Workload XM 2.3.0

Workload XM Cluster Optimization

Date published: 2020-12-04
Date modified: 2023-01-26

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Contents

Working with Workload XM

Tasks for identifying and troubleshooting job and query abnormalities and failures, optimizing workloads, and improving job performance with Workload XM.

Specifying a time range

Choose a time period in which your workload results are displayed in Workload XM for analysis and troubleshooting.

About this task

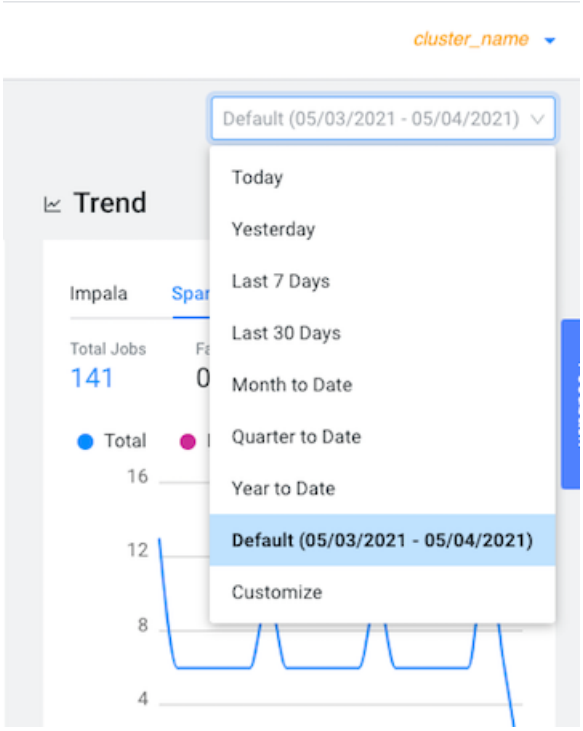
Describes how to specify a time period for displaying current or historical data about your cluster and the jobs performed in that cluster.

By default, Workload XM displays workload data for the last 24 hours. If there is no data available during that time, Workload XM displays the nearest date range that is available.

Procedure

1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. In the Clusters page, select the cluster required for analysis.

3. From the time-range list in the Cluster Summary page, do one of the following:
   - For a predefined period, select one of the default periods of time that meets your requirements.
   - For an exact date and time range, select Customize and then either, enter the date and time range using the YYYY/MM/DD HH:MM:SS format for the beginning and the ending time period, or in the calendar element, select the beginning and ending time period.



4. Click Ok, which clears any existing workload data from the chart and table components and updates your workload data for the chosen time period.

Results
All charts and tables in Workload XM are updated to reflect the workload data for the chosen time period.

Analyzing Your Workload Cluster Costs with Workload XM Cost Centers

Define customized cost centers based on user or pool resource criteria and CPU and memory consumption with the Chargeback feature. Once defined Workload XM visually displays a Workload cluster's current and historical costs. With these cost insights you can then plan and forecast budgets and future workload environments and/or justify current user groups and resources.

Creating a Workload XM Cost Center

Create Workload XM cost centers that enable you to display your current and historical workload cluster and resource costs that can be used for planning, budgeting, and forecasting future workload environments.

About this task
Describes how to configure your Workload XM Chargeback settings, which define your cost centers and the unit costs of your resources, and create a Workload XM cost center.

Note: To avoid cost duplication, resources must only be assigned one cost center.

Procedure

1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. From the Workload XM Navigation side-bar, select Chargeback.
3. Globally define your cost center criteria and memory and CPU costs by clicking Chargeback Setup.
4. From the Setup page, do the following:
   a. From the Select Chargeback criteria section, select your cluster's chargeback criteria.

   Note: Cost centers are associated with a specific criteria. If you later change the Chargeback criteria setting the cost centers associated with the previous selection are hidden. You can revert back to these cost centers by reselecting their Chargeback criteria.

   To revert back to previous cost centers, from the Actions list on the Chargeback page, select Chargeback Settings and then reselect their criteria option.
   b. From the Cluster list of the Cluster Selection section, select the clusters required for your cost centers. Where, the cost calculations use resource utilization for each of your chosen clusters.
   c. In the CPU field of the Unit cost section, enter the amount, in dollars, for each CPU core hour.
   d. In the Memory field of the Unit cost section, enter the amount, in dollars, for each Gigabyte hour.
   e. Click Complete Setup.
5. From the Chargeback page, create a new cost center by clicking Create a Cost Center.
   a. In the Name field, enter a unique name for your cost center.
   b. (Optional) In the Description field, enter a meaningful description for the cost center.
   c. Depending on the Chargeback criteria value you selected when you configured your Chargeback settings, do one of the following:
      - If you selected Pool, in the Add Pools field, enter one or multiple resource pools.
      - If you selected User, in the Add Users field, enter one or multiple users.
   d. Click Create.

Results

Once you have configured your Chargeback settings and created a cost center you can view your job costs associated with a cost center cluster.

Displaying Your Job Costs Associated with a Cost Center Cluster

Steps for displaying your Workload cluster jobs associated with a cost center cluster.

About this task

Describes how to view your workload costs associated with a cluster.

Procedure

1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. From the Workload XM Navigation side-bar, select Chargeback.
3. In the Chargeback page, select a cost center.

   Your cost center page opens displaying the costs, and the CPU and memory usage associated with the cost center.
4. To view more details about the pool, user, and job costs for a specific cluster in the cost center, from the Cluster column, locate the cluster and then either click its name or click the greater-than arrow (>) at the end of its row.

Assigning Uncategorized Resources to a Cost Center

Steps for moving unassigned resources into an existing or a new Workload XM cost center.

About this task

Describes how to locate and move uncategorized resources into an existing or a new Workload XM cost center.

Note: To avoid cost duplication, resources must only be assigned one cost center.

Procedure

1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. From the Workload XM Navigation side-bar, select Chargeback.
3. In the Chargeback page, select a cost center and then a cluster.
4. From the Overview tab, scroll down and click inside the Uncategorized section.

   The Uncategorized page opens.
5. Select the required uncategorised resource tab.
6. From either the Pools, Users, or Clusters page, select the check boxes of the resources you require for your cost center.

   The Assign Cost Center button becomes visible.
7. Click Assign Cost Center.
8. From the Select Cost Center list, do one of the following:
   a. To add the uncategorized resource/s in a new cost center, select New Cost Center and then click Create a new cost center.
   b. To add the uncategorized resource/s in an existing cost center, select an existing cost center and then click Assign to Cost Center.
9. (Optional) Repeat steps 4-8 until all your uncategorized resources are placed in your Workload XM cost centers.

Triggering Actions across Jobs and Queries

You can trigger action alerts, that are defined by you, across your workload applications, jobs, and queries whilst they are running with the Workload XM Auto Actions feature. When a workload application, job, or query matches the action's defined threshold value, the auto action event is triggered. For example, you may have a scenario where too much memory is being allocated to specific jobs and you would like to take an action before a problem occurs, such as avoiding memory exhaustion. In this case, you can create an auto action that triggers a notification alert when a job is identified as having an over-allocation of memory. You can then either manually take steps to alleviate the problem or include the Kill action option that stops the job in question.

Important: Before you can use the Auto Actions feature you must set the required auto actions configuration properties in Telemetry Publisher. For information on how to enable the Telemetry Publisher Auto Actions property settings, click the Related Information link below.

Considerations and Limitations

The following describes consideration decisions and limitations when using Auto Actions:

- Killing a workload application, job, or query could impact other workloads. Especially when another workload is dependent on the results of the killed workload application, job, or query. Before triggering a Kill type action, Cloudera recommends using the Notification only action alert until you have verified that no issues will arise.
- By default, the Workload XM UI limits the amount of displayed audit events to 500 and sorts them in ascending order (newest time stamp first). To display older audit events, change the date range duration and/or the time range duration from the time-range list on the Auto Actions Events page.
- Too Fast To Collect: The minimum polling interval is one minute. If you have jobs or queries that overlap or start before the minimum polling interval is completed there may not be enough time for Workload XM to evaluate your auto action’s definition.

  For example, if Workload XM starts polling at 1:00:00 PM and polling finishes by 1:00:10 PM (10 seconds) and then a job starts at 1:00:12 PM and finishes before 1:01:00 PM, there is not enough of a time lapse for Workload XM to evaluate and trigger your action alert.
- Too Fast to Kill: Under normal conditions the evaluation and invocation phases of an auto action is within the span of one minute. If you have jobs or queries whose run time is less than one minute, Workload XM may complete the evaluation phase but not have time to complete the invocation phase, such as killing the job. Depending on the context of your auto action, this may or may not be an issue. But if, for example, you have a workload cluster that is dedicated for specific jobs and a rogue job is run before the action is triggered, then this could be an issue

Related Information

Enabling the Collection of Auto Action Data by Telemetry Publisher

Creating an Auto Action Event

The steps to create a Workload XM auto action definition, which is triggered when a workload application, job, or query matches the auto action’s definition threshold. For example, when a job is taking too long to run it may delay other jobs waiting in the queue. With Auto Actions, you can create an auto action alert that informs you through an email when a job is exceeding its usual runtime so that you can decide whether to manually take steps to alleviate the problem or have an auto action that will kill the job or query process for you.

About this task

Describes how to create a Workload XM Auto Action definition.

Note: These instructions assume that you have set the required auto actions configuration properties in Telemetry Publisher. For information about the properties and how to enable the Telemetry Publisher Auto Actions property settings, click the Related Information links below.

Procedure

1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. From the Workload XM Navigation side-bar, select Auto Actions.
3. Do one of the following:
   - If no other auto actions exist, click Auto Actions Setup.
   - If other auto actions exist, click Create an Auto Action.

   The Auto Actions Create page opens.
4. In the Auto Action Name field, enter a unique name that is easily identifiable.
5. From the Scope list, select the workload component service that is to be monitored by the action.

   For example, if you want your action to only evaluate Spark related applications, you will select Spark Application.
6. Define the conditions for the auto action by doing at least one of the following:
   - Specify the Criteria:
     a. From the Criteria list, select a criteria item.
     b. From the Operator list, select the required operator.

     Important: Workload XM does not validate regular expressions when using the matches regex operator for string criteria types, such as User, Pool, or Query Name. Neither does it display help for poor syntax. Cloudera recommends validating your code and syntax before entering your regular expression in the Value field.
     c. In the Value field, enter the value for this filter.

     Tip: You can define multiple AND filters for the Criteria by clicking the plus sign.

     Note: An Auto Action does not require the Criteria filter as long as a Trigger condition is defined:
     - When included, only those workloads that are run on the selected workload component service and meet the criteria conditions are tested by the Trigger.
     - When not included, all workloads that are run on the selected workload component service are tested by the Trigger.
   - Specify the trigger for the auto action by doing the following:
     a. From the Metric list, select a metric item.
     b. From the Operator list, select the required operator.

     Note: The in between operator is inclusive.
     c. In the Value field, enter the value for this trigger condition.

     Tip: You can define multiple OR conditions for the trigger by clicking the plus sign.

     Note: An Auto Action does not require the Trigger filter as long as a Criteria condition is defined:
     - When included, workloads that are run on the selected workload component service and meet the criteria conditions are tested by the Trigger.
     - When not included, only those workloads that are run on the selected workload component service and meet the criteria conditions are evaluated.

7. From the Select Action options, select the action that is to be performed when the condition is met.

   Warning: Killing a workload application, job, or query could impact other workloads. Especially when another workload is dependent on the results of the killed workload application, job, or query. Before triggering a Kill type action, Cloudera recommends using the Notification only action until you have verified that no issues will arise if the workload application, job, or query is killed.

8. From the Notification section do the following:
   a. In the Emails field, enter the email address that you use to log into Workload XM.
   b. In the Subject field, enter the subject for the email that distinguishes the subject matter from other auto action emails.
9. Click Create, which creates the action and its audit log, adds it on the Auto Actions Events and Management pages, and displays its status as Enabled and its most recent event type as Create.

Results

When a workload application, job, or query meets the auto action's criteria and trigger conditions the action event is triggered.

Related Information

Enabling the Collection of Auto Action Data by Telemetry Publisher

Telemetry Publisher Configuration Settings for Auto Actions

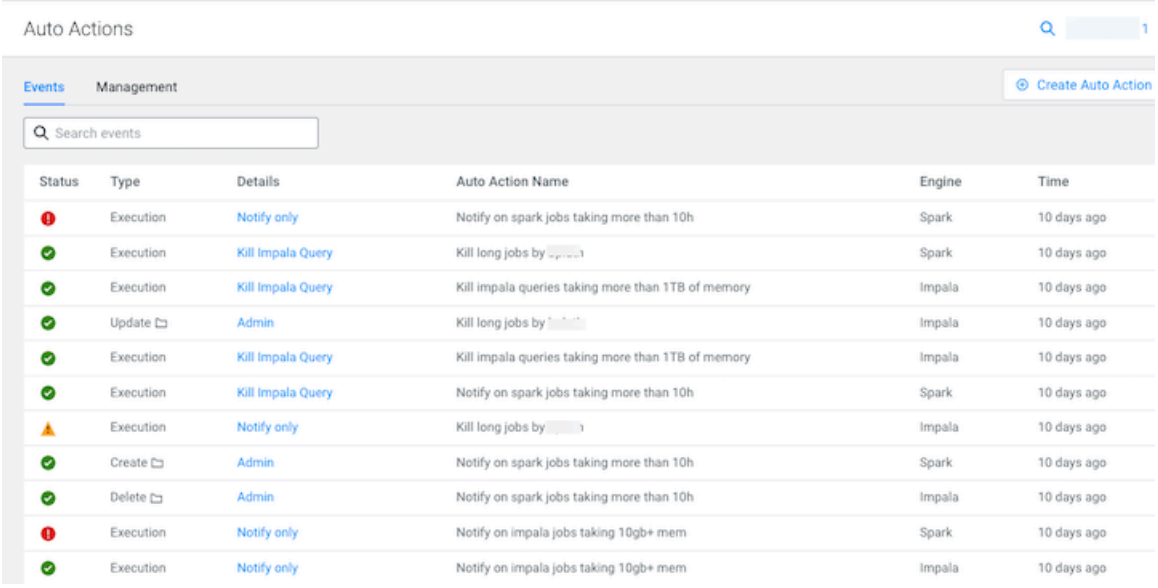
Understanding the Events and Management Fields

Describes the fields in the Auto Actions Events and Management pages.

The Events and Management pages help you monitor, manage, and troubleshoot your Auto Actions.

Events

The Auto Actions Events page displays information about your auto action audit events:



It contains the following audit entry fields:

- Status, which displays the results of the auto action event as an icon. Where, green indicates that the action was successful (SUCCEEDED). All the other icons indicate that the action was unsuccessful (FAILED), such as the action timed out.
- Type, which displays the auto action's audit event category type, such as Create or Update.

- Details, which displays the type of action. When clicked the auto action's Event Details audit log opens, as shown in the following Invoke and Update event type example images:

Figure 1: Invoke Event Type Example

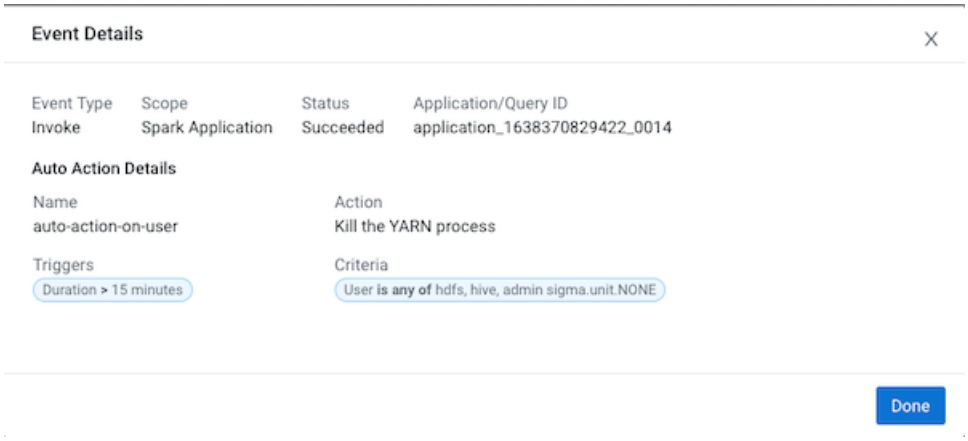
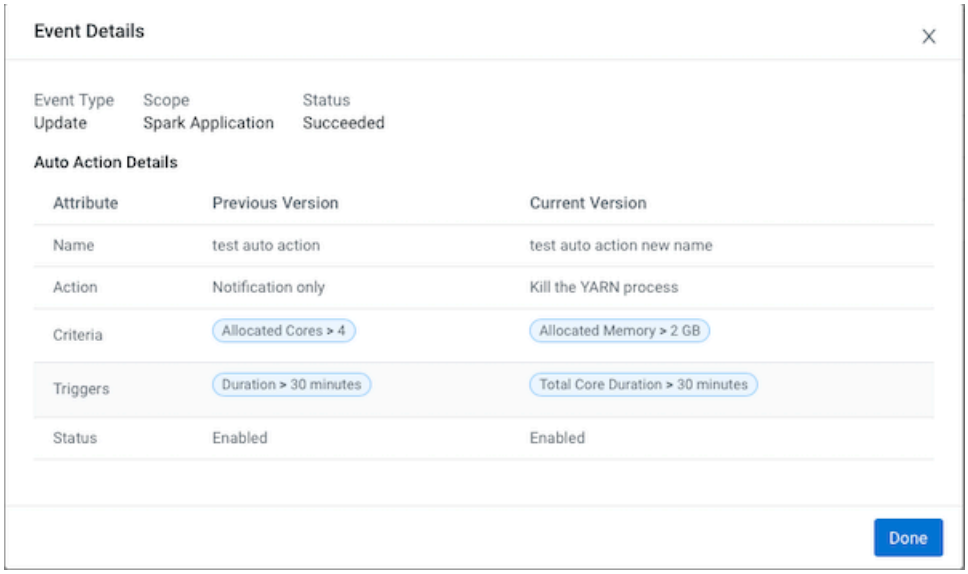


Figure 2: Update Event Type Example

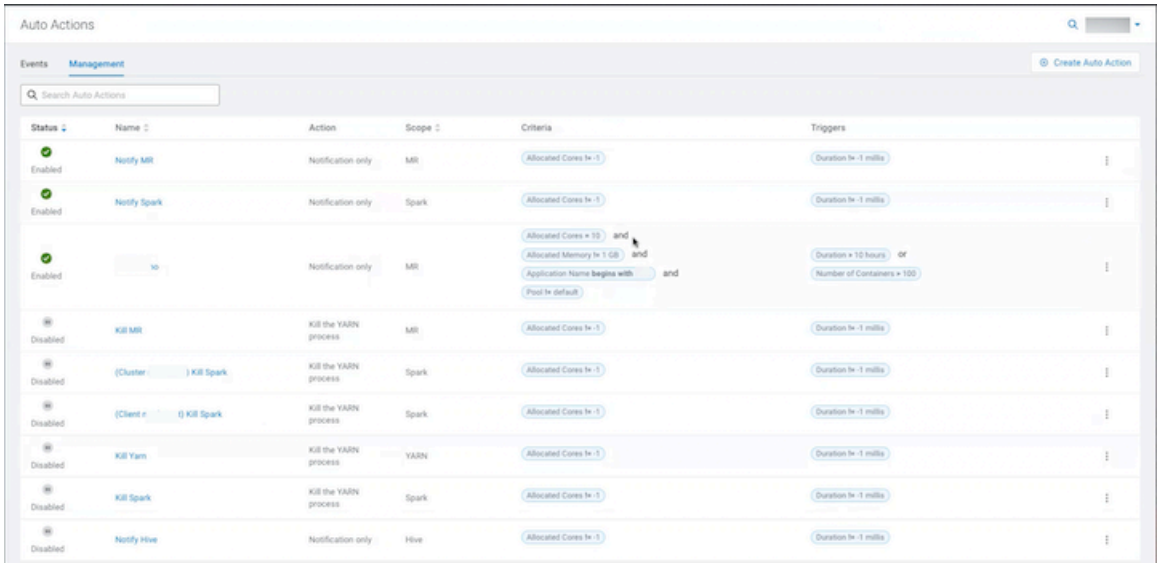


- Auto Action Name, which displays the unique name you entered for the auto action.
- Scope, which displays the workload component service that is monitored by the action.
- Time, which displays the time stamp of when the auto action's audit event occurred.

