



Informatica® PowerExchange for Amazon S3
10.5.9

User Guide for PowerCenter

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.

Table of Contents

Preface	5
Informatica Resources.	5
Informatica Network.	5
Informatica Knowledge Base.	5
Informatica Documentation.	5
Informatica Product Availability Matrices.	6
Informatica Velocity.	6
Informatica Marketplace.	6
Informatica Global Customer Support.	6
 Chapter 1: Introduction to PowerExchange for Amazon S3.....	 7
PowerExchange for Amazon S3 Overview.	7
PowerCenter Integration Service and Amazon S3 Integration.	8
Introduction to Amazon S3.	8
Amazon S3 Objects.	8
Amazon S3 Object Format.	8
 Chapter 2: PowerExchange for Amazon S3 Configuration.....	 10
PowerExchange for Amazon S3 Configuration Overview.	10
Prerequisites.	10
IAM Authentication.	11
Create a Minimal Amazon IAM Policy.	11
Registering the Plug-in.	12
Configure the Java Heap Memory.	12
 Chapter 3: Amazon S3 Sources and Targets.....	 13
Amazon S3 Sources and Targets Overview.	13
Import Amazon S3 Objects.	13
 Chapter 4: Amazon S3 Sessions.....	 16
Amazon S3 Session Overview.	16
Amazon S3 Connections.	16
PowerExchange for Amazon S3 Connections.	16
Configuring an Amazon S3 Connection.	18
Configuring the Source Qualifier.	19
Amazon S3 Source Sessions.	19
Working with Multiple Files.	19
Source Partitioning.	20
Amazon S3 Source Sessions.	20
Amazon S3 Target Sessions.	21

Data Encryption in Amazon S3 Targets.	21
Distribution Column.	22
Object Tag.	23
Target Partitioning.	24
Amazon S3 Target Session Configuration.	25
Appendix A: Data Type Reference.	27
Data Type Reference Overview.	27
Index.	28

Preface

Use the Informatica® PowerExchange® for Amazon S3 User Guide to learn how to read from or write to Amazon S3 by using PowerCenter Client. Learn to create an Amazon S3 connection, develop mappings, and run sessions in an Informatica domain.

Informatica Resources

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

Informatica Network

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

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

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

Informatica Knowledge Base

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

To search the Knowledge Base, visit <https://search.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

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

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

Informatica Product Availability Matrices

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

Informatica Velocity

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

You can find Informatica Velocity resources at <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 extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at <https://marketplace.informatica.com>.

Informatica Global Customer Support

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

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

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

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

CHAPTER 1

Introduction to PowerExchange for Amazon S3

This chapter includes the following topics:

- [PowerExchange for Amazon S3 Overview, 7](#)
- [PowerCenter Integration Service and Amazon S3 Integration, 8](#)
- [Introduction to Amazon S3, 8](#)

PowerExchange for Amazon S3 Overview

You can use PowerExchange for Amazon S3 to connect PowerCenter and Amazon S3.

Amazon S3 is a cloud-based store that stores many objects in one or more buckets. You can read from or write to multiple Amazon S3 sources and targets. Use PowerExchange for Amazon S3 to read delimited file data from and write delimited file data to Amazon S3. You can upload or download a large object as a set of multiple independent parts.

Use PowerExchange for Amazon S3 to read delimited file data from and write delimited file data to Amazon S3. You can use Amazon S3 objects as sources and targets in mappings. When you use Amazon S3 objects in mappings, you must configure properties specific to Amazon S3. When you write to Amazon S3, you can enable data encryption to protect data.

Example

You are a medical data analyst in a medical and pharmaceutical organization who maintains patient records. A patient record can contain patient details, doctor details, treatment history, and insurance from multiple data sources.

You use PowerExchange for Amazon S3 to collate and organize the patient details from multiple input sources and write the data to Amazon S3.

PowerCenter Integration Service and Amazon S3 Integration

The PowerCenter Integration Service uses the Amazon S3 connection to connect to Amazon S3.

When you run an Amazon S3 session, the PowerCenter Integration Service reads data from Amazon S3 based on the session and Amazon S3 connection configuration. The PowerCenter Integration Service connects and reads data from Amazon S3 through a TCP/IP network. The PowerCenter Integration Service then stores data in a staging directory on the PowerCenter Integration Service machine and writes to any target.

When you run the Amazon S3 session, the PowerCenter Integration Service writes data to Amazon S3 based on the session and Amazon S3 connection configuration. The PowerCenter Integration Service reads from any source and stores data in a staging directory on the PowerCenter Integration Service machine. The PowerCenter Integration Service then connects and writes data to Amazon S3 through a TCP/IP network.

Introduction to Amazon S3

Amazon Simple Storage Service (Amazon S3) is storage service in which you can copy data from source and simultaneously move data to any target. You can use Amazon S3 to store and retrieve any amount of data at any time, from anywhere on the web. You can accomplish these tasks using the AWS Management Console web interface.

