=libverticaodbc.so
[ODBC]
InstallDir=/export/home/vertica/Informatica/9.6.1/ODBC7.1
TraceFile=/export/home/vertica/Informatica/9.6.1/odbctrace.out
TraceDll=/export/home/vertica/Informatica/9.6.1/ODBC7.1/lib/DWtrc27.so
[vertica odbc]
Description = Vmart Database
Driver = /opt/vertica/lib64/libverticaodbc.so
Database = <database name>
Servername = <server name>
UID= <user ID>
PWD= <password>
Port = <port_number>
ConnSettings=
SSLKeyFile=
SSLCertFile=
Locale=UTF-8
```

3. Set the ODBCINI environment variable to the path where the odbc.ini file is stored.

# Importing Vertica Source and Target Definitions

To import a Vertica source or target definition, use an ODBC data source.

- In the Source Analyzer, click Sources > Import from Database. Alternatively, in the Target Designer, click Targets > Import from Database.
  - The **Import Tables** dialog box appears.
- 2. Select the ODBC data source that you created for the Vertica database.

3. Enter the user name and password to connect to the Vertica database.

If you are not the owner of the table that you want to use, specify the owner name.

4. Click Connect.

A list of Vertica tables appears.

- 5. Drill down through the list of database objects to view the available tables.
- 6. Select the Vertica table that you want to import into the PowerCenter repository.

You can press and hold the Shift key to select a block of tables, or press and hold the Ctrl key to make non-contiguous selections. You can also use the **Select all** and **Select none** buttons to select or clear all available targets.

#### 7. Click OK.

If you imported a Vertica source, the selected source definition appears under the Sources node in the Navigator. If you imported a Vertica target, the selected target definition appears under the Targets node in the Navigator.

The database type in the source and target definitions appears as Vertica.

# CHAPTER 4

# Vertica Mappings

This chapter includes the following topics:

- Vertica Mappings Overview, 17
- Vertica Mapping Example, 17

# Vertica Mappings Overview

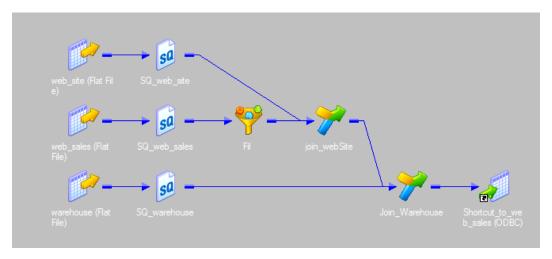
After you import a Vertica source or target definition into the PowerCenter repository, you can create a mapping to read data from a Vertica source or write data to a Vertica target.

You can read data from multiple Vertica sources and write data to multiple Vertica targets.

# Vertica Mapping Example

Your organization sells electronic products through multiple websites and warehouses. You want to determine the sales details for a particular warehouse and write the sales information to a Vertica database.

The following image shows the Vertica mapping example:



The Vertica mapping contains the following objects:

#### Sources

The mapping contains the following flat file sources:

- web site. Contains details of the websites through which customers can place orders for products.
- web sales. Contains details of the sales done through websites and warehouses.
- warehouse. Contains details of the warehouses from where the products are shipped.

#### **Transformations**

The  ${\tt Fil}$  filter transformation filters the sales details for a particular warehouse.

The  $join\_webSite$  Joiner transformation joins the sales details with the website details. The sales record contains the website key. The transformation uses the website key to retrieve website details from the web site source.

The  $Join\_Warehouse$  Joiner transformation joins the output of the  $join\_webSite$  Joiner transformation with the warehouse details.

#### **Target**

The target named Web\_sales is a Vertica database. Specify the columns in the Vertica table to which you want to write the data.

When you run the mapping, the PowerCenter Integration Service writes the sales information of the specified warehouse to the Vertica target database. You can then run queries on the Vertica target database to analyze the data.

# CHAPTER 5

# **Vertica Sessions**

This chapter includes the following topics:

- Vertica Sessions Overview, 19
- PowerExchange for Vertica Connections, 20
- Copy Method, 22
- Exception File, 22
- Rejected File, 23
- Non-Data Nodes, 24
- · Staging File and Compression, 24
- · Session Configuration for Vertica Targets in Bulk Mode, 24
- Partitioning for Vertica Targets, 26

# Vertica Sessions Overview

After you create a Vertica mapping in the Designer, you can create a session and configure the session to read data from or write data to a Vertica database. You can also run sessions on a grid.

