Suigeth Shetty 07/18/2020

BUAN 6346 Big Data

Homework 13

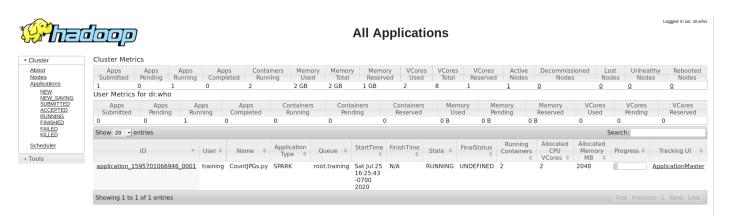
Write and run a Spark application

Spark Application in Python:

```
[training@localhost ~]$ spark-submit CountJPGs.py /loudacre/weblogs/*
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/zookeeper/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/flume-ng/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
20/07/25 16:21:05 INFO spark.SparkContext: Running Spark version 1.3.0
20/07/25 16:21:52 INFO scheduler.TaskSetManager: Finished task 310.0 in stage 0.0 (TID 310) in 103 ms on localhost (311/311)
20/07/25 16:21:52 INFO scheduler.DAGScheduler: Stage 0 (count at /home/training/CountJPGs.py:15) finished in 32.098 s
20/07/25 16:21:52 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
20/07/25 16:21:52 INFO scheduler.DAGScheduler: Job 0 finished: count at /home/training/CountJPGs.py:15, took 32.298435 s
Number of JPG requests: 64978
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/metrics/json,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/stage/kill,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/static,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors/threadDump/json,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors/threadDump,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors/json,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/environment/json,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/environment,null}
```

Submit a Spark Application to Cluster:

[training@localhost ~]\$ spark-submit \
> --master yarn-client \
> CountJPGs.py /loudacre/weblogs/*
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/zookeeper/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/flume-ng/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]



Sujeeth Shetty 07/18/2020

