Usage Statistics Privacy Advisory 2016-03-30

A bug introduced in Jenkins 1.645 and 1.642.2 caused Jenkins to send anonymous usage statistics data to the Jenkins project even if the administrator opted out of it.

- Which versions are affected?
- What data is being collected?
- How is the data transferred and processed? Who can access the data?
- What are you doing with the usage data? Why are you even collecting it?
- What about private data, like names of plugins we wrote ourselves?
- What are you doing to fix this?
- What about the data you have already collected?
- What can I do about this? Is there a workaround to really disable submission of usage statistics?
- This was fixed in 1.653, why are you only notifying us now?

Which versions are affected?
The following versions of Jenkins do not properly honor the setting to opt out of usage statistics collection, resulting in data being submitted even when disabled:

- Main line: 1.645 (inclusive) to 1.652 (inclusive)
- LTS: 1.642.2 and 1.642.3

What data is being collected?

- The public instance ID
- Name and version of the servlet container
- The version of Jenkins
- For all connected nodes, including master,
  - The values of the system properties `java.vm.vendor`, `java.vm.name`, `java.version`
  - The number of executors
  - The operating system as shown on the "Manage Nodes" page in Jenkins, e.g. Linux (amd64)
- For all installed and enabled plugins,
  - the short name/artifactId (e.g. `git`, or `email-ext`)
  - the version
- For all item/job types,
  - the class name (e.g. `hudson.model.FreestyleProject`)
  - the number of items/job of that type in your Jenkins

For an example of this data, see [this wiki page](https://jenkins-ci.org/census/).

We do not collect information like job names, user names, host names, IP addresses, build logs, build names or descriptions, etc.

How is the data transferred and processed? Who can access the data?
The data is encrypted, base64-encoded, and transferred via an AJAX request piggybacking on a user's browser.

The data is stored encrypted on the Jenkins project infrastructure, from where it is downloaded, decrypted, and filtered locally by one of the board members, and uploaded again. Only Jenkins board members have access to the decryption key to read the full, raw data.

Partially decrypted (see below) data was made available to the public at [https://jenkins-ci.org/census/](https://jenkins-ci.org/census/)

What are you doing with the usage data? Why are you even collecting it?

We're currently providing the statistics at [http://stats.jenkins-ci.org/jenkins-stats/](http://stats.jenkins-ci.org/jenkins-stats/) to everyone. They show various statistics, e.g. number of known Jenkins instances, which plugins are most popular, etc.

We're also providing mostly plugin-related statistics at [http://stats.jenkins-ci.org/plugin-installation-trend/](http://stats.jenkins-ci.org/plugin-installation-trend/) that are used on the Jenkins wiki to indicate plugin popularity, as well as help plugin developers make choices about which versions of Jenkins to target as minimum compatible release.

Data such as operating system, servlet container and Java version help us make decisions on what platforms need to be supported by Jenkins.

In the future, we'd like to use the statistics to help tell us which versions of Jenkins/plugins are unpopular and downgraded from, to help us spot problems.

What about private data, like names of plugins we wrote ourselves?

Operating system names, servlet container names, and Java VM vendor, name, and version are always fully decrypted.

Item/job types and plugin names are only decrypted based on what's essentially a whitelisted:

- In the case of job types, all types starting with `hudson`, `org.jvnet.hudson`, and `org.jenkinsci` are currently considered public. If you follow the convention to use package names corresponding to your organization (e.g. `com.yourcorp.whatever`), these types will not be decrypted.
- In the case of plugins, only plugin names for plugins available on the Jenkins community update site are decrypted.
While the processed data still contains entries for all plugins and item types, those determined to be private are still encrypted, and not part of output such as the stats site.

Again: We do not collect information like job names, user names, host names, IP addresses, build logs, build names or descriptions, etc.

What are you doing to fix this?

As soon as we found out about this bug, we disabled third party access to the partially decrypted data at http://jenkins-ci.org/census

We also changed the decryption/filter tool to discard all data submitted by the affected Jenkins versions, to make sure we're not using data from users not upgrading Jenkins and not applying the workaround described below. We can't do anything about the encrypted data we will continue to receive, as we don't know what version it's about until we decrypt it. So we discard it in the same processing step as decryption.

We'll also release an unscheduled LTS release, 1.642.4, later this week, so users on the LTS line can upgrade to that version to get the bug fix. This LTS release will not impact the regular LTS schedule, i.e. we still plan to release 1.651.1 in two weeks.

What about the data you have already collected?

We regenerated all statistics data for January and February, applying the version filter mentioned above, to make sure we remove all traces of the affected versions' usage data from the data we have.

What can I do about this? Is there a workaround to really disable submission of usage statistics?

If you are running one of the affected versions, the best medium-term solution is to upgrade. The bug does not affect Jenkins 1.653 or newer, or Jenkins LTS 1.642.4 or newer.

In the short term, you can immediately disable submission of usage statistics to the Jenkins project by running the following script in Manage Jenkins » Script Console:

```java
hudson.model.UsageStatistics.DISABLED = true
```

This will immediately disable usage data submission until you restart Jenkins. To make this permanent, change your Jenkins startup script so it passes a system property to the java process:

```java
java -Dhudson.model.UsageStatistics.disabled=true -jar .../jenkins.war
```

For information how to do this when using one of the installers/packages, see the installer/package documentation here.

To verify that usage stats submission is disabled, run the following script in Manage Jenkins » Script Console and confirm the result is true:

```java
println hudson.model.UsageStatistics.DISABLED
```

This was fixed in 1.653, why are you only notifying us now?

We noticed an unusual drop in usage data for the month of February. We were initially unable to determine the cause, but reverted a robustness-improving commit that changed code related to usage data submission anyway. This change went into 1.653. While again looking for the cause of the drop of reported usage data earlier this week, we determined that it was really only affecting versions that included this commit, so we took a closer look at it -- and found that it actually inverted the meaning of the usage data option: Only instances opted out of it did now submit data.