

# Altibase Hadoop Connector User's Manual

---

## Altibase 7.3

Altibase® Tools & Utilities

Altibase Tools & Utilities Altibase Hadoop Connector User's Manual  
Release 7.3

Copyright © 2001~2023 Altibase Corp. All Rights Reserved.

This manual contains proprietary information of Altibase® Corporation; it is provided under a license agreement containing restrictions on use and disclosure and is also protected by copyright patent and other intellectual property law. Reverse engineering of the software is prohibited.

All trademarks, registered or otherwise, are the property of their respective owners.

**Altibase Corp**

10F, Daerung PostTower II,  
306, Digital-ro, Guro-gu, Seoul 08378, Korea

Telephone : +82-2-2082-1000

Fax : +82-2-2082-1099

Customer Service Portal : <http://support.altibase.com/en/>

Homepage : <http://www.altibase.com>



# Table Of Contents

---

- [Preface](#)
- [1. Introduction to Altibase Hadoop Connector](#)
  - [Background Knowledge](#)
  - [What is the Altibase Hadoop Connector?](#)
- [2. Installing the Altibase Hadoop Connector](#)
  - [Software Requirements](#)
  - [Installing the Altibase Hadoop Connector](#)
  - [Executing & Testing](#)
- [3. Functions](#)
  - [Command-line Options](#)
  - [Import](#)
  - [Export](#)
  - [list-databases](#)
  - [list-tables](#)
- [Appendix A: Data Types](#)
  - [Supported Data Type](#)

# Preface

---

About This Manual This manual explains how to use the Altibase Hadoop Connector with Altibase and Hadoop.

## Audience

This manual has been prepared for the following Altibase users:

- Database administrators
- Performance administrators
- Database users
- Application developers
- Technical Supporters

It is recommended for those reading this manual possess the following background knowledge:

- Basic knowledge in the use of computers, operating systems, and operating system utilities
- Experience in using relational database and an understanding of database concepts
- Computer programming experience
- Experience in database server management, operating system management, or network administration

## Organization

The manual is organized as follows:

- Chapter 1: Introduction to Altibase Hadoop Connector  
This chapter introduces the Altibase Hadoop Connector and provides background knowledge.
- Chapter 2: Installing the Altibase Hadoop Connector  
This chapter explains how to install the Altibase Hadoop Connector and required software.
- Chapter 3: Functions  
This chapter describes the functionality of the Altibase Hadoop connector with examples.
- Appendix A: Data Types  
This appendix describes the data types supported by the Altibase Hadoop Connector.

## Documentation Convention

This section describes the convention used in this manual. Understanding this convention will make it easier to find information in this manual and in the other manuals in the series.

This convention described here is as follow:

- Sample Code Convention

## Sample Code Conventions

The code examples explain SQL statements, stored procedures, iSQL statements, and other command line syntax.

The following table describes the printing conventions used in the code examples.

Rules	Meaning	Example
[ ]	Indicates an optional item	VARCHAR [(size)] [[FIXED   ] VARIABLE]
{ }	Indicates a mandatory field for which one or more items must be selected.	{ ENABLE   DISABLE   COMPILE }
	A delimiter between optional or mandatory arguments.	{ ENABLE   DISABLE   COMPILE } [ ENABLE   DISABLE   COMPILE ]
...	Indicates that the previous argument is repeated, or that sample code has been omitted.	SQL> SELECT ename FROM employee; ENAME ----- SWNO HJNO HSCHOI . . . 20 rows selected.
Other Symbols	Symbols other than those shown above are part of the actual code.	EXEC :p1 := 1; acc NUMBER(11,2)
Italics	Statement elements in italics indicate variables and special values specified by the user.	SELECT * FROM <i>table_name</i> ; CONNECT <i>userID/password</i> ;
Lower case words	Indicate program elements set by the user, such as table names, column names, file names, etc.	SELECT ename FROM employee;
Upper case words	Keywords and all elements provided by the system appear in upper case.	DESC SYSTEM.SYS_INDICES;

## Related Documentations

For more detailed information, please refer to the following documents.

- Installation Guide
- Getting Started Guide

- Administrator's Manual
- General Reference
- Error Message Reference

## **Altibase Welcomes Your Comments and Feedbacks**

Please let us know what you like or dislike about our manuals. To help us with better future versions of our manuals, please tell us if there is any corrections or classifications that you would find useful.

