Using Ambari Views

Date of Publish: 2018-04-30

Contents

Using an Ambari View

You use a view to interact with your capacity scheduler or file system. The Ambari Pig View, which you can use to interact with Pig scripts, is deprecated in HDP 3.0 and later. Ambari does not enable Pig View. To enable Pig View in HDP 3.0 and later, you need to contact Hortonworks support for instructions that include how to install WebHCat using an Ambari management pack.

Related Information

Understanding Ambari Views

Using YARN Queue Manager View

About this task

The Yarn Capacity Scheduler allows for multiple tenants in an HDP cluster to share compute resources according to configurable workload management policies. The YARN Queue Manager View is designed to help Hadoop operators configure these policies for YARN. In the View, operators can create hierarchical queues and tune configurations for each queue to define an overall workload management policy for the cluster.

Set up YARN workflow queues

To set up Capacity Scheduler queues on a view instance.

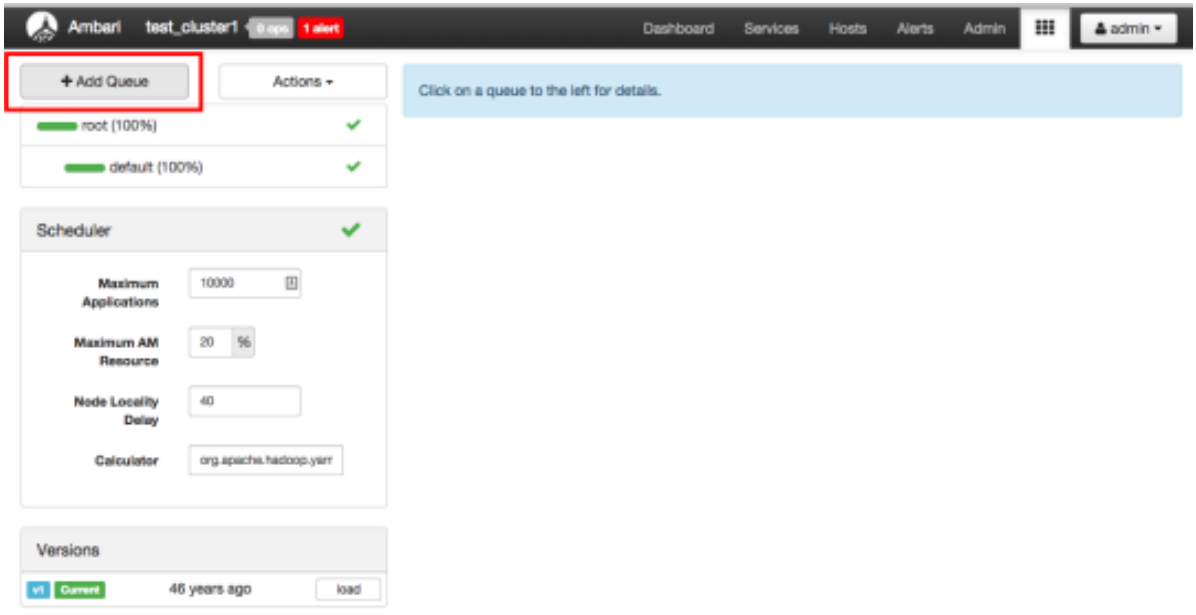
Procedure

1. On the YARN Queue Manager view instance configuration page, click **Add Queue**.

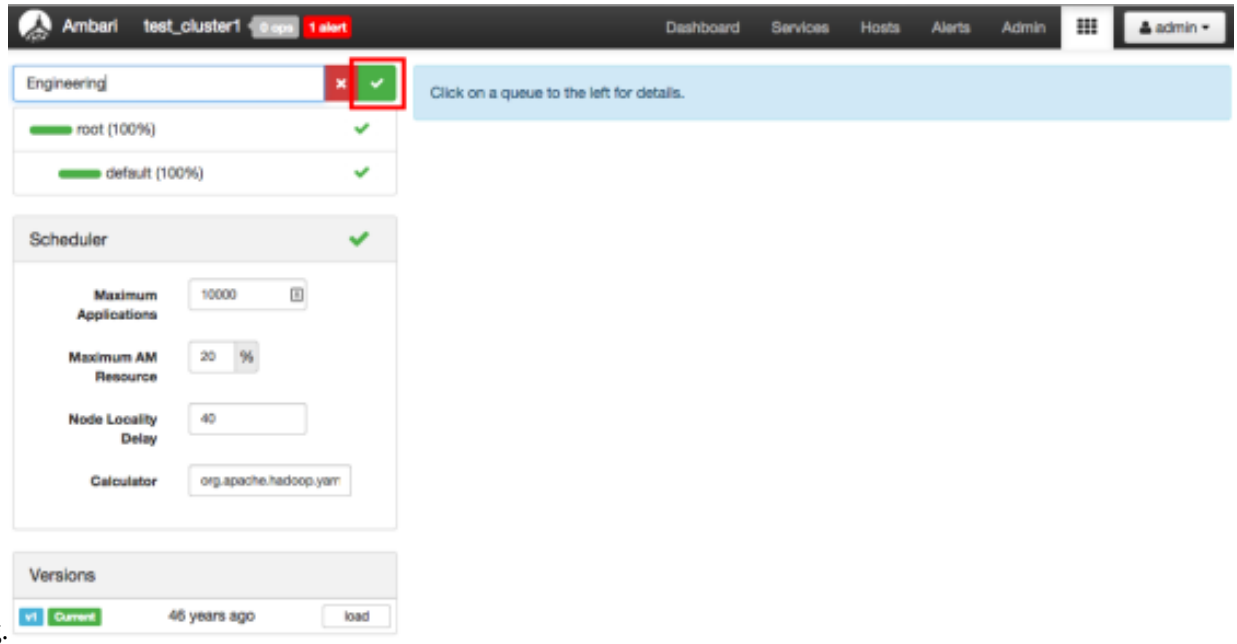
   To return to a previously created YARN Queue Manager view instance:

   a) Click **Views** on the Manage Ambari page.
   b) Click **CAPACITY-SCHEDULER**.
   c) Click the applicable YARN Queue Manager view instance, then click **Go to instance** at the top of the page.

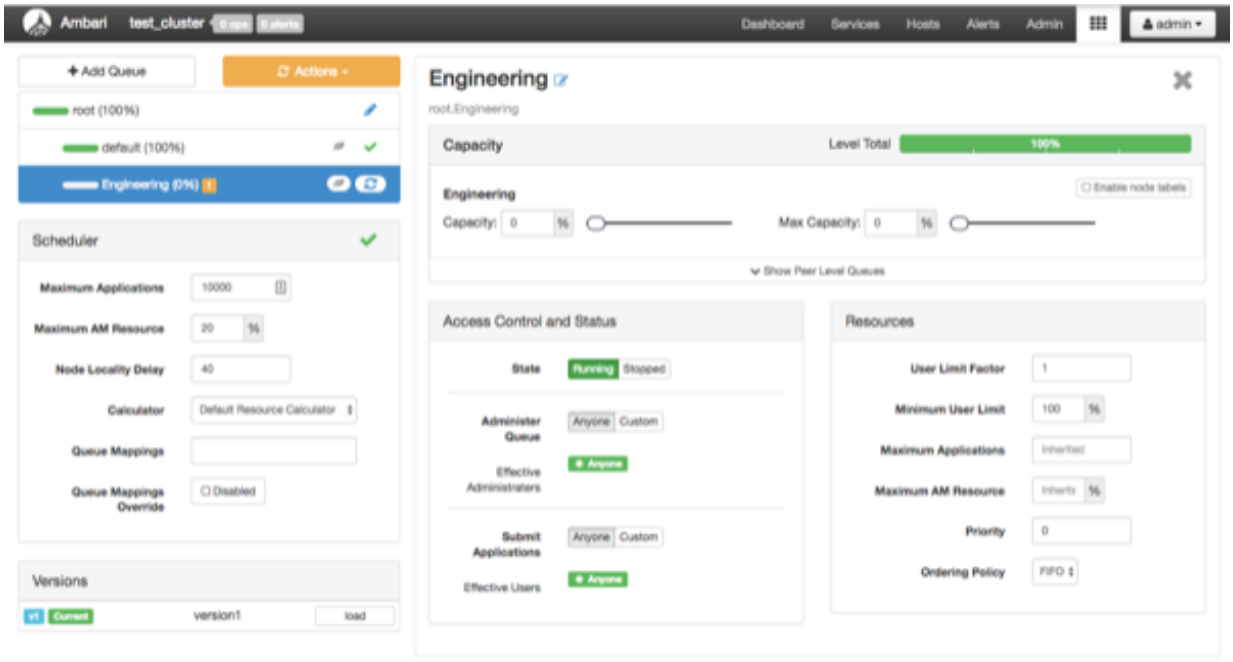
   The queue will be added under the top-level, or root queue. A default queue already exists under the root queue.



