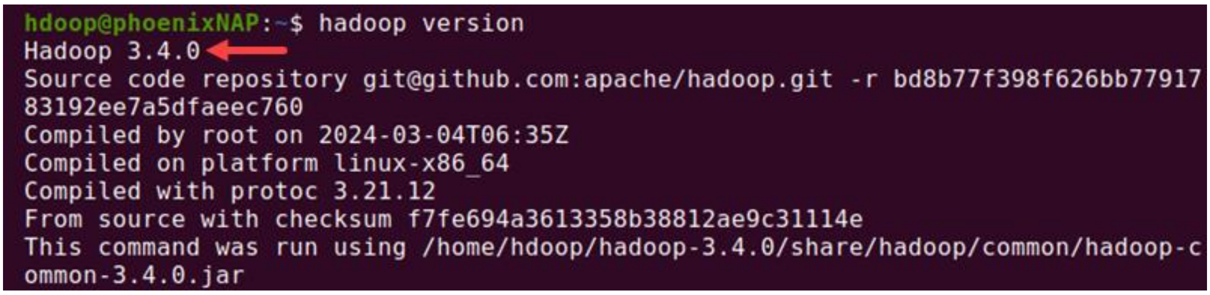
Install Apache Hive

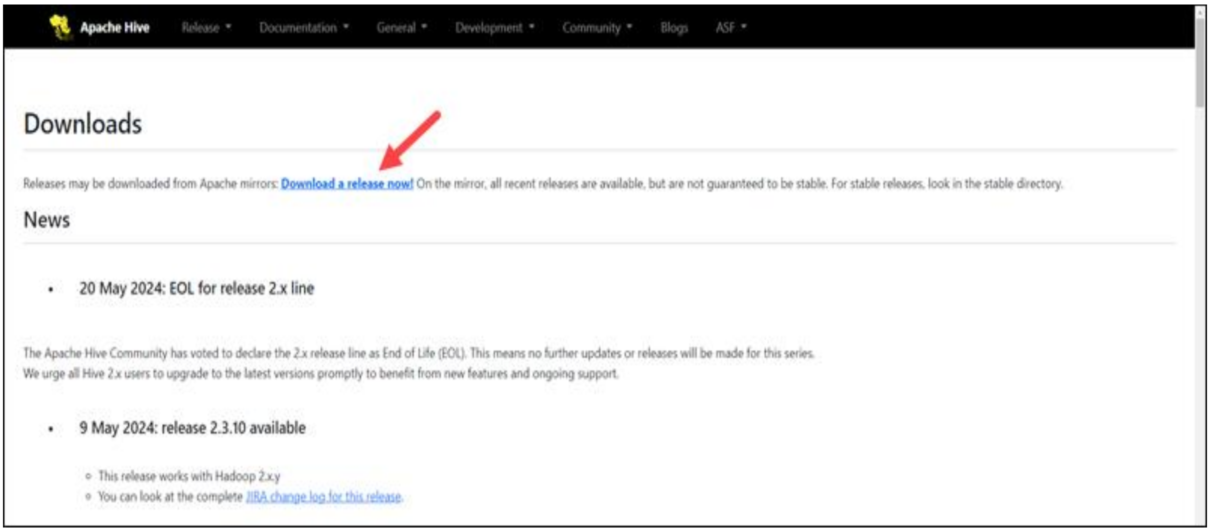
Step 1: Download and Untar Hive

1. Visit the Apache Hive official download page and determine which Hive version is compatible with the local Hadoop installation. To check the Hadoop version, run the following in the terminal:



```
hdoop@phoenixNAP:~$ hadoop version
Hadoop 3.4.0
Source code repository git@github.com:apache/hadoop.git -r bd8b77f398f626bb77917
83192ee7a5dfaeec760
Compiled by root on 2024-03-04T06:35Z
Compiled on platform linux-x86_64
Compiled with protoc 3.21.12
From source with checksum f7fe694a3613358b38812ae9c31114e
This command was run using /home/hdoop/hadoop-3.4.0/share/hadoop/common/hadoop-c
ommon-3.4.0.jar
```

2. Click the **Download a release now!** link to access the mirrors page.



Downloads

Releases may be downloaded from Apache mirrors: Download a release now! On the mirror, all recent releases are available, but are not guaranteed to be stable. For stable releases, look in the stable directory.

