

Hive PTest2 Infrastructure

Background

Hive build jobs include [Hive PreCommit Testing](#) and scheduled builds of various branches with various flags. They are run on the Hive PTest2 Infrastructure, which is an EC2 cluster currently sponsored by Cloudera. The Hive PTest2 Infra runs each build sequentially, but splits the testing across all the resources of the cluster to achieve much higher throughput.

Location

The infrastructure master is at `ec2-174-129-184-35.compute-1.amazonaws.com`. Committers can be granted access to this host by request, see [HIVE-4739](#).

Processes

The infrastructure master hosts two processes:

1. Jenkins: <http://ec2-174-129-184-35.compute-1.amazonaws.com/jenkins/>
 - This is the admin for Hive builds. The dashboard shows the state of the build queue, test results of past builds, etc.
 - Builds are submitted either manually in the dashboard UI, automatically by patch upload (see [Hive PreCommit Testing](#)), or by Jenkins schedule (branch builds).
2. Hive PTest2 WebServer
 - Exposes a REST API to take test requests from the Jenkins builds. Each request specifies a profile property-file, which contains a set of properties such as what branch to build and what tests to run.
 - On a test request:
 - Creates EC2 slaves, which are spot-instances.
 - Reads the specified property-file and compiles Hive using it.
 - Distributes the compiled artifacts across the EC2 slaves, makes SSH calls to run the tests remotely in parallel, and gathers the results.
 - If no further test request comes in 30 minutes, the slaves are shutdown.

Jenkins Configuration/Debug

- Jenkins logs: `/var/log/jenkins/jenkins.log`
- Jenkins config: `/etc/sysconfig/jenkins`
- Jenkins home: `/var/lib/jenkins` (job info is located at `/var/lib/jenkins/jobs`)

Hive PTest2 WebServer Configuration/Debug

- Test Logs: <http://ec2-174-129-184-35.compute-1.amazonaws.com/logs/> (redirects to `/data/hive-ptest/logs/`)
- Profile properties files: `/usr/local/hiveptest/etc/public/`
- Webservice logs: `/home/hiveptest/apache-tomcat-7.0.41/logs/`
- Working Directory: `/data/hive-ptest/working/` (where libs, source, etc. are stored)
- Scripts: `/usr/local/hiveptest/bin/`
- Relevant JIRAs: [HIVE-4675](#) and [HIVE-4739](#)

FAQ:

How do I restart Jenkins?

```
sudo service jenkins restart
```

How do I setup a new branch build?

1. Log into build machine, go to `/usr/local/hiveptest/etc/public`, create a new profile properties in the form `{newprofile.properties}`, where "newprofile" is the name of the new profile. Easiest way is to copy an existing one, for example "trunk-mr2.properties" which are used to run pre-commit tests.
2. In the new profile property, set the "repository" property in the new profile to point to the new branch, and "repository-name" property to any unique value (will be used as the directory name of the checked-out code). Set any other properties accordingly, as they will be invoked by the jenkins build.
3. In Jenkins, create a new job by copying template from existing job.
4. In new Jenkins job, go in the build config modify the profile argument in the Hive PTest2 WebServer call (it will be like: "java ... org.apache.hive.ptest.api.client.PTestClient --profile trunk-mr2") to point to the new profile ie, ("java ...--profile newprofile).

How can I read or make changes to Hive PTest2 Infra code?

Code is located in the Hive git repo. It is under `/hive/testutils/ptest2`. As the Hive PTest2 WebServer is a running process, if changes are made, it needs to be restarted.

How do I stop, start, restart the Hive PTest2 WebServer, or have it use the latest test infra code?

```
$ sudo /usr/local/hiveptest/bin/stop-server.sh
$ sudo /usr/local/hiveptest/bin/start-server.sh
$ sudo /usr/local/hiveptest/bin/restart-server.sh
$ sudo /usr/local/hiveptest/bin/update.sh
```