Important: By default, the Workload XM UI limits the amount of displayed audit events to 500 and sorts them in ascending order (newest time stamp first). To display older audit events, from the time-range list on the Auto Actions Events page, change the date range duration and/or the time range duration.

Management

The Auto Actions Management page displays your auto action's defined settings and state:



It contains the following entry fields:

- Status, which displays the current state of the action, as either Enabled or Disabled.
- Name, which contains the name of the auto action. When clicked the auto action's definition settings page opens.
- Action, which displays the name of the action that is invoked when the auto action is triggered, such as Notify Only.
- Scope, which displays the workload component service that is monitored by the action.
- Criteria, which displays the action's Criteria filters. These are attributes with static values that remain the same during the execution of a job or query.
- Triggers, which displays the action's Trigger conditions. These are attributes with dynamic values that change during the execution of a job or query.

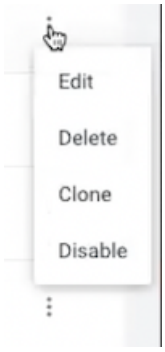
Managing your Auto Actions

Steps for updating, deleting, duplicating, and disabling an auto action.

The following Auto Actions management tasks are performed in the Auto Actions Management page, which is accessed by selecting Auto Actions in the Workload XM Navigation side-bar.

Updating your Auto Action

In the Auto Actions Management page, click the action’s vertical ellipsis, as shown in the following image, and select Edit. Make your changes and then click Update.



Deleting an Auto Action

In the Auto Actions Management page, click the action's vertical ellipsis, and select Delete. In the confirmation message, click OK to confirm. The action is permanently removed.

Note: Unless the action is no longer required, Cloudera recommends disabling the action, as you may require the action at another time.

Duplicating an Auto Action

In the Auto Actions Management page, click the action's vertical ellipsis, and select Clone. Replace the exisiting name with a new unique name for the cloned auto action, make any other changes, and then click Create. A new auto action is created and is displayed on the Auto Actions Management page.

Note: You must change the name of the cloned auto action before a new one can be created.

Disabling an Auto Action

In the Auto Actions Management page, click the action's vertical ellipsis, and select Disable. In the confirmation message, click OK to confirm. The action is no longer active and the Disabled state is displayed in the action's Status column on the Auto Actions Management page.

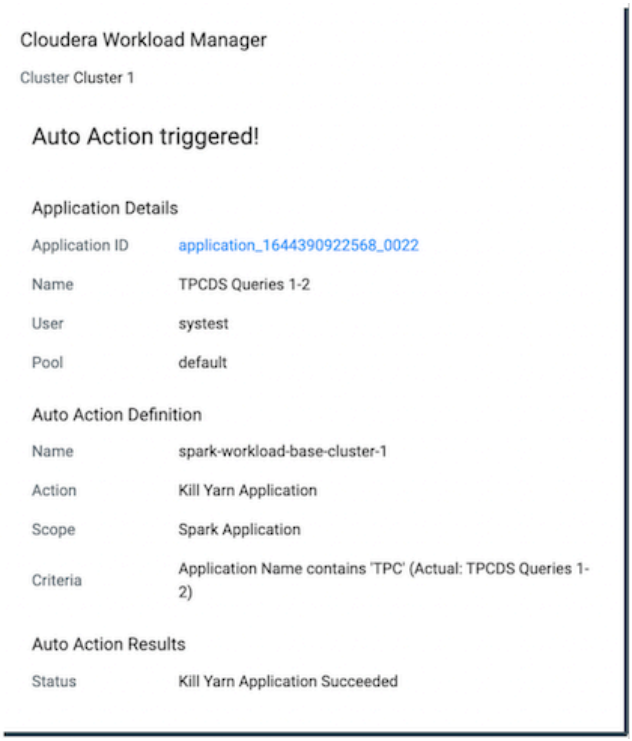
Auto Action Email Notification Examples

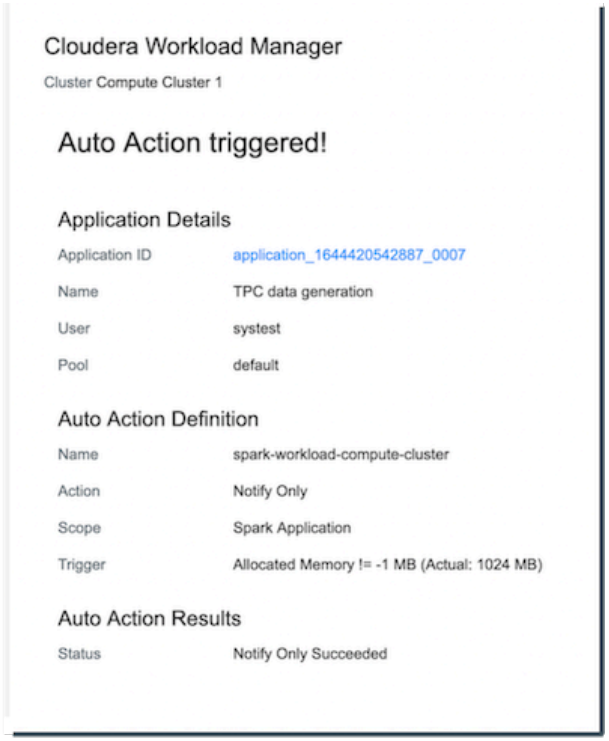
Examples of a Workload XM Auto Actions alert notification email.

The following email notification examples were sent when the listed application met the action's criteria and the trigger conditions, which are also included in the email notification.

Where, as in these examples:

- In the Application Details section, the Application ID contains a link to the workload application, job, or query.
- In the Auto Action Definition section, the Trigger and the Criteria definition display both the value and file size type that you defined and in brackets the Actual value, in megabytes, that was captured by the engine.
- In the Auto Action Results section, the results of the invoked auto action is displayed.





Classifying Workloads for Analysis with Workload Views

The Workload View feature enables you to analyze workloads with much finer granularity. For example, you can analyze how queries that access a particular database or that use a specific resource pool are performing against your SLAs. Or you can examine how all the queries are performing on your cluster that are sent by a specific user.

Working with Auto Generated Workload Views

Steps for using the Workload XM default workload views.

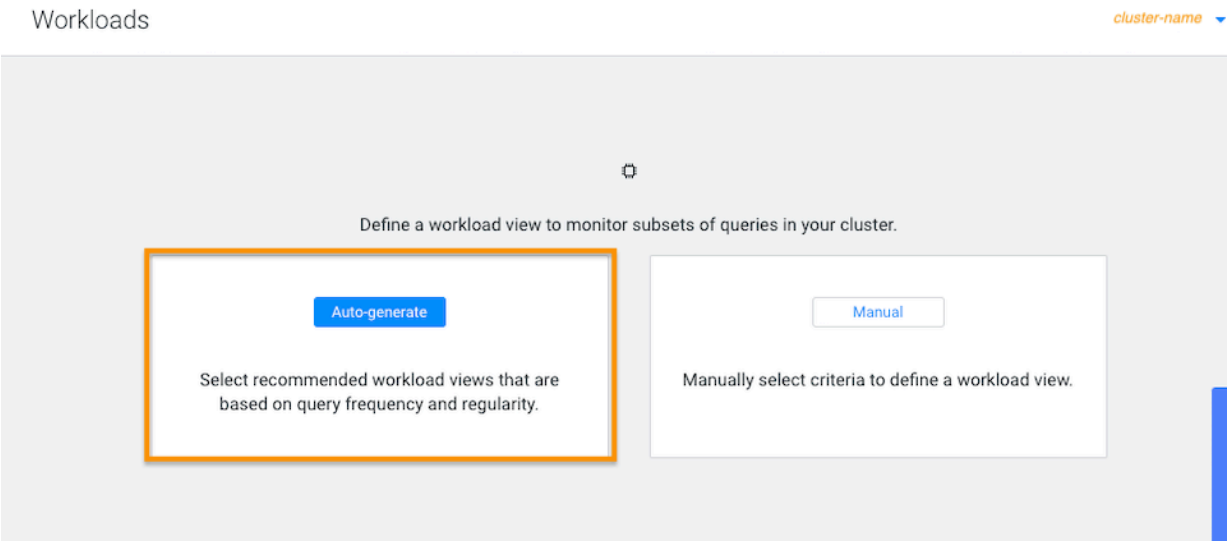
About this task

Describes how to use the workload views that Workload XM automatically generates.

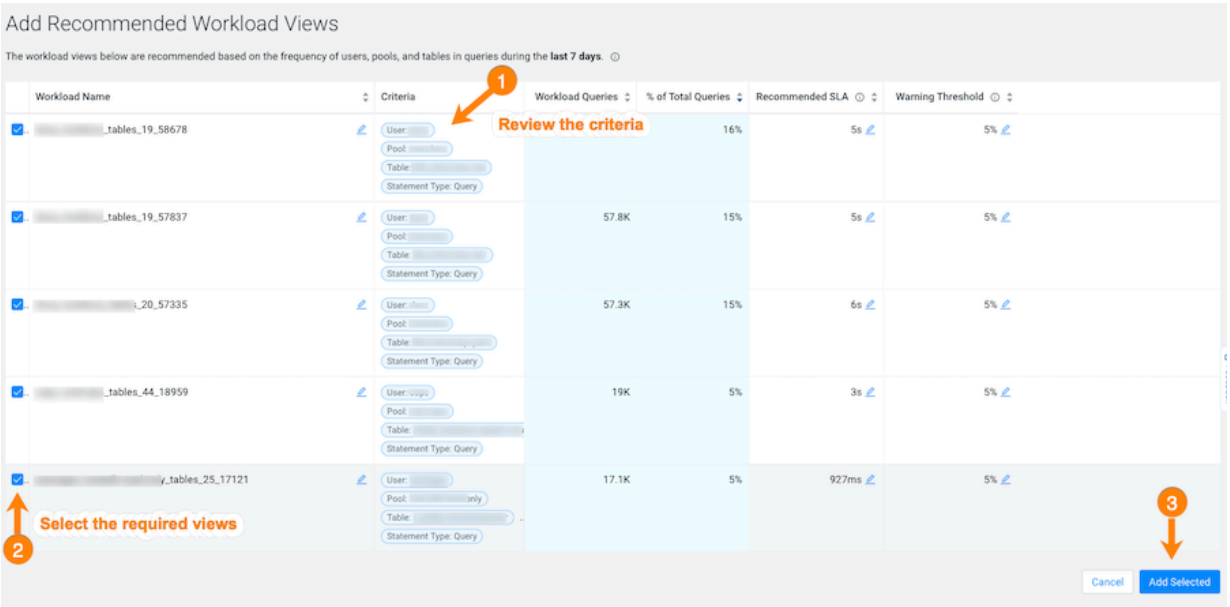
Procedure

1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. In the Clusters page do one of the following:
   - In the Search field, enter the name of the cluster whose workloads you want to analyze.
   - From the Cluster Name column, locate and click on the name of the cluster whose workloads you want to analyze.
3. From the time-range list in the Cluster Summary page, select a time period that meets your requirements.
4. From the navigation panel, select Workloads.

5. In the Workloads page, click Auto-generate:

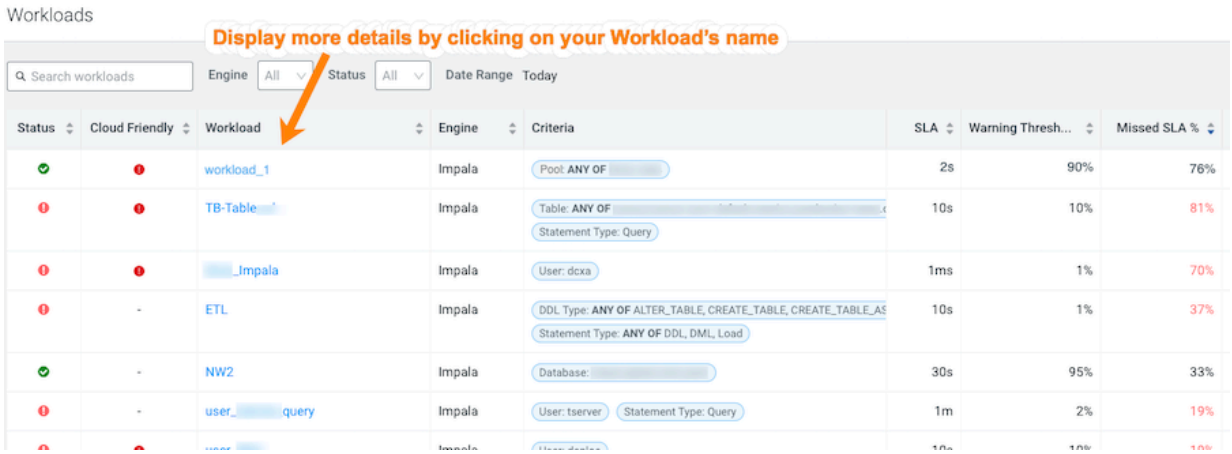


6. From the Criteria column, examine the criteria that is used for each workload view, select the required workload views, and then click Add Selected:



The workload views you selected are saved and displayed on the Workloads page.

7. To verify your workload views, from the navigation panel, select Workloads and then on the Workload page locate the workload view you added. When verified, click the workload to view its details:



Defining Workload Views Manually

Steps for manually defining your workload views.

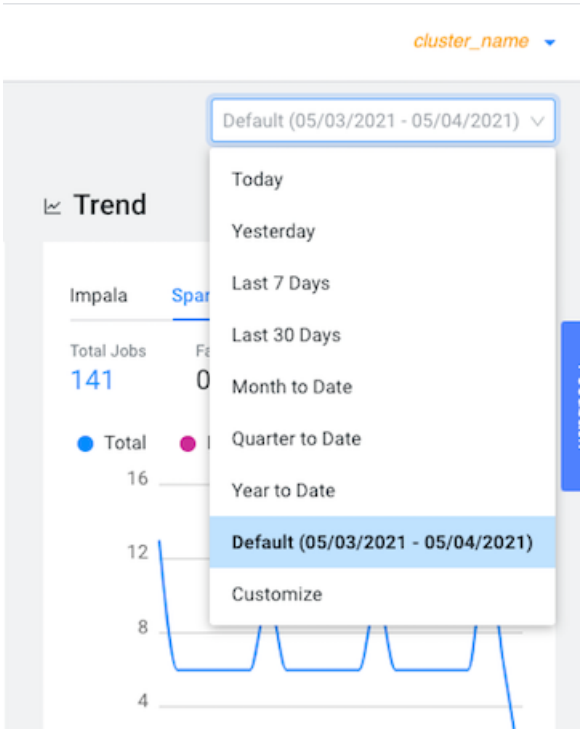
About this task

This task describes how to manually define your Workload Views.

Procedure

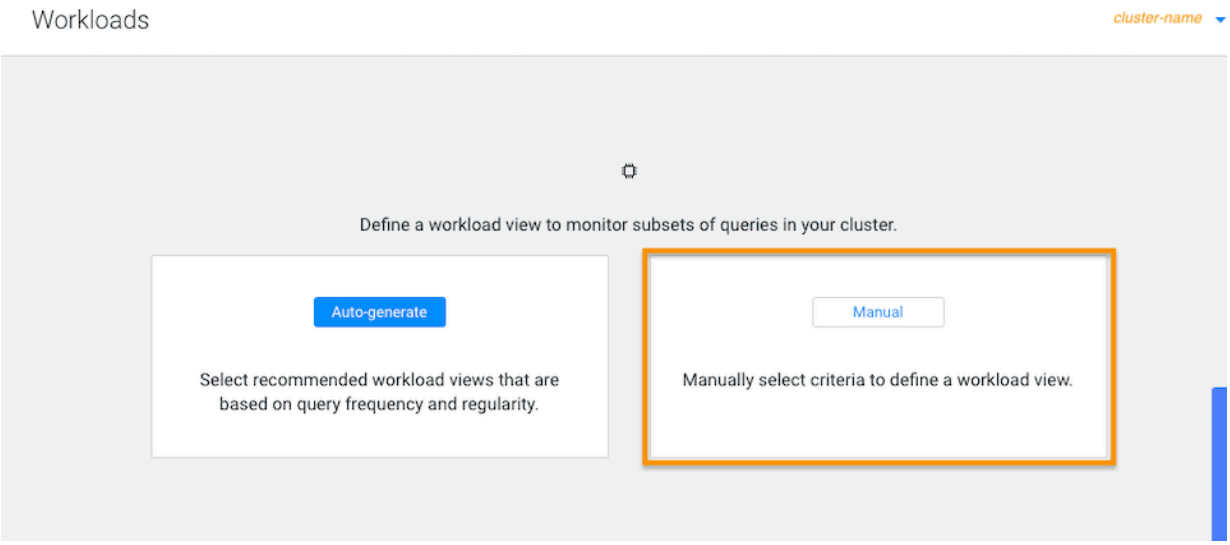
1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. In the Search field of the Clusters page, enter the name of the cluster whose workloads you want to analyze.