You can use relational mode or bulk mode to process Vertica data. To read data, you must use relational mode. To write data, you can use relational mode or bulk mode.

#### **Relational Mode**

Use relational mode when you want to read large volumes of data from a Vertica source or when you want to update or delete data in a Vertica target.

To read data in relational mode, you must create a Vertica relational connection and configure the session to use a relational reader. Then, define the session properties for the relational reader. The session properties for relational mode are the same as that of relational sources.

Similarly, to write data in relational mode, you must create a Vertica relational connection and configure the session to use a relational writer. Then, define the session properties for the relational writer. The session properties for relational mode are the same as that of relational targets.

#### **Bulk Mode**

Use bulk mode when you want to insert large volumes of data into a Vertica target.

To write data in bulk mode, you must create a Vertica bulk connection and configure the session to use a Vertica bulk writer. Then, define the session properties that the copy command must use to write data to the Vertica target.

# PowerExchange for Vertica Connections

You must create a Vertica connection for each Vertica source or target that you want to access. When you configure a Vertica connection, you specify the connection attributes that the PowerCenter Integration Service uses to connect to Vertica.

The type of Vertica connection that you need to create depends on the mode that you use to read or write data. To read or write data in relational mode, you must create a Vertica relational connection. To write data in bulk mode, you must create a Vertica bulk connection.

# **Vertica Relational Connection Properties**

The following table describes the properties that you must configure for a Vertica relational connection:

Property	Description
Name	Enter a name for the connection.
Туре	The connection type is set by default. You cannot edit this value.
User Name	Enter the user name to connect to the Vertica database.  The user must have read permissions on the schema where the Vertica source table resides. Similarly, the user must have write permissions on the schema where the Vertica target table resides.
Password	Enter the password to connect to the Vertica database.
Use Parameter in Password	Indicates that the password for the repository user name is a session parameter, \$ParamName. Define the password in the workflow or session parameter file, and encrypt it by using the pmpasswd CRYPT_DATA option.  Default is disabled.
Connect String	Enter the name of the ODBC data source that you created for the Vertica database.
Code Page	Select the code page that the PowerCenter Integration Service must use to read or write data.
Connection Environment SQL	Runs an SQL command with each database connection.  Default is disabled.
Transaction Environment SQL	Runs an SQL command before the initiation of each transaction.  Default is disabled.
Connection Retry Period	The number of seconds that the PowerCenter Integration Service attempts to reconnect to the database if the connection fails. If the PowerCenter Integration Service cannot connect to the database in the retry period, the session fails.  Default value is 0.
Load Balance	Select this option to enable load balancing on the Vertica database. When you enable load balancing, the PowerCenter Integration Service requests a load balanced connection with the Vertica database.  Default is enabled.
	For more information on load balancing, see the Vertica database documentation.

# **Vertica Bulk Connection Properties**

The following table describes the properties that you must configure for a Vertica bulk connection:

Property	Description
Name	Enter a name for the connection.
Туре	The connection type is set by default. You cannot edit this value.
User Name	Enter the user name to connect to the Vertica database.
	The user must have write permissions on the schema where the Vertica target table resides.
Password	Enter the password to connect to the Vertica database.
Use Parameter in Password	Indicates that the password for the repository user name is a session parameter, \$ParamName.  Define the password in the workflow or session parameter file, and encrypt it by using the  pmpasswd CRYPT_DATA option.  Default is disabled.
Connect String	Enter the name of the ODBC data source that you created for the Vertica target.
Code Page	The code page is set by default. You cannot edit this value.
Schema	Enter the name of the schema where the Vertica target table resides.  Default is public.
Load Balance	Select this option to enable load balancing on the Vertica database. When you enable load balancing, the PowerCenter Integration Service requests a load balanced connection with the Vertica database.
	Default is enabled.
	For more information on load balancing, see the Vertica database documentation.

# Configuring a Vertica Connection

Before you run a Vertica session, create a Vertica connection.

- 1. In the Workflow Manager, connect to a repository.
- 2. Click Connections > Relational.

The **Relational Connection Browser** dialog box appears.

3. Click New.

The **Select Subtype** dialog box appears.

- 4. Create a Vertica connection.
  - To create a Vertica relational connection, select Vertica from the Select Subtype list.
  - To create a Vertica bulk connection, select Vertica Bulk Connection from the Select Subtype list.
- Click **OK**.

The Connection Object Definition dialog box appears.

