Cloudera Runtime 1.0.0

Administering Hue

Date published: 2021-12-01
Date modified: 2024-12-05

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Contents

Hue configurations in Cloudera Data Warehouse

You can customize the settings for Hue at a Vitrual Warehouse level by specifying the configurations in the hue-safety-valve field in the Cloudera Data Warehouse (CDW) UI.

The following table lists the safety valve parameters supported by Hue in Cloudera Data Warehouse:

Parameter	Description

`[notebook]`
`  [[interpreters]]`
`  [[[hive]]]`
`  name=Hive`
`  interface=hiveserver2`
`  [[[hplsql]]]`
`  name=Hplsql`
`  interface=hiveserver2`
	Used to activate and enable switching between Hive and HPL/SQL interpreters for your queries.
	 
	Note: Hue enables you to switch between Hive and HPL/SQL interpreters. By default, the Hive interpreter is enabled when you create a Hive Virtual Warehouse. To enable the HPL/SQL interpreter, you must update the configuration in the hue-safety-valve field in your Hive Virtual Warehouse. However, updating hue-safety-valve overrides the default configuration. Therefore, to use both Hive and HPL/SQL interpreters, you must enable both by updating the configuration in the hue-safety-valve field.

`[[desktop]]` `app_blacklist`	Used to add or remove applications, such as the File Browser, Impala, Hive, Oozie, and so on from the blocked list of applications.

`[desktop]`
`  [[session]]`
`  ttl=[***NUMBER-OF-SECONDS***]`
	Used to configure the duration of a user session. The ttl property determines the length of time that the cookie with the user's session ID lives before expiring. After the ttl setting is reached, the user's session expires whether it is active or not.
	The default setting for ttl is 1,209,600 seconds, which equals two weeks.

`[jobbrowser]`
`[[query_store]]`
`server_url=[***QUERY-PROCESSOR-URL***]`
	Used to display the **Queries** tab for Hive and Impala on the **Job Browser** page. This configuration is enabled by default and is not exposed in the Hue safety valve.
	However, to hide the **Queries** tab, you can override the configuration by adding the following lines in Virtual Warehouse ⋮ Edit CONFIGURATIONS Hue hue-safety-valve :
	`[jobbrowser]`
	`[[query_store]]`
	`is_enabled=false`
	To enable the **Queries** tab, set is_enabled to true.

`[aws]`
`[[aws_accounts]]`
`[[[default]]]`
`access_key_id=[***AWS-ACCESS-KEY***]`
`secret_access_key=[***SECRET-ACCESS-KEY***]`
`region=[***AWS-REGION***]`
	(Non-RAZ) Used to enable the S3 File Browser for Hue in Cloudera Data Warehouse to access an S3 bucket from Hue.

`[azure]`
`  [[azure_accounts]]`
`  [[[default]]]`
`  client_id=[***AZURE-ACCOUNT-CLIENT-ID***]`
`  client_secret=[***AZURE-ACCOUNT-CLIENT-SECRET***]`
`  tenant_id=[***AZURE-ACCOUNT-TENANT-ID***]`
 	
`  [[abfs_clusters]]`
`  [[[default]]]`
`  fs_defaultfs=abfs://[***CONTAINER-NAME***]@[***AZURE-STORAGE-ACCOUNT-NAME***].dfs.core.windows.net`
`  webhdfs_url=https://[***AZURE-STORAGE-ACCOUNT-NAME***].dfs.core.windows.net/`
	(Non-RAZ) Used to enable the ABFS File Browser for Hue in Cloudera Data Warehouse to access ADLS Gen2 containers.

Hue supported browsers

Hue works with the two most recent LTS (long term support) or ESR (extended support release) browsers. Cookies and JavaScript must be enabled.

The lists the minimum tested versions of the most common browsers:

- Chrome: (Version history)
- Firefox: (Version history)
- Safari (Mac only): Version history
- Microsoft Edge: (Version history)

Hue can display in other browsers and in older versions of the common browsers, but you might not have access to all features.

