Informatica

Informatica® Cloud Data Integration
July 2024

PostgreSQL CDC Connector

Informatica Cloud Data Integration PostgreSQL CDC Connector
July 2024

© Copyright Informatica LLC 2020, 2024

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.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica, the Informatica logo, Informatica Cloud, and PowerCenter 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. Required third party notices are included with the product.

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: 2024-07-12

Table of Contents

Preface 4

Chapter 1: Introduction to PostgreSQL CDC Connector 5

PostgreSQL CDC Connector usage scenario 6

Integration with the PowerExchange CDC environment 6

PostgreSQL CDC Connector asset types 8

Software requirements 8

Supported targets 9

PostgreSQL CDC Connector rules and guidelines 9

Prerequisite tasks for using PostgreSQL CDC Connector 11

Task flow for using PostgreSQL CDC Connector 12

Chapter 2: PostgreSQL CDC connections 13

PostgreSQL CDC connection properties 13

Chapter 3: Mappings and mapping tasks with PostgreSQL CDC Connector 16

Mapping configuration 16

PostgreSQL CDC source properties in mappings 17

Targets in mappings with PostgreSQL CDC sources 20

Mapping task configuration 22

Properties for parameterized sources and connections 23

CDC runtime properties in mapping tasks 24

Configuring a restart point for change data extraction 25

Advanced session properties in mapping tasks 26

Running a mapping task 27

Monitoring a mapping task 27

Chapter 4: PostgreSQL CDC administration 29

Upgrading to a new version of PowerExchange 29

Recovery and restart processing 29

Handling changes to extraction maps 30

Chapter 5: Data type reference 32

Data type reference overview 32

PostgreSQL and transformation data types 32

Chapter 6: Connection overrides reference 36

CDC connection overrides 36

Index 39

Preface

Learn how to use *PostgreSQL CDC Connector* to retrieve PostgreSQL change records from a PowerExchange change data capture (CDC) environment and write the changes to a supported target. The CDC connector works with Cloud Data Integration mappings, mapping tasks, and taskflows. Learn how to configure CDC connections, mappings, and mapping tasks and to run and monitor mapping tasks.

CHAPTER 1

Introduction to PostgreSQL CDC Connector

You can use PostgreSQL CDC Connector to connect to a PowerExchange CDC environment to retrieve change records for registered PostgreSQL source tables from the PowerExchange Logger for Linux, UNIX, and Windows.

PostgreSQL CDC Connector extracts change records from PowerExchange Logger for Linux, UNIX, and Windows log files and sends the change records to Data Integration. Data Integration can then transmit the change records to a supported target.

When you define mappings in Data Integration, you select the extraction maps that PowerExchange generated for PostgreSQL source tables when you registered the tables. The PostgreSQL CDC Connector metadata adapter can then retrieve the source metadata from the extraction maps and store this information in the cloud.

To extract source change records from PowerExchange Logger log files, you must run mapping tasks that are associated with mappings. A mapping task runs to the end-of-log (EOL) in the PowerExchange Logger log files and then stops. You can manually restart the mapping task from a specific point in the logs, create a schedule that runs the mapping task on a recurring basis, or run the mapping task continuously. If the mapping task fails, PostgreSQL CDC Connector can resume processing from the point of interruption based on recovery information that is stored on the target.

PostgreSQL CDC Connector must be able to connect to an active PowerExchange CDC environment that includes the PowerExchange Logger. You do not need to copy the PowerExchange dbmover configuration file or PowerExchange Logger pwxccl configuration file to the PostgreSQL CDC Connector system. For more information about configuring PowerExchange CDC for PostgreSQL sources and the PowerExchange Logger, see the *PowerExchange CDC Guide for Linux, UNIX, and Windows*.

PostgreSQL CDC Connector works with the Informatica Cloud Secure Agent. When you install and start the Secure Agent, it downloads the PostgreSQL CDC Connector components and PowerExchangeClient packages. When you define a runtime environment, you configure a Secure Agent group.

Note: PostgreSQL CDC Connector does not work with the Hosted Agent.

PostgreSQL CDC Connector usage scenario

Your organization needs to replicate real-time change data from a mission-critical PostgreSQL production system to minimize intrusive, non-critical work, such as offline reporting or analytical operations, on that system.

You can use PostgreSQL CDC Connector to connect to an on-premises PowerExchange CDC system from another Linux or Windows system where the Secure Agent is installed. PostgreSQL CDC Connector components retrieve metadata about PostgreSQL source tables and the change records that PowerExchange captured for these sources. The source metadata is stored in the Data Integration cloud. You can use the browser-based Informatica Intelligent Cloud Services interfaces to define PostgreSQL CDC connections, mappings, and mapping tasks. When you run the mapping tasks, PostgreSQL CDC Connector components in conjunction with Data Integration retrieve and process the change records that PowerExchange captured from the production database and transmit these change records to one or more targets. Multiple users can then access the latest data on the target systems.

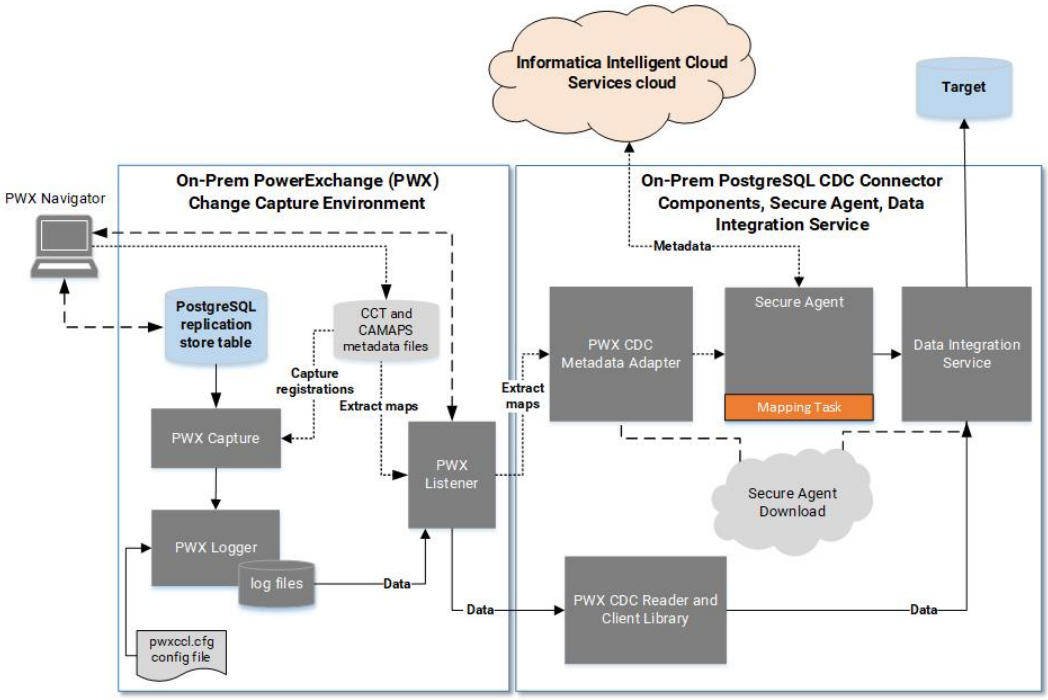
By using PostgreSQL CDC Connector, you reduce the workload on PostgreSQL production systems while making accurate and up-to-date PostgreSQL change data available to multiple users on less critical target systems. Also, you gain access to easy-to-use, browser-based interfaces for defining mappings and mapping tasks and for running and monitoring mapping tasks.

Integration with the PowerExchange CDC environment

You must have a functional on-premises PowerExchange CDC environment that can capture change records from PostgreSQL source tables and log them to PowerExchange Logger for Linux, UNIX, and Windows log files. Also, you must have a license that allows your organization to use the PostgreSQL CDC Connector and PowerExchangeClient packages on a Linux or Windows system.

On each system from which you plan to use PostgreSQL CDC Connector, you must install the Secure Agent. After you start the Secure Agent the first time, the PostgreSQL CDC Connector and PowerExchangeClient components are installed locally. You then can use the Cloud Data Integration, Administrator, and Monitor interfaces to configure PostgreSQL CDC connections, mappings, and mapping tasks and to run and monitor mapping tasks.