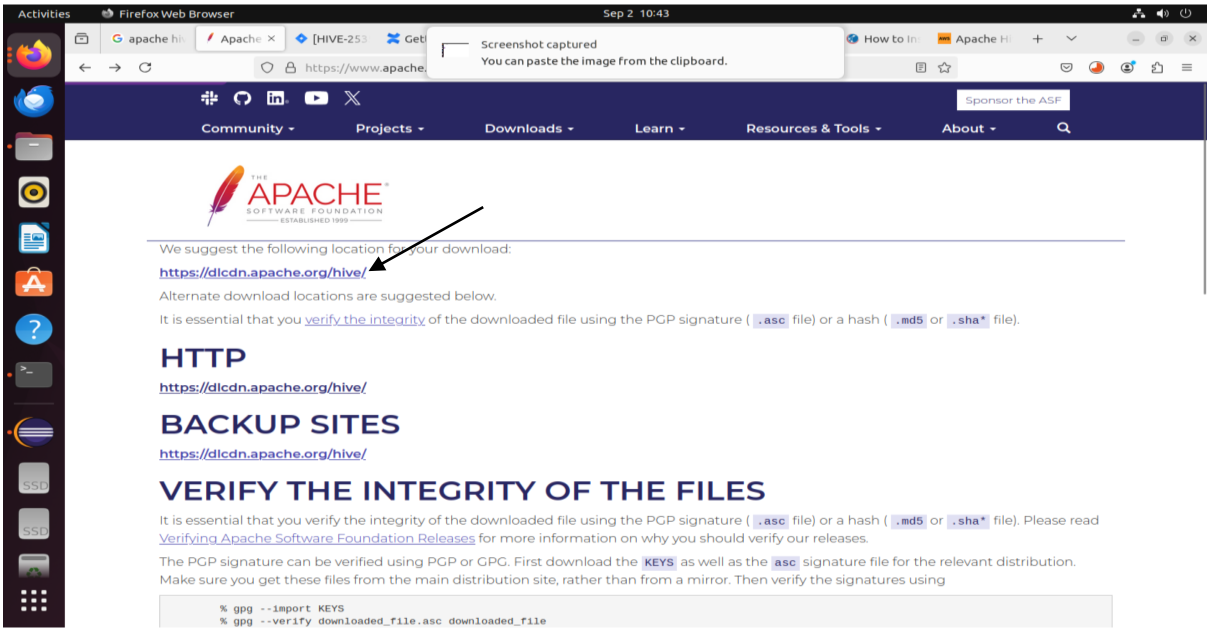
News

- 20 May 2024: EOL for release 2.x line

The Apache Hive Community has voted to declare the 2.x release line as End of Life (EOL). This means no further updates or releases will be made for this series. We urge all Hive 2.x users to upgrade to the latest versions promptly to benefit from new features and ongoing support.

- 9 May 2024: release 2.3.10 available
  - This release works with Hadoop 2.x.y
  - You can look at the complete JIRA change log for this release.

3. Choose the default mirror link.



We suggest the following location for your download:

https://dlcdn.apache.org/hive/

Alternate download locations are suggested below.

It is essential that you verify the integrity of the downloaded file using the PGP signature (.asc file) or a hash (.md5 or .sha* file).

HTTP

https://dlcdn.apache.org/hive/

BACKUP SITES

https://dlcdn.apache.org/hive/

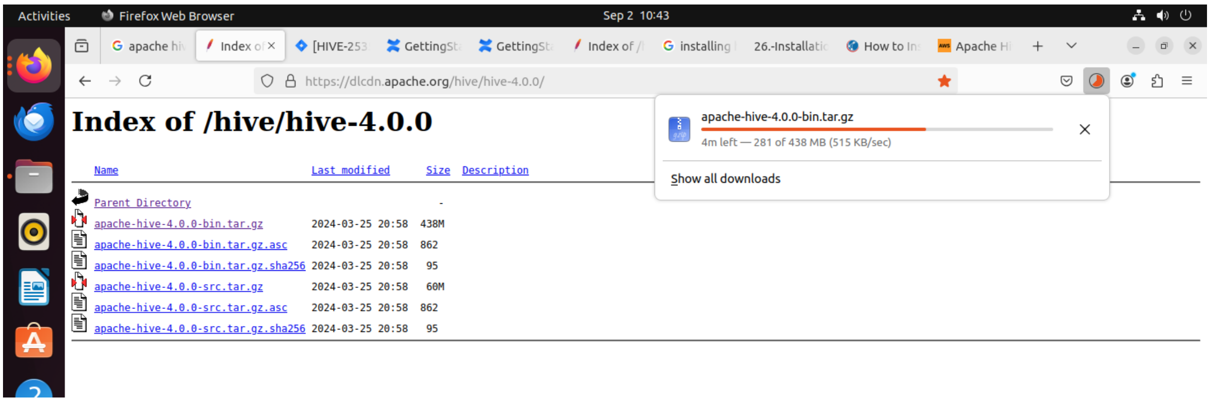
VERIFY THE INTEGRITY OF THE FILES

It is essential that you verify the integrity of the downloaded file using the PGP signature (.asc file) or a hash (.md5 or .sha* file). Please read Verifying Apache Software Foundation Releases for more information on why you should verify our releases.