- 6. Enter the connection properties.
- 7. Click OK.

The Vertica connection appears in the **Connection Browser** list.

# Copy Method

The copy method defines how the copy command writes data to the Vertica target. It defines the type of container to which the copy command writes the data.

You can write data to the following types of containers in Vertica:

- Write Optimized Store in memory (WOS) container. WOS is a memory-resident data structure that enables high-speed data loading. WOS stores data without any compression or indexing.
- Read Optimized Store on disk (ROS) container. ROS is a highly optimized, read-oriented, disk storage structure that compresses and indexes data.

You can set the copy method to one of the following values:

#### Auto

The copy command writes the data continuously to a WOS container. When the WOS container is full, the copy command writes the data to a ROS container.

Select this method when you want to write multiple data sets that are lesser than 100 MB in size. Since the WOS container does not compress and index data, you can write multiple data sets at high speed and then consolidate the data sets into a ROS container.

#### **Direct**

The copy command directly writes the data to a ROS container without writing it first to a WOS container.

Select this method when you want to write data sets that are greater than 100 MB in size, at a time. Since the ROS container compresses and indexes data, you can reduce the disk storage and improve query execution.

#### Trickle

The copy command writes the data incrementally to a WOS container. When the WOS container is full, the copy command displays an error and rolls back the entire data load.

Select this method when you have written your initial data and want to write additional data on an ongoing basis. For example, you can write 10,000 to 100,000 rows per transaction into the WOS container.

# **Exception File**

When you run a session in bulk mode, you can configure the PowerCenter Integration Service to log the copy command exceptions in an exception file. Use the exception file to understand why the exceptions occurred and take corrective action.

The exception file contains the following details:

- The record number that was rejected, along with the reason for the exception
- The file path and name of the rejected file that contains the rejected record
- · The pipe location from where the copy command read the record

If you use partitions, the PowerCenter Integration Service creates an exception file for each partition.

By default, the PowerCenter Integration Service creates the exception file in the following path:

\$PMRootDir/Temp/

If you want to change the default path, you can define the path in the session properties.

The PowerCenter Integration Service uses the following naming convention for the exception file:

```
Exceptions_PartNum_[Partition_Number]_[Table Name]_[Schema
Name]_[Counter]_<milliseconds*1000>.dat
```

#### Sample Exception File

```
COPY: Input record 8 has been rejected (int8 out of range '-9223372036854775808' for column 1 (COL1_INTEGER)). See /export/qa_adp/PMATT_96/work_dir/Vertica/Linux.64_RH6.1/exec_dir/Temp/RejectedData_PartNum_0_V0001_INTEGER_TGT_adpqa_schema_0_805000.dat, record 1 for the rejected record. This record was read from V0001_INTEGER_TGT_adpqa_schema_0_pipe_0

COPY: Input record 9 has been rejected (int8 out of range '-9223372036854775808' for column 1 (COL1_INTEGER)). See /export/qa_adp/PMATT_96/work_dir/Vertica/Linux.64_RH6.1/exec_dir/Temp/RejectedData_PartNum_0_V0001_INTEGER_TGT_adpqa_schema_0_805000.dat, record 2 for the rejected record. This record was read from V0001_INTEGER_TGT_adpqa_schema_0_pipe_0

COPY: Input record 10 has been rejected (int8 out of range '-9223372036854775808' for column 1 (COL1_INTEGER)). See /export/qa_adp/PMATT_96/work_dir/Vertica/Linux.64_RH6.1/exec_dir/Temp/
```

RejectedData PartNum 0 V0001 INTEGER TGT adpqa schema 0 805000.dat, record 3 for the rejected

record. This record was read from V0001\_INTEGER\_TGT\_adpqa\_schema\_0\_pipe\_0

# Rejected File

When you run a session in bulk mode, you can configure the PowerCenter Integration Service to write the rejected records to a rejected file. The rejected file contains all the records that were rejected by the copy command and not written to the Vertica target.

The rejected file displays the record details of each column separated by a | character. Therefore, if you have sensitive data, you must encrypt the data before you run the session.

If you use partitions, the PowerCenter Integration Service creates a rejected file for each partition.

By default, the PowerCenter Integration Service creates the rejected file in the following path:

```
$PMRootDir/Temp/
```

If you want to change the default path, you can define the path in the session properties.