The following image shows the general architecture of the PostgreSQL CDC Connector components with Cloud Data Integration and a remote PowerExchange CDC and PostgreSQL source system:



In this configuration, the PowerExchange change capture process, PowerExchange Logger, and PostgreSQL source reside on one Linux, UNIX, or Windows system and the PostgreSQL CDC Connector components reside on another Linux or Windows system. The PowerExchange Navigator runs on a Windows system apart from both environments. Based on this image, the following processing occurs:

1. PowerExchange reads change events, including data changes, from the PostgreSQL replication store table. PowerExchange captures change records for registered source tables and sends the change records to the PowerExchange Logger on a continuous basis.

   To capture change records, the PowerExchange capture process uses the capture registrations that are recorded in the PowerExchange CCT file. You defined a capture registration for each PostgreSQL source table in the PowerExchange Navigator. The PowerExchange Navigator generated a corresponding extraction map, which will be used to extract the change records.

2. The PowerExchange Logger logs the change records in its log files.

   The pwxccl.cfg file defines the PowerExchange Logger configuration.

3. When you design a mapping in Data Integration, you associate the Source transformation with a PostgreSQL CDC connection. You can define the connection in either Data Integration or Administrator. The PWX CDC Metadata Adapter then uses the PostgreSQL CDC connection to connect to the PowerExchange Listener and retrieve extraction-map metadata for the PostgreSQL source from the PowerExchange CAMAPS file.

   In the mapping, you must also define a Target transformation and connection and also map the source fields to target fields.

4. The Secure Agent in conjunction with Data Integration sends the extraction-map metadata to the Informatica Intelligent Cloud Services repository in the cloud.

5. When you run the mapping task that is associated with the mapping, all of the CDC metadata, connection properties, mapping information, advanced source properties, CDC runtime properties, and advanced session properties are pushed down to the on-premises Secure Agent in the form of a session .xml file.

6. The PWX CDC Reader on the Secure Agent machine communicates with the PowerExchange Listener to retrieve the change records from PowerExchange Logger log files.
7. Data Integration transmits the change records to the mapped target.

Alternative configurations of PowerExchange with PostgreSQL CDC Connector components and Cloud Data Integration are possible. For example:

- The PowerExchange capture process, PowerExchange Logger, PostgreSQL source, PostgreSQL CDC Connector components, and Secure Agent can all run on the same Linux or Windows system.
- The PowerExchange Logger and its log files can reside on a system that is remote from the PostgreSQL source system and remote from the Linux or Windows system where the PostgreSQL CDC Connector components and Secure Agent run. By logging change data to a remote PowerExchange Logger, you can reduce resource consumption on the source system and possibly reduce the network overhead of data transfer. For more information, see "Remote Logging of Data" in the *PowerExchange CDC Guide for Linux, Unix, and Windows*.
- The PowerExchange Navigator can run on any supported Windows system, including a Windows system where the Secure Agent runs.

PostgreSQL CDC Connector asset types

The following table lists the Data Integration asset types that the PostgreSQL CDC Connector uses:

Asset Type	Required or Optional?
Mapping	Required
Mapping task	Required
Taskflow	Optional

Software requirements

Verify that your PowerExchange system and PostgreSQL CDC Connector system meet the following software requirements for CDC integration with Informatica Cloud Data Integration:

- You must have a local or remote PowerExchange installation. The following PowerExchange versions are supported:
  - PowerExchange 10.5.*x*
  - PowerExchange 10.4.1
  - PowerExchange 10.4.0 HotFix 1
  - PowerExchange 10.4.0
- PostgreSQL CDC Connector components run on the same Linux and Windows operating systems that are supported for the Informatica Secure Agent. For more information, see the Product Availability Matrix (PAM) for Informatica Intelligent Cloud Services (IICS) *<current_version>*, which is available on the Informatica Network at

https://network.informatica.com/community/informatica-network/product-availability-matrices. You must have a login to access this location.

- The supported versions of PostgreSQL source databases and operating systems are the same as those that PowerExchange supports. For more information, see the **PWX Classic** tab in the Product Availability Matrix (PAM) for Informatica <*pwx_version*>, which is available on Informatica Network at https://network.informatica.com/community/informatica-network/product-availability-matrices.

  Note: PowerExchange does not support PostgreSQL source databases deployed in a cloud environment.

Supported targets

PostgreSQL CDC Connector is certified with the following target types:

- Amazon Redshift
- Aurora PostgreSQL
- IBM Db2 for z/OS
- IBM Db2 for i
- Google BigQuery
- Microsoft SQL Server and Amazon Relational Database Service (RDS) for SQL Server
- MySQL
- Oracle and Amazon RDS for Oracle
- PostgreSQL
- Snowflake
- Teradata

Other Cloud Data Integration relational database targets are expected to work provided that the PM recovery tables can be created in the target schema. However, these targets have not been fully certified with this source connector. Use them in your test environment first.

PostgreSQL CDC Connector rules and guidelines

Review the PostgreSQL CDC Connector rules and guidelines before you begin creating connections, mappings, and mapping tasks.

- PostgreSQL CDC connections are for sources only. Do not specify a PostgreSQL CDC connection for a target.
- PostgreSQL CDC Connector connects to a PowerExchange Listener on the PowerExchange system by using the **Listener Location** property in the PostgreSQL CDC connection definition. On the Connector system, you do not need to create a local copy of the DBMOVER configuration file that contains a NODE statement pointing to the PowerExchange Listener.
- If you upgrade your system to a later PowerExchange version, you must stop the Secure Agent that connects to the older PowerExchange Listener before you try to connect to the upgraded PowerExchange Listener. After you edit the connection properties to specify the upgraded PowerExchange Listener, restart

the Secure Agent before you try to establish the connection again. For more information, see "Upgrading to a new version of PowerExchange" on page 29.

- For registered PostgreSQL source tables, PowerExchange captures DML changes (inserts, updates, and deletes) but does not capture DDL changes.
- Each PostgreSQL source table must have a primary key before it can be registered for CDC.
- For PostgreSQL CDC Connector, use of the PowerExchange Logger for Linux, UNIX, and Windows is required.
- Each PostgreSQL source table must have a primary key before it can be registered for CDC.
- PostgreSQL Connector does not support EnterpriseDB (EDB) PostgreSQL.
- The PowerExchange Listener can compress or encrypt the change data that it sends to the PWX CDC Reader component of PostgreSQL CDC Connector, depending on how you set the **Compression** and **Encryption** properties for the connection.
- PostgreSQL CDC Connector supports PowerExchange Logger group definition files. A *group definition file* creates a separate set of PowerExchange Logger log files for each defined group of registered PostgreSQL source tables. When a mapping task runs, the PWX CDC Reader reads only the subset of log files for the group that contains the source table.
- If you include multiple PostgreSQL source tables in a mapping, the PWX CDC Reader automatically uses group source processing to read change records for all of the tables in a single pass.
- The PWX CDC Metadata Adapter retrieves information for registered PostgreSQL source tables from the PowerExchange extraction-map metadata file, CAMAPS. The Adapter retrieves information only from PowerExchange-generated extraction maps that have names beginning with "d." The extraction maps provide information, such as the incoming fields, that will be used when you create a mapping. The Adapter does not retrieve information from user-defined extraction maps that have names beginning with "u" for user-defined.
- The PowerExchange change data extraction modes, as defined in the *PowerExchange CDC Guide for Linux, UNIX, and Windows*, do not pertain to the PostgreSQL CDC Connector. The PWX CDC Reader component of PostgreSQL CDC Connector extracts change records until it reaches the end-of-log in the PowerExchange Logger log files and then stops. However, the PowerExchange configuration that the PWX CDC Reader requires is similar to the configuration for continuous extraction mode in the following ways:
  - A CAPX CAPI_CONNECTION statement must exist in the PowerExchange DBMOVER configuration file.
  - The capture registrations for PostgreSQL source tables must specify **Part** for the **Condense** option.
- The maximum network buffer size that the PowerExchange Listener can use for sending data is 8 MB. If you have a mapping with an extremely large PostgreSQL source table, this maximum buffer size might be exceeded. In this case, the mapping task fails with the following error:

  ```
  PWXPC_10034 [date time] [ERROR] CDCDispatcher cannot fetch changed data from the
  PowerExchange Listener. Error code: [271]
  ```

- If you change the source object in a mapping that has been used for CDC processing, create a new mapping task for the mapping. If you need to retain the mapping task that is currently associated with the mapping, increment the revision number in the CDC Runtime properties for the mapping task.

Prerequisite tasks for using PostgreSQL CDC Connector

Before you begin configuring PostgreSQL CDC connections, mappings, and mapping tasks, you must complete several prerequisite tasks in PowerExchange and in Informatica Intelligent Cloud Services Administrator. Make sure you have a user role that allows you to perform these tasks.

