Simba Athena JDBC Driver with SQL Connector
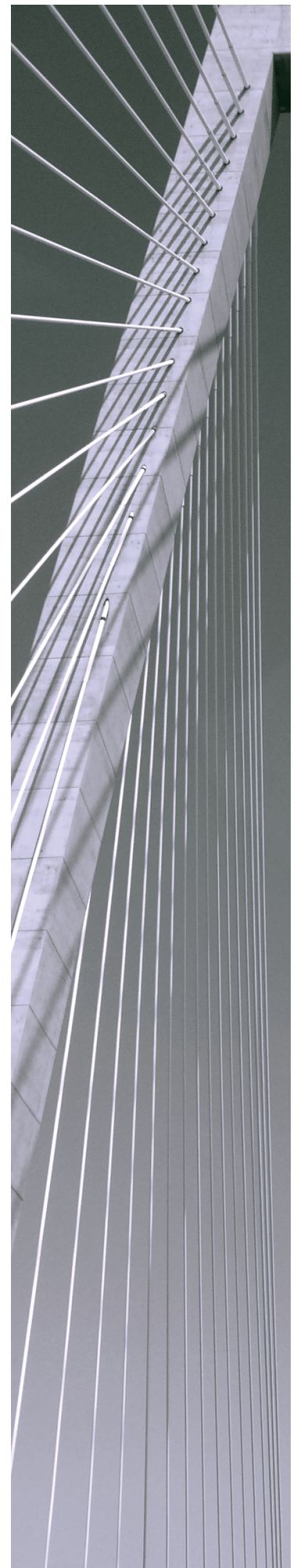
Migration Guide

Simba Technologies Inc.

Version 2.0.14 (Preview)

July 22, 2020

Contact Us

Simba Technologies Inc.
938 West 8th Avenue
Vancouver, BC Canada
V5Z 1E5

Tel: +1 (604) 633-0008

Fax: +1 (604) 633-0004

www.simba.com

About This Guide

Purpose

This document lists differences between the drivers that may disrupt workflows when you migrate from the 1.x driver to the 2.x driver, and provides recommendations on how to recreate those workflows for a successful migration.

Audience

The guide is intended for end users of the Simba Athena JDBC Driver.

Document Conventions

Italics are used when referring to book and document titles.

Bold is used in procedures for graphical user interface elements that a user clicks and text that a user types.

`Monospace font` indicates commands, source code or contents of text files.

> ✎ **Note:**
>
> A text box with a pencil icon indicates a short note appended to a paragraph.

> ❗**Important:**
>
> A text box with an exclamation mark indicates an important comment related to the preceding paragraph.

Table of Contents

Upgrading From 1.x to 2.0.x

JDBC Driver Class Name

The drivers use different class names.

Version 1.x	Version 2.x
`com.amazonaws.athena.jdbc.Athena Driver`	`com.simba.athena.jdbc.Dr iver`

If you are using the following line in your code to explicitly load the driver class in your source code:
`Class.forName("com.amazonaws.athena.jdbc.AthenaDriver");`,
then you will need to change it to:
` Class.forName("com.simba.athena.jdbc.Driver");`

Connection URL

Specifying the Host and Port

The 2.x version provides an alternative way to specify the AWS region.

1.x Version	2.x Version

`jdbc:awsathena://athena.{REGION}.amazonaws.com:443`
 	
Where `{REGION}` is a region identifier, such as `us-west-2`
	`jdbc:awsathena://athena.{REGION}.amazonaws.com:443`
	Or
	`jdbc:awsathena://AwsRegion={REGION}`
	
	Where `{REGION}` is a region identifier, such as `us-west-2`. If `{REGION}` is specified using both endpoint URL and `AwsRegion`, the value specified in `AwsRegion` takes precedence.

Changes are not required in this case, but be aware the 2.x version provides an alternative way to specify the AWS region in the connection URL.

Connection String Attributes Separator

The drivers use different attribute separators in their connection URLs.

1.x Version	2.x Version
`&` and `?`	`;`

The following is an example connection URL using the 1.x version syntax:
```
jdbc:awsathena://athena.us-west-1.amazonaws.com:443?s3_
staging_dir=s3://query-resultsbucket/folder/&query_results_
encryption_option=SSE_S3
```

The following shows the equivalent URL constructed using the 2.x version syntax:
```
jdbc:awsathena://athena.us-west-1.amazonaws.com:443;s3_
staging_dir=s3://query-resultsbucket/folder/;query_results_
encryption_option=SSE_S3
```

Driver Configuration Options

There are some differences in the supported connection properties for the drivers.