2. Type in a name for the new queue, then click the green check mark to create the queue.
   In the following example, we're creating a queue named



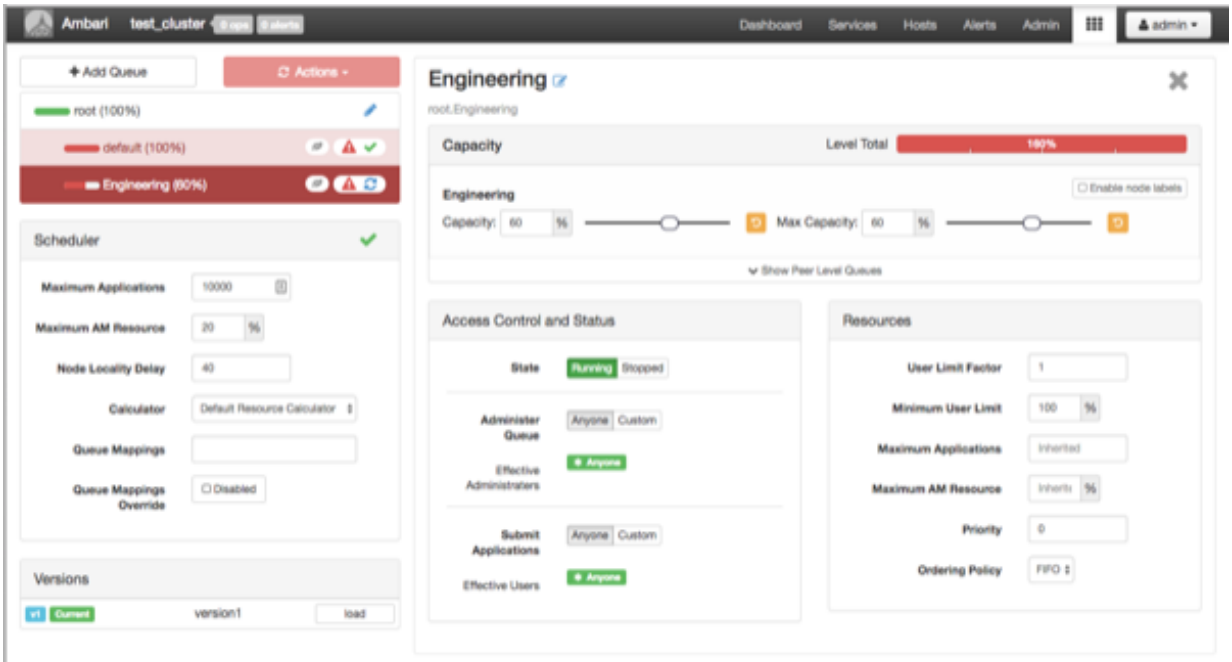
   Engineering.
   The Engineering queue is added, and its configuration page



   appears.

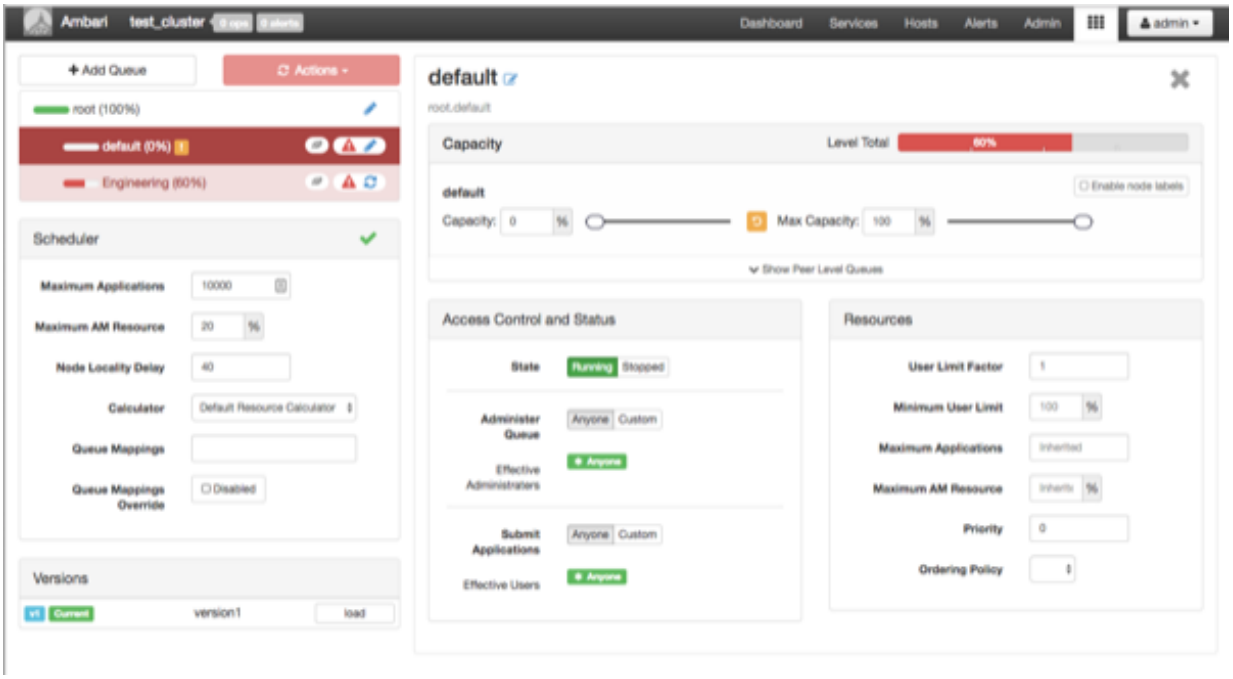
3. Set the capacity for the Engineering queue to 60%.
   The sum of queue capacities at any level in the YARN Queue Manager configuration must total 100%. Here the default queue is already set to 100%. Therefore, if we try to set the

Engineering queue capacity to 60%, error messages appear warning that the total at this level is



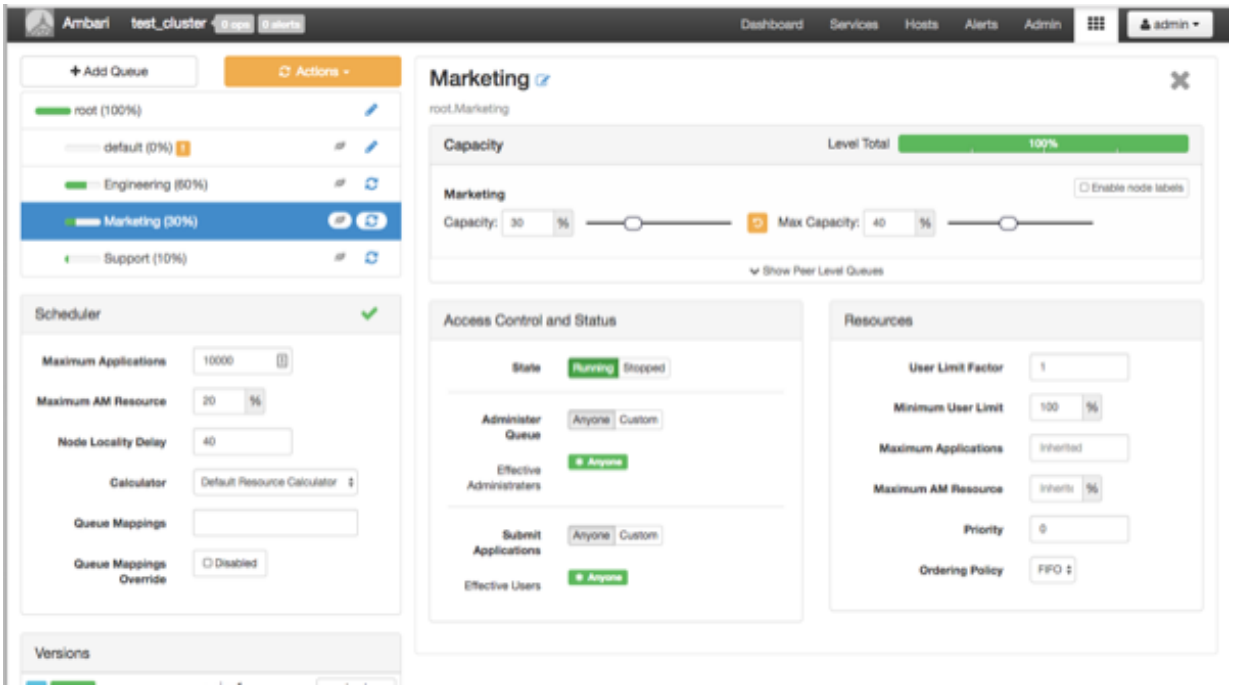
160%.

4. If we click the **Default** queue and set its capacity to 0%, the **Level Total** bar at the top of the page lists the total queue capacity at this level as 60%.

