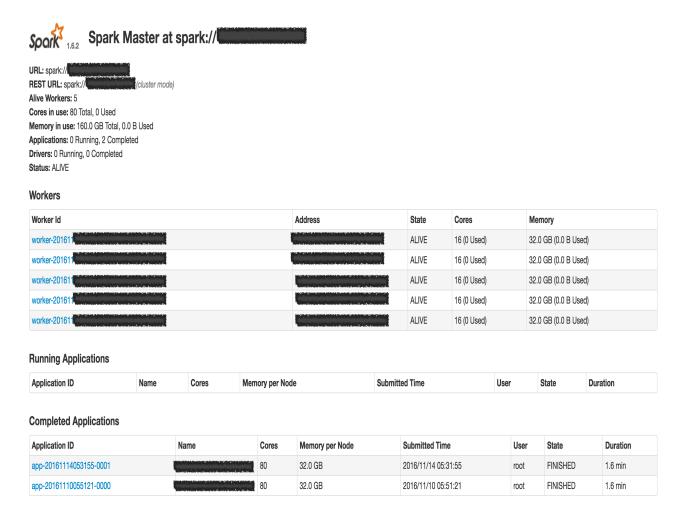
## Robin Jha 1.125 Deliverable #2

For this project I had to gather the data myself for which I used ~50 node Mesos cluster. On the cluster I have agents running to collect the current disk used percentage, which is then written to Elasticsearch as shown in the screenshot below (extra fields have been covered for security reasons):



Once the agents started pushing data, I wrote a module in Java to get this data from Elasticsearch and do some processing on it. I then compiled this module into a jar and submitted this as a job to the Spark cluster that was setup as a part of deliverable #1. Here is the UI for Spark Master (some fields have been covered for security reasons):



To submit the Spark job, I logged into the container for master and ran the command below:

./spark-submit --class com.ml.pie.run.PieExecutor /mnt/mesos/sandbox/pie-1.0-SNAPSHOT-jar-with-dependencies.jar DiskUsagePredictor

Jobs currently in progress is listed under "Running Applications" on the Spark UI.

Currently the output for my Spark job looks as follows. I am able to get the IP and disk usage on the host. The next step is to use Machine Learning techniques on this dataset:

```
Running PIE Job for predicting : Disk Usage
2016-11-17 18:04:48 INFO DiskUsageStats:183 - processing Host IP: 100.200.100.50 Disk Name : /
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/boot
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr/share/oem
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/boot
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr/share/oem
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr
2016-11-17 18:04:48 INFO DiskUsageStats:183 - processing Host IP: 100.200.101.50 Disk Name : /
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/boot
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr/share/oem
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/boot
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr/share/oem
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr
2016-11-17 18:04:49 INFO DiskUsageStats:183 - processing Host IP: 100.200.102.50 Disk Name : /
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/boot
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr/share/oem
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/boot
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr/share/oem
2016-11-17 18:04:49 INFO ESHistoryProcessor:116 - Disk Name is :/usr
2016-11-17 18:04:50 INFO DiskUsageStats:183 - processing Host IP: 100.200.103.50 Disk Name : /
2016-11-17 18:04:50 INFO ESHistoryProcessor:116 - Disk Name is :/
2016-11-17 18:04:50 INFO ESHistoryProcessor:116 - Disk Name is :/boot
2016-11-17 18:04:50 INFO ESHistoryProcessor:116 - Disk Name is :/usr/share/oem
2016-11-17 18:04:50 INFO ESHistoryProcessor:116 - Disk Name is :/usr
2016-11-17 18:04:50 INFO ESHistoryProcessor:116 - Disk Name is :/boot
2016-11-17 18:04:50 INFO ESHistoryProcessor:116 - Disk Name is :/usr/share/oem
2016-11-17 18:04:50 INFO ESHistoryProcessor:116 - Disk Name is :/usr
```