

Hue User Guide

- [Introducing Hue](#)
- [About Hue](#)
- [Beeswax Hive Editor](#)
- [Impala Query UI](#)
- [Pig Editor](#)
- [File Browser](#)
- [Metastore Manager](#)
- [Sqoop UI](#)
- [Job Browser](#)
- [Job Designer](#)
- [Oozie Editor and Dashboard](#)
- [Solr Search](#)
- [ZooKeeper Browser](#)
- [HBase Browser](#)
- [User Admin](#)

Introducing Hue

Hue is a set of web applications that enable you to interact with a Hadoop cluster. Hue applications let you browse HDFS and jobs, manage a Hive metastore, run Hive, Cloudera Impala queries and Pig scripts, browse HBase, export data with Sqoop, submit MapReduce programs, build custom search engines with Solr, and schedule repetitive workflows with Oozie.

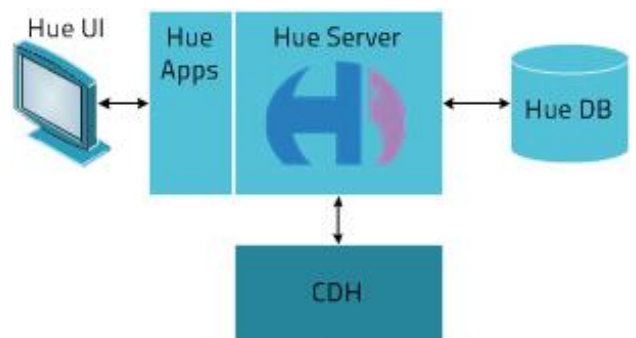
Hue Architecture

Hue applications run in a Web browser and require no client installation.

The following figure illustrates how Hue works. Hue Server is a "container" web application that sits in between CDH and the browser. It hosts all the Hue web applications and communicates with CDH components.

Starting Applications

To open a Hue application, click the appropriate icon in the navigation bar at the top of the Hue web browser window. To open a second application concurrently (or a second instance of the same application), right-click the icon and select **Open link in new tab**.



Displaying Help for the Hue Applications

To display the help text for a Hue application, click the **Help** () tab in the Hue navigation bar, then click the appropriate link in the Help navigation bar at the top of the Help window.



Logging In and Out

To log out of Hue, click **Sign Out** from the pull-down list under the logged-in user name (at the right of the Hue navigation bar).

Notice of Misconfiguration

If Hue detects a misconfiguration, an indicator appears in the navigation bar at the top of the page. Clicking this indicator takes you to the [Check Configuration](#) screen which will indicate the potential misconfiguration(s) with hints about fixing them.



Changing Your Password

You can go directly to your own information by selecting username > Profile at the right of the Hue navigation bar.

If authentication is managed by Hue (that is, authentication is not managed via some external mechanism), and you are a superuser, you can use the [User Admin](#) application to change password and personal details.

Seeking Help, Reporting Bugs, and Providing Feedback

The Hue team strongly values your feedback. The best way to contact us is to send email to hue-user@cloudera.org.

If you're experiencing transient errors (typically an error message saying a service is down), contact your system

administrator first.

About Hue

The **About Hue** application displays the version of Hue you are running. If you are a superuser, it lets you perform Hue setup tasks, and lets you view configuration and logs.

Starting About Hue

To start the About Hue application, click in the navigation bar at the top of the Hue browser page. It opens to the Quick Start Wizard screen.



Quick Start Wizard

The Quick Start wizard allows you to perform the following Hue setup operations by clicking the tab of each step or sequentially by clicking Next in each screen:

1. **Check Configuration** validates your Hue configuration. It will note any potential misconfiguration and provide hints as to how to fix them. You can edit the configuration file described in the next section or use Cloudera Manager, if installed, to manage your changes.
2. **Examples** contains links to install examples into the Beeswax, Cloudera Impala, Metastore Manager, Job Designer, Oozie Editor/Dashboard, and Pig Editor applications.
3. **Users** contains a link to the User Admin application to create or import users and a checkbox to enable and disable collection of usage information.
4. **Go!** - displays the Hue home screen, which contains links to the different categories of applications supported by Hue: Query, Hadoop, and Workflow.

Configuration

Displays a list of the installed Hue applications and their configuration. The location of the folder containing the Hue configuration files is shown at the top of the page. Hue configuration settings are in the hue.ini configuration file.

Click the tabs under **Configuration Sections and Variables** to see the settings configured for each application. For information on configuring these settings, see Hue Configuration in the Hue installation manual.

Server Logs

Displays the Hue Server log and allows you to download the log to your local system in a zip file.

Beeswax

The Beeswax application enables you to perform queries on Apache Hive, a data warehousing system designed to work with Hadoop. For information about Hive, see [Hive Documentation](#). You can create Hive databases, tables and partitions, load data, create, run, and manage queries, and download the results in a Microsoft Office Excel worksheet file or a comma-separated values file.

Beeswax and Hive Installation and Configuration

Beeswax is installed and configured as part of Hue. For information about installing and configuring Hue, see the Hue Installation manual.

Beeswax assumes an existing Hive installation. The Hue installation instructions include the configuration necessary for Beeswax to access Hive. You can view the current Hive configuration from the **Settings** tab in the Beeswax application.

By default, a Beeswax user can see the saved queries for all users - both his/her own queries and those of other Beeswax users. To restrict viewing saved queries to the query owner and Hue administrators, set the `share_saved_queries` property under the `[beeswax]` section in the Hue configuration file to `false`.

Starting Beeswax

Click the **Beeswax** icon () in the navigation bar at the top of the Hue browser page.


Managing Databases, Tables, and Partitions

You can create databases, tables, partitions, and load data by executing [Hive data manipulation statements](#) in the Beeswax application.

You can also use the [Metastore Manager](#) application to manage the databases, tables, and partitions and load data.

Installing Example Queries and Tables

Note: You must be a superuser to perform this task.

1. Click  . The Quick Start Wizard opens.
2. Click **Step 2: Examples**.
3. Click **Beeswax (Hive UI)**.

Query Editor

The Query Editor view lets you create, save, and submit queries in the [Hive Query Language \(HQL\)](#), which is similar to Structured Query Language (SQL). When you submit a query, the Beeswax Server uses Hive to run the queries. You can either wait for the query to complete, or return later to find the queries in the History view. You can also request to receive an email message after the query is completed.

In the box to the left of the Query field, you can select a database, override the default Hive and Hadoop settings, specify file resources and user-defined functions, enable users to enter parameters at run-time, and request email notification when the job is complete. See [Advanced Query Settings](#) for details on using these settings.

Creating Queries

1. In the Query Editor window, type a query or multiple queries separated by a semicolon ";". To be presented with a drop-down of autocomplete options, type CTRL+spacebar when entering a query.
2. To save your query and advanced settings to use again later, click **Save As**, enter a name and description, and then click **OK**. To save changes to an existing query, click **Save**.
3. If you want to view the execution plan for the query, click **Explain**. For more information, see <http://wiki.apache.org/hadoop/Hive/LanguageManual/Explain>.

Loading Queries into the Query Editor

1.
 - Click the My Queries tab.
 1. Click the Recent Saved Queries or Recent Run Queries tab to display the respective queries.
- Do one of the following:
- Click the Saved Queries tab.
2. Click a query name. The query is loaded into the Query Editor.

Running Queries

Note: To run a query, you must be logged in to Hue as a user that also has a Unix user account on the remote server. ■

1. To execute a portion of the query, highlight one or more query statements.
2. Click **Execute**. The Query Results window appears with the results of your query.
 - To view a log of the query execution, click **Log** at the top of the results display. You can use the information in this tab to debug your query.
 - To view the query that generated these results, click **Query** at the top of the results display.
 - To view the columns of the query, click **Columns**.
 - To return to the query in the Query Editor, click **Unsaved Query**.
3. If there are multiple statements in the query, click Next in the Multi-statement query pane to execute the remaining statements.

Note: Under **MR JOBS**, you can view any MapReduce jobs that the query generated. ■

Downloading and Saving Query Results



Important: ■

- You can only save results to a file when the results were generated by a MapReduce job.
- This is the preferred way to save when the result is large (for example > 1M rows).

- Do any of the following to download or save the query results:
 - Click **Download as CSV** to download the results in a comma-separated values file suitable for use in other applications.
 - Click **Download as XLS** to download the results in a Microsoft Office Excel worksheet file.
 - Click **Save** to save the results in a table or HDFS file.
 - To save the results in a new table, select **In a new table**, enter a table name, and then click **Save**.
 - To save the results in an HDFS file, select **In an HDFS directory**, enter a path and then click **Save**. You can then download the file with [File Browser](#).