- Verify that a supported PowerExchange version is installed and that PowerExchange is configured and able to capture change records for PostgreSQL source tables.
- In PowerExchange Navigator, check that an active PowerExchange capture registration and extraction map exist for each PostgreSQL source table. The extraction map name must begin with "d" for default, which indicates the default extraction map that PowerExchange generates.
- If your site uses TLS or SSL security, verify that valid certificates are available for the source and target servers and that the PWX_CONFIG environment variable points to the dbmover configuration file that includes the SSL configuration for the PowerExchange Listener. For more information about configuring TLS or SSL security, see the *PowerExchange Reference Manual*.
- If you localize PowerExchange messages in Japanese, before you start the Secure Agent, create a copy of the PowerExchange dbmover_ja.cfg file. Then set the PWX_CONFIG environment variable to point to the copy. For more information about the dbmover_ja.cfg configuration file, see the *PowerExchange Reference Manual*.
- Create the target tables and materialize them with data, if materialized target tables do not already exist. When users are ready to run mapping tasks the first time, they should start the tasks close to the time when the initial materialization of the target tables occurred to avoid missing changes.
- Verify that the PostgreSQL source system is configured for CDC. For more information, see the *PowerExchange CDC Guide for Linux, UNIX, and Windows*.
- In Administrator, verify that the **PostgreSQL CDC** Connector is available.
- Download the Secure Agent. The Secure Agent can then download the PostgreSQL CDC Connector components and PowerExchangeClient packages based on your licensing. The download also creates a Secure Agent group that contains a Secure Agent as the runtime environment. You can add another Secure Agent to the group. You must have at least one runtime environment in each organization so that users in the organization can run tasks.

  Note: PostgreSQL CDC Connector does not support the Hosted Agent.
- In Administrator, configure user roles if the predefined roles are not sufficient. Then assign one or more roles to each user or user group. The roles must provide the privileges that the users require to perform all of their tasks. Usually, the organization administrator configures user roles.
- If you do not want users to run scheduled mapping tasks during a certain period, define a blackout period for all schedules on the **Schedules** page in Administrator.

Task flow for using PostgreSQL CDC Connector

To use PostgreSQL CDC Connector, create a source connection, mapping, and mapping task. Also create a target connection if one is not already defined.

Before you begin, in the Administrator, you must complete the prerequisite tasks. See "Prerequisite tasks for using PostgreSQL CDC Connector" on page 11.

1. In Administrator, create a source connection with the type of **PostgreSQL CDC**. Test the connection.

   Alternatively, you can create a source connection in Data Integration when you configure a mapping. When you create a connection, you must set some connection properties that are specific to PostgreSQL CDC Connector. For more information, see "PostgreSQL CDC connection properties" on page 13.

   PostgreSQL CDC Connector uses this connection to retrieve source metadata and change data records from PowerExchange.
2. Create a target connection with a supported target type, if one is not already defined.
3. In Data Integration, configure a mapping. Validate the mapping and then save it.
4. Create a mapping task that is associated with the mapping.

   You can set optional CDC runtime properties and advanced session properties. For more information, see "CDC runtime properties in mapping tasks" on page 24 and "Advanced session properties in mapping tasks" on page 26.
5. Run the mapping task manually, or create a schedule to run it on a recurring basis. For more information, see "Running a mapping task" on page 27.
6. Monitor the job status for the mapping task instance. For more information, see "Monitoring a mapping task" on page 27.

   If a job has the status of Failed or Warning, check the session log for related error or informational messages. Then try to correct the issues that caused the failure and restart the mapping task from the appropriate point in the change stream.

CHAPTER 2

PostgreSQL CDC connections

You must create a PostgreSQL CDC connection to connect to a stream of change records that PowerExchange CDC for PostgreSQL captured from one or more PostgreSQL source tables.

The PWX CDC Reader that is on the system where the Secure Agent runs requires the connection information to retrieve PostgreSQL change records from PowerExchange.

You can create a PostgreSQL CDC connection in Administrator or Data Integration, provided that your assigned user role allows this action. Use one of the following methods:

- In Administrator, click **Connections** on the navigation bar. On the **Connections** page, click **New Connection** and then enter connection details and properties for a connection that has the type of **PostgreSQL CDC**.
- In Data Integration, when you create a mapping, click **Source** on the mapping canvas and then click **Source** in the **Source Properties** panel. To create a connection, click **New Connection**. Alternatively, select a connection from the **Connection** list.

When you configure a connection, you must specify a runtime environment for the connection. The runtime environment must contain a Secure Agent with the status "Up and Running."

After you create a PostgreSQL CDC connection, it becomes available for use within the organization and can be used in multiple mappings.

PostgreSQL CDC connection properties

When you configure a PostgreSQL CDC connection, you must set the connection properties.

The following table describes PostgreSQL CDC connection properties:

Property	Description
Connection Name	A name for the PostgreSQL CDC connection. This name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + - Spaces at the beginning or end of the name are trimmed and are not saved as part of the name. Maximum length is 100 characters. Connection names are not case sensitive.
Description	Description of the PostgreSQL CDC connection. Maximum length is 4000 characters.
Type	Type of connection. For PostgreSQL CDC, the type must be **PostgreSQL CDC**.

Property	Description
Runtime Environment	Name of the runtime environment that contains the Secure Agent that you want to use to run mapping tasks.

Listener Location	Host name or IP address of the system where both the PowerExchange Listener that processes PWX CDC Reader requests for PostgreSQL change data and the PowerExchange Logger for Linux, UNIX, and Windows run. Also include the Listener port number. Enter the value in the following format, where *host_name* can be a host name or IP address:
	`host_name:port_number`
	For example:
	`MYSCDC1A:1467`

User Name	A user name that can be used to access the PowerExchange Listener when PowerExchange Listener security is enabled. If you enabled PowerExchange LDAP user authentication, the user name is an enterprise user name. For more information, see the SECURITY statement in the *PowerExchange Reference Manual*.
Password	Password that is associated with the user name that is specified in the **User Name** property.
Collection Name	PostgreSQL instance name that is specified in the **Instance** field of the registration group that contains capture registrations for the PostgreSQL source tables. This value is used to filter the extraction-map metadata that the PWX CDC Metadata Adapter imports when using this connection.
CAPI Connection Name	Name of a CAPX CAPI_CONNECTION statement that is defined in the PowerExchange dbmover configuration file. This statement includes parameters that the PWX CDC Reader uses to extract change data from PowerExchange Logger for Linux, UNIX, and Windows log files. The PWX CDC Reader requires this property value and ignores any default CAPI_CONNECTION statement that is defined in the dbmover configuration file.
Connection Retry Period	Number of seconds that the PWX CDC Reader tries to reconnect to the PowerExchange Listener after the initial connection attempt fails. If a connection cannot be established within the retry period, the mapping task fails. The default is 0, which disables connection retries.
Compression	Controls whether the PowerExchange Listener compresses change data before sending the data over the network to the PWX CDC Reader. Select this property to compress the data. By default, this property is not selected.
Encryption	Controls whether the PowerExchange Listener encrypts change data before sending it over the network to the PWX CDC Reader. Also specifies the type of encryption to use. Select one of the following options: - **None**. Do not use encryption. - **AES 128-bit**. Use a 128-bit encryption key. - **AES 192-bit**. Use a 192-bit encryption key. - **AES 256-bit**. Use a 256 encryption key. The default is **None**.
Pacing Size	Amount of data, in rows or kilobytes, that the source system passes to the PowerExchange Listener before pausing to wait for another PWX CDC Reader request for more data. Decrease this value to improve session performance. Use 0 for maximum performance. The default and minimum value is 0.
Pacing Units	Type of units to use with the **Pacing Size** property. Select either **Rows** or **Kilobytes**.

Property	Description

Map Location	Host name or IP address of the system that contains the extraction maps. Also include the port number.
	This value is required when the PowerExchange Listener runs on a PowerExchange Logger for Linux, UNIX, and Windows machine that is remote from the extraction maps. The Listener requires access to the extraction maps to process change data extraction requests.
	Enter the value in the following format, where *host_name* can be a host name or IP address:
	`host_name:port_number`
	For example:
	`PSQCDC2B:25100`
	Note: The **Map Location** value takes precedence over the **Listener Location** value for testing connections and importing extraction-map metadata.

