YARN Log Aggregation

The YARN Log Aggregation option aggregates and moves log files for completed applications from the local file system to the MapR-FS. This allows users to view the entire set of logs for a particular application using the HistoryServer UI or by running the `yarn logs` command.

By default, YARN container logs are not aggregated on the MapR-FS. Instead, the logs are retained for 3 hours on the local file system before they are deleted. To enable YARN log aggregation or to edit the configuration of YARN log aggregation, you must edit the yarn-site.xml file in the following directory: `/opt/mapr/hadoop/hadoop-2.x.x/etc/hadoop/`

This section contains information about how to complete the following tasks:
- Enabling YARN Log Aggregation
- Viewing Logs for Completed Applications
  - Using the Command Line to View Logs for Completed Applications:
  - Using the HistoryServer UI to View Logs for Completed Applications:
- Editing the Retention Settings of Aggregated Logs

Enabling YARN Log Aggregation

To enable YARN log aggregation, add or edit the following properties in `yarn-site.xml`:

- Set the value of `yarn.log-aggregation-enable` to `true`.
- Optional: Set the value of `yarn.nodemanager.remote-app-log-dir` to a location in the MapR-FS. By default, the location is `maprfs:///tmp/logs`.
- Optional: Set the value of `yarn.nodemanager.remote-app-log-dir-suffix` to the name of the folder that should contain the logs for each user. By default, the folder name is `logs`.

On a non-secure cluster, you also need to add the following property to `/opt/mapr/hadoop/hadoop-2.x/etc/hadoop/yarn-env.sh` on the Node Manager nodes:

```
export MAPR_IMPERSONATION_ENABLED=1
```

Then restart the Node Manager services. This setting enables impersonation for Node Manager processes so that log files can be created with the correct user ownership.

Aggregated logs are owned by the user who runs the job. For example, when user `admin` runs a job, the logs are stored to `maprfs:///tmp/logs/admin`. If user `analyst` runs a job, the logs are stored to `maprfs:///tmp/logs/analyst`. If these two users do not share the same UNIX group, they will not be able to see each other's logs.

> ⓘ If centralized logging and YARN log aggregation are enabled, the logs for MapReduce v2 applications will be managed by Centralized Logging while the logs for non-MapReduce applications will be managed by YARN log aggregation.

Viewing Logs for Completed Applications

With YARN log aggregation, you can use yarn commands or the HistoryServer UI to access logs for completed applications.

Using the Command Line to View Logs for Completed Applications:

1. Determine the application ID for the application that you want to view the logs for.
   For example, run the following command to list the applications:

   ```
   yarn application -list
   ```

2. Run the `yarn logs` command to view the logs for the application.
   For example, run the following command to view the log files for application `application_1415822090718`

   ```
   yarn logs -applicationId application_1415822090718
   ```

Using the HistoryServer UI to View Logs for Completed Applications:

1. Log on to the MapR Control System.

2. In the Navigation Pane, click **JobHistoryServer**
3. Click the **Job ID** link for the job that you want to view the logs for.
4. In the Logs column of the Application Master section, click the **logs** link.

Editing the Retention Settings of Aggregated Logs

By default, aggregate logs are stored on the MapR-FS for 30 days. The retention time for aggregated logs also applies to centralized logs.

To edit the retention settings, add or edit the following properties in yarn-site.xml:

1. Set the value of `yarn.log-aggregation.retain-seconds` to set the duration that the logs are maintained. If you set a negative value for `yarn.log-aggregation.retain-seconds`, logs will not be deleted.

   > ⓘ The duration specified by `yarn.log-aggregation.retain-second`s starts from the time that the application starts running. Therefore, when you configure the duration, consider how long you want the log to remain in addition to the amount of time that the application will take to run. For example, if you expect most applications to take 20 seconds to run, do not set the value of this property to 20 seconds because the log might be deleted as soon as the applications completes.

2. Optionally, set the `yarn.log-aggregation.retain-check-interval-seconds to` specify how often the log retention check should be run. By default, it is one-tenth of the log retention time.

For more details about the properties that impact the YARN container logs and the aggregation option, see yarn-site.xml.