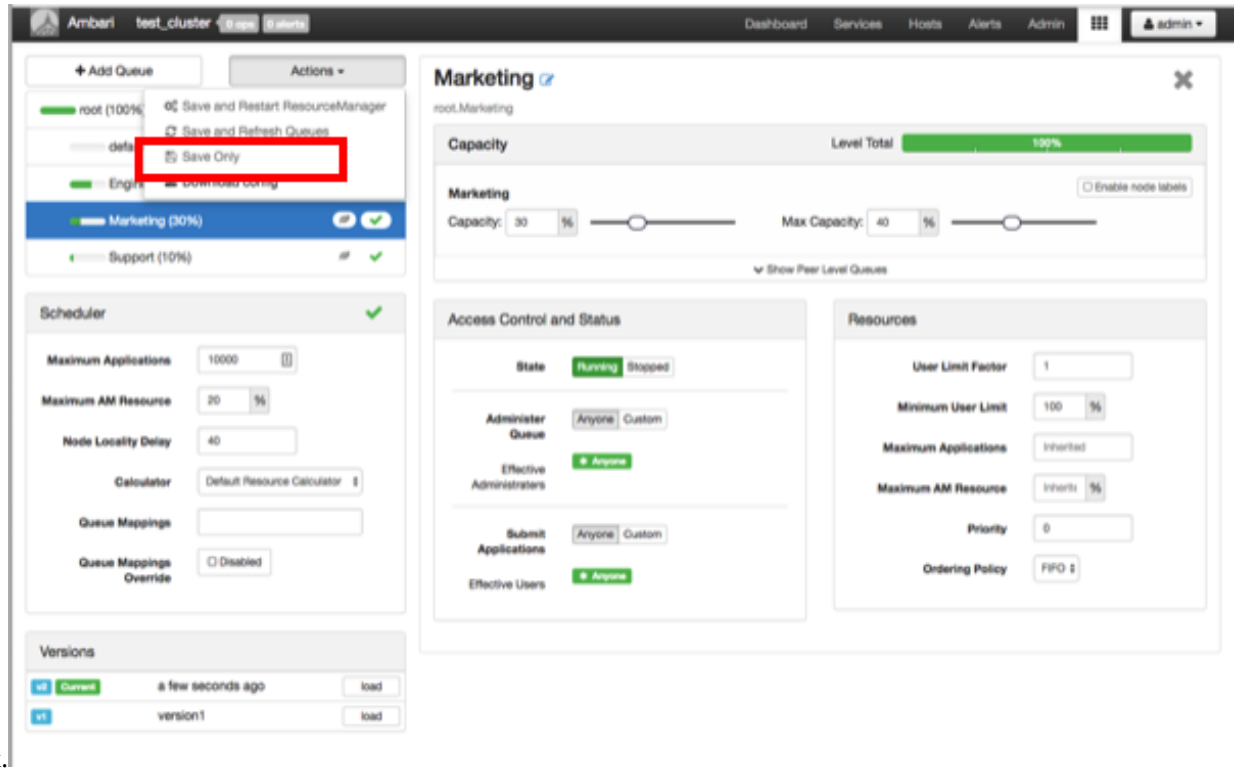


5. To add more queues at the root level, click the **root** queue, then click **Add Queue**.

In the following example, we have added a Support queue set to 10% of the level capacity, and a Marketing queue set to 30%. The root-level queue capacities now total 100%.

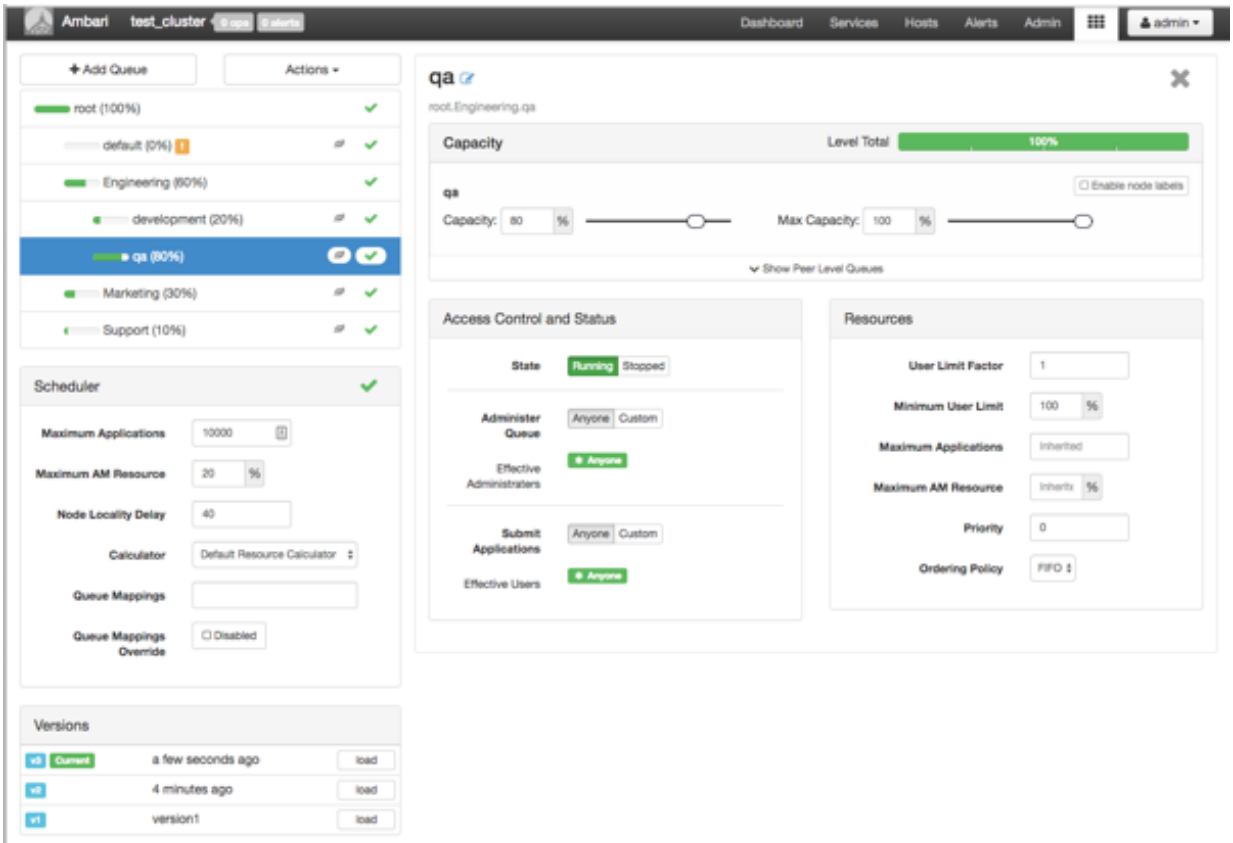


6. To save your configuration, click **Actions** > **Save Only**.
7. On the **Notes** pop-up, enter an optional description of your changes, then click **Save**. Each version is retained and listed in the Versions box.



8. To build a queue hierarchy, click a top-level queue, then click **Add Queue**.

In the following example, the qa and development queues have been added under the Engineering queue.



Configure YARN workflow queues

To configure a queue:

Procedure

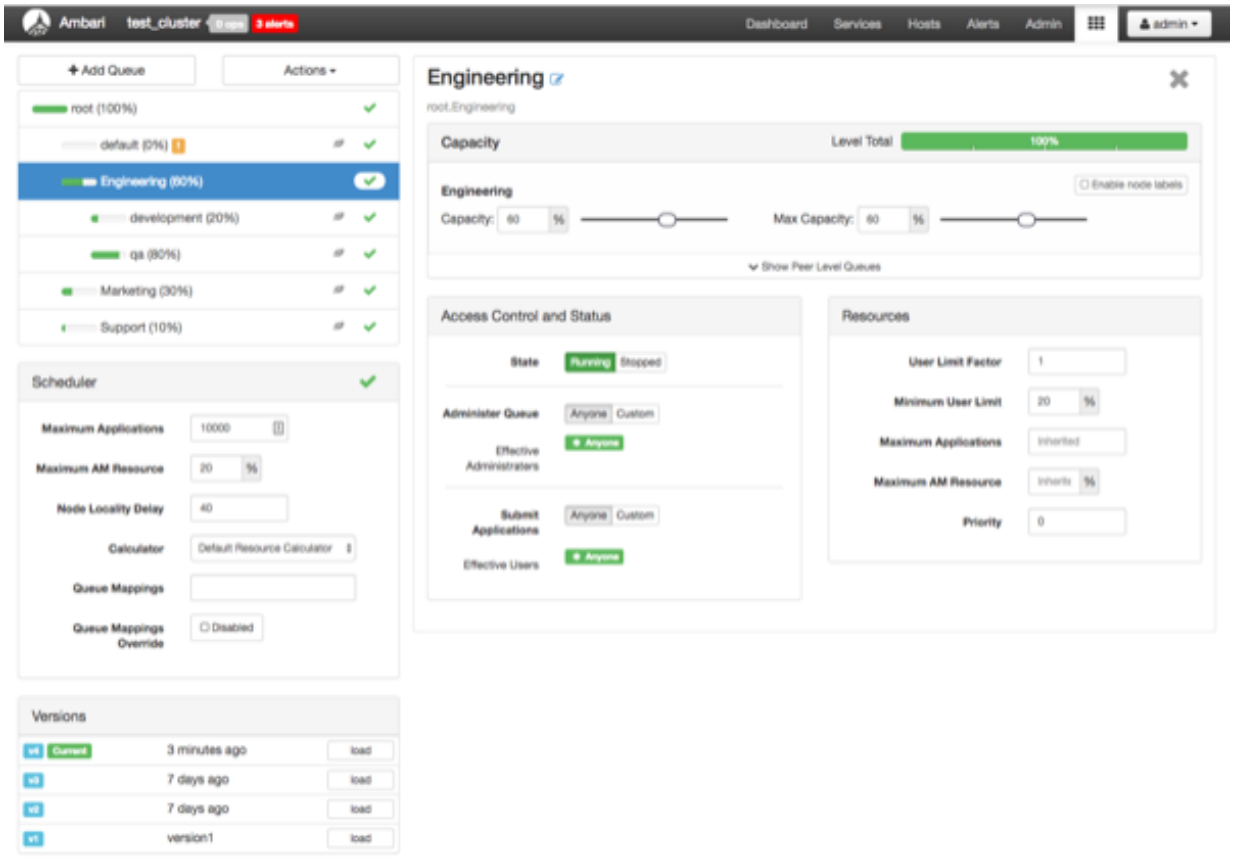
1. Click the queue name.
2. Set the following queue parameters:

   Hover the cursor over a parameter name to display a description of the parameter.