Advanced Query Settings

The pane to the left of the Query Editor lets you specify the following options:

DATABASE	The database containing the table definitions.
SETTINGS	<p>Override the Hive and Hadoop default settings. To configure a new setting:</p> <ol style="list-style-type: none"> 1. Click Add. 2. For Key, enter a Hive or Hadoop configuration variable name. 3. For Value, enter the value you want to use for the variable. For example, to override the directory where structured Hive query logs are created, you would enter hive.querylog.location for Key, and a path for Value. <p>To view the default settings, click the Settings tab at the top of the page. For information about Hive configuration variables, see: [http://wiki.apache.org/hadoop/Hive/AdminManual/Configuration] (http://wiki.apache.org/hadoop/Hive/AdminManual/Configuration). For information about Hadoop configuration variables, see: [http://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/mapred-default.xml] (http://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/mapred-default.xml).</p>
FILE RESOURCES	<p>Make files locally accessible at query execution time available on the Hadoop cluster. Hive uses the Hadoop Distributed Cache to distribute the added files to all machines in the cluster at query execution time.</p> <ol style="list-style-type: none"> 1. Click Add to configure a new setting. 2. From the Type drop-down menu, choose one of the following: <ul style="list-style-type: none"> ◦ jar - Adds the specified resources to the Java classpath. ◦ archive - Unarchives the specified resources when distributing them. ◦ file - Adds the specified resources to the distributed cache. Typically, this might be a transform script (or similar) to be executed. ◦ For Path, enter the path to the file or click  to browse and select the file. <p> Note: It is not necessary to specify files used in a transform script if the files are available in the same path on all machines in the Hadoop cluster.</p>

USER-DEFINED FUNCTIONS	Specify user-defined functions. Click Add to configure a new setting. Specify the function name in the Name field, and specify the class name for Classname. You <i>must</i> specify a JAR file for the user-defined functions in FILE RESOURCES. To include a user-defined function in a query, add a \$ (dollar sign) before the function name in the query. For example, if MyTable is a user-defined function name in the query, you would type: SELECT \$MyTable
PARAMETERIZATION	Indicate that a dialog box should display to enter parameter values when a query containing the string \$parametername is executed. Enabled by default.
EMAIL NOTIFICATION	Indicate that an email message should be sent after a query completes. The email is sent to the email address specified in the logged-in user's profile.

Viewing Query History

You can view the history of queries that you have run previously. Results for these queries are available for one week or until Hue is restarted.

1. Click **History**. A list of your saved and unsaved queries displays in the Query History window.
2. To display the queries for all users, click **Show everyone's queries**. To display your queries only, click **Show my queries**.
3. To display the automatically generated actions performed on a user's behalf, click **Show auto actions**. To display user queries again, click **Show user queries**.

Viewing, Editing, Copying, and Deleting Saved Queries

You can view a list of saved queries of all users by clicking **My Queries** and then selecting either Recent Saved Queries or Recent Run Queries tab to display the respective queries or clicking **Saved Queries**. You can copy any query, but you can edit, delete, and view the history of only your own queries.

Edit

1. Click **Saved Queries**. The Queries window displays.
2. Check the checkbox next to the query and click **Edit**. The query displays in the Query Editor window.
3. Change the query and then click **Save**. You can also click **Save As**, enter a new name, and click **OK** to save a copy of the query.

Copy

1. Click **Saved Queries**. The Queries window displays.
2. Check the checkbox next to the query and click **Copy**. The query displays in the Query Editor window.
3. Change the query as necessary and then click **Save**. You can also click **Save As**, enter a new name, and click **OK** to save a copy of the query.

Copy in Query History

1. Click **History**. The Query History window displays.
2. To display the queries for all users, click **Show everyone's queries**. The queries for all users display in the History window.
3. Click the query you want to copy. A copy of the query displays in the Query Editor window.
4. Change the query, if necessary, and then click **Save As**, enter a new name, and click **OK** to save the query.

Delete

1. Click **Saved Queries**. The Queries window displays.
2. Check the checkbox next to the query and click **Delete**.
3. Click **Yes** to confirm the deletion.

Cloudera Impala Query UI

The Cloudera Impala Query UI application enables you to perform queries on Apache Hadoop data stored in HDFS or HBase using Cloudera Impala. For information about Cloudera Impala, see [Installing and Using Cloudera Impala](#). You can create, run, and manage queries, and download the results in a Microsoft Office Excel worksheet file or a comma-separated values file.

Cloudera Impala Query UI Installation and Configuration

The Cloudera Impala Query UI application is one of the applications installed as part of Hue. For information about installing and configuring Hue, see the Hue Installation manual..

The Cloudera Impala Query UI assumes an existing Cloudera Impala installation. The Hue installation instructions include the configuration necessary for Impala. You can view the current configuration from the **Settings** tab.

Starting Cloudera Impala Query UI

Click the **Cloudera Impala Query UI** icon () in the navigation bar at the top of the Hue browser page.



Managing Databases, Tables, and Partitions

You can create databases, tables, partitions, and load data by executing [Hive data manipulation statements](#) in the Beeswax application.

You can also use the [Metastore Manager](#) application to manage the databases, tables, and partitions and load data.

When you change the metastore using one of these applications, you must click the Refresh button under METASTORE CATALOG in the pane to the left of the Query Editor to make the metastore update visible to the Cloudera Impala server.

Installing Example Queries and Tables

Note: You must be a superuser to perform this task.

1. Click . The Quick Start Wizard opens.
2. Click **Step 2: Examples**.
3. Click **Cloudera Impala Query UI**.



Query Editor

The Query Editor view lets you create queries in the Cloudera Impala Query Language, which is based on the Hive Standard Query Language (HiveQL) and described in the Cloudera Impala Language Reference topic in [Installing and Using Cloudera Impala](#).

You can name and save your queries to use later.

When you submit a query, you can either wait for the query to complete, or return later to find the queries in the **History** view.

In the box to the left of the Query field, you can select a database, override the default Cloudera Impala settings,

enable users to enter parameters at run-time. See [Advanced Query Settings](#) for details on using these settings.

Creating Queries

1. In the Query Editor window, type a query or multiple queries separated by a semicolon ";". To be presented with a drop-down of autocomplete options, type CTRL+spacebar when entering a query.
2. To save your query and advanced settings to use again later, click **Save As** , enter a name and description, and then click **OK**. To save changes to an existing query, click **Save**.


Loading Queries into the Query Editor

1.
 - Click the My Queries tab.
 1. Click the Recent Saved Queries or Recent Run Queries tab to display the respective queries.

Do one of the following:

- Click the Saved Queries tab.
2. Click a query name. The query is loaded into the Query Editor.

Running Queries

Note: To run a query, you must be logged in to Hue as a user that also has a Unix user account on the remote server. 

1. To execute a portion of the query, highlight one or more query statements.
2. Click **Execute**. The Query Results window appears with the results of your query.
 - To view a log of the query execution, click **Log** at the top of the results display. You can use the information in this tab to debug your query.
 - To view the query that generated these results, click **Query** at the top of the results display.
 - To view the columns of the query, click **Columns**.
 - To return to the query in the Query Editor, click **Unsaved Query**.
3. If there are multiple statements in the query, click Next in the Multi-statement query pane to execute the remaining statements.

Advanced Query Settings

The pane to the left of the Query Editor lets you specify the following options:

DATABASE	The database containing the table definitions.
SETTINGS	<div>Override the Cloudera Impala default settings. To configure a new setting:<ol style="list-style-type: none">1. Click Add.2. For Key, enter a Hive or Hadoop configuration variable name.3. For Value, enter the value you want to use for the variable. For example, to override the directory where structured Hive query logs are created, you would enter hive.querylog.location for Key, and a path for Value.</div> <div>To view the default settings, click the Settings tab at the top of the page.</div>

PARAMETERIZATION	Indicate that a dialog box should display to enter parameter values when a query containing the string \$parametername is executed. Enabled by default.
METASTORE CATALOG	Refresh metadata. It is best to refresh metadata after making changes to databases such as adding or dropping a table.

Viewing Query History

You can view the history of queries that you have run previously. Results for these queries are available for one week or until Hue is restarted.

1. Click **History**. A list of your saved and unsaved queries displays in the Query History window.
2. To display the queries for all users, click **Show everyone's queries**. To display your queries only, click **Show my queries**.
3. To display the automatically generated actions performed on a user's behalf, click **Show auto actions**. To display user queries again, click **Show user queries**.

Viewing, Editing, Copying, and Deleting Saved Queries

You can view a list of saved queries of all users by clicking **My Queries** and then selecting either Recent Saved Queries or Recent Run Queries tab to display the respective queries or clicking **Saved Queries**. You can copy any query, but you can edit, delete, and view the history of only your own queries.

Edit

1. Click **Saved Queries**. The Queries window displays.
2. Check the checkbox next to the query and click **Edit**. The query displays in the Query Editor window.
3. Change the query and then click **Save**. You can also click **Save As**, enter a new name, and click **OK** to save a copy of the query.

Copy

1. Click **Saved Queries**. The Queries window displays.
2. Check the checkbox next to the query and click **Copy**. The query displays in the Query Editor window.
3. Change the query as necessary and then click **Save**. You can also click **Save As**, enter a new name, and click **OK** to save a copy of the query.

Copy in Query History

1. Click **History**. The Query History window displays.
2. To display the queries for all users, click **Show everyone's queries**. The queries for all users display in the History window.
3. Click the query you want to copy. A copy of the query displays in the Query Editor window.
4. Change the query, if necessary, and then click **Save As**, enter a new name, and click **OK** to save the query.

Delete

1. Click **Saved Queries**. The Queries window displays.
2. Check the checkbox next to the query and click **Delete**.
3. Click **Yes** to confirm the deletion.

Pig Editor


The Pig Editor application allows you to define Pig scripts, run scripts, and view the status of jobs. For information about Pig, see [Pig Documentation](#).

Pig Editor Installation and Configuration

Pig Editor is one of the applications installed as part of Hue. For information about installing and configuring Hue, see the Hue Installation manual.

Pig Editor assumes an existing Pig installation. The Hue installation instructions include the configuration necessary for Pig Editor to access Pig.

