



Informatica™

Informatica® Cloud Data Integration

Amazon Athena Connector

© Copyright Informatica LLC 2020, 2021

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

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

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

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

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

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

NOTICES

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

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

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

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

Publication Date: 2021-08-06

Table of Contents

Preface	4
Informatica Resources.	4
Informatica Documentation.	4
Informatica Intelligent Cloud Services web site.	4
Informatica Intelligent Cloud Services Communities.	4
Informatica Intelligent Cloud Services Marketplace.	4
Data Integration connector documentation.	5
Informatica Knowledge Base.	5
Informatica Intelligent Cloud Services Trust Center.	5
Informatica Global Customer Support.	5
Chapter 1: Introduction to Amazon Athena Connector	6
Introduction to Amazon Athena.	6
Amazon Athena Connector overview.	6
Administration of Amazon Athena Connector.	7
Create minimal Amazon S3 bucket policy.	7
Create AWS Glue data catalog policy.	7
Create Amazon Athena policy.	8
Chapter 2: Amazon Athena connections	9
Amazon Athena connections overview.	9
Amazon Athena connection properties.	9
Chapter 3: Amazon Athena sources	11
Amazon Athena source features.	11
Data encryption in Amazon Athena sources.	12
Chapter 4: Mappings and mapping tasks with Amazon Athena Connector	13
Amazon Athena sources in mappings.	13
Amazon Athena sources in mapping tasks.	15
Chapter 5: Data type reference	16
Data type reference overview.	16
Amazon Athena and transformation data types.	16
Index	18

Preface

Use *Amazon Athena Connector* to learn how to read from Amazon Athena by using Cloud Data Integration. Learn to create an Amazon Athena connection and develop and run mapping tasks and mappings in Cloud Data Integration.

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 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 Intelligent Cloud Services web site

You can access the Informatica Intelligent Cloud Services web site at <http://www.informatica.com/cloud>. This site contains information about Informatica Cloud integration services.

Informatica Intelligent Cloud Services Communities

Use the Informatica Intelligent Cloud Services Community to discuss and resolve technical issues. You can also find technical tips, documentation updates, and answers to frequently asked questions.

Access the Informatica Intelligent Cloud Services Community at:

<https://network.informatica.com/community/informatica-network/products/cloud-integration>

Developers can learn more and share tips at the Cloud Developer community:

<https://network.informatica.com/community/informatica-network/products/cloud-integration/cloud-developers>

Informatica Intelligent Cloud Services Marketplace

Visit the Informatica Marketplace to try and buy Data Integration Connectors, templates, and mapplets:

<https://marketplace.informatica.com/>

Data Integration connector documentation

You can access documentation for Data Integration Connectors at the Documentation Portal. To explore the Documentation Portal, visit <https://docs.informatica.com>.

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 Intelligent Cloud Services Trust Center

The Informatica Intelligent Cloud Services Trust Center provides information about Informatica security policies and real-time system availability.

You can access the trust center at <https://www.informatica.com/trust-center.html>.

Subscribe to the Informatica Intelligent Cloud Services Trust Center to receive upgrade, maintenance, and incident notifications. The [Informatica Intelligent Cloud Services Status](#) page displays the production status of all the Informatica cloud products. All maintenance updates are posted to this page, and during an outage, it will have the most current information. To ensure you are notified of updates and outages, you can subscribe to receive updates for a single component or all Informatica Intelligent Cloud Services components. Subscribing to all components is the best way to be certain you never miss an update.

To subscribe, go to <https://status.informatica.com/> and click **SUBSCRIBE TO UPDATES**. You can then choose to receive notifications sent as emails, SMS text messages, webhooks, RSS feeds, or any combination of the four.

Informatica Global Customer Support

You can contact a Customer Support Center by telephone or online.

For online support, click **Submit Support Request** in Informatica Intelligent Cloud Services. You can also use Online Support to log a case. Online Support requires a login. You can request a login at <https://network.informatica.com/welcome>.

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site at <https://www.informatica.com/services-and-training/support-services/contact-us.html>.

CHAPTER 1

Introduction to Amazon Athena Connector

This chapter includes the following topics:

- [Introduction to Amazon Athena, 6](#)
- [Amazon Athena Connector overview, 6](#)
- [Administration of Amazon Athena Connector, 7](#)

Introduction to Amazon Athena

Amazon Athena is a query service that you can use to analyze data in Amazon Simple Storage Service (Amazon S3) by using the standard SQL. Amazon Athena executes multiple queries simultaneously, and fetches the result quickly even with large datasets and complex queries.

Amazon Athena helps you analyze unstructured, semi-structured, and structured data stored in Amazon S3.

