Cloudera Data Visualization 7.2.9

Logs and Usage

Date published: 2020-10-30
Date modified: 2024-12-18

https://docs.cloudera.com/

Legal Notice

Contents

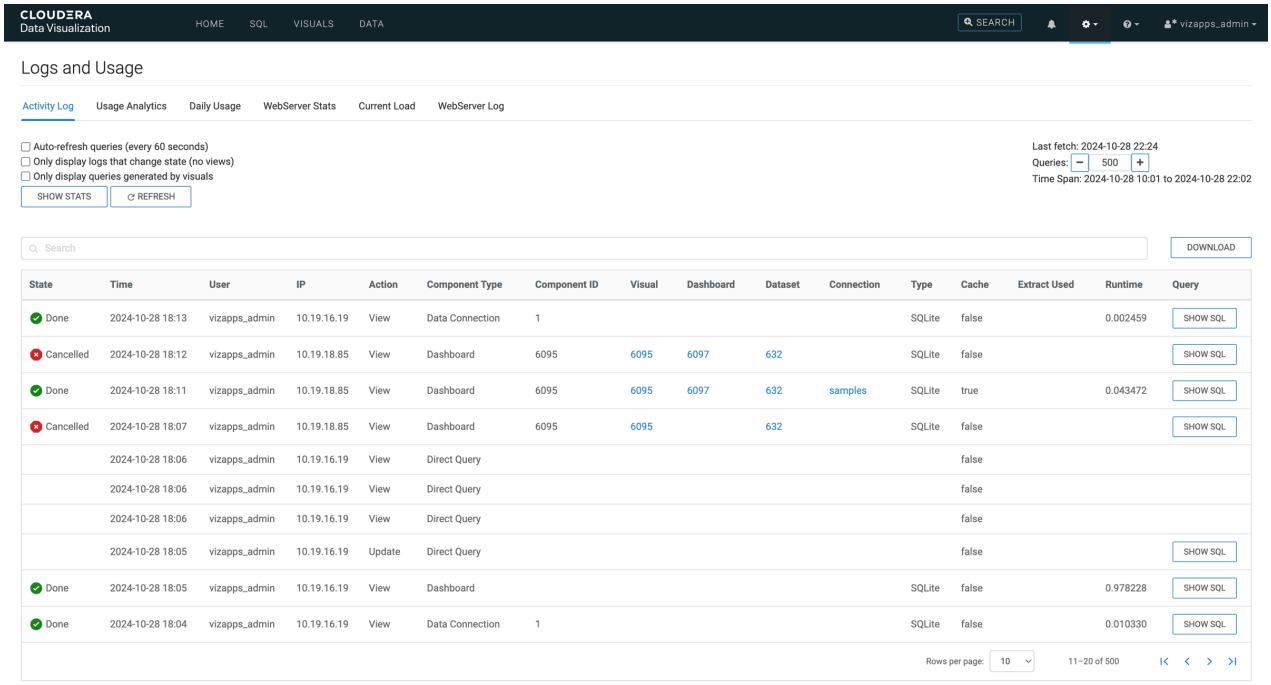
Activity Log

Logs and Usage is a powerful monitoring interface that can help you better understand how users are engaging with Cloudera Data Visualization in your organization. The different tabs provide you detailed information on various aspects, including user activities, system load, and runtime performance. The Activity Log tab of the interface provides access to the audit logs stored in query log files.

You can access the Activity Log through Site Administration Logs and Usage .

Note:

The Logs and Usage interface is only accessible to system administrators (or to users who have a role with View activity logs or Admin user permissions).



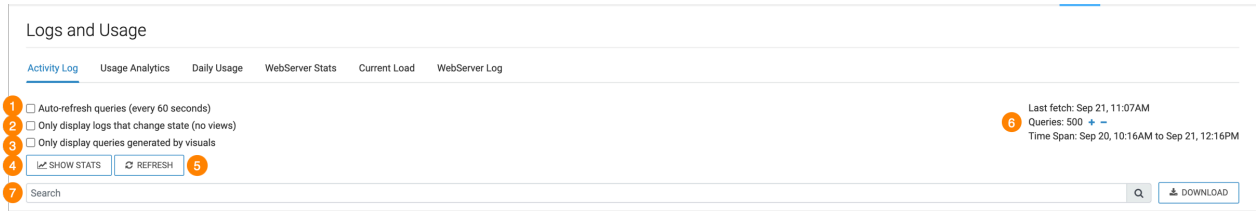
When you open the Activity Log interface, you can see a list view of the activity logs by default. This view provides comprehensive details about each activity.

Note:

As the data shown is persisted in files, restarting the Cloudera Data Visualization instance does not impact the information presented.

Key functions

Within Activity Log, you have access to several essential functions and information:



1. Auto-refresh queries

   Enabling this option automatically updates the information displayed in the interface every 60 seconds.
2. Only display logs that change state

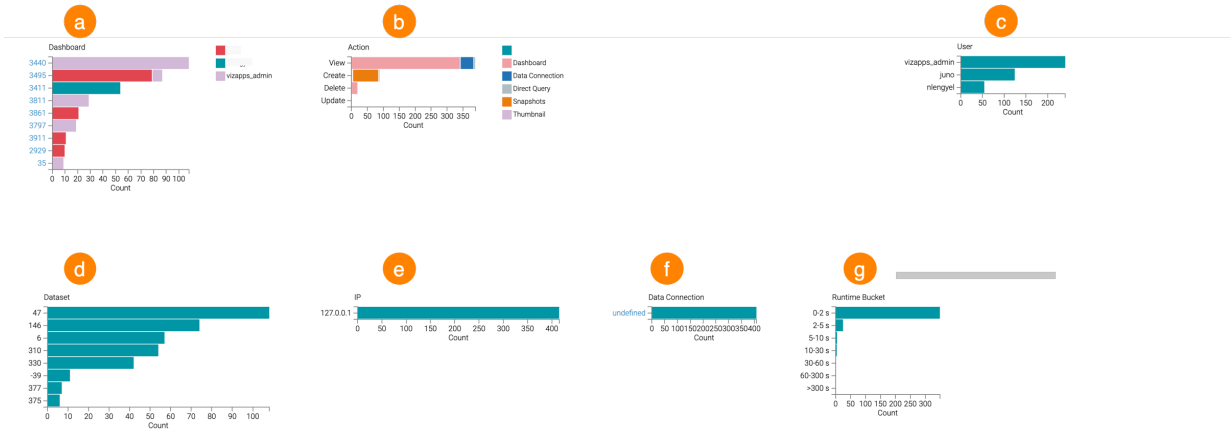
   You can use this option to filter the logs, showing only actions that alter the state of artifacts.
3. Only display queries generated by visuals

   You can use this option to narrow the reporting to actions related to visuals. It excludes actions on datasets, connections, and analytical views from the logs.

4. Show Stats/Hide Stats

You can use this option to display or hide statistical visualizations of the activity logs based on object type.

Note: When you choose to display the visualizations, the traditional list view remains accessible below the visuals. Simply scroll down in your browser window to view it.



a. Dashboard

The horizontal bar chart shows the most frequently accessed dashboards, listing them by their unique IDs.

You can click the IDs to navigate to the respective dashboard, which opens in another browser tab. You can hover over a bar to display detailed information including dashboard ID, activity count, dataset ID, dataset name.

b. Action

The stacked bar chart provides a breakdown of requests by type, arranged along the vertical axis. Individual components within each bar and the legend show the breakdown by component type, such as Dashboard, Data Connection, Dataset, Snapshots, Thumbnail, or Workspaces.