Customizing the Hue web interface

You can customize the page logo and set the cache timeout value by configuring the parameters in the Virtual Warehouse which is running Hue.

Adding a custom banner in Hue

You can add a custom banner to the Hue web interface by adding your custom HTML to the hue-safety-valve configuration for your Virtual Warehouse.

Procedure

1. Log in to the Data Warehouse service as DWAdmin.
2. Go to your Virtual Warehouse tile and click ⋮ Edit .
3. Go to CONFIGURATIONS Hue , select hue-safety-valve from the Configuration files drop-down menu and add the following lines:

```
[desktop]
[[custom]]
banner_top_html=<H1>Your company's custom Hue Web UI banner</H1>
```

4. Click Apply Changes.

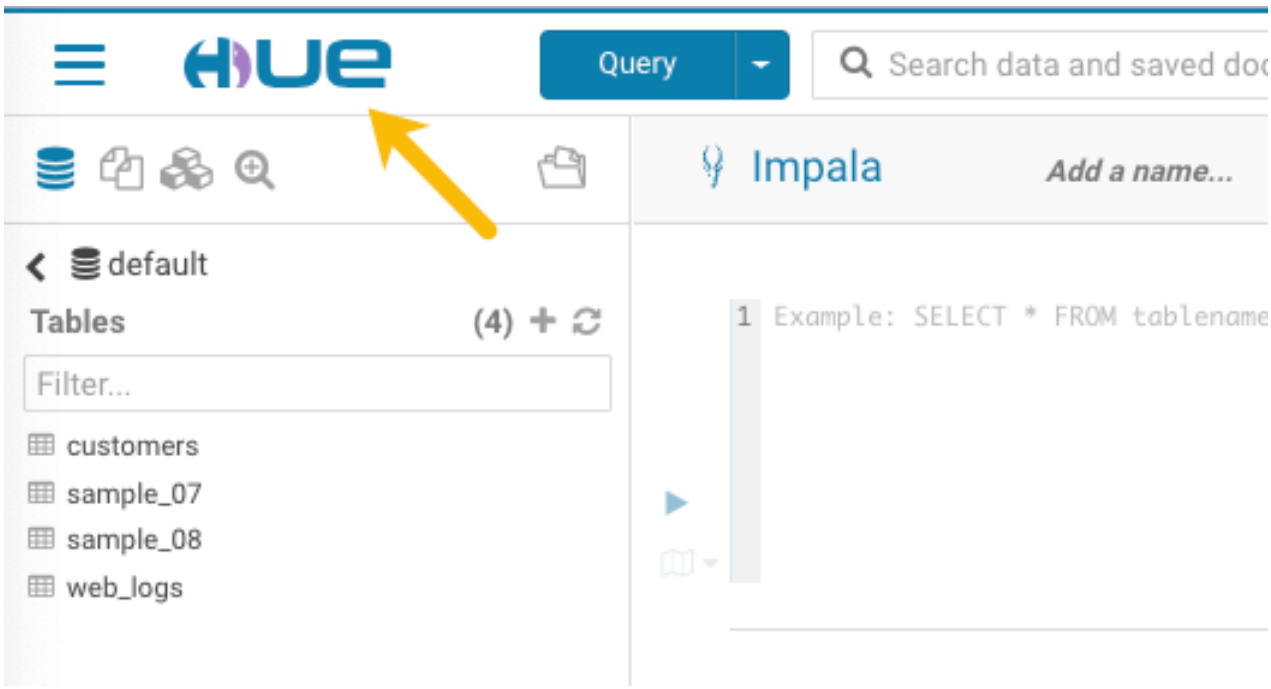
   The Virtual Warehouse goes into an "Updating" state. Wait for the update to complete, and then open or refresh Hue.

Changing the page logo in Hue

You can replace the Hue web interface logo with a custom log that is created with an SVG code. Add any type of logo you want, but your custom logo should be designed to fit into a 160 x 40 pixel space.

