

Release Notes

Amazon JDBC Driver for Apache Hive 1.0.6

The release notes provide details of enhancements, features, and known issues in Amazon JDBC Driver for Apache Hive 1.0.6, as well as the version history.

Enhancements & New Features

The following are highlights of the new features and functionalities that have been added to Amazon JDBC Driver for Apache Hive 1.0.6.

AsyncExecPollInterval Configuration option

You can now configure how often the driver polls the server for query execution status.

Support added for specifying how the driver obtains the Kerberos Subject

When using Kerberos authentication, you can now configure the `KrbAuthType` connection property to specify how the driver obtains the Kerberos Subject. For more information, see the [Installation and Configuration Guide](#).

Resolved Issues

The following issues have been resolved in Amazon JDBC Driver for Apache Hive 1.0.6.

In some configurations users would receive a `NullPointerException` system message when attempting to connect to the server.

The driver sometimes returns fewer rows than expected if the JVM is running low on memory, even though no errors are reported.

In some cases, when the heap size is restricted, the driver retrieves fewer rows than expected but does not return any exceptions or errors.

The driver now retrieves the correct number of rows when the heap size is restricted.

Returns fewer rows than expected with restricted heap size

The driver fetches fewer rows than expected without producing exceptions or system messages when heap size was restricted.

The driver throws a null pointer exception upon calling `Driver.getPropertyInfo()`.

The buffer length and consequently char octet length defaults to 0 for CHAR and VARCHAR types.

Query translation drops brackets around OR clauses.

Version History

Version 1.0.5

Enhancements & New Features

The following are highlights of the new features and functionalities that have been added to Amazon JDBC Driver for Apache Hive 1.0.5.

Updated handling of socket timeouts

The driver now uses 30 as the default value for the `SocketTimeout` property. If a socket read operation takes longer than 30 seconds to complete, then the driver closes the connection. Also, error messages for socket timeouts have been improved for clarity.

Support added for connecting through the Apache Knox Gateway

You can now use the new `transportMode` and `httpPath` connection properties to configure the driver to connect through the Knox Gateway. The default UID and PWD values in the driver have been updated for improved compatibility with these new connection properties, and the default `AuthMech` value now depends on the `transportMode` setting. For more information, see the [Simba JDBC Driver for Apache Hive Installation and Configuration Guide](#).

Support added for configuring host name verification for SSL connections

You can now use the `AllowAllHostNames` connection property to specify whether host name verification is enabled for SSL connections.

Resolved Issues

The following issues have been resolved in Amazon JDBC Driver for Apache Hive 1.0.5.

Returns fewer rows than expected with restricted heap size

[13981] The driver fetches fewer rows than expected without producing exceptions or system messages when heap size was restricted.

The driver throws a null pointer exception upon calling

`Driver.getPropertyInfo()`.

The buffer length and consequently char octet length defaults to 0 for CHAR and VARCHAR types.

Query translation drops brackets around OR clauses.

Driver returns error when casting to TIMESTAMP and DATE in SQL statement

Inserting a false Boolean value as 0 inserts TRUE

When the `setQueryTimeout()` method is called and the query processing time exceeds the query timeout value, the driver returns a socket timeout error.

The driver now returns the correct error for query timeouts (`SqlTimeoutException`).

`DatabaseMetaData.getColumns()` does not return complex type fields through the JDBC driver.

The driver now partially supports Array, Struct, Map, and Union SQL types. These types are mapped to Varchar.

Hive Server 1 and Hive Server 2 drivers cannot run at the same time in the same JVM

This issue has been resolved. Before, the `HS1Driver` and `HS2Driver` could not coexist in the same JVM; loading both drivers at the same time caused a Hive Server 2 connection failure. You can now run both drivers at the same time in the same JVM.

Version 1.0.4

Enhancements & New Features

Support added for write-back when connected to Hive 0.14 or later

You can now execute INSERT, UPDATE, and DELETE statements when connected to Apache Hive 0.14 or later.

Support added for dynamic service discovery with Apache ZooKeeper

You can now connect to Hive servers that are registered against a ZooKeeper service by connecting to the ZooKeeper service.

SSL now configured separately from authentication

Before, you would enable SSL in the driver by setting the authentication mechanism to "User Name and Password with Secure Sockets Layer" (AuthMech=4) or "No Authentication with Secure Sockets Layer" (AuthMech=5). You can now use the new SSL property to enable or disable SSL connections, and use the AuthMech property solely to configure authentication.