You can hover over a bar to display in-depth insights into each action, including its type, activity count, and associated component.

c. User

The bar chart compiles data on all users who initiated actions. You can hover over a bar to display comprehensive information about each user, including their user name, activity count.

d. Dataset

The horizontal bar chart highlights the most frequently accessed datasets. You can hover over a bar to display detailed dataset information, including its ID, activity count, and name.

e. IP

The bar chart reports server addresses hosting various actions. You can hover over a bar to display detailed information about each connection, including its IPnumber and activity count.

f. Data Connection

The bar chart tracks activity on each live data connection. You can hover over a bar to display detailed information about each connection, including its name and activity count.

g. Runtime Bucket

The bar chart functions as a histogram, categorizing possible runtimes into distinct buckets. Each bar represents a bucket of duration times.You can hover over a bar to provide a breakdown of runtime bucket and activity count for further analysis.

5. Refresh

   This option triggers an immediate refresh operation, bypassing the waiting period for auto-refresh.
6. Information panel on the right side of the page:
   - Last fetch

     This timestamp captures the moment when the query was generated.
   - Queries

     This indicates the number of queries currently reported.

     The default value is 500. You can increase the number of queries by clicking the + icon incrementing them in multiples of 500. Conversely, clicking the - icon decreases the number of queries in increments of 500.
   - Time Span

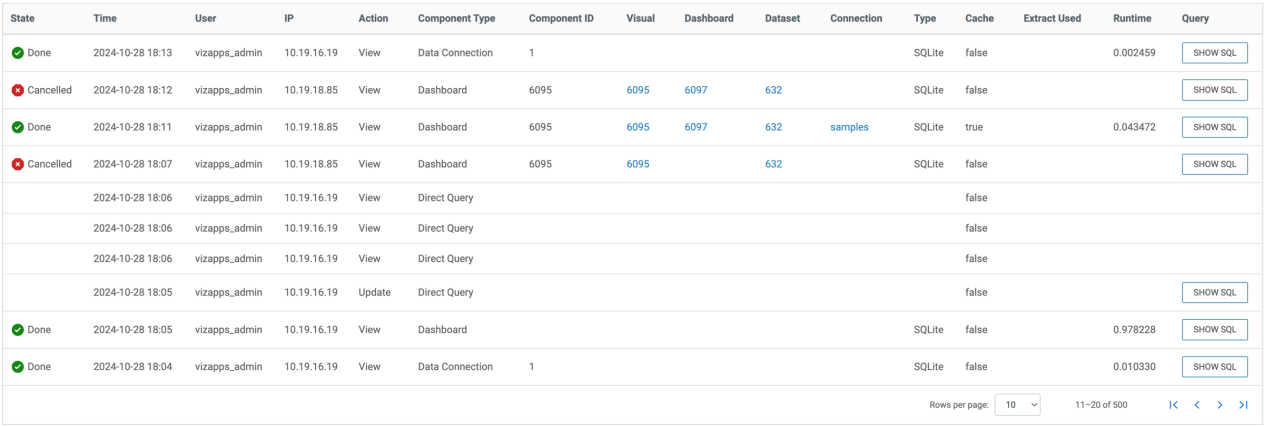
     This time period reports the interval covered by the log data.
7. Search

   You can look up multiple terms, separated by spaces. The search results will show records that contain all the specified terms.

List view

You can check the list view of the Activity Log to get more detailed information about past activities.

By default, the log displays the last 500 activities in reverse chronological order, with the most recent action appearing at the top.



State	Time	User	IP	Action	Component Type	Component ID	Visual	Dashboard	Dataset	Connection	Type	Cache	Extract Used	Runtime	Query
Done	2024-10-28 18:13	vizapps_admin	10.19.16.19	View	Data Connection	1					SQLite	false		0.002459	SHOW SQL
Cancelled	2024-10-28 18:12	vizapps_admin	10.19.18.85	View	Dashboard	6095	6095	6097	632		SQLite	false			SHOW SQL
Done	2024-10-28 18:11	vizapps_admin	10.19.18.85	View	Dashboard	6095	6095	6097	632	samples	SQLite	true		0.043472	SHOW SQL
Cancelled	2024-10-28 18:07	vizapps_admin	10.19.18.85	View	Dashboard	6095	6095		632		SQLite	false			SHOW SQL
	2024-10-28 18:06	vizapps_admin	10.19.16.19	View	Direct Query							false			
	2024-10-28 18:06	vizapps_admin	10.19.16.19	View	Direct Query							false			
	2024-10-28 18:06	vizapps_admin	10.19.16.19	View	Direct Query							false			
	2024-10-28 18:05	vizapps_admin	10.19.16.19	Update	Direct Query							false			SHOW SQL
Done	2024-10-28 18:05	vizapps_admin	10.19.16.19	View	Dashboard						SQLite	false		0.978228	SHOW SQL
Done	2024-10-28 18:04	vizapps_admin	10.19.16.19	View	Data Connection	1					SQLite	false		0.010330	SHOW SQL

Rows per page: 10 11–20 of 500

- State indicates the completion status of the SQL query. Possible values are: Started, Done, Queued, Cancelled, Abandoned, Error.
- Time displays the date and time of the audited action.
- User shows the username of the person who initiated the activity (if available).
- IP provides the IP address of the computer used by the user carrying out the action (if available).
- Action indicates the type of event that triggered the logged activity (View, Create, Update, Delete).
- Component Type specifies the category of the activity (Dashboard, Data Connection, Dataset, User, App, Custom Style, Custom Color, Snapshots, Thumbnail, Static Asset, Role, Custom Email, Custom Date Range, Workspace, Search, Direct Query, Segment, Filter Association, Job, Extract).
- Component ID reflects the ID of the component involved in the action.
- Visual displays the ID of the visual that generated the action or is the target of the action. You can click this number to navigate to the visual.
- Dashboard displays the ID of the dashboard that generated the action or is the target of the action. You can click this number to navigate to the dashboard.
- Dataset displays the ID of the dataset that generated the action or is the target of the action. You can click this number to access the Dataset Detail interface for that dataset.

- Connection displays the ID of the data connection associated with the activity. You can click this name to navigate to the Datasets interface for that connection.
- Type specifies the type of data connection (for example: SQLite, Hive, Impala).
- Cache indicates whether the query was served from the cache instead of executing the SQL query. Possible values are true or false.
- Extract Used shows when SQL query rewriting is enabled. A link to a data extract indicates that the query was redirected to a different data table.
- Runtime shows the execution time of the SQL query in seconds (if SQL execution is involved).
- Query shows the SQL query statement (when available). By default, Cloudera Data Visualization hides the SQL statements. To view a specific query, click Show SQL. To hide statements, click Hide SQL.

Usage Analytics [Technical Preview]

Logs and Usage is a powerful monitoring interface that can help you better understand how users are engaging with Cloudera Data Visualization in your organization. The different tabs provide you detailed information on various aspects, including user activities, system load, and runtime performance. The Usage Analytics tab is a dynamic dashboard featuring visualizations generated from the audit logs presented on the Activity Log tab.