About this task

For example, here is the Hue logo shown in the following image:



You can change this Hue logo by adding the appropriate SVG code to the logo_svg property under the [desktop] [[custom]] section in the hue_safety_valve configuration parameter in Cloudera Data Warehouse.

Procedure

1. Log in to the Cloudera Data Warehouse service as an administrator.
2. Go to the Virtual Warehouses Edit CONFIGURATIONS Hue and select hue-safety-valve from the Configuration files drop-down list.
3. Add the custom logo SVG code in the [desktop] [[custom]] section as shown in the following example:

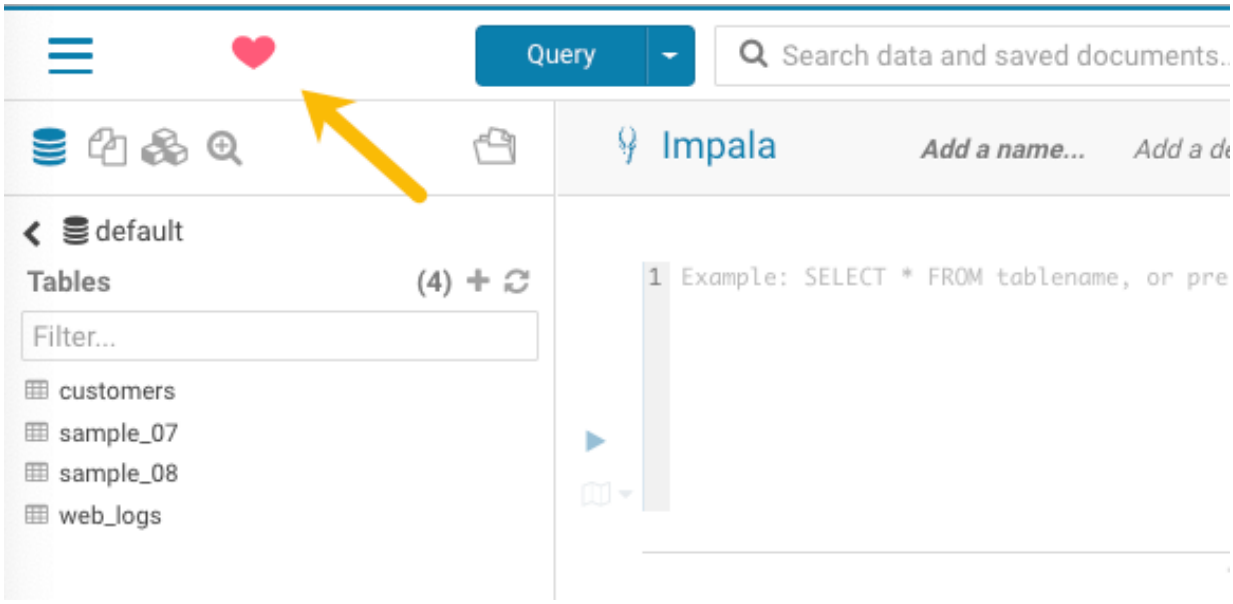
```
[desktop]
[[custom]]
logo_svg='[***SVG-CODE-FOR-CUSTOM-LOGO***]'
```

