Simple batchtools example with SGE

Approximating the value of π

This is a description of how to run the piApprox example on a CfnCluster using SGE scheduler and batchtools. The piApprox example is described in `vignettes/batchtools.Rmd` in the batchtools repository.

Cfncluster can run the SGE schedule on an Ubuntu 16.04 system, but it can't run slurm. The reason is outlined in the note at the end of this document.

I. Configuration file:

```
[aws]
aws_access_key_id = ### Your AWS Access Key ###
aws_secret_access_key = ### Your AWS Secret Access Key ###
aws_region_name = us-east-1

[cluster tiny]
vpc_settings = public
key_name = ### a private key from a pair AWS generated ###
base_os = ubuntu1604
initial_queue_size = 2
max_queue_size = 4
maintain_initial_size = false
master_instance_type = t2.micro
compute_instance_type = t2.micro
scheduler = sge

[vpc public]
vpc_id = ### Your default VPC ###
master_subnet_id = ### Any subnet in your default VPC ###

[global]
sanity_check = true
update_check = true
cluster_template = tiny
```

The default location for the configuration file is `~/.cfncluster/config`, or the `--config` option can be used to indicate another file location. In cfncluster 1.3.2, slurm must be run on a CentOS6 system. (see note on the systemd bug.) To launch the cluster, execute:

```
$ cfncluster create tiny
```

Creation of the cluster usually takes about ten minutes. To watch the creation process, open the AWS CloudFormation console, click on the stack name and open the Events and/or Resources sections. When the cluster is created, the instances will be visible in the EC2 console.

II. Files for the pi Approximation Example

The `/home` and `/shared` directories are NFS exported on the master node and mounted on the compute nodes. This means either of these is a good place to put a batchtools repository. The simplest place to run the tiny example is in `/home/centos`.

To run the example, we need to upload or create (vim is installed) four files. First, we need a one-line `batchtools.conf.R` file, like this:

```
cluster.functions = makeClusterFunctionsSGE("./simple.tmpl")
```

Then, we need the brew template, `simple.tmpl`, like this:

```
<%
    # make two additional directories in registry
    rdir <- paste(file.dir, "/rexe", sep="")
    if (!dir.exists(rdir))  {
      dir.create(rdir, recursive = TRUE)
    }

    jdir <- paste(file.dir, "/exe", sep="")
    if (!dir.exists(jdir))  {
      dir.create(jdir, recursive = TRUE)
    }

    # make R script for job to run - rscript that runs JobCollection
    rscript = paste(
          "library(batchtools)",
          sprintf("jc = readRDS('%s')", uri),
          sprintf("batchtools:::doJobCollection(jc, output = '%s')", log.file),
        sep=";")

    rf <- fp(file.dir, "rexe", sprintf("%s.R", job.hash))
    cat(rscript,file=rf,sep="\n")
    Sys.chmod(rf, mode = "0777")

    # make script to run job
    jscript <- fp(file.dir, "exe", sprintf("%s.sh", job.hash))
    cat(paste0("R CMD BATCH ", rf),file=jscript,sep="\n")
    Sys.chmod(jscript, mode = "0777")

%>

```

```
#$ -N <%= job.name %>
#$ -j y
#$ -S /bin/bash
#$ -V
#$ -o <%= log.file %>
#$ -cwd

<%= jscript %>
```

We need to set the environment variables `SGE_ROOT` in the file `.Renviron`. The path gets truncated when launcing R also, so we need to put the SGE directories back in. (The R library in `/shared` will be created in the next section.)

```
SGE_ROOT="/opt/sge"
PATH="/opt/sge/bin:/opt/sge/bin/lx-amd64:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/snap/bin"
R_LIBS="/shared/R/Library"
```

Finally, as a convenience, we make an R file, `piApprox.R`, with the function in it.

```
piApprox = function(n)  {
  nums = matrix(runif(2 * n), ncol = 2)    # random point in [0,1] x [0,1]
  d = sqrt(nums[,1]^2 + nums[,2]^2)        # distance from origin to point
  4 * mean(d <= 1)                         # count points in unit circle
}
```

III. Install batchtools package

R-3.4.0 is installed on the cfncluster AMI's but we need to install batchtools, and (important) we want to install it in a place that all nodes can access. The `/shared` directory is a good place for this.

```
$ sudo mkdir -p /shared/R/Library
$ sudo chmod -R 777 /shared/R/Library
$ R
> install.packages("batchtools", repos="http://cran.us.r-project.org")
```

IV. Run the example

At this point, it is a good idea to open another ssh session to the master node. This is so we can watch the jobs go to the queue in one window and type commands at the R prompt in the other. At the R prompt, run this:

```
> library(batchtools)
> source("piApprox.R")
> reg <- makeRegistry("./registry")
> n <- 10
> batchMap(fun = piApprox, n = rep(1e5,n))
> submitJobs()
> waitForJobs()                            # should be very fast
> reduceResults(function(x, y) x + y)/n    # should be about 3.141
```

After executing `submitJobs()` in the R session, execute `qstat` in the other session. This should show the queue and the nodes running the jobs. When the queue is empty, the `waitForJobs()` should return `TRUE`.

V. Autoscaling

The cluster will autoscale to four compute nodes (the maximum we allowed in the config file) if the conditions are right. We just need to put enough jobs in the queue to make the cluster monitor notice the extra work, launch instances, and bring them online before the queue empties.

```
> system("rm -Rf ./registry")
> reg <- makeRegistry("./registry")
> n <- 50                                  # more jobs
> batchMap(fun = piApprox, n = rep(1e6,n))  # bigger jobs
> submitJobs()
> waitForJobs()                            # should be very fast
> reduceResults(function(x, y) x + y)/n     # should be about 3.141
```

VI. Delete the cluster

To delete the cluster execute (on the local machine):

```
$ cfncluster delete tiny
```

Note: Slurm on Ubuntu 16.04

The initialization system (i.e., the utility that launches services) on Ubuntu 16.04 is the newish systemd. Cfncluster has a bug in the configuration files. The configuration files launch services by running the scripts in /etc/init.d. This is how services were launched in the classic sysvinit system. There is a systemd service (systemd-sysv-generator) which is supposed to provide backwards compatability, but apparently it doesn't work. The result is that slurm daemons are not launched correctly on systemd systems, including Ubuntu 16.04 and Centos7.

Cfncluster is aware of the bug and it will be fixed in cfncluster-1.3.3. However, no release date has been determined for cfncluster-1.3.3 and until then cfncluster's advice is either (1) use SGE with Ubuntu or (2) use Centos6 with slurm.

The initialization files are managed by the Chef configuration management system. When a cfncluster is created, the Chef client is installed on the nodes and the client downloads configuration data from the Chef server. The Chef server is an AWS resource. This means users can't fix this bug on their own.

The subject is discussed in Issue 209 at cfncluster github: https://github.com/awslabs/cfncluster/issues/209

References:

CfnCluster documentation: http://cfncluster.readthedocs.io/en/latest/index.html

CfnCluster main GitHub repository: https://github.com/awslabs/cfncluster