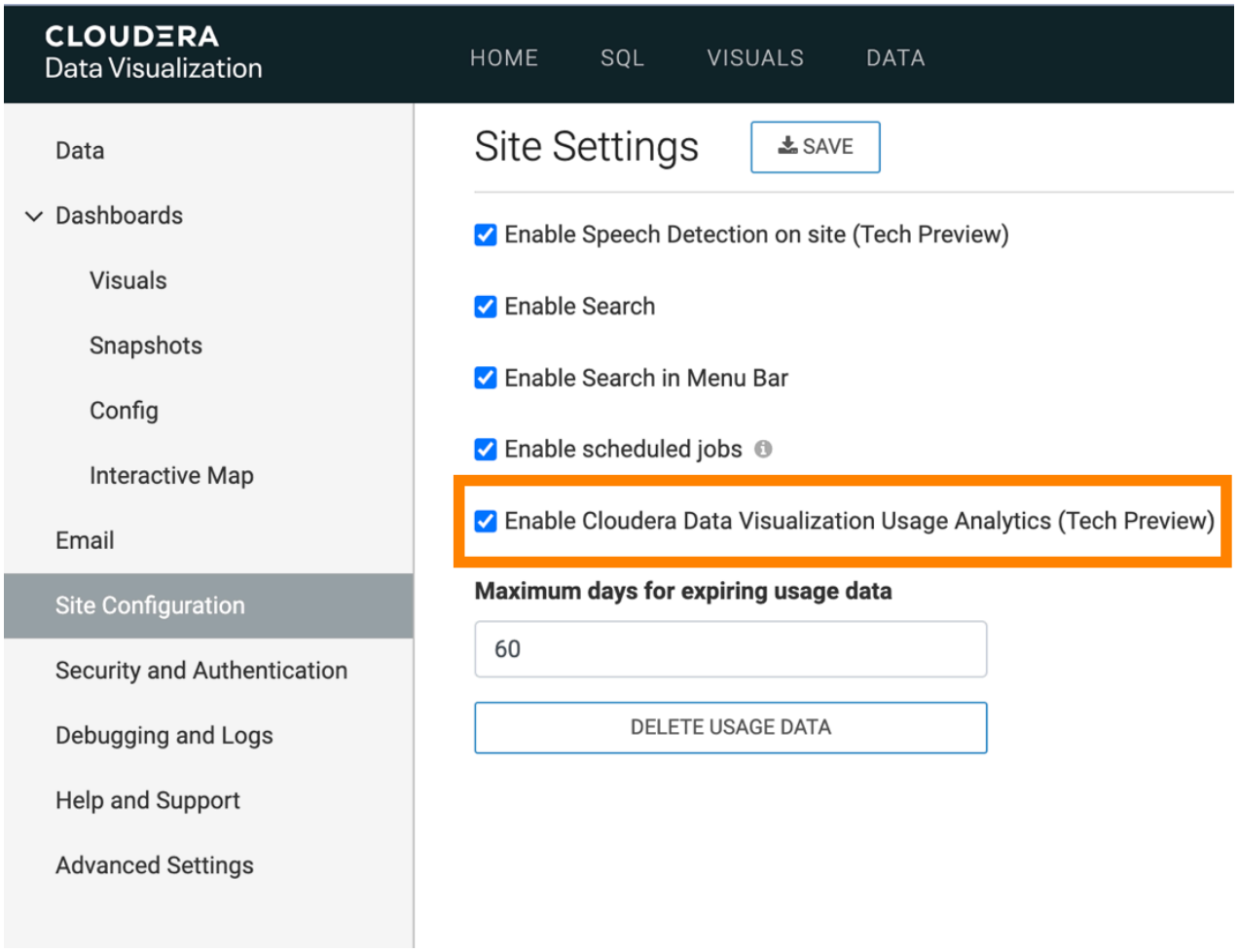
You can use Usage Analytics to get valuable insights into how users engage with Cloudera Data Visualization. You can get information on the most active users, frequently used datasets and visual types, peak activity hours, and other critical aspects of Cloudera Data Visualization operations, all through pre-built visualizations.

You can access the Usage Analytics through Site Administration Logs and Usage .

Note:

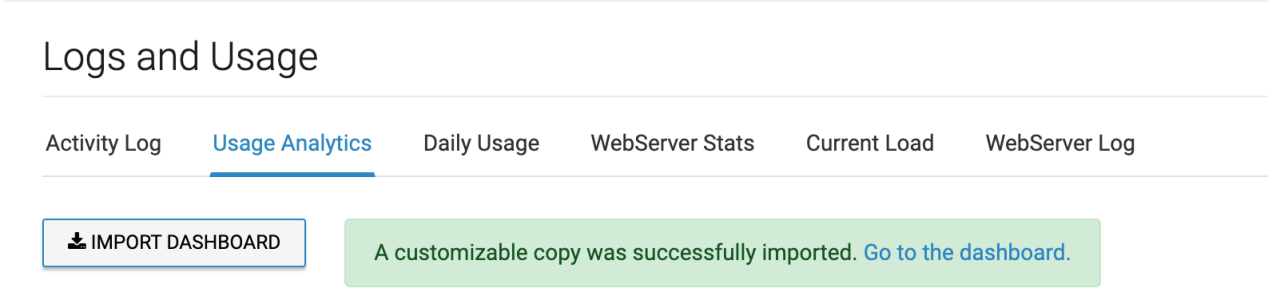
The Logs and Usage interface is only accessible to system administrators (or to users who have a role with View activity logs or Admin user permissions).

You have to enable Usage Analytics in Site Settings to add it to the Logs and Usage interface. For more information, see *Site Configuration*.



The dashboard offers a time range selector that enables you to specify the data used for the visualization. Filters are available for action type, visual type, and username to refine the data displayed.

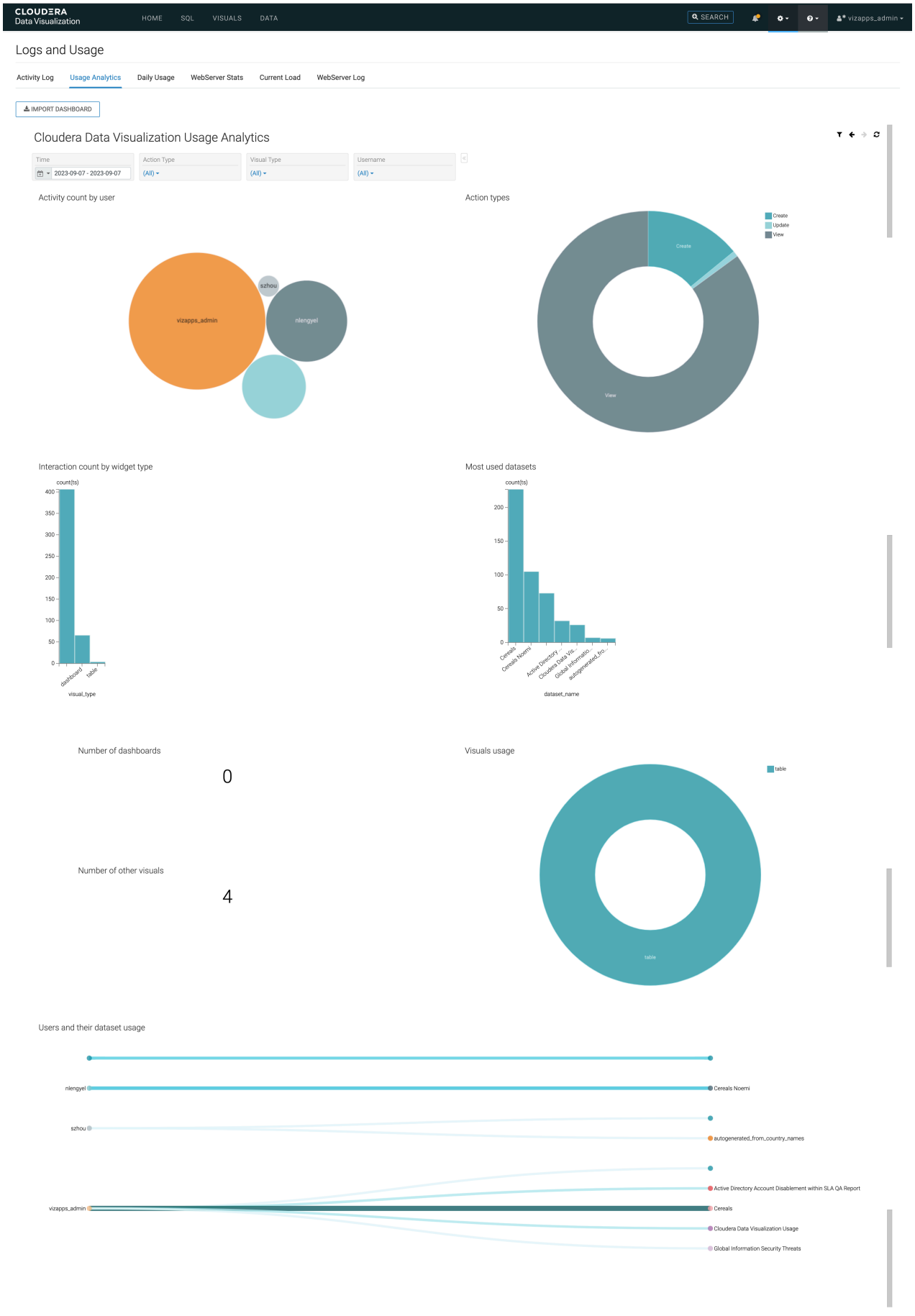
The Usage Analytics view is further customizable. You can click IMPORT DASHBOARD to add an editable version labeled Cloudera Data Visualization Usage Analytics to your collection of dashboards.