3. From the time-range list in the Cluster Summary page, select a time period that meets your requirements.



4. From the navigation panel, select Workloads.

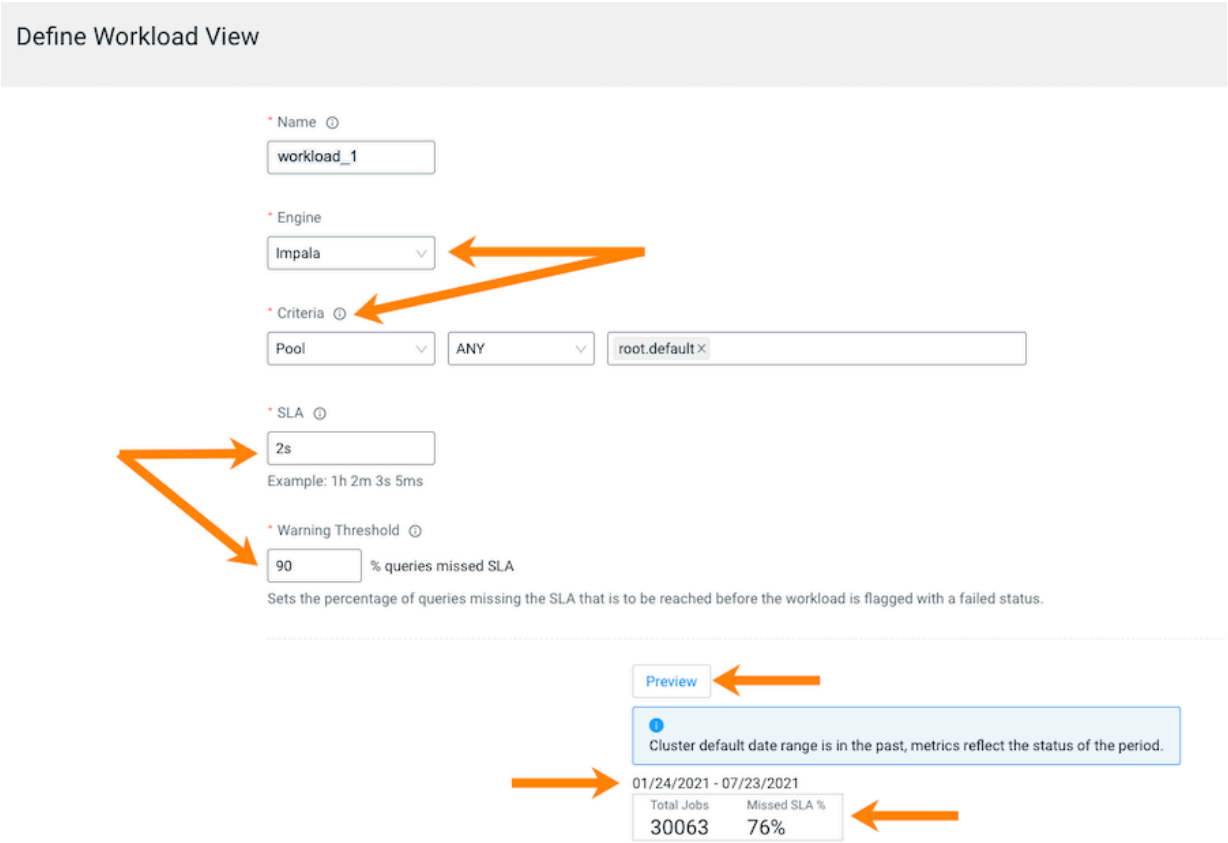
5. In the Workloads page, click Manual:



The Define Workload View widget opens, where you define a set of criteria that enables you to analyze a specific set of queries.

For example, as shown in the image below, you can list the total amount of failed queries, as a percentage, from a specific engine that are subject to a two second SLA.

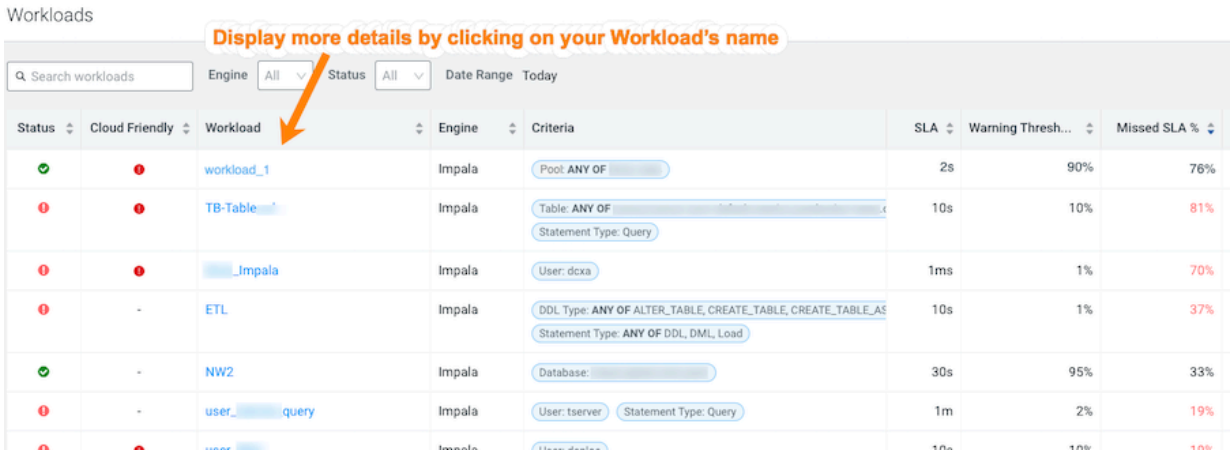
Where, as defined by the criteria condition, Workload XM will monitor all query jobs from the Impala engine. When the total query execution time exceeds 2 seconds, as defined by the SLA condition, for 90 percent of these queries, as defined by the Warning Threshold, the workload is flagged with a failed state:



6. (Optional) To display a summary of the queries matching your criteria, click Preview. Which displays the date range, the number of queries that match the criteria, and the number of queries that missed the SLA condition.

7. When you are satisfied with the results, click Save.

   The Workloads page opens and your workload view appears in the Workload column.

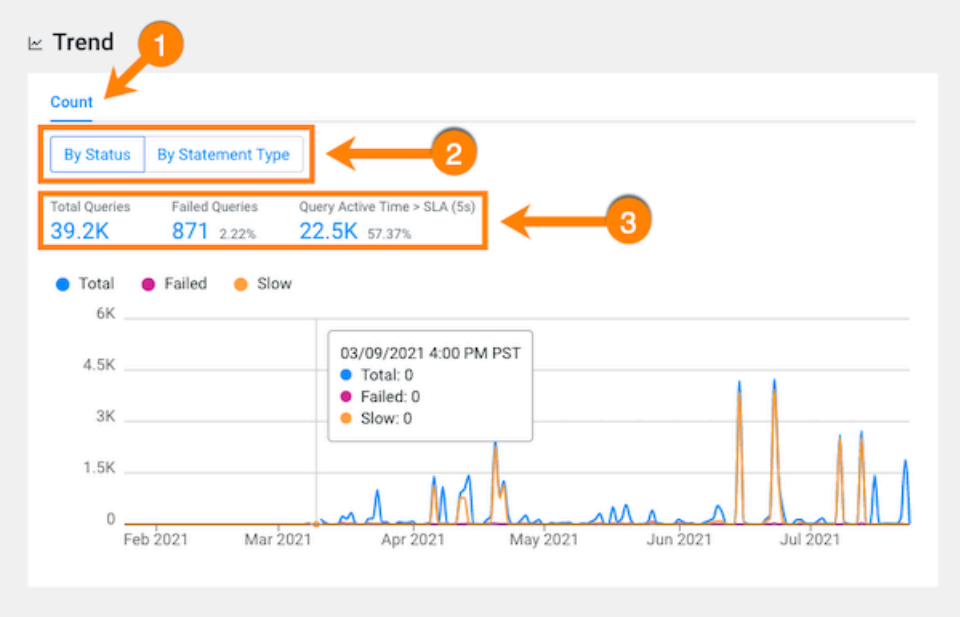


Tip: When you have a long list of Workload views, sorting the Workload column alphabetically in ascending or descending order by clicking the up or down arrows, helps locate the workload.

8. (Optional) To view more information about the workloads using the view's formula, open the Summary page by clicking the name of the workload view in the Workload column, which visually displays the view's details as chart widgets that you can use to further analyze the results.

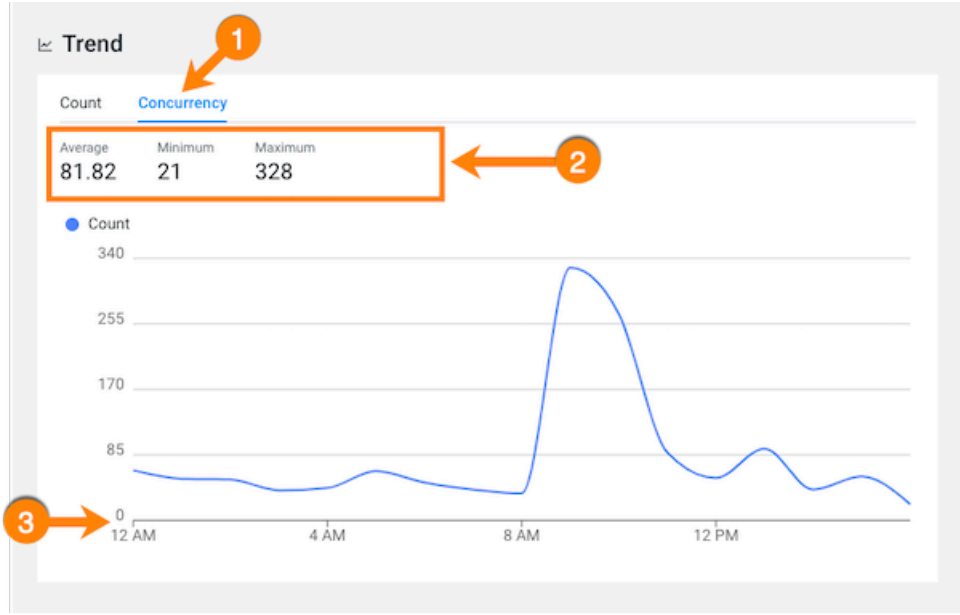
   The following examples, display how this group of queries are meeting the Workload view's SLA in the Trend chart, where:

   - The Count tab, displays the number of executing queries, either By Status or By Statement Type. To view further details, click the Total Queries, the Failed Queries, or the Query Active Time value.



   - The Concurrency tab (which is not available for CDP), displays the number of queries executing concurrently.

     In the following example, the maximum concurrency for this view is 328. This indicates that for the queries monitored by this view, 328 queries accessed the same data at the same time during the specified time period. The graph view displays how the concurrency fluctuates over the date range specified for the workload view.



Assigning Access Roles in Workload XM

Workload XM supports cluster privilege role types that define who is entitled to access jobs and queries that are created by the user, who is entitled to create and administer cost centers and view cluster costs, and who is entitled to access and administer jobs and queries within either a specific cluster or across all clusters within the Workload XM environment.

Limiting the trust boundary for jobs, queries, cluster costs, and administrative management at the cluster level, enables more control over the security and access management of your Workload XM environment.

Understanding the Workload XM Access Roles

Describes the Workload XM access roles.

Important: Customers are responsible for managing and reviewing access credentials for their Workload XM accounts and activities. All user privileges and access rights should periodically be reviewed and monitored, including who should access Workload XM, its services, and components. For example, access rights should be reviewed when a user moves to another business unit.

Workload XM supports cluster privilege roles that define Workload XM users as a:

- System Admin
- Cluster Admin
- Cluster User

The following tables describe these cluster privilege roles, also known as access roles:

System Admin Access Role

An authentic Workload XM user who is assigned the System Admin access role has full access rights and system administrator privileges across all clusters within the Workload XM environment. Where they can view, edit, and create cost centers, view, edit, and create auto actions, and view all the jobs and queries in all the Workload clusters. These users have the least restrictive access permissions.

Table 1: System Admin

Resource	Actions
Access Management page	View and manage all the Workload XM cluster policies and user access from the Access Management page
Cluster	• View all the workload clusters on the Clusters page • Rename a workload cluster • Delete a workload cluster
Workloads	• Create workloads • View all the workloads in a cluster • Update all the workloads in a cluster • Delete all the workloads in a cluster
Queries	View all the queries in all the clusters of the Workload XM environment
Jobs	View all the jobs in all the clusters of the Workload XM environment

Resource	Actions
Chargeback	• Create cost centers • Update cost centers • List cost centers • Delete cost centers • View all the Chargeback related dashboards
Auto Actions	• Create auto actions • View auto actions • Update auto actions • Disable auto actions • Delete auto actions • Enable an auto action email

Cluster Admin Access Role

An authentic Workload XM user who is assigned the Cluster Admin access role has full access rights and cluster administrator privileges across an assigned cluster within the Workload XM environment. Where they can view all the jobs and queries in the assigned Workload cluster.

Table 2: Cluster Admin

Resource	Actions
Cluster	• View the assigned Workload cluster on the Clusters page • Rename the Workload cluster • Delete the Workload cluster
Workloads	• Create workloads • View all workloads in the assigned cluster • Update all workloads in the assigned cluster • Delete all workloads in the assigned cluster
Queries	View all the queries in the assigned cluster
Jobs	View all the jobs in the assigned cluster

Cluster User Access Role

An authentic Workload XM user who is assigned the Cluster User access role has limited access rights across an assigned cluster within the Workload XM environment. Where they can view only those jobs and queries they created and executed in the assigned Workload cluster.

Table 3: Cluster User

Resource	Actions
Cluster	View their assigned cluster on the Clusters page.
Workloads	View their assigned workloads on the Workloads page
Queries	View their queries in the assigned cluster
Jobs	View their jobs in the assigned cluster

The Cluster User access role type has the most restricted access permissions, where the user may only view their own jobs and queries.

This access role further restricts the Cluster User to one cluster per policy. For users who are responsible for jobs and queries in more than one cluster they must also be assigned access rights to those clusters. You can either add them to the Cluster Policy for that cluster or include the pool that contains those workloads in the Cluster Policy in which they are assigned.

Also, for users who require access to jobs and queries executed by other users, you can create a Custom Policy as part of the Cluster Policy. This policy includes the user names of the users who execute those jobs and queries and/or the pool names in which they are executed.

For example, though user A and user B have been granted the same Cluster User role type their access to jobs and queries is different. This is due to the conditions of the Cluster Policy in which they are assigned. Where:

- The cluster policy that defines user A's Cluster User role type does not permit the user to view workloads within a pool or view other user workloads. In this case, user A is restricted to only view their own jobs and queries within their policy's assigned cluster.
- The cluster policy that defines user B's Cluster User role type contains a Custom Policy that permits the user to view workloads within a pool and view other user workloads. In this case, user B can view the jobs and queries executed by other users and the jobs and queries executed in the pool.

Understanding a Workload XM Cluster Policy

Describes the Workload XM Cluster Policy criteria that is used to assign Workload XM access roles to your users.

Access to your Workload jobs and queries is determined by a Workload XM Cluster Policy, which comprises two or more of the following conditions:

- One or more LDAP Group identifier account names.
- One or more user names. By default, Workload XM authenticates user access by checking that the user is a member of an LDAP group.
- A Workload XM access role type. The access role is assigned to the users that you provide in the Users field and/or the users who are part of the groups you provide in the Groups field and is defined by the conditions in the Cluster Policy.
- (Cluster User and Cluster Admin only) The cluster associated with the access role.
- (Cluster User only) A custom policy whose criteria is defined from the provided user names and/or the provided pools. A custom policy enables the user or users defined in the Cluster Policy to view the jobs and queries executed by other users and/or the jobs and queries executed in a pool.

Workload XM Cluster Policies are created, managed, and maintained from the Access Management page. Only users who have been granted the System Admin access role type can view and manage your Workload XM cluster policies.

Configuring a Default Systems Administrator for Workload XM

Pre-tasks that are required before you can start enabling role based access in Workload XM.

About this task

Describes how to enable role based access in Workload XM and configure a Workload XM default systems administrator.

Before you can assign access roles in Workload XM you must first enable role based access and configure a default systems administrator. Both tasks are completed in Cloudera Manager. Once configured, the default administrator (also known as a superuser) can log into the Workload XM UI and assign the System Admin access policy role to one or more users.

Procedure

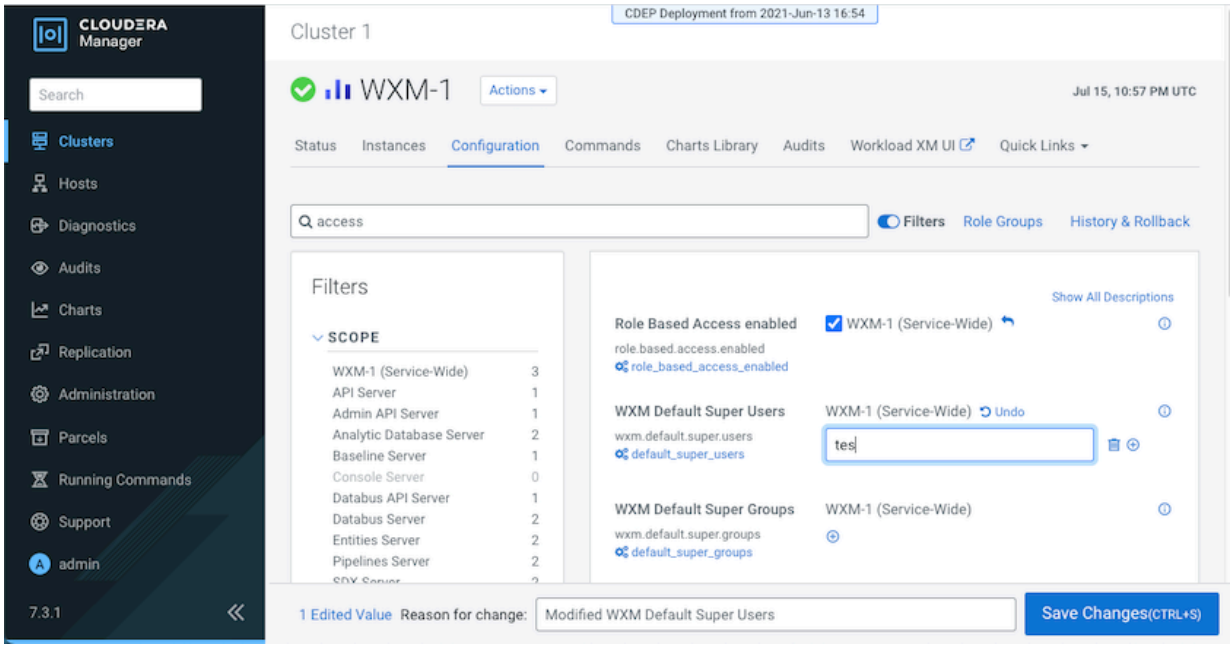
1. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select Clusters, WXM, and then click the Configuration tab.
3. In the Configuration page, search for the Role Based Access enabled property and then select its WXM (Service-Wide) check box.

4. According to your requirements, do one of the following:
   a. In the WXM (Service Wide) field of the WXM Default Super Users property, enter either the user name or the account name of a system administrator who is to be granted access to perform administration tasks in Workload XM. By default, admin.

      Tip: If the WXM (Service Wide) field is not displayed, click the plus sign circle icon.

   b. In the WXM (Service Wide) field of the WXM Default Super Groups property, enter the group account name of your LDAP admin group. For example, admin_grp.

   The following image shows the configuration properties:



5. Click Save Changes.
6. Navigate to the top of the Workload XM service page and from the Actions menu, restart the Workload XM service, by selecting Restart.