The PGP signature can be verified using PGP or GPG. First download the KEYS as well as the asc signature file for the relevant distribution. Make sure you get these files from the main distribution site, rather than from a mirror. Then verify the signatures using

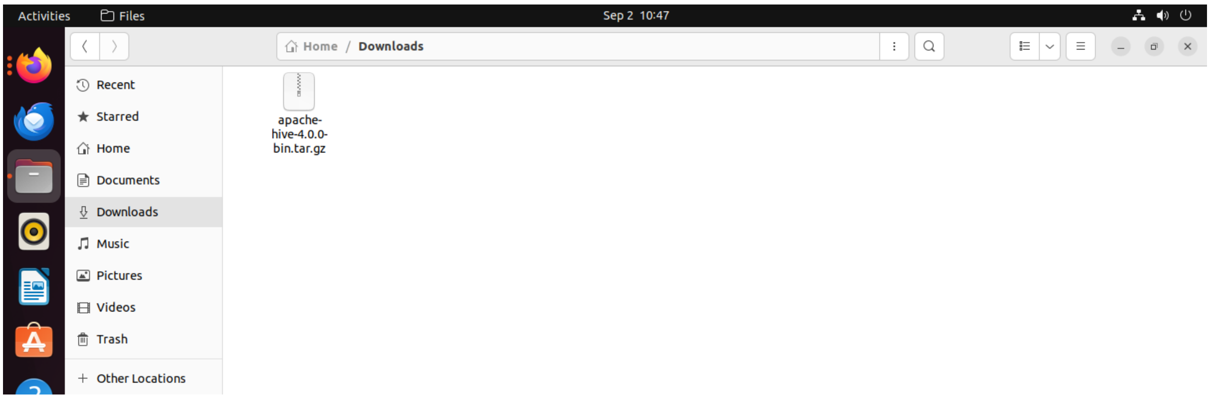
```
% gpg --import KEYS
% gpg --verify downloaded_file.asc downloaded_file
```

The link leads to a downloads listing page

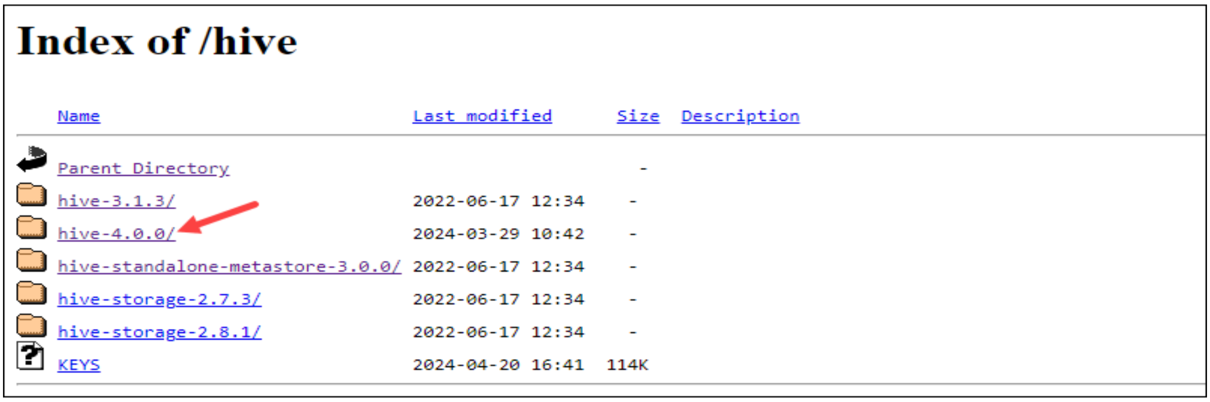


Index of /hive/hive-4.0.0

Name	Last modified	Size	Description
Parent Directory		-	
apache-hive-4.0.0-bin.tar.gz	2024-03-25 20:58	438M	
apache-hive-4.0.0-bin.tar.gz.asc	2024-03-25 20:58	862	
apache-hive-4.0.0-bin.tar.gz.sha256	2024-03-25 20:58	95	
apache-hive-4.0.0-src.tar.gz	2024-03-25 20:58	60M	
apache-hive-4.0.0-src.tar.gz.asc	2024-03-25 20:58	862	
apache-hive-4.0.0-src.tar.gz.sha256	2024-03-25 20:58	95	



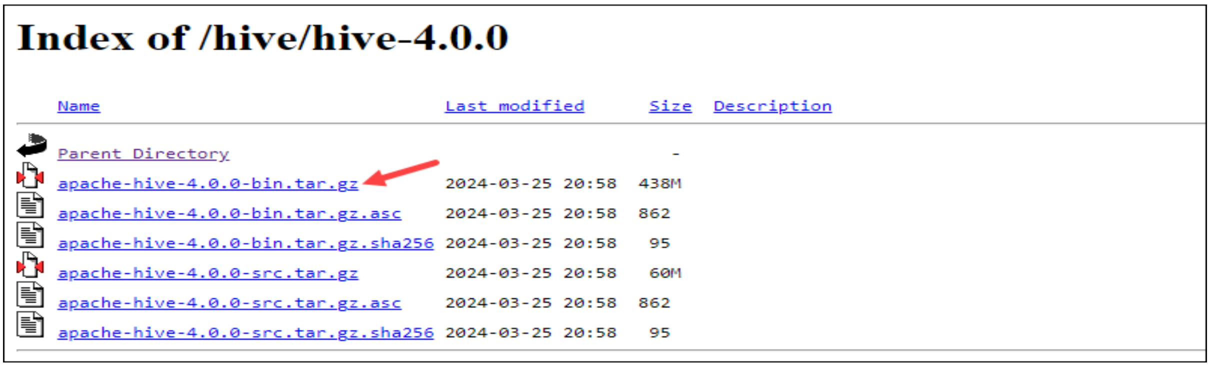
4. Open the directory for the desired Hive version.