By default, the Usage Analytics dashboard includes the following visualizations:

- Activity count by user: Tracks user activity and provides insights into user engagement.
- Action types: Displays a breakdown of various action types, helping you understand user interactions.
- Interaction count by widget type: Provides insights into widget interactions and usage.
- Most used datasets: Highlights frequently accessed datasets.
- Number of dashboards: Shows the count of dashboards in use.
- Visuals usage: Offers insights into the usage of visual components.
- Number of other visuals: Indicates the usage count of other types of visuals.

- Users and their dataset usage: Provides information about users and their dataset usage patterns.
- Dataset usage over day's hours: Shows dataset usage trends throughout the day.


CLOUDERA
Data Visualization
HOME
SQL
VISUALS
DATA
SEARCH
vizapps_admin
Logs and Usage
Activity Log
Usage Analytics
Daily Usage
WebServer Stats
Current Load
WebServer Log
IMPORT DASHBOARD
Cloudera Data Visualization Usage Analytics
Time
2023-09-07 - 2023-09-07
Action Type
(All)
Visual Type
(All)
Username
(All)
Activity count by user
vizapps_admin
szhou
nlengyel
Action types
Create
Update
View
Interaction count by widget type
count(ts)
dashboard
table
visual_type
Most used datasets
count(ts)
Cereals
Cereals Noemi
Active Directory ...
Cloudera Data Vis...
Global Informatio...
autogenerated_fro...
dataset_name
Number of dashboards
0
Number of other visuals
4
Visuals usage
table
Users and their dataset usage
nlengyel
Cereals Noemi
szhou
autogenerated_from_country_names
Active Directory Account Disablement within SLA QA Report
vizapps_admin
Cereals
Cloudera Data Visualization Usage
Global Information Security Threats
Dataset usage over day's hours
Record Count

Related Information
Site Configurations

Daily Usage

Logs and Usage is a powerful monitoring interface that can help you better understand how users are engaging with Cloudera Data Visualization in your organization. The different tabs provide you detailed information on various aspects, including user activities, system load, and runtime performance. The Daily Usage tab of the interface provides you information on activities, users, and dashboards from the audit logs grouped by day.

Note:

The Logs and Usage interface is only accessible to system administrators (or to users who have a role with View activity logs or Admin user permissions).

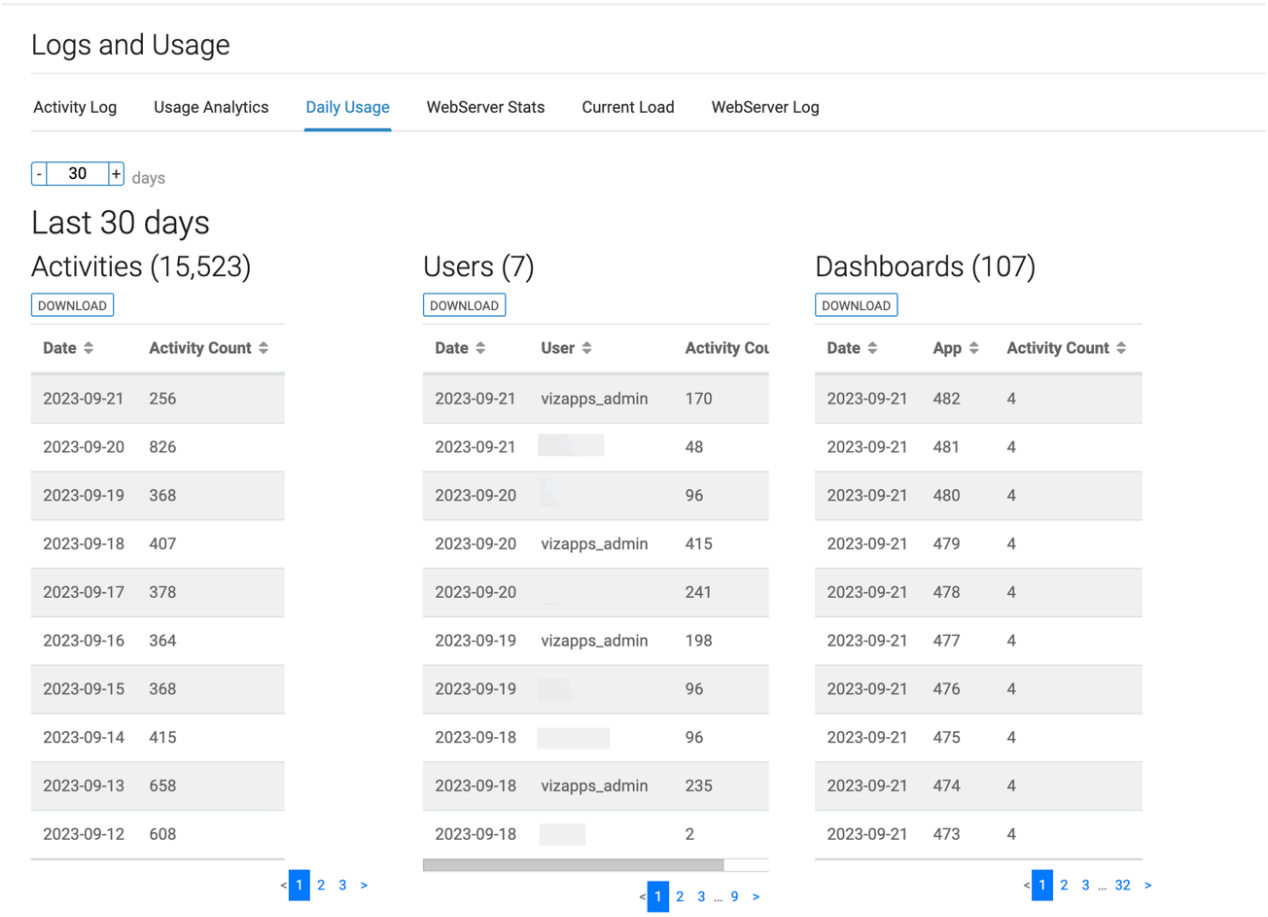
You can access Daily Usage through Site Administration Logs and Usage .

