

Sqoop Installation Guide for Hadoop 3.3.6 Cluster

This document provides step-by-step instructions for installing Apache Sqoop 1.4.7 on a Hadoop 3.3.6 cluster.

Prerequisites

Before installing Sqoop, ensure you have:

- Hadoop 3.3.6 installed and configured
- Java 11 installed
- User with sudo privileges
- Internet connection for downloading dependencies

Installation Steps

Step 1: Verify Hadoop Installation

First, verify that your Hadoop cluster is running correctly:

```
# Check Hadoop version
hadoop version

# Check Hadoop processes
jps

# Verify HDFS status
hdfs dfsadmin -report
```

Expected output should show:

- Hadoop 3.3.6 version
- Running Hadoop daemons (NameNode, DataNode, ResourceManager, etc.)
- Active DataNodes in the cluster

Step 2: Download Sqoop

Download Apache Sqoop 1.4.7 binary distribution:

```
# Navigate to home directory
cd ~

# Download Sqoop 1.4.7 binary distribution
wget https://archive.apache.org/dist/sqoop/1.4.7/sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz
```

Step 3: Extract and Install Sqoop

Extract the downloaded archive and install Sqoop:

```
# Extract Sqoop to /usr/local directory
sudo tar -xzf sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz -C /usr/local

# Rename the extracted directory for easier access
sudo mv /usr/local/sqoop-1.4.7.bin__hadoop-2.6.0 /usr/local/sqoop

# Change ownership to your user
sudo chown -R ubuntu:ubuntu /usr/local/sqoop
```

Step 4: Configure Environment Variables

Add Sqoop environment variables to your `.bashrc` file:

```
# Edit bashrc file
nano ~/.bashrc
```

Add the following lines at the end of the file:

```
# Sqoop Environment Variables
export SQOOP_HOME=/usr/local/sqoop
export PATH=$PATH:$SQOOP_HOME/bin
export HADOOP_CLASSPATH=$(hadoop classpath)
```

Reload the environment variables:

```
# Apply changes to current session
source ~/.bashrc

# Or manually set for current session
export SQOOP_HOME=/usr/local/sqoop
export PATH=$PATH:$SQOOP_HOME/bin
export HADOOP_CLASSPATH=$(hadoop classpath)
```

Step 5: Configure Sqoop Environment

Configure Sqoop to work with your Hadoop installation:

```
# Copy the environment template
cp $SQOOP_HOME/conf/sqoop-env-template.sh $SQOOP_HOME/conf/sqoop-env.sh
```

```
# Edit the configuration file
nano $SQOOP_HOME/conf/sqoop-env.sh
```

Uncomment and update the following lines in the file:

```
#Set path to where bin/hadoop is available
export HADOOP_COMMON_HOME=/usr/local/hadoop

#Set path to where hadoop-*-core.jar is available
export HADOOP_MAPRED_HOME=/usr/local/hadoop
```

Step 6: Install Database JDBC Drivers

Download and install JDBC drivers for the databases you want to connect to. Here's an example for MySQL:

```
# Download MySQL JDBC driver
wget https://repo1.maven.org/maven2/mysql/mysql-connector-
java/8.0.28/mysql-connector-java-8.0.28.jar

# Move the driver to Sqoop lib directory
mv mysql-connector-java-8.0.28.jar $SQOOP_HOME/lib/
```

For other databases, download their respective JDBC drivers and place them in `$SQOOP_HOME/lib/`:

- **PostgreSQL:** `postgresql-xx.x.x.jar`
- **Oracle:** `ojdbcx.jar`
- **SQL Server:** `mssql-jdbc-x.x.x.jre.x.jar`
- **MySQL:** `mysql-connector-java-x.x.x.jar`

Step 6.1: Install Required Dependencies (IMPORTANT)

Sqoop requires Apache Commons Lang library to avoid `NoClassDefFoundError: org/apache/commons/lang/StringUtils` errors:

```
# Install Apache Commons Lang library
sudo apt-get update
sudo apt-get install -y libcommons-lang-java

# Copy the commons-lang jar to Sqoop lib directory
cp /usr/share/java/commons-lang-2.6.jar $SQOOP_HOME/lib/

# Verify the jar is installed
ls -la $SQOOP_HOME/lib/ | grep commons-lang
```

Note: Sqoop comes with commons-lang3 but requires commons-lang (version 2.x) for the StringUtils class.

Step 7: Verify Installation

Test the Sqoop installation:

```
# Check Sqoop version
sqoop version

# List available commands
sqoop help
```

Expected output should show:

- Sqoop version 1.4.7
- List of available commands (codegen, import, export, etc.)

You may see some warnings about HBase, HCatalog, Accumulo, and Zookeeper not being installed. These are optional components and can be ignored for basic Sqoop functionality.

Post-Installation Configuration

MySQL Server Verification

Ensure you have MySQL server setup and configured for remote connections. If you need help setting up MySQL, visit: <https://github.com/SoyVitouPro/MySQL-Only-No-MySQL-Admin> for complete setup instructions.

