

DevelopmentHints

This page documents various framework development tricks.

Server Log Scanning

When trying to find how many rpc calls were dropped per a second, I use:

```
grep ' discard' < logfile | sed -e 's/[^ ]* //' -e 's/,.*//' | sort | uniq -c
```

the output looks like:

```
1 00:00:34
1 00:03:05
798 00:04:34
600 00:04:35
85 00:04:36
132 00:04:37
1347 00:04:38
```

When RPCs timeout on the server side, the server dumps the thread stacks, but the stacks have newlines in them so they don't work well with the normal unix tools. So I wrote a little sed script that scans a log file and produces one line for each thread:

```
sed -n -e ':start' -e '/^Thread /{h;b inner;}' -e 'd' \
    -e ':inner' -e n -e '/^ /{H;b inner;}' -e '{x;s/\n/ /g;p;x;b start}' < logfile
```

Once you have each thread on a separate line you can do things like look at what is blocking the various threads:

```
grep 'BLOCKED' < stacks | sed -e 's/. *Blocked on //' -e 's/ .*//' | sort | uniq -c
```

and get lists like:

```
146 java.lang.Class@10bd9dd
59 java.lang.Class@8aca13
4 org.apache.hadoop.mapred.JobInProgress@135af93
1 org.apache.hadoop.mapred.JobInProgress@1448b3
3 org.apache.hadoop.mapred.JobInProgress@1f54f6d
3 org.apache.hadoop.mapred.JobInProgress@4d39a3
1749 org.apache.hadoop.mapred.JobTracker@1da74f0
714 org.apache.hadoop.mapred.JobTracker@e0e9f5
```