The PowerCenter Integration Service uses the following naming convention for the rejected file:

```
RejectedData_PartNum_[Partition_Number]_[Table Name]_[Schema
Name] [Counter] <milliseconds*1000>.dat
```

#### Sample Rejected File

```
"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"|"-9223372036854775808"
```

# Non-Data Nodes

When you run a session in bulk mode, you can configure the copy command to connect to a non-data node in the Vertica cluster. A non-data node is a high-performance node that is used for writing data.

# Staging File and Compression

When you write large volumes of data to a Vertica target in bulk mode, you can configure the session to create a staging file. You can also choose to compress the data to improve the session performance.

If you configure the session to create a staging file, the PowerCenter Integration Service writes the data to the staging file and then runs the copy command. The copy command reads the data from the staging file and writes the data to the Vertica target.

On UNIX operating systems, when you enable file staging, you can also choose to compress the data in a GZIP format. By compressing the data, you can reduce the size of data that is transferred over the network and improve session performance.

You can specify the compression factor to determine the level to which the data must be compressed. You can enter a value in the range of 1 to 9.

A value of 1 indicates that the PowerCenter Integration Service takes minimal time to compress the data, but the data compression level is low. A value of 9 indicates that the PowerCenter Integration Service takes maximal time to compress the data, but the data compression level is high in this case. For more information about configuring the data compression factor, see the GNU operating system documentation.

# Session Configuration for Vertica Targets in Bulk Mode

To write data in bulk mode, configure the session to use a Vertica bulk connection and Vertica bulk writer. The copy command uses the session properties to write data to the Vertica target.

You can configure the session properties for Vertica targets in the **Transformations** view on the **Mapping** tab. Define the properties for each target instance in the session.

#### Method

The method that the copy command uses to write data to the Vertica target.

The method is set to INSERT because you can only insert data to a Vertica target. You cannot update or delete data.

#### **Copy Method**

The copy method that the copy command uses to write data to the Vertica target.

You can specify the copy method as auto, direct, or trickle.

Default is auto.

#### **Delimiter**

The character that separates successive input fields.

Default is I.

#### **Escape As**

The character that treats special characters in the data as regular characters.

Default is \.

#### **Null AS**

The character that the PowerCenter Integration Service replaces null values with when it writes data to the pipe. When the copy command reads the data from the pipe and writes it to the Vertica target, it converts the Null AS value to NULL.

Default is \n.

#### Quote

The character within which each column value gets enclosed. This option helps improve the readability of the data.

Default is ".

#### **Exception File Path**

The file path where the PowerCenter Integration Service creates the exception file.

Default is \$PMRootDir/Temp/.

#### Rejected File Path

The file path where the PowerCenter Integration Service creates the rejected file.

Default is \$PMRootDir/Temp/.

#### **Abort On Errors**

Select this option to configure the PowerCenter Integration Service to fail the session when it encounters an error.

#### Truncate target option

Select this option to truncate the Vertica target table before the PowerCenter Integration Service writes the data to the target.

#### Windows pipe buffer size

The size that the PowerCenter Integration Service allocates to buffer data before it writes the data to the Vertica target.

Enter a value between 1 KB and 2096128 KB.

Default is 4 KB.

You might need to test different settings for optimal performance. This attribute is applicable if the Informatica servers run on Windows operating systems.

#### **Pipe Location**

The file path where the PowerCenter Integration Service creates the pipes that it uses for data transfer.

Default is \$PMRootDir/Temp/.

This attribute is applicable if the Informatica servers run on UNIX operating systems.

#### **Reject Max Records**

Enter the maximum number of records that the PowerCenter Integration Service can reject after which it must fail the session.

#### **Connect Non-Data Nodes**

Select this option to connect to a non-data node to improve performance. Non-data nodes are high-performance nodes that are used for data loading.

#### **Enable File Staging**

Select this option to create a staging file in a local directory.

Default is disabled.

#### Stage File Directory

The file path where the PowerCenter Integration Service creates the staging file.

#### **Data Compression mode**

On a UNIX operating system, use this option to compress the data before you write it to the staging file. The PowerCenter Integration Service uses this attribute only if you enable file staging. Otherwise, it ignores this attribute.

You can specify the data compression mode as GZIP.

#### **GZIP Compression factor**

If you select the GZIP data compression mode, set a compression factor in the range of 1 to 9.

Default is 6.

#### **Miscellaneous Attribute**

Use this field to configure any Vertica attribute that is currently not available in the session properties.