Index of /hive

Name	Last modified	Size	Description
Parent Directory		-	
hive-3.1.3/	2022-06-17 12:34	-	
hive-4.0.0/	2024-03-29 10:42	-	
hive-standalone-metastore-3.0.0/	2022-06-17 12:34	-	
hive-storage-2.7.3/	2022-06-17 12:34	-	
hive-storage-2.8.1/	2022-06-17 12:34	-	
KEYS	2024-04-20 16:41	114K	

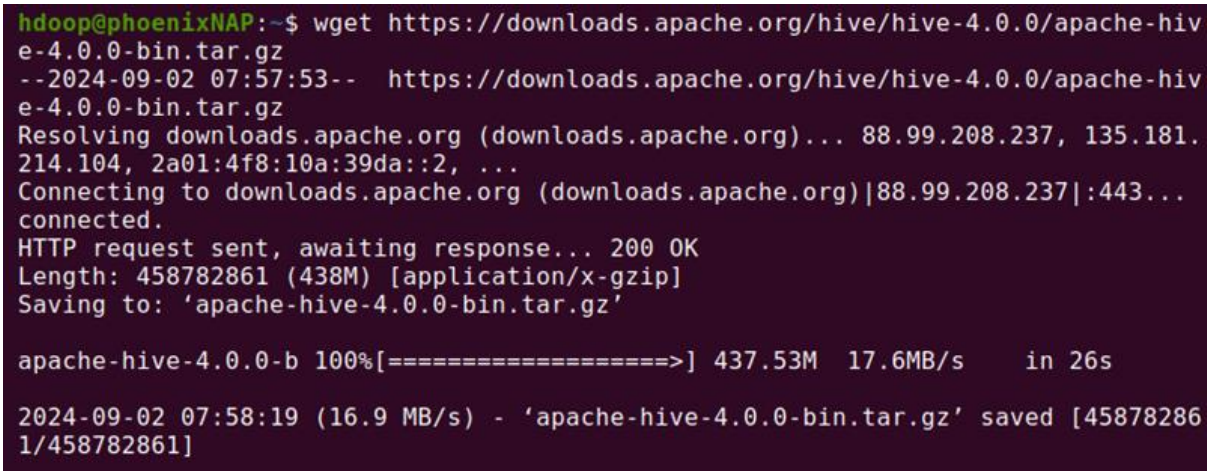
5. Select the *bin.tar.gz* file to begin the download.



Index of /hive/hive-4.0.0

Name	Last modified	Size	Description
Parent Directory		-	
apache-hive-4.0.0-bin.tar.gz	2024-03-25 20:58	438M	
apache-hive-4.0.0-bin.tar.gz.asc	2024-03-25 20:58	862	
apache-hive-4.0.0-bin.tar.gz.sha256	2024-03-25 20:58	95	
apache-hive-4.0.0-src.tar.gz	2024-03-25 20:58	60M	
apache-hive-4.0.0-src.tar.gz.asc	2024-03-25 20:58	862	
apache-hive-4.0.0-src.tar.gz.sha256	2024-03-25 20:58	95	

Alternatively, copy the URL and use the wget command to download the file:

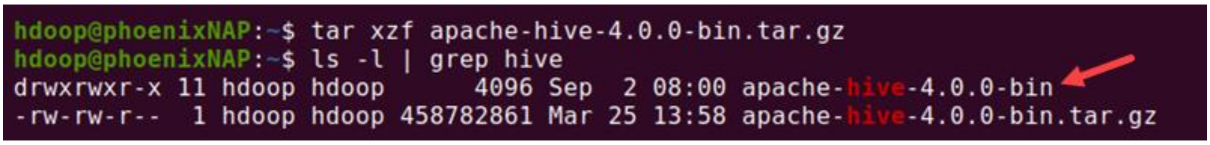


```
hdoop@phoenixNAP:~$ wget https://downloads.apache.org/hive/hive-4.0.0/apache-hiv
e-4.0.0-bin.tar.gz
--2024-09-02 07:57:53--  https://downloads.apache.org/hive/hive-4.0.0/apache-hiv
e-4.0.0-bin.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 88.99.208.237, 135.181.
214.104, 2a01:4f8:10a:39da::2, ...
Connecting to downloads.apache.org (downloads.apache.org)|88.99.208.237|:443...
connected.
HTTP request sent, awaiting response... 200 OK
Length: 458782861 (438M) [application/x-gzip]
Saving to: ‘apache-hive-4.0.0-bin.tar.gz’

apache-hive-4.0.0-b 100%[===================>] 437.53M  17.6MB/s    in 26s

2024-09-02 07:58:19 (16.9 MB/s) - ‘apache-hive-4.0.0-bin.tar.gz’ saved [45878286
1/458782861]
```