The following information is available on the Daily Usage tab:

- Activities – This table provides a comprehensive view of all audit events that occurred on a specific day, encompassing actions like creates, reads, writes, and deletes
- Users – This table presents audit events categorized by day and user, showing the number of audited actions associated with each user.
- Dashboards – This table shows the number of activities impacting or related to visual components, grouped by date and dashboard/visual ID.

By default, Daily Usage shows data for the last 30 days, but you can customize the time period to suit your analytical needs.

You can download the collected data in CSV format, enabling further analysis and integration with other tools and systems.



WebServer Stats

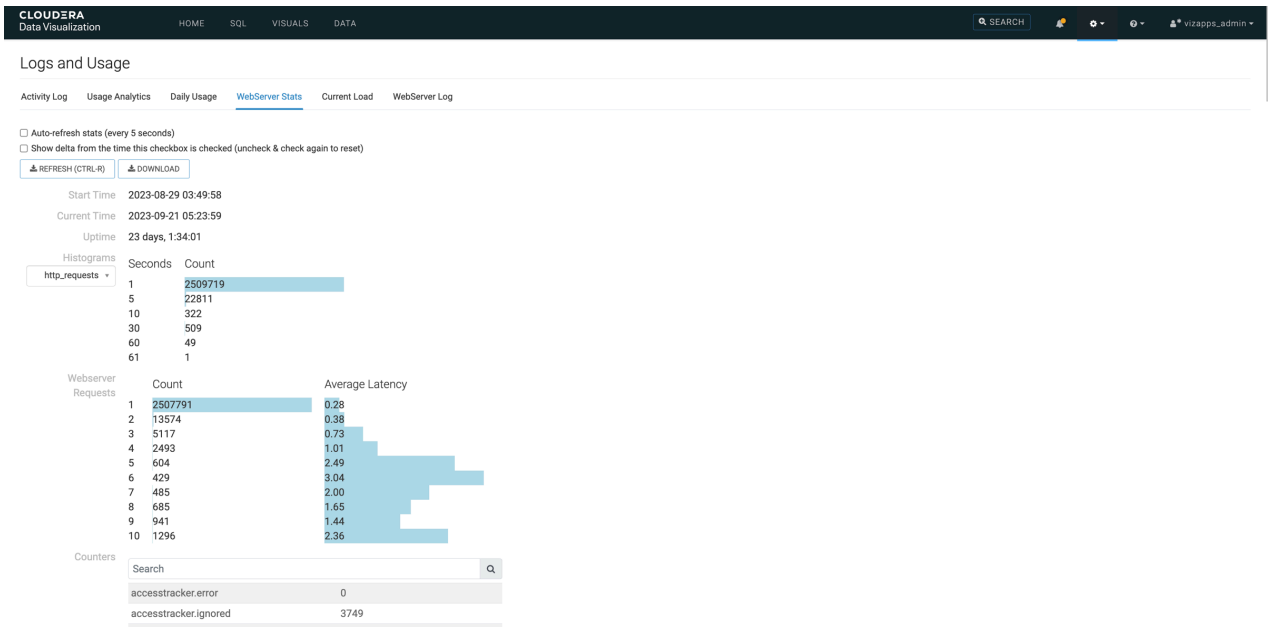
Logs and Usage is a powerful monitoring interface that can help you better understand how users are engaging with Cloudera Data Visualization in your organization. The different tabs provide you detailed information on various aspects, including user activities, system load, and runtime performance. The WebServer Stats tab of the interface provides runtime information. You can use it for performance tuning and troubleshooting. Monitoring runtime can help you identify performance bottlenecks, and plan how to optimize system performance.

Note:

The Logs and Usage interface is only accessible to system administrators (or to users who have a role with View activity logs or Admin user permissions).

You can access WebServer Stats through Site Administration Logs and Usage .

Most of the information displayed as webserver statistics is not permanently stored. Restarting the Cloudera Data Visualization instance significantly impacts the data presented.



Key features and data:

- Auto-Refresh

  You can set the interface to auto-refresh at 5-second intervals for real-time data updates.
- Delta

  You have the option to view delta information from a specific point in time.
- Start Time

  This timestamp indicates when Cloudera Data Visualization was initially started.

  The default time zone for all time data received from the server is Pacific Standard Time (PST). You can configure the time zone in Advanced Site Settings. For instructions, see *Time zone* in the Advanced Settings documentation.
- Current Time

  It shows the time when the current page was retrieved.
- Uptime

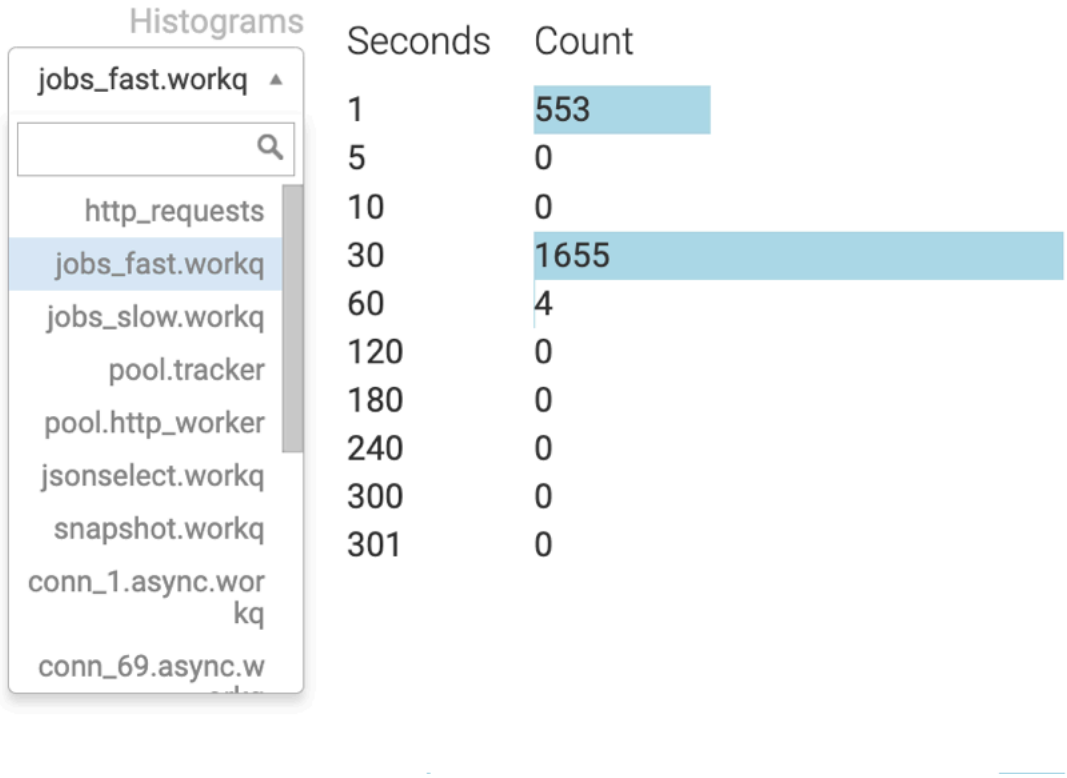
  It shows the time elapsed since Cloudera Data Visualization was started, presented in the format of days, hours:minutes:seconds.
- Histograms

  This section provides insights into execution times within worker pools collected in several buckets (6 or 10).