Starting Pig Editor


Click the **Pig Editor** icon () in the navigation bar at the top of the Hue browser page. The Pig Editor opens with three tabs:

- Editor - editor where you can create, edit, run, save, copy, and delete scripts and edit script properties.
- Scripts - script manager where you can create, open, run, copy, and delete scripts.
- Dashboard - dashboard where you can view running and completed scripts and view the log of a job.

Pig Scripts

Installing the Example Scripts

Note: You must be a superuser to perform this task.

1. Click  . The Quick Start Wizard opens.
2. Click **Step 2: Examples**.
3. Click **Pig Editor**.

Creating a Script

1. In either the Editor or Scripts screen, click New script. Edit the script as desired.
2. Click Edit Properties. In the Script name field, type a name for the script.
3. Click Save.

Opening a Script

1. Click the Scripts tab.
2. In the list of scripts, click a script.

Running a Script

1. Do one of the following:
 - In the Editor screen, click the Run button.
 - In the Scripts screen, check the checkbox next to a script and click the Run button.

Viewing the Result of Running a Script

1. Click the Dashboard tab.
2. Click a job.

File Browser

The File Browser application lets you browse and manipulate files and directories in the Hadoop Distributed File System (HDFS) while using Hue. With File Browser, you can:

- Create files and directories, upload and download files, upload zip archives, and rename, move, and delete files and directories. You can also change a file's or directory's owner, group, and permissions. See [Files and Directories](#).
- Search for files, directories, owners, and groups. See [Searching for Files and Directories](#).
- View and edit files as text or binary. See [Viewing and Editing Files](#).

File Browser Installation and Configuration

File Browser is one of the applications installed as part of Hue. For information about installing and configuring Hue, see the Hue Installation manual.

Starting File Browser

Click the **File Browser** icon () in the navigation bar at the top of the Hue browser page.

Files and Directories

You can use File Browser to view the input and output files of your MapReduce jobs. Typically, you can save your output files in /tmp or in your home directory if your system administrator set one up for you. You must have the proper permissions to manipulate other user's files.

Creating Directories

1. In the File Browser window, select **New > Directory**.
2. In the **Create Directory** dialog box, enter a directory name and then click **Submit**.

Changing Directories

- Click the directory name or parent directory dots in the **File Browser** window.
- Click the icon, type a directory name, and press **Enter**.

To change to your home directory, click **Home** in the path field at the top of the **File Browser** window.

Note:

The **Home** button is disabled if you do not have a home directory. Ask a Hue administrator to create a home directory for you.

Creating Files

1. In the File Browser window, select **New > File**.
2. In the **Create File** dialog box, enter a file name and then click **Submit**.

Uploading Files

You can upload text and binary files to the HDFS.

1. In the **File Browser** window, browse to the directory where you want to upload the file.
2. Select **Upload > Files**.
3. In the box that opens, click **Upload a File** to browse to and select the file(s) you want to upload, and then click **Open**.

Copying a File

1. In the **File Browser** window, check the checkbox next to the file you want to copy.
2. Click the Copy button.



Downloading Files

You can download text and binary files to the HDFS.

1. In the **File Browser** window, check the checkbox next to the file you want to download.
2. Click the **Download** button.

Uploading Zip Archives

You can upload zip archives to the HDFS. The archive is uploaded and extracted to a directory named archivename.

1. In the **File Browser** window, browse to the directory where you want to upload the archive.
2. Select **Upload > Zip file**.
3. In the box that opens, click **Upload a zip file** to browse to and select the archive you want to upload, and then click **Open**.

Trash Folder

File Browser supports the HDFS trash folder (*home directory/.Trash*) to contain files and directories before they are permanently deleted. Files in the folder have the full path of the deleted files (in order to be able to restore them if needed) and checkpoints. The length of time a file or directory stays in the trash depends on HDFS properties.

Open

1. In the **File Browser** window, click .



Move Files and Directories To

1. In the **File Browser** window, check the checkbox next to one or more files and directories.
2. Select **Delete > Move to trash**.

Empty

1. In the **File Browser** window, click .
2. Click **Empty**.



Renaming, Moving, Deleting, and Restoring Files and Directories

Rename

1. In the **File Browser** window, check the checkbox next to the file or directory you want to rename.
2. Click the **Rename** button.
3. Enter the new name and then click **Submit**.

Move

1. In the **File Browser** window, check the checkbox next to the file or directory you want to move.
2. Click the **Move** button.
3. In the **Move** dialog box, browse to or type the new directory, and then click **Submit**.

Delete

1. In the **File Browser** window, check the checkbox next to the file or directory you want to delete. If you select a directory, all of the files and subdirectories contained within that directory are also deleted.
2. Choose one of the following:
 - **Delete > Move to trash**
 - **Delete > Delete forever**
3. Click **Yes** to confirm. When you move a file to trash it is stored in the `.Trash` folder in your home directory.

Restore

1. In the **File Browser** window, open the `.Trash` folder.
2. Navigate to the folder containing the file you want to restore.
3. Check the checkbox next to the file.
4. Click **Restore**.

Changing a File's or Directory's Owner, Group, or Permissions

Note:

Only the Hadoop superuser can change a file's or directory's owner, group, or permissions. The user who starts Hadoop is the Hadoop superuser. The Hadoop superuser account is not necessarily the same as a Hue superuser account. If you create a Hue user (in User Admin) with the same user name and password as the Hadoop superuser, then that Hue user can change a file's or directory's owner, group, or permissions.

Owner or Group

1. In the **File Browser** window, check the checkbox next to the select the file or directory whose owner or group you want to change.
2. Choose **Change Owner/Group** from the Options menu.
3. In the **Change Owner/Group** dialog box:
 - Choose the new user from the **User** drop-down menu.
 - Choose the new group from the **Group** drop-down menu.
 - Check the **Recursive** checkbox to propagate the change.
4. Click **Submit** to make the changes.

Permissions

1. In the **File Browser** window, check the checkbox next to the file or directory whose permissions you want to change.
2. Click the **Change Permissions** button.
3. In the **Change Permissions** dialog box, select the permissions you want to assign and then click **Submit**.

Searching for Files and Directories

To search for files or directories by name using the query search box, enter the name of the file or directory in the query search box. File Browser lists the files or directories matching the search criteria.

Viewing and Editing Files

You can view and edit files as text or binary.

View

1. In the **File Browser** window, click the file you want to view. File Browser displays the first 4,096 bytes of the file in the **File Viewer** window.
 - If the file is larger than 4,096 bytes, use the Block navigation buttons (First Block, Previous Block, Next Block, Last Block) to scroll through the file block by block. The **Viewing Bytes** fields show the range of bytes you are currently viewing.
 - To switch the view from text to binary, click **View as Binary** to view a hex dump.
 - To switch the view from binary to text, click **View as Text**.

Edit

1. If you are viewing a text file, click **Edit File**. File Browser displays the contents of the file in the **File Editor** window.
2. Edit the file and then click **Save** or **Save As** to save the file.

View Location in HDFS

Click **View File Location**. File Browser displays the file's location in the **File Browser** window.

Metastore Manager

The Metastore Manager application enables you to manage the databases, tables, and partitions of the [Hive](#) metastore shared by the ([Beeswax](#) and [Cloudera Impala Query UI](#)) applications. You can use Metastore Manager to perform the following operations:

- Databases
- Tables

Metastore Manager Installation and Configuration

Metastore Manager is one of the applications installed as part of Hue. For information about installing and configuring Hue, see the Hue Installation manual.


Starting Metastore Manager

Click the **Metastore Manager** icon () in the navigation bar at the top of the Hue browser page.



Installing Sample Tables

Note: You must be a superuser to perform this task.

1. Click . The Quick Start Wizard opens.
2. Click **Step 2: Examples**.
3. Click **Beeswax (Hive UI)** or **Cloudera Impala Query UI**.



Importing Data

If you want to import your own data instead of installing the sample tables, follow the procedure in [Creating Tables](#).

Selecting a Database

1. In the pane on the left, select the database from the DATABASE drop-down list.

Creating a Database

1. Click .

 Databases

2.

1. Specify a database name and optional description. Database names are not case-sensitive. Click **Next**.
2. Do one of the following:
 - Keep the default location in the Hive warehouse folder.
 - Specify an external location within HDFS:
 1. Uncheck the **Location** checkbox.
 2. In the External location field, type a path to a folder on HDFS or click to browse to a folder and click **Select this folder**.



Click **Create a new database**.

1. Click the **Create Database** button.

Dropping Databases

1. Click .
2. In the list of databases, check the checkbox next to one or more databases.
3. Click the Drop button.
4. Confirm whether you want to delete the databases.

 **Databases**



Creating Tables

Although you can create tables by executing the appropriate Hive HQL DDL query commands, it is easier to create a table using the Metastore Manager table creation wizard.

There are two ways to create a table: from a file or manually. If you create a table from a file, the format of the data in the file will determine some of the properties of the table, such as the record and file formats. The data from the file you specify is imported automatically upon table creation. When you create a file manually, you specify all the properties of the table, and then execute the resulting query to actually create the table. You then import data into the table as an additional step.

From a File

1. In the ACTIONS pane in the Metastore Manager window, click **Create a new table from a file**. The table creation wizard starts.
2. Follow the instructions in the wizard to create the table. The basic steps are:
 - Choose your input file. The input file you specify must exist. Note that you can choose to have Beeswax create the table definition only based on the import file you select, without actually importing data from that file.
 - Specify the column delimiter.
 - Define your columns, providing a name and selecting the type.
3. Click **Create Table** to create the table. The new table's metadata displays on the right side of the **Table Metadata** window. At this point, you can view the metadata or a sample of the data in the table. From the ACTIONS pane you can import new data into the table, browse the table, drop it, or go to the File Browser to see the location of the data.