Assigning Workload XM Access Roles

Role based access to your Workload jobs and queries requires a Workload XM Cluster Policy that defines the conditions for the role based access type and assigns it to your users. You can have multiple Cluster Policies that define the access criteria for all of your workloads.

Assigning a Workload XM System Admin Access Role

Steps for assigning a System Admin access role to your Workload XM users.

About this task

Describes how to assign a Workload XM Role Based Access (RBAC) role for a system administrator. This access role has full access rights and system administrator privileges across all clusters within the Workload XM environment and can create your Workload XM Cluster Policies that define your access roles.

Note: Generally, only a user assigned the System Admin access role can create a Workload XM Cluster Policy. But until the first System Admin access role is assigned, a Cluster Policy can only be created by a default systems administrator, also known as a default super user.

Before you begin

This task assumes that you have:

- Enabled role based access in Cloudera Manager.
- Created a default systems administrator, also known as a default super user, in Cloudera Manager.

Procedure

1. In a supported web browser log in to Workload XM as the user with default systems administrator privileges.
2. From the Workload XM Navigation side-bar, select Access Management.
3. In the Access Management page, click New Cluster Policy.

   The Create Cluster Policy page opens.
4. Do one or more of the following:

   a. In the Groups field, enter the name of the LDAP administration group account whose users will be assigned this cluster policy's access role.
   b. In the Users field, enter the user name or user names who will be assigned this cluster policy's access role.
5. From the Assign Roles list, select System Admin.
6. Click Create.

Note: Workload XM will take at least 60 minutes to assign the access role to the user, users, and/or groups provided in the Cluster Policy.

Results

The Successfully created access policy message appears when the Cluster Policy is created and the policy is displayed in the Access Management's home page.

Assigning a Workload XM Cluster Admin Access Role

Steps for assigning a Cluster Admin access role to your Workload XM users.

About this task

Describes how to assign a Workload XM Role Based Access (RBAC) role for a cluster administrator.

Note: Only a user assigned the System Admin access role can create a Workload XM Cluster Policy.

Procedure

1. In a supported web browser log in to Workload XM as a user that has been granted the System Admin access role.
2. From the Workload XM Navigation side-bar, select Access Management.
3. In the Access Management page, click New Cluster Policy.

   The Create Cluster Policy page opens.
4. Do one or more of the following:

   a. In the Groups field, enter the name of the LDAP group account whose users will be assigned this cluster policy's access role.
   b. In the Users field, enter the user name or user names who will be assigned this cluster policy's access role.
5. From the Assign Roles list, select Cluster Admin.
6. From the Cluster list, select the name of the cluster that will be assigned to this policy's access role.
7. Click Create.

Note: Workload XM will take at least 60 minutes to assign the access role to the user, users, and/or groups provided in the Cluster Policy.

Results

The Successfully created access policy message appears when the Cluster Policy is created and the policy is displayed in the Access Management's home page.

Assigning a Workload XM Cluster User Access Role

Steps for assigning a Cluster User access role to your Workload XM users.

About this task

Describes how to assign a Workload XM Role Based Access (RBAC) role for a cluster user.

Note: Only a user assigned the System Admin access role can create a Workload XM Cluster Policy.

Procedure

1. In a supported web browser log in to Workload XM as a user that has been granted the System Admin access role.
2. From the Workload XM Navigation side-bar, select Access Management.
3. In the Access Management page, click New Cluster Policy.

   The Create Cluster Policy page opens.
4. Do one or more of the following:
   a. In the Groups field, enter the name of the LDAP group account whose users will be assigned this cluster policy's access role.
   b. In the Users field, enter the user name or user names who will be assigned this cluster policy's access role.
5. From the Assign Roles list, select Cluster User.
6. From the Cluster list, select the name of the cluster that will be assigned to this policy's access role.
7. (Optional) Enable the user or users defined in this cluster policy to view executed workloads from other users or executed workloads from a pool by doing the following:
   a. In the Users field, enter the user name or user names whose jobs and queries can be viewed by the user or users defined in this cluster policy.
   b. In the Pools field, enter the pool name or pool names whose jobs and queries can be viewed by the user or users defined in this cluster policy.
8. Click Create.

   Note: Workload XM will take at least 60 minutes to assign the access role to the user, users, and/or groups provided in the Cluster Policy.

Results

The Successfully created access policy message appears when the Cluster Policy is created and the policy is displayed in the Access Management's home page.

Managing Your Workload XM Access Roles

Describes how to manage your Workload XM cluster policies and access roles.

Information about your Workload XM Cluster Policies are displayed on the Access Management page, which are viewed and managed by the user with the System Admin access role.

Each row displays a Cluster Policy and its conditions, where:

- The Status column displays the state of the policy, as either Enabled or Disabled.
- The Clusters column displays the name of the cluster assigned to the Workload XM access role.
- The Role column displays the Workload XM access role type.

- The Groups column displays the LDAP group users who are assigned the Cluster Policy's access role.
- The Users column displays the user names who are assigned the Cluster Policy's access role.
- The Custom Policy column displays the user and pool filter conditions.
- The Last Updated column displays the date when the policy was last updated.
- The Actions column's vertical ellipses, when selected, lists the management tasks that can be performed.

The following management tasks are performed from the Access Management home page by a user with the System Admin access role, which is accessed by selecting Access Management from the Workload XM Navigation side-bar.

Updating a Cluster Policy

In the Access Management page, click the cluster policy’s vertical ellipsis in the Actions column, and select Edit. In the Cluster Policy, make your changes and then click Update.

Deleting a Cluster Policy

In the Access Management page, click the cluster policy's vertical ellipsis in the Actions column, and select Delete. In the confirmation message, click OK to confirm the action. The policy is permanently removed.

Tip: Cloudera recommends disabling rather than deleting a Cluster Policy.

Disabling a Cluster Policy

In the Access Management page, click the cluster policy's vertical ellipsis in the Actions column, and select Disable. In the confirmation message, click OK to confirm the action. The Status column displays the state of the policy as Disabled.

Troubleshooting an Abnormal Job Duration

Identify areas of risk from jobs running on your cluster that complete within an unusual time period.

About this task

Describes how to locate and troubleshoot an abnormal job duration.

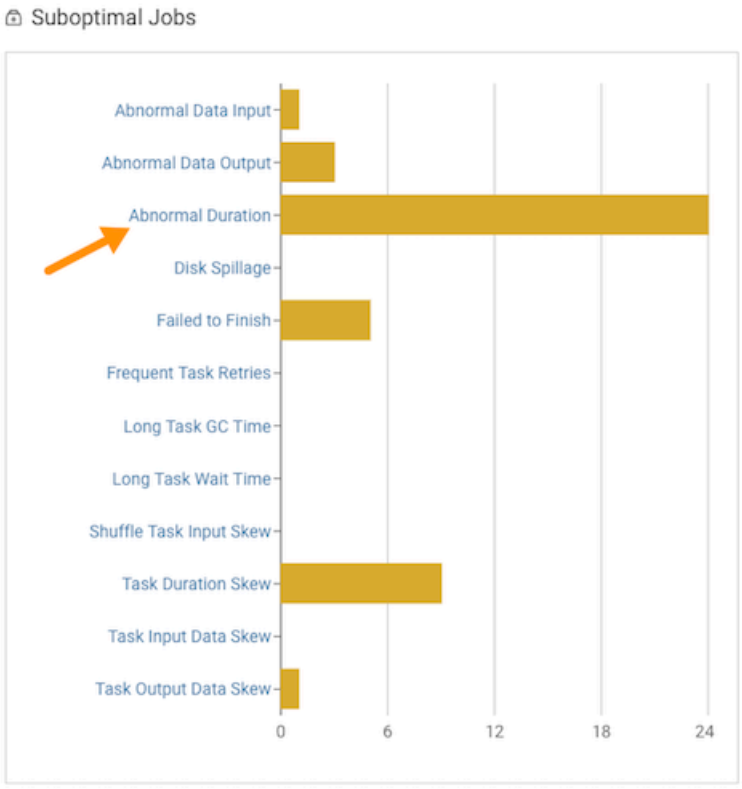
The following procedure uses examples from a Spark engine to explain how to further investigate and troubleshoot the cause of an abnormal job duration.

Procedure

1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. In the Clusters page do one of the following:
   - In the Search field, enter the name of the cluster whose workloads you want to analyze.
   - From the Cluster Name column, locate and click on the name of the cluster whose workloads you want to analyze.
3. From the time-range list in the Cluster Summary page, select a time period that meets your requirements.
4. In the Usage Analysis chart, click the engine whose Failed column displays the number of jobs that did not complete.

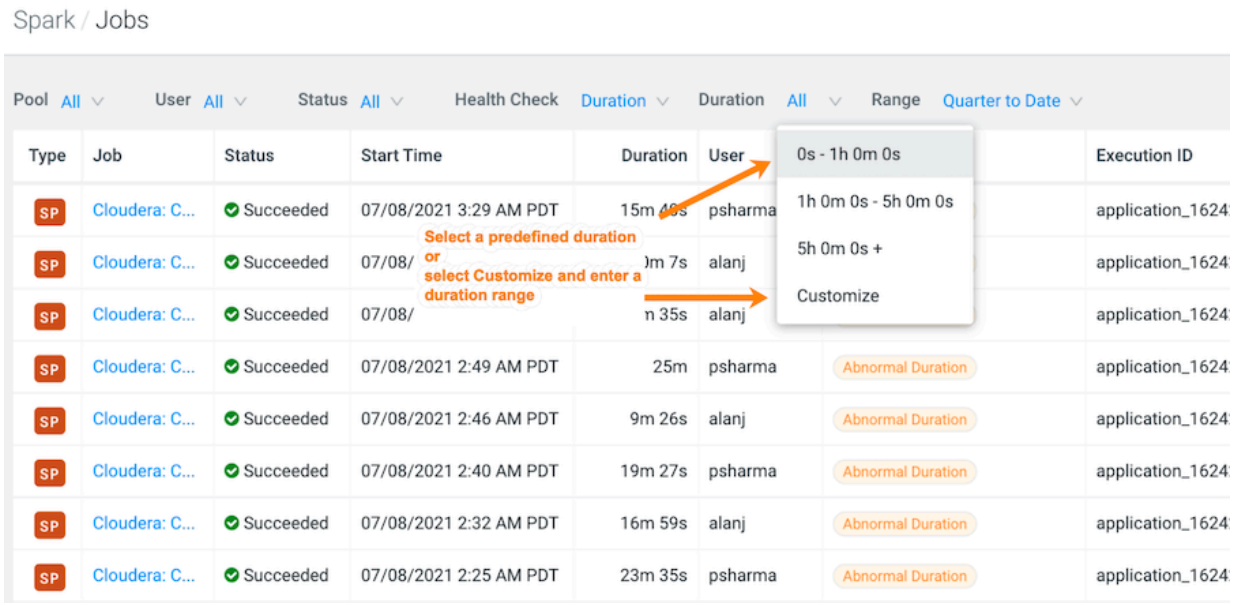
5. Depending on the engine you selected, in the engine's page that opens scroll down to either the Suboptimal Jobs or the Suboptimal Queries chart widget and click the Abnormal Duration health check bar.

   The Jobs or Queries page opens, listing all the jobs or queries that have triggered the Abnormal Duration Health check.



Tip: Any jobs or queries that fall outside of their baseline are counted. You can hover over each bar within the chart to view how many jobs or queries triggered each health check.

6. Specify a specific amount of time in which the job either ran less than or more than the Health check rule by either selecting a predefined time duration or selecting Customize and enter the minimum or maximum time period from the Duration list.

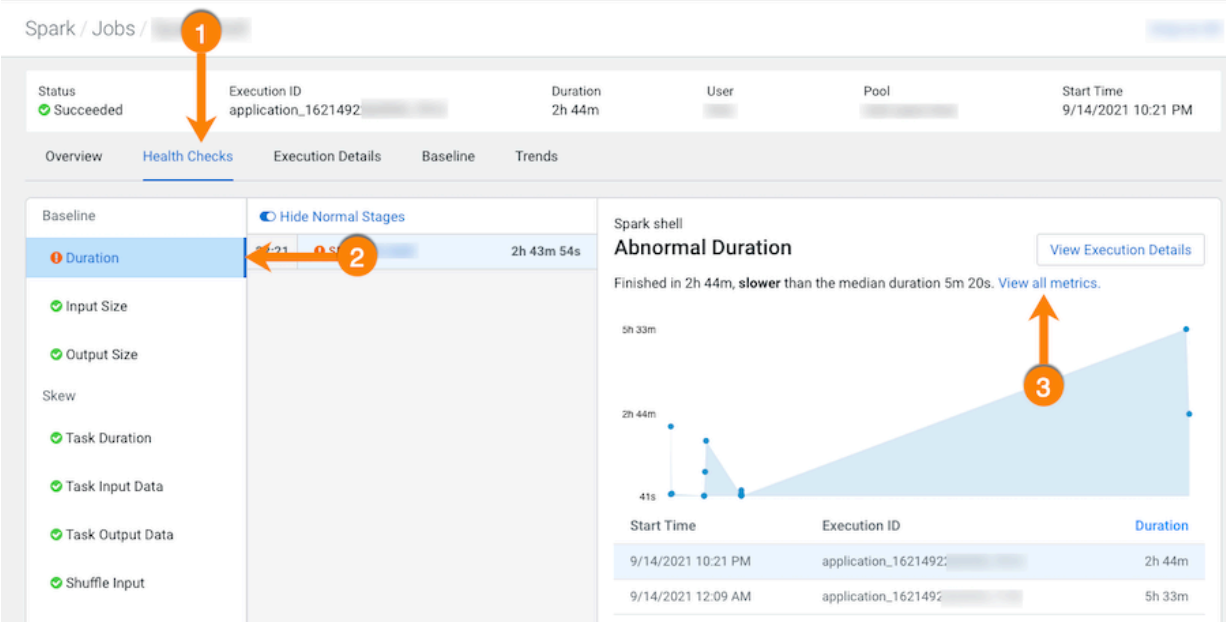


7. View more details about a job by selecting a job's name from the Job column and then clicking the Health Checks tab.

   The Baseline Health checks are displayed.

8. Display more information about the job's duration by selecting Duration from the Baseline section. As shown in the image below.

   In the following example, the job finished much slower than the baseline duration, which is the aggregate calculated over multiple jobs.

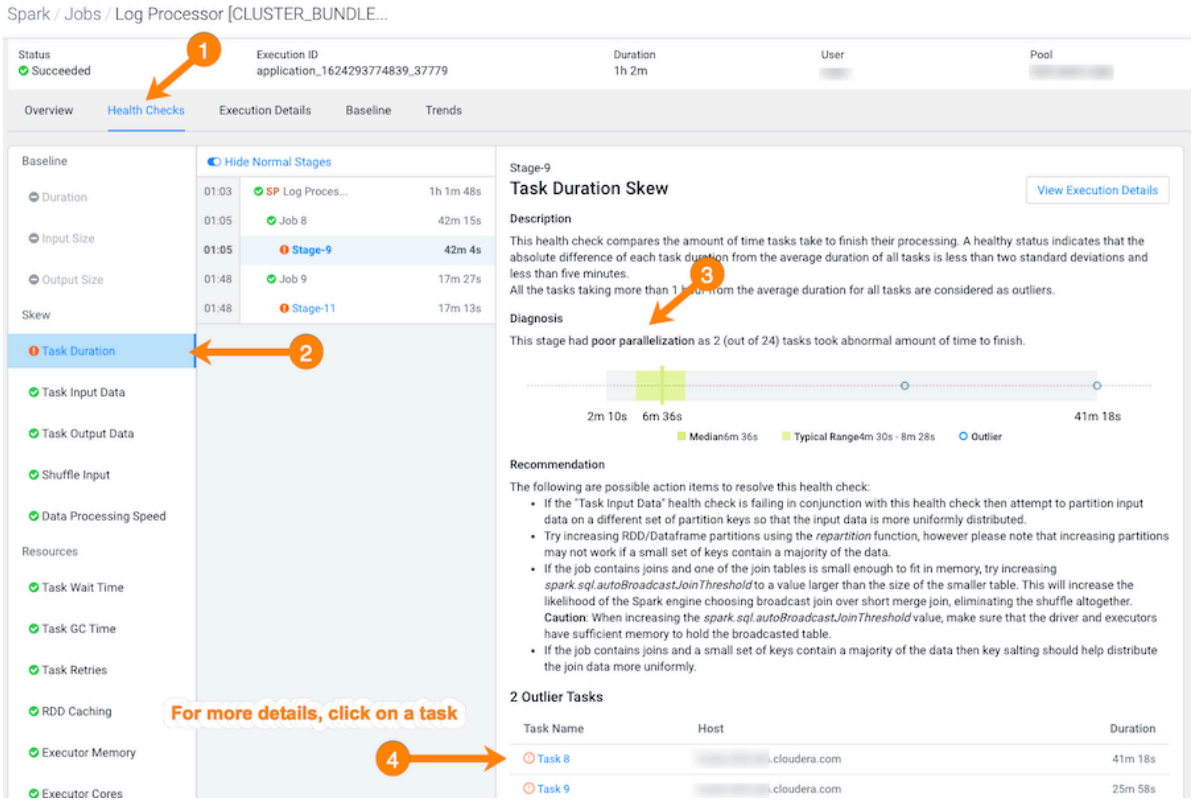


9. Compare and analyze this job against other baseline metrics by clicking View all metrics.

10. Continue to analyze and search for probable causes by doing one or more of the following:

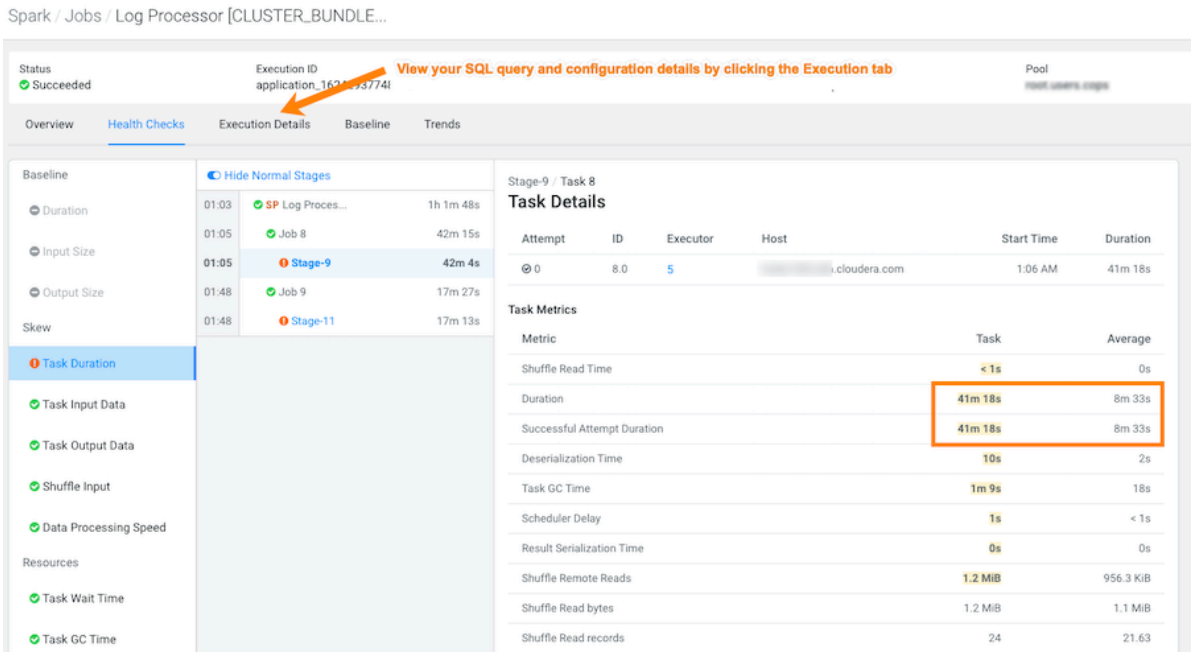
- To display more information about the length of time the processing tasks took within a job, select Task Duration, which opens a panel that describes the health check, displays information about the possible causes, and lists recommended solutions.