6. When the download completes, extract the tar.gz archive by providing the command with the exact file name:



```
hdoop@phoenixNAP:~$ tar xzf apache-hive-4.0.0-bin.tar.gz
hdoop@phoenixNAP:~$ ls -l | grep hive
drwxrwxr-x 11 hdoop hdoop      4096 Sep  2 08:00 apache-hive-4.0.0-bin
-rw-rw-r--  1 hdoop hdoop 458782861 Mar 25 13:58 apache-hive-4.0.0-bin.tar.gz
```

Step 2: Configure Hive Environment Variables (.bashrc)

Set the **HIVE_HOME** environment variable to direct the client shell to the *apache-hive-4.0.0-bin* directory and add it to **PATH**:

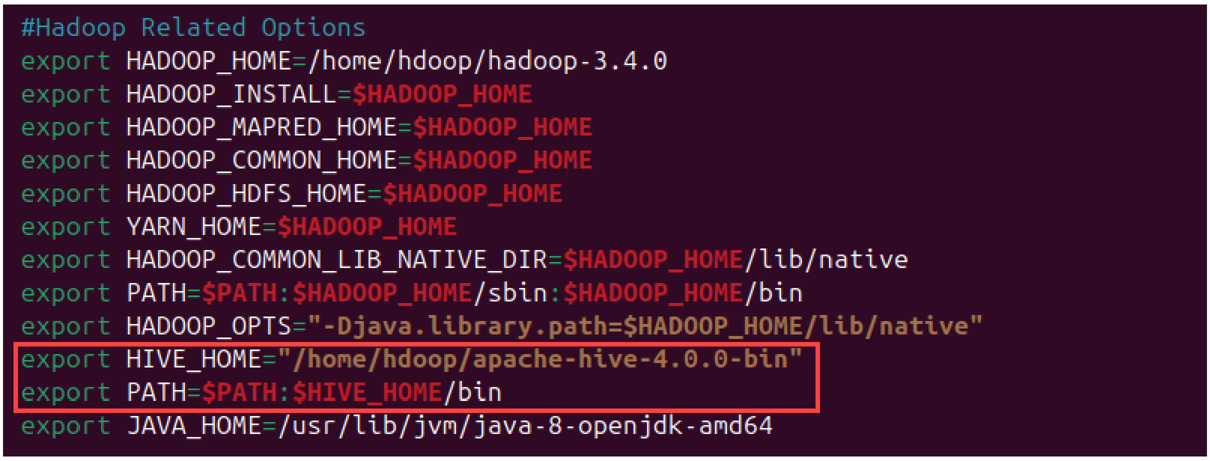
1. Edit the *.bashrc* shell configuration file using a text editor (we will use nano):

nano .bashrc

2. Append the following Hive environment variables to the *.bashrc* file and ensure you provide the correct Hive program version:

export HIVE_HOME="/home/hdoop/apache-hive-4.0.0-bin"

export PATH=$PATH:$HIVE_HOME/bin



```
#Hadoop Related Options
export HADOOP_HOME=/home/hdoop/hadoop-3.4.0
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HIVE_HOME="/home/hdoop/apache-hive-4.0.0-bin"
export PATH=$PATH:$HIVE_HOME/bin
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

The Hadoop environment variables are in the same file.

3. Save and exit the *.bashrc* file.

4. Apply the changes to the current environment:

```
source ~/.bashrc
```

The variables are immediately available in the current shell session.

Step 3: Edit core-site.xml File

Adjust the settings in the *core-site.xml* file, which is part of the Hadoop configuration:

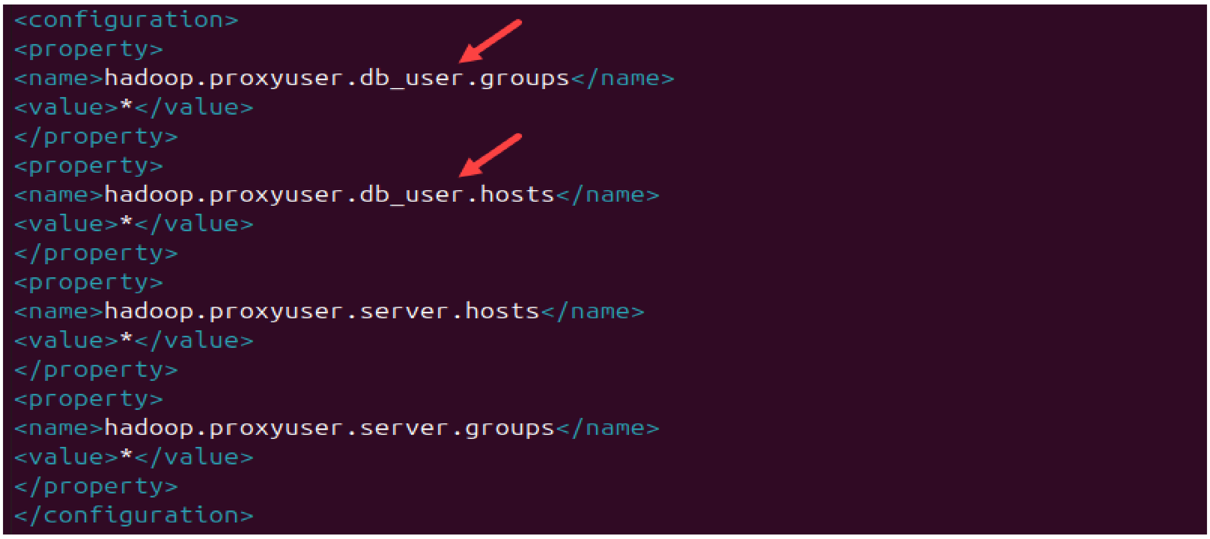
1. Open the *core-site.xml* file in a text editor:

```
nano $HADOOP_HOME/etc/hadoop/core-site.xml
```

Change the path if the file is in a different location or if the Hadoop version differs.

2. Paste the following lines in the file:

```
<configuration>
<property>
<name>hadoop.proxyuser.db_user.groups</name>
<value>*</value>
</property>
<property>
<name>hadoop.proxyuser.db_user.hosts</name>
<value>*</value>
</property>
<property>
<name>hadoop.proxyuser.server.hosts</name>
<value>*</value>
</property>
<property>
<name>hadoop.proxyuser.server.groups</name>
<value>*</value>
</property>
</configuration>
```



```
<configuration>
<property>
<name>hadoop.proxyuser.db_user.groups</name>
<value>*</value>
</property>
<property>
<name>hadoop.proxyuser.db_user.hosts</name>
<value>*</value>
</property>
<property>
<name>hadoop.proxyuser.server.hosts</name>
<value>*</value>
</property>
<property>
<name>hadoop.proxyuser.server.groups</name>
<value>*</value>
</property>
</configuration>
```

The **db_user** is the username used to connect to the database.

3. Save the file and close **nano**.

Step 4: Create Hive Directories in HDFS

The directory is within the HDFS storage layer. It will contain the intermediary data Hive sends to the HDFS. Follow the steps below:

1. Create a */tmp* directory:

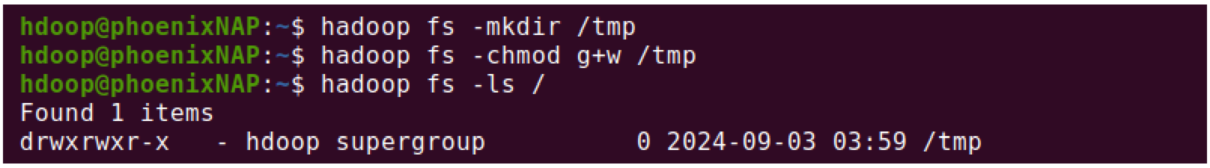
hadoop fs -mkdir /tmp

2. Add write and execute permissions to group members with:

hadoop fs -chmod g+w /tmp

3. Check the permissions with:

hadoop fs -ls /



```
hdoop@phoenixNAP:~$ hadoop fs -mkdir /tmp
hdoop@phoenixNAP:~$ hadoop fs -chmod g+w /tmp
hdoop@phoenixNAP:~$ hadoop fs -ls /
Found 1 items
drwxrwxr-x   - hdoop supergroup          0 2024-09-03 03:59 /tmp
```

Create /user/hive/warehouse Directory

Create the *warehouse* subdirectory within the */user/hive/* parent directory:

1. Create the directories one by one. Start with the */user* directory:

hadoop fs -mkdir /user

2. Make the */user/hive* directory:

hadoop fs -mkdir /user/hive

3. Create the */user/hive/warehouse* directory:

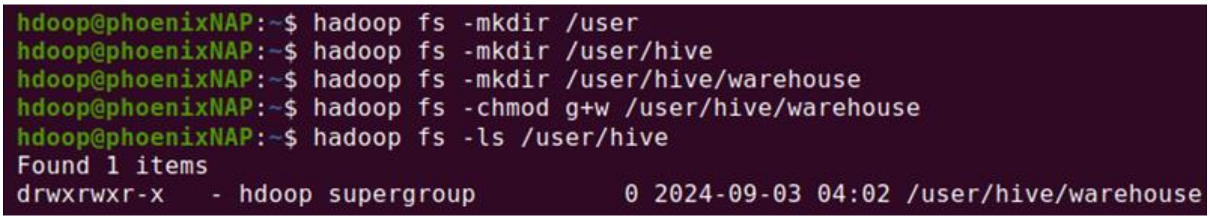
hadoop fs -mkdir /user/hive/warehouse

4. Add **write** and **execute** permissions to group members:

hadoop fs -chmod g+w /user/hive/warehouse

5. Check if the permissions applied correctly:

hadoop fs -ls /user/hive