For example, the following SVG code replaces the Hue logo with a red heart:

```
[desktop]
[[custom]]
logo_svg='<g><path stroke="null" id="svg_1" d="m44.41215,11.43463c-4.05
017,-10.71473
-17.19753,-5.90773 -18.41353,-0.5567c-1.672,-5.70253 -14.497,-9.95663
-18.411,0.5643c-4.35797,11.71793 16.891,22.23443 18.41163,23.95773c1.518
1,-1.36927 22.7696,-12.43803
18.4129,-23.96533z" fill="#ffffff"/> <path stroke="null" id="svg_2"
d="m98.41246,10.43463c-4.05016,-10.71473 -17.19753,-5.90773 -18.41353,-
0.5567c-1.672,-5.70253
-14.497,-9.95663 -18.411,0.5643c-4.35796,11.71793 16.891,22.23443 18.4116
4,23.95773c1.5181,-1.36927
22.76959,-12.43803 18.41289,-23.96533z" fill="#FF5A79"/> <path stroke="nu
ll" id="svg_3"
d="m154.41215,11.43463c-4.05016,-10.71473 -17.19753,-5.90773 -18.41353,-0
.5567c-1.672,-5.70253
-14.497,-9.95663 -18.411,0.5643c-4.35796,11.71793 16.891,22.23443 18.41164
,23.95773c1.5181,-1.36927 22.76959,-12.43803 18.41289,-23.96533z" fill="
#ffffff"/> </g>'
```

4. Click APPLY.
5. Restart the Virtual Warehouse.
6. Verify your changes by opening Hue.

   If you added the sample SVG code that defines a red heart as the logo, then your Hue web interface looks as shown in the following image:



Adding a splash screen in Hue

You can add a custom splash screen to the Hue web interface by adding your custom HTML to the hue-safety-valve configuration for your Virtual Warehouse.

Procedure

1. Log in to the Cloudera Data Warehouse as DWAdmin.
2. Go to your Virtual Warehouse tile and click ⋮ Edit .
3. Go to CONFIGURATIONS Hue , select hue-safety-valve from the Configuration files drop-down menu and add the following lines:

```
[desktop]
[[custom]]
login_splash_html=[***CUSTOM-HTML***]
```

```
[desktop]
[[custom]]
login_splash_html=<h1>Hue, the next-gen SQL Assistant</h1>
```

4. Click Apply Changes.

   The Virtual Warehouse goes into an "Updating" state. Wait for the update to complete, and then open or refresh Hue.

Setting the cache timeout

Enable Hue UI caching by setting a timeout value in milliseconds. The default is 10 days or 864000000 milliseconds. Set the timeout to 0 to disable caching.

About this task

When you browse tables using the left assist panel or run queries, Hue caches this information for fetching information faster and query autocompletion. You can configure the time for which you want to Hue to cache this information by setting the value of the cacheable_ttl property under the [desktop][[custom]] section in the hue_safety_valve configuration property in Cloudera Data Warehouse.

Procedure

1.
2. Log in to the Cloudera Data Warehouse service as an administrator.
3. Go to the Virtual Warehouses ⋮ Edit CONFIGURATIONS Hue and select hue-safety-valve from the Configuration files dropdown menu.
4. Add the following parameters with the cache timeout value to the hue_safety_valve configuration parameter:

   ```
   [desktop]
   [[custom]]
   cacheable_ttl=[***VALUE-IN-MILLISECONDS***]
   ```

   For example, the following configuration sets the cache timeout to the default value of 86400000 milliseconds:

   ```
   [desktop]
   [[custom]]
   cacheable_ttl=86400000
   ```
5. Click APPLY.
6. Restart the Virtual Warehouse.

Enabling or disabling anonymous usage date collection

Hue tracks anonymized pages and application versions to gather information about application usage levels. The data collected does not include hostnames or IDs. For example, the data collected has the format /2.3.0/pig or /2.5.0/beeswax/execute.

About this task

To enable or disable anonymous usage data collection:

Procedure

1. In the Cloudera Manager Admin Console, select Clusters Hue Configuration to navigate to the configuration page for Hue.
2. In the Search text box, type usage to locate the Enable Usage Data Collection check box:
   - To enable anonymous data collection, check the box, which is the default setting.
   - To disable anonymous data collection, clear the check box.
3. Enter a Reason for change..., and then click Save Changes at the bottom of the page to save the configuration change.

4. Refresh the browser page and click the restart icon  at the top of the page so the new configuration changes can be read by the server and the new data collection setting takes effect.