Amazon Athena stores the schema in a data catalog or AWS Glue data catalog and uses it when you run queries. Amazon Athena uses Apache Hive to create tables and databases.

Amazon Athena Connector overview

You can use Amazon Athena Connector to read data from Amazon Athena by using Cloud Data Integration. Use Amazon Athena Connector to read flat files and Parquet files from Amazon S3 using Amazon Athena.

You can use Amazon Athena Connector to read data from views and external tables in the Athena data catalog and AWS Glue data catalog. Use Amazon Athena Connector to read and query Amazon S3 files that are mapped to the Amazon S3 location in an external table.

You can run queries in Amazon Athena on encrypted data in Amazon S3. You must run the Amazon Athena queries in the region where Amazon S3 is hosted. You can also encrypt the Amazon Athena query result stored on Amazon S3.

You can create an Amazon Athena connection and use the connection in mappings and mapping tasks.

Create a mapping task to process data based on the data flow logic defined in a mapping or integration template.

Administration of Amazon Athena Connector

You can use Amazon Athena Connector after the organization administrator performs the following tasks:

- Manages authentication by creating an access key and a secret key.
Provide the values for access key and secret key when you configure the Amazon Athena connection.
- Creates an AWS Key Management Service (AWS KMS)-managed customer master key if you want to enable server-side encryption or client-side encryption.
- Creates the minimal Amazon S3 bucket policy, AWS Glue data catalog policy, and Amazon Athena policy for Amazon Athena Connector.

Create minimal Amazon S3 bucket policy

The minimal Amazon S3 bucket policy restricts user operations and user access to specific Amazon S3 buckets by assigning an AWS Identity and Access Management (IAM) policy to users. You can configure the AWS IAM policy through the AWS console.

You can use the following minimum required permissions to successfully store the Amazon Athena query result on Amazon S3:

- PutObject
- GetObject
- DeleteObject
- ListBucket

Sample Policy:

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

Create AWS Glue data catalog policy

You can use AWS IAM to define policies and roles that are needed to access resources used by AWS Glue.

You can use the following sample policy for AWS Glue data catalog:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "glue:*"
      ],

```

```

        "Resource": [
            "*"
        ]
    }
]
}

```

Create Amazon Athena policy

Specify the minimum required permissions for Amazon Athena Connector to read data from views and external tables in the AWS Glue data catalog and to read and query Amazon S3 files.

You can use the following minimum required permissions:

- GetWorkGroup
- GetTableMetadata
- StartQueryExecution
- GetQueryResultsStream
- ListDatabases
- GetQueryExecution
- GetQueryResults
- GetDatabase
- ListTableMetadata
- GetDataCatalog

Sample Policy:

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "athena:GetWorkGroup",
        "athena:GetTableMetadata",
        "athena:StartQueryExecution",
        "athena:GetQueryResultsStream",
        "athena:ListDatabases",
        "athena:GetQueryExecution",
        "athena:GetQueryResults",
        "athena:GetDatabase",
        "athena:ListTableMetadata",
        "athena:GetDataCatalog"
      ],
      "Resource": [
        "arn:aws:athena:*:*:workgroup/*",
        "arn:aws:athena:*:*:datacatalog/*"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "athena:ListDataCatalogs",
        "athena:ListWorkGroups"
      ],
      "Resource": "*"
    }
  ]
}

```


CHAPTER 2

Amazon Athena connections

This chapter includes the following topics:

- [Amazon Athena connections overview, 9](#)
- [Amazon Athena connection properties, 9](#)

Amazon Athena connections overview

You can use an Amazon Athena connection to read data from Amazon Athena. You can use the connections to specify sources in mappings and mapping tasks.

Create the connection on the **Connections** page and use it to read data from Amazon Athena tables and views, preview data, and run mappings. Define the source properties to read data from Amazon Athena.

Amazon Athena connection properties

When you set up an Amazon Athena connection, you must configure the connection properties.

The following table describes the Amazon Athena connection properties:

Connection property	Description
Runtime Environment	Name of the runtime environment where you want to run the tasks. Specify a Secure Agent or Hosted Agent.
Authentication Type	The authentication mechanism to connect to Amazon Athena. Select Permanent IAM Credentials .
Access Key	Optional. The access key to connect to Amazon Athena.
Secret Key	Optional. The secret key to connect to Amazon Athena.