Capacity **Capacity**	The percentage of cluster resources available to the queue. For a sub-queue, the percentage of parent queue resources.
Max Capacity	The maximum percentage of cluster resources available to the queue. Setting this value tends to restrict elasticity, as the queue will be unable to utilize idle cluster resources beyond this setting.
Enable Node Labels	Select this check box to enable node labels for the queue.
Access Control and Status **State**	Running is the default state. Setting this to Stopped lets you gracefully drain the queue of jobs (for example, before deleting a queue).

Administer Queue	Click **Custom** to restrict administration of the queue to specific users and groups.
Submit Applications	Click **Custom** to restrict the ability to run applications in the queue to specific users and groups.
Resources **User Limit Factor**	The default value of "1" means that any single user in the queue can at maximum only occupy the queue's configured capacity. This prevents users in a single queue from monopolizing resources across all queues in a cluster. Setting the value to "2" would restrict the queue's users to twice the queue's configured capacity. Setting it to a value of 0.5 would restrict any user from using resources beyond half of the queue capacity.
Minimum User Limit	This property can be used to set the minimum percentage of resources allocated to each queue user. For example, to enable equal sharing of the queue capacity among five users, you would set this property to 20%.
Maximum Applications	This setting enables you to override the Scheduler Maximum Applications setting. The default setting is Inherited (no override).
Maximum AM Resource	This setting enables you to override the Scheduler Maximum AM Resource setting. The default setting is Inherited (no override).
Priority	An integer value that sets a relative priority for a queue. The default value is 0 for all queues. Setting this to a higher value gives a queue access to cluster resources ahead of queues with lower priorities. In order for YARN Queue Priorities to be applied, you must enable preemption. For more information see Setting YARN Queue Priorities.

The following image shows the example Engineering queue with these settings



specified:

Enable preemption for YARN workflow queues

With Preemption enabled, under-served queues can begin to claim their allocated cluster resources almost immediately, without having to wait for other queues' applications to finish running.

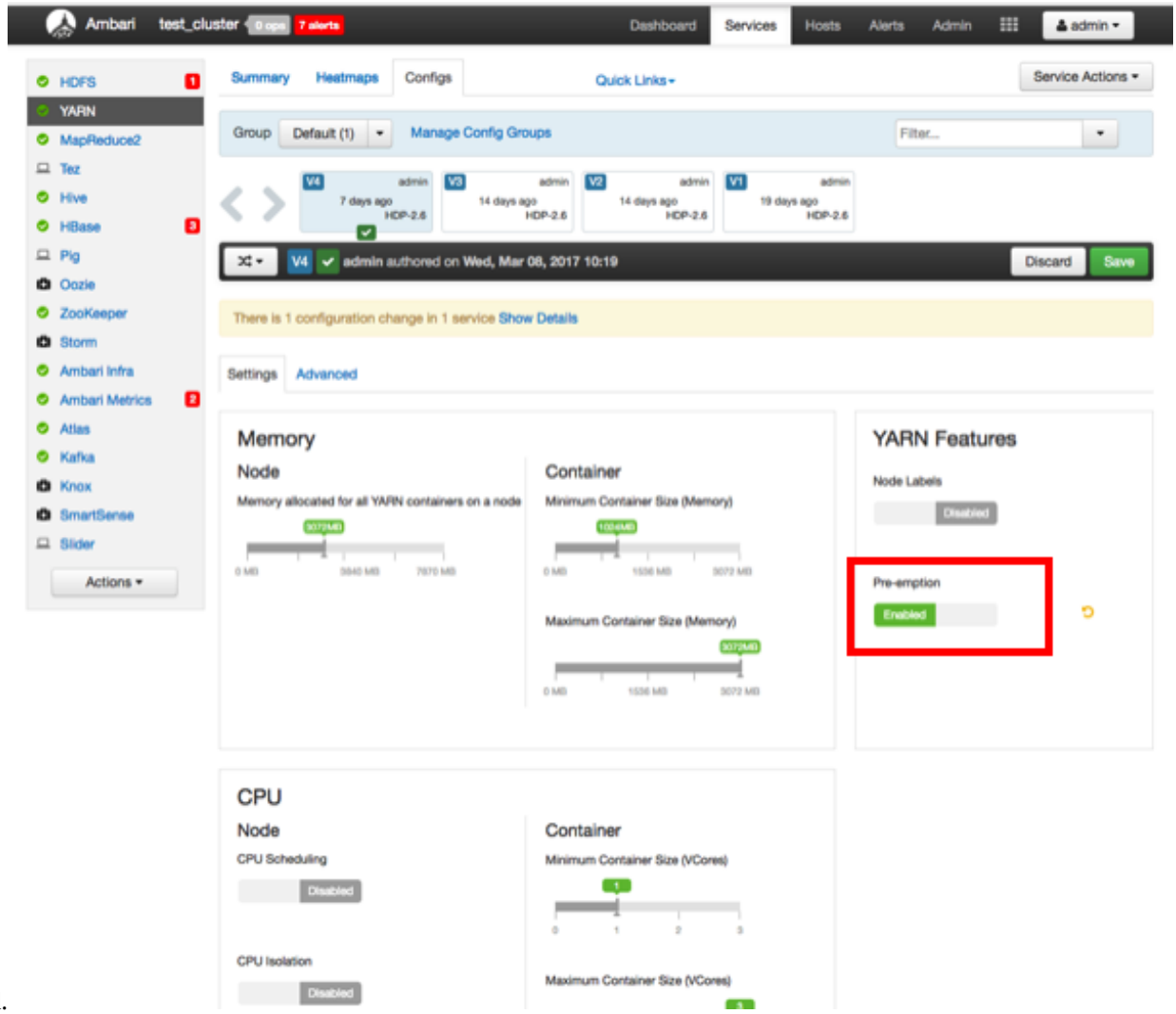
About this task

When using YARN queues, a scenario can occur in which a queue has a guaranteed level of cluster resources, but must wait to run applications because other queues are utilizing all of the available resources. If Preemption is enabled, higher priority applications do not have to wait because lower priority applications have taken up the available capacity.

Procedure

1. In **Ambari Web**, browse to **Services** > **YARN** > **Configs**. In **YARN Features**, click **Pre-emption**.

The button label changes to indicate that Preemption is



enabled.

2. Browse and scroll to **Advanced** > **Custom yarn-site**. Click **Add Property** and use the **Add Property** pop-up to add a new property in the following format:
   yarn.resourcemanager.monitor.capacity.preemption.total_preemption_per_round=<(memory-of-one-NodeManager)/(total-cluster-memory)>

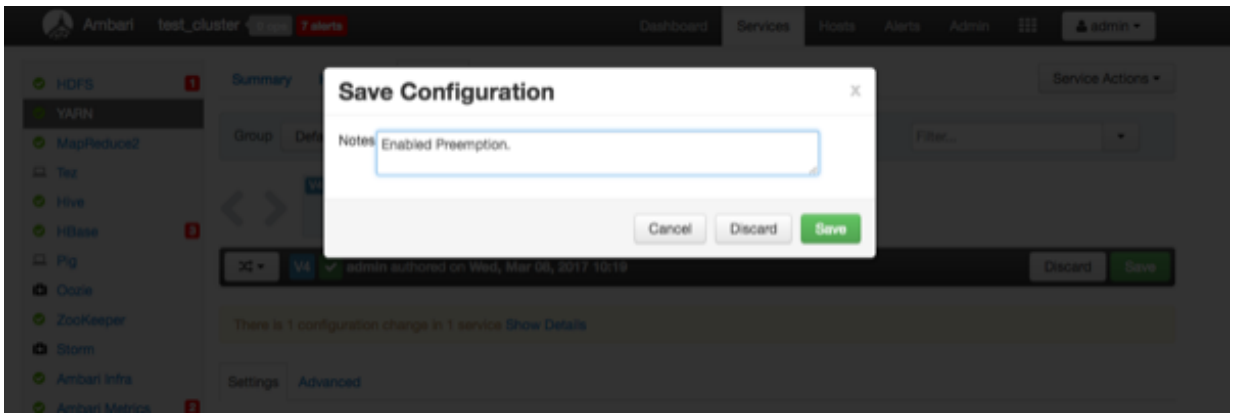
   This is the maximum percentage of resources preempted in a single round. You can use this value to restrict the pace at which Containers are reclaimed from the cluster. After computing the total desired preemption, the policy scales it back to this limit. This should be set to (memory-of-one-NodeManager)/(total-cluster-memory). For example, if one NodeManager has 32 GB, and the total cluster resource is 100 GB, the total_preemption_per_round should set to 32/100 = 0.32. The default value is 0.1 (10%):