Disabling the share option in Hue

Hue allows you to share documents, queries, and workflows with other users, either for viewing only or viewing and modifying in any Hue instances across all Virtual Warehouses within a Database Catalog. Sharing is enabled by default in the Hue UI. For added privacy and control, you can disable sharing by setting the enable_sharing property to false in the hue-safety-valve.

About this task

The sharing option is always available to the admin users. To disable the share option:

Procedure

1. Log in to the Cloudera Data Warehouse service as an administrator.
2. Go to the Virtual Warehouses Edit CONFIGURATIONS Hue and select hue-safety-valve from the Configuration files drop-down list.
3. Add the following lines in the hue-safety-valve:

```
[desktop]
enable_sharing=false
```

4. Click APPLY.
5. Restart the Virtual Warehouse.

Configuring idle session timeout for Hue

By configuring idle session timeout, you can prevent unauthorized access to your data by automatically logging out users when they are not actively using the Hue web interface. Hue warns the users and then logs them out after a set period of inactivity that you specify in the Hue configuration.

About this task

By default, the value of the idle_session_timeout property is set to “-1”, which means that the users are not logged out until they manually log out of Hue.

Procedure

1. Log in to the Cloudera Data Warehouse service as an administrator.
2. Go to the Virtual Warehouse in which Hue is running and click Edit.
3. Go to CONFIGURATIONS Hue and select hue-safety-valve from the Configuration files drop-down list and add the following lines:

```
[desktop]
[[auth]]
idle_session_timeout=[***TIME-IN-SECONDS***]
```

4. Click APPLY.

   The Virtual Warehouse restarts to update the configuration.

Adding Query Processor Administrator users and groups in Cloudera Data Warehouse

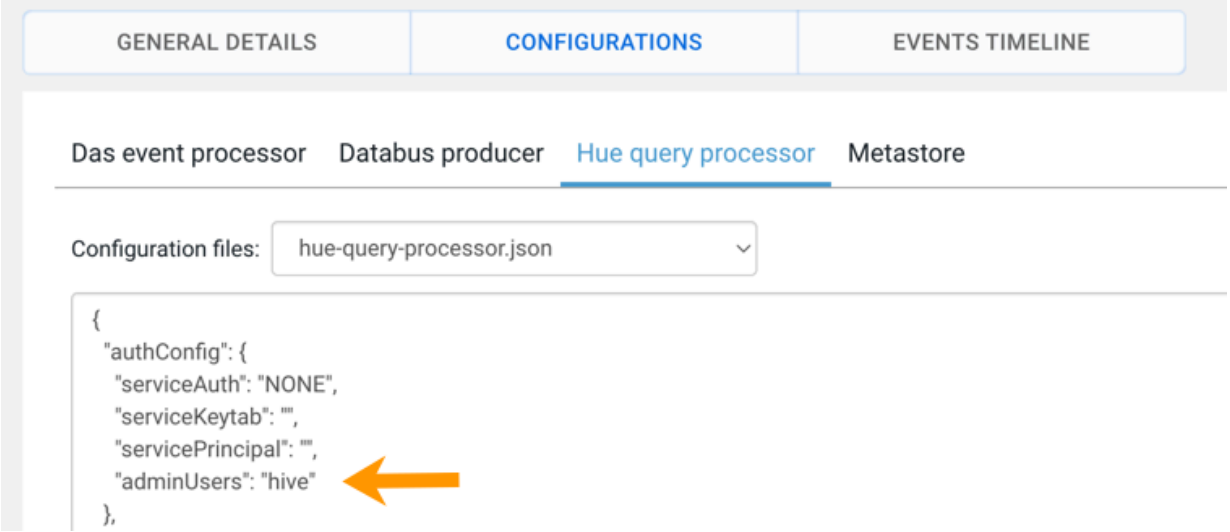
The Query Processor Administrators have special privileges that enable them to view and monitor queries from all users, including the ones that were submitted from query interfaces, such as Beeline, Hive Warehouse Connector (HWC), Tableau, Impala-shell, Impyla, and other JDBC/ODBC clients.