Enter any attribute that Vertica supports along with the attribute value.

Use the following format:

```
ATTRIBUTE NAME 1, ATTRIBUTE VALUE 1; ATTRIBUTE NAME 2, ATTRIBUTE VALUE 2; <and so on>
```

#### **Table Name Prefix**

Enter a target owner name or prefix to override the table name prefix in the mapping.

#### **Target Table Name**

Enter a target table name to override the table name in the mapping.

# Partitioning for Vertica Targets

If you need to write large amounts of data to a Vertica target in bulk mode, you can configure partitioning to improve session performance.

You can configure pass-through partitioning for Vertica sessions.

When you configure partitioning, the PowerCenter Integration Service runs multiple copy commands based on the number of partitions that you configured. For example, if you configured four partitions, the

PowerCenter Integration Service creates four pipes, and runs four copy commands. It then writes data into four pipes. Each copy command reads from one pipe and writes the data to the Vertica target.

If you configure partitioning, the PowerCenter Integration Service creates the exception file and the rejected file for each partition. Similarly, in the session log, the PowerCenter Integration Service prints the load statistics such as the applied rows and affected rows for each partition.

## APPENDIX A

# Vertica Data Type Reference

This appendix includes the following topics:

- Data Type Reference Overview, 28
- Vertica and Transformation Data Types, 28

# Data Type Reference Overview

PowerCenter uses the following data types in Vertica mappings:

- Vertica native data types. Vertica data types appear in Vertica definitions in a mapping.
- Transformation data types. Set of data types that appear in the transformations. They are internal data types based on ANSI SQL-92 generic data types, which the PowerCenter Integration Service uses to move data across platforms. They appear in all transformations in a mapping.

When the PowerCenter Integration Service reads source data, it converts the native data types to the comparable transformation data types before transforming the data. When the PowerCenter Integration Service writes to target, it converts the transformation data types to the comparable native data types.

# Vertica and Transformation Data Types

The following table compares the Vertica data types and the transformation data types:

Vertica Data Type	Transformation Data Type	Description
Bigint	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
Binary	Binary	1 to 104,857,600 bytes
Boolean	String	1 to 104,857,600 characters
Bytea	Binary	1 to 104,857,600 bytes
Char	String	1 to 104,857,600 characters

Vertica Data Type	Transformation Data Type	Description
Date	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
Datetime	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
Decimal	Decimal	Precision 1 to 28, scale 0 to 28
Double precision	Double	Precision 15
Float	Double	Precision 15
Float (N)	Double	Precision 15
Float8	Double	Precision 15
Int	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
Int8	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
Integer	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
Interval	String	1 to 104,857,600 characters
Money	Numeric (P, S)	Precision 1 to 28, scale 0 to 28
Number	Decimal	Precision 1 to 28, scale 0 to 28
Numeric	Decimal	Precision 1 to 28, scale 0 to 28
Raw	Binary	1 to 104,857,600 bytes
Real	Double	Precision 15
Smalldatetime	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
Smallint	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0
Time	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
Timestamp	Date/Time	Jan 1, 0001 A.D. to Dec 31, 9999 A.D. (precision to the nanosecond)
Tinyint	Bigint	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807 Precision 19, scale 0

Vertica Data Type	Transformation Data Type	Description
Varbinary	Binary	1 to 104,857,600 bytes
Varchar	String	1 to 104,857,600 characters

# INDEX

C	V
configuring ODBC data sources on Linux $\underline{15}$ on Windows $\underline{14}$	Vertica configuration file <u>11</u> configuration options <u>11</u> data types overview <u>28</u> mappings overview 17
P	sample configuration file <u>12</u> sources overview <u>14</u>
PowerExchange for Vertica	targets overview <u>14</u>
adapter overview <u>8</u>	Vertica data types
configuration overview <u>10</u>	mapping to transformation data types $\underline{2}$
sessions overview <u>19</u>	Vertica mappings
upgrading <u>13</u>	example <u>17</u>
PowerExchange for Vertica configuration	Vertica session properties
prerequisites <u>11</u>	configuration <u>24</u>
PowerExchange for Vertica connections	copy method <u>22</u>
bulk properties <u>21</u>	exception file 22
configuring 21	rejected file 23
relational properties <u>20</u>	Vertica sources
	importing 15
	Vertica targets
R	importing 15
TX.	partitioning 26
registering plug-ins	· - <del>-</del>
PowerExchange for Vertica 12	