Manually

1. In the ACTIONS pane in the Metastore Manager window, click **Create a new table manually**. The table creation wizard starts.
2. Follow the instructions in the wizard to create the table. The basic steps are:
 - Name the table.
 - Choose the record format.
 - Configure record serialization by specifying delimiters for columns, collections, and map keys.
 - Choose the file format.
 - Specify the location for your table's data.
 - Specify the columns, providing a name and selecting the type for each column.
 - Specify partition columns, providing a name and selecting the type for each column.
3. Click **Create table**. The Table Metadata window displays.

Browsing Tables

To browse table data:

In the Table List window, check the checkbox next to a table name and click **Browse Data**. The table's data displays in the Query Results window.

To browse table metadata:

Do one of the following:

- In the Table List window, click a table name.
- Check the checkbox next to a table name and click **View**.
- The table's metadata displays in the **Columns** tab. You can view the table data by selecting the **Sample** tab.
- If the table is partitioned, you can view the partition columns by clicking the **Partition Columns** tab and display the partitions by clicking **Show Partitions(n)**, where n is the number of partitions in the ACTIONS pane on the left.

Importing Data into a Table

When importing data, you can choose to append or overwrite the table's data with data from a file.

1. In the Table List window, click the table name. The Table Metadata window displays.
2. In the ACTIONS pane, click **Import Data**.
3. For **Path**, enter the path to the file that contains the data you want to import.
4. Check **Overwrite existing data** to replace the data in the selected table with the imported data. Leave unchecked to append to the table.
5. Click **Submit**.

Dropping Tables

1. In the Table List window, click the table name. The Table Metadata window displays.
2. In the ACTIONS pane, click **Drop Table**.
3. Click **Yes** to confirm the deletion.

Viewing a Table's Location

1. In the Table List window, click the table name. The Table Metadata window displays.
2. Click **View File Location**. The file location of the selected table displays in its directory in the File Browser window.

Sqoop UI

The Sqoop UI enables transferring data from a relational database to Hadoop and vice versa. The UI lives uses Apache Sqoop to do this. See the [Sqoop Documentation](#) for more details on Sqoop.

Installation and Configuration

The Sqoop UI is one of the applications installed as part of Hue. For information about installing and configuring Hue, see the Hue Installation manual.

Starting

Click the **Sqoop** icon () in the navigation bar at the top of the Hue browser page.



Sqoop Jobs

Sqoop UI is oriented around jobs in Apache Sqoop.

Creating a New Job

1. Click the **New job** button at the top right.
2. In the Name field, enter a name.
3. Choose the type of job: import or export. The proceeding form fields will change depending on which type is chosen.
4. Select a connection, or create one if it does not exist.
5. Fill in the rest of the fields for the job. For importing, the "Table name", "Storage type", "Output format", and "Output directory" are necessary at a minimum. For exporting, the "Table name" and "Input directory" are necessary at a minimum.
6. Click **save** to finish.

Editing a Job

1. In the list of jobs, click on the name of the job.
2. Edit the desired form fields in the job.

Copying a Job

1. In the list of jobs, click on the name of the job.
2. On the left hand side of the job editor, there should be a panel containing actions. Click **Copy**.

Removing a Job

1. In the list of jobs, click on the name of the job.
2. On the left hand side of the job editor, there should be a panel containing actions. Click **Delete**.

Running a Job

There's a status on each of the items in the job list indicating the last time a job was ran. The progress of the job should dynamically update. There's a progress bar at the bottom of each item on the job list as well.

1. In the list of jobs, click on the name of the job.
2. On the left hand side of the job editor, there should be a panel containing actions. Click **Run**.

Creating a New Connection

1. Click the **New job** button at the top right.
2. At the connection field, click the link titled **Add a new connection**.
3. Fill in the displayed fields.
4. Click **save** to finish.

Editing a Connection

1. Click the **New job** button at the top right.
2. At the connection field, select the connection by name that should be edited.
3. Click **Edit**.
4. Edit the any of the fields.
5. Click **save** to finish.

Removing a Connection

1. Click the **New job** button at the top right.
2. At the connection field, select the connection by name that should be deleted.
3. Click **Delete**.

NOTE: If this does not work, it's like because a job is using that connection. Make sure not jobs are using the connection that will be deleted.

Filtering Sqoop Jobs

The text field in the top, left corner of the Sqoop Jobs page enables fast filtering of sqoop jobs by name.

Job Browser

The Job Browser application lets you to examine the Hadoop MapReduce jobs running on your Hadoop cluster. Job Browser presents the job and tasks in layers. The top layer is a list of jobs, and you can link to a list of that job's tasks. You can then view a task's attempts and the properties of each attempt, such as state, start and end time, and output size. To troubleshoot failed jobs, you can also view the logs of each attempt.

Job Browser Installation and Configuration

Job Browser is one of the applications installed as part of Hue. For information about installing and configuring Hue, see the Hue Installation manual.

Job Browser can display both MRv1 and MRv2 jobs, but must be configured to display one type at a time.

Starting Job Browser

Click the **Job Browser** icon () in the navigation bar at the top of the Hue web page.

If there are no jobs that have been run, the **Welcome to the Job Browser** page opens, with links to the Job Designer and Beeswax.

If there are jobs running, then the Job Browser list appears.

Filtering the Job Browser List

- To filter the jobs by their state (such as **Running** or **Completed**), choose a state from the **Job status** drop-down menu.
- To filter by a user who ran the jobs, enter the user's name in the **User Name** query box.
- To filter by job name, enter the name in the **Text** query box.
- To clear the filters, choose **All States** from the **Job status** drop-down menu and delete any text in the **User Name** and **Text** query boxes.
- To display retired jobs, check the **Show retired jobs** checkbox. Retired jobs show somewhat limited information – for example, information on maps and reduces and job duration is not available. Jobs are designated as Retired by the JobTracker based on the value of `mapred.jobtracker.retirejob.interval`. The retired jobs no longer display after the JobTracker is restarted.

Viewing Job Information and Logs

Note: At any level you can view the log for an object by clicking the icon in the Logs column.

To view job information for an individual job:

1. In the **Job Browser** window, click **View** at the right of the job you want to view. This shows the **Job** page for the job, with the recent tasks associated with the job are displayed in the **Tasks** tab.
2. Click the **Metadata** tab to view the metadata for this job.
3. Click the **Counters** tab to view the counter metrics for the job.

To view details about the tasks associated with the job:

1. In the **Job** window, click the **View All Tasks** link at the right just above the **Recent Tasks** list. This lists all the tasks associated with the job.
2. Click **Attempts** to the right of a task to view the attempts for that task.

To view information about an individual task:

1. In the **Job** window, click the **View** link to the right of the task. The attempts associated with the task are displayed.
2. Click the **Metadata** tab to view metadata for this task. The metadata associated with the task is displayed.
3. To view the Hadoop counters for a task, click the **Counters** tab. The counters associated with the task are displayed.
4. To return to the **Job** window for this job, click the job number in the status panel at the left of the window.

To view details about a task attempt:

1. In the **Job Task** window, click the **View** link to the right of the task attempt. The metadata associated with the attempt is displayed under the **Metadata** tab.
2. To view the Hadoop counters for the task attempt, click the **Counters** tab. The counters associated with the attempt are displayed.
3. To view the logs associated with the task attempt, click the **Logs** tab. The logs associated with the task attempt are displayed.
4. To return to the list of tasks for the current job, click the task number in the status panel at the left of the window.

Viewing Job Output

1. In the **Job Browser** window, click the link in the ID column.
2. To view the output of the job, click the link under **OUTPUT** in the panel at the left of the window. This takes you to the job output directory in the **File Browser**.

Job Designer

The Job Designer application enables you to create and submit jobs to the Hadoop cluster. You can include variables with your jobs to enable you and other users to enter values for the variables when they run your job. The Job Designer supports the actions supported by Oozie: MapReduce, Streaming, Java, Pig, Hive, Sqoop, Shell, Ssh, DistCp, Fs, and Email.

Job Designer Installation and Configuration

Job Designer is one of the applications installed as part of Hue. For information about installing and configuring Hue, see the Hue Installation manual..

In order to run DistCp, Streaming, Pig, Sqoop, and Hive jobs, Oozie must be configured to use the Oozie ShareLib. See the Oozie Installation manual.

Starting Job Designer

Click the **Job Designer** icon () in the navigation bar at the top of the Hue web page. The **Job Designs** page opens in the browser.



Installing the Example Job Designs

Note: You must be a superuser to perform this task.

1. Click . The Quick Start Wizard opens.
2. Click **Step 2: Examples**.
3. Click **Job Designer**.



Job Designs

A job design specifies several meta-level properties of a job, including the job design name, description, the executable scripts or classes, and any parameters for those scripts or classes.

Filtering Job Designs

You can filter the job designs that appear in the list by owner, name, type, and description.

To filter the Job Designs list:

1. In the **Job Designs** window, click **Designs**.
2. Enter text in the Filter text box at the top of the **Job Designs** window. When you type in the Filter field, the designs are dynamically filtered to display only those rows containing text that matches the specified substring.

Creating a Job Design

1. In the **Job Designs** window, click New Action > Action, where Action is MapReduce, Streaming, Java, Pig, Hive, Sqoop, Shell, Ssh, DistCp, Fs, or Email.
2. In the **Job Design (Action type)** window, specify the common and job type specific information.
3. Click **Save** to save the job settings.