Before you begin

Make sure that the Virtual Warehouse to which you want to add the Hue Query Processor Administrators users is in the stopped state.

Procedure

1. Log in to the Cloudera Data Warehouse web interface as a DWAdmin.
2. Click ⋮ Edit on the Database Catalog for which you want to add Hue Query Processor Administrators users.
3. On the **Database Catalog** details page, click CONFIGURATIONS Hue query processor and select hue-query-processor.json from the Configuration files drop-down menu, as shown in the following image:



4. In the “authConfig” section, add the list of users to the “adminUsers” key.
   For example: "adminUsers": "hive, [***USER-1***], [***USER-2***]"
   You can also add a list of admin groups as follows:

```
"adminGroups": "admin-group, [***GROUP-1***], [***GROUP-2***]"
```

5. Click Apply.
   The Hue service will be unavailable for approximately 5 minutes to make the update.

Ways to clean up old queries from the Query Processor tables

Learn how to schedule a query clean-up and how to use the API to manually clean up queries from the following Query Processor tables: vertex_info, dag_details, dag_info, query_details, hive_query, tez_app_info.

Scheduling query clean-up

Both Hive and Impala queries are retained in the backend database for 30 days by default, after which they are cleaned up. You can change the clean-up interval from the Database Catalog configurations. Go to Database Catalogs ⋮ Edit CONFIGURATIONS Hive query processor and select the hue-query-processor.json from the Configuration files drop-down menu.

Add the following line under the dasConf section and specify the time interval in seconds:

```
 "dasConf": {
    "hue.query-processor.event-pipeline.cleanup-interval-secs": "[***TIME-
INTERVAL-IN-SECONDS***]",
    “hue.query-processor.event-pipeline.cleanup.cron.expression” : “[***CR
ON-EXPRESSION***]”
  },
```

For example:

```
“hue.query-processor.event-pipeline.cleanup.cron.expression” : “0 0 2 * * ?”
,
“hue.query-processor.event-pipeline.cleanup-interval-secs” : “2592000”
```

Manually cleaning up queries using an API

The ability to clean up queries manually in addition to the scheduled clean-up routines is useful when you have a high load of queries in a particular week that are hogging resources that you must free up. The API also runs a VACUUM command on the Query Processor table to reclaim storage that is occupied by dead tuples.

You can send an API request using tools such as cURL or Postman.

API format: `[***X-Do-As:[COMPONENT/PROCESS USER ID]***][***QUERY-PROCESSOR-ADDRESS***]/api/admin/cleanup/[***EPOCH-TIME***]`

Where,

- [***QUERY-PROCESSOR-ADDRESS***] is the query processor host address
- [***EPOCH-TIME***] is the Unix epoch time in seconds

Queries that were run before the specified epoch time are purged.

For example:

```
curl "http://machine1.company.com:30700/api/admin/cleanup/1670006742"
```

Downloading debug bundles

The debug bundle is a ZIP file that contains the query details in JSON format and an error-reports.json file, which is created only if an error occurs while the query is run.

About this task

Note: This feature is available only for Hive queries.

If Tez is used to run a query, then the debug bundle also contains DAG and Tez JSON files, as shown in the following image:

Procedure

1. Go to the Cloudera Data Warehouse web interface and open Hue from your Virtual Warehouse.
2. Click Jobs from the left assist panel.
   The **Job Browser** page is displayed.
3. Click Queries.
   The Hive queries that were run are displayed.
4. Select a query for which you want to download the debug bundle.
5. Click Download and save the ZIP file on your computer.
   The filename is in the following format:

```
hive_[***HIVE-QUERY-ID***]_[***USER-ID***]_[***UNIQUE-INDEX***]
```

Enabling Spark 3 engine in Hue

Hue leverages Apache Livy 3 to support Spark SQL queries in Hue on the Apache Spark 3 engine. To enable the Spark 3 engine, specify the Livy server URL in the Hue Advanced Configuration Snippet using Cloudera Manager, and enable the Spark SQL notebook. Livy for Spark 3 and Spark 3 services are installed when you create a Cloudera Data Hub cluster with the Cloudera Data Engineering cluster template.

