

bin/nutch domainstats

domainstats is an alias for org.apache.nutch.util.domain.[DomainStatistics](#)

In short its a tool which provides information about which domains have been fetched.

usage

```
$ bin/nutch DomainStatistics inputDirs outDir host|domain|suffix [numOfReducer]
```

example

```
$ bin/nutch DomainStatistics hdfs://nn:9000/user/otis/crawl/crawldb/current hdfs://nn:9000/user/otis/ds-host host
8
```

You can then -cat ds-host file from DFS and pipe it to sort -nrk1 for sorting by count, higher count first.

[CommandLineOptions](#)