Map Location User	A user name that can access the PowerExchange Listener at the location that is specified in the **Map Location** property.
Map Location Password	Password associated with the user name that is specified in **Map Location User** property.
Event Table	If you created an event table to stop change data extraction based on user-defined events, enter the name of the PowerExchange extraction map for the event table. The event table must be a PostgreSQL table on the CDC source system.
Replace Low Values with Spaces	Controls whether to replace embedded nulls in character data with spaces. Select this property to replace embedded nulls. By default, this property is selected.
Custom Properties	Custom properties that you can specify to override PowerExchange default settings. You can enter multiple properties by using a semicolon (;) as the separator. Normally, you set custom properties only at the direction of Informatica Global Customer Support. **Note:** These properties are equivalent to the **PWX Override** options for the PowerExchange Client for PowerCenter (PWXPC) CDC connections in PowerCenter. To specify connection property overrides through a parameter file, you set a parameter in the form of `$<ParameterName>`, where you prefix a user-defined parameter name with a dollar sign character ($). Then configure a mapping task for the mapping to use a parameter file that contains the user-defined parameter definition by specifying the parameter file name in the **Parameter File Name** field on the **Runtime Options** tab. In the parameter file, set the connection overrides for the parameter in the format of name value pairs, where multiple values are delimited with a semicolon. For example: `$UserPass="User Name=jdoe;Password=mypassword"` The parameter name you specify in this field must match an entry defined in the parameter file. For more information, see the "Connection overrides reference" chapter.

CHAPTER 3

Mappings and mapping tasks with PostgreSQL CDC Connector

In Data Integration, create a mapping that includes the PostgreSQL CDC source and a target of a supported type. Then create a mapping task that is associated with the mapping to process change data based on the mapping.

The mapping and mapping task information is stored in the Informatica Intelligent Cloud Services cloud.

To process the PostgreSQL change records, run the mapping task in the runtime environment. The mapping task runs on premises, on the same machine where the Secure Agent and PowerExchange client components run. The mapping task uses the mapping and mapping task information pushed down from the Informatica Intelligent Cloud Services cloud. During a mapping task run, the PWX CDC Reader retrieves the captured PostgreSQL change records from the PowerExchange Logger log files, and Data Integration transmits the change records to the target based on the mapping.

After a mapping task ends, you can determine its job status and view the session log from either the Data Integration or Monitor interface.

Mapping configuration

You must configure a mapping that contains a Source transformation for a PostgreSQL CDC source and map it to a Target transformation.

To begin creating a mapping in Data Integration, click **New** > **Mappings** > **Mapping** and then click **Create**. In the **New Mapping** dialog box, click **Mapping** > **Continue**.

A mapping must contain at least one Source transformation mapped to one Target transformation. Alternatively, a mapping can contain multiple PostgreSQL CDC sources and multiple targets. For example, you can define multiple one-to-one mappings, map one source to multiple targets, or map multiple sources to one target. To map multiple sources to the same target, you can create multiple one-to-one mappings that map each source separately to the same target. In this case, each source must have a different primary key value.

Important: If you include multiple CDC sources in a mapping, all of the Source transformations must use the same PostgreSQL CDC connection, which connects to the same PowerExchange Listener and source location.

A mapping can also contain other transformation types, such as a Joiner transformation or Lookup transformation, as needed. If you include a Lookup transformation, make sure that you select either the **Lookup Source is Static** or **Lookup Cache Persistent** advanced property for the transformation. Otherwise,

any associated mapping task for which the **Recovery Strategy** advanced session property is set to **Resume from last checkpoint**, as recommended, will fail.

When you define a Source transformation in a mapping, you must specify a name for the transformation, a PostgreSQL CDC connection that can connect to the PowerExchange Listener on the PowerExchange Logger system, the source type, and the extraction map for a PostgreSQL source object. You can also set some additional source properties. The PWX CDC Metadata Adapter retrieves metadata for the source table from the PowerExchange-generated extraction map, including the field definitions.

When you define a Target transformation, specify a name for the transformation, a connection to the target, the target type, and the operation type of **Data Driven**. In the user's environment, the target can reside on premises or in a cloud. If you do not want to send all of the incoming fields to the target, you can define one or more exclusion field rules. As part of configuring the target, you must also map the incoming fields to the target fields. You can map fields automatically based on the field names or map them manually.

Instead of selecting specific connections and objects in the Source and Target transformations, you can define a parameter for the connection, the source or target object, or both. You then must enter a specific connection or object in the associated mapping tasks. If you want to use a parameter for the source or target connection, you must perform one of the following actions before the mapping can be successfully validated:

- Set the **Source Type** or **Target Type** field to **Parameter** and then enter a parameter for the source or target object in the **Parameter** field. On the **Field Mapping** tab of the Target transformation, you must set the map option to **Automatic** for the mapping to be validated.
- If you want to use a specific source or target object with a connection parameter, first select a specific connection, set the **Source Type** or **Target Type** field to **Single Object**, select a source or target object in the **Object** field, validate the mapping, and then enter the connection parameter.

  Tip: If you want to validate the source-to-target field mappings, map the fields in the Target transformation before validating the mapping. After the mapping is confirmed as valid, enter connection parameters.

You can also completely or partially parameterize field mappings, and then define specific mappings in the mapping tasks.

PostgreSQL CDC source properties in mappings

When you create a mapping, you configure a Source transformation for a PostgreSQL CDC source and set some source properties for the Source transformation.

Click the Source transformation box on the mapping canvas. Then configure the source properties in the **Source Properties** panel below the mapping canvas.

If you have multiple Source transformations in a mapping, configure source properties for each source. If you use a specific connection instead of a connection parameter, enter the same connection for all sources.

Note: If a mapping includes source tables or columns that have special characters in their names, the associated mapping task fails because it cannot import the source metadata. Special characters include s #, $, @, %, *, !, and ~. Therefore, ensure that none of the source tables and mapped columns have names with any special character. Rename the tables and columns if necessary.

The **Source Properties** panel contains the following tabs on which you can enter information:

- **General** tab. Specify a user-defined name and optional description for the PostgreSQL CDC source.
- **Source** tab. Configure the source properties under **Details** and under **Advanced**. The advanced properties are available only if you enter a specific CDC connection for the source. If you enter a parameter for the source connection, the advanced properties are not displayed. However, you can define these properties in the associated mapping task after you select a specific source connection for the task.

  Note: The Filter and Sort options under **Query Options** are not supported for CDC connections.

The following table describes the PostgreSQL CDC source properties under **Details**:

Property	Description
Connection	*Required*. Select a source connection of the type **PostgreSQL CDC**, or click **New Connection** to create one. Alternatively, you can define a connection parameter in the mapping and enter a specific connection in each mapping task that is associated with the mapping. If you want to use a connection parameter with a specific source object, you must first select a specific connection so that you can access the source to select the source object in the **Object** field. After you select the source object, define the connection parameter. If you want to use parameters for both the connection and source object, you do not need to first select a specific connection. You can just define the parameters in either order. To define a connection parameter, click **New Parameter** next to the **Connection** property and enter the following information: - Enter a parameter name that begins with a letter and is up to 200 characters in length, including @, #, _, and alphanumeric characters. - In the **Type** property, verify that **connection** is selected. - Leave the **Connection Type** property blank. After you click **OK**, the parameter appears in the **Connection** property.
Source Type	*Required*. Select **Single Object** to be able to select an extraction map for a PostgreSQL source table in the **Object** property. Alternatively, you can define a parameter for the source object in the mapping and enter a specific source object in each mapping task that is associated with the mapping. To define a parameter, select **Parameter** as the source type, and then specify the parameter in the **Parameter** property.
Object	Click **Select**. In the **Select Source Object** dialog box, under **Packages**, click a listed value. This value is the first part of the generated extraction map name. The right pane lists extraction maps that have names matching the selected package value. Select an extraction map. You can use only extraction maps that PowerExchange generated by default for capture registrations. These extraction maps have names that begin with "d." The extraction-map metadata is stored in the Informatica Intelligent Cloud Services repository in the cloud. The **Object** property appears only if you select **Single Object** as the source type. **Note:** The **Preview Data** feature is not supported for CDC source objects.
Parameter	Select a parameter for the source object, or click **New Parameter** to define a new parameter for the source object. The **Parameter** property appears only if you select **Parameter** as the source type.

The following table describes the optional source properties under **Advanced**:

Advanced Property	Description
Map Schema Override	A schema name that overrides the schema name in the PowerExchange data map.
Map Name Override	A map name that overrides the name defined in the PowerExchange data map.

