



Informatica® PowerExchange for Kafka
10.2 HotFix 1

User Guide for PowerCenter

Informatica PowerExchange for Kafka User Guide for PowerCenter
10.2 HotFix 1
December 2018

© Copyright Informatica LLC 2018, 2019

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.

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.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

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: 2019-01-21

Table of Contents

Preface	5
Informatica Resources.	5
Informatica Network.	5
Informatica Knowledge Base.	5
Informatica Documentation.	5
Informatica Product Availability Matrixes.	6
Informatica Velocity.	6
Informatica Marketplace.	6
Informatica Global Customer Support.	6
Chapter 1: Introduction to PowerExchange for Kafka	7
PowerExchange for Kafka Overview.	7
Introduction to Kafka.	7
Chapter 2: PowerExchange for Kafka Installation and Configuration	8
PowerExchange for Kafka Installation and Configuration Overview.	8
Installing and Configuring PowerExchange for Kafka.	8
Prerequisites.	9
Installing PowerExchange for Kafka.	9
Installing the Server Component.	10
Installing the Client Component.	11
Registering the PowerExchange for Kafka Plug-in.	11
Registering the PowerExchange for Kafka Plug-in from the Administrator Tool.	11
Registering the PowerExchange for Kafka Plug-in from the Command Line Interface.	12
Configuring the krb5.conf file to Read Data from or Write to a Kerberised Kafka Cluster (Optional). .	12
Configuring JVM Options to Access a Kerberised Kafka Cluster (Optional).	14
Chapter 3: Kafka Sources and Targets	16
Kafka Sources and Targets Overview.	16
Import Kafka Source and Target Definitions.	16
Sample JSON Schema.	20
Chapter 4: Kafka Mappings	21
Kafka Mappings Overview.	21
Kafka Mapping Example.	21
Chapter 5: Kafka Sessions	24
Kafka Sessions Overview.	24
Kafka Connections Overview.	25
Kafka Connection Properties.	25

Configuring a Kafka Connection.	25
Configure Kafka Source Session Properties.	25
Configure Kafka Target Session Properties.	26
Configure Message Recovery Strategy for a Kafka Session.	27
Rules and Guidelines for Kafka Sessions.	28
Appendix A: Kafka Datatype Reference.	29
Data Type Reference Overview.	29
Kafka and Transformation Data Types.	29
JSON Kafka Data Types and Transformation Data Types.	30
Index.	31

Preface

The *Informatica PowerExchange for Kafka User Guide for PowerCenter* provides information about extracting data from and loading data to Kafka topics. This guide is written for PowerCenter administrators and developers who are responsible for developing mappings, sessions, and workflows that extract data from and load data to Kafka. This guide assumes you have knowledge of Kafka and PowerCenter.

Informatica Resources

Informatica Network

Informatica Network hosts Informatica Global Customer Support, the Informatica Knowledge Base, and other product resources. To access Informatica Network, visit <https://network.informatica.com>.

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.
- View product availability information.
- 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 search Informatica Network for product resources such as documentation, how-to articles, best practices, and PAMs.

To access the Knowledge Base, visit <https://kb.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

To get the latest documentation for your product, browse the Informatica Knowledge Base at https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx.

If you have questions, comments, or ideas about this documentation, contact the Informatica Documentation team through email at infa_documentation@informatica.com.

Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. If you are an Informatica Network member, you can access PAMs at

<https://network.informatica.com/community/informatica-network/product-availability-matrixes>.

Informatica Velocity

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

If you are an Informatica Network member, you can access Informatica Velocity resources at <http://velocity.informatica.com>.

If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

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

Informatica Global Customer Support

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

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

<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>.

If you are an Informatica Network member, you can use Online Support at <http://network.informatica.com>.

CHAPTER 1

Introduction to PowerExchange for Kafka

This chapter includes the following topics:

- [PowerExchange for Kafka Overview, 7](#)
- [Introduction to Kafka, 7](#)

PowerExchange for Kafka Overview

You can use PowerExchange for Kafka for connectivity between PowerCenter and Kafka.

PowerExchange for Kafka sources and targets represents data in a Kafka stream.

You can use Kafka topics as sources and targets in mappings. When you use Kafka topics in mappings, you must configure properties specific to Kafka.

Example

You run the IT department of a major bank that has millions of customers. You want to monitor network activity in real time. You need to collect network activity data from various sources such as firewalls or network devices to improve security and prevent attacks. The network activity data includes Denial of Service (DoS) attacks and failed login attempts made by customers. The network activity data is written to Kafka queues.

You can use PowerExchange for Kafka to read network activity data from Kafka topics in JSON format and write the data to a target for processing the network activity data..

Introduction to Kafka

Kafka is an open-source scalable, fault tolerant, and distributed data streaming platform. You can process large amounts of data quickly by using Kafka. It can serve as an interim staging area for data that will be consumed by various types of downstream consumer applications.

Kafka runs as a cluster comprised of one or more servers each of which is called a broker. Kafka brokers stream data in the form of messages. These messages are published to a topic.

Kafka topics are divided into partitions. This gives better throughput and could be used to scale the number of messages processed. Message ordering is guaranteed only within partitions.

CHAPTER 2

PowerExchange for Kafka Installation and Configuration

This chapter includes the following topics:

- [PowerExchange for Kafka Installation and Configuration Overview, 8](#)
- [Installing and Configuring PowerExchange for Kafka, 8](#)
- [Prerequisites, 9](#)
- [Installing PowerExchange for Kafka, 9](#)
- [Registering the PowerExchange for Kafka Plug-in, 11](#)
- [Configuring the krb5.conf file to Read Data from or Write to a Kerberised Kafka Cluster \(Optional\), 12](#)
- [Configuring JVM Options to Access a Kerberised Kafka Cluster \(Optional\), 14](#)

PowerExchange for Kafka Installation and Configuration Overview

You can install PowerExchange for Kafka on a Windows 64-bit or Red Hat Enterprise Linux 64-bit machine.

