

# Reporting/Analytics

## Why is my spark work directory filling up the hard drive?

There are circumstances that can cause the /opt/spark-2.1.2/work directory to contain so much data that it fills the hard drive. This happens because every time spark is restarted it creates two large folders of around 100 megabytes each. One will be prefixed with "driver-" and the other "app-"

Unfortunately spark does not clean up the folders from the last restart. There are error conditions that can cause spark to restart every few minutes. When these errors happen spark will restart many times through the day creating more copies of the app- and driver- folders. Over time this can fill the hard drive.

To see these you can run the following command:

```
=====
[root]# ls /opt/spark-2.1.2/work
driver-20191205124436-0000
app-20191205124441-0000
=====
```

To see how much disk space the spark work directory is using run you can use the "du" command. For example:

```
=====
[root]# du --max-depth 1 -h /opt/spark-2.1.2 | grep work
760G /opt/spark-2.1.2/work
=====
```

To remedy the full hard drive you can stop the spark worker and master services, remove all folders in the spark work directory, and restart spark as shown below.

On the spark master host

```
=====
[root]# systemctl stop spark-master
=====
```

On each of the spark worker hosts

```
=====
[root]# systemctl stop spark-worker
=====
```

Remove all folders in the spark work directory that start with "app-" or "driver-".

```
=====
[root]# cd /opt/spark-2.1.2/work/
[root]# rm -rf app-* driver-*
=====
```

# Reporting/Analytics

On the spark master host

```
=====
[root]# systemctl start spark-master
=====
```

On each of the spark worker hosts

```
=====
[root]# systemctl start spark-worker
=====
```

On the host where Reporting Web Services you will need to restart tomcat so that the reporting application is redeployed to spark.

```
=====
[root]# systemctl restart tomcat
=====
```

Although this should remedy the full hard drive problem, you should then try to determine the underlying cause for the frequent spark restarts. Otherwise the hard drive will likely fill up again.

To do this look for errors in the log files stored in /opt/spark-2.1.2/logs and /var/log/spark/. One common reasons for spark restarts is `kafka.common.OffsetOutOfRangeException`. How to fix this issue is described in <http://docs.adaptivecomputing.com/9-1-3/installGuide/RH7/installRH7.htm#topics/hpcSuiteInstall/Troubleshooting/reportingIssues.htm#outofrange>

Unique solution ID: #1224

Author: Nate Seeley

Last update: 2020-03-09 20:09