  Worker pools (thread pools) are used for executing tasks asynchronously, such as database queries. This view provides information on the performance characteristics of tasks executed by worker pools. It helps assess the

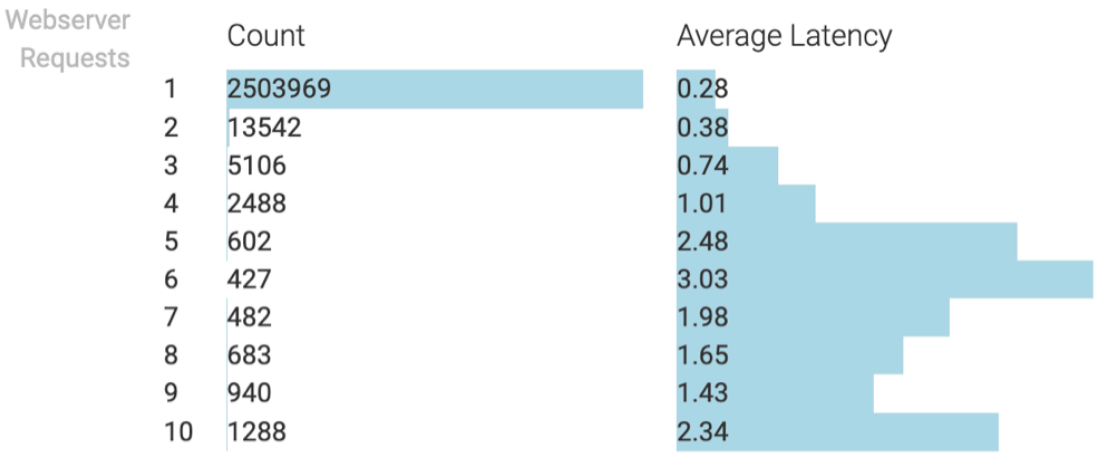
access speed of specific data sources. The drop-down menu allows you to select different worker pools to gauge resource access speed.

Note: The histogram is constructed based on task (query) execution time, excluding the time spent in the queue.



- Webserver Requests

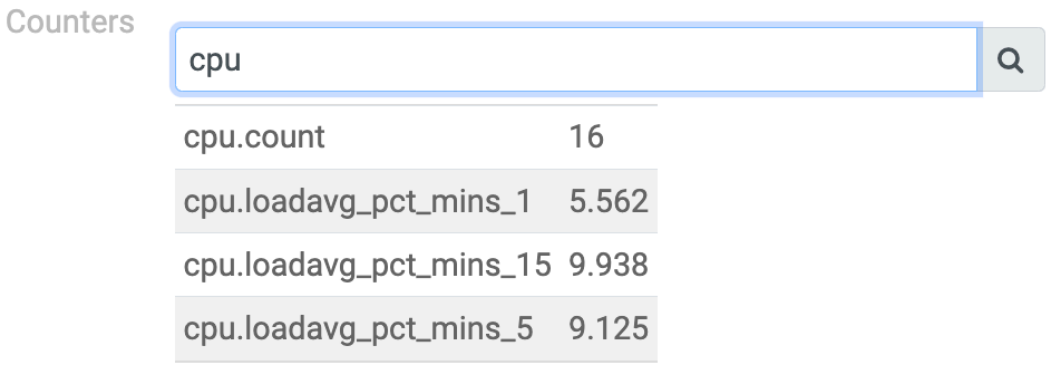
  This section provides information regarding concurrent HTTP request handling. Since there are 10 HTTP worker threads, a maximum of 10 rows are displayed. The Count column shows the number of cases where the server had to manage a specific number of parallel requests (ranging from 1 to 10). The Average Latency column shows the average response time for requests within a given category.



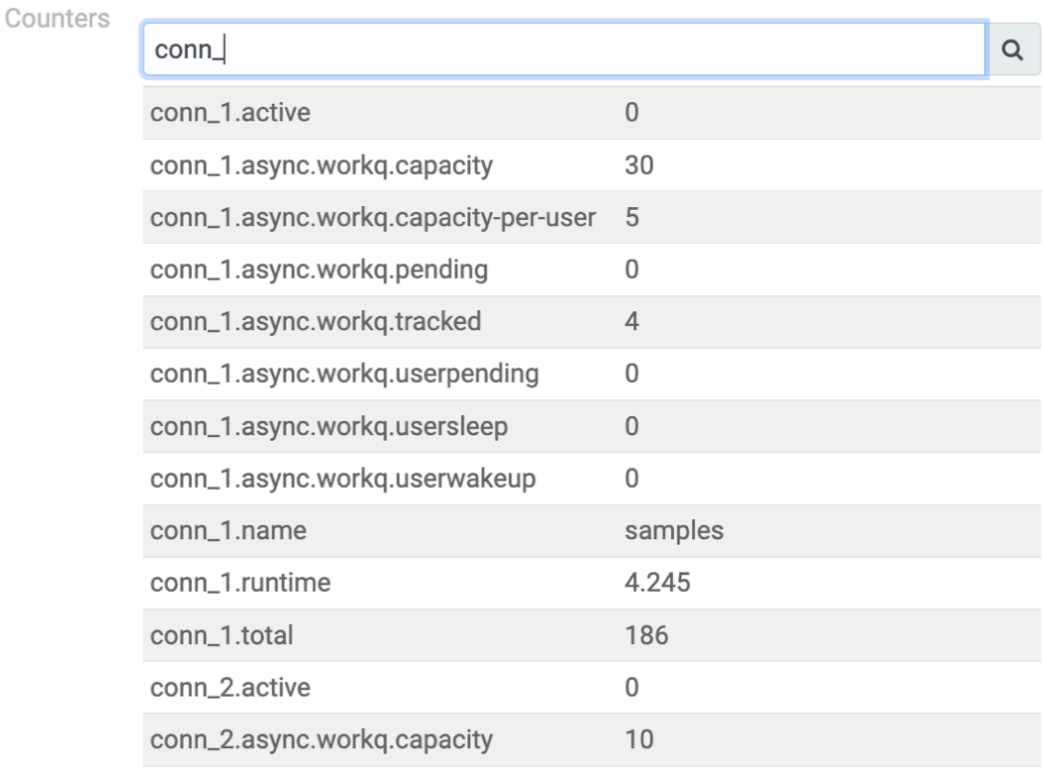
- Counters

  This section provides data for debugging purposes. It lists key-value pairs, where keys are dot-separated strings.

  - Keys starting with cpu offer CPU-related information:



  - Keys starting with conn_ are related to a particular data connection:



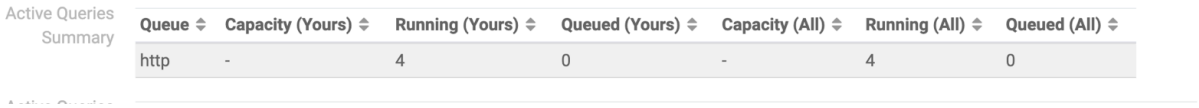
  - Explanation of the connection related counters:
    - active: It represents the number of queries currently being executed. A value of 0 indicates an idle connection, with no queries being executed.
    - async.workq counters:
      - capacity, capacity-per-user display the values of concurrency and concurrency per user settings for the data connection. These values remain constant unless you edit the data connection.
      - pending shows the number of queries waiting in the queue to be processed. For an idle connection, this value is 0.
      - Tracked shows the number of queries that were queued and are still awaiting to be delivered to the UI. For an idle connection, this value should be 0.