Installing and Configuring PowerExchange for Kafka

To install and configure PowerExchange for Kafka, complete the following steps:

1. Complete the prerequisite tasks.
2. Install the PowerExchange for Kafka server and client components.
3. Register the PowerExchange for Kafka plug-in.
4. If you use a Kerberised Kafka cluster, configure the `krb5.conf` file to read data from or write data to a Kerberised Kafka cluster.
5. If you use a Kerberised Kafka cluster, configure JVM options for the PowerCenter Client and the PowerCenter Integration Service to access a Kerberised Kafka cluster.

Prerequisites

Before you install PowerExchange for Kafka, complete the following tasks:

1. Install or upgrade to Informatica Services 10.2 HotFix 1.
2. Install and configure the Informatica Developer and PowerCenter. You can install the Informatica Developer and PowerCenter when you install or upgrade to Informatica Client 10.2 HotFix 1.
Note: You must install the Informatica Developer if you want to successfully import a Kafka Data Object in JSON format.
3. Verify that you have read and write permissions on the following directories on each machine that runs the PowerCenter Integration Service and PowerCenter Repository Service:
 - `<Informatica installation directory>\server\connectors\cci\plugins`
 - `<Informatica installation directory>\server\bin\Plugin`
 - `<Informatica installation directory>\connectors\thirdparty`The installer must be able to add and overwrite files in these directories.
4. Verify that you have read and write permissions on the following directories on each machine where you installed the PowerCenter Client:
 - `<Informatica installation directory>\clients\PowerCenterClient`
 - `<Informatica installation directory>\clients\PowerCenterClient\bin\Help\<language>`The installer must be able to add and overwrite files in these directories.
5. Download the Informatica EBF-13264 and EBF-13366 at <https://tsftp.informatica.com/> from the following location:
`/updates/Informatica10/10.2.0 HotFix1`
For more information, contact Informatica Global Customer Support.
6. Ensure that you install the Apache Kafka cluster and the Informatica Services in different nodes.

For more information about product requirements and supported platforms, see the Product Availability Matrix on Informatica Network:

<https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Installing PowerExchange for Kafka

When you install PowerExchange for Kafka, you install the following components that allow PowerCenter to access the Kafka topics:

Client component

Allows you to import definitions, create mappings, and create connection objects with the PowerCenter Client.

Server component

Allows the PowerCenter Repository Service to store and access the Kafka metadata in the repository and the PowerCenter Integration Service to run Kafka sessions.

Installing the Server Component

The PowerExchange for Kafka server component installs the Informatica service components.

When you install the PowerExchange for Kafka server component, you enable the PowerCenter Integration Service to read data from or write data to Kafka.

If the PowerCenter Integration Service or PowerCenter Repository Service is configured to run on primary and backup nodes, install the PowerExchange for Kafka server component on each node configured to run the PowerCenter Integration Service or PowerCenter Repository Service.

If the PowerCenter Integration Service is configured to run on a grid, install the PowerExchange for Kafka server component on each node configured to run on the grid. If you cannot install the PowerExchange for Kafka server component on each node in the grid, create a resource in the domain and assign it to each node where you installed the PowerExchange for Kafka server component. When you create a session, configure the session to use the resource.

For example, create a custom resource called Kafka. When you create a session, you can configure the session to use the Kafka resource. The Load Balancer dispatches the session to nodes that has the resource.

Installing the Server Component on Windows

Install the PowerExchange for Kafka server component on Windows when the PowerCenter Integration Service or PowerCenter Repository Service runs on Windows.

1. Run `install.bat` from the installation package.
2. Click **Next**.
3. Select the Informatica installation directory.

By default, the server components are installed in the following directory:

```
C:\<Informatica installation directory>\<version folder>
```

4. Click **Next**.
5. Click **Install** to begin the installation.
6. Click **Done** when the installation is complete.
7. To successfully run mappings with JSON files, install the server component of the Informatica EBF-13366.

Installing the Server Component on Linux

Install the PowerExchange for Kafka server component on Linux when the PowerCenter Integration Service or PowerCenter Repository Service runs on Linux.

To install the PowerExchange for Kafka server component on Linux platforms that support the graphical user interface, perform the same steps that you use to install the server components on Windows.

To install the PowerExchange for Kafka server component on Linux platforms that use the command line interface, perform the following steps:

1. Enter `install.sh` at the prompt.
2. Enter the path to the Informatica installation directory.

By default, the server components are installed in the following directory:

```
<User Home Directory>/Informatica/<version folder>
```

3. To successfully run mappings with JSON files, install the server component of the Informatica EBF-13366.

Installing the Client Component

Install the Client component on each PowerCenter Client machine where you want to create or access the Kafka topics.

1. Run `install.bat` from the installation package.
2. Click **Next**.
3. Select the Informatica installation directory.

By default, the client is installed in the following location:

```
C:\Informatica\<version folder>
```

4. Click **Next**.
5. Click **Install** to begin the installation.
6. Click **Done** when the installation is complete.
The client component is installed.
7. To successfully run mappings with JSON files, install the client component of the Informatica EBF-13366.

Registering the PowerExchange for Kafka Plug-in

To register the plug-in, the repository must be running in exclusive mode. Use the Administrator tool or the `pmrep RegisterPlugin` command line program to register the plug-in. If you do not have the correct privileges to register the plug-in, contact the user who manages the PowerCenter Repository Service.