Amazon S3 stores data as objects within buckets. An object consists of a file and optionally any metadata that describes that file. To store an object in Amazon S3, you upload the file you want to store to a bucket. Buckets are the containers for objects. You can have one or more buckets. When using the AWS Management Console, you can create folders to group objects, and you can nest folders.

Amazon S3 Objects

PowerExchange for Amazon S3 sources and targets represent delimited file data objects that are read from or written to Amazon S3 buckets as CSV files.

Use PowerExchange for Amazon S3 to read delimited files from Amazon S3 and to insert data to delimited files in Amazon S3 buckets.

Amazon S3 Object Format

Amazon S3 objects are CSV or delimited text files. All fields in an Amazon S3 file are of string data type with a data format that you cannot change and with a defined precision of 256. Data in Amazon S3 text files is written in String 256 format.

PowerExchange for Amazon S3 accepts target data with a precision greater than 256. You do not need to change the precision in the Target transformation.

To read source data with a precision greater than 256, increase the precision in the Source transformation to view the complete data.

To write Amazon S3 source data to any relational target data source, you can specify field expressions in the Fields page. The PowerCenter Integration Service converts the Amazon S3 string data to the target data format.

An Amazon S3 file uses the following data format:

- The delimiter is a colon.
- The qualifier is a double-quote.
- The escape character is a backslash.

CHAPTER 2

PowerExchange for Amazon S3 Configuration

This chapter includes the following topics:

- [PowerExchange for Amazon S3 Configuration Overview, 10](#)
- [Prerequisites, 10](#)
- [IAM Authentication, 11](#)
- [Registering the Plug-in, 12](#)
- [Configure the Java Heap Memory, 12](#)

PowerExchange for Amazon S3 Configuration Overview

You can use PowerExchange for Amazon S3 on Windows or Linux. You must configure PowerExchange for Amazon S3 before you can extract data from or load data to Amazon S3.

Prerequisites

Before you can use PowerExchange for Amazon S3, perform the following tasks:

1. Install or upgrade to PowerCenter.
2. Create an access key ID and secret key in AWS. You can provide these key values when you create an Amazon S3 connection.
For more information about creating an access key ID and secret key, see the AWS documentation.
3. To encrypt data inserted in Amazon S3 target objects, enable client-side encryption.
4. Verify that you have read, write, and execute permissions on the following directory: `<Informatica installation directory>/server/bin`.

IAM Authentication

Optional. You can configure Amazon Identity and Access Management (IAM) authentication when the PowerCenter Integration Service runs on an Amazon Elastic Compute Cloud (EC2) system. Use IAM authentication for secure and controlled access to Amazon S3 resources when you run a session.

Use IAM authentication when you want to run a session on an EC2 system. Perform the following steps to configure IAM authentication:

1. Create a minimal Amazon S3 Bucket Policy. For more information, see [“Create a Minimal Amazon IAM Policy” on page 11](#).
2. Create the Amazon EC2 role. Associate the minimal Amazon IAM policy while creating the EC2 role. The Amazon EC2 role is used when you create an EC2 system in the S3 bucket. For more information about creating the Amazon EC2 role, see the AWS documentation.
3. Create an EC2 instance. Assign the Amazon EC2 role that you created in step #2 to the EC2 instance.
4. Install the PowerCenter Integration Service on the EC2 system.

Create a Minimal Amazon IAM Policy

You can configure an IAM policy through the AWS console. Use Amazon IAM authentication to securely control access to Amazon S3 resources.

If you have valid AWS credentials and you want to use IAM authentication, you do not have to specify the access key and secret key when you create an Amazon S3 connection.

You can use the following minimum required actions for users to successfully read data from and write data to Amazon S3 bucket:

- PutObject
- GetObject
- DeleteObject
- ListBucket
- GetBucketPolicy

You can use the following sample minimal Amazon IAM policy:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:PutObject",
        "s3:GetObject",
        "s3:DeleteObject",
        "s3:ListBucket",
        "s3:GetBucketPolicy"
      ],
      "Resource": [
        "arn:aws:s3:::<bucket_name>/*",
        "arn:aws:s3:::<bucket_name>"
      ]
    }
  ]
}
```

The Amazon S3 bucket policy must contain GetBucketPolicy to connect to Amazon S3.

Registering the Plug-in

After you install or upgrade to PowerCenter, you must register the plug-in for PowerExchange for Amazon S3 with the PowerCenter repository.

A plug-in is an XML file that defines the functionality of PowerExchange for Amazon S3. To register the plug-in, the repository must be running in exclusive mode. Use the Administrator tool or the pmrep RegisterPlugin command to register the plug-in.

The plug-in file for PowerExchange for Amazon S3 is `AmazonS3Plugin.xml`. When you install PowerExchange for Amazon S3, the installer copies the `AmazonS3Plugin.xml` file to the following directory: `<Informatica Installation Directory>\server\bin\Plugin`.

Note: If you do not have the correct privileges to register the plug-in, contact the user who manages the PowerCenter Repository Service.

Configure the Java Heap Memory

When the source or target contains a large amount of data, configure the memory for the Java heap size on the node that runs the PowerCenter Integration Service. You must ensure that the minimum physical memory available on the server machine for each session is in the range of 300 to 500 MB.

You must set a maximum heap size value based on the amount of data that you want to process.

1. In the Administrator tool, navigate to the PowerCenter Integration Service for which you want to change the Java heap size.
2. On the **Processes** tab, edit the **General Properties** section.
3. Specify the maximum heap size in Java SDK Maximum Memory limit based on the amount of data you want to process.

The default value of the maximum heap size is 64 MB and the minimum size is 32 MB.

4. Click **OK**.
5. Recycle the PowerCenter Integration Service.

CHAPTER 3

Amazon S3 Sources and Targets

This chapter includes the following topics:

- [Amazon S3 Sources and Targets Overview, 13](#)
- [Import Amazon S3 Objects, 13](#)

Amazon S3 Sources and Targets Overview

Create a mapping with an Amazon S3 source to read delimited file data from Amazon S3 and write to a target.

Create a mapping with any source and an Amazon S3 target to write delimited file data to Amazon S3. When you write data to an Amazon S3 file, if there is a single or double quote in the source data, an extra quote is added to the target.

When you create a mapping to read data from an Amazon S3 file, if the backslash (\) is present in the source data, the backslash is skipped in the target. Backslash is the default escape character in the formatting options. You must specify a different escape character.

If you specify a delimiter in the **Others** option in **File Formatting** options, the PowerCenter Integration Service might display an error or the fields or metadata are not fetched as expected. You must follow the below guidelines when you use the **Others** option for delimiter in File Formatting Options:

- You must use a single special character excluding comma as the delimiter such as, #, @, &, or ~.
- You must not use multibyte characters as the delimiter.
- You must not use text qualifiers such as, ' and " as the delimiter.

Import Amazon S3 Objects

You can import Amazon S3 source and target objects before you create a mapping.

Ensure that you have valid AWS credentials before you create a connection. You can also use IAM authentication if you install the PowerCenter client on an EC2 system.

1. Start the PowerCenter Designer and connect to a PowerCenter repository.
2. Open a source or target folder.
3. Select **Source Analyzer** or **Target Designer**.

4. Click **Sources** or **Targets**, and then click **Import from AmazonS3Plugin**.
The **Establish Connection** dialog box appears.
5. Specify the following information and click **Connect**.

Connection Property	Description
Access Key	The access key ID used to access the Amazon account resources. Required if you do not use AWS Identity and Access Management (IAM) authentication. Note: Ensure that you have valid AWS credentials before you create a connection.
Secret Key	The secret access key used to access the Amazon account resources. This value is associated with the access key and uniquely identifies the account. You must specify this value if you specify the access key ID. Required if you do not use AWS Identity and Access Management (IAM) authentication.
Folder Path	The complete path to the Amazon S3 objects and must include the bucket name and any folder name. Ensure that you do not use a forward slash at the end of the folder path. For example, <bucket name>/<my folder name>
Master Symmetric Key	Optional. Provide a 256-bit AES encryption key in the Base64 format when you enable client-side encryption. You can generate a key using a third-party tool. If you specify a value, ensure that you specify the Encryption Type as Client Side Encryption in the target session properties.
Customer Master Key ID	Optional. Specify the customer master key ID or alias name generated by AWS Key Management Service (AWS KMS). You must generate the customer master key for the same region where Amazon S3 bucket reside. You can specify any of the following values: Customer Generated Customer Master Key Enables client-side or server-side encryption. Default Customer Master Key Enables client-side or server-side encryption. Only the administrator user of the account can use the default customer master key ID to enable client-side encryption.
Code Page	The code page compatible with the Amazon S3 source. Select one of the following code pages: <ul style="list-style-type: none"> - MS Windows Latin 1. Select for ISO 8859-1 Western European data. - UTF-8. Select for Unicode and non-Unicode data. - Shift-JIS. Select for double-byte character data. - ISO 8859-15 Latin 9 (Western European). - ISO 8859-2 Eastern European. - ISO 8859-3 Southeast European. - ISO 8859-5 Cyrillic. - ISO 8859-9 Latin 5 (Turkish). - IBM EBCDIC International Latin-1.

Connection Property	Description
Region Name	<p>The name of the region where the Amazon S3 bucket is available. Select one of the following regions:</p> <ul style="list-style-type: none"> - Asia Pacific (Mumbai) - Asia Pacific (Seoul) - Asia Pacific (Singapore) - Asia Pacific (Sydney) - Asia Pacific (Tokyo) - AWS GovCloud - Canada (Central) - China (Beijing) - EU (Ireland) - EU (Frankfurt) - South America (Sao Paulo) - US East (Ohio) - US East (N. Virginia) - US West (N. California) - US West (Oregon) <p>Default is US East (N. Virginia).</p>
Formatting Options	<p>Select a delimiter, text qualifier, or an escape character.</p> <p>If you want to specify a delimiter other than comma, tab, colon, and semi-colon, you must choose Other in File Formatting options.</p> <p>When you specify a delimiter in the Other option, the PowerCenter Integration Service might display an error message or the fields or metadata are not fetched as expected. You must follow the below guidelines when you use the Other option for delimiter in File Formatting Options:</p> <ul style="list-style-type: none"> - You must use a single special character excluding comma as the delimiter such as, #, @, &, or ~. - You must not use multibyte characters as the delimiter. - You must not use text qualifiers such as, ' and " as the delimiter. <p>Note: You cannot use this property when you configure an Amazon S3 connection in the Workflow Manager.</p>

