

Running Benchmarks

Whirr provides a way to run benchmarks on a Hadoop cluster. (This was introduced in [WHIRR-92](#).) This allows you to see how a particular cluster performs on a particular provider, and can provide a basis for tuning.

The following command will start a Hadoop cluster then run a suite of benchmarks (currently Terasort as DFSIO) against it, before tearing it down:

```
mvn verify -Pintegration \  
  -DargLine="-Dwhirr.test.provider=<cloud-provider> -Dwhirr.test.identity=<cloud-provider-user> -Dwhirr.test.credential=<cloud-provider-secret-key>" \  
  -Dit.test=HadoopBenchmarkSuite
```

If you want to provide extra properties, you can do so as follows by specifying the `config` option. Here we use a file called `.whirr-test.properties` in our home directory:

```
mvn verify -Pintegration \  
  -DargLine="-Dwhirr.test.provider=<cloud-provider> -Dwhirr.test.identity=<cloud-provider-user> -Dwhirr.test.credential=<cloud-provider-secret-key> -Dconfig=.whirr-test.properties" \  
  -Dit.test=HadoopBenchmarkSuite
```

You can run a single benchmark as follows (note that this launches and tears down a cluster for this one test):

```
mvn verify -Pintegration \  
  -DargLine="-Dwhirr.test.provider=<cloud-provider> -Dwhirr.test.identity=<cloud-provider-user> -Dwhirr.test.credential=<cloud-provider-secret-key>" \  
  -Dit.test=HadoopServiceTestDFSIOBenchmark
```

Some tests take properties to control their behavior. E.g. you can specify the amount of data to sort in Terasort by specifying `terasortBytesPerNode`:

```
mvn verify -Pintegration \  
  -DargLine="-Dwhirr.test.provider=<cloud-provider> -Dwhirr.test.identity=<cloud-provider-user> -Dwhirr.test.credential=<cloud-provider-secret-key> -DterasortBytesPerNode=100000000" \  
  -Dit.test=HadoopBenchmarkSuite
```