   yarn.resourcemanager.monitor.capacity.preemption.total_preemption_per_round=0.1

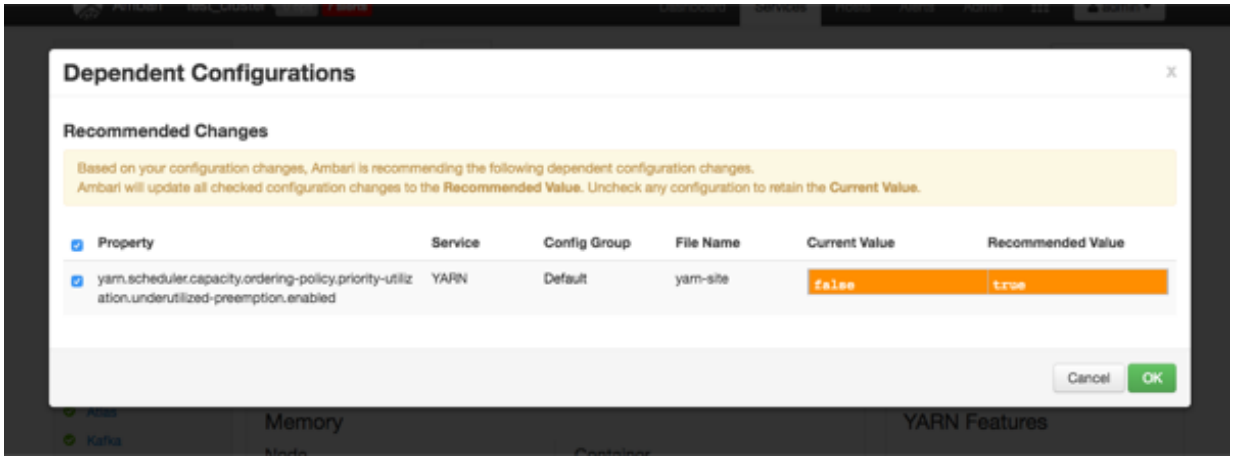
3. Click **Add Property** again and use the Add Property pop-up to add the following **Custom yarn-site** property:
   yarn.resourcemanager.monitor.capacity.preemption.natural_termination_factor=1.0

   Similar to total_preemption_per_round, you can apply this factor to slow down resource preemption after the preemption target is computed for each queue (for example, “give me 5 GB back from queue-A”). For example, if 5 GB is needed back, in the first cycle preemption takes back 1 GB (20% of 5GB), 0.8 GB (20% of the remaining 4 GB) in the next, 0.64 GB (20% of the remaining 3.2 GB) next, and so on. You can increase this value to speed up resource reclamation. The recommended value for this parameter is 1.0, meaning that 100% of the target capacity is preempted in a cycle.

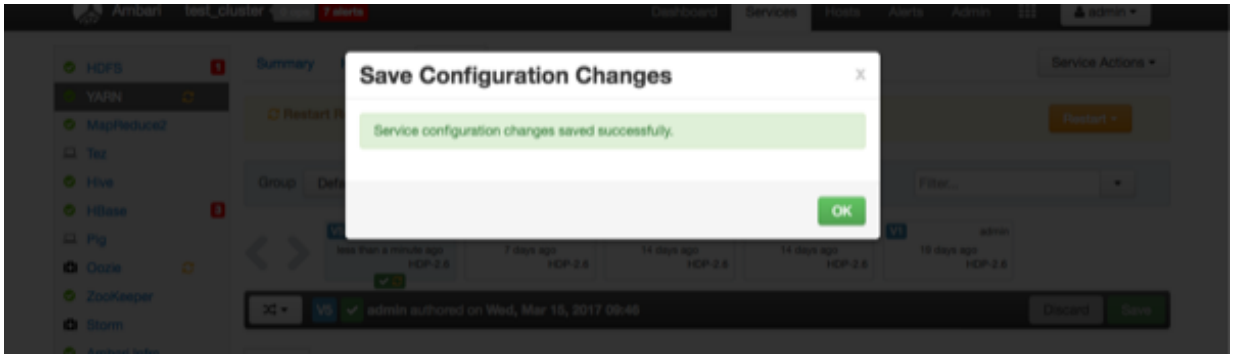
4. Click **Save** to save the new configuration settings. On the **Save Configuration** pop-up, type a description of the changes in the **Notes** box, then click **Save**.



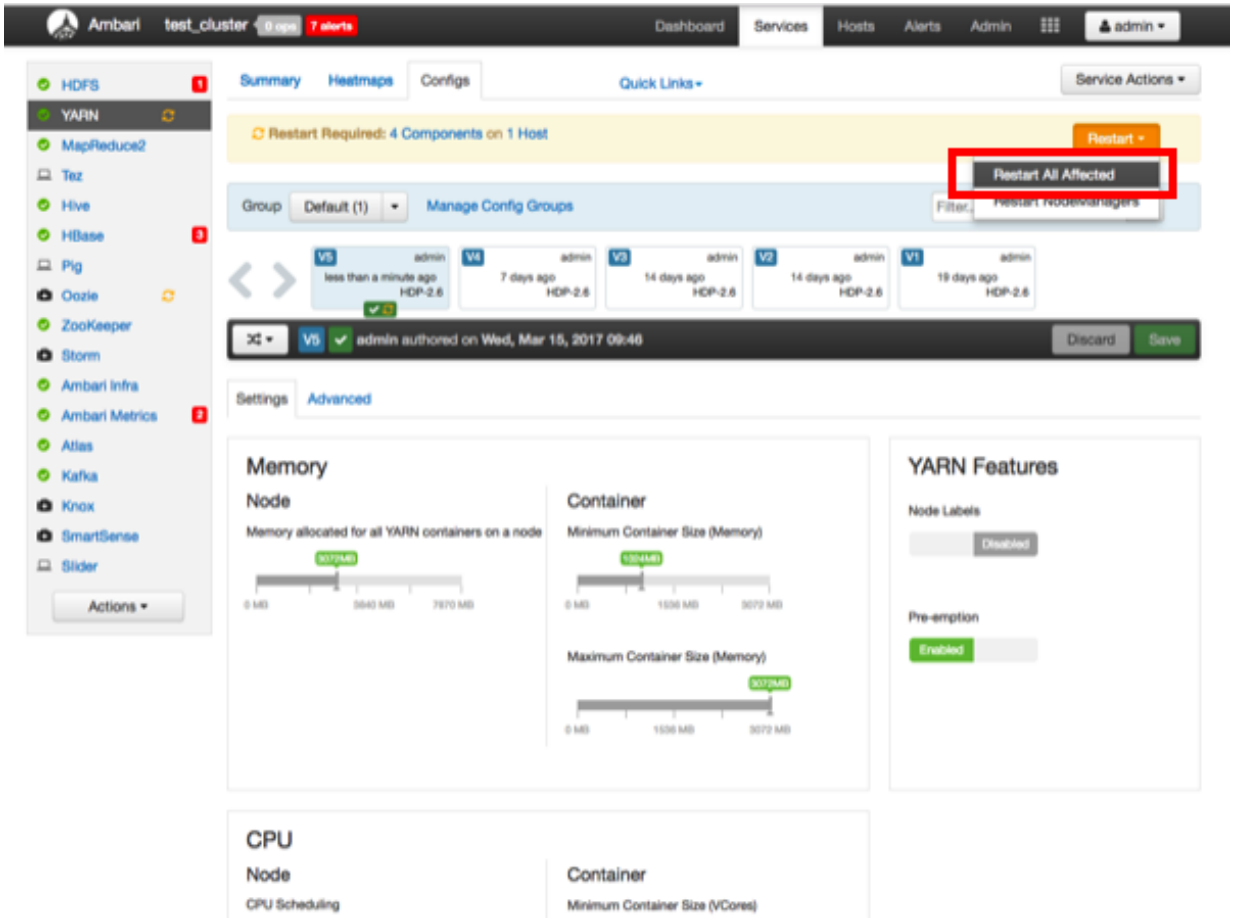
5. On the **Dependent Configurations** pop-up, click **OK** to accept the recommended value of true for the yarn.scheduler.capacity.ordering-policy.priority-utilization.underutilized-preemption.enabled property.



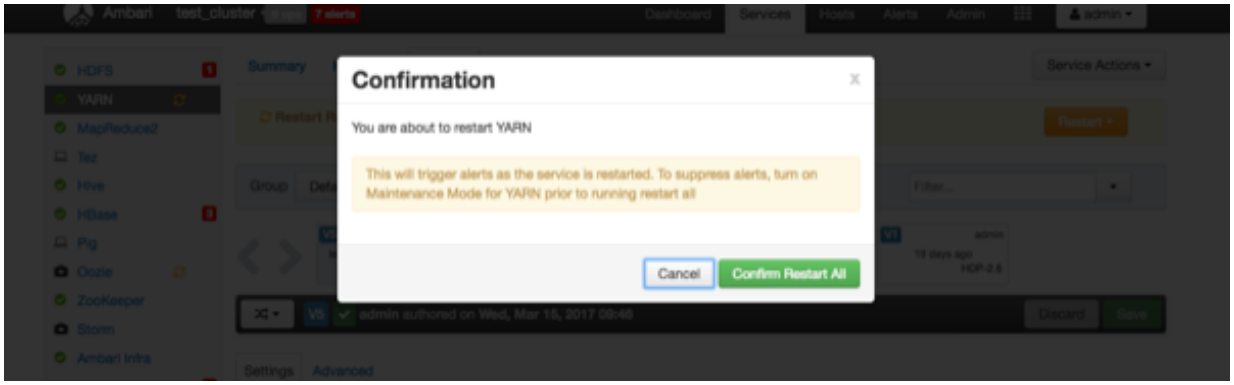
6. Click **OK** on the **Save Configuration Changes** pop-up.