The plug-in file is an `.xml` file that defines the functionality of the adapter. When you install the server component, the installer copies the plug-in file to the following directory:

```
<Informatica installation directory>/server/bin/Plugin
```

The name of the plug-in file for PowerExchange for Kafka is `kafka_Plugin.xml`.

Registering the PowerExchange for Kafka Plug-in from the Administrator Tool

Register a repository plug-in to add its functionality to the repository.

1. Run the PowerCenter Repository Service in exclusive mode.
2. In the **Navigator**, select the PowerCenter Repository Service to which you want to add the plug-in.
3. In the **Contents** panel, click the **Plug-ins** view.
4. In the **Actions** menu of the **Domain** tab, select **Register Plug-in**.
5. On the **Register Plugin** page, click the **Browse** button to locate the plug-in file.
6. Enter your user name and password.
7. Click **OK**.

The PowerCenter Repository Service registers the plug-in with the repository. The results of the registration operation appear in the activity log.

8. Run the PowerCenter Repository Service in normal mode.

Registering the PowerExchange for Kafka Plug-in from the Command Line Interface

You can use the `pmrep RegisterPlugin` command to register the plug-in from the command line program.

1. Run the PowerCenter Repository Service in exclusive mode.
2. Run the `pmrep Connect` command to connect to the Repository Service using a user account with Administrator Repository privilege.

The `RegisterPlugin` command uses the following syntax:

```
pmrep connect -r <repository name> -d <domain_name> -n <domain user name> -x  
<domain_password>
```

3. Find `kafka_Plugin.xml` in the following directory:

```
<Informatica installation directory>\server\bin\Plugin
```

4. Run the `pmrep RegisterPlugin` command to update the repository.

The `RegisterPlugin` command uses the following syntax:

```
pmrep registerplugin -i <Informatica installation directory>\server\bin\Plugin  
\kafka_Plugin.xml -e -N
```

Configuring the `krb5.conf` file to Read Data from or Write to a Kerberised Kafka Cluster (Optional)

To read from or write to a Kerberised Kafka cluster, configure the default realm, KDC, and Kafka source or target session properties.

Before you read from or write to a Kerberised Kafka cluster, perform the following tasks:

1. Ensure that you have the `krb5.conf` file for the Kerberised Kafka cluster.
2. Configure the default realm and KDC. If the default `/etc/krb5.conf` file is not configured or you want to change the configuration, add the following lines to the `/etc/krb5.conf` file:

```
[libdefaults]  
default_realm = <REALM NAME>  
dns_lookup_realm = false  
dns_lookup_kdc = false  
ticket_lifetime = 24h  
renew_lifetime = 7d  
forwardable = true  
  
[realms]  
<REALM NAME> = {  
kdc = <Location where KDC is installed>  
admin_server = <Location where KDC is installed>  
}  
  
[domain_realm]  
.<domain name or hostname> = <KERBEROS DOMAIN NAME>  
<domain name or hostname> = <KERBEROS DOMAIN NAME>
```

3. To pass a static JAAS configuration file into the JVM using the `java.security.auth.login.config` property at runtime, perform the following tasks:
 - a. Ensure that you have JAAS configuration file.

For information about creating JAAS configuration and configuring Keytab for Kafka clients, see the Apache Kafka documentation at <https://kafka.apache.org/0101/documentation/#security>

For example, the JAAS configuration file can contain the following lines of configuration:

```
//Kafka Client Authentication. Used for client to kafka broker connection
KafkaClient {
com.sun.security.auth.module.Krb5LoginModule required
doNotPrompt=true
useKeyTab=true
storeKey=true
keyTab="<path to keytab file>/<keytab file name>"
principal="<principal name>"
client=true
};
```

- b. Place the JAAS config file and keytab file in the same location on all the nodes. Informatica recommends that you place the files in a location that is accessible to all the nodes in the cluster. Example: /etc or /temp
- c. Configure the following properties:

Kafka connection

Configure the **Kerberos Configuration Properties** property in a Kafka connection and specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI
```

Sources

Configure the **Consumer Configuration Properties** property in the source session properties to override the value specified in the **Kerberos Configuration Properties** property in a Kafka connection. Specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI
```

Targets

Configure the **Producer Configuration Properties** property in the target session properties to override the value specified in the **Kerberos Configuration Properties** property in a Kafka connection. Specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI
```

4. To embed the JAAS configuration in the `sasl.jaas.config` configuration property, configure the following properties:

Kafka connection

Configure the **Kerberos Configuration Properties** property in a Kafka connection and specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI,
sasl.jaas.config=com.sun.security.auth.module.Krb5LoginModule required
useKeyTab=true
storeKey=true doNotPrompt=true serviceName="<service_name>" keyTab="<location of keytab file>"
client=true principal="<principal_name>;
```

Sources

Configure the **Consumer Configuration Properties** property in the source session properties to override the value specified in the **Kerberos Configuration Properties** property in a Kafka connection. Specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=
GSSAPI,
sasl.jaas.config=com.sun.security.auth.module.Krb5LoginModule required
useKeyTab=true
storeKey=true doNotPrompt=true serviceName="<service_name>" keyTab="<location of
keytab file>"
client=true principal="<principal_name>;
```

Targets

Configure the **Producer Configuration Properties** property in the target session properties to override the value specified in the **Kerberos Configuration Properties** property in a Kafka connection. Specify the value in the following format:

```
security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=
GSSAPI,
sasl.jaas.config=com.sun.security.auth.module.Krb5LoginModule required
useKeyTab=true
storeKey=true doNotPrompt=true serviceName="<service_name>" keyTab="<location of
keytab file>"
client=true principal="<principal_name>;
```

Configuring JVM Options to Access a Kerberised Kafka Cluster (Optional)

If you configure a session to read data from or write data to a Kafka topic in a Kerberised Kafka cluster, you must configure the JVM Options for the PowerCenter Client and the PowerCenter Integration Service to specify the path of the `krb5.conf` and `kafka_client_jaas.conf` file.

Configuring JVM Options for the PowerCenter Client

You can configure the `ccijvmoptions.ini` file for the PowerCenter Client.

1. Ensure that you have the `krb5.conf` and `kafka_client_jaas.conf` file for the Kerberised Kafka cluster.
2. Access the `ccijvmoptions.ini` file from the following location:
<PowerCenter Installation Directory>\clients\PowerCenterClient\client\bin
3. Add the following properties to the `ccijvmoptions.ini` file:

```
-Djava.security.auth.login.config=complete path and name of the
kafka_client_jaas.conf
-Djava.security.krb5.conf=complete path and name of the krb5.conf
```

4. Restart the PowerCenter Client.

Configuring JVM Options for the PowerCenter Integration Service

You can configure the JVM Options for the PowerCenter Integration Service to specify the path of the `krb5.conf` and `kafka_client_jaas.conf` file from the Administrator Console.

1. Ensure that you have the `krb5.conf` and `kafka_client_jaas.conf` file for the Kerberised Kafka cluster.

2. In the Administrator tool, navigate to the PowerCenter Integration Service for which you want to set the kerberised Kafka settings and then add the following JVM options in the custom properties section:

```
JVMOption1=-Djava.security.auth.login.config=<complete path and name of the  
kafka_client_jaas.conf file>  
JVMOption2=-Djava.security.krb5.conf=<complete path and name of the krb5.conf file>
```

3. Restart the PowerCenter Integration Service.

CHAPTER 3

Kafka Sources and Targets

This chapter includes the following topics:

- [Kafka Sources and Targets Overview, 16](#)
- [Import Kafka Source and Target Definitions, 16](#)
- [Sample JSON Schema, 20](#)

Kafka Sources and Targets Overview

You can create a mapping with a Kafka source to read data from a Kafka topic. You can create a mapping with any source and a Kafka target to write data to a Kafka topic.

A Kafka source or target represents data in a Kafka stream. Create a PowerExchange for Kafka source or target to read from or write to Apache Kafka brokers.

Kafka runs as a cluster comprised of one or more servers each of which is called a broker. Kafka brokers stream data in the form of messages. These messages are published to a topic.

When you import a Kafka source or target, select the topic name that you want to read from or write to. You can specify the topic name or use a regular expression for the topic name pattern only when you read from Kafka. To subscribe to multiple topics that match a pattern, you can specify a regular expression. When you run the application on the cluster, the pattern matching is done against topics before the application runs. If you add a topic with a similar pattern when the application is already running, the application will not read from the topic.

When you import a Kafka topic, you can specify the format in which the PowerExchange for Kafka reads data from or writes data to a Kafka topic. You can specify JSON as format. When you specify JSON format, you must provide a sample JSON schema file.

Import Kafka Source and Target Definitions

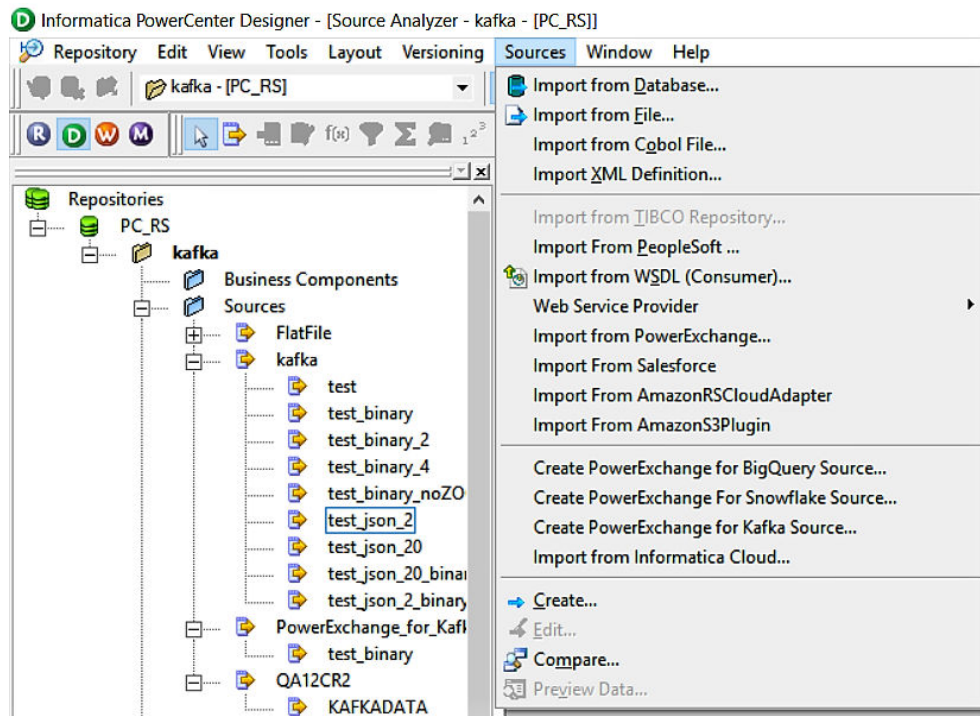
Use the **Create PowerExchange for Kafka Source** or **Create PowerExchange for Kafka Target** wizard to import Kafka source and target definitions into the PowerCenter repository.

You must import Kafka source and target objects before you create a mapping.

1. Start PowerCenter Designer, and connect to a PowerCenter repository configured with a Kafka instance.
2. Open a source or target folder.

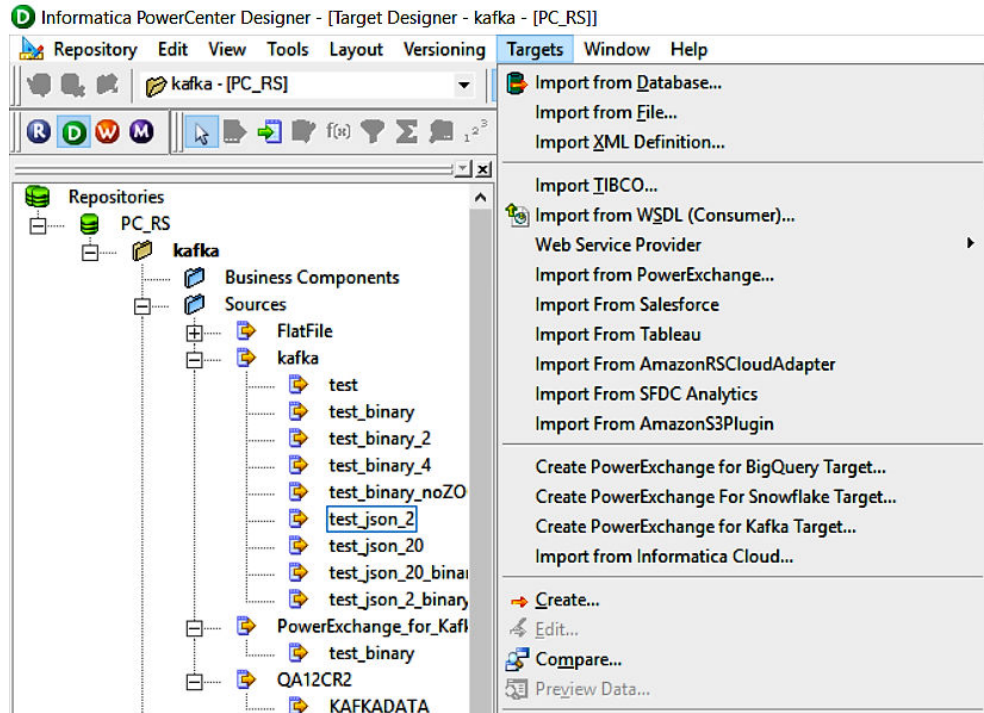
3. Select **Source Analyzer** or **Target Designer**.

- In the Source Analyzer, click **Sources > Create PowerExchange for Kafka Source** as shown in the following image:



The **Kafka Connection** wizard appears.

- In the Target Analyzer, click **Targets > Create PowerExchange for Kafka Target** as shown in the following image:



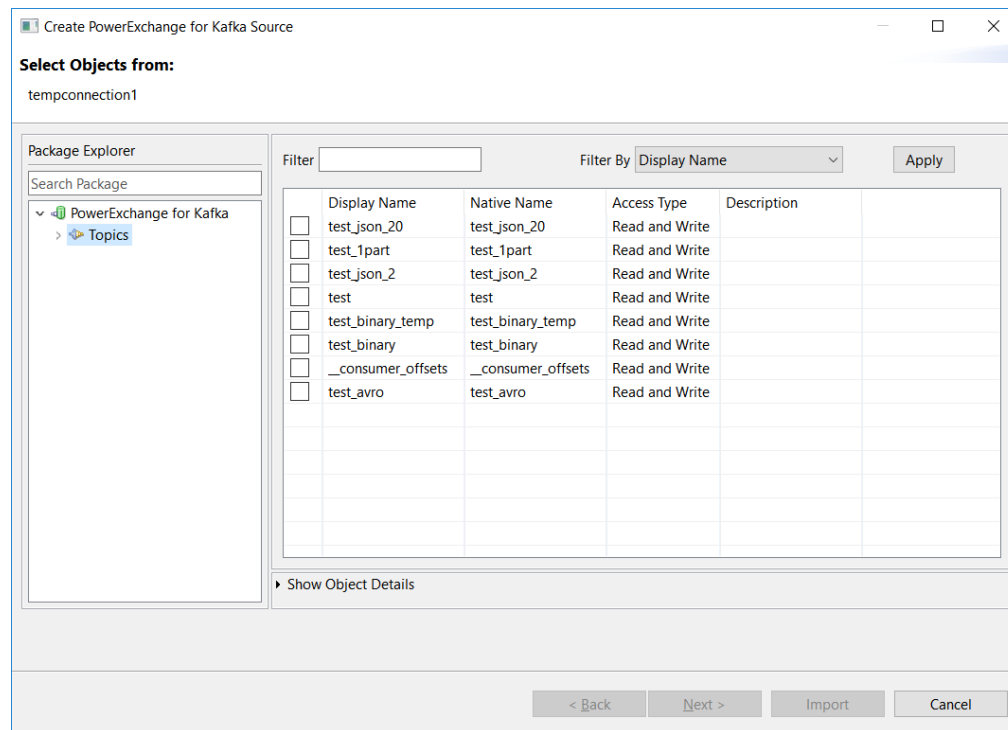
The **Kafka Connection** dialog box appears.

4. Configure the following connection parameters:

Connection Parameter	Description
Kafka Broker List	The IP address and port combinations of the Kafka messaging system broker list. The IP address and port combination has the following format: <IP Address>:<port> You can enter multiple comma-separated IP address and port combinations.
Kerberos Configuration Properties	The configuration properties required to connect to a Kerberised Kafka cluster. Specify the following property: <code>security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI</code>

5. Click **Test** to test the connection.
6. Click **Finish** to add the connection.
The **Select Objects from** tab appears.
7. Select the **Topics** package in **Package Explorer**.

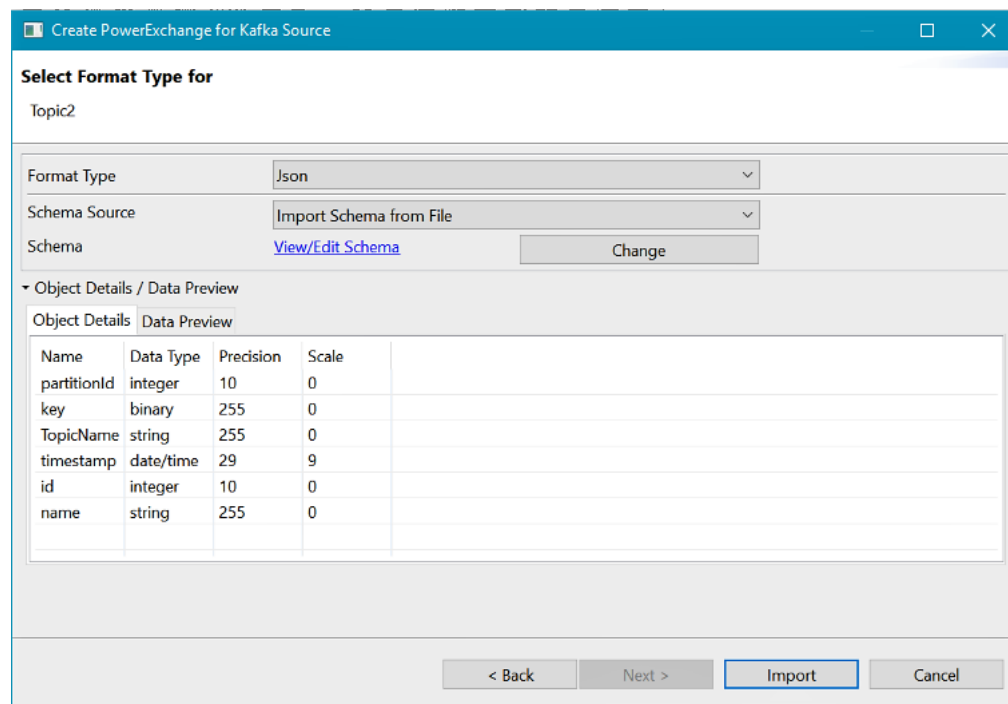
A list of topic appears as shown in the following image:



8. Select the topic that you want to import, and then click **Next**.

The **Select Format Type for** dialog box appears.

The following image shows the **Select Format Type for** dialog box:



9. In the **Format Type** list, select **Json** to import data from Kafka topics in JSON format.

10. If you select **Json** as the **Format Type**, you must configure the following format properties:

Property	Description
Schema Source	Specifies the mode to import schema for the Kafka topic. Select Import Schema from File to import schema from a schema definition file in your local machine. You cannot import a schema for a Kafka topic using the Read Data from File option.
Schema	When you select the Import Schema from File option in the Schema Source drop-down list, you must choose a schema definition file on your local machine.

11. Click **Import**.

Note: If you click **Import** in the **Select Objects from** dialog box, PowerExchange for Kafka imports the data from the Kafka topic in binary format.

To view the object details or preview data, expand the **Object Details/Data Preview** option.

12. Select the Kafka topic that you want to import, and then click **Import**.

To view the topic metadata, select the topic, and double-click the topic name.

Note: You cannot select multiple topics in a Kafka source or target definition.

Sample JSON Schema

The Kafka sources and targets in a session read and write data in JSON format. When you import a Kafka source or target definition, specify the format in which the PowerExchange for Kafka reads or writes data. When you specify JSON format, provide a sample JSON file.

The following file is a sample JSON schema:

```
{"id" : 1, "name" : "sample"}
```

CHAPTER 4

Kafka Mappings

This chapter includes the following topics:

- [Kafka Mappings Overview, 21](#)
- [Kafka Mapping Example, 21](#)

Kafka Mappings Overview

After you import a Kafka source or target definition into the PowerCenter repository, you can create a mapping to extract data from a Kafka source or load data to a Kafka target.

Note: You cannot preview data of a Kafka source or target definition.

Kafka Mapping Example

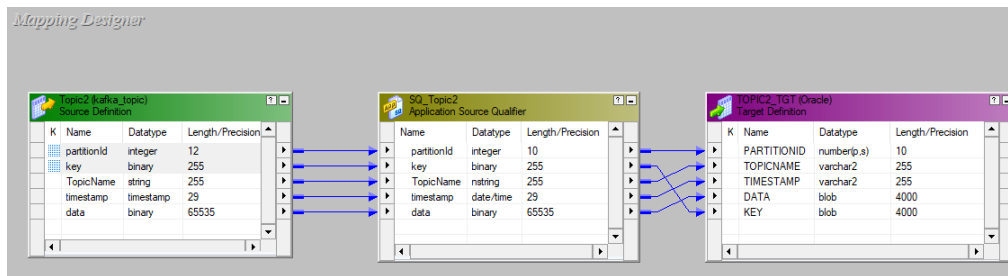
You run the IT department of a major bank that has millions of customers. You want to monitor network activity in real time. You need to collect network activity data from various sources such as firewalls or network devices to improve security and prevent attacks. The network activity data includes Denial of Service (DoS) attacks and failed login attempts made by customers. The network activity data is written to Kafka queues.

Create a mapping to read the network activity data from Kafka topics and write the data to Oracle for processing the network activity data.

The following procedure shows how to move data from the Kafka topic to Oracle:

1. Import a Kafka source.
2. Import an Oracle target.

3. Create a mapping with the Kafka source and an Oracle target.
The following image shows the example mapping:



The mapping contains the following objects:

Source Definition

The mapping source definition is a Kafka topic that contains the network activity data.

In the Source Analyzer, import the Kafka topic that you want to read. The PowerCenter Integration Service reads network activity data from the Kafka topic.

The following table describes the structure of the Kafka source definition:

Field Name	Data Type
partitionId	Integer
key	Binary
TopicName	String
timestamp	Timestamp
data	binary

Target Definition

The mapping contains an Oracle target definition. In the Target Designer, import an Oracle target definition.

The following table describes the structure of the Oracle target definition:

The following table describes the structure of the Oracle target definition:

Field Name	Data Type
PARTITIONID	Number(p,s)
TOPICNAME	varchar2
TIMESTAMP	Varchar2
DATA	Blob
KEY	Blob

Link ports between the Kafka Source Qualifier and Oracle target definition to create a flow of data. In the **Workflow Manager**, create a session and add the mapping to the session. Create a workflow and add the session to the workflow.

When you run the workflow, the data is read from the Kafka queue and written to the Oracle target. You can then run queries on the Oracle database to analyze the network activity data.

CHAPTER 5

Kafka Sessions

This chapter includes the following topics:

- [Kafka Sessions Overview, 24](#)
- [Kafka Connections Overview, 25](#)
- [Configure Kafka Source Session Properties, 25](#)
- [Configure Kafka Target Session Properties, 26](#)
- [Configure Message Recovery Strategy for a Kafka Session, 27](#)
- [Rules and Guidelines for Kafka Sessions, 28](#)

Kafka Sessions Overview

After you create mappings, you can create a session to extract and load data from Kafka brokers.

You must configure a Kafka connection in the **Workflow Manager** to extract data from or load data to Kafka topics. You can define properties in a session to determine how the PowerCenter Integration Service must extract data from or load data to a Kafka topic.

Kafka Connections Overview

The Kafka connection is a Messaging connection. Use the Kafka connection to access an Apache Kafka broker as a source or a target. You can create and manage a Kafka connection in the PowerCenter Workflow Manager or infacmd. Create a Kafka connection to read data from and write data to Apache Kafka brokers.

Kafka Connection Properties

When you configure a Kafka connection, you define the connection attributes that the PowerCenter Integration Service uses to connect to the Kafka cluster.

The following table describes the Kafka connection properties:

Property	Description
Kafka Broker List	The IP address and port combinations of the Kafka messaging system broker list. The IP address and port combination has the following format: <IP Address>:<port> You can enter multiple comma-separated IP address and port combinations.
Kerberos Configuration Properties	The configuration properties required to connect to a Kerberised Kafka cluster. Specify the following property: <code>security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI</code>

Configuring a Kafka Connection

Configure a Kafka connection in the **Workflow Manager** to define the connection attributes that the PowerCenter Integration Service uses to connect to a Kafka cluster.

1. In the **Workflow Manager**, click **Connections > Application**.
The **Application Connection Browser** dialog box appears.
2. Click **New**.
The **Select Subtype** dialog box appears.
3. Select **PowerExchange for Kafka** and click **OK**.
The **Application Connection Editor** dialog box appears.
4. Enter a name for the Kafka connection.
5. Enter the Kafka connection attributes.
6. Click **OK** to create a Kafka connection.

Configure Kafka Source Session Properties

You can configure the session properties for a Kafka source on the **Mapping** tab. Define the properties for the source instance in the session.

The following table describes the session properties that you can configure for a Kafka source session:

Property	Description
Start position offset	The time from which the Kafka source starts reading Kafka messages from a Kafka topic. You can select one of the following options: <ul style="list-style-type: none"> - Custom. Read messages from a specific time. - Earliest. Read the earliest messages available on the Kafka topic. - Latest. Read messages received by the Kafka topic after the mapping has been deployed.
Custom Start Position Timestamp	The time in GMT from which the Kafka source starts reading Kafka messages from a Kafka topic. Specify a time in the following format: dd-MM-yyyy HH:mm:ss.SSS The milliseconds are optional.
Consumer Configuration Properties	The configuration properties for the consumer. Note: If you read data from a Kafka cluster that is configured for Kerberos authentication, specify the following property to override the Kerberos Configuration Properties specified in the Kafka connection: <code>security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI</code>
Recover from last state	If you select this option, the PowerCenter Integration Service recovers a stopped, aborted, or terminated session from the last checkpoint. Default is selected.
Topic Pattern	Specify a regular expression pattern for the topic name that you want to read from. Use the regular expression syntax guidelines to specify the pattern. For more information about how to use topic patterns to read data from a Kafka source, see the Knowledge Base article 565201 .
Tracing Level	Amount of detail displayed in the session log for the transformation. You can choose Normal, Verbose Initialization, or Verbose Data. Default is Normal.

Configure Kafka Target Session Properties

You can configure the session properties for a Kafka target on the **Mapping** tab. Define the properties for the target instance in the session.

The following table describes the session properties that you can configure for a Kafka target session:

Property	Description
Metadata Fetch Timeout in milliseconds	The time after which the metadata is not fetched.