  In the following example, issues arose during Stage-9 of the job due to poor parallelization. The Recommendation section lists items for you to complete that may resolve the problem and the specific outlier tasks that produced the unusual results:

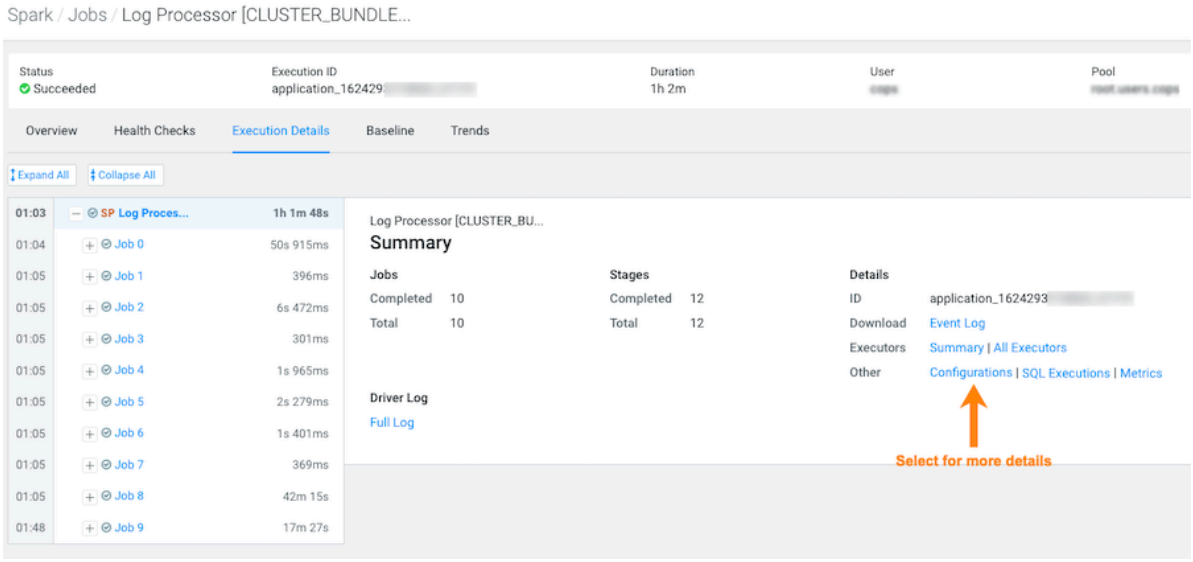


- To display more details about an outlier, click the outlier task, which opens the Task Details panel.

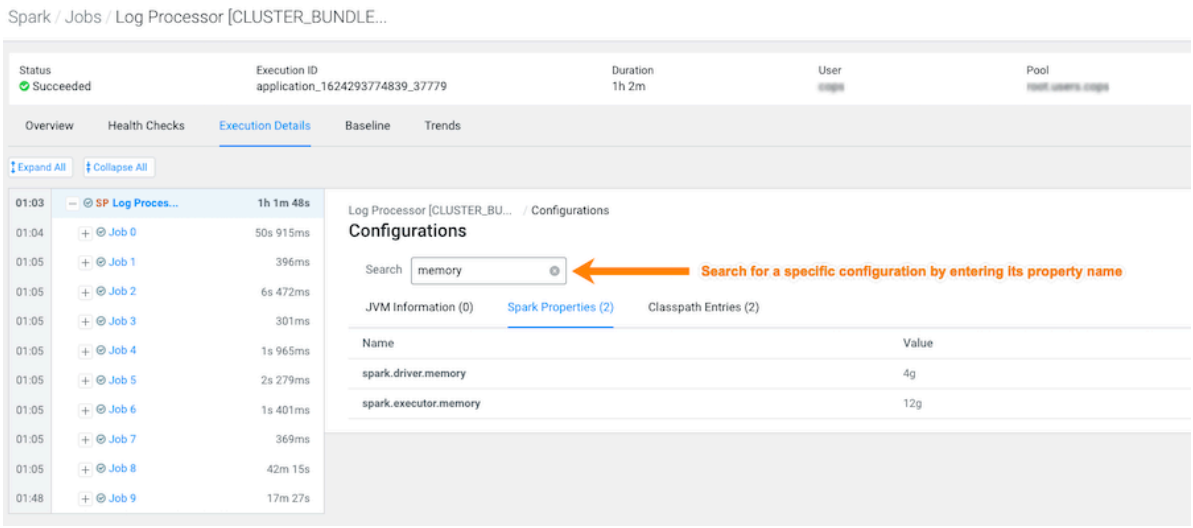
  In the following example, the Task Details show that the outlier task took significantly more time to complete compared to previous runs. In this case, 41 minutes as compared to 8 minutes:



- To gain more insights about the task's duration, such as checking memory allocation, click the Execution Details tab and then in the Summary panel, click Configurations:



- In the Configurations panel, click the Spark Properties tab and search for the memory configuration settings and their values. If memory is less than the recommended value, increasing its value will improve cluster performance:



Troubleshooting Failed Jobs

Steps for troubleshooting incomplete jobs running on your cluster.

About this task

Describes how to locate and troubleshoot jobs that have failed to complete.

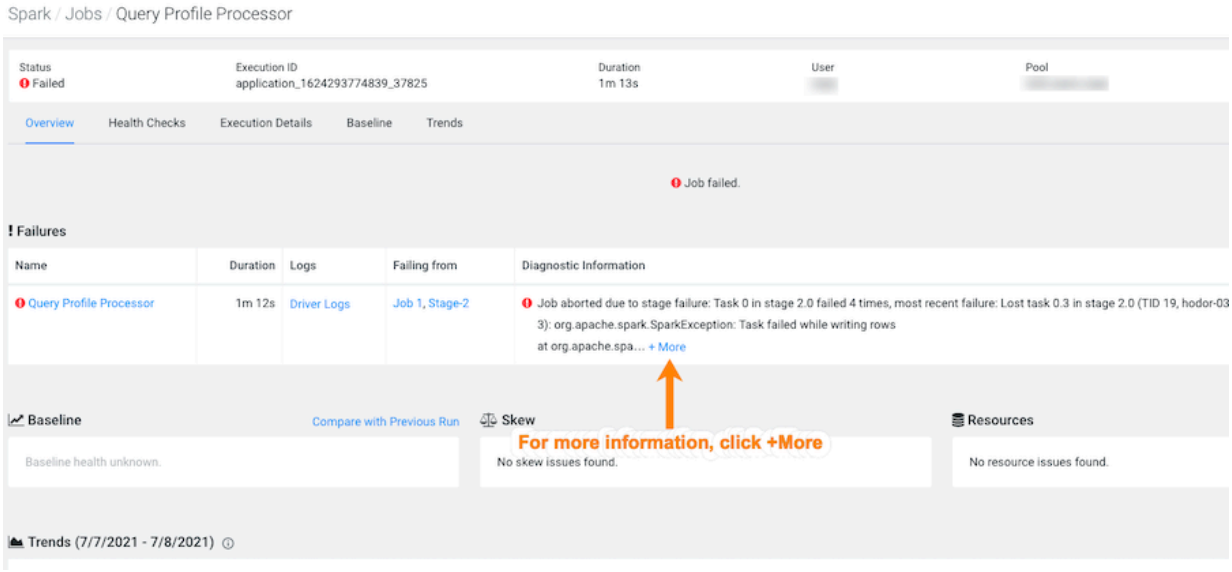
The following procedure uses examples from a Spark engine to explain how to further investigate and troubleshoot the root cause of an uncompleted job

Procedure

1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. In the Clusters page do one of the following:
   - In the Search field, enter the name of the cluster whose workloads you want to analyze.
   - From the Cluster Name column, locate and click on the name of the cluster whose workloads you want to analyze.
3. From the time-range list in the Cluster Summary page, select a time period that meets your requirements.
4. In the Usage Analysis chart widget, notice which engine's are displaying Failed jobs and then from the Trend widget, select the tab of an engine whose failed jobs you wish to analyze and then click its Total Jobs value.

   The engine's Jobs page opens.
5. From the Health Check list, select Failed to Finish, which filters the list to display a list of jobs that did not complete.

6. To view more details about why a job failed to complete, from the Job column select a job's name. The job's page opens displaying information about the job you selected and where the failure happened.



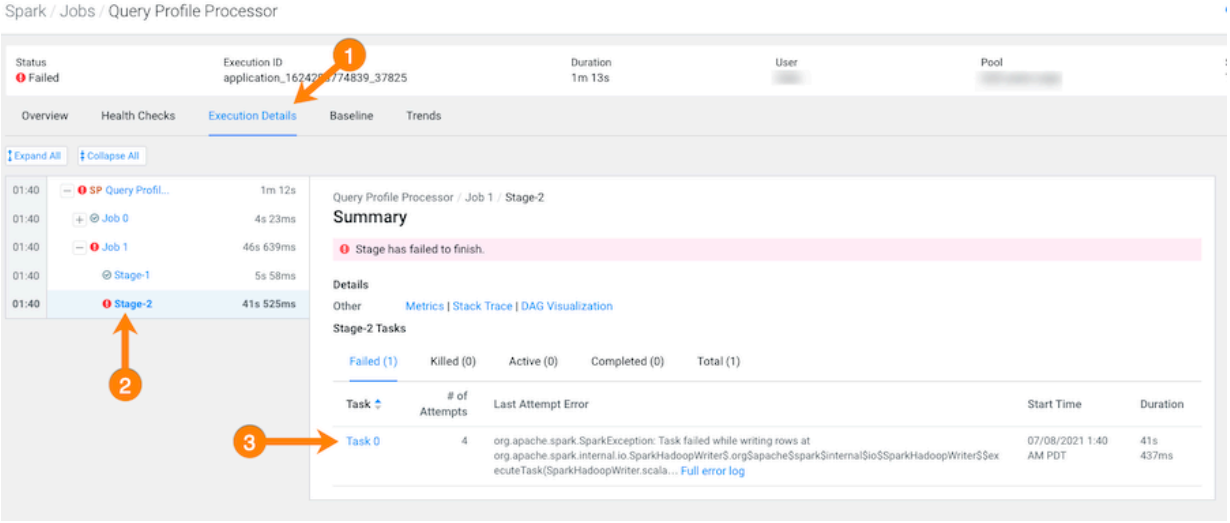
7. From the Failures section, in the Diagnostic Information column, click More.

   The Diagnostic Information dialog box opens, which describes more details about why the job aborted. In the following example's case, the job was aborted whilst writing rows due to an out of bounds java exception:

8. Click Close, to close the dialog box.

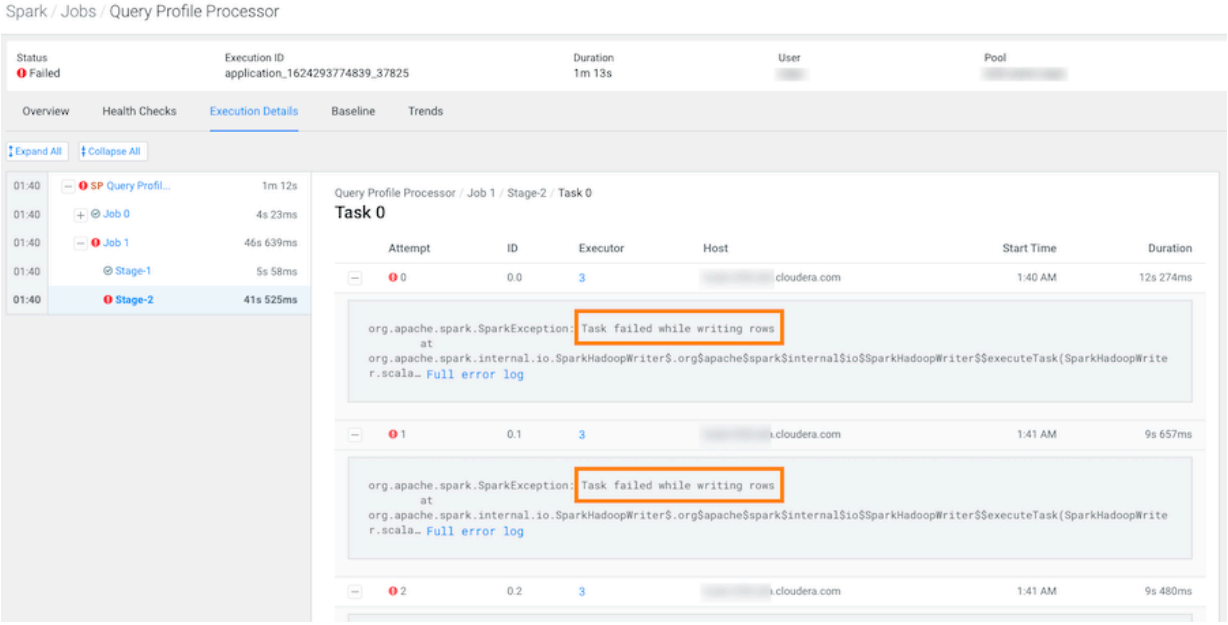
9. To display more information about the stage where the job failed, in this case the Stage-2 process, in the Failing from column, click the stage's link. Or select the Execution Details tab and then click the failed stage's link.

   In the following example's Summary panel, it shows that Task 0 was attempted 4 times:



10. To display more information about all the failed attempts, in the Summary panel, click the Failed task value.

    In the following example, the job aborted when Task 0 was writing rows. To understand more about what triggered the SparkException error message and to further troubleshoot the root cause, you can open the associated log file by clicking Full error log.



Determining the Cause of Slow and Failed Queries

Identifying the cause of slow query run times and queries that fail to complete.

About this task

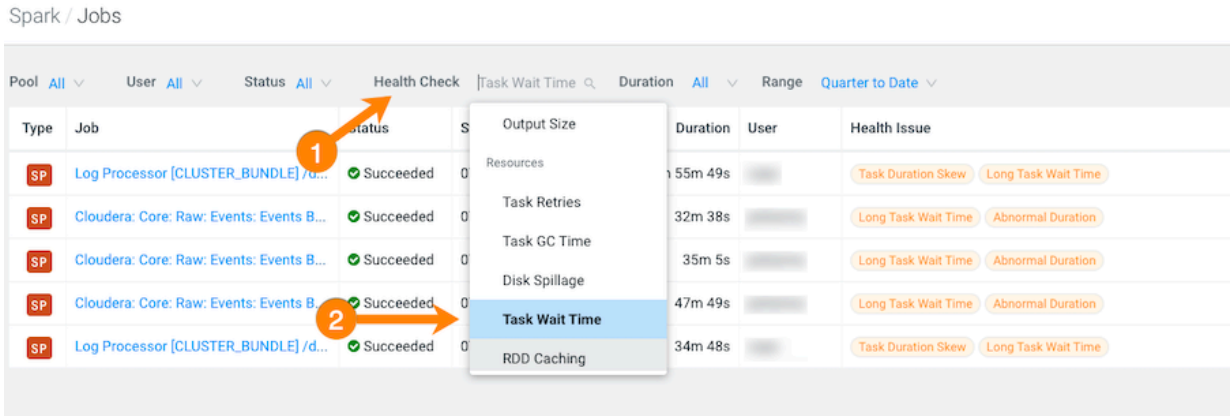
Describes how to determine the cause of slow and failed queries.

The following procedure uses examples from a Spark engine to explain how to further investigate and troubleshoot the cause of a slow and failed query.

Procedure

1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. In the Clusters page do one of the following:
   - In the Search field, enter the name of the cluster whose workloads you want to analyze.
   - From the Cluster Name column, locate and click on the name of the cluster whose workloads you want to analyze.
3. From the time-range list in the Cluster Summary page, select a time period that meets your requirements.
4. From the Trend widget, select the tab of an engine whose jobs you wish to analyze and then click its Total Jobs value.

   The engine's Jobs page opens.
5. From the Health Check list in the Jobs page, select Task Wait Time, which filters and displays a list of jobs with longer than average wait times before the process was executed.

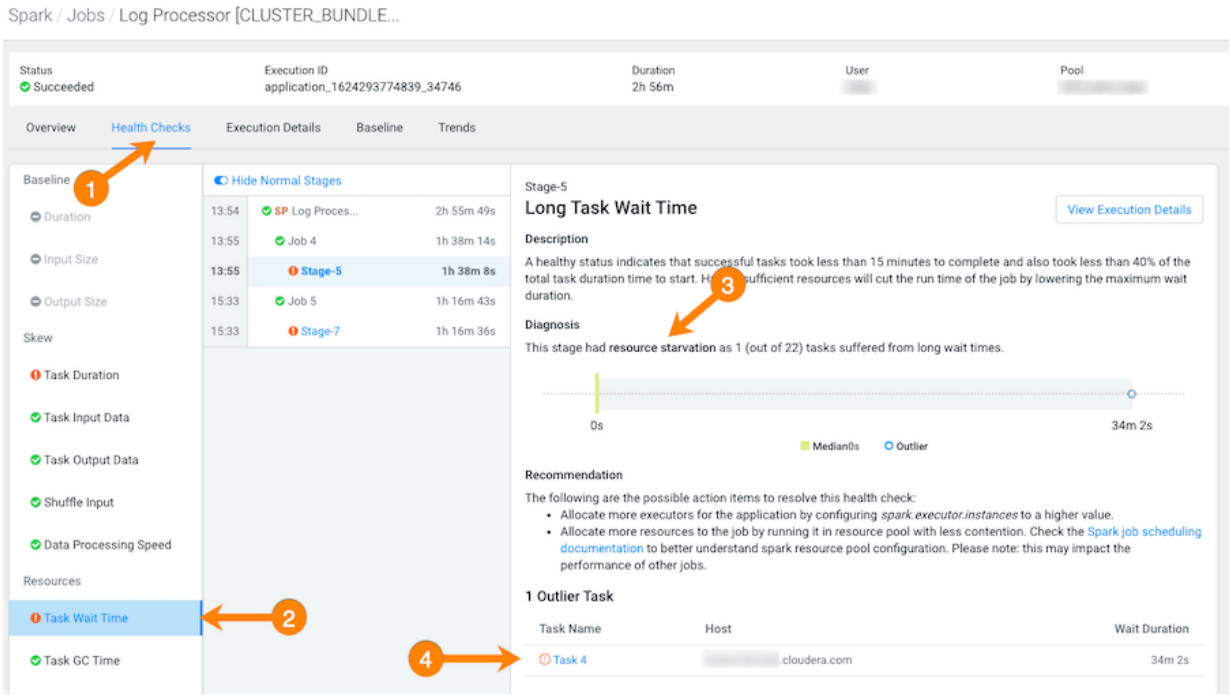


6. Display more details by selecting a job's name from the Job column and then clicking the Health Checks tab.

   The Baseline Health checks are displayed.