7. Select **Restart** > **Restart All Affected** to restart the YARN service and load the new configuration.



8. Click **Confirm Restart All** on the confirmation pop-up to confirm the YARN restart.



Results

After YARN restarts, Preemption will be enabled.

What to do next

Other components may also require a restart.

For more information about Preemption, see Better SLAS Via Resource Preemption in the YARN Capacity Scheduler.

Set YARN queue priorities

To ensure that applications can access cluster resources.

Before you begin

In order for YARN Queue Priorities to be applied, you must enable preemption.

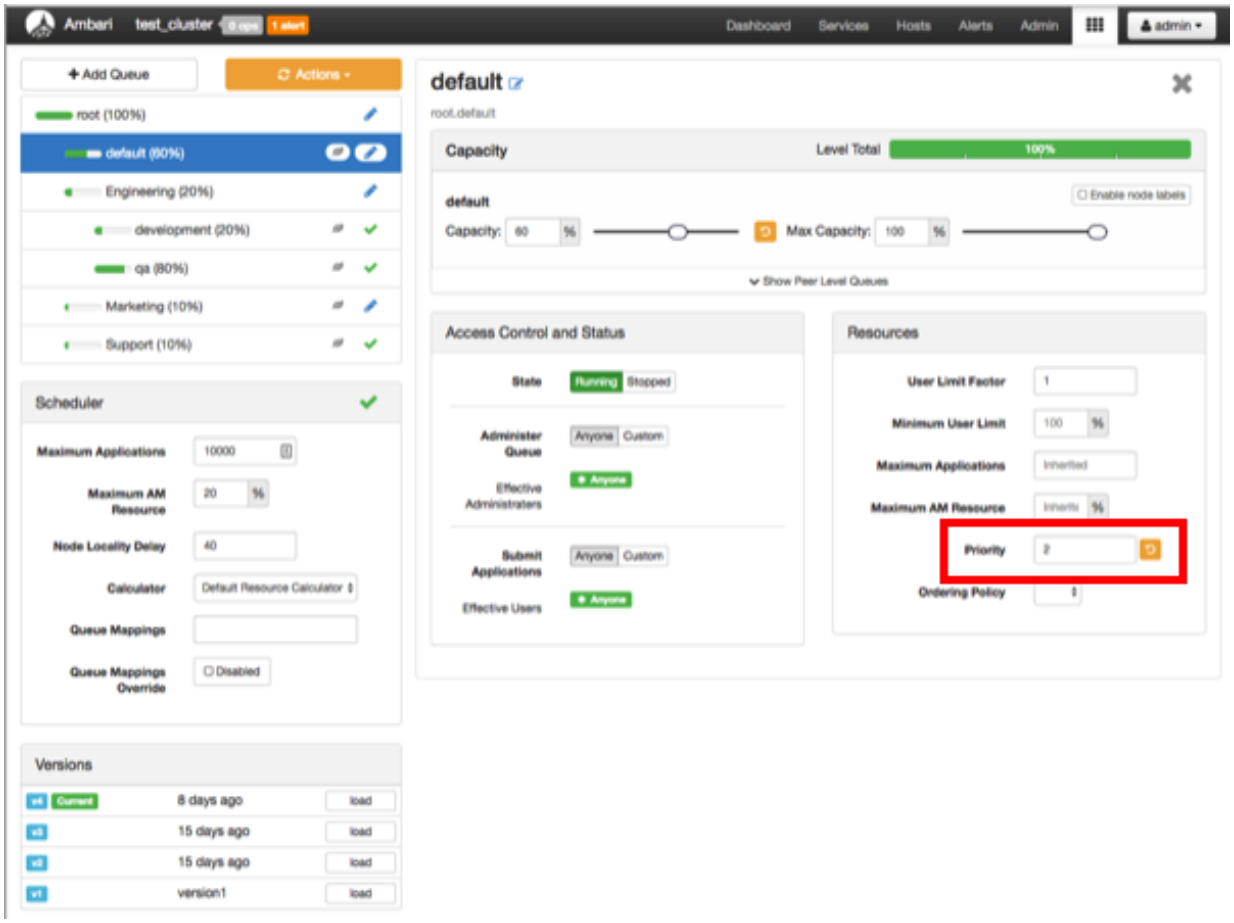
About this task

Even with preemption enabled, there are some use cases where applications might not have access to cluster resources without setting priorities.

Long-running applications	Without setting priorities, long-running applications in queues that are under capacity and with lower relative resource usage may not release cluster resources until they finish running.
Applications that require large containers	The issue with long-running applications is exacerbated for applications that require large containers. With short-running applications, previous containers may eventually finish running and free cluster resources for applications with large containers. But with long-running services in the cluster, the large containers may never get sufficiently large resources on any nodes.
Hive LLAP	Hive LLAP (Low-Latency Analytical Processing) enables you to run Hive queries with low-latency in near real-time. To ensure low-latency, you should set the priority of the queue used for LLAP to a higher priority, especially if your cluster includes long-running applications.

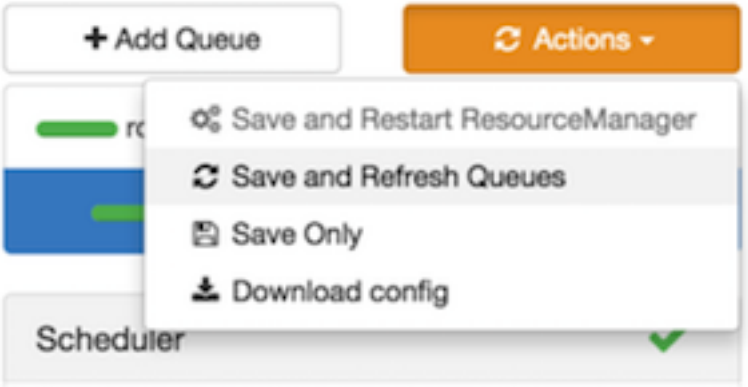
Procedure

1. On the **YARN Queue Manager** view, select a queue, then enter a priority in the **Priority** box under **Resources**.



All queues are set to a priority of 0 by default. Higher numbers indicate higher priority.

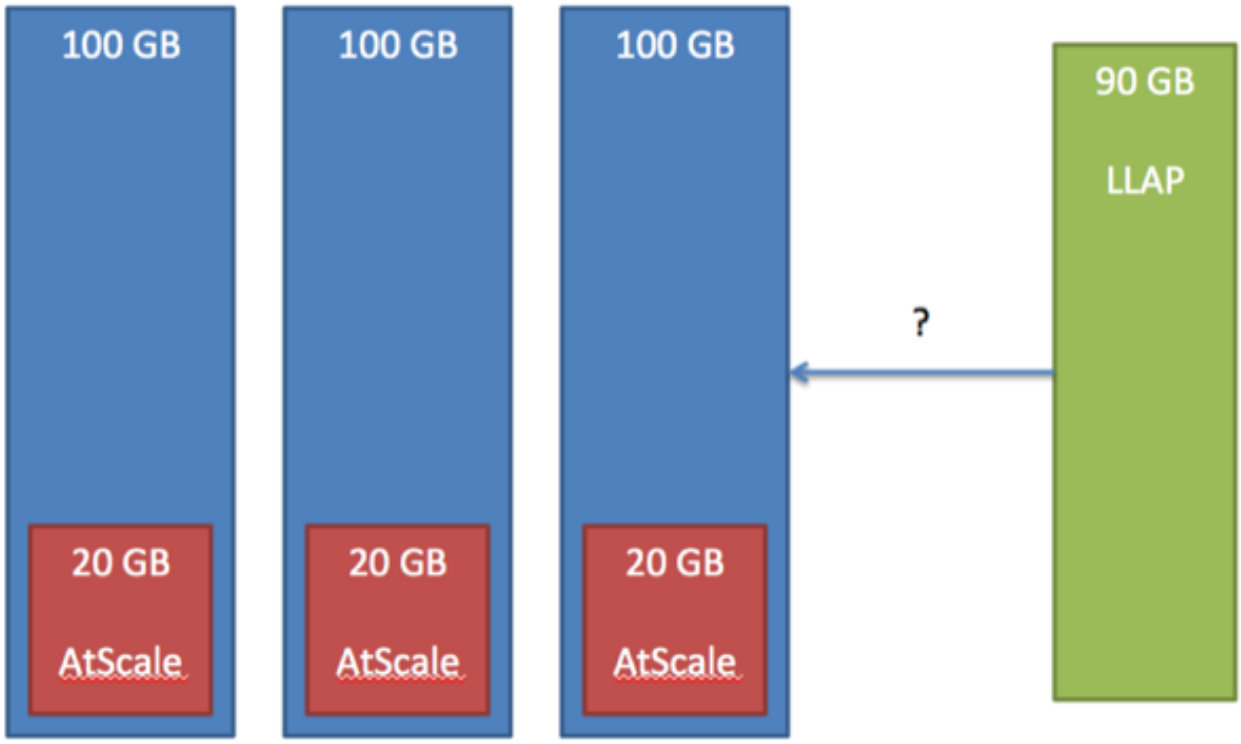
2. Select **Actions** > **Save and Refresh Queues** to save the priority setting.