Advanced Property	Description

Filter Overrides	One or more conditions that filter the source change records that the PWX CDC Reader retrieves from the PowerExchange Logger log files. Based on the filter conditions, a WHERE clause is appended to the default SQL SELECT query that the PWX CDC Reader uses to read the change records.
	You must use the nonrelational SQL syntax that PowerExchange supports for comparison and LIKE operators. For more information, see the "PowerExchange Nonrelational SQL" chapter in the *PowerExchange Reference Manual*. If the filter includes a date or timestamp column, ensure that the range of years in the column data is within the range of years for PowerExchange data filtering, as controlled by the DATERANGE statement in the DBMOVER configuration file. The default range is 1800 to 2200.
	For example, to select change records in which the TYPE column has a value of A or D, specify the following condition:
	`TYPE='A' or TYPE='D'`
	Maximum length of the entire SELECT statement that PowerExchange builds, including the WHERE clause with the filter conditions, is 8192 bytes.
	Note: If you specify the **Filter Overrides** attribute and also specify an **SQL Query Override** attribute value that contains a filtering WHERE clause, the resulting SELECT statement contains a WHERE clause that uses the AND operator to associate the Filter Overrides filter conditions with the SQL Query Override conditions. For example:
	`SELECT * from schema.table WHERE Filter_Overrides_conditions AND SQL_Query_Override_conditions`

SQL Query Override	An SQL statement that overrides the default SQL query that the PWX CDC Reader uses to retrieve change records from PowerExchange.
	You must use the nonrelational SQL syntax that PowerExchange supports. For more information, see the "PowerExchange Nonrelational SQL" chapter in the *PowerExchange Reference Manual*. If the override statement includes a date or timestamp column, ensure that the range of years in the column data is within the range of years for PowerExchange data checking and filtering, as controlled by the DATERANGE statement in the DBMOVER configuration file. The default range is 1800 to 2200.
	For example, to select change records for the USER source table when the TYPE column has the value of A or D, specify the following query:
	`SELECT ID, NAME from schema.extract_map.USER where TYPE='A' or TYPE='D';`
	Maximum length of the SELECT query override statement is 8192 bytes.
	Note: If you specify an **SQL Query Override** attribute value that contains a filtering WHERE clause and also specify the **Filter Overrides** attribute, the resulting SELECT statement contains a WHERE clause that uses the AND operator to associate the SQL Query Override conditions with the Filter Overrides filter conditions. For example:
	`SELECT * from schema.table WHERE Filter_Overrides_conditions AND SQL_Query_Override_conditions`

Advanced Property	Description

Connection Overrides	A variable or parameter file that overrides attributes at the mapping level for one or more sources. Use a semicolon (;) as the separator.
	For example, to override attributes by using a parameter file, use the following syntax:
	`RestartToken FileName=D:\abcagent\apps\Data_Integration_Server\data \restart\test.dat;APPLICATION=testapp`
	Overrides can be parameterized and can refer to other parameters. For example:
	`RestartToken File Name=$PMRootDir/restart/restartfile.txt`
	If you have more than one CDC source in the mapping, you only need to define the override in one of the sources. The override is automatically applied to all sources.
	Note: If you provide an override value for more than one source, all values must match to avoid a runtime error.
	For more information, see Chapter 6, "Connection overrides reference" on page 36.

Tracing Level	The level of detail that appears in the session log for a Source transformation. Select one of the following options, which are listed in order of increasing detail: - **Terse** - **Normal** - **Verbose Initialization** - **Verbose Data** Default is **Normal**.

- **Fields** tab. Select the fields that you want to use in the data flow. You can edit field metadata such as precision and scale if necessary.

Note: Ignore the **Partitions** tab. Key-range partitioning is not supported for CDC connections.

Targets in mappings with PostgreSQL CDC sources

To configure a target in a mapping, click the Target transformation box on the mapping canvas. Then configure the target properties in the **Target Properties** panel below the mapping canvas.

Note: To determine which target types are supported by the PostgreSQL CDC Connector, see "Supported targets" on page 9.

In the **Target Properties** panel, enter information on the following tabs:

- **General** tab. Specify a user-defined name and optional description for the target.
- **Incoming Fields** tab. Optionally review the included fields and define field exclusion rules and renaming criteria.
- **Target** tab. Configure the target properties under **Details** and under **Advanced**. The advanced properties are available only if you enter a specific connection for the target.

The following table describes the target properties under **Details**:

Property	Description
Connection	Select an existing target connection, or create a target connection. Alternatively, you can use a connection parameter in the mapping and then enter a specific connection in each mapping task that is associated with the mapping. If you want to use a connection parameter with a specific target object, you must first select a specific connection so that you can access the target to select the target object in the **Object** field. After you select the target object, define the connection parameter. If you want to use parameters for both the connection and target object, you do not need to first select a specific connection. You can just define the parameters in either order. To define a connection parameter, click **New Parameter** and enter the following information: - Enter a parameter name that begins with a letter and is up to 200 characters in length, including @, #, _, and alphanumeric characters. - In the **Type** property, verify that **connection** is selected. - Leave the **Connection Type** property blank.
Target Type	Select **Single Object** if you want to specify a target table. Alternatively, select **Parameter** and then specify an input parameter for the target object in the **Parameter** property. When you use a parameter, you specify the target table for the parameterized object on the **Targets** page of each mapping task associated with the mapping.
Object	Click **Select**. In the **Target Object** dialog box, select **Existing** and then select an existing target table. Alternatively, select **Create New at Runtime** and enter the name of a target table to generate at runtime. However, Informatica recommends that you do not use this option. The generated table might not be completely materialized with data the first time the mapping task runs. Also, with this setting, you cannot set the **Operation** property to **Data Driven**, which is required to properly handle subsequent insert, update, and delete records. After the table is generated, you must return to the **Object** field, select **Existing** and the generated target table, and then set the **Operation** property to **Data Driven**. **Note:** A generated target table contains the PowerExchange-generated DTL__ metadata fields from the source. If you do not want these columns on the target, you can configure a field rule that excludes the columns that have the "DTL__" prefix.
Parameter	Select or create a parameter for the target object. Appears only when the target type is **Parameter**.
Operation	Select **Data Driven** to properly handle insert, update, and delete records from the source. **Note:** By default, inserts, updates, and deletes are applied as inserts, updates, and deletes, respectively.
Data Driven Condition	Enables you to define expressions that flag rows for an insert, update, delete, or reject operation. Appears only when the operation type is **Data Driven** but is not required for this operation type.
Update Columns	Click **Add** to select the columns that you want to use as a logical primary key for performing update, upsert, and delete operations on the target. Appears only when the operation type is **Data Driven**. This field is not required if the target table already has a primary key.
Truncate target	Select this option if you want to clear the target table each time the mapping task runs. However, if you select this option and set the **Operation** property to **Data Driven**, update operations on the target will fail.

The following table describes the advanced target properties that appear under **Advanced**:

Advanced Property	Description
Forward Rejected Rows	Select this option to have the Target transformation pass rejected records to a reject file that is generated in the following directory: `\agent_install_directory\apps\Data_Integration_Server\data\error` Clear this option to ignore the rejected records. **Note:** This field is not displayed if a parameter is specified for the target connection.
Pre SQL	SQL statements that run on the target before the extracted data is written to a target. Maximum length is 5000 characters.
Post SQL	SQL statements that run on the target after the extracted data is written to a target. Maximum length is 5000 characters.
Update Override	An override SQL update statement to update the data in a target table. The override SQL update statement you specify overrides the default update statements that the Secure Agent uses to update targets based on key columns. You define an override update statement to update target tables based on non-key columns. In the override statement, you must enclose all reserved words in quotation marks.

- **Target Fields** tab. Optionally add fields, delete fields, or edit field metadata such as native data type, precision, and scale.
- **Field Mapping** tab. Map the incoming fields to the target fields. You can manually map an incoming field to a target field or automatically map fields based on the field names. If you entered a connection parameter, you must select **Automatic** as the field map option to be able to validate the mapping. Alternatively, you can completely or partially parameterize field mappings, and then define the specific field mappings in the mapping task. For more information, see Transformations > Target Transformation > Target Transformation Field Mappings in the Data Integration Help.

Mapping task configuration

Configure a mapping task for a mapping with a PostgreSQL CDC source. When you run the mapping task, the PostgreSQL change records are transmitted to the target. You cannot run mappings for PostgreSQL CDC.

Note: SQL ELT optimization and Visio templates are not supported for PostgreSQL CDC mapping tasks.

To begin configuring a mapping task, perform one of the following actions:

- In Data Integration, click **New** > **Mapping Task** and then click **Create**.
- While the mapping is open, select **New Mapping Task** from the **Actions** menu.