7. From the Health Checks panel on the left, click the Task Wait Time health check, which opens a panel that describes the health check, displays information about the possible causes, and lists recommended solutions.

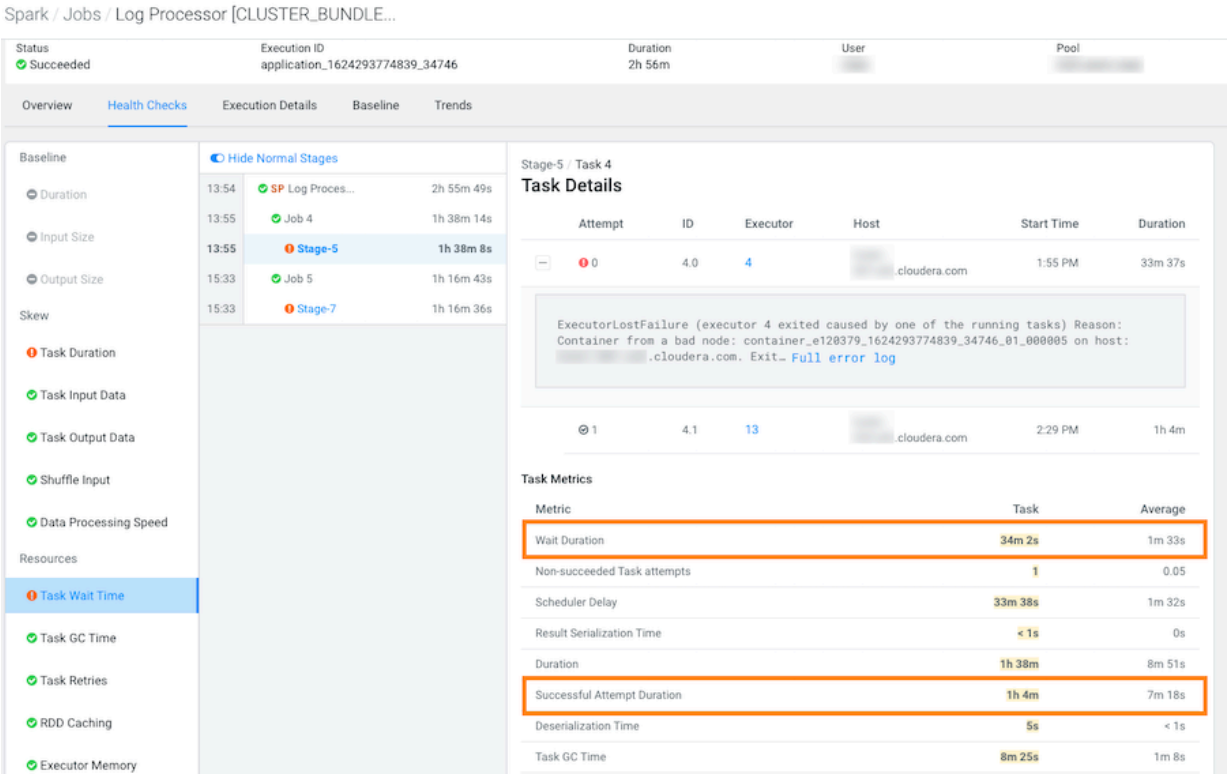
   In the following example, the long wait time occurred in Stage-5 of the job process due to insufficient resources. The Recommendation section lists items for you to complete that may resolve the problem and the specific outlier tasks that produced the unusual results:



8. To display more details about why this job is experiencing longer than average wait times, click one of the tasks listed under Outlier Tasks.

   In the following example, the Task Metrics section shows higher than average criteria measurement results and the Task Details reveal an ExecutorLostFailure error. This indicates a probable memory issue, where the container

is exceeding the memory limits. In this case, more details maybe found by clicking Full error log and reviewing the log:



Troubleshooting with the Job Comparison Feature

Steps for comparing two different runs of the same job, which is especially useful when you notice unexpected changes, such as when you have a job that consistently completes within a specific amount of time and then it starts taking longer. Comparing two runs of the same job enables you to analyze the performance and differences so that you can troubleshoot the cause.

About this task

Describes how to compare any two runs of a job using the Job Comparison tool.

The following procedure uses examples from a Spark engine to explain how to further investigate and troubleshoot.

Note: When a job is flagged as slow, you can select the slow job from the Slow Jobs widget in the job's engine page and then in the details page, click Compare with Previous Run. The job is compared with its last run and the results are displayed in the Job Comparison page for you to analyze.

Procedure

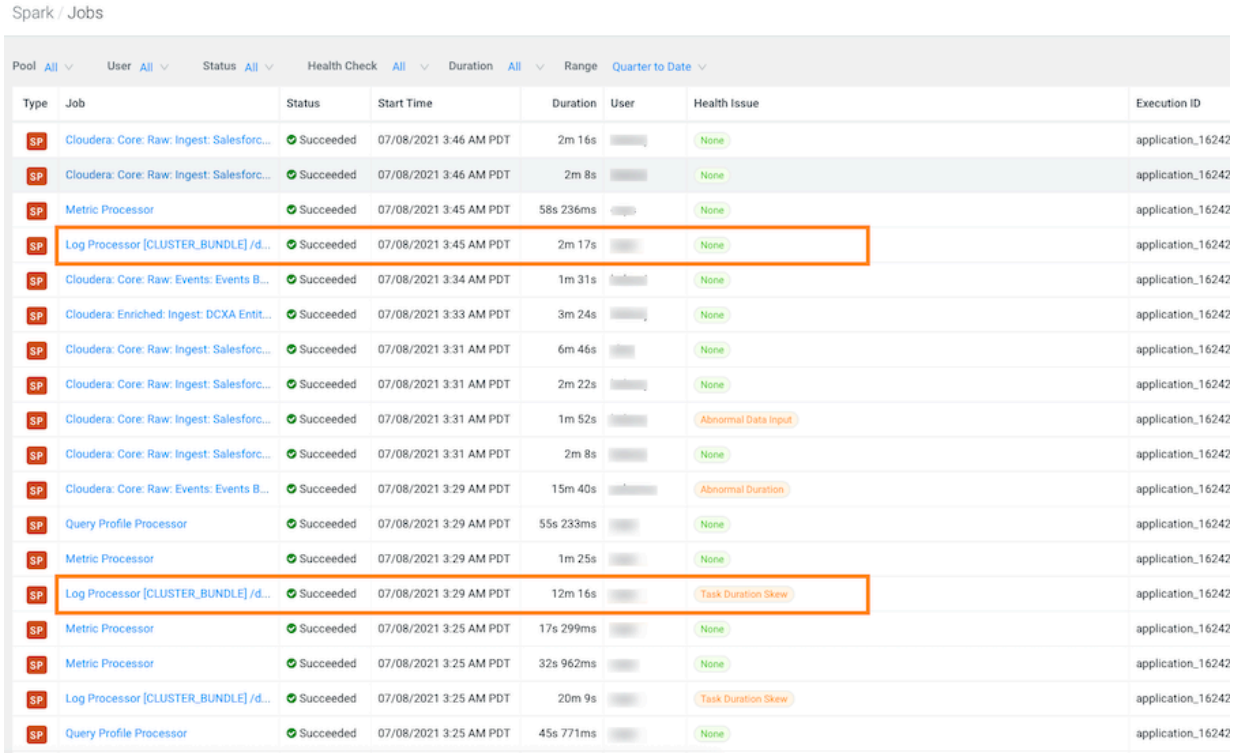
1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. In the Search field of the Clusters page, enter the name of the cluster whose workloads you want to analyze.
3. From the time-range list in the Cluster Summary page, select a time period that meets your requirements.

4. In the Trend widget, select the tab of an engine whose jobs you want to analyze and then click its Total Jobs value.

   The engine's Jobs page opens.

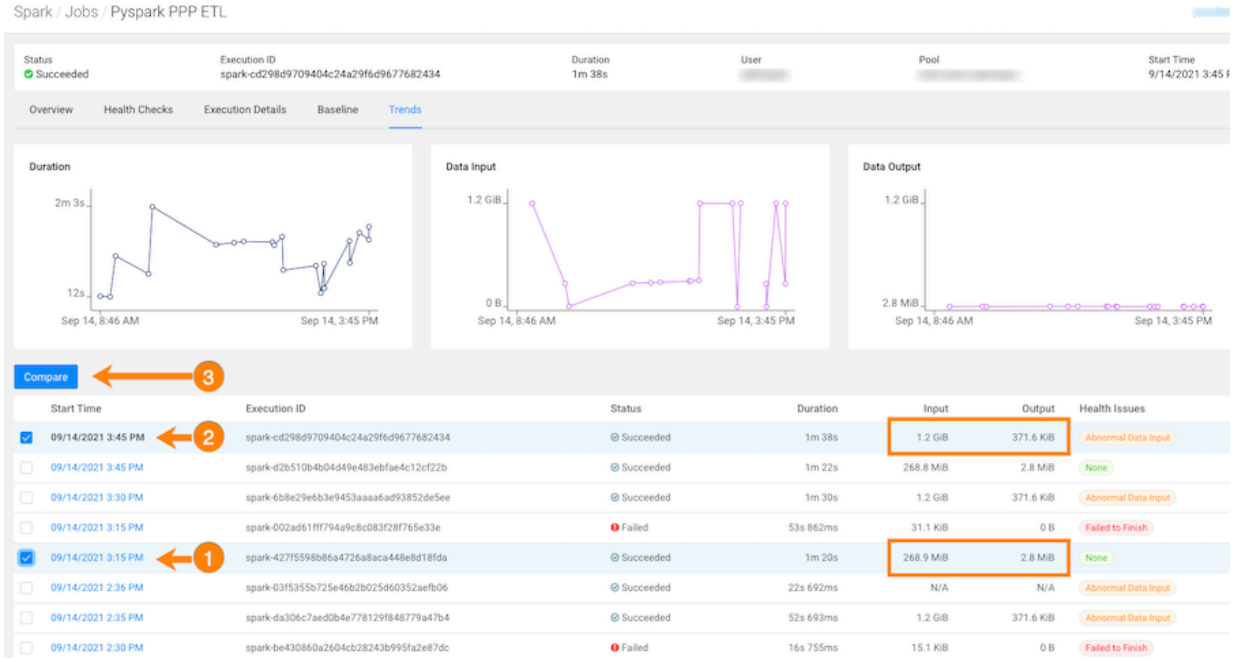
5. Examine the list of jobs that have executed during the selected time period and manually compare runs of the same job.

   For example, as shown in the following image, when manually comparing the last two runs of the Log Processor job we can see that there are duration differences. In this example, the older run had a Task duration skew health issue, which appears to be fixed:



6. List and display details of all the runs of a specific job of interest by selecting one of the job runs and then in its jobs details page, click the Trends tab.

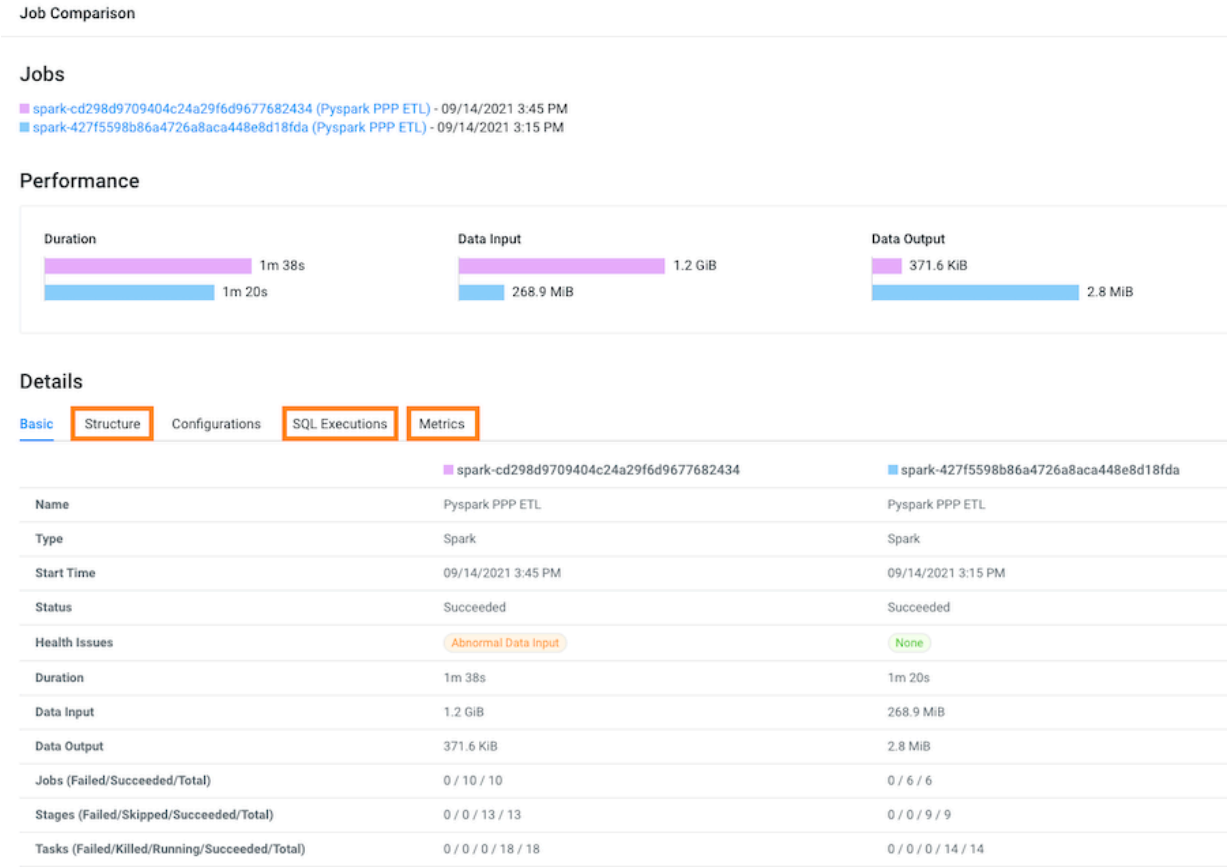
   In the following example, notice how the amount of Input and Output data changes between runs. The Job Comparison tool enables you to examine more details about two runs to determine why the amount of data changed. In this case we will compare a run with no health issues with the last run of the job:



7. To compare two job runs, select the check boxes adjacent to the job runs you require and then click Compare.

   The Job Comparison page opens displaying more details about each job.

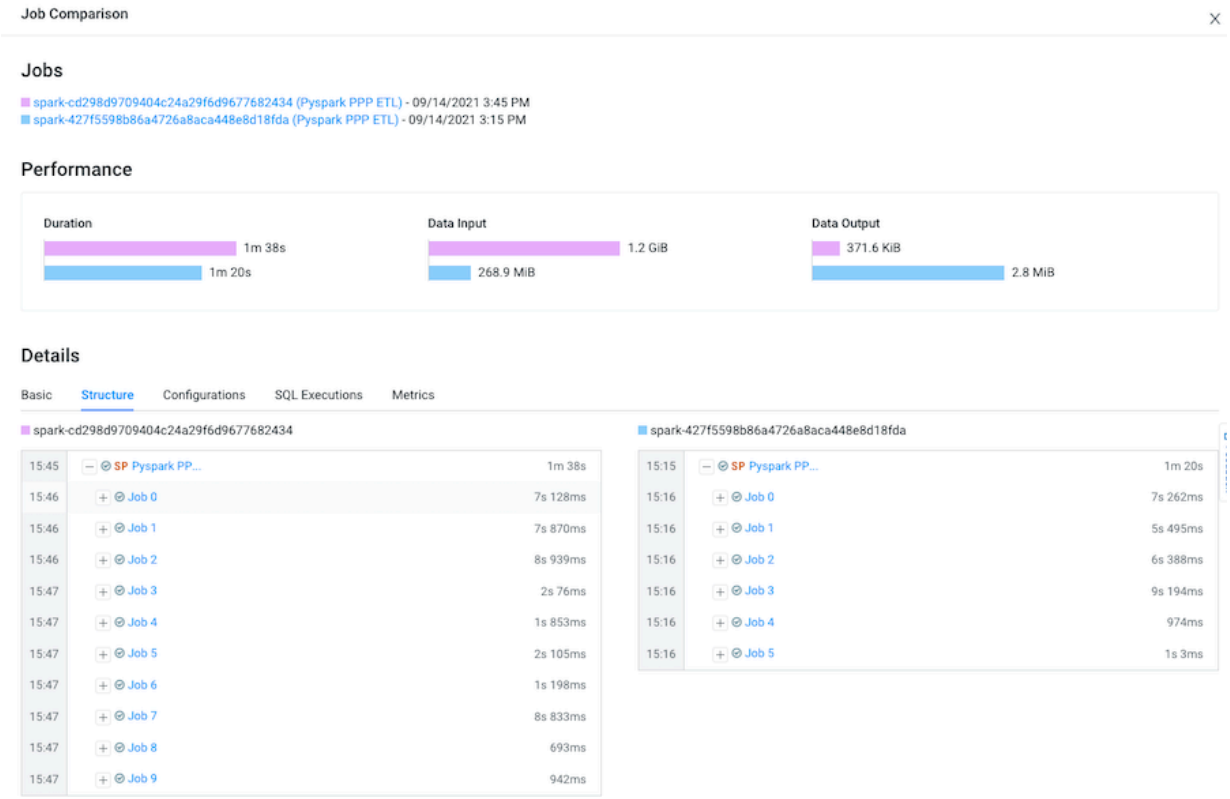
   For this example's comparison, the tabs that contain more information about the job runs are the Structure, SQL Executions, and the Metrics tabs:



Note: The SQL Executions tab is only available for Spark jobs.