```
hdoop@phoenixNAP:~$ hadoop fs -mkdir /user
hdoop@phoenixNAP:~$ hadoop fs -mkdir /user/hive
hdoop@phoenixNAP:~$ hadoop fs -mkdir /user/hive/warehouse
hdoop@phoenixNAP:~$ hadoop fs -chmod g+w /user/hive/warehouse
hdoop@phoenixNAP:~$ hadoop fs -ls /user/hive
Found 1 items
drwxrwxr-x   - hdoop supergroup          0 2024-09-03 04:02 /user/hive/warehouse
```

Step 5: Configure hive-site.xml File (Optional)

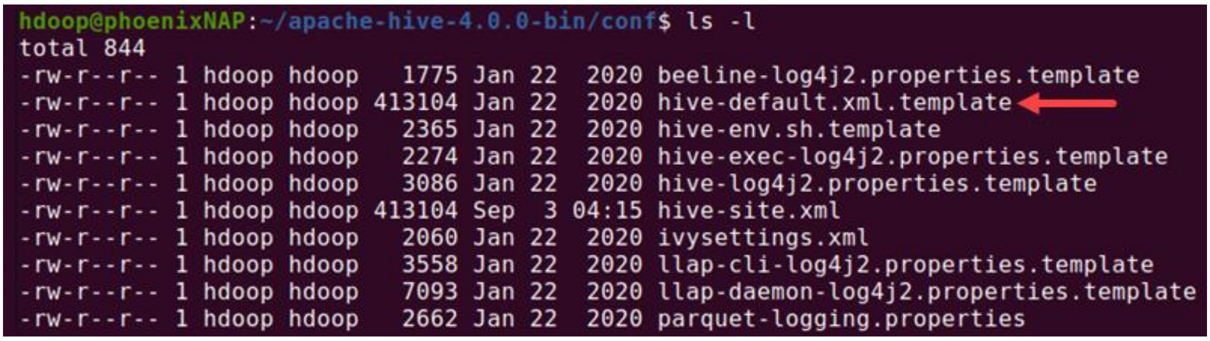
Apache Hive distributions contain template configuration files by default. The template files are located within the Hive *conf* directory and outline default Hive settings:

1. Navigate to the */conf* directory in the Hive installation:

cd $HIVE_HOME/conf

2. List the files contained in the folder using the ls command:

ls -l



```
hdoop@phoenixNAP:~/apache-hive-4.0.0-bin/conf$ ls -l
total 844
-rw-r--r-- 1 hdoop hdoop   1775 Jan 22  2020 beeline-log4j2.properties.template
-rw-r--r-- 1 hdoop hdoop 413104 Jan 22  2020 hive-default.xml.template
-rw-r--r-- 1 hdoop hdoop   2365 Jan 22  2020 hive-env.sh.template
-rw-r--r-- 1 hdoop hdoop   2274 Jan 22  2020 hive-exec-log4j2.properties.template
-rw-r--r-- 1 hdoop hdoop   3086 Jan 22  2020 hive-log4j2.properties.template
-rw-r--r-- 1 hdoop hdoop 413104 Sep  3 04:15 hive-site.xml
-rw-r--r-- 1 hdoop hdoop   2060 Jan 22  2020 ivysettings.xml
-rw-r--r-- 1 hdoop hdoop   3558 Jan 22  2020 llap-cli-log4j2.properties.template
-rw-r--r-- 1 hdoop hdoop   7093 Jan 22  2020 llap-daemon-log4j2.properties.template
-rw-r--r-- 1 hdoop hdoop   2662 Jan 22  2020 parquet-logging.properties
```

Locate the *hive-default.xml.template* file.

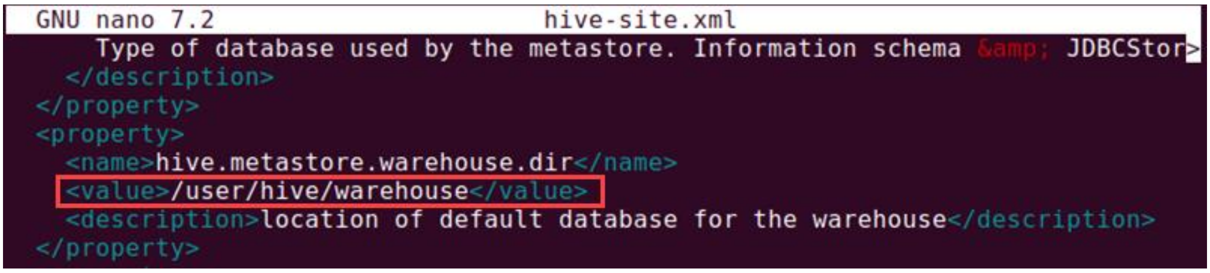
3. Create a copy of the file and change its extension using the cp command:

cp hive-default.xml.template hive-site.xml

4. Open the *hive-site.xml* file using **nano**:

nano hive-site.xml

5. Configure the system to use the local storage.