Deleting and Restoring Job Designs

You can move job designs to the trash and later restore or permanently delete them.

Deleting Job Designs

1. In a Manager screen, check the checkbox next to one or more job designs.
2. Choose one of the following:
 - Delete > Move to trash
 - Delete > Delete forever

Restoring Job Designs

1. In a Manager screen, click **Trash**.
2. Check the checkbox next to one or more job designs.
3. Click Restore.



Job Design Settings

Job Design Common Settings

Most job design types support all the settings listed in the following table. For job type specific settings, see: [MapReduce](#), [Streaming](#), [Java](#), [Pig](#), [Hive](#), [Sqoop](#), [Shell](#), [Ssh](#), [DistCp](#), [Fs](#), and [Email](#).

All job design settings except Name and Description support the use of variables of the form `$variable_name`. When you run the job, a dialog box will appear to enable you to specify the values of the variables.

Parameters to pass to a script or command. The parameters are expressed using the [JSP 2.0 Specification (JSP.2.3) Expression Language](http://jcp.org/aboutJava/communityprocess/final/jsr152/), allowing variables, functions, and complex expressions as parameters.

Name	Identifies the job and its collection of properties and parameters.
Description	A description of the job. The description is displayed in the dialog box that appears if you specify variables for the job.
Advanced	Advanced settings: <ul style="list-style-type: none">• Is shared- Indicate whether to share the action with all users.• Oozie parameters - parameters to pass to Oozie
Prepare	Specifies paths to create or delete before starting the workflow job.
Params	
Job Properties	Job properties. To set a property value, click Add Property . <ol style="list-style-type: none">1. Property name - a configuration property name. This field provides autocompletion, so you can type the first few characters of a property name and then select the one you want from the drop-down list.2. Value the property value.

Files	Files to pass to the job. Equivalent to the Hadoop -files option.
Archives	Files to pass to the job. Archives to pass to the job. Equivalent to the Hadoop -archives option.

MapReduce Job Design

A MapReduce job design consists of MapReduce functions written in Java. You can create a MapReduce job design from existing mapper and reducer classes without having to write a main Java class. You must specify the mapper and reducer classes as well as other MapReduce properties in the Job Properties setting.

Jar path	The fully-qualified path to a JAR file containing the classes that implement the Mapper and Reducer functions.
----------	--

Streaming Job Design

Hadoop streaming jobs enable you to create MapReduce functions in any non-Java language that reads standard Unix input and writes standard Unix output. For more information about Hadoop streaming jobs, see [Hadoop Streaming](#).

Mapper	The path to the mapper script or class. If the mapper file is not on the machines on the cluster, use the Files option to pass it as a part of job submission. Equivalent to the Hadoop -mapper option.
Reducer	The path to the reducer script or class. If the reducer file is not on the machines on the cluster, use the Files option to pass it as a part of job submission. Equivalent to the Hadoop -reducer option.

Java Job Design

A Java job design consists of a main class written in Java.

Jar path	The fully-qualified path to a JAR file containing the main class.
Main class	The main class to invoke the program.
Args	The arguments to pass to the main class.
Java opts	The options to pass to the JVM.

Pig Job Design

A Pig job design consists of a Pig script.

Script name	Script name or path to the Pig script.
-------------	--

Hive Job Design

A Hive job design consists of a Hive script.

Script name	Script name or path to the Hive script.
-------------	---

Sqoop Job Design

A Sqoop job design consists of a Sqoop command.

Command	The Sqoop command.
---------	--------------------

Shell Job Design

A Shell job design consists of a shell command.

Capture output

Command	The shell command.
---------	--------------------

	Indicate whether to capture the output of the command.
--	--

Ssh Job Design

A Ssh job design consists of an ssh command.

Capture output

User	The name of the user to run the command as.
------	---

Host	The name of the host to run the command on.
------	---

Command	The ssh command.
---------	------------------

	Indicate whether to capture the output of the command.
--	--

DistCp Job Design

A DistCp job design consists of a DistCp command.

Fs Job Design

A Fs job design consists of a command that operates on HDFS.

Create directory

Delete path	The path to delete. If it is a directory, it deletes recursively all its content and then deletes the directory.
-------------	--

	The path of a directory to create.
--	------------------------------------

Move file	The source and destination paths to the file to be moved.
-----------	---

Change permissions	The path whose permissions are to be changed, the permissions, and an indicator of whether to change permission recursively.
--------------------	--

Email Job Design

A Email job design consists of an email message.

To addresses	The recipient of the email message.
--------------	-------------------------------------

CC addresses (optional)	The cc recipients of the email message.
Subject	The subject of the email message.
Body	The body of the email message.

Submitting a Job Design

Note:

A job's input files must be uploaded to the cluster before you can submit the job.

To submit a job design:

1. In the **Job Designs** window, click **Designs** in the upper left corner. Your jobs and other users' jobs are displayed in the **Job Designs** window.
2. Check the checkbox next to the job you want to submit.
3. Click the **Submit** button.
 1. If the job contains variables, enter the information requested in the dialog box that appears. For example, the sample grep MapReduce design displays a dialog where you specify the output directory.
 2. Click **Submit** to submit the job.

After the job is complete, the Job Designer displays the results of the job. For information about displaying job results, see [Displaying the Results of Submitting a Job](#).

Copying, Editing, and Deleting a Job Design

If you want to edit and use a job but you don't own it, you can make a copy of it and then edit and use the copied job.

Copy

1. In the **Job Designs** window, click **Designs**. The jobs are displayed in the **Job Designs** window.
2. Check the checkbox next to the job you want to copy.
3. Click the **Copy** button.
4. In the **Job Design Editor** window, change the settings and then click **Save** to save the job settings.

Edit

1. In the **Job Designs** window, click **Designs**. The jobs are displayed in the **Job Designs** window.
2. Check the checkbox next to the job you want to edit.
3. Click the **Edit** button.
4. In the **Job Design** window, change the settings and then click **Save** to save the job settings.

Delete



1. In the **Job Designs** window, click **Designs**. The jobs are displayed in the **Job Designs** window.
2. Check the checkbox next to the job you want to delete.
3. Click the **Delete** button.
4. Click **OK** to confirm the deletion.

To display the Job Submission History:

In the **Job Designs** window, click the **History** tab. The jobs are displayed in the **Job Submissions History** listed by Oozie job ID.

To display Job Details:

In the **Job Submission History** window, click an Oozie Job ID. The results of the job display:

- Actions - a list of actions in the job.
- Click to display the action configuration. In the action configuration for a MapReduce action, click the value of the mapred.output.dir property to display the job output. 
- In the root-node row, click the Id in the External Id column to view the job in the Job Browser.
- Details - the job details. Click to display the Oozie application configuration. 
- Definition - the Oozie application definition.
- Log - the output log.

Oozie Editor and Dashboard

The Oozie Editor/Dashboard application allows you to define Oozie workflow, coordinator, and bundle applications, run workflow, coordinator, and bundle jobs, and view the status of jobs. For information about Oozie, see [Oozie Documentation](#).

A workflow application is a collection of actions arranged in a directed acyclic graph (DAG). It includes two types of nodes:

- Control flow - start, end, fork, join, decision, and kill
- Action - [MapReduce](#), [Streaming](#), [Java](#), [Pig](#), [Hive](#), [Sqoop](#), [Shell](#), [Ssh](#), [DistCp](#), [Fs](#), and [Email](#). In order to run DistCp, Streaming, Pig, Sqoop, and Hive jobs, Oozie must be configured to use the Oozie ShareLib. See the Oozie Installation manual.

A coordinator application allows you to define and execute recurrent and interdependent workflow jobs. The coordinator application defines the conditions under which the execution of workflows can occur.

A bundle application allows you to batch a set of coordinator applications.

Oozie Editor/Dashboard Installation and Configuration

Oozie Editor/Dashboard is one of the applications installed as part of Hue. For information about installing and configuring Hue, see the Hue Installation manual.

Starting Oozie Editor/Dashboard

Click the **Oozie Editor/Dashboard** icon () in the navigation bar at the top of the Hue browser page. **Oozie Editor/Dashboard** opens with the following screens:



- [Dashboard](#) - shows the running and completed workflow, coordinator, and bundle jobs and information about Oozie instrumentation and configuration. The screen is selected and opened to the Workflows page.
- [Workflow Manager](#) - shows available workflows and allows you to create and import workflows.
- [Coordinator Manager](#) - shows available coordinators and allows you to create coordinators.
- [Bundle Manager](#) - shows available bundles and allows you to create bundles.

Installing Oozie Editor/Dashboard Examples

Note: You must be a superuser to perform this task.

1. Click . The Quick Start Wizard opens.
2. Click **Step 2: Examples**.
3. Click **Oozie Editor/Dashboard**.



Filtering Lists in Oozie Editor/Dashboard

Many screens contain lists. When you type in the Filter field on screens, the lists are dynamically filtered to display only those rows containing text that matches the specified substring.

Permissions in Oozie Editor/Dashboard

In the Dashboard workflows, coordinators, and bundles can only be viewed, submitted, and modified by their owner or a superuser.

Editor permissions for performing actions on workflows, coordinators, and bundles are summarized in the following table:

	Action	Superuser or Owner	All
View	Y.	Only if "Is shared" is set	
Submit	Y.	Only if "Is shared" is set	
Modify	Y.	N	

Deleting and Restoring Workflows, Coordinators, and Bundles

You can move workflows, coordinators, and bundles to the trash and later restore or permanently delete them.