Version 1.x Option	Version 2.x Option	Possible Values	
`log_path`	`LogPath`	No difference.	
`log_level`	`LogLevel`	**1.x**	**2.x**
		OFF	0
		FATAL	1
		ERROR	2
		WARNING	3
		INFO	4
		DEBUG	5
		TRACE	6
`retry_base_delay`	Not configurable.		
`retry_max_backoff_time`	Not configurable.		

The following is an example connection URL for enabling logging using the syntax for version 1.x:

```
jdbc:awsathena://athena.us-west-1.amazonaws.com:443?s3_
staging_dir=s3://query-resultsbucket/folder/&log_
level=TRACE&log_path=/tmp
```

The following is the equivalent connection URL using the syntax for version 2.x:

```
jdbc:awsathena://athena.us-west-1.amazonaws.com:443;s3_
staging_dir=s3://query-
resultsbucket/folder/;LogLevel=6;LogPath=/tmp
```

ResultSetMetaData Differences for API Calls

The drivers return different matadata for the following API calls.

getCatalogs

Column Name	Version 1.x		Version 2.x	
TABLE_CAT	**Metadata**	**Value**	**Metadata**	**Value**
	Type Name	varchar	Type Name	VARCHAR
	Type ID	-16	Type ID	12
	Display Size	1073741824	Display Size	128
	Precision	1073741824	Precision	128
	Scale	0	Scale	0

getColumns

Column Name	Version 1.x		Version 2.x	

`TABLE_CAT`
`TABLE_SCHEM`
`TABLE_NAME`
`COLUMN_NAME`
`TYPE_NAME`
`IS_AUTOINCREMENT`
`IS_`
`GENERATEDCOLUMN`
	Metadata	**Value**	**Metadata**	**Value**

	Type Name	varchar	Type Name	VARCHAR
	Type ID	-16	Type ID	12
	Display Size	1073741824	Display Size	128
	Precision	1073741824	Precision	128
	Scale	0	Scale	0

`REMARKS`
`COLUMN_DEF`
`IS_NULLABLE`
`SCOPE_CATALOG`
`SCOPE_SCHEMA`
`SCOPE_TABLE`
	Metadata	**Value**	**Metadata**	**Value**

	Type Name	varchar	Type Name	VARCHAR
	Type ID	-16	Type ID	12
	Display Size	1073741824	Display Size	254
	Precision	1073741824	Precision	254
	Scale	0	Scale	0
`SOURCE_DATA_TYPE`	**Metadata**	**Value**	**Metadata**	**Value**
	Type Name	smallint	Type Name	INGEGER
	Type ID	5	Type ID	4
	Display Size	6	Display Size	11
	Precision	5	Precision	10
	Scale	0	Scale	0

getSchemas

Column Name	Version 1.x		Version 2.x	
`TABLE_SCHEM` `TABLE_CATALOG`	**Metadata**	**Value**	**Metadata**	**Value**
	Type Name	varchar	Type Name	VARCHAR
	Type ID	-16	Type ID	12
	Display Size	1073741824	Display Size	128
	Precision	1073741824	Precision	128
	Scale	0	Scale	0

getTables

Column Name	Version 1.x		Version 2.x	

`TABLE_CAT`
`TABLE_SCHEM`
`TABLE_NAME`
`TABLE_TYPE`
`TYPE_CAT`
`TYPE_SCHEM`
`TYPE_NAME`
`SELF_REFERENCING_`
`COL_NAME`
`REF_GENERATION`
	Metadata	**Value**	**Metadata**	**Value**

	Type Name	varchar	Type Name	VARCHAR
	Type ID	-16	Type ID	12
	Display Size	1073741824	Display Size	128
	Precision	1073741824	Precision	128
	Scale	0	Scale	0

Column Name	Version 1.x		Version 2.x	
REMARKS	**Metadata**	**Value**	**Metadata**	**Value**
	Type Name	varchar	Type Name	VARCHAR
	Type ID	-16	Type ID	12
	Display Size	1073741824	Display Size	254
	Precision	1073741824	Precision	254
	Scale	0	Scale	0

Data Type for TIME Literal in Query Result

For a query such as `SELECT TIME '12:00:00'`, the drivers use different data types in the query result set for the TIME literal column.

Version 1.x	Version 2.x
TIME	VARCHAR

Upgrading from 2.0.2 to 2.0.5 or later

Result Set Streaming Support

Starting with version 2.0.5, the driver uses the result set streaming API to improve the performance in fetching query results. To take advantage of this feature, you must do the following:

- Include and allow the `athena:GetQueryResultsStream` action in your IAM policy statement. For details on managing Athena IAM policies, see https://docs.aws.amazon.com/athena/latest/ug/access.html.
- If you are connecting to Athena through a proxy server, make sure that the proxy server does not block port 444. The result set streaming API uses port 444 on the Athena server for outbound communications.