- Click **Connect**.
- Click **Next**.
- Select the Amazon S3 object that you want to import.
- Optionally, click **Data Preview** to view the resource metadata.
- Click **Finish**.

When you configure an object in a mapping or in the Amazon S3 object connection at run time, the folder path you specify while you import Amazon S3 objects from PowerCenter Designer cannot be overwritten.

CHAPTER 4

Amazon S3 Sessions

This chapter includes the following topics:

- [Amazon S3 Session Overview, 16](#)
- [Amazon S3 Connections, 16](#)
- [Amazon S3 Source Sessions, 19](#)
- [Amazon S3 Target Sessions, 21](#)

Amazon S3 Session Overview

You can configure an Amazon S3 connection in the Workflow Manager to read delimited file data from or write delimited file data to an Amazon S3. Ensure that you have write access to the Amazon S3 bucket you want to access.

When you write to Amazon S3 targets, you can only insert data to Amazon S3 targets. You cannot update or delete data. Any data in the target is overwritten when you select an existing Amazon S3 target. To protect data, you can also enable data encryption before writing data to Amazon S3 targets.

Note: When you run a PowerExchange for Amazon S3 session, you cannot edit the connection properties within the session.

Amazon S3 Connections

Amazon S3 connections enable you to read data from or write data to Amazon S3. The PowerCenter Integration Service uses the connection when you run an Amazon S3 session.

PowerExchange for Amazon S3 Connections

When you configure an Amazon S3 connection, you define the connection attributes that the PowerCenter Integration Service uses to connect to Amazon S3.

The following table describes the Amazon S3 connection properties:

Property	Description
Name	The name of the Amazon S3 connection.
Type	The Amazon S3 connection type.
Access Key	The access key ID used to access the Amazon account resources. Required if you do not use AWS Identity and Access Management (IAM) authentication. Note: Ensure that you have valid AWS credentials before you create a connection.
Secret Key	The secret access key used to access the Amazon account resources. This value is associated with the access key and uniquely identifies the account. You must specify this value if you specify the access key ID. Required if you do not use AWS Identity and Access Management (IAM) authentication.
Folder Path	The complete path to the Amazon S3 objects and must include the bucket name and any folder name. Ensure that you do not use a forward slash at the end of the folder path. For example, <bucket name>/<my folder name> Note: To read data from the folder path, set the <code>-DS3UseConnectionFolderPath=true</code> in the JVM option for the PowerCenter Integration Service. Ignore the mentioned JVM option when you write data to Amazon S3.
Master Symmetric Key	Optional. Provide a 256-bit AES encryption key in the Base64 format when you enable client-side encryption. You can generate a key using a third-party tool. If you specify a value, ensure that you specify the Encryption Type as Client Side Encryption in the target session properties.
Customer Master Key ID	Optional. Specify the customer master key ID or alias name generated by AWS Key Management Service (AWS KMS). You must generate the customer master key for the same region where Amazon S3 bucket reside. You can specify any of the following values: Customer Generated Customer Master Key Enables client-side or server-side encryption. Default Customer Master Key Enables client-side or server-side encryption. Only the administrator user of the account can use the default customer master key ID to enable client-side encryption.
Code Page	The code page compatible with the Amazon S3 source. Select one of the following code pages: <ul style="list-style-type: none"> - MS Windows Latin 1. Select for ISO 8859-1 Western European data. - UTF-8. Select for Unicode and non-Unicode data. - Shift-JIS. Select for double-byte character data. - ISO 8859-15 Latin 9 (Western European). - ISO 8859-2 Eastern European. - ISO 8859-3 Southeast European. - ISO 8859-5 Cyrillic. - ISO 8859-9 Latin 5 (Turkish). - IBM EBCDIC International Latin-1.

Property	Description
Region Name	<p>The name of the region where the Amazon S3 bucket is available. Select one of the following regions:</p> <ul style="list-style-type: none"> - Asia Pacific (Mumbai) - Asia Pacific (Seoul) - Asia Pacific (Singapore) - Asia Pacific (Sydney) - Asia Pacific (Tokyo) - AWS GovCloud - Canada (Central) - China (Beijing) - EU (Ireland) - EU (Frankfurt) - South America (Sao Paulo) - US East (Ohio) - US East (N. Virginia) - US West (N. California) - US West (Oregon) <p>Default is US East (N. Virginia).</p>
Formatting Options	<p>Select a delimiter, text qualifier, or an escape character.</p> <p>If you want to specify a delimiter other than comma, tab, colon, and semi-colon, you must choose Other in File Formatting options.</p> <p>When you specify a delimiter in the Other option, the PowerCenter Integration Service might display an error message or the fields or metadata are not fetched as expected. You must follow the below guidelines when you use the Other option for delimiter in File Formatting Options:</p> <ul style="list-style-type: none"> - You must use a single special character excluding comma as the delimiter such as, #, @, &, or ~. - You must not use multibyte characters as the delimiter. - You must not use text qualifiers such as, ' and " as the delimiter. <p>Note: You cannot use this property when you configure an Amazon S3 connection in the Workflow Manager.</p>

Configuring an Amazon S3 Connection

Configure an Amazon S3 connection in the Workflow Manager to define the connection attributes that the PowerCenter Integration Services uses to connect to Amazon S3.

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 **AmazonS3** and click **OK**.
The **Connection Object Definition** dialog box appears.
4. Enter a name for the Amazon S3 connection.
5. Enter the Amazon S3 connection properties.
6. Click **OK**.

Configuring the Source Qualifier

After you import a source to create a Mapping for Amazon S3 source, you must configure the source qualifier.

1. In a mapping, double-click the Source Qualifier.
2. Select the **Configure** tab and click **Configure**.
The Establish Connection dialog box appears.
3. Specify the Amazon S3 connection properties and click **Connect**.
4. Click **Finish**.
5. Save the mapping.

Amazon S3 Source Sessions

Create a session and associate it with a mapping that you created to read data from Amazon S3. Define the session properties to read data from Amazon S3.

You can create a mapping with an Amazon S3 source to read delimited file data from Amazon S3 and write to a target.

When you create a mapping to read data from an Amazon S3 file, if the backslash (\) is present in the source data, the backslash is skipped in the target. Backslash is the default escape character in the formatting options. You must specify a different escape character.

Working with Multiple Files

You can read from multiple Amazon S3 sources and write to a single target.

To read multiple files, all files must be available in the same Amazon S3 bucket. When you want to read from multiple sources in the Amazon S3 bucket, you must create a `.manifest` file that contains all the source files with the respective absolute path or directory path. You must specify the `.manifest` file name in the following format: `<file_name>.manifest`

For example, the `.manifest` file contains source files in the following format:

```
{
  "fileLocations": [{
    "URIs": [
      "dirl/dir2/file_1.csv",
      "dirl/dir2/dir47file_2.csv",
      "dirA/dirB/file_3.csv",
      "dirA/dirB/file_4.csv" ] }, {
    "URIPrefixes": [
      "dirl/dir2/",
      "dirl/dir2/"] } ],
  "settings": {
    "stopOnFail": "true"
  }
}
```

You can configure the `stopOnFail` property to display error messages while reading multiple files. Set the value to `true`, if you want the PowerCenter Integration Service to display error messages if the read operation fails for any of the source files. If you set the value to `false`, the error messages appear only in the session log. The PowerCenter Integration Service skips the file that generated the error and continues to read other files.

The **Data Preview** tab displays the data of the first file available in the URI specified in the `.manifest`. If the URI section is empty, the first file in the folder specified in `URIPrefixes` is displayed.

You can specify an asterisk (*) wildcard in the file name to fetch files from the Amazon S3 bucket. You can specify the asterisk (*) wildcard to fetch all the files or only the files that match the name pattern. Specify the wildcard character in the following format:

```
abc*.txt
abc.*
```

For example, if you specify `result*.txt`, all the file names starting with the term `result` and ending with the `.txt` file extension are read. If you specify `result.*`, all the file names starting with the term `result` are read regardless of the extension.

Use the wildcard character to specify files from a single folder. For example,

```
{
  "fileLocations": [{
    "URIs": [
      "dir1/dir2/file_1.csv",
      "dir1/dir2/dir4/file_2.csv",
    ]
  }, {
    "URIPrefixes": [
      "dir1/dir2/",
      "dir1/dir2/"
    ]
  }
],
  { "WildcardURIs": [ "multiread_wildcard/file_1/*.csv" ] }
}
"settings": {
  "stopOnFail": "true"
}
```

You cannot use the wildcard characters to specify folder names. For example,