Include the following information:

- The name and version of the manual that you are using
- Any comments about the manual
- Your name, address, and phone number

If you need immediate assistance regarding any errors, omissions, and other technical issues, please contact [Altibase's Support Portal](#).

Thank you. We always welcome your feedbacks and suggestions.

# 1. Introduction to Altibase Hadoop Connector

---

This chapter introduces the Altibase Hadoop Connector and provides background knowledge.

## Background Knowledge

This chapter introduces the basic concept of transferring data between Hadoop and Altibase with Sqoop.

### Hadoop

Hadoop is a system that is suitable for the analysis and management of large amounts of data. Hadoop is one of the solutions seeing increased interest as demands for parallel processing of data warehouses and the management of 'big data' increase and the popularity of cloud and distributed computing grows.

Hadoop is an open source software based on Java, and is a framework which distributively processes large amounts of data spread across clusters composed of multiple computers. Hadoop is composed of HDFS(Hadoop Distributed File System) and MapReduce.

### Sqoop

Sqoop is a tool for data transfer between Hadoop and relational databases, and is open source software. Using Sqoop, the user can import data from RDBMS to HDFS and export it back to RDBMS. For the import and export of data, Sqoop uses MapReduce. MapReduce offers Fault Tolerance and Parallel Operation.

Sqoop automates most of this process by relying on a database to represent the data to be imported as a schema. Sqoop uses MapReduce to import and export data.

MapReduce provides fault tolerance and parallel operation.

## What is the Altibase Hadoop Connector?

The Altibase Hadoop Connector enables the efficient transfer of data between Hadoop and Altibase servers. It also allows Altibase to handle data management and Hadoop to handle data analysis. In other words, the Altibase Hadoop Connector lets the user connect to the Altibase server and export data into HDFS or HIVE for the purpose of processing data from Hadoop.

The Altibase Hadoop Connector operates on Sqoop and supports nearly all functions provided by it. Also, users that have experience with Sqoop will easily be able to use the Altibase Hadoop Connector, as it uses a command line argument structure similar to Sqoop



## 2. Installing the Altibase Hadoop Connector

---

This chapter explains how to install the Altibase Hadoop Connector and required software.

### Software Requirements

The following software must be installed in order to install and run the Altibase Hadoop Connector.

- JRE (Java Runtime Environment) or JDK (Java Development Kit) version 1.6 or later
- Hadoop Version 1.0
- Sqoop version 1.4.4 or later
- Altibase version 5.0 or later

This section describes how to install the Hadoop and Swap required to run the Altibase Hadoop connector. In addition, it describes how to install the Altibase JDBC driver in Hadoop and Scan environment for interworking with the Altibase server.

### Installing Hadoop

Install Hadoop in the following order and set the Hadoop operating environment.

1. Go to <https://www.cloudera.com/content/support/en/downloads.html>, download Hadoop version 1.0.
2. Install Hadoop according to the Hadoop installation method provided by [<http://www.cloudera.com>], and set environment variables such as HADOOP\_HOME.

### Installing Sqoop

The Altibase Hadoop Connector is based on Apache Sqoop, which is distributed under the Apache License version 2.0.

Install Sqoop in the following order and install the Altibase JDBC driver into the Sqoop environment to enable sqoop to connect to Altibase.

1. Download the Sqoop (package Sqoop-1.4.4.bin\_hadoop-1.0.0.tar.gz) that supports Hadoop 1.0 from <http://Mirror.apache.kr.org/sqoop/1.4.4>. The current Altibase Hadoop Connector officially supports Sqoop 1.4.4 (Sqoop 1.9 will be supported soon).
2. Install Sqoop using the installation manual from the Sqoop homepage

### Installing the JDBC Driver

If both Hadoop and Scoop are installed, follow the instruction below to install the JDBC driver.