Deleting Workflows, Coordinators, and Bundles

1. In a Manager screen, check the checkbox next to one or more workflows, coordinators or bundles.
2. Choose one of the following:
 - Delete > Move to trash
 - Delete > Delete forever

Restoring Workflows, Coordinators, and Bundles

1. In a Manager screen, click **Trash**.
2. Check the checkbox next to one or more workflows, coordinators or bundles.
3. Click Restore.



Dashboard

The Dashboard shows a summary of the running and completed workflow, coordinator, and bundle jobs.

You can view jobs for a period up to the last 30 days.

You can filter the list by date (1, 7, 15, or 30 days) or status (Succeeded, Running, or Killed). The date and status buttons are toggles.

Workflows

Click the **Workflows** tab to view the running and completed workflow jobs for the filters you have specified.

Click a workflow row in the Running or Completed table to view detailed information about that workflow job.

In the left pane contains a link to the workflow and the variable values specified.

For the selected job, the following information is available in the right area.

- **Graph** tab shows the workflow DAG.
- **Actions** tab shows you details about the actions that make up the workflow.
 - Click the **Id** link to see additional details about the action.
 - Click the **External Id** link to view the job in the Job Browser.
- **Details** tab shows job statistics including start and end times.
- **Configuration** tab shows selected job configuration settings.
- **Logs** tab shows log output generated by the workflow job.
- **Definition** tab shows the Oozie workflow definition, as it appears in the workflow.xml file (also linked under the application path properties in the **Details** tab and the **Configuration** tab).

For each action in the workflow you can:

- Click the icon to view the action screen, which contains:
 - **Details** tab shows job statistics including start and end times.
 - **Configuration** tab shows the action configuration settings.
 - **Child Jobs** tab lists jobs generated by the action.
- Click the icon to view the log in the Job Browser.



WORKFLOW
Pig - sample

SUBMITTER
cloudera

STATUS
SUCCEEDED

PROGRESS
100%

ID
0000029-
130515125804945-
oozie-oozi-W

VARIABLES
 input
 output

MANAGE
Rerun

Coordinators

Click the **Coordinators** tab to view the running and completed coordinator jobs for the filters you have specified.

For the selected job, the following information is available.

- The **Calendar** tab shows the timestamp of the job. Click the timestamp to open the workflow DAG.
- The **Actions** tab shows you details about the actions that make up the coordinator.
 - Click the **Id** link to see additional details about the action.
 - Click the **External Id** link to view the job in the Job Browser.
- The **Configuration** tab shows selected job configuration settings.
- The **Logs** tab shows log output generated by the coordinator.
- The **Definition** tab shows the Oozie coordinator definition, as it appears in the coordinator.xml file (also linked under the oozie.coord.application.path property in the **Configuration** tab).

Bundles

Click the **Bundles** tab to view the running and completed bundle jobs for the filters you have specified.

Oozie

The Oozie tab provides subtabs that give you access to Oozie instrumentation and configuration settings.

Instrumentation

For information on the instrumentation metrics supported by Oozie, see [Oozie Monitoring](#).

Configuration

For information on the configuration properties supported by Oozie, see [Oozie Configuration](#).

Workflow Manager

In Workflow Manager you create Oozie workflows and submit them for execution.

Click the **Workflows** tab to open the Workflow Manager.

Each row shows a workflow: its name, description, timestamp of its last modification. It also shows:

- **Steps** - the number of steps in the workflow execution path. This is the number of execution steps between the start and end of the workflow. This will not necessarily be the same as the number of actions in the workflow, if there are control flow nodes in the control path.
- **Status** - who can run the workflow. **shared** means users other than the owner can access the workflow. **personal** means only the owner can modify or submit the workflow. The default is personal.
- **Owner** - the user that created the workflow.

In Workflow Editor you edit workflows that include MapReduce, Streaming, Java, Pig, Hive, Sqoop, Shell, Ssh, DistCp, Fs, Email, Sub-workflow, and Generic actions. You can configure these actions in the Workflow Editor, or you can import job designs from Job Designer to be used as actions in your workflow. For information about defining workflows, see the [Workflow Specification](#).

Opening a Workflow

To open a workflow, in Workflow Manager, click the workflow. Proceed with [Editing a Workflow](#).

Creating a Workflow

1. Click the **Create** button at the top right.
2. In the Name field, type a name.
3. Check the Is shared checkbox to allow all users to access the workflow.
4. Click **advanced** to specify the deployment directory or a job.xml file.
5. Click **Save**. The Workflow Editor opens. Proceed with [Editing a Workflow](#).

Importing a Workflow

1. Click the **Import** button at the top right.
2. In the Name field, type a name.
3. In the **Local workflow.xml file** field, click **Choose File** and select a workflow file.
4. Click **advanced** to specify whether the workflow is shared, the deployment directory, or a job.xml file.
5. Click **Save**. The Workflow Editor opens. Proceed with [Editing a Workflow](#).

Submitting a Workflow

To submit a workflow for execution, do one of the following:

- In the Workflow Manager, click the radio button next to the workflow, and click the **Submit** button.
- In the Workflow Editor, click the **Submit** button.

The workflow job is submitted and the Dashboard displays the workflow job.

To view the output of the job, click **View the logs**.



Suspending a Running Job

In the pane on the left, click the **Suspend** button.

1. Verify that you want to suspend the job.

Resuming a Suspended Job

In the pane on the left, click the **Resume** button.

1. Verify that you want to resume the job.

Rerunning a Workflow

In the pane on the left, click the **Rerun** button.

1. Check the checkboxes next to the actions to rerun.
2. Specify required variables.
3. Click **Submit**.

Scheduling a Workflow

To schedule a workflow for recurring execution, do one of the following:

- In the Workflow Manager, check the checkbox next to the workflow and click the **Schedule** button.
- In the Workflow Editor, click the **Schedule** button.

A coordinator is created and opened in the Coordinator Editor. Proceed with [Editing a Coordinator](#).

Editing a Workflow

In the Workflow Editor you can easily perform operations on Oozie action and control nodes.

Action Nodes

The Workflow Editor supports dragging and dropping action nodes. As you move the action over other actions and forks, highlights indicate active areas. If there are actions in the workflow, the active areas are the actions themselves and the areas above and below the actions. If you drop an action on an existing action, a fork and join is added to the workflow.

- Add actions to the workflow by clicking an action button and drop the action on the workflow. The Edit Node screen displays.



1. Set the action properties and click **Done**. Each action in a workflow must have a unique name.

- Copy an action by clicking the **Copy** button.
- The action is opened in the Edit Node screen.
- Edit the action properties and click **Done**. The action is added to the end of the workflow.
- Delete an action by clicking the button.
- Edit an action by clicking the button.
- Change the position of an action by left-clicking and dragging an action to a new location.



Control Nodes

- Create a fork and join by dropping an action on top of another action.
- Remove a fork and join by dragging a forked action and dropping it above the fork.
- Convert a fork to a decision by clicking the button.
- To edit a decision:
 1. Click the button.
 2. Fill in the predicates that determine which action to perform and select the default action from the drop-down list.
 3. Click **Done**.



Uploading Workflow Files

In the Workflow Editor, click the **Upload** button.

The workspace of the workflow is opened in the File Browser application. Follow the procedure in [Uploading Files](#) to upload the files. You must put JAR files in a lib directory in the workspace.

Editing Workflow Properties

1. In the Workflow Editor, click the link under the Name or Description fields in the left pane.
2. To share the workflow with all users, check the **Is shared** checkbox.
3. To set advanced execution options, click **advanced** and edit the deployment directory, add parameters and job properties, or specify a job.xml file.
4. Click **Save**.

Displaying the History of a Workflow

1. Click the **Dashboard** tab.
2. Click the **Workflows** tab.
3. Click a workflow.
4. Click the **Actions** tab.

Coordinator Manager

In Coordinator Manager you create Oozie coordinator applications and submit them for execution.

Click the **Coordinators** tab to open the Coordinator Manager.

Each row shows a coordinator: its name, description, timestamp of its last modification. It also shows:

- **Workflow** - the workflow that will be run by the coordinator.
- **Frequency** - how often the workflow referenced by the coordinator will be run.
- **Status** - who can run the coordinator. **shared** means users other than the owner can access the workflow. **personal** means only the owner can modify or submit the workflow. The default is personal.
- **Owner** - the user that created the coordinator.

In Coordinator Editor, you edit coordinators and the datasets required by the coordinators. For information about defining coordinators and datasets, see the [Coordinator Specification](#).

Opening a Coordinator

To open a coordinator, in Coordinator Manager, click the coordinator. Proceed with [Editing a Coordinator](#).

Creating a Coordinator

To create a coordinator, in Coordinator Manager:

1. Click the **Create** button at the top right. The Coordinator wizard opens. Proceed with [Editing a Coordinator](#).

Submitting a Coordinator

To submit a coordinator for execution, check the checkbox next to the coordinator and click the **Submit** button.

Editing a Coordinator

In the Coordinator Editor you specify coordinator properties and the datasets on which the workflow scheduled by the coordinator will operate by stepping through screens in a wizard. You can also advance to particular steps and revisit steps by clicking the Step "tabs" above the screens. The following instructions walk you through the wizard.