```
{ "WildcardURIs": [ "multiread_wildcard/dir1*/", "multiread_wildcard/*/ " ] }
```

Note: PowerExchange for Amazon S3 supports only asterisk (*) wildcard character.

Source Partitioning

You can configure partitioning to optimize the mapping performance at run time when you read data from Amazon S3 sources.

The partition type controls how the PowerCenter Integration Service distributes data among partitions at partition points. You can define the partition type as `passthrough` partitioning. With partitioning, the PowerCenter Integration Service distributes rows of source data based on the number of threads that you define as partition.

The PowerCenter Integration Service enables the partition according to the size of the Amazon S3 source file. The file name is appended with a number starting from 0 in the following format: `<file name>_<number>`

Note: If you enable partitioning and the precision for the source column is less than the maximum data length in that column, you might receive unexpected results. To avoid unexpected results, the precision for the source column must be equal to or greater than the maximum data length in that column for partitioning to work as expected.

Amazon S3 Source Sessions

Create a mapping with an Amazon S3 source and a target to read data from Amazon S3. Define the properties for each session instance in the session.

The following table describes the source session properties:

Source Property	Description
Enable Downloading S3 File in Multiple Parts	If the file size of an Amazon S3 object is greater than 8 MB, you can enable the Enable Downloading S3 Files in Multiple Parts option to download the object in multiple parts in parallel.
Staging File Location	<p>Amazon S3 staging directory.</p> <p>When you specify the directory path, the PowerCenter Integration Service create folders depending on the number of partitions that you specify in the following format:</p> <pre>Infas3Staging<<00/11><processID>_<partition number>>></pre> <p>where, 00 represents read operation and 11 represents write operation. For example,</p> <pre>Infas3Staging<< 115348_0>></pre> <p>Note: The temporary files are created within the new directory.</p> <p>When you specify a directory name, if a folder with same name already exists, the PowerCenter Integration Service deletes the contents of the folder. You must have the write permission for the specified location.</p> <p>If you do not specify a directory path, the PowerCenter Integration Service uses a temporary directory as the staging file location.</p>
Part Size	<p>Specifies the part size of an object.</p> <p>Default is 5 MB.</p>
Header Line Number	<p>Specify the line number that you want to use as the header when you read data from Amazon S3.</p> <p>Default is 0.</p>
Read Data From Line	<p>Specify the line number from where you want the PowerCenter Integration Service to read data.</p> <p>Default is 1.</p> <p>Note: To read data from the header, the value of the Header Line Number and the Read Data From Line fields should be the same.</p>
Tracing Level	<p>Sets the amount of detail that appears in the log file.</p> <p>You can choose terse, normal, verbose initialization, or verbose data. Default is normal.</p>

Amazon S3 Target Sessions

Create a session and associate it with a mapping that you created to write data to Amazon S3. Define the session properties to write data to Amazon S3.

You can create a mapping with any source and an Amazon S3 target to write delimited file data to Amazon S3. When you write data to an Amazon S3 file, if there is a single or double quote in the source data, an extra quote is added to the target.

When you write to Amazon S3 targets, you can only insert data to Amazon S3 targets. You cannot update or delete data. Any data in the target is overwritten when you select an existing Amazon S3 target. To protect data, you can also enable data encryption before writing data to Amazon S3 targets.

Data Encryption in Amazon S3 Targets

To protect data, you can enable server-side encryption or client-side encryption to encrypt data inserted in Amazon S3 buckets.

You can encrypt data by using the master symmetric key or customer master key. Do not use the master symmetric key and customer master key together. Customer master key is a user managed key generated by AWS Key Management Service (AWS KMS) to encrypt data.

Master symmetric key is a 256-bit AES encryption key in the Base64 format that is used to enable client-side encryption. You can generate master symmetric key by using a third-party tool.

Note: You cannot read KMS encrypted data when you use the IAM role with an EC2 system that has a valid KMS encryption key and a valid Amazon S3 bucket policy.

Server-side Encryption

Enable server-side encryption if you want to use Amazon S3-managed encryption key or AWS KMS-managed customer master key to encrypt the data while uploading the CSV files to the buckets. To enable server-side encryption, select **Server Side Encryption** as the encryption type in the target session properties.

Client-side Encryption

Enable client-side encryption if you want the PowerCenter Integration Service to encrypt the data while uploading the CSV files to the buckets. To enable client-side encryption, perform the following tasks:

1. Provide a master symmetric key or customer master key ID when you create an Amazon S3 connection. Ensure that you provide a 256-bit AES encryption key in Base64 format.

Note: The administrator user of the account can use the default customer master key ID to enable the client-side encryption.

2. Select **Client Side Encryption** as the encryption type in the target session properties.
3. Ensure that an organization administrator updates the security policy .jar files on the machines that host the PowerCenter Integration Service. Update the `local_policy.jar` and the `US_export_policy.jar` files in the following directory: `<Informatica installation directory>\java\jre\lib\security`.
From the Oracle website, download the .jar files that are supported by the JAVA environment on the machine that hosts the PowerCenter Integration Service.

Distribution Column

You can write multiple files to Amazon S3 target from a single source. Configure the **Distribution Column** option in the target session properties.

You can specify one column name in the **Distribution Column** field to create multiple target files during run time. When you specify the column name, the PowerCenter Integration Service creates multiple target files in the column based on the column values that you specify in **Distribution Column**.

Each target file name is appended with the **Distribution Column** value in the following format:

```
<Target_fileName>+_<Distribution column value>+<file extension>
```

Each target file contains all the columns of the table including the column that you specify in the **Distribution Column** field.

For example, the name of the target file is `Region.csv` that contains the values North America and South America. The following target files are created based on the values in the **Distribution Column** field:

- `Region_North America.csv`
- `Region_South America.csv`

You cannot specify two column names in the **Distribution Column** field. If you specify a column name that is not present in target field column, the task fails.

When you specify a column that contains value with special characters in the **Distribution Column** field, the PowerCenter Integration Service fails to create target file if the corresponding Operating System do not support the special characters.

For example, the PowerCenter Integration Service fails to create target file if the column contains date value in the following format: YYYY/MM/DD

Object Tag

You can add a tag to the object stored on the Amazon S3 bucket. Each tag contains a key value pair.

Tagging an object helps to categorize the storage. You can add the object tags in the **Object Tags** field under the target session properties. Enter the object tag in the `Key=Value` format.

You can also enter multiple object tags. To add multiple object tags, enter the object tags in the following format inside a file that is present on the machine on which the PowerCenter Integration Service runs:

```
Key1=Value1
Key2=Value2
```

You can either enter the key value pairs or the specify the file path that contains the key value pairs. For example, you can specify the file path in the `C:\object\tags.txt` format. You can specify any file path on which the PowerCenter Integration Service is installed.

When you upload new objects in the Amazon S3 bucket, you can add tags to the new objects or add tags to the existing objects. If the PowerCenter Integration Service overrides a file that contains a tag in the Amazon S3 bucket, the tag is not retained. You must add a new tag for the overridden file. If you upload multiple files to the Amazon S3 bucket, each file that you upload must have the same set of tags associated with the multiple objects.

To add tags in the Amazon S3 target object, you must add the `s3:PutObjectTagging` permission in the Amazon S3 policy. Following is the sample policy:

```
{
  "Version": "2012-10-17",
  "Id": "Policy1500966932533",
  "Statement": [
    {
      "Sid": "Stmt1500966903029",
      "Effect": "Allow",
      "Action": [
        "s3:DeleteObject",
        "s3:GetBucketPolicy",
        "s3:GetObject",
        "s3:ListBucket",
        "s3:PutObject",
        "s3:PutObjectTagging"
      ],
      "Resource": [
        "arn:aws:s3:::<bucket_name>/*",
        "arn:aws:s3:::<bucket_name>"
      ]
    }
  ]
}
```

The following table lists the special characters that PowerExchange for Amazon S3 supports during entering the key value pair:

Special Characters	Support
+	Yes
-	Yes
=	No
.	Yes
_	Yes
:	Yes
/	Yes

Rules and Guidelines for Tagging an Object

Use the following rules and guidelines for tagging an object:

- You can add maximum 10 tags for each object.
- When you enter a tag for an object, the tag must contain a unique tag key.
- The tag key can contain maximum 128 unicode characters in length and tag values can contain maximum 256 unicode characters in length.
- The key and values are case sensitive.

Target Partitioning

You can configure partitioning to optimize the mapping performance at run time when you write data to Amazon S3 targets.

The partition type controls how the PowerCenter Integration Service distributes data among partitions at partition points. You can define the partition type as passthrough partitioning. With partitioning, the PowerCenter Integration Service distributes rows of target data based on the number of threads that you define as partition.

You can configure the **Merge Partition Files** options in the target session properties. You can specify whether the PowerCenter Integration Service must merge the number of partition files as a single file or maintain separate files based on the number of partitions specified to write data to the Amazon S3 targets.

If you do not select the **Merge Partition Files** option, separate files are created based on the number of partitions specified. The file name is appended with a number starting from 0 in the following format: <filename>_<number>

For example, the number of threads for the `Region.csv` file is three. If you do not select the **Merge Partition Files** option, the PowerCenter Integration Service writes three separate files in the Amazon S3 target in the following format:

- <Region_0>
- <Region_1>

- <Region_2>

If you configure the **Merge Partition Files** option, the PowerCenter Integration Service merges all the partitioned files as a single file and writes the file to Amazon S3 target.

Amazon S3 Target Session Configuration

You can configure a session to write data to Amazon S3. Define the properties for each target instance in the session.

The following table describes the target session properties:

Property	Description
Encryption Type	<p>Method you can use to encrypt data. Select one of the following values:</p> <p>None</p> <p>The data is not encrypted.</p> <p>Client Side Encryption</p> <p>The PowerCenter Integration Service encrypts data while uploading the CSV files to Amazon buckets. You must select Client Side Encryption if you specify a master symmetric key or customer master key ID in the Amazon S3 connection properties.</p> <p>Server Side Encryption</p> <p>Amazon S3 encrypts data while uploading the CSV files to Amazon buckets. If you select the Server Side Encryption in the target session properties and do not specify the customer master key ID in the connection properties, Amazon S3-managed encryption keys are used to encrypt data.</p>
Folder Path	<p>The complete path to the Amazon S3 objects and must include the bucket name and any folder name. Ensure that you do not use a forward slash at the end of the folder path.</p> <p>For example, <bucket name>/<my folder name></p> <p>The folder path specified at run-time overrides the path specified while creating a connection.</p>
Compression Type	<p>Compress the data in GZIP format when you write the data to Amazon S3. The target file in Amazon S3 will have .gz extension. The PowerCenter Integration Service compresses the data and then sends the data to Amazon S3 bucket.</p> <p>Default is None.</p>
Part Size	<p>Specifies the part size of an object.</p> <p>Default is 5 MB.</p>
TransferManager Thread Pool Size	<p>Specifies the number of the threads to write data in parallel.</p> <p>PowerExchange for Amazon S3 uses the AWS TransferManager API to upload a large object in multiple parts to Amazon S3.</p> <p>When the file size is more than 5 MB, you can configure multipart upload to upload object in multiple parts in parallel. Default is 10.</p>
Merge Partition Files	<p>Specifies whether the PowerCenter Integration Service must merge all the partition files into a single file or maintain separate files based on the number of partitions specified to write data to the Amazon S3 targets.</p> <p>Default is not selected.</p>

Property	Description
Distribution Column	Specify the name of the column that is used to create multiple target files during run time.
Staging File Location	<p>Amazon S3 staging directory.</p> <p>When you specify the directory path, the PowerCenter Integration Service create folders depending on the number of partitions that you specify in the following format: <code>Infas3Staging<<00/11><processID>_<partition number>>></code> where, 00 represents read operation and 11 represents write operation. For example, <code>Infas3Staging<< 115348_0>></code></p> <p>Note: The temporary files are created within the new directory.</p> <p>When you specify a directory name, if a folder with same name already exists, the PowerCenter Integration Service deletes the contents of the folder. You must have the write permission for the specified location.</p> <p>If you do not specify a directory path, the PowerCenter Integration Service uses a temporary directory as the staging file location.</p>
Object Tags	<p>You can add a single or multiple tags to the objects stored on the Amazon S3 bucket.</p> <p>You can either enter the key value pairs or specify the file path that contains the key value pairs.</p>
INSERT	<p>Inserts the source data in the Amazon S3 target. Overwrites any existing data in the target object.</p> <p>Note: You can only insert data to Amazon S3 objects. You cannot perform delete or update operations on Amazon S3 targets.</p>

Note: PowerExchange for Amazon S3 does not support the **Success File Directory** and **Error File Directory** properties.

APPENDIX A

Data Type Reference

This appendix includes the following topic:

- [Data Type Reference Overview, 27](#)

Data Type Reference Overview

PowerExchange for Amazon S3 uses only delimited files in PowerCenter sessions.

PowerExchange for Amazon S3 uses the following data types in PowerCenter sessions with Amazon S3 objects:

Amazon S3 native data types

Amazon S3 data types appear on the Datatype tab for source qualifiers and target definitions when you edit metadata for the fields.

Transformation data types

Set of data types that appear in the remaining transformations. They are internal data types based on ANSI SQL-92 generic data types, which PowerCenter uses to move data across platforms.

Transformation data types appear in all remaining transformations in a PowerCenter sessions.

When PowerExchange for Amazon S3 reads source data, it converts the native data types to the comparable transformation data types before transforming the data. When PowerExchange for Amazon S3 writes to a target, it converts the transformation data types to the comparable native data types.

The following table lists the Amazon S3 data types that PowerExchange for Amazon S3 supports and the corresponding transformation data types:

Amazon S3 Native Data Type	Transformation Data Type	Description
String	String	1 to 104,857,600 characters

INDEX

A

administration

 IAM authentication [11](#)

 minimal Amazon S3 bucket policy [11](#)

Amazon S3

 source sessions [20](#)

 connections [16](#)

 integration with PowerCenter [8](#)

 object format [8](#)

 objects [8](#)

 overview [8](#)

 source session [19](#)

 target sessions [21](#)

Amazon S3 connection

 configuring [18](#)

 properties [16](#)

Amazon S3 objects

 importing [13](#)

Amazon S3 session

 overview [16](#)

Amazon S3 source

 session properties [20](#)

Amazon S3 sources and targets

 overview [13](#)

Amazon S3 target

 session configuration [25](#)

D

data encryption

 client-side [22](#)

 server-side [22](#)

data type reference

 overview [27](#)

distribution column [22](#)

J

Java heap memory [12](#)

O

object tag

 rules and guidelines [24](#)

object tags [23](#), [24](#)

P

partitioning [24](#)

PowerExchange for Amazon S3

 overview [7](#)

prerequisites

 Amazon S3 [10](#)

S

source partitioning [20](#)

source qualifier

 configuring [19](#)

W

Working with Multiple Files [19](#)