Before you begin

Note: Livy v0.6 supports Python versions upto Python 3.7. If you install Python 3.8 or higher, then you may see the following error: TypeError: required field "type_ignores" missing from Module.

Procedure

1. Log in to Cloudera Manager as an Administrator.

2. Go to Clusters HDFS Configuration and add the following lines in the Cluster-wide Advanced Configuration Snippet (Safety Valve) for core-site.xml field:

```
<property>
    <name>hadoop.proxyuser.hue.groups</name>
    <value>*</value>
</property>
<property>
    <name>hadoop.proxyuser.hue.hosts</name>
    <value>*</value>
</property>
<property>
    <name>hadoop.proxyuser.spark.groups</name>
    <value>*</value>
</property>
<property>
    <name>hadoop.proxyuser.spark.hosts</name>
    <value>*</value>
</property>
<property>
    <name>hadoop.proxyuser.livy.groups</name>
    <value>*</value>
</property>
<property>
    <name>hadoop.proxyuser.livy.hosts</name>
    <value>*</value>
</property>
```

3. Click Save Changes.
4. Go to Clusters Livy for Spark 3 service Configuration and add the following configurations:
   a) Add the hue user in the Admin Users (livy.superusers) field.
   b) Go to the HMS Service field and select Hive.
   c) Click Save Changes.
5. Go to Clusters SPARK_ON_YARN Configuration Admin Users , add hue to the list of admin users (spark.history.ui.admin.acls) and click Save Changes.
6. Go to Clusters SPARK Configuration Admin Users , add hue to the list of admin users (spark.history.ui.admin.acls) and click Save Changes.
7. Go to Clusters SPARK 3 Configuration Admin Users , add hue to the list of admin users (spark.history.ui.admin.acls) and click Save Changes.
8. Go to Clusters Hue Configuration and enter the following lines in the Hue Service Advanced Configuration Snippet (Safety Valve) for hue_safety_valve.ini field and click Save Changes:

```
[desktop]
app_blacklist=zookeeper, pig #custom list of blocked apps
[spark]
#This is not a thrift server port
#If this TLS/SSL is enabled then check to see whether the livy url is on h
ttps or http and modify the url accordingly.
livy_server_url=http(s)://[***LIVY-FOR-SPARK3-SERVER-HOST***]:[***LIVY-
FOR-SPARK3-SERVER-PORT***]
ssl_cert_ca_verify=false
security_enabled=true
[notebook]
[[interpreters]]
[[[sparksql]]]
name=Spark SQL
```

```
interface=livy
```

Attention:

- Ensure that the Spark application is not on the blocked list.
- Set ssl_cert_ca_verify=false if an SSL certificate is not present in Hue's truststore. ssl_cert_ca_verify=true if your environment is a secured environment.
- Set security_enabled=true for Kerberized clusters.

9. Restart the affected services.

Results

You can now select the “Spark SQL” dialect on the Hue editor and run Spark queries from Hue.

Note: Starting a Livy session can take 30-45 seconds. If you cancel a running Spark SQL query and rerun it, the Livy session expires, and a new session is created.

About the Hue task server in Cloudera Data Warehouse

The Hue task server helps you schedule tasks to clean up Hue documents and the /tmp directory, improving cluster maintenance experience and performance. The task server also improves the file-upload experience, allowing you to upload larger files up to 5 GB.

The task server is enabled by default. Hue superusers or administrators can schedule cleanup jobs from the **Task Server** page in Hue. They can enable or disable the task server from the Advanced Configuration Snippet in Cloudera Data Warehouse. They can also modify the default behavior, such as the frequency of the cleanup jobs, threshold of the disk space, and so on.