1. Type a name, select the workflow, check the **Is shared checkbox** to share the job, and click **Next**. If the Coordinator Editor was opened after scheduling a workflow, the workflow will be set.
2. Select how many times the coordinator will run for each specified unit, the start and end times of the coordinator, the timezone of the start and end times, and click **Next**. Times must be expressed as UTC times. For example, to run at 10 pm PST, specify a start time of 6 am UTC of the following day (+8 hours) and set the Timezone field to America/Los_Angeles.
3. Click **Add** to select an input dataset and click **Next**. If no datasets exist, follow the procedure in [Creating a Dataset](#).
4. Click **Add** to select an output dataset. Click **Save coordinator** or click **Next** to specify advanced settings.
5. To share the coordinator with all users, check the **Isshared** checkbox.
6. Fill in parameters to pass to Oozie, properties that determine how long a coordinator will wait before timing out, how many coordinators can run and wait concurrently, and the coordinator execution policy.
7. Click **Save coordinator**.

Creating a Dataset

1. In the Coordinator Editor, do one of the following:
 - Click **here** in the Inputs or Outputs pane at the top of the editor.
 - In the pane at the left, click the **Create new** link. Proceed with [Editing a Dataset](#).

Displaying Datasets

1. In the Coordinator Editor, click **Show existing** in pane at the left.
2. To edit a dataset, click the dataset name in the Existing datasets table. Proceed with [Editing a Dataset](#).

Editing a Dataset

1. Type a name for the dataset.
2. In the Start and Frequency fields, specify when and how often the dataset will be available.
3. In the URI field, specify a URI template for the location of the dataset. To construct URIs and URI paths containing dates and timestamps, you can specify the variables `${YEAR}`, `${MONTH}`, `${DAY}`, `${HOUR}`, `${MINUTE}`. For example:
`hdfs://foo:9000/usr/app/stats/${YEAR}/${MONTH}/data`.
4. In the Instance field, click a button to choose a default, single, or range of data instances. For example, if `frequency==DAY`, a window of the last rolling 5 days (not including today) would be expressed as start: -5 and end: -1. Check the advanced checkbox to display a field where you can specify a coordinator [EL function](#).
5. Specify the timezone of the start date.
6. In the Done flag field, specify the flag that identifies when input datasets are no longer ready.

Displaying the History of a Coordinator

1. Click the **Dashboard** tab.
2. Click the **Coordinators** tab.
3. Click a coordinator.
4. Click the **Actions** tab.

Bundle Manager

In Bundle Manager you create Oozie bundle applications and submit them for execution.

Click the **Bundle** tab to open the Bundle Manager.

Each row shows a bundle: its name, description, timestamp of its last modification. It also shows:

- **Coordinators** - the coordinators that will be run by the bundle.
- **Kick off** - the UTC time when the coordinators referenced by the bundle will be started.
- **Status** - who can run the bundle. **shared** means users other than the owner can access the workflow. **personal** means only the owner can modify or submit the workflow. The default is personal.
- **Owner** - the user that created the bundle.

For information about defining bundles, see the [Bundle Specification](#).

Opening a Bundle

To open a bundle, in Bundle Manager, click the bundle. Proceed with [Editing a Bundle](#).

Creating a Bundle

1. Click the **Create** button at the top right.

2. In the Name field, type a name.
3. In the Kick off time field, choose a kick off time.
4. Check the Is shared checkbox to allow all users to access the workflow.
5. Click **Save**. The Bundle Editor opens. Proceed with [Editing a Bundle](#).

Submitting a Bundle

To submit a bundle for execution, check the checkbox next to the bundle and click the **Submit** button.

Editing a Bundle

In the Bundle Editor, you specify properties by stepping through screens in a wizard. You can also advance to particular steps and revisit steps by clicking the Step "tabs" above the screens. The following instructions walk you through the wizard.

1. Click **Add** to select a coordinator that the bundle will kick off.
2. Choose the kick off time. The time must be expressed as a UTC time. For example, to run at 10 pm PST, specify a start time of 6 am UTC of the following day (+8 hours).
3. To share the bundle with all users, check the **Is shared** checkbox.
4. Click **Next** to specify advanced settings or click **Save bundle**.
5. Fill in parameters to pass to Oozie.
6. Click **Save bundle**.

Displaying the History of a Bundle

1. Click the **Dashboard** tab.
2. Click the **Bundles** tab.
3. Click a bundle.
4. Click the **Actions** tab.


Solr Search

The Solr Search application, which is based on [Apache Solr](#), allows you to perform keyword searches across Hadoop data. A wizard lets you style the result snippets, specify facets to group the results, sort the results, and highlight result fields.

Solr Search Installation and Configuration

Solr Search is one of the applications installed as part of Hue. For information about installing and configuring Hue, see the Hue Installation manual.

Starting Solr Search

Click the **Solr Search** icon () in the navigation bar at the top of the Hue browser page. **Solr Search** opens to the [Collection Manager](#). If there are no collections, the [Import Collections and Cores](#) dialog displays.

Collection Manager

In Collection Manager you import, copy, and delete [collections](#).


Displaying the Collection Manager

When you start Solr Search, the Collection Manager displays. You navigate to the Collection Manager by clicking **Collection manager** in the Search page or the Template Editor.

Filtering Collections

When you type in the Filter field, the list of collections is dynamically filtered to display only those rows containing text that matches the specified substring.

Importing Collections



1. If there are existing collections, click the **Import** button at the top right. The Import Collections and Cores dialog displays. 
2. Check the checkboxes next to the collections to import.
3. Click **Import Selected**. The collection is added to the Collection Manager.

Editing Collection Properties

1. In the Collection Manager, click a collection.
2. In the **COLLECTION** area on the left, click **Properties**.
3. Edit a property and click **Save**.

Searching a Collection

1. In the Collection Manager, click **Search page** or click **Search it** in the Collection area on the left. The Search page displays.
2. Select a collection from the **Search in** drop-down list.
3. Type a search string in the **Search...** text box.

4. Press **Enter** or click the icon. 
5. If you have defined [facets](#), click a facet to display only those results in the group defined by the facet.
6. If you have defined [sorting fields](#), select from the **Sort by** drop-down list to sort the results.
7. Click to clear the search string. 

Styling Search Results

Do one of the following:



- In the Collection Manager, click a collection.
- In the Search page, select a collection from the **Search in** drop-down list and click **Customize this collection**. The Template Editor displays.

Template Editor

The Template Editor provides four features:

- [Snippet editor](#) - Specify the layout of the search result snippet, which fields appear in the snippet, and style the results.
- [Facet editor](#) - Define buckets in which to group results.
- [Sort editor](#) - Specify on which fields and order the results are sorted.
- [Highlighting editor](#) - Enable highlighting of search fields.

Snippet Editor

1. In the Snippet Editor, click a tab to choose the method for editing the search snippet fields and styling:
2.
 - **Visual editor** - Click to choose an overall layout for the snippet. 
3.
 - - Select the fields and functions from the drop-down lists on the right and click . 
4.
 - - Select fields, right-click, and select **Cut** and **Paste** to place the fields on the canvas.
5.
 - - Select fields and apply styling using the buttons on top.
6.
 - **Source** -

7.
 - - Select the data fields and functions from the drop-down lists on the right.
8.
 - - Specify layout and styling using HTML tags.
9.
 - **Preview** - Preview the snippet.
10.
 - **Advanced** - Specify styles for CSS classes specified in the Source tab.
11. Click **Save**.

Facet Editor

By default, faceting search result fields is disabled. Click **Enabled** to enable faceting.

1. In the Template Editor, click **2. Facets**. You can move between the facet tabs by clicking each **Step** tab, or by clicking **Back** and **Next**.
2. In the General tab, specify
3.
 - **Limit** - the maximum number of values for each facet.
4.
 - **Mincount** - the minimum number of search results that fall into a group for the facet to display on the Search page.
5. In the Field, Range, and Date Facet tabs, specify the facet properties and click **Add**.
6. In the Facets Order tab, drag and drop the facet to specify the order in they appear in the Search page.
7. Click **Save**. When you display the Search page, the facets display on the left.



Sorting Editor

By default, sorting on search result fields is disabled. Click **Enabled** to enable sorting.

1. In the Template Editor, click **3. Sorting**.
2. In the Field drop-down, select a field. Optionally specify a label for the field.
3. The default order is ascending. Click the arrows to change the order.
4. Click **Add**.
5. Click **Save**. When you display search results, the results are sorted by the fields in the order that they appear left to right.



Highlighting Editor

By default, highlighting search result fields is disabled. Click **Enabled** to enable highlighting.

1. In the Template Editor, click **3. Highlighting**.
2. Select the fields to be highlighted.
3. Click **Save**. When you display search results, the selected fields are displayed with the style of the **em** class defined in the Advanced tab of the [Snippet editor](#).

ZooKeeper Browser

The main two features are:

- Listing of the ZooKeeper cluster stats and clients
- Browsing and edition of the ZNode hierarchy

ZooKeeper Browser requires the [ZooKeeper REST](#) service to be running. Here is how to setup this one:

First get and build ZooKeeper:

```
git clone https://github.com/apache/zookeeper
cd zookeeper
ant
Buildfile: /home/hue/Development/zookeeper/build.xml

init:
    [mkdir] Created dir: /home/hue/Development/zookeeper/build/classes
    [mkdir] Created dir: /home/hue/Development/zookeeper/build/lib
    [mkdir] Created dir: /home/hue/Development/zookeeper/build/package/lib
    [mkdir] Created dir: /home/hue/Development/zookeeper/build/test/lib

...
```

And start the REST service:

```
cd src/contrib/rest
nohup ant run&
```

If ZooKeeper and the REST service are not on the same machine as Hue, go update the [Hue settings](#) and specify the correct hostnames and ports:

```
[zookeeper]

[[clusters]]

[[[default]]]
    # Zookeeper ensemble. Comma separated list of Host/Port.
    # e.g. localhost:2181,localhost:2182,localhost:2183
    ## host_ports=localhost:2181

# The URL of the REST contrib service
    ## rest_url=http://localhost:9998
```

HBase Browser

We'll take a look at the new [HBase Browser App](#) added in Hue 2.5 and improved significantly since.