```
 GNU nano 7.2                          hive-site.xml
      Type of database used by the metastore. Information schema &amp; JDBCStor>
    </description>
  </property>
  <property>
    <name>hive.metastore.warehouse.dir</name>
    <value>/user/hive/warehouse</value>
    <description>location of default database for the warehouse</description>
  </property>
```

Set the **hive.metastore.warehouse.dir** parameter value to the Hive warehouse directory (*/user/hive/warehouse*).

6. Save the file and close **nano**.

Step 6: Initiate Derby Database

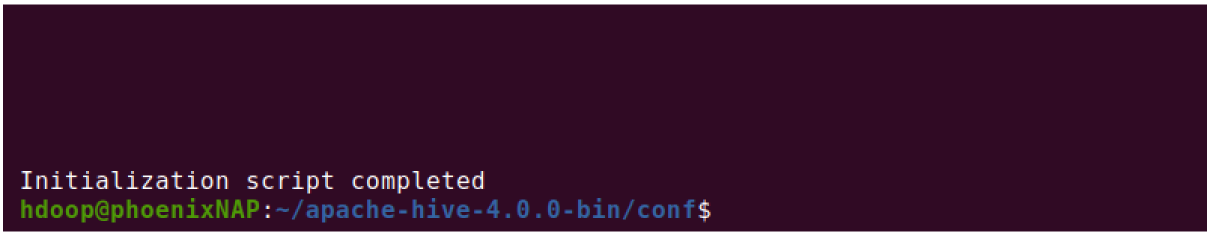
Apache Hive uses the Derby database to store metadata. Initiate the Derby database from the Hive *bin* directory:

1. Navigate to the Hive base directory:

cd $HIVE_HOME

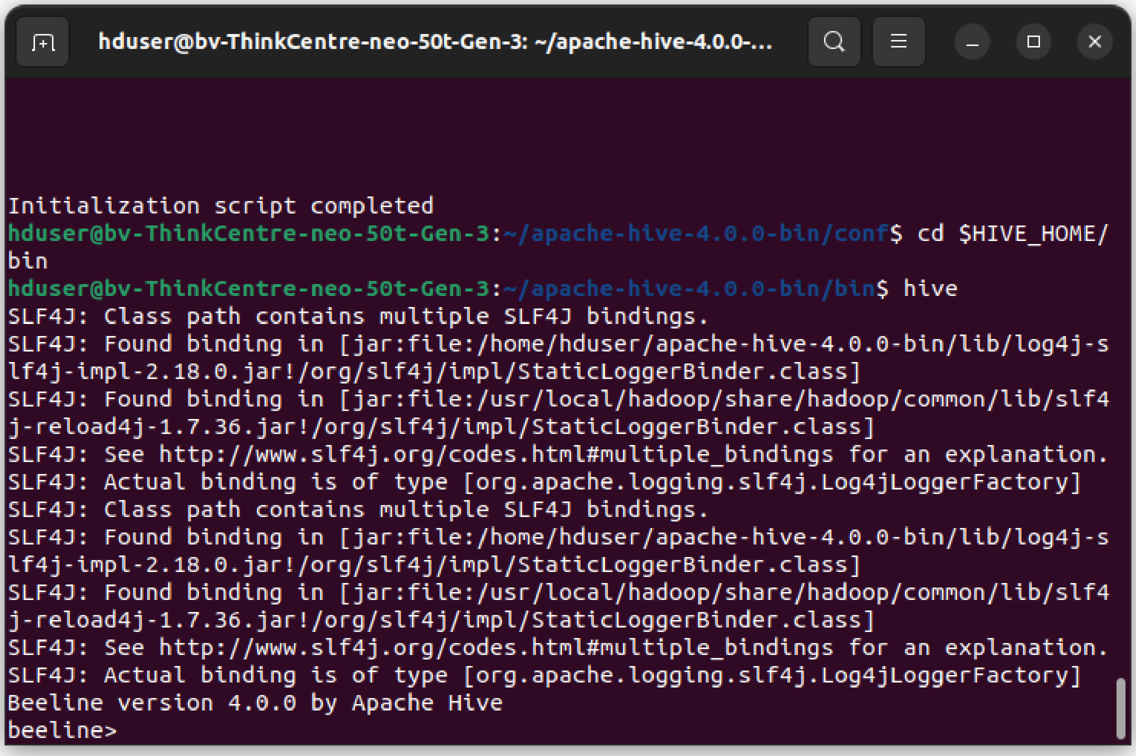
2. Use the **schematool** command from the */bin* directory:

bin/schematool -dbType derby -initSchema



```
Initialization script completed
hdoop@phoenixNAP:~/apache-hive-4.0.0-bin/conf$
```

The process takes a few moments to complete.



hduser@bv-ThinkCentre-neo-50t-Gen-3: ~/apache-hive-4.0.0-...

```
Initialization script completed
hduser@bv-ThinkCentre-neo-50t-Gen-3:~/apache-hive-4.0.0-bin/conf$ cd $HIVE_HOME/
bin
hduser@bv-ThinkCentre-neo-50t-Gen-3:~/apache-hive-4.0.0-bin/bin$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hduser/apache-hive-4.0.0-bin/lib/log4j-s
lf4j-impl-2.18.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4
j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hduser/apache-hive-4.0.0-bin/lib/log4j-s
lf4j-impl-2.18.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4
j-reload4j-1.7.36.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Beeline version 4.0.0 by Apache Hive
beeline>
```