Then enter information on some or all of the following pages in the mapping task wizard:

- **General Properties**. Enter a name for the mapping task, a project or folder location for the mapping task, and the runtime environment that contains the Secure Agent or Secure Agent Group under which the mapping task will run. If you began creating the mapping task without first opening the mapping, you must also select the mapping with which to associate the mapping task. The description is optional.
- **Sources**. This page appears only if you defined a parameter for the source connection or object in the mapping. If you defined a connection parameter, select or create the specific connection that you want

the mapping task to use. If you defined a source object parameter, select the extraction map for a specific source object. For more information, see "Properties for parameterized sources and connections" on page 23.

- **CDC Runtime**. This page appears only for CDC sources. Enter CDC-related properties that are used when the mapping task runs. From this page, you can reset the point in the PowerExchange Logger for Linux, UNIX, and Windows logs from which the PWX CDC Reader starts extracting change data. For more information, see "CDC runtime properties in mapping tasks" on page 24 and "Configuring a restart point for change data extraction" on page 25.
- **Targets**. This page appears only if you defined a parameter for the target connection or object in the mapping. For more information about configuring targets in mapping tasks, click the help icon on this page.
- **Runtime Options**. From this page you can optionally define a schedule for running the mapping tasks, email notifications, advanced options, parameter file location, SQL ELT optimization, execution mode, and advanced session properties. For more information, see "Advanced session properties in mapping tasks" on page 26.

Properties for parameterized sources and connections

The **Sources** page appears in the mapping task wizard if you defined a parameter for the PostgreSQL source connection or object in the associated mapping.

If you defined a parameter for the source connection in the mapping, you must select or create a specific CDC connection for the mapping task. Ensure that the connection type is consistent with the source object type. After a connection is selected, you can click **Advanced Options** to configure the optional **Filter Overrides** or **SQL Query Override** advanced property. For more information about these advanced properties, see "PostgreSQL CDC source properties in mappings" on page 17.

Note: If the mapping uses a specific source connection and a parameter for the source object, you can specify these properties in the mapping and also specify them in the mapping task. In this case, the properties in the mapping task override those in the mapping.

If you defined a parameter for the source object in the mapping, you must select a specific source object for the mapping task. After you select a source object, the **Preview All Columns** button appears under **Data Preview**. However, you cannot preview data for a CDC source object.

When you finish entering information on the **Sources** page, click **Validate** to validate your entries and then click **Next** to proceed to the **CDC Runtime** page.

Note: If the mapping specifies a connection parameter and you edit the mapping to change the source object, check the specific connection in the mapping task to determine if you need to reset it and then **Validate** the mapping task again.

CDC runtime properties in mapping tasks

When you create or edit a mapping task that processes change data from a CDC source, you can set optional properties on the **CDC Runtime** page. These properties are used when the mapping task runs.

The following table describes the optional CDC runtime properties:

Property	Description
Maximum Rows Per Commit	Maximum number of change records in a UOW that the mapping task can process. When this maximum is reached, the UOW change records are committed on the target. Valid values are 0 through 999999999. Default is 0, which causes this property to be ignored.
Minimum Rows Per Commit	Minimum number of change records that must exist in a UOW before the PWX CDC Reader can request a real-time flush of the change records to the target. Valid values are 0 through 999999999. Default is 0, which causes this property to be ignored.
Maximum Latency in Seconds	Maximum number of seconds that the PWX CDC Reader waits for more change records before flushing change data from partially filled or empty blocks and returning control to Data Integration. Valid values are 2 through 360. Default is 2.
Real-time Flush Latency in Milliseconds	Number of milliseconds that must elapse before the PWX CDC Reader requests a real-time flush of uncommitted data in the change data stream to the targets defined in the mapping. Valid values are -1 through 999999999. Default is 2000. **Note:** To control commit processing, you can also use the **UOW Count**, **Maximum Rows Per Commit**, and **Minimum Rows Per Commit** properties.
Restart Point	Displays the restart point in read-only mode. The restart point indicates the position in the change data stream from which the PWX CDC Reader will start reading change records for a source table. Usually, you set the restart point before running the mapping task the first time to coincide with the time when the target table was materialized. You might also need to set the restart point to reprocess change records or to perform some testing. To set the restart point, click **Select**. If you do not specify a restart point, the mapping task starts or restarts from the end of log by default.
Restart Revision	Displays the restart-point revision number in read-only mode. When you navigate to the **CDC Runtime** page, open the **Select Restart Point** dialog box, and click **OK**, the revision number is incremented by 1, regardless of whether you changed the restart point. **Note:** During a single mapping-task edit session, the revision number is incremented by 1 only once, even if you return to the **Select Restart Point** dialog box multiple times. When a new revision is created, the CDC Connector ignores the restart information in the PM_REC_STATE recovery table and uses the specified restart point the next time you run the mapping task. **Note:** If you run the mapping task and then change to a different source object, you must either increment the restart revision number or create a new mapping task. Initial value is 0.
UOW Count	Number of units of work (UOWs) in the change data stream that the PWX CDC Reader processes before it requests a real-time flush of change records to the targets in the mapping. Valid values are -1 through 999999999. The values -1 and 0 cause this property to be ignored. Default is -1.

Property	Description
Update as Delete and Insert	Controls whether each SQL update that is captured from a source table is processed as an update or as a delete followed by an insert. You can select this property to handle changes to primary keys in source tables, in which case the insert record contains the key changes. Otherwise, leave this property cleared for normal processing of inserts, updates, and deletes. By default, this check box is not selected, which causes updates to be processed as update operations.
Continuous Extraction	Controls whether the mapping task runs continuously to capture change data from the source. Selecting this property indicates that the mapping task runs until it is stopped in the Monitor. By default, this check box is not selected and the mapping task stops after a single execution.

Note: You can also use the Informatica Intelligent Cloud Services Data Integration REST API to view and update CDC runtime options. The **mttask** API resource lets you retrieve or update mapping task options. For example, you might use **mttask** to change a selected option in multiple mapping tasks by using the API endpoint provided for the connector. For more information, see the *REST API Reference*.

Configuring a restart point for change data extraction

When you create or edit a mapping task, you can specify a restart point. The restart point indicates the point in the PowerExchange Logger log files from which the PWX CDC Reader starts extracting change records that were captured from a CDC source table. By default, the restart point is the end of log for the first run. Thereafter, the mapping task restarts from where it last left off by default, unless you specify a restart point.

Usually, you set the restart point before you start the mapping task the first time to coincide with the time at which the mapped target was materialized with data. However, you might need to set the restart point at other times, such as when you need to reprocess change records or perform tests.

Note: The restart point applies to all source tables in the mapping that is associated with the mapping task.

1. On the **CDC Runtime** page, click **Select** next to the **Restart Point** property.

   The **Select Restart Point** dialog box appears.
2. To set a restart point, select one of the following options:
   - **End of log**. The latest change record in the Logger logs.
   - **Earliest available**. The oldest available change record in the Logger logs.
   - **Timestamp**. The logged change record associated with a timestamp you set. To set a timestamp, select a date from the date picker. Then enter the time in HH:MM:SS.s format in the adjacent list boxes, where *s* represents fractional seconds.
   - **PWX Token**. The PowerExchange restart token that was automatically generated during change data capture. To obtain the restart token, use one of the following methods:
     - Run a CAPX row test on the extraction map in PowerExchange Navigator. In the row test results window, select **Show Auto Generated Columns** to see the generated DTL__ columns. Take the value from the DTL__CAPXRESTART1 field and add eight zeroes at the end. Enter the edited value in the **PWX Token** field. For more information about running a CAPX row test, see the *PowerExchange Navigator User Guide*.
     - Run the PWXUCDCT utility with the REPORT_CDCT command. The report output shows the restart token for each PowerExchange Logger log file. For more information about the PWXUCDCT utility, see the *PowerExchange Utilities Guide*.

3. Click **OK**.

   On the **CDC Runtime** page, the **Restart Point** property displays the new restart point in read-only mode and the **Restart Revision** value is incremented by 1.

   Warning: Anytime the revision number is incremented, the restart point option is reapplied during the next mapping task run, even if you did not change the option. In this case, the mapping task does not start from the ending log position of the previous run and might read duplicate records or skip records. Ensure that this result is acceptable.

Advanced session properties in mapping tasks

When you create or edit a mapping task that processes change records from a PowerExchange CDC source, set some advanced session properties that pertain to CDC.

In the **Advanced Session Properties** section on the **Runtime Options** page, click **+** to add each advanced session property. The following table describes the advanced session property values to add:

Advanced Property	Required Value
Commit on End of File	**No** This setting avoids duplicate data from occurring on the target after a mapping task job is restarted. The default is **No**.
Recovery Strategy	**Resume from last checkpoint** This setting is the only valid value. It is required to enable the recovery of change data and CDC restart information on the target after a mapping task ends abnormally or is stopped before completing change data processing. With this setting, the mapping task can resume processing change data from the point of interruption in the change data stream. By default, recovery is disabled and the task status remains as Failed. If the task is included in a taskflow, the taskflow continues to run.