After installing MySQL or using an existing MySQL setup, run these commands to verify connectivity:

```
# Test network connectivity to MySQL server
ping mysql_server_ip

# Test Sqoop connection
sqoop list-databases --connect jdbc:mysql://mysql_server_ip:port --username
username --password password
```

Testing Sqoop with a Sample Database Connection

To test Sqoop with a database, you can use the following command structure:

```
# List databases on a MySQL server (replace with actual IP and port)
sqoop list-databases \
  --connect jdbc:mysql://mysql_server_ip:port/ \
  --username username \
  --password password

# Example with working configuration:
sqoop list-databases \
  --connect jdbc:mysql://172.17.199.56:4987 \
```

```
--username ikhode \  
--password ikhode2357.!  
  
# List tables in a specific database  
sqoop list-tables \  
--connect jdbc:mysql://mysql_server_ip:port/database_name \  
--username username \  
--password password  
  
# Example listing tables:  
sqoop list-tables \  
--connect jdbc:mysql://172.17.199.56:4987/mydb \  
--username ikhode \  
--password ikhode2357.!
```

Common Sqoop Commands

Here are some common Sqoop operations:

```
# Import a table from MySQL to HDFS  
sqoop import \  
--connect jdbc:mysql://hostname:3306/database_name \  
--username username \  
--password password \  
--table table_name \  
--target-dir /user/hadoop/table_name \  
--m 1  
  
# Export data from HDFS to MySQL  
sqoop export \  
--connect jdbc:mysql://hostname:3306/database_name \  
--username username \  
--password password \  
--table table_name \  
--export-dir /user/hadoop/table_name \  
--m 1
```

Troubleshooting

Common Issues and Solutions

1. Command not found error

```
# Ensure environment variables are set  
echo $SQOOP_HOME  
echo $PATH  
  
# Set manually if needed
```

```
export SQOOP_HOME=/usr/local/sqoop
export PATH=$PATH:$SQOOP_HOME/bin
```

2. Hadoop classpath issues

```
# Set Hadoop classpath
export HADOOP_CLASSPATH=$(hadoop classpath)
```

3. NoClassDefFoundError: org/apache/commons/lang/StringUtils

```
# This error occurs because Sqoop needs commons-lang (not commons-
lang3)
sudo apt-get install -y libcommons-lang-java
cp /usr/share/java/commons-lang-2.6.jar $SQOOP_HOME/lib/

# Verify installation
ls -la $SQOOP_HOME/lib/ | grep commons-lang
```

4. Database connection errors

- Verify JDBC driver is in `$SQOOP_HOME/lib/`
- Check database connectivity with `nc -zv mysql_server_ip 3306`
- Verify connection string format (use correct port, especially for Docker)
- Ensure database server is accessible from the cluster
- Check if MySQL user has remote access privileges
- Verify MySQL bind-address configuration (should be 0.0.0.0 for remote access)

5. Connection timeout errors

```
# Test network connectivity
ping mysql_server_ip

# Check port accessibility
telnet mysql_server_ip 3306

# For Docker containers, verify port mapping
docker ps | grep mysql

# Check MySQL configuration for remote access
# In MySQL config: bind-address = 0.0.0.0
```

6. Permission issues

```
# Check permissions
ls -la /usr/local/sqoop/

# Fix ownership if needed
sudo chown -R ubuntu:ubuntu /usr/local/sqoop
```





Expected Warnings

The following warnings are normal and can be ignored for basic functionality:

```
Warning: /usr/local/sqoop/../hbase does not exist! HBase imports will fail.
Warning: /usr/local/sqoop/../hcatalog does not exist! HCatalog jobs will
fail.
Warning: /usr/local/sqoop/../accumulo does not exist! Accumulo imports will
fail.
Warning: /usr/local/sqoop/../zookeeper does not exist! Accumulo imports
will fail.
```

Integration with Hadoop Cluster

Sqoop is now configured to work with your existing Hadoop 3.3.6 cluster. The installation:

1.  Uses the existing Hadoop configuration
2.  Integrates with HDFS for data storage
3.  Works with YARN for resource management
4.  Supports MapReduce jobs for data transfer

Next Steps

1. **Configure Database Connections:** Set up connections to your source databases
2. **Test Data Transfers:** Perform sample import/export operations
3. **Schedule Jobs:** Set up automated data transfer workflows
4. **Monitor Performance:** Optimize parallel operations and resource usage

Files Modified/Created

- `/usr/local/sqoop/` - Sqoop installation directory
- `/usr/local/sqoop/conf/sqoop-env.sh` - Sqoop environment configuration
- `~/.bashrc` - Updated with Sqoop environment variables
- `/usr/local/sqoop/lib/mysql-connector-java-8.0.28.jar` - MySQL JDBC driver

Verification Commands

Sqoop Installation Verification

```
# Verify Sqoop installation
sqoop version

# Check environment variables
echo $SQOOP_HOME
echo $HADOOP_CLASSPATH

# Verify Hadoop integration
hadoop classpath

# Test Sqoop help
sqoop help

# Verify required dependencies are installed
ls -la $SQOOP_HOME/lib/ | grep commons-lang
ls -la $SQOOP_HOME/lib/ | grep mysql-connector
```

MySQL Connection Verification

```
# 1. Test basic network connectivity
ping mysql_server_ip

# 2. Test MySQL port accessibility
nc -zv mysql_server_ip mysql_port

# 3. For Docker MySQL containers, check port mapping
docker ps | grep mysql

# 4. Test MySQL connection with MySQL client
mysql -h mysql_server_ip -P mysql_port -u username -p

# 5. Verify Sqoop-MySQL connection
sqoop list-databases \
  --connect jdbc:mysql://mysql_server_ip:mysql_port \
  --username username \
  --password password

# 6. Test actual data import (optional)
sqoop import \
  --connect jdbc:mysql://mysql_server_ip:mysql_port/database_name \
  --username username \
  --password password \
  --table table_name \
  --target-dir /user/hadoop/test_import \
  --m 1
```

Complete Working Example


```
# Working example with specific configuration
sqoop list-databases \
  --connect jdbc:mysql://172.17.199.56:4987 \
  --username ikhode \
  --password ikhode2357.!

# Expected output:
# information_schema
# performance_schema
# mydb
```

Sqoop is now successfully installed and configured to work with your Hadoop 3.3.6 cluster!