## When using Altibase

1. Copy the JDBC driver file for Altibase (\$ALTIBASE\_HOME/lib/Altibase.jar) to the \$SQOOP\_HOME/lib.

```
% cp $ALTIBASE_HOME/lib/Altibase.jar $SQOOP_HOME/lib
```

## Installing the Altibase Hadoop Connector

Install the Altibase Hadoop Connector in the following order:

1. Download the Altibase Hadoop Connector Package from <http://altibase.com/supportcenter/en/>.

Package for Altibase: altibase\_sqoop14\_connector.jar

2. Copy the Altibase Hadoop connector package to \$SQOOP\_HOME/lib.

## Executing & Testing

The following is a command which retrieves the table list from the Altibase server and can check whether or not the Altibase Hadoop Connector is installed properly.

```
% sqoop list-tables
--connect jdbc:Altibase://127.0.0.1:20300/mydb
--driver Altibase.jdbc.driver.AltibaseDriver
--username SYS
--password MANAGER
--connection-manager com.altibase.sqoop.manager.AltibaseManager
```

If the following table list is printed out with the log below after execution, it means that the Altibase Hadoop Connector has been loaded properly.

```
13/10/02 13:48:15 INFO manager.AltibaseManager: init default option autocommit false
13/10/02 13:48:15 INFO manager.SqlManager: Using default fetchSize of 1000
13/10/02 13:48:15 INFO manager.AltibaseManager: Altibase manager 1.0 connector
create
```

## 3. Functions

This chapter explains Altibase Hadoop Connector functions with examples.

### Command-line Options

Because the Altibase Hadoop connector is based on the sqoop, it supports all the functions provided by the sqoop. Since this chapter only describes some of the features, please refer to <http://sqoop.apache.org/docs/1.4.4/SqoopUserGuide.html>.

### Syntax

Listed below is an example of the Altibase Hadoop Connector syntax:

```
sqoop <command> --connect <url>
--driver <driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
```

### Options

Options	Description
<command>	Since only the following commands are described in this document, refer to the sqoop document mentioned above for other commands. export import list-databases list-tables
--connect <url>	Specifies the Altibase URL: jdbc:Altibase://ip:port/mydb
--driver <driver>	Specifies the Altibase JDBC driver: class:Altibase.jdbc.driver.AltibaseDriver
--username <user>	Specifies the database user.
--password <password>	Specifies the password.
--connection-manager <connector>	Specifies the Altibase Hadoop Connector: com.altibase.sqoop.manager.AltibaseManager

Additional commands, import and export options can be found at: <http://sqoop.apache.org/docs/1.4.4/SqoopUserGuide.html>.

# Import

Import is a function which imports the data of Altibase to HDFS or Hive. This section provides explanations of these functions with examples.

## Importing to HDFS in a Text File

The following options can be used with the Sqoop command to store data of a certain table existing in Altibase to a configured directory on HDFS in text file format.

```
% sqoop import
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
--table <table_name>
--split-by <split_name>
```

By default, text files are stored as CSV. Separators, such as field termination characters, line termination characters can be specified with the options listed in the following table. If the user does not specify one of the options, default values are taken.

Option	Description	Default Value
--enclosed-by	The character which encloses the field; this option is required.	\"
--escaped-by	The escape character	
--fields-terminated-by	The field separator character	,
--lines-terminated-by	The line termination character	\n
--mysql-delimiters	The MySQL default delimiter; the following are available: fields-terminated-by: , lines-terminated-by: \n escaped-by: \ optionally-enclosed-by: '	
--optionally-enclosed-by	The character which encloses the field; this option is optional.	

Note: If a line separator or a field separator is included in the field, it must be enclosed in punctuation marks, such as quotation marks(""). If a quotation mark is included in the field, an escape character must be prefixed to the quotation mark to indicate that it is not a punctuation mark. Due to a sqoop bug found at the time of writing(sqoop operates abnormally when the

punctuation mark and the escape character are duplicate), the Altibase Hadoop Connector only provides quotation marks as the default value for punctuation marks and offers support for the user to directly specify the escape character

## Importing to HDFS in a Sequence File

The following options can be used with the Sqoop command to store data of a certain table existing in Altibase to a specified directory on HDFS in sequence file format.

```
% sqoop import
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
--table <table_name>
--split-by <split_name>
--as-sequencefile
```

## Importing to HDFS in an Avro File

The following options can be used with the Sqoop command to store data of a certain table existing in Altibase to a specified directory on HDFS in Avro file format.

```
% sqoop import
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
--table <table_name>
--split-by <split_name>
--as-avrodatafile
```

## Importing to HDFS Using a Query

The following options can be used with the Sqoop command to store data of a certain table existing in Altibase into HDFS using a user-specified query statement.

```
% sqoop import
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
--table <table_name>
--split-by <split_name>
--boundary-query <query>
```

## Importing to Hive

The following options can be used with the Sqoop command to store data of a certain table existing in Altibase into Hive.

```
% sqoop import
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
--table <table_name>
--split-by <split_name>
--hive-import
```

## Export

Export is a function which exports HDFS data to Altibase. Depending on the option, data is inserted into the database or existing data is updated.