For information about other advanced session properties, see the topic "Advanced Session Properties" under Tasks > Data Integration Tasks in the Informatica Cloud Data Integration Help. The following advanced session properties are not supported or required for CDC:

- **$Source Connection Value**
- **$Target Connection Value**
- **Incremental Aggregation**
- **Reinitialize Aggregate Cache**
- **SQL ELT Optimization options**
- **Session Sort Order**
- **Cache LOOKUP() Function**
- **Pre-build Lookup Cache**
- **On Stored Procedure Error**
- **Data Column Delimiter**
- **Enable cross-schema SQL ELT optimization** (Clear this check box.)

Note: If you want to run a CDC mapping with restart capability, make sure that the **Allow the mapping task to be executed simultaneously** check box is cleared.

Running a mapping task

A mapping task runs on premises under control of the Secure Agent.

By default, a mapping task runs to the end of log in the PowerExchange Logger log files and then stops. You must restart the mapping task to resume change data extraction processing. Alternatively, you can create a schedule that runs the mapping task on a recurring or continuous basis without manual intervention. You can define the schedule to minimize the time between mapping task runs.

To run the mapping task, use any of the following methods:

- After you finish configuring the mapping task, click **Run**.
- Manually start the mapping task from the **Explore** page.
- Create a schedule that runs the mapping task based on a date, time, and repeat frequency, such as hourly, daily, weekly, monthly, or a specific number of minutes.
- Select the **Continuous Extraction** runtime property to run the mapping task continuously. The mapping task runs until you stop the task in the Monitor.
- Include the mapping task in a taskflow and then run the taskflow.

Monitoring a mapping task

After you run a mapping task, you can check the job status for the task to determine if the job succeeded, completed with warnings, or failed. If the job failed or ended with a warning, you can view the session log for the job to identify the error or warning conditions. For continuous mapping tasks, you can stop the mapping task from the Monitor.

Each time a user or schedule starts a mapping task, Informatica Intelligent Cloud Services creates a job to run the task.

1. To view a list of mapping task jobs, use one of the following methods:
   - In Data Integration, select **My Jobs**. The **My Jobs** page lists all jobs that you, the user who is currently logged in, started. It does not include jobs that were started by a schedule.
   - In Monitor, select **All Jobs** or **Running Jobs**. The **All Jobs** page lists all of the running and completed jobs that members of your organization started or that were started by a schedule. The **Running Jobs** page lists only the jobs that are currently running or that have completed within the last 5 minutes.
2. On the **My Jobs**, **All Jobs**, or **Running Jobs** page, review the information about the jobs you're interested in. The following information is displayed by default and shown for each job instance:
   - Start date and time
   - End date and time
   - Number of rows processed
   - Current state, which can be one of the following values: Queued, Starting, Running, Success, Stopped, Suspended, Warning, or Failed.

   Tip: You can change the columns that are displayed by right-clicking the column-heading area and selecting or deselecting column names in the list.

   For more information about a job, click the job name to open a page that contains job details and a download link for the session log. You can also download the session log by moving the cursor to the right end of the row for a job and clicking the **Download Log** icon.

3. To stop a Running job, hover the cursor over the Running job status and select the Clean Stop or Stop icon. Clean Stop attempts to complete the any change capture processing that is currently in progress before stopping the task. Stop ends the task immediately, regardless of whether change capture processing is complete.
4. If the job state is Failed or Warning, gather more information about the error or warning condition in the following ways:
   - For a Failed job, hover the cursor over the Failed icon in the **State** column. A brief description of the error condition appears, which you can copy.
   - Download the session log and look for error or warning messages that begin with the message ID prefix of "PWXPC" or "PWX."

   If possible, correct the error or warning condition and then restart the job. To restart the job from the **Jobs** or **My Jobs** page, move the cursor to the right end of the row for the job and click the **Restart** icon. If you set the **Recovery Strategy** session property to **Resume from last checkpoint**, the job restarts extraction processing from the recovery point that is recorded in the state table or state file on the target system to ensure that no change data is lost.

CHAPTER 4

PostgreSQL CDC administration

As the organization administrator of a PostgreSQL CDC Connector implementation, you might need to perform some administrative tasks to keep CDC processing of change records running smoothly.

To learn about some common administrative issues and tasks, read the following topics:

- "Upgrading to a new version of PowerExchange" on page 29
- "Recovery and restart processing" on page 29
- "Handling changes to extraction maps" on page 30

Upgrading to a new version of PowerExchange

The connector can replicate data from only a single version of PowerExchange. If you upgrade your on-premise environment to a later supported PowerExchange version, use this task to move your Informatica Intelligent Cloud Services Data Integration connections to the new PowerExchange environment.

1. Upgrade the PowerExchange installation. For more information, see the *PowerExchange Installation and Upgrade Guide*.
2. In Monitor, open the **Running Jobs** page.
3. Make sure that any running mapping tasks that connect to the previous PowerExchange version run to completion, or stop the task if necessary.
4. Stop the Secure Agent that is specified in the connection configuration for any connection to the previous PowerExchange version.
5. In Data Integration, edit the connections to point to the new PowerExchange version.
6. If necessary, edit the mapping task configurations to specify the appropriate connection to the new version of the PowerExchange environment.
7. Start the mapping tasks again, when appropriate.

Recovery and restart processing

Each type of CDC Connector supports recovery and restart processing. If a mapping task job fails or is stopped before the end of log, the Data Integration service in conjunction with the CDC client can use recovery information that is stored on the target system to resume the extraction of change records in the

PowerExchange Logger log files from the point of interruption. This functionality prevents change data loss and inconsistencies between the source and target.

To enable recovery and restart processing, you must set the **Recovery Strategy** advanced session property to **Resume from last checkpoint** on the **Runtime Options** page when you create or edit a mapping task. With this recovery strategy, the Data Integration service provides for the recovery of target tables, and the CDC client and PowerExchange provide for the recovery of the CDC restart information that is used to resume extraction processing in the log files.

CDC restart information consists of PowerExchange-generated restart tokens from the source system. For a relational target, the Data Integration service stores CDC restart information in the PM_REC_STATE recovery table in the target database. Data Integration in conjunction with the CDC client automatically creates this PM_REC_STATE table on the target.

For databases where the recovery tables are not automatically created, use the template script for creating the recovery tables:

- To replicate data to a target, you must create a recovery table on the target. To do so, you can modify the recovery_template.sql template provided in the PowerExchange root installation directory for the Linux, UNIX, and Windows installers. Update the template according to the target database requirements. For more information, see *PowerExchange Interfaces for PowerCenter*.
- To use recovery tables, set the **Recovery Strategy** advanced session property to **Resume from last checkpoint** on the **Runtime Options** page when you create or edit a mapping task. Set **Commit on End of File** to **No**.

In special situations, you can specify a restart point for a mapping task. Typically, the first time you start a mapping task, you specify a restart point that corresponds to the target materialization time so that no change records in the change stream are skipped. The default restart point is the end of log (EOL), which is the current point of CDC processing in the log. You can specify that extraction processing starts from the earliest available record in the log or from a specific date and time. When you use a time-based restart point, extraction processing starts in the log that contains the first unit-of-work (UOW) that has an end time later than the restart time.

When you specify a restart point, consider the following points:

- The restart point applies to all sources in the mapping that is associated with the mapping task.
- If you set a restart point that is too early, it might correspond to a expired log file. In this case, the PWX CDC Reader uses the earliest available record in the available log files.
- If you set a restart point that is later than the latest record in the log files, an error message is issued.

Note: Restart information is associated with a mapping task and a specific source and target combination. If you change the source object in a mapping, you must either create a new mapping task for the mapping or increment the restart revision number for the existing mapping task. To increment the restart revision number, navigate to the **CDC Runtime** page for the mapping task, open the **Select Restart Point** dialog box, and click **OK**. If you do not take one of these actions, the mapping task will fail the next time you run it.

Handling changes to extraction maps

If you add or drop columns in a PostgreSQL source table, you must create a new capture registration. PowerExchange generates an associated extraction map by default. Use the generated extraction map

instead of creating a user-defined extraction map. The PWX CDC Reader on the PostgreSQL CDC Connector system will retrieve information from the new generated extraction map.