Example

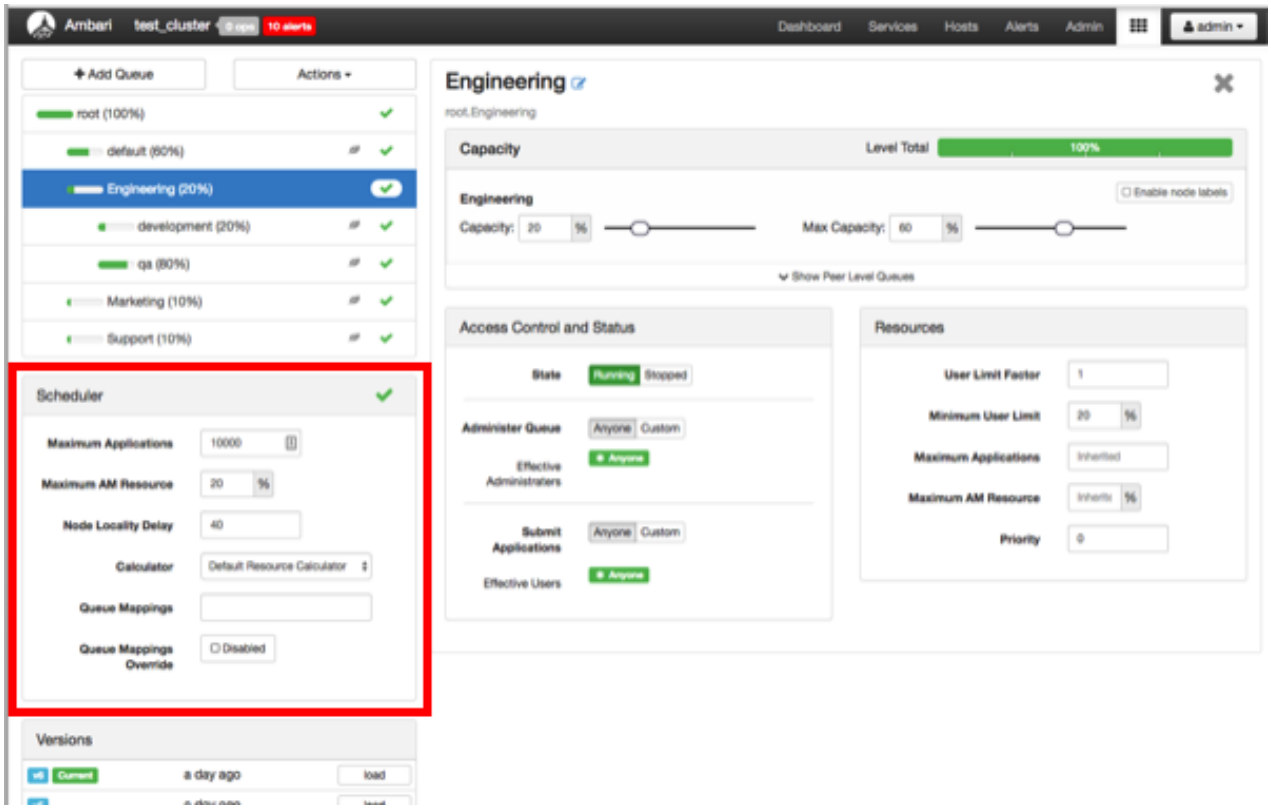
To set the queue used for Hive LLAP, in **Ambari Web**, browse to **Hive** > **Config** > **Settings**, then select a queue using the **Interactive Query Queue** drop-down. For example, the following figure shows a 3-node cluster with long-running 20 GB containers. The LLAP daemons require 90 GB of cluster resources, but preemption does not occur because the available queues are under capacity with lower relative resource usage. With only 80 GB

available on any of the nodes, LLAP must wait for the long-running applications to finish before it can access cluster resources.



Configure capacity scheduler settings for a cluster

You can use the Scheduler box to set global capacity scheduler settings that apply to all queues.



The following Scheduler global parameters are available:

- Maximum Applications – To avoid system-thrash due to an unmanageable load -- caused either by malicious users, or accidentally -- the Capacity Scheduler enables you to place a static, configurable limit on the total number of concurrently active (both running and pending) applications at any one time. This property is used to set this limit, with a default value of 10,000.
- Maximum AM Resource – The limit for running applications in any specific queue is a fraction of this total limit, proportional to its capacity. This is a hard limit, which means that once this limit is reached for a queue, any new applications submitted to that queue will be rejected, and clients will have to wait and retry later.
- Node Locality Delay – The number of missed scheduling cycles after which the scheduler attempts to schedule rack-local containers.
- Calculator – The method by which the scheduler calculates resource capacity across resource types.
- Queue Mappings – You can use this box to specify default queue mapping based on user or group.
- Queue Mappings Override – Select this box to enable override of default queue mappings.

Related Information

Define Queue Mapping Policies

Apply capacity scheduler queue hierarchy changes

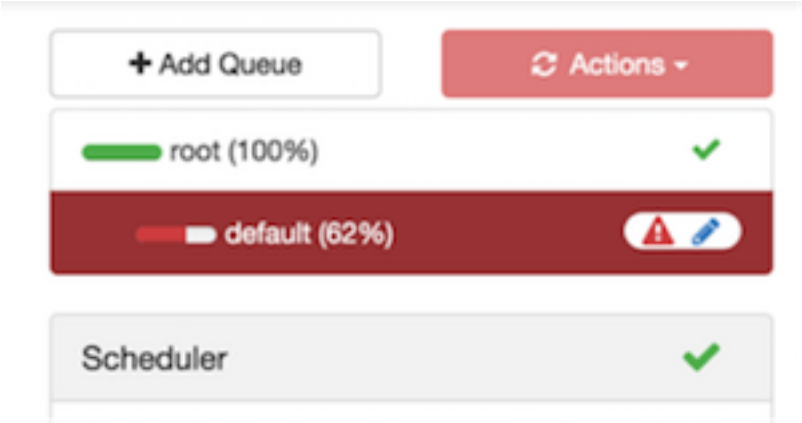
Using the Actions menu, to ensure that all components pick up applied configuration changes.

About this task

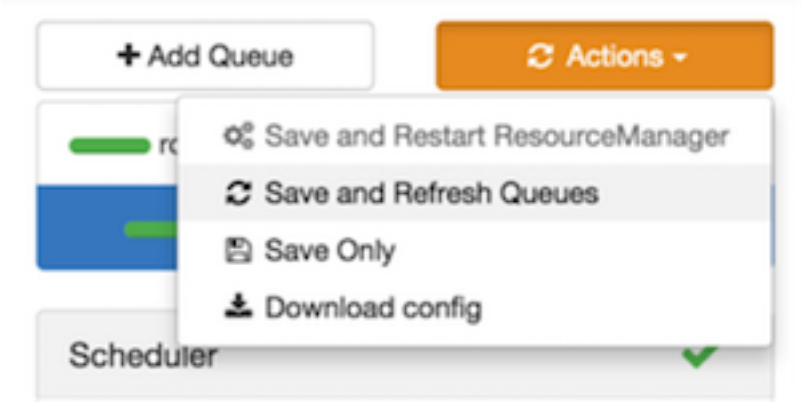
You can use the Actions menu to apply configuration changes made to the queue hierarchy. Depending on the configuration changes made, the Actions menu will guide you to the options available to apply the changes.

Procedure

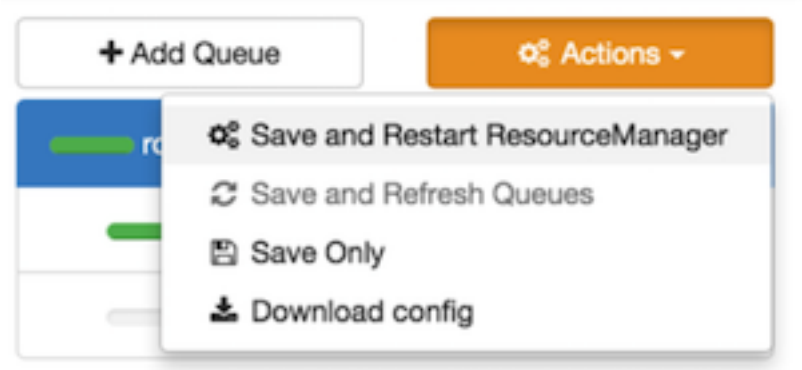
- For changes that are not valid and cannot be applied, the **Actions** button will turn red, and the menu will not appear.



- For configuration changes that can be applied dynamically (without restarting the YARN ResourceManager), the Actions Menu will guide you to **Save and Refresh Queues**.



- For configuration changes that require a restart of the YARN ResourceManager, the Actions Menu will guide you to **Save and Restart ResourceManager**.



Using Files View

Ambari files view lets you to upload, move, copy, and modify permissions to the files or folders in your HDFS file system.

Understanding Files View

You can access files and folders in a cluster's file system using the Ambari web user interface.

Ambari automatically creates one Files View, pointed to the HDFS file system, during cluster deployment. You can also create a custom view from the Files View instance that points to an Amazon Web Services (AWS) S3 file system bucket.

To create a custom Files View to work with files in S3, provide the bucket URL, and valid AWS credentials in the **Settings** > **View Configs** field when you create the view instance that accesses an S3 bucket. You must enter the AWS credentials as a single string containing no space characters using the following syntax:

fs.defaultFS=s3a://[BUCKET_URL]/; fs.s3a.access.key=[VALID_ACCESS_KEY_STRING]; fs.s3a.secret.key[VALID_SECRET_KEY_STRING]

Related Information

About Access Keys
Move files or folders within your file system
Copy files or folders within your file system
Upload files from a local system
Modify permissions of files and folders

Move files or folders within your file system

You can move files or folders within your HDFS file system.