Support added for case-insensitive treatment of catalog names, schema names, table names, and column names in catalog function calls

The driver is now able to work with catalogs, schemas, tables, and columns as expected regardless of whether the names are spelled with upper- or lower-case characters.

Optimized metadata retrieval

The driver is now designed to push catalog function restrictions down to the server for processing when possible, which improves driver performance during metadata retrieval.

Support added for working with the REMARKS column in Hive Server 1 instances

When connected to a Hive Server 1 instance that is running Hive 0.10 or later, you can now query and write to the REMARKS column.

Resolved Issues

Driver does not report an error when a MapReduce job fails on the server side during query execution

This issue has been resolved. The driver now reports an error if a MapReduce job fails.

The `getColumns()` function returns an empty result set if it uses both schema and table restrictions, and the table restriction contains an escaped wildcard character

This issue has been resolved. The `getColumns()` function now returns the correct result set.

Query translation does not use the database context specified by a `USE <database>` query

This issue has been resolved. Now, when you change the database context by executing a `USE <database>` query, any subsequent query translation will use that database context.

Driver does not support the `LENGTH` scalar function for non-character columns

This issue has been resolved. You can now call the `LENGTH` function on any data type that can be converted to `STRING`.

Driver cannot execute `"SET <key>=<value>"` statements using `executeUpdate()`

This issue has been resolved.

Driver cannot execute `SET` statements with leading or trailing spaces

This problem has been resolved. Before, the driver returned an error if a `SET` statement started or ended with spaces. The driver is now able to parse the statement and execute it successfully.

TCP connections do not close when `connection.close()` is called

This issue has been resolved. Before, TCP connections did not close when `connection.close()` was called, and only closed when the client application closed. Now, TCP connections close immediately after `connection.close()` is called.

When handling multiple connections at the same time, in some cases the driver will retrieve the wrong data

This issue has been resolved. Driver support for multithreading has been fixed and the driver will now return the correct data when there are multiple connections.

Driver returns duplicate data for the getColumns JDBC API when the table restriction contains a wildcard

This issue has been resolved.

In some cases, the driver fails to retrieve columns metadata from tables that contain partitioned columns

This issue has been resolved.

Driver sends a separate DESC <table> query for each DECIMAL, VARCHAR, and CHAR column in the queried table

This issue has been resolved. Now, the driver sends only one DESC <table> query for the queried table.

The RowsFetchedPerBlock setting limits the amount of schemas, tables, and columns metadata retrieved from Hive Server 2 instances

This issue has been resolved. The RowsFetchedPerBlock setting now works as expected, and only limits the maximum number of rows retrieved per fetch call.

In some cases, the getTables catalog function returns table types incorrectly

This issue has been resolved.

The REMARKS table metadata column does not contain the correct data

This issue has been resolved. Previously, the REMARKS table metadata column was hard-coded, so the data in the column could not be changed. You can now work with the column as expected.

In some cases, driver performance is slow when using catalog functions to retrieve metadata

This issue has been resolved. Previously, the way that the driver retrieved metadata involved retrieving and processing more data than was necessary. Now, the behavior in the driver is optimized and performance has improved significantly.

In CDH 4.x or earlier, in some cases nested calls cause the driver to return a "Read a negative frame size" error

This issue has been resolved.

Driver does not use the default User Name value as expected

This issue has been resolved. Before, if the driver is configured to connect using the User Name authentication mechanism but the user name is not specified, the driver returns an error. Now, the driver uses the default value "anonymous" if the user name is not specified.

Version 1.0.3

Resolved Issues

The following issues were resolved in Amazon JDBC Driver for Apache Hive 1.0.2.

Using Java services to load the driver classes may cause a Hive Server 2 connection failure

When working in Java 6 or later, you have the option of loading the required driver classes using Java services instead of using `Class.forName()`. Java services sometimes loads HS1Driver before loading HS2Driver, which prevents HS2Driver from operating correctly. As a short-term workaround for this issue, this release of the driver prevents Java services from performing the loading.

Version 1.0.2

Enhancements & New Features

The following are highlights of the new features and functionalities that were added to Amazon JDBC Driver for Apache Hive 1.0.2.

JDBC 4.1 now supported

The driver now supports JDBC 4.1. To use the driver with JDBC 4.1, use the `Amazon_HiveJDBC41_version` package, where *version* is the driver version number.