All users having access to Hue can leverage the improved file-upload capabilities from the Hue task server.

Note: The task server is available only for Hue that you access from the Virtual Warehouses. It is not available for the shared Hue service in Cloudera Data Warehouse.

Scheduling tasks to clean up Hue documents

If the task server is enabled, then you can clean up data from the backend Hue tables from the Hue web interface. You no longer need to run the document cleanup shell commands or queries for this purpose.

About this task

Note: Only Hue Superusers or Administrators can access the **Task Server** page to cleanup documents and /tmp directory.

Procedure

1. Log in to the Hue web interface as a superuser.
2. Click Administer Server from the left assist panel and then click the Task Server tab.
3. Click Schedule Task.
4. On the **Schedule Task** modal, select document cleanup from the dropdown menu.

   In the keep-days field, specify the number of days for which you want to retain the data in the Hue tables. For example, if you specify 30, the tables are cleaned up every 30 days.
5. Click Submit.

What to do next
You can monitor the logs by clicking the job ID. By default, only INFO-level logs are displayed. You can change this to DEBUG-level from the Hue Advanced Configuration Snippet for troubleshooting purposes.

Scheduling tasks to clean up Hue's tmp directory

Hue writes temporary files to a temporary block space, also used by other services such as Hive and Impala. If the tmp space is full, the Hue pod gets evicted disrupting operational continuity. Set up a cleanup threshold in the Hue task server to prevent pod eviction.

About this task
Files older than 60 minutes are cleaned up from the /tmp directory. This is a configurable parameter. The default cleanup threshold is 90%.

Note: Only Hue Superusers or Administrators can access the **Task Server** page to cleanup documents and /tmp directory.

Procedure

1. Log in to the Hue web interface as a superuser.
2. Click Administer Server from the left assist panel and then click the Task Server tab.
3. Click Schedule Task.
4. On the **Schedule Task** modal, select tmp clean up from the dropdown menu.

   In the threshold for clean up field, specify the threshold value in percentage.

   The cleanup job is triggered when the disk space reaches the specified threshold value. For example, if you specify 70, the temporary disk space is cleaned up when it is 70% full.
5. Click Submit.

What to do next
You can monitor the logs by clicking the job ID. By default, only INFO-level logs are displayed. You can change this to DEBUG-level from the Hue Advanced Configuration Snippet for troubleshooting purposes.

Advanced configuration snippet parameters for Hue task server in Cloudera Data Warehouse

The Hue task server is enabled by default in Cloudera Data Warehouse with preset cleanup job schedules. The preset values can work in most scenarios. If needed, you can override the values from the Hue Advanced Configuration Snippet.

Configuration parameter	Description	Default value
enabled	A setting to enable or disable the Task Server in Hue. To turn off using the task, set the value of this property to False.	True
check_disk_usage_and_clean_task_periodic_interval	The time interval in seconds at which the cleanup job is run.	1000
disk_usage_cleanup_threshold	A setting to specify the threshold value of disk space in percentage. For example, the cleanup job is triggered when the disk is 90% full.	90

Configuration parameter	Description	Default value
disk_usage_and_clean_task_time_delta	A setting to specify the time (in minutes) after which the files should be cleaned up. For example, if this is set to 60, then only the files with a timestamp older than 60 minutes will be cleaned up from the /tmp directory. This setting helps avoid race condition issues.	60
cleanup_stale_uploads_task_time_delta	The time interval in seconds to run the peroidic cleaner which cleans up failed upload tasks. Only successfully completed upload task keys get cleaned automatically. The failed upload task's keys are retained.	900
cleanup_stale_uploads_task_time_delta	A setting to clean up keys of uploads tasks older than the time delta. The unit of time delta is minutes.	60

How to override the default presets for the Hue task server

Go to Virtual Warehouses ⋮ Edit Configurations , select hue-safety-valve from the Configuration files drop-down menu, and add the following lines:

```
[desktop]
[[task_server_v2]]
```