## Inserting Data

The following options can be used with the Sqoop command to store HDFS data in a certain table existing in Altibase.

```
% sqoop export
-D sqoop.export.records.per.statement=<size>
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
--table <table_name>
--export-dir <dir>
```

If the value set for sqoop.export.records.per.statement is larger than 1, the Altibase Hadoop Connector operates in batch mode. Batch mode inserts multiple records in one EXECUTE statement. This option must precede other options. If the value set for sqoop.export.records.per.statement is 1, it is possible that the Altibase Hadoop Connect will not operate in batch mode.

## Inserting Data From a CSV File

When exporting with a CSV file, options which specify delimiters are available for use. For more detailed information on options available for use, please refer to “Importing to HDFS in a Text File”.

## Inserting Data in Batch Mode

The following options can be used with the Sqoop command to store HDFS data to a certain table existing in Altibase using the batch mode feature of the Altibase JDBC driver.

```
% sqoop export
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
--table <table_name>
--export-dir <dir>
--batch
```

Regardless of whether or not the batch option has been specified, the Altibase Hadoop Connector runs in batch mode and 100 records are inserted by default for one Execute operation. To disable the Altibase Hadoop Connector from running in batch mode, the -D sqoop.export.records.per.statement option should be set to 1.

## Updating Data

The following options can be used with the Sqoop command to update data of a certain table existing in Altibase to HDFS data

```
% sqoop export
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
--table <table_name>
--export-dir <dir>
--update-key <column>
```

## Updating or Inserting Data

This function performs an update if HDFS data exists in Altibase, and performs an insert if not. The following options can be used with the Sqoop command:.

```
% sqoop export
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
--table <table_name>
--export-dir <dir>
--update-key <column>
--update-mode allowinsert
```

Note: Since this function uses the MERGE statement of Altibase, it is only supported when the Altibase Hadoop Connector is used with Altibase version 6.3.1 or above.

## Inserting Data Using a Staging-Table

Since Sqoop exports with multiple transactions, some data can fail to be committed. The -- staging-table option is used to prevent such failure and first inserts HDFS data to the table specified for this option, and then moves this data to a target table in Altibase. The following options can be used with the Sqoop command:

```
% sqoop export
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
--table <table_name>
--export-dir <dir>
--update-key <column>
--staging-table <table_name>
```

## list-databases

The following options can be used with the Sqoop command to retrieve the database list of Altibase.

```
% sqoop list-databases
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
```

## list-tables

The following options can be used with the Sqoop command to retrieve the tables existing in Altibase.

```
% sqoop list-tables
--connect <url>
--driver <jdbc_driver>
--username <user>
--password <password>
--connection-manager com.altibase.sqoop.manager.AltibaseManager
```



# Appendix A: Data Types

This appendix describes data types supported by Altibase Hadoop Connector.

## Supported Data Type

The following is the table that shows the conversion of data type between Sqoop and Altibase when importing and exporting using the Altibase Hadoop Connector. Also it shows whether it supports for import or export depending on the data types.

Altibase Data Type	Sqoop Data Type	Import	Export
CHAR	String	O	O
VARCHAR	String	O	O
NCHAR	String	O	O
NVARCHAR	String	O	O
INTEGER	Integer	O	O
BIGINT	Long	O	O
SMALLINT	Integer	O	O
NUMBER	Double	O	O
NUMERIC	java.math.BigDecimal	O	O
DECIMAL	java.math.BigDecimal	O	O
FLOAT	Double	O	O
DOUBLE	Double	O	O
REAL	Float	O	O
DATE	java.sql.Timestamp	O	O
BLOB	com.cloudera.sqoop.lib.BlobRef	O	X
CLOB	com.cloudera.sqoop.lib.ClobRef	O	X

Nope: BLOB and CLOB data types are supported only for import in Sqoop so Altibase Hadoop Connector also supports import as well. The export for BLOB and CLOB data types is to be supported soon.