

coreservlets.com – Hadoop Course

MapReduce Overview and Install

In this exercise, you will have a chance to become familiar with YARN installation as well as to execute a MapReduce job and monitor its status.

Approx. Time: 45 minutes

Answer

1. What is the location of YARN configuration files?
2. Where is the data persisted by Node Managers? What property determines the location? What file specifies the property?
3. Where does YARN persist its logs? What property determines the location? What file specifies the property?

Perform

1. Start YARN and MapReduce History Server
2. Look at YARN Resource Manager Web based UI. How many active Nodes does it show? What is the status of each node?
3. Execute `pi` sample job. Monitor the job on Resource Manager UI. How many mappers did the job execute?

Answer Solution

1. YARN configuration files location is specified by the environment variable `$YARN_CONF_DIR`
2. In our installation the Node Manager persists its data to `/home/hadoop/Training/hadoop_work/mapred/nodemanager`; the location is specified by `yarn.nodemanager.local-dirs` property in `$YARN_CONF_DIR/yarn-site.xml`
3. In our installation YARN persist its logs to `/home/hadoop/Training/logs/yarn`; the location is specified by `YARN_LOG_DIR` variable in `$YARN_CONF_DIR/yarn-env.sh`
`export YARN_LOG_DIR=/home/hadoop/Training/logs/yarn`

Perform Solution

1. Use command line to start YARN and MapReduce history server:

```
$ cd $HADOOP_HOME/sbin
$ ./start-yarn.sh
$ ./mr-jobhistory-daemon.sh start historyserver
```

2. You can find Resource Manager at `http://localhost:8088`; It should show 1 active node; click on the '1' link and you will see a list of all the active nodes. There is a column labeled 'Health-status' which should have a value of 'Healthy'

3. You can execute this job on the command line:

```
$ yarn jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.0.0-cdh4.0.0.jar pi 5 5
```

In this case there will be 5 mappers but the number of mappers is a parameter. You can also find number of Map tasks by browsing Resource Manager Management UI.