        - Userpending displays the number of queries waiting for execution due to exceeding the per-user concurrent execution limit.
        - Usersleep is the cumulative count of events when a query could have been executed because there were free workers in the pool, but was postponed because of the user-specific limit.
        - Userwakeup is the cumulative count of events when a query was rescheduled after the user-specific concurrency dropped below the configured limit.
    - name: It denotes the name of the data connection.
    - runtime: It indicates the cumulative runtime of queries, excluding queue time.
    - total: It represents the number of queries sent to the database backend. You can calculate the average execution time as runtime/total.
- Explanation of other counters:
    - accesstracker.*: It pertains to the Access tracking feature, which is no longer in use. It always displays only ignored items.
    - arcengine.create: It indicates the total number of legacy Arcengine, legacy Impala, or legacy Hive connections defined among the data connections.
    - arcengine_v2.create: It indicates the total number of non-legacy Arcengine connections. Non-legacy Impala and Hive connections have their respective counters: 'impyla' and 'hyve'.
    - arcengine_v2.connect: It tracks the number of times a non-legacy Arcengine connection was initiated to some of the defined data connection targets.
    - async.*: It provides metrics on specific HTTP endpoints marked as 'asynchronous'. These endpoints use a worker thread ('http_worker') to run incoming requests, returning a task handle for client-side result polling.
        - Async.jsonselect_parallel_async measures the cumulative number of scheduled tasks, typically data requests sent by dashboard visuals.
        - async.asyncstatus_async captures the cumulative number of client queries checking the status of previous requests. A considerably higher count compared to 'jsonselect_parallel_async' may indicate slow data retrieval (multiple checks are needed before getting the results to the client).
        - async.asyncresults_async records the number of times results were fetched by the client, which is typically low or 0, as status checks often yield data as a sign of completed processing.
        - async.asynccancel_async indicates the cumulative number of query cancellations not initiated explicitly by the user, often occurring when closing the browser page before data retrieval is complete.
        - Async.asynccancel_qt tracks the total number of tasks canceled explicitly by the user by clicking CANCEL ALL MY QUERIES on the current load page.
        - pool.http_worker.concurrency represents the number of concurrent task executions at the time of data collection, typically 0 in an idle system.
    - hive.create: It reflects the number of legacy Hive connections.
    - hs2.*: (hs2.conncreate, hs2.opensession, hs2.poolcreate, hs2.queries_retry, hs2.queries_submit): It contains legacy Hive Server 2 connection-related information (no longer maintained).
    - http_async.*: It offers information on the asynchronous HTTP-handling infrastructure.
        - Http_async.calls shows all HTTP client requests handled by the async framework.
        - Http_async.tasks shows the number of requests that resulted in a task being scheduled.
        - Http_async.retries shows the number of requests responded immediately (without scheduling a task).
    - hyve.create: It tracks the number of non-legacy Hive data connections defined by users.
    - hyve.connect: It monitors the number of connections established to non-legacy Hive data connections.
    - Impyla.create: It tracks the number of non-legacy Impala data connections defined by users.
    - impyla.connect: It monitors the number of connections established to non-legacy Impala connections.
    - jobs_fast.workq.*: It contains information related to email jobs.
        - Jobs_fast.workq.capacity shows the number of worker threads assigned for email jobs.
        - Jobs_fast.workq.pending shows the number of email jobs waiting to be processed. In an idle system this should be 0.

    - tracked, userpending, usersleep, userwakeup all have values that are constant zeros. They are inherited from the base worker queue implementation.
- jobs_slow.workq.*: It offers data extracts related information, similar to the information jobs_fast.workq provides for email jobs.
- jsonselect.workq.*: It provides details on an additional worker pool managing data queries. Entries have the same meaning as for jobs_fast and jobs_slow counters above.
- memory.rss_mb: It displays the memory resident set size (in megabytes).
- pool.tracker.concurrency: It indicates the concurrency level, related to access tracking. This counter is currently unused.
- query_stat.*: It encompasses cumulative data for existing data connections. In contrast to other counters, these values are persisted in the metadata store and are not reset by a server restart.
    - Query_stat.total_number_of_queries_returned shows the sum of all queries issued for all currently defined data connections since the time they were created.
    - Query_stat.returned_from_backend shows the number of queries that were actually sent to the backing data source.
    - Query_stat.returned_from_cache shows the number of queries that were served from the local cache.
- requests.concurrency_overflow: It reflects the number of times an HTTP request was ignored when calculating webserver requests statistics. Typically 0 due to the current concurrent limit of 10. This counter is currently unused.
- response.totalsize: It represents the total number of bytes returned by the HTTP requests, excluding streaming responses.
- rowdata.count: It captures the total number of data rows returned by non-streaming data queries.
- snapshot.queue: It shows the number of times a snapshot was created either when creating/updating a visual or when saving a dashboard.
- snapshot.workstarted: It tracks the total number of snapshoting tasks started, including snapshots created by email jobs and other actions.
- snapshot.workcompleted: It records the total number of completed snapshot tasks, typically matching snapshot.workstarted in an idle system.
- snapshot.workq.*: It contains snapshot worker queue counters, with meanings similar to jobs_fast.workq.*.
- workq.error: It shows the number of unhandled exceptions that occurred when executing tasks in worker pools. A value of 0 indicates stability.
- workq.markInterrupted: It captures the number of tasks not executed because they spent more time waiting in the queue than the specified execution timeout.
- workq.querycleanup.create: It indicates the number of times a cleanup thread was created for worker queues associated with data connections with a query timeout set. If the query timeout is modified, the cleanup thread is recreated.
- workq.querycleanup.cancels: It reflects the number of times a worker queue cleanup thread was canceled due to user changes in connection timeout settings.

- Active Queries Summary

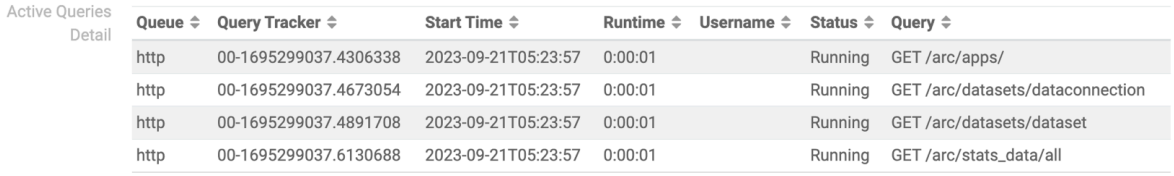
  It provides a summary of queries/tasks executed or queued by various worker pools.


Active Queries Summary

Queue	Capacity (Yours)	Running (Yours)	Queued (Yours)	Capacity (All)	Running (All)	Queued (All)
http	-	4	0	-	4	0

  - Queue is the name of the worker pool queue.
  - Capacity (Yours) shows the user-specific concurrency limit, if the queue supports it (data connection queues support it, others do not (like the http worker pool)).
  - Running (Yours) shows the number of tasks/queries running on behalf of the logged-in user.
  - Queued (Yours) shows the number of tasks queued for the current user.
  - Capacity (All) shows the overall number of tasks that can be executed in parallel by the worker pool.
  - Running (All) shows the number of all tasks executed by the pool across all users.
  - Queued (All) shows all tasks waiting in the pool to be processed for all users.
- Active Queries Detail

  It provides information about tasks being executed or queued.


Active Queries Detail

Queue	Query Tracker	Start Time	Runtime	Username	Status	Query
http	00-1695299037.4306338	2023-09-21T05:23:57	0:00:01		Running	GET /arc/apps/
http	00-1695299037.4673054	2023-09-21T05:23:57	0:00:01		Running	GET /arc/datasets/dataconnection
http	00-1695299037.4891708	2023-09-21T05:23:57	0:00:01		Running	GET /arc/datasets/dataset
http	00-1695299037.6130688	2023-09-21T05:23:57	0:00:01		Running	GET /arc/stats_data/all

  - Queue is the name of the worker pool queue.
  - Query tracker is a task identifier.
  - Start Time shows the time when the task was submitted to the queue.
  - Runtime shows the time elapsed since the task was submitted.
  - Username displays the user to whom the task belongs.
  - Status indicates whether the task is Pending or Running.
  - Query displays the path of the SQL query or HTTP request, if available.