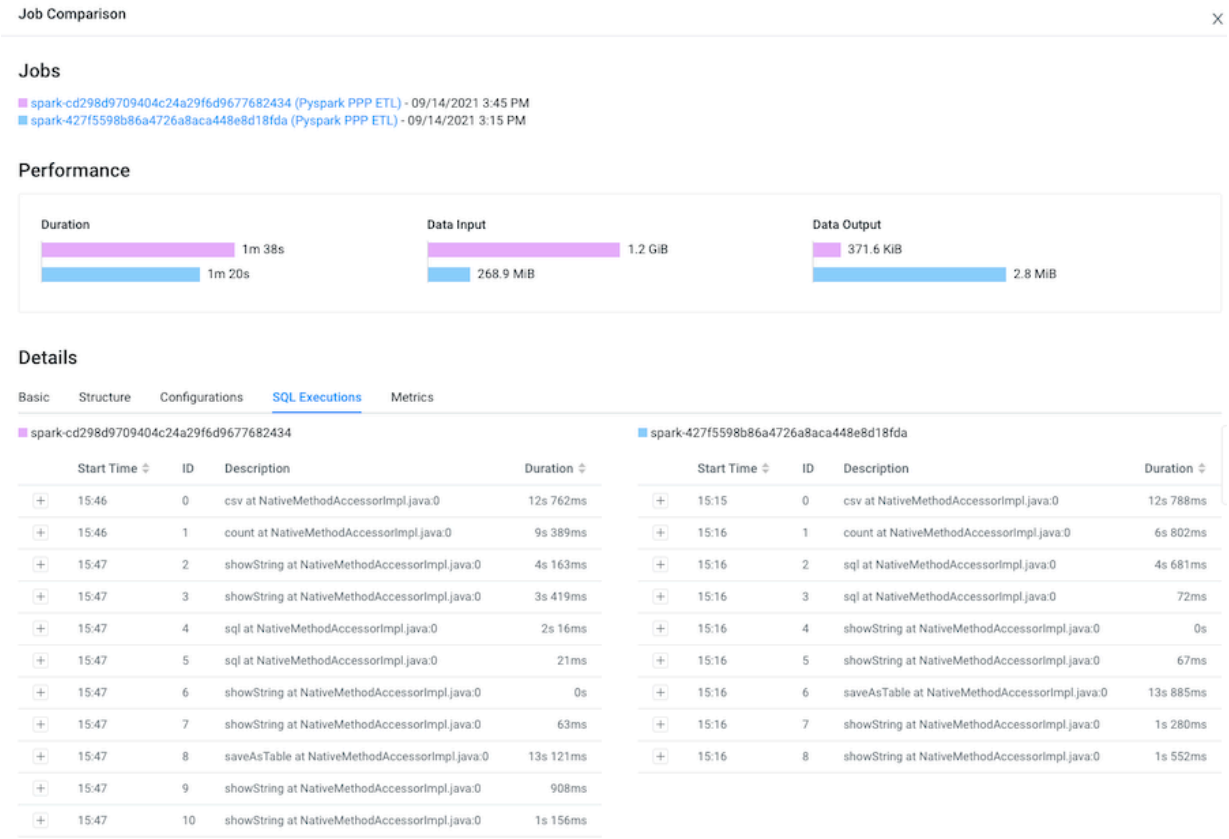
8. Display and compare the sub-jobs executed for both of your selected job runs by selecting the Structure tab.

For example, as shown in the following image, the last run of the job (with health issues) completed in 1minute and 38 seconds and executed 9 sub-jobs and the run that had no health issues took 1 minute and 20 seconds but only executed 5 sub-jobs. Clicking any of the listed sub-jobs displays more details.

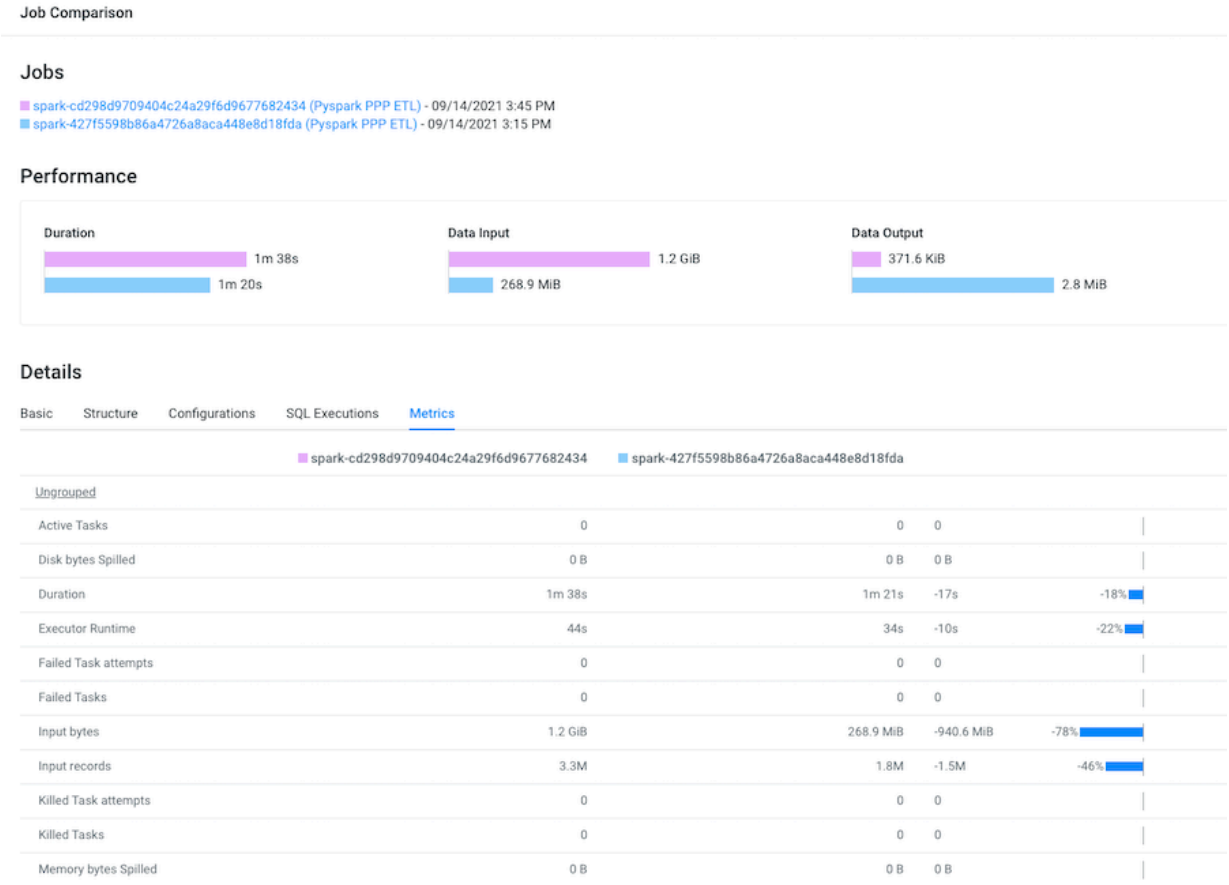


9. Display and compare what Spark SQL was run and how long they ran for both of your selected job runs by selecting the SQL Executions tab.

For example, as shown in the following image, more Spark SQL queries were performed on the data in the last job run.



10. Display and compare what metrics were performed on both of your selected job runs by selecting the Metrics tab. For example, as shown in the following image, more input records were digested in the last job run.



Identifying File Size Storage Issues

Data stored in small files or partitions may create performance issues. The File size reporting feature helps you identify data that is stored inefficiently in small files or partitions.

Important: At this time the Workload XM File Size Report feature is only supported on CDH Workload clusters, version 6.3 to version 7.0, with Cloudera Navigator enabled. CDP Workload clusters are not supported.

A table's data maybe stored in a large number of files, perhaps millions of files. For example, the first time you run an Impala query it loads the metadata for each file, which can cause processing delays. In addition, every time you change a query, refresh the metadata, or add a new file or partition, Impala reloads the metadata. This puts pressure on the NameNode, which stores each file's metadata.

The Workload XM file size reporting enables you to identify tables that have a large number of files or partitions. For example, for queries that run slowly or when an Impala cluster crashes, you can view a table's metadata to determine whether a large number of files or partitions are causing the problem.

Note: Before you can view the file size metadata in Workload XM, you must enable file size reporting in Cloudera Manager. Once enabled, the file size metadata is saved in HDFS, which is then forwarded to Workload XM by Telemetry Publisher.

Displaying File Size Metadata

Steps for displaying a table's File Size report and the metadata that describes the table's file size distribution.

About this task

Describes how to open a table's File Size report and display the metadata.

Important: At this time the Workload XM File Size Report feature is only supported on CDH Workload clusters, version 6.3 to version 7.0, with Cloudera Navigator enabled. CDP Workload clusters are not supported.

Procedure

1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. In the Search field of the Clusters page, enter the name of the cluster whose workloads you want to analyze.
3. From the Navigation side-bar, select File Size Report.
4. In the File Size Report page, either search for a specific table, or locate the table by sorting the tables by the number of files, the number of partitions, or the table size.

   For example, the File Size Reports shows that the Animantarx table has 7 million files and 913 partitions.

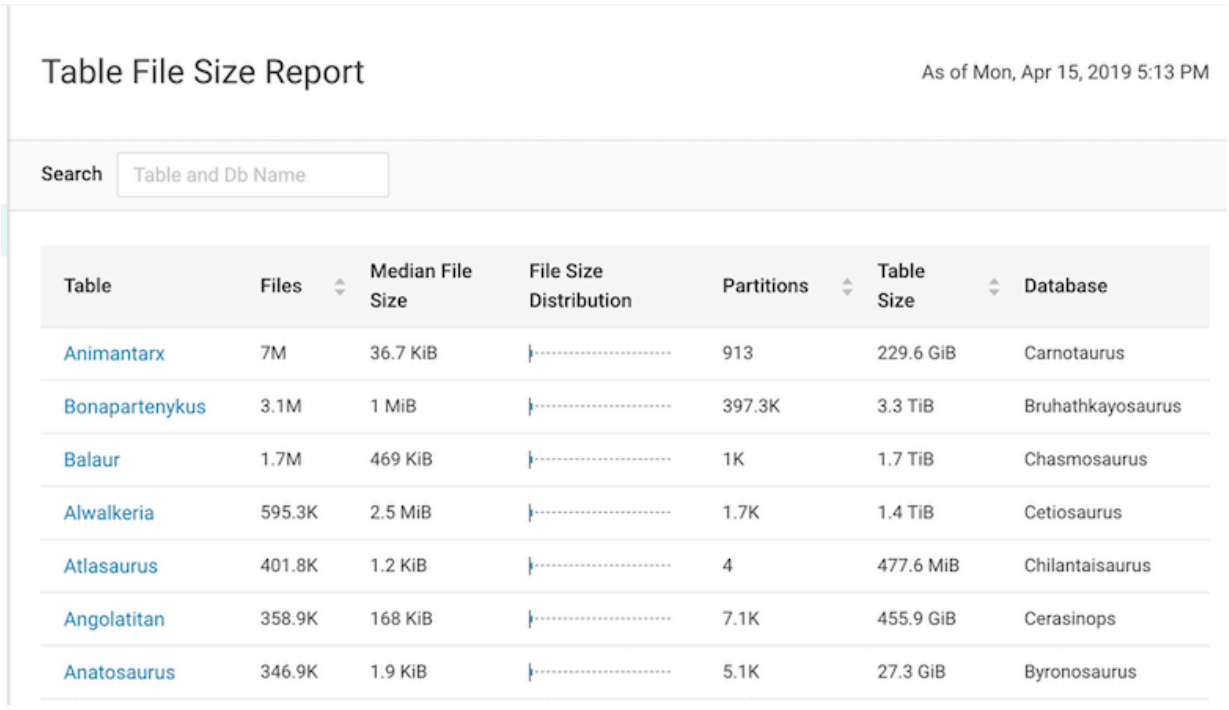
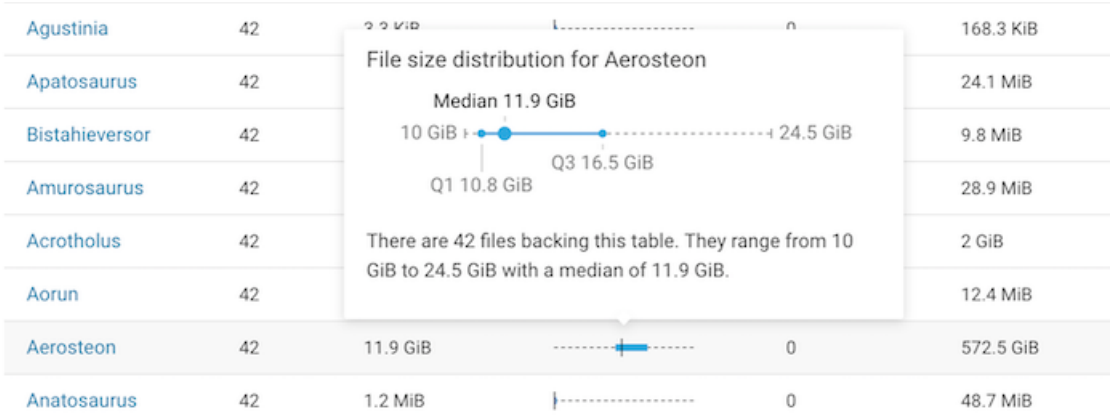


Table	Files	Median File Size	File Size Distribution	Partitions	Table Size	Database
Animantarx	7M	36.7 KiB		913	229.6 GiB	Carnotaurus
Bonapartenykus	3.1M	1 MiB		397.3K	3.3 TiB	Bruhathkayosaurus
Balaur	1.7M	469 KiB		1K	1.7 TiB	Chasmosaurus
Alwalkeria	595.3K	2.5 MiB		1.7K	1.4 TiB	Cetiosaurus
Atlasaurus	401.8K	1.2 KiB		4	477.6 MiB	Chilantaisaurus
Angolatitan	358.9K	168 KiB		7.1K	455.9 GiB	Cerasinops
Anatosaurus	346.9K	1.9 KiB		5.1K	27.3 GiB	Byronosaurus

5. To display details about the table's file size distribution, select a table name.

   For example, the following details window shows that the Aerosteon table uses 42 data files that range from 10 to 24.5 GiB and the graph displays the Q1 and Q3 file size distribution.



Displaying the Metadata of a Table

Steps for displaying a table's metadata that could be causing a query to run slowly.

About this task

Describes how to display the metadata of a table used in your query, such as the table's file size distribution that could be causing your query statement to run slowly.

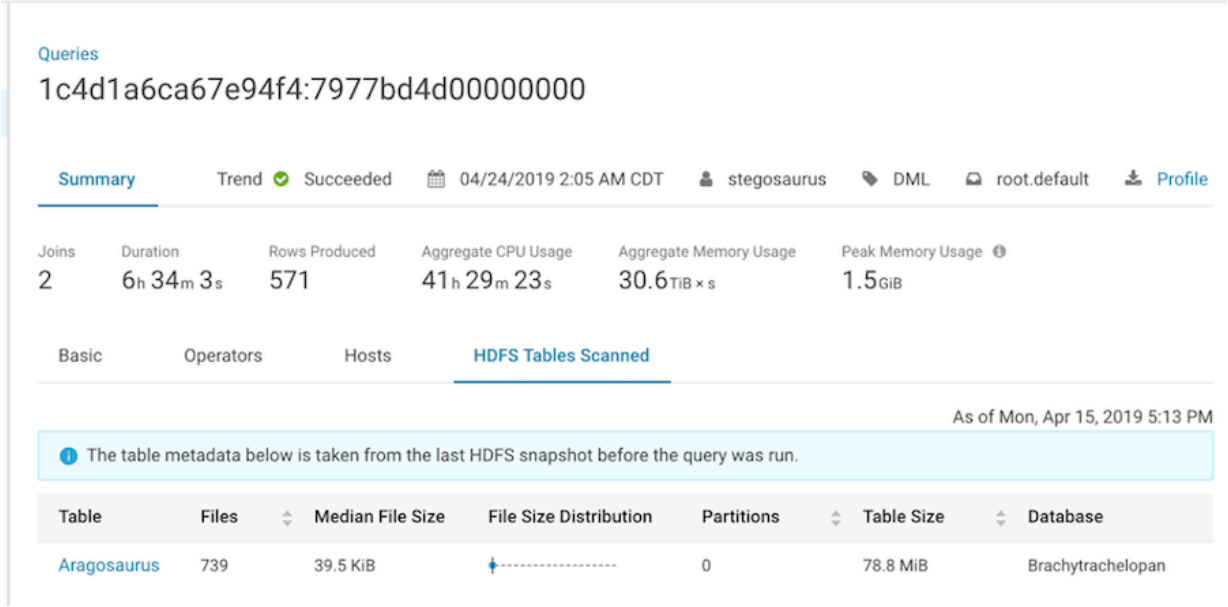
Procedure

1. Verify that you are logged in to the Workload XM web UI.
   a) In the URL field of a supported web browser, enter the Workload XM URL that you were given by your system administrator and press Enter.
   b) When the Workload XM Log in page opens, enter your Workload XM user name and password access credentials.
   c) Click Log in.
2. In the Search field of the Clusters page, enter the name of the cluster whose workloads you want to analyze.
3. From the Navigation side-bar, select Summary.

4. In the Queries page, select the query of interest and then select the HDFS Tables Scanned tab.

   For example, the Duration column shows that the query took over six hours to run and the HDFS Tables Scanned section displays the metadata for the tables that were scanned.

   Note: This is not the number of files accessed, but the total number of files that were in the table the last time a HDFS snapshot was taken before the query was run.



5. To display details about the table's file size distribution, select a table name.

Understanding the Workload XM Cluster Services Metrics

Describes the Workload XM cluster services metrics, which are visually displayed in a series of charts that show the state, activity, and performance of the Workload XM cluster services. Accessed from Cloudera Manager they help you monitor the health, performance, and workload usage of your Workload XM Cluster Services for identifying and troubleshooting existing and potential problems.

Note: The Workload XM cluster services metrics are viewable in Cloudera Manager version 7.5.3 and above. They also require Workload XM version 2.2.2 or 2.3.0 and the latest version of Telemetry Publisher.

The Workload XM cluster services metric charts are displayed on the Workload XM Cluster and the Workload XM Services pages located in the Cloudera Manager Admin Console. For further analysis, each metric chart can be opened to display more detailed information.

The metrics displayed are dependent on the selected Workload XM element. But, whether you view a chart from the Workload XM Cluster Status tab page, the Charts Library tab page, or a Workload XM Service's page, the basic functionality works in the same way.

For example, you can:

- Change the size of the chart by dragging it's lower-right corner.
- View detailed information about elements of interest in the chart by hovering your mouse over the element. When you move your mouse horizontally across the chart, the data values will change according to the time represented.
- For additional information, you can enlarge the pop-up window by clicking Click to expand.

- When the pop-up window is fully expanded you can view:
  - The Workload XM service associated with the chart by clicking View Service.
  - Display the chart on its own page by clicking View Entity Chart.

    Note: If the chart displays more than one stream of data, the new chart displays only the stream that was selected.

For more information about charts in Cloudera Manager, click the Related Information link below.

Workload XM Cluster Chart Library Categories

The Workload XM Status Page visually displays a limited set of metrics that are based on historical Workload XM user payload analysis.

The Charts Library displays a much larger set of metric charts, which are organized into categories.

