Starting Hadoop Cluster at boot time

The following describes how to edit/setup the attached example Hadoop startup script in CentOS 5.6. When this script is run from the master Hadoop node, all other Hadoop services start on the datanodes. This script is used in the data trawlers. These should work similarly on other Linux distributions.

An example of shell script code is below.

Prerequisites

Hadoop should be installed and running successfully from the hadoop master node. (i.e. <path to hadoop>/bin/start-all.sh)

Hadoop

1. Create the script on the Linux machine (in my example the script is named "*hadoop-master*").
2. Move the `das.server` script to `/etc/init.d/`
     1. These scripts should be owned by root.
3. Edit the `hadoop-master` script.
     1. Change the value of `HPATH` to the correct path.
          1. *export HPATH=/opt/hadoop*
     2. Change `HLOCK` to a location where the hadoop pid file can be stored.
          1. *export HLOCK=/var/lock/subsys*
4. Add das server to the list of available services in Linux.
     1. *chkconfig --add hadoop-master*
5. Check what init levels are active for `hadoop-master`
     1. *chkconfig --list |grep hadoop* (Note init 3,4,5 have hadoop-master "on")

       hadoop-master   0:off   1:off   2:off   3:on   4:on   5:on   6:off

This automates the start of Hadoop at bootup.

```bash
#!/bin/bash
#
#
# Starts a Hadoop Master
#
# chkconfig: 2345 90 10
# description: Hadoop master

. /etc/rc.d/init.d/functions
. /opt/hadoop/conf/hadoop-env.sh
export HPATH=/opt/hadoop
export HLOCK=/var/lock/subsys

RETVAL=0
PIDFILE=$HLOCK/hadoop-hdfs-master.pid
desc="Hadoop Master daemon"

start() {
  echo -n $"Starting $desc (hadoop): "
  daemon --user root $HPATH/bin/start-all.sh $1
  RETVAL=$?
  echo
  [ $RETVAL -eq 0 ] && touch $HLOCK/hadoop-master
  return $RETVAL
}
```

```bash
stop() {
   echo -n $"Stopping $desc (hadoop): "
   daemon --user root $HPATH/bin/stop-all.sh
   RETVAL=$?
   sleep 5
   echo
   [ $RETVAL -eq 0 ] && rm -f $HLOCK/hadoop-master $PIDFILE
}

checkstatus(){
   jps |grep NameNode
}

restart() {
   stop
   start
}

format() {
   daemon --user root $HPATH/bin/hadoop master -format
}

case "$1" in
   start)
      start
      ;;
   upgrade)
      upgrade
      ;;
   format)
      format
      ;;
   stop)
      stop
      ;;
   status)
      checkstatus
      ;;
   restart)
      restart
      ;;
   *)
      echo $"Usage: $0 {start|stop|status|restart|try-restart}"
      exit 1
esac
```

```
exit $RETVAL
```