Support added for Kerberos authentication through AccessControlContext

The driver now provides an alternative method for obtaining Kerberos tickets. Instead of obtaining a ticket-granting ticket (TGT) from the ticket cache, the driver can now check whether there are any Subjects associated with the AccessControlContext and whether those Subjects have a TGT. If so, the driver can use the TGT from the Subject.

Support added for Kerberos authentication on IBM Java 1.6

You can now use Kerberos authentication when running IBM Java 1.6.

Support added for the `Connection.isValid()` and `Connection.getClientInfo()` methods for JDBC 4 and JDBC 4.1

The driver now supports the `Connection.isValid()` and `Connection.getClientInfo()` methods.

Support added for connection pooling

The driver now supports the JDBC ConnectionPoolDataSource interface via the following classes:

- `com.amazon.hive.jdbc3.DataSource`
- `com.amazon.hive.jdbc4.DataSource`
- `com.amazon.hive.jdbc41.DataSource`

SSLTrustStore and SSLTrustStorePwd configuration parameters implemented

The SSLTrustStore and SSLTrustStorePwd parameters are optional parameters that you can use in the connection string. Use these parameters to configure the driver to use a specific TrustStore when connecting through SSL. If these parameters are not set, then the driver uses the default TrustStore located in `jre\lib\security\cacerts`

Support for Hive 0.13.0 data types added to the Hive Server 1 client

Support for the CHAR(n) and DECIMAL(precision, scale) data types have been added to the Hive Server 1 client. The client now fully supports all the data types in Hive 0.13.0.

Support added for PreparedStatements with parameters

Previously, the driver did not support the use of PreparedStatements with parameters because the Thrift API did not support this functionality. You can now use parameters in PreparedStatements. However, the workaround for supporting this feature is not a typical method for doing PreparedStatements. The driver replaces the question mark (?) in the query with the actual parameter value and executes the query during the execution stage. It is not recommended that you use `prepareStatement.getParameterMetadata` or `prepareStatement.getResultSetMetadata` before `prepareStatement.executeQuery`, because the PreparedStatement does not return the expected values.

Direct integration with Kerberos Key Distribution Center added

You can now configure the driver to get a ticket from your Key Distribution Center directly. To do this, in the JVM environment, configure a JAAS configuration file that directs the driver to use a generated keytab file as the credentials.

Resolved Issues

The following issues were resolved in Amazon JDBC Driver for Apache Hive 1.0.2.

In some cases, debug logging causes connections to fail

Fixed an issue within the debug logging infrastructure where connections would fail under multithreaded conditions.

Server errors are not always exposed to the user

Fixed an issue where some errors that were delivered from the server to the driver were not getting passed on to the client application.

Connection properties fail to load if the connection string contains a space

This issue has been resolved. The driver can now remove spaces and ensure that the connection properties load successfully.

Column metadata fails to be retrieved from Hive 0.9 servers

Fixed an issue where column metadata failed to be retrieved from servers running Hive 0.9.

Query cancellation requests fail when schema is not defined in the connection string

Fixed an issue where the query cancellation from version 2.5.5 does not work properly if the schema is not defined in the connection string.

Queries that include comments fail to execute

This issue has been resolved. Now, queries that contain comments can be processed.

Query cancellation request fails to stop queries

Fixed an issue where the driver ignores query cancellation requests from the application and continues to execute the query. Now, when the application sends a query cancellation request, the query stops executing and the server stops processing the query.

REPLACE commands that include a meta-character as a parameter are not translated correctly

Fixed an issue where queries that use the REPLACE command "replace(expression, pattern, replacement)" with a meta-character as a parameter are not translated correctly. The solution for the issue enables escaping of all meta-characters, including the following:

. ^ \$ * + ? { } [] | 0

Certain known connection string properties could be incorrectly "SET" on the server

In order to support server property configuration in Hive at startup, any unrecognized properties specified in the connection string would cause a "SET" command to be executed automatically on the server side. When this happened, certain combinations of known properties stopped working correctly. This issue has been resolved.

Version 1.0.1

Version 1.0.1 was an internal release of the driver.

Version 1.0.0

Version 1.0.0 was the initial release of Amazon JDBC Driver for Apache Hive.

Contact Us

For support, check the EMR Forum at <https://forums.aws.amazon.com/forum.jspa?forumID=52> or open a support case using the AWS Support Center at <https://aws.amazon.com/support>