

yarn-site.xml

YARN configuration options are stored in the `/opt/mapr/hadoop/hadoop-2.x.x/etc/hadoop/yarn-site.xml` file and are editable by the `root` user. This file contains configuration information that overrides the default values for YARN parameters. Overrides of the default values for core configuration properties are stored in the [Default YARN Parameters](#) file.

To override a default value for a property, specify the new value within the `<configuration>` tags, using the following format:

```
<property>
  <name> </name>
  <value> </value>
  <description> </description>
</property>
```

The following configuration tables describe the possible entries you can place between the `<name>` tags and between the `<value>` tags. The `<description>` tag is optional but recommended for maintainability.

Configurations for Resource Manager

Parameter	Description
<code>yarn.resourcemanager.hostname</code>	The hostname of the ResourceManager. The <code>configure.sh</code> command automatically sets this value to the IP address that you provide with the <code>-RM</code> option. Default value: {IP Address}
<code>yarn.resourcemanager.scheduler.address</code>	<i>hostname</i> is the hostname of the ResourceManager and port 8030 is the port on which the Applications in the cluster talk to the ResourceManager. Example value: <code>\${yarn.resourcemanager.hostname}:8030</code>
<code>yarn.resourcemanager.resource-tracker.address</code>	Example value: <code>\${yarn.resourcemanager.hostname}:8031</code>
<code>yarn.resourcemanager.address</code>	The hostname of the ResourceManager and the port on which the clients can talk to the Resource Manager. Example value: <code>\${yarn.resourcemanager.hostname}:8032</code>

Configurations for NodeManager

Parameter	Description
<code>yarn.nodemanager.container-executor.class</code>	Identifies who will launch the containers. Set to <code>LinuxContainerExecutor</code> so that jobs can run as the user that submits the job. Note: If a system user (a user with <code>userID<500</code>) wants to submit a job, you must add the user in the <code>container-executor.cfg</code> file. The user <code>mapr</code> is already configured as an allowed system user. Default value: <code>org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor</code>

yarn.nodemanager.aux-services	<p>Selects a shuffle service that needs to be set for MapReduce to run.</p> <p>This property, in conjunction with other properties, sets "direct shuffle" as the default shuffle for MapReduce.</p> <p>Default value: <code>mapreduce_shuffle, mapr_direct_shuffle</code></p>
yarn.nodemanager.aux-services.mapreduce_shuffle.class	<p>This property, in conjunction with other properties, sets "direct shuffle" as the default shuffle for MapReduce.</p> <p>Default value: <code>org.apache.hadoop.mapred.ShuffleHandler</code></p>
yarn.nodemanager.aux-services.mapr_direct_shuffle.class	<p>This property, in conjunction with other properties, sets "direct shuffle" as the default shuffle for MapReduce.</p> <p>Default value: <code>com.mapr.hadoop.mapred.LocalVolumeAuxService</code></p>

Configurations for MapReduce

Parameter	Description
mapreduce.job.shuffle.provider.services	<p>This property, in conjunction with other properties, sets "direct shuffle" as the default shuffle for MapReduce.</p> <p>Default value: <code>mapr_direct_shuffle</code></p>

Configurations for Container Logs

Parameter	Description
yarn.nodemanager.log-dirs	<p>The location to store container logs on the node. An application's log directory is <code>\${yarn.nodemanager.log-dirs}/application_\${appid}</code>. Individual containers' log directories are named <code>container_\${contid}</code>. Each container directory will contain the files <code>stderr</code>, <code>stdin</code>, and <code>syslog</code> generated by that container.</p> <p>By default the log files are located in the following directory: <code>/opt/mapr/hadoop/hadoop-<version>/logs/userlogs/<applicationID>/<containerID>/<filename</code></p> <p>Note: You can find the application ID associated with your job in the MCS.</p>
yarn.log-aggregation-enable	Determines whether the logs are aggregated. Defaults to false.
yarn.nodemanager.log.retain-seconds	When log aggregation is disabled, this value determines the duration that user logs are maintained. The default is 10800 or 3 hours.
yarn.log-aggregation.retain-seconds	When log aggregation is enabled, this property determines the number of seconds to retain. The default, -1, disables the deletion of logs.
yarn.log-aggregation.retain-check-interval-seconds	The interval between aggregated log retention checks. If set to 0 or a negative value then the value is computed as one-tenth of the aggregated log retention time. Note: Setting this to a value may cause unnecessary log retention checks.
yarn.nodemanager.remote-app-log-dir	The location on the filesystem where the logs are aggregated. The default is <code>/tmp/logs</code> .
yarn.nodemanager.remote-app-log-dir-suffix	The suffix for the directory that will contain the aggregated logs for each user. The default is