Connection property	Description
JDBC URL	<p>The URL of the Amazon Athena connection.</p> <p>Enter the JDBC URL in the following format:</p> <pre>jdbc:awsathena://AwsRegion=<region_name>;S3OutputLocation=<S3_Output_Location>;</pre> <p>You can use pagination to fetch the Amazon Athena query results. Set the property <code>UseResultsetStreaming=0</code> to use pagination.</p> <p>Enter the property in the following format:</p> <pre>jdbc:awsathena:// AwsRegion=<region_name>;S3OutputLocation=<S3_Output_Location>;UseResultsetStreaming=0;</pre> <p>You can also use streaming to improve the performance and fetch the Amazon Athena query results faster. When you use streaming, ensure that port 444 is open.</p> <p>By default, streaming is enabled.</p>
Customer Master Key ID	<p>Optional. Specify the customer master key ID generated by AWS Key Management Service (AWS KMS) or the Amazon Resource Name (ARN) of your custom key for cross-account access.</p> <p>You must generate the customer master key ID for the same region where your Amazon S3 bucket resides. You can either specify the customer-generated customer master key ID or the default customer master key ID.</p>

CHAPTER 3

Amazon Athena sources

This chapter includes the following topics:

- [Amazon Athena source features, 11](#)
- [Data encryption in Amazon Athena sources, 12](#)

Amazon Athena source features

You can use an Amazon Athena object as a source in a mapping. You can use tables and views as Amazon Athena sources.

When you configure the advanced source properties, configure properties specific to Amazon Athena.

The following table lists the Amazon Athena source features that you can use in mappings:

Feature	Applicable in Mappings
Server-side encryption	Yes
Server-side encryption with KMS	Yes
Client-side encryption with KMS	Yes
Filter	Yes
Sort	Yes
Data Preview	Yes
Table name override	Yes
Schema name override	Yes

Data encryption in Amazon Athena sources

You can run queries in Amazon Athena on encrypted data in Amazon S3. You must run the Amazon Athena queries in the region where Amazon S3 is hosted. You can also encrypt the query results in Amazon S3.

Select the type of the encryption in the **Encryption Type** field under the Amazon Athena advanced source properties.

Use the customer master key ID generated by AWS Key Management Service for server-side encryption.

You can configure the following types of encryption:

None

The data is not encrypted.

SSE-S3

If you select the **SSE-S3** encryption type, Amazon Athena encrypts the file using the Amazon S3-managed key for server-side encryption.

SSE-KMS

If you select the **SSE-KMS** encryption type, Amazon Athena encrypts the file using the AWS KMS-managed key or Amazon Resource Name (ARN) for server-side encryption.

CSE-KMS

If you select the **CSE-KMS** encryption type, Amazon Athena encrypts the file using the AWS KMS-managed key for client-side encryption.

CHAPTER 4

Mappings and mapping tasks with Amazon Athena Connector

This chapter includes the following topics:

- [Amazon Athena sources in mappings, 13](#)
- [Amazon Athena sources in mapping tasks, 15](#)

Amazon Athena sources in mappings

In a mapping, you can configure a Source transformation to represent an Amazon Athena source.

The following table describes the Amazon Athena source properties that you can configure in a Source transformation:

Property	Description
Connection	Name of the source connection. Select a source connection, or click New Parameter to define a parameter for the source connection.
Source type	Type of the source object. Select any of the following source object types: <ul style="list-style-type: none">- Single Object. Select to specify a single Amazon Athena object.- Parameter. Select to specify a parameter name.
Object	Name of the source object. Select a single source object.

Note: When you preview a column in a table that uses binary data type, the column displays blank values.

The following table describes the query options that you can configure in a Source transformation:

Property	Description
Filter	<p>Filter value in a read operation. Click Configure to add conditions to filter records and reduce the number of rows that the Secure Agent reads from the source.</p> <p>You can specify the following filter conditions:</p> <ul style="list-style-type: none"> - Not parameterized. Use a basic filter to specify the object, field, operator, and value to read specific records. - Completely parameterized. Use a parameter to represent the filter conditions. - Advanced. Use an advanced filter to define a complex filter condition.
Sort	<p>Conditions to sort records.</p> <p>You can specify the following sort conditions:</p> <ul style="list-style-type: none"> - Not parameterized. Select the fields and type of sorting to use. - Parameterized. Use a parameter to specify the sort condition. - Sort Order. Specify whether you want to sort data in ascending or descending order.