The following lists the Workload XM Chart Library categories available on the Charts Library page, accessed by clicking the Charts Library tab:

- Status Page Charts, whose charts display a consolidated view of the overall Workload XM cluster metrics.
- Zookeeper Queue, whose charts display the ZooKeeper service metrics, including the number of queues and shards for all streams.

  When the number of messages in a Zookeeper queue exceeds the defined threshold limits, a Workload XM health check alert is triggered. For more information about the Zookeeper Elevated Queue Count alerts, click the Related Information link below.
- Counters, whose charts display the number of jobs received and the number of jobs that failed. Counter metrics are also separated into Pipeline, Analytic Database, and SDX service categories.
- Processing Timers, whose charts display the average job processing time and the average rate across servers. They are calculated using the 75th and 95th percentiles. Processing Timer metrics are also separated into Pipelines and Analytic Database service categories.

  When less than 75% of the service's audit payloads are processing slower than the defined yellow and red timer threshold limits, a Workload XM health check alert is triggered. For more information about the Slow Payload Processing Timer alerts, click the Related Information link below.
- Events, whose charts display the number of important and informational alerts.

Workload XM Services Categories

The Workload XM Services chart categories are accessed by selecting the Workload XM service in the Status Summary section of the Workload XM Status page.

The following lists the Workload XM service's chart categories. As they are dependent on the service you are viewing, not all of the following categories will be displayed:

- Status Page Charts, whose charts display a limited set of Workload XM service's metrics.
- Counters, whose charts display the number of jobs or queries received and the number of jobs or queries that failed.
- Processing Timers, whose charts display the average job processing time and the average rate across servers. They are calculated using the 75th and 95th percentiles. Processing Timer metrics are also separated into Pipelines and Analytic Database service categories.
- Payload Size, whose histogram charts display the average, maximum, minimum, and 75th percentile processing payload sizes.
- Process Resources, whose charts display metrics about the service's processing resources, such as the amount of resident memory used.
- Host Resources, whose charts display metrics about the service's host, which are broken down, depending on the service, into CPU, Memory, Disk Aggregates, Disk Comparison, Network Aggregates, Network Interface Comparison, File Descriptors, and Entropy categories.
- Liveness, whose chart displays metrics about the service's processing performance.

- Events, whose charts display the number of important and informational alerts

Related Information

Understanding the Workload XM Services Health Check Alerts
Viewing Charts for Cluster, Service, Role, and Host Instances

Understanding the Workload XM Services Health Check Alerts

Describes the Workload XM cluster services health check alerts and thresholds, and what actions are required to resolve the problem.

Understanding the Health Check Alert Threshold Colors

Workload XM Health Check Alerts and suggested actions are located on the Workload XM service's health test page. When completed they are compared to their defined thresholds that determine if the service element is Good, Concerning, or Bad. For example, when a service's ZooKeeper queue size has exceeded the Critical threshold (Red Alert) limit, the health check will trigger an alert and display an alert message, the cause, and corrective actions.

For descriptions of the health checks performed on each Workload XM cluster service, click the Related Information link below.

To help you recognize the severity level of the Workload XM health check, the health check results include the following colors:

Table 4: Health Check Alert Colors

Alert Color	Severity
Green	Good - The health check result is normal and within the acceptable range.
Yellow	Concerning - The health check result has exceeded the Warning threshold limit and indicates a potential problem, which eventually must be resolved but does not have to be completed at this time. See the corrective actions in the Actions and Advice sections.
Red	Bad - The health check result has exceeded the Critical threshold limit and indicates a serious problem, which must be resolved immediately. See the corrective actions in the Actions and Advice sections. For example, the Hive Audit Zookeeper queue size has exceeded the Critical threshold limit and can no longer process messages. Possible actions are: • Change the Hive Audit Zookeeper queue size for this role instance, which will reduce the number of messages in the queue. • View the log for the role instance at the time of the health test to see what changed.

Elevated Queue Count

A Workload XM health check alert is triggered when the number of messages in the workload queue exceeds the defined yellow and red threshold limits.

Table 5: ZooKeeper Elevated Queue Count

Queue Name	Default Yellow Alert Threshold	Default Red Alert Threshold
SparkEventLog	100K	200K
PSE	400K	800K
Other services	200K	400K

To address an alert consider the following:

- Check the status of Telemetry Publisher, specifically did it restart after a long pause, as this will create a sudden influx of pending workload data records and increase the size of the queue.
- Check whether any pipelines or ADB services are down, as this will prevent the queues from clearing and workloads from being processed.
- Check whether any new environment, cluster, or workloads are now publishing to your Workload XM cluster, as this could result in new jobs sending large amounts of data at the same time as your jobs.
- Check the health of the Zookeeper service.

  Note: The Zookeeper service is used by Workload XM to manage workload queues.

- Check whether the maximum number of Zookeeper connections is configured correctly for your environment.

If none of the above corrects the problem, contact Cloudera Support and create a support ticket.

Slower Payload Processing times

A health check alert is triggered when less than 75% of the service's audit payloads are processing slower than the defined yellow and red timer threshold limits.

Table 6: Slower Payload Processing Times

Payload Type	Default Yellow Alert Threshold	Default Red Alert Threshold
All services	30 seconds	60 seconds

To address an alert consider the following:

- Check the number of items in the ZooKeeper queue, as too many items can slow down processing.
- Check that the HBase Region Servers are in good health.
- Check that the Phoenix Query Server (PSQ) instances are up and running.
- Check that the Pipeline server instances are up and running.
- Check the Pipeline Server payload size metric, which denotes the size of each job and how much data is being sent. An increase in the average payload size will lead to longer processing times.

If none of the above corrects the problem, contact Cloudera Support and create a support ticket.

Related Information

Workload XM Cluster Services Health Checks

Accessing the Workload XM Cluster Services Charts

Describes where to view the Workload XM cluster services charts in Cloudera Manager that show the state, activity, and performance of the Workload XM services.

About this task

Steps for accessing the Workload XM cluster services charts in Cloudera Manager.

Procedure

1. In a supported web browser, log in to Cloudera Manager as a user with full system administrative privileges.
2. From the Navigation panel, select Clusters and then WXM.

   A subset of the most commonly used Workload XM Cluster services metrics are displayed as charts in the Charts section.

3. Do one or more of the following:
   - To display more Workload XM metrics, select the Charts Library tab and then select a category.
   - To display metrics for a specific Workload XM service, in the Status Summary section of the Status page, click a server name.

What to do next

Manually create your own Workload XM charts using the Cloudera Manager Chart builder and the Workload XM service metric name. For more information on how to build your own chart, click the Related Information links below.

Related Information

Building Your Own Workload XM Services Metric Chart

Workload XM Cluster Services Metrics

Building Your Own Workload XM Services Metric Chart

Describes the steps to manually build a Workload XM metric chart in Cloudera Manager using the Cloudera Manager Chart builder and the Workload XM services metric name.

About this task

Steps for building your own Workload XM Services metrics chart.

Note: Displaying the predefined Workload XM Services metric charts in Cloudera Manager requires Cloudera Manager version 7.5.3 and above. The metrics also require Workload XM version 2.2.2 or 2.3.0 and the latest version of Telemetry Publisher.

Note: These instructions assume that you have read and recorded the required service metric name for your chart from the predefined Workload XM Cluster Services Metrics.

For more information about the metrics collected from each server by Workload XM, click the Related Information link below.

Procedure

1. In a supported web browser, log in to Cloudera Manager as a user with full system administrative privileges.
2. From the Navigation panel, select Charts and then Chart Builder.
3. In the Search field, enter search and then the metric name:

   search *metric_name*

   For example, search wxm_dbus_api_service_heap_used
4. Click Build Chart.

Related Information

Workload XM Cluster Services Metrics

Purging HDFS Data

Reduce bottlenecks between Telemetry Publisher and Workload XM, free up storage space, and increase job and query runtime efficiency by removing obsolete HDFS data that exceeds the maximum retention limit.

Note: Cloudera recommends performing regular purge events for HDFS files that are no longer required.

Understanding the Purge Date used by the Purge Event

Describes the Workload XM purge event's criteria that is based on the file's data group and the data group's retention limit and how the purge date is calculated.

The purge event's criteria is based on the maximum data retention policy, described in days, for the following HDFS data groups:

- Temporary data, when the retention period exceeds 8 days
- Staging data, when the retention period exceeds 31 days
- Detailed data, when the retention period exceeds 181 days
- Summarized data, when the retention period exceeds 731 days

The purge date is calculated by subtracting the retention days, specified by the maximum data retention period policy, from the current date and comparing the resultant date with the data's timestamp date. If the data's timestamp date is less than or equal to the resultant date the data is removed.

The data's timestamp date is determined by where the data resides:

- If the data resides in the cloudera-bus root directory, the timestamp date is extracted from the subdirectory name. For example, if the directory name is /cloudera-dbus/HiveAudit/2021030623. The timestamp date extracted by the purge event is 2021/03/06, using the YYYY/MM/DD date format.

  Important: The purge event deletes files from the cloudera-dbus directory as follows:

  - If the date is successfully extracted and is less than or equal to the resultant date, all the files in the directory are removed and are counted as one file by the maximum deletion limit.
  - If the date is successfully extracted, is less than or equal to the resultant date, and a file or files are set in the blobstore.purger.paths.to.keep parameter, all the files except the file or files set in the blob store.purger.paths.to.keep parameter are removed and each file that is removed is counted by the maximum deletion limit.
- If the data resides in a cloudera-sigma-olap-impala, cloudera-sigma-partial-pse, cloudera-sigma-pse-extended, or cloudera-sigma-sdx-payloads root directory, the timestamp date is extracted from the file's last modified time.

Obsolete data can be purged from the following HDFS root directories:

- cloudera-dbus
- cloudera-sigma-olap-impala
- cloudera-sigma-partial-pse
- cloudera-sigma-pse-extended
- cloudera-sigma-sdx-payloads

Workload XM Purge Event Parameters

Lists the Workload XM purge event parameter settings that enable you to set the event's execution time, frequency, and maximum purge duration. You can also exclude files and directories from being purged with the blobstore.purger .paths.to.keep parameter setting.

Table 7: Purge Event Parameters

Parameter	Description	Example
blobstore.purger.frequency	The purge event's recurring schedule, based on one of the following values: • None. By default, the purge process is set to none. • Daily. When this value is set for the first time, files are automatically deleted the next day at 1am. • Weekly. By default, files are automatically deleted every Saturday at 1am. • Monthly. When this value is set for the first time, files are automatically deleted the last Saturday of the month at 1am. Thereafter, files are deleted every 28th day. The monthly parameter uses the 28 day calendar format	blobstore.purger.frequency = none
blobstore.purger.start.time	The purge event's start time, based on the 24-hour time format. Where, 01:00 and 0:00 are valid time values, and 24:00, 1:0, and 01:0 are not valid time values By default, Workload XM schedules the purge process when it will cause the least amount of disruption to users. **Note:** Cloudera recommends scheduling a time during non-peak working hours or job execution hours.	blobstore.purger.start.time = 01:00
blobstore.purger.paths.to.keep	Lists the files and directories that are to be excluded from the purge event. Where each file and/or directory is separated by a comma and where: • a file value must use its full path, directory name, and file name. • a directory value must use its full path and directory name.	`blobstore.purger.paths.to.keep= /cloudera-dbus/ImpalaQueryProfile/20210302170217/7d2bcefa-8819-4fa1-be0c-4529ee4eb98f, /cloudera-dbus/HiveAudit, /cloudera-sigma-olap-impala/02f54999-b9a4-4dca-8237-d1b047755efb, /cloudera-sigma-sdx-payloads/2bc85719-7a3e-4438-96a4-8fc0f77ff`

Parameter	Description	Example
blobstore.purger.delete.request.limit	The maximum deletion limit. By default, the maximum number of files that can be deleted by the purge process is 500,000. This ensures that a purge cycle is not overloaded, does not introduce bugs, or takes up too much time. When the deletion limit is met, the purge process: • Stops processing for a daily scheduled value. • Stops processing and restarts the next day for all other scheduled values. **Note:** The purge event's maximum deletion limit calculates all the files in a dbus directory as one file. When you exclude a file or files that reside in the dbus directory from the purge process, the purge event's maximum deletion limit condition calculates all the files in the directory minus those files you have excluded.	blobstore.purger.delete.request.limit=500000

Configuring the Workload XM Purge Event

Steps for scheduling and configuring a purge event.

About this task

Describes how to schedule and configure the Workload XM purge event.

Procedure

1. In a supported web browser, log in to Cloudera Manager as a user with full system administrator privileges.
2. From the Navigation panel, select Clusters and then WXM.
3. In the Status Summary panel of the WXM page, select Admin API Server.
4. Click the Configuration tab.
5. Search for the Admin API Server Advanced Configuration Snippet (Safety Valve) for the wxm-conf/sigmaadminapi.properites option.
6. In the text field enter your purge event's parameter settings, using the *Purge Event Parameters* table.

   For example,

```
blobstore.purger.delete.request.limit=9990000
blobstore.purger.paths.to.keep=/cloudera-dbus/ImpalaQueryProfile/202103021
7/7d2bcefa-8819-4fa1-be0c-4529ee4eb98f,/cloudera-dbus/HiveAudit,/cloudera-
sigma-olap-impala/02f54999-b9a4-4dca-8237-d1b047755efb,/cloudera-sigma-sdx
-payloads/2bc85719-7a3e-4438-96a4-8fc0f77ff79e
blobstore.purger.frequency=daily
blobstore.purger.start.time = 0:00
```

7. Click Save Changes, which sets and schedules the purge process.
8. From the Actions menu, select Restart this Admin API Server.
9. In the Restart this Admin API Server message, confirm your changes by clicking Restart this Admin API Server.
10. When the Restart API Server step window displays Completed, click Close.

Manually Executing a Workload XM Purge Event

You can manually run your purge event immediately with a one-time operation, rather than scheduling a purge event.

About this task

Describes how to manually run a Workload XM purge event.

A one-time purge event is based on the maximum data retention policy using the Workload XM purge event's parameter values, without the frequency value.

Procedure

1. In a supported web browser, log in to Cloudera Manager as a user with full system administrator privileges.
2. From the Navigation panel, select Clusters and then WXM.
3. In the Status Summary panel of the WXM page, select Admin API Server.
4. Click the Configuration tab.
5. Search for the Admin API Server Advanced Configuration Snippet (Safety Valve) for the wxm-conf/sigmaadminapi.properites option.
6. In the text field enter your purge event's parameter settings, using the *Purge Event Parameters* table.

   For example,

   ```
   blobstore.purger.delete.request.limit=9990000
   blobstore.purger.paths.to.keep=/cloudera-dbus/ImpalaQueryProfile/202103021
   7/7d2bcefa-8819-4fa1-be0c-4529ee4eb98f,/cloudera-dbus/HiveAudit,/cloudera-
   sigma-olap-impala/02f54999-b9a4-4dca-8237-d1b047755efb,/cloudera-sigma-sdx
   -payloads/2bc85719-7a3e-4438-96a4-8fc0f77ff79e
   blobstore.purger.frequency=none
   blobstore.purger.start.time = 0:00
   ```
7. Click Save Changes.
8. From the Actions menu, select Restart this Admin API Server.
9. In the Restart this Admin API Server message, confirm your changes by clicking Restart this Admin API Server.
10. When the Restart API Server step window displays Completed, click Close.
11. When a manual purge event run is required, do the following:
    a) Log in to Cloudera Manager.
    b) From the Navigation panel, select Clusters and then WXM.
    c) From the Actions menu, select Purge HDFS Bucket Data.
    d) In the Purge HDFS Bucket Data confirmation message, confirm the purge event by clicking Purge HDFS Bucket Data.
    e) When the Purge HDFS Bucket Data window displays Completed, click Close.

Managing your Workload XM Purge Event

Steps for updating, stopping, and troubleshooting your Workload XM Purge event.

The following management tasks can be performed:

Updating your Workload XM Purging Event

To update your purge event:

1. In a supported web browser, log in to Cloudera Manager as a user with full system administrator privileges.
2. From the Navigation panel, select Clusters and then WXM.

3. From the Status Summary panel, select Admin API Server.
4. Click the Configuration tab.
5. Search for the Admin API Server Advanced Configuration Snippet (Safety Valve) for the wxm-conf/sigmaadminapi.properites option field.
6. In the text field, change the required values.
7. Click Save Changes.
8. From the Actions menu, select Restart this Admin API Server.
9. In the Restart this Admin API Server message, confirm your changes by clicking Restart this Admin API Server.
10. When the Restart API Server step window displays Completed, click Close.

Stopping the Workload XM Purge Event

You can stop a recurring purge event or stop a scheduled purge event whilst still running.

- To stop a recurring purge event:
  1. In a supported web browser, log in to Cloudera Manager as a user with full system administrator privileges.
  2. From the Navigation panel, select Clusters and then WXM.
  3. From the Status Summary panel, select Admin API Server.
  4. Click the Configuration tab.
  5. Search for the Admin API Server Advanced Configuration Snippet (Safety Valve) for the wxm-conf/sigmaadminapi.properites option field.
  6. In the text field, replace the blobstore.purger.frequency value with none.
  7. Click Save Changes.
  8. From the Actions menu, select Restart this Admin API Server.
  9. In the Restart this Admin API Server message, confirm your changes by clicking Restart this Admin API Server.
  10. When the Restart API Server step window displays Completed, click Close.
- To stop a scheduled purge event whilst still running:
  1. In a supported web browser, log in to Cloudera Manager as a user with full system administrator privileges.
  2. From the Navigation panel, select Clusters and then WXM.
  3. From the Status Summary panel, select Admin API Server.
  4. From the Actions menu, select Stop this Admin API Server.
  5. Still in the Admin API Server page, click the Configuration tab.
  6. Search for the Admin API Server Advanced Configuration Snippet (Safety Valve) for the wxm-conf/sigmaadminapi.properites option field.
  7. Replace the blobstore.purger.frequency value with none.
  8. Click Save Changes.
  9. From the Actions menu, select Restart this Admin API Server.
  10. In the Restart this Admin API Server message, confirm your changes by clicking Restart this Admin API Server.
  11. When the Restart API Server step window displays Completed, click Close.

Troubleshooting

The Workload XM purge event does not delete directories and files that do not have the full wxm owner and file permissions. Files and directories may revert back to the hdfs owner when a restore is created from a snapshot. In this case and before creating an automatic or manual purge event you must verify the owner and file permissions of the required files to be purged.

To reset your HDFS files and directories as the wxm owner with full administrative permissions do the following:

1. In a terminal go to the /etc directory and open the hdfs password file by entering:

   vim passwd

2. Search for the kafka parameter.
3. Replace /sbin/nologin with /bin/hash.
4. Save the file.
5. Grant full wxm access permissions to the hdfs password file by using the chown command.

Tracking your Purge Event from Log Entries

You can determine if the purge event was successful or identify potential problems from the Cloudera Manager Admin API Server log files.

The Admin API Server log file entries also list the names of the files and directories that were deleted and provide details about how many files and directories were deleted, the sum total size of the files and directories that were deleted, and the time they were deleted.