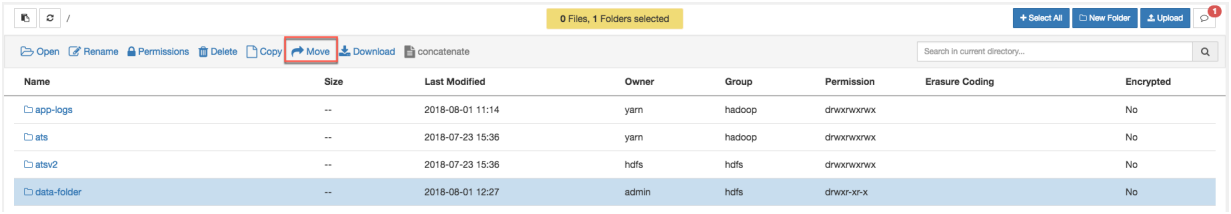
Before you begin

Ensure that you have the necessary folder permissions to move files from a source to a target directory within your HDFS file system.

Procedure

1. In the Ambari dashboard, browse to **Views** > **Files View**.

2. Select a file or a folder that you want to move, and then click



Move.
The **Move to** window appears.

3. Select a destination folder, and then click **Move**.

Copy files or folders within your file system

You can copy files or folders to a different location in your HDFS file system.

Before you begin

Ensure that you have the necessary permissions to copy files from a source to a target directory in your HDFS file system.

Procedure

1. In the Ambari dashboard, browse to **Views** > **Files View**.
2. Select a file or a folder that you want to move, and then click

Copy.
The **Copy to** window appears.

3. Select a destination folder, and then click **Copy**.

Upload files from a local system

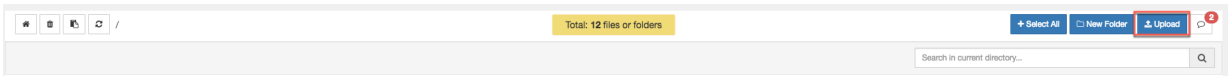
You can upload a file from your local file system to your HDFS file system.

Before you begin

Ensure that you have the necessary permissions to upload files to your HDFS file system.

Procedure

1. In the Ambari dashboard, browse to **Views** > **Files View**.
2. Click



Upload.
The **Upload file to** window appears.

3. Drag a file into the **Upload file to** window to upload, or click to browse to the file location.
   You can upload only one file at a time. You cannot upload a directory.

Modify permissions of files and folders

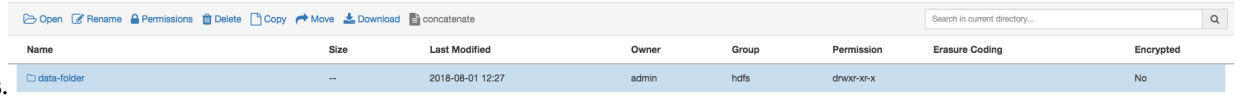
You can modify permissions for specific files and folders in your HDFS file system.

Before you begin

Ensure that you have the necessary permissions to modify file and folder permissions in your HDFS file system.

Procedure

1. In the Ambari dashboard, browse to **Views** > **Files View**.
2. Select a file or a folder for which you want to modify permissions.
3. Click



Permissions.

4. Edit the file or folder permissions and then click **Save**.

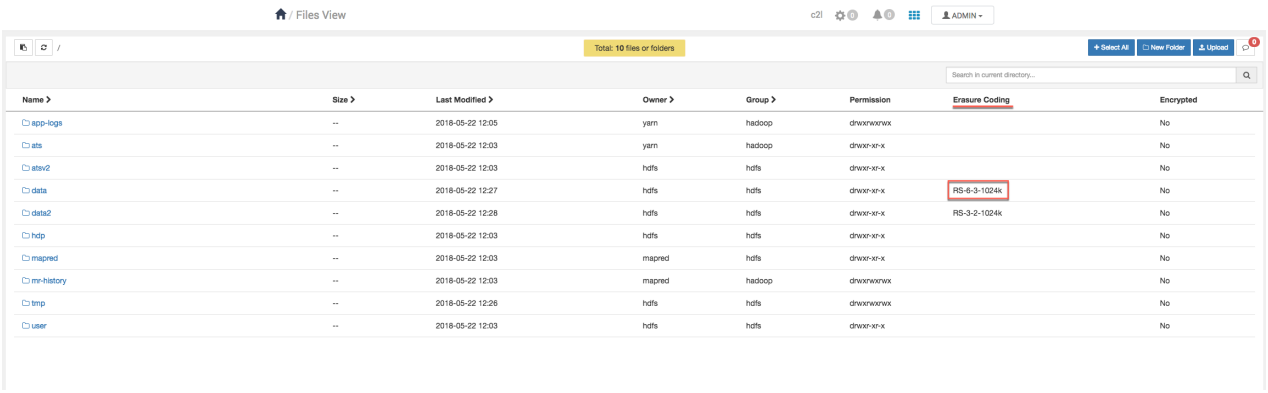
View HDFS erasure coding and encryption status

You can view HDFS erasure coding and encryption status in the Ambari Files View. Erasure coding is used to reduce the amount of storage space required for replication. Encryption is a form of data security required in industries such as healthcare.

Erasure coding

From the Ambari Files view, you can view if an erasure coding policy is set for a specific directory in HDFS. For example, in the following figure, RS-6-3-1024k erasure coding policy is set for the data directory.

Figure 1: HDFS erasure coding



The supported erasure coding policies are:

RS-3-2-1024k	Reed-Solomon code with 3 data blocks, 2 parity blocks, and 1024 KB cell size.
RS-6-3-1024k	Reed-Solomon code with 6 data blocks, 3 parity blocks, and 1024 KB cell size.
RS-LEGACY-6-3-1024k	Reed-Solomon legacy code with 6 data blocks, 3 parity blocks, and 1024 KB cell size.
XOR-2-1-1024k	XOR with 2 data blocks, 1 parity block, and 1024 KB cell size.

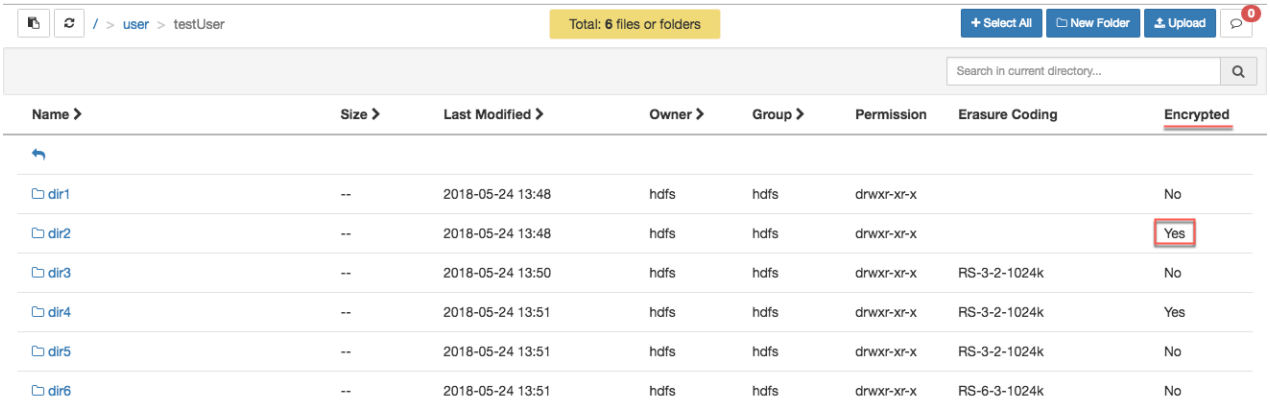
For more information about erasure coding, see the *Data Storage* documentation.

Encryption

From the Ambari Files view, you can view if encryption is enabled for a specific directory in HDFS.

Yes is displayed when encryption is enabled for a directory, and No is displayed when encryption is disabled. For example, in the following figure, encryption is enabled for the dir2 directory.

Figure 2: HDFS encryption



For more information about HDFS encryption, see the *Security* documentation.

Related Information

HDFS Erasure Coding
HDFS Encryption

Using SmartSense View

SmartSense View allows Hortonworks support subscription customers to capture diagnostic data for two purposes:

- To receive recommendations on performance, security, and operational changes based on your server hardware, HDP services deployed, and your use cases.
- To quickly capture diagnostic information about services and hosts when working with support to troubleshoot a support case.

Use the SmartSense View to:

- Capture a bundle for support troubleshooting or for analysis
- Set a bundle capture schedule
- Access SmartSense recommendations

For information on how to perform these tasks, refer to SmartSense documentation.

Related Information

SmartSense Documentation

Using Workflow Manager View

Ambari includes Workflow Manager View, which supports creating, scheduling, and monitoring jobs on a Hadoop cluster. Hadoop administrators can easily design and visualize workflows in the UI as flow graphs.

Workflow Manager is based on the Apache Oozie workflow engine that allows users to connect and automate the execution of big data processing tasks into a defined workflow. Workflow Manager integrates with the Hortonworks Data Platform (HDP) and supports Hadoop jobs for Hive, Sqoop, Pig, MapReduce, Spark, and more. In addition, it can be used to perform Java, Linux shell, distcp, SSH, email, and other operations.

You can access the Workflow Manager documentation on the Hortonworks documentation website.

Related Information

Workflow Management