Current Load

Logs and Usage is a powerful monitoring interface that can help you better understand how users are engaging with Cloudera Data Visualization within your organization. The different tabs provide you detailed information on various aspects, including user activities, system load, and runtime performance. The Current Load tab of the interface offers the convenience of viewing essential information in real-time and provides a quick way to manage ongoing queries by canceling them if needed.

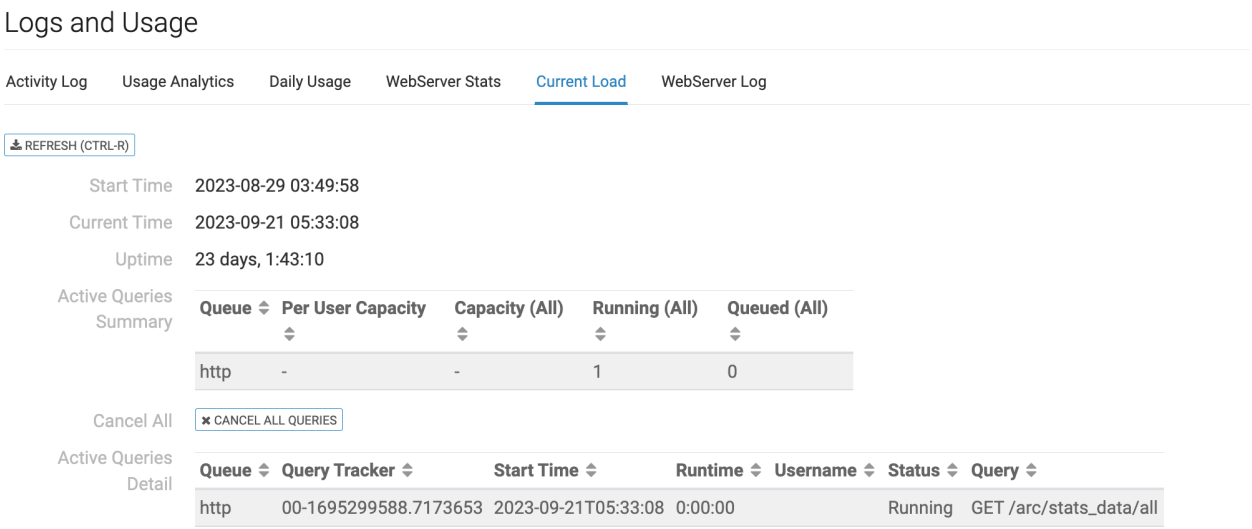
Note:

The Logs and Usage interface is only accessible to system administrators (or who have a role with View activity logs or Admin user permissions).

You can access the Current Load view through Site Administration Logs and Usage .

Current Load provides the following information:

- Data duplicated from WebServer Stats: Start Time, Current Time, Uptime, and Active Queries Detail.
- Active Queries Summary presenting aggregated data for all users. It is a consolidated view of the system's load but does not provide user-specific details.
- Option to cancel all queries currently in progress.



WebServer Log

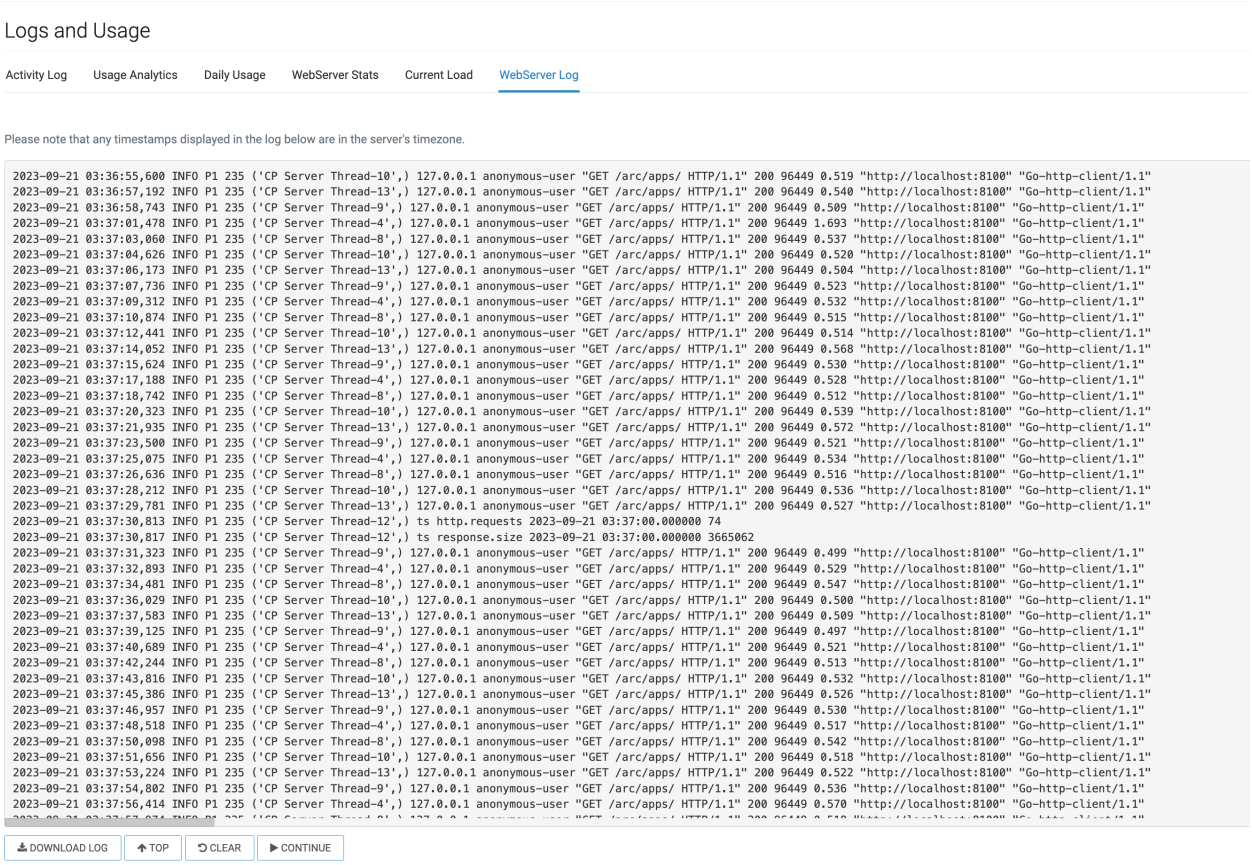
Logs and Usage is a powerful monitoring interface that can help you better understand how users are engaging with Cloudera Data Visualization in your organization. The different tabs provide you detailed information on various aspects, including user activities, system load, and runtime performance. The WebServer Log tab provides access to the log messages generated by the Cloudera Data Visualization application.

You can use the technical details as a resource for debugging Cloudera Data Visualization issues or download the log files and share them with Cloudera Support.

Note:

The Logs and Usage interface is only accessible to system administrators (or to users who have a role with View activity logs or Admin user permissions).

You can access the WebServer Log view through Site Administration Logs and Usage .



1. WebServer Activity log window

   The activity log is presented as a flat file, featuring timestamps for each event, followed by technical details.
2. DOWNLOAD LOG

   You can save log files to your local system. The downloaded files follow the naming convention arcv iz_yyyy_mm_dd_hh_mm_ss.log. For example, a file saved on April 17, 2018 at 1:46:42 P.M. has the name arcv iz_2018_04_17_13_46_42.log.
3. TOP

   You can use this option to swiftly return to the top of the activity list after scrolling down.
4. CLEAR

   You can use this option to empty the log window ensuring a clean slate for new log data.
5. PAUSE/CONTINUE

   You can use this option to temporarily suspend/resume log reporting.