To switch to the new table definition and preserve access to previously captured data, you must perform several steps in both the PowerExchange and PostgreSQL CDC Connector environments.

Note: You do not need to perform these steps if you selectively capture change data for a subset of columns, and the DDL changes do not affect any of these columns.

1. Stop data change activity (inserts, updates, deletes) on the PostgreSQL source table.
2. Verify that any change data that was captured under the current table definition has completed PostgreSQL CDC Connector extraction processing. Then stop all mapping tasks that extract change data for the table.
3. Shut down the PowerExchange Logger for Linux, UNIX, and Windows.
4. In the PowerExchange Navigator, open the original capture registration and set its status to **History**.
5. Add or drop columns in the source table.
6. In the PowerExchange Navigator, create a new capture registration for the table that reflects the column changes. Include the following settings:
   - In the **Condense** list, select **Part**.
   - In the **Status** list, select **Active**.
7. Warm start the PowerExchange Logger.
8. Alter the target table to reflect the source table changes, if necessary.
9. In Data Integration, for each mapping that includes the source table, perform the following steps:
   - Select the Source transformation. Under Source Properties, either click the **Synchronize** button on the **Fields** tab or select the table again in the **Object** field on the **Source** tab. Either action causes the PWX CDC Metadata Adapter to retrieve the updated extraction-map metadata.
   - If you changed the target table, select the Target transformation. Under Target Properties, click the **Synchronize** button on the **Target Fields** tab or select the table again in the **Object** field on the **Target** tab.
   - Edit the field mapping to reflect the added or removed columns.
10. If necessary, rematerialize the target tables.
11. In the mapping task, set a time-based restart point that coincides with the time of target rematerialization or let the mapping task restart from the last checkpoint that was recorded in the state table on the target.
12. Resume change activity on the source table.
13. Start the mapping task again.

CHAPTER 5

Data type reference

This chapter includes the following topics:

- Data type reference overview, 32
- PostgreSQL and transformation data types, 32

Data type reference overview

Data Integration uses the following data types in mappings and tasks with PostgreSQL CDC:

PostgreSQL CDC native data types

PostgreSQL CDC data types appear in the source and target transformations when you choose to edit metadata for the fields.

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 Secure Agent uses to move data across platforms. Transformation data types appear in all transformations in a mapping.

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

PostgreSQL and transformation data types

The following table lists the PostgreSQL CDC data types that Data Integration supports and the corresponding transformation data types:

PostgreSQL data type	Transformation data type
abstime	timestamp
array types	varchar
bigint	num64
bit	char

PostgreSQL data type	Transformation data type
bit varying	varchar
boolean	num8
box	varchar
bytea - hex and escape formats	varbin.
character	char
character varying	varchar
cid	num64
cidr	varchar
circle	varchar
date	varchar
daterange	varchar
double precision	double.
inet	varchar
int4range	varchar
int8range	varchar
integer	num32
interval	varchar
json types	varchar
line	varchar
lseg	varchar
macaddr and macaddr8	varchar
money	varchar
name	varchar
numeric	numchar, or If the precision is greater than 100 or the scale is greater than 50, varchar
numrange	varchar
oid	num64
path	varchar

PostgreSQL data type	Transformation data type
point	varchar
polygon	varchar
pg_lsn	varchar
real	float.
regclass	varchar
regconfig	varchar
regdictionary	varchar
regnamespace	varchar
regoper	varchar
regoperator	varchar
regproc	varchar
regprocedure	varchar
regrole	varchar
regtype	varchar
reltime	varchar
smallint	num16
text	varchar
time with time zone	time
time without time zone	time
timestamp with time zone	varchar
timestamp without time zone	varchar
tsquery	varchar
tsrange	varchar
tstzrange	varchar
tsvector	varchar
user-defined types, including enum	varchar
uuid	bin

PostgreSQL data type	Transformation data type
xid	num64
xml	varchar

CHAPTER 6

Connection overrides reference

When you create a mapping that uses a CDC connector type, you can specify connection overrides in the **Connection Overrides** field in the mapping's Source properties panel.

CDC connection overrides

The following table lists the connection overrides that the Data Integration CDC connectors support.

Connection Override	Description
User Name	A user name for connecting to the database.
Password	A password for the specified user or a valid passphrase.
Compression	Specifies whether or not to compress source data. Valid values are: - Y. - N. Default is N.
Encryption Type	Specifies the encryption level to use. Valid values are: AES This attribute has no default. **Note:** You must specify both Encryption Type and Encryption Level.
Encryption Level	Specifies the encryption level to use. Valid values are: 1. AES 128, 2. AES 192, and 3. AES 256. Default is 1. **Note:** You must specify both the Encryption Type and the Encryption Level.
Pacing Size	Specifies the pacing size. Valid values are: 0 to 999999999. Default is 0.
Interpret as Rows	Specifies whether to represent the pacing size in number of rows. Valid values are: - Y. - N. Default is Y.

Connection Override	Description
Image Type	Indicates whether Data Integration should send captured Updates as Updates, or as Deletes followed by Inserts. Valid values are: - AI. Process Updates as update operations. - BA. Process Updates as Deletes followed by Inserts. Default is BA.
Application Name	Uniquely identifies the GMD application.
RestartToken File Name	Fully qualified restart file name.
Number of Runs to Keep RestartToken File (*)	The maximum number of backup copies to keep of the restart token file. Valid values are: 0 to 9999. Default is 0.
Recovery Cache Folder	The name of the file cache folder that is used for recovery of the session. Default is `$PMRootDir/Cache`.
UOW Count	The number of units of work (UOWs) that Data Integration processes before it sends a commit request to commit data to all targets in the CDC session. Valid values are: 1 to 999999999. Default is 1.
Convert character data to string	Controls whether to convert character fields to string fields so that embedded nulls in data are processed as spaces. Valid values are: - Y. - N. Default is N.
Retrieve PWX log entries	Controls whether Data Integration writes informational messages related to a CDC session in the session log. Valid values are: - Y. - N. Default is Y.
Continuous Extraction	Valid values are: - 0. Stop extraction at the end of the log. - 1. Perform continuous extraction. Default is 0.
Real-time Flush Latency in milli-seconds	The number of milliseconds that must elapse before Data Integration sends a commit request to commit data to all targets in the CDC session. When this period elapses, Data Integration continues to read the changes in the current UOW until it reaches the end of the UOW. Then Data Integration sends a commit request so that the data can be committed to the targets. Valid values are: 0 to 999999999. Default is 0. The Real-time Flush Latency in milliseconds, Maximum Rows Per Commit, and Minimum Rows Per Commit attributes work together to control commit frequency.

Connection Override	Description
Maximum Rows Per Commit	The maximum number of change records that Data Integration processes before it sends a commit request to commit data to all targets in the CDC session. Data Integration does not wait for a UOW boundary to commit the change data. Valid values are: 0 to 999999999. Default is 0. The Real-time Flush Latency in milliseconds, Maximum Rows Per Commit, and Minimum Rows Per Commit attributes work together to control commit frequency.
Minimum Rows Per Commit	The minimum number of change records that the PowerExchange Listener must read from the change stream before it passes a commit record to the target. Before reaching this minimum, the PowerExchange Listener passes only change records to Data Integration and discards any commit records. Valid values are: 0 to 999999999. Default is 0. The Real-time Flush Latency in milliseconds, Maximum Rows Per Commit, and Minimum Rows Per Commit attributes work together to control commit frequency.
Event Table	The source event table.
CAPI Connection Name Override	A name that overrides the default CAPI connection name.

INDEX

C

CDC connectors
  advanced session properties in mapping tasks 26
connection
  overrides 36
connection overrides reference
  overview 36
connections
  PostgreSQL CDC connection properties 13

D

data type reference
  overview 32

M

mapping tasks
  advanced session properties for CDC 26
  CDC runtime properties 24
  monitoring mapping tasks 27
  PostgreSQL CDC 22
  running mapping tasks 27
mappings
  PostgreSQL CDC 16

P

PostgreSQL CDC
  mapping task configuration 22
  source properties in mappings 17
  task and asset types 8
  administration 29
  connection properties 13
  connections overview 13
  Connector architecture 6
  Connector software requirements 8
  Connector usage scenario 6
  data types 32
  handling extraction map changes 31
  introduction to the CDC Connector 5
  mapping configuration overview 16
  mapping task properties for parameterized sources 23
  prerequisite tasks for using the Connector 11
  rules and guidelines for Connector use 9
  supported target types 9
  target properties mappings 20
  task flow for using the Connector 12

R

restart and recovery
  CDC connectors 30
  CDC restart point for a mapping task 25