Prerequisites before using the app:

1. Have HBase and Thrift Service 1 initiated (Thrift can be configured)
2. Configure your list of HBase Clusters in [hue.ini](#) to point to your Thrift IP/Port

SmartView

The smartview is the view that you land on when you first enter a table. On the left hand side are the row keys and hovering over a row reveals a list of controls on the right. Click a row to select it, and once selected you can perform batch operations, sort columns, or do any amount of standard database operations. To explore a row, simply scroll to the right. By scrolling, the row should continue to lazily-load cells until the end.

Adding Data

To initially populate the table, you can insert a new row or bulk upload CSV/TSV/etc. type data into your table.

On the right hand side of a row is a '+' sign that lets you insert columns into your row

Mutating Data

To edit a cell, simply click to edit inline.

If you need more control or data about your cell, click "Full Editor" to edit.

In the full editor, you can view cell history or upload binary data to the cell. Binary data of certain MIME Types are detected, meaning you can view and edit images, PDFs, JSON, XML, and other types directly in your browser!

Hovering over a cell also reveals some more controls (such as the delete button or the timestamp). Click the title to select a few and do batch operations:

If you need some sample data to get started and explore, check out this [howto create HBase table tutorial](#).

Smart Searchbar

The "Smart Searchbar" is a sophisticated tool that helps you zero-in on your data. The smart search supports a number of operations. The most basic ones include finding and scanning row keys. Here I am selecting two row keys with:

```
domain.100, domain.200
```

Submitting this query gives me the two rows I was looking for. If I want to fetch rows after one of these, I have to do a scan. This is as easy as writing a '+' followed by the number of rows you want to fetch.

```
domain.100, domain.200 +5
```

Fetches domain.100 and domain.200 followed by the next 5 rows. If you're ever confused about your results, you can look down below and the query bar and also click in to edit your query.

The Smart Search also supports column filtering. On any row, I can specify the specific columns or families I want

to retrieve. With:

```
domain.100[column_family:]
```

I can select a bare family, or mix columns from different families like so:

```
domain.100[family1:, family2:, family3:column_a]
```

Doing this will restrict my results from one row key to the columns I specified. If you want to restrict column families only, the same effect can be achieved with the filters on the right. Just click to toggle a filter.

Finally, let's try some more complex column filters. I can query for bare columns:

```
domain.100[column_a]
```

This will multiply my query over all column families. I can also do prefixes and scans:

```
domain.100[family: prefix* +3]
```

This will fetch me all columns that start with prefix* limited to 3 results. Finally, I can filter on range:

```
domain.100[family: column1 to column100]
```

This will fetch me all columns in 'family:' that are lexicographically \geq column1 but \leq column100. The first column ('column1') must be a valid column, but the second can just be any string for comparison.

The Smart Search also supports prefix filtering on rows. To select a prefixed row, simply type the row key followed by a star *. The prefix should be highlighted like any other searchbar keyword. A prefix scan is performed exactly like a regular scan, but with a prefixed row.

```
domain.10* +10
```

Finally, as a new feature, you can also take full advantage of the [HBase filtering](#) language, by typing your filter string between curly braces. HBase Browser autocompletes your filters for you so you don't have to look them up every time. You can apply filters to rows or scans.

```
domain.1000 {ColumnPrefixFilter('100-') AND ColumnCountGetFilter(3)}
```

This doc only covers a few basic features of the Smart Search. You can take advantage of the full querying language by referring to the help menu when using the app. These include column prefix, bare columns, column range, etc. Remember that if you ever need help with the searchbar, you can use the help menu that pops up while typing, which will suggest next steps to complete your query.

User Admin

The User Admin application lets a superuser add, delete, and manage Hue users and groups, and configure group permissions. Superusers can add users and groups individually, or import them from an LDAP directory. Group permissions define the Hue applications visible to group members when they log into Hue and the application features available to them.

Starting User Admin

Click the **User Admin** icon () in the navigation bar at the top of the Hue browser page. The Hue Users page opens.



Users

The User Admin application provides two levels of user privileges: superusers and users.

- Superusers — The first user who logs into Hue after its initial installation becomes the first superuser. Superusers have permissions to perform administrative functions:
 - Add and delete users
 - Add and delete groups
 - Assign permissions to groups
 - Change a user into a superuser
 - Import users and groups from an LDAP server
- Users — can change their name, e-mail address, and password and log in to Hue and run Hue applications, subject to the permissions provided by the Hue groups to which they belong.

Adding a User

1. In the **User Admin** page, click **Add User**.
2. In the **Credentials** screen, add required information about the user. Once you provide the required information you can click the wizard step tabs to set other information.

Username	A user name that contains only letters, numbers, and underscores; blank spaces are not allowed and the name cannot begin with a number. The user name is used to log into Hue and in file permissions and job submissions. This is a required field.
Password and Password confirmation	A password for the user. This is a required field.
Create home directory	Indicate whether to create a directory named /user/username in HDFS. For non-superusers, the user and group of the directory are username. For superusers, the user and group are username and supergroup.

1. Click **Add User** to save the information you specified and close the **Add User** wizard or click **Next**.
2. In the **Names and Groups** screen, add optional information.

First name and Last name	The user's first and last name.
E-mail address	The user's e-mail address. The e-mail address is used by the Job Designer and Beeswax applications to send users an e-mail message after certain actions have occurred. The Job Designer sends an e-mail message after a job has completed. Beeswax sends a message after a query has completed. If an e-mail address is not specified, the application will not attempt to email the user.
Groups	The groups to which the user belongs. By default, a user is assigned to the **default** group, which allows access to all applications. See [Permissions](#permissions).
<ol style="list-style-type: none"> 1. Click Add User to save the information you specified and close the Add User wizard or click Next. 2. In the Advanced screen, add status information. 	
Active	Indicate that the user is enabled and allowed to log in. Default: checked.
Superuser status	Assign superuser privileges to the user.

1. Click **Add User** to save the information you specified and close the **Add User** wizard.

Deleting a User

1. Check the checkbox next to the user name and click **Delete**.
2. Click **Yes** to confirm.

Editing a User

1. Click the user you want to edit in the **Hue Users** list.
2. Make the changes to the user and then click **Update user**.

Importing Users from an LDAP Directory

Hue must be configured to use an external LDAP directory (OpenLDAP or Active Directory). See Hue Installation in [CDH4 Installation](#).

Note:

Importing users from an LDAP directory does not import any password information. You must add passwords manually in order for a user to log in.

To add a user from an external LDAP directory:

1. Click **Add/sync LDAP user**.
2. Specify the user properties:

Username	The user name.
Distinguished name	Indicate that Hue should use a full distinguished name for the user. This imports the user's first and last name, username, and email, but does not store the user password.
Create home directory	Indicate that Hue should create a home directory for the user in HDFS.

1. Click **Add/sync user**.

If the user already exists in the User Admin, the user information in User Admin is synced with what is currently in the LDAP directory.

Syncing Users and Groups with an LDAP Directory

You can sync the Hue user database with the current state of the LDAP directory using the **Sync LDAP users/groups** function. This updates the user and group information for the already imported users and groups. It does not import any new users or groups.

1. Click **Sync LDAP users/groups**.
2. The **Create Home Directories** checkbox creates home directories in HDFS for existing imported members that don't have home directories.
3. In the **Sync LDAP users and groups** dialog, click **Sync** to perform the sync.

Groups

Superusers can add and delete groups, configure group permissions, and assign users to group memberships.

Adding a Group

You can add groups, and delete the groups you've added. You can also import groups from an LDAP directory.

1. In the **User Admin** window, click **Groups** and then click **Add Group**.
2. Specify the group properties:

Name	The name of the group. Group names can only be letters, numbers, and underscores; blank spaces are not allowed.
Members	The users in the group. Check user names or check Select all.
Permissions	The applications the users in the group can access. Check application names or check Select all.

1. Click **Add group**.

Adding Users to a Group

1. In the **User Admin** window, click **Groups**.
2. Click the group.
3. To add users to the group, check the names in the list provided or check **Select All**.
4. Click **Update group**.

Deleting a Group

1. Click **Groups**.
2. Check the checkbox next to the group and click **Delete**.
3. Click **Yes** to confirm.

Importing Groups from an LDAP Directory

1. From the **Groups** tab, click **Add/sync LDAP group**.

2. Specify the group properties:

Name	The name of the group.
Distinguished name	Indicate that Hue should use a full distinguished name for the group.
Import new members	Indicate that Hue should import the members of the group.
Import new members from all subgroups	Indicate that Hue should import the members of the subgroups.
Create home directories	Indicate that Hue should create home directories in HDFS for the imported members.

1. Click **Add/sync group**.

Permissions

Permissions for Hue applications are granted to groups, with users gaining permissions based on their group membership. Group permissions define the Hue applications visible to group members when they log into Hue and the application features available to them.

1. Click **Permissions**.
2. Click the application for which you want to assign permissions.
3. Check the checkboxes next to the groups you want to have permission for the application. Check **Select all** to select all groups.
4. Click **Update permission**. The new groups will appear in the Groups column in the **Hue Permissions** list.