Note: When you configure an advanced filter on a Timestamp column, you must specify the value as `TIMESTAMP` in the filter condition. For example,

```
SELECT * FROM <dbname>.<tablename> where col_date < TIMESTAMP '2030-06-22 18:30:00.000';
```

The following table describes the Amazon Athena advanced source properties that you can configure in a Source transformation:

Property	Description
Retain Athena Query Result On S3 File	<p>Specifies whether you want to retain the Amazon Athena query result on the Amazon S3 file. Select the check box to retain the Amazon Athena query result on the Amazon S3 file.</p> <p>The Amazon Athena query result is stored in the CSV file format.</p> <p>By default, the Retain Athena Query Result on S3 File check box is not selected.</p>
S3OutputLocation	<p>Specifies the location of the Amazon S3 file that stores the result of the Amazon Athena query.</p> <p>You can also specify the Amazon S3 file location in the <code>S3OutputLocation</code> parameter in the JDBC URL connection property.</p> <p>If you specify the Amazon S3 output location in both the connection and the advanced source properties, the Secure Agent uses the Amazon S3 output location specified in the advanced source properties.</p>
Fetch Size	<p>Determines the number of rows to read in one result set from Amazon Athena.</p> <p>Default is 10000.</p>
Encryption Type	<p>Encrypts the data in the Amazon S3 staging directory.</p> <p>You can select the following encryption types:</p> <ul style="list-style-type: none"> - None - SSE-S3 - SSE-KMS - CSE-KMS <p>Default is None.</p>

Property	Description
Schema Name	Overrides the schema name of the source object.
Source Table Name	Overrides the table name used in the metadata import with the table name that you specify.

You can set the tracing level in the Amazon Athena advanced source properties to determine the amount of details that logs contain.

The following table describes the tracing levels that you can configure:

Tracing Level	Description
Terse	The Secure Agent logs initialization information, error messages, and notification of rejected data.
Normal	The Secure Agent logs initialization and status information, errors encountered, and skipped rows due to transformation row errors. Summarizes session results, but not at the level of individual rows.
Verbose Initialization	In addition to normal tracing, the Secure Agent logs additional initialization details, names of index and data files used, and detailed transformation statistics.
Verbose Data	In addition to verbose initialization tracing, the Secure Agent logs each row that passes into the mapping. Also notes where the Secure Agent truncates string data to fit the precision of a column and provides detailed transformation statistics. When you configure the tracing level to verbose data, the Secure Agent writes row data for all rows in a block when it processes a transformation.

Amazon Athena sources in mapping tasks

For Amazon Athena source connections used in template-based mapping tasks, you can configure advanced properties in the **Sources** page.

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

You can configure all the advanced source properties, except **Tracing Levels**, for a mapping task that you configured for a mapping. For more information, see [“Amazon Athena sources in mappings” on page 13](#).

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

CHAPTER 5

Data type reference

This chapter includes the following topics:

- [Data type reference overview, 16](#)
- [Amazon Athena and transformation data types, 16](#)

Data type reference overview

Data Integration uses the following data types in mappings and mapping tasks with Amazon Athena:

Amazon Athena native data types

Amazon Athena data types appear in the **Fields** tab of the Source and Target transformations when you edit metadata for the fields.

Transformation data types

Set of data types that appear in the transformations. They are internal data types based on ANSI SQL-92 generic data types, which the Secure Agent uses to move data across platforms. Transformation data types appear in all transformations in a mapping.

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

Amazon Athena and transformation data types

The following table lists the Amazon Athena native data types that Data Integration supports and the corresponding transformation data types:

Amazon Athena Data Type	Transformation Data Type	Description
Bigint	Bigint	Signed eight-byte integer
Binary	Binary	1 to 104,857,600 bytes
Boolean	Integer	Logical Boolean (true/false)

Amazon Athena Data Type	Transformation Data Type	Description
Char	String	Fixed-length character string
Date	Date/Time	Calendar date (year, month, day)
Decimal	Decimal	Exact numeric of selectable precision
Double	Double	Precision 15
Float	Double	Precision 15
Int	Integer	Signed four-byte integer
Smallint	Integer	Signed two-byte integer
String	String	-1 to 104,857,600 characters
Timestamp	Date/Time	Date and time (without time zone)
Tinyint	Integer	Signed one-byte integer
Varchar	String	Variable-length character string with a user-defined limit

Note: You can use the Binary data type only with Parquet files.

INDEX

A

Amazon Athena
 connection properties [9](#)
Amazon Athena Connector
 overview [6](#)
Amazon Athena sources
 mappings [13](#)

C

Cloud Application Integration community
 URL [4](#)
Cloud Developer community
 URL [4](#)
connections
 Amazon Athena [9](#)

D

Data Integration community
 URL [4](#)
data type reference
 overview [16](#)

I

Informatica Global Customer Support
 contact information [5](#)

Informatica Intelligent Cloud Services
 web site [4](#)

M

maintenance outages [5](#)

S

status
 Informatica Intelligent Cloud Services [5](#)
system status [5](#)

T

trust site
 description [5](#)

U

upgrade notifications [5](#)

W

web site [4](#)