# Usage Statistics

## Usage Statistics in Jenkins

Jenkins collects anonymous usage statistics from installations, and this data is processed and used in our infrastructure. The following is the big picture of how this happens:

- Jenkins core has a code that submits their usage statistics daily to our central server.
- Submission arrives to `updates.jenkins.io` (or its past name `updates.jenkins-ci.org`) in an encrypted form and gets stored on this server at `/srv/usage/apache-logs`. We'll call this the encrypted log.
- Kohsuke has automation set up that downloads the encrypted log daily, unencrypt it, anonymize it (code), and send back up to `updates.jenkins.io` at `/srv/usage/usage-stats`. We'll call this the anonymized data.
- The anonymized data gets aggregated into the monthly data (code) by a Jenkins job and placed at `/srv/census/census`.
- The same script and Jenkins job above generates `http://stats.jenkins.io`.

## Access to data

Anonymized data and monthly data are available to those who are interested in analyzing it and share the result with the community. Due to the size, these data are only made available to people on the need basis. Please get in touch with the infra team to arrange access.

## Format of the anonymized data

Anonymized data is a daily text file where each line is a JSON entry of the following form:

```json
{
  "nodes" : [ 
    {
      "os" : "Linux (amd64)",
      "jvm-name" : "OpenJDK 64-Bit Server VM",
      "jvm-vendor" : "Oracle Corporation",
      "jvm-version" : "1.8.0_45",
      "executors" : 2,
      "master" : true
    },
    {
      "timestamp" : "14/Jul/2015:20:00:18 -0400",
      "install" : "36fec69e48a31db4a6d08ab7bb79bfe9c1cc1f436fec69e48a31db4a6d08ab",
      "servletContainer" : "jetty/winstone-2.8",
      "version" : "1.617",
      "jobs" : {
        "hudson-model-ExternalJob" : 0,
        "hudson-model-FreeStyleProject" : 0,
        "hudson-maven-MavenModuleSet" : 1,
        "hudson-matrix-MatrixProject" : 0
      }
    }
  ],
  "plugins" : [ 
    {
      "name" : "maven-plugin",
      "version" : "2.7.1"
    },
    {
      "version" : "2.11",
      "name" : "cvs"
    },
    {
      "name" : "ant",
      "version" : "1.2"
    }
  ],
  "stat" : 1
}
```

"nodes" refer to the master and all the agents that are connected, "install" refers to the hashed installation ID to track different submissions over time. The rest of the fields should be self explanatory. See the code for further details.