Batch Flush Time in milliseconds	The interval after which the data is published to the target.

Property	Description
Batch Flush Size in bytes	The batch size of the events after which the PowerCenter Integration Service writes data to the target.
Producer Configuration Properties	The configuration properties for the producer. Note: If you write data to a Kafka cluster that is configured for Kerberos authentication, include the following property to override the Kerberos Configuration Properties specified in the Kafka connection: <code>security.protocol=SASL_PLAINTEXT,sasl.kerberos.service.name=kafka,sasl.mechanism=GSSAPI</code>

Configure Message Recovery Strategy for a Kafka Session

When you configure message recovery for a real-time session, the PowerCenter Integration Service can recover unprocessed messages from a failed session. When you enable message recovery for a real-time session, the PowerCenter Integration Service stores source messages or message IDs in a recovery file, recovery table, recovery queue, or recovery topic. If the session fails, run the session in recovery mode to recover the messages that the PowerCenter Integration Service did not process. For sessions with Kafka topics, the PowerCenter Integration Service connects to the recovery topic as a durable subscriber so messages persist even if the session ends.

When you configure a session for recovery, you can recover the session when you recover a workflow, or you can recover the session without running the rest of the workflow.

When you configure a session, you can configure the **Recovery Strategy** property on the **Properties** page of the Session task.

You can select one of the following recovery strategies:

- **Resume from the last checkpoint.** The PowerCenter Integration Service saves the session state of operation and maintains target recovery tables. If the session aborts, stops, or terminates, the PowerCenter Integration Service uses the saved recovery information to resume the session from the point of interruption. You cannot configure a session with a resume strategy if it uses mapping variables.
- **Restart task.** The PowerCenter Integration Service runs the session again when it recovers the workflow. When you recover with restart task, you might need to remove the partially loaded data in the target or design a mapping to skip the duplicate rows.
- **Fail task and continue workflow.** When the PowerCenter Integration Service recovers a workflow, it does not recover the session. The session status becomes failed, and the PowerCenter Integration Service continues running the workflow.

Steps to Enable Message Recovery

Complete the following steps to enable message recovery for sessions:

1. In the session properties, select Resume from Last Checkpoint as the recoverystrategy.
2. Specify a recovery cache directory in the session properties at each partition point.

The PowerCenter Integration Service stores messages in the location indicated by the recovery cache directory. The default value recovery cache directory is \$PMCacheDir.

Rules and Guidelines for Kafka Sessions

Use the following rules and guidelines when you configure a Kafka session:

- You cannot configure partitioning to read data from a Kafka source.
- When you select JSON as the format type, you cannot read hierarchical data from a Kafka source or write hierarchical data to a Kafka target.
- When you use a Kafka source or target definition in a mapping, you cannot configure multiple pipelines.
- When you configure a session to write data to a Kafka target, the PowerCenter Integration Service does not display the target statistics in the session log until the session completes successfully.
- When you write data to a Kafka target and enable message recovery for the session, the PowerCenter Integration Service loads data from the first checkpoint instead of the last checkpoint during message recovery.
- When you read data from a Kafka topic or use a topic pattern and the offset of the last checkpoint is deleted during message recovery, you must provide the following property in the **Consumer Configuration Properties** to recover the messages from the next available offset:
`auto.offset.reset=earliest`. Otherwise, the mapping reads data from the latest offset available.
- When you remove the link for a port between the Kafka source definition and the Kafka Source Qualifier transformation, you must remove the port from the Kafka Source Qualifier transformation. Otherwise, the mapping fails with the following error:
`com.informatica.powercenter.sdk.SDKException`

APPENDIX A

Kafka Datatype Reference

This appendix includes the following topics:

- [Data Type Reference Overview, 29](#)
- [Kafka and Transformation Data Types, 29](#)
- [JSON Kafka Data Types and Transformation Data Types, 30](#)

Data Type Reference Overview

PowerCenter uses the following data types in Kafka mappings:

- Kafka native data types. Kafka data types appear in Kafka 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 a target, it converts the transformation data types to the comparable native data types.

Kafka and Transformation Data Types

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 a target, it converts the transformation data types to the comparable native data types.

The following table lists the Kafka data types that PowerCenter supports and the corresponding transformation data types:

Kafka Data Type	Transformation Data Type	Range and Description
BINARY	Binary	1 to 104,857,600 bytes
INTEGER	Integer	Precision 10, scale 0

Kafka Data Type	Transformation Data Type	Range and Description
STRING	Nstring	1 to 104,857,600 characters
TIMESTAMP	Date/Time	Date and time values. (precision to the nanosecond)

JSON Kafka Data Types and Transformation Data Types

JSON Kafka data types map to transformation data types that the PowerCenter Integration Service uses to move data across platforms.

The following table lists the JSON Kafka data types that the PowerCenter Integration Service supports and the corresponding transformation data types:

JSON Kafka Data Type	Transformation Data Type	Range and Description
BOOLEAN	Integer	TRUE (1) or FALSE (0)
DOUBLE	Double	Precision 15
INTEGER	Integer	-2,147,483,648 to 2,147,483,647 Precision of 10, scale of 0
LONG	Bigint	Precision of 19 digits, scale of 0
STRING	Nstring	1 to 104,857,600 characters

INDEX

C

client component installation [11](#)
configuration
overview [8](#)

F

filter [21](#)

I

Import
connection properties [16](#)
source definition [16](#)
target definition [16](#)
installation and configuration
client component [11](#)
EBF [9](#)
prerequisites [9](#)
product availability matrix [9](#)
server component [10](#)

J

JSON Kafka data types
transformation data types [30](#)

K

Kafka
connections [25](#)
mappings [21](#)
overview [7](#)
source [16](#)

Kafka (*continued*)
target [16](#)
Kafka connections
properties [25](#)
connection configuration [25](#)
Kerberos Kafka
prerequisites [12](#)

M

mapping
filter [21](#)
source [21](#)
target [21](#)

P

plug-in registration
administrator tool [11](#)
PowerExchange for Kafka
configuration [8](#)
overview [7](#)
plug-in registration [11](#)

S

server component installation
on Linux [10](#)
on Windows [10](#)
session
overview [24](#)
source properties [25](#)
target properties [26](#)
sessions
message recovery [27](#)